SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Mathematische Zeitschrift
Verlag: Springer
Jahr: 1972
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN266833020_0129
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN266833020_0129

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

Herausgegeben von **H. Wielandt**, Tübingen

unter Mitwirkung von **T. tom Dieck**, Saarbrücken
E. Heinz, Göttingen
A. Pfluger, Zürich
K. Zeller, Tübingen

Wissenschaftlicher Beirat
B. Eckmann, Zürich
P. Gabriel, Bonn
K. H. Hofmann, New Orleans
W. Klingenberg, Bonn
H. Kneser, Tübingen
U. Krengel, Göttingen
W. Magnus, New York
G. Pickert, Gießen
H. Salzmann, Tübingen

Band 129 · 1972

Springer-Verlag · Berlin · Heidelberg · New York

The exclusive copyright for all languages and countries, including the right for photomechanical and any other reproductions, even in microform, is transferred to the publisher

Alle Rechte, einschließlich das der Übersetzung in fremde Sprachen und das der fotomechanischen Wiedergabe oder einer sonstigen Vervielfältigung, auch in Mikroform, vorbehalten Jedoch wird gewerblichen Unternehmen für den innerbetrieblichen Gebrauch nach Maßgabe des zwischen dem Börsenverein des Deutschen Buchhandels e V und dem Bundesverband der Deutschen Industrie abgeschlossenen Rahmenabkommens die Anfertigung einer fotomechanischen Vervielfältigung gestattet Wenn für diese Zeitschrift kein Pauschalabkommen mit dem Verlag vereinbart worden ist, ist eine Wertmarke im Betrage von DM 0,40 pro Seite zu verwenden *Der Verlag läßt diese Beträge den Autorenverbänden zufließen*

Springer-Verlag Berlin · Heidelberg New York
Printer: Universitätsdruckerei H Stürtz AG, Wurzburg
Printed in Germany © by Springer-Verlag Berlin Heidelberg 1972

NIEDERSACHS. STAATS- U. UNIV.-BIBLIOTHEK GÖTTINGEN

Z 1973. 4186

Inhalt/Contents

Indexed in Current Contents

A 21267 D

Mathematische Zeitschrift

Herausgegeben von **H. Wielandt**, Tübingen

unter Mitwirkung von **T. tom Dieck**, Saarbrücken
E. Heinz, Göttingen
A. Pfluger, Zürich
K. Zeller, Tübingen

Wissenschaftlicher Beirat
B. Eckmann, Zürich
P. Gabriel, Bonn
K. H. Hofmann, New Orleans
W. Klingenberg, Bonn
H. Kneser, Tübingen
U. Krengel, Göttingen
W. Magnus, New York
G. Pickert, Gießen
H. Salzmann, Tübingen

Band 129 · Heft 1 · 1972

Springer-Verlag · Berlin · Heidelberg · New York

Math. Z. — MAZEAX 129(1) 1–94 (1972) 6. XII. 1972

Niedersächsische Staats- u. Universitätsbibliothek Göttingen
16. Dez. 1972

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. Kampke* und *F. K. Schmidt*, 1936 durch *R. Nevanlinna* und 1950 durch *H. Wielandt*, der 1952 die Herausgabe übernahm.

„**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.

*

Die Mathematische Zeitschrift dient der Pflege der reinen Mathematik, doch werden auch Beiträge aus den Gebieten der theoretischen Physik und Astronomie Aufnahme finden, soweit sie mathematisch von Interesse sind. Besprechungen, Aufgaben u. dgl. werden nicht zugelassen.

Grundsätzlich dürfen nur Arbeiten eingereicht werden, die vorher weder im Inland noch im Ausland veröffentlicht worden sind. Der Autor verpflichtet sich, seinen Beitrag auch nachträglich nicht an anderer Stelle zu publizieren. Mit der Annahme des Manuskriptes und seiner Veröffentlichung durch den Verlag geht auch das Recht der fotomechanischen Wiedergabe oder einer sonstigen Verfielfältigung, auch in Mikroform, an den Verlag über. Jedoch wird gewerblichen Unternehmen für den innerbetrieblichen Gebrauch nach Maßgabe des zwischen dem Börsenverein des Deutschen Buchhandels e. V. und dem Bundesverband der Deutschen Industrie abgeschlossenen Rahmenabkommens die Anfertigung einer fotomechanischen Vervielfältigung gestattet. Wenn für diese Zeitschrift kein Pauschalabkommen mit dem Verlag vereinbart ist, ist eine Wertmarke im Betrage von DM 0.40 pro Seite zu verwenden. *Der Verlag läßt diese Beträge den Autorenverbänden zufließen.*

Von jeder Arbeit werden 75 Sonderdrucke kostenlos zur Verfügung gestellt, weitere können zum Selbstkostenpreis bezogen werden.

"Mathematische Zeitschrift" is devoted primarily to pure mathematics; papers on theoretical physics and astronomy may be accepted if they present interesting mathematical results. Reviews, problems etc. will not be published.

It is a fundamental condition that manuscripts submitted should not have been published elsewhere, in this or any other country. The author must undertake not to publish elsewhere at a later date. With the acceptance of a manuscript for publication, the publishers acquire the sole copyright for all languages and countries. Unless special permission has been granted by the publishers, no photographic reproductions microform or any other reproductions of a similar nature may be made of the journal, of individual contributions contained therein or of extracts therefrom.

75 reprints of each paper are provided free of charge; additional copies may be ordered at cost price.

Manuskripte nehmen entgegen / Manuscripts may be sent to:

H. Wielandt, 74 Tübingen, Math. Inst. d. Universität

T. tom Dieck, 66 Saarbrücken, Math. Inst. d. Universität

E. Heinz, 34 Göttingen, Math. Inst. d. Universität

A. Pfluger, Zürich (Schweiz), Eidgen. Technische Hochschule

K. Zeller, 74 Tübingen, Math. Inst. d. Universität

*

Die Zeitschrift erscheint, um eine rasche Publikation zu ermöglichen, in einzelnen Heften, die zu Bänden vereinigt werden. Ein Band besteht im allgemeinen aus 4 Heften. Der Preis eines Bandes beträgt DM 112.—.

In the interest of speedy publications, this journal is issued at frequent intervals according to the material received. As a rule four numbers constitute one volume. The price is DM 112.— per volume.

Springer-Verlag

D-6900 Heidelberg 1
Postfach 1780
Tel. (06221) 49101
Telex 04-61723

D-1000 Berlin 33
Heidelberger Platz 3
Tel. (0311) 822001
Telex 01-83319

Springer-Verlag
New York Inc.
175 Fifth Avenue
New York, N.Y. 10010

Math. Z. 129, 1 – 19 (1972)
© by Springer-Verlag 1972

Banach Lattices of Compact Maps

John Chaney

Introduction

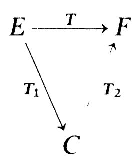
This is a study of some of the properties of a subspace of the space of compact linear maps where the domain or range is a Banach lattice. It is known that the space of compact maps from a Banach lattice into a $C(X)$ is a lattice, and the nuclear maps from a Banach lattice into the space L^1 for a finite measure space is a lattice. These two examples of subspaces of the space of compact linear maps which are lattices will be shown to be special cases of a general subspace of compact maps which is always a Banach lattice when the domain and range are Banach lattices. This space of compact maps, denoted $M_*(E', F)$, is shown to represent a completed normed tensor product $E \tilde{\otimes}_M F$ defined when E is a Banach space and F is a Banach lattice.

Throughout this paper, E denotes a Banach space and F a Banach lattice, unless otherwise stated. Define the $|\mu|$-norm on $E \otimes F$ by

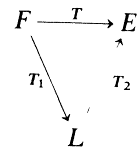
$$\|u\|_{|\mu|} = \inf \left\{ \left\| \sum_{i=1}^{n} \|x_i\| \, y_i \right\| : \; u = \sum_{i=1}^{n} x_i \otimes y_i \text{ in } E \otimes F, \; y_i \geqq 0 \right\}.$$

In the first section, we introduce the M-tensor product norm on $E \otimes F$ and show that this norm is actually an alternate description of the $|\mu|$-norm. The completion $E \tilde{\otimes}_M F$ of the M-tensor product $E \otimes_M F$ is characterized as a space $M_*(E', F)$ of compact maps from E' into F in (1.3). In (1.7) it is shown that if both E and F are Banach lattices then $M_*(E', F)$ is also a Banach lattice of compact maps.

In the second section, we show that $M_*(E'', F)$ is canonically isomorphic to the space $M(E, F)$ of all maps $T: E \to F$ that can be factored as:



where T_1 is a compact map into an (AM)-space C and T_2 is a positive map from C into F. We also show that $M_*(E', F')$ is canonically isomorphic to the space $\Lambda(F, E)$ of all maps $T: F \to E$ that can be factored as:



where T_1 is a positive map from F into an (AL)-space L and T_2 is a compact map from L into E. It is then shown that the dual of $E \otimes_M F$ is equal to $E' \tilde{\otimes}_M F'$ for all Banach spaces E if and only if F' is a generalized sequence space.

The general representation theorem (1.3) is applied in the third section to obtain a characterization of the space $L^p(E)$ for $1 \leqq p < \infty$ as a space of compact linear maps from L^q into E. An extension of the Dunford-Pettis theorem to the L^p spaces for $p > 1$ and a proof of the equivalence of the Radon-Nikodym property for a Banach space E with the property that every integral map from L^∞ into E is nuclear is also given in this section.

For general terminology and notation concerning functional analysis we will use [8]; our references for ordered spaces will be taken from [6]. The background material concerning the π and ε-tensor product norms can be found in [3] and [7].

We shall now introduce some special results and terminology that will be used throughout this paper. If z is a positive element in Banach lattice F then define the space F_z to be the linear hull of $[-z, z]$ equipped with the Minkowski functional of $[-z, z]$. By Kakutani's theorem (see (8.5) of Chapter V in [8]), F_z is a $C(X)$ space. If z' is a positive element of F', then define $F_{[-z', z']^0}$ to be the completion of the quotient space $F/p^{-1}(0)$, where p is the Minkowski functional of $[-z', z']^0$, equipped with the norm induced by p.

Let $S_+(F, E)$ denote the space of continuous linear maps from F into E which map positive summable sequences into absolutely summable sequences and equip this space with the norm defined for T in $S_+(F, E)$ by

$$\|T\|_L = \sup\left\{\sum_{i=1}^{k} \|Ty_i\| : \; y_i \geqq 0 \text{ in } F \text{ and } \left\|\sum_{i=1}^{k} y_i\right\| \leqq 1\right\}.$$

In [4], Jacobs defined a norm symmetric to the $|\mu|$-norm and showed the dual of $E \otimes_{|\mu|} F$ is norm isomorphic to $S_+(F, E')$. An element of $S_+(F, E')$ is termed an *order summable linear map*. The triple bar norm defined by Dinculeanu in [1] is the L-norm defined on order summable operators, and the operators studied by Uhl in [11] are order summable operators. These maps were also studied by Schlotterbeck in [9].

The author would like to thank Professors A.L. Peressini and H.P. Lotz for their interest and suggestions.

1. A Representation for the M-Tensor Product

A crossnorm, the M-norm, is defined on the tensor product $E \otimes F$ of a Banach space E with a Banach lattice F. The completion of $E \otimes F$ equipped with this norm is shown to be represented by $M_*(E', F)$, a space of compact linear maps from E' into F. We will show the M-norm is equal to the $|\mu|$-norm. The M-norm has the advantage its value is not dependent on all the representations of an element as a member of the tensor product. When both E and F are Banach lattices, we will show $M_*(E', F)$ is a Banach lattice for the usual notion of a lattice of order bounded linear operators. See p. 21–23 in [6] for the formulas for the absolute value of an order bounded linear operator.

Definition. Let $M_*(E', F)$ be the space of linear maps T from E' into F which have the decomposition $T = T_2 \circ T_1$, where T_1 is weak*–weak ($w^* - w$) continuous and compact and maps E' into some F_z and T_2 is the canonical map from F_z into F.

Remarks. (1) In the definition above, since the spaces F_z are (AM)-spaces with units they can be replaced by arbitrary $C(X)$ spaces, by Kakutani's theorem, when T_2 is restricted to be a positive map from $C(X)$ into F.

(2) $M_*(E', F)$ can be identified under transposition with the space of maps T from F' into E which have the decomposition $T = T_2 \circ T_1$ where T_1 is the adjoint of some canonical map from F_z into F and T_2 is a $w^* - w$ continuous compact map from $(F_z)'$ into E.

(3) In the definition of $M_*(E', F)$ the requirement T_1 be $w^* - w$ continuous can be replaced by the condition that T be $w^* - w$ continuous. It suffices to show that if T is $w^* - w$ continuous then the compact map T_1 is $w^* - w$ continuous. Since T_2 is one-to-one then T_2' has $\sigma((F_z)', F_z)$-dense range in $(F_z)'$. Since T_1 is compact then T_1' is $w^* - w$ continuous. Let Cl denote closure. Because $\mathrm{Cl}(T_1'(T_2'(F')))$ contains $T_1'((F_z)')$, $T'(F')$ is contained in E, and $T' = T_1' \circ T_2'$ then $T_1'((F_z)')$ is contained in E. Therefore, T_1 is $w^* - w$ continuous.

If $T = T_2 \circ T_1$ is an element of $M_*(E', F)$ and U is the unit ball in E then $T_1(U^0)$ is relatively compact in F_z. By (3.10) on p. 95 in [6], $\sup T_1(U^0)$ exists in F_z. Since F_z is an ideal in F then $\sup T(U^0)$ exists in F. Equip $M_*(E', F)$ with the norm $\|T\|_M = \|\sup T(U^0)\|_F$. Denote $E \otimes F$ considered as a subspace of $M_*(E', F)$ by $E \otimes_M F$, and its completion by $E \tilde{\otimes}_M F$.

Remarks. (4) It follows from the inequalities

$$\sup_{x' \in U^0} \left\|\sum \langle x_i, x'\rangle y_i\right\| \leq \left\|\sup_{x' \in U^0} \sum \langle x_i, x'\rangle y_i\right\|$$
$$\leq \left\|\sum \|x_i\| \, |y_i|\right\| \leq \sum \|x_i\| \, \|y_i\|$$

for an element $\sum x_i \otimes y_i$ in $E \otimes F$ that the identity maps $E \otimes_\pi F \to E \otimes_M F$, $E \otimes_{|\mu|} F \to E \otimes_M F$, $E \otimes_M F \to E \otimes_\varepsilon F$ are continuous and the M-norm is a crossnorm.

(5) If $y_1, \ldots, y_n$ are disjoint elements of an order complete Banach lattice F and $x_1, \ldots, x_n$ are from E then $\left\| \sum_{i=1}^n x_i \otimes y_i \right\|_M = \left\| \sum_{i=1}^n \|x_i\| \, |y_i| \right\|$. To see this, let B be the band in F generated by $\{y_1, \ldots, y_n\}$ and let B_j be the principal band generated by $\{y_j\}$. By propositions (4.6) and (4.7) on p. 39 in [6], B is the order direct sum of the B_j's. Let p_j be the positive projection of B onto B_j. If $s = \sup \sum_{i=1}^n \langle x_i, U^0 \rangle y_i$ then s is in B and $s = \sum_{j=1}^n p_j(s)$. Choose x' in U^0 to satisfy $\|x_k\| = \langle x_k, x' \rangle$ for the fixed index k. Since $s \geqq \sum_{j=1}^n \langle x_j, \pm x \rangle y_j$ then $p_k(s) \geqq \pm p_k \left(\sum_{j=1}^n \langle x_j, x \rangle y_j \right) = \pm \|x_k\| \, y_k$. Therefore, $p_k(s) \geqq \|x_k\| \, |y_k|$ for each index k, and $s \geqq \sum_{k=1}^n \|x_k\| \, |y_k|$. Since in general $s \leqq \sum_{j=1}^n \|x_j\| \, |y_j|$ then $\left\| \sum_{i=1}^n x_i \otimes y_i \right\|_M = \|s\| = \left\| \sum_{j=1}^n \|x_j\| \, |y_j| \right\|$.

(1.1) **Lemma.** *Let A be a subset of E such that $A = -A$ and let T be a continuous linear map from E into F such that* $\sup T(A)$ *exists in F. If $T(A)$ is a bounded subset of F_z then* $\sup \|T(A)\|_{F_z} = \|\sup T(A)\|_{F_z}$.

Proof. Let $s = \sup \|T(A)\|_{F_z}$. For every $\delta > 0$, $T(A) \subset (s+\delta)[-z, z]$ and $-(s+\delta)z \leqq \sup T(A) \leqq (s+\delta)z$. Therefore $\|\sup T(A)\|_{F_z} \leqq (s+\delta)$ for each $\delta > 0$ and hence $\|\sup T(A)\|_{F_z} \leqq s$. Conversely, since $A = -A$ then $T(A)$ is contained in $[-\sup T(A), \sup T(A)]$ and $s \leqq \|\sup T(A)\|_{F_z}$.

(1.2) **Corollary.** *If X is a compact Hausdorff topological space then the norms on $E \otimes_M C(X)$ and $E \otimes_\varepsilon C(X)$ coincide.*

Proof. The uniform operator norm on $\mathscr{L}(E', C(X))$ induces the norm $E \otimes_\varepsilon C(X)$. Hence, if we let A be the unit ball in E', the result follows directly from (1.1).

Remarks. (6) Since each F_y has the approximation property then (see the Corollary of Prop. 36 in Chapt. 1 in [3]) every $w^* - w$ continuous compact map from E' into F_y is given by an element of $E \tilde{\otimes}_\varepsilon (F_y)$. Therefore, $M_*(E', F) = \cup \{E \tilde{\otimes}_\varepsilon (F_y) \colon y \geqq 0\}$.

(7) Let u be an element of $E \otimes F_y$, then $\|u\|_{E \otimes_M F} \leqq \|y\|_F \, \|u\|_{E \otimes_M F_y}$. Hence, in view of (1.2) the canonical map from $E \otimes_\varepsilon F_y$ into $E \otimes_M F$ is continuous and $M_*(E', F) = \bigcup_{y \geqq 0} E \tilde{\otimes}_\varepsilon F_y$ then $E \otimes F$ is dense in $M_*(E', F)$.

(1.3) **Theorem.** $E \tilde{\otimes}_M F = M^*(E', F)$.

Proof. In view of remark (7) it suffices to show $M_*(E', F)$ is complete. Let (T_n) be a Cauchy sequence in $M_*(E', F)$ and suppose $\|T_{n+1} - T_n\|_M \leqq 2^{-2n}$. Let $z_0 = \sup T_1(U^0)$ and $z_n = \sup(T_{n+1} - T_n)(U^0)$; note $\|z_n\|_F \leqq 2^{-2n}$ for $n \geqq 1$. Define $w = \sum_{n=0}^{\infty} 2^n z_n$, then $z_n \leqq w 2^{-n}$. Let $z = \sum_{n=0}^{\infty} z_n$, then $z - \sum_{n=0}^{N} z_n \leqq 2^{-N} w$. Thus, z_n is summable to z in F_w.

Choose positive elements y_n in F so that $\|y_n\| \leqq 2^{-n}$ and T_1 and $T_{n+1} - T_n$ have the decompositions $T_1 : E' \to F_{y_0} \to F$ and

$$T_{n+1} - T_n:\ E' \xrightarrow{R_n} F_{y_n} \to F,$$

where R_n is $w^* - w$ continuous and compact. Let $y = \left(\sum_{n=0}^{\infty} y_n\right) \vee w$. There exist decompositions $T_{n+1} - T_n$: $E' \xrightarrow{S_n} F_y \xrightarrow{i} F$, where S_n is $w^* - w$ continuous and compact and z_n is summable to z in F_y. We will now see that (S_n) is a summable sequence in $\mathscr{L}(E', F_y)$.

$$\begin{aligned}
\left\|\sum_{n=M}^{N} S_n\right\| &= \sup_{x' \in U^0} \left\|\sum_{n=M}^{N} S_n x'\right\|_{F_y} \\
&\leqq \left\|\sup_{x' \in U^0} \sum_{n=M}^{N} S_n x'\right\|_{F_y} \\
&\leqq \left\|\sum_{n=M}^{N} (\sup S_n(U^0))\right\|_{F_y} \\
&= \left\|\sum_{n=M}^{N} z_n\right\|_{F_y}.
\end{aligned}$$

Let $S = \sum_{n=0}^{\infty} S_n$ in $\mathscr{L}(E', F_y)$. Since $E \tilde{\otimes}_\varepsilon (F_y)$ represents the space of $w^* - w$ continuous compact maps in $\mathscr{L}(E', F_y)$ and $E \tilde{\otimes}_\varepsilon (F_y)$ is a closed subspace of $\mathscr{L}(E', F_y)$ then S is in $E \tilde{\otimes}_\varepsilon (F_y)$ and $T = i \circ S$ is in $M_*(E', F)$ where i is the canonical map from F_y into F. Finally, T_N converges to T in $M_*(E', F)$ since $T - T_N = i \circ \sum_{n=N+1}^{\infty} S_n$ and

$$\begin{aligned}
\|T - T_N\|_M &= \|\sup(T - T_N)(U^0)\|_F \\
&\leqq \|y\|_F \left\|\sup\left(\sum_{n=N+1}^{\infty} S_n\right)(U^0)\right\|_{F_y} \\
&\leqq \|y\|_F \left\|\sum_{n=N+1}^{\infty} z_n\right\|_{F_y}.
\end{aligned}$$

Remark. (8) If T is in $M_*(E', F)$ then

$$\|T\|_M = \inf\{\|\sup T(U^0)\|_{F_z}: z \geqq 0, \|z\| = 1, T(U^0) \text{ is bounded in } F_z\}.$$

For if, $w = \sup T(U^0)$ and $u = w/\|w\|$ then $\|T\|_M = \|w\|_F = \|w\|_{F_u}$ and $T(U^0)$ is bounded in F_u. Also, if $T(U^0)$ is bounded in F_z, $z \geqq 0$, and $\|z\| = 1$ then $\|w\|_{F_z} \geqq \|w\|_F$ because the unit ball of F contains $[-z, z]$.

(1.4) **Theorem.** *The M-norm and the $|\mu|$-norm are equal.*

Proof. We first consider the case F is order complete. For each positive y in F let D_y be the collection of elements u in $E \otimes (F_y)$ such that $u = \sum_{i=1}^{n} x_i \otimes y_i$ where $y_1, \ldots, y_n$ is a set of disjoint elements of F_y, and let $D = \bigcup\{D_y: y \geqq 0\}$ where each D_y is regarded as a subset of $E \otimes F$. By remarks (5) and (4) the two norms coincide on D. We will show D is dense in $E \otimes_{|\mu|} F$. Since F is order complete then the spaces F_y are also order complete. By Kakutani's theorem (see (8.5) in Chapt. V in [8]) and Nakano's result (see p. 16 in [6]), F_y is norm and lattice isomorphic to a $C(X)$ where X is an extremally disconnected compact Hausdorff topological space. Let S be the space of simple functions defined on the algebra of clopen subsets of X. It is known that S is dense in $C(X)$; therefore, $E \otimes S$ is dense in $E \otimes_\varepsilon C(X)$. If we consider S as a subset of F_y then $E \otimes S$ is contained in D_y. Therefore, D_y is dense in $E \otimes_\varepsilon F_y$. Since $S_+(F_y, E') = (E \otimes_\varepsilon F_y)'$ (see (4.1) in [9]) and $(E \otimes_{|\mu|} F_y)' = S_+(F_y, E')$ (see (1.6) in [4]) then the identity map $E \otimes_{|\mu|} F_y \to E \otimes_\varepsilon F_y$ has a continuous inverse. Since D_y is dense in $E \otimes_\varepsilon F_y$ then it is dense in $E \otimes_{|\mu|} F_y$. Since the inclusion map from $E \otimes_{|\mu|} F_y$ into $E \otimes_{|\mu|} F$ is continuous then D is dense $E \otimes_{|\mu|} F$. Since the two norms coincide on D (see remark (5)), D is dense in $E \otimes_{|\mu|} F$, and the identity map $E \otimes_{|\mu|} F \to E \otimes_M F$ is continuous (see remark (4)) then the two norms are equal on $E \otimes F$.

Now consider the case F is an arbitrary Banach lattice. We will obtain our result by using the first case and the fact that F'' is order complete.

The inclusion map from $M_*(E', F)$ into $M_*(E', F'')$ defined by $T \to i \circ T$ where i is the canonical map from F into F'' is a norm isomorphism. In fact, suppose that $T = T_2 \circ T_1$ is in $M_*(E', F)$ where T_1 is a $w^* - w$ continuous compact map from E' into F_z and T_2 is the canonical map from F_z into F. Let $w = \sup T_1(U^0)$. Since w is the limit in F_z (see the proof of (3.10), p. 95, in [6]) of a sequence (w_n) of suprema of finite subsets of $T_1(U^0)$ and the map $i \circ T_2$ preserves the supremum of a finite set and is continuous then w is the supremum of $T(U^0)$ in F and F''. Hence, $i(\sup T(U^0)) = \sup (i \circ T)(U^0)$, and so $\|T\|_M = \|i \circ T\|_M$. Consequently the canonical map J from $E \otimes_M F$ into $E \otimes_M F''$ is a norm isomorphism.

Next we show the map from $S_+(F'', E')$ into $S_+(F, E')$ defined by $T \to T \circ i$ is a metric homomorphism. If R is in $S_+(F, E')$ let $S = j \circ R''$ where j is the adjoint of the canonical map from E into E''. Note $S \circ i = R$. By Corollary 5, p. 23, in [9], $\|S\|_L = \|j \circ R''\|_L \leqq \|j\| \, \|R''\|_L \leqq \|R''\|_L$, and by (1.6) and (1.7) in [9], $\|R''\|_L = \|R\|_L$. Therefore, $\|S\|_L \leqq \|R\|_L$. Since the restriction of S to F equals R and the L-norm is a supremum over positive summable sequences whose sum lies in the unit ball then $\|S\|_L \geqq \|R\|_L$. Therefore, $\|S\|_L = \|R\|_L$, and the map is a metric homomorphism. Since the dual of $E \otimes_{|\mu|} F$ is norm isomorphic to $S_+(F, E')$ equipped with the L-norm (see (1.5) in [4]) then the canonical map I from $E \otimes_{|\mu|} F$ into $E \otimes_{|\mu|} F''$ is a norm isomorphism.

By the first case the identity map from $E \otimes_{|\mu|} F''$ into $E \otimes_M F''$ is a norm isomorphism. Since the diagram in Fig. 1 commutes, and the maps I, J, K are norm isomorphisms then H is a norm isomorphism.

$$\begin{array}{ccc} E \otimes_M F & \xrightarrow{\;J\;} & E \otimes_M F'' \\ \uparrow{\scriptstyle H} & & \uparrow{\scriptstyle K} \\ E \otimes_{|\mu|} F & \xrightarrow{\;I\;} & E \otimes_{|\mu|} F'' \end{array}$$

Fig. 1

(1.5) **Proposition.** *If F is an (AM)-space then $E \otimes_\varepsilon F$ and $E \otimes_M F$ are norm isomorphic.*

Proof. Consider the diagram in Fig. 2, where the maps I, J, H, K

$$\begin{array}{ccc} E \otimes_M F & \xrightarrow{\;J\;} & E \otimes_M F'' \\ \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle K} \\ E \otimes_\varepsilon F & \xrightarrow{\;I\;} & E \otimes_\varepsilon F'' \end{array}$$

Fig. 2

are canonical. By the proof of (1.4) J is a norm isomorphism. I is a norm isomorphism since the ε-tensor product preserves norm isomorphisms. By (1.2) K is a norm isomorphism. Since the diagram commutes then H is a norm isomorphism.

Suppose both E and F are Banach lattices and define an element of $M_*(E', F)$ to be *positive* if it is a positive linear map. We will now show $M_*(E', F)$ is a Banach lattice.

Definition. If G and H are Banach lattices define the *projective cone* in the tensor product $G \otimes H$ to be the set of all elements $\sum_{i=1}^{n} g_i \otimes h_i$ where $g_1, \ldots, g_n$ are positive in G and $h_1, \ldots, h_n$ are positive in H.

(1.6) **Lemma.** *If X is a compact Hausdorff topological space then $E\tilde{\otimes}_\varepsilon C(X)$ is a Banach lattice for the closure of the projective cone. Moreover, the absolute value of an element in $E\tilde{\otimes}_\varepsilon C(X)$ is equal to its absolute value as an order bounded linear map when $E\tilde{\otimes}_\varepsilon C(X)$ is identified with the space of w^*-w continuous compact maps from E' into $C(X)$.*

Proof. From the usual proof that $E\tilde{\otimes}_\varepsilon C(X)$ is norm isomorphic to $C(X;E)$, the space of continuous functions from X into E one can also see that the closure of K_p coincides with the positive continuous functions. Since $C(X;E)$ is a Banach lattice then so is $E\tilde{\otimes}_\varepsilon C(X)$.

We will now see the notion of absolute value in this space coincides with that of an order bounded linear map. Let T be a w^*-w continuous compact map in $\mathscr{L}(E', C(X))$. Define f in $C(X;E)$ by $f(t)=T'(\varepsilon_t)$ where ε_t is the point evaluation at t. The function $|f|$ defined by $|f|(t)=|T'(\varepsilon_t)|$ is also an element of $C(X;E)$, and it defines the w^*-w continuous compact map S in $\mathscr{L}(E', C(X))$ by $Sx'=\langle |T'(\varepsilon_t)|, x'\rangle$. It will be shown that $(\sup T[-x',x'])(t)=(Sx')(t)$ for all positive x' in E' and t in X. Suppose z is in $C(X)$ and $z\geqq Ty'$ for all y' in $[-x',x']$, then $z(t)\geqq (Ty')(t)=\langle T'\varepsilon_t, y'\rangle$ and so

$$z(t)\geqq \sup_{|y'|\leqq x'} \langle T'\varepsilon_t, y'\rangle = \langle |T'\varepsilon_t|, x'\rangle = (Sx')(t).$$

Moreover, for y' in $[-x',x']$,

$$(Sx')(t)=\langle |T'\varepsilon_t|, x'\rangle \geqq \langle T'\varepsilon_t, y'\rangle = (Ty')(t).$$

Therefore, $(Sx')(t)=(\sup T[-x',x'])(t)$.

Definition. A subspace G of $\mathscr{L}(E,F)$ is an *operator lattice* if it is a lattice and the absolute value of T in G is given by $|T|(x)=\sup T[-x,x]$, for all positive x in E.

(1.7) **Theorem.** *If E and F are Banach lattices then $E\tilde{\otimes}_M F$ is a Banach lattice for the closure of the projective cone. $\overline{K_p}$ coincides with the cone of positive elements in $M_*(E',F)$. Moreover, $M_*(E',F)$ is an operator lattice.*

Proof. From remark (6), (1.3), and (1.6), $\overline{K_p}$ coincides with the cone of positive linear maps in $M_*(E',F)$ and $M_*(E',F)=\bigcup\{E\tilde{\otimes}_\varepsilon F_z\colon z\geqq 0\}$. Hence, by (1.6), $M_*(E',F)$ is an operator lattice.

We now check the unit ball in $M_*(E',F)$ is solid. If T is in $M_*(E',F)$ then note that $T(U^0)$ is bounded in F_z if and only if $|T|(U^0)$ is bounded in F_z. Let $B=\{z\colon \|z\|=1,\ z\geqq 0$, and $T(U^0)$ is bounded in $F_z\}$.

$$\begin{aligned}
\| |T| \|_M &= \| \sup |T| (U^0) \|_F \\
&= \inf\{ \| \sup |T| (U^0) \|_{F_z} : z \in B\} && \text{(remark 8)} \\
&= \inf\{\sup\{ \| |T| x' \|_{F_z} : x' \geqq 0 \text{ in } U^0\} : z \in B\} && (1.1) \\
&= \inf\{\sup\{ \| \sup T[-x', x'] \|_{F_z} : x' \geqq 0 \text{ in } U^0\} : z \in B\} && (1.6) \\
&= \inf\{\sup \| T(U^0) \|_{F_z} : z \in B\} && (1.1) \\
&= \| T \|_M . && \text{(remark 8)}
\end{aligned}$$

Since the norm is monotone on the positive cone then $M_*(E', F)$ is a Banach lattice.

(1.8) **Proposition.** *Let E and F be Banach lattices, then $S_+(F, E')$ when considered as the dual of $E \otimes_M F$ is an operator lattice.*

Proof. The cone in $S_+(F, E')$ dual to K_p in $E \otimes F$ is the cone of positive linear operators in $S_+(F, E')$. Since the dual cone equips $S_+(F, E')$ with the natural ordering then from (2.2) in [9] it follows that $S_+(F, E')$ is an operator lattice.

2. The Spaces $M(E, F)$ and $\Lambda(F, E)$

Definition. Let $M(E, F)$ be the space of all linear maps T from E into F which have the decomposition $T = T_2 \circ T_1$ where T_1 is a compact map from E into some (AM)-space C and T_2 is a positive map from C into F. Let $\Lambda(F, E)$ denote the space of linear maps T from F into E which have the decomposition $T = T_2 \circ T_1$ where T_1 is a positive map from F into some (AL)-space L and T_2 is a compact map from L into E.

Let $T = T_2 \circ T_1$ be an element of $M(E, F)$ and let U denote the unit ball in E. Since $T_1(U)$ is a relatively compact subset of the (AM)-space C then $\sup T_1(U)$ exists in C'' by (3.10) p. 95 [6], and $\sup T_1(U)$ is the limit of a sequence of suprema of finite subsets of $T_1(U)$. Since the canonical map from C into C'' preserves the supremum of a finite subset and C is closed in C'' then the supremum of $T_1(U)$ in C'' is an element of C. Let $z = T_2(\sup T_1(U))$. Since T_2 is positive then $T(U)$ is contained in $[-z, z]$. Therefore, T has the decomposition $T = S_2 \circ S_1$ where S_1 is a compact map from E into F_z and S_2 is the natural map from F_z into F. This shows $M(E, F)$ is a vector space, and we can equip it with the norm $\| T \|_M = \| \sup T(U) \|$.

Note that if F is an (AM)-space then $M(E, F) = C(E, F)$, the space of compact maps from E into F.

If T is in $\Lambda(F, E)$ then T' is in $M(E', F')$. Equip $\Lambda(F, E)$ with the norm $\| T \|_L = \| T' \|_M$. Note that if F is an (AL)-space then $\Lambda(F, E) = C(F, E)$.

If E and F are Banach lattices then define an element of either $M(E, F)$ of $\Lambda(F, E)$ to be *positive* if it is a positive linear map.

(2.1) **Theorem.** (1) *$M(E,F)$ is norm isomorphic to $M_*(E'',F)$ and $E'\tilde{\otimes}_M F$. The isomorphism I from $M_*(E'',F)$ onto $M(E,F)$ is given by $I(T)=T\circ S$ where S is the canonical map from E into E''.*

(2) *$\Lambda(F,E)$ is norm isomorphic to $M_*(E',F')$ and $E\tilde{\otimes}_M F'$. The isomorphism J from $M_*(E',F')$ onto $\Lambda(F,E)$ is given by $J(T)=T'\circ P$ where P is the canonical map from F into F''.*

(3) *If E is also a Banach lattice then $M(E,F)$ and $\Lambda(F,E)$ are Banach and operator lattices.*

Proof. (1) Let T be in $M_*(E'',F)$ and let $T=T_2\circ T_1$ be a decomposition of T where T_1 is a w^*-w continuous compact map from E'' into F_z and T_2 is the canonical map of F_z into F. Since $S(U)$ is $\sigma(E'',E')$-dense in U^{00} and T_1 is w^*-w continuous then $T_1\circ S(U)$ is $\sigma(F_z,(F_z)')$-dense in $T_1(U^{00})$. Since $T_1(SU)$ is convex then $T_1(SU)$ is norm dense in $T_1(U^{00})$, and because the cone is closed in F_z then $\sup T_1(U^{00})=\sup T_1(SU)$. Therefore, I is a norm isomorphism. To see that I is onto, let R be an element of $M(E,F)$ and let $R=R_2\circ R_1$ where R_1 is a compact map from E into some (AM)-space C and R_2 is a positive map from C into F. Since $R_1(U)$ is relatively compact in C then by Lemma 1, p. 111, in [8] there exists a compact subset B of C such that $R_1(U)$ is compact in C_B and contained in B. By (3.10), p. 95, in [6], $\sup R_1(U)$ exists in C; let $z=R_2(\sup R_1(U))$. Consider the diagram in Fig. 3. Define

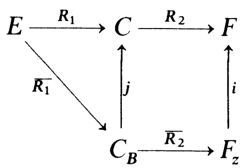


Fig. 3

$\bar{R}_1$ by $R_1=j\circ\bar{R}_1$ where j is the canonical map. $\bar{R}_1$ is compact since $R_1(U)$ is compact in C_B. Define $\bar{R}_2$ by $R_2\circ j=i\circ\bar{R}_2$ where i is the canonical map. By Schauder's theorem, $i\circ\bar{R}_2\circ(\bar{R}_1'')$ is in $M_*(E'',F)$, and its image under I is R.

(2) Let T be in $M_*(E',F')$. Consider the diagrams:

$$F\xrightarrow{P}F''\xrightarrow{T'}E'',\qquad E'\xrightarrow{S}E'''\xrightarrow{T''}F'''\xrightarrow{P'}F'.$$

Since $\|T'\circ P\|_L=\|(T'\circ P)'\|_M$ and the map I from $M_*(E''',F')$ into $M(E',F')$ is a norm isomorphism by part (1) then

$$\|J(T)\|_L=\|P'\circ T''\|_M=\|P'\circ T''\circ S\|_M=\|T\|_M,$$

and J is a norm isomorphism. Let R be an element of $\Lambda(F,E)$ and $R=R_2\circ R_1$ where R_1 is a positive map from F into some (AL)-space L and R_2 is a compact map from L into E. Let e be the unit in L' and let

$x' = R_1'(e)$. Since $R_2''(L')$ is contained in E then R_2' is $w^* - w$ continuous and since R_1' factors through $(F')_{x'}$ then $R' = R_1' \circ R_2'$ is in $M_*(E', F')$.

(3) We know that both $M(E, F)$ and $\Lambda(F, E)$ are Banach lattices because $M_*(E'', F)$ and $M_*(E', F')$ are lattices and the positive cones correspond under the isomorphisms I and J. It is not immediately clear they are operator lattices. Let S and P be the canonical maps from E into E'' and F into F'', respectively. If T is in $M(E, F)$ then its absolute value is $|T'|' \circ P$ (see (1.7)). The problem is to show that if T is in $M(E, F)$ then $|T| = |T''| \circ S$, and if T is in $\Lambda(F, E)$ then $|T| = |T'|' \circ P$.

Let T be an element of $M(E, F)$, then T'' is $M_*(E'', F)$. Note for positive x in E, $S[-x, x]$ is $\sigma(E'', E')$-dense in $[-Sx, Sx]$. Since T'' is $w^* - w$ continuous then $T'' \circ S([-x, x])$ is dense in $T''[-Sx, Sx]$, and

$$|T''|(Sx) = \sup T''[-Sx, Sx] = \sup(T'' \circ S)([-x, x]) = |T|(x).$$

For T in $\Lambda(F, E)$, first consider the case E is order complete. If y is positive in F then $T([-y, y])$ is majorized in E by $|T'|' Py$, and so $|T|$ exists since E is order complete. Since $|T|\, y \leqq |T'|' \circ P y \leqq |T|'' P y = |T|\, y$ then $|T| = |T'|' \circ P$. Now suppose E is arbitrary and T is in $\Lambda(F, E)$. Since $S \circ T$ is in $\Lambda(F, E'')$ and E'' is order complete then $|S \circ T| = |(S \circ T)'|' \circ P$. Since $S'([-z, z]) = [-S'z, S'z]$ for z positive in E''' then $|(S \circ T)'| = |T'| \circ S'$. Therefore, $|S \circ T| = S'' \circ |T'|' \circ P$. Since $|T'|'$ has its range in E then $|S \circ T|$ has its range in E. Therefore, $|S \circ T| = |T|$ and $|T| = |T'|' \circ P$.

Let F_0 denote F equipped with the topology $0(F, F')$ of uniform convergence on the order intervals of F' (see p. 126 in [6]). A map from F_0 into E is *compact* if it maps some 0-neighborhood into a relatively compact subset of E. Let $C(F_0, E)$ denote the space of compact linear maps from F_0 into E.

(2.2) **Proposition.** *$C(F_0, E)$ coincides with $\Lambda(F, E)$, $M_*(E', F')$, and $E \tilde{\otimes}_M F'$ under the canonical identifications.*

Proof. By (2.1) it suffices to show that $C(F_0, E)$ coincides with $M_*(E', F')$ under transposition. Let T be in $C(F_0, E)$, then for some positive y' in F' the induced map $\bar{T}$ from the quotient space $F_{[-y', y']^0}$ into E is compact. Since $\bar{T}'$ maps E' into $F'_{y'}$ and $\bar{T}'$ is compact and $w^* - w$ continuous then T' is an element of $M_*(E', F')$. Conversely, if S is in $M_*(E', F')$ then let $S = S_2 \circ S_1$ where S_1 is a $w^* - w$ continuous compact map from E' into $F'_{y'}$ and S_2 is the canonical map from $F'_{y'}$ into F'. Let T_2 equal S_1' restricted to $F_{[-y', y']^0}$ and T_1 be the canonical map from F into $F_{[-y', y']^0}$, then $T = T_2 \circ T_1$ is in $C(F_0, E)$ and $S = T'$.

Definition. A subset of F is *relatively uniformly compact* if it is a compact subset of some F_y.

(2.3) **Theorem.** *The following conditions on the Banach lattice F are equivalent:*

(1) *For all Banach spaces E, $(E \otimes_M F)' = E' \tilde{\otimes}_M F'$.*

(2) *For all Banach spaces E, $S_+(F, E') = \Lambda(F, E')$.*

(3) *For all Banach spaces E, $\mathscr{L}(F_0, E') = C(F_0, E')$.*

(4) *Each order interval in F' is relatively uniformly compact.*

(5) *Each order interval in F' is compact.*

(6) *F' is isomorphic as a vector lattice to a generalized sequence space S. For each element $x = (x_\alpha)$ in S, at most countably many x_α are non-zero, and the filter of sections of (x_α) converges to x in the topology on S induced by the norm from F'.*

Proof. Since $(E \otimes_M F)' = S_+(F, E')$ (*see* (1.4)), $\mathscr{L}(F_0, E) = S_+(F, E)$ (see remark (1.9) in [4]), and $E \tilde{\otimes}_M F' = \Lambda(F, E') = C(F_0, E')$ (see (2.2)) then (1), (2), and (3) are equivalent.

(3) implies (4). Suppose $\mathscr{L}(F_0, E') = C(F_0, E')$ for all Banach spaces E. For each positive y' in F', define $T_{y'}$ as the natural map from F_0 into $(F'_{y'})'$. Then $F_{[-y', y']^0}$ is a normed subspace of $(F'_{y'})'$ and the range of $T_{y'}$ is contained in $F_{[-y', y']^0}$. $T_{y'}$ is continuous; therefore by (3) it is compact. Since $T_{y'}$ is compact there exists an x' in F' such that $T_{y'}([-x', x']^0)$ is compact in $F_{[-y', y']^0}$. Therefore, the map S induced by $T_{y'}$ from $F_{[-x', x']^0}$ into $F_{[-y', y']^0}$ is compact. Since S' is compact and maps $F'_{y'}$ into $F'_{x'}$ then $[-y', y']$ is compact in $F'_{x'}$.

(4) implies (3). Suppose each order interval in F' is relatively uniformly compact and let T be an element of $\mathscr{L}(F_0, E')$. T has the decomposition $T = T_2 \circ T_1$ where T_1 is the canonical map from F_0 into some $F_{[-y', y']^0}$ and T_2 is a continuous map from $F_{[-y', y']^0}$ into E'. By hypothesis, there exists an x' in F' such that the canonical map S from $F'_{y'}$ into $F'_{x'}$ is compact. Let P be the canonical map from F_0 into $F_{[-x', x']^0}$, and let R be the canonical map from $F_{[-x', x']^0}$ into $F_{[-y', y']^0}$. Since $R' = S$ then R is compact by Schauder's theorem, and $T = T_2 \circ R \circ P$ is in $C(F_0, E)$.

(4) implies (5) is immediate, and (5) implies (6) follows from Theorem 1 in [12].

(6) implies (4). Let $z = (z_\alpha)$ be a positive element of S, and let (α_n) be the sequence of indices for which $z_{\alpha_n} \neq 0$. Let e^β be the element of F' which corresponds to the generalized sequence $(\delta_{\alpha\beta})$ in S (where $\delta_{\alpha\beta}$ is the Kronecker delta). Since, by hypothesis, $(z_{\alpha_n} e^{\alpha_n})$ is summable to z in F' then choose a sequence of numbers (c_n) which are greater than 1 and increase to infinity such that $(c_n z_{\alpha_n} e^{\alpha_n})$ is still summable in F'. Let y' be the sum of this sequence in F'. We will show that $[-z, z]$ is compact in $F'_{y'}$. Let $w^k = (w^k_\alpha)$ be a sequence from $[-z, z]$. Observe

that if α is not an α_n then $w^k_\alpha=0$. By the compactness of $\Pi[-z_{\alpha_n}, z_{\alpha_n}]$, choose a subsequence (w^{k_j}) and a generalized sequence (w_α) so that $w^{k_j}_\alpha$ converges to w_α as j tends infinity, for each α. Since $|w_\alpha|\leqq z_\alpha$ then $w=\sum w_{\alpha_n} e^{\alpha_n}$ is in $[-z, z]$. Given $\eta>0$, choose N such that $n\geqq N$ implies $2<\eta\, c_n$. Let $S_N=\min\{c_n z_{\alpha_n}\colon 1\leqq n\leqq N\}$. Choose J so that $j\geqq J$ implies $|w^{k_j}_{\alpha_n}-w_{\alpha_n}|<\eta\,\delta_N$ for $1\leqq n\leqq N$. Thus, for any $j\geqq J$ and any n, $|w^{k_j}_{\alpha_n}-w_{\alpha_n}|<\eta\, c_n\, z_{\alpha_n}$, and $\|w^{k_j}-w\|_{F'_{y'}}<\eta$. Hence, $[-z, z]$ is compact in $F'_{y'}$.

3. Nuclear Maps and the Spaces $L^p(E)$

In this section we obtain a characterization of the space $L^p(E)$ as a space of compact linear maps from L^q into E. This result yields a property of nuclear maps defined on an (AM)-space which neatly distinguishes them from the integral linear maps. In addition, the classical Dunford-Pettis theorem is extended to the L^p spaces for $p>1$ and a proof is given of the equivalence of the Radon-Nikodym property of a Banach space E and the property that every integral map from L^∞ into E is nuclear.

Definition. Let (Ω, Σ, μ) be a finite measure space. A function $f\colon \Omega\to E$ is *measurable* if it is the almost uniform limit of a sequence of E-valued Σ-simple functions. For $1\leqq p\leqq\infty$, define $L^p(E)$ to be the space of all equivalence classes of measurable functions f identified almost everywhere and such that $\|f\|_E$ is in L^p. Equip $L^p(E)$ with the norm $\|f\|=\big\|\,\|f\|_E\big\|_{L^p}$. Let S_E denote the Σ-simple functions from Ω into E.

(3.1) **Lemma.** *For* $1\leqq p<\infty$, S_E *is dense in* $L^p(E)$.

Proof. Let f be an element of $L^p(E)$. Given $\eta>0$, find a $\delta>0$ such that if $\mu(A)<\delta$ then $\int_A \|f(t)\|^p\, d\mu(t)<\eta$. Since f is measurable then for some A in Σ such that $\mu(A)<\delta$ and some ϕ in S_E, which vanishes on A, we have $\|f(t)-\phi(t)\|^p<\eta\big(\mu(\Omega)^{-1}\big)$ for all t in $\Omega\setminus A$. Therefore,

$$\|f-\phi\|^p< \int_{\Omega\setminus A} \|f(t)-\phi(t)\|^p\, d\mu(t)+\int_A \|f(t)\|^p\, d\mu(t)<2\eta.$$

(3.2) **Theorem.** (1) *For* $1\leqq p<\infty$, $L^p(E)$ *is norm isomorphic to* $E\tilde{\otimes}_M L^p$.

(2) *For* $1<p<\infty$ *and* $1/p+1/q=1$, *the space* $\Lambda(L^q, E)$ *is norm isomorphic to* $L^p(E)$.

(3) *If* M *is an* (AM)-*space then* $\Lambda(M, E)$ *is norm isomorphic to* $M'\tilde{\otimes}_\pi E$.

Proof. Equip S_E with the norm from $L^p(E)$. Define the map I from S_E into $E\otimes_M L^p$ by $I\left(\sum_{i=1}^n x_i\,\chi_{A_i}\right)=\sum_{i=1}^n x_i\otimes\chi_{A_i}$. I is a norm isomorphism since if $\phi=\sum_{i=1}^n x_i\,\chi_{A_i}$ where the A_i are disjoint then by remark (5),

$\|I(\phi)\|_M = \left\| \sum_{i=1}^{n} \|x_i\| \chi_{A_i} \right\| = \|\phi\|_{L^p(E)}$. Since $S_{\mathbb{R}}$ is dense in L^p then the range of I is dense in $E \otimes_M L^p$. Since S_E is dense in $L^p(E)$ then I extends by continuity to a norm isomorphism of $L^p(E)$ onto $E \tilde{\otimes}_M L^p$.

The second statement follows from the first and (2.1). The third statement follows from Kakutani's representation of the (AL)-space M', (2.1), and the first statement.

Remark. (9) It is known that the space of integral maps from an (AM)-space M into E is $S_+(M, E)$ (see (4.1) in [9]). The nuclear maps from M into E are characterized in (3.2)(3) as $\Lambda(M, E)$, a closed subspace of $S_+(M, E)$.

We now show that an order summable map defined on an L^p space with range in a dual Banach space is determined by a scalar measurable function. This gives a generalization of the classical Dunford-Pettis theorem without the condition that E be separable.

In our discussion, we shall use $\mathscr{L}^q$ to denote the space of all scalar-valued measurable functions f such that $|f|^q$ is integrable or, if $q=\infty$, such that f is essentially bounded. If $f \in \mathscr{L}^q$, we shall denote its equivalence class in L^q by $\dot{f}$. $\mathscr{L}^q$ is lattice ordered by the positive cone of functions f such that $f(t) \geqq 0$ for *all* t in Ω.

(3.3) **Theorem.** (1) *Let $1 \leqq p < \infty$ and $1/p + 1/q = 1$. If T is in $S_+(L^p, E')$ then there exists a function $g\colon \Omega \to E'$ such that $\langle g, x\rangle$ is in L^q for all x in E, $\|g\|_{E'}$ is in $\mathscr{L}^q$, and for each x in E and h in L^p,*

$$\langle Th, x\rangle = \int_{\Omega} h(t) \langle g(t), x\rangle \, d\mu(t).$$

(2) *Let X be a compact Hausdorff topological space. If T is in $S_+(C(X), E')$ then there exists a positive Radon measure v and a function g from X into E' such that $\langle g, x\rangle$ is in $\mathscr{L}^1(X, v)$ for all x in E, $\|g\|_{E'}$ is in $\mathscr{L}^1(X, v)$, and for each x in E and h in $C(X)$,*

$$\langle Th, x\rangle = \int_{X} h(t) \langle g(t), x\rangle \, dv(t).$$

Proof. (1) If T is in $S_+(L^p, E')$ then (see (1.4) in [9]) $T = T_2 \circ T_1$ where T_1 is a positive map from L^p into some (AL)-space L and T_2 is a continuous map from L into E'. The adjoint map T_1' from L' into L^q is positive, so T_1' maps the unit ball in L' into an order interval $[-\dot{f}, \dot{f}]$ in L^q. T_1' has the decomposition $T_1' = I \circ R$ where R is the positive map from L' into $(L^q)_{\dot{f}}$ induced by T_1', and I is the canonical map from $(L^q)_{\dot{f}}$ into L^q. Let $S = R \circ T_2' \circ Q$, where Q is the canonical map from E into E''. Observe that the range of S is contained in $(L^q)_{\dot{f}}$. Let f be a non-negative finite-valued element of $\dot{f}$. We will now define a map from $(L^q)_{\dot{f}}$ into $(\mathscr{L}^q)_f$. For $\dot{g}$ in $(L^q)_{\dot{f}}$ and g in $\dot{g}$ define the function $\bar{g}\colon \Omega \to R$ by $\bar{g}(t)=0$ if $f(t)=0$ and $\bar{g}(t) = g(t)/f(t)$ if $f(t) \neq 0$. Note if g_1 and g_2 belong to $\dot{g}$

then $\bar{g}_1$ and $\bar{g}_2$ belong to the same equivalence in $L^{\infty}(\mu f)$. Also note that g and $\bar{g}f$ belong to the same equivalence class L^q. Define the map ϕ: $(L^q)_{\dot{f}} \to L^{\infty}(\mu f)$ by $\phi(\dot{g}) = \dot{\bar{g}}$. It can be seen that ϕ is a norm isomorphism. By Theorem 3 of Chapter 4 in [9], there exists a positive linear map ρ: $L^{\infty}(\mu f) \to \mathscr{L}^{\infty}(\mu f)$ such that $\rho(1) = 1$ and $\pi \circ \rho$ is the identity on $L^{\infty}(\mu f)$, where π: $\mathscr{L}^{\infty}(\mu f) \to L^{\infty}(\mu f)$ is the quotient map. Now define σ: $\mathscr{L}^{\infty}(\mu f) \to (\mathscr{L}^q)_f$ by $\sigma(g) = fg$. Observe that $\lambda = \sigma \circ \rho \circ \phi$ is a positive linear map from $(L^q)_{\dot{f}}$ into $(\mathscr{L}^q)_f$ such that $\lambda(\dot{f}) = f$ and $\zeta \circ \lambda$ is the identity on $(L^q)_{\dot{f}}$, where ζ is the quotient map from $(\mathscr{L}^q)_f$ into $(L^q)_{\dot{f}}$. Define $g: \Omega \to E'$ by $\langle g(t), x \rangle = (\lambda(Sx))(t)$ for all x in E and t in Ω. Since $\lambda([-\dot{f}, \dot{f}])$ is contained in $[-f, f]$ then for every t in Ω and x in E, $|\langle g(t), x \rangle| \leqq \|Sx\| f(t) < K \|x\| f(t)$ where K is the norm of S. Therefore, $g(t)$ is in E', $\|g\|_{E'}$ is in L^q, and for $\dot{h}$ in L^p and x in E

$$\langle T\dot{h}, x \rangle = \langle \dot{h}, T'x \rangle = \int_{\Omega} h(t)(T'x)(t)\, d\mu(t)$$
$$= \int_{\Omega} h(t)(\lambda Sx)(t)\, d\mu(t) = \int_{\Omega} h(t) \langle g(t), x \rangle\, d\mu(t).$$

(2) Since T is in $S_+(C(X), E')$ then (see (1.4) in [9]) $T = T_2 \circ T_1$ where T_1 is a positive map from $C(X)$ into some (AL)-space L and T_2 is a continuous map from L into E'. Since T_1' maps the unit ball in L' into an order interval $[-v, v]$ in $C(X)'$ then T_1' has the decomposition $T_1' = I \circ R$ where R is the positive map from L' into $(C(X)')_v$ and I is the canonical map from $(C(X)')_v$ into $C(X)'$. By (8.6) of Chapter V in [8], the closure of the range of I is $L^1(X, v)$. Let P be the positive projection of $C(X)'$ onto $L^1(X, v)$ given by (4.7) in Chapter I of [6], and define $S = P \circ I \circ R \circ T_2' \circ Q$ where Q is the canonical map from E into E''. Now procede as in part (1) to obtain g.

(3.4) **Proposition.** *If E' is a separable dual Banach space or if E' is a reflexive Banach space then $S_+(L^p, E') = L^q(E')$ for $1 \leqq p < \infty$, and $S_+(L^{\infty}, E') = \Lambda(L^{\infty}, E')$.*

Proof. Let T be in $S_+(L^p, E')$. If E' is separable then apply (8.15.2) in [2] and (3.3) to obtain T as an element of $L^q(E')$. Now suppose E' is reflexive. By (1.4) in [9], $T = T_2 \circ T_1$ where T_1 is a positive map from L^p into some (AL)-space L and T_2 is a continuous map from L into E'. Since E' is reflexive then T_2 is weakly compact and by (5.13) on p. 63 in [5] T_2 has separable range. Let H be the closure of the range of T_2 in E'. Since E' is reflexive then H is a separable dual space. Therefore, T is determined by an element of $L^q(H)$, a subspace of $L^q(E')$.

Let T be in $S_+(L^{\infty}, E')$. By Kakutani's theorem we can write $L^{\infty} = C(X)$. If E' is separable then apply (8.15.2) in [2], (3.3) and (3.2)(3) to obtain T as an element of $\Lambda(L^{\infty}, E')$. The case for E' reflexive is similar to the argument above.

From (2.3) and (3.2) it follows that $S_+(L^p, E) = L^q(E)$ for all Banach spaces E and $1 < p < \infty$ if and only if L^p is discrete. Therefore, some, and not all, Banach spaces E satisfy $S_+(L^p, E) = L^q(E)$. We will now develop some characteristics of Banach spaces with this property.

Definition. Let (Ω, Σ, μ) be a finite measure space. A *μ-continuous vector measure* G with values in E is a countably additive set function $G: \Sigma \to E$ such that whenever $\mu(A) = 0$ then $G(A) = 0$. G has *finite variation* if there exists a constant K such that for every partition (A_i) of Ω, $\sum \|G(A_i)\| < K$. E has the *Radon-Nikodym property* if for each finite measure space (Ω, Σ, μ) and each μ-continuous vector measure G of finite variation there exists a function g in $L^1(E)$ such that for each A in Σ and x' in E' $\langle G(A), x'\rangle = \int\limits_\Omega \chi_A(t) \langle g(t), x'\rangle \, d\mu(t)$.

(3.5) **Theorem.** *The following conditions on the Banach space E are equivalent:*

(1) *E has the Radon-Nikodym property.*

(2) *For all finite measure spaces, every integral map from L^∞ into E is nuclear.*

(3) *For all finite measure spaces, $S_+(L^\infty, E) = \Lambda(L^\infty, E)$.*

(4) *For all finite spaces and all p such that $1 \leqq p < \infty$, $S_+(L^p, E) = L^q(E)$.*

(5) *For all finite measure spaces and a particular p such that $1 \leqq p < \infty$, $S_+(L^p, E) = L^q(E)$.*

Proof. (2) implies (1). Let $G: \Sigma \to E$ be a μ-continuous vector measure of finite variation. Using G, we define a map T from L^∞ into E, then we show that T is a nuclear map which is represented by a function g in $L^1(E)$. If A is in Σ then define $T(\chi_A) = G(A)$. Since G is μ-continuous then T is well-defined on the equivalence classes of characteristic functions. Extend T by linearity to simple functions. If $\phi = \sum_{i=1}^n c_i \chi_{A_i}$ where the A_i's are disjoint then $\|T\phi\| \leqq \sum_{i=1}^n |c_i| \, \|G(A_i)\| \leqq K \|\phi\|$, where K is the total variation of G. Therefore T can be extended to all of L^∞ by continuity. Next we show that T is order summable. Let (ϕ_i) be a positive summable sequence of simple functions in L^∞, and for $i = 1, \ldots, n$ write $\phi_i = \sum_{j=1}^m c_{ij} \chi_{A_j}$ where the A_j's are disjoint and the c_{ij} are positive. Since

$$\begin{aligned} \sum_{i=1}^n \|T\phi_i\| &= \sum_{i=1}^n \left\| \sum_{j=1}^m c_{ij} G(A_j) \right\| \\ &\leqq \sum_{j=1}^m \left(\sum_{i=1}^n c_{ij} \right) \|G(A_j)\| \leqq K \left\| \sum_{i=1}^n \phi_i \right\|, \end{aligned}$$

T is continuous, and since simple functions are dense in L^∞ then T is order summable. It follows from (4.1) in [9] that T is integral, and so by hypothesis (2) T is nuclear. Since T is a nuclear map then it is given by $u=\sum_{i=1}^{\infty} y_i \otimes x_i$ in $(L^\infty)' \tilde{\otimes}_\pi E$ (see (6.5) and (7.1) in Chapter III in [8]). By the Radon-Nikodym theorem, the map $R: E' \to (L^\infty)'$, which associates with each x' in E' the scalar measure $A \to \langle G(A), x' \rangle$, has range in L^1. Since $R=T'$ then R is nuclear, and it is given by $\sum_{i=1}^{\infty} x_i \otimes y_i$ in $E \tilde{\otimes}_\pi (L^\infty)'$. Let P be the band projection of $(L^\infty)'$ onto L^1 (see (8.6) of Chapter V in [8] and (4.7) of Chapter I in [6]). Since the range of R is contained in L^1 then $(P \circ R)' = Q \circ T$ where Q is the canonical map from E into E''. Since $(P \circ R)'$ is given by $\sum_{i=1}^{\infty} (P y_i) \otimes x_i$ then T is given by $\sum_{i=1}^{\infty} (P y_i) \otimes x_i$ in $L^1 \tilde{\otimes}_\pi E$. Let $g=\sum_{i=1}^{\infty} (P y_i)\, x_i$ in $L^1(E)$ then for A in Σ and x' in E'

$$\langle G(A), x' \rangle = \langle T(\chi_A), x' \rangle = \left\langle \sum_{i=1}^{\infty} \langle P y_i, \chi_A x_i, \rangle\, x' \right\rangle$$
$$= \int_{\Omega} \chi_A(t) \langle g(t), x' \rangle \, d\mu(t).$$

Therefore, E has the Radon-Nikodym property.

The equivalence of (3) and (2) follows directly from (4.1) in [9] and (3.2) (3).

(1) implies (2). Let T be an integral linear map from L^∞ into E. Since T is order summable then it has the decomposition $T = T_1 \circ T_2$ (see (1.6) in [9]) where T_1 is a positive map from L^∞ into some (AL)-space L and T_2 is a continuous map from L into E. By Kakutani's theorem, there exists a finite measure space $(\Omega_1, \Sigma_1, \nu)$ such that L is norm and lattice isomorphic to $L^1(\Omega_1, \Sigma_1, \nu) = L^1_\nu$. Since T_1 is positive then $T_1([-1, 1])$ is contained in some order interval $[-f, f]$ in L^1_ν. It was shown in the proof of (3.3) that $(L^1_\nu)_f$ is norm and lattice isomorphic to $L^\infty(\Omega_1, \Sigma_1, \nu f) = L^\infty_{\nu f}$. Therefore, the map T_1 has the decomposition given by the diagram in Fig. 4, where each map in this diagram is either canonical or induced by T_1. Define

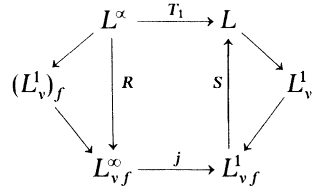


Fig. 4

$G: \Sigma_1 \to E$ by $G(A) = T_2 \circ S \circ j(\chi_A)$ for A in Σ_1. Since $(\Omega_1, \Sigma_1, \nu f)$ is a finite measure space and $T_2 \circ S \circ j$ is order summable then G is countably additive, νf-continuous, and has finite variation. By hypothesis (1), there exists $u = \sum_{i=1}^{\infty} f_i \otimes x_i$ in $L^1_{\nu f} \tilde{\otimes}_\pi E$ (see (6.5) in Chapter III in [8]) such that (f_i) is a norm bounded sequence in $L^1_{\nu f}$ and (x_i) is an absolutely summable sequence in E and $\langle G(A), x' \rangle = \int_{\Omega_1} \chi_A \langle g, x' \rangle f d\nu$ where $g = \sum_{i=1}^{\infty} f_i x_i$ in $L^1_{\nu f}(E)$. If ϕ is a Σ_1-simple function then $\langle T_2 \circ S \circ j(\phi), x' \rangle = \int_{\Omega_1} \langle g, x' \rangle \phi f d\nu = \sum_{i=1}^{\infty} \langle f_i, \phi \rangle \langle x_i, x' \rangle$. Using the absolute summability of the x_i's one can show for an arbitrary h in $L^\infty_{\nu f}$ that $T_2 \circ S \circ j(h) = \sum_{i=1}^{\infty} \langle f_i, h \rangle x_i$. Therefore, $T_2 \circ S \circ j$ is nuclear (see (7.1) in Chapter III in [8]), and hence $T = T_2 \circ S \circ j \circ R$ is nuclear.

(2) implies (4). Let T be an element of $S_+(L^p, E)$ where $1 \leqq p < \infty$, and let S be the canonical map from L^∞ into L^p. Since $T \circ S$ is in $S_+(L^\infty, E)$ then by (4.1) in [9] and (2) it is nuclear. Since T' maps E' into L^q then the range of $(T \circ S)'$ lies by L^1. By the proof that (2) implies (1), $T \circ S$ is given by an element f in $L^1(E)$. Note that f is measurable. $(T \circ S)'$ is the map from E' into L^1 defined by $(T \circ S)'(x') = \langle f, x' \rangle$. By (1.5) and (1.6) in [9], $T'(U^0)$ is contained in an order interval $[-g, g]$ in L^q, where U is the unit ball in E. Since S' is the canonical map from L^q into $L^1 \subseteq (L^\infty)'$ then $\langle f, x' \rangle$ is in $[-g, g]$ for all x' in U^0. Therefore, $\|f\| = \sup \{\langle f, x' \rangle : x' \in U^0\}$ is an element of L^q, and since f is measurable then f is in $L^q(E)$. Hence, $S_+(L^p, E) = L^q(E)$.

(4) implies (5) is immediate.

(5) implies (2). It follows from an argument similar to the proof of (2) implies (4) that if $1 < p < \infty$ and $S_+(L^p, E) = L^q(E)$ then $S_+(L^1, E) = L^\infty(E)$. Therefore we can assume that $S_+(L^1, E) = L^\infty(E)$ and let T be an integral map from L^∞ into E. T has the decomposition $T = T_2 \circ T_1$ and T_1 has the decomposition given in the diagram of Fig. 4 as shown in the proof of (1) implies (2). By hypothesis, $T_2 \circ S$ is given by an element g in $L^\infty_{\nu f}(E)$. Since $T_2 \circ S \circ j$ is given by the same function g which can be considered as an element of $L^1_{\nu f}(E)$ then $T_2 \circ S \circ j$ is nuclear. Therefore, $T_2 \circ S \circ j \circ R = T$ is nuclear.

References

1. Dinculeanu, N.: Integral representation of linear operation. Queen's Papers in Pure and Applied Mathematics, no. 10, 1967, pp. 2-41.
2. Edwards, R.: Functional analysis. New York: Holt, Rinehart and Winston 1965.

3. Grothendieck, A.: Produits tensoriels topologique et espace nucléaires. Mem. Amer. Math. Soc. No. 16, 1955.
4. Jacobs, H.: Ordered topological tensor products. Thesis, University of Illinois, 1969.
5. Lotz, H. P.: Topological tensor products, linear mappings, and nuclear spaces. Lecture notes prepared by A. L. Peressini, University of Illinois, 1971.
6. Peressini, A. L.: Ordered topological vector spaces. New York: Harper and Row 1967.
7. Pietsch, A.: Nukleare lokalkonvexe Räume. Berlin: Akademie Verlag 1965.
8. Schaefer, H.: Topological vectors spaces. New York: Macmillan 1966.
9. Schlotterbeck, U.: Über Klassen majorisierbarer Operatoren auf Banachverbänden. Dissertation, Eberhard-Karls-Universität zu Tübingen, 1969.
10. Tulcea, A. I., Tulcea, C. I.: Topics in the theory of lifting. New York: Springer 1969.
11. Uhl, J. J.: On a class of operators on Orlicz spaces. Studia Math. **40**, 17–22 (1971).
12. Walsh, B.: On characterizing Köthe sequence spaces. Math. Ann. **175**, 253–256 (1968).

Dr. J. W. Chaney
273 Altgeld Hall
University of Illinois
Urbana, Illinois 61801
USA

(Received April 6, 1972)

Math. Z. 129, 21–23 (1972)
© by Springer-Verlag 1972

Non-Replaceable Matrices

Andrew K. Snyder and Albert Wilansky

A sufficient condition for non-replaceability (defined below), based on a formula in [4], was given in [1] and used to advantage. When this condition applies, it gives a very simple proof of non-replaceability. It does indeed apply to the noted original example of K. Zeller and so yields a much simpler proof than the one given in [3]. It was our hope that the condition would turn out to be necessary; however, we are able to construct a coregular conservative non-replaceable matrix not satisfying the sufficient condition. (Such an example was also outlined in a letter from Professor K. Zeller to the authors.)

Let A be an arbitrary infinite matrix and $c_A = \{x: Ax \in c\}$ where c is the set of convergent sequences. We call A *replaceable* if there exists B with $c_B = c_A$ and $b_k = \lim_n b_{nk} = 0$ for each k. Clearly a replaceable matrix has convergent columns, and we shall consider only such matrices; an equivalent condition is $c_A \supset E^\infty$, the set of finite sequences. As usual, c_A is made into an FK space, [5], and following [4], we define $F = \{x \in c_A: \sum f(\delta^k)\, x_k$ is convergent for all $f \in c'_A\}$, $B = \{x \in c_A: \{x^{(n)}\}$ is bounded$\}$ where $x^{(n)} = (x_1, x_2, \ldots, x_n, 0, 0, \ldots)$, $I = \{x \in c_A: \sum a_k x_k$ is convergent$\}$.

It is proved in [4], Lemma 4.3, that $F = B \cap I$. Since F and B are invariant, i.e. the same for all A such that c_A is the same ([5], p. 203, Corollary 1), it follows that a sufficient condition for non-replaceability is $F \neq B$.

This test applies to the matrix of [2], p. 209 (pointed out to us by Grahame Bennett). It also applies to the matrix of [3], p. 657 as we now show. The matrix is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 1 & a_2 & 0 & 0 & 0 & 0 & \ldots \\ 0 & a_2 & 1 & 0 & 0 & 0 & \ldots \\ 0 & a_2 & 1 & a_4 & 0 & 0 & \ldots \\ 0 & a_2 & 0 & a_4 & 1 & 0 & \ldots \\ 0 & a_2 & 0 & a_4 & 1 & a_6 & \ldots \\ \ldots & & & & & & \end{pmatrix}$$

where $a_{2k} \neq 0$ for all k, and $\sum |a_{2k}| < \infty$. Now let $\{x_{2n-1}\}$ be bounded, divergent, and satisfy $x_{2n+1} - x_{2n-1} \to 0$. Define x_{2n} by $\sum_{k=1}^{n} a_{2k} x_{2k} = -x_{2n-1}$. It is obvious that $x \in c_A$ and $x \notin I$. However $x \in B$. To see this, note that c_A is a Banach space with norm h, where $h(x) = \sup |(Ax)_n|$. Thus it is sufficient to prove that $\left|\sum_{i=1}^{k} a_{ni} x_i\right| < M$. But this quantity is $\left|\sum_{i=1}^{r} a_{2i} x_{2i} + \varepsilon x_s\right|$ where r, s are suitably chosen, s is odd, and $\varepsilon = 0$ or 1. So it is equal to $|(Ax)_r - x_{2r-1} + \varepsilon x_s| \leqq h(x) + 2 \sup |x_{2i-1}|$. So, finally, $x \in B \setminus I$ hence $B \neq B \cap I = F$.

We shall now show a non-replaceable coregular conservative matrix which fails the test, i.e. has $F = B$.

Lemma 1. *For $x \in B$ we have $x_k = O(\|\lambda^k\|_\infty^{-1})$ where λ^k is the k-th column of A.*

We have $|a_{nk} x_k| = |(A x^{(k)})_n - (A x^{(k-1)})_n| \leqq h(x^{(k)}) + h(x^{(k-1)})$ where $h(x) = \sup |(Ax)_n|$. Since h is a continuous seminorm the hypothesis implies that $|a_{nk} x_k| < M$ for all n, k. Taking the supremum over n yields the result.

Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 1 & 1 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{4} & 1 & 1 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{4} & 0 & \frac{1}{8} & 1 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

It follows from Lemma 1 that every member of B is bounded, hence is in I. Thus $B \subset B \cap I = F \subset B$, so that $B = F$. To show that A is not replaceable we note that if it were, since it is coregular and conservative, lim would be continuous on c_A and so 1 would not belong to the closure of E^∞. But indeed 1 does belong to this closure for let i be a positive integer and define $v = v^i$ by $v_n = 0$ for $n < 2i$, $v_n = 1$ for $n > 2i+1$, $v_{2i} = 2^{i+2}$, $v_{2i+1} = -2^{i+2}$. Now $(Av)_n = 0$ for $n \leqq i+2$, $(Av)_n = 2^{-i-1} + 2^{-n+1}$ for $n \geqq i+3$, hence $h(v) \to 0$ as $i \to \infty$. Moreover $v_n^i \to 0$ as $i \to \infty$ for each n, so by [5], p. 228, Theorem 2, we have $v^i \to 0$. Since $1 - v^i \in E^\infty$ and $1 - v^i \to 1$ as $i \to \infty$ the proof is complete.

References

1. Bennett, G.: Distinguished subsets and summability invariants. To appear in Studia Math.
2. Chang, S.C., MacPhail, M.S., Snyder, A.K., Wilansky, A.: Consistency and replaceability for conull matrices. Math. Z. **105**, 208–212 (1968).

3. Wilansky, A.: Summability: the inset, replaceable matrices, the basis in summability space. Duke Math. J. **19**, 647–660 (1952).
4. Wilansky, A.: Distinguished subsets and summability invariants. J. Analyse Math. **12**, 327–350 (1964).
5. Wilansky, A.: Functional analysis. Waltham, Mass.: Ginn Blaisdell 1964.

Prof. A. K. Snyder and Prof. A. Wilansky
Department of Mathematics
Lehigh University
Bethlehem, PA 18015
USA

(Received December 6, 1971)

Math. Z. 129, 25 – 42 (1962)
© by Springer-Verlag 1972

Normen auf freien topologischen Gruppen

Herbert Abels

$F(X)$ sei die freie topologische Gruppe über dem vollständig regulären topologischen Hausdorff-Raum X. Es wird bewiesen, daß die reduzierte Länge $l(w)$ eines Wortes $w \in F(X)$ gleich dem Supremum der Zahlen $N(w)$ ist, wobei N über alle stetigen Normen auf $F(X)$ läuft, deren Werte auf dem Bild von X in $F(X)$ kleiner oder gleich 1 sind. Der Hauptsatz 2.2 sagt genauer aus, daß sich die Normen N ihrem Supremum l in einer gewissen Weise gleichmäßig nähern.

Als Folgerung erhält man (Korollar 3.2), daß jede folgenkompakte Teilmenge einer solchen freien topologischen Gruppe beschränkte reduzierte Länge hat. Das folgt auch aus einem Satz von Graev ([5], Satz 4 = [6], Satz 26) mit einem Satz von Samuel ([13], 3, Satz 2, S. 596). Aus Korollar 3.2 folgt, daß zwei naheliegende Definitionen der kompakten Definierbarkeit für hausdorffsche – nicht notwendig lokal kompakte – topologische Gruppen übereinstimmen (s. [1], Proposition 2.2 für den Fall lokal kompakter Gruppen).

Wir bemerken noch, daß eine freie topologische Gruppe $F(X)$ über einem kompakten unendlichen Hausdorff-Raum folgende eigenartige Kombination von Eigenschaften hat: Sie ist abzählbare Vereinigung von kompakten Teilmengen, k-Raum im Sinne von Kelley ([8], S. 230, Beweis: [5], Satz 4 = [6], Satz 26) und vollständig in ihrer rechten oder linken uniformen Struktur ([5], Satz 6), aber nicht lokal folgenkompakt (Korollar 3.4) und kein Baire-Raum (Korollar 3.5) und kein Punkt von $F(X)$ hat eine abzählbare Umgebungsbasis (Korollar 3.6).

Der Hauptsatz wird bewiesen, indem geeignete Homomorphismen $f: F(X) \to G$ in die Liegruppe G der biholomorphen Homöomorphismen des Inneren des Einheitskreises in der komplexen Ebene $\mathbb{C}$ konstruiert werden und man auf G die Norm

$$N(g) = d(0, g(0))$$

betrachtet, wo d die Metrik der hyperbolischen Geometrie ist. Die stetigen Normen $N \circ f$ auf $F(X)$ müssen noch so normiert werden, daß ihre

Einschränkungen auf das Bild von X in $F(X)$ Werte $\leqq 1$ haben. Daß die so entstehenden stetigen Normen gegen die reduzierte Länge gewisser Wörter $w \in F(X)$ konvergieren, läßt sich mit Hilfe von hyperbolischer Geometrie und elementarer Analysis zeigen.

Zum Aufbau der Arbeit: Die Paragraphen 1 und 2 bringen Vorbereitungen und im Paragraphen 2 wird der Hauptsatz formuliert. Am Ende des Paragraphen werden Verallgemeinerungen des Hauptsatzes auf andere Typen von freien topologischen Gruppen diskutiert. Im Paragraphen 3 werden Folgerungen gezogen. Die restlichen Paragraphen dienen dem Beweis.

1. Freie topologische Gruppen (s. [13, 7, 4] und [9])

X sei ein topologischer Raum. Es gibt eine hausdorffsche topologische Gruppe $F(X)$ und eine stetige Abbildung f_X: $X \to F(X)$ mit der folgenden universellen Eigenschaft: Zu jeder stetigen Abbildung f: $X \to G$ in eine hausdorffsche topologische Gruppe G gibt es genau einen stetigen Homomorphismus $\bar{f}$: $F(X) \to G$, so daß $f = \bar{f} \circ f_X$. Offenbar sind $F(X)$ und f_X hierdurch bis auf Isomorphie eindeutig bestimmt und $f_X(X)$ erzeugt $F(X)$. f_X: $X \to F(X)$ heißt die *freie topologische Gruppe über X*.

Man kann $F(X)$ folgendermaßen konstruieren (s. [7, 13]): M sei die Menge aller hausdorffschen topologischen Gruppen G, die als abstrakte Gruppen Faktorgruppen der freien Gruppe mit der Basis X sind und für die die natürliche Abbildung i_G: $X \to G$ stetig ist. M ist tatsächlich eine Menge. Die Untergruppe $F(X)$ von $\prod_{G \in M} G$, die von der Teilmenge $(\prod_{G \in M} i_G)(X)$ erzeugt wird, zusammen mit $f_X := \prod_{G \in M} i_G$: $X \to F(X)$ ist offenbar eine freie topologische Gruppe über X.

Als vollständige Regularisierung eines topologischen Raumes X wollen wir eine stetige Abbildung g: $X \to X_u$ bezeichnen mit der folgenden universellen Eigenschaft: Wenn h: $X \to Y$ eine stetige Abbildung in einen vollständig regulären Hausdorff-Raum ist, dann gibt es genau eine stetige Abbildung h_u: $X_u \to Y$ mit $h_u \circ g = h$. Wenn j: $X \to \hat{X}$ die Stone-Čech-Kompaktifizierung des topologischen Raumes X ist, dann ist $h_u := j|X$: $X \to j(X)$ offenbar eine vollständige Regularisierung von X. Da jede hausdorffsche topologische Gruppe G eine uniforme Struktur besitzt (s. [8], Kapitel 6, Problem 0) und daher vollständig regulär ist (s. [8], 6.17), läßt sich jede stetige Abbildung f: $X \to G$ durch X_u faktorisieren und daher kann man $F(X) = F(X_u)$ und $f_X = f_{X_u} \circ g$ annehmen. Wir wollen deshalb im folgenden nur vollständig reguläre Hausdorff-Räume betrachten. Falls X solch ein Raum ist, sieht man sofort aus der Definition der freien topologischen Gruppe über X, daß f_X: $X \to f_X(X)$ ein Homöomorphismus ist, weil es genügend viele stetige Abbildungen von X nach $\mathbb{R}$ gibt. Der grundlegende Satz über freie topologische Gruppen ist der folgende

Satz 1.1. *Falls X vollständig regulär ist, ist die $F(X)$ zugrundeliegende abstrakte Gruppe die freie Gruppe mit der Basis X.*

Für Beweise siehe die angegebenen Arbeiten. Satz 1.1 wird auch aus dem in dieser Arbeit bewiesenen Satz 2.2 folgen.

2. Normen (s. [9])

G sei eine Gruppe. Eine Norm N auf G ist eine Funktion $N\colon G \to \mathbb{R}$ mit den beiden Eigenschaften:

(N 1) $N(1)=0$.

(N 2) $N(x \cdot y^{-1}) \leqq N(x)+N(y)$ für je zwei Elemente x und y aus G.

Aus (N 2) folgt mit $x=y$, daß

$$N(x) \geqq 0 \qquad \text{für alle } x \in G$$

gilt. Weiter folgt aus (N 2) mit $x=1$: $N(y^{-1}) \leqq N(y)$. Vertauschen wir die Rollen von y und y^{-1}, so folgt:

$$N(y)=N(y^{-1}) \qquad \text{für alle } y \in G. \tag{2.1}$$

Aus (N 2) folgt weiter:

$$-N(y^{-1}) \leqq N(x \cdot y^{-1})-N(x) \leqq N(y). \tag{2.2}$$

Wenn G eine topologische Gruppe ist, ist eine Norm also stetig, wenn sie an der Stelle $1 \in G$ stetig ist. Sie ist dann sogar gleichmäßig stetig bezüglich der linken uniformen Struktur.

Wenn f eine beschränkte reellwertige Funktion auf einer Gruppe G ist, dann ist

$$N(x) := \sup_{y \in G} |f(y\,x)-f(y)|$$

eine Norm. Denn $N(1)=0$ und

$$\begin{aligned} N(x \cdot y^{-1}) &= \sup_{z \in G} |f(z \cdot x \cdot y^{-1})-f(z)| \\ &\leqq \sup_{z \in G} |f(z \cdot x \cdot y^{-1})-f(z \cdot x)| + \sup_{z \in G} |f(z \cdot x)-f(z)| \\ &= N(y)+N(x). \end{aligned}$$

Satz 2.1. *G sei eine topologische Gruppe. Dann ist das System der Mengen $N^{-1}([0,1))$, wo N alle stetigen Normen $N\colon G \to \mathbb{R}$ durchläuft, eine Umgebungsbasis von 1 in G.*

Beweis. Wir verwenden die folgende Tatsache: Zu jeder Umgebung V eines Punktes x eines uniformen Raumes X gibt es eine gleichmäßig stetige beschränkte Funktion $f\colon X\to \mathbb{R}$, so daß $f(x)=0$ und $f^{-1}([0,1))\subset V$ (vgl. [8], 6.12).

Wenn f eine beschränkte reellwertige Funktion auf einer topologischen Gruppe G ist, die bezüglich der linken uniformen Struktur gleichmäßig stetig ist, dann ist

$$N(x)=\sup_{y\in G}|f(yx)-f(y)|$$

eine stetige Norm. Denn zu jeder reellen Zahl $\varepsilon>0$ gibt es dann eine Umgebung U von 1 in G, so daß $|f(yx)-f(y)|<\varepsilon$ für alle $y\in G$ und alle $x\in U$ gilt. N ist also stetig bei 1, also überall.

Die wichtigste Norm auf einer freien Gruppe mit der Basis X ist die reduzierte Länge l der Worte in Elementen von X. Sie ist durch die folgende Eigenschaft charakterisiert:

$$l(w)=\sup_N N(w),$$

wobei N über alle Normen läuft, deren Werte auf X kleiner oder gleich 1 sind. Es erhebt sich die Frage, ob dieselbe Formel gilt, wenn N über alle *stetigen* Normen auf der freien topologischen Gruppe über einem topologischen Raum X läuft. Daß sie gilt, falls X vollständig regulär hausdorffsch ist, folgt aus dem Hauptresultat Satz 2.2 der vorliegenden Arbeit, das nun formuliert werden soll.

Die freie topologische Gruppe über einem diskreten topologischen Raum mit n Punkten $x_1,\ldots,x_n$ ist offenbar die freie Gruppe $F(x_i;\, i=1,\ldots,n)$ mit der Basis $x_1,\ldots,x_n$ versehen mit der diskreten Topologie. Wenn $x_1,\ldots,x_n$ Punkte eines vollständig regulären Hausdorff-Raumes X sind, dann induziert die Einbettung $\{x_i\}\to X \xrightarrow{f_X} F(X)$ einen Homomorphismus $g\colon F(x_i;\, i=1,\ldots,n)\to F(X)$. Unter diesen Voraussetzungen gilt

Satz 2.2. *Zu jeder reellen Zahl* $\varepsilon>0$ *gibt es eine stetige Norm* $N\colon F(X)\to\mathbb{R}$ *mit*

$$N\circ f_X(x)\leqq 1 \qquad \textit{für alle } x\in X$$

und

$$N\circ g(w)\geqq(1-\varepsilon)\cdot l(w) \qquad \textit{für alle } w\in F(x_i;\, i=1,\ldots,n).$$

Insbesondere ist g injektiv; daraus folgt Satz 1.1.

Ferner folgt die oben behauptete Formel:

Korollar 2.3. *Für jeden vollständig regulären topologischen Raum* X *gilt*

$$l(w)=\sup_N N(w) \qquad \textit{für alle } w\in F(X),$$

wobei N *über alle stetigen Normen auf* $F(X)$ *läuft, deren Werte auf* $f_X(X)$ *kleiner oder gleich* 1 *sind.*

Der Satz 2.2 wird in den Paragraphen 5ff. bewiesen.

Bemerkungen. 1. Nakayama hat in [11] die *uniforme freie topologische* Gruppe $F_l(X)$ über einem uniformen Hausdorff-Raum X definiert: Eine gleichmäßig stetige Abbildung $X \xrightarrow{f_l} F_l(X)$ von X nach $F_l(X)$, versehen mit der linken uniformen Struktur, heißt uniform freie topologische Gruppe über X, wenn es zu jeder gleichmäßig stetigen Abbildung von X in eine Hausdorffsche topologische Gruppe G, versehen mit der linken uniformen Struktur, einen stetigen Homomorphismus $F_l(X) \to G$ gibt, der das Diagramm

$$\begin{array}{ccc} X & \longrightarrow & F_l(X) \\ & \searrow \quad \swarrow & \\ & G & \end{array}$$

kommutativ macht.

Auch für uniform freie topologische Gruppen über einem uniformen Hausdorff-Raum gilt Satz 2.2 entsprechend, wobei man solche Normen N zur Konkurrenz zuläßt, für die $N \circ f_l$: $X \to \mathbb{R}$ gleichmäßig stetig sind.

Zum Beweis bemerken wir, daß es in jedem uniformen Hausdorff-Raum X zu je zwei Punkten x und y eine gleichmäßig stetige Abbildung f von X in das Einheitsintervall $[0, 1] \subset \mathbb{R}$ gibt mit $f(x)=0$ und $f(y)=1$ (s. Beweis von Satz 2.1). Satz 2.2 für uniform freie topologische Gruppen wird nun auf Satz 4.1 zurückgeführt, wie am Ende des Paragraphen 4. Die dort auftretende Abbildung f_1 können wir gleichmäßig stetig wählen, die übrigen Abbildungen sind sowieso gleichmäßig stetig.

2. Für die *freie präkompakte topologische Gruppe* (s. [7, 13]) über einem vollständig regulären Hausdorff-Raum X ist das Analogon zu Satz 2.2 falsch. Denn jede stetige Norm auf einer Gruppe ist gleichmäßig stetig, das Bild einer Norm auf einer präkompakten – oder äquivalent: total beschränkten – Gruppe ist also auch total beschränkt, also beschränkt. Das widerspricht aber der Aussage in 2.2.

3. Für die *freie abelsche topologische Gruppe* $A(X)$ (s. [9, 7, 13, 4]) über einem vollständig regulären Hausdorff-Raum X gilt Satz 2.2 analog, wobei die reduzierte Länge durch die Summe der Exponentenbeträge ersetzt wird:

$$e\left(\prod x_i^{a_i}\right)=\sum |a_i|.$$

Zum Beweis beachte man, daß es auf dem vollständig regulären Hausdorff-Raum X eine stetige Abbildung $X \to \mathbb{R}^n$ mit $f(x_i)=e_i$ gibt, wo $\{e_i\}$ die kanonische Basis des $\mathbb{R}^n$ ist. Es sei

$$N_1\colon \mathbb{R}^n \to \mathbb{R}$$

die Norm mit

$$N_1(y_1, \dots, y_n) = \sum_{i=1}^{n} |y_i|.$$

Der von f induzierte stetige Homomorphismus sei $\bar{f}: A(X) \to \mathbb{R}^n$. Dann erfüllt $N := N_1 \circ \bar{f}$ die Gleichung

$$N \circ g = e,$$

wo $g: A(x_i; i=1, \dots, n) \to A(X)$ der von der Einbettung $x_i \to X$ induzierte Homomorphismus ist.

3. Topologie der freien topologischen Gruppen

Eine Norm N auf der freien topologischen Gruppe $F(X)$ über dem vollständig regulären topologischen Raum X heißt *zulässig*, wenn für jedes Element $a \in F(X)$ und jedes Element $y \in X$ die Funktion $x \mapsto N(a^{-1} \cdot x \cdot y^{-1} \cdot a)$ von X nach $\mathbb{R}$ stetig ist an der Stelle $y \in X$.

Lemma 3.1. *Jede zulässige Norm ist stetig.*

Dieses Lemma folgt aus der Konstruktion von Markov [9], allerdings mit einem etwas engeren Begriff von „zulässig". Wir geben hier einen einfacheren Beweis: Z sei der Vektorraum der zulässigen Normen auf $F(X)$. Die Menge $\mathfrak{B} := \{N^{-1}([0, r)) : r > 0, r \in \mathbb{R}, N \in Z\}$ ist eine Filterbasis auf $F(X)$. Wenn $V = N^{-1}([0, r)) \in \mathfrak{B}$ und $N \in Z$ ist, dann gilt für $W := N^{-1}([0, r/2))$ wegen (N2): $W \cdot W^{-1} \subset V$. Wenn α ein innerer Automorphismus von $F(X)$ ist und $N \in Z$ ist, dann ist auch $N \circ \alpha \in Z$. Daher ist mit $V \in \mathfrak{B}$ auch $\alpha(V) \in \mathfrak{B}$. Es gibt daher genau eine Topologie auf $F(X)$, für die $F(X)$ eine topologische Gruppe ist und $\mathfrak{B}$ eine Basis des Umgebungsfilters von 1 ist. Wir nennen diese Topologie die zulässige Topologie. Sie ist offenbar feiner als die Topologie von $F(X)$. Eine Abbildung f eines topologischen Raumes Y nach $F(X)$ mit $f(y) = 1$ ist dann und nur dann stetig bezüglich der zulässigen Topologie an der Stelle y, wenn $N \circ f$ für alle zulässigen Normen N an der Stelle y stetig ist.

Ich behaupte, daß $X \xrightarrow{f_X} F(X)$ stetig ist in der zulässigen Topologie von $F(X)$. Sei $y \in X$, $R_{y^{-1}}: F(X) \to F(X)$ sei die Rechtstranslation $w \mapsto w \cdot y^{-1}$ mit y^{-1}. Da nach Voraussetzung die Abbildungen $N \circ R_{y^{-1}} \circ f_X: X \to R$ mit $x \to N(x \cdot y^{-1})$ für alle zulässigen Normen N an der Stelle y stetig sind, ist f_X an jeder Stelle $y \in X$ stetig. $X \xrightarrow{f_X} F(X)$ hat daher auch in der zulässigen Topologie die universelle Eigenschaft einer freien topologischen Gruppe. Die Topologie von $F(X)$ und die zulässige Topologie stimmen also überein. Insbesondere ist jede zulässige Norm stetig an der Stelle 1, also überall, q.e.d.

Korollar 3.2. *Sei w_m eine Folge von Elementen einer freien topologischen Gruppe über einem vollständig regulären Hausdorff-Raum. Wenn die Folge* *der reduzierten Längen $l(w_m)$ gegen unendlich divergiert, dann besitzt die Folge keinen Häufungspunkt.*

Beweis. X sei ein vollständig regulärer Hausdorff-Raum. Wenn es eine Folge w_m von Elementen von $F(X)$ gibt, die einen Häufungspunkt w besitzt, für die aber $l(w_m)$ gegen unendlich divergiert, dann gibt es eine Teilfolge w_n mit dem Häufungspunkt w und $l(w_n) \geqq 4^n + 1$. Dann gibt es nach Korollar 2.3 zu jedem w_n eine stetige Norm N_n auf $F(X)$, so daß

$$N_n(w_n) > 2^n \tag{3.1}$$

und

$$N_n | f_X(X) \leqq \frac{1}{2^n}. \tag{3.2}$$

Aus (3.2) folgt $N_n \leqq \frac{l}{2^n}$. Die Reihe $\sum_{n=1}^{\prime} N_n$ konvergiert daher gleichmäßig auf jeder Menge, auf der l beschränkt ist. Die Summe S der Reihe ist folglich eine Norm, die auf jeder Menge stetig ist, auf der l beschränkt ist. S ist insbesondere eine zulässige Norm. Nach Lemma 3.1 ist S also stetig. Dann ist $S(w)$ Häufungspunkt von $S(w_n) \geqq N_n(w_n)$ im Gegensatz zu (3.1).

Bemerkung. Man kann Korollar 3.2 auch aus Resultaten von Graev ([5] oder [6]) und Samuel [13] herleiten. Die Details sind einem Kürzungsvorschlag des Referenten zum Opfer gefallen.

Die folgenden Aussagen dieses Paragraphen beschäftigen sich weiter mit der eigenartigen Topologie der freien topologischen Gruppen.

Lemma 3.3. *Wenn X ein vollständig regulärer und nicht-diskreter Hausdorff-Raum ist, dann gibt es in jeder Umgebung von $1 \in F(X)$ Elemente beliebiger Länge.*

Beweis. Es sei x ein Punkt aus X, so daß jede Umgebung von x noch wenigstens einen weiteren Punkt enthält. V sei eine beliebige Umgebung von $1 \in F(X)$. Zu jeder natürlichen Zahl n gibt es dann eine Umgebung W von x in $F(X)$, so daß $W^n \cdot x^{-n} \subset V$ gilt. Falls y ein von x verschiedener Punkt aus X mit $f_X(y) \in W$ ist, dann hat $y^n \cdot x^{-n} \in V$ die reduzierte Länge $2n$.

Aus Korollar 3.2 folgt nun

Korollar 3.4. *Wenn eine freie topologische Gruppe lokal kompakt (lokal folgenkompakt) ist, dann ist sie diskret.*

Korollar 3.5[1]. *Wenn eine freie topologische Gruppe ein Baire-Raum ist, dann ist sie diskret.*

Beweis. In der freien topologischen Gruppe $F(X)$ über dem vollständig regulären Hausdorff-Raum X sind die Mengen $X_n := (f_X(X) \cup \{1\} \cup$

[1] Die entsprechende Aussage für die freie abelsche topologische Gruppe $A(X)$ über einem kompakten Hausdorff-Raum X findet sich in [3], Proposition 1, woraus Korollar 3.5 für kompaktes X folgt, da $F(X) \to A(X)$ offen, stetig und surjektiv ist.

$f_X(X)^{-1})^n$ abgeschlossen. Für kompaktes X ist das klar. Für beliebiges X folgt das dann aus der Stetigkeit und Injektivität der Abbildung $F(X) \to F(\hat{X})$, die von der Einbettung $X \to \hat{X}$ von X in seine Stone-Čech-Kompaktifizierung $\hat{X}$ induziert wird (s. [13], 3, Theorem 2).

Wenn nun $F(X)$ ein Baire-Raum ist, enthält eine der Mengen X_n eine offene nicht leere Teilmenge. Dann ist X_{2n} eine Umgebung der 1. Nach Lemma 3.3 ist X dann diskret und folglich auch $F(X)$.

Der Beweis des Korollars 3.5 hängt nicht von den Ergebnissen des Paragraphen 2 ab.

Korollar 3.6. *Wenn in einer freien topologischen Gruppe ein Punkt eine abzählbare Umgebungsbasis hat, dann ist sie diskret.*

Beweis. Wenn ein Punkt eine abzählbare Umgebungsbasis hat, dann auch $1 \in F(X)$. Wenn X nicht diskret ist, können wir wegen Lemma 3.3 aus jeder Umgebung U_n einer Umgebungsbasis von 1 ein w_n mit $l(w_n) > n$ auswählen. Die Folge dieser w_n konvergiert dann gegen 1 im Gegensatz zu Korollar 3.2.

4. Beweisidee

Die Idee des Beweises von Satz 2.2 ist folgende: Wenn die Gruppe G auf dem metrischen Raum E mit der Metrik d als Gruppe von Isometrien operiert, dann ist

$$N(g) := d(g(x), x)$$

für jeden Punkt $x \in E$ eine Norm. Denn

$$\begin{aligned} N(g \cdot h^{-1}) &= d(g h^{-1}(x), x) \\ &= d(h^{-1}(x), g^{-1}(x)) \\ &\leqq d(h^{-1}(x), x) + d(g^{-1}(x), x) \\ &= d(x, h(x)) + d(x, g(x)) \\ &= N(h) + N(g). \end{aligned}$$

Wenn G eine hausdorffsche topologische Gruppe ist und die Operation von G auf E stetig ist, dann ist N eine stetige Norm. Jede stetige Abbildung $f: X \to G$ induziert dann einen stetigen Homomorphismus $\bar{f}: F(X) \to G$. Dann ist $N \circ \bar{f}$ eine stetige Norm auf $F(X)$.

Zum Beweis des Satzes 2.2 nehmen wir für E das Innere des Einheitskreises der komplexen Ebene mit der hyperbolischen Metrik d, für G die Liesche Gruppe aller biholomorphen Abbildungen von E auf sich. Wir werden zeigen:

Satz 4.1. *Zu jedem $\varepsilon>0$ und jeder natürlichen Zahl n gibt es n Elemente $g_1, \ldots, g_n \in G$ mit $g_i(0) \neq 0$, so daß für jedes Wort $w \in F(x_i, i=1, \ldots, n)$ gilt*

$$d(0, w(g_i)(0)) > (1-\varepsilon) \cdot l(w) \max_{i=1,\ldots,n} d(0, g_i(0)).$$

Dabei bedeutet $w(g_i)$ das Bild von w unter dem von $x_i \to g_i$ induzierten Homomorphismus $F(x_i;\, i=1, \ldots, n) \to G$.

Aus diesem Satz folgt der Satz 2.2. Denn wenn X vollständig regulär ist, dann gibt es eine stetige Abbildung $f_1\colon X \to [1, n]$ mit $f_1(x_i) = i$ für $i = 1, \ldots, n$. Da die Menge der $g \in G$ mit $d(0, g(0)) \leqq r$ für alle $r>0$ wegzusammenhängend ist, gibt es eine stetige Abbildung $f_2\colon [1, n] \to G$ mit $f_2(i) = g_i$ und

$$\max d(0, f_2(s)(0)) \leqq \max_{i=1,\ldots,n} d(0, g_i(0)) \qquad \text{für } s \in [1, n].$$

Die stetige Abbildung $f := f_2 \circ f_1\colon X \to G$ setzt sich zu einem stetigen Homomorphismus $\bar{f}\colon F(X) \to G$ fort. Die im Satz 2.2 gesuchte Norm ist nun

$$N(w) = \frac{d(0, \bar{f}(w)(0))}{\max\limits_{i=1,\ldots,n} d(0, g_i(0))}.$$

5. Reduktion auf ein Lemma

Das Innere des Einheitskreises der komplexen Ebene

$$E = \{z \in \mathbb{C};\, |z| < 1\}$$

versehen wir mit der hyperbolischen Metrik

$$d(a, b) = \operatorname{arcosh} \frac{|1 - \bar{a} b|}{\sqrt{(1 - a\bar{a})(1 - b\bar{b})}}$$

$$= \frac{1}{2} \log \frac{1 + \left|\dfrac{a-b}{1-\bar{a}b}\right|}{1 - \left|\dfrac{a-b}{1-\bar{a}b}\right|}$$

(vgl. [2], IV, § 5, (5) und (14), [12], § 1). Setzt man zur Abkürzung

$$v(z) = \sqrt{1 - z\bar{z}}$$

und

$$t = \frac{|a-b|}{|1 - \bar{a} b|},$$

so wird der Übergang zwischen beiden Formeln durch

$$d(a,b)=\frac{1}{2}\log\frac{1+t}{1-t}$$
$$=\operatorname{artanh} t$$
$$=\operatorname{arcosh} v(t)$$

und

$$v(t)=\frac{v(a)\cdot v(b)}{|1-\bar{a}\,b|}$$

gegeben.

Sei $a\in E$, $a\neq 0$. Die Menge der Punkte $x\in E$ mit

$$d(x,a)=d(x,0), \tag{5.1}$$

das Mittellot von a und 0, ist der Teil in E des euklidischen Orthokreises mit dem Mittelpunkt $\frac{1}{\bar{a}}$ und dem Radius $r=\frac{1}{a\bar{a}}-1=\frac{v(a)}{|a|}$. Die Menge der Punkte $K_a=\{x\in E;\, d(x,a)<d(x,0)\}$ ist daher der Durchschnitt des Inneren dieses Orthokreises mit E.

Jede biholomorphe Abbildung von E auf sich ist eine Isometrie bezüglich dieser Metrik. Es seien a und b zwei Punkte aus E mit $|a|=|b|\neq 0$. Dann gibt es genau eine biholomorphe Abbildung g mit

$$g(a)=0$$

und

$$g(0)=b,$$

nämlich

$$g(z)=\frac{b}{a}\cdot\frac{z-a}{\bar{a}z-1}.$$

Das Mittellot ∂K_a wird dabei auf das Mittellot ∂K_b abgebildet und es gilt:

$$g(K_a)=C\bar{K}_b$$

und

$$g(C\bar{K}_a)=K_b.$$

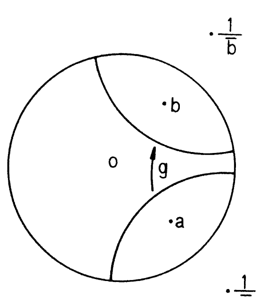


Fig. 1

Wir betrachten von jetzt an nur noch den topologischen Raum E, die abgeschlossene Hülle einer Teilmenge von E ist daher immer in der Relativtopologie von E zu verstehen. Wir betrachten nun die $2n$ Punkte

$$a_j = s \cdot e^{\frac{2\pi i}{2n} \cdot j}, \qquad 0 < s < 1, j = 0, \ldots, 2n-1$$

und die n biholomorphen Abbildungen

$$g_j = e^{\frac{2\pi i}{n}} \frac{z - a_{2j}}{a_{2j} z - 1}, \qquad j = 0, \ldots, n-1$$

mit

$$g_j(a_{2j}) = 0$$

und

$$g_j(0) = a_{2j+1}.$$

Wenn s hinreichend nahe bei 1 liegt, dann sind die Kreisscheiben $K_{a_j} =: K_j$ disjunkt. Wir wollen von jetzt an annehmen, daß die K_j disjunkt sind. Dann gilt

$$g_j(\complement \overline{K_{2j}}) = K_{2j+1}$$

und

$$g_j^{-1}(\complement \overline{K_{2j+1}}) = K_{2j}.$$

Hieraus folgt durch Induktion nach der reduzierten Länge: Wenn w ein reduziertes Wort in den Symbolen g_j ist, das mit g_j beginnt: $w = g_j \ldots$, dann ist $w(0) \in K_{2j+1}$. Wenn w mit g_j^{-1} beginnt, dann ist $w(0) \in K_{2j}$. Insbesondere ist die von den g_j erzeugte Untergruppe von G frei.

Wenn w ein reduziertes Wort in den Symbolen g_j ist, das nicht mit g_j^{-1} beginnt, dann gibt es folgende Möglichkeiten:

$$\begin{aligned} &w = \varnothing, \qquad w(0) = 0, \qquad g_j \cdot w(0) = a_{2j+1}, \\ &d(0, g_j w(0)) - d(0, w(0)) = d(0, a_{2j+1}) = d(0, s), \end{aligned} \tag{5.2}$$

oder $w \neq \varnothing$, $w(0) \in K_i$, $i \neq 2j$. Dann gilt

$$\begin{aligned} d(0, g_j w(0)) - d(0, w(0)) &= d(g_j^{-1}(0), w(0)) - d(0, w(0)) \\ &\geqq \inf_{\substack{z \in K_i \\ i \neq 2j}} \{d(a_{2j}, z) - d(0, z)\}. \end{aligned} \tag{5.3}$$

Wenn w ein reduziertes Wort in den Symbolen g_j ist, das nicht mit g_j beginnt, dann gibt es die analogen Möglichkeiten

$$w = \varnothing: \ d(0, g_j^{-1} w(0)) - d(0, w(0)) = d(0, s); \tag{5.4}$$

$$w \neq \varnothing: \ d(0, g_j^{-1} w(0)) - d(0, w(0)) \geqq \inf_{\substack{z \in K_i \\ i \neq 2j+1}} \{d(a_{2j+1}, z) - d(0, z)\}. \tag{5.5}$$

Wegen der Symmetrie des Problems sind die auftretenden Infima alle gleich und hängen nur von s ab. Wir setzen

$$D(s) := \inf_{\substack{z \in K_i \\ i \neq 0}} \{d(s,z) - d(0,z)\}.$$

Wegen der Dreiecksungleichung ist

$$D(s) \leqq d(0,s). \tag{5.6}$$

Der Satz 4.1 ergibt sich, wenn wir das folgende Lemma bewiesen haben.

Lemma 5.1.

$$\lim_{s \to 1} \frac{D(s)}{d(0,s)} = 1.$$

Aus den Ungleichungen (5.2) bis (5.6) folgt nämlich durch Induktion nach der Länge eines Wortes w in den Symbolen g_j:

$$d(0, w(0)) \geqq l(w) \cdot D(s).$$

Aus dem Lemma 5.1 folgt dann, daß es zu jedem $\varepsilon > 0$ ein $s < 1$ gibt, so daß

$$D(s) \geqq (1-\varepsilon) \cdot d(0,s) = (1-\varepsilon) \cdot d(0, g_j(0))$$

gilt, woraus der Satz 4.1 folgt.

6. Beweis des Lemmas, erster Teil

Es bleibt also das Lemma 5.1 zu beweisen. In dieser Nummer wird $D(s)$ berechnet. In der nächsten Nummer wird der Limes $\frac{D(s)}{d(0,s)}$ für s gegen 1 berechnet. Das sind Aufgaben der hyperbolischen Geometrie und der elementaren Analysis. Die verwendeten Formeln der hyperbolischen Geometrie findet man in [10].

Die Zahl s wird in diesem Paragraphen fest sein, daher oft weggelassen. Wir wollen

$$D(s) = \inf_{\substack{z \in K_i \\ i \neq 0}} \{d(s,z) - d(0,z)\}$$

berechnen. Wegen der Symmetrie des Problems genügt es, $i = 1, \ldots, n$ zu betrachten. Auf den Linien, wo $d(0,z)$ konstant ist – das sind

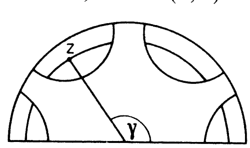


Fig. 2

euklidische Kreise um den Nullpunkt – hängt $d(s, z)$ monoton vom Winkel γ ab, der von den – hyperbolischen und euklidischen – Strecken $0s$ und $0z$ bei 0 eingeschlossen wird; denn nach dem Kosinussatz der hyperbolischen Geometrie gilt

$$\cosh d(z, s) = \cosh d(0, z) \cdot \cosh d(0, s) - \sinh d(0, z) \cdot \sinh d(0, s) \cdot \cos \gamma .$$

Zur Berechnung des Infimums $D(s)$ braucht man daher nur Punkte z aus der K_0 zugewandten Seite von ∂K_1 zu betrachten.

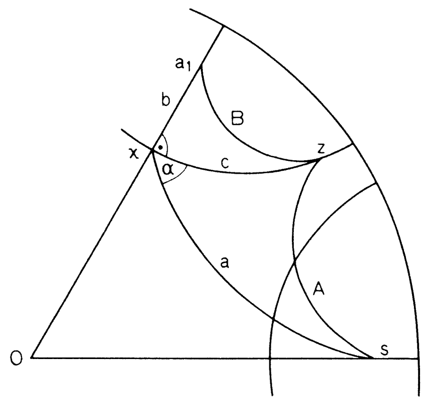


Fig. 3

Zur Abkürzung setzen wir

$$A := d(s, z)$$
$$B := d(0, z) = d(g_0(0), z) \qquad (\text{s. } (5.1)).$$

Es sei x der Schnittpunkt der Geraden $0\,a_1$, $a_1 = g_0(0)$, mit dem Orthokreis ∂K_1. Wir setzen

$$c := d(x, z)$$
$$b := d(x, a_1)$$
$$a := d(x, s).$$

Offenbar hängt z analytisch von c ab und damit auch A und B.

Hilfssatz 6.1. *$A - B$ ist eine monoton fallende Funktion von c.*

Beweis. Nach dem Kosinussatz der hyperbolischen Geometrie gilt

$$\cosh A = \cosh a \cdot \cosh c - \sinh a \cdot \sinh c \cdot \cos \alpha, \tag{6.1}$$

$$\cosh B = \cosh b \cdot \cosh c, \tag{6.2}$$

wobei α der von den hyperbolischen Geraden xz und xs eingeschlossene Winkel ist. Differenzieren wir nach c:

$$\sinh A \cdot A' = \cosh a \cdot \sinh c - \sinh a \cdot \cosh c \cdot \cos\alpha, \tag{6.3}$$

$$\sinh B \cdot B' = \cosh b \cdot \sinh c. \tag{6.4}$$

Bei nochmaligem Differenzieren ergibt sich:

$$(\sinh A \cdot A')' = \cosh A$$

$$(\sinh B \cdot B')' = \cosh B.$$

Wir beweisen die Behauptung $A' - B' \leqq 0$, indem wir zeigen, daß

$$f := (A' - B') \cdot \sinh A \cdot \sinh B \leqq 0.$$

Es gilt

$$\begin{aligned} f' &= \cosh A \cdot \sinh B - \sinh A \cdot A' \cdot \cosh B \cdot B' \\ &\quad - \cosh B \cdot \sinh A - \sinh B \cdot B' \cdot \cosh A \cdot A' \\ &= -\sinh(A-B)(1 - A' \cdot B'), \end{aligned}$$

da $\sinh(A-B) = \sinh A \cdot \cosh B - \cosh A \cdot \sinh B$. Aus der Dreiecksungleichung folgt

$$|A(c + \Delta c) - A(c)| \leqq \Delta c,$$

also $|A'| \leqq 1$. Ebenso erhält man $|B'| \leqq 1$. Der Faktor $1 - A'B'$ ist also nicht negativ. Da $z \notin \bar{K}_0$ ist $d(z,s) > d(z,0)$ und daher $A > B$, folglich $f' \leqq 0$. Wegen (6.3) und (6.4) ist

$$f(0) = -\sinh a \cdot \cos\alpha \cdot \sinh B(0) < 0,$$

also $f(c) < 0$ für $c > 0$ und daher $A' - B' < 0$ für $c \geqq 0$, q.e.d.

Daher gilt für $D(s) = \inf\limits_{c>0} A - B$ das

Korollar 6.2. $D(s) = \lim\limits_{c\to\infty} A - B$.

Das Resultat dieses Paragraphen ist:

Hilfssatz 6.3.

$$\cosh D(s) = \frac{1+d^2}{2d},$$

wo

$$d = \frac{\cosh a - \sinh a \cdot \cos\alpha}{\cosh b}.$$

Beweis. Nach (6.2) strebt mit $c \to +\infty$ auch $\cosh B \to +\infty$ und damit $B \to +\infty$. Wegen (6.1) und (6.2) gilt

$$\lim_{c\to\infty} \frac{\cosh A}{\cosh B} = \lim_{c\to\infty} \frac{\cosh a - \sinh a \cdot \tanh c \cdot \cos\alpha}{\cosh b} = d.$$

Es ist

$$d \geqq 1,$$

da

$$A \geqq B.$$

Deshalb strebt auch $\cosh A$ gegen $+\infty$ für $c \to +\infty$ und damit auch A. Nach der Regel von de l'Hôpital gilt also

$$\begin{aligned}
\cosh D(s) &= \cosh \lim_{c \to \infty} A - B \\
&= \lim_{c \to \infty} \cosh(A-B) = \lim_{c \to \infty} \cosh A \cdot \cosh B - \sin A \cdot \sinh B \\
&= \lim_{c \to \infty} \frac{1 - \tanh A \cdot \tanh B}{\dfrac{1}{\cosh A \cdot \cosh B}} \\
&= \lim_{c \to \infty} \frac{-\dfrac{A'}{\cosh^2 A} \cdot \tanh B - \dfrac{B'}{\cosh^2 B} \cdot \tanh A}{\dfrac{-(\sinh A \cdot \cosh B \cdot A' + \cosh A \cdot \sinh B \cdot B')}{\cosh^2 A \cdot \cosh^2 B}} \\
&= \lim_{c \to \infty} \frac{A' \cdot \tanh B \cdot \cosh^2 B + B' \cdot \tanh A \cdot \cosh^2 A}{A' \cdot \sinh A \cdot \cosh B + B' \cdot \cosh A \cdot \sinh B} \\
&= \lim_{c \to \infty} \frac{A' \cdot \tanh B \cdot \dfrac{\cosh B}{\sinh B} + B' \cdot \tanh A \cdot \dfrac{\cosh A}{\cosh B} \cdot \dfrac{\cosh A}{\sinh B}}{A' \cdot \dfrac{\sinh A}{\sinh B} + B' \cdot \dfrac{\cosh A}{\cosh B}} \\
&= \frac{1+d^2}{2d},
\end{aligned}$$

denn

$$\lim_{c \to \infty} \frac{\cosh A}{\sinh B} = \lim_{c \to \infty} \frac{\cosh A}{\cosh B} \cdot \coth B = d$$

und

$$\lim_{c \to \infty} \frac{\sinh A}{\sinh B} = \lim_{c \to \infty} \frac{\cosh A}{\cosh B} \cdot \coth B \cdot \tanh A = d,$$

und nach (6.3) gilt

$$\begin{aligned}
\lim_{c \to \infty} A' &= \lim_{c \to \infty} \frac{\cosh a \cdot \sinh c - \sinh a \cdot \cosh c \cdot \cos \alpha}{\sinh A} \\
&= \lim_{c \to \infty} \frac{\cosh a \cdot \sinh c - \sinh a \cdot \cosh c \cdot \cos \alpha}{\cosh a \cdot \cosh c - \sinh a \cdot \sinh c \cdot \cos \alpha} \cdot \frac{\cosh A}{\sinh A} = 1
\end{aligned}$$

und

$$\lim_{c\to\infty} B' = \lim_{c\to\infty} \frac{\cosh b \cdot \sinh c}{\sinh B} = \lim_{c\to\infty} \frac{\cosh b \cdot \sinh c}{\cosh b \cdot \cosh c} \cdot \frac{\cosh B}{\sinh B} = 1 .$$

7. Beweis des Lemmas, zweiter Teil

Aus

$$\frac{1}{2}\log\frac{1+s}{1-s} = d(0,s) = d(g_0(0), g_0(s)) = d(g_0(0), 0) = 2b$$

folgt, daß wir die Funktion $b(s)$ umkehren können und s als analytische Funktion von b erhalten. $\lim_{b\to\infty} s = 1$. Der Kosinussatz der hyperbolischen

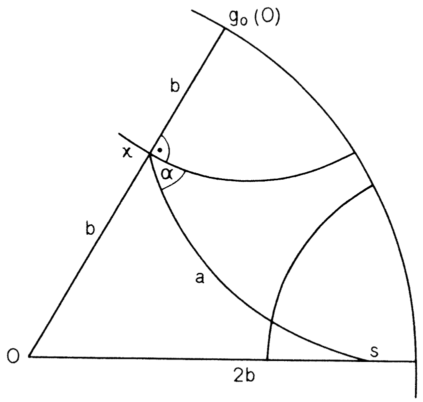


Fig. 4

Geometrie liefert für das Dreieck $x0s$:

$$\cosh a = \cosh b \cdot \cosh 2b - \sinh b \cdot \sinh 2b \cos\phi \tag{7.1}$$

mit $\phi = \frac{2\pi}{2n}$. Der Sinussatz ergibt

$$\frac{\sinh a}{\sin\phi} = \frac{\sinh 2b}{\cos\alpha}.$$

Insbesondere sind a und α analytische Funktionen von b (b hinreichend groß!) und

$$\lim_{b\to\infty} \frac{\cosh a}{\cosh b \cdot \cosh 2b} = 1 - \cos\phi > 0 .$$

Für d erhalten wir

$$d = \frac{\cosh a - \sinh a \cdot \cos\alpha}{\cosh b}$$

$$= \frac{\cosh a - \sinh 2b \cdot \sin\phi}{\cosh b},$$

also $\lim_{b\to\infty} \frac{d}{\cosh 2b} = 1-\cos\phi$. Folglich ist $\lim_{b\to\infty} d=\infty$ und daher auch

$$\lim_{b\to\infty} D = \lim_{b\to\infty} \operatorname{arcosh} \frac{1+d^2}{2d} = \infty .$$

Nach der Regel von de l'Hôpital folgt durch Differentiation nach b

$$\lim_{s\to\infty} \frac{D(s)}{d(0,s)} = \lim_{b\to\infty} \frac{D(b)}{2b}$$

$$= \lim_{b\to\infty} \frac{4d^2-2(1+d^2)}{2\sqrt{\left(\frac{1+d^2}{2d}\right)^2-1}\cdot 4d^2} \cdot d'$$

$$= \lim_{b\to\infty} \frac{2(d^2-1)\cdot d'}{\frac{2(d^2-1)}{2d}\cdot 4d^2} = \lim_{b\to\infty} \frac{d'}{2d}.$$

Aus (7.1) folgt

$$\lim_{b\to\infty} \frac{(\cosh a)'}{\cosh b\cdot\cosh 2b} = 3(1-\phi)$$

und daher

$$\lim_{b\to\infty} \frac{d'}{2d} = \lim_{b\to\infty} \frac{\cosh b((\cosh a)' - 2\sinh 2b\cdot\sin\phi)}{\cosh^2 b}$$

$$= \lim_{b\to\infty} \frac{\{\cosh b((\cosh a)'-2\sinh 2b\cdot\sin\phi) - \sinh b(\cosh a - \sinh 2b\cdot\sin\phi)\}}{2\cosh b(\cosh a - \sinh 2b\cdot\sin\phi)}$$

$$= \lim_{b\to\infty} \frac{\frac{(\cosh a)'}{\cosh a} - 2\sin\phi\frac{\sinh 2b}{\cosh a} - \tanh b + \sin\phi\frac{\sinh 2b}{\cosh b\cdot\cosh a}}{2\left(1-\sin\phi\frac{\sinh 2b}{\cosh b\cdot\cosh a}\right)}$$

$$= \frac{3-0-1+0}{2(1-0)} = 1, \quad \text{q.e.d.}$$

Literatur

1. Abels, H.: Kompakt definierbare topologische Gruppen. Math. Ann. **197**, 221–233 (1972).
2. Behnke, H., Sommer, F.: Theorie der analytischen Funktionen einer komplexen Veränderlichen. Berlin-Göttingen-Heidelberg: Springer 1962.
3. Eisenberg, M.: Embedding a transformation group in an automorphism group. Proc. Amer. Math. Soc. **23**, 276–281 (1969).
4. Gelbaum, B.R.: Free topological groups. Proc. Amer. Math. Soc. **12**, 737–743 (1961).

5. Graev, M.J.: Free topological groups. Izvestija Akad. Nauk. SSSR, Ser. Mat 12, 279–324 (1948), (russ.); engl. Übersetzung: Amer. Math Soc. Translat. Ser 1, 8, 305–364 (1962).
6. Graev, M.J.: Theorie der topologischen Gruppen I. Normen und Metrik auf Gruppen. Vollständige Gruppen. Freie topologische Gruppen. (russ.) Uspehi Mat. Nauk 5, 2(36), 1–56 (1950).
7. Kakutani, S.: Free topological groups and infinite direct products of topological groups. Proc. Imper. Acad. Japan 20, 595–598 (1944).
8. Kelley, J.L.: General topology. Princeton: van Nostrand 1955.
9. Markov, A.A.: On free topological groups. Izvestija Akad. Nauk. SSSR, Ser. Mat. 9, 3–64 (1945) (russ.); engl. Übersetzung: Amer. Math. Soc. Translat. Ser. 1, 8, 195–272 (1962).
10. Meschkowski, H.: Nichteuklidische Geometrie. Braunschweig: Vieweg 1954.
11. Nakayama, T.: Note on free topological groups. Proc. Imper. Acad. Japan 19, 417–475 (1943).
12. Peschl, E., Erwe, F.: Über beschränkte Systeme von Funktionen. Math. Ann. 126, 185–220 (1953).
13. Samuel, P.: On universal mappings and free topological groups. Bull. Amer. Math. Soc. 54, 591–598 (1948).

H. Abels
Fakultät für Mathematik
Universität Bielefeld
D-4800 Bielefeld
Kurt-Schumacher-Str. 6
Bundesrepublik Deutschland

(Eingegangen am 27. Oktober 1971)

Math. Z. 129, 43 – 48 (1972)
© by Springer-Verlag 1972

π-Groups that are M-Groups

Murray Schacher* and Gary M. Seitz**

An M-group is a finite group G such that each irreducible complex character of G is induced from a linear character of a subgroup of G. In this note we prove

Theorem. *There is an infinite set of primes π such that all π-groups are M-groups.*

The existence of an infinite set of primes π such that for each $p \neq q$ in π, p has odd order (mod q) and q has odd order (mod p) is proved in § 1. Combining this with some standard techniques for proving groups are M-groups, we have the above theorem. The class of groups described in the theorem turns out to be a subclass of the class $\mathscr{I}$ of groups G such that $|G|$ is odd and all chief factors of subgroups of G have odd rank. It is proved in §2 that $\mathscr{I}$ is a saturated formation of M-groups.

§ 1. A Set of Primes π

Our objective is to prove that there is an infinite set of primes S satisfying: $p, q \in S \Rightarrow$ order of p (mod q) is odd. We will construct the set S inductively. Suppose then $S_r = \{p_1, \ldots, p_r\}$ is a set of primes satisfying: $p_i \equiv 5 \pmod 8$ $i = 1, \ldots, r$ and the order of $p_i \pmod{p_j}$ is odd for $i, j = 1, \ldots, r$.

Suppose p_{r+1} is a prime satisfying:

(1) p_i is a 4-th power (mod p_{r+1}) for $i = 1, \ldots, r$.

(2) $p_{r+1} \equiv 1 \pmod{4 p_1 \ldots p_r}$

(3) $p_{r+1} \not\equiv 1 \pmod 8$.

Considering the set $S_{r+1} = \{p_1, \ldots, p_{r+1}\}$ we have: the order of $p_{r+1} \pmod{p_i}$ is 1 for $i = 1, \ldots, r$ by (2), while for $i \leqq r$ the order of p_i (mod p_{r+1}) divides $\frac{p_{r+1}-1}{4}$ by (1), and this is odd by (3). Hence the order of p_i (mod p_j) is odd for $i, j = 1, \ldots, r+1$, and the set S_{r+1} is an

* Research supported in part by NSF Grant GP 28696.
** Research supported in part by NSF Grants GP 29041X and GP 20308.

enlargement of our original set S_r. Continuing inductively, or letting r go to infinity, we could construct the infinite set S. Therefore it is sufficient to prove that for any r there is a prime p_{r+1} satisfying conditions (1), (2), and (3).

For notational purposes we write $\sqrt[4]{p_i}$ ($i=1, \ldots, r$) for the real 4-th root of p_i and $Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_r})$ for the real field obtained by adjoining these elements to the rational field Q. If n is a positive integer we denote by ζ_n a primitive n-th root of unity over Q. In what follows we use freely the theory of valuations over algebraic number fields.

Lemma 1. *If p is a rational prime, then p_i is a 4-th power* (mod p) *for $i=1, \ldots, r$ if p splits completely in $Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_r})=L$.*

Proof. If p splits completely in L, then the polynomials x^4-p_i have 4 distinct roots mod p for $i=1, \ldots, r$.

Since infinitely many primes split completely in an algebraic number field, Lemma 1 insures that infinitely many primes satisfy condition (1).

Lemma 2. *If a prime p splits completely in $M=Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_r}, \zeta_{4p_1 \ldots p_r})$ then:*

(a) $p \equiv 1 \pmod{4p_1 \ldots p_r}$,

(b) *p splits completely in L.*

Proof. This follows from Lemma 1 and 7-2-5 of [7].

Lemma 2 guarantees an infinite set of primes satisfying (1) and (2).

Let T be the set of rational primes which split completely in M. We are finished if there is some element $p \in T$ so that $p \not\equiv 1 \pmod 8$. Assume then that all primes in T are 1 (mod 8); we show this leads to a contradiction. The primes that split completely in M are then the same as the primes that split completely in $M(\zeta_8)$; by Bauer's theorem [2, Theorem 9-1-3] $M(\zeta_8)=M$. But $M(\zeta_8)=M(\sqrt{2}, i)=M(\sqrt{2})$ where $i^2=-i$. Thus $M(\zeta_8)=M \Leftrightarrow \sqrt{2} \in M$. We establish our contradiction by showing $\sqrt{2} \notin M$.

To begin with we show

$$\sqrt{2} \notin Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_j}) \quad \text{for } j=1, \ldots, r.$$

Clearly

$$\sqrt{2} \notin Q(\sqrt[4]{p_1}), \quad \text{for} \quad \sqrt{2} \in Q(\sqrt[4]{p_1}) \Rightarrow Q(\sqrt[4]{p_1}) = Q(\sqrt{p_1}, \sqrt{2}) \Rightarrow Q(\sqrt[4]{p_1})$$

is normal over Q, a contradiction since $i \in$ the splitting field of x^4-p_1 over Q and $Q(\sqrt[4]{p_1})$ is real. Assume we have proved

$$\sqrt{2} \notin Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_j}) = K_j.$$

If

$$\sqrt{2} \in Q(\sqrt[4]{p_1}, \ldots, \sqrt[4]{p_{j+1}}) = K_{j+1},$$

then

$$K_{j+1}=K_j\,(\sqrt{2},\sqrt{p_{j+1}})\Rightarrow K_{j+1}$$

is normal over K_j. This is a contradiction as before.

For $j=r$, this establishes that

$$\sqrt{2}\notin Q(\sqrt[4]{p_1},\ldots,\sqrt[4]{p_r})=L.$$

Since L is a real field it is clear that $\sqrt{-2}\notin L$. It follows now that

$$\sqrt{2}\notin Q(\sqrt[4]{p_1},\ldots,\sqrt[4]{p_r},\zeta_{p_1\ldots p_r})=L(\zeta_{p_1,\ldots,p_r})=M_0,$$

since $L(\zeta_{p_1\ldots p_r})$ is not ramified at primes of L extending 2 by [7, 7-2-5]. For the same reason $\sqrt{-2}\notin M_0$, and $i\notin M_0$. We show finally that

$$\sqrt{2}\notin M=M_0(i)=Q(\sqrt[4]{p_1},\ldots,\sqrt[4]{p_r},\zeta_{4p_1\ldots p_r}).$$

If $\sqrt{2}\in M$, then $(a+b\,i)^2=2$ with $a,b\in M_0$. Then: $(a^2-b^2)+2ab\,i=2\Rightarrow a=0$ or $b=0$ as $1, i$ is a basis of M/M_0. If $b=0$, then $a^2=2\Rightarrow\sqrt{2}\in M_0$, a contradiction. If $a=0$, $-b^2=2\Rightarrow\sqrt{-2}\in M_0$, again a contradiction. This establishes:

Theorem 3. *There is an infinite set of primes π satisfying: $p,q\in\pi\Rightarrow$ order p* (mod q) *is odd. The element of π can be chosen all congruent to* 5 (mod 8).

Proof. We have proved that there are solutions p_{r+1} satisfying (1), (2), and (3). As $p_{r+1}\equiv 1 \pmod 4$ and $p_{r+1}\not\equiv 1 \pmod 8$, we conclude $p_{r+1}\equiv 5 \pmod 8$.

§2. A Formation of M-Groups

In this section we derive results about M-groups. If G is a solvable group then chief factors of G may be viewed as vector spaces and it makes sense to speak of the rank of a chief factor. Let $\mathcal{I}$ be the largest class of (solvable) groups of odd order G such that all chief factors of subgroups of G have odd rank.

Lemma 4. *Let G be a solvable group of odd order operating irreducibly on a module V over a finite field K of characteristic p. Suppose that for each prime divisor $q\neq p$ of $|G|$ that p has odd order* (mod q). *Then V has odd dimension.*

Proof. We proceed by induction on $|G|$. We may assume that G is faithfully represented on V. Write $|G|=p^a p_1^{a_1}\ldots p_k^{a_k}$ where the primes $p, p_1,\ldots,p_k$ are distinct. From the hypothesis it easily follows that p

has odd order (mod $p_i^{a_i}$) for each i, say d_i. Let $d=\prod_i d_i$, so that d is odd and $|x|\,|p^d-1$ for each p'-element x of G. Write $K=\mathbb{F}_{p^b}$ and set $L=\mathbb{F}_{p^{bd}}$. Then $\mathbb{F}_{p^d}\leqq L$ so that L is a splitting field for the modular representations of G in characteristic p ([1], (41.1) and (83.7)).

We have $L\otimes V=V_1\oplus\cdots\oplus V_k$ where the V_i are all absolutely irreducible and conjugate under the action of $G(L/K)$ ([1], (70.15) and (70.23)). Since $|L:K|=d$ and d is odd we may assume that $k=1$. That is we may assume that V is absolutely irreducible. Let M be a maximal normal subgroup of G and apply Clifford's theorem to M. Since $|G:M|=q$ with q a prime the restriction of V to M is either the direct sum of q non-isomorphic irreducible M-modules or there is just one homogeneous component. In the first case we are done by induction since q is odd. Suppose then that the restriction of V to M contains just one homogeneous component. Then $V_M=V_1\oplus\cdots\oplus V_t$ with the V_i all isomorphic, absolutely irreducible, and of odd dimension. It remains to show that t is odd. Let $\bar{V}=\bar{L}\otimes V$ where $\bar{L}$ is an algebraic closure of L. Then $\bar{V}_M=\bar{V}_1\oplus\cdots\oplus\bar{V}_t$ where $\bar{V}_i=\bar{L}\otimes V_i$ and the $\bar{V}_i$ are still isomorphic and irreducible. By 17.5 of [3] $\bar{V}\cong R_1\otimes R_2$ where R_1 and R_2 are irreducible projective representations of G, $\dim R_2=\dim\bar{V}_1=\dim V_1$, and R_1 is a projective representation of G/M. As G/M is cyclic, R_1 is linear and $t=1$. This completes the proof of Lemma 4.

Corollary 5. *Let π be a set of odd primes such that for each $p\neq q$ in π, p has odd order* (mod q) *and q has odd order* (mod p). *Then $\mathscr{I}$ contains the class of all π-groups.*

Proof. Apply Lemma 4 to the chief factors of a π-group.

Proposition 6. *$\mathscr{I}$ is a saturated formation.*

Proof. Clearly $\mathscr{I}$ is closed under homomorphic images. Suppose $G/H, G/K\in\mathscr{I}$. As $H/H\cap K$ is G-isomorphic to HK/K it is clear that $G/H\cap K$ has all chief factors of odd rank. Let $H\cap K\leqq L\leqq G$. Since $LH/H\cong L/L\cap H$ and $LK/K\cong L/L\cap K$, both $L/L\cap H$ and $L/L\cap K$ have chief factors with odd rank and consequently $L/H\cap K$ has chief factors with odd rank. We have shown that $\mathscr{I}$ is a formation.

To show $\mathscr{I}$ is saturated it suffices to show that if N is the unique minimal normal subgroup of G, $G/N\in\mathscr{I}$, and $N\leqq\Phi(G)$, then $G\in\mathscr{I}$. Let $|N|=p^a$ for p a prime. Since $N\leqq\Phi(G)$, the Frattini argument implies that $O_{p'}(G/N)=1$. Then $G/O_p(G)$ is represented faithfully on $O_p(G)/N\Phi(O_p(G))$. Let $x\in G$ have prime order $q\neq p$. As $G/N\in\mathscr{I}$, $O_p(G)\langle x\rangle/N$ has all chief factors of odd rank and it follows that p has odd order (mod q). Let $H\leqq G$. Since $G/N\in\mathscr{I}$ all chief factors of $HN/N\cong H/H\cap N$

have odd rank. Using the above and Lemma 4 we also get the H chief factors in $H \cap N$ to be of odd rank. Thus $G \in \mathscr{I}$ and the proof is complete.

Proposition 7. *If $G \in \mathscr{I}$, then G is an M-group.*

Proof. Let G be a minimal counterexample. Then G has a faithful irreducible representation T which is not monomial. Since $\mathscr{I}$ is subgroup closed, G is primitive and the structure of Fit(G) is known (Rigby [5]). If P is p-Sylow in Fit(G) then $P = Z(P) P_0$ where $Z(P)$ is cyclic and P_0 is either trivial or extraspecial. Moreover not all Sylow subgroups of Fit(G) are central so we may choose $P_0 \neq 1$. Let $L/Z(P)$ be a chief factor of G with $L \leqq P$. Then $L = Z(P)(P_0 \cap L)$ and $L/Z(P)$ has odd rank. Then $P_0 \cap L$ is not extraspecial and since $L/Z(P)$ is chief, $(P_0 \cap L)' = 1$ and L is abelian. As G is primitive $L \leqq Z(P)$ a contradiction.

We remark that essentially the same proof shows that if $N \leqq G$ with $G/N \in \mathscr{I}$ and N an A-group, then G is an M-group.

Corollary 8. *There is an infinite set of primes π such that all π-groups are M-groups.*

Proof. This is immediate from Theorem 1, Corollary 5, and Proposition 7.

Proposition 9. *Let $\mathscr{H}$ be a class of M-groups of odd order such that $\mathscr{H}$ is closed under homomorphic images, subgroups, and such that if $|G|$ is odd and $G/Z(G) \in \mathscr{H}$, then $G \in \mathscr{H}$. Then $\mathscr{H} \leqq \mathscr{I}$.*

Proof. Let G be a minimal counterexample. Then $G \in \mathscr{H}$, $G \notin \mathscr{I}$, there is a unique minimal normal subgroup N of G, and $G/N \in \mathscr{I}$. Let D be an $\mathscr{I}$-normalizer of G, so that $ND = G$, $N \cap D = 1$. Thus $N = C_G(N)$ and if $|N| = p^a$, then $O_p(D) = 1$. Since $G \notin \mathscr{I}$ and all proper subgroups of G are in $\mathscr{I}$, N must have even rank.

Let $x \in D$ have prime order $q \neq p$. If $N\langle x \rangle < G$, then $N\langle x \rangle$ has chief factors of odd rank and it follows that p has odd order (mod q). Lemma 4 implies that $G = N\langle x \rangle$ for some such x. If $|N| = p^{2l}$ then x fixes a symplectic form on N and thus there is an extraspecial group L of order p^{2l+1} such that x acts on L, centralizes $Z(L)$, and acts irreducibly on $L/Z(L)$. Let $\bar{G} = L\langle x \rangle$. Then $\bar{G}/Z(G) \cong G \in \mathscr{H}$, so that $\bar{G} \in \mathscr{H}$. But $\bar{G}$ has a character of degree p^l and no subgroup of index p^l. Thus $\bar{G}$ is not an M-group, and this is a contradiction.

We remark that Proposition 8 shows that $\mathscr{I}$ is the unique largest subgroup closed saturated formation consisting of M-groups of odd order. Also $\mathscr{I}$ is a Fitting class. If $\mathscr{C}$ is the class of M-groups described in §2 of [6], then $\mathscr{I}$ is precisely the class of groups in $\mathscr{C}$ having odd order. For further results along these lines the reader is referred to [4].

References

1. Curtis, C. W., Reiner, J.: Representation theory of finite groups and associative algebras. New York: Interscience 1962.
2. Goldstein, L. J.: Analytic number theory. New Jersey: Prentice-Hall 1971.
3. Huppert, B.: Endliche Gruppen I. Berlin-Heidelberg-New York: Springer 1967.
4. Price, D. J.: A generalization of M-groups. Thesis, University of Chicago, 1971.
5. Rigby, J. F.: Primitive linear groups containing a normal nilpotent subgroup larger than the centre of the group. J. London Math. Soc. **35**, 389–400 (1960).
6. Seitz, G. M.: M-groups and the supersolvable residual. Math. Z. **110**, 101–122 (1969).
7. Weiss, E.: Algebraic number theory. New York: McGraw-Hill 1963.

Prof. M. Schacher
Department of Mathematics
University of California
Los Angeles, California 90024
USA

Prof. Gary M. Seitz
Department of Mathematics
University of Oregon
Eugene, Oregon 97403
USA

(Received November 15, 1971)

Math. Z. 129, 49–53 (1972)
© by Springer-Verlag 1972

Cross-Ratios and Projectivities of a Line

Nirmala B. Limaye

Introduction

In [3] a generalisation to local rings of the classical von Staudt's theorem was proved. In this note we extend this result to a larger class of rings.

Let A be a commutative ring with 1 and M a free A-module. Let $\mathbb{P}(M)$ be the set of all projective direct summands of rank 1 of M. If spec A, the set of all prime ideals of A, is connected in the Zariski topology, we show in § 1 that the cross-ratio of any four collinear points in $\mathbb{P}(M)$ can be defined. Thus we extend the notion of the cross-ratio introduced in [3]. In § 2, we prove that if A is a semi-local ring satisfying certain conditions and if M is a free A-module of rank 2, then every projectivity of $\mathbb{P}(M)$ is given by a semi-linear map.

§ 1. Cross-Ratios

Let A be a commutative ring with 1. Let M be a free A-module. Let $\mathbb{P}(M)$ be the set of all projective direct summands of rank 1 of M. $\mathbb{P}(M)$ will be called the *projective space* associated to M. The elements of $\mathbb{P}(M)$ will be called the *points of the projective space* associated to M.

Definition. Points $P_1, \dots, P_n$ in $\mathbb{P}(M)$ are said to be *collinear* if there exists a projective direct summand L of rank 2 of M such that for $1 \leqq i$, $j \leqq n$ and $i \neq j$, $P_i + P_j = L$.

Proposition. *Let A be a commutative ring with* 1 *such that* spec A *is connected in the Zarisky topology. Given any four collinear points P_1, P_2, P_3, P_4 in $\mathbb{P}(M)$ there exists a unique element x in A such that for every prime ideal p in* spec A, *there exist unimodular elements e_1 and e_2 in $S_p^{-1} M$ such that $S_p^{-1} P_1 = A_p e_1$, $S_p^{-1} P_2 = A_p e_2$, $S_p^{-1} P_3 = A_p(e_1 + e_2)$ and $S_p^{-1} P_4 = A_p(e_1 + x_p e_2)$, where $A_p = S_p^{-1} A$ and x_p is the image of x under the natural homomorphism from A to A_p.*

Proof. First note that since A_p is a local ring, every element in $\mathbb{P}(S_p^{-1} M)$ is a free A_p-module. Now consider the four points $S_p^{-1} P_i$, $1 \leqq i \leqq 4$, in $\mathbb{P}(S_p^{-1} M)$. Clearly these are collinear and hence, as in [3], their cross-ratio is well defined and is a unit in A_p. Denote it by x_p. We can write $S_p^{-1} P_i$, $1 \leqq i \leqq 4$, as $A_p e_1$, $A_p e_2$, $A_p(e_1 + e_2)$, $A_p(e_1 + x_p e_2)$. If p and p_1

are any two prime ideals in spec A such that $p_1 \subset p$ then we have a natural homomorphism $A_p \to A_{p_1}$ and $S_{p_1}^{-1} P_1, S_{p_1}^{-1} P_2, S_{p_1}^{-1} P_3, S_{p_1}^{-1} P_4$ are isomorphic to $A_p e_1 \underset{A_p}{\otimes} A_{p_1}$, $A_p e_2 \underset{A_p}{\otimes} A_{p_1}$, $A_p(e_1+e_2) \underset{A_p}{\otimes} A_{p_1}$, $A_p(e_1+x_p e_2) \underset{A_p}{\otimes} A_{p_1}$ respectively. Hence $x_{p_1} = x_p$ in A_{p_1}.

Now let p_1 and p_2 be any two prime ideals in Spec A. Since Spec A is connected there exist sequences $q_0, \ldots, q_n$ of prime ideals and $m_1, \ldots, m_n$ of maximal ideals of A such that $p_1 = q_0$, $p_2 = q_n$, while q_i and q_{i+1} are contained in m_{i+1}, $0 \leqq i \leqq n-1$. But then $x_{q_i} = x_{m_{i+1}} = x_{q_{i+1}}$, $0 \leqq i \leqq n-1$; that is, $x_{q_0} = x_{q_n}$. Hence $x_{p_1} = x_{p_2}$ for any p_1 and p_2 in Spec A.

Thus, given any four collinear points P_1, P_2, P_3, P_4, we get an element which is a unit in A_p for every p in Spec A and hence it is a unit in A. This element serves the purpose.

Definition. Given any four collinear points P_1, P_2, P_3, P_4 in $\mathbb{P}(M)$, the *unique element* in A given by the above proposition will be called the cross-ratio of P_1, P_2, P_3, P_4.

§ 2. A Generalisation of the Von Staudt's Theorem

Let A be a semi-local ring, and $m_1, \ldots, m_n$ be its maximal ideals. We denote the natural homomorphism $A \to \prod_{1 \leqq i \leqq n} A/m_i$ by n. By the Chinese Remainder Theorem, n is onto. For an element a in A, we denote by a^i, $1 \leqq i \leqq n$, the i-th component of $n(a)$ in $\prod_{1 \leqq i \leqq n} A/m_i$. Note that a is a unit in A if and only if $n(a)$ is a unit in $\prod_{1 \leqq i \leqq n} A/m_i$; that is, a is a unit in A if and only if $a^i \neq 0$ in A/m_i for each i, $1 \leqq i \leqq n$. For a semi-local ring A and a A-free module M, $\mathbb{P}(M)$ is just the set $P(M)$ of all free direct summands of rank 1 of M. Hence we can use the properties of cross-ratios given in [3]. Any four collinear points in $\mathbb{P}(M)$ are said to be harmonic if their cross-ratio is -1. As in [3], we define a projectivity from $\mathbb{P}(M)$ to $\mathbb{P}(N)$, where M and N are free modules of rank 2 over commutative rings A and B respectively, to be a bijective map $\alpha: \mathbb{P}(M) \to \mathbb{P}(N)$ such that, any four points P_1, P_2, P_3, P_4, in $\mathbb{P}(M)$ are harmonic if and only if $\alpha(P_1)$, $\alpha(P_2)$, $\alpha(P_3)$, $\alpha(P_4)$ are harmonic.

Theorem. *Let M and N befree modules of rank 2 over semi-local rings A and B respectively. Let A satisfy the following conditions.* (I) *The field A/m has a least 7 distinct elements for every maximal ideal m of A.* (II) *If the number of distinct maximal ideals of A is n, then $2, 3, \ldots, n+1$ are invertible in A. Then given a projectivity α: $\mathbb{P}(M) \to \mathbb{P}(N)$, there exists an isomorphism σ: $A \to B$ and a σ-semilinear isomorphism f: $M \to N$ such that $\alpha(P) = f(P)$ for every P in $\mathbb{P}(M)$; that is, $\alpha = P(f)$. If σ_i: $A \to B$, $i = 1, 2$, are isomorphisms and f_i: $A \to B$ are σ_i-semilinear isomorphisms such that $P(f_1) = P(f_2)$, then there exists a unit u in A such that $f_1 = f_2 \cdot u$ and $\sigma_1 = \sigma_2$.*

Proof. If (e_1, e_2) is any basis of M, then as in the theorem of [3] we get a bijective map $\sigma\colon A \to B$ such that $\sigma(0)=0, \sigma(1)=1$ and

$$\alpha A(e_1 + a\, e_2) = B(f_1 + \sigma(a) f_2)$$

for every a in A, where (f_1, f_2) is a basis of N. We also have

$$\sigma(a+b) = \sigma(a) + \sigma(b)$$

for all a, b in A such that $a+b$ and $a-b$ are units, as in the lemma of [3].

Let x and y be any two elements of A such that either $x+y$ or $x-y$ is not a unit. Because of the condition (I), we can choose an element $(z_1, \ldots, z_n)$ in $\prod_{1 \leqq i \leqq n} A/\mathfrak{m}_i$ such that $z_i \neq \pm y^i, \pm(x+y)^i, (x-y)^i$, for each i, $1 \leqq i \leqq n$. Since ν is onto, there exists an a in A such that $\nu(a)=(z_1, \ldots, z_n)$; that is, $a^i = z_i$, $1 \leqq i \leqq n$. Since $\nu(x+y\pm a), \nu(x-y-a), \nu(y \pm a)$ are units in $\prod_{1 \leqq i \leqq n} A/\mathfrak{m}_i$, $x+y\pm a$, $x-y-a$, $y\pm a$ are units in A. Now we have

$$\sigma(x+y)+\sigma(a) = \sigma(x+y+a) = \sigma(x)+\sigma(y+a) = \sigma(x)+\sigma(y)+\sigma(a).$$

Hence $\sigma(x+y)=\sigma(x)+\sigma(y)$ for all x, y in A.

We can similarly define a map $\tau\colon A \to B$ such that $\tau(0)=0$, $\tau(1)=1$, $\alpha A(a\, e_1 + e_2) = B(\tau(a) f_1 + f_2)$, for every a in A and show that τ is additive. That $\sigma(a)=\tau(a)$ for any a in A such that $1+a$ and $1-a$ are units follows as in [3].

If a in A is such that either $a+1$ or $a-1$ is not a unit, then as before we choose $(z_1, \ldots, z_n)$ in $\prod_{1 \leqq i \leqq n} A/\mathfrak{m}_i$ such that $z_i \neq \pm 1, (1-a)^i, -(1+a)^i$, $1 \leqq i \leqq n$. But then there exists an element x in A such that $\nu(x)=(z_1, \ldots, z_n)$ and $(a+x\pm 1)$, $(x \pm 1)$ are units in A. Hence $\sigma(a+x)=\tau(a+x)$; that is $\sigma(a)+\sigma(x)=\tau(a)+\tau(x)$. This, with $\sigma(x)=\tau(x)$, gives $\sigma(a)=\tau(a)$ for all a in A.

For a unit u in A, we have $A(e_1 + u\, e_2) = A(u^{-1} e_1 + e_2)$. Hence $B(f_1 + \sigma(u) f_2) = B(\sigma(u^{-1}) f_1 + f_2)$; that is, $\sigma(u^{-1}) = \sigma(u)^{-1}$ for any unit u in A.

Let now a be any unit in A. Consider $\nu(a)=(a^1, \ldots, a^n)$. Let $a^{i_1}, \ldots, a^{i_r}$ belong to the prime fields of $A/\mathfrak{m}_{i_1}, \ldots, A/\mathfrak{m}_{i_r}$, where $0 \leqq r \leqq n$. Because of the condition (II), we can choose an integer m such that m is invertible in A and $m \neq a^i$, $1 \leqq i \leqq n$; that is, such that m and $m-a$ are units in A. Now using the identity $a^2 = m\left[a - (a^{-1} + (m-a)^{-1})^{-1}\right]$ and the fact that σ is additive we get

$$\sigma(a^2) = m\left[\sigma(a) + (\sigma(a)^{-1} + (m - \sigma(a)^{-1})^{-1}\right];$$

that is $\sigma(a^2) = [\sigma(a)]^2$ for any unit a in A. If a is not a unit then we choose an integer m such that $a-m$ is a unit in A. Hence $\sigma((a-m)^2) = [\sigma(a)-m]^2$, which gives $\sigma(a^2)=[\sigma(a)]^2$ for all a in A. Now for any a, b in A using the above fact and the identity $ab=\frac{1}{2}[(a+b)^2-a^2-b^2]$ we get $\sigma(ab)=\sigma(a)\sigma(b)$. Thus σ is an isomorphism of rings.

We define $f\colon M\to N$ by $f(a_1 e_1+a_2 e_2)=\sigma(a_1)f_1+\sigma(a_2)f_2$. Clearly f is a σ-semilinear isomorphism.

Let $a_1 e_1+a_2 e_2$ be a unimodular element in M such that a_1 is a unit in A. Then we have

$$\begin{aligned}\alpha A(a_1 e_1+a_2 e_2)&=\alpha A(e_1+a_1^{-1} a_2 e_2)=B(f_1+\sigma(a_1)^{-1}\sigma(a_2)f_2)\\&=B(\sigma(a_1)f_1+\sigma(a_2)f_2).\end{aligned}$$

Similarly if a_2 is a unit in A then

$$\alpha A(a_1 e_1+a_2 e_2)=B(\sigma(a_1)f_1+\sigma(a_2)f_2).$$

Let now $(a_1 e_1+a_2 e_2)$ be a unimodular element in M such that a_1 and a_2 are non-units. Since a_1 and a_2 generate the ring A, and A is semi-local, there exists an element x in A such that $a_1 x+a_2=u$ for some unit u in A. Choose an element $(z_1,\dots,z_n)$ in $\prod_{1\leqq i\leqq n} A/\mathfrak{m}_i$ such that $z_i\neq 0,\ \pm a_1^i,\ 1\leqq i\leqq n$. Since ν is surjective there exists a unit u' in A such that $\nu(u')=(z_1,\dots,z_n)$. Now $a_1\pm u'$ are units in A. Since

$$\begin{vmatrix}a_1+u' & a_2-u'x\\ a_1-u' & a_2+u'x\end{vmatrix}=2a_1 x u'+2a_2 u'=2uu',$$

which is a unit in A, the unimodular elements $(a_1+u')e_1+(a_2-u'x)e_2$ and $(a_1-u')e_1+(a_2+u'x)e_2$ form a basis of M. Now consider the points

$$A[(a_1+u')e_1+(a_2-u'x)e_2],\quad A[(a_1-u')e_1+(a_2+u'x)e_2],$$
$$A(a_1 e_1+a_2 e_2),\quad A(u'e_1-u'x_2).$$

Since they are harmonic, their images

$$B[(\sigma(a_1)+\sigma(u'))f_1+(\sigma(a_2)-\sigma(u')\sigma(x))f_2],$$
$$B[(\sigma(a_1)-\sigma(u'))f_1+(\sigma(a_2)+\sigma(u')\sigma(x))f_2],\quad \alpha A(a_1 e_1+a_2 e_2),$$
$$B[\sigma(u')f_1-\sigma(u')\sigma(x)f_2]$$

are harmonic. It can be easily seen that given any three collinear points Q_1, Q_2, Q_4 a point Q_3 is uniquely determined such that Q_1, Q_2, Q_3, Q_4 are harmonic. Hence $\alpha A(a_1 e_1+a_2 e_2)=B(\sigma(a_1)f_1+\sigma(a_2)f_2)$. Thus $\alpha=P(f)$.

The proof of the second part of the theorem is the same as in the classical case which can be found for instance in [1].

Remark. The conditions (I) and (II) of the theorem are not necessary. For example, if $A = B = \mathbb{Z}_3 \times \mathbb{Z}_3$ then one can easily verify that the only isomorphisms from A to B are the identity and σ, where $\sigma(a, b) = (b, a)$; and the only projectivities are those which are given by semi-linear isomorphisms. Thus the theorem holds in this case.

References

1. Artin, E.: Geometric algebra. New York: Interscience 1957.
2. Ojanguren, M., Sridharan, R.: A note on the fundamental theorem of projective geometry. Commentarii Math. Helvet. **44**, 310–315 (1969).
3. Limaye, N. B.: Projectivities over local rings. Math. Z. **121**, 175–180 (1971).

N. B. Limaye
School of Mathematics
Tata Institute of Fundamental Research
Coloba, Bombay 5
India

(Received June 7, 1972)

Math. Z. 129, 55 (1972)
© by Springer-Verlag 1972

On Normal Subgroups of p-Solvable Groups

Morton E. Harris

Theorem. *Let G be a finite p-solvable group, let S be a Sylow p-subgroup of S and let T be a subgroup of S. Then the following two conditions are equivalent:*

(i) *G contains a normal subgroup N such that T is a Sylow p-subgroup of N;*

(ii) *T is a strongly closed subgroup of S with respect to G (i.e., no element of T is conjugate in G to an element of $S-T$).*

Proof. Clearly (i) implies (ii) in any group. In order to prove (ii) implies (i), assume (ii) and proceed by induction on $|G|$.

Suppose that $O_{p'}(G) \neq \{1\}$. Setting $\bar{G}=G/O_{p'}(G)$, we have $|\bar{G}|<|G|$, $\bar{G}$ is p-solvable, $\bar{S}$ is a Sylow p-subgroup of $\bar{G}$ and $\bar{T}$ is a strongly closed subgroup of $\bar{S}$ with respect to $\bar{G}$. Since the theorem holds for $\bar{G}$, it follows that G satisfies (i).

Suppose that $O_{p'}(G)=\{1\}$ and set $P=O_p(G)$. Then $\{1\} \neq P \trianglelefteq G$ and $Z(S) \leqq P$ by a well known lemma of P. Hall and G. Higman. Since T is strongly closed in S with respect to G and $P \leqq S$, we have $T \trianglelefteq S$ and $P \cap T \trianglelefteq G$. Whence $T \cap Z(S) \neq \{1\}$, so that $P \cap T \neq \{1\}$. Setting $\bar{G}=G/P \cap T$, we conclude as above that G satisfies (i) and the theorem follows.

Finally, note that (i) implies (ii) for arbitrary groups G but (ii) does not in general imply (i). For example, if $G=Sz(q)$ with $q=2^{2n+1}>2$ (a Suzuki simple group) and if S is a Sylow 2-subgroup of G, then $T=\Omega_1(S)$ is a strongly closed proper subgroup of S while G is simple.

Dr. Morton E. Harris
Mathematics Department
The University of Illinois
at Chicago Circle
Chicago, Illinois 60680, USA

(Received July 31, 1972)

Math. Z. 129, 57–63 (1972)
© by Springer-Verlag 1972

Kompakt erzeugte Räume und Limesräume

Alfred Frölicher *

Einleitung

Verschiedene Probleme führten in den letzten Jahren dazu, die mengentheoretische Topologie durch das Studium der Limesräume (cf. [8]) einerseits, der kompakt erzeugten Räume (cf. [14, 12, 18]) andererseits zu erweitern. Der Hauptvorteil dieser Räume besteht darin, daß sie, mit den stetigen Abbildungen als Morphismen, Kategorien $\mathcal{L}\mathcal{R}$ bzw. $\mathcal{K}\mathcal{E}$ bilden, welche kartesisch abgeschlossen (cf. [15]) sind, was für die Kategorie der topologischen oder der hausdorffschen Räume nicht zutrifft. Unter den Gebieten, auf denen dadurch bedeutende Fortschritte ermöglicht wurden, seien erwähnt: die Differentialrechnung für Vektorräume ohne Norm (cf. [1, 2, 13, 10, 17]); die Dualitätstheorie (cf. [5, 11, 6]); die Theorie der Funktionenalgebren (cf. [3, 4, 6, 7, 9, 16]).

In dieser Arbeit wird gezeigt, daß die Verwendung kompakt erzeugter Räume äquivalent ist zur Verwendung gewisser Limesräume. Wir ordnen jedem kompakt erzeugten Raum X einen im allgemeinen nicht topologischen Limesraum lX zu und zeigen, daß man aus lX den Raum X dadurch zurückerhält, daß man die Limitierung von lX durch die zugehörige Topologie (cf. [8]) ersetzt. Man erhält so einen Funktor $l\colon \mathcal{K}\mathcal{E} \to \mathcal{L}\mathcal{R}$, durch den $\mathcal{K}\mathcal{E}$ isomorph auf eine volle Unterkategorie $\mathcal{L}\mathcal{R}\ell^0$ von $\mathcal{L}\mathcal{R}$ abgebildet wird.

Als isomorphe Kategorien haben $\mathcal{K}\mathcal{E}$ und $\mathcal{L}\mathcal{R}\ell^0$ die gleichen kategorischen Eigenschaften: ebenso wie $\mathcal{L}\mathcal{R}$ sind sie vollständig, covollständig und kartesisch abgeschlossen. Das gleiche wird auch nachgewiesen für die volle Unterkategorie $\mathcal{L}\mathcal{R}\ell$ von $\mathcal{L}\mathcal{R}$, welche durch die sogenannten lokalkompakten Limesräume (cf. [16]) gebildet wird und welche zwischen $\mathcal{L}\mathcal{R}\ell^0$ und $\mathcal{L}\mathcal{R}$ liegt: $\mathcal{K}\mathcal{E} \cong \mathcal{L}\mathcal{R}\ell^0 \subset \mathcal{L}\mathcal{R}\ell \subset \mathcal{L}\mathcal{R}$. Die Funktoren, welche diese verschiedenen Kategorien kartesisch abschließen, werden explizit beschrieben, wodurch auch die zwischen ihnen bestehenden Beziehungen ersichtlich werden.

---

* Diese Arbeit entstand während eines Aufenthaltes am Forschungsinstitut für Mathematik der Eidgenössischen Technischen Hochschule Zürich.

Allgemeines über Limesräume

Sind $\mathfrak{X}$, $\mathfrak{Y}$ Filter auf einer Menge M, so schreiben wir $\mathfrak{X} \leqq \mathfrak{Y}$, falls $\mathfrak{X}$ feiner ist als $\mathfrak{Y}$. Wir verzichten ferner auf die Bedingung, daß die leere Menge $\varnothing$ nicht zu einem Filter gehören dürfe. Dadurch erhalten wir einen zusätzlichen Filter auf M, nämlich den durch $\varnothing$ erzeugten Filter $[\varnothing]$. Dafür bilden dann alle Filter auf M einen vollständigen Verband.

Ist E ein Limesraum (cf. [8]), so bedeute $\mathfrak{X} \downarrow_a E$, daß $\mathfrak{X}$ ein Filter auf der E zugrunde liegenden Menge $\underline{E}$ sei, welcher gegen a konvergiere. Ein Punkt $a \in E$ heißt Adhärenzpunkt eines Filters $\mathfrak{X}$ auf E, falls es einen Filter $\mathfrak{Y}$ mit $[\varnothing] \neq \mathfrak{Y} \leqq \mathfrak{X}$ und $\mathfrak{Y} \downarrow_a E$ gibt. Ist speziell $\mathfrak{X} = [A]$ der durch eine Teilmenge A von E erzeugte Filter, so wird die Menge aller Adhärenzpunkte von $[A]$ mit $\bar{A}$ bezeichnet. Der Operator $A \mapsto \bar{A}$ ist im allgemeinen nicht idempotent; aber die Mengen A, für die $A = \bar{A}$ gilt, erfüllen die Axiome für die abgeschlossenen Mengen einer Topologie. Den mit der so erhaltenen Topologie versehenen Raum $\underline{E}$ bezeichnen wir mit tE. Ist $f\colon E \to F$ eine stetige Abbildung zwischen Limesräumen, so ist auch $f\colon tE \to tF$ stetig. Man erhält somit einen Funktor

$$t\colon \mathscr{LR} \to \mathscr{TOP}$$

der Kategorie der Limesräume in die Kategorie der topologischen Räume. Die identische Abbildung $1\colon E \to tE$ ist dabei stets stetig. Ferner ist eine Teilmenge U von E genau dann in tE offen, wenn U zu jedem Filter auf E gehört, welcher gegen irgendeinen Punkt von U konvergiert.

Die Kategorie $\mathscr{LR}$ besitzt induzierte (initiale) und coinduzierte (finale) Strukturen; daraus folgt, daß sie vollständig und covollständig ist. Sie ist aber auch kartesisch abgeschlossen (cf. [15]) mittels des Funktors

$$C\colon \mathscr{LR}^{\mathrm{op}} \times \mathscr{LR} \to \mathscr{LR};$$

dabei wird $C(E,F)$ gebildet, indem man die Menge der stetigen Abbildungen von E nach F mit der sogenannten Limitierung der stetigen Konvergenz versieht, für welche genau dann $\mathfrak{F} \downarrow_f C(E,F)$ gilt, wenn aus $\mathfrak{X} \downarrow_a E$ folgt, daß $\mathfrak{F}(\mathfrak{X}) \downarrow_{f(a)} F$. Die kartesische Abgeschlossenheit folgt unmittelbar aus der folgenden universellen Eigenschaft von $C(E,F)$: eine Abbildung $f\colon G \to C(E,F)$ ist genau dann stetig, wenn die durch $\hat{f}(z,x) = (f(z))(x)$ definierte Abbildung $\hat{f}\colon G \times E \to F$ stetig ist.

Daraus, daß der Funktor $t\colon \mathscr{LR} \to \mathscr{TOP}$ zum Inklusionsfunktor $\mathscr{TOP} \to \mathscr{LR}$ coadjungiert ist, oder durch direkte Verifikation, ergibt sich leicht

Lemma 1. *Besitzt F die durch die Abbildungen $f_i\colon E_i \to F$ coinduzierte Limitierung, dann hat tF die durch die Abbildungen $f_i\colon tE_i \to tF$ coinduzierte Topologie.*

Die Kategorie der lokalkompakten Limesräume

Eine Teilmenge K eines Limesraumes E heißt kompakt, falls jeder Filter auf E, der die Menge K enthält, einen zu K gehörigen Adhärenzpunkt besitzt.

Das folgende Ergebnis stammt von Schroder (cf. [16]).

Lemma 2. *Es sei K eine kompakte Teilmenge von E und $\mathfrak{S}$ ein System von Teilmengen von E derart, daß jeder gegen einen Punkt von K konvergente Filter auf E eine Menge aus $\mathfrak{S}$ enthält. Dann kann K durch endlich viele Mengen von $\mathfrak{S}$ überdeckt werden.*

Beweis. Sei $\mathfrak{F}$ der Filter auf E, der durch die Komplementärmengen $\complement S$, $S \in \mathfrak{S}$, erzeugt wird. Die Behauptung ist äquivalent zur Aussage $\complement K \in \mathfrak{F}$. Angenommen es wäre $\complement K \notin \mathfrak{F}$. Dann gilt $\mathfrak{G} := \mathfrak{F} \wedge [K] \neq [\varnothing]$, und wegen $K \in \mathfrak{G}$ existiert dann ein Filter $\mathfrak{H}$ auf E mit $[\varnothing] \neq \mathfrak{H} \leqq \mathfrak{G}$, welcher gegen einen Punkt von K konvergiert. Folglich existiert ein $S \in \mathfrak{H}$ mit $S \in \mathfrak{S}$. Da anderseits $\complement S \in \mathfrak{F} \subset \mathfrak{H}$, folgt $\varnothing = S \cap \complement S \in \mathfrak{H}$, im Widerspruch zu $\mathfrak{H} \neq [\varnothing]$.

Für Limesräume gilt auch ein Satz von Tychonof (cf. [16]), der wie im klassischen Fall bewiesen werden kann:

Satz 3. *Ein Produkt von nicht leeren Limesräumen ist genau dann kompakt, wenn alle Faktoren kompakt sind.*

Definition 4. Ein Limesraum E heißt *lokalkompakt*, wenn jeder konvergente Filter auf E eine kompakte Menge enthält.

Jedem Limesraum E kann ein neuer Limesraum lE zugeordnet werden, indem man $\underline{E}$ mit der durch die Inklusionen der kompakten Teilräume von E coinduzierten Limitierung versieht. Man zeigt leicht, daß dann E und lE die gleichen kompakten Teilräume haben und daß eine Abbildung $f\colon lE \to F$ genau dann stetig ist, wenn $f\colon E \to F$ „stetig auf Kompakta" ist, d.h. wenn die Restriktion $f|K\colon K \to F$ für jeden kompakten Teilraum K von E stetig ist. Für die Konvergenz auf lE erhält man: $\mathfrak{X} \downarrow_a lE$ genau dann, wenn $\mathfrak{X}$ eine kompakte Teilmenge von E enthält und $\mathfrak{X} \downarrow_a E$ gilt. Somit ist lE stets ein lokalkompakter Limesraum. Weil anderseits das stetige Bild eines Kompaktums wieder kompakt ist, impliziert die Stetigkeit einer Abbildung $f\colon E \to F$ jene von $f\colon lE \to lF$. Man erhält somit einen Funktor

$$l\colon \mathscr{LR} \to \mathscr{LR}\ell,$$

wobei $\mathscr{LR}\ell$ die aus den lokalkompakten Limesräumen gebildete volle Unterkategorie von $\mathscr{LR}$ bezeichnet. Weil offenbar die identische Abbildung $lE \to E$ stets stetig ist, ist $l\colon \mathscr{LR} \to \mathscr{LR}\ell$ zum Inklusionsfunktor $i\colon \mathscr{LR}\ell \to \mathscr{LR}$ adjungiert. Ferner ist $l \circ i$ der identische Funktor von $\mathscr{LR}\ell$. Zusammenfassend hat man also:

Satz 5. *$\mathcal{LRl}$ ist eine reflektive Unterkategorie von $\mathcal{LR}$.*

Korollar 6. *Die Kategorie $\mathcal{LRl}$ ist auch vollständig und covollständig. Dabei fallen Colimites in $\mathcal{LRl}$ mit den in $\mathcal{LR}$ gebildeten zusammen, während man Limites in $\mathcal{LRl}$ erhält, indem man auf die in $\mathcal{LR}$ gebildeten Limites den Funktor l ausübt.*

Im Spezialfall eines in $\mathcal{LRl}$ zu bildenden Produktes ist es allerdings nicht nötig, l anzuwenden; aus Satz 3 folgt sofort, daß jedes in $\mathcal{LR}$ gebildete Produkt lokalkompakter Räume bereits lokalkompakt ist, also mit dem in $\mathcal{LRl}$ gebildeten Produkt übereinstimmt.

Satz 7. *Die Kategorie $\mathcal{LRl}$ wird durch den Funktor $(F, G) \mapsto l\,C(F, G)$ kartesisch abgeschlossen.*

Beweis. Es ist zu zeigen, daß für $E, F, G \in \mathcal{LRl}$ gilt:

$$f\colon E \to l\,C(F, G) \text{ stetig} \Leftrightarrow \hat{f}\colon E \times F \to G \text{ stetig}.$$

Dabei bezeichnet $E \times F$ das in $\mathcal{LR}$ gebildete Produkt von E und F, welches, wie oben erwähnt, mit dem in $\mathcal{LRl}$ gebildeten übereinstimmt, und $\hat{f}$ ist wiederum durch $\hat{f}(x, y) = (f(x))(y)$ definiert. Gemäß der universellen Eigenschaft von $C(F, G)$ ist $\hat{f}\colon E \times F \to G$ genau dann stetig, wenn $f\colon E \to C(F, G)$ stetig ist; letzteres ist aber gemäß Satz 5 zur Stetigkeit von $f\colon E \to l\,C(F, G)$ äquivalent.

Die den kompakt erzeugten Räumen entsprechenden Limesräume

Durch Restriktion auf die betreffenden Unterkategorien erhalten wir Funktoren $l\colon \mathcal{TOP} \to \mathcal{LRl}$ und $t\colon \mathcal{LRl} \to \mathcal{TOP}$. Man verifiziert leicht den

Satz 8. *$l\colon \mathcal{TOP} \to \mathcal{LRl}$ ist zu $t\colon \mathcal{LRl} \to \mathcal{TOP}$ adjungiert, und es gilt $l \circ t \circ l = l$ und $t \circ l \circ t = t$.*

Korollar 9. *$t\,l\colon \mathcal{TOP} \to \mathcal{TOP}$ und $l\,t\colon \mathcal{LRl} \to \mathcal{LRl}$ sind idempotente Endofunktoren. Die aus den gegenüber $t\,l$ bzw. $l\,t$ invarianten Objekten gebildeten vollen Unterkategorien $\mathcal{TOP}^*$ bzw. $\mathcal{LRl}^*$ von $\mathcal{TOP}$ bzw. $\mathcal{LRl}$ sind isomorph. Die Funktoren $t\,l$ bzw. $l\,t$ induzieren Retraktionsfunktoren $t\,l\colon \mathcal{TOP} \to \mathcal{TOP}^*$ bzw. $l\,t\colon \mathcal{LRl} \to \mathcal{LRl}^*$, welche zu den entsprechenden Inklusionsfunktoren adjungiert bzw. coadjungiert sind.*

Korollar 10. *$\mathcal{TOP}^*$ und $\mathcal{LRl}^*$ sind vollständig und covollständig.*

Lemma 11. *Ist $E = l\,t\,E$ und $F = l\,t\,F$, dann ist $1\colon t\,C(E, F) \to CO(t\,E, t\,F)$ stetig. Dabei bezeichnet $CO(t\,E, t\,F)$ den mit der kompakt-offenen Topologie versehenen Raum aller stetigen Abbildungen $t\,E \to t\,F$.*

Beweis. Auf Grund der Funktorialität von l und t und der Voraussetzungen bezüglich E und F sind die zugrunde liegenden Mengen von $t\,C(E, F)$ und $CO(t\,E, t\,F)$ identisch. Die behauptete Stetigkeit folgt, wenn gezeigt wird, daß jede Subbasis-offene Teilmenge von $CO(tE, tF)$ auch in $tC(E, F)$ offen ist. Eine Subbasis für die offenen Mengen von $CO(tE, tF)$ wird gebildet durch die Mengen der Form $W(K, U) := \{f \in CO(tE, tF); f(K) \subset U\}$, wobei K in tE kompakt und U in tF offen ist. Sei also $W(K, U)$ eine solche Menge; und sei $f \in W(K, U)$ und $\mathfrak{F} \downarrow_f C(E, F)$. Es ist zu zeigen, daß dann $W(K, U) \in \mathfrak{F}$ gilt. Ist nun $x \in K$ und $\mathfrak{X} \downarrow_x E$, dann gilt $\mathfrak{F}(\mathfrak{X}) \downarrow_{f(x)} F$; somit. weil $f(x) \in U$ und U in tF offen: $U \in \mathfrak{F}(\mathfrak{X})$. Wir können daher zu $x \in K$ und $\mathfrak{X} \downarrow_x E$ Mengen $F = F_{x, \mathfrak{x}} \in \mathfrak{F}$ und $X = X_{x, \mathfrak{x}} \in \mathfrak{X}$ wählen, für die $U \supset F(X)$ gilt. Diese Mengen X bilden dann ein System $\mathfrak{S}$, welches die Voraussetzungen von Lemma 2 erfüllt; es folgt daher, daß K durch endlich viele solcher Mengen X überdeckt werden kann: $K \subset X_1 \cup \ldots \cup X_n$. Es seien $F_1, \ldots, F_n$ die entsprechenden Mengen von $\mathfrak{F}$, für welche also $F_i(X_i) \subset U$ gilt. Dann ist $F := F_1 \cap \ldots \cap F_n \in \mathfrak{F}$ und es folgt $F \subset W(K, U)$; in der Tat, ist $f \in F$ und $x \in K$, so existiert $i \in \{1, \ldots, n\}$ mit $x \in X_i$, und wegen $f \in F_i$ folgt $f(x) \in F_i(X_i) \subset U$.

Lemma 12. *Ist $E = ltE$ und $F = ltF$, dann ist $1: lCO(tE, tF) \to C(E, F)$ stetig.*

Beweis. Bekanntlich hat die kompakt-offene Topologie die Eigenschaft, daß die Evaluationsabbildung e: $CO(tE, tF) \times tE \to tF$ „stetig auf Kompakta" ist. Also ist e: $l(CO(tE, tF) \times tE) \to ltF$ stetig. Benützt man, daß l als adjungierter Funktor mit Produkten verträglich ist, und berücksichtigt man die Voraussetzung betreffend E und F, ergibt dies e: $lCO(tE, tF) \times E \to F$ ist stetig. Die universelle Eigenschaft von $C(E, F)$ liefert daraus direkt die Behauptung.

Korollar 13. *Ist $E = ltE$ und $F = ltF$, dann ist $1: ltC(E, F) \to C(E, F)$ stetig.*

Satz 14. *Die Kategorie $\mathscr{L}\mathscr{R}\ell^*$ wird durch den Funktor $(F, G) \mapsto ltC(F, G)$ kartesisch abgeschlossen.*

Beweis. Man bemerkt zuerst, daß für $F, G \in \mathscr{L}\mathscr{R}\ell^*$ wegen Satz 8 auch $ltC(F, G) \in \mathscr{L}\mathscr{R}\ell^*$ gilt. Der Rest des Beweises ist analog zu jenem von Satz 7. Man benützt dabei einerseits Korollar 13; andererseits die Tatsache, daß Produkte in $\mathscr{L}\mathscr{R}\ell^*$ mit jenen von $\mathscr{L}\mathscr{R}\ell$ (und somit mit jenen von $\mathscr{L}\mathscr{R}$) übereinstimmen, weil ja der Inklusionsfunktor $\mathscr{L}\mathscr{R}\ell^* \to \mathscr{L}\mathscr{R}\ell$ gemäß Korollar 9 ein adjungierter Funktor ist und somit mit Produkten verträglich ist.

Satz 15. *Ist X ein hausdorffscher*[1] *topologischer Raum, dann ist tlX der zugehörige (in* [18] *mit kX bezeichnete) kompakt erzeugte Raum.*

---

[1] Kompakt erzeugte Räume werden in der Regel hausdorffsch vorausgesetzt. Satz 15 gilt auch, wenn man von dieser Konvention absieht; man vergleiche Bemerkung 2 am Schluß dieser Arbeit.

Beweis. Die Behauptung folgt aus Lemma 1, da bekanntlich die Topologie von kX die durch die Inklusionen der kompakten Teilräume von X coinduzierte Topologie ist.

Satz 16. *Der Funktor l induziert einen Isomorphismus der Kategorie $\mathcal{KE}$ der kompakt erzeugten Räume auf die volle Unterkategorie $\mathcal{LR\ell}^\circ$ von $\mathcal{LR\ell}^*$, die aus denjenigen Objekten E gebildet wird, für die tE hausdorffsch ist. Die Kategorie $\mathcal{LR\ell}^\circ$ wird durch den Funktor $(F, G) \mapsto l\,t\,C(F, G)$ kartesisch abgeschlossen.*

Beweis. Die erste Aussage folgt aus Satz 15 und Korollar 9. Für die zweite Aussage benützt man folgendes:

a) Ist $E, F \in \mathcal{LR\ell}^\circ$, so gilt $t(E \times F) = t(ltE \times ltF) = tl(tE \times tF)$, und dies ist hausdorffsch, da $1: tl(tE \times tF) \to tE \times tF$ stetig ist und sogar $tE \times tF$ hausdorffsch ist. Also ist $E \times F \in \mathcal{LR\ell}^\circ$; das in $\mathcal{LR}$ gebildete Produkt $E \times F$ ist also auch jenes von $\mathcal{LR\ell}^\circ$.

b) Ist $F, G \in \mathcal{LR\ell}^\circ$, so gilt gemäß den Lemmata 11 und 12: $t\,l\,t\,C(F, G) = t\,l\,CO(tF, tG)$. Da $CO(tF, tG)$ bekanntlich hausdorffsch ist, folgt wie bei a), daß $l\,t\,C(F, G) \in \mathcal{LR\ell}^\circ$.

Bemerkungen. 1) Aus dem Funktor, der $\mathcal{LR\ell}^\circ$ kartesisch abschließt, erhält man mittels der Isomorphie von $\mathcal{LR\ell}^\circ$ und $\mathcal{KE}$ auch den Funktor, der $\mathcal{KE}$ kartesisch abschließt, nämlich $(X, Y) \mapsto t\,l\,t\,C(lX, lY)$. Gemäß den Lemmata 11 und 12 ergibt sich, daß dieser mit dem wohlbekannten $\mathcal{KE}$ kartesisch abschließenden Funktor übereinstimmt: für $X, Y \in \mathcal{KE}$ gilt $t\,l\,t\,C(lX, lY) = t\,l\,CO(t\,lX, t\,lY) = k\,CO(X, Y)$.

2) $\mathcal{TOP}^*$ ist die Kategorie der „kompakt erzeugten Räume ohne Trennungsaxiom". Man gelangt von ihr zu $\mathcal{KE}$, indem man die Objekte X von $\mathcal{TOP}^*$ durch die Bedingung „X hausdorffsch" einschränkt. Geht man in die zu $\mathcal{TOP}^*$ isomorphe Kategorie $\mathcal{LR\ell}^*$, so übersetzt sich diese Bedingung in „tE hausdorffsch"; von diesem Gesichtspunkt aus würde die Trennungsbedingung „E separiert" (im Sinne der Limesräume) natürlicher erscheinen; ihr entspräche in $\mathcal{TOP}^*$ die Bedingung „lX separiert", und man würde eine zwischen $\mathcal{KE}$ und $\mathcal{TOP}^*$ gelegene Kategorie erhalten.

Literatur

1. Bastiani, A.: Applications différentiables et variétés différentiables de dimension infinie. J. Analyse Math. **13**, 1–114 (1964).
2. Binz, E.: Ein Differenzierbarkeitsbegriff in limitierten Vektorräumen. Commentarii Math. Helvet. **41**, 137–156 (1966).
3. Binz, E.: Kompakte Limesräume und limitierte Funktionenalgebren. Commentarii Math. Helvet. **43**, 195–203 (1968).
4. Binz, E.: Notes on a characterization of function algebras. Math. Ann. **186**, 314–326 (1970).
5. Binz, E., Butzmann, H.-P., Kutzler, K.: Über den c-Dual eines topologischen Vektorraumes. Math. Z. **127**, 70–74 (1972).

6. Butzmann, H.-P.: Über die c-Reflexivität von $C_c(X)$. Manuskript Nr. 16 (1971), Math. Fak. Univ. Mannheim.
7. Dubuc, E. J., Porta, H.: Convenient categories of topological algebras. Bull. Amer. Math. Soc. **77**, 975–979 (1971).
8. Fischer, H. R.: Limesräume. Math. Ann. **137**, 269–303 (1959).
9. Frölicher, A.: Sur l'anneau des fonctions continues d'un espace à génération compacte. C. R. Acad. Sci. Paris **275**, 25–27 (1972).
10. Frölicher, A., Bucher, W.: Calculus in vector spaces without norm. Lecture Notes in Math. 30. Berlin-Heidelberg-New York: Springer 1966.
11. Frölicher, A., Jarchow, H.: Zur Dualitätstheorie kompakt erzeugter und lokalkonvexer Vektorräume; wird erscheinen in Commentarii Math. Helvet.
12. Gabriel, P., Zisman, M.: Calculus of fractions and homotopy theory. Ergebn. der Math. 35. Berlin-Heidelberg-New York: Springer 1967.
13. Keller, H. H.: Über Probleme, die bei einer Differentialrechnung in topologischen Vektorräumen auftreten. Festband zum 70. Geburtstag von Rolf Nevanlinna. Berlin-Heidelberg-New York: Springer 1966.
14. Kelley, J. L.: General topology. Princeton, N. J.: Van Nostrand 1955.
15. MacLane, S.: Categories. Berlin-Heidelberg-New York: Springer 1971.
16. Schroder, M.: Continuous convergence in a Gelfand theory for topological algebras. Ph. D. thesis, Queens University, Kingston, Ontario (1971).
17. Seip, U.: Kompakt erzeugte Vektorräume und Analysis. Lecture Notes in Math. **273**. Berlin-Heidelberg-New York: Springer 1972.
18. Steenrod, N.: A convenient category of topological spaces. Michigan Math. J. **14**, 133–152 (1967).

Dr. Alfred Frölicher
Institut de Mathématiques
Université de Genève
CH-1200 Genève
Schweiz

(Eingegangen am 3. August 1972)

Math. Z. 129, 65–73 (1972)
© by Springer-Verlag 1972

Isomorphism of Modular Group Algebras

Inder Bir S. Passi and Sudarshan K. Sehgal

1. Introduction

For a group G, let $M_i(G)$ be the i-th term in its Brauer-Jennings-Zassenhaus series which is defined inductively by $M_1(G)=G$, $M_i(G)=(G, M_{i-1}(G))\, M_{(i/p)}(G)^p$ for $i\geqq 2$ where (i/p) is the least integer $\geqq i/p$ and $(G, M_{i-1}(G))$ denotes the subgroup generated by all commutators $(x, y)=x^{-1}y^{-1}x\,y$, $x\in G$, $y\in M_{i-1}(G)$. It is known ([1, 5, 6, 11, 13]) that if k is a field of characteristic $p>0$ and $\Delta_k(G)$ is the augmentation ideal of the group algebra $k(G)$, then

$$M_i(G)=\{g\in G \mid g-1\in \Delta_k^i(G)\}.$$

It is also known [6] that

$$M_n(G)=\prod_{i\,p^j\geqq n} G_i^{p^j}$$

where G_i is the i-th term in the lower central series

$$G=G_1\geqq G_2\geqq \cdots \geqq G_i\geqq \cdots \qquad \text{of } G.$$

Our first main result is that if G and H are two groups with isomorphic group algebras over the field Z_p of p elements for a prime p, then they have isomorphic M-series for that prime. We exploit the fact that the natural monomorphism

$$0\to M_i(G)/M_{i+1}(G)\to \Delta_{Z_p}^i(G)/\Delta_{Z_p}^{i+1}(G)$$

splits over Z_p to deduce that, in fact, the subquotients $M_i(G)/M_{i+2}(G)$ are also isomorphism invariants of the group algebra $Z_p(G)$ of G with coefficients in Z_p.

We next consider the case when the coefficients are in the ring Z_{p^h} of integers mod p^h, p prime. Corresponding to the M-series, we have the series [6]

$$M_{n,h,p}(G)=\prod G_i^{p^j}$$

where the product is taken over all i, j such that $i\geqq n$ or $i\,p^{(j-h+1)}\geqq n$. We denote by $D_{i,h,p}(G)$ the i-th dimension subgroup of G with coefficients in the ring Z_{p^h} i.e. $D_{i,h,p}(G)=\{g \mid g-1\in \Delta_{Z_{p^h}}^i(G)\}$ where $\Delta_{Z_{p^h}}(G)$ is the augmentation ideal of $Z_{p^h}(G)$. By a result of Moran [8] $M_{i,h,p}(G)=$

$D_{i,h,p}(G)$ for $i \leqq p$. We show that the counterpart of the splitting mentioned in the previous paragraph is available for the natural embedding

$$0 \to G_2/G_2 \cap D_3(G, R) \to \Delta_R^2(G)/\Delta_R^3(G)$$

where R is the ring of integers mod n, n odd, or R is the ring of integers and G/G_2 is of odd exponent. $D_3(G, R)$ stands for the 3rd dimension subgroup of G over R. Thus, for example, we show that whenever for two groups G and H

$$Z_{p^n}(G) \cong Z_{p^n}(H), \qquad p \text{ prime} \neq 2,$$

then $G/M_{3,n,p}(G) \cong H/M_{3,n,p}(H)$ and moreover if, in addition, G and H are each of exponent p^n and class 2, then $G \cong H$.

2. Notations

Let G be a group, R a commutative ring with unity. We adopt the following notation:

$R(G)$ = the group ring of G with coefficients in R.

$\Delta_R(G)$ = the augmentation ideal of $R(G)$.

$Q_{n,R}(G) = \Delta_R^n(G)/\Delta_R^{n+1}(G)$.

$\Delta_R^{[i]}(G) = \Delta_R(G)$ for $i=1$ and $\Delta_R^{[i]}(G) = [\Delta_R(G), \Delta_R^{[i-1]}(G)]$

the additive subgroup generated by all elements of the type

$$\alpha\beta - \beta\alpha, \ \alpha \in \Delta_R(G), \ \beta \in \Delta_R^{[i-1]}(G) \text{ if } i > 1.$$

$\Delta_R(G, N)$ = Kernel of the natural homomorphism $R(G) \to R(G/N)$, where N is a normal subgroup of G.

$G = G_1 \geqq G_2 \geqq \cdots \geqq G_i \geqq \cdots$ denotes the lower central series of G. Z_n = the ring of integers mod n.

3. M-Series

Lemma 1. *$\Delta_R^{[i]}(G) + \Delta_R^{i+1}(G) = \Delta_R(G, G_i) + \Delta_R^{i+1}(G)$ for all $i \geqq 1$.*

Proof. Induct on i or see [2, Theorem 6].

Following Zassenhaus [13] we define a series of ideals $L_{n,h,p}(G)$ of $Z_{p^h}(G)$ by setting

$$L_{n,h,p}(G) = \sum p^{h-1} (\Delta^{[i]})^{p^j} + \Delta^{[n]}(G) + \Delta^{n+1}(G)$$

where $\Delta = \Delta_{Z_{p^h}}$, the summation is over all i, j such that $i < n$, $i p^j \geqq n$ and by S^{p^r} for a subset S of $Z_{p^h}(G)$ we understand the Z_{p^h}-submodule of $Z_{p^h}(G)$ generated by s^{p^r}, $s \in S$.

Lemma 2. *$M_{n,h,p}(G)/D_{n+1,h,p}(G)\cap M_{n,h,p}(G)\cong L_{n,h,p}(G)/\Delta^{n+1}(G)$ for all $n\geqq 1$.*

Proof. By definition $M_{n,h,p}(G)=\prod_{i\,p^{(j-h+1)^+}\geqq n} G_i^{p^j}$.

For $g\in G_i$, we have $g-1\in\Delta(G_i)$ and by Lemma 1 $g-1=\alpha+\beta$ for some $\alpha\in\Delta^{[i]}(G)$, $\beta\in\Delta^{i+1}(G)$.

If for some integer j, $i\,p^{(j-h+1)^+}\geqq n$, then $(j-h+1)^+=0$ would mean $i\geqq n$, $g^{p^j}\in G_n$ and so $g^{p^j}-1\in\Delta^{[n]}(G)+\Delta^{n+1}(G)$. If $j-h+1>0$, then consider

$$g^{p^j}-1=\sum_{r=1}^{p^j}\binom{p^j}{r}(g-1)^r.$$

Now p^h divides $\binom{p^j}{r}$ for $r<p^{j-h+1}$ and for $r>p^{j-h+1}$ we have $i\,r>i\,p^{j-h+1}\geqq n$. Hence, module $L_{n,h,p}(G)$

$$\begin{aligned} g^{p^j}-1 &\equiv \binom{p^j}{p^{j-h+1}}(g-1)^{p^{j-h+1}} \\ &\equiv a\,p^{h-1}(g-1)^{p^{j-h+1}} \qquad \text{where a is an integer [9]} \\ &\equiv a\,p^{h-1}(\alpha+\beta)^{p^{j-h+1}} \\ &\equiv 0. \end{aligned}$$

The equation

$$x\,y-1=(x-1)+(y-1)+(x-1)(y-1) \tag{$*$}$$

allows us to conclude that

$$m\in M_{n,h,p}(G)\Rightarrow m-1\in L_{n,h,p}(G).$$

Define a map θ: $M_{n,h,p}(G)\to L_{n,h,p}(G)/\Delta^{n+1}(G)$

$$m\to m-1+\Delta^{n+1}(G).$$

It follows from $(*)$ that θ is a homomorphism and clearly the kernel of θ is $D_{n+1,h,p}(G)$. It remains to prove that θ is an epimorphism. In view of Lemma 1, we have only to verify that if $g\in G_i$ and $i\,p^j\geqq n$, then $p^{h-1}(g-1)^{p^j}$ has a preimage.

Now

$$\begin{aligned} g^{p^{j+h-1}}-1 &\equiv \binom{p^{j+h-1}}{p^j}(g-1)^{p^j} \bmod \Delta^{n+1}(G) \\ &\equiv a\,p^{h-1}(g-1)^{p^j} \qquad \text{where } a \text{ is an integer [9].} \end{aligned}$$

Since there exists an integer b such that $a\,b\equiv 1\ (\mathrm{mod}\ p^h)$, we have

$$b(g^{p^{j+h-1}}-1)\equiv p^{h-1}(g-1)^{p^j} \bmod \Delta^{n+1}(G).$$

Consequently

$$g^{b p^{j+h-1}}-1 \equiv p^{h-1}(g-1)^{p^j} \bmod \Delta^{n+1}(G)$$

and θ is onto as $g^{p^{j+h-1}} \in G_i^{p^{j+h-1}} \subseteq M_{n,h,p}(G)$.

Theorem 3. *Let θ: $Z_p(G) \to Z_p(H)$ be an isomorphism. Then*

(i) $M_i(G)/M_{i+1}(G) \cong M_i(H)/M_{i+1}(H)$,

(ii) $M_i(G)/M_{i+2}(G) \cong M_i(H)/M_{i+2}(H)$

for all $i \geqq 1$, where $M_1(G) = G \supseteq M_2(G) \supseteq \cdots \supseteq M_i(G) \supseteq \cdots$ is the M-series with respect to the prime p.

Proof. We remark first that one can assume without loss of generality that θ is normalized in the sense that the sum of the coefficients of $\theta(g)$ is 1 for all $g \in G$. Thus the expression on the right hand side of the isomorphism in Lemma 2 is a ring invariant and (i) follows from the fact that $D_{i,1,p}(G) = M_i(G) = M_{i,1,p}(G)$.

Consider the embedding

$$M_i(G)/M_{i+1}(G) \to \Delta^i(G)/\Delta^{i+1}(G), \qquad \Delta = \Delta_{Z_p},$$

given by

$$m + M_{i+1}(G) \to m - 1 + \Delta^{i+1}(G), \qquad m \in M_i(G).$$

Since $\Delta^i(G)/\Delta^{i+1}(G)$ is a vector space over Z_p, the embedding splits over Z_p and we have

$$\Delta^i(G)/\Delta^{i+1}(G) = \Delta(G, M_i) + \Delta^{i+1}(G)/\Delta^{i+1}(G) \oplus K_i(G)/\Delta^{i+1}(G), \quad (**)$$

where $K_i(G)$ is a subspace of $Z_p(G)$. Moreover, since $\Delta^{i+1}(G) \subseteq K_i(G) \subseteq \Delta^i(G)$, we can conclude that $K_i(G)$ is an ideal of $Z_p(G)$. We claim that $M_i(G)/M_{i+2}(G)$ is isomorphic to the group of units $M_i(G) + K_{i+1}(G)/K_{i+1}(G)$ of $Z_p(G)/K_{i+1}(G)$. Define the map

$$\lambda\colon\ M_i(G) \to M_i(G) + K_{i+1}(G)/K_{i+1}(G)$$
$$m \to m + K_{i+1}(G).$$

Clearly λ is an epimorphism and $M_{i+2}(G)$ is contained in the kernel of λ. Suppose $\lambda(m) = 1 + K_{i+1}(G)$ i.e. $m - 1 \in K_{i+1}(G)$. Then $m - 1 \in \Delta^{i+1}(G)$ and so $m \in M_{i+1}(G)$. But because $\Delta(G, M_{i+1}) \cap K_{i+1}(G) \subseteq \Delta^{i+2}(G)$, we conclude that $m \in M_{i+2}(G)$ proving the isomorphism.

By Lemma 4(a) below, (**) yields the splitting

$$\Delta^i(H)/\Delta^{i+1}(H) = \Delta(H, M_i(H)) + \Delta^{i+1}(H)/\Delta^{i+1}(H) \oplus \theta(K_i(G))/\Delta^{i+1}(H).$$

Therefore

$$\begin{aligned} M_i(H)/M_{i+2}(H) &\cong M_i(H) + \theta(K_{i+1}(G))/\theta(K_{i+1}(G)) \\ &\cong M_i(G) + K_{i+1}(G)/K_{i+1}(G) \qquad \text{(Lemma 4(b) below)} \\ &\cong M_i(G)/M_{i+2}(G). \end{aligned}$$

Lemma 4. *Under the assumptions of Theorem* 3, *we have*

(a) $\theta(\Delta(G, M_n(G)) + \Delta^{n+1}(G)) = \Delta(H, M_n(H)) + \Delta^{n+1}(H)$

(b) $\theta(M_n(G) + K_{n+1}(G)) = M_n(H) + \theta(K_{n+1}(G))$

where $K_{n+1}(G)$ *is an ideal of* $Z_p(G)$ *yielding a splitting of* $\Delta^n(G)/\Delta^{n+1}(G)$ *as in* (**).

Proof. (a) Every element x of $\Delta(G, M_n(G)) + \Delta^{n+1}(G)$ can be written as $x = m - 1 + \alpha$, $m \in M_n(G)$, $\alpha \in \Delta^{n+1}(G)$, since $M_n(G) \subseteq 1 + \Delta^n(G)$. As in the proof of Lemma 2, $m - 1 \in L_{n,1,p}(G)$ and therefore $\theta(m-1) \in L_{n,1,p}(H)$. Also $\theta(\alpha) \in L_{n,1,p}(H)$. Hence $\theta(x) \in L_{n,1,p}(H)$. By Lemma 2 again $\theta(x) = h - 1 + \gamma$, $h \in M_n(H)$, $\gamma \in \Delta^{n+1}(H)$ and we are done.

(b) Let $m \in M_n(G)$. Then $\theta(m-1) = h - 1 + \gamma$ for some $h \in M_n(H)$ and $\gamma \in \Delta^{n+1}(H)$. As $\theta(K_{n+1}(G))$ yields a splitting, $\gamma = u - 1 + v$ for some $u \in M_{n+1}(H)$, $v \in \theta(K_{n+1}(G))$.

Therefore,

$$\begin{aligned}\theta(m-1) &= h - 1 + u - 1 + v \\ &= hu - 1 + w \qquad \text{where } w = v - (h-1)(u-1).\end{aligned}$$

As $(h-1)(u-1) \in \Delta^{n+2}(H) = \theta(\Delta^{n+2}(G)) \subseteq \theta(K_{n+1}(G))$, $w \in \theta(K_{n+1}(G))$. Since $hu \in M_n(H)$, the proof of the Lemma is complete.

Corollary 5. *Let* $Z_p(G) \cong Z_p(H)$. *Then* $M_i(G) = (1)$ *if and only if* $M_i(H) = (1)$.

Proof. Suppose $M_i(G) = (1)$. Then, by Theorem 3, $M_i(H) = M_{i+j}(H)$ for all $j \geqq 0$. $M_i(G) = (1)$ implies that G is a nilpotent p-group of bounded exponent. Thus we can conclude [4, Theorem E] that $\bigcap_l \Delta^l(G) = (0)$ where $\Delta = \Delta_{Z_p}$ and, therefore, $\bigcap_l \Delta^l(H) = (0)$. It follows that $M_i(H) = (1)$. The converse follows by symmetry.

Related to Corollary 5 is the following remark:

Remark 6. *Let* $Z_p(G) \cong Z_p(H)$. *Then* $\bigcap_i M_i(G) = (1)$ *if and only if* $\bigcap_i M_i(H) = (1)$.

Proof. Suppose $\bigcap_i M_i(G) = (1)$. Then G is residually nilpotent p-group of bounded exponent. Therefore, $\bigcap_i \Delta^i_{Z_p}(G) = (0)$ ([4, Theorem E] and [7, Lemma 5.2]). Thus $\bigcap_i \Delta^i_{Z_p}(H) = (0)$ and, therefore, $\bigcap_i M_i(H) = (1)$.

The converse follows by symmetry.

Corollary 5 and Theorem 3 give the following:

Corollary 7. *Suppose* G *has* M-*length* $\leqq 2$ *i.e.* $M_3(G) = (1)$. *Then* $Z_p(G) \cong Z_p(H)$ *implies* $G \cong H$.

The next result has been proved by Dieckmann [3] for finite p-groups of class 2, $p \neq 2$.

Corollary 8. *Let G and H be two groups such that*

(i) $Z_p(G) \cong Z_p(H)$

(ii) *G is of class $<p$ and exponent p^n.*

Then H is of exponent p^n.

Proof. We observe that $M_{p^n}(G) = \prod_{i p^j \geqq p^n} G_i^{p^j} = (1)$ and that $M_k(G) \neq (1)$ for $k < p^n$. It follows by Corollary 5 that $M_{p^n}(H) = (1)$ and $M_k(H) \neq (1)$ for $k < p^n$. Hence H is of exponent p^n.

4. Groups of Class Two

Let R be a commutative ring with unity, G a group. Consider the natural epimorphism

$$\alpha\colon R(G) \to R(G/G_2)$$

$$\alpha\Big(\sum_{\substack{r \in R\\ g \in G}} r g\Big) = \sum r \bar{g} \qquad \text{where } \bar{g} = g G_2 .$$

α induces an epimorphism $\alpha\colon Q_{2,R}(G) \to Q_{2,R}(G/G_2)$ whose kernel is evidently $\Delta_R(G, G_2) + \Delta_R^3(G)/\Delta_R^3(G)$.

Consider the map

$$\tau\colon G/G_2 \to Q_{2,R}(G/G_2) \qquad \text{given by } \tau(x G_2) = (x-1)^2 + \Delta_R^3(G).$$

τ is well-defined. For, if $y \in G_2$, then

$$\begin{aligned}(x y - 1)^2 &= (x - 1 + y - 1 + (x-1)(y-1))^2\\ &\equiv (x-1)^2 \bmod \Delta_R^3(G), \qquad \text{since } G_2 \subseteq 1 + \Delta_R^2(G).\end{aligned}$$

Extend τ by linearity to $R(G/G_2)$. We assert that τ vanishes on $\Delta_R^3(G/G_2)$. For, let $x, y, z \in G$. Then, writing $\bar{x} = x G_2$ etc., we have

$$\begin{aligned}\tau\big((\bar{x}-1)(\bar{y}-1)(\bar{z}-1)\big) &= (x y z - 1)^2 - (x y - 1)^2 - (x z - 1)^2 - (y z - 1)^2\\ &\quad + (x-1)^2 + (y-1)^2 + (z-1)^2 + \Delta_R^3(G)\\ &= (x - 1 + y - 1 + z - 1)^2 - (x - 1 + y - 1)^2\\ &\quad - (x - 1 + z - 1)^2 - (y - 1 + z - 1)^2\\ &\quad + (x-1)^2 + (y-1)^2 + (z-1)^2 + \Delta_R^3(G)\\ &= 0 .\end{aligned}$$

Thus τ induces an R-homomorphism

$$\tau\colon R(G/G_2)/\Delta_R^3(G/G_2) \to Q_{2,R}(G).$$

Restricting to $Q_{2,R}(G/G_2)$ we obtain a homomorphism

$$\tau\colon\ Q_{2,R}(G/G_2)\to Q_{2,R}(G).$$

Theorem 9. $\alpha\circ\tau=2I$ *where* I *is the identity homomorphism.*

Proof. Let $\bar{x}=xG_2$, $\bar{y}=yG_2$ be elements of G/G_2.

$$\begin{aligned}\tau\big((\bar{x}-1)(\bar{y}-1)\big)&=(xy-1)^2-(x-1)^2-(y-1)^2+\Delta_R^3(G)\\&=(x-1)(y-1)+(y-1)(x-1)+\Delta_R^3(G).\end{aligned}$$

Hence $\alpha\circ\tau\big((\bar{x}-1)(\bar{y}-1)+\Delta_R^3(G/G_2)\big)=2(\bar{x}-1)(\bar{y}-1)+\Delta_R^3(G/G_2)$.

As the elements $(\bar{x}-1)(\bar{y}-1)+\Delta_R^3(G/G_2)$ generate $Q_{2,R}(G/G_2)$ as an R-module, the result follows.

Corollary 10. *If division by 2 is uniquely defined in* $Q_{2,R}(G/G_2)$, *then the exact sequence*

$$0\to\Delta(G,G_2)+\Delta_R^3(G)/\Delta_R^3(G)\to Q_{2,R}(G)\xrightarrow{\ \alpha\ }Q_{2,R}(G/G_2)\to 0$$

splits.

Proof. Define $\tau'\colon Q_{2,R}(G/G_2)\to Q_{2,R}(G)$ by

$$\tau'(z)=\tau(y)\qquad\text{where } 2y=z,\ y,z\in Q_{2,R}(G/G_2).$$

Then we have $\alpha\circ\tau'=I$.

Let us examine the case when $R=Z_n$, the ring of integers mod n, n odd, or Z the ring of integers. The case $n=0$ in the following discussion corresponds to Z as the ring of coefficients. For every group G we have the exact sequence

$$0\to G_2/G_2\cap D_3(G,Z_n)\xrightarrow{\ i\ }Q_{2,Z_n}(G)\xrightarrow{\ \alpha\ }Q_{2,Z_n}(G/G_2)\to 0 \tag{4.1}$$

where i is the natural map induced by

$$\begin{aligned}i\colon\ &G_2\to Q_{2,Z_n}(G)\\&x\to x-1+\Delta_{Z_n}^3(G)\end{aligned}$$

and $D_r(G,Z_n)$ denotes the r-th dimension subgroup of G over Z_n. We also have the natural isomorphism

$$G/G_2G^n\cong\Delta_{Z_n}(G)/\Delta_{Z_n}^2(G) \tag{4.2}$$

induced by $x\to x-1+\Delta_{Z_n}^2(G)$, $x\in G$.

Suppose that the exact sequence (4.1) splits over Z_n. Then we have a Z_n-submodule K of $\Delta_{Z_n}^2(G)$ such that $K\supseteq\Delta_{Z_n}^3(G)$ and therefore is, in fact, an ideal of $Z_n(G)$ and

$$Q_{2,Z_n}(G)=\Delta_{Z_n}(G,G_2)+\Delta_{Z_n}^3(G)/\Delta_{Z_n}^3(G)\oplus K/\Delta_{Z_n}^3(G). \tag{4.3}$$

Consider the ring $S=Z_n(G)/K$. $U=1+\Delta_{Z_n}(G)/K$ is a group of units of S and we have a homomorphism $\lambda: G\to U$ given by $\lambda(x)=x+K$.

Let $\delta\in\Delta_{Z_n}(G)$. By (4.2)

$$\begin{aligned}\delta&=g-1+\delta_2, \quad \delta_2\in\Delta^2_{Z_n}(G)\\&=g-1+g_2-1+k, \quad g_2\in G_2,\ k\in K \quad \text{(by (4.3))}\\&=g\,g_2-1 \bmod K.\end{aligned}$$

Hence $\lambda(g\,g_2)=1+\delta+K$. This proves that λ is an epimorphism. Let $x\in\operatorname{Ker}\lambda$. Then $x-1\in K\subseteq\Delta^2_{Z_n}(G)$. Therefore, $x\in G_2\,G^n$, say $x=y\,z$, $y\in G_2$, $z\in G^n$. Now

$$y\,z-1=y-1+z-1+(y-1)(z-1).$$

The last term on the right hand side is in K, since $K\supseteq\Delta^3_{Z_n}(G)$. For $g\in G$,

$$\begin{aligned}g^n-1&\equiv n(g-1)+\binom{n}{2}(g-1)^2 \bmod \Delta^3_{Z_n}(G)\\&\equiv 0 \quad \text{since } n \text{ is odd or } 0.\end{aligned}$$

Thus $G^n\subseteq D_3(G,Z_n)$.

Hence $z-1\in K$. Consequently $y-1\in K$. As $y\in G_2$, (4.3) implies that $y-1\in\Delta^3_{Z_n}(G)$ i.e. $y\in D_3(G,Z_n)$ and so $x\in D_3(G,Z_n)$. We have proved the following

Theorem 11. *If n is an odd integer or 0 and the sequence*

$$0\to G_2/G_2\cap D_3(G,Z_n)\xrightarrow{\ i\ }Q_{2,Z_n}(G)\xrightarrow{\ \alpha\ }Q_{2,Z_n}(G/G_2)\to 0$$

splits over Z_n, then

$$G/D_3(G,Z_n)\cong 1+\Delta_{Z_n}(G)/K$$

where $K/\Delta^3_{Z_n}(G)$ is a complement of $\operatorname{Im} i$ *in $Q_{2,Z_n}(G)$.*

If n is an odd integer or G is a group of odd exponent, then Corollary 10 yields that the hypothesis of Theorem 11 is satisfied. The case $n=0$, i.e. the integral case, for finite groups is due to Sandling ([10, 11]). By Lemma 1 $\Delta_{Z_n}(G,G_2)+\Delta^3_{Z_n}(G)$ can be characterized ring theoretically. Thus if, for some group H, $Z_n(G)\cong Z_n(H)$, then the sequence corresponding to (4.2) for H would also split. We thus have

Theorem 12. *If n is an odd integer or G is a group of odd exponent, and $Z_n(G)\cong Z_n(H)$ for some group H, then*

$$G/D_3(G,Z_n)\cong H/D_3(H,Z_n).$$

Corollary 13. *If G and H are p-groups of exponent p^n, $p\neq 2$ and class 2 with $Z_{p^n}(G)\cong Z_{p^n}(H)$, then*

$$G\cong H.$$

Proof. By Theorem 12

$$G/D_3(G, Z_{p^n}) \cong H/D_3(H, Z_{p^n}).$$

Now

$$D_3(G, Z_{p^n}) = G^{p^n} G_2^{p^n} G_3 \quad [8]$$
$$= (1) \quad \text{(since } G \text{ is of exponent } p^n \text{ and class 2)}$$

and similarly $D_3(H, Z_{p^n}) = (1)$. Hence $G \cong H$.

Remark 14. It has been proved in [12] that for finite groups G and H with $|G| = |H| = p^m b$, $(p, b) = 1$ the isomorphism $Z_{p^{2m+1}}(G) \cong Z_{p^{2m+1}}(H)$ implies the isomorphism $Z_p^{\hat{}}(G) = Z_p^{\hat{}}(H)$ where $Z_p^{\hat{}}$ is the ring of p-adic integers. Hence it follows [12] that if $b = 1$ and $Z_{p^{2m+1}}(G) \cong Z_{p^{2m+1}}(H)$ then $G/G_3 \cong H/H_3$. The last two results replace the modulus p^{2m+1} by a much smaller power of p.

References

1. Bovdi, A.A.: Dimension subgroups. Proceedings of the Riga Seminar on Algebra 5–7 (1969).
2. Cohn, P.M.: Generalization of a theorem of Magnus. Proc. London Math. Soc. (3) **2**, 297–310 (1952).
3. Dieckmann, E.M.: Isomorphism of group algebras of p-groups. Ph.D. dissertation, Washington University, St. Louis, Missouri, 1967.
4. Hartley, B.: The residual nilpotence of wreath products. Proc. London Math. Soc. (3) **20**, 365–392 (1970).
5. Jennings, S.A.: The structure of the group ring of a p-group over a modular field. Trans. Amer. Math. Soc. **50**, 175–185 (1941).
6. Lazard, M.: Sur les groupes nilpotents et les anneaux de Lie. Ann. Sci. Ecole Norm Sup. (3) **71**, 101–190 (1954).
7. Mital, J.N.: On residual nilpotence. J. London Math. Soc. (2), **2**, 337–345 (1970).
8. Moran, S.: Dimension subgroups modulo n. Proc. Cambridge Philos. Soc. **68**, 579–582 (1970).
9. Passi, I.B.S.: Polynomial maps on groups. J. Algebra **9**, 121–151 (1968).
10. Sandling, R.: Note on the integral group ring problem. Math. Z. **124**, 255–258 (1972).
11. Sandling, R.: The modular group rings of p-groups. Thesis, University of Chicago, 1969.
12. Sehgal, S.K.: Isomorphism of p-adic group rings. J. Number Theory **2**, 500–509 (1970).
13. Zassenhaus, H.: Ein Verfahren jeder endlichen p-Gruppe einen Lie-Ring mit der Charakteristik p zuzuordnen. Abh. Math. Sem. Univ. Hamburg **13**, 200–207 (1940).

Dr. I.B.S. Passi
Department of Mathematics
University of Kurukshetra
Kurukshetra
India

Prof. S.K. Sehgal
Mathematisches Institut der Universität
D-6900 Heidelberg
Federal Republic of Germany

(Received May 2, 1972)

Math. Z. 129, 75–82 (1972)
© by Springer-Verlag 1972

A Limit-Point Criterion for Separated Dirac Operators and a Little Known Result on Riccati's Equation

Hubert Kalf

Introduction

There exists a vast body of literature on the limit-point or limit-circle behaviour of Sturm-Liouville operators, and even Dunford-Schwartz's long list of results [1, p. 1604–1607] is by no means complete. For systems of two first order differential operators the situation is quite different. Indeed, there seems to be only one result, the assumptions of which are mild enough not to exclude the most important example occurring in applications, i.e., the three-dimensional separated Dirac operator. This is Titchmarsh's criterion [9, p. 233] (cited here as Theorem 1) for the differential expression

$$Du(x) := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u'(x) + \begin{pmatrix} q_1(x) & \frac{k}{x} \\ \frac{k}{x} & q_2(x) \end{pmatrix} u(x) \tag{1}$$

$$(x \in (0, a],\ a > 0 \text{ arbitrary};\ k \in \mathbb{Z},\ k \neq 0),$$

stating that the limit-point behaviour of (1) with a Coulomb potential, remains unaltered under perturbations which are integrable down to $x=0$. We shall derive a related result (Theorem 2)—bearing, as does Titchmarsh's theorem, some resemblance to Sears's criterion for Sturm-Liouville operators (see [1, p. 1606, 7(a), (b)])—which allows even certain nonintegrable perturbations of the Coulomb potential (and at the same time certain integrable perturbations of the separation term k/x [Remark 2]).

Our method of proof is entirely different from that of Titchmarsh, who constructs a square integrable solution of

$$Du = \lambda u \quad (\lambda \in \mathbb{C}) \tag{2}$$

by means of Picard-Lindelöf's method of successive approximation and then shows that this is the sole solution of integrable square apart from constant factors. We use the fact that the nontrivial real solutions of (2) with $\lambda=0$ may be obtained from the real solutions of the nonlinear system

$$\Theta'=-\left[q_1(x)\cos^2\Theta+2\frac{k}{x}\sin\Theta\cos\Theta+q_2(x)\sin^2\Theta\right]\quad(x\in(0,a])\qquad(3)$$

$$(\log\rho)'=-\left\{[q_1(x)-q_2(x)]\tfrac{1}{2}\sin 2\Theta-\frac{k}{x}\cos 2\Theta\right\}\quad(x\in(0,a])\qquad(4)$$

by setting

$$u=\rho\begin{pmatrix}\cos\Theta\\ \sin\Theta\end{pmatrix}$$

(Kamke [6, p. 121 f.]). This Prüfer transformation was applied by Kamke in 1939 to obtain comparison theorems of Sturm's type for first order systems (*loc. cit.*, p. 126 f.; see also Weidmann [10, p. 255 ff.]). (3) in turn is equivalent to the Riccati equation (whose directional field is easier to discuss than that of (3))

$$y'=-[x\,q_1(x)+2k\,y+x\,q_2(x)\,y^2]\frac{1}{x}\quad(x\in(0,a])\qquad(5)$$

by the transformation $y=\tan\Theta$ (cf. Kamke [6, p. 309, Eq. 1.79]).

Under certain conditions the non-integrability of $1/x$ makes all solutions of (5) tend to one of the two roots of an algebraic equation associated with (5). Such an observation was first made by Poincaré [Amer. J. Math. **7**, 204 f. (1885)] under very special assumptions for the coefficients of the Riccati equation, $+\infty$ being the singular point. Later, the behaviour of the solutions of Riccati's equation near a singular point was studied by Petrovitch [7, p. 78 ff.], Horn [5, § 1] and most thoroughly by Falkenhagen [2], but their results seem to be little known[1]. Those results of Falkenhagen we need for our Theorem 2 are stated in the appendix (Theorem 3 and its corollary). The framework of the very simple though somewhat lengthy proof is outlined[2].

A Criterion of Sears's Type

For the differential expression (1) Titchmarsh [9, p. 233] obtained the following result.

---

[1] E.g., Hille, giving a very detailed account of the Riccati equation [4, p. 273–276, p. 456–470] derives a similar result for a special case (*loc. cit.*, p. 464 f.), but does not mention their work on the subject.

[2] There exist results similar to those of Theorem 3 for quite general nonlinear systems of dimension $n>2$ in which case one has to call on general topological principles (see Hartman [3, p. 279 f.]; cf. also *loc. cit.*, p. 205 f.). This was pointed out to us by J. Walter whom we should also like to thank for many stimulating discussions.

Theorem 1. *Suppose there exist constants*[3] $c_1, c_2 \in \mathbb{R}$ *and real-valued functions* $\tilde{q}_1, \tilde{q}_2 \in L^1(0, a]$ *such that*

$$q_j(x) = \frac{c_j}{x} + \tilde{q}_j(x) \qquad (j=1,2)$$

for all $x \in (0, a]$. *Then we have limit-point case (l.p.c.) for* (1) *at* $x=0$ *if* $c_1 c_2 < k^2 - \frac{1}{4}$. *The constant* $k^2 - \frac{1}{4}$ *is sharp.*

Incidentally we note that Theorem 1 is an immediate consequence of a well-known theorem on perturbed linear systems, at least if we furthermore assume $\tilde{q}_j \in C^0(0, a]$. In fact, change of the independent variable by setting $t = -\log x$, subsequent application of Theorem 5.4.5 in [4], p. 168 (see also [3, p. 311 f.] where much more general systems than (2) are dealt with), and resubstitution show that (2) with $\lambda = 0$ has a fundamental system of the form

$$u_{\pm}(x) := x^{\pm\sqrt{k^2 - c_1 c_2}} \left\{ \begin{pmatrix} 1 \\ \dfrac{k \mp \sqrt{k^2 - c_1 c_2}}{-c_2} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} o(1) \right\} \qquad (x \in (0, a])$$

whence the assertion of Theorem 1 follows. Besides, this result is unaffected by integrable perturbations of the separation term k/x.

By the method indicated in the introduction we shall prove the following limit-point criterion.

Theorem 2. *Suppose there exist constants* $c_1, c_2 \in \mathbb{R}$ *and real-valued functions* $\tilde{q}_1, \tilde{q}_2 \in C^0(0, a]$ *with*

$$\tilde{q}_1 - \tilde{q}_2 \in L^1(0, a] \qquad \textit{and} \qquad \lim_{x \to 0+} x\, \tilde{q}_j(x) = 0 \qquad (j=1,2)$$

such that

$$q_j(x) = \frac{c_j}{x} + \tilde{q}_j(x) \qquad (j=1,2)$$

for all $x \in (0, a]$. *Then we have l.p.c. for* (1) *at* $x=0$ *if* $c_1 c_2 < k^2 - \frac{1}{4}$.

Remark 1. *The potential*

$$q_1(x) = q_2(x) = \frac{c}{x} + \frac{1}{x \log x} \qquad (x \in (0, \tfrac{1}{2}],\ c \in \mathbb{R})$$

satisfies the conditions of our but not those of Titchmarsh's theorem. On account of our regularity condition, there are clearly also examples for which Titchmarsh's but not our assumptions hold.

---

[3] Titchmarsh himself has $c_1 = c_2 =: \alpha$ though this is not necessary for his argument.

Proof of Theorem 2. Suppose $c_2 \neq 0$, $\Delta := k^2 - c_1 c_2 > 0$. According to Theorem 3 of the appendix there exist solutions y_1, y_2 of (5) such that

$$y_{1,2}(0) := \lim_{x\to 0+} y_{1,2}(x) = -\frac{k}{c_2} \pm \frac{1}{|c_2|}\sqrt{\Delta}.$$

Then $\Theta_{1,2} := \arctan y_{1,2}$ are solutions of (3) satisfying

$$\begin{aligned}\lim_{x\to 0+} \cos 2\Theta_{1,2}(x) &= \frac{1-y_{1,2}^2(0)}{1+y_{1,2}^2(0)} \\ &= \frac{c_2^2 - c_1^2 \pm 4k \cdot \operatorname{sign} c_2 \cdot \sqrt{\Delta}}{4k^2 + (c_2 - c_1)^2},\end{aligned} \tag{6}$$

$$\begin{aligned}\lim_{x\to 0+} \frac{1}{2} \sin 2\Theta_{1,2}(x) &= \frac{y_{1,2}(0)}{1+y_{1,2}^2(0)} \\ &= \frac{-k(c_1+c_2) \mp (c_1 - c_2)\cdot \operatorname{sign} c_2 \cdot \sqrt{\Delta}}{4k^2 + (c_2 - c_1)^2}.\end{aligned} \tag{7}$$

Thus (3) has a solution Θ_+ for which

$$\lim_{x\to 0+} \left(\frac{c_1 - c_2}{2} \sin 2\Theta_+(x) - k \cos 2\Theta_+(x) \right) = \sqrt{\Delta}$$

holds, no matter whether $c_2 > 0$ or $c_2 < 0$. Consequently there exists for every $\varepsilon > 0$ an $x_0 \in (0, a]$ such that

$$\frac{c_1 - c_2}{2} \sin 2\Theta_+(x) - k \cos 2\Theta_+(x) \geqq \sqrt{\Delta} - \varepsilon$$

for all $x \in (0, a]$. Hence

$$\begin{aligned}\rho_+^2(x) &= \rho_+^2(x_0) \exp 2 \int_x^{x_0} \left\{ [q_1(t) - q_2(t)] \frac{1}{2} \sin 2\Theta_+(t) - \frac{k}{t} \cos 2\Theta_+(t) \right\} dt \\ &\geqq \rho_+^2(x_0) \exp\left\{ - \int_x^{x_0} |\tilde{q}_1(t) - \tilde{q}_2(t)|\, |\sin 2\Theta_+(t)|\, dt \right\} . \\ &\quad \cdot \exp\left\{ 2 \int_x^{x_0} \left[\frac{c_1 - c_2}{2} \sin 2\Theta_+(t) - k \cos 2\Theta_+(t) \right] \frac{dt}{t} \right\} \\ &\geqq \text{const} \left(\frac{x_0}{x} \right)^{2(\sqrt{\Delta} - \varepsilon)}\end{aligned}$$

for the corresponding solution ρ_+ of (4). If $c_1 c_2 < k^2 - \frac{1}{4}$ then $\sqrt{\Delta} > \frac{1}{2}$. Choosing $\varepsilon = \dfrac{2\sqrt{\Delta} - 1}{2}$ we obtain

$$\rho_+^2(x) \geqq \frac{\text{const}}{x},$$

which means that (2) with $\lambda=0$ possesses a non-square integrable solution.

In the case $c_2=0$ the corollary to Theorem 3 establishes the existence of solutions y_1, y_2 of (5) satisfying

$$\lim_{x\to 0+} y_1(x) = -\frac{c_1}{2k},$$

$$\liminf_{x\to 0+} |y_2(x)| = +\infty .$$

The corresponding solutions $\Theta_{1,2}$ of (3) have the property

$$\lim_{x\to 0+} \cos 2\Theta_1(x) = \frac{1-\left(\frac{c_1}{2k}\right)^2}{1+\left(\frac{c_1}{2k}\right)^2},$$

$$\lim_{x\to 0+} \frac{1}{2} \sin 2\Theta_1(x) = \frac{-\frac{c_1}{2k}}{1+\left(\frac{c_1}{2k}\right)^2},$$

$$\lim_{x\to 0+} \cos 2\Theta_2(x) = -1,$$

$$\lim_{x\to 0+} \tfrac{1}{2} \sin 2\Theta_2(x) = 0,$$

i.e., (6) and (7) are again satisfied with $\operatorname{sign} c_2$ replaced by 1. Consequently we have l.p.c. at $x=0$.

Remark 2. *It is clear from the above proof that Theorem 2 holds true for systems slightly more general than* (2), *viz., differential expressions of the form* (1) *with* $\frac{k}{x}$ *replaced by* $\frac{k}{x}+f(x)$ *where* $f\in C^0(0,a]\cap L^1(0,a]$ *is real-valued and* $\lim_{x\to 0+} xf(x)=0$.

Appendix: A Result of Falkenhagen on Riccati's Equation

Falkenhagen [2] studied the Riccati equation

$$y' = [\alpha(x)+\beta(x)\,y+\gamma(x)\,y^2]\,F(x) \qquad (-\infty \leqq l < x \leqq x_0) \tag{8}$$

under the following conditions on its coefficients:

(i) $\alpha, \beta, \gamma, F \in C^0(0, x_0]$;

(ii) $\alpha_l := \lim_{x\to l+} \alpha(x)$, $\beta_l := \lim_{x\to l+} \beta(x)$, $\gamma_l := \lim_{x\to l+} \gamma(x) \neq 0$ *exist;*

(iii) sign $F = const.$ *near* l[4], $\int_l^{x_0} |F(x)|\, dx = +\infty$;

(iv) *for* $x = l$ *the roots* $Y_1(x)$, $Y_2(x)$ *of the equation*

$$\alpha(x) + \beta(x)\, y + \gamma(x)\, y^2 = 0 \tag{9}$$

are real and distinct (*let them be numbered such that* $Y_2(l) < Y_1(l)$).

The essential statement of his theorem is that in the case $\gamma_l \neq 0$, all solutions of (8) cannot but behave continuously at $x = l$ and tend to one of the two roots of the algebraic equation (9) on account of the strong singularity of F at l. In the case $\gamma_l = 0$ this dichotomy in the behaviour of the solutions still holds provided $\beta_l \neq 0$, but now at least one solution of (8) becomes infinite at l. (For the case that (9) has two complex roots at $x = l$ see Horn [5, p. 262–265] and Picard [8, p. 391 ff.]. If (9) has a double root at $x = l$—this is the most complicated case—we refer the reader to Falkenhagen's paper [2, p. 228–240].)

Theorem 3. *Under the assumptions* (i)–(iv) *there exists exactly one solution* $\tilde{Y}$ *of* (8) *for which*

$$\lim_{x \to l+} \tilde{Y}(x) = \begin{cases} Y_1(l) & \text{if } \gamma F < 0 \text{ near } l \\ Y_2(l) & \text{if } \gamma F > 0 \text{ near } l \end{cases}$$

holds whereas all other solutions Y *of* (8) *defined near* l *satisfy*

$$\lim_{x \to l+} Y(x) = \begin{cases} Y_2(l) & \text{if } \gamma F < 0 \text{ near } l \\ Y_1(l) & \text{if } \gamma F > 0 \text{ near } l. \end{cases}$$

Corollary. *Suppose* (i)–(iv) *hold with the modifications* $\beta_l \neq 0$, $\gamma_l = 0$. *Then there exists exactly one solution* $\tilde{Y}$ *of* (8) *for which*

$$\lim_{x \to l+} \tilde{Y}(x) = -\frac{\alpha_l}{\beta_l} \quad \text{if } \beta F < 0 \text{ near } l,$$

$$\liminf_{x \to l+} |\tilde{Y}(x)| = +\infty \quad \text{if } \beta F > 0 \text{ near } l$$

holds whereas all other solutions Y *of* (8) *satisfy*

$$\liminf_{x \to l+} |Y(x)| = +\infty \quad \text{if } \beta F < 0 \text{ near } l,$$

$$\lim_{x \to l+} Y(x) = -\frac{\alpha_l}{\beta_l} \quad \text{if } \beta F > 0 \text{ near } l.$$

---

[4] Falkenhagen himself (*loc. cit.*, p. 227) assumes $\liminf_{x \to l+} |F(x)| = +\infty$, but it is readily seen that the above assumption is sufficient for his proof of Theorem 3.

Remark 3. *If we consider* (8) *in the interval* $x_0 \leqq x < m \leqq +\infty$ *under the appropriately modified assumptions* (i)–(iv) *the assertions of Theorem* 3 *and its corollary remain valid if we replace* "$\gamma F \lesseqgtr 0$ *near* l" *and* "$\beta F \lesseqgtr 0$ *near* l" *by* "$\gamma F \gtreqless 0$ *near* m" *and* "$\beta F \gtreqless 0$ *near* m", *respectively.*

Since Falkenhagen's proof of Theorem 3 is rather lengthy [2, p. 216–228], a sketch of a slightly modified proof may be useful. The details can easily be carried out by the reader.

Because of (i)–(iv) it is possible to choose x_0 so close to l that

$$\sup_{x \in (l, x_0]} Y_2(x) < \inf_{x \in (l, x_0]} Y_1(x) =: m_1,$$

and either $\gamma(x) F(x) < 0$ or $\gamma(x) F(x) > 0$ for all $x \in (l, x_0]$. We shall assume $\gamma F < 0$ in the sequel. It is convenient to proceed in five steps.

1. *There exists a unique number* $\tilde{y}$ *with the following property: every solution* Y *of* (8) *with* $Y(x_0) > \tilde{y}$ $(Y(x_0) \leqq \tilde{y})$ *has a domain of definition properly contained in* $(l, x_0]$ (*equal to* $(l, x_0]$). (This can at once be read off from (8) written in the form $y' = [y - Y_1(x)]\,[y - Y_2(x)]\,\gamma(x) F(x)$.)

2. *For every solution* Y *of* (8) *defined near* l *the existence of* $Y_l := \lim_{x \to l+} Y(x)$ *implies* $Y_l = Y_1(l)$ *or* $Y_l = Y_2(l)$. (Integrate both sides of (8).)

3. *For every solution* Y *of* (8) *defined near* l $\lim_{x \to l+} Y(x)$ *exists.* (The possibilities $Y_- := \liminf_{x \to l+} Y(x) < \limsup_{x \to l+} Y(x) =: Y_+$ or $Y_- = Y_+ = -\infty$ are immediately ruled out by the directional field of (8). $Y_- = Y_+ = +\infty$ is impossible since $Z := (K - Y)^{-1}$ $(K \neq Y_1(l), Y_2(l))$ would be a solution of

$$y' = \{\gamma(x) - [\beta(x) + 2K\gamma(x)]\,y + [\alpha(x) + K\beta(x) + K^2\gamma(x)]\,y^2\}\,F(x) \qquad (x \in (l, x_0]) \tag{10}$$

(cf. Hille [4, p. 275 f.]) with $Z_l := \lim_{x \to l+} Z(x) = 0$ which contradicts 2.)

4. *The solution* $\tilde{Y}$ *of* (8) *with* $\tilde{Y}(x_0) = \tilde{y}$ *satisfies* $\lim_{x \to l+} \tilde{Y}(x) = Y_1(l)$. (It suffices to show $\tilde{Y}(x) \geqq m_1$ for all $x \in (l, x_0]$.)

5. *All solutions* Y *of* (8) *defined near* l *and different from* $\tilde{Y}$ *satisfy* $\lim_{x \to l+} Y(x) = Y_2(l)$. (This is clear for the solutions with $Y(x_0) < m_1$ and those solutions defined near l that tend to $-\infty$ in the interval $(l, x_0]$. For the solutions with $Y(x_0) \in [m_1, \tilde{y})$ the desired result follows from the Riccati equation for $Z := Y - \tilde{Y}$ after taking account of 2., 3., and 4.)

The corollary can be proved by applying Theorem 3 to the Riccati equation (10) with $K \neq -\dfrac{\alpha_l}{\beta_l}$ and subsequent transformation back to (8).

References

1. Dunford, N., Schwartz, J.T.: Linear operators. Part II. 3rd printing. New York-London-Sydney: Interscience 1967.
2. Falkenhagen, J.H.M.: Über das Verhalten der Integrale einer Riccatischen Gleichung in der Nähe einer singulären Stelle. Nieuw Arch. Wisk. (2) **6**, 209–248 (1905).
3. Hartman, Ph.: Ordinary differential equations. New York-London-Sydney: John Wiley 1964.
4. Hille, E.: Lectures on ordinary differential equations. Reading, MA: Addison-Wesley 1969.
5. Horn, J.: Über das Verhalten der Integrale von Differentialgleichungen bei der Annäherung der Veränderlichen an eine Unbestimmtheitsstelle. J. Reine Angew. Math. **118**, 257–274 (1897).
6. Kamke, E.: Differentialgleichungen, Lösungsmethoden und Lösungen. Bd. I. 7. Aufl. Leipzig: Akademische Verlagsgesellschaft 1961.
7. Petrovitch, M.: Remarques algébriques sur les fonctions définies par les équations différentielles du premier ordre. Bull. Soc. Math. France **24**, 58–80 (1896).
8. Picard, E.: Traité d'analyse. T. III. 2ième éd. Paris: Gauthiers-Villars 1908.
9. Titchmarsh, E.C.: On the nature of the spectrum in problems of relativistic quantum mechanics. Quart. J. Math. Oxford, Ser. **12**, 227–240 (1961).
10. Weidmann, J.: Oszillationsmethoden für Systeme gewöhnlicher Differentialgleichungen. Math. Z. **119**, 349–373 (1971).

Dr. Hubert Kalf
Institut für Mathematik
der Rheinisch-Westfälischen
Technischen Hochschule
D-5100 Aachen
Templergraben 55
Federal Republic of Germany

(Received August 10, 1972)

Math. Z. 129, 83–94 (1972)
© by Springer-Verlag 1972

Absolute Continuity of the Essential Spectrum of $-\frac{d^2}{dt^2}+q(t)$ without Monotony of q

Johann Walter

1. Introduction

Let q be a twice continuously differentiable real function defined in $[0, \infty)$ such that $q(\infty) := \lim_{t\to\infty} q(t)$ exists. We consider the singular Sturm-Liouville differential equation

$$-u''+q\,u=\lambda\,u, \qquad 0\leqq t<\infty \tag{1}$$

and the corresponding differential operator H defined by

$$\begin{gathered} H\,u=-u''+q\,u, \qquad u\in D(H), \\ D(H)=\{u|u, u' \text{ locally absolutely continuous in } [0,\infty); \\ u(0)\cos\alpha-u'(0)\sin\alpha=0;\ u,\ -u''+q\,u\in L^2(0,\infty)\}. \end{gathered} \tag{2}$$

Operators of this kind are used in quantum mechanics as energy operators (Schroedinger operators) of hydrogen-like one particle-systems where they are required to be selfadjoint. H is for example selfadjoint if in the case $q(\infty)=-\infty$

$$\int^{\infty} \frac{dt}{\sqrt{|q|}}=\infty \tag{3}$$

and

$$\int^{\infty} |Q(t,\lambda)|\,dt<\infty\,^{1} \qquad \text{for some } \lambda\in(q(\infty),\infty), \tag{4}$$

([5; p. 1409, Th. 20] with $\lambda=0$).

It is the aim of this paper to show that all regularity properties of the spectrum of H usually required in applications can be deduced from (3) and (4) without any additional assumption. This is especially interesting in the case $q(\infty)=-\infty$ which at present could only be

[1] Here we have used the abbreviation

$$Q(t,\lambda):=\frac{1}{4}\,\frac{q''}{(\lambda-q)^{\frac{3}{2}}}+\frac{5}{16}\,\frac{(q')^2}{(\lambda-q)^{\frac{5}{2}}}.$$

treated by assuming monotony of q. (For $q(\infty) > -\infty$ better results are given in [20; p. 311, Satz 5], [21; p. 118], [23; p. 293, Satz 5.1].)

Remark 1. It is not self-evident but follows immediately from a lemma of Atkinson-Coppel ([1; p. 108], [4; p. 121, Lemma 6]) that (4) implies

$$\int^{\infty} |Q(t,\lambda)|\, dt < \infty \qquad \text{for all } \lambda \in (q(\infty), \infty). \tag{5}$$

With the help of the lemma just mentioned one proves first that for the same λ as in (4)

$$\int^{\infty} \left(\frac{|q''|}{(\lambda - q)^{\frac{3}{2}}} + \frac{|q'|^2}{(\lambda - q)^{\frac{5}{2}}} \right) dt < \infty \tag{6}$$

and

$$\lim_{t \to \infty} \left(\frac{1}{(\lambda - q)^{\frac{1}{2}}} \right)' = 0. \tag{7}$$

(6) obviously implies (5). (7) will be used later. (The relations (5), (6) and (7) even hold locally uniform for $\lambda \in (q(\infty), \infty)$.)

In all examples of quantum mechanical problems with specific "potentials" q the structure of the spectrum of H has turned out to be sufficiently simple. More precisely the energy operator H of a hydrogen-like one-particle-system should satisfy the following three requirements:

(i) The essential spectrum $\sigma_e(H)$ should coincide with the closure of $(q(\infty), \infty)$.

(ii) Eigenvalues larger than $q(\infty)$ should not exist.

(iii) It should be possible to avoid the use of the Stieltjes integral in the expansion formula for the interval $(q(\infty), \infty)$. Or equivalently: the spectral function should be absolutely continuous in the same interval.

Remark 2. The requirements (i) and (ii) can easily be motivated by spectroscopic experience (cf. [17; Sec. 30]). Requirement (iii) is less restrictive than it seems to be at first sight. As a matter of fact the invalidity of (iii) has never been seriously taken into account (cf. e.g. the special character of the linear aggregates used for the computation of the spectral family or of the formulas used for the normalization of the socalled eigendifferentials in [3; p. 584], [11; p. 153, 215], [17; p. 164], [19; p. 68]). Moreover (iii) holds in all physically relevant applications (this has been stressed recently by several authors, cf. [7; p. 211], [16; p. 518], [18; p. 77], [23; p. 269]). In this connection it is also interesting to note that Titchmarsh [21; p. 123, Th. 5.10] and Neumark [20; p. 331, Satz 9] actually prove (iii) (this can be seen from [21; p. 125, line 1] and [20; p. 331, (84)]) while only asserting (ii).

Now the problem arises to find sufficiently mild conditions on q which imply (i), (ii) and (iii).

Let $\sigma_d(H)$ denote the subset of all isolated elements of the spectrum $\sigma(H)$ of H and $\sigma_0(H)$ the set of all real λ such that (1) has no nontrivial solution in $L^2(0,\infty)$. Then $\sigma_e(H):=\sigma(H)\setminus\sigma_d(H)$ is called the essential spectrum of H. (In our case this purely set-theoretic definition of the essential spectrum is equivalent to the usual one since eigenvalues of infinite multiplicity do not exist.)

Ad (i). Because of the nonoscillatory character of (1) for $\lambda\notin[q(\infty),\infty)$ we have

$$\sigma_e(H)\subset[q(\infty),\infty), \tag{8}$$

[5; p. 1464, Lemma 39]. On the other hand

$$\sigma_0(H)\subset\sigma_e(H), \tag{9}$$

[23; p. 279, Satz 3.1 (b)]. For $q(\infty)>-\infty$ we even have $\sigma_e(H)=[q(\infty),\infty)$ [5; p. 1448, Th. 16]. In the case $q(\infty)=-\infty$ the relation $\sigma_e(H)=(-\infty,\infty)$ holds if $q'=o(\sqrt{|q|})$ and

$$q'(t)<0, \qquad \operatorname{sign} q''(t)=\text{const} \tag{10}$$

are assumed [7; p. 116, Th. 28]. This result has been considerably sharpened by Hinton [14] who has recently shown that the monotony condition (10) can be dropped. Note that Hinton's condition $q'=o(\sqrt{|q|})$ is not implied by and does not imply our conditions (3), (4). This can be seen from the examples $q=-t^2$, $q=-t+\frac{\sin t^2}{t}$.

Let H_s and H_p denote the spectrally singular and discontinuous part of H respectively (cf. [16; p. 514–518], and define[2]

$$\sigma_c(H):=(\sigma_e(H)\setminus\sigma(H_p))^0, \qquad \sigma_{ac}(H):=(\sigma_e(H)\setminus\sigma(H_s))^0$$

(cf. [23; p. 271]). Under assumption (i) requirement (ii) is equivalent to

$$\sigma_c(H)=(q(\infty),\infty) \tag{11}$$

and (iii) is equivalent to

$$\sigma_{ac}(H)=(q(\infty),\infty). \tag{12}$$

By abuse of language we call the spectrum of H "continuous" resp. "absolutely continuous" in $(q(\infty),\infty)$ if (11) or (12) hold.

Ad (ii). Let $t_0(\lambda)$ be a continuous function defined in $(q(\infty),\infty)$ such that $q(t)<\lambda$ for $t_0(\lambda)\leqq t$ and assume (5). Then it is known (cf. [2; p. 31, § 1.22], [4; p. 122, Th. 14(i)], [10; p. 371, Ex. 8.3(a)]) that for every $\lambda\in(q(\infty),\infty)$ the relation

$$u(t)=\rho(\lambda-q)^{-\frac{1}{4}}\left\{\sin\left(\int_{t_0(\lambda)}^{t}(\lambda-q)^{\frac{1}{2}}\,d\tau+\varphi\right)+o(1)\right\} \tag{13}$$

[2] M^0 is the interior of M.

defines a one-to-one correspondence between the set of all nontrivial real solutions of (1) and the set of all pairs $\{\rho, \varphi\} \in (0, \infty) \times [0, 2\pi)$. Together with (3) (in the case $q(\infty) = -\infty$) this obviously implies that *at least one* solution of (1) is not in $L^2(0, \infty)$. In order to prove (11) (for every α occuring in (2)), however, it is necessary to show that for every $\lambda \in (q(\infty), \infty)$ *no* nontrivial solution of (1) is in $L^2(0, \infty)$. For $q(\infty) > -\infty$ this is an immediate consequence of (13) too. But for $q(\infty) = -\infty$ it cannot be excluded prima vista that for a suitable choice of $\varphi \in [0, 2\pi)$ the solution (13) is an element of $L^2(0, \infty)$. If q is *monotone*, the nonexistence of a nontrivial square integrable solution of (1) follows from (3) and (13) by means of a criterion due to Hardy[3] [8] (cf. also [6; p. 594, Nr. 6]). This result is essentially contained in [5; p. 1412, Corollary 21(b)]. But the proof is not complete because only (4) is assumed (l.c., p. 1410, Th. 20, assumption (b)) although (5) is needed in the proof (p. 1414, line 2 and 3).

We shall derive the same result for q not necessarily monotone (our Theorem 2) by using a new inequality (Theorem 1) which may be of independent interest. This inequality can be considered as an inversion of inequalities like

$$(c, d) \subset (a, b) \Rightarrow \int_c^d |f|\, dt \leqq \int_a^b |f|\, dt .$$

Its proof is based on an inversion of Schwarz's inequality we have derived in [22; Satz 2].

Ad (iii). In the mainly interesting case $q(\infty) = -\infty$ Titchmarsh [21; p. 123, Th. 5.10] has proved (12) under the assumptions (3), (10) and

$$q' = O(|q|^\alpha) \qquad \text{for some } \alpha \in (0, \tfrac{3}{2}). \tag{14}$$

In his proof he makes use of complex values of λ and gets (12) by letting $\operatorname{Im}(\lambda) \to 0$. A strictly "real" proof of (12) under the same assumptions has been given by Neumark [20; p. 331, Satz 9]. Titchmarsh needs the monotony assumption (10) *only once*, namely to infer from (14) that the convergence of (5) is locally uniform in λ [21; p. 121]. Neumark needs (10) *a first time* [20; p. 287] to obtain (6) from (14). In both cases the conditions (10) and (14) can be replaced by (4) on account of Remark 1. But Neumark uses the monotony of q at a crucial point *a second time* [20; p. 330, line 1].

This second use cannot be eliminated as easily as in the former two cases.

It is the purpose of this paper to show (Theorem 3) that the abovementioned inequality allows to give a "real" proof of (12) in which the monotony-assumption for q can be dropped.

[3] This was pointed out to the author by H. Kalf, Aachen.

2. Asymptotic Behaviour and Explicit Bounds for the Solutions of (1)

As we have seen it is sufficient for (11) that the property of (1) of not having a nontrivial solution in $L^2(0,\infty)$ holds pointwise in the λ-interval $(q(\infty),\infty)$. In contrast to this the relation (12) requires certain properties to hold uniformly with respect to λ. More precisely we have (12) if for every finite interval $[\lambda_1,\lambda_2]\subset(q(\infty),\infty)$ and for every $u\in L^2(0,\infty)$ the function $\langle E_\lambda u,u\rangle$ is absolutely continuous in $[\lambda_1,\lambda_2]$.[4]

(For this it is sufficient to show that a constant K exists such that

$$\langle (E_{\lambda''}-E_{\lambda'})\,u,u\rangle \leqq K(\lambda''-\lambda') \tag{15}$$

for all $\lambda',\lambda''\in[\lambda_1,\lambda_2]$, $\lambda'\leqq\lambda''$.)

Information concerning uniformity with respect to λ cannot be deduced from (13) until the λ-dependence of $o(1)$ is made explicit. For the convenience of the reader this more explicit information will be given in the following.

Estimates Near Infinity. Let matrices B, Y and a function F be defined by

$$B^{\pm 1}=\begin{pmatrix}(\lambda-q)^{\pm\frac14} & 0\\ \pm\frac14(\lambda-q)^{-\frac54}\cdot(\lambda-q)' & (\lambda-q)^{\mp\frac14}\end{pmatrix},\qquad Y=\begin{pmatrix}\cos s & \sin s\\ -\sin s & \cos s\end{pmatrix},$$
$$F(t,\lambda)=\int_t^\infty |Q(\tau,\lambda)|\,d\tau\cdot\exp\left\{\int_{t_0(\lambda)}^\infty |Q(t,\lambda)|\,dt\right\}. \tag{16}$$

Let u be a solution of (1); then $\begin{pmatrix}u\\u'\end{pmatrix}$ is a solution of

$$v'=\begin{pmatrix}0 & 1\\ q-\lambda & 0\end{pmatrix}v. \tag{17}$$

In the interval $[t_0(\lambda),\infty)$ we introduce new independent and dependent variables by

$$s=\int_{t_0(\lambda)}^t (\lambda-q)^{\frac12}\,d\tau,\qquad x=Y\overset{-1}{B}v$$

(Liouville transformation and variation of constants). Derivations with respect to the variable s are indicated by a dot. Let $t(s)$ be the inverse function of $s(t)$. A simple calculation shows that

$$\dot x=Q(t,\lambda)\,(\lambda-q)^{-\frac12}\begin{pmatrix}\sin s\cos s & \sin^2 s\\ -\cos^2 s & -\sin s\cos s\end{pmatrix}x. \tag{18}$$

---

[4] E_λ denotes the spectral family of H;

$$\langle u,v\rangle := \int_0^\infty u\bar v\,dt.$$

Because of (4) there exists a fundamental matrix X of (18) such that

$$|x_{ik}(s)-\delta_{ik}|\leqq F(t(s),\lambda) \quad \text{for } s\in[0,\infty),\ i,k=1,2$$

(cf. [10; p. 275, (1.14)]). The matrix $\overset{-1}{B}YX$ is a fundamental matrix of (17) in $[t_0(\lambda),\infty)$. Hence it follows from (16) that for every nontrivial real solution of (1) there exists exactly one pair $\{\rho,\varphi\}\in(0,\infty)\times[0,2\pi)$ and functions $f_i(t,\lambda)$, $i=1,2,3$ such that

$$u(t)=\rho(\lambda-q)^{-\frac{1}{4}}\left\{\sin\left(\int_{t_0(\lambda)}^{t}(\lambda-q)^{\frac{1}{2}}\,d\tau+\varphi\right)+f_1(t,\lambda)\right\}, \tag{19}$$

$$u'(t)=\rho(\lambda-q)^{\frac{1}{4}}\left\{\cos\left(\int_{t_0(\lambda)}^{t}(\lambda-q)^{\frac{1}{2}}\,d\tau+\varphi\right)+f_2(t,\lambda)+f_3(t,\lambda)\right\}, \tag{20}$$

$$|f_i(t,\lambda)|\leqq 4F(t,\lambda), \qquad i=1,2, \tag{21}$$

$$|f_3(t,\lambda)|\leqq\left|\left(\frac{1}{(\lambda-q)^{\frac{1}{4}}}\right)'\right|(1+2F(t,\lambda)) \tag{22}$$

for $t\in[t_0(\lambda),\infty)$.

Estimates Near the Origin. (19), (20), (21) and (22) give estimates for $u(t)$ near infinity. In the following an estimate for $u(t)$ near the origin is also needed. Let t_1 be any positive number.

For $t\in[0,t_1]$ the inequality

$$|u(t)|\leqq\{|u(t_1)|+|u'(t_1)|\}\cdot\exp\left\{\int_0^{t_1}(1+|\lambda-q|)\,dt\right\} \tag{23}$$

can easily be derived (cf. [10; p. 54, Lemma 4.1]).

3. An Inequality for Positive Functions

Theorem 1. *Let g be a continuously differentiable positive function defined in some interval $[t_0,\infty)$. Let a,b,c,x,y,z,w be numbers with $0<a\leqq b$, $t_0\leqq x\leqq y\leqq z\leqq w$ such that*

$$\int_y^z\frac{dt}{g}=a, \tag{24}$$

$$\int_x^w\frac{dt}{g}\leqq b, \tag{25}$$

$$|g'(t)|\leqq c \qquad \text{for } t\in[t_0,\infty). \tag{26}$$

We define

$$h(a,b,c):=\left(\frac{a}{c}\left[\left(\frac{1}{a}+c\right)^2\exp\{c(b-a)\}-\frac{1}{a^2}\right]\right)^2\left(\frac{1}{a}+c\right)a. \tag{27}$$

Then we have

$$\int_x^w g\,dt \leqq h(a,b,c)\int_y^z g\,dt. \tag{28}$$

Remark 3. An inequality of type (28) does not seem to have been formulated as yet. But a qualitative analysis of Theorem 1 shows that comparable results exist. Qualitatively Theorem 1 states that on intervals the length of which is roughly speaking proportional to g the variation of g is small in a certain sense. In other words the ratio

$$\frac{g(t+\text{const.}\,g(t))}{g(t)}$$

does not differ too much from 1. Results of this kind can be found in Hartman [9] and Hinton [13; p. 127, (6)] (cf. also [10; p. 349, Ex. 5.6(b)], [12; p. 473], [15; p. 129]).

Proof of Theorem 1. From the mean value theorem and (24) we obtain the existence of a number $\xi\in(y,z)$ such that

$$g(\xi)=\frac{1}{a}(z-y).$$

(26) then implies

$$g(t)\leqq\begin{cases}\frac{1}{a}(z-y)+c(z-t), & t\in(x,y)\\ \frac{1}{a}(z-y)+c(t-y), & t\in(z,w).\end{cases}$$

Because of (24) and (25) we have

$$\int_x^y \frac{dt}{\frac{z-y}{a}+c(z-t)}+\int_z^w \frac{dt}{\frac{z-y}{a}+c(t-y)}\leqq b-a,$$

which gives after simple calculations

$$\frac{w-x}{z-y}\leqq\frac{a}{c}\left[\left(\frac{1}{a}+c\right)^2\exp\{c(b-a)\}-\frac{1}{a^2}\right]. \tag{29}$$

In [22; Satz 2] we have proved an inversion of Schwarz's inequality a special case of which is

$$\begin{aligned}\int_x^w g\,dt &\leqq\left\{\left(\int_x^w\frac{dt}{g}\right)^{-1}+\sup_{x\leqq t\leqq w}|g'(t)|\right\}(w-x)^2\\ &\leqq\left\{\frac{1}{a}+c\right\}(w-x)^2.\end{aligned} \tag{30}$$

On the other hand Schwarz's inequality yields

$$\int_y^z g\,dt \geqq \left(\int_y^z \frac{dt}{g}\right)^{-1} (z-y)^2 = \frac{1}{a}(z-y)^2. \tag{31}$$

(28) now follows from (27), (29), (30) and (31).

4. Continuity of the Spectrum

Theorem 2. *Under the assumptions* (3) *and* (4) *the spectrum of* H *is continuous in* $(q(\infty), \infty)$.

Proof. We have to show that for every $\lambda \in (q(\infty), \infty)$ and every $\varphi \in [0, 2\pi)$ the solution u of (1) represented by (19) is not in $L^2(0, \infty)$. Suppose $\lambda \in (q(\infty), \infty)$ and $\varphi \in [0, 2\pi)$. Without loss of generality we may assume that the ρ appearing in (19) is normalized by

$$\rho = 1. \tag{32}$$

From (19) we obtain

$$|u(t)|^2 \geqq \tfrac{1}{2}(\lambda-q)^{-\frac{1}{2}} \sin^2\left(\int_{t_0(\lambda)}^t (\lambda-q)^{\frac{1}{2}}\,d\tau + \varphi\right) - (\lambda-q)^{-\frac{1}{2}}\,|f_1(t,\lambda)|^2 \tag{33}$$

for $t \in [t_0(\lambda), \infty)$. Now we use Theorem 1 with

$$a = \frac{\pi}{4}, \quad b = \pi, \quad c = 1, \quad g = (\lambda - q)^{-\frac{1}{2}},$$

writing $M_1 := h\left(\frac{\pi}{4}, \pi, 1\right)$.

In view of (7) and (21) it is possible to define in $(q(\infty), \infty)$ a continuous function $t_1(\cdot)$ with $t_0(\lambda) \leqq t_1(\lambda)$ for $\lambda \in (q(\infty), \infty)$ such that

$$\left|\left(\frac{1}{(\lambda-q)^{\frac{1}{2}}}\right)'\right| \leqq 1, \tag{34}$$

$$|f_1(t,\lambda)| \leqq \frac{1}{8M_1} \tag{35}$$

for $t \in [t_1(\lambda), \infty)$. Let $\omega \in [t_1(\lambda), \infty)$ be sufficiently large. It is then possible to define a partition

$$t_1(\lambda) =: x_1 \leqq y_1 \leqq z_1 \leqq x_2 \leqq \cdots \leqq x_n := \omega$$

of the interval $[t_1(\lambda), \omega]$ such that

$$\int_{x_k}^{x_{k+1}} (\lambda-q)^{\frac{1}{2}}\,dt \leqq \pi, \qquad \int_{y_k}^{z_k} (\lambda-q)^{\frac{1}{2}}\,dt = \frac{\pi}{4},$$

$$\sin^2\left(\int_{t_0(\lambda)}^t (\lambda-q)^{\frac{1}{2}}\,d\tau + \varphi\right) \geqq \tfrac{1}{2} \quad \text{for } t \in (y_k, z_k)$$

$(k=1, \ldots, n-1)$. From this, Theorem 1, and (34) we infer

$$\begin{aligned}\frac{1}{4M_1}\int_{t_1(\lambda)}^{\omega}\frac{dt}{(\lambda-q)^{\frac{1}{2}}} &= \frac{1}{4M_1}\sum_{k=1}^{n-1}\int_{x_k}^{x_{k+1}}\frac{dt}{(\lambda-q)^{\frac{1}{2}}}\\ &\leqq \tfrac{1}{4}\sum_{k=1}^{n-1}\int_{y_k}^{z_k}\frac{dt}{(\lambda-q)^{\frac{1}{2}}}\\ &\leqq \tfrac{1}{2}\sum_{k=1}^{n-1}\int_{y_k}^{z_k}\frac{1}{(\lambda-q)^{\frac{1}{2}}}\sin^2\left(\int_{t_0(\lambda)}^{t}(\lambda-q)^{\frac{1}{2}}\,d\tau+\varphi\right)dt.\end{aligned}$$

Replacing the integration intervals (y_k, z_k) by (x_k, x_{k+1}) on the right hand of the last inequality we finally obtain [5]

$$\frac{1}{4M_1}\int_{t_1(\lambda)}^{\omega}\frac{dt}{(\lambda-q)^{\frac{1}{2}}} \leqq \frac{1}{2}\int_{t_1(\lambda)}^{\omega}\frac{1}{(\lambda-q)^{\frac{1}{2}}}\sin^2\left(\int_{t_0(\lambda)}^{t}(\lambda-q)^{\frac{1}{2}}\,d\tau+\varphi\right)dt. \tag{36}$$

Integrating (33) over $[t_1(\lambda), \omega]$ and using (35) and (36) we get

$$\int_{t_1(\lambda)}^{\omega} u^2(t)\,dt \geqq \frac{1}{8M_1}\int_{t_1(\lambda)}^{\omega}\frac{dt}{(\lambda-q)^{\frac{1}{2}}}. \tag{37}$$

Letting $\omega\to\infty$ the assertion of Theorem 2 follows because of (3)[6].

5. Absolute Continuity of the Spectrum

Theorem 3. *Under the conditions of Theorem 2 the spectrum of H is absolutely continuous in $(q(\infty), \infty)$.*

Proof. It is sufficient to show (15) for every $u\in L^2(0,\infty)$ with compact support in $[0,\infty)$ (cf. [23; p. 295, Lemma 5.5]).

Let $[\lambda_1, \lambda_2]$ be an interval in $(q(\infty), \infty)$ and u a function with compact support. Let M_2 be a number such that

$$u(t)=0 \quad \text{for } t\in[M_2, \infty). \tag{38}$$

We consider a sequence of selfadjoint operators H_m with $D(H_m)=\{v\in L^2(0,\infty)\,|\, v, v'$ *absolutely continuous in* $[0,m]$; $-v''+qv\in L^2(0,m)$; $v(0)\cos\alpha - v'(0)\sin\alpha = v(m)=0\}$ and

$$(H_m v)(t)=\begin{cases}-v''(t)+q(t)\,v(t), & 0\leqq t\leqq m\\ 0, & m<t\end{cases}$$

[5] If q is monotone (36) is a direct consequence of the above-mentioned criterion of Hardy.

[6] Another (indirect) proof of Theorem 2 can be based on a result of Hille: Let u be a solution of (1) in $L^2(0,\infty)$. Then by a theorem of Levinson (cf. [12; p. 503, (10.1.27)]) $(\lambda-q)^{-\frac{1}{4}}u'\in L^2(0,\infty)$ too. This is not compatible with (3) because of [12; p. 472, (9.5.6)]. It would be interesting to know if by this method a proof of Theorem 3 can also be given.

for $v\in D(H_m)$, $m=1,2,3,\ldots$. Let E_λ^m be the spectral family of H_m. Then for every $v\in L^2(0,\infty)$ and for every $\lambda\in(q(\infty),\infty)$

$$\lim_{m\to\infty}\langle E_\lambda^m v, v\rangle=\langle E_\lambda v, v\rangle$$

(cf. [23; p. 273, Satz 1.2]). Because of the regularity of the corresponding boundary value problems the spectra $\sigma(H_m)$ are purely discrete. Let λ_{mn} be the n-th eigenvalue of H_m and v_{mn} the corresponding eigenfunction. Evidently we have $v_{mn}(t)=0$ for $t\geqq m$. Without loss of generality we may assume v_{mn} to be normalized as indicated in (32). We define

$$t_2:=\max_{\lambda\in[\lambda_1,\lambda_2]} t_1(\lambda).$$

Applying Theorem 1 as in (36) and (37) we obtain for $\lambda_{mn}\in[\lambda_1,\lambda_2]$ and for sufficiently large $m\geqq t_2$

$$\|v_{mn}\|^2=\int_0^m v_{mn}^2\,dt\geqq\int_{t_1(\lambda_{mn})}^m v_{mn}^2\,dt\geqq\frac{1}{8M_1}\int_{t_2}^m\frac{dt}{(\lambda_{mn}-q)^{\frac{1}{2}}}. \tag{39}$$

From (19), (20), (21), (22), (23) and (32) follows the existence of a constant M_3 such that

$$\lambda_{mn}\in[\lambda_1,\lambda_2]\Rightarrow|v_{mn}(t)|\leqq M_3\quad\text{for } t\in[0,\infty). \tag{40}$$

Now assume $\lambda',\lambda''\in[\lambda_1,\lambda_2]$, $\lambda'\leqq\lambda''$. On account of the spectral theorem we have

$$\langle(E_{\lambda''}^m-E_{\lambda'}^m)u,u\rangle=\sum_{\lambda'<\lambda_{mn}\leqq\lambda''}\left|\left\langle u,\frac{v_{mn}}{\|v_{mn}\|}\right\rangle\right|^2. \tag{41}$$

We define

$$K:=\frac{1}{\pi}M_1 M_3^2\left(\int_0^{M_2}|u|\,dt\right)^2. \tag{42}$$

From (38), (39), (40), (41) and (42) we deduce

$$\langle(E_{\lambda''}^m-E_{\lambda'}^m)u,u\rangle\leqq\pi K\left(\int_{t_2}^m\frac{dt}{(\lambda''-q)^{\frac{1}{2}}}\right)^{-1}\sum_{\lambda'<\lambda_{mn}\leqq\lambda''}1. \tag{43}$$

Let $N(t',t'',\lambda)$ denote the maximum number of zeros of a real non-trivial solution of (1) in $[t',t'']$. Then it is known that

$$\left|N(t_2,m,\lambda)-\frac{1}{\pi}\int_{t_2}^m(\lambda-q)^{\frac{1}{2}}\,dt\right|\leqq 1+\frac{1}{\pi}\int_{t_2}^m|Q(t,\lambda)|\,dt \tag{44}$$

(cf. [10; p. 348, Ex. 5.5(b)]),

$$N(0,t_2,\lambda)\leqq 1+\frac{t_2}{\pi}\sup_{0\leqq t\leqq t_2}|\lambda-q(t)|^{\frac{1}{2}} \tag{45}$$

(Sturm's comparison theorem) and

$$\sum_{\lambda'<\lambda_{mn}\leqq\lambda''} 1 \leqq N(0,m,\lambda'')-N(0,m,\lambda')+1 \tag{46}$$
$$\leqq N(0,t_2,\lambda'')-N(0,t_2,\lambda')+N(t_2,m,\lambda'')-N(t_2,m,\lambda')+2$$

(cf. [21; p. 143]). (44), (45) and (46) imply the existence of a constant M_4 with

$$\sum_{\lambda'<\lambda_{mn}\leqq\lambda''} 1 \leqq M_4+\frac{1}{\pi}\int_{t_2}^{m}(\lambda''-q)^{\frac{1}{2}}\,dt-\frac{1}{\pi}\int_{t_2}^{m}(\lambda'-q)^{\frac{1}{2}}\,dt$$
$$\leqq M_4+\frac{\lambda''-\lambda'}{\pi}\int_{t_2}^{m}\frac{dt}{(\lambda''-q)^{\frac{1}{2}}}.$$

Inserting this into (43) taking (3) into consideration and letting $m\to\infty$ we finally get

$$\langle (E_{\lambda''}-E_{\lambda'})\,u,u\rangle \leqq K\cdot(\lambda''-\lambda'),$$

q.e.d.

References

1. Atkinson, F.V.: Asymptotic formulae for linear oscillations. Proc. Glasgow Math. Assoc. **3**, 105–111 (1956–58).
2. Bellman, R.: Methods of nonlinear analysis. New York and London: Academic Press 1970.
3. Blochinzew, D.J.: Grundlagen der Quantenmechanik. Frankfurt a.M.: Verlag Harri Deutsch 1963.
4. Coppel, W.A.: Stability and asymptotic behavior of differential equations. Boston: Heath & Company 1965.
5. Dunford, N., Schwartz, J.T.: Linear operators. Part II: Spectral theory. New York and London: Interscience Publishers 1963.
6. Fichtenholz, G.M.: Differential- und Integralrechnung II. Berlin: VEB Deutscher Verlag der Wissenschaften 1964.
7. Glazman, J.M.: Direct methods of qualitative spectral analysis. Jerusalem: Israel Program for Scientific Translations 1965.
8. Hardy, G.H.: Notes on some points in the integral calculus IV. Messenger of Math. **31**, 6–8 (1902).
9. Hartman, P.: On the ratio $f(t+c\cdot f^{-\alpha}(t))/f(t)$. Boll. Un. Mat. Ital. (3) **14**, 57–61 (1959).
10. Hartman, P.: Ordinary differential equations. New York: John Wiley & Sons 1964.
11. Hellwig, G.: Differentialoperatoren der mathematischen Physik. Berlin-Göttingen-Heidelberg: Springer 1964.
12. Hille, E.: Lectures on ordinary differential equations. Reading, Ma.: Addison Wesley Publishing Company 1969.
13. Hinton, D.B.: Some stability conditions for $y''+qy=0$. J. Math. Analysis Appl. **21**, 126–131 (1968).
14. Hinton, D.B.: Continuous spectra of second order differential operators. Pacific J. Math. **33**, 641–643 (1970).
15. Kamke, E.: Differentialgleichungen, Lösungsmethoden und Lösungen I (7. Auflage). Leipzig: Akademische Verlagsgesellschaft Geest & Portig K.G. 1961.
16. Kato, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966.

17. Kemble, E.C.: The fundamental principles of quantum mechanics. New York: Dover Publications, Inc. 1958.
18. Ludwig, G.: Die Grundlagen der Quantenmechanik. Berlin-Heidelberg-New York: Springer 1954.
19. Neumann, J. von: Mathematische Grundlagen der Quantenmechanik. Berlin-Heidelberg-New York: Springer 1968.
20. Neumark, M.A.: Lineare Differentialoperatoren. Berlin: Akademie-Verlag 1963.
21. Titchmarsh, E.C.: Eigenfunction expansions associated with second order differential equations, Part I (second edition). Oxford: At the Clarendon Press 1962.
22. Walter, J.: Bemerkungen zu dem Grenzpunktfallkriterium von N. Levinson. Math. Z. **105**, 345–350 (1968).
23. Weidmann, J.: Zur Spektraltheorie von Sturm-Liouville-Operatoren. Math. Z. **98**, 268–302 (1967).

Prof. Dr. Johann Walter
Institut für Mathematik
der RWTH Aachen
D-5100 Aachen
Templergraben 55
Federal Republic of Germany

(Received August 2, 1972)

Math. Z. 129, 95 – 107 (1972)
© by Springer-Verlag 1972

Surfaces of Constant Mean Curvature Which Have a Simple Projection

Robert Gulliver and Joel Spruck

1. Introduction

According to a result of Radó [7, p. 35], if a minimal surface in E^3 of the type of the disk spans a curve Γ, which projects simply onto the boundary of a convex plane domain D, then the surface itself may be represented as the graph of a function over D. It is the purpose of the present work to give a generalization of the above statement for surfaces of constant mean curvature H; we shall accomplish this under the assumption that ∂D is of class C^2, that Γ is the graph of a C^2 function on ∂D, that the curvature of ∂D is $\geqq 2H$, and that the surface is contained in a slab of thickness $1/H$ (Corollary 4.2). As in Radó's result, the surface in question is not required to be immersed.

A similar result has been proven by Serrin [8; cf. proof of Theorem 2] under the additional hypothesis that the surface is embedded, but requiring Γ to be only continuous. The proof rested on Serrin's results in [9] on the Dirichlet problem for the constant mean curvature equation. The present work does not use those results, but leads to an independent proof for them, under additional smoothness hypotheses, using the authors' solution of the Plateau problem for surfaces of prescribed mean curvature in a cylindrical region (Corollary 4.7).

Section 2 is devoted to a preliminary result of independent interest: if $x\colon \bar{B} \to \bar{\Omega}$ gives a surface of constant mean curvature H in conformal parameters, where Ω is a domain in E^3 whose boundary is C^2 and has inward mean curvature $\geqq |H|$, then x is transversal to $\partial\Omega$. Section 3 gives our main result, and some applications are worked out in Section 4.

2. Transversality to the Boundary

In this section, we shall consider a generalized surface of constant mean curvature H, that is, a mapping $x\colon \bar{B} \to E^3$, $x \in C^0(\bar{B}) \cap C^2(B)$, which satisfies in B the systems

$$x_{uu} + x_{vv} = 2H\, x_u \wedge x_v, \tag{2.1}$$

$$x_u^2 - x_v^2 = x_u \cdot x_v = 0, \tag{2.2}$$

where H is a nonnegative constant. Here, $B = B_1(0)$, where $B_r(p)$ denotes the open disk in $\mathbb{R}^2$ of radius r and center p. We write $C^{k+\alpha}$ for the class of functions whose k-th partial derivatives satisfy a Hölder condition with exponent α.

2.1. Lemma. *Suppose x maps ∂B into a Jordan curve Γ of class C^2. Then, $x \in C^{1+\alpha}(\bar{B})$ for each $\alpha < 1$. Moreover, if, for some $w_0 \in \bar{B}$,*

$$x_u(w_0) \wedge x_v(w_0) = 0,$$

then x satisfies an asymptotic representation,

$$x_w(w) = a(w - w_0)^{m-1} + o(|w - w_0|^{m-1}), \tag{2.3}$$

for some $m \geqq 2$ and $a \in \mathbb{C}^3 \setminus \{0\}$ with $\sum (a^k)^2 = 0$.

Proof. This was proved in [4].

Observe that if $x_u(w_0) \wedge x_v(w_0) \neq 0$, an asymptotic representation of the form (2.3) holds with $m = 1$; the condition $\sum (a^k)^2 = 0$ is just (2.2). Thus, for any point w_0, we may make the

2.2. Definition. $N_x(w_0) = \dfrac{b \wedge c}{|b \wedge c|}$, where $b, c \in \mathbb{R}^3$ and x satisfies a representation of the form (2.3) with $a = b + i\,c \neq 0$.

It may be readily seen from (2.3) that N_x is a continuous mapping of $\bar{B}$ into the unit sphere.

2.3. Lemma. *Suppose ζ: $\bar{B} \to \mathbb{C}$ is of class C^1 near $w_0 \in \bar{B}$ and satisfies*

$$\zeta(w) = (w - w_0)^m + \sigma(w),$$

where $\sigma(w_0) = 0$ and $D\sigma(w) = o(|w - w_0|^{m-1})$ as $w \to w_0$. Then there is a neighborhood U of w_0 in $\bar{B}$ and a C^1-diffeomorphism F of U onto its image with $F(w_0) = 0$ and

$$\zeta(w) = [F(w)]^m \quad \textit{for } w \in U.$$

If, moreover, ζ is of class C^2 and $D^2\sigma(w) = o(|w - w_0|^{m-2})$, then $D^2 F = o(|w - w_0|^{-1})$.

Proof. This may be seen from the proof of Lemma 2.2 in [1].

2.4. Lemma. *For some $c \in \mathbb{R}^2$, $c \neq 0$, let K be the disk $\{x \in \mathbb{R}^2 : |x - c| < |c|\}$, and suppose given two functions φ, $\bar{\varphi}$: $K \cup \{0\} \to \mathbb{R}$ of class $C^2(K) \cap C^1(K \cup \{0\})$, with $\varphi(0) = \bar{\varphi}(0)$, $\varphi \geqq \bar{\varphi}$ in K, $\varphi \not\equiv \bar{\varphi}$ in K. Suppose*

$$L[\varphi] = (1 + |D\varphi|^2)\,\Delta\varphi - \sum_{i,j} \varphi_{x^i}\varphi_{x^j}\varphi_{x^i x^j} - 2H(1 + |D\varphi|^2)^{\frac{3}{2}} \leqq 0, \tag{2.4}$$

while

$$L[\bar{\varphi}] \geqq 0. \tag{2.5}$$

Then, $c \cdot \operatorname{grad} \varphi(0) > c \cdot \operatorname{grad} \bar{\varphi}(0)$.

Proof. The difference $\varphi - \bar{\varphi}$ satisfies a linear, homogeneous, uniformly elliptic partial differential inequality (cf. [6, p. 150]). The conclusion then follows from the boundary-point lemma of E. Hopf (cf. [6, p. 67]). Q.E.D.

2.5. Theorem. *Let Ω be a region in E^3 such that $\partial\Omega$ is of class C^2 and has mean curvature $\geqq H$ at each point with respect to the inward normal. Let $\Gamma \subset \bar{\Omega}$ be a Jordan curve of class C^2. Suppose $x(\bar{B}) \subset \bar{\Omega}$, $x(\partial B) \subset \Gamma$ and $x(\bar{B}) \not\subset \partial\Omega$. Then, for $w_0 \in \bar{B}$ with $x(w_0) \in \partial\Omega$, $N_x(w_0)$ is independent of the normal to $\partial\Omega$ at $x(w_0)$.*

Proof. By Lemma 2.1, for some $m \geqq 1$,

$$x_w(w) = (b + i\,c)(w - w_0)^{m-1} + o(|w - w_0|^{m-1}), \tag{2.6}$$

where $b, c \in \mathbb{R}^3$, $b + i\,c \neq 0$, $b^2 - c^2 = b \cdot c = 0$. Thus there exists an orthonormal system of coordinates (x^1, x^2, x^3) at $x(w_0)$ in terms of which $b = (\lambda, 0, 0)$, $c = (0, -\lambda, 0)$ and $N_x(w_0) = (0, 0, -1)$. Integrating (2.6), we have

$$x(w) = \operatorname{Re}\left\{2\,\frac{b + i\,c}{m}\,(w - w_0)^m\right\} + o(|w - w_0|^m).$$

Introducing a new variable, $w' = \left(\dfrac{2\lambda}{m}\right)^{1/m}(w - w_0)$, this gives

$$x^1 + i\,x^2 = w'^m + o(|w'|^m) = \zeta(w'), \tag{2.7}$$
$$x^3 = o(|w'|^m) = \chi(w').$$

Denote $B' = \{w' \colon w \in B\}$. Applying Lemma 2.3, there is a neighborhood U of 0 in $\bar{B}'$ and a C^1-diffeomorphism F of U onto its image with $F(0) = 0$ and such that

$$x^1 + i\,x^2 + [F(w')]^m \qquad \text{for } w' \in U.$$

Let $\varepsilon > 0$ be such that $\bar{B}_\varepsilon(0) \cap F(\partial U \cap B') = \varnothing$.

First, suppose $w_0 \in B$. Then we may assume $\bar{U} \subset B'$. Let K be any open disk in the (x^1, x^2)-plane with $0 \in \partial K$, small enough that $K \subset B_{\varepsilon^m}(0)$. There is a well-defined mapping $\rho\colon \bar{K} \to \bar{B}_\varepsilon(0)$ such that $x^1 + i\,x^2 = [\rho(x^1, x^2)]^m$. Define $\varphi\colon K \to \mathbb{R}$ by $\varphi = \chi \circ F^{-1} \circ \rho$. Then $x^3 = \varphi(x^1, x^2)$, $(x^1, x^2) \in K$ is a nonparametric description of a portion of our original surface, a surface of constant mean curvature $-H$ with respect to the increasing x^3-direction. Thus φ satisfies $L[-\varphi] = 0$, where L is defined in (2.4).

If $w_0 \in \partial B$, we must go into more detail. Denote by $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ the projection onto the (x^1, x^2)-plane. Since $F(\partial B')$ is a C^1 curve, and since $m \geqq 1$, the directions at 0 which point into the set $[F(U)]^m$ include a sector of angle π. Therefore, since Γ is a C^2 curve, we may choose an open disk $K \subset B_{\varepsilon^m}(0) \setminus \pi(\Gamma)$ to contain at least one point $P' \in [F(U)]^m$. We assume ε is small enough that $\pi(\Gamma)$ separates $B_{\varepsilon^m}(0)$ into two components. It is readily seen that there is a well-defined map ρ from the component of $B_{\varepsilon^m}(0) \setminus \pi(\Gamma)$ containing P' to $F(U)$ such that $x^1 + i x^2 = [p(x^1, x^2)]^m$. In particular, this gives us the map ρ: $K \to F(U)$ which we need to define φ as in the preceding paragraph.

In either case, since $\lim_{w \to w_0} N_x(w) = N_x(w_0) = (0, 0, -1)$, we have $D\varphi(x^1, x^2) \to 0$ as $(x^1, x^2) \to 0$.

Now let N be the interior normal to $\partial\Omega$ at $x(w_0)$, and suppose, for contradiction, $N = \sigma N_x(w_0)$, where $\sigma = \pm 1$. Then $\partial\Omega$ may be represented nonparametrically in some neighborhood of 0 in the (x^1, x^2)-plane by $x^3 = \bar{\varphi}(x^1, x^2)$, and $D\bar{\varphi}(x^1, x^2) \to 0$ as $(x^1, x^2) \to 0$. The hypothesis on $\partial\Omega$ implies $L[-\sigma\bar{\varphi}] \geqq 0$, while the hypothesis $x(\bar{B}) \subset \bar{\Omega}$ implies $-\sigma(\varphi - \bar{\varphi}) \geqq 0$. Meanwhile, $x^3 = -\sigma\varphi(x^1, x^2)$ represents a surface with mean curvature $\sigma H \leqq H$ so that $L[-\sigma\varphi] \leqq 0$. With these facts, and recalling that $\varphi(0) = 0 = \bar{\varphi}(0)$ and $D\varphi(0) = 0 = D\bar{\varphi}(0)$, Lemma 2.4 implies that $\varphi \equiv \bar{\varphi}$ on K, i.e., $x(w) \in \partial\Omega$ whenever $[F(w')]^m \in K$.

Now, by hypothesis, there is $w_1 \in B$ with $x(w_1) \notin \partial\Omega$. Choose a curve γ in $B \cup \{w_0\}$ from w_0 to w_1, such that for some $\varepsilon > 0$, $[F(w')]^m \in K$ whenever $w = \gamma(t)$ for $0 \leqq t \leqq \varepsilon$. Let w_2 be the last point on γ with $x(w_2) \in \partial\Omega$. Then $w_2 \in B$ and $N_x(w_2)$ is normal to $\partial\Omega$. We may therefore carry out the above discussion with w_2 in place of w_0; since $w_2 \in B$, K may be taken as any circle in the tangent plane at $x(w_2)$ with $0 \in \partial K$ and with sufficiently small radius. This shows $x(w) \in \partial\Omega$ for all w in a neighborhood of w_2, contradicting the choice of w_2. Q.E.D.

2.6. Corollary. *Let Ω, x be as in Theorem 2.5. Then $x(B) \subset \Omega$.*

Proof. Suppose $w_0 \in B$ and $x(w_0) \in \partial\Omega$. Let N be a normal to $\partial\Omega$ at $x(w_0)$. By Theorem 2.5, $N_x(w_0)$ is independent of N. Now, given a vector T perpendicular to $N_x(w_0)$, there is a ray $w = \gamma(t)$ starting from w_0 such that $x \circ \gamma$ may be re-parameterized as a regular C^1 curve with tangent vector T at $x(w_0)$; this may be seen from Eq. (2.7) and Lemma 2.1. In particular, T could be a direction pointing out of Ω, contradicting the hypothesis $x(B) \subset \Omega$. Q.E.D.

2.7. Corollary. *Let Ω, x be as in Theorem 2.5. Then $x_u \wedge x_v(w_0) \neq 0$ for $w_0 \in \partial B$.*

Proof. If $x_u \wedge x_v(w_0) = 0$, then we have $m \geqq 2$ in the representations (2.3) and (2.7). As in the proof of Corollary 2.6, consider rays $w = \gamma(t)$ emanating from w_0 and going initially into B. Their directions comprise a sector of angle π. Their images may be re-parametrized to form C^1 curves, whose initial directions will form a sector of angle $m\pi > \pi$ in the plane perpendicular to $N_x(w_0)$. Since $N_x(w_0)$ is independent of N, some of these images must leave Ω initially, contradicting the hypothesis $x(B)\subset\Omega$. Q.E.D.

2.8. *Remarks.*

(1) The conclusions of Theorem 2.5 and Corollary 2.7 together are equivalent to the transversality of x to $\partial\Omega$.

(2) All results of this section remain valid with the constant H replaced by a continuous function $H(x)$. The hypothesis on $\partial\Omega$ in Theorem 2.5 and its two corollaries would then be that the mean curvature of $\partial\Omega$ at $x\in\partial\Omega$ (with respect to the inward normal) should be $\geqq |H(x)|$.

3. Transversality to All Vertical Lines

For the purposes of this section, suppose given $x\colon \bar{B}\to E^3$, $x\in C^0(\bar{B})\cap C^2(B)$ satisfying Eqs. (2.1)–(2.2). Let D be a domain of $\mathbb{R}^2$ with ∂D of class C^2 and having curvature $\geqq 2H$ at each point. Thus, without loss of generality, we may assume $D\subset B_{1/(2H)}(0)$. Let $\psi\colon \partial D\to\mathbb{R}$ be a C^2 function and $\Gamma\subset E^3=\mathbb{R}^2\times\mathbb{R}$ its graph. We assume x maps ∂B monotonically onto Γ and that $x(\bar{B})\subset\bar{D}\times\mathbb{R}$.

3.1. Definitions. A vector in E^3 will be called *horizontal* if it is parallel to $\mathbb{R}^2\times\{0\}$. Given a horizontal vector N_0 and given $\varepsilon>0$, we define a transformation $T\colon E^3\to E^3$ as follows: introduce orthogonal coordinates (x^1, x^2) in the plane $\mathbb{R}^2$ so that N_0 is parallel to the x^1-axis. Write x^3 for the remaining coordinate in E^3. Define $T\colon E^3\to E^3$ by $T(x^1, x^2, x^3)=(x^1-\varphi_0(x^2, x^3), x^2, x^3)$, where

$$\varphi_0(x^2, x^3)=-[(2H)^{-2}-(x^2)^2]^{\frac{1}{2}}$$

for $|x^2|\leqq\dfrac{1}{(2H)}-\varepsilon$ and φ_0 is extended linearly for $|x^2|\geqq\dfrac{1}{(2H)}-\varepsilon$ so that $\varphi_0\in C^1(\mathbb{R}^2)$. Let $\pi\colon E^3\to\mathbb{R}^2$ denote the projection on the first two coordinates.

3.2. Corollary. *$N_x(w_0)$ is not horizontal for $w_0\in\partial B$.*

Proof. Follows from Theorem 2.5 with $\Omega=D\times\mathbb{R}$. Q.E.D.

For $w_0\in B$, a regular point of x, denote by $K(w_0)$ the Gauss curvature of the surface $Tx(B)$ at $Tx(w_0)$.

3.3. Lemma. *Let N_0 be a horizontal vector, (x^1, x^2, x^3) the corresponding coordinates and $T\colon E^3\to E^3$ the corresponding transformation. Suppose $w_0\in B$ with $N_{Tx}(w_0)=N_0$ and $|x^2(w_0)|<\dfrac{1}{(2H)}-\varepsilon$. Then, for any sufficiently small neighborhood V of w_0, $N_{Tx}(V)$ is a neighborhood of*

N_0 on S^2 and $N_{Tx}(w) \neq N_0$ for $w \in V \setminus \{w_0\}$. If, moreover, w_0 is a regular point of x, then $K(w) < 0$ for $w \in V \setminus \{w_0\}$.

Proof. Applying Lemma 2.1 as in the proof of Theorem 2.5, there is $m \geqq 1$ and a similarity transformation $w \mapsto \omega$ such that

$$x^2 + i x^3 = \omega^m + o(\omega^m),$$
$$x^1 = o(\omega^m).$$

As is well known, a mapping satisfying (2.1) must be real-analytic; therefore, we actually have

$$x^2 + i x^3 = \omega^m + \sigma(\omega),$$
$$x^1 = \chi(\omega),$$

where $D^k \sigma(\omega)$ and $D^k \chi(\omega)$ are $O(\omega^{m+1-k})$ for all $k \geqq 0$. We may now apply Lemma 2.3 to find a C^1 diffeomorphism F of a neighborhood of 0 with a neighborhood W of 0, with $F(0)=0$ and $D^2 F(\omega) = o(\omega^{-1})$ such that

$$x^2 + i x^3 = [F(\omega)]^m. \tag{3.1}$$

Write $\zeta = \xi^1 + i \xi^2 = F(\omega)$. Denote $x^1 = \psi(\zeta)$; that is, $\psi = \chi \circ F^{-1}$. Then $\psi \in C^2(W)$ and for $k = 0, 1, 2$, we have $D^k \psi(\zeta) = O(\zeta^{m+1-k})$. Further, because it describes a surface of constant mean curvature H in combination with (3.1), ψ satisfies the elliptic partial differential equation,

$$\sum_i [\psi_{\xi^i}(1 + A^{-2} |D\psi|^2)^{-\frac{1}{2}}]_{\xi^i} = 2 H A^2, \tag{3.2}$$

where $A = m|\zeta|^{m-1}$. Define $\psi_0(\zeta) = \varphi_0(\zeta^m)$ where the pair (x^2, x^3) is identified with $x^2 + i x^3$. Then ψ_0 is also a solution to (3.2), and $D^k \psi_0(\zeta) = O(|\zeta|^{2m-k}) = O(|\zeta|^{m+1-k})$ for all $k \geqq 0$. Applying techniques analogous to those used for Lemma 2.4, $\Psi = \psi - \psi_0$ satisfies a linear, homogeneous elliptic partial differential equation

$$a^{ij} \Psi_{\xi^i \xi^j} + a^i \Psi_{\xi^i} = 0, \tag{3.3}$$

where $a^{ij} \in C^1(W)$ and $a^i \in C^0(W)$. For details, see [1, Section 3]. We may now apply a result of Hartman and Wintner [3, Theorems 1* and 2*] to show that $\Psi_\zeta(\zeta) = c\,\zeta^n + o(\zeta^n)$ for some $n > m$ and $c \in \mathbb{C} \setminus \{0\}$.

In some neighborhood of any $\zeta_0 \in W \setminus \{0\}$ the surface $x(B)$ may be represented nonparametrically by $x^1 = \varphi(x^2, x^3)$; then $\Phi = \varphi - \varphi_0$ gives the nonparametric representation $x^1 = \Phi(x^2, x^3)$ for $Tx(B)$. Note $\Psi(\zeta) = \Phi(\zeta^m)$. The unit normal may be described by

$$N_{Tx} = (1 + |D\Phi|^2)^{-\frac{1}{2}}(1, -\Phi_{x^2}, -\Phi_{x^3}). \tag{3.4}$$

Now, with $z=x^2+i\,x^3=\zeta^m$, we have

$$\Phi_z=\Psi_\zeta\frac{d\zeta}{dz}=\frac{c}{m}\,\zeta^{n-m+1}+o\,(\zeta^{n-m+1}).$$

From this it may be readily seen, e.g., by way of Lemma 2.3 and well-known properties of conformal mappings, that for ζ in any sufficiently small punctured neighborhood U' of 0, the image of $D\Phi$ is again a punctured neighborhood of 0. Hence, from (3.4), the image of U' under N_{Tx} is a punctured neighborhood of N_0 on the unit sphere, as was to be shown.

Now suppose w_0 is a regular point, i.e., $m=1$. Then $z=\zeta$, the functions Φ and Ψ coincide and are real-analytic and the coefficients in Eqs. (3.2) and (3.3) are real-analytic. It now follows from a result of Lewy [5, pp. 259–260] that

$$\Phi_{x^2x^2}\,\Phi_{x^3x^3}-(\Phi_{x^2x^3})^2=K(1+|D\Phi|^2)^2<0$$

holds in a punctured neighborhood of 0. Q.E.D.

Remark. It is reasonable to suppose that the assumption of regularity is not necessary for the second part of the conclusion.

Recall that at a point $w_0\in B$ with $K(w_0)\neq 0$, the sign of $K(w_0)$ is the degree of the mapping N_{Tx}: $B\to S^2$ at w_0, where the unit sphere S^2 is given the outward orientation. In general, for a C^1 map $F\colon M\to M'$, where M and M' are oriented surfaces with boundary, we write $\deg F(p)$ for the sign of the Jacobian of F at $p\in M$. For $p'\in M'$ which is a regular value of F, we write

$$\deg(F;p')=\sum_{p\in F^{-1}(p')}\deg F(p).$$

It is well known that if p' is a regular value of two C^1 maps, $F, G\colon M\to M'$ which are homotopic via a homotopy which leaves the mapping fixed on ∂M, then $\deg(F;p')=\deg(G;p')$. It follows that as a function of p', $\deg(F;p')$ is constant on components of $M'\setminus F(\partial M)$. Thus $\deg(F;p')$ may be defined by continuity for all $p'\in M'\setminus F(\partial M)$.

3.4. Lemma. *There exists a mapping f: $B\to E^3$, $f\in C^2(B)$, which agrees with x on a neighborhood of ∂B, such that $\pi\circ f$: $B\to D$ diffeomorphically.*

Proof. Applying Corollaries 2.7 and 3.2 with the continuity of N_x, we see that the set of points $w\in B$ which are critical points for x or which have $N_x(w)$ horizontal lie in a compact subdomain of B. Thus (via Corollary 2.6), their images under $\pi\circ x$ lie in a compact subdomain D' of D. Choose $\varepsilon>0$ but less than the distance from D' to ∂D, and also less than the smallest radius of curvature of ∂D. Let D'' be the set of points in D and within ε of ∂D. Then $B''=(\pi\circ x)^{-1}(D'')$ is connected: any interior component would contain a point w for which the distance from $x(w)$ to $\partial D\times\mathbb{R}$ would be a local minimum, hence $N_x(w)$ would be horizontal. Observe further that each point of B'' is a regular point of $\pi\circ x$. Thus $\deg\pi\circ x(w)$ is constant for $w\in B''$. Meanwhile, since $\pi\circ x(\partial B)=\partial D$, $\deg(\pi\circ x;p)$ is constant for $p\in D$. By Corollary 2.7, $\deg(\pi\circ x;p)=\pm 1$ for p near ∂D, hence for all $p\in D$. It follows that each point of D'' has exactly one pre-image under $\pi\circ x$, i.e., $x(B'')$ lies non-parametrically over D''.

Now, denote by Γ' the intersection of $x(B)$ with the parallel surface at distance $\varepsilon/2$ from $\partial D\times\mathbb{R}$. Thus Γ' is a smooth curve which may be represented nonparametrically over a curve in $\mathbb{R}^2$. Clearly, Γ' may be spanned by a smooth nonparametric surface. This surface may be patched to that part of $x(B'')$ which describes a nonparametric surface bounded by Γ and Γ', and then smoothed to give a mapping $f\colon \bar{B}\to E^3$ with the desired properties. Q.E.D.

3.5. Lemma. *For any horizontal vector* N, $\deg(N_{Tx};N)=0$.

Proof. Let f be as in the conclusion of Lemma 3.4. Observe N_f is nowhere horizontal; since $\dfrac{\partial\varphi_0}{\partial x^3}(x^2,x^3)=0$, it follows that N_{Tf} is nowhere horizontal. Now for $t\in[0,1]$ define $y_t(w)=t\,x(w)+(1-t)f(w)$, and let $F_t=N_{Ty_t}$. We have just shown that $F_0=N_{Tf}$ does not assume any horizontal values; therefore $\deg(F_0;N)=0$. But, for $w\in\partial B$, $F_t(w)=F_0(w)$ for all $t\in[0,1]$, i.e., the homotopy leaves the boundary mapping fixed; therefore $\deg(N_{Tx};N)=\deg(F_1;N)=\deg(F_0;N)=0$. Q.E.D.

3.6. Lemma. N_x *is nowhere horizontal.*

Proof. Suppose $N_x(w_0)$ is horizontal. Consider the continuous function $N_1\colon \bar{B}\to\mathbb{R}^2$ defined by $N_1(w)=\pi\left[x(w)+\dfrac{1}{(2H)}N_x(w)\right]$, where $\pi\colon\mathbb{R}^3\to\mathbb{R}^2$ is the projection on the first two coordinates. For any $w\in B$ with $N_x(w)$ horizontal, we may take $N_0=N_1(w)$ and define the corresponding transformation T; then assuming $|x^2(w)|<\dfrac{1}{(2H)}-\varepsilon$, this choice gives $N_{Tx}(w)=N_0$. Applying Lemma 3.3, w_0 has a neighborhood V such that $N_1(w)\neq N_1(w_0)$ for $w\in V\setminus\{w_0\}$. Now, by Lemma 2.1 and Corollary 2.7, there are only finitely many critical points of x. Therefore we may choose $w_1\in V\setminus\{w_0\}$ such that $N_x(w_1)$ is horizontal and $N_1(w_1)\neq N_1(w_2)$ for any critical point w_2 of x.

Having chosen w_1, fix $N_0=N_1(w_1)$ and let T be the deformation of E^3 associated with this N_0 and with $\varepsilon>0$ as defined in the proof of Lemma 3.4. Thus whenever $N_x(w)$ is horizontal, there holds $|x^2(w)|<(2H)^{-1}-\varepsilon$.

Further, whenever $N_{Tx}(w)=N_0$, we have $N_1(w)=N_0=N_1(w_1)$, so that w must be a regular point of x. Thus, by Lemma 3.3, there are only finitely many points w with $N_{Tx}(w)=N_0$, and each has a neighborhood V so that $\deg(N_{Tx}|_V; N_0)<0$. Since there is at least one such point, namely, w_1, it follows that $\deg(N_{Tx}; N_0)<0$, contradicting Lemma 3.5. Q.E.D.

3.7. Theorem. *Let π: $\mathbb{R}^3 \to \mathbb{R}^2$ denote the projection onto the first two coordinates. Under the hypotheses set out in the first paragraph of this section, $\pi\circ x$: $\bar{B} \to \bar{D}$ is a diffeomorphism.*

Proof. Since x maps ∂B homeomorphically onto $\Gamma=\operatorname{graph}(\bar{\psi})$, $\pi\circ x$ maps ∂B homeomorphically onto ∂D. It follows that for $y\in D$, $\deg(\pi\circ x; y)=\pm 1$. By Lemma 3.6, N_x is never horizontal, so that $\pi\circ x$ is a local diffeomorphism except on the finite set of critical points. In particular, the local degree (i.e., orientation) of $\pi\circ x$ is constant. It follows that each regular value in D has exactly one pre-image. Using the asymptotic representation (2.3) with the fact that N_x is nowhere horizontal, it is readily seen that this excludes critical points. Thus $\pi\circ x$ is a homeomorphism which is also a local diffeomorphism, hence a diffeomorphism: $\bar{B} \to \bar{D}$. Q.E.D.

3.8. Corollary. *x is an embedding.*

3.9. Corollary. *Suppose $\pi\circ x|_{\partial B}$: $\partial B \to \partial D$ is orientation-preserving (orientation-reversing). Then $x(\bar{B})$ is the graph of a function ψ: $\bar{D}\to\mathbb{R}$, $\psi\in C^1(\bar{D})\cap C^2(D)$, satisfying $L[\psi]=0$ $(L[-\psi]=0)$ and $\psi=\bar{\psi}$ on ∂D. Here L is the operator defined in* (2.4).

Proof. That $x(\bar{B})$ is the graph of a function ψ is clear from Theorem 3.7, and the smoothness of ψ follows using Lemma 2.1. Now x satisfies Eqs. (2.1)–(2.2), and so describes a surface of constant mean curvature H with respect to the normal N_x, whose component in the x^3-direction has constant sign; this sign is the same as the sign of the x^3-component of $(\pi\circ x)_u \wedge (\pi\circ x)_v$. Since $\pi\circ x$: $\partial B \to \partial D$ preserves (reverses) orientation, this sign is positive (negative) at the boundary and hence everywhere. Therefore, $\psi(-\psi)$ satisfies the constant mean curvature equation, $L[\varphi]=0$. Q.E.D.

4. Further Consequences

In this section, we suppose given a constant $H\geqq 0$; a domain $D\subset B_{1/(2H)}(0)\subset\mathbb{R}^2$ with ∂D of class C^2 and the curvature of $\partial D\geqq 2H$; a C^2 function $\bar{\psi}$: $\partial D \to\mathbb{R}$, Γ the graph of $\bar{\psi}$. Denote the cylinder $Z=\bar{B}_{1/(2H)}(0)\times\mathbb{R}\subset E^3$ and the slab $S=\left[\frac{-1}{(2H)}, \frac{1}{(2H)}\right]\times\mathbb{R}^2\subset E^3$.

The following preliminary result is of independent interest.

4.1. Lemma. *(Inclusion principle) Suppose* $x: \bar{B} \to S$, $x \in C^2(B) \cap C^0(\bar{B})$, *$x$ satisfies Eqs.* (2.1)–(2.2) *in B, and that* $x(\partial B) \subset \bar{D} \times \mathbb{R}$. *Then* $x(\bar{B}) \subset \bar{D} \times \mathbb{R}$.

Proof. We shall first show that we may assume $x(\bar{B}) \subset Z$. Let (y^1, y^2, y^3) be coordinates in terms of which S is described by $|y^1| \leqq \frac{1}{(2H)}$. Let $\Sigma_{\pm} = \{y \in \partial Z: \pm y^2 \geqq 0\}$. Since $x(\bar{B})$ is compact, we may translate Σ_+ and Σ_- in the direction of increasing and decreasing y^2, respectively, until we reach a point of last contact with $x(\bar{B})$, which we denote $x(w_1)$. Denote the translated surface which makes this contact by Σ. If $w_1 \in \partial B$, then since $x(\partial B) \subset \bar{D} \times \mathbb{R} \subset Z$, we would have $\Sigma = \Sigma_+$ or Σ_-, so that $x(\bar{B}) \subset Z$ as desired. If $w_1 \in B$, we consider two cases: (1) If $x(w_1) \in \partial S$, then since $x(B) \subset S$, $N = N_x(w_1)$ is normal to ∂S, as in the proof of Corollary 2.6. Σ and ∂S are tangent in this case, so that N is normal to Σ. (2) If $x(w_1) \in \operatorname{int} S$, then $x(w_1)$ is an interior point of Σ, so that N is again normal to Σ since $x(B)$ lies on one side of Σ. In either case, there is a closed disk K in the plane orthogonal to N, with $x(w_1) \in \partial K$, such that a piece of Σ may be represented nonparametrically over K by a function $\bar{\varphi}: K \to \mathbb{R}$. Assume the direction of increasing $\bar{\varphi}$ is toward the interior of Σ_j; then $L[\bar{\varphi}] = 0$. As in the proof of Theorem 2.5, if K is chosen sufficiently small, a piece of $x(B)$ may be described nonparametrically over K, by a function $\varphi: K \to \mathbb{R}$. Thus φ satisfies in K either $L[\varphi] = 0$ or the same equation with the sign of H changed. In either case, $L[\varphi] \leqq 0$. Now $\varphi = \bar{\varphi}$ and $D\varphi = D\bar{\varphi}$ at $x(w_1)$; while $\varphi \geqq \bar{\varphi}$ in K since $x(w_1)$ is a point of last contact. By Lemma 2.4, this implies $\varphi = \bar{\varphi}$ in K; that is, $x(U) \subset \Sigma$ for some open subset U of B. Denote the central axis of Σ by $y^1 = 0$, $y^2 = a$; define $h: \mathbb{R}^3 \to \mathbb{R}$ by $h(y^1, y^2, y^3) = (y^1)^2 + (y^2 - a)^2 - \frac{1}{(2H)^2}$. Then $h \circ x = 0$ on U. But h and x are real-analytic, so that $h \circ x = 0$ on $\bar{B}$. In particular, $h \circ x \leqq 0$. Thus if we replace Z by the cylinder $h \leqq 0$, we have $x(B) \subset Z$, with no loss of generality.

Now suppose, for some $w_0 \in B$, $x(w_0) \notin \bar{D} \times \mathbb{R}$. There is a continuous one-parameter family of circular cylinders $Z_t = \bar{B}_{1/(2H)}(p_t) \times \mathbb{R}$ with $\bar{D} \times \mathbb{R} \subset \operatorname{int} Z_t$ for $0 \leqq t \leqq 1$, $Z_0 = Z$ and $x(w_0) \in \partial Z_1$; in fact, the center curve $t \mapsto p_t$ may be taken to be a straight segment. There must be a first value s such that ∂Z_s meets $x(\bar{B})$. Let $w_1 \in \bar{B}$ be such that $x(w_1) \in \partial Z_s$. Then $w_1 \in B$. Observe that $N_x(w_1)$ is normal to ∂Z_s, since otherwise s would not be the first value for which contact occurs (cf. proof of Corollary 2.6). As in the proof of Theorem 2.5, we may show that there is a circular disk K in the plane passing through $x(w_1)$ and tangent to Z_s, with $x(w_1) \in \partial K$, such that portions of $x(B)$ and ∂Z_s may be represented nonparametrically over K by giving the signed distance from K in terms of functions φ and $\bar{\varphi}: K \to \mathbb{R}$. Since $N_x(w_1)$ is normal to ∂Z_s, $D\varphi = D\bar{\varphi} = 0$

at $x(w_1)$. Since $x(w_1)$ is a point of first contact, $\varphi \geqq \bar{\varphi}$ while $\varphi = \bar{\varphi}$ at $x(w_1)$. But $L[\bar{\varphi}]=0$ and $L[\varphi]\leqq 0$, where L is defined in (2.4), which implies $\varphi \equiv \bar{\varphi}$ by Lemma 2.4. As above, it follows that $x(\bar{B}) \subset \partial Z_s$, which contradicts $\bar{D}\times \mathbb{R} \subset \operatorname{int} Z_s$. Q.E.D.

4.2. Corollary. *Suppose* $x\colon \bar{B}\to S$, $x\in C^2(B)\cap C^0(\bar{B})$, x *satisfies Eqs.* (2.1)–(2.2) *in* B, *and* x *maps* ∂B *monotonically onto* Γ. *Then* $x(\bar{B})$ *is the graph of a function* $\psi\colon \bar{D}\to \mathbb{R}$, $\psi\in C^2(D)\cap C^1(\bar{D})$, *satisfying* $L[\psi]=0$ *or* $L[-\psi]=0$.

Proof. Immediate from Corollary 3.9 and Lemma 4.1.

4.3. Lemma. *Let* Γ *be given either orientation. There exists a mapping* $x\colon \bar{B}\to \bar{D}\times\mathbb{R}$ *of class* $C^2(B)\cap C^1(\bar{B})$ *which maps* ∂B *diffeomorphically and in an orientation-preserving fashion onto* Γ, *satisfies Eqs.* (2.1)–(2.2) *in* B, *and minimizes the functional*

$$E[y]=\iint_B (y_u^2+y_v^2+2H\pi(y)\cdot y_u\wedge y_v)\,du\,dv. \tag{4.1}$$

among smooth mappings $y\colon \bar{B}\to Z$ *such that* $y|_{\partial B}$ *is a monotone mapping onto* Γ; *here* $\pi\colon \mathbb{R}^3\to\mathbb{R}^2$ *is the projection on the first two coordinates.*

Proof. The existence of $x\colon \bar{B}\to Z$, $x\in C^2(B)\cap C^0(\bar{B})$, which satisfies (2.1)–(2.2) in B, minimizes E in the class described, and maps ∂B homeomorphically onto Γ, is proven in [2; Theorem 3]. Lemma 4.1 shows $x(\bar{B})\subset \bar{D}\times\mathbb{R}$. That $x\in C^1(\bar{B})$ is shown in Lemma 2.1, and critical points on ∂B are ruled out by Corollary 2.7. It follows that x maps ∂B diffeomorphically onto Γ.

It remains to show that either orientation may be chosen for the boundary mapping. To accomplish this, we return to the proof of the existence of a solution to the variational problem briefly described above. In that proof, the boundary values of a minimizing sequence were required to satisfy a three-point condition [2; p. 174]. The choice of the three points may be made in such a way that any monotone mapping: $\partial B\to\Gamma$ satisfying the three-point condition will be of degree 1 (over the integers). The limiting mapping $x\colon \partial B\to\Gamma$ satisfies the three-point condition, has degree 1, and hence is orientation-preserving. Q.E.D.

To this existence statement we may now add the following uniqueness result.

4.4. Corollary. *Let* Γ *be given either orientation. Suppose given two mappings* $x_1, x_2\colon \bar{B}\to S$, $x_i\in C^2(B)\cap C^0(\bar{B})$, *satisfying Eqs.* (2.1)–(2.2) *in* B, *and mapping* ∂B *monotonically and in an orientation-preserving fashion onto* Γ. *Then there exists a diffeomorphism* $f\colon \bar{B}\to\bar{B}$ *such that* $x_2=x_1\circ f$.

Proof. Let the orientation of Γ be inherited from the positive (negative) orientation of ∂D. From Corollary 4.2 we know that for $i=1,2$, $x_i(\bar{B})$

is the graph of ψ_i: $\bar{D} \to \mathbb{R}$, where $L[\psi_i]=0$ $(L[-\psi_i]=0)$, and $\psi_i=\bar{\psi}$ on ∂D. Now suppose $\psi_1-\psi_2$ assumes a maximum M (a minimum m) at $p_0 \in D$. So, $D\psi_1(p_0)=D\psi_2(p_0)$. Further, $\psi_1 \leqq \psi_2+M$ $(\psi_1 \geqq \psi_2+m)$ on a neighborhood of p_0, with equality holding at p_0. But $L[\psi_2+\text{const.}]=L[\psi_2]=0$ $(L[-\psi_2-\text{const}]=L[-\psi_2]=0)$; applying Lemma 2.4, we have $\psi_1 \equiv \psi_2+M$ $(\psi_1 \equiv \psi_2+m)$ on a neighborhood of p_0, and by extending the process, on all of D. This forces $M(m)$ to be zero. Thus $\psi_1-\psi_2$ cannot assume a positive maximum or a negative minimum, so $\psi_1=\psi_2$, and defining $f=(\pi\circ x_1)^{-1}\circ\pi\circ x_2$ gives $x_2=x_1\circ f$ as required. Q.E.D.

4.5. Corollary. *Let Γ be given either orientation. Denote by $C(\Gamma)$ the class of mappings y: $\bar{B} \to Z$ of class $C^2(B)\cap C^0(\bar{B})$, which map ∂B monotonically and with degree* 1 *onto Γ. Then $x \in C(\Gamma)$ minimizes the functional E (defined in* (4.1)) *for the class $C(\Gamma)$ if and only if x satisfies Eqs.* (2.1)–(2.2) *in B.*

Proof. That a minimizing mapping must satisfy the Euler equations (2.1) and the conformality relations (2.2) is well known. Now suppose $x=x_1$ satisfies (2.1)–(2.2). By Lemma 4.3, there is $x_2 \in C(\Gamma)$ which minimizes E for the class $C(\Gamma)$ and satisfies (2.1)–(2.2). Applying Corollary 4.4, there is a diffeomorphism f: $\bar{B} \to \bar{B}$ such that $x_2=x_1\circ f$. Since both x_1 and x_2 satisfy the conformality relations (2.2), f is conformal and orientation-preserving. This implies $E[x_2]=E[x_1]$, so that x_2 is also minimizing. Q.E.D.

4.6. *Remark.* A simple example shows the conditions $x(\bar{B})\subset S$ and $x(\bar{B})\subset Z$ are necessary for the two preceding corollaries. If Γ is a small plane circle, each of the two parts of a sphere of radius $1/H$, separated by Γ, in conformal parameterization, defines a smooth mapping $x: \bar{B} \to E^3$ satisfying (2.1)–(2.2) and taking ∂B diffeomorphically onto Γ. Let one of them be rotated about a diameter of Γ so that the boundary orientations agree. Then these two mappings satisfy all hypotheses of Corollary 4.4 except the requirement that $x(\bar{B})\subset S$. Further, the larger segment of the sphere is not in stable equilibrium for E: as the radius of the sphere is increased, E decreases monotonically toward $-\infty$.

The following result has been proven by Serrin [9] with weaker hypotheses on the smoothness of ∂D and $\bar{\psi}$, using entirely different methods.

4.7. Corollary. *There exists a function ψ: $\bar{D} \to \mathbb{R}$ of class $C^2(D)\cap C^1(\bar{D})$, satisfying $\psi=\bar{\psi}$ on ∂D and $L[\psi]=0$ $(L[-\psi]=0)$ in D, where L is defined in* (2.4).

Proof. Let x: $\bar{B} \to \bar{D}\times\mathbb{R}$ be as given in Lemma 4.3, where Γ is given the orientation induced from the positive (negative) orientation of ∂D. Then, by Corollary 3.9, there exists ψ as claimed. Q.E.D.

Added in Proof. The statement of Lemma 3.5 is incorrect; its proof fails in that $\{F_t\}$ need not be a continuous family. For each critical point ω of x, satisfying an asymptotic representation (2.3), we call the integer $m-1$ the *order of branching* at ω. Let s denote the sum of the orders of branching at critical points of x. The statement of Lemma 3.5 should be that for any horizontal vector N, deg $(N_{Tx}; N)=\frac{1}{2}s$. Since $s \geqq 0$, this modification does not affect our main results (see the proof of Lemma 3.6). To prove this corrected statement, we first extend the oriented surface $Tx(B)$ to a C^2 compact surface Σ of the type of the sphere, such that on the added piece of surface, each horizontal normal direction is assumed exactly once, at a point of positive Gauss curvature. In doing this, we may modify $Tx(\bar{B})$ in a closed subset of $Tx(B'' \cup \partial B)$, where B'' is as in the proof of Lemma 3.4. Σ will have critical points corresponding exactly to those of x. The Gauss-Bonnet formula may be applied to Σ, after removing small neighborhoods of its critical points, to show that the integral curvature of Σ – which is 4π times the degree of its unit-normal map – equals $4\pi+2\pi s$ (see the proof of Lemma 3 of [Heinz and Hildebrandt: The number of branch points of surfaces of bounded mean curvature. J. Differential Geometry **4**, 227–235 (1970)]). But that contribution to the degree of the unit-normal map of Σ at N which arises from the surface added to $Tx(B)$ is exactly 1.

Bibliography

1. Gulliver, R.: Regularity of minimizing surfaces of prescribed mean curvature. Dissertation, Stanford University 1971.
2. Gulliver, R., Spruck, J.: The Plateau problem for surfaces of prescribed mean curvature in a cylinder. Inventiones Math. **13**, 169–178 (1971).
3. Hartman, P., Wintner, A.: On the local behavior of solutions of nonparabolic partial differential equations. Amer. J. Math. **75**, 449–476 (1953).
4. Heinz, E.: Über das Randverhalten quasilinearer elliptischer Systeme mit isothermen Parametern. Math. Z. **113**, 99–105 (1970).
5. Lewy, H.: On differential geometry in the large, I. Trans. Amer. Math. Soc. **43**, 258–270 (1938).
6. Protter, M., Weinberger, H.: Maximum principles in differential equations. Englewood Cliffs, N. J.: Prentice-Hall 1967.
7. Rado, T.: On the problem of Plateau. Berlin: Julius Springer 1933 (reprint: New York: Springer 1971).
8. Serrin, J.: On the surfaces of constant mean curvature which span a given space curve. Math. Z. **112**, 77–88 (1969).
9. Serrin, J.: The Dirichlet problem for surfaces of constant mean curvature. Proc. London Math. Soc. **21**, 361–384 (1970).

Dr. R. Gulliver
Department of Mathematics
University of California
Berkeley, Ca. 94720
USA

Dr. J. Spruck
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, New Mexico 87106
USA

(Received March 27, 1972)

Math. Z. 129, 109 – 136 (1972)
© by Springer-Verlag 1972

On Planes of Lenz-Barlotti Class I 6 and Planes of Lenz Class III*

Jill C. D. S. Yaqub

It has been recognized for some time (Lüneburg [11, 12]), that there is a group-theoretic connection between planes of class I 6 and planes of class III. In each case the group Δ which is generated by the given perspectivities fixes a special line f, and there exists a set S of points on f such that the permutation group Δ^* induced on S by Δ has the following properties: i) Δ^* is doubly transitive on S, ii) if $X \in S$ then the stabilizer Δ_X^* contains a normal subgroup which is transitive and regular on $S - \{X\}$. (For class III, $S = f$; for class I 6, f contains a point F which is fixed by Δ, and $S = f - \{F\}$.) It has been shown, in the first instance by the combined work of several authors who used a variety of methods, all depending on deep results in finite group theory, that there is no finite plane of class I 6 or of class III. (An account of this work, up to 1968, is given in Dembowski [5], § 4.3.) A unified proof, which in particular completes the result for class III 1, has now been given by Hering and Kantor [7], who use the recent theorem of Hering, Kantor and Seitz [6], in which all finite permutation groups satisfying i) and ii) are determined explicitly.

In the infinite case, where such powerful group-characterization theorems are not available, the existence of planes of class I 6 and planes of class III can be studied by means of Hall's ternary ring. The ternary ring of a plane of class I 6 has been investigated by Pickert [15], Jónsson [9], Yaqub [17] and Jagannathan [8]: the results suggest that there is probably no plane of this class. The only known planes of class III are the "generalized Moulton planes", which furnish examples of both class III 1 and class III 2 (Moulton [13], Pickert [14], p. 93, Yaqub [16], André [1], Yaqub [18]), but it seems not unlikely that there exist others.

Here we continue the study of (infinite) planes of class I 6 and of class III by coordinate methods, and at the same time attempt to exploit the properties of the permutation groups Δ^*. In § 1 we consider permutation groups which satisfy i) and ii) and which also satisfy a further

* This work was supported by U.S. National Science Foundation Grant GP-8933, and was extended while the author held an Alexander von Humboldt Fellowship at the University of Tübingen.

condition, fulfilled by the groups Δ^* associated with a plane of class I 6 or of class III. § 2 is concerned with planes of class I 6: § 2.1 contains the required known theorems, in § 2.2 we consider the additive loop, extending results in [9] and [17], in § 2.3 we apply § 1 to the multiplicative group and in § 2.4 we assume that the plane admits a certain type of duality. § 3 is concerned with planes of class III: § 3.1 contains known theorems, in § 3.2 we simplify the criterion given in [16] for a cartesian group to coordinatize a plane of class III, § 3.3 contains some scattered results connected with the "distributor" defined in [1] and [18], and in § 3.4 we apply § 1 to the additive group.

For the sake of a uniform notation in the present work, the notation differs in places from that of [17] and [18]; such differences are pointed out in the foot-notes. Basic definitions and theorems concerning projective planes can be found in Pickert [14], and Dembowski [5].

I am grateful to Professor G. Pickert for reading my manuscript and for suggesting a number of improvements in the exposition.

1. A Class of Doubly Transitive Permutation Groups

Throughout this section we make the following assumption:

(A) *Let Γ be a group which is doubly transitive on the set S (finite or infinite), and suppose that the stabilizer Γ_x of an element $x \in S$ contains a normal subgroup N_x which is transitive and regular on $S-\{x\}$.*

Elements of Γ will be denoted by Greek letters. If $x \in S$ and $\gamma \in \Gamma$, then $x \cdot \gamma$ will denote the image of x under γ. Let $0, 1 \in S, 0 \neq 1$. By hypothesis Γ_0 contains a normal subgroup N_0 which is transitive and regular on $S-\{0\}$. If $a \in S$ and $a \neq 0$, let ϕ_a be the unique element in N_0 which maps 1 on a. If $x, a \in S$ and if $a \neq 0$, define $x * a = x \cdot \phi_a$; define $x * 0 = 0$ for all x. Since N_0 is regular on $S-\{0\}$, $\phi_{(a*b)} = \phi_a \phi_b$ for all $a, b \neq 0$ in S. Thus $\bar{N}_0 = (S-\{0\}, *)$ is a group, permutation-isomorphic to N_0, with neutral element 1; also $x * 0 = 0 * x = 0$ for all $x \in S$. Let N_1 be the unique subgroup of Γ_1 which is conjugate to N_0 in Γ. Then N_1 is normal in Γ_1 and is transitive and regular on $S-\{1\}$. If $a \in S$ and $a \neq 1$, let ψ_a be the unique element in N_1 which maps 0 on a. Define $x \circ a = x \cdot \psi_a$ if $x, a \in S$ and $a \neq 1$, and define $x \circ 1 = 1$ for all x. Since N_1 is regular on $S-\{1\}$, $\psi_{(a \circ b)} = \psi_a \psi_b$ for all $a, b \neq 1$ in S. Thus $\bar{N}_1 = (S-\{1\}, \circ)$ is a group, permutation-isomorphic to N_1, with neutral element 0; also $x \circ 1 = 1 \circ x = 1$ for all $x \in S$. If $a \neq 0$, let a^{-1} denote the inverse of a in $\bar{N}_0$. Then $\phi_{a^{-1}} = (\phi_a)^{-1}$ for each $a \neq 0$. Henceforth we make the following additional assumption:

(B) *If $a \in S$ and if $a \neq 0, 1$, then $\psi_{a^{-1}} = (\psi_a)^{-1}$; equivalently, $a \circ a^{-1} = 0$ if $a \neq 0, 1$.*

Theorem 1.1. *If* $a \neq 0$, 1, *let* $\chi_a = (\phi_a)^{-1} \psi_a (\phi_a)^{-1}$. *Then* i) $0 \cdot \chi_a = 1$ *and* $1 \cdot \chi_a = 0$, ii) $(b * c) \cdot \chi_a = (b \cdot \chi_a) \circ (c \cdot \chi_a)$ *for all* $b, c \in S$, iii) $(b \circ c) \cdot \chi_a = (b \cdot \chi_a) * (c \cdot \chi_a)$ *for all* $b, c \in S$, iv) $a \cdot \chi_a = a^{-1}$.

Proof. i) By definition, $0 \cdot \chi_a = 0 \cdot (\phi_a)^{-1} \psi_a (\phi_a)^{-1} = 0 \cdot \psi_a (\phi_a)^{-1} = a \cdot (\phi_a)^{-1} = 1$, while $1 \cdot \chi_a = 1 \cdot (\phi_a)^{-1} \psi_a (\phi_a)^{-1} = a^{-1} \cdot \psi_a (\phi_a)^{-1} = 0 \cdot (\phi_a)^{-1}$ by (B), $= 0$. ii) If either b or c is 0, then at least one of $b \cdot \chi_a$, $c \cdot \chi_a$ is 1, whence $(b * c) \cdot \chi_a = 0 \cdot \chi_a = 1 = (b \cdot \chi_a) \circ (c \cdot \chi_a)$. If $d \neq 0$, then $(\chi_a)^{-1} \phi_d \chi_a = \psi_{(d \cdot \chi_a)}$, since $(\chi_a)^{-1} \phi_d \chi_a \in N_1$ and maps 0 on $d \cdot \chi_a$. Thus, if $b, c \neq 0$, the result follows from the fact that $(\chi_a)^{-1} \phi_{(b*c)} \chi_a = [(\chi_a)^{-1} \phi_b \chi_a][(\chi_a)^{-1} \phi_c \chi_a]$. iii) can be proved similarly. iv) By definition, $a \cdot \chi_a = a \cdot (\phi_a)^{-1} \psi_a (\phi_a)^{-1} = 1 \cdot \psi_a (\phi_a)^{-1} = 1 \cdot (\phi_a)^{-1} = a^{-1}$.

If $a \neq 0$, 1 and if n is a positive integer, let a^n, $a^{[n]}$ denote the n^{th} powers of a in $\bar{N}_0$, $\bar{N}_1$ respectively, and define $a^{-n} = (a^{-1})^n$, $a^{[-n]} = (a^{-1})^{[n]}$; define $a^0 = 1$, $a^{[0]} = 0$. Then for all integers m, n, $a^{m+n} = a^m * a^n$, $(a^m)^n = a^{mn}$, $a^{[m+n]} = a^{[m]} \circ a^{[n]}$ and $(a^{[m]})^{[n]} = a^{[mn]}$.

Theorem 1.2. *Let* $a, b \neq 0$, 1. *Then for every integer* n, i) $b^n \cdot \chi_a = (b \cdot \chi_a)^{[n]}$, ii) $b^{[n]} \cdot \chi_a = (b \cdot \chi_a)^n$, iii) $a^n \cdot \chi_a = a^{[-n]}$, iv) $a^{[n]} \cdot \chi_a = a^{-n}$. v) *The element* a *has finite order* n *in* $\bar{N}_0$ *if and only if it has order* n *in* $\bar{N}_1$. vi) *The element* $b \cdot \chi_a$ *has finite order* n *in* $\bar{N}_0$ *if and only if* b *has order* n *in* $\bar{N}_0$.

Proof. By Theorem 1.1, i)–iii), the map $x \to x \cdot \chi_a$, $x \in S$, $a \neq 0$, 1, induces an isomorphism of $\bar{N}_0$ onto $\bar{N}_1$ and an isomorphism of $\bar{N}_1$ onto $\bar{N}_0$. Statements i), ii) follow immediately from this and from (B), and iii), iv) are obtained by setting $b = a$ and applying Theorem 1.1 iv). Statement v) holds because χ_a: $\bar{N}_0 \to \bar{N}_1$ maps a on a^{-1} and a^{-1} is the inverse of a in $\bar{N}_1$. Finally b has finite order n in $\bar{N}_0$ if and only if $b \cdot \chi_a$ has order n in $\bar{N}_1$, i.e., by v), in $\bar{N}_0$.

Theorem 1.3. *Let* n *be a positive integer. Let* a, $\neq 0$, 1, *be an element such that* $a^r \neq 1$ *for* $0 < r \leqq n$. *Then* i) $(a^n)^{[n]} = a = (a^{[n]})^n$, ii) $x^n = a \Rightarrow x = a^{[n]}$; $y^{[n]} = a \Rightarrow y = a^n$.

Proof. We show first that i) implies ii). Suppose i) holds. Let n be a positive integer, and let $a \neq 0$, 1, $a^r \neq 1$ for $0 < r \leqq n$. If $x^n = a$ and if $x^r = 1$ for some r with $0 < r \leqq n$, then $a^r = 1$ contrary to hypothesis. Hence i) can be applied to x, and $x = (x^n)^{[n]} = a^{[n]}$. If $y^{[n]} = a$ and if $y^r = 1$ with $0 < r \leqq n$, then by Theorem 1.2 v), $y^{[r]} = 0$, whence $a^{[r]} = 0$ and $a^r = 1$. Hence by i) applied to y, $y = (y^{[n]})^n = a^n$.

We shall prove i) by induction on n. The statement is true for $n = 1$. Let $n > 1$ and assume as induction hypothesis that for every $c \neq 0$, 1 and every $m < n$ with $c^r \neq 1$ for $r = 1, \ldots, m$ the equations $(c^m)^{[m]} = c = (c^{[m]})^m$ hold. Let $a \neq 0$, 1 and let $a^r \neq 1$ for $0 < r \leqq n$. By Theorem 1.2 iii), $a^{-(n-1)} \cdot \chi_a = a^{[n-1]}$. By definition, $a^{-(n-1)} \cdot \chi_a = a^{-(n-1)} \cdot (\phi_a)^{-1} \psi_a (\phi_a)^{-1} = (a^{-n} \circ a) * a^{-1}$. Hence $a^{[n-1]} = (a^{-n} \circ a) * a^{-1}$, i.e. $a = a^n \circ (a^{[n-1]} * a)$. Let $b = a^{[n-1]}$.

Then by the induction hypothesis with $c=a$ and $m=n-1$, $b^{n-1}=a$. Hence

(1) $$b^{n-1}=b^{n(n-1)}\circ b^{n}.$$

If $b^r=1$ for some r with $0<r\leqq n$, then $a^r=1$, contrary to hypothesis. If $b^{ns}=1$ for some s with $0<s\leqq n-1$, then $a^s=b^{-s}$, whence by the induction hypothesis with $c=a$, b^{-1} and $m=s$, $a=(a^s)^{[s]}=[(b^{-1})^s]^{[s]}=b^{-1}$. But this implies that $a^n=1$. Thus the induction hypothesis can be applied with $c=b^n$ and $m=n-1$, whence $b^n=(b^{n(n-1)})^{[n-1]}$. Now, by (1), $b^{n-1}=b^{n(n-1)}\circ(b^{n(n-1)})^{[n-1]}$, $=(b^{n(n-1)})^{[n]}$, i.e. $a=(a^n)^{[n]}$. Then $a^{-1}=a\cdot\chi_a$, by Theorem 1.1 iv), $=(a^n)^{[n]}\cdot\chi_a=(a^n\cdot\chi_a)^n$ by Theorem 1.2 ii), $=[(a\cdot\chi_a)^{[n]}]^n$ by Theorem 1.2 i), $=(a^{[-n]})^n$ by Theorem 1.1 iv) $=(a^{[n]})^{-n}$ by (B). Hence $a=(a^{[n]})^n$. Thus, by induction, i) holds for all positive integers n.

Theorem 1.4. *If a, $\neq 0, 1$, has finite order n in $\bar{N}_0$, then n is prime.*

Proof. Suppose that a, $\neq 0$, 1, has finite order n in $\bar{N}_0$. Let A be the cyclic subgroup of $\bar{N}_0$ which is generated by a, and let p be the smallest prime which divides n. If $x\in A$, $x\neq 1$, then $x^r\neq 1$ for $0<r<p$. If $n>p$, there exists $b\in A$ with $b^p\neq 1$. By Theorem 1.3, $(b^p)^{[p]}=b$. There also exists $c\in A$ with $c\neq 1$, $c^p=1$. Since $(b*c)^p=b^p\neq 1$, $[(b*c)^p]^{[p]}=b*c$ by Theorem 1.3. But this implies the contradiction $b*c=b$, since $(b*c)^p=b^p$. Hence $n=p$.

Theorem 1.5. *If a, $\neq 0, 1$, has order p in $\bar{N}_0$, and if $n\not\equiv 0 \pmod p$,* then i) $(a^n)^{[n]}=a=(a^{[n]})^n$, ii) $x^n=a\Rightarrow x=a^{[n]}$ *and* $y^{[n]}=a\Rightarrow y=a^n$, iii) $a^{[n]}\in A$, *the subgroup of $\bar{N}_0$ which is generated by a.*

Proof. Since $n\neq 0 \pmod p$, $n=kp+r$ where k is an integer and $0<r<p$. Since $a^p=1$, $a^{[p]}=0$ by Theorem 1.2 v). i) By Theorem 1.3, $(a^r)^{[r]}=a=(a^{[r]})^r$, whence $(a^n)^{[n]}=(a^{kp+r})^{[kp+r]}=(a^r)^{[r]}=a$ and similarly $(a^{[n]})^n=a$. ii) If $x^n=a$, then $x^{np}=1$. Since x has prime order by Theorem 1.4 and since $x^n\neq 1$, it follows that $x^p=1$, whence $x^r=a$. By Theorem 1.3 ii), this implies that $x=a^{[r]}$, $=a^{[n]}$. Similarly, if $y^{[n]}=a$ then $y^{[p]}=0$ and $y=a^n$. iii) Since n is relatively prime to p, A contains a solution x of the equation $x^n=a$. By ii), $x=a^{[n]}$.

Theorem 1.6. *If $a\neq 0$ and if a is not of finite order in $\bar{N}_0$, then there is an isomorphism σ of $(Q, +)$, the additive group of the rational numbers, into $\bar{N}_0$ with $(m/n).\sigma=(a^{[n]})^m$ for all integers m, n with $n\neq 0$; the image of $(Q, +)$ under σ will be written as $\bar{N}_0(a)$.*

Proof. If $a\neq 0$ and if a is not of finite order in $\bar{N}_0$, then a satisfies the conclusions of Theorem 1.3 for every positive integer n. Note that if m, n are integers with $n\neq 0$, then $(a^m)^{[n]}=(a^{[n]})^m$. For if $a^{[n]}=b$, then $a=b^n$, whence $(a^m)^{[n]}=(b^{mn})^{[n]}=b^m$. A similar argument shows that if m, n, k, l are integers with $n, l\neq 0$, then $(a^m)^{[n]}=(a^k)^{[l]}$ only if $ml=kn$. Moreover if

m, n, k, l are integers with $n, l \neq 0$, then $(a^m)^{[n]} * (a^k)^{[l]} = (a^{ml+kn})^{[nl]}$; for if $a^{[nl]} = b$ then $a = b^{nl}$, and $(a^m)^{[n]} * (a^k)^{[l]} = b^{ml} * b^{kn} = b^{ml+kn} = (a^{ml+kn})^{[nl]}$. For each positive integer n, let A_n be the subgroup of $\bar{N}_0$ which is generated by $a^{[n!]}$. Then $A_n \subset A_{n+1}$ for every n. Let $\bar{N}_0(a) = \bigcup_1^\infty A_n$; then $\bar{N}_0(a)$ is a subgroup of $\bar{N}_0$, and contains the element a. If r is rational and $r = m/n$, where m, n are integers and $n \neq 0$, define $r.\ \sigma = (a^{[n]})^m$. Then it is easily verified that σ: $(Q, +) \to (\bar{N}_0(a), *)$ is an isomorphism.

In view of Theorem 1.6, we shall write $a^{m/n}$ for $(a^{[n]})^m$ if $a \neq 0, 1$ and m, n are integers with $n \neq 0$. It is to be understood that the domain of exponents for the element a is $GF(p)$ if a has finite (prime) order p in $\bar{N}_0$, and is Q if a is not of finite order.

Theorem 1.7. *Let* $a \neq 0, 1$. i) *If* $u \neq 0$, *then* $a^u \cdot \chi_a = a^{-1/u}$, ii) *if* $r, s, r+s \neq 0$, *then* $a^r \circ a^s = a^{rs/(r+s)}$.

Proof. Suppose first that a is not of finite order in $\bar{N}_0$. i) Let $u = m/n$, where m, n are integers and $n \neq 0$. Let $a^{1/n} = b$, so that $b^n = a$. Let $w = a^u \cdot \chi_a,\ = b^m \cdot \chi_a$. By Theorem 1.2 i), $w^{1/n} = b^{mn} \cdot \chi_a = a^m \cdot \chi_a,\ = a^{-1/m}$ by Theorem 1.2 iii). Hence $w = a^{-n/m} = a^{-1/u}$. ii) If $r, s, r+s \neq 0$, then by i), $a^r \circ a^s = (a^{-1/r} \cdot \chi_a) \circ (a^{-1/s} \cdot \chi_a) = (a^{-1/r} * a^{-1/s}) \cdot \chi_a$ by Theorem 1.1 ii), $= (a^{(-1/r)+(-1/s)}) \cdot \chi_a$ by Theorem 1.6, $= a^{rs/(r+s)}$ by i). If a has finite order p, then $u,\ \neq 0,\ \in GF(p)$ and i) is a restatement of Theorem 1.2 iii). Note that by Theorem 1.5, $a^{-1/u} \in A$, the group generated by a in $\bar{N}_0$. Thus, in the above proof of ii), the statement $a^{-1/r} * a^{-1/s} = a^{(-1/r)+(-1/s)}$ is now justified by the fact that $a^{-1/r}, a^{-1/s} \in A$.

Theorem 1.8. *If* $a, b \in \bar{N}_0$, $a, b \neq 1$, *and if for some prime* p, $a^p = 1$ *but* $b^p \neq 1$, *then* $a * b \neq b * a$.

Proof. Suppose that $a^p = 1, b^p \neq 1$ and that $a * b = b * a$. By Theorem 1.3 (or by Theorem 1.5 if b has finite order $< p$), $b = (b^p)^{[p]}$. Since $a * b = b * a$, $(a * b)^p = b^p \neq 1$. If $(a * b)^r = 1$ for some r with $0 < r < p$, then $1 = (a * b)^{pr} = b^{pr}$. By Theorem 1.4, the order of b is prime, whence, since $b^p \neq 1$, $b^r = 1$. But then $a^r = 1$, contrary to hypothesis. Hence, by Theorem 1.3, $(a * b) = [(a * b)^p]^{[p]}$, and this yields the contradiction $a * b = b$.

Theorem 1.9. *If* $\bar{N}_0$ *is abelian, then either* a) $\bar{N}_0$ is *of exponent* p *for some prime* p, *or* b) $\bar{N}_0$ *contains no element of finite order and is the direct product of groups isomorphic to* $(Q, +)$.

Proof. If $\bar{N}_0$ contains an element of order p, then by Theorem 1.8 every element has order p. If $\bar{N}_0$ contains no element of finite order, then for each $a \neq 1$ in $\bar{N}_0$ and for every positive integer n, the equation $x^n = a$ has a unique solution x in $\bar{N}_0$, by Theorem 1.3. Since $\bar{N}_0$ is abelian and contains no element of finite order, this implies that $\bar{N}_0$ is the direct product of groups isomorphic to $(Q, +)$. (Kurosch [10], § 23.)

Definition. Let (Γ, S, N_0), $(\Gamma', S', N'_{0'})$ satisfy (A) and (B). Then (Γ, S, N_0) is *permutation-isomorphic* to $(\Gamma', S', N'_{0'})$ if and only if i) there exists a bijection β: $S \to S'$ with $0\beta = 0'$ and an isomorphism σ: $\Gamma \to \Gamma'$, with $\gamma^\sigma = \beta^{-1}\gamma\beta$ as the image of γ under σ, ii) $(N_0)^\sigma = N'_{0'}$.

If (Γ, S, N_0) is permutation-isomorphic to $(\Gamma', S', N'_{0'})$ and if $1\beta = 1'$, then $N_1^\sigma = N'_{1'}$, since N_1 is defined as the subgroup of Γ_1 which is conjugate to N_0 in Γ and $N'_{1'}$ is defined similarly in Γ'. Moreover if $a \in S$, $a \neq 0$, then $(\phi_a)^\sigma = \phi'_{a\beta}$, since $(\phi_a)^\sigma \in N'_{0'}$ and maps $1'$ on $a\beta$; similarly, if $a \neq 1$, then $(\psi_a)^\sigma = \psi'_{a\beta}$. It follows that β: $S \to S'$ induces an isomorphism of N_0 onto $N'_{0'}$ and an isomorphism of N_1 onto $N'_{1'}$, and that if $a \neq 0, 1$ then $(\chi_a)^\sigma = \chi'_{a\beta}$. The bijection β can be chosen so that $1 \cdot \beta = t'$, where t' is any element $\neq 0'$ in S'. For, by (A), there exists $\tau' \in \Gamma'$ with $0' \cdot \tau' = 0'$ and $1' \cdot \tau' = t'$; it is easily verified that β, σ can be replaced by $\beta\tau'$ and $\sigma\hat{\tau}$ respectively, where $\hat{\tau}$ is the automorphism of Γ' in which $\gamma' \to (\tau')^{-1}\gamma'\tau'$ for each $\gamma' \in \Gamma'$.

Let $(K, +, \ )$ be a skew-field. Let $\bar{K} = K \cup \{\infty\}$. For each $a \in K$, define $\infty \cdot \phi_a = \infty$, $x \cdot \phi_a = x + a$ if $x \in K$, and let $E^*_\infty(K) = \{\phi_a | a \in K\}$. For each $a \in \bar{K} - \{0\}$, define $0 \cdot \psi_a = 0$ and $x \cdot \psi_a = x \oplus a$ if $x \in \bar{K}$, $x \neq 0$, where $x \oplus \infty = x$ and, if $a \neq \infty$, $\infty \oplus a = a$, $(-a) \oplus a = \infty$ and $x \oplus a = x(x+a)^{-1}a$ if $x \neq \infty, -a$. Let $E^*_0(K) = \{\psi_a | a \in \bar{K} - \{0\}\}$. We denote by $PSL(2, K)$ the permutation group on $\bar{K}$ which is generated by $E^*_\infty(K)$ and $E^*_0(K)$. It can be shown, either by direct calculation or by suitable interpretation in the desarguesian plane over K (cf. §3), that $(PSL(2, K), \bar{K}, E^*_\chi(K))$ satisfies (A) and that $E^*_0(K)$ is conjugate to $E^*_\infty(K)$. Moreover, if $\infty, 0$ are chosen as the special symbols previously denoted by 0, 1, then (B) is also satisfied, since if $a \neq \infty, 0$ then $(\phi_a)^{-1} = \phi_{-a}$ and $(\psi_a)^{-1} = \psi - a$. The operations $+, \oplus$ (extended by the definitions $x + \infty = \infty$ and $x \oplus 0 = 0$ for all $x \in \bar{K}$), now coincide with the operations $*, \circ$ respectively. If $a \neq \infty, 0$, let $\chi_a = (\phi_a)^{-1}\psi_a(\phi_a)^{-1}$ as before. Then we find:

(2) *If* $a, b \neq \infty, 0$, *then* $\chi_b\chi_a$ *fixes* $\infty, 0$ *and if* $x \neq \infty, 0$, $x \cdot \chi_b\chi_a = ab^{-1}xb^{-1}a$ *(evaluated in K).*

Thus if $a \neq \infty, 0$, then $(\chi_a)^2$ is the identity. Furthermore, if $a, b \neq \infty, 0$ and if $ab^{-1} = z$ is in ZK, the centre of K, then given any $d \neq \infty, 0$, there exists $c \neq \infty, 0$ such that $\chi_b\chi_a = \chi_d\chi_c$, namely $c = zd, = dz$. The following theorem will be required in § 2.

Theorem 1.10. *Let K be a skew-field with centre ZK, and suppose that* (Γ, S, N_0) *is permutation-isomorphic to* $(PSL(2, K), \bar{K}, E^*_\infty(K))$, *with associated bijection* β: $S \to \bar{K}$ *and isomorphism* σ: $\Gamma \to PSL(2, K)$ *chosen so that* $0 \cdot \beta = \infty$ *and* $1 \cdot \beta = 0$. *Then* i) N_0 *is abelian,* ii) *if* $a \in S$, $a \neq 0, 1$, $(\chi_a)^2$ *is the identity,* iii) *if* $a, b, d \in S$, $\neq 0, 1$ *and if* $(a\beta)(b\beta)^{-1} \in ZK$, *then there exists* $c \in S$, $c \neq 0, 1$ *such that* $\chi_b\chi_a = \chi_d\chi_c$.

Proof. i) By hypothesis N_0 is isomorphic to the abelian group $E^*_\infty(K)$. ii) If $a \neq 0, 1$ then $a\beta \neq \infty, 0$. Thus since $(\chi_a)^\sigma = \chi'_{a\beta}$ and since $(\chi'_{a\beta})^2$ is the identity, $(\chi_a)^2$ is the identity. iii) If $a, b, d \neq 0, 1$ then $a\beta, b\beta, d\beta \neq \infty, 0$. Hence, since $(a\beta)(b\beta)^{-1} \in ZK$, there exists $c' \neq \infty, 0$ such that $\chi'_{b\beta}\chi'_{a\beta} = \chi'_{d\beta}\chi'_{c'}$. Let $c\beta = c'$. Then $c \neq 0, 1$ and $\chi_b\chi_a = \chi_d\chi_c$.

Remark. By the theorem of Hering, Kantor and Seitz [6], if Γ, S satisfy (A) and if S is finite, then there exists $H \lhd \Gamma \subseteq \text{Aut}\, H$ such that H is isomorphic to one of the following, in its usual representation: i) a sharply doubly transitive group, ii) $PSL(2, q)$, iii) $Sz(q)$, iv) $PSU(3, q)$, v) a group of Ree type. In the last three cases the subgroup N_0 contains elements which are not of prime order, whence by Theorem 1.4 these groups do not satisfy (B). Moreover if $|S| > 4$, a sharply doubly transitive group also does not satisfy (B). For such a group is the group of all permutations $x \to xa + b$, $a \neq 0$, where the elements and operations are those of a right near-field. (See [5], p. 33.) With our previous notation, if $a \neq 0, 1$ then $x * a = xa$ and $x \circ a = x(1-a) + a$. Hence, if (B) holds, $a^{-1} \circ a = a^{-1}(1-a) + a = 0$ for all $a \neq 0, 1$. Then $1 - a = a(-a)$, and with $b = -a$, $1 + b = (-b)b = -b^2$ for all $b \neq 0, -1$. If $1 + 1 \neq 0$, then with $b = 1$, $1 + 1 + 1 = 0$. If $|S| > 3$ and if $b \neq 0, 1, -1$, $b^2 + b + 1 = 0$, whence $b^2 - b = b - 1$, i.e. $(b-1)b = b - 1$. But this implies the contradiction $b = 1$. If $1 + 1 = 0$, then $b^2 = b + 1$ for all $b \neq 0, 1$, whence $b^3 = 1$ and $b^{-1} = b + 1$. If $|S| > 4$, there exist $b, c \neq 0, 1$ with $b \neq c$, $bc \neq 1$. Then $bcb = (c^{-1}b^{-1} + 1)b = c^{-1} + b = c + 1 + b$, and similarly $cbc = b + 1 + c$. Since addition is commutative, $bcb = cbc$, whence $(bc)^2 = cbc^{-1}$. But $(bc)^2 = c^{-1}b^{-1}$, so that $(c^{-1}b)^2 = 1$ and we obtain the contradiction $b = c$. Hence if $4 < |S| < \infty$, condition (B) in fact characterizes the groups with $H \simeq PSL(2, q)$ among the finite groups which satisfy (A).

2. Planes of Class I 6

2.1. Known Results

The projective plane π is *of class* I 6 if and only if it contains a point-line pair (F, f) with F on f such that i) π is $(X, \theta(X))$ transitive for each $X \neq F$ on f, where θ is a bijection between the points $\neq F$ on f and the lines $\neq f$ on F, ii) π is not (P, l) transitive for any other pair (P, l). Note that every collineation of π must fix F, f and preserve θ.

We use the coordinate system of [17]. Let $R+$ denote the additive loop and R^* the multiplicative loop of the ternary ring R. Let $x = a \diagdown b$ if and only if $a + x = b$ and let $y = b/a$ if and only if $y + a = b$. In particular, for each a let $-a = a \diagdown 0$ and let $\sim a = 0/a$, so that $a + (-a) = 0 = (\sim a) + a$. We write $b - a$ for $b + (-a)$. $R+$ has the *left inverse property* if and only if $a + (-a + x) = x$ for all a, x. In general $\sim a \neq -a$, but if $R+$ has the left inverse property then $\sim a = -a$ for all a. The *right nucleus* of $R+$, denoted by RN, is the set $\{x \mid a + (b + x) = (a + b) + x \text{ for all } a, b \in R\}$. The *left* and *middle nuclei*, LN and MN, are defined analogously. Each of the right, left and middle nuclei is a group. Their intersection is the *nucleus*, N.

If π is a plane of class I 6, the fundamental coordinate quadrangle $VUOE$ can be chosen so that $V=F$, $VU=f$, $\theta(U)=VO$ and $\theta((1))=[1]$ (Pickert [15]). Since π is (U, OV) transitive, R is linear (i.e. the point (x, y) is on the line $[m, c]$ if and only if $y=mx+c$), and multiplication is associative. The (U, OV) homology which maps (1) on (a), $a\neq 0$, is given by $(x, y)\to(a^{-1}x, y)$, $[m, c]\to[ma, c]$ ([14], p. 102). Neither distributive law holds and addition is not associative ([14], pp. 104, 105, 100). π admits the (V, UV) elation a^* which maps $(0, 0)$ on $(0, a)$ if and only if $a\in RN$; in this case $(x, y)\,a^*=(x, y+a)$, $[m, c]\,a^*=[m, c+a]$. (Note that collineations are regarded as acting on the right.)

Theorem A (Pickert [15]). *If π is of class* I 6 *and is represented as above, then* i) $\theta((p))=[p^{-1}]$ *for all* $p\neq 0$ *in* R, ii) $p^{-1}+(p+x)=1+x$ *for all* $p\neq 0, 1$ *and all* x *in* R, iii) $(a+b)+1=a+(b+1)$ *for all* a, b *in* R, iv) $a+1=1+a$ *for all* a *in* R, v) π *admits the* (V, UV) *elation* 1^*.

Theorem B (Jónsson [9]). *Let π be of class* I 6, *represented as above. Let* $a\circ b=a+(ab\backslash b)$. *Then* i) *for each* $p\neq 0, 1$, *the* $((1), [1])$ *perspectivity which maps* (p) *on* U *is given by* $(x, y)\to(p\circ x,\ p+(px\backslash y))$, $[m, c]\to[m\circ p^{-1}, p^{-1}\backslash(mp^{-1}+c)]$, ii) *if* $p\neq 0, 1$, *then* $p+(px\backslash\{mx+c\})=(m\circ p^{-1})(p\circ x)+(p^{-1}\backslash\{mp^{-1}+c\})$, iii) *the elements other than* 1 *form a group with respect to the operation* $\circ$, *with neutral element* 0; $p\circ 1=1\circ p=1$ *for all* p; *for all* $p\neq 0, 1$, $p\circ p^{-1}=p^{-1}\circ p=0$, iv) *if* $p, x\neq 0$, $p\neq 1$ *and if* $px\neq 1$, *then* $p^{-1}(p\circ x)=p+(x^{-1}p^{-1}-p)$, v) *if* $p, x, y\neq 0, 1$ *and if* $xy\neq 1$, *then* $[p+(x-p)]\circ[p+(y-p)]=p+(yx-p)$ *and* $[p+(x-p)][p+(y-p)]=p+[(y\circ x)-p]$, vi) *if* $p\neq 0, 1$ *and if* $p^2\neq 1$, *then* $p^2\circ p^2=p=(p\circ p)^2$; *moreover* $x^2=p$ *only if* $x=p\circ p$ *and* $x\circ x=p$ *only if* $x=p^2$, vii) *if* $a\circ b=ba$ *for all* a, b *with* $a, b\neq 0, 1$, $ab\neq 1$, *then* $1+1+1=0$ *and* $x^2=1$ *for all* $x\neq 0$, viii) R *contains no element* $a\neq 0, 1$ *such that* $a+(x-a)=x$ *for all* x *in* R.

Theorem C (Lüneburg [11], Cofman [4]). *Let π be of class* I 6. *Let* $S=f-\{F\}$. *If* $X\in S$, *let* H_X *denote the group of all* $(X, \theta(X))$ *homologies and let* H_X^* *denote the permutation group induced on* S *by* H_X. *Let* Δ *be the group generated by all* H_X *with* $X\in S$, *and let* Δ^* *be the permutation group induced on* S *by* Δ. *Let* $Z\Delta$ *denote the centre of* Δ. *Then* i) Δ^* *is doubly transitive on* S, ii) *if* $X\in S$, H_X^* *is normal in the stabilizer* Δ_X^* *and is transitive and regular on* $S-\{X\}$, iii) $\Delta/Z\Delta\simeq\Delta^*$, iv) $Z\Delta$ *consists of those* (F, f) *elations of* π *which are contained in* Δ.

It is easily verified that in fact Δ is generated by any two distinct homology groups H_X, H_Y, with $X, Y\in S$, $X\neq Y$.

Theorem D (Yaqub [17]). *Let π be of class* I 6. *Let π' be a subplane of π such that* $F, f\in\pi'$ *and if* $X\in f$ *then* $X\in\pi'$ *if and only if* $\theta(X)\in\pi'$. *Then π' is either of class* I 6 *or desarguesian. If π' is desarguesian, it is of finite order* 2, 3 *or* 4.

Theorem D is proved in [17], Lemma 2, under the hypothesis that π' is fixed by some non-identical collineation of π: however, this hypothesis is used only to show that π' satisfies the assumptions of Theorem D. Note that the theorem applies to any subplane π' which contains the coordinate quadrangle $VUOE$ (chosen as above). For $V, VU, U, VO \in \pi'$ and if $p \neq 0$ then $X=(p)$ is in π' if and only if $\theta(X)=[p^{-1}]$ is in π'.

Theorem E ([17]). *Let R be the ternary ring of a plane of class* I 6 *represented as above. Then* i) $p+(z+c)=[p+(z-p)]+[p+c]$ *for all* p, z, c, ii) *if* $z \neq 0$, $[p+(z-p)]^{-1}=p+(z^{-1}-p)$, iii) *if* $p^2 \neq 1$, $p+(p+z)=p^2+(p^2+z)$ *for all z in R,* iv) *$R+$ has the left inverse property if and only if* $1 \in N$, v) *if, for some p,* $x=p+(y-p)$, *then* $x^3=1$ *if and only if* $y^3=1$, vi) *if ϕ_a denotes the (U, OV) homology which maps U on* (a), $a \neq 0$, *if* ψ_a[1] *denotes the* ((1), [1]) *homology which maps U on* (a), $a \neq 1$, *if*

$$\chi_a=(\phi_a)^{-1}\psi_a(\phi_a)^{-1}, \quad a \neq 0, 1,$$

and if $\phi_{a,b}=(\chi_b\chi_a)^{-1}$, $a, b \neq 0, 1$, *then* $(x, y)\phi_{a,b}=(\{b+(a+x)\}/(b+a), b+(a+y-1))$, $[m, c]\phi_{a,b}=[\{b+(a+m)\}/(b+a), b+(a+c-1)]$, vii) *for each* $t \neq 0, 1, -1$ *in R, π admits the collineation Φ_t where* $(x, y)\Phi_t=(t+(-t+x), t+(-t+y))$, $[m, c]\Phi_t=[t+(-t+m), t+(-t+c)]$, viii) *if* $a \in RN$ *and if* $t \neq 0, 1, a, a-1$, *π admits the collineation $\Phi_{t,a}$, where*

$$(x, y)\Phi_{t,a}=(t+\{(-t+a)+x\}-a, t+\{(-t+a)+y\}-a),$$

$$[m, c]\Phi_{t,a}=[t+\{(-t+a)+m\}-a, t+\{(-t+a)+c\}-a],$$

ix) *either* a) $1+1=0$ *or* b) $1+1+1=0$ *and* $(-1)^2=1$.

Theorem F (Burn [3]). *If π is a projective plane of Lenz-Barlotti class* $>$ I 2, *then π cannot admit both an involutory homology and an involutory elation.*

Theorem G ([17]). *Let R be the ternary ring of a plane of class* I 6, *represented as above, with* $1+1=0$. *Then* i) *if* $a \neq 0, 1$, *if* $a^3=1$ *and if* $a+(a+x)=x$ *for all x, then* $a+b=b+a$ *only if* $b=0, 1, a$ *or* $a+1$, ii) *if $R+$ has the left inverse property, if* $a \neq 0, 1$ *and if* $y+a=a+y$ *or* $a+y+1$ *for all y, then either* $a^3=1$ *or* $a^5=1$, iii) *if the conditions of* ii) *are satisfied and if also* $a+a=0$, *then* $a+b=b+a$ *if and only if* $b=0, 1, a$ *or* $a+1$.

Theorem G i) can be proved by the argument given in [17], (12): the fact that $a+(w-a)=z$ implies $a+(z-a)=w$ follows easily from Theorem E i) and the assumption $a+(a+x)=x$, while $a^3=1$ and Theorem E iii) together imply that a^{-1} satisfies the same hypotheses as a. Theorem G ii) and iii) are proved in [17], (7) and (8), under the assumption that π is finite. However, finiteness is used only to show that $p^2 \neq 1$ for any

[1] ψ_a, χ_a are the inverses of the ψ_a, χ_a defined in [17]; $\phi_{a,b}$ is the $\phi_{b,a}$ of [17].

$p \neq 1$, and this is also true in the infinite case, since by Theorem A v) π admits the involutory elation 1^*, and hence by Theorem F cannot admit an involutory homology.

Theorem H ([9, 11, 4, 17, 7]). *There is no finite plane of class* I 6.

2.2. The Additive Loop

Throughout this section, R denotes the ternary ring of a plane of class I 6, represented as in § 2.1.

Theorem 2.1. i) *For all* p, z, $(p+z)/p = p+(z-p)$, ii) $p+z=z+p$ *if and only if* $p+(z-p)=z$, iii) $p+(-p+c) = \sim p+(p+c)$ *for all* p, c, iv) $p+[p x \diagdown (x+c)] = (p \circ x)+c$ *for all* p, x, c, v) $[b+(a+x)]/(b+a) = b+[\{a+(x-a)\}-b]$ *for all* a, b, x.

Proof. By Theorem E i), $p+(z+c) = [p+(z-p)]+[p+c]$ for all p, z, c. Statements i) and ii) follow immediately, by taking $c=0$. With $z=-p$ and $c=0$, we find $\sim p = p+(-p-p)$. Hence with $z=-p$ and arbitrary c, $p+(-p+c) = \sim p+(p+c)$. If $p=0$ or 1, iv) follows from Theorems A and B iii); if $p \neq 0, 1$, take $m=1$ in Theorem B ii). Finally, by Theorem E i), $(b+[\{a+(x-a)\}-b])+(b+a) = b+[\{a+(x-a)\}+a] = b+(a+x)$, whence we obtain v).

Theorem 2.2. i) $LN = MN$, ii) $MN \subseteq \{1, 0\}$, iii) *if* $1 \in MN$ then $1+1=0$.

Proof. i) If $p \in MN$, then for all z, c, $p+(z+c) = [p+(z-p)]+[p+c]$ by Theorem E i), $=[\{p+(z-p)\}+p]+c$ since $p \in MN$, $=(p+z)+c$ by Theorem 2.1 i). Hence $p \in LN$. Conversely, if $p \in LN$ then for all z, c, $p+(z+c) = [p+(z-p)]+[p+c] = p+[(z-p)+(p+c)]$. Write $p+c=d$. Then $c=-p+d$ and $z+(-p+d) = (z-p)+d$ for all z, d. Thus $-p \in MN$, whence $p \in MN$. ii) Suppose that $a \in MN$, $a \neq 0, 1$. Choose $b \neq 0, 1, -a, -a+1$ and write $b+a=u$; then $u \neq 0, 1$. Since $a \in MN$ and since $(u+x)/u = u+(x-u)$ by Theorem 2.1 i), the collineation $\phi_{a,b}$ defined in Theorem E vi) gives $(x, y)\,\phi_{a,b} = (u+(x-u),\ u+y-1)$, $[m, c]\,\phi_{a,b} = [u+(m-u),\ u+c-1]$. Choose $c=-u$. Then since $\phi_{a,b}$ preserves incidence,

$$[u+(m-u)]\,[u+(x-u)] = u+(mx-u) \qquad \text{for all } m, x.$$

But by Theorem B v), $[u+(m-u)]\,[u+(x-u)] = u+[(x \circ m)-u]$ if $m, x \neq 0, 1$ and $mx \neq 1$. Hence $x \circ m = mx$ for all $m, x \neq 0, 1$ with $mx \neq 1$, and by Theorem B vii), $1+1+1=0$ and $x^2=1$ for all $x \neq 0$. In particular $a^2=1$, whence $a+a=1$ by Theorem A ii). Thus, since $a \in MN$, $1 \in MN$. But then for all x, $x-1 = -1+(1+x-1) = -1+x$ by Theorem A iv). This contradicts Theorem B viii), since $-1 \neq 1$. Hence if $a \in MN$ then $a=0$ or 1. iii) If $1 \in MN$, then $1+1 \in MN$, whence $1+1=0$ by ii).

In Theorems 2.3–2.6 we consider the case $1+1+1=0$, and in Theorems 2.7–2.10 the case $1+1=0$.

Theorem 2.3. *If* $1+1+1=0$, *then* $RN=\{0, 1, -1\}$.

Proof. By Theorem A iii), $\{0, 1, -1\}\subseteq RN$. Suppose that $a\in RN$, $a\neq 0, 1, -1$. Then for each $t\neq 0, 1, -1$ the collineation Φ_t defined in Theorem E vii) fixes the subplane π' generated by V, U, O, E and (a, a), since $t+(-t+a)=a$. By Theorem D, π' is strictly of class I 6, since it is clearly of order >4. Let R' be the subring of R which coordinatizes π'. The equation $z+(-1+x)=x$ has a solution $z\in R'$ for each $x\in R'$. If $z\neq 0, 1, -1$ then $z+(-z+x)=x$ for each $x\in R'$. Hence the required solution z is 0, 1 or -1. But certainly $z\neq 0$, and $z\neq -1$ since by Theorem A ii), $-1+(-1+x)=x+1$. Hence $1+(-1+x)=x$ for each $x\in R'$. But then, by Theorem A ii) and iii), $-1+x=x-1$ for each $x\in R'$. This contradicts Theorem B viii) in π'. Hence $RN=\{0, 1, -1\}$.

Theorem 2.4. *If* $1+1+1=0$ *and if* $s^2=1$, $s\neq 1$, *then* $s+a=a+s$ *if and only if* $a=s$, 0 *or* 1.

Proof. Certainly $s+a=a+s$ if $a=s, 0$ or 1. Suppose that $s+a=a+s$ where $a\neq s, 0, 1$. By Theorem B iv), $a^{-1}(a\circ a^{-1}s^{-1})=a+(s-a)$, $=s$ by Theorem 2.1 ii). Hence $a=as\circ sa$. Similarly $s=sa\circ as$. Since $s^2=1$, $0=(sa\circ as)\circ(sa\circ as)$, $\Rightarrow(as\circ sa)\circ(as\circ sa)=0$, whence $a^2=1$ by Theorem B iii). But then $(sa)(as)=1$, contradicting the fact that $sa\circ as=s$, $\neq 0$.

Theorem 2.5. *If* $1+1+1=0$, *then* $a+(-1-a)=b+(-1-b)$ *if and only if* $a=b$, $b+1$ *or* $b-1$.

Proof. Write $a=b+c$. By Theorem E i),

$$a+(-1-a)=b+(-1-b)\Leftrightarrow a-1=[b+(-1-b)]+a\Leftrightarrow(b+c)-1$$
$$=b+(-1+c)\Leftrightarrow c-1=-1+c,$$

since $-1\in RN$, $\Leftrightarrow c=0, 1$ or -1 by Theorem 2.4. (This argument does not apply to a general involution s, since we have used the fact that $-1\in RN$.)

Corollary. *If* $1+1+1=0$, *then* R^* *contains infinitely many involutions.*

Proof. The element -1 is an involution in R^*, whence by Theorem E ii) $a+(-1-a)$ is an involution for each $a\in R$. Since R contains infinitely many elements, R^* contains infinitely many involutions by Theorem 2.5.

Theorem 2.6. *Suppose that* $1+1+1=0$ *and that* $x^2=1$ *for all* $x\in R^*$. *Let* π_0 *be the subplane generated by* V, U, O, E. *Let* $t\neq 0, 1, -1$. *Then* i) $-t=t-1$, $\sim t=t+1$ *and* $t=-(-t)$, ii) *the collineation* Φ_t *defined in Theorem* E vii) *is of order* 3; *it fixes the elements of a subplane* π_t *such that* π_t *properly contains* π_0 *and* $(t, t)\notin\pi_t$.

Proof. i) By Theorem A ii), $t+t=1$, whence $-t=t-1$. Then $t=-(\sim t)=\sim t-1$, so that $\sim t=t+1$. Also $-(-t)=t-1-1=\sim t$. ii) The collineation Φ_t fixes V, U, O, E and hence fixes the points and lines of a subplane π_t such that $\pi_t \supseteq \pi_0$. Now Φ_t induces the automorphism τ on R, where $x \cdot \tau = t+(-t+x)$ for all x. By i) and Theorem 1.1 iii), we may also write $x \cdot \tau = -t+(\sim t+x)=\sim t+(t+x)$ for all x. On composing the three expressions for τ in a suitable order and applying Theorem A ii), we find $x \cdot \tau^3 = x$ for all x. Since $t \cdot \tau = -t, \neq t$, it follows that Φ_t has order 3 and that $(t, t) \notin \pi_t$. Let $w=t+(-1-t)$. Then

$$w \cdot \tau = -t+[-1-(-t)],$$

$=w$ by i) and Theorem 2.5. Since $w \neq 0, 1, -1, \pi_t \neq \pi_0$.

Theorem 2.7. *If* $1+1=0$ *and if* $u \in RN$, $u \neq 0, 1$, *then* $u+u=0$, $u^3=1$ *and* $RN=\{0, 1, u, u+1\}$.

Lemma. *If* $1+1=0$ *and if* $u \in RN$, $u \neq 0, 1$, *then there exists* $a \in RN$ *with* $a \neq 0, 1, a+a=0$ *and* $a^3=1$.

Proof of Lemma. Let $u \in RN$, $u \neq 0, 1$. Let π_1 be the subplane generated by $V, U, O, E, (u, u)$ and let R_1 coordinatize π_1. For each $t \neq 0, 1$ the collineation Φ_t defined in Theorem E vii) fixes the points $V, U, O, E, (u, u)$ and hence fixes all points and lines of π_1. Thus $t+(-t+y)=y$ for all $y \in R_1$ and for all $t \neq 0, 1$. Since also $1+(1+y)=y$ for all y, R_1+ has the left inverse property; then by Theorem E iv) the element 1 is in the nucleus of R_1+. If π_1 is of order 4, then it is desarguesian and $u+u=0$, $u^3=1$. Suppose π_1 is of order >4. For each $t \neq 0, 1, u, u+1$, the collineation $\Phi_{t,u}$ defined in Theorem E viii) fixes the elements of π_1. Hence, by the left inverse property, $(-t+u)+y=-t+(y+u)$ for all $y \in R_1$ and all $t \neq 0, 1, u, u+1$. For each fixed $y \in R_1$, the maps $t \to (-t+u)+y$ and $t \to -t+(y+u)$, $t \in R_1$, are bijections of R_1 onto R_1. If t takes all values in R_1 except $0, 1, u, u+1$, then $(-t+u)+y$ takes all values except $u+y$, $u+y+1$, y and $y+1$ while $-t+(y+u)$ takes all values except $y+u$, $y+u+1$, $-u+y+u$ and $-u+y+u+1$. Hence for each $y \in R_1$, either $y+u=u+y$ or $y+u=u+y+1$, since certainly $y+u \neq y$ or $y+1$. Thus by Theorem G ii) applied to u in R_1, either $u^3=1$ or $u^5=1$. In either case, by Theorem E iii), $u+u=u^{-1}+u^{-1}$, whence $u+u+u+u=0$ by Theorem A ii). Since $u^2 \neq 1$, $u+u \neq 1$. Hence either $u+u=0$ or the element $a=u+u$ satisfies the conditions $a \in RN$, $a \neq 0, 1$ and $a+a=0$. In the latter case we can apply the above argument to the element a. Thus, without loss of generality, assume that $u+u=0$. If $u^5=1$, then π_1 has order >4 and by Theorem G iii) applied to u in R_1, $u+y \neq y+u$ if $y \in R_1$ and if $y \neq 0, 1, u, u+1$. Hence $y+u=u+y+1$ for all $y \in R_1$ with $y \neq 0, 1$, $u, u+1$, i.e., since 1 is in the nucleus of R_1+, $u+(y-u)=y+1$. But if $u+(y-u)=y+1$ and if $u+(z-u)=z+1$, then $u+(y+z)-u=y+z$ by

Theorem E i) and the fact that $u \in RN$. Hence π_1 has order 8, and this contradicts Theorems D and H. Thus $u^3=1$. This completes the proof of the Lemma.

Proof of Theorem 2.7. Let $u \in RN$, $u \neq 0, 1$. By the Lemma, there exists $a \in RN$ with $a+a=0$ and $a^3=1$. Suppose that $b \in RN$, $b \neq 0, 1, a, a+1$. Let π_2 be the plane generated by $V, U, O, E, (a, a)$ and (b, b) and let R_2 coordinatize π_2. For each $t \neq 0, 1$, the collineation Φ_t fixes π_2 pointwise, whence R_2+ has the left inverse property. In particular $a+(a+y)=y$ for all $y \in R_2$, whence by Theorem G i) applied to a in R_2, $a+b \neq b+a$. For each $t \neq 0, 1, a, a+1$, the collineation $\Phi_{t,a}$ fixes $V, U, O, E, (a, a)$ and interchanges the points (b, b) and $(a+b-a, a+b-a)$. Thus $\Phi_{t,a}$ induces a Baer involution on π_2, whose fixed elements form a subplane π_3 which is maximal in π_2. Note that π_3 is independent of t, since $\Phi_{t,a}$ is completely determined in π_2 by its effect on the generating points. Let R_3 coordinatize π_3. If π_3 has order 4, then π_2 has order 16, which contradicts Theorems D and H. If π_3 has order >4, then, by the left inverse property, $(-t+a)+y = -t+(y+a)$ for all $t, y \in R_3$ with $t \neq 0, 1, a, a+1$. It follows, as in the proof of the Lemma, that $y+a=a+y$ or $a+y+1$ for all $y \in R_3$. By Theorem G iii), $y+a=a+y$ for $y \in R_3$ only if $y=0, 1, a$ or $a+1$. If $y+a= a+y+1$, $z+a=a+z+1$, then $(y+z)+a=a+(y+z)$, as before. Hence π_3 has order 8, and this again contradicts Theorems D and H. Thus RN contains only 4 elements. Since $u \in RN$ by hypothesis, $RN=\{0, 1, u, u+1\}$, $u+u=0$ and $u^3=1$.

Theorem 2.8. *If* $1+1=0$ *and if* $RN=\{0, 1, a, a+1\}$, *then* $b+a=a+b$ *if and only if* $b=0, 1, a$ *or* $a+1$.

Proof. Let $1+1=0$, $a \in RN$, $a \neq 0, 1$. Then $a+a=0$ and $a^3=1$ by Theorem 2.7. If $b=0, 1, a$ or $a+1$ then $b+a=a+b$. Suppose that $b+a= a+b$ where $b \neq 0, 1, a, a+1$. Then $a+(a+b)=a+(b+a)=(a+b)+a= (b+a)+a=b$, since $a \in RN$ and $a+a=0$. Let π_2 be the subplane generated by $V, U, O, E, (a, a)$ and (b, b), and let R_2 coordinatize π_2. The collineation Φ_a, defined in Theorem E vii), fixes the elements of π_2, since it fixes the generating points. Hence $a+(a+y)=y$ for all $y \in R_2$. But then a satisfies the conditions of Theorem G i) in R_2, whence $a+b \neq b+a$, a contradiction.

Theorem 2.9. *If* $1+1=0$ *and if* $RN=\{0, 1, a, a+1\}$, *then* $b+(a-b)= c+(a-c)$ *if and only if* $b=c$, $c+1$, $c+a$ *or* $c+a+1$.

Proof. Write $b=c+d$. By Theorem E i),

$$b+(a-b)=c+(a-c) \Leftrightarrow b+a=[c+(a-c)]+b \Leftrightarrow (c+d)+a$$
$$=c+(a+d) \Leftrightarrow d+a=a+d,$$

since $a \in RN$, $\Leftrightarrow d=0, 1, a$ or $a+1$ by Theorem 2.8.

Corollary. *If* $1+1=0$ *and if* $RN=\{0,1,a,a+1\}$, *then* R^* *contains infinitely many elements of order* 3.

Proof. By Theorem 2.7, $a^3=1$, whence, by Theorem E v), $[b+(a-b)]^3=1$ for every $b\in R$. Since R contains infinitely many elements, R^* contains infinitely many elements of order 3.

Theorem 2.10. *If* $1+1=0$ *and if* $RN=\{0,1,a,a+1\}$, *then there exists* $x\in R^*$ *with* $x^3\neq 1$.

Proof. Suppose that $RN=\{0,1,a,a+1\}$ and that $x^3=1$ for all $x\in R^*$. Let $b\in R^*$, $b\neq 1$. By Theorem E iii), $b+(b+y)=b^{-1}+(b^{-1}+y)$ for all y, whence $b+[b+\{b+(b+y)\}]=y$ for all y by Theorem A ii). Consider the collineation $\phi_{b,b}$ defined in Theorem E vi). Since $\phi_{b,b}^2$ fixes all lines on U and also fixes the lines $[0]$, $[1]$ on V, $\phi_{b,b}^2$ is the identity, whence $\phi_{b,b}$ is either the identity or an involution. Since $b+(b+y)\neq y+1$ for any y, by Theorem A ii), $\phi_{b,b}$ is neither the identity nor a Baer involution. Hence $\phi_{b,b}$ is an elation, by Baer's Theorem and Theorem F, and is necessarily a (V, UV) elation. Since $RN=\{0,1,a,a+1\}$ and since $b+b\neq 1$, $\phi_{b,b}=1^*, a^*$ or $(a+1)^*$, and for all y, $b+(b+y)=y$, $y+a+1$, or $y+a$ respectively. In particular, since $a+a=0$, $a+(a+y)=y$ for all y. By Theorem 2.9 we may choose $t\in R$ so that $t+(a-t)\neq 0,1,a,a+1$. Let $b=t+(a-t)$. Then $b+b=[t+(a-t)]+[t+(a-t)]=t+[a+(a-t)]$ by Theorem E i), $=0$. It follows that $b+(b+y)=y$ for all y. Now $b+a\neq 0,1$, and $(b+a)+(b+a)=0$, a or $a+1$. If $(b+a)+(b+a)=0$, then $(b+a)+b=a=b+(b+a)$. This contradicts Theorem G i), applied to the element b. If $(b+a)+(b+a)=a$, then $(b+a)+b=0$, contradicting the fact that $b+b=0$. Finally, if $(b+a)+(b+a)=a+1$, then $b+a=b^{-1}=b+1$. Hence there exists $x\in R^*$ with $x^3\neq 1$.

Remark. With the notation of Theorem C, the group $Z\Delta$ consists of those (V, UV) elations which are contained in Δ. By Theorems 2.3, 2.7, $|Z\Delta|\leqq 4$. If $RN=\{0,1,a,a+1\}$, then, as has been shown in the proof of Theorem 2.10, $\phi_{a,a}=1^*$. Since $\phi_{a,a}\in\Delta$, $|Z\Delta|\geqq 2$ in this case. If $RN=\{0,1\}$ and if $R+$ has the left inverse property, then $\phi_{-a,a}=1^*$ for every $a\neq 0,1$, and $|Z\Delta|=2$. In general it is not evident that $|RN|=|Z\Delta|$.

2.3. *The Multiplicative Group*

Let π be of class I 6 and let $S=f-\{F\}$. Then, with the notation of Theorem C, Δ^*, S, H_X^* satisfy condition (A) of § 1. Moreover, if $X, Y\in S$, then H_X^* is conjugate to H_Y^* in Δ^*, since H_X is conjugate to H_Y in Δ. If the coordinate system is chosen as before, $S=\{(m)\mid m\in R\}$. Thus we can identify S with R, by means of the map $(m)\to m$; in particular we write H_m^* for $H_{(m)}^*$. If $\delta\in\Delta$, let δ^* denote the permutation induced on S by δ. (Note that in this section, and in § 3.4, Greek letters which denote permutations of S will always carry an asterisk.) If $a \in R$, $a \neq 0$, let ϕ_a be the (U, OV) homology which maps (1) on (a), and if $a \in R$, $a \neq 1$, let ψ_a be the ((1), [1]) homology which maps U on (a). Then, with the notation of § 1, if $m \in R$, $m * a = m \cdot \phi_a^*$ if $a \neq 0$, and $m * 0 = 0$, while $m \circ a = m \cdot \psi_a^*$ if $a \neq 1$ and $m \circ 1 = 1$. Hence $m * a = m\, a$, and by Theorem B i), iii), $m \circ a = m + (m a \setminus a)$ for all $m, a \in R$. By Theorem B iii), if $a \neq 0, 1$ then $a \circ a^{-1} = 0$. Hence Δ^*, R, H_0^* also satisfy condition (B) of § 1, and the results of § 1 can be applied. In the present case $(\overline{H}_0^*, *)$ is the multiplicative group $(R^*, \ )$, while $(\overline{H}_1^*, 0)$ is the "circle-group" $(R^0, \circ)$ of Theorem B iii), where $R^0 = R - \{1\}$.

If $a \neq 0, 1$ and if $\chi_a = (\phi_a)^{-1} \psi_a (\phi_a)^{-1}$, $\chi_a^* = (\phi_a^*)^{-1} \psi_a^* (\phi_a^*)^{-1}$. Thus if $m \in R$, $m \cdot \chi_a^* = (m a^{-1} \circ a) a^{-1}$, and from Theorems B iv) and E ii) we find $0 \cdot \chi_a^* = 1$, $1 \cdot \chi_a^* = 0$ and $m \cdot \chi_a^* = a^{-1} + (m^{-1} - a^{-1})$ if $m \neq 0, 1$. On taking inverses, we see that the basic identities in Theorem B v) correspond to Theorem 1.1 ii), iii). By Theorem 1.2 v) and vi):

Theorem 2.11. i) *If* $a \neq 0, 1$ *then a has finite order n in* R^* *if and only if a has order n in R.* ii) *If* $b \neq 0, 1$, *the element* $a + (b - a)$ *has finite order n in* R^* *if and only if b has order n in* R^*.

Theorem 2.11 ii) is a generalization of Theorems E ii), v). Theorems 1.3–1.9 can be restated as follows, where now $a^n, a^{[n]}$ denote the n^{th} powers of $a \neq 0, 1$ in R^*, R^0 respectively. Theorem 2.12 is seen to be a generalization of Theorem B vi).

Theorem 2.12. *Let n be a positive integer. Let* $a, \in R^*$, *be an element such that* $a^r \neq 1$ *for* $0 < r \leq n$. *Then* i) $(a^n)^{[n]} = a = (a^{[n]})^n$, ii) $x^n = a \Rightarrow x^{[n]} = a^{[n]}$; $y^{[n]} = a \Rightarrow y = a^n$.

Theorem 2.13. *If* $a \neq 0, 1$ *has finite order n in* R^*, *then n is prime.*

Theorem 2.14. *If* $a \neq 0, 1$ *has finite order p in* R^* *and if* $n \not\equiv 0 \pmod{p}$, *then* i) $(a^n)^{[n]} = a = (a^{[n]})^n$, ii) $x^n = a \Rightarrow x = a^{[n]}$; $y^{[n]} = a \Rightarrow y = a^n$, iii) $a^{[n]} \in A$, *the group generated by a in* R^*.

Theorem 2.15. *If* $a \neq 0$ *and if a is not of finite order in* R^*, *then there is an isomorphism* σ *of* $(Q, +)$ *into* R^* *with* $(m/n)\sigma = (a^{[n]})^m$ *for all integers m, n with* $n \neq 0$; *the image of* $(Q, +)$ *under* σ *will be written as* $R^*(a)$.

Theorem 2.16. i) *If* $u \neq 0$, *then* $a + (a^u - a) = a^{1/u}$, ii) *if* $r, s, r+s \neq 0$, *then* $a^r \circ a^s = a^{rs/(r+s)}$.

Theorem 2.17. *If* $a, b \in R^*$, $a, b \neq 1$, *and if for some prime p*, $a^p = 1$ *but* $b^p \neq 1$, *then* $ab \neq ba$.

Theorem 2.18. *If* R^* *is abelian, then either* a) R^* *is of exponent p for some prime p or* b) R^* *contains no element of finite order and is the direct product of groups isomorphic to* $(Q, +)$.

In fact we can say a little more here if R^* is abelian.

Theorem 2.19. *Suppose that R^* is abelian.* i) *If* $1+1+1=0$ *then R^* is of exponent* 2, ii) *if* $1+1=0$ *then* $RN=\{0,1\}$ *and either R^* is of exponent p for some odd prime p or R^* is the direct product of at least two groups isomorphic to* $(Q,+)$.

Proof. i) If $1+1+1=0$, the element -1 has order 2 in R^* by Theorem E ix), whence R^* is of exponent 2. ii) If $1+1=0$ and if $a\in RN$ with $a\neq 0,1$, then by Theorem 2.7 $a^3=1$, whence R^* is of exponent 3: this is impossible by Theorem 2.10. Hence $RN=\{0,1\}$. If R^* is of exponent p, then p is odd by Theorem F. Suppose that R^* is isomorphic to $(Q,+)$. Let $a\in R^*$, $a\neq 1$. By Theorem 2.15, $R^*\supseteq R^*(a)$ where $R^*(a)$ is isomorphic to $(Q,+)$. Hence $R^*=R^*(a)$. Let $r,s\in Q-\{0\}$. The collineation ϕ_{a^s,a^r} defined in Theorem E vi) maps (x,y) on $(\{a^r+(a^s+x)\}/(a^r+a^s), a^r+(a^s+y-1))$ and $[m,c]$ on $[\{a^r+(a^s+m)\}/(a^r+a^s), a^r+(a^s+c-1)]$, and in particular fixes the points U, (1). If $m\neq 0,1$ then $m=a^t$ for some $t\neq 0$. By Theorem 2.1 v), $\{a^r+(a^s+a^t)\}/(a^r+a^s)=a^r+[\{a^s+(a^t-a^s)\}-a^r]$. By Theorem 2.16 i), $a^s+(a^t-a^s)$, $=a^s+[(a^s)^{t/s}-a^s]$, $=a^{s^2/t}$, and similarly $a^r+(a^{s^2/t}-a^r)=a^{tr^2/s^2}$. With $r=s=1$, the collineation $\phi_{a,a}$ fixes all points on VU and all lines on V, whence, since $RN=\{0,1\}$, $a+(a+y)=y$ or $y+1$ for all y. Since $a+a\neq 1$, $a+a=0$. Moreover if $r,s,u,v\in Q-\{0\}$ and if $r/s=u/v$, then $(\phi_{a^s,a^r})(\phi_{a^v,a^u})^{-1}$ fixes all points on VU and all lines on V, whence $a^r+a^s=a^u+a^v$ or a^u+a^v+1, by considering the y-coordinate. Since $a+a=0$, $a+a^2\neq 0$. Thus $a+a^2=a^s$ for some $s\in Q$; clearly $s\neq 0,\pm 1,\pm 2$. Let $r=s/2$. With $u=1$, $v=2$ in the above, $a^r+a^s=a+a^2+1$. But this yields the contradiction $a^r=0$ or 1. Hence R^* is not isomorphic to $(Q,+)$, and it follows that if R^* contains no element of finite order then R^* is the direct product of at least two groups isomorphic to $(Q,+)$.

In the excluded case $R^*\simeq(Q,+)$, it is not hard to show that (Δ^*,R,H_0^*) would be permutation-isomorphic to $(PSL(2,Q),\bar{Q},E_\infty^*(Q))$, by using Theorem 2.16 ii) and the fact that $R^*=R^*(a)$, $a\neq 0,1$. It is natural to ask whether there exists any skew-field K such that (Δ^*,R,H_0^*) is permutation-isomorphic to $(PSL(2,K),\bar{K},E_\infty^*(K))$. A partial answer is given in Theorem 2.21. We first note a more general result.

Theorem 2.20. *If* $a,b\neq 0,1$, *if* $b\neq a,a^{-1}$ *and if* $a+b=b+a$, *then* i) *a and b are conjugate in R^0*, ii) *b is of finite order p in R^* if and only if a is of finite order p in R^**, iii) $ab\neq ba$.

Proof. i) By Theorem B iv), $a^{-1}(a\circ a^{-1}b^{-1})=a+(b-a)$, $=b$ by Theorem 2.1 ii). Hence $a=ab\circ ba$, and similarly $b=ba\circ ab$. Thus a and b are conjugate in R^0. ii) By i), a and b have the same order in R^0, and hence in R^* by Theorem 2.11 i). iii) If $ab=ba$, then the proof of i) yields the contradiction $a=b$.

Theorem 2.21. *Let K be a skew-field, with centre ZK, and suppose that (Δ^*, R, H_0^*) is permutation-isomorphic to $(PSL(2,K), \bar{K}, E_\infty^*(K))$. Then* i) *if* $1+1=0$ *in* R, $ZK=GF(3)$, ii) *if* $1+1+1=0$ *in* R, $ZK=GF(2)$.

Proof. Let the associated bijection $\beta: R \to \bar{K}$ be chosen so that $1\cdot\beta=0$. By Theorem E vi), if $a, b \neq 0, 1$, Δ contains the collineation $\phi_{a,b}$, $=(\chi_b\chi_a)^{-1}$, where $(x,y)\phi_{a,b}=(\{b+(a+x)\}/(b+a),\ b+(a+y-1))$ and $[m,c]\phi_{a,b}=[\{b+(a+m)\}/(b+a), b+(a+c-1)]$. i) Suppose $1+1=0$ in R. Since R^* is abelian by Theorem 1.10 i), $RN=\{0,1\}$ by Theorem 2.19. By Theorem 1.10 ii), $(\chi_a^*)^2$ is the identity for each $a\neq 0,1$, whence either $a+(a+y)=y$ for all y or $a+(a+y)=y+1$ for all y. By Theorem F, $a+a\neq 1$. Hence $a+a=0$ for all a and $R+$ has the left inverse property. Suppose that $a, b\neq 0, 1$ and that $(a\beta)(b\beta)^{-1}\in ZK$. Then if $d\neq 0,1$ there exists $c\neq 0, 1$ such that $\chi_b^*\chi_a^*=\chi_d^*\chi_c^*$, by Theorem 1.10 iii). Thus $\phi_{a,b}^*=\phi_{c,d}^*$, whence, since $RN=\{0,1\}$, either $b+(a+y)=d+(c+y)$ for all y or $b+(a+y)=d+(c+y)+1$ for all y. In the latter case, since $1\in N$ by Theorem E iv), $b+(a+y)=d+[(c+1)+y]$ for all y. So we can assume without loss of generality that there exists $c\neq 0,1$ such that $b+(a+y)=d+(c+y)$ for all y. Setting $y=0$, $c=d+(b+a)$ by the left inverse property, whence $d+[b+(a+y)]=[d+(b+a)]+y$ for all y. With $y=a+(b+d)$, we find:

(3) *If* $a, b\neq 0, 1$ *and if* $(a\beta)(b\beta)^{-1}\in ZK$, *then* $d+(b+a)=a+(b+d)$ *for all* $d\neq 0, 1$.

Since the maps $t\to t+(b+a)$ and $t\to a+(b+t)$ are bijections of R, it follows from (3) that $b+a=a+b$ or $a+b+1$. By Theorem 2.20, $b+a\neq a+b$ if $b\neq a, a^{-1}$. Hence:

(4) *If* $a, b\neq 0, 1$, *if* $b\neq a, a^{-1}$ *and if* $(a\beta)(b\beta)^{-1}\in ZK$, *then* $b+a=a+b+1$.

Since R^* is isomorphic to $\bar{E}_\infty^*(K)$, $a^{-1}\beta=-(a\beta)$ if $a\neq 0, 1$. Calculating in K, $-(a\beta)=(a\beta)(-1)$, whence $(a\beta)(a^{-1}\beta)^{-1}=-1$. (Note that $1+1\neq 0$ in K, since R^* contains no involutions.) Suppose that there exists $z\in ZK$ with $z^2\neq 1, -1$. Let $a\in R$, $a\neq 0, 1$ and let b, c be determined by $a\beta=(b\beta)z=(c\beta)z^2$. Then, $b, c\neq 0, 1, a, a^{-1}$ and $b\neq c, c^{-1}$, so that, by (4), $b+a=a+b+1$, $b+c=c+b+1$ and $a+c=c+a+1$. Since $1\in N$, $[c+(a-c)]+[c+(b-c)]=(a+1)+(b+1)=a+b$. But by Theorem E i) and the fact that $-c=c$, $[c+(a-c)]+[c+(b-c)]=c+[a+(b+c)]$, $=c+[c+(b+a)]$ by (3), $=b+a$. Thus we have the contradiction $b+a=a+b$, and it follows that if $z\in ZK$ then $z^2=1$ or -1. This implies that the prime field of K is $GF(3)$ or $GF(5)$. In the former case, if $z\in ZK$ and $z^2=-1$, $ZK\supseteq GF(9)$, whence ZK contains an element of order 8, which is impossible; thus $ZK=GF(3)$. If $ZK=GF(5)$, let $a\in R$, $a\neq 0, 1$ and consider the collineation $\phi_{a^2,a}$. By definition $\phi_{a^2,a}^*=(\chi_a^*\chi_{a^2}^*)^{-1}=\chi_{a^{-2}}^*\chi_{a^{-1}}^*$. If $a\beta=a_1$, then

$a^{-1}\beta = -a_1$ and $a^{-2}\beta = -2a_1$, since R^* is isomorphic to $\bar{E}^*_\alpha(K)$. Hence, by (2) in § 1, if $m \neq 0, 1$ then

$$[m \cdot \phi^*_{a^2, a}]\,\beta = [(-a_1)(-2a_1)^{-1}(m\beta)(-2a_1)^{-1}(-a_1)], \qquad = -m\beta.$$

Since this is independent of a, it follows that $a+a^2 = b+b^2+1$ for all $a, b \neq 0, 1$. Let $a+a^2=t$. Then $t \neq 0, 1$ and we have the contradiction $t=t+t^2$ or $t+t^2+1$. Hence $ZK \neq GF(5)$.

ii) Suppose $1+1+1=0$ in R. Then $RN=\{0, 1, -1\}$ and R^* is of exponent 2 by Theorem 2.19, whence K is of characteristic 2. Suppose that $a, b \neq 0, 1$ and that $(a\beta)(b\beta)^{-1} \in ZK$. Then if $d \neq 0, 1$ there exists $c \neq 0, 1$ such that $\chi^*_b\,\chi^*_a = \chi^*_d\,\chi^*_c$, by Theorem 1.10 iii). Since $RN=\{0, 1, -1\}$, it follows from considering the collineations $\phi_{a,b}, \phi_{c,d}$ that either $b+(a+y)=d+(c+y)$ for all y, or $b+(a+y)=d+(c+y)+1$ for all y or $b+(a+y)=d+(c+y)-1$ for all y. In the first case $c=d+(b+a)-1$ and $d+[b+(a+y)]=[d+(b+a-1)]+y+1$ for all y, by Theorem A. With $y=a+(b+d)$, $0=[d+(b+a)-1]+[a+(b+d)+1]$, whence $d+(b+a)=a+(b+d)$ by Theorem 2.6. Similarly, in the second and third cases $d+(b+a)=a+(b+d)-1$, $a+(b+d)+1$ respectively. Hence if $d \neq 0, 1$, then $d+(b+a)=a+(b+d)+k(d)$, where $k(d)$ may depend on d and can take the values 0, 1 and -1. Define $x \sim y$ if and only if $x=y$, $y+1$ or $y-1$. Then "$\sim$" is an equivalence relation on R. The maps $t \to t+(b+a)$, $t \to a+(b+t)$ are bijections of R, and if $t \neq 0, 1$ the images of a given t belong to the same equivalence class. Moreover the excluded images $a+b$, $a+b+1$ belong to the same class. Hence $(b+a) \sim (a+b)$. Then $[-1+(b+a)] \sim [(a+b)-1] \sim (b+a)$. But $-1+(b+a) \neq b+a$ or $b+a+1$, and $-1+(b+a)=(b+a)-1$ only if $b+a=0, 1$ or -1, by Theorem 2.4. Thus if $(a\beta)(b\beta)^{-1} \in ZK$ and if $b \neq a$, then $b+a=0$ or -1. But if $(a\beta)(b\beta)^{-1}=z, \in ZK$, then for $m \neq 0, 1$, $(m \cdot \chi^*_b\,\chi^*_a)\,\beta = z(m\beta)\,z$, by (2) in § 1, whence $\chi^*_b\,\chi^*_a$ has no fixed elements $\neq 0, 1$ if $b \neq a$, since $z^2=1$ only if $z=1$. Now if $b+a=0$ or -1, $\phi^*_{a,b}, =(\chi^*_b\,\chi^*_a)^{-1}$, fixes the point (-1). Thus $(a\beta)(b\beta)^{-1} \notin ZK$ if $b+a=0$ or -1, and it follows that $ZK=GF(2)$.

We conclude this section with a theorem which gives a little information about the possible finite orders of elements in R^*.

Theorem 2.22. *Let p be an odd prime such that 2 is a generator of the multiplicative group of $GF(p)$. Then* i) *if* $1+1+1=0$, *R^* contains no element of order p,* ii) *if* $1+1=0$ *and if a has order p in R^*, then for all y, $a+(a+y)=y+u$, where $u \in RN$ and $u \neq 1$.*

Proof. i) Suppose that $1+1+1=0$ and that $a^p=1$, $a \neq 1$. By Theorem E iii) and the condition imposed on p, $a+(a+y)=a^{-1}+(a^{-1}+y)$, whence $a+[a+\{a+(a+y)\}]=y-1$ by Theorem A ii). It follows that the collineation $\phi_{a,a}$ defined in Theorem E vi) is either the identity or an involution. Since $a+(a+y)\neq y+1$ for any y by Theorem A ii), $\phi_{a,a}$ cannot be the identity or a Baer involution. Hence $\phi_{a,a}$ is an involutory homology by Baer's Theorem and Theorem F: this is impossible, since $\phi_{a,a}$ fixes $V, U, (1), [0]$ and $[1]$. ii) Suppose that $1+1=0$ and that $a^p=1, a\neq 1$. Then, as in i), $a+(a+y)=a^{-1}+(a^{-1}+y)$, whence now $a+[a+\{a+(a+y)\}]=y$ for all y, so that again $\phi_{a,a}$ is either the identity or an involution. Since $a+(a+y)\neq y+1$ for any y, $\phi_{a,a}$ cannot be the identity or a Baer involution, and hence is a (V, UV) elation. Thus $a+(a+y)=y+u$, for all y, with $u\in RN$, $u\neq 1$.

2.4. (U, OV) Homogeneity

Let π be a projective plane, and let (P, l) be a non-incident pointline pair in π. A (P, l) *duality* of π is a duality δ such that $X\delta=PX$ for all X on l and $x\delta=l\cap x$ for all x on P. π is (P, l) *homogeneous* if and only if for each Y, $\neq P$ and $\notin l$, and for each y, $\neq l$ and $\notin P$, such that $y\in PY\cap l$, there exists a (P, l) duality δ such that $Y\delta=y$. If π is (P, l) homogeneous then π is (P, l) transitive; if π is (P, l) transitive and admits a (P, l) duality then π is (P, l) homogeneous. If π is (P, l) homogeneous and if ϕ is any collineation of π, then π is also $(P\phi, l\phi)$ homogeneous (Baer [2]). For a plane of class I 6, the only possible (P, l) homogeneity pairs are the $(X, \theta(X))$ with $X\in S$; if π admits a single $(X, \theta(X))$ duality for some $X\in S$, then it is $(X, \theta(X))$ homogeneous for all $X\in S$.

Theorem 2.23. *Let R be the ternary ring of a plane of class* I 6, *represented as above. Then π is (U, OV) homogeneous if and only if* i) $R+$ *has the left inverse property,* ii) $a+a=0$ *for all a,* iii) *multiplication is abelian.*

Lemma. *Let R be the ternary ring of a projective plane π. Then π is (U, OV) homogeneous if and only if* i) *R is linear and multiplication is associative,* ii) *$R+$ has the left inverse property,* iii) $(-a)(-1)^{-1}(-b)+ba=0$ *for all a, b.*

Proof of Lemma. If π is (U, OV) homogeneous, then π is (U, OV) transitive, whence R satisfies i). Consider the (U, OV) duality δ which maps (1) on $[1]$: $V\delta=UV$, $(UV)\delta=V$, $(x, y)\delta=[x^\sigma, y]$ and $[m, c]\delta=(m^\tau, c)$, where σ, τ are bijections of R such that $0^\sigma=0=0^\tau$, $1^\tau=1$. Since incidence is preserved, $y=mx+c$ implies $c=x^\sigma m^\tau+y$, i.e. $c=x^\sigma m^\tau+(mx+c)$ for all m, x, c. With $c=0$, $x^\sigma m^\tau=\sim(mx)$ for all m, x, whence $R+$ has the left inverse property. It follows that $\sim a=-a$ for all a. Since $1^\tau=1$, $x^\sigma=-x$ for all x. With $x=1$, $(-1)m^\tau=-m$, i.e. $m^\tau=(-1)^{-1}(-m)$ for all m. Hence $(-x)(-1)^{-1}(-m)+mx=0$ for all m, x. Thus R satisfies i)–iii). Conversely, if R satisfies i)–iii) then π is (U, OV) transitive and admits the (U, OV) duality δ, where $(x, y)\delta=[-x, y]$, $[m, c]\delta=((-1)^{-1}(-m), c)$.

Proof of Theorem. Let π be of class I6, represented as above. If π is (U, OV) homogeneous then $R+$ has the left inverse property, by the Lemma, whence $-a = \sim a$ for all a and, by Theorem E iv) and Theorem 2.2, $1+1=0$. With $b=-a$ in condition iii) of the Lemma, $(-a)a+(-a)a=0$ for all a. Let $a \neq 0, 1$. Then $-a \neq a$ and by Theorem B v), $a \circ (-a) = [a+(a-a)] \circ [a+(-a-a)] = a + [(-a)a-a]$. Also, by definition, $a \circ (-a) = a + [a(-a) \diagdown -a]$. Hence $a(-a)+[(-a)a-a] = -a$, so that, by the left inverse property, $a(-a)+(-a)a=0$. Thus $a(-a) = (-a)a$. By Theorem 2.20, since a commutes with $-a$ both additively and multiplicatively, and since $-a \neq 0, 1, a^{-1}$, we have $-a=a$. Hence $a+a=0$ for all a, and by condition iii) of the Lemma, multiplication is abelian. Conversely, if R satisfies i)–iii) of Theorem 2.22, then π admits the (U, OV) duality δ with $(x, y)\delta = [x, y]$, $[m, c]\delta = (m, c)$.

3. Planes of Class III

3.1. Known Results

The plane π is *of class* III 1 if and only if i) π contains (F, f) with $F \notin f$ such that π is (X, XF) transitive for each $X \in f$, ii) π is (P, l) transitive for no other pair (P, l); π is *of class* III 2 if and only if it satisfies i) and ii)′ π is also (F, f) transitive but is (P, l) transitive for no other pair (P, l). The plane π is *at least of class III* if it satisfies i), and is *at least of class* III 2 if it satisfies i) and is also (F, f) transitive.

Theorem I (Lüneburg [12]). *Suppose π contains (F, f) with $F \notin f$ such that π is (X, XF) transitive for all $X \in f$. If $X \in f$, let E_X denote the group of all (X, XF) elations and let E_X^* denote the permutation group induced on f by E_X. Let Δ be the group generated by all E_X with $X \in f$ and let Δ^* be the permutation group induced on f by Δ. Let $Z\Delta$ denote the centre of Δ. Then* i) *Δ^* is doubly transitive on f,* ii) *E_X^* is normal in the stabilizer Δ_X^* and is transitive and regular on $f - \{X\}$,* iii) *$\Delta / Z\Delta \simeq \Delta^*$,* iv) *$Z\Delta$ consists of those (F, f) homologies which are contained in Δ.*

Note that in fact Δ is generated by any two distinct elation groups E_X, E_Y with $X, Y \in f$, $X \neq Y$ and that π is at least of class III provided that it is (X, XF) and (Y, YF) transitive for 3 non-collinear points X, Y, F.

If π is at least of class III, we may choose the coordinate quadrangle $VUOE$ so that $U=F$ and $V, O \in f$. Since π is (V, VU) transitive, the corresponding ternary ring is a cartesian group, i.e. is linear and has associative addition; π is also (O, OU) transitive. Conversely, if π is (V, VU) and (O, OU) transitive, then π is at least of class III. With this choice of coordinate quadrangle, if π is strictly of class III then neither distributive law holds; π is of class III 2 if multiplication is associative, and is otherwise of class III 1 ([14], pp. 100, 103, 104).

Let $(R, +, \ )$ be a ternary ring. In this section, since addition will now be associative and multiplication in general non-associative, we define the symbols $\diagdown, /$ as follows: if $a, b \in R$ and if $a \neq 0$, then $a \diagdown b = c$ if and only if $b = a c$ and $b/a = c$ if and only if $b = c a$. In particular, if $a \neq 0$, define $a \diagdown 1 = a^R$, $1/a = a^L$, so that $a a^R = 1 = a^L a$. In general $a^L \neq a^R$, but if $a^L = a^R$ we denote the common value by a^{-1}. If R is linear then π admits a (U, OV) homology α mapping $(1, 1)$ on $(a, 1)$, $a \neq 0$, if and only if $x(a y) = (x a) y$ for all $x, y \in R$, and in this case $(x, y)\alpha = (a x, y)$, $[m, c]\alpha = [m a^{-1}, c]$ ([14], p. 103).

Theorem J (Yaqub [16]). *Let C denote the cartesian group with respect to a chosen quadrangle $VUOE$ of a (V, VU) transitive plane π. Let $C' = [C - \{0\}] \cup \{\infty\}$, where the symbol ∞ corresponds to the line UV in the pencil on U. Let the operation $\oplus$ be defined on C' as follows: $a \oplus \infty = \infty \oplus a = a, a \oplus (-a) = \infty, (a \neq \infty), a \oplus b = a/[b \diagdown (a+b)], (a, b \neq \infty, a+b \neq 0)$. Then π is (O, OU) transitive if and only if* i) *the operation $\oplus$ is associative,* ii) *if $u, x, c \in C \cap C'$ then $u' x = u x - u c$ and $u x = (u c) d$ together imply that $(u' c) d = u' x$.*

Theorem K (Yaqub [18]). *Let π be (V, VU) and (O, OU) transitive for a given quadrangle $VUOE$, and let $(R, +, \ )$ be the corresponding ternary ring. If $a \in R$, let ϕ_a*[2] *be the (V, UV) elation which maps O on $(0, a)$ and if $a \neq 0$ let ψ_a be the (O, OU) elation which maps V on $(0, a)$. If $a \neq 0$, let $\chi_a = (\phi_a)^{-1} \psi_a (\phi_a)^{-1}$ and if $a, b \neq 0$ let $\phi_{a,b} = \chi_a \chi_b$. Then* i) *if $y \neq 0, -a$, $(0, y)\psi_a = (0, \{(y+a)/y\} \diagdown a)$, $(0, -a)\psi_a = V$,* ii) *$V\chi_a = O$, $O\chi_a = V$ and if $y \neq 0$, $(0, y)\chi_a = (-a)(y \diagdown a)$,* iii) *if $x, y, m, c \neq 0$, $(x, y)\phi_{a,b} = (\{(-a)/x\} \diagdown b, (-b)[\{(-a)(y \diagdown a)\} \diagdown b])$, $[m, c]\phi_{a,b} = [(-b)/(m \diagdown a), (-b)(\{(-a)(c \diagdown a)\} \diagdown b)]$; if x, y, m or $c = 0$, the corresponding coordinate in the image is 0.*

If π is at least of class III and is represented as above, let R^* denote the multiplicative loop of R and define $H = \{x \in R^* | x(a+b) = x a + x b$ and $x(a b) = (x a) b$ for all $a, b \in R\}$, $H^* = \{y \in R^* | (a+b) y = a y + b y$ and $(a b) y = a(b y)$ for all $a, b \in R\}$. Note that H, H^* are subgroups of R^*.

Theorem L ([18]). *Let π be at least of class* III, *represented as above.* i) *π admits a (V, OU) homology α which maps $(0, 1)$ on $(0, a)$, $a \neq 0$, if and only if $a \in H$, and in this case $(x, y)\alpha = (x, a y)$, $[m, c]\alpha = [a m, a c]$,* ii) *$\pi$ admits an (O, UV) homology β which maps $(0, 1)$ on $(0, b)$ if and only if $b \in H^*$, and in this case $(x, y)\beta = (x b, y b)$, $[m, c]\beta = [m, c b]$,* iii) *$H = H^*$,* iv) *$[(-1) v] v^L = -1$ for each $v \neq 0$.*

Theorem M (André [1]). *Let π be at least of class* III, *represented as above.* i) *If there exists $s \in H$ such that $s^2 = 1$, $s \neq 1$, then $s = -1$ and*

[2] ϕ_a is the inverse of the ϕ_a defined in [18].

$a(-1)=-a=(-1)a$ *for all* $a\in R$, ii) *if* $-1\in H$, $-1\neq 1$, *then* $(-1)^2=1$, $a(-1)=-a=(-1)a$ *for all* $a\in R$ *and addition is abelian.*

Corollary. *If* $-1\in H$, $-1\neq 1$, *then* $[a(-1)]\,b=a[(-1)\,b]$ *for all* a, b *and* π *admits the involutory* (U, OV) *homology* $(x, y)\to((-1)\,x, y)$, $[m, c]\to[m(-1), c]$.

Theorem M i) is proved in [1] under the hypothesis that multiplication is associative, but the same proof applies with the present assumption. The Corollary follows from Theorem L i)–iii) and Theorem M ii), either by direct calculation or by Ostrom's Theorem ([5], p. 120, 8 a).

3.2. A Simplified Ternary Ring Criterion

When expanded and viewed purely as ternary ring identities, the algebraic conditions in Theorem J are quite complicated. In Theorem 3.3 we derive an equivalent condition, which does not involve the additional symbol ∞ and which is somewhat simpler, particularly if multiplication is associative.

Theorem 3.1. *Let* $(C, +, \ )$ *be a cartesian group. If* $a, b\in C$ *and if* a, b, $a+b\neq 0$, *let* $a\oplus b=a/[b\diagdown(a+b)]$. *Then condition* ii) *of Theorem* J *is equivalent to:* ii)′ *if* $a, b, x\in C$ *and if* $a, b, a+b, x\neq 0$, *then*

$$(a/x)([(a+b)/x]\diagdown b)=a\oplus b.$$

Proof. Suppose first that Theorem J ii) holds. Let $a, b, a+b, x\neq 0$. Let u', u, d, c be defined by $u'x=a$, $ux=a+b=bd$, $uc=b$. Then $u'x=ux-uc$ and $ux=(uc)\,d$, whence by ii), $(u'c)\,d=u'x$. But $u'=a/x$, $u=(a+b)/x$, $c=[(a+b)/x]\diagdown b$, $d=b\diagdown(a+b)$. Hence $[(a/x)([(a+b)/x]\diagdown b)]\,[b\diagdown(a+b)]$ $=(a/x)\,x=a$, i.e. $(a/x)([(a+b)/x]\diagdown b)=a/[b\diagdown(a+b)]$, $=a\oplus b$. Thus ii) implies ii)′. Conversely, if ii)′ holds, let $u, x, c\neq 0$ and suppose that $u'x=ux-uc$, $ux=(uc)\,d$. Then $d\neq 0$, and we can assume that $u'\neq 0$, since if $u'=0$ then certainly $(u'c)\,d=u'x$. Let $a=u'x$, $b=uc$. Then $ux=a+b=bd$ and $a, b, x, a+b\neq 0$. Hence by ii)′, since $a/x=u'$, $(a+b)/x=u$, $u\diagdown b=c$ and $b\diagdown(a+b)=d$, we have $u'c=(u'x)/d$, i.e. $(u'c)\,d=u'x$.

Corollary. *If* C *satisfies Theorem* J ii), *then if* $a, b, a+b\neq 0$, $a\oplus b=a[(a+b)\diagdown b]=[a/(a+b)]\,b=[(a+b)/a]\diagdown b=(a/[y\diagdown(a+b)])(y\diagdown b)$ *for each* $y\neq 0$.

Proof. In ii)′ take $x=1, a+b, a, y\diagdown(a+b)$ respectively.

Theorem 3.2. *If* C *is a cartesian group such that* $C, C', \oplus$ *satisfy conditions* i) *and* ii) *of Theorem* J, *then* i) *if* $a, b\in C$ *and if* $a, b, a+b\neq 0$, *then* $a\oplus b=\{(-b)/[b\diagdown(a+b)]\}+b$; *if* $b, c\in C$ *and if* $b, c, b+c\neq 0$, *then* $b\oplus c=b+\{[(b+c)/b]\diagdown(-b)\}$, ii) *if* $a, b, c\in C$ *and if* $a, b, c, a+b, b+c\neq 0$, *then* $a+(b\oplus c)=0$ *if and only if* $(a\oplus b)+c=0$.

Proof. i) By Theorem J i), $(a\oplus b)\oplus(-b)=a$. By the Corollary to Theorem 3.1,

$$(a\oplus b)\oplus(-b)=(a\oplus b)([(a\oplus b)+(-b)]\searrow(-b))$$
$$=(a/[v\searrow(a+b)])([(a\oplus b)+(-b)]\searrow(-b)).$$

Hence $[(a\oplus b)+(-b)]\searrow(-b)=b\searrow(a+b)$, i.e. $a\oplus b=\{(-b)/[b\searrow(a+b)]\}+b$. The other equation can be deduced similarly from the fact that $(-b)\oplus(b\oplus c)=c$. ii) Let $a,b,c\in C$, $a,b,c,a+b,\ b+c\neq 0$. If $a+(b\oplus c)=0$, then $a\oplus(b\oplus c)=\infty$, whence $(a\oplus b)\oplus c=\infty$ by Theorem J i). Since $a+b\neq 0$, $a\oplus b\neq\infty$, whence $(a\oplus b)+c=0$. The converse can be shown similarly.

Theorem 3.3. *Let C be a cartesian group. Then C, C', $\oplus$ satisfy conditions* i) *and* ii) *of Theorem* J *if and only if C satisfies the following condition:* iii) *Let $b,x,y\in C-\{0\}$. Let $X=\{(-b)/x\}+yb$, $Y=bx+\{y\searrow(-b)\}$. If $r,s\in C-\{0\}$, then $[(bx)/r][s\searrow X]=b\Leftrightarrow[Y/r][s\searrow(yb)]=b$.*

Proof. a) We first show that Theorem J i), ii) imply iii). Let $b,x,y\in C-\{0\}$. If $x=y=1$, then $X=Y=0$, and since neither equation in iii) has a solution (r,s), iii) is satisfied vacuously. If $x=1$ and $y\neq 1$, then $[(bx)/r][s\searrow X]=b$ if and only if $r=s\searrow X$, i.e. if and only if $sr=-b+yb$. Let $c=-b+yb$. Then $c,b+c\neq 0$. By Theorem 3.2 i),

$$b\oplus c=b+\{[(b+c)/b]\searrow(-b)\},\quad =Y,$$

since $(b+c)/b=y$ and $x=1$. If $sr=c$, then

$$[Y/r][s\searrow(yb)]=[(b\oplus c)/(s\searrow c)][s\searrow(b+c)].$$

But, by the Corollary to Theorem 3.1, $b\oplus c=[b/\{s\searrow(b+c)\}][s\searrow c]$, whence $[Y/r][s\searrow(yb)]=b$. Conversely, if $r,s\neq 0$ and if $[Y/r][s\searrow(yb)]=b$, then since this equation has only one solution r for each given s, it follows that $sr=c$, whence $[(bx)/r][s\searrow X]=b$. Hence iii) is fulfilled when $x=1$, $y\neq 1$. A similar argument applies to the case $x\neq 1$, $y=1$.

Suppose that $x,y\neq 1$. Let a,c be defined by $a+b=bx$, $b+c=yb$. Then $a,c,a+b,b+c\neq 0$. By Theorem 3.2 i),

$$(a\oplus b)+c=(\{(-b)/[b\searrow(a+b)]\}+b)+(-b+yb),\quad =\{(-b)/x\}+yb=X,$$

and similarly $a+(b\oplus c)=Y$. Thus, by Theorem 3.2 ii), $X=0$ if and only if $Y=0$. If $X=Y=0$ then neither equation in iii) has a solution, and iii) is satisfied vacuously. Hence we can assume $X,Y\neq 0$. By Theorem 3.1 and its Corollary, if $r\neq 0$ then

$$a\oplus(b\oplus c)=(a/r)([\{a+(b\oplus c)\}/r]\searrow[b\oplus c])=[a/r][(Y/r)\searrow(b\oplus c)],$$

while if $s\neq 0$, $(a\oplus b)\oplus c=[(a\oplus b)/(s\searrow X)][s\searrow c]$. If $[(bx)/r][s\searrow X]=b$, then $[(a+b)/r][s\searrow X]=b$, whence $a\oplus b,\ =(a/r)([(a+b)/r]\searrow b)$ by ii)',

$=(a/r)(s\setminus X)$. Thus $[(a\oplus b)/(s\setminus X)]=(a/r)$, and $(a\oplus b)\oplus c=(a/r)(s\setminus c)$. Then, since $(a\oplus b)\oplus c=a\oplus(b\oplus c)$ by Theorem J i), $(Y/r)(s\setminus c)=b\oplus c$. But by the Corollary to Theorem 3.1, $b\oplus c=[b/\{s\setminus(b+c)\}][s\setminus c]$. Hence, recalling that $b+c=yb$, $[Y/r][s\setminus(yb)]=b$. Thus we have shown that if $x, y\neq 1$, then $[(bx)/r][s\setminus X]=b$ implies $[Y/r][s\setminus(yb)]=b$. The converse can be proved similarly.

b) Suppose now that C satisfies Theorem 3.3 iii). Note first that for given $b, x, y\neq 0$, $X=0$ if and only if $Y=0$. For if $X\neq 0$ and $Y=0$ then the first equation in iii) has solutions but the second has none; similarly iii) is contradicted if $X=0$ and $Y\neq 0$. Let the operation $\oplus$ be defined on $C'=(C-\{0\})\cup\{\infty\}$ as in Theorem J. It follows immediately from the definition that the element ∞ associates with any two elements of C' and that if $b\neq\infty$ then $(-b)\oplus[b\oplus(-b)]=-b=[(-b)\oplus b]\oplus(-b)$. Hence in verifying Theorem J i) we can assume that $a, b, c\in C-\{0\}$ and that at least one of $a+b, b+c\neq 0$.

Let $a, b\in C$, $a, b, a+b\neq 0$. Let x be defined by $a+b=bx$; then $x\neq 0, 1$. Choose $y=1$. Then $Y=bx-b=a$. Let $r, s\neq 0$, be such that $[Y/r][s\setminus(yb)]=b$, i.e. let $sr=a$. Then, by iii), $[(bx)/r][s\setminus X]=b$, whence

$$X=(a/r)([(a+b)/r]\setminus b)$$

for all $r\neq 0$. With $r=b\setminus(a+b)$, $X=a\oplus b$. Hence C satisfies ii)$'$, and equivalently Theorem J ii), and it follows that $a\oplus b$ can be written in each of the forms given in the Corollary to Theorem 3.1. By definition $X=\{(-b)/x\}+b$ (where we still assume $y=1$), whence, since $X=a\oplus b$ and $x=b\setminus(a+b)$, $[(a\oplus b)+(-b)][b\setminus(a+b)]=\{(-b)/x\}=-b$. Thus C satisfies the first identity in Theorem 3.2 i). Moreover

$$(a\oplus b)\oplus(-b)=[a\oplus b][\{(a\oplus b)+(-b)\}\setminus(-b)]=[a\oplus b][b\setminus(a+b)]=a.$$

Similarly we can show by applying iii) with $x=1$, $y\neq 1$ that C satisfies the second identity in Theorem 3.2 i), and that if $b, c, b+c\neq 0$ then $(-b)\oplus(b\oplus c)=c$.

Let $a, b, c\in C-\{0\}$, $a+b$, $b+c\neq 0$. Let x, y be defined by $bx=a+b$, $yb=b+c$; then $x, y\neq 0, 1$ and $x=b\setminus(a+b)$, $y=(b+c)/b$. By Theorem 3.2 i), $X=\{(-b)/x\}+yb=[(a\oplus b)+(-b)]+(b+c)=(a\oplus b)+c$. Similarly $Y=a+(b\oplus c)$. As noted above, iii) implies that for any $b, x, y\neq 0$, $X=0$ if and only if $Y=0$. Hence here $(a\oplus b)+c=0\Leftrightarrow a+(b\oplus c)=0$, and it follows that $(a\oplus b)\oplus c=\infty\Leftrightarrow a\oplus(b\oplus c)=\infty$. So assume $X, Y\neq 0$. Choose r, s so that $[(a+b)/r][s\setminus X]=b$. Then by iii), since $a+b=bx$ and $b+c=yb$, $[Y/r][s\setminus(b+c)]=b$. Now

$$a\oplus(b\oplus c)=[a/r][(Y/r)\setminus(b\oplus c)],$$

by ii)$'$ and the fact that $Y=a+(b\oplus c)$, and similarly

$$(a\oplus b)\oplus c=[(a\oplus b)/(s\setminus X)][s\setminus c].$$

But $(a\oplus b)/(s\diagdown X)=(a\oplus b)/([(a+b)/r]\diagdown b)$, $=a/r$ by ii)′, and similarly $(Y/r)\diagdown(b\oplus c)=s\diagdown c$ by the Corollary to Theorem 3.1. Hence $(a\oplus b)\oplus c=(a/r)(s\diagdown c)=a\oplus(b\oplus c)$. This completes the proof of Theorem 3.3.

Corollary. *Let C be a cartesian group with associative multiplication. Then C, C' satisfy conditions* i) *and* ii) *of Theorem* J *if and only if for each* $b, x, y\in C-\{0\}$, $[(-b)x^{-1}+yb]\,x=y[bx+y^{-1}(-b)]$.

Proof. If multiplication is associative, condition iii) becomes: If $b,x,y,r,s\neq 0$, then if $X=(-b)x^{-1}+yb$ and $Y=bx+y^{-1}(-b)$, $bxr^{-1}s^{-1}X=b$ if and only if $Yr^{-1}s^{-1}yb=b$. On cancelling b and eliminating $r^{-1}s^{-1}$, this reduces to $Xx=yY$.

3.3. The Distributor H

Let π be at least of class III, coordinatized as in § 3.1, and let H, H^* be defined as in § 3.1. Let R^* denote the multiplicative loop of the ternary ring R.

Theorem 3.4. *If* $-1\in H$ *and if* $-1\neq 1$, *then* i) $x^L=x^R$ *for each* $x\neq 0$, ii) *if* $b, m, x\neq 0$ *then* $b[(mx)^{-1}\diagdown b]=(b/m^{-1})(x^{-1}\diagdown b)$, iii) if $mx\neq 0$ then $(mx)^{-1}=x^{-1}m^{-1}$.

Proof. If $-1\in H$ and if $-1\neq 1$, then by Theorem L iii), $-1\in H^*$ and by Theorem M and its Corollary $(-1)^2=1$, $a(-1)=-a=(-1)a$ for all a and $[a(-1)]\,b=a[(-1)\,b]$ for all a, b. Thus -1 commutes and associates with all elements of R. i) By Theorem L iv), $x^L=x^R$ for each $x\neq 0$: denote the common value by x^{-1}. ii) For each $b\neq 0$, π admits the collineation $\phi_{-1,b}$ defined in Theorem K. By the above properties of -1, if $x, y, m, c\neq 0$ then $(x, y)\,\phi_{-1,b}=([-1][x^{-1}\diagdown b], b[y^{-1}\diagdown b])$, $[m, c]\,\phi_{-1,b}=[(b/m^{-1})(-1), b(c^{-1}\diagdown b)]$, while if x, y, m or $c=0$, the corresponding coordinate of the image is 0. Since $\phi_{-1,b}$ preserves incidence, we find with $y=mx$ that $b[(mx)^{-1}\diagdown b]=(b/m^{-1})(x^{-1}\diagdown b)$ for all $m, x, b\neq 0$. iii) In particular, with $b=x^{-1}m^{-1}$, $(x^{-1}m^{-1})[(mx)^{-1}\diagdown(x^{-1}m^{-1})]=x^{-1}m^{-1}$, whence $(mx)^{-1}=x^{-1}m^{-1}$.

Theorem 3.5. *Let* $-1\in H$, $-1\neq 1$. *If* $a, b\in R^*$, *let* $a*b=a^{-1}\diagdown b$. *Then* $(R^*, *)$ *is a (right) Bol loop, i.e.* $a*[(b*c)*b]=[(a*b)*c]*b$ *for all* $a, b, c\in R^*$.

Proof. R^* satisfies the conditions given in Theorem 3.4. If $a*b=a^{-1}\diagdown b$ for each $a, b\in R^*$, it is easily verified that $(R^*, *)$ is a loop with neutral element 1 and that $a^{-1}*a=1=a*a^{-1}$ for each a. Moreover, since $a^{-1}(a*b)=b$, $xy=z$ if and only if $y=x^{-1}*z$. By Theorem 3.4 iii), if $t=x^{-1}m^{-1}$ then $t^{-1}=mx$. Thus if $m^{-1}=x*t$ then $x=m^{-1}*t^{-1}$, whence $m^{-1}=(m^{-1}*t^{-1})*t$ for all m, t, i.e. $(R^*, *)$ has the right inverse property. Let $b, m, w\in R^*$ and let x be defined by $w=b[(mx)^{-1}\diagdown b]$. Then by

Theorem 3.4 ii), $w=(b/m^{-1})(x^{-1}\setminus b)$. If $mx=u$ then $x=m^{-1}*u$. If $b/m^{-1}=v$ then $b=vm^{-1}$, whence $b^{-1}=mv^{-1}$ by Theorem 3.4 iii) and $v^{-1}=m^{-1}*b^{-1}$. Since $w=b[u^{-1}\setminus b]$, $u*b=b^{-1}*w$, whence

$$u=(b^{-1}*w)*b^{-1}$$

by the right inverse property. Since $w=(b/m^{-1})(x^{-1}\setminus b)$, $=v(x*b)$, $x*b=v^{-1}*w$, i.e. $(m^{-1}*u)*b=(m^{-1}*b^{-1})*w$. Substituting for u and using the right inverse property, we find that $m^{-1}*[(b^{-1}*w)*b^{-1}]=[(m^{-1}*b^{-1})*w]*b^{-1}$. Thus, since b, m, w were arbitrary elements of R^*, $(R^*, *)$ is a (right) Bol loop.

Theorem 3.6. *Let π be at least of class* III 2, *represented as above. Let ZR^* denote the centre of R^*. Then* i) *if $a\in ZR^*$, $(-a)(-1)^{-1}a\in H$,* ii) *for each $a, b\in R^*$, $[-\{b^{-1}(-a)\}](-1)^{-1}(-b)\,a^{-1}\in H$.*

Proof. i) Since multiplication is associative, for each $a\in R^*$ the collineation $\phi_{-1,a}$ defined in Theorem K is given by $(x,y)\,\phi_{-1,a}=(xa, (-a)(-1)^{-1}ya)$, $[m,c]\,\phi_{-1,a}=[(-a)(-1)^{-1}m, (-a)(-1)^{-1}ca]$. Since π is (U, OV) transitive, it admits the (U, OV) homology α given by $(x,y)\,\alpha=(a^{-1}x, y)$, $[m,c]\,\alpha=[ma, c]$. Let $h=(-a)(-1)^{-1}a$. Since $a\in ZR^*$, $(x,y)\,\alpha\,\phi_{-1,a}=(x, hy)$ and $[m,c]\,\alpha\,\phi_{-1,a}=[hm, hc]$. Hence $\alpha\,\phi_{-1,a}$ is a (V, OU) homology, and $h\in H$ by Theorem L i). ii) Let $a, b\in R^*$. The collineation $\phi_{a,b}$ defined in Theorem K is given by $(x,y)\,\phi_{a,b}=(x(-a)^{-1}b, (-b)\,a^{-1}y(-a)^{-1}b)$, $[m,c]\,\phi_{a,b}=[(-b)\,a^{-1}m, (-b)\,a^{-1}c(-a)^{-1}b]$. Let $d=b^{-1}(-a)$, $h=(-d)(-1)^{-1}(-b)\,a^{-1}$ and let $\gamma=\phi_{a,b}\,\phi_{-1,a}$. Then $(x,y)\,\gamma=(x, hy)$, $[m,c]\,\gamma=[hm, hc]$, whence γ is a (V, OU) homology and $h\in H$.

Corollary. *If π is at least of class* III 2, *represented as above, and if $-1\in H$, $-1\neq 1$, then* i) *if $a\in ZR^*$, $a^2\in H$,* ii) *the derived group of R^* is contained in H.*

Proof. i) If $-1\in H$ and $-1\neq 1$ then $(-a)(-1)^{-1}a=a^2$ by Theorem M ii), whence $a^2\in H$ by Theorem 3.6 i). ii) If $a, b\in R^*$ then, since $-1\in H$, $[-\{b^{-1}(-a)\}](-1)^{-1}(-b)\,a^{-1}=b^{-1}aba^{-1}$. Hence $b^{-1}aba^{-1}\in H$ by Theorem 3.6 ii). Thus since all commutators in R^* belong to H and since H is a subgroup of R^*, the derived group of R^* is contained in H.

No example is known of a plane of class III in which $-1\in H$, $-1\neq 1$.

3.4. The Additive Group

Let π be at least of class III. Then, with the notation of Theorem I, Δ^*, f, E_X^* satisfy condition (A) of § 1. Moreover if $X, Y\in f$ then E_X^* is conjugate to E_Y^* in Δ^*, since E_X is conjugate to E_Y in Δ. If the coordinate system is chosen as before, $f=\{(0,c)\mid c\in R\}\cup V$. Thus we can identify f with $\bar{R}=R\cup\{\infty\}$, where $(0,c)$ is represented by c for all $c\in R$ and V by the additional symbol ∞; in particular we write $E_V^* = E_\infty^*$, $E_U^* = E_0^*$. Let $R' = \bar{R} - \{0\}$. If $\delta \in \Delta$, let δ^* denote the permutation induced on f by δ. If $a \in R$, let ϕ_a be the (V, UV) elation which maps O on $(0, a)$ and if $a \in R'$ let ψ_a be the (O, OU) elation which maps V on $(0, a)$. With the notation of § 1, if $\infty, 0$ are chosen as the "special" symbols there denoted by 0, 1 $y * a = y \cdot \phi_a^*$ if $a \in R$ and $y * \infty = \infty$, while $y \circ a = y \cdot \psi_a^*$ if $a \in R'$ and $y \circ 0 = 0$. Hence if $y, a \in R$ then $y * a = y + a$, and if $y, a \in R'$ then $y \circ a = y \oplus a$, where by Theorem K i), $\infty \oplus a = a \oplus \infty = a$ and if $a \neq \infty$, $(-a) \oplus a = \infty$ and $y \oplus a = \{(y+a)/y\} \setminus a$. (The last expression can also be written in the various forms given in Theorem 3.1 and its Corollary.) If $a \in \bar{R}$, $a \neq \infty, 0$, then $(\phi_a^*)^{-1} = \phi_{-a}^*$ and $(\psi_a^*)^{-1} = \psi_{-a}$, since $a \oplus (-a) = \infty$. Hence $\Delta^*, \bar{R}, \bar{E}_\infty^*$ also satisfy condition (B) of § 1. In the present case $\bar{E}_\infty^*$ is the additive group $(R, +)$ and $\bar{E}_0^*$ is the group $(R', \oplus)$, which we abbreviate to $R+$, R' respectively.

If $a \in \bar{R}$, $a \neq \infty, 0$, let $\chi_a = (\phi_a)^{-1} \psi_a (\phi_a)^{-1}$. Then by Theorem K ii), $\infty \cdot \chi_a^* = 0$, $0 \cdot \chi_a^* = \infty$ and if $y \neq \infty, 0$, $y \cdot \chi_a^* = (-a)(y \setminus a)$. Since we now have an additive notation, if $a \in \bar{R}$, $a \neq \infty, 0$ and if n is an integer, let $a n$, $a[n]$ denote the n^{th} powers of a in $R+$, R' respectively. The main theorems in § 1 can then be restated as follows, where it is assumed throughout that π is at least of class III and is represented as above.

Theorem 3.7. i) *If* $a, b, c, b+c \neq \infty, 0$, *then*

i) $(-a)[(b+c) \setminus a] = [(-a)(b \setminus a)] \oplus [(-a)(c \setminus a)]$,

ii) $(-a)[(b \oplus c) \setminus a] = [(-a)(b \setminus a)] + [(-a)(c \setminus a)]$.

Theorem 3.8. *Let* $a, b \neq \infty, 0$. *Then* i) *the element* a *has finite order* n *in* $R+$ *if and only if it has finite order* n *in* R', ii) *the element* $(-a)(b \setminus a)$ *has finite order* n *in* $R+$ *if and only if* b *has order* n *in* $R+$.

Theorem 3.9. *Let* n *be a positive integer. Let* a, $\neq \infty, 0$ *be such that* $ar \neq 0$ *for* $0 < r \leq n$. *Then* i) $(an)[n] = a = (a[n])\, n$, ii) $x n = a \Rightarrow x = a[n]$ *and* $y[n] = a \Rightarrow y = a n$.

Theorem 3.10. *If* a, $\neq \infty, 0$, *has finite order* n *in* $R+$ *then* n *is prime.*

Theorem 3.11. *If* a, $\neq \infty, 0$, *has order* p *in* $R+$ *and if* $n \not\equiv 0 \pmod p$, *then* i) $(a n)[n] = a = (a[n])\, n$, ii) $x n = a \Rightarrow x = a[n]$ *and* $y[n] = a \Rightarrow y = a n$, iii) $a[n] \in A$, *the group generated by* a *in* $R+$.

Theorem 3.12. *If* a *is not of finite order in* $R+$, *then there exists an isomorphism* σ *of* $(Q, +)$ *into* $R+$ *with* $(m/n)\sigma = (a[n])\, m$ *for all integers* m, n *with* $n \neq 0$.

Theorem 3.13. *Let* $a \neq \infty$, 0. i) *If* $u \neq 0$ (*where* $u \in Q$ *or* $GF(p)$), *then* $(-a)(a u \setminus a) = a(-1/u)$, ii) *if* $r, s, r+s \neq 0$ ($r, s \in Q$ *or* $GF(p)$), *then*

$$(a r) \oplus (a s) = a\big(r s/(r+s)\big),$$

where the last bracket is evaluated in Q *or* $GF(p)$ *respectively.*

Theorem 3.14. *If* $a, b \in R$, $a, b \neq 0$, *and if for some prime* p, $ap = 0$ *but* $bp \neq 0$, *then* $a + b \neq b + a$.

Theorem 3.15. *If* $R+$ *is abelian, then either* a) $R+$ *is of exponent* p *for some prime* p *or* b) $R+$ *is the direct sum of groups isomorphic to* $(Q, +)$.

The possibilities given in Theorem 3.15 all occur in suitable desarguesian planes. No plane of class III is known in which $R+$ is of exponent p. If π is a generalized Moulton plane over the skew-field K, with ternary ring R, then since Δ^* is generated by the permutations $x \to x + a$ and $x \to x \oplus a$, and since the operations $+, \oplus$ are identical in R and K ([16], p. 149), $(\Delta^*, \bar{R}, E^*_\infty(R))$ is permutation isomorphic to $(PSL(2, K), \bar{K}, E^*_\infty(K))$.

References

1. André, J.: Über projektive Ebenen vom Lenz-Barlotti-Typ III 2. Math. Z. **84**, 316–328 (1964).
2. Baer, R.: Homogeneity of projective planes. Amer. J. Math. **64**, 137–152 (1942).
3. Burn, R. P.: The coexistence of involutory elations and homologies. Math. Z. **103**, 195–200 (1968).
4. Cofman, J.: On the non-existence of finite projective planes of Lenz-Barlotti type I 6. J. Algebra **4**, 64–70 (1966).
5. Dembowski, P.: Finite geometries. New York: Springer 1968.
6. Hering, C., Kantor, W.M., Seitz, G.M.: Finite groups with a split *BN* pair of rank 1. To appear in J. Algebra.
7. Hering, C., Kantor, W.M.: On the Lenz-Barlotti classification of projective planes. Arch. der Math. **22**, 221–223 (1971).
8. Jagannathan, T.V.S.: The Lenz-Barlotti class I 6. Math. Z. **115**, 350–356 (1970).
9. Jónsson, W.: Transitivität und Homogenität projektiver Ebenen. Math. Z. **80**, 269–292 (1963).
10. Kurosch, A.G.: The theory of groups, vol. 2. New York: Chelsea 1960.
11. Lüneburg, H.: Endliche projektive Ebenen vom Lenz-Barlotti Typ I 6. Abh. Math. Sem. Univ. Hamburg **27**, 75–79 (1964).
12. Lüneburg, H.: Charakterisierungen der endlichen desarguesschen projektiven Ebenen, Math. Z. **85**, 419–450 (1964).
13. Moulton, F.R.: A simple non-desarguesian plane geometry. Trans. Amer. Math. Soc. **3**, 192–195 (1902).
14. Pickert, G.: Projektive Ebenen. Berlin-Göttingen-Heidelberg: Springer 1955.
15. Pickert, G.: Eine Kennzeichnung desarguesscher Ebenen. Math. Z. **71**, 99–108 (1959).
16. Yaqub, J.: On projective planes of class III. Arch. der Math. **12**, 146–150 (1961).
17. Yaqub, J.: On projective planes of Lenz-Barlotti class I 6. Math. Z. **95**, 60–70 (1966).
18. Yaqub, J.: A geometric characterization of generalized Moulton planes. Math. Z. **99**, 385–391 (1967).

Prof. J. C. D. S. Yaqub
Department of Mathematics
The Ohio State University
Columbus, Ohio 43210, U.S.A.

(Received June 14, 1971)

Math. Z. 129, 137 – 156 (1972)
© by Springer-Verlag 1972

Detecting Stable Homotopy Classes by Primary *BP* Cohomology Operations

Larry Smith and Raphael Zahler

For each odd prime p, Toda [13] has discovered an infinite family $\{\alpha_t: t \geqq 1\}$ of non-trivial elements in the p-primary component of the stable homotopy of spheres. These are complemented by the 2-primary families $\{\alpha_t: t \geqq 1\}$ and $\{\mu_{8t+1}: t \geqq 0\}$ of Adams [1; §12]. In [12; I] it was shown how to construct and detect these elements using the spectrum MU representing complex cobordism. In fact, however, the methods of [12] imply that these elements are all detected by primary operations of the Brown-Peterson spectrum BP for the prime p [2, 3, 10]. Moreover, these operations detect the entire p primary part of the image of the J-homomorphism for an odd prime p.

In [12; I] there was also introduced for each prime $p > 3$ an infinite family $\{\beta_t: t \geqq 1\}$ of non-trivial elements of order p in the p component of the stable homotopy of spheres, generalizing the elements $\{\beta_i: 1 \leqq i < p\}$ already discovered by Toda [13] for all odd primes. It can also be shown that the entire family $\{\beta_t: t \geqq 1\}$ is detected by secondary BP operations when $p > 3$, and that $\{\beta_1, \beta_2\}$ are so detected when $p = 3$ [16].

The manner in which the classes α_t are detected by BP cohomology operations is analogous to, and generalizes, the way in which Steenrod's definition of the Hopf invariant detects the element α_1 of Hopf invariant 1 mod p in the $(2p-3)$-stem for an odd prime p. Recall that in the case of Z_p cohomology we say that an element $\zeta \in \pi_r^s$ of the stable r-stem, is *detected* by the stable cohomology operation φ if when we represent ζ by a mapping

$$\zeta: S^{n+r} \to S^n,$$

the operation

$$\varphi: H^n(C(\zeta); Z_p) \to H^{n+r+1}(C(\zeta); Z_p)$$

is non-zero, where $C(\zeta)$ is the mapping cone of ζ. If this happens ζ must be essential, since all cohomology operations are zero in $S^n \vee S^{n+r+1}$.

The solution of the Hopf-invariant-one problem by Adams $(p=2)$ and Liulevicius, Shimada and Yamanoshita (p odd) implies that the only elements detected by ordinary primary cohomology operations are the Hopf maps $\eta \in \pi_1^s$, $\nu \in \pi_3^s$, $\sigma \in \pi_7^s$ of 2-primary order, and, for each odd prime p, the element $\alpha_1 \in \pi^s_{2p-3}$. In order to detect more stable homotopy elements we can try to carry through the above procedure using a generalized cohomology theory in place of Z_p cohomology. This was first explicitly done by Adams [1] and Toda[1] using complex K theory and their results were very influential on further developments.

We will use the theory given by the stable Brown-Peterson spectrum BP for a fixed prime p. This is closely related to the Thom spectrum MU, which defines complex cobordism. Localized at the prime p MU becomes a sum of suspensions of BP [5]. Hence MU and BP define the same p primary Adams spectral sequence and detect the same homotopy elements. Our choice of BP in place of MU is to fit our results better with [16] and really only a matter of taste.

The paper is divided as follows. In the first section we carefully introduce the BP Steenrod-Hopf invariant. There are some added complications not present in the classical case that must be handled precisely. Section 2 shows how this invariant detects the elements $\{\alpha_t : t \geqq 1\}$, modulo technicalities postponed to the last section. In Section 3 we connect up our invariant with the J homomorphism and the BP Adams-Novikov spectral sequence. Section 4 is devoted to applications.

§ 1. The *BP* Steenrod-Hopf Invariant

We denote by p a fixed odd prime and BP the Brown-Peterson spectrum for p [3] as presented by Quillen [10, 2]. (We shall assume familiarity with the notations of [2, 15] upon which we will draw freely.) There is [10, 2] the coalgebra $\mathscr{R}$ of BP cohomology operations, free as Q_p-module on generators r_E, where $E = (e_1, e_2, \ldots)$ runs over all finitely non-zero sequences of non-negative integers and

$$\deg r_E = 2\left(\sum_{i=1}^{\infty} (p^i - 1) e_i\right) = |E|.$$

The comultiplication is given by

$$\Delta^*(r_E) = \sum_{E_1 + E_2 = E} r_{E_1} \otimes r_{E_2},$$

addition of sequence being componentwise and $r_{(0,\ldots,0)} = 1$ by convention. The coefficient ring $\pi_* BP = BP_*$, a polynomial algebra on classes x_i of degree $2p^i - 2$, $i = 1, 2, 3, \ldots$ is an algebra over $\mathscr{R}$.

Let us suppose that we are given a mapping

$$f \colon S^{n+2t(p-1)-1} \to S^n \qquad (n, t > 0)$$

---

[1] Toda's work appeared in an article written in Japanese for the Journal Sugakū. An English translation by Mi-Soo Bae Smith is to appear soon in the journal *Advances in Mathematics*.

and let $C(f)$ be the mapping cone of f. There is then the cofibring

$$S^n \xrightarrow{i} C(f) \xrightarrow{j} S^{n+2t(p-1)}$$

from which we obtain an exact sequence

$$0 \leftarrow \widetilde{BP}^*(S^n) \xleftarrow{i^*} \widetilde{BP}^*(C(f)) \xleftarrow{j^*} \widetilde{BP}^*(S^{n+2t(p-1)}) \leftarrow 0,$$

the shortness of which follows from the fact that

$$f^*: \widetilde{BP}^*(S^n) \to \widetilde{BP}^*(S^{n+2t(p-1)-1})$$

must vanish for degree reasons. Thus $\widetilde{BP}^*(C(f))$ is seen to be a free BP^*-module on generators

$$\lambda_n \in \widetilde{BP}^n(C(f) \quad \text{and} \quad \lambda_{n+2t(p-1)} \in \widetilde{BP}^{n+2t(p-1)}(C(f)).$$

The generator $\lambda_{n+2t(p-1)}$ is uniquely determined by the requirement that $\lambda_{n+2t(p-1)} = j^* \sigma_{n+2t(p-1)}$ where $\sigma_{n+2t(p-1)} \in \widetilde{BP}^{n+2t(p-1)}(S^{n+2t(p-1)})$ is the canonical class. The generator λ_n depends on choices and is only determined up to addition of an element of the form $v\lambda_{n+2t(p-1)}$ where $v \in BP^{-2t(p-1)} = BP_{2t(p-1)}$. Choosing generators $\lambda_{n+2t(p-1)}$ and λ_n as above we find for each element $\rho \in \mathscr{R}$ of degree $2t(p-1)$ that

$$\rho(\lambda_n) = S_\rho(f)\, \lambda_{n+2t(p-1)}$$

for a unique element $S_\rho(f) \in Q_p$. The element $S_\rho(f)$ is, however, not uniquely determined by the mapping f. For if $\lambda_n' = \lambda_n + v\lambda_{n+2t(p-1)}$ is some different choice of a generator of degree $2t(p-1)$ for $\widetilde{BP}^*(C(f))$ we find by the Cartan formula

$$\begin{aligned}
\rho(\lambda_n') &= \rho(\lambda_n + v\lambda_{n+2t(p-1)}) \\
&= \rho\lambda_n + \rho(v\lambda_{n+2t(p-1)}) \\
&= \rho\lambda_n + \sum \rho' v \otimes \rho'' \lambda_{n+2t(p-1)} \\
&= S_\rho(f)\, \lambda_{n+2t(p-1)} + (\rho v)\, \lambda_{n+2t(p-1)} \\
&= (S_\rho(f) + \rho v)\, \lambda_{n+2t(p-1)}
\end{aligned}$$

where

$$\Delta^* \rho = \sum \rho' \otimes \rho''$$

and we have used the fact that

$$\rho'' \lambda_{n+2t(p-1)} \in \widetilde{BP}^{n+2t(p-1)+\deg \rho''}(C(f)) = 0 \qquad \text{for } \deg \rho'' \neq 0.$$

Introduce the cokernel of the mapping

$$\rho: BP_{2t(p-1)} \to BP_0 = Q_p$$

which is seen to be a cyclic group $Z_{p^{\phi(\rho)}}$ for some non-negative integer $\phi(\rho)$. The above equations show that the residue class $S_\rho(f) \in Z_{p^{\phi(\rho)}}$ is a well defined homotopy invariant of f. We call it the BP Steenrod-Hopf invariant of f, or just the S invariant of f, by analogy with Steenrod's definition of Hopf's invariant using primary cohomology operations in Z_2 cohomology.

The S invariant is readily seen to be stable under suspension and so passes to a function (homomorphism)

$$S_\rho\colon \pi^S_{2t(p-1)-1} \to Z_{p^{\phi(\rho)}}$$

defined on the stable stem $\pi^S_{2t(p-1)-1} = \lim_{n\to\infty} \pi_{n+2t(p-1)-1}(S^n)$. It is apparent that practical calculations with this invariant will depend on the evaluation of the number $\phi(\rho)$ for suitable operations ρ. This brings us to:

Proposition 1.1. *For each positive integer t there exists an operation $\rho_t \in \mathscr{R}$ of degree $2t(p-1)$ such that $\phi(p_t) = t$, and $\rho_t(v_1^t) = p^t$ for a suitable generator $v_1 \in \pi_{2p-2} BP \simeq Q_p$.*

The proof of (1.1) is rather technical and appears in Section 5. The first few operations ρ_t are as follows:

$$\rho_t = r_t, \qquad 1 \leqq t \leqq p;$$
$$\rho_{p+1} = r_{p+1} + p^{p-1} r_{\Delta_2};$$
$$\rho_{p+2} = r_{p+2} + p^{p-1} r_{(1,1)}.$$

Our next task is to evaluate the S invariant

$$S\colon \pi^s_{2t(p-1)-1} \to Z_p t$$

determined by the operation $\rho_t \in \mathscr{R}$. Before doing so we wish to amplify our remark that the S invariant is the BP analog of Steenrod's mod p Hopf invariant. It is clear that we have merely mimicked the definition of the Z_p Hopf invariant in BP cohomology. This analogy however, has more to recommend it than sheer formalism.

Let $\mu\colon BP \to KZ_p$ be the natural morphism of spectra. Let $\rho \in \mathscr{R}$ and write $\rho = \sum \lambda(E) r_E$, $\lambda(E) \in Q_p$. It is shown in [11, 15] that the diagram

$$\begin{array}{ccc} BP^*(X) & \xrightarrow{\rho} & BP^*(X) \\ \downarrow{\scriptstyle\mu} & & \downarrow{\scriptstyle\mu} \\ H^*(X; Z_p) & \xrightarrow{\sum \lambda(E)\chi(P^E)} & H^*(X; Z_p) \end{array}$$

commutes for all finite complexes X where P^E is the usual element of Milnor's basis for the Steenrod algebra for p odd (the usual modification thereof for $p=2$), and χ denotes the canonical antiautomorphism of the Steenrod algebra. Thus choosing $\rho=r_1$, the BP Steenrod-Hopf invariant and Z_p cohomology Steenrod-Hopf invariant are seen to agree up to a sign. Suitable other choices of ρ may be made to evaluate higher order cohomology operations in Z_p cohomology [11, 12; III].

§ 2. The S Invariant and the α Sequence

Recall the infinite family of elements $\alpha_t \in \pi^S_{2t(p-1)-1}$ introduced by Toda [13]. These elements are known to be of order p. The objective of the present section is to give another proof that the elements α_t are essential by using the BP Steenrod-Hopf invariant introduced in the preceding section. While it has been clear since the appearance of [9] that Toda's elements α_t were detected by primary cobordism operations specific formulas have been lacking. Using the results of [10, 15] it becomes apparent that primary BP cohomology operations will also detect. We make this explicit.

Theorem 2.1. *Let p be an odd prime and t a positive integer. Then* $S(\alpha_t)=p^{t-1}\in Z_{p^t}$ *for the operation* ρ_t *of* (1.1).

In more naive terms (2.1) says that the mapping α_t may be seen to be essential by comparing the action of the BP cohomology operation ρ_t in the complex

$$C(\alpha_t)=S^0\cup_{\alpha_t} e^{2t(p-1)}$$

to its action in the complex

$$S^0 \vee S^{2t(p-1)}$$

and that these two actions are fundamentally different. (Examples of such computations may be found at the end of Section 5.) In view of our discussion in Section 1 we might paraphrase (2.1) by saying that a suitable BP Steenrod-Hopf invariant of α_t is $p^{t-1} \bmod p^t$.

The proof of (2.1) is not difficult, but requires that we fix carefully our notations. We will of course draw freely on the notations of [15, 11] and [4].

We will employ the explicit construction of the mapping α_t given originally in [14], but in the context of [11; I, § 2]. Let $V(0)=S^0\cup_p e^1$ and note that duality applied to [12; I.2.1] yields a stable mapping

$$\phi_t\colon S^{2t(p-1)-1}\, V(0)\to S^0$$

such that

$$(\phi_t)^*(\sigma_0)=[\mathbf{C}P(p-1)]^t\,\gamma_{2t(p-1)}$$

where $\sigma_0 \in \widetilde{MU}^0(S^0)$ and $\gamma_{2t(p-1)} \in \widetilde{MU}^{2t(p-1)}(S^{2t(p-1)}\,V(0))$ are the canonical classes. Let [10, 2]

$$\begin{array}{ccc} MU\,Q_p & \xrightarrow{\;\varepsilon\;} & MU\,Q_p \\ & {\scriptstyle\pi}\searrow \quad \nearrow{\scriptstyle i} & \\ & BP & \end{array}$$

be Quillen's presentation of the Brown-Peterson spectrum and set $v_1 = \pi[CP(p-1)] \in \pi_{2p-2}(BP)$. (This is consistent with the choice of v_1 in (1.1).) By naturality we obtain

$$\phi_t^*(\sigma_0) = v_1^t\,\gamma_{2t(p-1)}$$

where now σ_0 and $\gamma_{2t(p-1)}$ denote the canonical classes in $\widetilde{BP}^0(S^0)$, $\widetilde{BP}^{2t(p-1)}(S^{2t(p-1)}\,V(0))$ respectively.

Introduce the following diagram:

$$\begin{array}{ccccccccc}
 & & & & S^{2t(p-1)} & = & S^{2t(p-1)} & & \\
 & & & & \downarrow{\scriptstyle e} & & \downarrow & & \\
S^{2t(p-1)-1} & \xrightarrow{\alpha_t} & S^0 & \xrightarrow{a} & C(\alpha_t) & \xrightarrow{b} & S^{2t(p-1)} & \xrightarrow{\Sigma\alpha_t} & S^1 \\
\downarrow{\scriptstyle i} & & \| & & \downarrow{\scriptstyle h} & & \downarrow & & \| \\
S^{2t(p-1)-1}\,V(0) & \xrightarrow[\phi_t]{} & S^0 & \xrightarrow[j]{} & C(\alpha_t) & \xrightarrow[k]{} & S^{2t(p-1)}\,V(0) & \xrightarrow[\Sigma\alpha_t]{} & S^1 \\
 & & & & \downarrow{\scriptstyle c} & & \downarrow & & \\
 & & & & S^{2t(p-1)+1} & = & S^{2t(p-1)+1} & &
\end{array}$$

where i is the canonical inclusion and all the remaining maps are defined by the requirement that the horizontal and vertical rows are cofibrings. The commutativity of the left-hand square may be taken as the definition of $\alpha_t \in \pi^s_{2t(p-1)-1}$ [1; 12.6].

Concerning this diagram we have the following readily verified facts:

$$\widetilde{BP}^*(S^r)\begin{cases}\text{generator:} & \sigma_r\\ \text{relations:} & \text{none,}\end{cases}$$

$$\widetilde{BP}^*(V(0))\begin{cases}\text{generator:} & \gamma_1\\ \text{relations:} & p\gamma_1 = 0,\end{cases}$$

$$\widetilde{BP}^*(C(\phi_t))\begin{cases}\text{generators: } \bar\gamma_0, \bar\gamma_{2t(p-1)+1}\\ \text{relations: } \quad A(\bar\gamma_{2t(p-1)-1})=(p, v_1^t)\\ \text{formulas: } \quad k^*\gamma_{2t(p-1)+1}=\bar\gamma_{2t(p-1)+1}\\ \qquad\qquad\quad j^*\bar\gamma_0=p\sigma_0,\end{cases}$$

$$\widetilde{BP}^*(C(\alpha_t))\begin{cases}\text{generators: } \lambda_0, \lambda_{2t(p-1)}\\ \text{relations: } \quad \text{none}\\ \text{formulas: } \quad b^*\sigma_{2t(p-1)}=\lambda_{2t(p-1)}\\ \qquad\qquad\quad a^*\lambda_0=\sigma_0\\ \qquad\qquad\quad e^*\lambda_{2t(p-1)}=p\sigma_{2t(p-1)}\\ \qquad\qquad\quad e^*\lambda_0=v_1^t\sigma_{2t(p-1)}.\end{cases}$$

With these *choices* of generators at hand we have:

Lemma 2.2. *The following formula holds in* $\widetilde{BP}^*(C(\alpha_t))$.

$$p\lambda_0-v_1^t\lambda_{2t(p-1)}=h^*(p\bar\gamma_0).$$

Proof. We have

$$\begin{aligned}e^*(p\lambda_0-v_1^t\lambda_{2t(p-1)})&=p\,e^*\lambda_0-v_1^t\,e^*\lambda_{2t(p-1)}\\&=p\,v_1^t\sigma_{2t(p-1)}-v_1^t\,p\sigma_{2t(p-1)}\\&=0.\end{aligned}$$

Therefore by exactness

$$p\lambda_0-v_1^t\lambda_{2t(p-1)}=h^*(A\bar\gamma_0+B\bar\gamma_{2t(p-1)+1}).$$

Then $B\in BP_{2t(p-1)+1}=0$ so $B=0$. Applying a^* to the preceding equation gives

$$\begin{aligned}p\sigma_0=a^*(p\lambda_0+v_1^t\lambda_{2t(p-1)})&=a^*h^*(A\bar\gamma_0)=j^*(A\bar\gamma_0)\\&=A\sigma_0\end{aligned}$$

and hence $A=p$ as required. $\square$

Lemma 2.3. *For all operations* $\rho\in\mathscr{R}$ *of positive degree,*

$$\rho(\lambda_0)=\frac{1}{p}\rho(v_1^t)\lambda_{2t(p-1)}.$$

Proof. By (2.2) we have

$$\begin{aligned}\rho(p\lambda_0-v_1^t\lambda_{2t(p-1)})&=\rho(h^*p\bar\gamma_0)\\&=p\,h^*(\rho\bar\gamma_0)=0\end{aligned}$$

since $\bar\gamma_0$ is spherical, while

$$\begin{aligned}\rho(p\lambda_0-v_1^t\lambda_{2t(p-1)})&=p\rho\lambda_0-\rho(v_1^t\lambda_{2t(p-1)})\\&=p\rho\lambda_0-\sum\rho' v_1^t\rho''\lambda_{2t(p-1)}\\&=p\rho\lambda_0-\rho(v_1^t)\lambda_{2t(p-1)}\end{aligned}$$

since $\rho''\lambda_{2t(p-1)}=0$ for ρ'' of positive degree. $\square$

Proof of (2.1). From (2.3) we obtain the formula

$$\rho_t \lambda_0 = \frac{1}{p} \rho_t(v_1^t) \lambda_{2t(p-1)}$$

and applying (1.1) we get

$$\rho_t \lambda_0 = \frac{1}{p} p^t \lambda_{2t(p-1)} = p^{t-1} \lambda_{2t(p-1)}$$

which is the desired conclusion. □

A similar analysis may be carried out mod 2 for the element $\alpha_t \in \pi_{8t-1}^s$ and $\mu_{8t+1} \in \pi_{8t+1}^s$ of order 2 [1; §12] by employing [11; II, §5]. In this way we obtain:

Theorem 2.4. *Let t be a positive integer. Then* $S(\alpha_t) = 2^{4t-1} \in Z_{2^{4t}}$ *for the operation* ρ_{4t} *of* (1.1), *and* $S(\mu_{8t+1}) = 2^{4t} \in Z_{2^{4t+1}}$ *for the operation* ρ_{4t+1} *of* (1.1). □

An application of the detection of the elements μ_{8t+1} by primary *BP* operations will be found in Section 4.

§ 3. Detecting the Image of *J*

Denote by

$$J\colon \pi_*(SO) \to \pi_*^s$$

the classical *J* homomorphism. Let *p* be an odd prime. It is known that the *p*-component of the image of *J* is zero for $* + 1 \not\equiv 0\,(p-1)$, and is cyclic of order $p^{1+v_p(t)}$ for $* = t(p-1)-1$, where $v_p(t)$ denotes the power of *p* in *t*. Let $\alpha_t' \in \pi_{2t(p-1)-1}$ be a generator of the *p*-component of image *J*. Write $1 + v_p(t) = n$. We may apply [12; II, §4] to construct a diagram (the notation with primes being less refined than [12; II, §4])

$$\begin{array}{ccc} S^{2t(p-1)} \cup_{p^n} e^{2t(p-1)+1} & \xrightarrow{\phi_t'} & S \cup_{p^n} e^1 \\ \uparrow & & \downarrow \\ S^{2t(p-1)} & \xrightarrow[\alpha_t']{} & S^1, \end{array}$$

which defines a particularly useful representative of α_t'. Here

$$(\phi_t')^*\colon \widetilde{BP}^*(S^0 \cup_{p^n} e^1) \to \widetilde{BP}^*(S^{2t(p-1)} \cup_{p^n} e^{2t(p-1)+1})$$

is given by

$$(\varphi_t')^*(\gamma_1) = v_1^t \gamma_{2t(p-1)+1}$$

for suitable generators

$$\gamma_1 \in \widetilde{BP}^1(S^0 \cup_{p^n} e^1) \text{ and } \gamma_{2t(p-1)+1} \in \widetilde{BP}^{2t(p-1)+1}(S^{2t(p-1)} \cup_{p^n} e^{2t(p-1)+1}).$$

Let the mapping cone of

$$\alpha_t' \colon S^{2t(p-1)-1} \to S^0$$

be $C(\alpha_t')$. Then $\widetilde{BP}^*(C(\alpha_t'))$ is a free BP^*-module on the two generators $\lambda_0 \in \widetilde{BP}^0(C(\alpha_t'))$, $\lambda_{2t(p-1)} \in \widetilde{BP}^{2t(p-1)}(C(\alpha_t'))$. Following the arguments of §2 it may be shown that

$$\rho_t \lambda_0 = p^{t-n} \lambda_{2t(p-1)}$$

for suitable choices of the λ's. Thus we have shown:

Theorem 3.1. *For an odd prime p and positive integer t there is*

$$\alpha_t' \in \pi^s_{2t(p-1)-1},$$

a generator of the p-component of the image of J, satisfying $S_{\rho_t}(\alpha_t') = p^{t-(1+\nu_p(t))} \in Z_{p^t}$. □

Corollary 3.2. *For an odd prime p and positive integer t the p-component of the subgroup* $\operatorname{Im} J \subset \pi^s_{2t(p-1)-1}$ *is detected by the S invariant for the operation* ρ_t.

Recall that for an odd prime $\operatorname{Im} J \otimes Q_p \subset \pi^s_* \otimes Q_p$ is a direct summand [1]. Hence any element $\alpha \in \pi^s_* \otimes Q_p$ may be written uniquely in the form

$$\alpha = \alpha_J \otimes \alpha_{J^\perp}$$

where $\alpha_J \in \operatorname{Im} J \otimes Q_p$ and $\alpha_{J^\perp}$ lies in the complimentary summand.

Proposition 3.3. *For an odd prime p and any* $\alpha \in \pi^s_* \otimes Q_p$ *we have* $S_p(\alpha) = S_p(\alpha_J) \in Z_{p^{\phi(p)}}$, $* = 2t(p-1)-1$, *and* $S_p(\alpha) = 0$, $*+1 \not\equiv 0\ (p-1)$. □

Before beginning the proof of (3.3) we return for the moment to the general situation of a mapping

$$f \colon S^{n+k-1} \to S^n,$$

where $k > 1$ and $C(f) = S^n \cup_f e^{n+k}$. One then has the exact sequence

$$0 \to \widetilde{BP}^*(S^{n+k}) \to \widetilde{BP}^*(C(f)) \to \widetilde{BP}^*(S^n) \to 0$$

of the cofibring

$$S^n \to C(f) \to S^{n+k}.$$

This sequence is split as a sequence of BP^* modules, and hence represents an element

$$E(f) \in \operatorname{Ext}^{1,k}_{BP^*(BP)}(BP^*, BP^*) = E_2^{1,k}$$

where $\{E_r, d_r\}$ denotes the BP Adams-Novikov spectral sequence [9, 15] for the sphere. The elements of $E_2^{1,*}$ represent the potential elements of $\pi^s_* \otimes Q_p$ that are detected by primary BP cohomology operations. In fact [9; §10, 11] Novikov has shown that if p is odd all elements in $E_2^{1,*}$ survive to E_∞ and represent the p-primary part of the image of the J homomorphism.

The groups

$$\operatorname{Ext}^{1,k}_{BP^*(BP)}(BP^*, BP^*)$$

have been computed by Novikov [9] and one has

$$\operatorname{Ext}^{1,k}_{BP^*(BP)}(BP^*, BP^*) \simeq \begin{cases} Z_{p^{1+\nu_p(t)}}: & k = 2t(p-1) \\ 0: & k \neq 0\,(p-1). \end{cases}$$

The invariant S_ρ can be defined for elements

$$E \in \operatorname{Ext}^{1,2t(p-1)}_{BP^*(BP)}(BP^*, BP^*)$$

exactly as in Section 1 and gives a homomorphism

$$S_\rho\colon \operatorname{Ext}^{1,2t(p-1)}_{BP^*(BP)}(BP^*, BP^*) \to Z_{p^{\Phi(\rho)}}$$

whose value on $E(f)$ is $S_\rho(f)$.

Proof of 3.3. The result is now immediate from the commutative diagram

$$\begin{array}{ccc} \pi^S_{2t(p-1)-1} & & \\ \Big\downarrow{\scriptstyle E(\ )} & \searrow^{S_\rho} & \\ \operatorname{Ext}^{1,2t(p-1)}_{BP^*(BP)}(BP^*, BP^*) & \xrightarrow[S_\rho]{} & Z_{p^{\Phi(\rho)}} \end{array}$$

and the preceding discussion. □

Theorem 3.4. *Let p be an odd prime, t a positive integer, $\rho \in \mathscr{R}$ an operation of degree $2t(p-1)$, and $s \in Z_{p^{\Phi(\rho)}}$. Then there exists an element $\alpha \in \pi^S_{2t(p-1)-1}$ such that $S(\alpha) = s \in Z_{p^{\Phi(\rho)}}$ iff s lies in the subgroup of $Z_{p^{\Phi(\rho)}}$ generated by $S_\rho(\alpha_t')$. In particular for $\rho = \rho_t$ and an element $s \in Z_{p^t}$ there is an $\alpha \in \pi^S_{2t(p-1)-1}$ with $S_{\rho_t}(\alpha) = s$ iff s lies in the subgroup generated by $S_{\rho_t}(\alpha_t') = p^{t-(1+\nu_p(t))}$.* □

From the stability of the invariant $S_\rho(\ )$ we now obtain:

Corollary 3.5. *Let p be an odd prime, t a positive integer and $s \in Z_{p^t}$. Then there exists a two-cell complex $C(f) = S^n \cup_f e^{n+2t(p-1)}$ with*

$$s = S_{\rho_t}(f) \in Z_{p^t}$$

iff s lies in the subgroup of Z_{p^t} generated by $p^{t-(1+\nu_p(t))}$. □

A similar discusssion can be carried out for the prime 2 modulo the usual technicalities and the fact that the elements $\mu_t \in \pi^S_{8t+1}$ of [1] are also detected by primary BP cohomology operations. In particular one finds for the three Hopf classes $\eta\in\pi_1^s$, $\nu\in\pi_3^s$, and $\sigma\in\pi_7^s$ that

$$S(\eta)=1\in Z_2 \qquad \text{(relative to } \rho_1)$$
$$S(\nu)=1\in Z_4 \qquad \text{(relative to } \rho_2).$$
$$S(\sigma)=1\in Z_{16} \qquad \text{(relative to } \rho_4).$$

Note that the S invariant detects the correct order for η and σ but fails to catch one power of 2 for ν as is apparent from [15; Table 2].

§ 4. Applications

In this section we will show how the preceding results may be used to deduce a result of Harris [5]. Let us suppose given a map

$$f\colon S^n\to B$$

where B is a CW complex, such that

$$f^*\colon \widetilde{BP}^*(B)\to\widetilde{BP}^*(S^n)$$

is surjective. According to Toda [13] the elements α_t may be defined on S^3. Let us suppose that $n\geqq 3$, and

$$\alpha_t\colon S^{n+2t(p-1)-1}\to S^n$$

is a mapping representing α_t. Let

$$f_t=f\alpha_t\colon S^{n+2t(p-1)-1}\to B$$

and suppose that $[f_t]=0\in\pi_{n+2t(p-1)-1}(B)$. Then we may construct an extension

$$F\colon S^n\cup_{\alpha_t} e^{n+2t(p-1)-1}\to B$$

of the mapping f. Let λ_n, $\lambda_{n+2t(p-1)}$ be generators of

$$\widetilde{BP}^*(S^n\cup_{\alpha_t} e^{n+2t(p-1)}),$$

elements chosen as in the proof of (2.1) so that $\rho_t\lambda_n=p^{t-1}\lambda_{n+2t(p-1)}$. (That this choice is possible follows at once from the fact that ρ_t is stable and such a choice is stably possible.) Let $\eta\in\widetilde{BP}^n(B)$ be a class such that $f^*\eta=\sigma_n$. Then $F^*\eta$ is also an acceptable generator of

$$\widetilde{BP}^*(S^n\cup_{\alpha_t} e^{n+2t(p-1)})$$

of degree n, so

$$F^*(\rho_t\eta)=\rho_t F^*\eta=p^{t-1}\lambda_{n+2t(p-1)}(\operatorname{mod} p).$$

In particular $F^* \rho_t \eta$ is a non-zero multiple of the class $\lambda_{n+2t(p-1)}$. Let

$$\mu_0 \colon BP^*(\ ) \to H^*(\ ; Q)$$

be the natural mapping. Then

$$F^* \mu_0(\rho_t \eta) = \mu_0 F^*(\rho_t \eta) \neq 0 \in H^{n+2t(p-1)}(S^n \cup_{\alpha_t} e^{n+2t(p-1)}; Q)$$

and hence the class $\mu_0(\rho_t \eta) \in H^{n+2t(p-1)}(B; Q)$ is a non-zero indecomposable class since $\mu_0 \lambda_{n+2t(p-1)}$ is.

Thus we have shown (compare [5; 2.1]):

Theorem 4.1. *Suppose given a mapping $f\colon S^n \to B$, $n \geqq 3$ satisfying $f^*\colon \widetilde{BP}^*(B) \to \widetilde{BP}^*(S^n)$ is surjective. Then either*

$$[f_t] = f_*[\alpha_t] \neq 0 \in \pi_{n+2t(p-1)-1}(B) \quad or \quad H^{n+2t(p-1)}(B; Q)$$

contains a non-zero indecomposable element. In particular if B is a finite complex then $[f_t] \neq 0 \in \pi_{n+2t(p-1)-1}(B)$ for sufficiently large t. □

It is not too difficult to combine the calculations of Petrie, the theorem of Hodgkin and results of [4] to show that for a compact semi-simple 1-connected Lie group G the Thom map

$$\mu\colon \widetilde{BP}^*(G) \to \tilde{H}^*(G; Q_p)$$

is surjective iff $H_*(G; Z)$ is free of p-torsion. If we now let $f\colon S^3 \to G$ be the canonical inclusion for such a Lie groups we obtain (compare [5; 3.1]):

Corollary 4.2. *Let G be a compact semi-simple 1-connected Lie group and p a prime for which $H_*(G; Z)$ is free of p-torsion. Let $f\colon S^3 \to G$ be the canonical inclusion. Then either $f_* \alpha_t \neq 0 \in \pi_{2t(p-1)+2}(G)$ or $\pi_{2t(p-1)+3}(G)$ contains an infinite cyclic summand. In particular $f_* \alpha_t \neq 0 \in \pi_{2t(p-1)+2}(G)$ for t sufficiently large.*

Proof. It only remains to recall that the Hurewicz map induces an isomorphism [8; Appendix]

$$\pi_*(G) \otimes Q \to PH_*(G; Q)$$

and that the primitives $PH_*(G; Q)$ of $H_*(G; Q)$ are dual to the indecomposables of $H^*(G; Q)$. □

Continuing to follow [5], we suppose that ξ is a vector bundle, over a complex B, that is orientable for BP cohomology theory (for example a complex bundle). Let

$$\sigma\colon S^n \to T\xi$$

be the orientation class. Of course

$$\sigma^*\colon \widetilde{BP}^*(T\xi) \to \widetilde{BP}^*(S^n)$$

is surjective. Hence (assuming $n \geqq 3$) we find either

$$\sigma_*(\alpha_i) \neq 0 \in \pi_{n+2i(p-1)-1}(T\xi)$$

or

$$QH^{n+2i(p-1)}(T\xi; Q) \neq 0.$$

A particularly interesting application occurs for $p=2$ where we use the fact that the classes $\mu_{8j+1} \in \pi_{8j+1}^s$ are detected by a BP cohomology operation $\rho_{4j+1} \in \mathscr{R}$ of degree $8j+2$.

Let ξ be the canonical bundle over $BSp(k)$ and

$$S^{4k} \to MSp(k)$$

the canonical mapping. Since ρ_{4i+1} detects μ_{8i+1} we may (assuming k large, so that μ_{8i+1} is defined on S^{4k}) conclude as above that

$$\mu_{8i+1} \neq 0 \in \pi_{4k+8i+1}(MSp(k)),$$

or

$$QH^{4K+8i+2}(MSp(k); Q) \neq 0.$$

But

$$H^*(MSp(k); Q) = 0, \quad * \not\equiv 0(4)$$

and hence we conclude that

$$\mu_{8i+1} \neq 0 \in \pi_{4k+8i+1}(MSp(k)).$$

Letting $k \to \infty$ we obtain the well-known fact

Theorem 4.3. *Let $S \to MSp$ be the natural map of spectra inducing $\Omega_*^{fr} \xrightarrow{\Phi} \Omega_*^{Sp}$. Then $\Phi(\mu_{8i+1}) \neq 0 \in \Omega_{8i+1}^{Sp}$ for all $i > 0$.* □

§ 5. Construction of Operations

The present section is devoted to the proof of Theorem 1.1. We will require a number of preliminary steps. The proof that we will give is constructive; however, we have been unable to write a closed formula for the operations ρ_t that we obtain. Perhaps some alternate construction will yield similar operations with more readily useful formulas.

Notations. We choose classes $v_i \in \pi_{2p^i-2} BP$ such that

$$\pi_*(B) \simeq Q_p[v_1, v_2, \ldots].$$

The Milnor criteria give

$$r_{\Delta_n} v_n = p$$

and

$$r_E v_n \equiv 0 \bmod p \colon |E| = 2(p^n - 1).$$

For the sake of computing some examples later in the section we will suppose that $v_1, v_2, \ldots$ are as in [7, 6] so that

$$h(v_r) = p v_r - \sum_{0<s<r} h(v_{r-s})^{p^s} m_s$$

where

$$h\colon \pi_* BP \to H_*(BP; Z)$$

is the Hurewicz map,

$$H_*(BP; Z) \simeq Q_p[m_1, m_2, \ldots],$$

and as usual [2, 15]

$$m_i = \frac{1}{p^i} h\pi[CP(p^i - 1)].$$

Lemma 5.1. *For* $|E| = 2q(p-1)$,

$$r_E(v_1^q) = \begin{cases} p^q\colon & E = q\Delta_1 \\ 0\colon & \textit{otherwise.} \end{cases}$$

Proof. By the Cartan formula

$$r_E(v_1^q) = \sum_{E_1 + \cdots + E_q = E} r_{E_1} v_1 \ldots r_{E_q} v_1,$$

while degree considerations show that $\deg F = 2p-2$ iff $F = \Delta_1$, in which case $r_{\Delta_1} v_1 = p$. Hence we find

$$r_{E_1} v_1 \ldots r_{E_q} v_1 = 0$$

unless each $E_i = \Delta_1$, in which case

$$E = \Delta_1 + \cdots + \Delta_1 = q\Delta_1$$

and

$$r_E(v_1^q) = (r_{\Delta_1} v_1)^q = p^q$$

as required. □

Definition. If $E = (e_1, \ldots, e_n)$ and $F = (f_1, \ldots, f_n)$ are index sequences with $|E| = |F|$ then $E > F$ iff there is an integer k such that

$$e_i = f_i \quad (i < k),$$

and

$$e_k > f_k.$$

(This is just the standard lexicographic ordering.)

Example. If $|F| = 2t(p-1)$, $F \neq t_{\Delta_1}$, then $F < t_{\Delta_1}$.

Proposition 5.2. *If* E *and* F *are index sequences of the same degree,* $|E| = 2t(p-1) = |F|$, *with* $E > F$, *then* $r_F v^E = 0$ *(here* $v^E = v_1^{e_1} v_2^{e_2} \ldots v_n^{e_n}$*).*

Proof. Let us proceed by induction on t. The case $t=1$ is trivial allowing us to start the induction. Set

$$E=(e_1, \ldots, e_n)$$
$$F=(f_1, \ldots, f_n).$$

Suppose first that $e_1 > f_1$. Applying the Cartan formula gives

$$r_F v^E = \sum_{F_1+F_2=F} (r_{F_1} v_1^{e_1})(r_{F_2} v^{E-e_1\Delta_1}).$$

By Lemma 5.1 $r_{F_1} v_1^{e_1}=0$ unless $F_1=e_1\Delta_1$. But if $F_1=e_1\Delta_1$ then $f_1 \geqq e_1$ contrary to assumption. Hence

$$r_{F_1} v_1^{e_1}=0 \quad \text{for all } F_1 \text{ with } F_1+F_2=F,$$

so that

$$r_F v^E = \sum 0 \cdot r_{F_2} v^{E-e_1\Delta_1}=0$$

as required. Having handled the case $e_1 > f_1$ we may safely suppose that $e_1=f_1$. Hence there exists a *positive* integer l such that

$$e_i=f_i\colon 1 \leqq i \leqq l$$
$$e_{l+1} > f_{l+1}.$$

There are two cases to consider. The first of these is when $e_i=0=f_i$: $1 \leqq i \leqq l$. Then we set

$$E'=e_{l+1}\Delta_{l+1},$$
$$E''=(0, \ldots, 0, e_{l+2}, \ldots).$$

Applying the Cartan formula we get

$$r_F v^E = r_F(v^{E'} v^{E''})$$
$$= \sum_{F'+F''=F} r_{F'} v_{l+1}^{e_{l+1}} r_{F''} v^{E''}.$$

Now as in the proof of Lemma 5.1 we obtain

$$r_{F'} v_{l+1}^{e_{l+1}}=0$$

unless

$$F'=e_{l+1} F'''.$$

Since

$$F'=(0, 0, \ldots, 0, f'_{l+1}, f'_{l+2}, \ldots)$$

degree considerations show this is possible iff $f'_{l+1}=e_{l+1}$. But then $f_{l+1} \geqq f'_{l+1}=e_{l+1}$ contrary to assumption. Therefore $r_{F'} v_{l+1}^{e_{l+1}}=0$ for all F' such that $F=F'+F''$, giving

$$r_F v^E = \sum r_{F'} v_{l+1}^{e_{l+1}} r_{F''} v^{E''}=0$$

as required.

In the remaining case there is an integer j with $1 \leqq j \leqq l$ and $e_j = f_j \neq 0$. Set

$$E' = (e_1, \ldots, e_l)$$
$$E'' = (0, \ldots, 0, e_{l+1}, \ldots, e_n).$$

Applying the Cartan formula we get

$$r_F v^E = r_F v^{E'} v^{E''} = \sum_{F'+F''=F} r_{F'} v^{E'} r_{F''} v^{E''}.$$

Now note that $F' \leqq E'$ for each F' with $F' + F'' = F$, and $\deg v^{E'} < 2t(p-1)$ since $l > 0$. Hence by the inductive assumption $r_{F'} v^{E'} = 0$ unless $F' = E'$, so we obtain

$$r_F v^E = r_{E'} v^{E'} r_{F-E'} v^{E''}.$$

But then $F - E' < E''$ and since $\deg v^{E''} < 2t(p-1)$ again because $l > 0$ we obtain $r_{F-E'} v^{E''} = 0$ from the inductive assumption. Therefore

$$r_F v^E = r_{E'} v^{E'} \cdot 0 = 0$$

as required. □

Lemma 5.3. *Let $F = (0, 0, \ldots, 0, f_n, f_{n+1}, \ldots, f_m)$ be an index sequence with $|F| = 2q(p^n - 1)$. Then*

$$r_F v_n^q = \begin{cases} p^q: & F = q\Delta_n \\ 0: & \text{otherwise}. \end{cases}$$

Proof. This is similar to Lemma 5.1. One applies the Cartan formula to get

$$r_F v_n^q = \sum_{F_1 + \cdots + F_q = F} r_{F_1} v_n \ldots r_{F_q} v_n.$$

By Proposition 5.2 $r_{F_i} v_n = 0$ unless $F_i \geqq \Delta_n$. Hence $F_i = (0, \ldots, 0, f_{i,n}, \ldots, f_{i,m})$ must have $f_{i,n} \geqq 1$ for all i or the right hand side will vanish. Degree considerations then force $F_i = \Delta_n$ for $i = 1, \ldots, q$, so that

$$r_F v_n^q = 0: \ F \neq q\Delta_n$$

while the Cartan formula gives

$$r_{q\Delta_n} v_n^q = (r_{\Delta_n} v_n)^q = p^q$$

as required. □

Lemma 5.4. *Let F be an index sequence with $|F| = 2t(p-1)$, and set $\|F\| = f_1 + f_2 + \cdots$. Then $r_F v^F = p^{\|F\|}$.*

Proof. We proceed by induction on t. The case $t = 1$ is as usual routine so we pass to the inductive step. Now suppose that $F = (0, \ldots, 0, f_n, \ldots, f_m)$.

Then by the Cartan formula

$$
\begin{aligned}
r_F v^F &= r_F(v_n^{f_n} v^{F-f_n\Delta_n}) \\
&= \sum_{F'+F''=F} r_{F'}(v_n^{f_n})\, r_{F''}(v^{F-f_n\Delta_n}) \\
&= r_{f_n\Delta_n}(v_n^{f_n})\, r_{F-f_n\Delta_n}(v^{F-f_n\Delta_n}) \\
&= p^{f_n}\, r_{F-f_n\Delta_n} v^{F-f_n\Delta_n},
\end{aligned}
$$

completing the proof by induction. □

Lemma 5.5. *Let E be an index sequence with* $|E|=2t(p-1)$ *and suppose that* $\rho\in\mathscr{R}$ *is an operation of degree* $2t(p-1)$. *Then* $\rho v^E \equiv 0 \bmod p^{\|E\|}$.

Proof. By repeated application of the Cartan formula we find that ρv^E is a sum of terms

$$(\sigma_1 v_1)^{e_1}\dots(\sigma_n v_n)^{e_n}$$

and recalling that each term $\sigma_i v_i$ is divisible by p we get

$$(\sigma_1 v_1)^{e_1}\dots(\sigma_n v_n)^{e_n} \equiv 0 \bmod p^{e_1} p^{e_2}\dots p^{e_n}$$

from which the result follows. □

Proof of Theorem 1.1. The monomials $v^E=v_1^{e_1}\dots v_n^{e_n}$ with indices E of degree $2t(p-1)$ from a basis for the free Q_p-module $\pi_{2t(p-1)}BP$. Let the elements of this basis be

$$v^{E_1}, v^{E_2}, \dots, v^{E_s};\ E_1>E_2>\dots>E_s$$

in the lexicographic ordering. We will construct inductively operations $\rho_{t,q}$, $q=1,\dots,s$, of degree $2t(p-1)$ such that

$$\rho_{t,q} v^{E_r} = \begin{cases} p^t: & r=1 \\ 0: & 1<r<q. \end{cases}$$

Since $E_1=t\Delta_1$ we may begin by setting $\rho_{t,1}=r_t$ since $r_t v_1^t=p^t$ by Lemma 5.1. Assuming inductively that we have constructed $\rho_{t,q}$ we proceed to construct $\rho_{t,q+1}$. Let

$$\rho_{t,q}(v^{E_{q+1}})=a p^\lambda, \qquad (a,p)=1,$$

and note that by Lemma 5.5 $\lambda \geqq \|E_{q+1}\|$. By Lemma 5.4

$$r_{E_{q+1}} v^{E_{q+1}} = p^{\|E_{q+1}\|}$$

while by Proposition 5.2

$$r_{E_{q+1}} v^{E_r} = 0:\ 1\leqq r\leqq q.$$

Now set

$$\rho_{t,q+1}=\rho_{t,q}-a p^{\lambda-\|E_{q+1}\|}\, r_{E_{q+1}}.$$

By Proposition 5.2

$$\rho_{t,q+1}\, v^{E_r} = \rho_{t,q}\, v^{E_r} \qquad 1 \leqq r \leqq q$$

while

$$\rho_{t,q+1}\, v^{E_{q+1}} = 0,$$

completing the inductive construction of $\rho_{t,q+1}$. It is clear that the required operation ρ_t may be taken to be $\rho_{t,s+1}$. □

It is perhaps instructive to examine a few examples to indicate why we have taken such care with the indeterminacies and constructions.

Example. $t = p+1$.

It is readily seen that

$$\pi_{2(p+1)(p-1)}\, BP \simeq Q_p \oplus Q_p$$

with generators for the two summands being given by v_1^{p+1} and v_2. From the formulae [7, 6]

$$h(v_2) = p\, m_2 - h(v_1)^p\, m_1$$

and [2, 15]

$$r_E\, m_n = \begin{cases} m_{n-i}\colon & E = p^{n-i}\, \Delta_i \\ 0\colon & \text{otherwise} \end{cases}$$

we find

$$r_{p+1}\, v_2 = -p^p.$$

Therefore, since

$$r_{p+1}\, v_1^{p+1} = p^{p+1}$$

the cokernel of

$$r_{p+1}\colon \pi_{2p^2-2}\, BP \to \pi_0\, BP$$

is the cyclic group Z_{p^p}, while

$$\rho = r_{p+1} - p^{p-1}\, r_{\Delta_2}\colon \pi_{2p^2-2}\, BP \to \pi_0\, BP$$

has cokernel $Z_{p^{p+1}}$.

Let us see how this would apply to detecting stable homotopy classes. Consider the complex

$$C(\alpha_{p+1}) = S^0 \cup_{\alpha_{p+1}} e^{2(p^2-1)}.$$

and choose generators $\lambda_0, \lambda_{2(p^2-1)} \in \widetilde{BP}^*(C(\alpha_{p+1}))$ as in Section 2. For the complex

$$S^0 \vee S^{2(p^2-1)}$$

let $\sigma_0 \in \widetilde{BP}^0(S^0 \vee S^{2(p^2-1)})$, $\sigma_{2(p^2-1)} \in \widetilde{BP}^{2(p^2-1)}(S^0 \vee S^{2(p^2-1)})$ be the canonical generators and set $\sigma_0' = \sigma_0 - v_2\, \sigma_{2(p^2-1)}$. Then using (2.3) and the preceding discussion we get the formulae

$$r_{p+1}\, \lambda_0 = p^p\, \lambda_{2(p^2-1)}$$

$$r_{p+1}\, \sigma_0' = p^p\, \sigma_{2(p^2-1)},$$

so that the operation r_{p+1} does not distinguish the space $C(\alpha_{p+1}) = S^0 \cup_{\alpha_{p+1}} e^{2(p^2-1)}$ from the space $S^0 \vee S^{2(p^2-1)}$, and hence r_{p+1} does not detect the homotopy element α_{p+1}.

On the other hand one may set $\lambda_0' = \lambda_0 + v_2 \lambda_{2(p^2-1)}$ in which case we find from (2.3) the formula

$$r_{p+1} \lambda_0' = (p^p - p^p) \lambda_{2(p^2-1)} = 0$$

$$r_{p+1} \sigma_0 = 0,$$

showing again that the operation r_{p+1} does not distinguish $C(\alpha_{p+1})$ from $S^0 \vee S^{2(p^2-1)}$.

Finally we see that the operation ρ_{p+1} does distinguish the space $C(\alpha_{p+1})$ from $S^0 \vee S^{2(p^2-1)}$. For

$$\rho_{p+1} \lambda_0 = p^p \lambda_{2(p^2-1)}$$

while for any $\sigma_0'' = \sigma_0 + (a v_1^{p+1} + b v_2) \cdot \sigma_{2(p^2-1)}$ one gets

$$\rho_{p+1}'' = (a p^{p+1} + b \cdot 0) \sigma_{2(p^2-1)} = a p^{p+1} \sigma_{2(p^2-1)}$$

so that ρ does distinguish $S^0 \vee S^{2(p^2-1)}$ from $C(\alpha_{p+1}) = S^0 \cup_{\alpha_{p+1}} e^{2(p^2-1)}$ and hence detects α_{p+1}.

References

1. Adams, J.F.: On the groups $J(X)$ IV. Topology **6**, 21–71 (1966).
2. Adams, J.F.: Quillen's work on formal group laws and complex Cobordism. Univ. of Chicago Notes 1970.
3. Brown, E.H., Peterson, F.P.: A spectrum whose Z_p cohomology is the algebra of reduced p^{th} powers. Topology **5**, 149–154 (1966).
4. Conner, P.E., Smith, L.: On the complex bordism of finite complexes. Publ. I.H.E.S. No. 37, 117–221 (1969).
5. Harris, B.: The Adams *e*-homomorphism and non-stable homotopy groups. Topology **9**, 337–344 (1969).
6. Hazewinkel, M.: Constructing formal groups over $Z_{(p)}$ algebras I. Preprint (Netherlands School of Economics).
7. Liulevicius, A.: On the Algebra $BP^*(BP)$. Symposium of Algebraic Topology, Lecture Notes in Math. **249**, Berlin-Heidelberg-New York: Springer 1972, S. 47–53.
8. Milnor, J.W., Moore, J.C.: On the structure of Hopf algebras. Ann. of Math. **81**, 211–264 (1965).
9. Novikov, S.P.: The methods of algebraic topology from the viewpoint of Cobordism theories. Math. U.S.S.R.-Izvestiia **1**, 827–913 (1967) (Izv. Akad. Nauk. SSSR. Seriia Matematicheskaia **31**, 855–951 (1967).)
10. Quillen, D.G.: On the formal group laws of unoriented and complex Cobordism theories. Bull. Amer. Math. Soc. **75**, 1293–1298 (1969).
11. Smith, L.: On Cobordism operations. Unpublished.
12. Smith, L.: On realizing complex bordism modules I, II and III. Amer. J. Math. **92**, 793–856 (1970); **93**, 226–263 (1971); preprint (University of Virginia).
13. Toda, H.: *p*-primary components of homotopy groups IV. Mem. Coll. Sci. Kyoto (Series A) **32**, 297–332 (1959).

14. Yamamoto, N.: Algebra of stable homotopy of Moore space. J. Math. Osaka City Univ. **14**, 45–68 (1963).
15. Zahler, R.: The Adams-Novikov spectral sequence for the spheres (to appear in Ann. of Math.).
16. Zahler, R.: Detecting stable homotopy with higher Cobordism operations. Preprint (University of California, Berkeley).

Dr. Larry Smith
Department of Mathematics
The University of Virginia
Charlottesville, Virginia 22903
USA

Dr. Raphael Zahler
Douglass College
Rutgers University
New Brunswick, New Jersey 08903
USA

Present Address
Matematisk Institut
8000 Aarhus C
Denmark

(Received April 13, 1972)

Math. Z. 129, 157–161 (1972)
© by Springer-Verlag 1972

Simple Flat Extensions. II

Wolmer V. Vasconcelos

Introduction

This note deals with the relationship between the flatness and the finite presentation of algebras over commutative rings. Specifically it is concerned with the question raised in [9] on when the flatness of the A-algebras $A[t]/I$ as an A-module implies the finite generation of I. This question was given an affirmative answer in [6] provided the subset of nonzero polynomials in I of least degree contains a regular element. In § 1 we give a relatively brief proof of this fact. § 2 contains the result that if $A[t]/I$ is an injective flat epimorphic image of A then I is finitely generated.

During the writing of this note the author benefited from conversations with W. Heinzer, J. Ohm and A. Viola-Prioli and had access to a manuscript with the details of the results announced in [6].

§ 1. The Theorem of Ohm-Rush

Let A be a commutative ring with identity, let t be an indeterminate, let I be an ideal of $A[t]$ and write $B=A[t]/I$. We shall be concerned here with conditions which ensure the $A[t]$-finiteness of I when B is a flat A-module *via* the natural map. The argument in [9] then shows that I is in this case a projective ideal.

The ideal of A generated by the coefficients of polynomials in I (the so-called content of I, notation: $c(I)$) is, in case B is A-flat, a pure ideal of A, *i.e.* $A/c(I)$ is a flat A-module (*e.g.* [4, p. 440]).

The following describes the local behaviour of I when B is A-flat and considers a fairly broad situation implying the finiteness of I.

Theorem 1. *Let $B=A[t]/I$ be flat as an A-module.*

(a) *For any prime ideal P of $A[t]$, I_P (=localization of I at P) is a principal ideal of $A[t]_P$.*

(b) *If $\operatorname{Min}(I)$ (=subset of polynomials in I of least positive degree) contains a regular element then I is finitely generated.*

Part (a) is a very special case of [8, p. 23] and has also been proved in [5] by other means. As for Part (b), it is proved in [6]. The proofs given here are somewhat shorter than its correspondents in [5, 6] and are very akin to the methods of the remarkable [8].

Proof. (a) Let P be a prime ideal of $A[t]$ and write $Q = A \cap P$. We may first localize $A[t]$ at the multiplicative set $A \setminus Q$, and thus, by abuse of notation, assume A local, of maximal ideal Q, with $A \cap P = Q$. Let (A', Q') be a henselinization of (A, Q); then A' is faithfully flat over A [3, p. 182] and thus $A'[t]$ is also faithfully flat over $A[t]$. This change of base ring allows then the substitution of A by A' and we may assume A to be henselian.

In the exact sequence

$$0 \to I \to A[t] \xrightarrow{\phi} B \to 0. \qquad (*)$$

We have that $c(I)$ is a pure ideal of A and thus $c(I) = (0)$ or A, A being a local ring. We assume $c(I) = A$, the other case being trivial. Then $\phi(P)$ is isolated over Q as $B \otimes A/Q$ is an artinian ring. According then to the Zariski Main Theorem [7, p. 41] we may write $B_f = C_f$, where C is the integral closure of A in B, $f \in C \setminus \phi(P)$ and $(\ )_f$ denotes localization with respect to the multiplicative set generated by f. As A is henselian and B is of finite type over A, C_f (and so B_f) is a finitely generated module over A. In particular the image u of t in B_f satisfies an integral equation over A, *i.e.* we have $g = t^n + a_{n-1} t^{n-1} + \cdots + a_0$ such that there is $h \notin P$ with $q = h \cdot g \in I$. Consider then the sequence

$$0 \to I/(q) \to A[t]/(q) \to B \to 0.$$

Since B_P is a finitely generated flat A-module, and $(A[t]/(q))_P$ is a free A-module on $\leqq n$ generators, $(I/(q))_P$ is finitely generated as an $A[t]_P$-module and so is I_P.

(b) Assume now that $\operatorname{Min}(I)$ contains a regular element. This means that there is $f \in I$ of least degree, with leading coefficient a regular element of A. Let C be the integral closure of A in its total ring of quotients. Then

$$0 \to I \otimes_A C \to C[t] \to B \otimes_A C \to 0$$

makes $B \otimes_A C = B'$ a flat C-algebra. It is clear that $\operatorname{Min}(I \otimes_A C)$ has the same regularity property of $\operatorname{Min}(I)$.

Let us first prove (b) under the condition that A be integrally closed and "descend" the general case.

Let $f = a \cdot t^n + \cdots + a_0$ be a regular element in $\operatorname{Min}(I)$ and let J denote the ideal generated by its coefficients. If g is another element in I we may write $a^m \cdot g = h \cdot f$ by the generalized euclidean algorithm. As $c(I) = A$, there is a polynomial g in I with $c(g) = A$. Using this polynomial in the above equation we get

$$a^m \cdot c(g) = (a^m) = c(h \cdot f) \subset c(h) \cdot c(f).$$

According to the "content formula" [2, p. 97], there is an integer k with $c(f) \cdot c(h)^k = c(f \cdot h) \cdot c(h)^{k-1}$ or $c(f) \cdot c(h) \cdot L = a^m \cdot L$ with $L = c(h)^{k-1}$. Thus $a^{-m} \cdot c(f) \cdot c(h)$ is integral over A since L contains a regular element and thus $c(f) \cdot c(h) \subset (a^m)$. But the equality $c(f) \cdot c(h) = (a^m)$ implies that $c(f) = J$ is an invertible ideal of A. With J^{-1} denoting its inverse, the elements in $J^{-1} \cdot f$ lie all in $A[t]$; since $a \cdot J^{-1} \cdot f \subset I$ and B is A-torsion-free, we have $J^{-1} \cdot f \subset I$. If J^{-1} is generated by $c_1, \ldots, c_l$ we claim that the polynomials $h_i = c_i \cdot f$ generate I. First notice that the contents $c(h_i)$'s generate the unite ideal. Let P be a prime ideal of $A[t]$ and $Q = A \cap P$. Then for some h_i (which we write simply as h) we have $c(h) \not\subset Q$. We claim that $(h)_P = I_P$. This would be clear by the content formula and the euclidean algorithm if we only knew that the image of h in $A_Q[t]$ actually lies in $\mathrm{Min}(I_Q)$. To see that this is really the case, let $g \in I_Q$; then $g = \sum a_i \cdot g_i$ $a_i \in A_Q$, $g_i \in I$. By multiplication by a high enough power of a and the euclidean algorithm we have $a^m \cdot g = q \cdot f$ for some $q \in A_Q[t]$. Thus degree $g \geqq$ degree f.

We are now ready to prove the general case of (b). Let C be the integral closure of A. By the preceding we know that $I \otimes_A C$ is finitely generated over $C[t]$. Let $g_1, \ldots, g_n$ be elements in I which generate $I \otimes_A C$ and let $I_0 = (g_1, \ldots, g_n)$. We claim that $I_0 = I$. Consider for that the exact sequence

$$0 \to I_0 \to I \to E \to 0$$

of $A[t]$-modules. Tensoring it with $C[t]$ we get $E \otimes_{A[t]} C[t] = (0)$. Also by Part (a) E is locally finitely generated. We are now in the following situation:

Lemma. *Let R' be integral over the subring R and let E be an R-module for which E_P is R_P-finitely generated for each prime P of R. Then if $E \otimes_R R' = (0)$, $E = (0)$ also.*

In fact we may assume R local; let

$$0 \to K \to F \to E \to 0$$

be a minimal projective cover of E. By tensoring this sequence with R' we get that $K \otimes_R R'$ maps onto $F \otimes_R R'$, which contradicts the Cohen-Seidenberg theorem, unless $F = (0)$.

§ 2. Simple Injective Flat Epimorphisms

In [5] an example of a simple flat epimorphic image of a local ring is given which is not of finite presentation. As we show here the difficulty resides only with the non faithfulness of the extension.

Theorem 2. *Let A be a commutative ring and let $B=A[t]/I$. Then if B is a faithful flat epimorphic image of A, I is finitely generated over $A[t]$.*

It might be possible that the conditions above imply the regularity of $\mathrm{Min}(I)$ and then the conclusion would follow from the Ohm-Rush result. The proof here relies on other ideas, however.

Proof. First notice that $c(I)=A$; otherwise, since $c(I)$ is a pure ideal, for some prime ideal P $c(I_P)=(0)$ and $A_P[t]$ would be an epimorphic image of A_P, which is impossible. Tensor the exact sequence $(*)$ with B to get

$$0\to B\otimes_A I\to B[t]\to B\otimes_A B\to 0.$$

Since $B\otimes_A B=B$, this sequence makes B a $B[t]$-algebra with the action of t given by $t\cdot b=u\cdot b$ where u is the image of t in $A[t]/I$. With this action of t there is another presentation of B

$$0\to(t-u)\to B[t]\to B\to 0.$$

By Schanuel's lemma we may then write $(t-u)\oplus B[t]\cong B\otimes_A I\oplus B[t]$, which by taking exterior powers yields $(t-u)\cong B\otimes_A I$ as $B[t]$-modules. Say that $B\otimes_A I$ is generated by $\sum b_i\otimes f_i$, $b_i\in B$, $f_i\in I$, $i=1,\dots,n$. Let now $J=\{r\in A \mid r\,u\in A\}$. According to [1, p. 206] $J\cdot B=B$, which says that $J\cdot A[t]+I=A[t]$. This enables us to write an equation

$$\sum r_j\cdot g_j+f=1 \qquad (**)$$

with $r_j\in J$, $f\in I$. We claim that $I=(f,f_1,\dots,f_n)$. Indeed, if $g\in I$, as $I\hookrightarrow B\otimes_A I$ ($A\to B$ is injective) $g=h(\sum b_i\otimes f_i)$, $h\in B[t]$. The coefficients of h and the b_i's can be written as "polynomials" in u of degree bounded by, say, k. Raising the equation $(**)$ to the power $2k$ we get an expression $p+q\cdot f=1$ where all the coefficients of p lies in J^{2k}. Thus $g=(q\cdot g)f+p\cdot g=(q\cdot g)f+p\cdot h(\sum b_i\otimes f_i)$. But $p\cdot h(\sum b_i\otimes f_i)$ is by the choice of k, a linear combination of the f_i's with coefficients in $A[t]$.

References

1. Akiba, T.: Remarks on generalized rings of quotients, III. J. Math. Kyoto Univ. **9**, 205–212 (1969).
2. Bourbaki, N.: Algèbre commutative, chap. VII. Paris: Hermann 1965.
3. Nagata, M.: Local rings. New York: Intersciences 1962.
4. Nagata, M.: Flatness of an extension of a commutative ring. J. Math. Kyoto Univ. **9**, 439–448 (1969).
5. Ohm, J., Heinzer, W.: The finiteness of I when $A[x]/I$ is flat (II). Preprint, Purdue University, 1971.
6. Ohm, J., Rush, D.: The finiteness of I when $A[x]/I$ is flat. Bull. Amer. Math. Soc. **77**, 793–796 (1971).

7. Raynaud, M.: Anneaux Locaux Henséliens. Lectures notes no. 169. Berlin-Heidelberg-New York: Springer 1970.
8. Raynaud, M., Gruson, L.: Critères de platitude et de projectivité. Inventiones Math. **13**, 1–89 (1971).
9. Vasconcelos, W. V.: Simple flat extensions. J. Algebra **16**, 105–107 (1970).

Prof. W. V. Vasconcelos
Department of Mathematics
Rutgers University
New Brunswick, New Jersey 08903
USA

(Received March 16, 1972)

Math. Z. 129, 163–171 (1972)
© by Springer-Verlag 1972

Ein Existenzsatz über maßtreue Abbildungen

Hans Martin Reimann

Maßtreue Abbildungen sind Homöomorphismen ϕ zwischen Gebieten G und G' des euklidischen Raumes R^n, welche das n-dimensionale Maß $|A|$ der meßbaren Mengen $A \subset G$ invariant lassen: $|\phi(A)| = |A|$ für alle $A \subset G$. Die vorliegende Untersuchung entstand aus der Frage, ob zwei diffeomorphe Gebiete in R^n durch eine diffeomorphe maßtreue Abbildung aufeinander abgebildet werden können.

Eingehend werden Diffeomorphismen ϕ zwischen Gebieten – und allgemeiner zwischen Mannigfaltigkeiten – konstruiert, die ein gegebenes Volumenelement τ in ein Volumenelement σ überführen; ein Sachverhalt, der sich durch die Beziehung

$$\int_{\phi(A)} \sigma = \int_A \tau$$

ausdrücken läßt, die für alle meßbaren Mengen A im Definitionsbereich von ϕ gelten soll. Im folgenden wird die Notation $\phi * \tau = \sigma$ verwendet. Wir stützen uns bei diesen Betrachtungen auf den Satz von Moser [2]:

Sind τ und σ zwei positive C^∞-Differentialformen vom Grade m (Volumenelemente) auf einer kompakten, m-dimensionalen, zusammenhängenden C^∞-Mannigfaltigkeit M, so existiert ein C^∞-Diffeomorphismus ϕ: $M \to M$, der τ in $c\sigma$ überführt ($\phi * \tau = c\sigma$), wobei die Konstante c durch $\int_M c\sigma = \int_M \tau$ gegeben ist.

Moser gibt für diesen Satz zwei Beweise an, die sich beide wesentlich auf die Kompaktheit der Mannigfaltigkeit M stützen. Aus dem zweiten Beweis ist ersichtlich, daß der Diffeomorphismus ϕ so konstruiert werden kann, daß er C^∞-differenzierbar von τ und σ abhängig ist. Der erste Beweis verwendet eine endliche Überdeckung von M durch geeignet gewählte Umgebungen und läßt sich wörtlich auf die folgende Situation übertragen: Korollar zum Satz von Moser:

Sind τ und σ zwei positive C^∞-Volumenelemente auf einer nicht notwendigerweise kompakten Mannigfaltigkeit M derart, daß $\int_M \sigma = \int_M \tau$ und daß $\tau = \sigma$ außerhalb einer kompakten zusammenhängenden Menge, so existiert ein C^∞-Diffeomorphismus ϕ: $M \to M$, der τ in σ überführt

$$(\phi * \tau = \sigma).$$

Im ersten Teil dieser Arbeit wird ein wichtiger Spezialfall behandelt: die offene n-dimensionale Einheitskugel. Es wird gezeigt, daß der Satz von Moser in diesem Falle immer noch gilt. Im zweiten und dritten Teil wird dasselbe Resultat auf Produktmannigfaltigkeiten und gewisse allgemeinere, ebenfalls nicht kompakte Mannigfaltigkeiten erweitert. Zum Abschluß werden die Ergebnisse zur Aufstellung von Existenzsätzen über maßtreue Abbildungen zwischen Gebieten in R^n und insbesondere in R^3 verwendet.

1. Ein Satz für die offene Einheitskugel

Satz 1. *Es seien τ und σ zwei positive C^∞-Volumenelemente auf*

$$B=\{x\in R^n \mid |x|<1\}, \qquad n\geqq 2,$$

welche die Gleichung $\int_B \sigma=\int_B \tau$ erfüllen. Dann existiert ein C^∞-Diffeomorphismus ϕ: $B\to B$ der τ in σ überführt: $\phi*\tau=\sigma$.

Das Euklidsche Volumenelement wird mit dm bezeichnet. σ läßt sich dann als $f\,dm$ schreiben, mit $f\in C^\infty(B)$. Es genügt offenbar, daß man einen Diffeomorphismus ϕ_σ von B_d -- der offenen Kugel vom Radius d – auf B konstruiert, welcher für alle meßbaren Mengen $A\subset B_d$ der Gleichung

$$\int_{\phi_\sigma(A)} \sigma=\int_A dm$$

genügt. d ist so bestimmt worden, daß $\int_{B_d} dm=\int_B \sigma$ $(0<d\leqq\infty)$. Der gesuchte Diffeomorphismus ϕ ergibt sich dann durch Komposition $\phi=\phi_\sigma\circ(\phi_\tau)^{-1}$.

f kann als Produkt zweier positiver C^∞-Funktionen g und h dargestellt werden derart, daß folgende Bedingungen erfüllt sind:

$$g=\begin{cases}1 & \text{in einer Umgebung von } x=0\\ f & \text{für } |x|\geqq\frac{1}{2}\end{cases}$$

$$\int_B g\,dm=\int_B f\,dm=\int_{B_d} dm.$$

Zuerst konstruieren wir eine Hilfsabbildung ψ_1: $B_d\to B$ der Form

$$\psi_1(x)=\frac{x}{|x|}\,r(|x|).$$

Die Funktion r wird durch die Gleichung

$$\int_{|x|<r(t)} g\,dm=\int_{|x|<t} dm$$

eindeutig festgelegt. ψ_1 ist ein C^∞-Diffeomorphismus mit positiver Funktionaldeterminante a. Es gilt

$$\int_A g\,dm = \int_{\psi_1^{-1}(A)} g\circ\psi_1\cdot a\,dm = \int_{\psi_1^{-1}(A)} g_1\,dm$$

mit $g_1=(g\circ\psi_1)\cdot a$. Insbesondere schließt man aus der Beziehung

$$\int_{|x|<r(t)} g\,dm = \int_{|x|<t} g_1\,dm = \int_{|x|<t} dm \qquad 0<t<d$$

daß

$$\int_{S_t} g_1\,ds = \int_{S_t} ds \qquad \text{für alle } t,\, 0<t<d.$$

ds bezeichnet das Oberflächenelement auf der Sphäre $S_t=\{x\in R^n \mid |x|=t\}$. Mit Hilfe des eingangs erwähnten Satzes von Moser, angewandt auf die Sphären S_t mit den Volumenelementen $g_1\,ds$ und ds, schließt man auf die Existenz von C^∞-Diffeomorphismen $\psi_{2,t}\colon S_t\to S_t$, welche (für alle meßbaren Mengen $A\subset S_t$) der Gleichung $\int_A g_1\,ds=\int_{\psi_{2,t}(A)} ds$ genügen. Die durch $\psi_2|_{S_t}=\psi_{2,t}$ gegebene Abbildung ist ein C^∞-Diffeomorphismus von B_d auf sich. Für $x\neq 0$ folgt dies aus der Bemerkung zum Satz von Moser, und in einer Umgebung von $x=0$ ist ψ_2 die Identität.

Für alle Kugeln $B_{x,r}$ mit Mittelpunkt x und Radius r, die in B_d liegen, gilt nun:

$$\begin{aligned}\int_{B_{x,r}} dm &= \int dt \int_{B_{x,r}\cap S_t} ds\\ &= \int_{\psi_2^{-1}(B_{x,r})} g_1\,dm = \int_{\psi_1\circ\psi_2^{-1}(B_{x,r})} g\,dm.\end{aligned}$$

Der Diffeomorphismus $\psi=\psi_1\circ(\psi_2)^{-1}$ erfüllt daher für beliebige $A\subset B_d$ die Gleichung

$$\int_{(A)} g\,dm = \int_A dm.$$

Es muß noch gezeigt werden, daß ein Diffeomorphismus $\chi\colon B_d\to B_d$ existiert, welcher das Volumenelement $(h\circ\psi)\,dm$ in das Euklidsche Volumenelement dm überführt. Es gilt dann

$$\int_{\psi(A)} f\,dm = \int_{\psi(A)} g\cdot h\,dm = \int_A (h\circ\psi)\,dm = \int_{\chi(A)} dm,$$

woraus man schließen kann, daß $\psi\circ\chi^{-1}=\phi_\sigma$ die Gleichung

$$\int_{\phi_\sigma(A)} \sigma = \int_A dm$$

erfüllt. Die Existenz von χ ist nun aber eine Konsequenz des Korollars zum Satz von Moser, denn die Voraussetzungen $g\circ\psi=1$ außerhalb einer kompakten Teilmenge von B_d und

$$\int_{B_d} (g \circ \psi)\, dm = \int_{B_d} dm$$

sind bei der ausgeführten Konstruktion erfüllt.

Abschließend sei bemerkt, daß die Abbildung ψ sich direkt aus der Funktion f (anstelle von g) definieren ließe. Es ist jedoch fraglich, ob die derart konstruierte Abbildung im Nullpunkt C^∞-differenzierbar wäre.

2. Produktmannigfaltigkeiten

M sei eine kompakte, zusammenhängende, m-dimensionale Mannigfaltigkeit mit einer C^∞-Struktur $\{(\alpha_i, O_i)\}$. Die offenen Mengen O_i bilden hier eine endliche Überdeckung von M und die α_i sind Homöomorphismen von O_i auf $U_i \subset R^m$. B ist die offene Einheitskugel in R^n $(n \geqq 2)$. Wir betrachten die Produktmannigfaltigkeit $M \times B$. Etwas allgemeiner formuliert können wir B durch eine Mannigfaltigkeit N ersetzen, deren (C^∞-)Struktur durch eine einzige Karte, nämlich durch einen Homöomorphismus β von N auf $B \subset R^n$, bestimmt ist. Ein Atlas auf $M \times N$ ist durch die offene Überdeckung $\{O_i \times N\}$ und durch die Homöomorphismen $\gamma_i = (\alpha_i, \beta)$ gegeben: $\gamma_i(y, z) = (\alpha_i(y), \beta(z)) \in U_i \times B$ für Punkte $(y, z) \in M \times N$. Dieser Atlas definiert eine C^∞-Struktur auf $M \times N$. In $M \times N$ sind die Untermannigfaltigkeiten $M \times \{z\}$ definiert. Sie bilden eine Parametrisierung von $M \times N$. Als Parameter wird $t = \beta(z) \in B$ verwendet. Mit den derart eingeführten Bezeichnungen gilt:

Satz 2. *Sind zwei positive C^∞-Volumenelemente τ, σ auf $M \times N$ vorgegeben, die $\int_{M \times N} \sigma = \int_{M \times N} \tau$ erfüllen, so existiert ein C^∞-Diffeomorphismus ϕ von $M \times N$ auf sich, der τ in σ überführt: $\phi * \tau = \sigma$. Ist $N = (-1, 1)$, so existiert ein derartiger Diffeomorphismus unter der zusätzlichen Voraussetzung $\int_{M \times (-1,1)} \sigma \left(= \int_{M \times (-1,1)} \tau \right) < \infty$.*

Das Volumenelement σ auf $M \times N$ induziert auf Grund der Beziehung $\sigma_B(A) = \int_{M \times \beta^{-1}(A)} \sigma$ ein C^∞-Volumenelement σ_B auf B. σ_B und τ_B erfüllen die Voraussetzungen zum Satz 1 und es existiert daher ein C^∞-Diffeomorphismus χ: $B \to B$, der τ_B in σ_B überführt: $\chi * \tau_B = \sigma_B$. Mit Hilfe von χ kann man einen C^∞-Diffeomorphismus von $M \times N$ auf sich definieren, indem man in lokalen Koordinaten

$$\psi(y, z) = \gamma_i^{-1}(\alpha_i(y), \chi \circ \beta(z)) \quad \text{für } (y, z) \in O_i \times N$$

setzt. ψ ist also die Identität in der ersten Komponente (M) und ist in der zweiten Komponente (N) durch die Transformation des Parameterbereichs bestimmt. Der Diffeomorphismus ψ transformiert das Volumenelement τ in ein Volumenelement $\rho=\psi*\tau$, für welches die Beziehung

$$\int_{M\times\beta^{-1}(A)}\sigma=\int_{M\times\beta^{-1}(A)}\rho$$

gilt (A ist eine beliebige Borelmenge in B).

Sind $u\in U$, $v\in B$ die Koordinaten für eine Karte mit dem Parameterbereich $U\times B$, so besitzen ρ und σ Darstellungen der Form

$$\rho=g\,du_1\dots du_m\,dv_1\dots dv_n$$
$$\sigma=h\,du_1\dots du_m\,dv_1\dots dv_n$$

mit positiven C^∞-Funktionen g und h. Soll ρ in einer anderen Karte durch

$$\rho=g'\,du'_1\dots du'_m\,dv'_1\dots dv'_n$$

dargestellt sein, so transformiert sich g nach der Beziehung

$$g'(u'_1,\dots,u'_m,v'_1,\dots,v'_n)\cdot\det\left(\frac{\partial u'_i}{\partial u_k}\right)=g(u_1,\dots,u_m,v_1,\dots,v_n),$$

denn die Koordinatentransformation für die v'_j hat die spezielle Form $v'_j=v_j$, $j=1,\dots,n$. Man schließt daraus, daß $g\,du_1\dots du_m$ für festes $v\in B$ ein m-dimensionales C^∞-Volumenelement ρ_v auf der Untermannigfaltigkeit $M\times\{\beta^{-1}(v)\}$ darstellt.

Aus der Definition für ρ_v bzw. σ_v ist ersichtlich, daß

$$\int_{C\times\beta^{-1}(A)}\rho=\int_A dv_1\dots dv_n\int_{C\times\{\beta^{-1}(v)\}}\rho_v$$

für beliebige Borelmengen $A\subset B$, $C\subset M$. Mit Hilfe der Gleichung

$$\int_{M\times\beta^{-1}(A)}\rho=\int_{M\times\beta^{-1}(A)}\sigma$$

läßt sich daraus die Integralbedingung

$$\int_{M\times\{\beta^{-1}(v)\}}\rho_v=\int_{M\times\{\beta^{-1}(v)\}}\sigma_v$$

herleiten.

Der Satz von Moser wird nun auf die Mannigfaltigkeit M mit den C^∞-Volumenelementen ρ_v und σ_v angewandt. Die Diffeomorphismen $\phi_v\colon M\to M$, welche ρ_v in $\sigma_v=\phi_v*\rho_v$ überführen, werden als Abbildungen von $M\times\{\beta^{-1}(v)\}$ auf sich interpretiert. Die Abbildung $\phi\colon M\times N\to M\times N$, welche auf $M\times\{\beta^{-1}(v)\}$ mit ϕ_v übereinstimmt, ist ein C^∞-Diffeomorphismus. Für Borelmengen $A\subset B$, $C\subset M$ erhält man:

$$\begin{aligned}\int_{C\times\beta^{-1}(A)}\rho&=\int_A dv_1\dots dv_n\int_{C\times\{\beta^{-1}(v)\}}\rho_v\\&=\int_A dv_1\dots dv_n\int_{\phi_v(C)\times\{\beta^{-1}(v)\}}\sigma_v=\int_{\phi(C\times\beta^{-1}(A))}\sigma,\end{aligned}$$

es gilt also $\phi * \rho = \sigma$. Der Diffeomorphismus $\phi \circ \psi$ erfüllt nun alle Forderungen von Satz 2.

3. Eine Verallgemeinerung

Unter einer Einbettung einer m-dimensionalen C^∞-Mannigfaltigkeit M' in die (m-dimensionale) C^∞-Mannigfaltigkeit M versteht man einen C^∞-Diffeomorphismus $\psi\colon M' \to M$. Wir betrachten zusammenhängende Mannigfaltigkeiten M, die „zerlegt" werden können: Es existieren endlich viele kompakte $(m-1)$-dimensionale C^∞-Mannigfaltigkeiten N_i $i=1, \dots, k$ und Einbettungen $\psi_i\colon N_i \times (-1, 1) \to M$ derart, daß für alle $a_i \in (-1, 1)$ $i=1, \dots, k$

$$M \setminus \bigcup_{i=1}^{k} \psi_i(N_i \times (a_i, 1))$$

kompakt und zusammenhängend ist und der relative Rand von

$$\bigcup_{i=1}^{k} \psi_i(N_i \times (a_i, 1))$$

in M aus

$$\bigcup_{i=1}^{k} \psi_i(N_i \times \{a_i\})$$

besteht.

Die im zweiten Abschnitt behandelten Produktmannigfaltigkeiten können alle zerlegt werden. Als weiteres Beispiel sei erwähnt: Ist M eine kompakte, zusammenhängende berandete C^∞-Mannigfaltigkeit mit den (endlich vielen) Randkomponenten ∂M_i $i=1, \dots, k$, dann ist

$$M \setminus \bigcup_{i=1}^{k} \partial M_i$$

eine C^∞-Mannigfaltigkeit, die zerlegt werden kann. Man vergleiche in diesem Zusammenhang auch mit der Arbeit von Banyaga [1], in welcher der Satz von Moser auf kompakte berandete zusammenhängende C^∞-Mannigfaltigkeiten ausgedehnt wird.

Satz 3. *Sind σ und τ zwei positive C^∞-Volumenelemente auf einer zusammenhängenden C^∞-Mannigfaltigkeit M, die zerlegt werden kann, so existiert ein C^∞-Diffeomorphismus ϕ mit $\phi * \tau = \sigma$, vorausgesetzt daß nur $\int_M \sigma = \int_M \tau < \infty$.*

Der erste Teil unseres Beweises stützt sich auf Satz 2, für den zweiten verwenden wir das Korollar zum Satz von Moser.

Zuerst wird ein positives C^∞-Volumenelement σ_1 mit den Eigenschaften

$$\sigma_1 = \begin{cases} \sigma & \text{in } \psi_1(N_1 \times (c_1, 1)) \\ \tau & \text{in } M \setminus \psi_1(N_1 \times (0, 1)) \end{cases}$$

$$\int_M \sigma_1 = \int_M \tau$$

konstruiert. Die Konstante $c_1 > 0$ wird so gewählt, daß

$$0 < \int_{\psi_1(N_1 \times (c_1, 1))} \sigma < \int_{\psi_1(N_1 \times (0, 1))} \tau.$$

Da $\int_M \sigma < \infty$, existieren immer derartige Konstanten c_1. Der Satz 2 kann nun auf die Mannigfaltigkeit $M_1 = \psi(N_1 \times (-1, 1))$ mit den durch σ_1 und τ induzierten Volumenelementen angewendet werden. Die Integralbedingung ist erfüllt und aus dem Beweis des Satzes 2 ist ersichtlich, daß die resultierende Abbildung ϕ_1 für eine klein genuge Konstante $b_1 > -1$ auf $\psi_1(N_1 \times (-1, b_1))$ die Identität ist. Da der relative Rand von $\psi_1(N_1 \times (b_1, 1))$ aus $\psi_1(N_1 \times \{b_1\})$ besteht, kann die Abbildung ϕ_1 durch die Identität zu einem (wiederum mit ϕ_1 bezeichneten) C^∞-Diffeomorphismus auf ganz M erweitert werden.

Diese Konstruktion läßt sich wiederholen: Zu vorgegebenen σ_{i-1} werden Diffeomorphismen ϕ_i: $M \to M$ konstruiert mit $\phi_i * (\sigma_{i-1}) = \sigma_i$, $i = 2, \dots, k$. Es ist nun wesentlich, daß das Volumenelement σ_k außerhalb einer kompakten zusammenhängenden Menge in M mit σ übereinstimmt. Aus dem Korollar schließt man daher auf die Existenz eines C^∞-Diffeomorphismus ϕ_0: $M \to M$ mit $\sigma = \phi_0 * (\sigma_k)$. Mit

$$\phi = \phi_0 \circ \phi_k \circ \cdots \circ \phi_1$$

ist der Satz 3 erfüllt.

4. Abbildungen zwischen Gebieten in R^n, insbesondere in R^3

Ist M eine zusammenhängende n-dimensionale C^∞-Mannigfaltigkeit, die zerlegt werden kann (s. Abschnitt 3), so bezeichnen wir das beschränkte Bild von M unter einer C^∞-Einbettung in R^n als Normalgebiet N. Wir setzen immer $n \geqq 2$ voraus. Beispiele von Normalgebieten sind:

die Kugel $\quad B = \{x \in R^n \,|\, |x| < 1\}$,

der Torus $\quad T = \left\{ x = (z_1, \dots, z_{n-2}, r, \varphi) \,\middle|\, (r-2)^2 + \sum_{i=1}^{n-2} z_i^2 < 1,\ 0 \leqq \varphi < 2\pi \right\}$, beschränkte Gebiete, die durch endlich viele kompakte C^∞-eingebettete $(n-1)$-dimensionale C^∞-Mannigfaltigkeiten berandet sind.

Ein Homöomorphismus ψ zwischen zwei Gebieten G und G' in R^n ist maßtreu, falls das n-dimensionale Maß der meßbaren Mengen A in G erhalten bleibt: $\int_A dm = \int_{\psi(A)} dm$ für alle $A \subset G$. Ein Diffeomorphismus ψ ist genau dann maßtreu, wenn die Funktionaldeterminante im ganzen Gebiet die Konstante ± 1 ist.

Satz 4. *Sind G und G' zwei zu einem Normalgebiet N (C^1-)diffeomorphe Gebiete in R^n mit gleichem, endlichem, n-dimensionalen Maß, so existiert ein maßtreuer C^∞-Diffeomorphismus ϕ von G auf G'.*

Korollar. *Ist $n=2$ oder $n=3$ und sind G und G' zwei zu einem Normalgebiet N homöomorphe Gebiete mit gleichem, endlichem Maß, so existiert ein maßtreuer C^∞-Diffeomorphismus von G auf G'.*

Das Korollar ergibt sich aus dem Satz durch Approximation der Homöomorphismen $\chi: N \to G$ und $\chi': N \to G'$ durch Diffeomorphismen von N auf G beziehungsweise auf G'. Daß derartige Approximationen existieren, ist wohl bekannt für $n=2$. Für $n=3$ wurde dieses Resultat von Munkres bewiesen (s. Munkres [3], Satz 6.3, p. 544; man vergleiche die Anmerkungen zur Entstehung dieses Satzes, loc. cit. p. 544/545).

Für beliebige n lassen sich die C^1-Diffeomorphismen $\chi: N \to G$ durch C^∞-Diffeomorphismen ψ von N auf G approximieren, s. Munkres [4]. Es kann also vorausgesetzt werden, daß N durch C^∞-Diffeomorphismen ψ und ψ' auf G beziehungsweise G' abgebildet werden kann. Mit J_ψ bezeichnen wir den Absolutbetrag der Funktionaldeterminante von ψ. Auf N betrachten wir sodann die C^∞-Volumenelemente $J_\psi\, dm$ und $c\, dm$, wobei die Konstante c durch die Bedingung $\int_N c\, dm = \int_N J_\psi\, dm$ eindeutig festgelegt ist. Nach den Sätzen 2 und 3 existiert nun ein C^∞-Diffeomorphismus $\vartheta: N \to N$ mit $\vartheta * (J_\psi\, dm) = c\, dm$. Wird ϑ' analog definiert, so erfüllt der Diffeomorphismus $\phi = \psi' \circ (\vartheta')^{-1} \circ \vartheta \circ \psi^{-1}$ die Forderungen von Satz 4. Für meßbare Mengen $A \subset N$ ist nämlich

$$\int_A c\, dm = \int_{\vartheta^{-1}(A)} J_\psi\, dm = \int_{\psi \circ \vartheta^{-1}(A)} dm$$

und für ψ' und ϑ' gelten die entsprechenden Beziehungen. ϕ ist also maßtreu.

Erhält man das Normalgebiet N durch Einbettung einer Produktmannigfaltigkeit (im Sinne von Abschnitt 2), so kann man in Satz 4 auf die Forderung, daß die n-dimensionalen Maße von G und G' endlich sein sollen, verzichten.

Literatur

1. Banyaga, A.: Formes-volume sur les variétés à bord. Université de Genève 1972.
2. Moser, J.: On the volume elements on a manifold. Trans. Amer. Math. Soc. **120**, 286–294 (1965).
3. Munkres, J.: Obstructions to the smoothing of piecewise differentiable homeomorphisms. Ann. of Math. **72**, 521–554 (1960).
4. Munkres, J.: Elementary differential topology. Princeton: University Press 1961, revised edition 1966.

Dr. H. M. Reimann
Mathematisches Institut der Universität
CH-3000 Bern, Siedlerstr. 5
Schweiz

(Eingegangen am 11. April 1972)

Math. Z. 129, 173–183 (1972)
© by Springer-Verlag 1972

A Coclassifying Map for the Inclusion of the Wedge in the Product

John W. Rutter

1.

The Whitehead product map $S(X \mathbin{⋇} Y) \to SX \vee SY$ is known to coclassify the inclusion $SX \vee SY \to SX \times SY$. I show in this article that, when B has a comultiplication with homotopy identity and homotopy inverse, the inclusion $SA \vee B \to SA \times B$ is coclassified by a map ω: $A \mathbin{⋇} B \to SA \vee B$ if, either SA and B are simply connected CW complexes, or the comultiplication on B is homotopy associative and A and B have nondegenerate base points. When B is an h-coloop the class of ω is uniquely determined and is the cooperator product class defined and investigated in [5].

Let $\tilde{B} = B \vee_1 I = B \cup I/\{*, 1\}$ and $\alpha: (\tilde{B}, 0) \to (B, *)$ the projection. Since $*$ is a nondegenerate base point there is a homotopy inverse $\beta: (B, *) \to (\tilde{B}, 0)$ to α. The comultiplication m on B induces a comultiplication $\bar{m}$ on $\tilde{B}$ by

$$\bar{m}: B \vee_1 I \xrightarrow{m \vee 1} (B \vee B) \vee_1 I \xrightarrow{1 \vee \beta \vee 1} \big(B \vee (B \vee_1 I)\big) \vee_1 I \to (B \vee_1 I) \vee (B \vee_1 I)$$

where $[0, \frac{1}{2}]$ in the I produced by β is doubled in length and folded back onto the original stalk I and $[\frac{1}{2}, 1]$ is doubled in length to make the stalk I in the second $\tilde{B}$ (see Diagram 1.1).

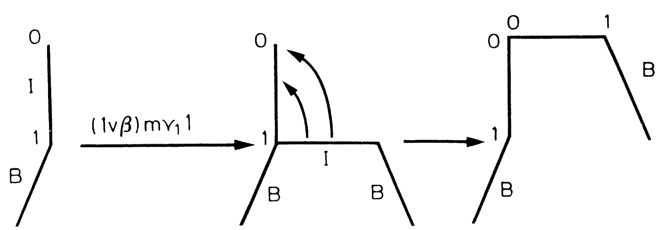


Diagram 1.1 (the comultiplication on $\tilde{B}$)

This comultiplication on $\tilde{B}$ clearly induces a comultiplication on $A \times B \cup CA = A \times \tilde{B}/A \times 0 \cup * \times I$.

Define

$$\sigma: (A \times B) \cup C(A \vee B) \to SA \vee B$$

by

$$\sigma | A \times B \cup CA = (1_{SA} \vee p_2)\,\phi - p_2$$

where ϕ: $A \times B \cup CA \to SA \vee (A \times B \cup CA)$ is the coaction and

$$p_2: \ (A \times B) \cup CA \to B$$

the projection, and

$$\sigma | CB = \theta: \ * \sim 1 - \alpha\beta$$

is a map obtained from some nulhomotopy. The map ω is the composite

$$\omega: \ A ⋇ B \to (A \times B) \cup C(A \vee B) \xrightarrow{\sigma} SA \vee B$$

where the first map is a homotopy inverse to the projection (fuller details for all the above constructions are given in [5]).

The main result in this article is the Theorem:

Theorem 1. *Let A and B have nondegenerate base points and let B have a comultiplication with homotopy identity and homotopy inverse, then there are maps h and k with k homotopic to the identity making commutative the following diagram of weak cofibration sequences*[1]*:*

$$\begin{array}{ccccccc} A ⋇ B & \xrightarrow{\omega} & SA \vee B & \longrightarrow & C_\omega & \longrightarrow & S(A ⋇ B) \\ & & \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle k} \\ & & SA \vee B & \longrightarrow & SA \times B & \longrightarrow & (SA) ⋇ B. \end{array}$$

Furthermore h is a homotopy equivalence[2,3] *if either, B is an h-cogroup, or SA and B are simply connected spaces having the based homotopy type of CW complexes.*

Remark 1.2. A necessary condition for $X \vee Y \to X \times Y$ to be coclassified in the manner specified is that $X ⋇ Y$ is a suspension. The sufficiency conditions given in the Theorem imply that $X ⋇ Y$ is the suspension of space having comultiplication with homotopy identity.

Remark 1.3. The class of h is functorial in A, and it is "functorial" in B for maps which preserve, in the topological category, the product and inverse and the maps β and θ.

---

[1] The lower sequence is the cofibre sequence induced by the inclusion $SA \vee B \to SA \times B$ modified by the projection $(SA \times B) \cup C(SA \vee B) \to SA ⋇ B$ which is a homotopy equivalence by Satz 16 and Satz 17 of [2].

[2] If $B = SK$ is a suspension this Theorem is equivalent to Corollary 2.3 of [6].

[3] If $SA \vee B \to SA \times B$ is a cofibration, then h is a cofibre equivalence by Theorem 5.3 of [4]. This is true for example if $(SA \times B, SA \vee B)$ is a CW pair.

Remark 1.4. The cofibration sequence

$$B \xrightarrow{i_2} A \times B \cup CA \xrightarrow{q} A ⋇ B$$

is a sequence of homomorphisms when B has a comultiplication and in the induced exact sequence of homomorphisms

$$[B, SA \vee B] \xleftarrow{i_2^*} [A \times B \cup CA, SA \vee B]$$
$$\xleftarrow{q^*} [A ⋇ B, SA \vee B] \leftarrow [SB, SA \vee B]$$

the class $\{(1_{SA} \vee p_2)\phi - p_2\}$ is in the kernel of i_2^* and hence is the image of an element $\{\omega\}$ of $[A ⋇ B, SA \vee B]$. The kernel of q^* is zero since

$$A ⋇ B \rightarrow SB$$

is nulhomotopic (cf. the proof of Proposition 3.4 of [5]). If B is an h-coloop then q^* is a homomorphism, with zero kernel, of loops and hence a monomorphism. In this case then the class of ω is uniquely determined and is the universal example for the cooperator product as defined in § 4 of [5].

Remark 1.5. For X or Y compact and regular the "unreduced" Whitehead product map was shown to coclassify the inclusion $S_0 X \vee S_0 Y \rightarrow S_0 X \times S_0 Y$ of unreduced suspensions by Cohen in Theorems 2.4 and 2.5 of [1] (cf. Proposition 2.1 and Corollary 2.3 of [6]). I note here that the maps $\tilde{h}$ and $\tilde{k}$ of Proposition 2.1 of [6] are homotopy equivalences without any restriction on the spaces, and it therefore follows, for *any* Hausdorff spaces X and Y, that the "unreduced" Whitehead product map $X * Y \rightarrow S_0 X \vee S_0 Y$ coclassifies the inclusion

$$S_0 X \vee S_0 Y \rightarrow S_0 X \times S_0 Y.$$

To show that $\tilde{h}$ is a homotopy equivalence, it is sufficient to prove that the bijection $S_0 X \bar{\times} S_0 Y \rightarrow S_0 X \times S_0 Y$ is a homotopy equivalence: this may be proved using the standard method (cf. the proof of Lemma 5.8 i) of [5]). It now follows that the reduced Whitehead product map coclassifies the inclusion $SX \vee SY \rightarrow SX \times SY$ of reduced suspensions provided X and Y have non degenerate base points.

2. Preliminaries

I consider hausdorff spaces X with base point usually denoted $*$. The unreduced cone, suspension and cylinder are $C_0 X = X \times I / X \times \{0\}$, $S_0 X = C_0 X / X \times \{1\}$ and $X \times I$ with base points $\{X \times \{0\}\}$, $\{X \times \{0\}\}$ and $(*, 0)$ respectively; and the reduced cone, suspension and cylinder are

$$CX = C_0 X / * \times I, \quad SX = S_0 X / * \times I \quad \text{and} \quad X ⋇ I = X \times I / * \times I.$$

The map $\beta: (X, *)\to(\tilde{X}, 0)$ and homotopy $\beta\alpha\sim 1$ are assumed chosen to satisfy the Lemma:

Lemma 2.1 (cf. Hilfsatz 14c) of [2]). *The map β and the homotopy $H: \beta\alpha\sim 1: (\tilde{X}, 0)\to(\tilde{X}, 0)$ can be chosen to satisfy $H_t(I)\subset I$.*

An M'-space is a space, X, with a comultiplication $m: X\to X\vee X$ having homotopy identity, an *h-coloop* is an M'-space whose comultiplication induces a loop[4] structure on $[X, Z]$ for each space Z[5], and an *h-cogroup* is a homotopy associative h-coloop.

It is convenient in the proof of the Theorem to modify the topology on certain quotient spaces in order to ensure that the functions defined are continuous. Since products of identification maps may not themselves be identifications, the maps

$$(A\times I)\times B\to S_0 A\times B \quad \text{and} \quad (A\times I)\times B\to C_0 A\times B$$

may not be identifications. The *sets* $S_0 A\times B$ and $C_0 A\times B$ with the quotient topology from these maps will be denoted $S_0 A\bar{\times} B$ and $C_0 A\bar{\times} B$; then $S_0 A\bar{\times} B\to S_0 A\times B$ and $C_0 A\bar{\times} B\to C_0 A\times B$ are continuous bijections and the first is a homotopy equivalence by Lemma 5.8 of [5].

3. Proof of Theorem

Define

$$h'=1\cup h_1\cup h_2:\ C_\sigma\doteq(SA\vee B)\cup_\sigma\big(C(A\times B\cup C(A\vee B))\cup(A\times B\cup C(A\vee B))\ltimes I\big)\to S_0 A\bar{\times}\tilde{B}/*\times I\times I$$

by[6]

$$h_1((a, b, s), t)=(0, \theta(b, st))$$

and

$$(h_2\,|\,CB\ltimes I)((b, s), t)=((*, t), \theta(b, s))$$

and otherwise the t-th level of h_2 by

$$(h_2\,|\,(A\times B\cup CA)\ltimes I)_t=(h_{21})_t-(h_{22})_t$$

where (see Diagram 3.1)

$$h_{21}((a, b, s), t)=\begin{cases}((a, 2st), 0) & 0\leqq s\leqq\frac{1}{2}\\ ((a, t), 2s-1) & \frac{1}{2}\leqq s<1\\ ((a, t), b) & s=1,\end{cases}$$

$$h_{22}((a, b, s), t)=((*, t), b).$$

---

[4] A loop is a set having a binary structure with two sided identity and unique solutions x and y for $xa=b$ and $ay=b$.

[5] Equivalently $[X, X\vee X]$ is a loop.

[6] In $A\times B\cup CA$, $(a, b)\in A\times B$ is denoted $(a, b, 1)$ and $(a, s)\in CA$ is denoted $(a, *, s)$.

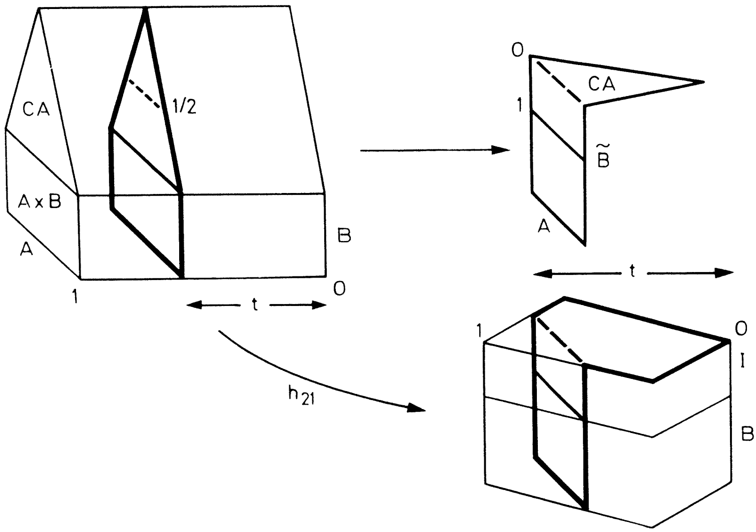


Diagram 3.1 (the map $h_{21}: (A\times B \cup CA)\divideontimes I \to S_0 A \bar{\times} B/* \times I \times I$)

In particular

$$h_2|((A\times B\cup C(A\vee B))\times 1)=\sigma\colon A\times B\cup C(A\vee B)\to SA\times 0\cup 1\times B.$$

The subspace $SA\times 0\cup 1\times B$ of $S_0 A\bar{\times}\tilde{B}/*\times I\times I$ is denoted $SA\vee B$. Define h to be the composite $C_\sigma \xrightarrow{h'} S_0 A\bar{\times}\tilde{B}/*\times I\times I\to SA\times B$. Then the composite $C_\sigma \xrightarrow{h} SA\times B \to SA\divideontimes B$ is constant on

$$(SA\vee B)\cup C(C(A\vee B))$$

and hence determines on taking the quotient a map[7]

$$k\colon S(A\divideontimes B)\to SA\divideontimes B$$

which satisfies $k=*+S(\alpha\beta\divideontimes((\alpha,*)\bar{m}\beta))\sim 1$. Since the projection map $S_0 A\bar{\times}\tilde{B}/*\times I\times I\to SA\times B$ is a homotopy equivalence by the standard arguments (cf. Proposition 2.2 and Lemma 5.8 of [5]), it is sufficient to show that $h'\colon C_\sigma\to S_0 A\bar{\times}\tilde{B}/*\times I\times I$ is a homotopy equivalence.

Given that $SA\vee B$ is simply connected and A and B have the homotopy type of CW complexes then h is a homotopy equivalence by, for example, Theorem 5.1 of [4] and homotopy type considerations.

To prove that h' is generally a homotopy equivalence when B is an h-cogroup I construct a homotopy inverse g to h'. Define

$$\begin{aligned} g=g_1\cup g_2\colon S_0 A\bar{\times}\tilde{B}/*\times I\times I &\doteq C_0 A\bar{\times}\tilde{B}/(*\times I\times I\cup A\times 1\times 0)\\ &\cup C_0 A\bar{\times}\tilde{B}/(*\times I\times I\cup C_0 A\times 0)\\ &\to C_\sigma = SA\vee B\cup_\sigma (C(A\times B\cup C(A\vee B))) \end{aligned}$$

[7] $S(A\divideontimes B)$ has the quotient topology from $A\times B\times I$.

as follows. The first map – the homeomorphism – is, on $S_0 A \bar{\times} B/* \times I \times 1$, induced by the usual decomposition

$$S_0 A = (A \times [0, \tfrac{1}{2}]/A \times 0) \cup (A \times [\tfrac{1}{2}, 1]/A \times 1) \doteq C_0 A \cup C_0 A$$

into the union of the first and second cones; on $C_0 A \times I/* \times I \times I$ it is induced by the homeomorphism which is inverse to the map

$$\gamma\colon I \times I/0 \times [\tfrac{1}{2}, 1] \to I \times I$$

given by

$$\gamma(s, t) = \begin{cases} (s, (2-s)t) & t \leqq \tfrac{1}{2} \\ (s, 1 - s(1-t)) & \tfrac{1}{2} \leqq t \end{cases}$$

(see Diagram 3.2).

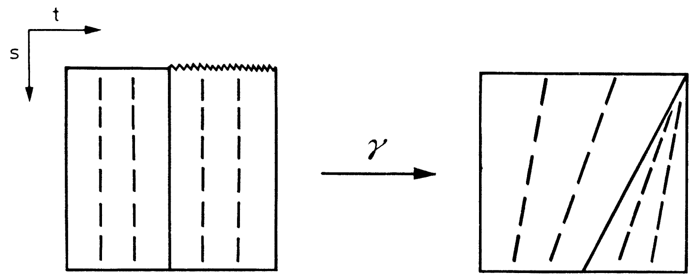


Diagram 3.2 (the homeomorphism $\gamma\colon I \times I/0 \times [\tfrac{1}{2}, 1] \to I \times I$)

The function g_1 is the composite

$$\begin{aligned} g_1\colon C_0 A \bar{\times} \tilde{B}/* \times I \times I \cup A \times 1 \times 0 \\ &\to (CA \mathbin{\divideontimes} I/A \times 1 \times 0) \cup (C_0 A \bar{\times} \tilde{B}/* \times I \times I) \\ &\xrightarrow{g_{11} \cup g_{12}} SA \vee (C_0 A \bar{\times} B/* \times I \times I \cup C_0 A \times 0) \\ &\xrightarrow{(\iota_1, \delta)} C_\sigma = (SA \vee B) \cup C((A \times B) \cup C(A \vee B)) \end{aligned}$$

where the first map is the homeomorphism induced by

$$\tilde{B} = (B \vee_1 [\tfrac{1}{2}, 1]) \vee_{\frac{1}{2}} [0, \tfrac{1}{2}] \doteq \tilde{B} \vee_1 \dot{I},$$

$$g_{11}((a, s), t) = (a, s(1-t)) \in SA,$$

g_{12} is the projection, ι_1 is the inclusion of SA into the space $SA \vee B$ at the base of the mapping cone and δ is obtained from the following map δ' by taking quotients: the map δ' is the composite[8]

$$\begin{aligned} \delta'\colon (A \times B \cup CA) \mathbin{\divideontimes} I &\xrightarrow{(\iota_2 + \pi_2) \divideontimes 1} (B \vee (A \times B \cup CA) \mathbin{\divideontimes} I) \\ &\doteq (B \mathbin{\divideontimes} I) \vee ((A \times B \cup CA) \mathbin{\divideontimes} I) \to B \vee C(A \times B \cup CA) \end{aligned}$$

[8] + is the structure induced on $A \times B \cup CA$ by the comultiplication on B.

where ι_2 is the inclusion into the second factor, π_2 is the composite $A \times B \cup CA \to B \subset B \vee (A \times B \cup CA)$ of the projection and the inclusion to the first factor, and $B ⋇ I \to B$ is the projection (see Diagram 3.3).

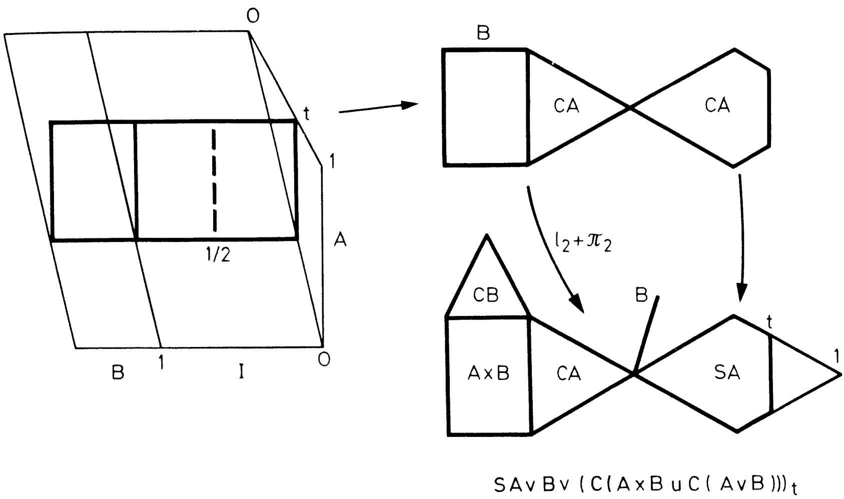


Diagram 3.3 (the t-th section of $g_1: C_0 A \bar{\times} \tilde{B}/* \times I \times I \cup A \times 1 \times 0 \to C_\sigma$)

Now, at $s=1$, g_1 is the function[9]

$$((-1_{SA}), ((1_{SA} \vee p_1)\phi - p_1) + p_1)\phi \sim p_1 : (A \times B) \cup CA \to SA \vee B,$$

and the inverse of this homotopy determines the function[10]

$$g_2: C_0 A \bar{\times} \tilde{B}/* \times I \times I \cup C_0 A \times 0 \doteq (A \times B \cup CA) ⋇ I/\sim \to SA \vee B$$

where the equivalence relation on the second term is induced by the projection $p_1: (A \times B \cup CA) \times 0 \to B$. This completes the definition of g.

Clearly g restricted to

$$SA \vee B = SA \times 0 \cup 1 \times B \subset S_0 A \bar{\times} \tilde{B}/* \times I \times I$$

is the identity map onto $SA \vee B \subset SA \vee B \cup C(A \times B \cup C(A \vee B))$.[11]

The Theorem is now immediate from the following Lemma.

Lemma 3.4. *g is a homotopy inverse to h'.*

[9] Homotopy associativity seems to be necessary at this point.
[10] The function g_2 and therefore g will depend on the choice of the homotopy above: however by Lemma 3.4 the homotopy class of g is unique, though the cofibre homotopy class may not be unique.
[11] Also $g|0 \times B$ is homotopic to the identity map of B, and g could if necessary be modified in the obvious way to give the identity map on $SA \times 0 \cup 0 \times B \cup 1 \times B$.

Proof. i) Consider the composite

$$u\colon S_0 A \bar{\times} \tilde{B}/* \times I \times I \xrightarrow{g} C_\sigma \xrightarrow{h'} S_0 A \bar{\times} \tilde{B}/* \times I \times I \to SA \times B$$

where the last map is the projection, and let u_1 and u_2 be the composites

$$u_1\colon S_0 A \bar{\times} \tilde{B}/* \times I \times I \xrightarrow{u} SA \times B \to SA$$

$$u_2\colon S_0 A \bar{\times} \tilde{B}/* \times I \times I \xrightarrow{u} SA \times B \to B$$

of u with the projections onto the factors. Now u_2 restricted to

$$S_0(*) \bar{\times} \tilde{B}/* \times I \times I = * \times I \times B/* \times I \times *$$

is homotopic to the composite

$$* \times I \times B/* \times I \times * \to * \times 1 \times B \xrightarrow{u_2} B$$

where the first map is the projection onto the subspace and the second map is already the identity. Since the construction of u is functorial in A it now follows, by considering the map $A \to *$, that u_2 is homotopic to the projection $r_2\colon S_0 A \bar{\times} \tilde{B}/* \times I \times I \to B$. Similarly u_1 restricted to

$$S_0 A \times (\tilde{*})/* \times I \times I = SA ⋇ I$$

is homotopic to the composite

$$SA ⋇ I \to SA \times 0 \xrightarrow{u_1} SA$$

where the first map is the projection and the second is again the identity. Now the map u is not functorial in B: however it is "functorial" with respect to maps $B \to B'$ which preserve, in the topological category, the map $\beta\colon B \to \tilde{B}$, and the comultiplication, coinverse and the homotopies which give the h-cogroup structure to B; such a map is $B \to *$. It follows now that u_1 is homotopic to the projection $r_1\colon S_0 A \bar{\times} \tilde{B}/* \times I \times I \to SA$. The projection $S_0 A \bar{\times} \tilde{B}/* \times I \times I \to SA \times B$ is a homotopy equivalence and since u is now shown to be homotopic to it, it follows that $h'g$ is homotopic to the identity.

ii) Consider next the map $gh'\colon C_\sigma \to C_\sigma$. Denote by

$$M_\sigma = (SA \vee B) \cup_\sigma (A \times B \cup C(A \vee B)) ⋇ I$$

the mapping cylinder of σ where the attachment is made at the 1-end of the cylinder, and, for convenience, let $P = A \times B \cup C(A \vee B)$. The map gh' may be written as the composite

$$C_\sigma \doteq M_\sigma \cup (P ⋇ I) \cup CP \xrightarrow{u \cup v \cup w} C_\sigma$$

where $u(M_\sigma) \subset SA \vee B$, $w(CP) \subset SA \vee B$ and $v(P \times t) \subset SA \vee B \vee (P \times t)$ for $0 < t < 1$. The map $v \cup w$ is now deformed such that the section $P \times t$ of $P ⋇ I$ is kept in $SA \vee B \vee (P \times t)$ throughout the deformation. First (see Diagram 3.5) the map v: $P ⋇ I \to (SA \vee B) \cup CP$ can be pulled off SA

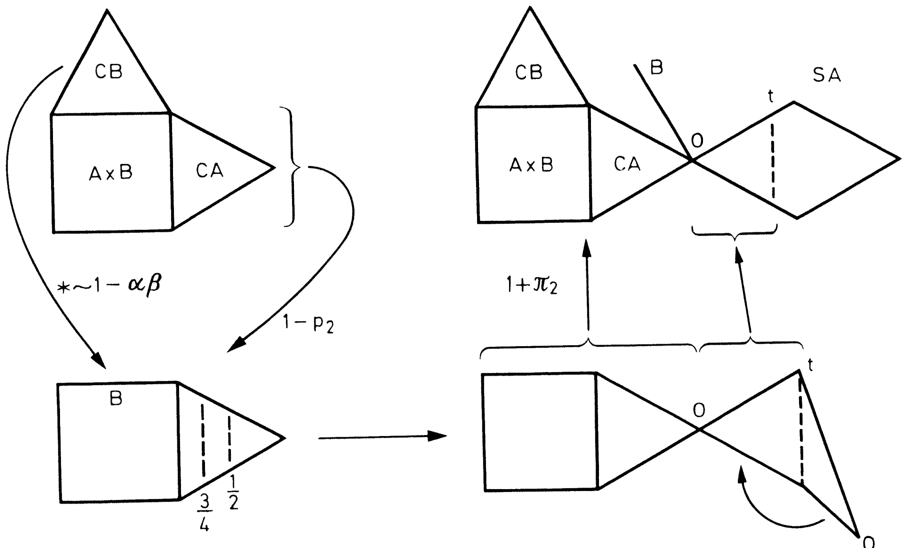


Diagram 3.5 (the map v: $P \times t \to SA \vee B \vee (P \times t)$)

(for $t < 1$) since the final map into SA clearly factors through CA; the resulting map is constant on the top three quarters of the cone CA, and using a further homotopy the second-vertical-map is deformed to the identity on CA. The combined homotopy V: $(P ⋇ I) ⋇ I \to (SA \vee B) \cup CP$, say, satisfies $V_\tau | P \times 0 = v | P \times 0$ for all τ, and therefore $u \cup V$ gives a homotopy $u \cup v = u \cup V_0 \sim u \cup V_1$. Now the t-th section of V_1 is independent of t and is given by

$$\begin{aligned}(V_1)_t | (A \times B \cup CA) &= (1 + \pi_2)(1 - p_2) \\ &= (1 + \pi_2) - (p_2 + \pi_2) \\ &\sim 1 - p_2 \\ &\sim 1\end{aligned}$$

since p_2 can be pushed out along the cone CB in the image; and using the restriction of this homotopy to B, a compatible homotopy

$$(V_1)_t | CB \sim \chi + 1_{CB}$$

is obtained, where $\chi\colon SB\to B$, since any part of the image lying in CB can be pushed off to the vertex, and where $+$ is given by the coaction $CB \to SB\vee CB$. In order to show that χ can be taken as the constant map by suitably modifying the homotopies above, it is sufficient to prove that the composite function[12]

$$\pi_1^{A\times B\cup CA}(A\times B\cup CA, 1)\xrightarrow{i_2^*}\pi_1^B(A\times B\cup CA, i_2)\xrightarrow{p_{2*}}\pi_1^B(B, 1)$$

is an epimorphism. However, since B and $A\times B\cup CA$ are h-cogroups, and $i_2\colon B\to A\times B\cup CA$ is a homomorphism, this is equivalent by Lemma 1.2.2 iv) and Coprimitivity Theorem 3.3.3 of [3] to showing that

$$[A\times B\cup CA, \Omega(A\times B\cup CA)]\xrightarrow{i_2^*}[B, \Omega(A\times B\cup CA)]\xrightarrow{(\Omega p_2)_*}[B, \Omega B]$$

is an epimorphism, which is easily seen to be the case; in fact $(\Omega i_1)_* p_2^*$ is a splitting. Thus $v\colon P\ltimes\!\rtimes I\to C_\sigma$ is homotopic by a level-preserving homotopy W to the projection

$$W_1\colon P\ltimes\!\rtimes I\to CP\to(SA\vee B)\cup CP.$$

The homotopy can be extended to a compatible homotopy of $w\colon CP\to SA\vee B$ and a compatible homotopy of $u\colon M_\sigma\to SA\vee B$ by using

$$W|P\times 0\times I$$

and respectively

$$W|P\times 1\times I$$

in the usual way. The end-function of this homotopy gives, on taking quotients, a function

$$(\bar{u}_1, 1, \bar{w}_1)\colon C_\sigma\to\overline{M}_\sigma\cup C_\sigma\vee SP\to C_\sigma$$

where $\overline{M}_\sigma$ is the double mapping cylinder obtained from $P\ltimes\!\rtimes I$ by attaching a copy of $SA\vee B$ at each end. Clearly $gh'\sim(\bar{u}_1, 1, \bar{w}_1)$ is a homotopy equivalence with homotopy inverse $(-\bar{u}_1, 1, -\bar{w}_1)$ and hence, since $h'g\sim 1$, it follows that $gh'\sim 1$. This completes the proof of the Lemma.

References

1. Cohen, D.E.: Products in carrier theory. Proc. London Math. Soc. **7**, 219–248 (1957).
2. Puppe, D.: Homotopiemengen und ihre induzierten Abbildungen I. Math. Z. **69**, 299–344 (1958).
3. Rutter, J.W.: A homotopy classification of maps into an induced fibre space. Topology **6**, 379–403 (1967).

---

[12] $\pi_1^X(Y, f)$ denotes the track group based at $f\colon X\to Y$.

4. Rutter, J.W.: Maps and equivalences into equalizing fibrations and from coequalizing cofibrations. Math. Z. **122**, 125–141 (1971).
5. Rutter, J.W.: A derived homotopy product. Preprint, University of Liverpool.
6. Rutter, J.W., Spencer, C.B.: On the James, Samelson and Whitehead products. Tohôku Math. J. **22**, 545–556 (1970).

Dr. J.W. Rutter
Department of Pure Mathematics
The University
P.O. Box 147
Liverpool, L69 3BX
England

(Received February 18, 1972)

Math. Z. 129, 185–193 (1972)
© by Springer-Verlag 1972

Kriterien für n-te Potenzreste

Horst von Lienen

§ 1. Überblick und Bezeichnungen

Unter n-ten Potenzresten $\bmod m$ $(n, m \in \mathbf{N})$ werden ganze Zahlen q verstanden, für die die Kongruenz

$$x^n \equiv q \quad \bmod m \tag{1}$$

mit ganzzahligen Werten x lösbar ist. Zerlegt man die Zahl m in Primzahlpotenzen, so ist (1) gleichwertig mit den simultanen Kongruenzen

$$p \text{ Primzahl}, \ \alpha \in \mathbf{N}\colon \ x^n \equiv q \quad \bmod p^\alpha \tag{2}$$

für alle Faktoren p^α in m.

Es ist üblich, noch die Bedingung $(q, m) = 1$ bzw. $p \nmid q$ vorauszusetzen. Wie wir in Satz 1 zeigen, läßt sich der Fall p/q aber sehr wohl allgemein behandeln. Für $n = 2$ findet sich ein Satz hierzu im Kapitel „Einige neueste Sätze über elementare Zahlentheorie" bei Holzer [12], Bd. III. Auch im Fall $p = 2$, der üblicherweise ausgeschlossen wird, kann man vollständig durch n und α ausdrücken, welche Restklassen q n-te Potenzreste mod 2^α sind (§ 2). Einen Ansatz hierzu unternahm Bachmann nach Vorlage von Arndt in [5].

In § 3 leiten wir zusammen mit dem öfter festgestellten Potenzrestkriterium, das nach der Teilbarkeit des Index von q fragt, ein verallgemeinertes Eulersches Kriterium her, das für alle Primzahlen p und alle dazu primen Restklassen q gilt. Für ungerade Primzahlen findet es sich bei Bachmann [5], Nagell [19] und Winogradow [25], für einen Spezialfall von $p = 2$ bei Scholz/Schoeneberg [21]. Mit ihm lassen sich die folgenden Sätze viel kürzer und vor allem für alle Primzahlen zugleich beweisen.

Gilt $p \nmid nq$ in (2), so ist die Zahl q genau dann n-ter Potenzrest mod p^α, wenn sie n-ter Potenzrest mod p ist. Dieser Satz, der für ungerade Primzahlen p in einer Richtung von Nagell aufgestellt wurde [19], wird in beiden Richtungen und für alle Primzahlen p bewiesen. Er bedeutet, daß sich n-te Potenzreste nicht durch Jacobi-Symbole beschreiben lassen. Neu ist der Satz, wie sich eine Potenz q reduzieren läßt, während die Rückführung eines Produkts n mit einem anderen Beweis bei Holzer vorkommt [12, Bd. I]. Sie verallgemeinert eine Arbeit von Storchi [22] in zwei Richtungen. Die umgekehrten Fälle, nämlich Produkte q und Potenzen n aus einfacheren Kriterien aufzubauen, sind dagegen nicht möglich (§ 4).

Ausführlichere Regeln für q als n-ten Potenzrest mit $n \geqq 3$ sind für einige Paare q, n, für $\alpha = 1$ und $p \equiv 1 \bmod n$ bekannt. Sie stützen sich stets auf Zerlegungen der Primzahl p durch quadratische Formen, die vom jeweiligen Wert n abhängig sind. Für $n = 3$ wurden sie von Euler, Jacobi, Nagell und – als Kongruenzen zum Berechnen der Koeffizienten für alle Primzahlen q – von Lehmer angegeben. Entsprechend sind für $n = 4$ Gauß, Cunningham/Gosset und Lehmer zu nennen. Ferner gibt es für $n = 5$ Kriterien von Lehmer und Alderson und für $n = 8$ von Western, Halter-Koch und dem Verfasser. Soweit handelt es sich um Primzahlen q. Auch auf Kriterien für einige zusammengesetzte Zahlen q, die vom Verfasser bewiesen wurden, wird hingewiesen. Eine Formel von Euler, für die Fueter einen Beweis gegeben zu haben glaubte, und eine Behauptung von Tihanyi werden richtiggestellt (§ 5).

Die Bezeichnungen von (2) sollen für die ganze Arbeit gelten.

§ 2. Sonderfälle

Satz 1. *Gilt in* (2) *die Beziehung* p/q, *so ist die Kongruenz für* $q \equiv 0 \bmod p^\alpha$ *immer lösbar. Gilt* $n \geqq \alpha$, *so ist sie es auch nur dann. Ist* p^β *im Fall* $n < \alpha$ *und* $q \not\equiv 0 \bmod p^\alpha$ *die größte in* q *enthaltene Potenz von* p, *so ist die Kongruenz* (2) *nur sinnvoll, wenn* β *durch* n *teilbar ist. Durch Dividieren von* q *und dem Modul* p^α *durch* p^β *wird dieser Fall auf* $p \nmid q$ *zurückgeführt: Die Kongruenz* (2) *ist genau dann lösbar, wenn die so entstandene es ist.*

Beweis. Schreibt man (2) in der Form

$$x^n = q_0 + r\,p^\alpha \qquad \text{mit } 0 \leqq q_0 < p^\alpha \text{ und } r \in \mathbf{Z}, \tag{3}$$

so folgt aus der Beziehung p/q_0 sofort p/x und p^n/x^n, wenn es eine Lösung x gibt. Für $n \geqq \alpha$ hat man jedenfalls p^α/x^n. Also muß auch p^α/q_0 oder $q \equiv 0 \bmod p^\alpha$ gelten. Dann gibt es die Lösung $x = 0$, die auch für $n < \alpha$ den Fall $q \equiv 0 \bmod p^\alpha$ löst.

Für $n < \alpha$ und $q \not\equiv 0 \bmod p^\alpha$ schreibt man die p-adische Darstellung

$$q_0 = a_{\alpha-1}\,p^{\alpha-1} + a_{\alpha-2}\,p^{\alpha-2} + \cdots + a_1\,p \qquad \text{mit } 0 \leqq a_i < p \text{ für } i = 1, \ldots, \alpha-1.$$

Soll es eine Lösung x geben, so muß p^n/q_0 gelten, die Werte

$$a_{n-1}, a_{n-2}, \ldots, a_1$$

müssen also verschwinden. (Für $n = 1$ wird hier nichts Zusätzliches festgestellt.) Nun werde $x = p\,y$ gesetzt. Durch Division mit p^n ergibt sich aus (3)

$$y^n = a_{\alpha-1}\, p^{\alpha-n-1} + a_{\alpha-2}\, p^{\alpha-n-2} + \cdots + a_n + r\, p^{\alpha-n}.$$

Ist $a_n \neq 0$, so liegt der Fall $p \nmid q$ vor.

Gilt $n \geqq \alpha - n$, so kann a_n nicht verschwinden, wenn es eine Lösung geben soll. Nach dem ersten Teil müßte sonst $q_0 = 0$ sein. Wenn im Fall $n < \alpha - n$ jedoch a_n verschwindet, müssen die Zahlen $a_{2n-1}, a_{2n-2}, \ldots, a_{n+1}$ ebenfalls 0 sein, wenn es eine Lösung geben soll. Dann setzt man $y = pz$ und dividiert erneut durch p^n, bis man schließlich insgesamt durch p^{kn} mit $k \in \mathbf{N}$ geteilt und $a_{kn} \neq 0$ gefunden hat. Wenn (3) lösbar ist, ist es auch die so entstandene Gleichung. Hat diese umgekehrt eine Lösung, so ist (3) bzw. (2) mit dem p^k-fachen Wert als x lösbar. Damit ist der Satz bewiesen.

Man kann ihn auch so formulieren: *Entweder muß* $q \equiv 0 \bmod p^\alpha$ *sein, oder die p-adische Darstellung von q muß ein Vielfaches von n Nullen am Ende haben. Dann entscheidet sich die Lösbarkeit von* (2) *am Fall* $p \nmid q$.

Künftig werden wir immer q teilerfremd zu p voraussetzen.

Satz 2. *Ist in* (2) $p = 2$, $\alpha \geqq 3$, q *ungerade und* $n = 2^a v$ *mit ungerader Zahl* v, *so gibt es drei Fälle:*

1. *Für* $a = 0$ *ist jede ungerade Zahl q n-ter Potenzrest.*

2. *Für* $1 \leqq a < \alpha - 2$ *sind genau die Restklassen* $q \equiv 1 + s \cdot 2^{a+2} \bmod 2^\alpha$ *mit* $s = 0, 1, 2, 3, \ldots, 2^{\alpha-a-2} - 1$ *n-te Potenzreste* $\bmod 2^\alpha$.

3. *Für* $a \geqq \alpha - 2$ *ist nur die Restklasse* $q \equiv 1 \bmod 2^\alpha$ *n-ter Potenzrest* $\bmod 2^\alpha$.

Beweis. Die Zahl x kann in (2) alle primen Restklassen mod 2^α durchlaufen, und wir betrachten, welche Restklassen nach dem Potenzieren mit n übrigbleiben. Die primen Restklassen mod p^α sind gegeben durch die beiden zyklischen Gruppen $A = \{5,\ 5^2,\ 5^3, \ldots, 5^{2^{\alpha-2}} \equiv 1\}$ und $B = \{-5,\ -5^2,\ -5^3, \ldots, -5^{2^{\alpha-2}} \equiv -1\}$. Das Potenzieren jedes Elements mit der ungeraden Zahl v bringt in A und B nur die Zählung von einer anderen Stelle an, da v teilerfremd zur Ordnung ist. Im ersten Fall des Satzes wird also jede prime Restklasse q mod 2^α als n-ter Potenzrest erreicht.

Enthält n den Faktor 2, so bedienen wir uns der Beziehung

$$x^{2^a} \equiv 1 \mod 2^{a+2} \qquad \text{für } a \geqq 1,\ x \text{ ungerade}, \tag{4}$$

die sich leicht mit vollständiger Induktion beweisen läßt. Durch Potenzieren mit v erfahren wir für $a + 2 < \alpha$, daß nur die im zweiten Teil des Satzes genannten Restklassen n-te Potenzreste sein können. Um zu zeigen, daß sie alle vorkommen, geben wir ebenso viele inkongruente Restklassen als n-te Potenzen an: $5^n, 5^{2n}, 5^{3n}, \ldots, 5^{2^{\alpha-a-2}n} \bmod 2^\alpha$. Wäre für $1 \leqq k_1 \neq k_2 \leqq 2^{\alpha-a-2}$ nämlich $5^{k_1 n} \equiv 5^{k_2 n} \bmod 2^\alpha$, so müßte $(k_1 - k_2)\, n \equiv 0$

mod $2^{\alpha-2}$ gelten, d.h. $2^{\alpha-a-2}/(k_1-k_2)\,v$. Da dies nicht möglich ist, ist der zweite Teil bewiesen.

Für $a+2\geqq\alpha$ schreiben wir in (4) nach dem Potenzieren mit v die Restklasse 1 mod 2^{a+2} als 1 mod 2^α. Sie ist n-ter Potenzrest zu $x=1$. Damit ist auch der dritte Teil des Satzes gezeigt.

§ 3. Allgemeine Kriterien

Satz 3. *Sei p ungerade oder* $p^\alpha=2$ *bzw.* 4 *und gelte* $p\nmid q$. *Nennt man den größten gemeinsamen Teiler* $(n,\varphi(p^\alpha))=d$, *so ist die Beziehung*

$$d/\operatorname{ind} q$$

notwendiges und hinreichendes Kriterium, daß q n-ter Potenzrest mod p^α *ist. Die Bedingung ist unabhängig davon, welche primitive Wurzel des Moduls* p^α *dem Index zugrundegelegt wurde.*

Der Beweis durch Übergang von (2) zu den Potenzen einer primitiven Wurzel findet sich bei Bachmann [5], Nagell [19], Winogradow [25] und Holzer [12]. Einen anderen Beweis, der die Kongruenz $d\equiv k\,n$ mod $\varphi(p^\alpha)$ mit $k\in\mathbf{Z}$ benutzt, kann man aus einem Ansatz bei Scholz/Schoeneberg [21] entwickeln. Die Unabhängigkeit von der primitiven Wurzel, die indirekt enthalten ist, wurde nie betrachtet. Es läßt sich jedoch zeigen, daß jeder Teiler von $\varphi(p^\alpha)$ alle Indizes zugleich teilt oder nicht teilt.

Man kann die Sätze 2 und 3 zu einem verallgemeinerten Eulerschen Kriterium für alle Primzahlen p zusammenfassen:

Satz 4. *Ist für* $p\nmid q$ *u der kleinste Exponent, der alle primen Restklassen zu* 1 mod p^α *macht, so ist q genau dann n-ter Potenzrest* mod p^α, *wenn gilt*

$$q^{\frac{u}{(n,u)}}\equiv 1 \mod p^\alpha. \tag{5}$$

Beweis. Mit den Voraussetzungen von Satz 3 ist $u=\varphi(p^\alpha)$. Bei der Darstellung von q in (5) mit einer primitiven Wurzel wird der Nenner (n,u) gerade weggekürzt, wie Bachmann [5], Nagell [19], Winogradow [25] und Scholz/Schoeneberg [21] genauer ausgeführt haben. Wir weiten den Beweis auf $p^\alpha=2$ bzw. 4 aus, denn Voraussetzung ist nur, daß es eine primitive Wurzel gibt. Die übrigen Fälle von $p=2$ brauchen einen gesonderten Beweis:

Betrachtet man p und α nach Satz 2, so ist $u=2^{\alpha-2}$. Die Zahl (n,u) sei mit d bezeichnet. Im Fall 1 ist $d=1$ und (5) wird zur Kongruenz (4) mit $a=\alpha-2$, die genau von den ungeraden Zahlen erfüllt wird. d ist 2^a im Teil 2 des Satzes. Ist q n-ter Potenzrest mod 2^α, so folgt $q^{u/d}\equiv x^{(vd)\,u/d}=x^{vu}\equiv 1$ mod 2^α, wie in (5) behauptet wird. Umgekehrt kann man die ungerade Zahl q nach Satz 2, 1. als $q\equiv\pm 5^{rv}$ mod 2^α schreiben. (5) bedeutet dann, daß r durch d teilbar sein muß, da nur ganze Vielfache von u als

Exponenten die Restklasse 1 erzeugen. Wegen des Faktors $n=dv$ im Exponenten ist q als n-ter Potenzrest dargestellt. Im dritten Fall des Satzes schließlich ist $d=u$ und (5) wird zur Kongruenz $q\equiv 1 \bmod 2^\alpha$, die nach Satz 2 genau die n-ten Potenzreste beschreibt.

§ 4. Reduktion von Kriterien

Satz 5. *Gilt $p\nmid q$ und $p\nmid n$, so ist die Zahl q genau dann n-ter Potenzrest* mod p^α, *wenn sie n-ter Potenzrest* mod p *ist.*

Beweis. Wendet man (5) an und setzt $f=p^{\alpha-1}$ für die Voraussetzungen von Satz 3, $=p^{\alpha-2}$ sonst, so ist $(n,u)=(n,f(p-1))=(n,p-1)$, da f nichts zum größten gemeinsamen Teiler beiträgt. Der von α unabhängige Bruch $\frac{p-1}{(n,u)}$ sei mit w bezeichnet.

Aus der Kongruenz $q^w\equiv 1 \bmod p$ bzw. der Gleichung $q^w=1+rp$ mit $r\in\mathbf{Z}$ folgt durch $(\alpha-1)$- bzw. $(\alpha-2)$-faches Potenzieren mit p (vgl. (4)), daß $q^{wf}=1+sp^\alpha$ mit $s\in\mathbf{Z}$ ist. Damit ist nach Satz 4 die eine Richtung gezeigt.

Gilt umgekehrt die letzte Gleichung, so folgt sofort die Kongruenz $q^{wf}\equiv 1 \bmod p$. Ist t der Exponent, zu dem die Restklasse q mod p gehört, so gilt danach t/wf. Wegen der Beziehungen $t/p-1$ und $(p-1,f)=1$ ist aber $(t,f)=1$ und daher t/w oder $q^w\equiv 1 \bmod p$. Das war zu beweisen.

Beschreibt man die n-ten Potenzreste mit dem verallgemeinerten Legendre-Symbol $(q/p)_n=\zeta_d^{\operatorname{ind} q}$, worin ζ_d eine d-te Einheitswurzel ist, so gilt nach Satz 5 für $p\nmid nq$ und alle Zahlen $\alpha\in\mathbf{N}$

$$\left(\frac{q}{p^\alpha}\right)_n=\left(\frac{q}{p}\right)_n.$$

Satz 6. *Es gelte $p\nmid q$, und u sei der kleinste Exponent, der alle primen Restklassen zu* 1 mod p^α *macht. Ist dann $q=y^h$, $g=(h,n,u)$ und $\frac{(n,u)}{g}=\left(\frac{n}{g},u\right)$, so ist q genau dann n-ter Potenzrest* mod p^α, *wenn y $\frac{n}{g}$-ter Potenzrest* mod p^α *ist.*

Beweis. Es sei $h=rg$, also $\left(r,\frac{(n,u)}{g}\right)=1$. Ist y $\frac{n}{g}$-ter Potenzrest, so ist $q=y^h$ durch Potenzieren in (2) nr-ter Potenzrest, also auch n-ter Potenzrest mod p^α. Setzt man das letztere nach (5) in der Form $y^{\frac{rgu}{(n,u)}}\equiv 1$ mod p^α voraus und gehört y zum Exponenten t mod p^α, so gilt $t\Big/\frac{ru}{(n,u)/g}$.
Da r prim zum Nenner, also auch zu u ist, t aber u teilt, ist der Faktor r

prim zu t, und man kann ihn fortlassen. Dann erhält man die Kongruenz $y^{\frac{u}{(n/g,\,u)}} \equiv 1 \bmod p^{\alpha}$: y ist $\frac{n}{g}$-ter Potenzrest. Damit ist der Satz für alle Primzahlen p bewiesen.

Bei den bekannten Kriterien wird immer n/u vorausgesetzt; dann ist die Voraussetzung $\frac{(n,u)}{g} = \left(\frac{n}{g}, u\right)$ erfüllt, und man kann $g=(h,n)$ setzen.

Satz 7. *Es gelte* $p \nmid q$, $n=rs$ *und* $(r,s)=1$. *Die Restklasse* q *ist genau dann* n*-ter Potenzrest* mod p^{α}, *wenn sie zugleich* r*-ter und* s*-ter Potenzrest* mod p^{α} *ist.*

Beweis. Setzt man $v=(r,u)$ und $w=(s,u)$ mit der bisherigen Bedeutung von u, so gilt $(n,u)=vw$ und $(v,w)=1$. Ist nun q r-ter und s-ter Potenzrest, d.h. nach (5) $q^{u/v} \equiv q^{u/w} \equiv 1 \bmod p^{\alpha}$, so ist $q^{u/v}\, q^{u/w} = q^{\frac{u}{vw}(v+w)} \equiv 1 \bmod p^{\alpha}$. Darin muß schon $\frac{u}{vw}$ ein Vielfaches des Exponenten sein, zu dem q gehört, denn der Faktor $v+w$ ist prim zu vw und zu u. Ohne ihn folgt aus (5), daß q auch n-ter Potenzrest mod p^{α} ist. Der umgekehrte Schluß ist trivial, da $q \equiv x^n \bmod p^{\alpha}$ als x^{rs} geschrieben werden kann.

Mit Satz 7 ist ein Resultat von Storchi [22] in zwei Richtungen verallgemeinert. Storchi behandelt den Fall $q=2$ für die Zahlen $n=6$, 12, 24 und 48. Man rechnet leicht nach, daß das Ergebnis die bekannten Kriterien für $r=3$ und $s=2, 4, 8$ und 16 kombiniert.

Die Sätze 6 und 7 zeigen, daß sich Kriterien für Potenzen q und Produkte n auf einfachere Formen zurückführen lassen. Das Umgekehrte, nämlich Produkte q und Potenzen n zu reduzieren, läßt sich nicht allgemein lösen: Im ersten Fall müßte man z.B. für ungerade Primzahlen aussagen können, wie sich die Teilbarkeit von $\operatorname{ind} q_1$ und $\operatorname{ind} q_2$ durch d zu der von $\operatorname{ind} q_1 + \operatorname{ind} q_2$ durch d verhält. Das geht natürlich, wenn d einen der Summanden teilt, aber nicht, wenn beide nicht durch d teilbar sind. Im zweiten Fall müßte man unter der üblichen Voraussetzung $n/p-1$ von $n_1/\operatorname{ind} q$ auf $n_1^k/\operatorname{ind} q$ schließen, was nicht möglich ist.

§ 5. Spezielle Kriterien

Die allgemeinen Potenzrestkriterien von § 3 galten bisher als unhandlich, weil man primitive Wurzeln nur durch Probieren findet, bevor man eine Indextabelle berechnen kann, und weil das Potenzieren im Eulerschen Kriterium sehr langwierig war. Heute läßt sich beides mit den elektronischen Rechenanlagen leicht bewältigen.

Spezielle Kriterien für $n>2$ benutzen stets Zerlegungen der Primzahlen p durch quadratische Formen und sagen aus, daß q genau dann

n-ter Potenzrest mod p ist, wenn bestimmte Kongruenzen zwischen den Koeffizienten der Darstellung von p erfüllt sind. Sie sind nur für $\alpha=1$ bekannt und werden für $n/p-1$, also $p\equiv 1 \bmod n$ formuliert, da man in der Beziehung $d=(n, p-1)$ beliebig große Zahlen n wählen könnte, ohne d zu verändern.

Die ältesten Formeln stammen von Euler [7]. Für $n=3$ und die Darstellung

$$p=a^2+3b^2$$

nannte er die Bedingungen für die Fälle $q=2, 3, 5, 6, 7$ und 10. Fueter gab im Vorwort zum entsprechenden Band von Eulers Werken (S. XXI – XXV) Beweise an, ohne zu bemerken, daß das Kriterium für $q=7$ nicht ganz richtig war. Von den Beziehungen

1) $21/b$; 2) $7/a$ und $3/b$; 3) $21/a\pm b$; 4) $7/a\pm 4b$; 5) $7/2a\pm b$

sind die vierte und fünfte gleich und müssen richtig heißen: $21/a\pm 4b$. Sonst müßte die Restklasse 7 für $p=103$ mit $a=10$ und $b=1$ kubischer Rest sein, was nicht zutrifft (Genaueres bei von Lienen [16], S. 56f.).

Jacobi benutzte eine andere Zerlegung, die bis heute üblich blieb:

$$4p=A^2+27B^2.$$

Er gab die empirischen Regeln für die Primzahlen bis $q=37$ an [13]. Bei dieser Primzahldarstellung nehmen die Kriterien für $q=2, 3, 5$ und 7 die einfache Form q/AB an. Sie wurden von Nagell [20] bewiesen, nachdem Tihanyi die beiden ersten Fälle gezeigt, sich aber mit der Behauptung geirrt hatte, für Zwillingsprimzahlen ergäben sich immer gleiche Formeln [23] (vgl. von Lienen [16], S. 60f.). Allgemeine Aussagen zu den Kriterien für kubische Reste finden sich bei Bachmann [4], Nagell [18] und Hasse [10]. 1958 gab Lehmer Regeln an, wie man für alle Primzahlen q die Koeffizienten der Kongruenzen in A und B berechnen kann, die Kriterien für q als kubischen Rest sind [15]. Für einige zusammengesetzte Zahlen q bewies der Verfasser Kriterien: für $q=6$, 12, 18, 10, 20 und 50 [16]. Auf die damit bekannten Fälle sind alle zusammengesetzten Werte q reduzierbar, die nur aus Potenzen von 2 und 3 bzw. 2 und 5 bestehen.

Auch für $n=4$ stammen die ersten Kriterien von Euler [7]. Mit der Zerlegung

$$p=a^2+b^2 \qquad (a \text{ ungerade})$$

nannte er die Regeln für $q=\pm 2, 3$ und 5. Gauß bewies sie und gab die Bedingungen für die Primzahlen bis $q=23$ an [8]. Cunningham und Gosset erweiterten die Liste bis $q=41$ [6]. Allgemeine Aussagen über Kriterien für biquadratische Reste finden sich bei Hasse [10]. Auch in diesem Fall wurden von Lehmer Formeln angegeben, mit denen man die

Koeffizienten der Kriterien für alle Primzahlen q berechnen kann [15]. Der Verfasser bewies Regeln für die zusammengesetzten Zahlen $q=6$ und $q=10$ [16] aus Ergebnissen von Aigner [2].

Im Fall $n=5$ macht die quadratische Form Mühe. Lehmer benutzt die Darstellung

$$16p=a^2+50b^2+50c^2+125d^2$$

mit einigen Zusatzbedingungen und bewies Kriterien für $q=2$ und $q=3$ [14]. Alderson verwandte eine andere Zerlegung und löste damit den Fall $q=2$ [3].

Für $n=8$ betrachtet man die Darstellungen

$$p=a^2+b^2=c^2+2d^2.$$

Damit bewies Western die Bedingungen für die Primzahlen bis $q=13$ [24]. Der Fall $q=\pm 2$ wurde klassenkörpertheoretisch von Aigner neu bewiesen [1] und ebenso $q=-3, 5, -7$ von Halter-Koch [9]. Der Verfasser zeigte, daß die unterschiedlich aussehenden Regeln übereinstimmen und leitete aus den Westernschen Formeln einige hinreichende Kriterien für alle Primzahlen q und die vollständigen Bedingungen für die Primzahlen bis $q=41$ her [17].

Schließlich wurde noch der Fall $n=16$, $q=\pm 2$ von Aigner [1] und anderen behandelt und für $n=32$, $q=2$ von Hasse keine geschlossene Form mehr gefunden [11].

Literatur

1. Aigner, A.: Kriterien zum 8. und 16. Potenzcharakter der Reste 2 und −2. Deutsche Mathematik **4**, 44−52 (1939).
2. Aigner, A.: Einige handliche Regeln für biquadratische Reste. Math. Nachr. **17**, 219−223 (1958).
3. Alderson, H. P.: On the quintic character of 2. Mathematica **11**, 125−130 (1964).
4. Bachmann, P.: Die Lehre von der Kreisteilung und ihre Beziehungen zur Zahlentheorie. Leipzig: Teubner 1872. Verweis: S. 143 u.a.
5. Bachmann, P.: Niedere Zahlentheorie, Bd. 1. Leipzig-Berlin: Teubner 1921.
6. Cunningham, A. J. C., Gosset, T.: On 4-tic and 3-bic residuacity tables. The Messenger of Mathematics **50**, 1−30 (1920).
7. Euler, L.: Opera omnia, ser. I, Band 5. Zürich: Orell Füßli, Leipzig: Teubner 1944. Verweis: S. 250ff. in Tractatus de numerorum doctrina capita sedecim quae supersunt.
8. Gauß, C. F.: Theoria residuorum biquadraticorum. Werke II. Göttingen: Königliche Gesellschaft der Wissensch. 1863.
9. Halter-Koch, F.: Kriterien zum 8. Potenzcharakter der Reste 3, 5 und 7. Math. Nachr. **44**, 129−144 (1970).
10. Hasse, H.: Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, Teil II: Reziprozitätsgesetz. 1930; 2. Aufl. Würzburg-Wien: Physica-Verlag 1965.
11. Hasse, H.: Der 2^n-te Potenzcharakter von 2 im Körper der 2^n-ten Einheitswurzeln. Rend. Circ. mat. Palermo (2) **7**, 185−244 (1958).

12. Holzer, L.: Zahlentheorie. Leipzig: Teubner. Bd. I 1958, Bd. III 1965.
13. Jacobi, J.G.D.: De residuis cubicis commentatio numerosa. J. reine angew. Math. **2**, 66–69 (1827).
14. Lehmer, E.: The quintic character of 2 and 3. Duke Math. J. **18**, 11–18 (1951).
15. Lehmer, E.: Criteria for cubic and quartic residuacity. Mathematica **5**, 20–29 (1958).
16. Lienen, H. von: Regeln für kubische und höhere Potenzreste. Dissertation, Bochum, 1970.
17. Lienen, H. von: Primzahlen als achte Potenzreste. Erscheint demnächst im J. reine angew. Math.
18. Nagell, T.: Zahlentheoretische Notizen. Skrifter utgitt av det Norske Videnskaps-Akademi i Oslo, Matem.-naturv. Klasse; 1923 (1924), No. 13; hier Teil IV. – Siehe auch: Sur les restes et les non-restes cubiques. Ark. Mat. **1**, 579–586 (1952).
19. Nagell, T.: Introduction to number theory. Stockholm: Almquist-Wiksell, New York: John Wiley 1951.
20. Nagell, T.: Sur quelques problèmes dans la théorie des restes quadratiques et cubiques. Ark. Mat. **3**, 211–222 (1958).
21. Scholz, A., Schoeneberg, B.: Einführung in die Zahlentheorie. 4. Aufl. Berlin: De Gruyter 1966.
22. Storchi, E.: Alcuni criteri di divisibilità per i numeri di Mersenne e il carattere 6^{co}, 12mo, 24mo, 48mo, dell' intero 2. Boll. Un. Mat. Ital. (3) **10**, 363–375 (1955).
23. Tihanyi, M.: Köbös maradékok ismertetö jele (Kennzeichen kubischer Reste). Matematikai és fizikai lapok **27**, 76–79 (1918).
24. Western, A.E.: Some criteria for the residues of eighth and other powers. Proc. London math. Soc. (2) **9**, 244–272 (1911).
25. Winogradow, I.M.: Elemente der Zahlentheorie. Übersetz. der 6. Aufl. aus dem Russ. München: Oldenbourg 1956.

Dr. H. von Lienen
Institut für Mathematik der
Ruhr-Universität
D-4630 Bochum
Postfach 2148
Bundesrepublik Deutschland

(Eingegangen am 7. Juni 1972)

Math. Z. 129, 195–206 (1972)
© by Springer-Verlag 1972

A Characterization of the Hopf Map by Stretch

Agnes Chi-Ling Hsu

1. Introduction

Let $f\colon M^n \to N^k$ be a C^1 map of Riemannian manifolds and let f_* be the differential of f. The *stretch* $\delta_f(x)$ *of* f *at* $x \in M^n$ is defined to be the maximum of $\|f_* u\|$ over all unit tangent vectors u at x. For compact M, let

$$\delta_f = \max_{x \in M} \delta_f(x);$$

δ_f is called *the stretch of* f.

In [8; p. 387, (3.2)] Olivier remarked that the Hopf bundle maps have stretch exactly 2. He also proved [8; p. 480]:

1.1. **Theorem.** *If* $f\colon S^n \to S^2$, $n > 2$, *is* C^1 *with* $\delta_f < 2$, *then* f *is null-homotopic.*

Moreover, he conjectured that, if $f\colon S^n \to S^2$, $n > 2$, is C^1 with $\delta_f = 2$ and f is not null-homotopic, then f is a fiber bundle map. We extend his theorem and a fortiori answer his conjecture affirmatively with the following theorem.

1.2. **Theorem.** *Let* $f\colon S^n \to S^2$, $n > 2$, *be a* C^1 *map, with* $\delta_f \leqq 2$, *and let* $U(1) = S^1$ *be the subgroup of* $O(4)$ *defined by scalar multiplication on* C^2. *Then either*

a) f *is null-homotopic; or*

b) $n = 3$, $\delta_f = 2$ *and there exists a unique right coset* Λ *of* $U(1)$ *in* $O(4)$ *such that for every* $\psi \in \Lambda$, $f = h \circ \psi$, *where* $h\colon S^3 \to S^2$ *is the Hopf bundle map.*

An analysis of the Hopf bundle maps suggests introducing a class $H_k^n(\eta)$ of maps, where $\eta > 0$, characterized as: $f \in H_k^n(\eta)$ if $f\colon S^n \to S^k$ is a C^1 map such that for all $q \in S^n$ there is a k-plane $\Gamma_f(q) \subset T_q S^n$, $\Gamma_f(q) \perp \ker f_{*q}$ and $\|f_{*q} v_q\| = \eta \|v_q\|$ for all $v_q \in \Gamma_f(q)$. The Hopf bundle maps $h \in H_k^{2k-1}(2)$ $(k = 2, 4)$ (for a detailed proof see [4; p. 19, Proposition 2.3.1]). A classification of $H_k^n(\eta)$ is obtained as following:

1.3. **Theorem.** $H_k^n(\eta) = \varnothing$ *unless*

a) $n = k$ *and* $\eta = 1 : H_n^n(1) = O(n+1)$;

b) $n=k=1$ *and* η *is a positive integer*: $H_1^1(\eta)=\{g_\eta \circ \psi : g_\eta(z)=z^\eta$ *and* $\psi \in O(2)\}$; *or*

c) $(n, k)=(3, 2), (7, 4)$ *or* $(15, 8)$ *and* $\eta=2$.

Thus our study focuses on $H_k^{2k-1}(2)$ $(k=2, 4, 8)$, and we prove the following theorem, which characterizes $H_2^3(2)$, and is the essential part of the proof of Theorem 1.2.

1.4. **Theorem.** *If* $f, g \in H_2^3(2)$ *and there exists a point* $x \in S^3$ *such that*

i) $f(x)=g(x)$, *and*

ii) $f_*|T_x S^3 = g_*|T_x S^3$.

Then there is a unique $\psi \in O(4)$ *with* $\psi(x)=x$ *and* $\psi^2=$ *identity such that* $f=g\circ\psi$. (*In fact, in some coordinates*, $\psi(x_1, x_2, x_3, x_4)=(\pm x_1, x_2, x_3, x_4)$.)

This is part of a Ph. D. dissertation at Syracuse University, directed by Professor P.T. Church, and supported by S.U. Fellowship.

2. Preliminaries

In this section we summarize results from Wolf's paper [10] that are needed and observe that the differentiability hypothesis can be replaced by C^1, the same proofs carrying through.

2.1. **Definition.** Let $f: M \to N$ be a C^1 map of Riemannian manifolds, everywhere of rank $\dim N$. The *associated Ehresmann connection* $=\{H_x\}_{x\in M}$ is defined by

H_x is the orthogonal complement of $V_x = \text{Kernel}\,(f_{*x}$ on $T_x M)$.

Thus, by the Rank Theorem f_{*x} is an isomorphism of H_x onto $T_{f(x)}N$.

We define a *sectionally smooth curve* to mean one which is piecewise regular, i.e., is an immersion except for a finite number of points, and the one-sided derivatives exist at these points.

2.2. **Definition.** A tangent vector $X \in T_x M$ is *horizontal* if $X \in H_x$; a C^1 curve in M is *horizontal* if each of its tangent vectors is horizontal.

Let $\alpha(t)$, $c \leqq t \leqq d$, be a sectionally smooth curve in $f(M) \subset N$. Given $x \in f^{-1}(\alpha(c))$, if there is a C^1 horizontal curve $\alpha_x(t)$, $c \leqq t \leqq d$, in M such that i) $\alpha_x(c)=x$ and ii) $f\circ\alpha_x=\alpha$, then α_x is called the *horizontal lift of* α *to* x. If α_x exists for every $x \in f^{-1}(\alpha(c))$, then we say that α *has horizontal lifts*.

2.3. **Definition.** Let $f: M \to N$ be a C^1 map of Riemannian manifolds. f is *metric-compatible* if there exists a continuous positive real-valued function λ on N such that if $q \in M$ and $Y \in T_{f(q)}N$, then there exists an $x \in T_q(M)$ such that i) $f_* X = Y$ and ii) $\|Y\| \geqq \lambda(f(q))\|X\|$.

2.4. **Proposition.** *Let f: $M \to N$ be a metric-compatible map of connected Riemannian manifolds. If M is complete, then $f(M)=N$, and every sectionally smooth curve in N has horizontal lifts relative to the associated Ehresmann connection* [10; p. 68, Proposition 3.4].

3. $H_k^n(\eta)$

In this section we develop some basic properties of maps in $H_k^n(\eta)$ which are used to characterize $H_k^n(\eta)$.

3.1. *Remarks.*

(i) *If $f \in H_k^n(\eta)$, then* rank $f=k$; *thus* $n \geqq k$.

(ii) *If $f \in H_k^n(\eta)$, then $\Gamma_f=\{\Gamma_f(q): q \in S^n\}$ is the associated Ehresmann connection.*

(iii) *If $f \in H_k^n(\eta)$, then f is metric-compatible.*

(iv) *If $f \in H_k^n(\eta)$, then $f(S^n)=S^k$ and f: $S^n \to S^k$ is a locally trivial fiber map* (cf. [3; p. 31]; see also [10; p. 67, Corollary 2.4]).

(v) *If $f \in H_k^n(\eta)$, then for every $y \in S^k$, $f^{-1}(y)$ is a regular embedded submanifold of S^n of dimension $n-k$; if $k \geqq 2$, then each $f^{-1}(y)$ is connected.*

(vi) *If $f \in H_k^n(\eta)$, then $\delta_f \leqq \eta$.*

(vii) *If $f \in H_k^n(\eta)$, then for any u, $v \in \Gamma_f(q)$ where $q \in S^n$, $\langle f_{*q}u, f_{*q}v\rangle = \eta^2\langle u, v\rangle$.*

(viii) $H_k^n(\eta)=\{f$: $S^n \to S^k \mid f$ is C^1, $\delta_f \leqq \eta$, *and for every $q \in S^n$, there exist k-independent vectors $v_{i,q} \in T_q S^n$, $i=1, 2, \ldots, k$, $\| f_{*q}(v_{i,q})\| = \eta \|v_{i,q}\|\}$. (In fact, $\{v_{i,q}\}$ is a basis for $\Gamma_f(q)$.)*

Proof. Except for (v), the proofs are immediate.

In (v) the first statement is from Rank Theorem. Now assume that $k \geqq 2$. Since $f(S^n)=S^k$, f has a unique monotone light factorization [11; p. 141, Theorem 4.1], say $f=h \circ g$, where g is a monotone map (i.e., $g^{-1}(g(q))$ is connected, for every $q \in S^n$) onto an intermediate space K and h is light (i.e., $\dim f^{-1}(y)=0$, for all $y \in S^k$). By [9; p. 64, Lemma 2.4], K is a k-manifold and h is a d-to-1 covering map of K onto S^k for some integer d. Since $\pi_1(S^k)=0$, $d=1$. Thus h is a homeomorphism and $f^{-1}(y)= g^{-1}(h^{-1}(y))$ is connected.

Thus, in the definition of $H_k^n(\eta)$, $\ker f_{*q}$ is the tangent plane of $f^{-1}(f(q))$ at q, and the perpendicularity condition becomes: $\Gamma_f(q) \perp f^{-1}(f(g))$.

3.2. **Lemma.** *If $f \in H_k^n(\eta)$, then the horizontal lift of any geodesic is a geodesic.*

Proof. Let σ_v be a geodesic in S^k (i.e., a great circle arc) through $f(x)$, where $x \in S^n$, with velocity vector $v \in T_{f(x)} S^k$, $0<\|v\|<\pi$. σ_v: $[0, 1] \to S^k$

is defined by $\sigma_v(t)=\exp_{f(x)} t\cdot v$, and $\left\|\frac{d\sigma_v}{dt}(a)\right\|=\left\|\frac{d\sigma_v}{dt}(0)\right\|=\|v\|$ for all $a\in[0,1)$ [6; p. 55]. It follows that

$$L(\sigma_v|_{[0,a]})=\int_0^a\left\|\frac{d\sigma_v}{dt}\right\|dt=a\|v\| \qquad \text{for all } a\in[0,1).$$

Let σ_{vx}: $[0,1)\to S^n$ be any horizontal lift of σ_v to x relative to Γ_f, i.e., $\sigma_{vx}(0)=x$, $f\circ\sigma_{vx}=\sigma_v$ and for all $a\in[0,1)$, $\frac{d\sigma_{vx}}{dt}(a)\in\Gamma_f(\sigma_{vx}(a))$; in other words, $\left\|f_*\left(\frac{d\sigma_{vx}}{dt}\right)(a)\right\|=\eta\left\|\frac{d\sigma_{vx}}{dt}(a)\right\|$.

For any $a\in[0,1)$ we have

1) length of

$$\sigma_{vx}|_{[0,a]}=L(\sigma_{vx}|_{[0,a]})=\int_0^a\left\|\frac{d\sigma_{vx}}{dt}\right\|dt=\int_0^a\frac{1}{\eta}\left\|f_*\left(\frac{d\sigma_{vx}}{dt}(t)\right)\right\|dt$$

which is equal to

$$(1/\eta)\int_0^a\left\|\frac{d}{dt}(f\circ\sigma_{vx})(t)\right\|dt=(1/\eta)\int_0^a\left\|\frac{d\sigma_v}{dt}(t)\right\|dt=(1/\eta)\int_0^a\|v\|\,dt=a\|v\|/\eta.$$

Thus $a=(\eta/\|v\|)\cdot L(\sigma_{vx}|_{[0,a]})$, that is to say, the parameter of σ_{vx} is equal to $(\eta/\|v\|)$ times the arc length parameter of σ_{vx}.

2) For any other piecewise smooth curve γ joining x and $\sigma_{vx}(a)$, $L(\gamma)\geqq L(\sigma_{vx}|_{[0,a]})$.

If not, we would have $L(\gamma)<L(\sigma_{vx}|_{[0,a]})$. Since $\sigma_v|_{[0,a]}$ is the minimal geodesic joining $f(x)$ and $\sigma_{vx}(a)$, $L(f\circ\gamma)\geqq L(\sigma_v|_{[0,a]})$. Hence

$$\frac{L(f\circ\gamma)}{L(\gamma)}>\frac{L(\sigma_v|_{[0,a]})}{L(\sigma_{vx}|_{[0,a]})}=\frac{a\|v\|}{(a\|v\|/\eta)}=\eta,$$

contradicting the fact that $\delta_f\leqq\eta$.

From 1) and 2) it follows that σ_{vx} is a geodesic [6; p. 61, Corollary 10.7].

3.3. **Lemma.** *If f: $S^n\to S^k$ is C^1 and $\delta_f\leqq 1$, then either f is null-homotopic, or $n=k$ and f is orthogonal.*

Proof. If $n<k$, then f is not surjective [2; p. 1037, Proposition 4]; thus f is null-homotopic. Thus we may suppose $n\geqq k$.

If there is a point $p\in S^n$ and a $v_p\in T_pS^n$ such that $\|f_{*p}v_p\|<\|v_p\|$, view $f(S^n)$ as the image of all geodesic joining p and $-p$ (antipodal point of p) under f. Since $\delta_f\leqq 1$, either $f(-p)=-f(p)$ or $-f(p)\notin f(S^n)$. Let σ_v be the geodesic joining p and $-p$ with velocity vector $\pi v_p/\|v_p\|$. Then $L(f(\sigma_v))<\pi$, and so $f(-p)\neq -f(p)$. Hence f is not surjective, and again f is null-homotopic.

The remaining case is that for every $v \in TS^n$, $\|f_* v\| = \|v\|$. It follows immediately that $k=n$, and f is an isometry. The only isometries on S^n are orthogonal maps.

3.4. *Proof of Theorem 1.3.* Let $f \in H_k^n(\eta)$. By Lemma 3.2, any horizontal lift of a geodesic σ in S^k is a geodesic τ in S^n, and $L(\tau) = \frac{1}{\eta} L(\sigma)$. Thus if σ is of length $2\eta\pi$, then τ is of length 2π, and hence a great circle in S^n. Therefore $f|\tau$ is a map of circle to circle with stretch η everywhere. This implies that η is a positive integer and $f|\tau$ is an η-to-1 covering map.

By 3.1 (iv) and (v), $f\colon S^n \to S^k$ is a locally trivial fiber map, with a compact $(n-k)$-manifold F for fiber. In particular, if $n=k$, then $f\colon S^n \to S^k$ is a covering map.

If $n=k \geqq 2$, then $\pi_1(S^n)=1$ and f is a C^1 homeomorphism; from the first paragraph, $\eta=1$, so that (by 3.3) $f \in O(n+1)$.

Suppose $n=k=1$. Let $S^1 = \{z = e^{i\theta} \in C\colon 0 \leqq \theta \leqq 2\pi\}$ and let $g_\eta\colon S^1 \to S^1$ be defined by $g_\eta(z) = z^\eta$. By the above argument, $f\colon S^1 \to S^1$ is an η-to-1 covering map, and the domain S^1 can be viewed as a horizontal lift of the range S^1 to some point $z_0 = e^{i\theta_0}$ with $f(z_0)=1$. Since $\eta \cdot L(\tau|_{[0,\theta]}) = L(\sigma|_{[0,\theta]})$ for any θ, $f(z) = e^{\pm i\eta(\theta-\theta_0)}$. The continuity of f shows that the sign $(\pm)$ is independent of z.

Denote by α either the identity map or the reflection along x-axis, according to whether the sign is positive or negative. Let $\psi \in O(2)$ be the rotation of negative angle θ_0 (i.e., clockwise rotation through θ_0), followed by α. Then $\psi(z_0)=1$ and

$$g_\eta \circ \psi(z) = g_\eta \circ \psi(e^{i\theta}) = g_\eta(e^{\pm i(\theta-\theta_0)}) = e^{\pm i\eta(\theta-\theta_0)} = f(z),$$

for any $z = e^{i\theta} \in S^1$, i.e., $f = g_\eta \circ \psi$.

Finally, suppose that $n>k$. Using a theorem of Browder [1; p. 353, Theorem 5.1], Timourian [9; p. 64, Lemma 27] proved that if $f\colon S^n \to S^k$ is a fiber bundle with fiber a compact $(n-k)$-manifold, then $(n,k)=(3,2)$, $(7,4)$, or $(15,8)$, and the fiber is a homotopy S^1, S^3, or S^7, respectively. Thus our $f\colon S^{2k-1} \to S^k$ has fiber a homotopy S^{k-1}, $k=2,4,8$. Suppose that f is null homotopic, i.e., $f_\# = 0$. From the homotopy sequence of a fibering, we obtain the short exact sequence

$$0 \to \pi_{j+1}(S^k) \to \pi_j(S^{k-1}) \to \pi_j(S^{2k-1}) \to 0$$

$(k=2, 4,$ or $8; j=1,2,\ldots)$. For $j=2k-2$, a contradiction results, since

$$\pi_6(S^3) \approx Z_{12}, \quad \pi_7(S^4) \approx Z+Z_{12}, \quad \pi_{14}(S^7) \approx Z_{120}, \quad \pi_{15}(S^8) \approx Z+Z_{120}$$

[5; pp. 329 and 332]. Thus f is essential.

If η were 1, f would be null homotopic from above lemma; thus $\eta = 2, 3, \ldots$.

The exponential map $\exp_x$ at the point $x \in S^n$ sends the open π-ball in $T_x S^n$ diffeomorphically onto an open subset of S^n, viz $S^n - \{-x\}$. For $\varepsilon \leqq \pi$, let

$$B_\varepsilon = \{w \in \Gamma_f(x) \subset T_x S^n\colon \|w\| < \varepsilon\};$$

then $\exp_x|_{B_\pi}$ is an embedding. Let $S_x^k = \overline{\exp_x(B_\pi)}$. S_x^k can be viewed as the union of horizontal lifts to x of closed geodesics in S^k with length $\eta\pi$. Note that for each $\ell = 0, 1, \ldots, [\eta - 1/2]$, $f(\partial \exp_x B_{(2\ell+1)\pi/\eta}) = -f(x)$, and $\partial \exp_x(B_{\pi/\eta})$ is diffeomorphic to S^{k-1}. Since $f^{-1}(-f(x))$ is a connected $(k-1)$-manifold, by Brower's Theorem on invariance of domain, $\partial \exp_x B_{\pi/\eta}$ is both open and closed in $f^{-1}(-f(x))$. Hence $\partial \exp_x B_{\pi/\eta} = f^{-1}(-f(x))$. For each $\ell = 1, 2, \ldots, [\eta - 1/2]$,

$$\partial \exp_x B_{(2\ell+1)\pi/\eta} \cap \partial \exp_x B_{\pi/\eta} = \varnothing$$

implies that $[\eta - 1/2] = 0$, i.e., $\eta = 2$. This completes the proof of the Theorem.

4. A Characterization of $H_2^3(2)$

This section is devoted to the proof of Theorem 1.4.

Let σ_v be a geodesic in S^2 through $f(x)$ $(=g(x))$ with unit velocity vector $v \in T_{f(x)} S^2$ and $L(\sigma_v) = 2\pi$. By Lemma 3.2, the horizontal lifts of σ_v to x relative to Γ_f and Γ_g are geodesics of length π passing through x with velocity vector $v' \in \Gamma_f(x)$ and $v'' \in \Gamma_g(x)$ respectively, and where $f_*(v') = v = g_*(v'')$. Since $\Gamma_f(x) = \Gamma_g(x)$ and $f_*|\Gamma_f(x) = g_*|\Gamma_g(x)$ is 1-1, $v'' = v'$. By the uniqueness of geodesic in S^3 passing through x with velocity vector v' and length $= \pi$ [6; p. 56, Lemma 10.2] σ_{vx} is the unique horizontal lift of σ_v to x relative to both Γ_f and Γ_g. Thus $f \circ \sigma_{vx} = \sigma_v = g \circ \sigma_{vx}$ or

$$f|\sigma_{vx} = g|\sigma_{vx}.$$

The exponential map $\exp_x$ at the point x sends the open π-ball in $T_x S^3$ diffeomorphically onto an open subset of S^3, viz. $S^3 - \{-x\}$. For $\varepsilon \leqq \pi$, let

$$B_\varepsilon = \{w \in \Gamma_f(x) \subset T_x S^3\colon \|w\| < \varepsilon\};$$

then $\exp_x|B_\pi$ is an embedding. Let $S_x = \overline{\exp_x(B_\pi)}$. Note that S_x is a geodesic 2-sphere; and it can be viewed as the union of the images of σ_{vx} where $v \in T_{f(x)} S^2$ with $\|v\| = 1$ and $L(\sigma_{vx}) = \pi$. Hence $f(q) = g(q)$ for all $q \in S_x$. Denote $\exp_x(B_{\pi/2})$ by N_x. Then $f(y) = g(y) = -f(x)$ for all $y \in \partial N_x$. ∂N_x is diffeomorphic to S^1; $\partial N_x \subset f^{-1}(f(y)) \cap g^{-1}(g(y))$ for every $y \in \partial N_x$; and both $f^{-1}(f(y))$ and $g^{-1}(g(y))$ are manifolds of dimension 1. By Brower's theorem on invariance of domain, ∂N_x is open in both $f^{-1}(f(y))$ and $g^{-1}(g(y))$. But ∂N_x is also closed in both $f^{-1}(f(y))$ and $g^{-1}(g(y))$. Since $f^{-1}(f(y))$ and $g^{-1}(g(y))$ are connected, we have $f^{-1}(f(y)) = \partial N_x = g^{-1}(g(y))$, that is, $f^{-1}(f(y)) = g^{-1}(g(y))$ for all $y \in \partial N_x$.

(1) Thus, we have shown that if $f, g \in H_2^3(2)$, $f(x)=g(x)$, and $f_{*x}=g_{*x}$ for some $x \in S^3$, then $f|\bar{N}_x = g|\bar{N}_x$ and $\partial N_x = f^{-1}(f(y)) = g^{-1}(g(y))$ for all $y \in \partial N_x$.

The commutativity of the diagram

$$\begin{array}{ccc} B_{\pi/2} & \xrightarrow{\exp_x|} & N_x \\ {\scriptstyle f_{*x}|}\downarrow & & \downarrow{\scriptstyle f|} \\ B'_\pi & \xrightarrow[\exp_{f(x)}|]{} & S^2 - \{-f(x)\} \end{array}$$

where $B'_\pi = \{v \in T_{f(x)}S^2 : \|v\| < \pi\}$ and the fact that $f_{*x}|B_{\pi/2}$, $\exp_x|B_{\pi/2}$, $\exp_{f(x)}|B'_\pi$ are diffeomorphism imply that $f|N_x(=g|N_x)$ is a diffeomorphism. Hence $f_*|T_qN_x(=g_*|T_qN_x)$ is an isomorphism of T_qN_x onto $T_{f(q)}S^2$, for every $q \in N_x$.

Since the diagram

$$\begin{array}{ccc} B_\varepsilon & \xrightarrow{\exp_x|} & \bar{N}_x \\ {\scriptstyle f_{*x}|}\downarrow & & \downarrow{\scriptstyle f|} \\ B'_{2\varepsilon} & \xrightarrow[\exp_{f(x)}|]{} & S^2 \end{array}$$

where $B'_{2\varepsilon} = \{v \in T_{f(x)}S^2 : \|v\| < 2\varepsilon\}$, commutes for every $\varepsilon \leqq \pi/2$, f maps $\partial \exp_x B_\varepsilon = \exp_x \partial B_\varepsilon$ onto $\partial \exp_{f(x)}(B'_{2\varepsilon}) = \exp_{f(x)} \partial B'_{2\varepsilon}$. $\partial \exp_x B_\varepsilon$ is a circle in N_x with radius $\sin\varepsilon$ and $\partial \exp_{f(x)} B'_{2\varepsilon}$ is a circle in S^2 with radius $\sin 2\varepsilon$. Thus the stretch of f on $\partial \exp_x B_\varepsilon$ at the point $q \in \partial \exp_x B_{\varepsilon_q}$ is $\sin 2\varepsilon_q / \sin \varepsilon_q = 2\cos\varepsilon_q$. So is the stretch of g at q.

For every $z \in S^2 - \{f(x), -f(x)\}$, let $u_1(z)$ be the tangent vector of length 2 along the great circle from $f(x)$ to $-f(x)$ containing z. Orient S^2. Let $u_2(z) \in T_zS^2$ be orthogonal to $u_1(z)$ with length 2 so that $u_1(z)$ and $u_2(z)$ give the positive orientation. For any meridian σ (with $f(x)$ as a pole) in S^2, $u_1|\sigma - \{f(x), -f(x)\}$ and $u_2|\sigma - \{f(x), -f(x)\}$ can be extended uniquely to σ, so that $u_1|\sigma$ and $u_2|\sigma$ are continuous.

For every $q \in N_x - \{x\}$, let $w'_2(q) \in T_qN_x$ such that $f_{*q}w'_2(q) = g_{*q}w'_2(q) = u_2(f(q))$. Then $w'_2(q)$ is tangent to $\partial \exp_x B_{\varepsilon_q}$ if $q \in \partial \exp_x B_{\varepsilon_q}$ [6; p. 60, Lemma 10.5]. Hence

$$(2) \qquad \|f_{*q}(w'_2(q))\| = \|g_{*q}(w'_2(q))\| = 2\cos\varepsilon_q \|w'_2(q)\|.$$

Since $f_*|T_qN_x(=g_*|T_qN_x)$ is an isomorphism for all $q \in N_x - \{x\}$, $u_2(f(q)) \neq 0$ implies $w'_2(q) \neq 0$. Let $w_2(q) = \dfrac{w'_2(q)}{\|w'_2(q)\|}$.

For every $q \in N_x - \{x\}$, let $w_1(q) \in T_qN_x$, $t_f(q) \in \Gamma_f(q)$, and $t_g(q) \in \Gamma_g(q)$ such that $f_* w_1(q) = g_* w_1(q) = u_1(f(q))$ and $f_* t_f(q) = u_2(f(q)) = g_* t_g(q)$.

For each half meridian σ_y on $\bar{N}_x$ beginning at x ending at $y \in \partial N_x$, $w_1|\sigma_y - \{x, y\}$, $w_2|\sigma_y - \{x, y\}$, $t_f|\sigma_y - \{x, y\}$ and $t_g|\sigma_y - \{x, y\}$ can be extended to σ_y uniquely and continuously. Thus

$$f_* \circ w_1|\sigma_y = g_* \circ w_1|\sigma_y = u_1|f \circ \sigma_y;$$

$$f_* \circ t_f|\sigma_y = g_* \circ t_g|\sigma_y = u_2|f \circ \sigma_y = 2\frac{f_* \circ w_2|\sigma_y}{\|f_* \circ w_2|\sigma_y\|} = 2\frac{g_* \circ w_2|\sigma_y}{\|g_* \circ w_2|\sigma_y\|};$$

$$\|w_1|\sigma_y(q)\| = 1 \quad \text{and} \quad \|t_f|\sigma_y(q)\| = \|t_g|\sigma_y(q)\| = \tfrac{1}{2}\|f_*(t_f|\sigma_y(q))\| = \tfrac{1}{2}\|u_2(f(q))\| = \tfrac{1}{2} \cdot 2 = 1$$

for every $q \in \sigma_y$. Moreover, since $u_1(f(q)) \perp u_2(f(q))$ for all $q \in \sigma_y$, by 3.1(vii) and [6; p. 60, Lemma 10.5], $t_f|\sigma_y(q)$, $t_g|\sigma_y(q)$ and $w_2(q)$ are all orthogonal to $w_1(q)$.

Alternatively $u_2|f \circ \sigma_y(f(x))$ of length 2 (resp. $w_2|\sigma_y(x)$ of length 1) is orthogonal to $f \circ \sigma_y$ at $f(x)$ (resp. σ_y at x) and translates parallel along $f \circ \sigma_y$ (resp. σ_y) to define $u_2|f \circ \sigma_y$ (resp. $w_2|\sigma_y$).

Now $T_q S^3 = \Gamma_f(q) \oplus V_f(q) = \Gamma_g(q) \oplus V_g(q)$, where $V_f(q) = \text{kernel of } f_{*q}$, $V_g(q) = \text{kernel of } g_{*q}$. For any point q on the half meridian σ_y on $\bar{N}_x$,

$$w_2|\sigma_y(q) = a_y(q)\, t_f|\sigma_y(q) + v_f(q), \qquad \text{for some } v_f(q) \in V_f(q);$$

and

$$w_2|\sigma_y(q) = b_y(q)\, t_g|\sigma_y(q) + v_g(q), \qquad \text{for some } v_g(q) \in V_g(q).$$

$$\begin{aligned}\|f_{*q}(w_2|\sigma_y(q))\|^2 &= \langle a_y(q) f_{*q}(t_f|\sigma_y(q)), a_y(q) f_{*q}(t_f|\sigma_y(q))\rangle \\ &= (a_y(q))^2 \langle f_{*q}(t_f|\sigma_y(q)), f_{*q}(t_f|\sigma_y(q))\rangle \\ &= (a_y(q))^2 \|u_2|f \circ \sigma_y(f(q))\|^2 \\ &= 4(a_y(q))^2.\end{aligned}$$

On the other hand, by (2), $\|f_{*q}(w_2|\sigma_y(q))\|^2 = (2\cos\varepsilon_q)^2$ where $q \in \partial \exp_x B_{\varepsilon_q}$ and $0 \leqq \varepsilon_q \leqq \pi$. Thus $a_y(q) = \pm\cos\varepsilon_q$. $a_y|\sigma_y - \{y\}$ is continuous and non-zero (since $\varepsilon_q \neq \pi/2$ for $q \neq y$); the connectedness of $\sigma_y - \{y\}$ implies that $a_y|\sigma_y - \{y\}$ is either positive or negative. Since $\Gamma_f(x) = T_x N_x$, $t_f|\sigma_y(x) = w_2|\sigma_y(x)$, so that $a_y(x) = 1 > 0$. Hence $a_y \geqq 0$ and $a_y(q) = \cos\varepsilon_q$ for all $q \in \sigma_y$. Same argument shows that $b_y(q) = \cos\varepsilon_q$.

Since for every $q \in \bar{N}_x$, $\Gamma_f(q) \perp V_f(q)$ and $\Gamma_g(q) \perp V_g(q)$;

$$\begin{aligned}\langle w_2|\sigma_y(q), t_f|\sigma_y(q)\rangle &= a_y(q)\langle t_f|\sigma_y(q), t_f|\sigma_y(q)\rangle = a_y(q) = \cos\varepsilon_q \\ &= b_y(q) = b_y(q)\langle t_g|\sigma_y(q), t_g|\sigma_y(q)\rangle = \langle w_2|\sigma_y(q), t_g|\sigma_y(q)\rangle,\end{aligned}$$

for every $q \in \sigma_y$ and $y \in \partial N_x$. Moreover

$$\cos \sphericalangle(w_2|\sigma_y(q), t_f|\sigma_y(q)) = \langle w_2|\sigma_y(q), t_f|\sigma_y(q)\rangle = \cos \varepsilon_q$$
$$= \langle w_2|\sigma_y(q), t_g|\sigma_y(q)\rangle = \cos \sphericalangle(w_2|\sigma_y(q), t_g|\sigma_y(q)) \geqq 0.$$

Hence $\sphericalangle(w_2|\sigma_y(q), t_f|\sigma_y(q)) = \pm \sphericalangle(w_2|\sigma_y(q), t_g|\sigma_y(q))$.

Since $w_2|\sigma_y(q)$, $t_f|\sigma_y(q)$ and $t_g|\sigma_y(q)$ are on the 2-plane in $T_q S^3$ orthogonal to $w_1|\sigma_y(q)$, either $t_f|\sigma_y(q) = t_g|\sigma_y(q)$ or $\sphericalangle(t_f|\sigma_y(q), t_g|\sigma_y(q)) = 2\varepsilon_q$. In particular, at y, either $t_f(y) = t_g(y)$ or $\langle t_f(y), t_g(y)\rangle = \cos\left(2 \cdot \frac{\pi}{2}\right) = -1$, i.e., $t_f(y) = -t_g(y)$. By continuity of the function $q \to \langle t_f|\sigma_y(q), t_g|\sigma_y(q)\rangle$ from σ_y to R, if $t_f(y) = t_g(y)$ then $t_f|\sigma_y = t_g|\sigma_y$; and if $t_f(y) = -t_g(y)$ then $\langle t_f|\sigma_y(q), t_g|\sigma_y(q)\rangle = \cos 2\varepsilon_q$ for all $q \in \sigma_y$.

On the other hand, by the continuity of the function $y \to \langle t_f(y), t_g(y)\rangle$ on ∂N_x; and the connectedness of ∂N_x, $\langle t_f(y), t_g(y)\rangle = 1$ for all $y \in \partial N_x$ or $\langle t_f(y), t_g(y)\rangle = -1$ for all $y \in \partial N_x$. Hence if $t_f(y) = t_g(y)$ for some $y \in \partial N_x$, then $t_f = t_g$ on ∂N_x. Thus, in that case, $t_f = t_g$ on $\bar{N}_x - \{x\}$ so that $f_{*q} = g_{*q}$ for all $q \in \bar{N}_x$.

(3) Now let t_f and t_g defined above be denoted by $t_{f,x}$ and $t_{g,x}$ to show the dependence on the initial point x. We have shown that if $f, g \in H_2^3(2)$, $f(x) = g(x)$ and $f_{*x} = g_{*x}$ for some $x \in S^3$, then $\langle t_{f,x}(y), t_{g,x}(y)\rangle = \pm 1$ for all $y \in \partial N_x$. Moreover, if $\langle t_{f,x}(y), t_{g,x}(y)\rangle = 1$ for some $y \in \partial N_x$, then $f_{*q} = g_{*q}$ for all $q \in \bar{N}_x$.

Now return to our proof, and pick a point $y_0 \in \partial N_x$. Then there exists a unique map $\psi \in O(4)$ such that ψ leaves the geodesic 2-sphere containing $\bar{N}_x$ pointwise fixed, and $\psi_*(t_f(y_0)) = t_g(y_0)$. Actually if $t_g(y_0) = t_f(y_0)$, then ψ is the identity; if $t_g(y_0) = -t_f(y_0)$, then ψ is a reflection such that $\psi_{*y_0}(t_f(y_0)) = -t_f(y_0) = t_g(y_0)$. Hence $g \circ \psi \in H_2^3(2)$; $g \circ \psi(q) = g(q) = f(q)$ for all $q \in \bar{N}_x$ (by (1)); and $(g \circ \psi)_{*x} = f_{*x}$. Since f, g and $g \circ \psi$ satisfy the hypothesis of this theorem, for every $y \in \partial N_x$,

$$\partial N_x = f^{-1}(f(y)) = g^{-1}(g(y)) = (g \circ \psi)^{-1}(g \circ \psi)(y)$$

by (1). Thus $\Gamma_f(y) = \Gamma_g(y) = \Gamma_{g \circ \psi}(y)$. Now

$$(g \circ \psi)_* t_f(y_0) = g_*(\psi_* t_f(y_0)) = g_*(t_g(y_0)) = u_2|\sigma_{y_0}(f(y_0)).$$

Therefore $t_f(y_0) = t_{g \circ \psi}(y_0)$, i.e., $\langle t_f(y_0), t_{g \circ \psi}(y_0)\rangle = 1$. By (3) ($f(q) = (g \circ \psi)(q)$ and) $f_{*q} = (g \circ \psi)_{*q}$ for all $q \in \bar{N}_x$.

For every $q \in \bar{N}_x$, we can define N_q and the vector fields $t_{f,q}$ and $t_{g \circ \psi, q}$ on $\bar{N}_q - \{q\}$ and for all $z \in \partial N_q$, $t_{f,q}(z) = \pm t_{g \circ \psi, q}(z)$. Define α: $\bar{N}_x \to \{-1, 1\}$ by

$$\alpha(q) = \langle t_{f,q}(z), t_{g \circ \psi, q}(z)\rangle = \pm 1$$

(independent of the choice of $z \in \partial N_q$). α is continuous; and $\alpha(x)=1$. Therefore $\alpha \equiv 1$, i.e., $t_{f,q}(z)=t_{g\circ\psi,q}(z)$ for all $z\in\partial N_q$ and all $q\in \bar{N}_x$.

Again by (3), $f_{*p}=(g\circ\psi)_{*p}$ for all $p\in \bar{N}_q$, where $q\in \bar{N}_x$. In particular, $f_{*p}=(g\circ\psi)_{*p}$ for all $p\in\partial N_q$. Since $q\in\partial N_p$, by (1), we have $f^{-1}(f(q))=(g\circ\psi)^{-1}(g\circ\psi(q))$. That is, for all $q\in \bar{N}_x$, $f^{-1}(f(q))=(g\circ\psi)^{-1}(g\circ\psi(q))$. Since $f(\bar{N}_x)=S^2$, given any $z\in S^2$, $z\in f^{-1}(f(q))$ for some $q\in \bar{N}_x$. Thus $g\circ\psi(z)=g\circ\psi(q)=f(q)=f(z)$, and the conclusion follows.

5. Proof of Theorem 1.2

We first extend Theorem 1.1 (Olivier) to the following lemma.

5.1. Lemma. *Let f: $S^n\to S^2$ be a C^1 map, with $n>2$ and $\delta_f\leqq 2$. If there exists a point $q\in S^n$ such that at most one tangent vector v_q in T_qS^n has $\|f_{*q}v_q\|=2\|v_q\|$, then f is null homotopic.*

Proof. We use the notation of the proof of [8] and extend that proof. Let

$$S^n=\{x\in R^{n+1}\colon\ x_1^2+\cdots+x_n^2+(x_{n+1}-1)^2=1\}.$$

We may assume q is the north pole, i.e., $q=(0,\ldots,0,2)$. Let v be any non-zero tangent vector at q along maximal stretch direction. Thus $\|f_* v\|\leqq 2\|v\|$. Visualize T_qS^n as $\{(x_1,\ldots,x_n,0)\in R^{n+1}\}$; then v has coordinates $(a_1,\ldots,a_n,0)$. Since $v\neq 0$, at least one of the $a_i\neq 0$, say $a_n\neq 0$. Then $v\notin\{(x_1,x_2,0,\ldots,0)\in R^{n+1}\}$. Let

$$S^{n-2}=\{x\in S^n\colon\ x_1=x_2=0\}.$$

Any point $P\in S^{n-2}-\{0\}$ together with x_1-x_2-plane determines a 3-dimensional linear subspace L_P in R^{n+1}, and $L_P\cap S^n=S_P^2$ a 2-sphere. Denote $S_0^2=\{0\}$. For any two distinct points P_1 and P_2 in S^{n-2}, $S_{P_1}^2\cap S_{P_2}^2=\{0\}$; and every point of $S^n-\{0\}$ lies on exactly one of the 2-spheres. Let $S^1=\{x\in R^{n+1}\colon\ x_1^2+x_2^2=1;$ and $x_i=0$ for all $i\geqq 3\}$. Any $P\in S^{n-2}$, and $u\in S^1$ determine a meridian $M_{P,u}$: $[0,1]\to S_P^2$ with reduced arc length parameter such that $M_{P,u}(0)=0$ and $M_{P,u}(1)=P$. If $P\neq q$, then $L(M_{P,u})<\pi$ for all $u\in S^1$; if $P=q$, then $L(M_{P,u})=\pi$ for all $u\in S^1$.

In Olivier's proof, as long as $L(f(M_{P,u}))<2\pi$ for all $P\in S^{n-2}$ and $u\in S^1$, the theorem results. In our case, when $P\neq q$,

$$L(f(M_{P,u}))\leqq 2\cdot L(M_{P,u})<2\pi,$$

for all $u\in S^1$. When $P=q$, remember that $v\notin\{(x_1,x_2,0,\ldots,0)\in R^{n+1}\}$, which can be visualized as $T_qS_q^2$ and $\|f_*(w_q)\|=2\|w_q\|$ only if w_q is a multiple of $v_q=v$. Thus the stretch of f along any meridian $M_{q,u}$ is less than 2, so that

$$L(f(M_{P,u}))<2\cdot L(M_{P,u})=2\pi.$$

The conclusion results.

5.2. *Remarks.* (i) *In a canonical way $SO(2)=U(1)$ is a subgroup of $O(4)$.* Specifically, identify C^2 with R^4, and let $S^1 \subset C$ act on C^2 by scalar multiplication. The subgroup thus is the set of matrices in the form

$$\begin{pmatrix} a & -b & 0 & 0 \\ b & a & 0 & 0 \\ 0 & 0 & a & -b \\ 0 & 0 & b & a \end{pmatrix}$$

where $a, b \in R$ satisfy $a^2+b^2=1$.

(ii) Let A denote either the complex numbers, the quoternions or the Cayley numbers. For $k=2, 4, 8$

$$S^{2k-1}=\{(z_1, z_2) \in A \times A, z_1 \bar{z}_1 + z_2 \bar{z}_2 = 1\},$$

and

$$S^k = \{(z, t) \in A \times R, z\bar{z} + t^2 = 1\}.$$

The map $h: S^{2k-1} \to S^k$ defined by

$$h(z_1, z_2) = (2 z_1 \bar{z}_2, |z_2|^2 - |z_1|^2)$$

is the Hopf map [7; p. 102].

5.3. **Lemma.** *Let $h: S^3 \to S^2$ be the Hopf map, and let $\lambda \in O(4)$. Then $h = h \circ \lambda$ if and only if $\lambda \in SO(2)$.*

Proof. For $\lambda \in C$ with $|\lambda|=1$ (i.e., $\lambda \in SO(2)$), $h(\lambda z_1, \lambda z_2) = h(z_1, z_2)$, proving "if".

Now suppose that $h = h \circ \lambda$. We first suppose, in addition, that $\lambda(0,1)=(0,1)$, and deduce that λ is the identity. Since λ maps fibers into fibers, $\lambda_{*(0,1)}(\Gamma_h(0,1)) = \Gamma_h(0,1)$. Since $h_{*(0,1)}|\Gamma_h(0,1)$ is an isomorphism, $\lambda_{*(0,1)}|\Gamma_h(0,1)$ is inclusion. Thus λ on the plane $C \times \{0\}$ is inclusion. Also $V_h(0,1) \xrightarrow{\lambda_*(0,1)} V_h(0,1)$ is plus or minus the identity (where $V_h(0,1)$ is the tangent space of the fiber of h at $(0,1)$). Thus $\lambda(z_1, z_2) = (z_1, \bar{z}_2)$ or λ is the identity. The former map does not satisfy $h = h \circ \lambda$, so that λ is the identity.

Finally, suppose $h = h \circ \lambda$, where $\lambda \in O(4)$. There is $\mu \in SO(2)$ (so that $h = h \circ \mu$) with $\mu(0,1) = \lambda(0,1)$. Since $\mu^{-1} \circ \lambda$ satisfies the previous property, it is the identity, so that $\lambda \in SO(2)$.

5.4. *Proof of Theorem 1.2.* If there exists a point q in S^n such that f has at most one direction with stretch 2 at q, then f is null-homotopic by 5.1.

Therefore we may suppose that for every $q \in S^n$, there are two independent unit vectors v_1 and v_2 in $T_q S^n$ with $\|f_{*q}(v_1)\| = \|f_{*q}(v_2)\| = 2$. By 3.1 (viii), $f \in H_2^3(2)$, and $h \in H_2^3(2)$ also; the existence of ψ follows from 1.4.

Suppose that $f=h\circ\varphi$, for another $\varphi\in O(4)$. Then $h=h\circ\psi\circ\varphi^{-1}$, so that $\psi\circ\varphi^{-1}\in SO(2)$ by 5.3, i.e., φ and ψ are in the same right coset.

5.5. *Remark. If* $f\in H_2^3(2)$ *— precisely if* $f=h\circ\psi$, *where* $\psi\in O(4)$ (Theorem 1.2), *then* $f\circ\varphi=f$ *if and only if* φ *is in the conjugate subgroup* $\psi^{-1}\circ SO(2)\circ\psi$.

References

1. Browder, W.: Higher torsion in H-spaces. Trans. Amer. Math. Soc. **108**, 353–375 (1963).
2. Church, P.T.: On points of Jacobian rank k, II. Proc. Amer. Math. Soc. **16**, 1035–1058 (1965).
3. Ehresmann, Ch.: Les connexions infinitésmales dans un espace fibré différentiable. Colloque de Topology, Bruxelles, 29–55 (1950).
4. Hsu, A.: The stretch of a differentiable map of Riemannian manifolds, Ph. D. Dissertation, Syracuse University, 1972.
5. Hu, S.T.: Homotopy theory. New York: Academic Press 1959.
6. Milnor, J.: Morse theory. Princeton, N.J.: Princeton Univ. Press 1963.
7. Milnor, J.: Singular points of complex hypersurfaces. Princeton, N.J.: Princeton Univ. Press 1963.
8. Olivier, R.: Über die Dehnung von Sphärenabbildungen. Inventiones Math. **1**, 380–390 (1966).
9. Timourian, J.G.: Fiber bundles with descrete singular set. J. Math. Mech. **18**, 61–70 (1968).
10. Wolf, J.A.: Differentiable fiber space and mappings compatible with Riemannian metrices. Michigan Math. J. **11**, 65–70 (1964).
11. Whyburn, G.T.: Analytic topology. Colloq. Publ. Vol. 28. Providence, R.I.: Amer. Math. Soc. 1942.

Dr. Agnes Chi-Ling Hsu
Department of Mathematics
University of Alberta
Edmonton 7, Alberta
Canada

(Received August 7, 1972)

Math. Z. 129, 207–230 (1972)
© by Springer-Verlag 1972

Decomposition of Modules over Right Uniserial Rings

Vlastimil Dlab and Claus Michael Ringel

One of the fundamental problems in the theory of rings is to characterize the rings of bounded representations type in terms of their structure. The problem has been solved for some classes of rings (for references, see Gabriel [8]); however, in general, it is still open. In the present paper, we give a complete solution of the problem for the class of (local) right uniserial rings with a commutative residue field.

Theorem. *Let R be a right uniserial ring with a commutative residue field $Q=R/W$. Then R is of bounded representation type if and only if R is also left uniserial or if $W^2=0$ and ${}_RW$ is of length $\leqq 3$. Moreover, if R is not uniserial and ${}_RW$ is of length 2 or of length 3, then there are just 3, or 5, isomorphism classes of finitely generated indecomposable left R-modules, respectively, and every left R-module is a direct sum of these. Similarly, if R is not uniserial and ${}_RW$ is of length 2 or of length 3, then there are just 3, or 5, isomorphism classes of finitely generated indecomposable right R-modules, respectively, and every right R-module is a direct sum of these.*

Here, by a *right uniserial* ring R we understand a right artinian ring whose right ideals are linearly ordered by inclusion. In particular, such a ring is local.

A ring R is said to be of *bounded, or finite, representation type* if the lengths of the finitely generated indecomposable left R-modules are bounded, or if there is only a finite number of finitely generated indecomposable left R-modules, respectively (cf. [4]). The apparent asymmetry in this definition is removed by Proposition 1.1. The fact that, for a semi-primary ring R, the number of the finitely generated indecomposable left R-modules equals the number of the finitely generated indecomposable right R-modules, implies in conjunction with a recent result of Tachikawa and Ringel [11] that, for a right artinian ring R of finite representation type, every right R-module is a direct sum of finitely generated indecomposable right R-modules, that the last statement of Theorem is a consequence of the preceeding one. We remark that Tachikawa-Ringel theorem is not used to derive direct decompositions of left

R-modules, but that these decompositions are constructed directly. Also, if $W^2=0$ and the length $\partial_R W=3$, five finitely generated indecomposable right R-modules can be described explicitely and an alternative proof of the fact that they are the only ones, based on Tachikawa's duality theory [10], can be given (Remark 6.4).

Now, semi-primary rings of finite representation type are necessarily left artinian, and thus of bounded representation type. And, Theorem asserts that, for right uniserial rings with a commutative residue field, also the converse, i.e. the Brauer-Thrall conjecture, holds. The proof of the essential part of this statement is given in § 5 and involves an extension of Roiter's method in [9]. There, Roiter proved the Brauer-Thrall conjecture for finite-dimensional algebras.

Throughout the paper, R denotes a right uniserial ring, W its radical and $Q=R/W$ the residue division ring which is always assumed to be commutative. Considering W/W^2 as a left or right vector space over Q, it is easy to see that if R is not left uniserial, Q must be infinite. In §§ 2, 3 and 4, we consider the case when $W^2=0$. First, the indecomposable injective left R-module is shown to be of length 2 in § 2. Then, in the case when the length $\partial_R W$ of ${}_R W$ is greater than 4, we construct an infinite family of non-isomorphic local left R-modules in § 3. And, in § 4, we give the description of the five indecomposable left R-modules in the case $\partial_R W=3$, and show that every left R-module is a direct sum of these modules. Let us remark that the case $\partial_R W=2$ was treated in [5]. Finally, in § 6, the investigation of rings R with $W^2\neq 0$ is reduced to the case $W^2=0$, which completes the proof of Theorem.

We conclude the introduction with a brief remark on our notation. If A is a ring (with unity), all A-modules are assumed to be unital. The symbols ${}_A M$ or M_A will be used to underline the fact that M is a left or a right A-module, respectively. The length of M will be denoted by ∂M, the socle of M by $\operatorname{Soc} M$.

§ 1. Correspondence Between Left and Right A-Modules

In [2], Auslander and Bridger have defined a duality functor by using projective resolutions of finitely presented A-modules. Starting with a finitely presented left A-module ${}_A M$ and a finite presentation f of ${}_A M$, that is a morphism which gives rise to an exact sequence

$$ {}_A P \xrightarrow{\ f\ } {}_A Q \to {}_A M \to 0 $$

with ${}_A P$, ${}_A Q$ finitely generated projective, they apply the functor

$$ * = \operatorname{Hom}_A({}_A -, {}_A A_A) $$

to f and consider the cokernel M'_A of f^*,

$$Q_A^* \xrightarrow{f^*} P_A^* \to M'_A \to 0.$$

Obviously, this is a finite presentation of the *right* A-module M'_A. Also, starting with a finitely presented right A-module and its finite presentation g, the application of the functor $\mathrm{Hom}_A(-_A, {}_AA_A)$, which we denote again by $*$, leads to a finitely presented left A-module, namely $\mathrm{Cok}\, g^*$. For a finitely generated projective module ${}_AP$, we have ${}_AP^{**} \cong {}_AP$, and also for a finite presentation f: ${}_AP \to {}_AQ$ of ${}_AM$ we get $f^{**} \cong f$ (as morphisms). Thus, ${}_AM = \mathrm{Cok} f \cong \mathrm{Cok} f^{**}$, that is to say, it is possible to get ${}_AM$ back. However, it should be noted that starting with ${}_AM$, the module M'_A is not uniquely determined, since we may use another presentation; it is only determined up to a "stable equivalence" [2].

Auslander and Bridger usually assume that A is noetherian, but the above procedure works obviously for arbitrary rings. Moreover, for a semi-perfect ring A, the existence of projective covers [3] enables us to define a one-to-one correspondence between finitely presented indecomposable left A-modules and finitely presented indecomposable right A-modules. For, if ${}_AM$ is finitely presented, we consider only minimal presentations

$${}_AP \xrightarrow{f} {}_AQ \xrightarrow{p} {}_AM \to 0,$$

that is p and f are projective covers of ${}_AM$ and $\mathrm{Ker}\, p$, respectively. Two such minimal presentations f_1 and f_2 of ${}_AM$ are isomorphic (as morphisms); therefore also $f_1^* \cong f_2^*$ and $\mathrm{Cok} f_1^* \cong \mathrm{Cok} f_2^*$. Moreover, if ${}_AM$ is indecomposable and not projective, then it is easily seen that f^* is a minimal presentation of $\mathrm{Cok} f^*$. Now, ${}_AM$ is indecomposable if and only if f cannot be decomposed as $f = f_1 \oplus f_2, f_i\colon P_i \to Q_i$, with $P = P_1 \oplus P_2$ and a non-trivial decomposition $Q = Q_1 \oplus Q_2$. In this case, in view of the minimality, there is not even a decomposition $f = f_1 \oplus f_2$ with, say, $f_2\colon P_2 \to 0$ and $P_2 \neq 0$. We have $f = f_1 \oplus f_2$ if and only if $f^* = f_1^* \oplus f_2^*$; thus, if ${}_AM$ is indecomposable, also $\mathrm{Cok} f^*$ is indecomposable. Altogether we get a one-to-one correspondence between the finitely presented indecomposable left A-modules which are not projective and the finitely presented indecomposable right A-modules which are not projective. But the finitely generated indecomposable projective left A-modules and the finitely generated indecomposable projective right A-modules are in a one-to-one correspondence using the functor $*$. By means of this correspondence we can easily sharpen the result of Eisenbud and Griffith [7] to the following

Proposition 1.1. *Let A be semi-primary.*

(a) *The lengths of the finitely generated indecomposable left A-modules are bounded if and only if the lengths of the finitely generated indecomposable right A-modules are bounded.*

(b) *There is only a finite number of finitely generated indecomposable left A-modules if and only if there is only a finite number of finitely generated indecomposable right A-modules and, in this case, the numbers are equal.*

Moreover, if (a) *or* (b) *holds, then A is left and right artinian.*

Proof. (a) If the lengths of the finitely generated indecomposable left A-modules are bounded, then A is obviously left artinian. Assume that A is not right artinian, and write $A_A = P_A \oplus P'_A$, with an indecomposable P_A which is not artinian. For every natural n, we find a submodule K_n of P_A which is generated by n elements and cannot be generated by less than n elements. Let

$$Q_n \xrightarrow{f_n} P_A \to P_A/K_n \to 0$$

be a minimal presentation of P_A/K_n. Calculating a bound for the length of the left A-module $\operatorname{Cok} f_n^*$, we get

$$\partial \operatorname{Cok} f_n^* \geqq \partial\, {}_AQ_n^* - \partial\, {}_AP^* \geqq n - \partial\, {}_AP^*,$$

and thus (since $\partial\, {}_AP^*$ is finite), $n - \partial\, {}_AP^*$ can be arbitrarily large. This shows that A is right artinian. It remains to prove that also the lengths of the finitely generated indecomposable right A-modules are bounded. But these modules are finitely presented (because A is right artinian), and therefore they occur as $\operatorname{Cok} f^*$, where f are minimal presentations of finitely generated indecomposable left A-modules ${}_AM$, say

$${}_A\bar{P} \to {}_A\bar{Q} \to {}_AM.$$

In particular, every $\operatorname{Cok} f^*$ is an epimorphic image of $\bar{P}_A^*$. But, by our assumption, there is a bound m such that every ${}_A\bar{Q}$ is generated by less than m elements, and consequently there is also a bound m' such that every submodule of ${}_A\bar{Q}$ is generated by less than m' elements (take, for example, $m' = m \cdot \partial\, {}_AA$). Therefore, also ${}_A\bar{P}$ and $\bar{P}_A^*$ are generated by less than m' elements. This proves (a).

(b) Now, assume that there is only a finite number of finitely generated indecomposable left A-modules. Then there is only a finite number of finitely presented indecomposable left A-modules, and therefore also only a finite number of finitely presented indecomposable right A-modules. Consequently, the semiprimary ring A is both left and right artinian. For, assume that A is not left (or right) artinian and let ${}_AP$ (or P_A) be an indecomposable direct summand of ${}_AA$ (or A_A) which is not artinian. If Q is a submodule of P of finite length, then P/Q is obviously finitely presented and $P \to P/Q$ is its projective cover. Now, every isomorphism $P/Q \to P/Q'$ can be lifted to an automorphism of P which maps Q onto Q' and thus $P/Q \cong P/Q'$ implies $\partial Q = \partial Q'$. Since P is not left (or right) artinian, there are submodules of P of arbitrarily large length and therefore there is an infinite number of non-isomorphic finitely presented indecomposable left (or right) A-modules, in contradiction to our hypothesis. Finally, finitely generated modules over an artinian ring are finitely presented, and hence (b) follows.

§ 2. Rings with $W^2=0$

As stated in the introduction, R stands always for a right uniserial ring, W for its radical and Q for its residue division ring R/W which is assumed to be commutative.

Here, in addition, we assume that $W^2=0$. Consequently, W can be considered as a left or a right vector space over Q. One of the main tools of our paper is the following generalization of the concept of an algebra over the field Q (cf. [5]).

Definition. A bimodule ${}_QV_Q$ over a field Q with a multiplication $\circ$ is said to be an *algebra*, if $(V, \circ)$ is a ring and if, for all $\kappa_1, \kappa_2 \in Q$ and $v_1, v_2 \in V$,

$$(\kappa_1 v_1) \circ (v_2 \kappa_2) = \kappa_1 (v_1 \circ v_2) \kappa_2 .$$

Thus, if $W^2=0$ and w_1 is an arbitrary (fixed) non-zero element of W, we can define a Q-isomorphism $\varphi\colon Q_Q \to W_Q$ by $\rho\varphi = w_1 \rho$ and define a multiplication $\circ$ on W as follows

$$(w_1 \rho_1) \circ (w_1 \rho_2) = w_1 \rho_1 \rho_2 \qquad \text{for all } \rho_1, \rho_2 \in Q .$$

One can see immediately that ${}_QW_Q$ is an algebra with respect to the operation $\circ$ and that $(W, \circ)$ is isomorphic to Q; it is therefore commutative and w_1 is its identity element. The proofs of the statements which follow will illustrate the use of this concept.

Lemma 2.1. *Let $W^2=0$ and $0 \neq w' \in W$ with*

$$\lambda w' = w' \rho \qquad \textit{for some } \lambda, \rho \in R .$$

Then $\lambda w = w \rho$ for all $w \in W$.

Proof. Obviously, we can assume that $\lambda \notin W$ and $\rho \notin W$. Then, we consider the algebra $(W, \circ)$ with the identity $w_1 = w'$ and, writing

$$\bar\lambda = \lambda + W \in Q, \qquad \bar\rho = \rho + W \in Q,$$

calculate

$$\begin{aligned}\lambda w = \bar\lambda w = (\bar\lambda w) \circ w' = \bar\lambda (w \circ w') = \bar\lambda (w' \circ w) = (\bar\lambda w') \circ w \\ = (w' \bar\rho) \circ w = w \circ (w' \bar\rho) = (w \circ w') \bar\rho = w \rho ,\end{aligned}$$

as required.

Lemma 2.2. *Let* $W^2=0$ *and*

$$T=\{\tau\in R \mid w\tau\in Rw \text{ for all } w\in W\}.$$

Then $\partial_R W=\partial Q_T$.

Proof. First, notice that

$$\partial_R W=\partial_Q W \quad\text{and}\quad \partial Q_T=\partial Q_{T/W}=\partial_{T/W} Q,$$

where T/W is obviously a subfield of Q. Thus, in order to establish the lemma, it is sufficient to show that

$$\partial_{T/W} Q=\partial_Q W.$$

Notice that, in view of Lemma 2.1,

$$T=\{\tau\in R \mid w_1\tau\in Rw_1 \text{ for a (fixed) non-zero } w_1\in W\}.$$

Now, writing $\bar\rho=\rho+W\in Q$, $\bar\lambda=\lambda+W\in Q$ and $\bar\tau=\tau+W\in T/W$, define the morphisms $\alpha\colon Q\to W$ and $\beta\colon T/W\to Q$ by

$$\bar\rho\alpha=w_1\rho$$

and

$$\bar\tau\beta=\bar\lambda \quad\text{with } \lambda \text{ satisfying } w_1\tau=\lambda w_1.$$

Clearly, both α and β are well-defined bijections. In fact, it is easy to verify that α is an isomorphism between the additive groups of Q and W and β is a ring isomorphism of T/W and Q. Moreover, if $\rho\in R$, $\tau\in T$ and $\lambda\in R$ with $w_1\tau=\lambda w_1$, then $(\bar\tau\bar\rho)\alpha=\overline{\tau\rho}\alpha=w_1\tau\rho=\lambda w_1\rho=\bar\lambda w_1\rho=(\bar\tau\beta)(\bar\rho\alpha)$, and this implies the required equality.

Proposition 2.3. *Let* $W^2=0$ *and* $\dim_Q W=s$. *Then the injective indecomposable left R-module* ${}_RE$ *is of length* 2.

Proof. First, let us show that the right action of Q on W is transitive on hyperplanes, i.e. on $(s-1)$-dimensional subspaces of ${}_QW$. Let us choose a basis

$$w_1, w_2, \ldots, w_s \quad \text{of } {}_QW,$$

take the hyperplane H generated by the vectors $w_1, w_2, \ldots, w_{s-1}$ and show that, for any given hyperplane H' generated by $s-1$ vectors

$$v_k=\sum_{j=1}^{s}\mu_{kj}w_j, \qquad 1\leqq k\leqq s-1,$$

there exists $\alpha\in Q$ such that $H\alpha=H'$; let

$$w_1\alpha=\sum_{j=1}^{s}\alpha_j w_j.$$

Notice that

$$w_i\,\alpha=(w_i\circ w_1)\,\alpha=(w_1\,\alpha)\circ w_i=\sum_{j=1}^{s}\alpha_j(w_i\circ w_j)=\sum_{j=1}^{s}\beta_{ij}\,w_j,$$

where β_{ij} are Q-linear combinations of α_j's. Now, for each $1\leqq i\leqq s-1$, α is required to satisfy

$$\sum_{j=1}^{s}\beta_{ij}\,w_j=w_i\,\alpha=\sum_{k=1}^{s-1}\kappa_{ik}\,v_k=\sum_{k=1}^{s-1}\sum_{j=1}^{s}\kappa_{ik}\,\mu_{kj}\,w_j,$$

yielding a homogeneous system of $s(s-1)$ linear equations

$$\beta_{ij}-\sum_{k=1}^{s-1}\kappa_{ik}\,\mu_{kj}=0$$

for unknowns α_j and κ_{ik} over Q. Since the number of the unknowns is $s+(s-1)^2=s(s-1)+1$, the system has a non-trivial solution. Moreover, it is easy to see that all α_j's cannot be zero and thus there exists a (non-zero) α with $H\,\alpha=H'$, as required.

Finally, in order to prove that ${}_RE=R/H$ is injective, it is sufficient to show that every morphism

$$\varphi\colon\ {}_RW\to{}_RE$$

can be extended to a morphism from ${}_RR$ to ${}_RE$. In view of the first part of our proof, we can assume that

$$w_i\,\varphi=0\quad\text{ for }\ 1\leqq i\leqq s-1\quad\text{ and }\quad w_s\,\varphi=\lambda\,w_s+H.$$

Then, taking $\rho\in Q$ such that $w_1\,\rho=\lambda\,w_1$ and making use of Lemma 2.1, it is easy to check that the right multiplication ${}_RR\xrightarrow{\rho}{}_RE$ is an extension of φ. The proof is completed.

§ 3. Rings with $W^2=0$ and $\dim_Q W\geqq 4$

The objective of this section is to prove the following

Proposition 3.1. *Let* $W^2=0$ *and* $\dim_Q W=s\geqq 4$. *Then there is an infinite number of non-isomorphic local left R-modules of the same length (equal to* $\partial\,{}_RR-2$).

Proof. In order to prove Proposition 3.1, we are going to investigate the right action of the field Q on the two-dimensional subspaces ${}_QP$ of ${}_QW$. Observe that two local R-modules of length $\partial_R R-2$, say R/P and R/P', are isomorphic if and only if there is $\alpha\in Q$ such that

$$P\,\alpha=P';$$

this follows immediately from the fact that an isomorphism between the modules can be lifted.

First assume that $s=4$. Consider the algebra $({}_QW_Q, \circ)$ and choose a basis for ${}_QW$. Obviously, there are two cases to be considered: We may assume that the basis is either of the form

$$w_1, w_2, w_3 = w_2^2, w_4 = w_2^3 \quad \text{with} \quad w_2^4 = \sum_{i=1}^{4} \pi_i w_i,$$

or of the form

$$w_1, w_2, w_3, w_4 = w_2 \circ w_3 \quad \text{with} \quad w_2^2 = \sum_{i=1}^{2} \rho_i w_i \quad \text{and} \quad w_3^2 = \sum_{i=1}^{4} \pi_i w_i.$$

Let us observe that there is an infinite subset $\{\kappa_1, \kappa_2, \ldots\}$ of non-zero elements κ_i of the field Q such that

$$\kappa_i + \kappa_j + \kappa_i \kappa_j \pi_4 \neq 0 \qquad \text{for every } i \neq j.$$

Correspondingly, there is an infinite number of distinct planes $P_i \subseteq {}_QW$ generated by the vectors

$$w_1 \quad \text{and} \quad w_2 + \kappa_i w_3, \qquad i = 1, 2, \ldots.$$

We are going to show that all R/P_i are non-isomorphic.

Assuming the contrary, take $i \neq j$ and $\alpha \in Q$ such that $P_i \alpha \subseteq P_j$; let

$$w_1 \alpha = \sum_{i=1}^{4} \alpha_i w_i \quad \text{with} \quad \alpha_i \in Q.$$

We are going to show that necessarily $\alpha = 0$. First, since $w_1 \alpha \in P_j$, we have $\alpha_4 = 0$ and

$$\sum_{i=1}^{3} \alpha_i w_i = \mu_1 w_1 + \nu_1 (w_2 + \kappa_j w_3).$$

From here, $\mu_1 = \alpha_1$, $\nu_1 = \alpha_2$ and, consequently,

$$\alpha_3 = \alpha_2 \kappa_j.$$

Furthermore,

$$(w_2 + \kappa_i w_3)\alpha = \left(\sum_{i=1}^{3} \alpha_i w_i\right) \circ (w_2 + \kappa_i w_3) = \sum_{i=1}^{4} \beta_i w_i \in P_j,$$

and thus

$$\beta_4 = \alpha_2 \kappa_i + \alpha_3 + \alpha_3 \kappa_i \pi_4 = 0.$$

Therefore,

$$\alpha_2 (\kappa_i + \kappa_j + \kappa_i \kappa_j \pi_4) = 0,$$

and hence,

$$\alpha_2 = 0.$$

Thus, also $\alpha_3=0$. Consequently,

$$(w_2+\kappa_i w_3)\alpha=\alpha_1 w_2+\alpha_1 \kappa_i w_3 \in P_j.$$

Therefore,

$$\alpha_i w_2+\alpha_1 \kappa_i w_3=\mu_2 w_1+\nu_2(w_2+\kappa_j w_3),$$

and we get

$$\mu_2=0, \quad \nu_2=\alpha_1 \quad \text{and} \quad \alpha_1(\kappa_i-\kappa_j)=0.$$

Thus, $\alpha_1=0$ as required.

Now, if $s>4$, then we can always choose a basis of ${}_QW$ which contains a subbasis of one of the following three forms:

either (i) $w_1, w_2, w_3=w_2^2, w_4=w_2^3, w_5=w_2^4$,
or (ii) $w_1, w_2, w_3, w_4=w_2 \circ w_3$ with $w_2^2=\rho_1 w_1+\rho_2 w_2$,
or (iii) $w_1, w_2, w_3, w_4=w_2 \circ w_3, w_5=w_2^2$.

In either of these three cases, it is a matter of routine to check that the above method applies and to complete the proof of Proposition 3.1.

§ 4. Rings with $W^2=0$ and $\dim {}_QW=3$

Throughout this section, we shall always assume that the ring R satisfies the conditions $W^2=0$ and $\dim {}_QW=3$. And, $\{w_1, w_2, w_3\}$ will be a fixed basis of ${}_QW$.

In view of Proposition 2.3, we can formulate

Lemma 4.1. *Let $W^2=0$ and $\dim {}_QW=3$. Then there are just 4 isomorphism classes of local left R-modules L_1, L_2, L_3 and L_4, represented by the simple module ${}_RR/W$, injective module ${}_RR/(Rw_1+Rw_2)$, by ${}_RR/Rw_1$ and by ${}_RR$, respectively.*

Lemma 4.2. *The left R-module*

$$X_5=({}_RR\oplus{}_RR)/(R(w_1,0)+R(0,w_1)+R(w_2,w_3))$$

does not possess epimorphic images of types L_3 or L_4; consequently, X_5 is an indecomposable module of length 5.

Proof. First, let $\varphi\colon X_5\to {}_RR$. Then, φ can be lifted to

$${}_RR\oplus{}_RR \xrightarrow{\binom{\alpha_1}{\alpha_2}} {}_RR$$

with $w_1\alpha_1=w_1\alpha_2=0$. Hence, α_1 and α_2 lie in W and φ cannot be surjective.

Second, let $\psi\colon X_5\to {}_RR/Rw_1$. Again, ψ can be lifted to

$${}_RR\oplus{}_RR \xrightarrow{\binom{\beta_1}{\beta_2}} {}_RR$$

with

$$w_1\beta_1 \in R w_1, \quad w_1\beta_2 \in R w_1 \quad \text{and} \quad w_2\beta_1 + w_3\beta_2 \in R w_1.$$

Now, in view of Lemma 2.1,

$$w_2\beta_1 = \lambda_1 w_2 \quad \text{and} \quad w_3\beta_2 = \lambda_2 w_2, \lambda_1, \lambda_2 \in R;$$

therefore, $w_2\beta_1 = w_3\beta_2 = 0$ and we deduce again that β_1 and β_2 belong to W. Thus ψ is not surjective.

Finally, the fact that X_5 is indecomposable follows easily. For, if X_5 were decomposable, then it would be a direct sum of two local left R-modules one of which would be of type L_3 or L_4.

Lemma 4.3. *Let N be a simple submodule of a direct sum M of modules of type L_3. Then either M/N contains a submodule of type L_2 or M/N is a direct sum of a copy of X_5 and several copies of L_3.*

Thus, in particular, if $\{x, y, z\}$ is another basis of ${}_QW$, then

$$X = ({}_RR \oplus {}_RR)/(R(x,0) + R(0,x) + R(y,z)) \cong X_5.$$

Proof. Obviously, we may assume that M is a finite direct sum. Let

$$M = (\bigoplus_n {}_RR)/D$$

with $D = \bigoplus_n R w_1$, be a representation of M. Let

$$N = R[(x_1, x_2, \ldots, x_n) + D].$$

Observe that the left ideal $L = R w_1 + \sum_{i=1}^{n} R x_i$ satisfies the relations $R w_1 \neq L \subseteq W$.

First, assume that $L \neq W$ and write, without loss of generality,

$$L = R w_1 + R x_1.$$

Then, for $2 \leqq i \leqq n$,

$$x_i = \kappa_i w_1 + \lambda_i x_1 \quad \text{with suitable } \kappa_i, \lambda_i \in R,$$

and we have

$$M/N = (\bigoplus_n {}_RR)/(D + R(x_1, x_2, \ldots, x_n)) = (\bigoplus_n {}_RR)/(D + R(x_1, \lambda_2 x_1, \ldots, \lambda_n x_1)).$$

Take $\rho_i \in R$ such that

$$x_1 \rho_i = -\lambda_i x_1 \quad \text{for } 2 \leqq i \leqq n,$$

and consider the isomorphism $\varphi\colon \bigoplus_n {}_R R \to \bigoplus_n {}_R R$ given by the following triangular matrix of right multiplication

$$\begin{pmatrix} 1 & \rho_2 & \rho_3 & \cdots & \rho_n \\ & 1 & & & 0 \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix}.$$

Clearly, in view of Lemma 2.1,

$$D_1 = [D + R(x_1, \lambda_2 x_1, \ldots, \lambda_n x_1)]\,\varphi = D + R(x_1, 0, \ldots, 0),$$

and thus

$$M/N \cong (\bigoplus_n {}_R R)/D_1$$

contains a submodule of type L_2, namely

$$R[(1,0,\ldots,0) + D] \cong {}_R R/L.$$

Thus, let $L = W$ and let

$$L = R\,w_1 + R\,x_1 + R\,x_2.$$

Then, there are elements $\varepsilon_j, \mu_j, \nu_j \in R$ such that

$$\varepsilon_j w_1 + \mu_j x_1 + \nu_j x_2 = w_{j+1} \qquad \text{for } j = 1, 2.$$

Moreover, take α_j, β_j $(j = 1, 2)$ such that

$$x_1 \alpha_j = \mu_j x_1 \quad \text{and} \quad x_2 \beta_j = \nu_j x_2,$$

and consider the isomorphism $\psi\colon \bigoplus_n {}_R R \to \bigoplus_n {}_R R$ given by the matrix

$$\begin{pmatrix} \alpha_1 & \alpha_2 & 0 & & \\ \beta_1 & \beta_2 & 0 & & \\ 0 & 0 & 1 & & 0 \\ & & & \ddots & \\ & & 0 & & 1 \end{pmatrix}.$$

Again, using Lemma 2.1, observe that

$$D' = [D + R(x_1, x_2, x_3, \ldots, x_n)]\,\psi = D + R(w_2, w_3, x_3, \ldots, x_n).$$

Now,

$$x_i = \varepsilon_i w_1 + \mu_i w_2 + \nu_i w_3 \quad \text{for } 3 \leqq i \leqq n;$$

take, for $3 \leqq i \leqq n$, α_i and β_i such that

$$w_2 \alpha_i = -\mu_i w_2 \quad \text{and} \quad w_3 \beta_i = -\nu_i w_3,$$

and consider the isomorphism η: $\bigoplus_n {}_R R \to \bigoplus_n {}_R R$ given by the triangular matrix

$$\begin{pmatrix} 1 & 0 & \alpha_3 & \dots & \alpha_n \\ & 1 & \beta_3 & \dots & \beta_n \\ & & 1 & & 0 \\ & & & \ddots & \\ & & 0 & & 1 \end{pmatrix}.$$

It is easy to verify that

$$D_2 = D' \eta = D + R(w_2, w_3, 0, \dots, 0),$$

and, consequently, that

$$M/N \cong (\bigoplus {}_R R)/D_2 \cong X_5 \oplus K,$$

where K is the direct sum of $n-2$ R-modules of type L_3.

The final statement of Lemma 4.3 follows trivially.

Lemma 4.4. *Let $x \in \operatorname{Soc} X_5$. Then there exists a submodule of type L_3 in X_5 containing x.*

Proof. Without loss of generality, assume that the basis $\{w_1, w_2, w_3\}$ of ${}_Q W$ satisfies with respect to the multiplication of the algebra $(W, \circ)$ the following conditions: w_1 is the identity, $w_3 = w_2 \circ w_2$ and $w_3 \circ w_2 = \sum_{i=1}^{3} \pi_i w_i$; notice that $\pi_1 + \pi_2 \pi_3 \neq 0$.

Refer to Lemma 4.2 for the definition of X_5 and put

$$x = (x_1, x_2) + D$$

with $D = R(w_1, 0) + R(0, w_1) + R(w_2, w_3)$, $x_1 = \kappa w_3$ and $x_2 = \mu w_2 + \nu w_3$. If $\kappa = 0$, then

$$x \in R[(0,1) + D] \cong {}_R R / R w_1.$$

Thus, assume $\kappa \neq 0$. Consider the homomorphism

$$\varphi\colon {}_R R \xrightarrow{(\alpha, \beta)} {}_R R \oplus {}_R R \xrightarrow{\varepsilon} X_5,$$

where

$$w_1 \alpha = (\kappa \pi_3 - \nu) w_1 + \kappa w_2,$$

$$w_1 \beta = (\mu - \kappa \pi_2) w_1 + \kappa w_3,$$

and ε is the natural epimorphism. Obviously,

$$w_1\varphi=[(\kappa\pi_3-\nu)w_1+\kappa w_2, (\mu-\kappa\pi_2)w_1+\kappa w_3]\varepsilon=0,$$

$$\begin{aligned} w_2\varphi &= [(\kappa\pi_3-\nu)w_2+\kappa w_3, \kappa\pi_1 w_1+\mu w_2+\kappa\pi_3 w_3]\varepsilon \\ &= [\kappa w_3+(\kappa\pi_3-\nu)w_2, (\mu w_2+\nu w_3)+\kappa\pi_1 w_1+(\kappa\pi_3-\nu)w_3]\varepsilon \\ &= (x_1, x_2)+D, \end{aligned}$$

and

$$w_3\varphi=(\ldots, \lambda_1 w_1+\kappa(\pi_1+\pi_2\pi_3)w_2+\lambda_3 w_3)\varepsilon\neq 0,$$

because $\kappa(\pi_1+\pi_2\pi_3)\neq 0$. Consequently, $R\varphi$ is of type L_3 and $x\in R\varphi$, as required.

Lemma 4.5. *Let $x\in\mathrm{Soc}(P\oplus Q)$, where $P\cong X_5$ and $Q\cong L_3$. Then there exists a submodule of type L_3 in $P\oplus Q$ containing x.*

Proof. Again, assume that the basis $\{w_1, w_2, w_3\}$ of ${}_QW$ satisfies with respect to the multiplication $\circ$ the same conditions as in the proof of Lemma 4.4.

Consider the following representation of $P\oplus Q$:

$$P\oplus Q\cong({}_RR\oplus{}_RR\oplus{}_RR)/D,$$

where $D=R(w_1, 0, 0)+R(0, w_1, 0)+R(w_2, w_3, 0)+R(0, 0, w_1)$; let

$$x=(x_1, x_2, x_3)+D.$$

In view of Lemma 4.4, we may assume that $x_3\neq 0$. Furthermore, we may obviously assume that $Rw_1+\sum_{i=1}^{3}Rx_i=W$. But then it is easy to follow the method of the proof of Lemma 4.4 and to verify that there is an automorphism φ of ${}_RR\oplus{}_RR\oplus{}_RR$ such that

$$D'=D\varphi=D+R(0, w_2, w_3).$$

Thus,

$$M'=(P\oplus Q)/Rx\cong({}_RR\oplus{}_RR\oplus{}_RR)/D'.$$

Now, consider the following mapping

$${}_RR\xrightarrow{(\alpha, \beta, \gamma)}{}_RR\oplus{}_RR\oplus{}_RR\xrightarrow{\varepsilon}M',$$

where $w_1\alpha=w_1$, $w_1\beta=\pi_3 w_1+w_2$, $w_1\gamma=-\pi_2 w_1+w_3$ and ε is the natural epimorphism. Obviously,

$$w_1\varphi=(w_1, \pi_3 w_1+w_2, -\pi_2 w_1+w_3)\varepsilon=0,$$

$$w_2\varphi=(w_2, \pi_3 w_2+w_3, \pi_1 w_1+\pi_3 w_3)\varepsilon=0$$

and

$$w_3\varphi=(w_3,\ldots,\ldots)\varepsilon\neq 0.$$

Thus, ${}_R R\varphi=R(m+Rx)$ is an injective submodule of $(P\oplus Q)/Rx$. But since $P\oplus Q$ does not contain any injective submodule, $x\in Rm$ and Rm is of type L_3, as required.

Lemma 4.6. *Let M be a direct sum of copies of X_5. Then every socle element of M is contained in a submodule of type X_5, and for that matter, in a submodule of type L_3.*

Proof. First, consider the case when $M=P\oplus Q$ is the direct sum of two copies P, Q of X_5. Let $x=(p,q)$ be an element of Soc M. By Lemma 4.4, q belongs to a submodule Q' of Q of type L_3 and thus x belongs to $P\oplus Q'$. Consequently, in accordance with Lemma 4.5, x belongs to a submodule $N\subseteq M$ of type L_3. Therefore, M/Rx contains an injective submodule N/Rx (of type L_2). It follows that

$$M/Rx=N/Rx\oplus C$$

for some complement C. We want to show that C also contains a submodule of type L_2. Let $P=Rm_1+Rm_2$, $Q=Rm_3+Rm_4$ with $\partial Rm_i=3$ for all $1\leqq i\leqq 4$. Then three of the m_i's generate together with N the entire module M, say

$$M=Rm_1+Rm_2+Rm_3+N=P+Rm_3+N,$$

and thus C is an epimorphic image of $P+Rm_3$. Since $\partial(P+Rm_3)=8$ and $\partial C=7$, there is a socle element $y\in\mathrm{Soc}(P+Rm_3)$ such that

$$C\cong(P+Rm_3)/Ry.$$

But, by Lemma 4.5, y belongs to a submodule of $P+Rm_3$ of type L_3, and therefore C contains a submodule of type L_2. Summarizing,

$$M/Rx=I_1\oplus I_2\oplus C' \quad \text{with } I_1 \text{ and } I_2 \text{ of type } L_2.$$

If we lift the homomorphism

$${}_R R\oplus{}_R R\xrightarrow{\varepsilon} I_1\oplus I_2\xrightarrow{\iota} M/Rx,$$

with the canonic epimorphism ε and imbedding ι, to a homomorphism ${}_R R\oplus{}_R R\to M$, then both copies of ${}_R R$ are mapped onto submodules of type L_3 which intersect in Rx; this follows from the fact that M has no submodules of type L_2. Now, according to Lemma 4.3, the latter two submodules of M generate a submodule of type X_5 which contains x. And, the lemma is proved in the case that M is the direct sum of two copies of X_5.

In the general case, we may assume that M is a finite direct sum; for, every element of M has only a finite number of non-zero components. Thus, let

$$x=(x_1, x_2, \ldots, x_n)\in \operatorname{Soc} M = \operatorname{Soc}(\bigoplus_n X_5).$$

Assuming, by induction, that $(x_1, x_2, \ldots, x_{n-1})\in \operatorname{Soc}(\bigoplus_{n-1} X_5)$ belongs to a submodule $P'\subseteq \bigoplus_{n-1} X_5$ of type X_5,

$$x=((x_1, x_2, \ldots, x_{n-1}), x_n)\in \operatorname{Soc}(P'\oplus X_5).$$

But then, by the first part of the proof, x belongs to a submodule of M of type X_5, as required.

The last statement of Lemma 4.6 is an immediate consequence of Lemma 4.4.

Lemma 4.7. *Let $x\in \operatorname{Soc}(P\oplus Q)$, where P is a direct sum of modules of type X_5 and Q is a direct sum of modules of type L_3. Then there exists either a submodule of type L_3 in $P\oplus Q$ containing x or there is an automorphism φ of $P\oplus Q$ which is the identity on P and which satisfies $x\varphi\in Q$.*

Proof. Since x has only a finite number of non-zero components and since an automorphism of a direct summand extends to an automorphism of the entire module, we may assume that both P and Q are finite direct sums. Write $x=p+q$ with $p\in P$ and $q\in Q$. Then, by Lemma 4.6, p belongs to a submodule P' of P of type X_5. Assuming that the lemma is true for $P'\oplus Q$, then either x belongs to a submodule of type L_3 in $P'\oplus Q\subseteq P\oplus Q$, or else there is an automorphism φ of $P'\oplus Q$ which is the identity on P' and which satisfies $x\varphi\in Q$. Obviously, we can extend φ to an automorphism of $P\oplus Q$ which is the identity on P. It follows that it is sufficient to consider the case when P is of type X_5 and Q is a finite direct sum of modules of type L_3. Let

$$P\oplus Q=(\bigoplus_n {}_RR)/D,$$

with $D=\bigoplus_n R w_1 + R(w_2, w_3, 0, \ldots, 0)$. In view of Lemmas 4.4 and 4.5, we may assume that $n\geqq 4$. Let

$$x=(x_1, x_2, \ldots, x_n)+D.$$

Without loss of generality, assume that

$$x_i\in R w_2 + R w_3 \quad \text{for } 1\leqq i\leqq n.$$

First, suppose that there are two linearly independent elements among $x_3, x_4, \ldots, x_n$; say, x_3 and x_4. Then

$$x_j=\kappa_j x_3=\lambda_j x_4 \quad \text{for } j=1,2,$$

and there are elements ρ_j, $\sigma_j \in R$ such that

$$x_3 \rho_j = -\kappa_j x_3 \quad \text{and} \quad x_4 \sigma_j = -\lambda_j x_4 \qquad (j=1,2).$$

Consider the automorphism of $\bigoplus_n {}_R R$ defined by the triangular matrix

$$\begin{pmatrix} 1 & & & & & & \\ 0 & 1 & & & & & \\ \rho_1 & \rho_2 & 1 & & & 0 & \\ \sigma_1 & \sigma_2 & 0 & 1 & & & \\ & & & & 1 & & \\ & & 0 & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

acting from the right. In view of Lemma 2.1, it maps D into D and thus induces an automorphism φ of $P \oplus Q$. It is easy to see that φ acts as the identity on P and that

$$[(x_1, x_2, \ldots, x_n) + D]\, \varphi = (0, 0, x_3, \ldots, x_n) + D \in Q.$$

Second, if all $x_3, x_4, \ldots, x_n$ are multiples of a certain x_k, then

$$x_i = \lambda_i x_k = x_k \rho_i \quad \text{with suitable } \lambda_i, \rho_i \in R \quad \text{for } 3 \leqq i \leqq n.$$

We may assume that not all ρ_i's belong to W. Consider

$$m = (0, 0, \rho_3, \rho_4, \ldots, \rho_n) + D.$$

Obviously, by Lemma 2.1,

$$x_k m = (0, 0, x_k \rho_3, x_k \rho_4, \ldots, x_k \rho_n) + D = (0, 0, x_3, x_4, \ldots, x_n) + D,$$

and thus $x \in P + Rm$. Since, again according to Lemma 2.1, $w_1 m = 0$ and since $P \oplus Q$ contains no injective submodule, necessarily $\partial R m = 3$; hence, Rm is of type L_3. Consequently, in view of Lemma 4.5, x belongs to a submodule of type L_3, as required.

Lemma 4.8. *Let M be a direct sum of submodules of types L_3, L_4 and X_5. Then every socle element of M is contained in a submodule of type L_3 or L_4.*

Proof. Write

$$M = \left(\bigoplus_I {}_R R\right)/D, \qquad \text{where } D \subseteq \bigoplus_I {}_R W \text{ with an index set } I.$$

Since, Soc $M = \operatorname{Rad} M$, we have

$$\operatorname{Soc} M = \left(\bigoplus_I {}_R W\right)/D,$$

and thus $x \in \operatorname{Soc} M$ can be written in the form

$$x = (w_i) + D \qquad \text{with } (w_i) \in \bigoplus_I {}_R W.$$

Take a fixed non-zero $w \in W$ and define the element

$$(\rho_i) \in \bigoplus_I {}_R R$$

as follows: If $w_i = 0$, then $\rho_i = 0$; if $w_i \neq 0$, then ρ_i satisfies the relation $w\rho_i = w_i$. Obviously,

$$w(\rho_i) = (w_i),$$

and thus

$$w[(\rho_i) + D] = (w_i) + D = x.$$

Since M contains no submodules of type L_2, the submodule

$$R[(\rho_i) + D] \subseteq M$$

containing x must be of type L_3 or L_4.

Proposition 4.9. *Let $W^2 = 0$ and $\partial_R W = 3$. Then there are 5 isomorphism classes of indecomposable left R-modules L_1, L_2, L_3, L_4 and X_5, defined in Lemmas 4.1 and 4.2. Moreover, every left R-module is a direct sum of these modules.*

Proof. To prove our proposition, we shall show that every left R-module M can be expressed as a direct sum of modules of types L_1, L_2, L_3, L_4 and X_5.

First, take a submodule I of M which is maximal with respect to the property of being a direct sum of modules of type L_2. Since I is injective, M is a direct sum of I and a submodule of M which contains no submodules of type L_2.

Thus, assume that M does not contain any submodule of type L_2 and let X be a submodule of M which is maximal with respect to the property of being a direct sum of modules of type X_5. Then, let Y_3 be a submodule of M which is maximal with respect to the property of being a direct sum of modules of type L_3 which intersects the submodule X trivially. Furthermore, let Y_4 be a submodule of M which is maximal with respect to the property of being a direct sum of modules of type L_4 which intersects the submodule $X \oplus Y_3$ trivially. And finally, let Z be a complement of the socle $\operatorname{Soc}(X \oplus Y_3 \oplus Y_4)$ in $\operatorname{Soc} M$. We want to show that

$$X \oplus Y_3 \oplus Y_4 \oplus Z = M.$$

Assume the contrary, i.e. that there is an element

$$m \in M \setminus (X \oplus Y_3 \oplus Y_4 \oplus Z).$$

Then Rm must be of type L_3 or L_4, because $m \notin \operatorname{Soc} M$ and M contains no submodules of type L_2.

First, consider the case that Rm is of type L_3 and that

$$\partial[(X \oplus Y_3) \cap Rm] = 1.$$

Then, the submodule $N = (X \oplus Y_3) + Rm \subseteq M$ is isomorphic to

$$(X \oplus Y_3 \oplus Q)/Rz,$$

where Q is of type L_3 and $z \in \operatorname{Soc}(X \oplus Y_3 \oplus Q)$. Hence, by Lemma 4.7, N either contains a submodule of type L_2, which is impossible, or there is an automorphism φ of $X \oplus Y_3 \oplus Q$ which is the identity on X and which satisfies $z\varphi \in Y_3 \oplus Q$. In the latter case,

$$(X \oplus Y_3 \oplus Q)/Rz \cong X \oplus [(Y_3 \oplus Q)/R(z\varphi)];$$

however, in view of Lemma 4.3, $(Y_3 \oplus Q)/R(z\varphi)$ contains either a submodule of type L_2 which is impossible, or a copy of X_5, in contradiction to maximality of X.

Thus, consider the second case when Rm is still of type L_3, but

$$\partial[(X \oplus Y_3) \cap Rm] = 2.$$

Writing

$$N' = (X \oplus Y_3) + Rm \cong (X \oplus Y_3 \oplus Q)/(Rz_1 \oplus Rz_2),$$

where Q is of type L_3 and $z_1, z_2 \in \operatorname{Soc}(X \oplus Y_3 \oplus Q)$, we deduce from Lemma 4.7 that there is an automorphism φ of $X \oplus Y_3 \oplus Q$ which is the identity on X and which satisfies $z_1\varphi \in Y_3 \oplus Q$. This follows from the obvious fact that z_1 cannot belong to a submodule of $X \oplus Y_3 \oplus Q$ of type L_3. Now,

$$N' \cong [(X \oplus Y_3 \oplus Q)/R(z_1\varphi)]/R(z_2\varphi).$$

By Lemma 4.3, $(X \oplus Y_3 \oplus Q)/R(z_1\varphi)$ contains either a submodule of type L_2 or a submodule of type X_5. By passage to the quotient module N', a submodule of type L_2 cannot avoid the kernel $R(z_2\varphi)$; but then it produces a simple direct summand of N', a contradiction. Thus, the only possibility is that $(X \oplus Y_3 \oplus Q)/R(z_1\varphi)$ contains a submodule of type X_5 which, in view of maximality of X, cannot avoid the kernel $R(z_2\varphi)$. Thus, N' contains a submodule isomorphic to a quotient module of X_5 by a simple submodule. But such a quotient module contains, according to Lemma 4.4, a submodule of type L_2. This contradiction completes the first part of the proof.

Second, consider the case when Rm is of type L_4. Since

$$(X \oplus Y_3 \oplus Y_4) \cap Rm \neq 0,$$

there is, according to Lemma 4.8, a submodule $Ra \subseteq X \oplus Y_3 \oplus Y_4$ of type L_3 or L_4 such that

$$Ra \cap Rm \neq 0.$$

Let $r_1 a = r_2 m \neq 0$ for some $r_1, r_2 \in R$. Obviously, r_1 and r_2 are non-zero elements of W and thus, since R is right uniserial, there exists $\rho \in R$ with $r_2 \rho = r_1$. Put $b = m - \rho a$. Then

$$r_2 b = r_2 m - r_2 \rho a = 0,$$

and thus $\partial R b \leqq 3$. Consequently, in view of the first part of the proof, $Rb \subseteq X \oplus Y_3 \oplus Y_4 \oplus Z$, and thus

$$Rm \subseteq Ra + Rb \subseteq X \oplus Y_3 \oplus Y_4 \oplus Z,$$

as required.

§ 5. Brauer-Thrall Conjecture

We have seen in § 1 that, under certain assumptions, R is not of finite representation type, and we want to deduce that R is not even of bounded representation type.

Roiter [9] has shown that every finite-dimensional algebra of bounded representation type is of finite representation type. In his paper, he remarked that the same conclusion holds for an arbitrary left artinian ring and that, to prove it, only the proof of a certain lemma requires slight modifications. However, in order to exclude at least the case where one of the indecomposable injectives is not finitely generated, it is clear that already the statement of that lemma has to be modified. And thus it remains open whether there are left artinian rings which are of bounded representation type, but not of finite representation type.

Here, we shall show that, for a right uniserial ring R with $W^2 = 0$ and commutative R/W, Roiter's method can be used. To this end, the following lemma is crucial.

Lemma 5.1. *Let R be left artinian and $W^2 = 0$. Given a finitely generated left R-module ${}_R M$, there is a local ring T and a bimodule structure ${}_R M_T$ such that, for every finitely generated left R-module ${}_R N$, the left T-module $\mathrm{Ext}^1_R({}_R M_T, {}_R N)$ is finitely generated and is annihilated by the radical* Rad T.

Proof. Assume that ${}_R M \neq 0$ and put

$$T = \{\tau \in R \mid w\tau \in Rw \text{ for all } w \in W\}.$$

We will show that M can be considered as a right T-module and, moreover, that M_T is finitely generated and ${}_R M_T$ is a bimodule. For, let

$${}_R M = (\bigoplus_n {}_R R)/D, \qquad \text{where } D \subseteq \bigoplus_n {}_R W.$$

Then the scalar $n \times n$ matrices

$$\begin{pmatrix} \tau & & 0 \\ & \tau & \\ 0 & & \ddots \\ & & & \tau \end{pmatrix} \quad \text{with } \tau \in T,$$

form a subring of $\operatorname{End}(\bigoplus_n {}_R R)$ isomorphic to T, and all these endomorphisms map, in view of Lemma 2.1, D into D. Now, according to our assumption, $\partial_R W$ is finite and thus R_T is finitely generated by Lemma 2.2. Consequently, also

$$M_T = (\bigoplus_n R_T)/D$$

is finitely generated.

Now, we want to prove that, for an arbitrary finitely generated R-module ${}_R N$, the left T-module $\operatorname{Ext}^1_R({}_R M_T, {}_R N)$ is finitely generated and annihilated by the radical $\operatorname{Rad} T$ of T. Denote by $\operatorname{E}{}_R N$ the injective hull of ${}_R N$ and consider the exact sequence

$$0 \to {}_R N \to \operatorname{E}{}_R N \to (\operatorname{E}{}_R N)/N \to 0,$$

which gives rise to the exact sequence of left T-modules

$$\operatorname{Hom}_R({}_R M_T, (\operatorname{E}{}_R N)/N) \to \operatorname{Ext}^1_R({}_R M_T, {}_R N) \to \operatorname{Ext}^1_R({}_R M_T, \operatorname{E}{}_R N).$$

Obviously, since $\operatorname{E}{}_R N$ is injective, the last term is zero. Consequently, the left T-module $\operatorname{Ext}^1_R({}_R M_T, {}_R N)$ is an epimorphic image of

$${}_T H = \operatorname{Hom}_R({}_R M_T, (\operatorname{E}{}_R N)/N),$$

and therefore it is sufficient to prove the assertion for ${}_T H$. Since the injective indecomposable module $\operatorname{E}{}_R Q$ is of finite length, also $(\operatorname{E}{}_R N)/{}_R N$ is of finite length; hence, $(\operatorname{E}{}_R N)/{}_R N$ is isomorphic to a finite direct sum $\bigoplus_m {}_R Q$. Also, using the above representation $M = (\bigoplus_n {}_R R)/D$, one can see easily that every homomorphism

$${}_R M \to (\operatorname{E}{}_R N)/{}_R N$$

factors through $\bigoplus_n ({}_R R/{}_R W) = \bigoplus_n {}_R Q$, which is again an R-T-bimodule decomposition. Thus, we get the T-isomorphisms

$${}_T H \cong \operatorname{Hom}_R(\bigoplus_n {}_R Q_T, \bigoplus_m {}_R Q) \cong \bigoplus_{mn} \operatorname{Hom}_R({}_R Q_T, {}_R Q),$$

where the last T-module $\operatorname{Hom}_R({}_R Q_T, {}_R Q)$ is obviously annihilated by $W = \operatorname{Rad} T$ and, according to Lemma 2.2, finitely generated.

Now, having proved that, for any two finitely generated left R-modules ${}_R M$ and ${}_R N$, always $\operatorname{Ext}^1_R({}_R M_T, {}_R N)$ can be considered as a finite-dimensional left vector space over $T/\mathrm{Rad}\,T$, only a slight modification of the proof in [10] (consisting in replacing the base field by the division ring $T/\mathrm{Rad}\,T$) is required to establish the following assertion.

If, under the conditions of Lemma 5.1, $M_1, M_2, \ldots, M_l, N$ are finitely generated left R-modules, then there exists a natural number n_0 such that, for every exact sequence of left R-modules

$$0 \to N \to X \to \bigoplus_{i=1}^{l} \bigoplus_{n_i} M_i \to 0 \text{ with } n_i \geqq 0 \text{ and } \partial_R X > n_0,$$

there is a direct decomposition of the left R-module X of the form $X = Y \oplus M_i$ for some i.

Now, using this assertion, the method of [9] yields that R is in this case of bounded representation type if and only if it is of finite representation type. Therefore, if R is left artinian, $W^2=0$ and $\partial_R W \geqq 4$, then, by Proposition 3.1, R is not of bounded representation type. But, if R is not left artinian (and $W^2=0$), then R has obviously local left R-modules of arbitrary finite length. Consequently, we get

Proposition 5.2. *If $W^2=0$ and $\partial_R W \geqq 4$, then R is not of bounded representation type.*

§ 6. Rings with $W^2 \neq 0$

In this final section, we are going to reduce the investigation of rings R with $W^2 \neq 0$ to the case when $W^2=0$. First, we need the following

Lemma 6.1. *Let $W^2 \neq 0$, $W^3=0$ and $\partial_R(W/W^2) \geqq 2$. Then $\partial_R W^2 \geqq 4$.*

Proof. Since R is right uniserial and $W^3=0$, R has only two proper right ideals, namely W and W^2, and these are the only two-sided ideals of R.

Now, for $a \in W \setminus W^2$, the left annihilator $l(a)$ is just W^2. For, since $a \in W$, obviously $W^2 \subseteq l(a)$. On the other hand, if $b \in l(a)$, then

$$bW = baR = 0,$$

because $W = aR$. Thus b belongs to the right socle of R which equals W^2.

Let $S = R/W^2$ and consider W as a left S-module. Evidently

$$\mathrm{Rad}_S W = \mathrm{Rad}_R W = W^2.$$

As we have shown, every element $a \in {}_S W \setminus \mathrm{Rad}_S W$ has zero-annihilator in S and thus,

$$Sa \cong {}_S S.$$

Therefore, if $\partial_S(W/W^2)=\partial_R(W/W^2)\geqq 4$, then

$$\partial\,\mathrm{Soc}(Sa)=\partial\,\mathrm{Soc}_S S\geqq 4$$

and hence

$$\partial_R W^2=\partial_S W^2=\partial\,\mathrm{Soc}_S W\geqq 4.$$

As a consequence, we may assume that

$$\partial_S(W/W^2)=2 \text{ or } 3.$$

But then, we can apply the decomposition theory to ${}_SW$. Since for every $a\in{}_SW\setminus\mathrm{Rad}\,{}_SW$, $Sa\cong{}_SS$, it follows from Proposition 3 of [5] or from Proposition 4.9 that ${}_SW$ is a free S-module. Moreover, it is obviously a free S-module on 2 or 3 generators, and thus $\partial_R W^2=\partial\,\mathrm{Soc}_S W\geqq 4$, as required.

Now, our intention is to construct, for every local ring R with $W^2\neq 0$, another local ring with radical-square zero and compare the indecomposable modules of both rings. This will be done in the following lemma which modifies arguments of Auslander [1]. Recall that, given a bimodule ${}_AM_A$ over a ring A, the split extension of ${}_AM_A$ by A is a ring whose additive structure is that of the direct sum $A\oplus M$ and whose multiplication is given by

$$(a,m)(a',m')=(aa',am'+ma').$$

We shall identify M with the two-sided ideal $0\oplus M$. Notice that, if A is a local ring, the split extension is a local ring as well.

Lemma 6.2. *Let A be a local ring with radical J and $S=\mathrm{Soc}\,{}_AA\cap\mathrm{Soc}\,A_A$. Let B be the split extension of ${}_{A/J}S_{A/J}$ by A/J. Then, if A is of bounded representation type, also B is of bounded representation type.*

Proof. We shall define a function F from the set of all isomorphism classes of finitely generated left B-modules into the set of all isomorphism classes of finitely generated left A-modules as follows. Given a finitely generated left B-module ${}_BM$, we consider its representation

$${}_BM=(\bigoplus_n {}_BB)/D \quad \text{with } D\subseteq\bigoplus_n {}_BS.$$

If $(\bigoplus_m {}_BB)/D'$ is another such representation of ${}_BM$, then obviously $m=n$ equals the minimal number of generators of ${}_BM$, and there is an automorphism $\varphi\in\mathrm{End}(\bigoplus_n {}_BB)$ with $D\varphi=D'$. We can write φ as an $n\times n$ matrix $\varphi=(b_{ij})$, where $b_{ij}\in B$. If $\varphi'=(b'_{ij})\in\mathrm{End}(\bigoplus_n {}_BB)$ with $b_{ij}-b'_{ij}\in S$ for all $1\leqq i$, $j\leqq n$, then, in view of $DS=0$, also $D\varphi'=D'$. Thus, we may take

$$b_{ij}=(a_{ij}+J,0) \qquad \text{for some } a_{ij}\in A.$$

Now, consider $(a_{ij}) \in \operatorname{End}(\bigoplus_n {}_AA)$. Since $D(a_{ij}) = D'$, (a_{ij}) induces an endomorphism of $(\bigoplus_n {}_AA)/D$. It is easy to verify that (a_{ij}) induces, in fact, an isomorphism between the left A-modules $(\bigoplus_n {}_AA)/D$ and $(\bigoplus_n {}_AA)/D'$. Denote the isomorphism class (or its representative) of these modules by $F(M)$. The above argument can be reversed which shows that F is an injective mapping.

Now, assume that ${}_BM = (\bigoplus_n {}_BB)/D$ is indecomposable. Then also $F(M) = (\bigoplus_n {}_AA)/D$ is indecomposable. For, assume that $F(M) = {}_AX \oplus {}_AY$. We can write

$${}_AX = (\bigoplus_{n_1} {}_AA)/D_1 \quad \text{and} \quad {}_AY = (\bigoplus_{n_2} {}_AA)/D_2 \quad \text{with } n_1 + n_2 = n.$$

Thus, there is an automorphism $\varphi = (a_{ij}) \in \operatorname{End}(\bigoplus_n {}_AA)$ mapping D onto $D_1 \oplus D_2$. Since $D \subseteq \bigoplus_n {}_AS$, also

$$D_1 \oplus D_2 = D\varphi \subseteq (\bigoplus_n {}_AS)\varphi = \bigoplus_n {}_AS,$$

and therefore $F(M)$ is also the image of

$$(\bigoplus_{n_1} {}_BB)/D_1 \oplus (\bigoplus_{n_2} {}_BB)/D_2$$

under F. Hence, the latter module is isomorphic to ${}_BM$. It follows that either $n_1 = 0$ or $n_2 = 0$. Consequently, the image $F(M)$ of an indecomposable B-module ${}_BM$ is an indecomposable A-module.

Finally, it is evident that $\partial_B M \leqq \partial_A F(M)$, and thus, if A is of bounded representation type, so is B.

Proposition 6.3. *If $W^2 \neq 0$ and R is of bounded representation type, then R is left uniserial.*

Proof. Obviously, we may assume that $W^3 = 0$. If R is not left uniserial, then $\partial_R W^2 \geqq 4$, by Lemma 6.1. Denote by B the split extension of ${}_QW^2_Q$ by $Q = R/W$. Note that $W^2 = \operatorname{Soc} {}_RR = \operatorname{Soc} R_R$. Now, B is a local ring with $(\operatorname{Rad} B)^2 = 0$, $B/\operatorname{Rad} B \cong Q$ and $\partial_B \operatorname{Rad} B \geqq 4$. Thus, in view of Proposition 5.2 and Lemma 6.2, R is not of bounded representation type, in contradiction to our assumption.

Remark 6.4. In this final remark, we would like to supplement the description of the indecomposable left R-modules given in Lemmas 4.1 and 4.2 by a similar description of the indecomposable right R-modules. It is easy to verify that, if $W^2 = 0$ and $\partial_R W = 3$, then the simple module

$C_1 = R_R/W$, the projective module $C_2 = R_R$, the module

$$C_3 = (R_R \oplus R_R)/(w_1, w_2)R,$$

the injective module

$$C_4 = (R_R \oplus R_R \oplus R_R)/((w_1, w_2, 0)R + (w_1, 0, w_3)R)$$

and the module

$$Y_5 = (R_R \oplus R_R \oplus R_R)/(w_1, w_2, w_3)R$$

are non-isomorphic indecomposable right R-modules.

Proposition 1.1 yields then that these are the only indecomposable right R-modules. The latter statement can be also derived by means of Tachikawa's duality theory [10], using the following statement the proof of which is rather technical: If $W^2 = 0$ and $\partial_R W = 3$, then the centralizer $\mathscr{C} = \mathrm{End}(C_4)$ is a local ring with radical $\mathscr{W}$ such that $\mathscr{W}^2 = 0$, $\mathscr{Q} = \mathscr{C}/\mathscr{W}$ is commutative, $\partial_{\mathscr{C}}\mathscr{W} = 3$ and $\mathscr{W}_{\mathscr{C}}$ is simple.

References

1. Auslander, M.: Representation dimension of artin algebras. Queen Mary College Math. Notes.
2. Auslander, M., Bridger, M.: Stable module theory. Mem. Amer. Math. Soc. **94**, Providence 1969.
3. Bass, H.: Finitistic dimension and a homological generalization of semi-primary rings. Trans. Amer. Math. Soc. **95**, 466–488 (1960).
4. Curtis, C.W., Reiner, I.: Representation theory of finite groups and associative algebras. New York: Interscience Publ. 1962.
5. Dlab, V., Ringel, C.M.: A class of balanced non-uniserial rings. Math. Ann. **195**, 279–291 (1972).
6. Dlab, V., Ringel, C.M.: Balanced rings. Lecture Notes in Math. **246**, 73–143. Berlin-Göttingen-Heidelberg: Springer 1972.
7. Eisenbud, D., Griffith, P.: The structure of serial rings. Pacific J. Math. **36**, 109–121 (1971).
8. Gabriel, P.: Indecomposable representations II. Manuscript.
9. Roiter, A.V.: Unboundedness of the dimensions of indecomposable representations of algebras having infinitely many indecomposable representations. Izvestija Akad. Nauk SSSR, Ser. Math. **32**, 1275–1282 (1968).
10. Tachikawa, H.: Duality theorem of character modules for rings with minimum condition. Math. Z. **68**, 479–487 (1958).
11. Tachikawa, H., Ringel, C.M.: QF-3 rings. To appear in J. Reine Angew. Math.

V. Dlab
Department of Mathematics
Carleton University
Ottawa, Canada

C.M. Ringel
Mathematisches Institut
Universität Tübingen
Federal Republic of Germany

(Received August 11, 1972)

Math. Z. 129, 231 – 234 (1972)
© by Springer-Verlag 1972

Projective and Free Modules

István Beck

1. Introduction

Kaplansky proved in [3, Theorem 2] that over a local[1] ring any projective module is free. Another large class of projective modules were shown to be free a few years later. Bass [2, Corollary 3.2] used another of Kaplansky's theorem to show that if R/N is a left Noetherian ring, any *uniformly big* projective R-module is free.

The definition of uniformly big is given at the end of this paper, and N denotes the Jacobson radical of R.

Our main result (Theorem 6) asserts that if P is a projective R-module for which P/NP is a free R/N-module, then P is a free R-module. As a corollary we get Kaplansky's theorem that over a local ring projective modules are free. Another corollary of Theorem 6 asserts that if uniformly big projective R/N-modules are R/N-free, then uniformly big projective R-modules are R-free. Hence, when proving Bass' theorem [2, Corollary 3.2] one may as well assume that R is left Noetherian.

We close the paper with an application of Theorem 6 to the case when R is a commutative semi-local ring.

Our proof of Theorem 6 uses Kaplansky's theorem [3, Theorem 1] that any projective module is a direct sum of countably generated modules. An elegant application of Eilenberg's Lemma given by Bass in [2] will also be useful:

Let P be a projective module which is a quotient of a nonfinitely generated free module F. If P has a direct summand isomorphic to F, then $P \cong F$.

Proof. $F \cong P \oplus Q$ and $P \cong F \oplus U$. We get

$$F \cong F \oplus F \oplus \cdots \cong P \oplus Q \oplus P \oplus Q \oplus \cdots \cong P \oplus F \oplus F \oplus \cdots \cong P \oplus F.$$

Also $P \oplus F \cong F \oplus F \oplus U \cong F \oplus U \cong P$. Hence $P \cong F$.

Most of the notation is standard. R denotes an associative ring with a unit element, and N denotes the Jacobson radical of R. Modules are unitary, and module = left module.

Let F be a free module with a basis $\{e_i\}_{i \in I}$. Let $z = \sum_{i \in I} r_i e_i$, $r_i \in R$ be an element in F. The *support* of z, $\operatorname{supp} z = \{i \mid r_i \neq 0\}$.

[1] No chain condition.

Finally, I am grateful to Peter J. Trosborg for several helpful discussions and for calling to my attention Bass' paper on projective modules.

2. Projective and Free Modules

Nakayama's Lemma asserts that if F is a finitely generated module with a submodule F' such that $F=F'+NF$ then $F'=F$. We prove that if we remove the finiteness condition on F and instead require F to be free, then F' has a direct summand isomorphic to F.

Theorem 1. *Let F be a free R-module and let F' be a submodule of F such that $F'+NF=F$. Then F has a direct summand F'', $F''\subseteq F'$ and $F''\cong F$.*

Proof. If F is finitely generated it follows from Nakayama's Lemma that $F'=F$ and we are done. So let $\{e_i\}_{i\in I}$ be a basis for F where I is an infinite set. Consider the set $\mathfrak{F}$ of all pairs (J,φ) where $J\subseteq I$ and φ is a function from J to F' such that $\operatorname{supp}(\varphi(j)-e_j)\cap J=\varnothing$ for all $j\in J$.

We note that $\mathfrak{F}$ is non-void. For let $i\in I$ and write

$$e_i=x+r_1\,e_{i_1}+\cdots+r_k\,e_{i_k}$$

where $x\in F'$ and $r_t\in N$ $(1\leqq t\leqq k)$. We get that $x=u_i e_i+z$ where u_i is a unit and $i\notin\operatorname{supp} z$. Let $J_0=\{i\}$ and define $\varphi_0\colon J_0\to F'$ by $\varphi_0(i)=u_i^{-1}x$. Then $(J_0,\varphi_0)\in\mathfrak{F}$.

We introduce a partial order $\leqq$ on $\mathfrak{F}$ by defining $(J,\varphi)\leqq(J',\varphi')$ if $J\subseteq J'$ and $\varphi'|J=\varphi$. It is easily proved, using Zorn's Lemma, that $\mathfrak{F}$ has a maximal member, say (J,φ). *We claim that* $I=\bigcup_{j\in J}\operatorname{supp}\varphi(j)$.

Let $F''=\sum_{j\in J}R\varphi(j)$. Since $\operatorname{supp}(\varphi(j)-e_j)\cap J=\varnothing$ for all $j\in J$ we see that F'' is a free module with $\{\varphi(j)\}_{j\in J}$ as a basis.

Define $\mu\colon F\to F$ by $\mu(e_k)=0$ when $k\notin J$ and $\mu(e_k)=\varphi(k)$ if $k\in J$. It follows that $\operatorname{im}\mu=F''$ and μ is a projection, i.e. $\mu^2=\mu$. Hence F'' is a direct summand of F with $\ker\mu$ as a complementary summand.

We have that $F=F'+NF=F'+NF''+N\ker\mu$. Since $F''\subseteq F'$ we get that $F=F'+N\ker\mu$. Let $i\in I$ and $i\notin\bigcup_{j\in J}\operatorname{supp}\varphi(j)$ and write $e_i=x+z$ where $x\in F'$ and $z\in N\ker\mu$. We get that $x=v_i e_i+z'$ where v_i is a unit, $i\notin\operatorname{supp} z'$ and $\operatorname{supp} z'\cap J=\varnothing$. We observe that $i\notin J$ since $J\subseteq\bigcup_{j\in J}\operatorname{supp}\varphi(j)$. Let $J'=J\cup\{i\}$ and define $\varphi'\colon J'\to F'$ by $\varphi'|J=\varphi$ and $\varphi'(i)=v_i^{-1}x$. It is easily seen that $(J',\varphi')\in\mathfrak{F}$. Further, $(J,\varphi)\leqq(J',\varphi')$ and $J\neq J'$. This contradicts the maximality of (J,φ) and we can conclude that $I=\bigcup_{j\in J}\operatorname{supp}\varphi(j)$. The set I is infinite so I and J are equipollent, hence $F''\cong F$. This completes the proof of Theorem 1.

Corollary 2. *Let F be a free R-module and let F' be a submodule of F such that $F'+NF=F$. Then F' has a surjection onto F.*

Let F be a free module with a basis $\{e_i\}$ and define $\varphi\colon F \to F$ by $\varphi(e_i) = e_i + z_i$ where $z_i \in NF$. It is a simple application of Nakayama's Lemma to show that φ is an injection. This idea is used of Bass when proving that $P \neq NP$ for a nonzero projective module P [1, Proposition 2.7]. We extend his result to projective modules.

Theorem 3. *Let P be a projective R-module and let $\varphi \in \mathrm{Hom}_R(P, P)$ be such that $\varphi(x) - x$ is contained in NP for all $x \in P$. Then φ is an injection.*

Proof. Bass has given a proof of Theorem 3 when P is a free module [1, Proposition 2.7]. Let $P \oplus Q = F$ where F is a free module and define $\bar{\varphi}\colon F \to F$ by $\bar{\varphi}(p+q) = \varphi(p) + q$ for all $p \in P$ and $q \in Q$. Then $\bar{\varphi}(x) - x \in NF$ for all $x \in F$ so $\bar{\varphi}$ is injective. Hence φ is an injection.

Corollary 4 (Bass). *If P is a nonzero projective R-module then $P \neq NP$.*

Let P and P' be two finitely generated projective R-modules. The uniqueness of projective covers implies that $P \cong P'$ if $P/NP \cong P'/NP'$. We use Theorem 3 to drop the assumption that both P and P' are finitely generated.

Theorem 5. *Let P and P' be two projective R-modules and assume that $P/NP \cong P'/NP'$. If P (or P') is finitely generated then $P \cong P'$.*

Proof. Let $\pi\colon P \to P/NP$ and $\pi'\colon P' \to P'/NP'$ be the canonical projections and let $\varphi\colon P/NP \to P'/NP'$ be an isomorphism with an inverse ψ. Since P and P' are projective we can find maps $\bar{\varphi}\colon P \to P'$ and $\bar{\psi}\colon P' \to P$ such that $\varphi\pi = \pi'\bar{\varphi}$ and $\pi\bar{\psi} = \psi\pi'$.

Since P is finitely generated and $\operatorname{im}\bar{\psi} + NP = P$ it follows that $\operatorname{im}\bar{\psi} = P$. We also get $\pi'\bar{\varphi}\bar{\psi} = \varphi\pi\bar{\psi} = \varphi\psi\pi' = \pi'$. This shows that $(\bar{\varphi}\bar{\psi})(x) - x \in \ker\pi' = NP'$ for all $x \in P'$. It follows from Theorem 3 that $\bar{\varphi}\bar{\psi}$ is injective, and therefore $\bar{\psi}$ is a bijection. Theorem 5 is proved.

We prove now the main theorem in this paper.

Theorem 6. *Let P be a projective R-module. If P/NP is a free R/N-module then P is a free R-module.*

Proof. If P/NP is a free R/N-module we have an isomorphism $P/NP \cong F/NF$ for some free R-module F. If P or F is finitely generated we get that $P \cong F$ (Theorem 5). We may therefore assume that both P and F are nonfinitely generated modules.

Since P is projective we get a map $\varphi\colon P \to F$ such that $\operatorname{im}\varphi + NF = F$. From Corollary 2 it follows that $\operatorname{im}\varphi$ has a surjection onto F, hence $P \cong F \oplus U$ for some module U. We wish to prove that P is a quotient of F and then use Eilenberg's Lemma. We distinguish two cases.

Case 1. P is countably generated but not finitely generated. We have assumed that F is not finitely generated and P is therefore a quotient of F.

Case 2. P is not countably generated. By Kaplansky's theorem $P=\bigoplus_{i\in I} P_i$ where each P_i is a countably generated nonzero projective module. Let $\aleph=\operatorname{card} I$. Then P can be generated by $\aleph$ elements, and P/NP also requires $\aleph$ generators since $P_i/NP_i \neq (0)$ for all $i\in I$ (Corollary 4). Hence F has a basis of cardinality $\aleph$ so P is a quotient of F.

Corollary 7 (Kaplansky). *Any projective module over a local ring is free.*

We recall a definition from Bass [2]. An R-module P is *uniformly* $\aleph$*-big*[2] if P can be generated by $\aleph$ elements and P/AP requires $\aleph$ generators for all two-sided ideals $A(\neq R)$.

Corollary 8. *If all uniformly $\aleph$-big projective R/N-modules are R/N-free, then all uniformly $\aleph$-big projective R-modules are R-free.*

Let us now assume that R is a commutative ring. Let P be a projective R-module and let $\mathfrak{m}\in\operatorname{spec} R$. Then $P_{\mathfrak{m}}$ is a free $R_{\mathfrak{m}}$-module of rank $r_{\mathfrak{m}}(P)$.

Theorem 9. *Let R be a commutative semi-local ring with maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_k$. A projective R-module P is free if and only if*

$$r_{\mathfrak{m}_1}(P)=r_{\mathfrak{m}_2}(P)=\cdots=r_{\mathfrak{m}_k}(P).$$

Proof. Let $\mathfrak{m}$ be a maximal ideal in R. Since $R/\mathfrak{m}\cong R_{\mathfrak{m}}/\mathfrak{m}R_{\mathfrak{m}}$ it is easily seen that $r_{\mathfrak{m}}(P)=[P/\mathfrak{m}P:R/\mathfrak{m}]$. Let P be a projective R-module and assume that $r_{\mathfrak{m}_1}(P)=\cdots=r_{\mathfrak{m}_k}(P)$. There exists then a free R-module F such that $P/\mathfrak{m}_i P\cong F/\mathfrak{m}_i F$ $(1\leqq i\leqq k)$. We get:

$$\begin{aligned} P/NP &\cong R/N\otimes P\cong(R/\mathfrak{m}_1\oplus\cdots\oplus R/\mathfrak{m}_k)\otimes P\\ &\cong(R/\mathfrak{m}_1\otimes P)\oplus\cdots\oplus(R/\mathfrak{m}_k\otimes P)\cong P/\mathfrak{m}_1 P\oplus\cdots\oplus P/\mathfrak{m}_k P\\ &\cong F/\mathfrak{m}_1 F\oplus\cdots\oplus F/\mathfrak{m}_k F\cong(R/\mathfrak{m}_1\otimes F)+\cdots+(R/\mathfrak{m}_k\otimes F)\\ &\cong(R/\mathfrak{m}_1\oplus\cdots\oplus R/\mathfrak{m}_k)\otimes F\cong R/N\otimes F\cong F/NF. \end{aligned}$$

Hence P/NP is a free R/N-module and it follows from Theorem 6 that P is a free R-module.

References

1. Bass, H.: Finitistic dimension and a homological generalization of semi-primary rings. Trans. Amer. Math. Soc. **95**, 466–488 (1960).
2. Bass, H.: Big projective modules are free. Illinois J. Math. **7**, 24–31 (1963).
3. Kaplansky, I.: Projective modules. Ann. of Math. **68**(2), 372–377 (1958).

[2] $\aleph$ is an infinite cardinal.

Dr. István Beck
Universitetet i Bergen
Matematisk institutt Avd. A
5014 Bergen, Norway

(*Received March 3, 1972*)

Math. Z. 129, 235–257 (1972)
© by Springer-Verlag 1972

Dualitäten mit zwei Geraden aus absoluten Punkten in projektiven Ebenen

Armin Herzer

Ist $\mathfrak{E}$ eine projektive Ebene, und sind g_1 und g_2 verschiedene Gerade von $\mathfrak{E}$, so heiße die Dualität δ eine (g_1, g_2)-Dualität, falls alle Punkte von g_1 und g_2 absolute Punkte von δ sind. Solche (g_1, g_2)-Dualitäten, ihr Zusammenhang mit bestimmten Schließungssätzen sowie ihr Einfluß auf die Struktur der Ebenen, in denen sie existieren, und deren Ternärkörper sollen hier untersucht werden.

Wie in Baer [1], wo die $p-L$-Dualitäten definiert und betrachtet werden, zeigt sich auch hier bei verwandter Fragestellung das ähnliche Ergebnis, daß man nämlich nicht allzuviele (g_1, g_2)-Dualitäten zuzulassen braucht, um als Ternärkörper bereits einen kommutativen Körper zu erhalten[1].

Andererseits ist die Existenz einer (g_1, g_2)-Dualität in direkt einzusehender Weise gekoppelt mit der Gültigkeit des Satzes von Pappos mit gewissen Festelementen. Der Satz von Pappos mit festen drei oder zwei Trägergeraden wurde von Pickert in [6] untersucht[2] mit dem Ergebnis, daß letzterer schon den allgemeinen Satz von Pappos nach sich zieht. Hier treten an deren Stelle als Festelemente unter den 9 Punkten und 9 Geraden der Papposkonfiguration ein inzidentes Punkt-Geradenpaar (P, h) sowie die beiden übrigen Geraden durch P und die beiden übrigen Punkte auf h. Läßt man den letzten dieser Punkte stattdessen variabel, so gilt bereits der Satz von Pappos.

1. Bezeichnungen und grundlegende Definitionen

Mit $\mathfrak{E}$ wird die zu betrachtende projektive Ebene bezeichnet. Sind g und h zwei verschiedene Gerade von $\mathfrak{E}$, so bezeichnet $g \cap h$ den Schnittpunkt von g mit h; sind A und B zwei verschiedene Punkte von $\mathfrak{E}$, so bezeichnet $A+B$ ihre Verbindungsgerade. Die Einführung der Koordinaten von $\mathfrak{E}$ bezüglich des nicht-ausgearteten Punktequadrupels O, U, V, E und die Definition des zugehörigen Ternärkörpers $\mathfrak{K}$ mit ternärer Verknüpfung $\mathbf{T}$ geschieht nach Pickert [7], S. 34ff.

---

[1] Untersuchungen ähnlicher Art über eine Verallgemeinerung der $p-L$-Dualität wurden von Jónsson in [5] unternommen.

[2] Von Burn in [2] mit anderen Beweismethoden und in erweiterter Form behandelt.

Mittels $\mathfrak{K}$ wird die Ebene $\mathfrak{E}$ folgendermaßen beschrieben: Die Punkte von $\mathfrak{E}$ außer V haben entweder die Form (x, y) oder die Form (z) für alle $x, y, z \in \mathfrak{K}$. Es ist $O=(0,0)$, $E=(1,1)$, $U=(0)$, weiter soll noch sein $W=(1)$. Die Geraden von $\mathfrak{E}$ außer $U+V$ haben entweder die Form $[m, k]$ oder die Form $[z]$ für alle $m, k, z \in \mathfrak{K}$. Es ist $(x, y)\,\mathrm{I}\,[m, k]$ genau dann, wenn $y=\mathbf{T}(m, x, k)$ gilt, und $(z)\,\mathrm{I}\,[m, k]$ g.d.w. $m=z$ gilt, und $(x, y)\,\mathrm{I}\,[z]$ g.d.w. $x=z$ gilt. Stets gilt $V\,\mathrm{I}\,[z]$ und $(z_1)\,\not{\mathrm{I}}\,[z_2]$ und $(x, y)\,\not{\mathrm{I}}\,(U+V)$ und $(z)\,\mathrm{I}\,(U+V)$. Insbesondere ist $O+V=[0]$ und $E+V=[1]$.

Ist $\mathfrak{E}$ bezüglich O, U, V, E der Ternärkörper $\mathfrak{K}$ zugeordnet und also $\mathfrak{E}$ durch $\mathfrak{K}$ auf die obige Weise beschrieben, so heißt $\mathfrak{E}$ auch projektive Ebene über $\mathfrak{K}$.

Ist g eine Gerade von $\mathfrak{E}$, so ist die durch g geschlitzte Ebene diejenige affine Ebene, die aus $\mathfrak{E}$ durch Entfernen der Geraden g und aller mit g inzidenten Punkte entsteht.

Ist δ eine Dualität von $\mathfrak{E}$, so heißt der Punkt P von $\mathfrak{E}$ *absoluter Punkt von* δ, falls $P\,\mathrm{I}\,P^{\delta}$ ist; andererseits heißt die Gerade g von $\mathfrak{E}$ *absolute Gerade von* δ, falls $g^{\delta}\,\mathrm{I}\,g$ ist.

Es bezeichnet $\Gamma(\mathfrak{E})$ die volle Kollineationsgruppe von $\mathfrak{E}$ und $\Gamma(Q, g)$ die Gruppe aller (Q, g)-Kollineationen von $\mathfrak{E}$.

Alle weiteren Bezeichnungen, soweit nicht nachstehend definiert, schließen sich an Pickert [7] an.

Definition 1. P, A_1, A_2 seien Punkte von $\mathfrak{E}$, und h, g_1, g_2 seien Geraden von $\mathfrak{E}$. Dann heißt $\mathfrak{D}=(P, h, A_1, A_2, g_1, g_2)$ ein *Gestell*, falls folgendes gilt: h, g_1, g_2 sind paarweise verschieden, und P, A_1, A_2 sind paarweise verschieden. Weiter ist $P=g_1 \cap g_2$ und $h=A_1+A_2$ und $P\,\mathrm{I}\,h$. – Zur Kennzeichnung eines Gestells $\mathfrak{D}$ schreiben wir auch kürzer $\mathfrak{D}=(P, h, A_i, g_i)$. Dabei wird stillschweigend vorausgesetzt, daß der Index i die Werte 1 und 2 annimmt.

Definition 2. g_1 und g_2 seien zwei verschiedene Gerade von $\mathfrak{E}$, und δ sei eine Dualität von $\mathfrak{E}$. Dann heißt δ eine *(g_1, g_2)-Dualität*, falls alle Punkte von g_1 und g_2 absolute Punkte von δ sind. Weiter heißt δ eine *(A_1, A_2, g_1, g_2)-Dualität*, falls δ eine (g_1, g_2)-Dualität mit $A_i^{\delta}=g_i$ ist für $i=1,2$. Ist (P, h) ein inzidentes Punkt-Geradenpaar, so heißt δ eine *(P, h)-Dualität*[3], falls es Punkte A_1, A_2 und Geraden g_1, g_2 gibt, so daß (P, h, A_i, g_i) ein Gestell und δ eine (A_1, A_2, g_1, g_2)-Dualität ist. In diesem Falle heißt δ auch *(P, h)-Dualität bezüglich* A_1, A_2, g_1, g_2. Mit der in

[3] Wegen $P\mathrm{I}h$ ist keine Verwechslung mit der Baerschen $p-L$-Dualität zu befürchten; denn dort ist stets $p \not{\mathrm{I}} L$ vorausgesetzt (vgl. [1]). Ebenso ist die (P, h)-Dualität von der (C, c)-Korrelation in [5] (einer Verallgemeinerung der $p-L$-Dualität, bei der auch $C\mathrm{I}c$ zugelassen ist) wohlunterschieden: Ist σ eine (C, c)-Korrelation und ist δ eine (P, h)-Dualität, dann ist σ^2 eine (C, c)-Kollineation, während δ^2 keine (P, h)-Kollineation (überhaupt keine zentrale Kollineation) ist, vgl. unten Satz 1.

Definition 1 getroffenen Vereinbarung über den Index i, werden dementsprechend die Abkürzungen g_i-*Dualität*, (A_i, g_i)-*Dualität* und (P, h)-*Dualität bezüglich* A_i, g_i eingeführt.

Da der Index i nur für 1 oder 2 steht, ist er in Ausdrücken wie g_{i+1} stets modulo 2 zu verstehen.

Für die folgenden Definitionen wird ein Gestell $\mathfrak{D}=(P, h, A_i, g_i)$ benötigt. Mit dessen Hilfe werden $\mathfrak{D}$-Sätze als Schließungssätze mit den Festelementen P, h, A_1, A_2, g_1, g_2 definiert. Ein solcher $\mathfrak{D}$-Satz kann dann auch als (A_i, g_i)-Satz und schließlich auch als (P, h)-Satz bezüglich A_i, g_i bezeichnet werden. Ein (P, h)-Satz bezüglich A_1, g_1, g_2 ist der entsprechende $\mathfrak{D}$-Satz für alle Gestelle $\mathfrak{D}=(P, h, A_1, X, g_1, g_2)$ mit festen P, h, A_1, g_1, g_2. Entsprechend sind ähnliche Bildungen zu verstehen. Insbesondere fällt der Begriff eines (P, h)-Satzes mit der üblichen Bezeichnungsweise zusammen: Ein (P, h)-Satz (ohne nähere Angaben) besagt den entsprechenden $\mathfrak{D}$-Satz für alle Gestelle $\mathfrak{D}=(P, h, A_i, g_i)$ mit diesen vorgegebenen Elementen P und h.

Definition 3. *Der* $\mathfrak{D}$-*Satz von Pappos* ist die folgende Spezialisierung des Satzes von Pappos vom Rang 3 (vgl. Fig. 1): Für jedes inzidente Punkt-Geradenpaar (Q, g) mit $Q \not\mathrm{I}\, h$ und $P \not\mathrm{I}\, g$ folgt aus $g \cap g_i = R_i$, $g_i \cap (A_i + Q) = S_i$ auch $(R_1 + A_2) \cap (R_2 + A_1)\,\mathrm{I}\,(S_1 + S_2)$.

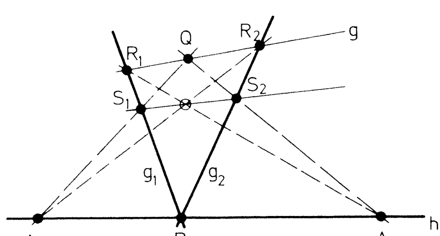


Fig. 1

Definition 4. *Der* $\mathfrak{D}$-*Satz von Desargues* ist die folgende Spezialisierung des Satzes von Desargues vom Rang 3 (vgl. Fig. 2): Ist g_0 eine Gerade mit $P\,\mathrm{I}\,g_0$ und $g_1 \neq g_0 \neq g_2$, sind weiter C_j, D_j für $j=0, 1, 2$ jeweils Punkte mit $C_j \not\mathrm{I}\, h$ und $D_j \not\mathrm{I}\, h$ und $C_j\,\mathrm{I}\,g_j$ und $D_j\,\mathrm{I}\,g_j$, und gilt ferner $A_i\,\mathrm{I}\,(C_0 + C_i)$ und $A_i\,\mathrm{I}\,(D_0 + D_i)$, dann gilt auch $[(C_1 + C_2) \cap h]\,\mathrm{I}\,(D_1 + D_2)$.

Definition 5. a) *Der kleine axiale* $\mathfrak{D}$-*Satz von Pappos* ist der kleine axiale Satz von Pappos mit Achse h, Diagonalpunkten P, A_1, A_2 und festem Geradenpaar g_1, g_2, also folgender Schließungssatz vom Rang 2: Ist C ein Punkt mit $C \not\mathrm{I}\, h$, und $R_i = (C + A_i) \cap g_i$, $S_i = (R_{i+1} + A_i) \cap g_i$, so gilt

$$[(S_1 + A_2) \cap (S_2 + A_1)]\,\mathrm{I}\,(C + P).$$

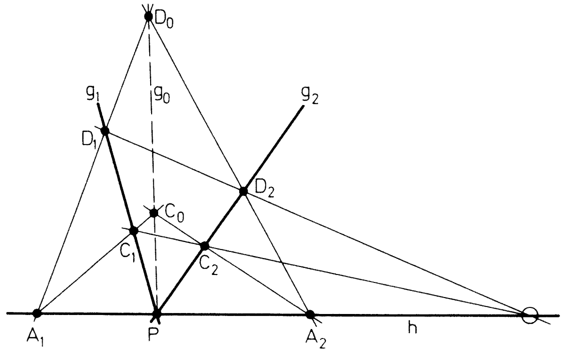


Fig. 2

b) *Der kleine zentrale* $\mathfrak{D}$*-Satz von Pappos* ist der kleine zentrale Satz von Pappos mit Zentrum P, Diagonalen h, g_1, g_2 und festem Punktepaar A_1, A_2, also folgender Schließungssatz vom Rang 2: Ist c eine Gerade mit $P \mathrm{I}\!\!\!/\, c$ und $r_i = (c \cap g_i) + A_{i+1}$, $s_i = (r_i \cap g_{i+1}) + A_i$, so gilt $(c \cap h) \mathrm{I} [(s_1 \cap g_1) + (s_2 \cap g_2)]$.

2. Die g_i-Dualität und mit ihr zusammenhängende Schließungssätze

Satz 1. *Seien g_1 und g_2 verschiedene Gerade von $\mathfrak{E}$. Sei δ eine g_i-Dualität von $\mathfrak{E}$ und sei $A_i^\delta = g_i$. Ferner sei $P = g_1 \cap g_2$ und $h = A_1 + A_2$. Dann gilt:*

a) *Alle Gerade durch A_i sind absolut. Es ist $A_i \mathrm{I}\!\!\!/\, g_i$.*

b) *Für Q mit $Q \mathrm{I}\!\!\!/\, h$ ist $Q^\delta = [(Q + A_1) \cap g_1] + [(Q + A_2) \cap g_2]$.*

c) *Es ist $g_i^\delta = A_{i+1}$; insbesondere vertauscht δ also P mit h.*

d) *Für g mit $P \mathrm{I}\!\!\!/\, g$ ist $g^\delta = [(g \cap g_1) + A_2] \cap [(g \cap g_2) + A_1]$.*

e) *δ ist durch A_i, g_i schon eindeutig bestimmt. Dabei bildet (P, h, A_i, g_i) ein Gestell.*

f) *Alle absoluten Punkte von δ liegen in g_1 oder in g_2, und alle absoluten Geraden von δ gehen durch A_1 oder durch A_2.*

Beweis. a) Sei $A_i \mathrm{I}\, g$; wegen $A_i^\delta = g_i$ folgt dann $g^\delta \mathrm{I}\, g_i$. Also ist g^δ absoluter Punkt und es gilt $g^\delta \mathrm{I}\, g^{\delta^2}$. Anwendung der Dualität δ^{-1} ergibt dann $g^\delta \mathrm{I}\, g$, d.h. g ist absolute Gerade. Ist nun $g \neq g_i$, dann folgt aus $g^\delta \mathrm{I}\, g$ und $g^\delta \mathrm{I}\, g_i$ sogar $g^\delta = g \cap g_i$. Also ist die Abbildung $g \mapsto g \cap g_i$ auf der Menge der von g_i verschiedenen Geraden durch A_i eine Injektion. Da es wenigstens 2 derartige Geraden gibt, muß $A_i \mathrm{I}\!\!\!/\, g_i$ gelten.

b) Sei $Q \mathrm{I}\!\!\!/\, h$. Wie im Beweis zu a) gezeigt, gilt dann

$$(A_i + Q)^\delta = (A_i + Q) \cap g_i.$$

Also ist

$$\begin{aligned}Q^\delta &= [(A_1+Q)\cap(A_2+Q)]^\delta\\ &= (A_1+Q)^\delta+(A_2+Q)^\delta\\ &= [(A_1+Q)\cap g_1]+[(A_2+Q)\cap g_2].\end{aligned}$$

c) Sei $B_1, B_2 \mathrm{I}\, g_i$, $B_1 \neq B_2$. Dann ist

$$\begin{aligned}B_j^\delta &= [(A_1+B_j)\cap g_1]+[(A_2+B_j)\cap g_2]\\ &= B_j+[(A_{i+1}+B_j)\cap g_{i+1}] = B_j + A_{i+1}.\end{aligned}$$

Daher ist auch

$$g_i^\delta = (B_1+B_2)^\delta = B_1^\delta \cap B_2^\delta = (B_1+A_{i+1})\cap(B_2+A_{i+1}) = A_{i+1}.$$

Ferner ist

$$P^\delta = (g_1\cap g_2)^\delta = g_1^\delta + g_2^\delta = A_2 + A_1 = h$$

und

$$h^\delta = (A_1+A_2)^\delta = A_1^\delta \cap A_2^\delta = g_1 \cap g_2 = P.$$

d) Sei $P \not\mathrm{I}\, g$. Nach Beweisteil c) gilt dann $(g\cap g_i)^\delta = (g\cap g_i) + A_{i+1}$. Also ist

$$\begin{aligned}g^\delta &= [(g\cap g_1)+(g\cap g_2)]^\delta\\ &= (g\cap g_1)^\delta \cap (g\cap g_2)^\delta\\ &= [(g\cap g_1)+A_2]\cap[(g\cap g_2)+A_1].\end{aligned}$$

e) Gemäß b) ist δ allein durch die Festlegung von A_i und g_i schon auf allen Punkten der durch h geschlitzten Ebene definiert. Dann läßt sich δ bekanntlich höchstens auf eine Weise als Kollineation von ganz $\mathfrak{E}$ (auf die zu $\mathfrak{E}$ duale Ebene) fortsetzen. Zum Beweis, daß $(P, h, A_i\, g_i)$ ein Gestell ist, muß noch gezeigt werden $g_1 \neq h \neq g_2$, $A_1 \neq P \neq A_2$ und $P \mathrm{I}\, h$. Nun ist nach Definition $P \mathrm{I}\, g_i$, also P absoluter Punkt, d.h. $P \mathrm{I}\, P^\delta$. Nach c) ist $P^\delta = h$, also $P \mathrm{I}\, h$. Nach a) gilt $A_i \not\mathrm{I}\, g_i$; daraus folgen die übrigen Aussagen.

f) Sei $C \not\mathrm{I}\, g_i$ für $i=1,2$. Ist $C \mathrm{I}\, h$, dann ist $C \neq P$, also $C^\delta \neq P^\delta = h = C+P$. Andererseits ist $h^\delta = P$ also $P \mathrm{I}\, C^\delta$. Beides zusammen erzwingt $C \not\mathrm{I}\, C^\delta$. Nun sei $C \not\mathrm{I}\, h$; insbesondere ist $A_1 \neq C \neq A_2$. Angenommen, es ist $C \mathrm{I}\, C^\delta$, dann folgt wegen $C \not\mathrm{I}\, g_i$, daß gilt $C^\delta = C + [(C+A_i)\cap g_i] = C + A_i$, also $C + A_1 = C^\delta = C + A_2$, woraus wegen $A_1 + A_2 = h$ der Widerspruch $C \mathrm{I}\, h$ erfolgt. Also liegt jeder absolute Punkt in g_1 oder g_2. Ist nun umgekehrt r eine absolute Gerade, so folgt aus $r^\delta \mathrm{I}\, r$, indem man δ^{-1} anwendet, $r^{\delta^{-1}} \mathrm{I}\, (r^{\delta^{-1}})^\delta$. Also ist $r^{\delta^{-1}}$ ein absoluter Punkt und liegt daher in g_1 oder in g_2. Nach c) liegt dann A_2 oder A_1 in r.

Satz 2. *Sei (P, h, A_i, g_i) ein Gestell von $\mathfrak{E}$. Genau dann existiert in $\mathfrak{E}$ die (A_i, g_i)-Dualität δ, wenn in $\mathfrak{E}$ der (A_i, g_i)-Satz von Pappos gilt.*

Beweis (vgl. Fig. 1). In $\mathfrak{E}$ existiere die (A_i, g_i)-Dualität δ. Sei (Q, g) ein inzidentes Punkt-Geradenpaar mit $Q \not{I} h$ und $P \not{I} g$. Sei $g \cap g_i = R_i$ und $(Q + A_i) \cap g_i = S_i$. Nach Satz 1b ist dann $Q^\delta = S_1 + S_2$, und nach Satz 1d ist $g^\delta = (R_1 + A_2) \cap (R_2 + A_1)$. Da δ Dualität ist, folgt aus $Q \mathrm{I} g$ auch $g^\delta \mathrm{I} Q^\delta$, d.h. $(R_1 + A_2) \cap (R_2 + A_1) \mathrm{I} (S_1 + S_2)$. Also gilt in $\mathfrak{E}$ der (A_i, g_i)-Satz von Pappos.

Nun sei umgekehrt die Gültigkeit des (A_i, g_i)-Satzes von Pappos in $\mathfrak{E}$ vorausgesetzt. Sei $\mathfrak{P}$ die Menge der Punkte von $\mathfrak{E}$, die nicht in h liegen, und sei $\mathfrak{G}$ die Menge der Geraden von $\mathfrak{E}$, die nicht durch P gehen. Auf $\mathfrak{P}$ wird eine Abbildung δ wie folgt definiert: Für $Q \not{I} h$ soll sein $Q^\delta = [(A_1 + Q) \cap g_1] + [(A_2 + Q) \cap g_2]$. Auf $\mathfrak{G}$ wird eine Abbildung η wie folgt definiert: Für r mit $P \not{I} r$ soll sein $r^\eta = [(r \cap g_1) + A_1] \cap [(r \cap g_2) + A_2]$. Man stellt nun fest, daß δ eine Abbildung von $\mathfrak{P}$ in $\mathfrak{G}$ ist, während η umgekehrt $\mathfrak{G}$ in $\mathfrak{P}$ abbildet. Ebenso sieht man schnell, daß $\delta\eta$ die Identität auf $\mathfrak{P}$ ergibt und $\eta\delta$ die Identität auf $\mathfrak{G}$ ergibt. Also ist δ eine Bijektion von $\mathfrak{P}$ auf $\mathfrak{G}$.

Wir zeigen, daß δ solche Punkte aus $\mathfrak{P}$, die mit einer Geraden aus $\mathfrak{G}$ inzidieren auf Geraden aus $\mathfrak{G}$ abbildet, die alle durch einen Punkt aus $\mathfrak{P}$ gehen: Sei g eine Gerade mit $P \not{I} g$, sei $R_i = g \cap g_i$. Nach Definition von δ ist dann (vgl. den Beweis zu 1c) $R_i^\delta = R_i + A_{i+1}$. Nun sei $Q \mathrm{I} g$, und sei $S_i = (Q + A_i) \cap g_i$. Da der (A_i, g_i)-Satz von Pappos gilt, ist $(R_1 + A_2) \cap (R_2 + A_1) \mathrm{I} (S_1 + S_2)$. Das heißt aber, es gilt $R_1^\delta \cap R_2^\delta \mathrm{I} Q^\delta$, und $R_1^\delta \cap R_2^\delta$ ist der gesuchte Punkt aus $\mathfrak{P}$.

Daher induziert δ auf den Geraden aus $\mathfrak{G}$ eine Abbildung in $\mathfrak{P}$, definiert durch $g^\delta = [(g \cap g_1) + A_2] \cap [(g \cap g_2) + A_1]$. Man sieht, daß die Abbildung η von $\mathfrak{P}$ in $\mathfrak{G}$, definiert durch

$$S^\eta = [(S + A_1) \cap g_2] + [(S + A_2) \cap g_1]$$

die zur vorigen Abbildung inverse ist; also induziert δ sogar eine Bijektion von $\mathfrak{G}$ auf $\mathfrak{P}$. Nun sei g eine Gerade mit $P \mathrm{I} g$, und R und S seien zwei von P und untereinander verschiedene Punkte von g. Sei $T = R^\delta \cap S^\delta$. Angenommen es ist $T \not{I} h$. Dann liegen nach dem eben Bewiesenen die Urbilder aller Geraden durch T unter δ auf genau einer Geraden r mit $P \not{I} r$. Da dann wenigstens einer der Punkte R oder S nicht auf r liegen kann, ist das ein Widerspruch. Also gilt $T \mathrm{I} h$, und es folgt $T \mathrm{I} Q^\delta$ für alle $Q \in \mathfrak{P}$ mit $Q \mathrm{I} g$. Daher ist nun gezeigt, daß δ eine Kollineation einer affinen Ebene (nämlich der durch h geschlitzten Ebene $\mathfrak{E}$) auf eine affine Ebene (nämlich die zur vorigen duale durch P geschlitzte Ebene $\mathfrak{E}$) ist. Dann läßt sich aber δ bekanntlich als Dualität auf ganz $\mathfrak{E}$ fortsetzen.

Bemerkung 1. Der (P, h)-Satz von Pappos bezüglich A_1, A_2, g_1, g_2 ist gleichwertig mit dem (P, h)-Satz von Pappos bezüglich A_2, A_1, g_1, g_2: Die Gültigkeit des (P, h)-Satzes von Pappos in $\mathfrak{E}$ bezüglich A_1, A_2, g_1, g_2 ist gleichwertig zur Existenz der (P, h)-Dualität δ in $\mathfrak{E}$ bezüglich A_1, A_2,

g_1, g_2. Diese ist aber gleichwertig zur Existenz einer (P, h)-Dualität in $\mathfrak{E}$ bezüglich A_2, A_1, g_1, g_2 (nämlich δ^{-1}), was schließlich gleichwertig zur Gültigkeit des (P, h)-Satzes von Pappos in $\mathfrak{E}$ bezüglich A_2, A_1, g_1, g_2 ist.

Satz 3. *Aus dem $\mathfrak{D}$-Satz von Pappos folgt der $\mathfrak{D}$-Satz von Desargues.*

Beweis (vgl. Fig. 2). Sei $\mathfrak{D}=(P, h, A_i, g_i)$ ein Gestell von $\mathfrak{E}$, und in $\mathfrak{E}$ gelte der $\mathfrak{D}$-Satz von Pappos. Nach Satz 2 besitzt dann $\mathfrak{E}$ die (A_i, g_i)-Dualität δ. Seien $g_0, C_j, D_j, j=0, 1, 2$, Elemente von $\mathfrak{E}$ welche die Voraussetzungen von Definition 4 erfüllen. Offenbar ist dann $C_0^\delta = C_1 + C_2$ und $D_0^\delta = D_1 + D_2$. Ferner ist $g_0^\delta = (P + C_0)^\delta = h \cap (C_1 + C_2)$. Wegen $D_0 \mathrm{I} g_0$ gilt auch $g_0^\delta \mathrm{I} D_0^\delta$, d.h. es gilt $[h \cap (C_1 + C_2)] \mathrm{I} (D_1 + D_2)$. Also gilt in $\mathfrak{E}$ der $\mathfrak{D}$-Satz von Desargues.

Der $(V, U+V)$-Satz von Desargues bezüglich $W, U, [0], [a]$ läßt sich leicht als Bedingung im Ternärkörper ausdrücken (vgl. [7], S. 97f.). Dies führt auf das folgende

Korollar 1. *Notwendige Bedingung für das Vorhandensein einer $(V, U+V)$-Dualität bezüglich $U, W, [0], [a]$ ist die Gültigkeit der Regel*

$$\mathbf{T}(m, a, k) = m\,a + k \qquad \text{für alle } m, k \in \mathfrak{K}.$$

Beweis. Besitzt $\mathfrak{E}$ die $(V, U+V)$-Dualität δ bezüglich $U, W, [0], [a]$, dann besitzt $\mathfrak{E}$ in δ^{-1} auch die $(V, U+V)$-Dualität bezüglich $W, U, [0], [a]$, also gilt nach Satz 3 in $\mathfrak{E}$ auch der $(V, U+V)$-Satz von Desargues bezüglich $W, U, [0], [a]$. Nun sei $C_1 = (0, 0)$, $C_2 = (a, m\,a)$, $C_0 = (m\,a, m\,a)$, $D_1 = (0, k)$, $D_2 = (a, m\,a + k)$, $D_0 = (m\,a, m\,a + k)$, dann sind mit $g_0 = [m\,a]$ alle Voraussetzungen dieses Satzes erfüllt. Insbesondere ist $(C_1 + C_2) \cap (U + V) = (m)$. Der Satz besagt dann, daß $D_1 + D_2 = [m, k]$ ist. $D_2 \mathrm{I} [m, k]$ ergibt aber gerade $\mathbf{T}(m, a, k) = m\,a + k$.

3. Weitere Schließungssätze bei speziellen g_i-Dualitäten

Sei (P, h, A_i, g_i) ein Gestell und δ die (P, h)-Dualität bezüglich A_i, g_i. Wie Satz 1 zeigt, vertauscht δ^2 gerade A_1 mit A_2 und g_1 mit g_2, während P und h von δ^2 festgelassen werden. Das legt die Frage nahe, unter welchen Bedingungen δ^4 eine zentrale Kollineation mit Zentrum P und Achse h ergibt. – Andererseits ist $\delta^4 \neq 1$: Sei $C \mathrm{I} g_1$ mit $C \neq P$, dann berechnet man: $C^{\delta^2} = (C + A_2) \cap g_2$, $C^{\delta^4} = (C^{\delta^2} + A_1) \cap g_1$. Wäre $C^{\delta^4} = C$, so wären C, A_1, A_2 kollinear, also $C \mathrm{I} h$.

Satz 4. *Sei δ eine (P, h)-Dualität von $\mathfrak{E}$ bezüglich A_i, g_i. Dann sind gleichwertig:*

a) *δ^4 ist eine zentrale Kollineation,*

b) *in $\mathfrak{E}$ gilt der kleine axiale (P, h)-Satz von Pappos bezüglich A_i, g_i,*

c) *in $\mathfrak{E}$ gilt der kleine zentrale (P, h)-Satz von Pappos bezüglich A_i, g_i.*

Beweis. Sei δ^4 eine zentrale Kollineation. Es ist $\delta^4 \neq 1$ und P hat die drei Fixgeraden g_1, g_2, h, während h die drei Fix-Punkte A_1, A_2 und P besitzt. Also hat δ^4 das Zentrum P und die Achse h.

Sei $C \,\mathrm{I}\!\!\!/\, h$, $R_i = (C + A_i) \cap g_i$, $S_i = (R_{i+1} + A_i) \cap g_i$; dann ist $C^\delta = R_1 + R_2$,

$$C^{\delta^2} = (R_1 + A_2) \cap (R_2 + A_1),$$
$$\begin{aligned} C^{\delta^3} &= [(C^{\delta^2} + A_1) \cap g_1] + [(C^{\delta^2} + A_2) \cap g_2] \\ &= [(R_2 + A_1) \cap g_1] + [(R_1 + A_2) \cap g_2] \\ &= S_1 + S_2, \end{aligned}$$
$$C^{\delta^4} = (S_1 + A_2) \cap (S_2 + A_1).$$

Ist δ^4 zentrale Kollineation mit Zentrum P, so folgt $C^{\delta^4} \mathrm{I} (C + P)$, d.h. $[(S_1 + A_2) \cap (S_2 + A_1)] \mathrm{I} (C + P)$. Daher gilt der kleine axiale (P, h)-Satz von Pappos bezüglich A_i, g_i in $\mathfrak{E}$.

Offenbar ist die Richtung dieses Beweises umkehrbar; also sind a) und b) äquivalent.

Wendet man nun δ auf eine Konfiguration des kleinen axialen (P, h)-Satzes von Pappos bezüglich A_i, g_i an, so erhält man gerade eine Konfiguration des kleinen zentralen (P, h)-Satzes von Pappos bezüglich A_i, g_i. Also sind auch b) und c) äquivalent.

Eine Äquivalenz der beiden kleinen $\mathfrak{D}$-Sätze von Pappos läßt sich auch mit Hilfe der in Korollar 1 gefundenen Bedingung für einen geeigneten Ternärkörper beweisen:

Satz 5. *$\mathfrak{E}$ besitze bezüglich O, U, V, E einen Ternärkörper $\mathfrak{K}$, welcher der Bedingung*

$$\mathbf{T}(m, a, k) = m a + k \qquad \text{für alle } m, k \in \mathfrak{K}$$

für ein festes $a \neq 0$ genügt. Dann sind gleichwertig:

a) $x + (a + z) = a + (x + z)$ *für alle* $x, z \in \mathfrak{K}$,

b) *in $\mathfrak{E}$ gilt der kleine axiale $(V, U+V)$-Satz von Pappos bezüglich $U, W, [0], [a]$,*

c) *in $\mathfrak{E}$ gilt der kleine zentrale $(V, U+V)$-Satz von Pappos bezüglich $U, W, [0], [a]$.*

Beweis. Der kleine axiale $(V, U+V)$-Satz von Pappos bezüglich U, W läßt sich als Thomsen-Bedingung im U, V, W-Gewebe auffassen. Von dorther ergibt sich die Äquivalenz von a) und b) unmittelbar (vgl. auch [6], S. 58). Wir leiten sie im folgenden noch einmal kurz her. Nach Vorgabe des Punktes $C = (x, x + z)$ sollen die Punkte R_i, S_i gemäß Definition 5a bestimmt werden (vgl. Fig. 3). Man erhält $R_1 = (0, x + z)$, $R_2 = (a, a + z)$, $S_1 = (0, a + z)$, $S_2 = (a, a + (x + z))$. Sei $D = (S_1 + W) \cap (S_2 + U)$ und $D + V = [x']$. Dann folgt aus $D \mathrm{I} (S_1 + W)$, daß $D = (x', x' + (a + z))$ ist, während $D \mathrm{I} (S_2 + U)$ andererseits $D = (x', a + (x + z))$ impliziert. Gilt nun a), so folgt $x' + (a + z) = a + (x + z) = x + (a + z)$, also $x = x'$, mithin $D \mathrm{I} [x]$. Mithin folgt b) aus a). Gilt umgekehrt b), dann gilt $D + V = [x]$, also $x + (a + z) = a + (x + z)$; demnach folgt auch a) aus b).

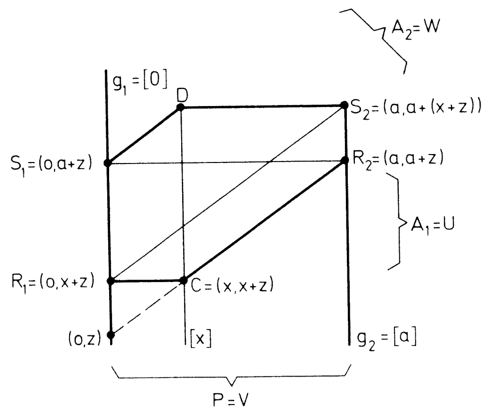


Fig. 3

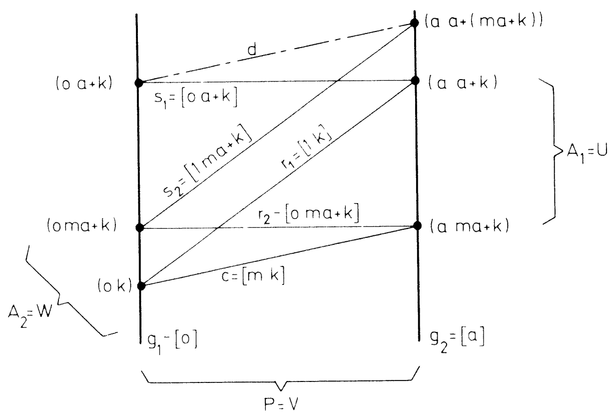


Fig. 4

Nun soll die Äquivalenz von a) mit c) gezeigt werden. Wir verwenden die Bezeichnungen von Definition 5b (vgl. Fig. 4). Sei $c = [m, k]$, dann ist $c \cap [0] = (0, k)$ und wegen der besonderen Voraussetzung über den Ternärkörper $\mathfrak{K}$ ist $c \cap [a] = (a, ma + k)$. Weiter ist $r_1 = [1, k]$, $r_2 = [0, ma + k]$, $r_1 \cap [a] = (a, a + k)$, $r_2 \cap [0] = (0, ma + k)$, $s_1 = [0, a + k]$, $s_2 = [1, ma + k]$, $s_1 \cap [0] = (0, a + k)$, $s_2 \cap [a] = (a, a + (ma + k))$. Sei $d = (0, a + k) + (a, a + (ma + k)) = [m', a + k]$. Dann ergibt $(a, a + (ma + k))\, \mathrm{I}\, d$, wenn man noch die vorausgesetzte besondere Eigenschaft von $\mathfrak{K}$ verwendet, die Gleichung

$$a + (ma + k) = m'a + (a + k).$$

Gilt nun a), dann folgt $m'a + (a + k) = ma + (a + k)$, also $m = m'$; das ist aber c). Gilt umgekehrt c), so folgt $m = m'$, also die Formel

$$a + (ma + k) = ma + (a + k).$$

Für ma und k können dabei beliebige Elemente von $\mathfrak{K}$ genommen werden, also gilt a).

Satz 4 und 5 zusammen ergeben unter Verwendung von Korollar 1 das Folgende

Korollar 2. *$\mathfrak{E}$ besitze eine $(V, U+V)$-Dualität δ bezüglich $U, W, [0], [a]$. Genau dann ist δ^4 eine zentrale Kollineation, wenn gilt*

$$a+(x+z)=x+(a+z) \quad \textit{für alle } x, z \in \mathfrak{K}.$$

Es ist möglich, den Sachverhalt von Satz 5 auch ohne Koordinaten zu formulieren und zu beweisen, was anhangsweise im folgenden noch geschehen soll:

Satz 6. *Sei $\mathfrak{D}=(P, h, A_i, g_i)$ ein Gestell von $\mathfrak{E}$. Gilt in $\mathfrak{E}$ der $\mathfrak{D}$-Satz von Desargues, so sind in $\mathfrak{E}$ der kleine axiale $\mathfrak{D}$-Satz von Pappos und der kleine zentrale $\mathfrak{D}$-Satz von Pappos äquivalent*[4].

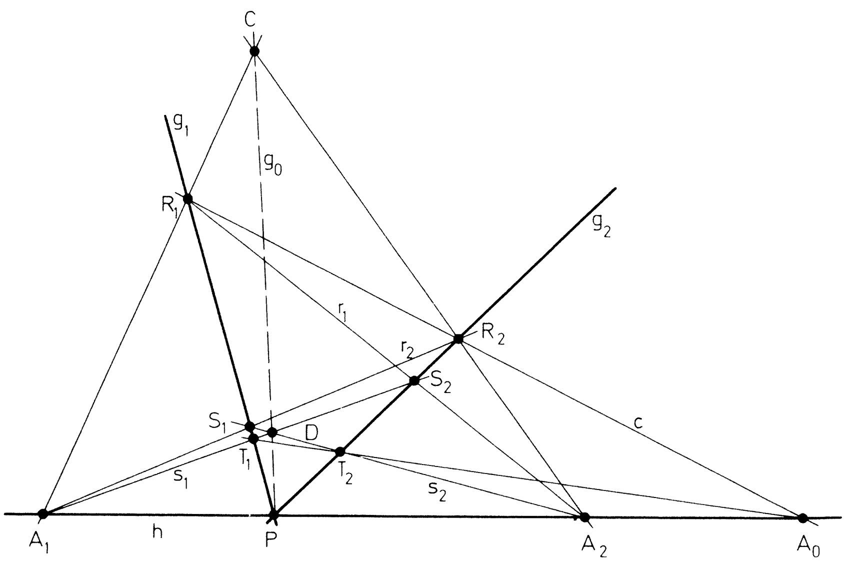


Fig. 5

Beweis (vgl. Fig. 5). In $\mathfrak{E}$ gelte der $\mathfrak{D}$-Satz von Desargues und der kleine axiale $\mathfrak{D}$-Satz von Pappos. Wir verwenden die Bezeichnungen von Definition 5b. Sei c eine Gerade von $\mathfrak{E}$ mit $P \not{I} c$. Sei $c \cap h = A_0$. Gilt $A_0 = A_i$ für $i=1$ oder $i=2$, so ist der kleine zentrale $\mathfrak{D}$-Satz von Pappos trivialerweise erfüllt. Sei also $A_1 \neq A_0 \neq A_2$. Sei $c \cap g_i = R_i$, also $r_i =$

[4] Daß die kleinen $\mathfrak{D}$-Sätze von Pappos nicht etwa aus dem $\mathfrak{D}$-Satz von Desargues folgen, erhellt schon aus der Existenz cartesischer Gruppen mit nicht-kommutativer Addition (vgl. [7], S. 91) zusammen mit Satz 5.

$R_i + A_{i+1}$. Weiter wird definiert $S_i = r_{i+1} \cap g_i$, was $s_i = S_{i+1} + A_i$ ergibt. Schließlich sei $s_i \cap g_i = T_i$ und $s_1 \cap s_2 = D$. Dann ist also zu zeigen $A_0 \mathrm{I} (T_1 + T_2)$. Sei noch $C = (A_1 + R_1) \cap (A_2 + R_2)$, dann erfüllen C, R_i, S_i die Voraussetzungen des kleinen axialen $\mathfrak{D}$-Satzes von Pappos gemäß Definition 5a. Also folgt $D \mathrm{I} (C + P)$. Setzt man nun $g_0 = C + P$, so erfüllen C, D, R_i, T_i alle Voraussetzungen des $\mathfrak{D}$-Satzes von Desargues, wenn man in Definition 4 nur $C_0 = C$, $D_0 = D$, und für $i = 1, 2$ noch $C_i = R_i$ und $D_i = T_i$ setzt. Also folgt $A_0 \mathrm{I} (T_1 + T_2)$.

Jetzt werde vorausgesetzt, daß in $\mathfrak{E}$ der $\mathfrak{D}$-Satz von Desargues und der kleine zentrale $\mathfrak{D}$-Satz von Pappos gilt. Sei C ein Punkt von $\mathfrak{E}$ mit $C \not\mathrm{I}\, h$, sei $g_0 = C + P$. Um triviale Fälle auszuschließen, dürfen wir wieder $g_1 \neq g_0 \neq g_2$ annehmen. Seien R_i, S_i gemäß Definition 5a bestimmt und $D = (S_1 + A_2) \cap (S_2 + A_1)$. Dann ist $D \mathrm{I}\, g_0$ zu zeigen.

Sei $T_i = (D + A_i) \cap g_i$. Setzt man nun $c = R_1 + R_2$, $r_i = R_i + A_{i+1}$, $s_i = D + A_i$, so erfüllen c, r_i, s_i die Voraussetzungen des kleinen zentralen $\mathfrak{D}$-Satzes von Pappos gemäß Definition 5b. Setzt man noch $A_0 = c \cap h$, so folgt also auch, daß $A_0 \mathrm{I} (T_1 + T_2)$ gilt.

Angenommen $D \not\mathrm{I}\, g_0$. Sei $(A_1 + T_1) \cap g_0 = D'$ und $(A_2 + D') \cap g_2 = T_2'$. Insbesondere ist $D \neq D'$, $T_2 \neq T_2'$, $T_1 + T_2 \neq T_1 + T_2'$. Aber die Punkte C, R_1, R_2, D', T_1, T_2' erfüllen die Voraussetzungen des $\mathfrak{D}$-Satzes von Desargues, wenn man sie gemäß Definition 4 der Reihe nach mit C_0, C_1, C_2, D_0, D_1, D_2 bezeichnet. Also muß auch gelten $A_0 \mathrm{I} (T_1 + T_2')$. Aber $A_0 \neq T_1$ ergibt nun den Widerspruch

$$T_1 + T_2 = T_1 + A_0 = T_1 + T_2'.$$

Also gilt $D \mathrm{I}\, g_0$.

4. (P, h)-Dualitäten in (P, h)-transitiven Ebenen

Wie in Abschnitt 2 und 3 wird auch hier ein Element $a \neq 0$ des Ternärkörpers von $\mathfrak{E}$ zur Definition des Gestells einer $(V, U+V)$-Dualität δ herangezogen. Jedoch soll a diesmal nicht zur Bestimmung der Geraden g_2 sondern des Punktes A_2 dienen; dadurch werden die Rechnungen einfacher, und die Ergebnisse gewinnen an Übersichtlichkeit[5]. — (P, h) bezeichnet stets ein inzidentes Punkt-Geradenpaar.

[5] Betrachtet man eine fest gewählte Dualität δ von $\mathfrak{E}$, so kann man sich Koordinaten von $\mathfrak{E}$ so bestimmt denken, daß δ die $(V, U+V)$-Dualität bezüglich U, W, $O+V$, $E+V$ ist. Dann kann man also, weil jetzt $a = 1$ ist, die Ergebnisse von Abschnit 2—4 gemeinsam betrachten. Ist $\mathfrak{E}$ außerdem als $(V, U+V)$-transitiv vorausgesetzt, so existiert in $\mathfrak{E}$ zu vorgegebenem $0 \neq a \in \mathfrak{K}$ genau dann die $(V, U+V)$-Dualität bezüglich (0), (a), $O+V$, $E+V$, wenn in $\mathfrak{E}$ auch die $(V, U+V)$-Dualität bezüglich U, V, $[0]$, $[a]$ existiert. Denn wendet man δ der Reihe nach auf (0), (a), $[0]$, $[1]$ an, so erhält man (z.B. nach den Formeln vor Korollar 3) $[0]$, $[a]$, (1), (0). Nun benutze man noch Bemerkung 1 und Hilfssatz 2 (im 6. Abschnitt). — Also ist auch für diesen Fall Abschnitt 4 mit den vorigen Abschnitten verbunden.

Satz 7. *$\mathfrak{E}$ sei $(V, U+V)$-transitiv. Genau dann besitzt $\mathfrak{E}$ eine $(V, U+V)$-Dualität bezüglich* (0), (a), $O+V$, $E+V$, *wenn gilt*

$$(a-a\,x)\,z = a\,z-(a\,z)\,x \qquad \textit{für alle } x, z\in\mathfrak{K}.$$

Beweis. Wegen der $(V, U+V)$-Transitivität von $\mathfrak{E}$ ist der Ternärkörper $\mathfrak{K}$ von $\mathfrak{E}$ (definiert bezüglich O, U, V, E) eine cartesische Gruppe. Nun besitze $\mathfrak{E}$ die $(V, U+V)$-Dualität δ bezüglich (0), (a), $O+V$, $E+V$. Sei $x, z\in\mathfrak{K}$, $g=[a\,z, 0]$, $y=(a\,z)\,x$, $Q=(x, y)$, also $Q\,\mathrm{I}\,g$. Mit Satz 1 b, d folgt

$$\begin{aligned} Q^\delta &= \big([(x,y)+(0)]\cap(O+V)\big)+\big([(x,y)+(a)]\cap(E+V)\big)\\ &= \big([0,y]\cap(O+V)\big)+\big([a,\,-a\,x+y]\cap(E+V)\big)\\ &= (0,y)+(1,\,a-a\,x+y)\\ &= [a-a\,x,\,y],\\ g^\delta &= \big[\big([a\,z,0]\cap(O+V)\big)+(a)\big]\cap\big[\big([a\,z,0]\cap(E+V)\big)+(0)\big]\\ &= [(0,0)+(a)]\cap[(1,a\,z)+(0)]\\ &= [a,0]\cap[0,a\,z]\\ &= (z,a\,z). \end{aligned}$$

Da δ eine Dualität ist, folgt aus $Q\,\mathrm{I}\,g$ auch $g^\delta\,\mathrm{I}\,Q^\delta$ also $(z, a\,z)\,\mathrm{I}\,[a-a\,x, y]$. Beachtet man noch $y=(a\,z)\,x$, ergibt das

$$a\,z=(a-a\,x)\,z+(a\,z)\,x.$$

Nun sei umgekehrt $\mathfrak{K}$ eine cartesische Gruppe, die der Bedingung

$$(a-a\,x)\,z=a\,z-(a\,z)\,x \qquad \text{für alle } x, z\in\mathfrak{K}$$

genügt, und $\mathfrak{E}$ sei die projektive Ebene über $\mathfrak{K}$. Sei $\mathfrak{P}$ die Menge der Punkte von $\mathfrak{E}$ der Form (x, y), und sei $\mathfrak{G}$ die Menge der Geraden von $\mathfrak{E}$ der Form $[m, k]$. Dann definieren wir eine Abbildung $\bar{\delta}$ auf $\mathfrak{P}\cup\mathfrak{G}$ durch

$$(x,y)^{\bar{\delta}}=[a-a\,x,\,y], \qquad [a\,z,k]^{\bar{\delta}}=(z,\,a\,z+k).$$

Man stellt leicht fest, daß $\bar{\delta}$ die Menge $\mathfrak{P}$ auf $\mathfrak{G}$, sowie $\mathfrak{G}$ auf $\mathfrak{P}$ bijektiv abbildet. Sei nun $(x, y)\,\mathrm{I}\,[a\,z, k]$, also $y=(a\,z)\,x+k$. Dann folgt $(a-a\,x)\,z+y$ $=a\,z-(a\,z)\,x+(a\,z)\,x+k=a\,z+k$, also $[a\,z, k]^{\bar{\delta}}\,\mathrm{I}\,(x, y)^{\bar{\delta}}$. Nun ist $\mathfrak{P}$ gerade die Menge der Punkte von $\mathfrak{E}$, die nicht in $U+V$ liegen, während $\mathfrak{G}$ die Menge derjenigen Geraden von $\mathfrak{E}$ ist, die nicht durch V gehen. Also können wir den Gedankengang im Beweis zu Satz 2 hier ebenfalls anwenden mit dem Ergebnis, daß durch $\bar{\delta}$ eine Dualität δ auf $\mathfrak{E}$ definiert ist. Weiter ist $(0, y)^\delta=[a, y]$ und $(1, y)^\delta=[0, y]$. Also sind alle Punkte aus $\mathfrak{P}$, die in $O+V$ oder $E+V$ liegen, absolute Punkte von δ, also muß auch der letzte noch übriggebliebene Punkt dieser Geraden, nämlich V, ein absoluter Punkt sein. Wegen $(0, y)^\delta = [a, y]$ für alle $y \in \mathfrak{K}$ ist dann $(O+V)^\delta = (a)$; entsprechend folgt aus $(1, y)^\delta = [0, y]$ für alle $y \in \mathfrak{K}$ auch $(E+V)^\delta = (0)$.

Nach Definition 2 und Satz 1c ist δ also die $(V, U+V)$-Dualität von E bezüglich (0), (a), $O+V$, $E+V$.

Man berechnet übrigens leicht die δ-Bilder der Punkte von $U+V$ und der Geraden durch V:

$$(a\,z)^\delta = [z], \qquad [x]^\delta = (a - a\,x).$$

Korollar 3. *Genau dann ist $\mathfrak{E}$ projektive Ebene über einem kommutativen Quasikörper* [6], *wenn es in $\mathfrak{E}$ ein inzidentes Punkt-Geradenpaar (P, h) derart gibt, daß $\mathfrak{E}$ eine (P, h)-Dualität besitzt und (P, P)- oder (h, h)-transitiv ist.*

Beweis. Sei (P, h) ein inzidentes Punkt-Geradenpaar von $\mathfrak{E}$ und δ sei die (P, h)-Dualität von $\mathfrak{E}$ bezüglich A_i, g_i. Dann bildet δ die Menge der Paare (X, h) mit $X \mathrm{I} h$ auf die Menge der Paare (P, x) mit $P \mathrm{I} x$ ab und umgekehrt. Das bedeutet: Ist $\mathfrak{E}$ (h, h)-transitiv, dann auch (P, P)-transitiv und umgekehrt. Also folgt aus der Voraussetzung, daß $\mathfrak{E}$ sowohl (P, P)- als auch (h, h)-transitiv ist. Wählt man nun O, U, V, E so, daß $O+V = g_1$, $E+V = g_2$, $A_1 = U$ und $A_2 = W$ ist, so ist der zugehörige Ternärkörper $\mathfrak{K}$ ein distributiver Quasikörper ([7], Satz 38 und 39, S. 101), der überdies nach Satz 7 (für $a=1$) die Regel

$$(1-x)\,z = z - z\,x \qquad \text{für alle } x, z \in K$$

erfüllt. Daraus folgt aber die Kommutativität der Multiplikation von $\mathfrak{K}$.

Sei nun umgekehrt $\mathfrak{E}$ die projektive Ebene über dem kommutativen Quasikörper $\mathfrak{K}$. Dann ist $\mathfrak{K}$ ein distributiver Quasikörper und nach den oben zitierten Sätzen ist $\mathfrak{E}$ (V, V)- und $(U+V, U+V)$-transitiv. Überdies erfüllt $\mathfrak{K}$ die Regel $(1-x)\,z = z - z\,x$ und besitzt daher nach Satz 7 die $(V, U+V)$-Dualität δ bezüglich U, W, $O+V$, $E+V$, definiert durch

$$(x, y)^\delta = [1-x, y]; \qquad (z)^\delta = [z],$$
$$[m, k]^\delta = (m, m+k); \qquad [z]^\delta = (1-z).$$

Da nach [7], Satz 6, S. 163 kommutative Alternativkörper bereits Körper sind, ergibt sich noch das folgende

Korollar 4. *In einer echten Moufang-Ebene gilt der $\mathfrak{D}$-Satz von Pappos für kein Gestell $\mathfrak{D}$.*

Zu Formulierung und Beweis des nächsten Satzes benötigen wir noch die folgende

---

[6] d.h. ein Quasikörper mit kommutativer Multiplikation.

Definition 6. Ein Punkt-Geradenpaar (S, r) heißt *ausgezeichnet in bezug auf die (P, h)-Dualität δ*, falls gilt

$$S^\delta = r, \qquad r^\delta = S, \qquad (S, r) \neq (P, h).$$

Sei δ die (P, h)-Dualität von $\mathfrak{E}$ bezüglich A_i, g_i und (S, r) ein ausgezeichnetes Punkt-Geradenpaar in bezug auf δ. Dann gilt $P \mathrm{I} r$ und also auch $S \mathrm{I} h$: Sei nämlich $r^{\delta^2} = r$, aber $P \not\mathrm{I} r$, und sei $R_i = r \cap g_i$. Dann folgt mit Satz 1

$$R_i^{\delta^2} = (r \cap g_i)^{\delta^2} = r^{\delta^2} \cap g_i^{\delta^2} = r \cap g_{i+1} = R_{i+1}.$$

Andererseits sind aber nach Satz 1 die Punkte R_i, $R_i^{\delta^2}$ und A_{i+1} kollinear, also A_1, R_1, R_2, A_2 kollinear. Das ergibt aber den Widerspruch $r = A_1 + A_2 = h$.

Das ausgezeichnete Paar (S, r) steht auch in merkwürdiger Beziehung zur Konstruktion harmonischer Punktequadrupel mittels eines vollständigen Vierecks (vgl. Fig. 6). Sei Q ein von P verschiedener Punkt

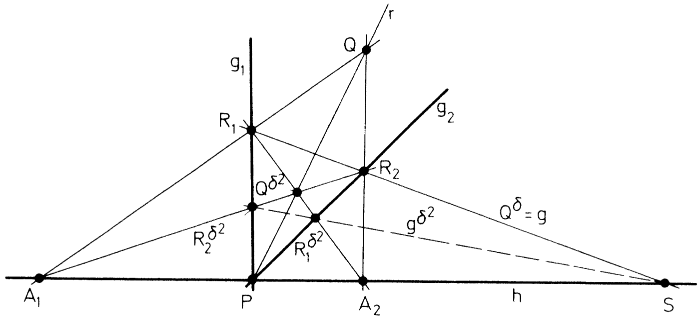


Fig. 6

mit $Q \mathrm{I} r$, dann gilt also $Q^{\delta^2} \mathrm{I} r$. Ist $R_i = (Q + A_i) \cap g_i$, so ist $Q^\delta = R_1 + R_2$, also $S = (R_1 + R_2) \cap h$. Weiter ist $Q^{\delta^2} = (R_1 + A_2) \cap (R_2 + A_1)$. Insbesondere ist $Q^{\delta^2} \neq Q$ (denn sonst wären R_1, R_2, A_1, A_2 kollinear, also $Q \mathrm{I} h$). Wegen $Q \not\mathrm{I} h$ gilt auch $R_i = (Q + A_i) \cap (Q^{\delta^2} + A_{i+1})$, und wir erhalten, daß S die *(Q, Q^{δ^2})-harmonische Konjugierte von P in bezug auf* (A_1, A_2) ist (vgl. die Definition auf S. 332 von [3]):

$$S = (A_1 + A_2) \cap ([(A_1 + Q) \cap (A_2 + Q^{\delta^2})] + [(A_2 + Q) \cap (A_1 + Q^{\delta^2})]).$$

Diese Aussage gilt für alle Paare (Q, Q^{δ^2}) mit $Q \neq P$, $Q \mathrm{I} r$. Durch Anwenden von δ erhalten wir die duale Aussage für r.

Es gilt aber auch umgekehrt: A_2 ist (R, R^{δ^2})-harmonisch konjugiert zu A_1 in bezug auf (P, S) für alle $R \mathrm{I} g_2$, $R \neq P$: Sei g eine Gerade von $\mathfrak{E}$ mit $S \mathrm{I} g$, $g \neq h$. Dann gilt also $S \mathrm{I} g^{\delta^2}$. Sei $R_i = g \cap g_i$. Dann ist $R_i^{\delta^2}=(R_i+A_{i+1})\cap g_{i+1}$. Daher ist $R_1=(R_2+S)\cap(R_2^{\delta^2}+P)$ und $R_1^{\delta^2}=(R_2^{\delta^2}+S)\cap(R_2+P)$ und $A_2=(R_1+R_1^{\delta^2})\cap(P+S)$. Wegen $A_1 \mathrm{I}(R_2+R_2^{\delta^2})$ ist daher A_2 die $(R_2, R_2^{\delta^2})$-harmonisch Konjugierte zu A_1 in bezug auf (P, S).

Die projektive Ebene $\mathfrak{E}$ sei nun als $(V, U+V)$-transitiv vorausgesetzt, und es existiere in $\mathfrak{E}$ die $(V, U+V)$-Dualität δ bezüglich $U, W, O+V, E+V$. Dann liest man aus der im Beweis zu Satz 7 explizit angegebenen Formel für δ (mit $a=1$) ab, daß ein ausgezeichnetes Punkt-Geradenpaar (S, r) in bezug auf δ genau dann existiert, wenn der zugehörige Ternärkörper $\mathfrak{K}$ ein Element z_0 mit $z_0+z_0=1$ enthält. Es ist dann $S=(z_0)$, $r=[z_0]$, mit einem Element z_0 der beschriebenen Eigenschaft.

Ist $\mathfrak{K}$ sogar ein kommutativer Quasikörper mit einer von 2 verschiedenen Charakteristik, so existiert ein solches z_0 und ist eindeutig bestimmt. Dann gibt es also zu δ ein eindeutig bestimmtes ausgezeichnetes Paar (S, r); diese Tatsache ließe sich auch aus der jetzt geltenden Eindeutigkeit der harmonischen Konjugierten bezüglich $(V,(m))$ für alle $m\in K$ erschließen (vgl. [3], Theorem 3.3).

Im folgenden Satz bezeichnen wir wie in [7], Satz 14, S. 70, mit $\mathbf{S}(\mathfrak{E})$ die Menge der inzidenten Punkt-Geradenpaare (Q, g), für die gilt: $\mathfrak{E}$ ist (Q, g)-transitiv, und benutzen die Klassifikation der Ebenen nach Lenz (vgl. [7], ebenda).

Satz 8. *Besitzt die nicht-desarguessche Ebene $\mathfrak{E}$ eine (P, h)-Dualität δ und ist $\mathbf{S}(\mathfrak{E})\neq\emptyset$, so ist $\mathfrak{E}$ von einem der folgenden Typen:*

II. $\mathbf{S}(\mathfrak{E})=\{(P, h)\}$.

III. $\mathbf{S}(\mathfrak{E})=\{(Q, g)\,|\,Q\mathrm{I}g, Q\mathrm{I}r, S\mathrm{I}g\}$, (S, r) *ein ausgezeichnetes Paar in bezug auf δ.*

V. $\mathbf{S}(\mathfrak{E})=\{(Q, g)\,|\,Q\mathrm{I}g, Q=P$ *oder* $g=h\}$.

Beweis. Wir untersuchen die möglichen Lagen von (Q, g) zu dem zu δ gehörigen Gestell (P, h, A_i, g_i) und die sich daraus mittels δ ergebenden Typen von Ebenen. Dabei wird es genügen, jeweils den niedrigsten Typ anzugeben.

1. Sei zuerst $g=h$, also $Q\mathrm{I}h$. Ist $Q=P$, so ergibt sich der oben angeführte Typ II. Sei also $Q\neq P$. Dann ist auch $(g^\delta, Q^\delta)=(P, Q^\delta)$ ein Element von $\mathbf{S}(\mathfrak{E})$. Sei T ein von P und Q verschiedener Punkt von h. Wegen der (P, Q^δ)-Transitivität von $\mathfrak{E}$ und wegen $Q^\delta\neq h$ gibt es dann $\alpha\in\Gamma(\mathfrak{E})$ mit $Q^\alpha=T$, $h^\alpha=h$. Daher ist $\mathfrak{E}$ auch (T, h)-transitiv. Die (Q, h)- und (T, h)-Transitivität zusammen ergibt nach [7], Satz 11, S. 67 die (h, h)-Transitivität von $\mathfrak{E}$. Nach Satz 7 folgt jetzt, daß $\mathfrak{E}$ von dem oben angegebenen Typ V ist.

2. Durch Anwendung von δ erledigen sich dual dazu alle Fälle mit $Q=P$.

3. Nun sei $g \neq h$ und $Q \neq P$. Dann gilt $(g^\delta, Q^\delta) \neq (Q, g)$. Denn Gleichheit würde bedeuten, daß Q ein absoluter Punkt, also $Q \mathrm{I} g_i$ ist, und ebenso, daß g absolute Gerade, also $A_j \mathrm{I} g$ ist. Dann prüft man aber leicht mit Satz 1 nach, daß $(Q, g) \neq (g^\delta, Q^\delta)$ oder $(Q, g) = (P, h)$ ist.

3.1. Sei $Q = g^\delta$ oder $Q^\delta = g$. Indem wir eventuell noch δ anwenden, können wir o.B.d.A. die erste Alternative annehmen. Da Q absolut ist, gilt etwa $Q \mathrm{I} g_1$ und daher $A_1 \mathrm{I} g$ (Fig. 7); Anwendung von δ^2 ergibt die

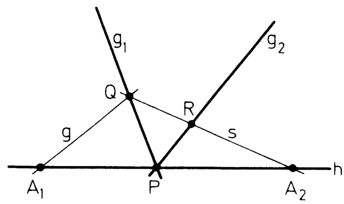


Fig. 7

(R, s)-Transitivität von $\mathfrak{E}$ mit $R = (Q + A_2) \cap g_2$ und $s = Q + A_2$; insbesondere ist $g \neq s$. Ist R' ein von R und Q verschiedener Punkt mit $R' \mathrm{I} s$, so gibt es demnach $\alpha \in \Gamma(Q, g)$ mit $R^\alpha = R'$, also ist $\mathfrak{E}$ auch (R', s)-transitiv. Aus der (R, s)- und (R', s)-Transitivität von $\mathfrak{E}$ folgt jetzt wieder wie vorhin die (s, s)-Transitivität von $\mathfrak{E}$. Wegen $s^{\delta^{-2}} = g$ und $s^\delta = R$ ist $\mathfrak{E}$ dann (g, g)-transitiv und (R, R)-transitiv; wegen $R \not\mathrm{I} g$ folgt daraus ersichtlich, daß $\mathfrak{E}$ eine Moufangebene ist, und nach Korollar 4 ist $\mathfrak{E}$ dann desarguessch.

3.2. Sei $Q \neq g^\delta$ und $Q^\delta \neq g$.

3.2.1. Sei zuerst $P \mathrm{I} (Q + g^\delta)$ und $(g \cap Q^\delta) \mathrm{I} h$. Wäre $Q \mathrm{I} h$, so folgte jetzt $P \mathrm{I} g$, im Widerspruch zu $Q \neq P$, $g \neq h$. Entsprechend erhält man, $P \not\mathrm{I} g$. Sei $S = g \cap h$ und $r = P + Q$. Dann ist $(S, r) \neq (P, h)$ und $S^\delta = (g \cap h)^\delta = g^\delta + P = Q + P = r$ und $r^\delta = (Q + P)^\delta = Q^\delta \cap h = g \cap h = S$; also ist (S, r) ein ausgezeichnetes Punkt-Geradenpaar in bezug auf δ. Wegen $A_1 \neq S \neq A_2$ ist einerseits $g^\delta \not\mathrm{I} g$, andererseits ist $Q \mathrm{I} g$; also sind Q und g^δ verschiedene Punkte mit $Q + g^\delta = r$. Ganz ebenso sind g und Q^δ verschiedene Gerade mit $g \cap Q^\delta = S$. Aus der (Q, g)-Transitivität von $\mathfrak{E}$ folgt auch die (g^δ, Q^δ)-Transitivität, und beide zusammen ergeben offenbar die (R, s)-Transitivität für alle inzidenten Paare (R, s) mit $R \mathrm{I} r$ und $S \mathrm{I} s$ – also ist der Typ III der Behauptung unter diesen Umständen die kleinstmögliche Stufe.

3.2.2. Schließlich sei $P \not\mathrm{I} (Q + g^\delta)$ oder $(g \cap Q^\delta) \not\mathrm{I} h$. O.B.d.A. kann $P \not\mathrm{I} (Q + g^\delta)$ angenommen werden. Sei $R = h \cap (Q + g^\delta)$. Ist $\mathfrak{E}$ nun (Q, g)-transitiv, dann auch (g^δ, Q^δ)-transitiv, wegen $Q \neq g^\delta$ und $Q^\delta \neq g$ ist $\mathfrak{E}$ dann (R, s)-transitiv für eine Gerade s mit $R \mathrm{I} s$. Ist $s = h$, so sind wir in Fall 1.

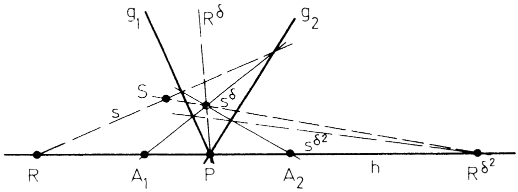


Fig. 8

Sei nun $s \neq h$. Wir dürfen wieder $R^\delta \neq s$ annehmen. Wegen $R \mathrm{I} h$, aber $R \neq P$ ist ohnehin $s^\delta \neq R$, aber $P \mathrm{I} R^\delta$ (vgl. Fig. 8). Wegen $P \not\mathrm{I} s$ ist $s^{\delta^2} \neq s$. Sei zuerst $R^{\delta^2} = R$. Aus der (R, s)-Transitivität und der (R, s^{δ^2})-Transitivität folgt, daß $\mathfrak{E}$ dann (R, R)-transitiv und schließlich auch (R^δ, R^δ)-transitiv ist. Wegen $R \not\mathrm{I} R^\delta$ ist dann $\mathfrak{E}$ eine Moufang-Ebene und nach Korollar 4 also desarguessch. Endlich sei $R^{\delta^2} \neq R$. Dann sind (R, s), $(R^{\delta^2}, s^{\delta^2})$, (s^δ, R^δ) drei verschiedene inzidente Punkt-Geradenpaare ohne weitere Inzidenzen, deren Punkte nicht kollinear und deren Geraden nicht konfluent sind (vgl. Fig. 8). Sei $S = s \cap (R^{\delta^2} + s^\delta)$. Dann ist $S \neq R$ und es gibt $\alpha \in \Gamma(R^{\delta^2}, s^{\delta^2})$ mit $s^{\delta\alpha} = S$. Ist $R^{\delta\alpha} \neq s$, so folgt aus der (R, s)-Transitivität, daß $\mathfrak{E}$ auch (S, S)-transitiv und insbesondere (S, s)-transitiv ist. Die (S, s)-Transitivität von $\mathfrak{E}$ folgt aber ebenso aus der Annahme, daß $R^{\delta\alpha} = s$ ist. Ist $\mathfrak{E}$ aber (R, s)- und (S, s)-transitiv, dann auch (s, s)- und also auch (s^δ, s^δ)-transitiv. Wie oben folgt jetzt, daß $\mathfrak{E}$ eine Moufangebene und damit desarguessch ist.

Damit sind jetzt alle Fälle erledigt.

Bemerkung 2. Für welche nicht-inzidenten Punkt-Geradenpaare kann $\mathfrak{E}$ noch zusätzlich transitiv sein? Aus der Liste der möglichen Lenz-Barlotti-Typen (vgl. z.B. [4], S. 126) geht hervor, daß höchstens noch Typ II.2 oder III.2 in Frage kommt, d.h. es handelt sich um ein einziges Paar (S, r) mit $P \mathrm{I} r$ und $S \mathrm{I} h$. Dann muß aber $(r^\delta, S^\delta) = (S, r)$ sein: Ist $\mathfrak{E}$ mit der (P, h)-Dualität δ vom Typ II oder III, dann ist $\mathfrak{E}$ zusätzlich (S, r)-transitiv höchstens für ein bezüglich δ ausgezeichnetes Punkt-Geradenpaar (S, r).

5. Beispiele

Beispiele distributiver Quasikörper mit kommutativer Multiplikation sind schon seit langem bekannt (man vergleiche die Zusammenstellung in [4], 5.3); daher ist auch an Ebenen vom Typ V mit einer (P, h)-Dualität kein Mangel.

Hier sollen nun Ebenen vom Typ II mit einer (P, h)-Dualität δ nachgewiesen werden. Dazu werden cartesische Gruppen $\mathfrak{K}$ mit kommutativer Addition und Multiplikation konstruiert, wobei $\mathfrak{K}$ als Menge mit der Menge $\mathbb{R}$ der reellen Zahlen übereinstimmt; auch die Addition in $\mathfrak{K}$

stimmt mit der in $\mathbb{R}$ überein, nur die Multiplikation in $\mathfrak{K}$ (geschrieben „$\circ$") wird neu erklärt. Unter diesen Umständen – die additive Gruppe von $\mathfrak{K}$ enthält kein Element der Ordnung 2 – ist es von Vorteil, eine etwas andere Lage des zu δ gehörigen $\mathfrak{D}$-Gestells zu den Bezugspunkten des Koordinatensystems zu wählen.

Die von uns konstruierten cartesischen Gruppen $\mathfrak{K}$ werden der folgenden Regel genügen:

$$(-x)\circ y = -(x\circ y) \qquad \text{für alle } x, y\in\mathfrak{K}.$$

Man weist nun leicht allgemein nach, daß für projektive Ebenen über cartesischen Gruppen mit $-1\neq 1$, deren Addition und Multiplikation kommutativ ist, und die überdies die obige Regel befolgen [7], die Abbildung δ, definiert durch

$$(x,y)^\delta = [-x, 1+y]; \qquad [x]^\delta = (-x),$$
$$[m,k]^\delta = (m, 1+k); \qquad (z)^\delta = [z],$$

eine Dualität von $\mathfrak{E}$ ist, derart, daß alle Punkte der Geraden $[1]$ und $[-1]$ absolute Punkte, und alle Geraden durch (1) und (-1) absolute Geraden sind (vgl. den ähnlichen Beweis zu Satz 7). Daher existiert in einer solchen projektiven Ebene die $(V, U+V)$-Dualität bezüglich $(1), (-1), [1], [-1]$. Zum Nachweis, daß $\mathfrak{K}(+,\circ)$ eine cartesische Gruppe ist (vgl. [7], S. 90), genügt es wegen der Kommutativität der Multiplikation, neben der Regel

$$1\circ x = x, \quad 0\circ x = 0 \qquad \text{für alle } x\in\mathfrak{K}$$

noch zu zeigen, daß die Funktion h mit $h(x) = -a\circ x + b\circ x$ für $a\neq b$ und $x\geqq 0$ streng monoton und stetig ist mit $\lim_{x\to\infty} h(x) = \pm\infty$. (Wegen $h(-x) = -h(x)$ folgt Entsprechendes für $x\leqq 0$.)

Konstruktion I. Sei φ eine auf dem Intervall $[1,\infty)$ definierte reellwertige, differenzierbare, monoton nicht-fallende Funktion mit $\varphi(1)=1$. Dann soll sein

$$a\circ b = \begin{cases} ab, & \text{falls } |a|\leqq 1 \text{ oder } |b|\leqq 1 \\ \operatorname{sign}(a)\operatorname{sign}(b)\,|a|^{\varphi(|b|)}\,|b|^{\varphi(|a|)} & \text{sonst.}\end{cases}$$

Damit ist eine stetige und kommutative Multiplikation erklärt mit $1\circ x = x$ und $0\circ x = 0$ für alle $x\in K$. O.B.d.A. sei nun $b>0$ und $|a|<b$. Die Funktion $x\mapsto b\circ x$ ist streng monoton wachsend mit $\lim_{x\to\infty} b\circ x = \infty$. Das-

[7] In Ebenen über solchen cartesischen Gruppen $\mathfrak{K}$ gilt, – sogar schon bei nichtkommutativer Multiplikation – daß $(m), (-m)$ eindeutig bestimmte harmonische Konjugierte in bezug auf U, V sind und U, V eindeutig bestimmte harmonische Konjugierte in bezug auf $(m), (-m)$, für alle $0\neq m\in K$.

selbe sieht man für $a \leqq 0$ sofort für die Funktion h mit $h(x) = -a \circ x + b \circ x$. Für $0 \leqq a \leqq 1 < b$ ist $h(x) \geqq (b-a)x$ und für $1 \leqq a < b$ ist $h(x) \geqq (b^{\varphi(x)} - a^{\varphi(x)})\, x^{\varphi(a)}$, so daß auch in diesen Fällen $\lim_{x \to \infty} h(x) = \infty$ ist. Schließlich erhält man nach kurzer Rechnung für diese beiden Fälle $h'(x) > 0$, was man nur für $x \geqq 1$ nachzuweisen braucht.

Seien a, b, c reelle Zahlen > 1 und $\varphi(c) > 1$. Dann ist

$$\begin{aligned} a \circ c + b \circ c &= a^{\varphi(c)}\, c^{\varphi(a)} + b^{\varphi(c)}\, c^{\varphi(b)} \\ &\leqq (a^{\varphi(c)} + b^{\varphi(c)})\, c^{\varphi(a+b)} \\ &< (a+b)^{\varphi(c)}\, c^{\varphi(a+b)} \\ &= (a+b) \circ c, \end{aligned}$$

also ist $\mathfrak{K}$ nicht distributiv.

Wählt man andererseits a, b, c als positive reelle Zahlen mit $b < 1$, $a > 1$, $c > 1$ derart, daß $a\,b < 1$ aber $b\,c > 1$ und sogar $\varphi(b\,c) > 1$ ist, dann wird

$$\begin{aligned} (a \circ b) \circ c &= a\,b\,c \\ &< a^{\varphi(b\,c)}\, (b\,c)^{\varphi(a)} \\ &= a \circ (b \circ c), \end{aligned}$$

also ist $\mathfrak{K}$ auch nicht assoziativ.

Schließlich sieht man aus der Definition von „$\circ$", daß das Gesetz $(-x) \circ y = -(x \circ y)$ für alle $x, y \in K$ erfüllt ist.

Konstruktion II (vgl. [4], S. 126, Fußnote). Sei f eine auf $\mathbb{R}$ definierte differenzierbare streng monoton steigende ungerade (reellwertige) Funktion, derart daß f' für $x \geqq 0$ monoton nicht-steigend, $f(1) = 1$, und $\lim_{x \to \infty} f(x) = \infty$ ist. Sei g die Umkehrfunktion von f. Dann soll sein

$$a \circ b = g(f(a)\, f(b)) \qquad \text{für alle } a, b \in \mathbb{R}.$$

Man sieht, daß die so definierte Multiplikation kommutativ und assoziativ ist mit $1 \circ x = x$ und $0 \circ x = 0$ für alle $x \in K$. Mit f ist auch g ungerade, also ist

$$\begin{aligned} (-x) \circ y &= g(f(-x)\, f(y)) \\ &= g(-f(x)\, f(y)) \\ &= -g(f(x)\, f(y)) \\ &= -(x \circ y). \end{aligned}$$

Für $a \leqq 0 < b$ sieht man wieder direkt, daß $h(x) = -a \circ x + b \circ x$ streng monoton steigend ist und gegen ∞ geht. O.B.d.A. sei jetzt $0 \leqq a < b$. Nun ist g eine schwach konkave Funktion mit $g(0) = 0$, d.h. für $u \geqq 0$, $v \geqq 0$ ist $g(u) + g(v) \leqq g(u+v)$, also ist für $x \to \infty$

$$\begin{aligned} h(x) &= -g(f(a)\, f(x)) + g(f(b)\, f(x)) \\ &\geqq g((f(b) - f(a))\, f(x)) \to \infty \end{aligned}$$

und für alle x

$$h'(x)=\left(\frac{-f(a)}{f'(g(f(a)f(x)))}+\frac{f(b)}{f'(g(f(b)f(x)))}\right)\cdot f'(x)>0.$$

Wir machen nun die (später nochmals benötigte) zusätzliche Voraussetzung, daß für $0\leqq x\leqq 1$ gilt $f(x)=qx$ $(q>0)$, daß aber andererseits g stellenweise echt konkav ist. Dann finden wir positive reelle Zahlen a, b, c mit $a+b\leqq 1$ und

$$\begin{aligned} a\circ c+b\circ c &= g(f(a)f(c))+g(f(b)f(c)) \\ &< g((f(a)+f(b))f(c)) \\ &= g(f(a+b)f(c)) \\ &= (a+b)\circ c, \end{aligned}$$

also ist $\mathfrak{K}$ nicht distributiv.

Nun soll noch ausgeschlossen werden, daß die projektive Ebene $\mathfrak{E}$ über $\mathfrak{K}$ vom Typ III ist. Das einzige in bezug auf δ ausgezeichnete Punkt-Geradenpaar von $\mathfrak{E}$ ist $(U, O+V)$. Daher genügt es nach Satz 8, zu zeigen, daß $\mathfrak{E}$ nicht $(O, O+U)$-transitiv ist.

Dazu setzen wir voraus, daß für $0\leqq x\leqq y\leqq 1+\varepsilon$ stets gilt $x\circ y=xy$, aber für hinreichend großes m die Kurve $y=m\circ x$ im Intervall $1\leqq x\leqq 1+\varepsilon$ echt konvex oder echt konkav ist. Im Falle der Konstruktion I ist das dadurch zu erreichen, daß man $\varphi(x)=1$ für $1\leqq x\leqq 1+\varepsilon$ und $\varphi(x)>1$ für $x>1+\varepsilon$ fordert. Im Falle der Konstruktion II setzt man wohl $f(x)=x$ für $0\leqq x\leqq 1+\varepsilon$ und verlangt für große m und $1<x\leqq 1+\varepsilon$, daß $f(mx)<f(m)f(x)$ gilt (z.B. setze man $f(x)=1+\varepsilon+\log(x-\varepsilon)$ für $x\geqq 1+\varepsilon$).

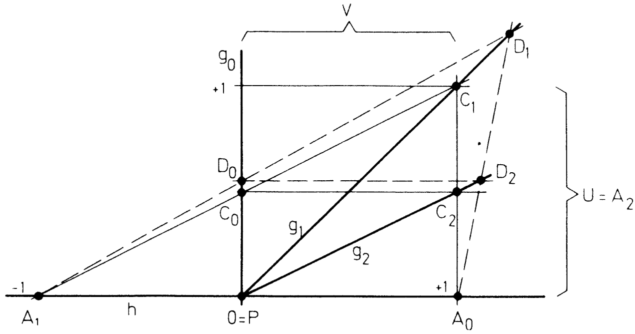


Fig. 9

Mit Verwendung der Bezeichnungen von Definition 4 sei $P=O$, $h=O+U$, $g_0=O+V$, $g_1=[1,0]$, $g_2=[\frac{1}{2},0]$, $A_1=(-1,0)$, $A_2=U$, $A_0=(C_1+C_2)\cap h=(1,0)$ (vgl. Fig. 9). Weiter sei $C_0=(0,\frac{1}{2})$, $C_1=(1,1)$, $C_2=(1,\frac{1}{2})$, $D_1=(1+\varepsilon, 1+\varepsilon)$, $D_0=g_0\cap(A_1+D_1)$, $D_2=g_2\cap(A_2+D_0)$. Dann sind alle Voraussetzungen des Desarguesschen $(O, O+U)$-Satzes (bezüglich A_i, g_i) erfüllt, sowohl in $\mathfrak{E}$ als auch nach den besonderen Voraussetzungen über $\mathfrak{K}$ in der Ebene über $\mathbb{R}$; daher liegen die Punkte A_0, D_1, D_2 in der reellen Ebene kollinear, aber nicht in $\mathfrak{E}$; denn die Gerade von $\mathfrak{E}$, welche A_0 und D_1 verbindet, ist – da es sich um eine große Steigung handelt – (in der reellen Ebene) eine echt konvexe oder konkave Kurve. Also ist $\mathfrak{E}$ nicht $(O, O+U)$-transitiv.

Da nach [7], Satz 45, S. 102, in projektiven Ebenen über cartesischen Gruppen $\mathfrak{K}$ die Assoziativität der Multiplikation in $\mathfrak{K}$ gleichwertig ist mit der $(U, O+V)$-Transitivität von $\mathfrak{E}$, liefert uns Konstruktion I Ebenen vom Typ II.1, während Konstruktion II Ebenen vom Typ II.2 liefert, – beide mit einer $(I, U+V)$-Dualität.

6. Der (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2

Dieser Abschnitt dient dem Beweis von

Satz 9. *Aus dem (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2 folgt der Satz von Pappos.*

Wir benötigen zuerst einige Hilfssätze.

Hilfssatz 1. *Aus dem (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2 folgt der Desarguessche (A_1, g_2)-Satz.*

Beweis. In $\mathfrak{E}$ gelte der (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2. Sei R ein von P und A_1 verschiedener Punkt mit $R \mathrm{I} h$, dann existiert in $\mathfrak{E}$ also die (P, h)-Dualität bezüglich A_1, R, g_1, g_2; diese werde mit δ_R bezeichnet. Wir zeigen, daß $\mathfrak{E}$ (A_1, g_2)-transitiv ist, indem wir für Punkte R, S von h, welche von A_1 und P verschieden sind, für $\alpha = \delta_R \delta_S^{-1}$ nachweisen, daß $\alpha \in \Gamma(A_1, g_2)$ und $R^\alpha = S$ gilt. Es ist $R^{\delta_R} = g_2 = S^{\delta_S}$, also $R^\alpha = S$. Ist s eine Gerade mit $A_1 \mathrm{I} s$, so ist nach Satz 1 nun $s^{\delta_R} = s \cap g_1 = s^{\delta_S}$, also $s^\alpha = s$; und ist T ein Punkt mit $T \mathrm{I} g_2$, so ist $T^{\delta_R} = T + A_1 = T^{\delta_S}$, also $T^\alpha = T$. Daher gilt $\alpha \in \Gamma(A_1, g_2)$.

Hilfssatz 2. *Ist δ eine (P, h)-Dualität und ist η die (P, h)-Dualität bezüglich A_i, g_i, so ist $\delta^{-1} \eta^{-1} \delta$ die (P, h)-Dualität bezüglich g_i^δ, A_i^δ.*

Beweis. Jedenfalls ist $\delta^{-1} \eta^{-1} \delta$ eine Dualität. Ist nun $Q \mathrm{I} A_i^\delta$, also $A_i \mathrm{I} Q^{\delta^{-1}}$, so ist $Q^{\delta^{\;1}}$ eine absolute Gerade in bezug auf η, d.h. es gilt $Q^{\delta^{-1}\eta} \mathrm{I} Q^{\delta^{-1}}$, woraus $Q \mathrm{I} Q^{\delta^{-1}\eta^{-1}\delta}$ folgt. Also ist Q absoluter Punkt in bezug auf $\delta^{-1} \eta^{-1} \delta$. Ferner ist $(g_i^\delta)^{\delta^{-1}\eta^{-1}\delta} = g_i^{\eta^{-1}\delta} = A_i^\delta$, $g_1^\delta + g_2^\delta = (g_1 \cap g_2)^\delta = P^\delta = h$, $A_1^\delta \cap A_2^\delta = (A_1 + A_2)^\delta = h^\delta = P$. Also ist $(P, h, g_i^\delta, A_i^\delta)$ das zu $\delta^{-1} \eta^{-1} \delta$ gehörige Gestell.

Hilfssatz 3. *Aus dem (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2 folgt der (P, h)-Satz von Pappos bezüglich A_1, g_1.*

Beweis. In $\mathfrak{E}$ gelte der (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2. Es ist nachzuweisen, daß zu vorgegebener Geraden s mit $g_1 \neq s \neq h$ und $P \mathrm{I} s$ und zu vorgegebenem Punkt R mit $A_1 \neq R \neq P$ und $R \mathrm{I} h$ in $\mathfrak{E}$ die (P, h)-Dualität bezüglich A_1, R, g_1, s existiert. Es sei δ_R wie im Beweis zu Hilfssatz 1 definiert. Dann ist $S = s^{\delta_R^{-1}}$ ein Punkt von $\mathfrak{E}$ mit $S \mathrm{I} h$ und $A_1 \neq S \neq P$, so daß auch δ_S existiert; ferner ist δ_S^{-1} die (P, h)-Dualität bezüglich A_1, S, g_2, g_1. Nun ist $g_2^{\delta_R} = A_1$, $g_1^{\delta_R} = R$, $A_1^{\delta_R} = g_1$, $S^{\delta_R} = s$. Nach Hilfssatz 2 ist daher $\delta_R^{-1} \delta_S \delta_R$ die (P, h)-Dualität von $\mathfrak{E}$ bezüglich A_1, R, g_1, s.

Beweis von Satz 9. In $\mathfrak{E}$ gelte der (P, h)-Satz von Pappos bezüglich A_1, g_1, g_2; nach Hilfssatz 3 gilt dann dieser Satz sogar bezüglich A_1, g_1. Mit Hilfssatz 1 folgt daraus, daß $\mathfrak{E}$ nun (A_1, g_2)- und (A_1, g_2')-transitiv für zwei untereinander und von h verschiedene Gerade g_2, g_2' mit $g_2 \cap g_2' = P$ ist, daher (A_1, P)- und insbesondere (A_1, h)-transitiv. In $\mathfrak{E}$ existiert die (P, h)-Dualität δ bezüglich A_1, A_2, g_1, g_2 (mit geeignet gewähltem A_2), und es ist $A_1^{\delta^2} = A_2$, $h^{\delta^2} = h$. Also ist $\mathfrak{E}$ auch (A_2, h)-transitiv und wegen $A_1 \neq A_2$ sogar (h, h)-transitiv. Nach Korollar 4 ist dann $\mathfrak{E}$ projektive Ebene über einem kommutativen Quasikörper $\mathfrak{K}$. Dabei ist das Koordinatenquadrupel O, U, V, E derart zu wählen, daß $V = P$, $U = A_1$, sowie etwa $O + V = g_1$ und $E + V = g_2$ gilt. Nach dem oben Bemerkten ist dann $\mathfrak{E}$ aber (U, V)-transitiv, was nach [7], Satz 46, S. 103 nun auch noch die Assoziativität der Multiplikation in $\mathfrak{K}$ erzwingt. Also ist $\mathfrak{K}$ ein (kommutativer) Körper und in $\mathfrak{E}$ gilt der Satz von Pappos.

Bemerkung 3. Den letzten Teil des Beweises von Satz 9 (Assoziativität der Multiplikation) kann man auch mit Hilfe von Satz 7 einsehen: Sind die Koordinaten wie oben gewählt, so besitzt $\mathfrak{E}$ die $(V, U + V)$-Dualität bezüglich (0), (a), $O + V$, $E + V$ für alle $0 \neq a \in \mathfrak{K}$, es gilt also die Gleichung

$$(a - a x) z = a z - (a z) x \qquad \text{für alle } a, x, z \in \mathfrak{K}.$$

Da $\mathfrak{K}$ distributiv ist mit kommutativer Multiplikation, folgt

$$(x a) z = (a x) z = (a z) x = x (a z) \qquad \text{für alle } x, a, z \in \mathfrak{K}.$$

Bemerkung 4. Hilfssatz 1 kann zugleich als Spezialisierung des Satzes von Hessenberg angesehen werden. Freilich wird hier nur der Desarguessche (Q, g)-Satz für $Q \not\mathrm{I} g$ ins Auge gefaßt. Um Entsprechendes für $Q \mathrm{I} g$ zu bringen, könnte man auch direkt beweisen, daß aus dem (P, h)-Satz von Pappos bezüglich A_1, g_1 der Desarguessche (P, h)-Satz folgt, indem man Satz 3 verwendet und benutzt, daß aus dem (P, h)-Satz von Desargues bezüglich A_1, g_1 schon der Desarguessche (P, h)-Satz folgt.

Literatur

1. Baer, R.: Homogeneity of projective planes. Amer. J. Math. **64**, 137—152 (1942).
2. Burn, R. P.: Bol quasifields and Pappus' theorem. Math. Z. **105**, 351—364 (1968).

3. Burn, R. P.: The uniqueness of the harmonic construction. Math. Z. **106**, 332—336 (1968).
4. Dembowski, P.: Finite geometries. New York: Springer 1968.
5. Jónsson, W. J.: Transitivität und Homogenität projektiver Ebenen. Math. Z. **80**, 269—292 (1963).
6. Pickert, G.: Der Satz von Pappos mit Festelementen. Arch. der Math. **10**, 56—61 (1959).
7. Pickert, G.: Projektive Ebenen. Berlin-Göttingen-Heidelberg: Springer 1955.

Dr. A. Herzer
Mathematisches Institut
Universität Mainz
D-6500 Mainz, Saarstraße 21
Bundesrepublik Deutschland

(Eingegangen am 21. Juni 1972)

Math. Z. 129, 259–267 (1972)
© by Springer-Verlag 1972

Automorphismen und Erzeugende für Gruppen mit einer definierenden Relation

Gerhard Rosenberger

1. Einleitung

A. Sei F_q eine freie Gruppe von endlichem Rang $q \geqq 1$, und sei Aut F_q die Automorphismengruppe von F_q. Sei G eine von q, aber nicht von $q-1$, Elementen erzeugte Gruppe (Rang$(G)=q$), und sei Aut G die Automorphismengruppe von G.

Wir nennen G eine *quasifreie Gruppe (vom Rang q)*, wenn Aut G epimorphes Bild von Aut F_q ist. Wir nennen G eine *fast quasifreie Gruppe (vom Rang q)*, wenn Aut G epimorphes Bild einer Untergruppe von Aut F_q ist. Insbesondere ist jede quasifreie Gruppe auch fast quasifreie Gruppe.

Gegenstand der vorliegenden Arbeit sind die Gruppen

$$F(n,g) .= \left\{ A_1, B_1, \ldots, A_g, B_g; \left(\prod_{i=1}^{g} [A_i, B_i] \right)^n = 1 \right\}$$

mit $n \geqq 1$, $g \geqq 1$.

Diese Gruppen sind auch deshalb von großem Interesse, weil sich $F(n,g)$ für $n \geqq 2$ als Fuchssche Gruppe mit kompaktem Fundamentalbereich vom Geschlecht g und für $n=1$ als Fundamentalgruppe einer orientierbaren geschlossenen Fläche vom Geschlecht g präsentieren läßt. Es sei noch bemerkt, daß $F(n,g)$ von $2g$, aber nicht von $2g-1$, Elementen erzeugt wird ([19]).

Die Gruppen $F(1,g)$ sind von Nielsen [12] und Zieschang ([18] und [19]) weitgehend untersucht worden. Es erscheint aber doch sinnvoll zu sein, nach einer Verallgemeinerung der von Nielsen und Zieschang erzielten Ergebnisse zu fragen.

Hier zeigen wir, daß die Gruppen $F(n,g)$ für $g=1$ quasifrei und für $g>1$ fast quasifrei sind (Satz 1 und Satz 2). Weiter charakterisieren wir die quasifreien Gruppen mit einer definierenden Relation (Satz 3).

Durch Satz 4 und Satz 5 lösen wir das Isomorphieproblem für die Gruppen $F(n,g)$, d.h. es ist nach endlich vielen Schritten entscheidbar, ob eine beliebige Gruppe mit einer definierenden Relation zu einer Gruppe $F(n,g)$ isomorph ist oder nicht. Satz 5 verallgemeinert das Resultat von Zieschang [19] über die Gruppen $F(1,g)$. Satz 4 zeigt, daß wir ein Resultat von Nielsen [10] über freie Gruppen vom Rang zwei auch für die Gruppen $F(n,1)$, $n \geqq 2$, aussprechen können, denn es gilt:

Zwei Elemente U, V von $F(n,1)$, $n \geqq 2$, sind genau dann Erzeugende von $F(n,1)$, wenn $[U,V]$ in $F(n,1)$ konjugiert ist zu $[A_1, B_1]^\varepsilon$ mit $\varepsilon = \pm 1$.

Beim Beweis von Satz 2 und Satz 5 greifen wir in starkem Maße auf die Ergebnisse von Nielsen und Zieschang über die Gruppen $F(1,g)$ zurück. Weil die Gruppe $F(1,1)$ abelsch ist, führen wir die Beweise für die Gruppen $F(n,1)$ gesondert.

B. Einige Definitionen und Vereinbarungen erweisen sich als nützlich. Es bedeute:

$F_q = \{a_1, \dots, a_q\}$ die von den $a_1, \dots, a_q$ erzeugte freie Gruppe.

$F_{2g} = \{a_1, b_1, \dots, a_g, b_g\}$ die von den $a_1, b_1, \dots, a_g, b_g$ erzeugte freie Gruppe.

$G = \{A_1, \dots, A_q\}$ die von den $A_1, \dots, A_q$ erzeugte Gruppe.

Aut G die Automorphismengruppe von G.

$I(G)$ die Menge der inneren Automorphismen von G.

G' die Kommutatorgruppe von G.

$G = \{A_1, \dots, A_q;\ R(A_1, \dots, A_q) = 1\}$ die von den $A_1, \dots, A_q$ erzeugte Gruppe mit der definierenden Relation $R(A_1, \dots, A_q) = 1$.

$[A,B] .= ABA^{-1}B^{-1}$ der Kommutator von A und B.

$$F(n,g) .= \left\{A_1, B_1, \dots, A_g, B_g;\ \left(\prod_{i=1}^{g} [A_i, B_i]\right)^n = 1\right\}$$

mit $n \geqq 1, g \geqq 1$.

$\bar{\varphi}_n$ der durch $a_i \mapsto A_i, b_i \mapsto B_i$ definierte natürliche Epimorphismus von F_{2g} auf $F(n,g)$.

$\bar{\psi}_n$ der durch $A_i \mapsto A_i$, $B_i \mapsto B_i$ definierte natürliche Epimorphismus von $F(n,g)$ auf $F(1,g)$.

Wir fassen eine Gruppe $G = \{A_1, \dots, A_q\}$ als epimorphes Bild der freien Gruppe $F_q = \{a_1, \dots, a_q\}$ unter dem durch $a_i \mapsto A_i$ definierten Epimorphismus $F_q \twoheadrightarrow G$ auf.

Einen Übergang von $(A_1, \dots, A_q)$ zu einem anderen Erzeugendensystem $(A'_1, \dots, A'_q)$ von G *nennen wir frei*, wenn es ein freies Erzeugendensystem $(a'_1, \dots, a'_q)$ von F_q gibt, so daß das Bild von a'_i bei dem durch $a_i \mapsto A_i$ definierten Epimorphismus von F_q auf G gleich A'_i ist (vgl. [19]). Durch $a_i \mapsto a'_i$ wird ein Automorphismus von F_q definiert, der aber keineswegs auf G einen Automorphismus zu induzieren braucht.

Wir identifizieren die $PSL(2,\mathbb{R})$ mit der Gruppe aller linearen Transformationen der oberen Halbebene auf sich. Es ist $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{E, -E\}$, d.h. die $PSL(2,\mathbb{R})$ besteht aus den Paaren $(W, -W)$ mit $W \in SL(2,\mathbb{R})$. Es ruft keine Mißverständnisse hervor, wenn wir kurz W statt $(W, -W)$ schreiben. Es bedeute:

Sp U die Spur von $U \in SL(2,\mathbb{R})$.

2. Automorphismen von $F(n, g)$ und die quasifreien Gruppen mit einer definierenden Relation

Lemma 1 ([10], vgl. auch [8] und [9]). *Sei $F_2=\{a, b\}$ eine freie Gruppe vom Rang zwei. Zwei Elemente u, v von F_2 sind genau dann Erzeugende von F_2, wenn $[u, v]$ in F_2 konjugiert ist zu $[a, b]^\varepsilon$ mit $\varepsilon=\pm 1$.*

Lemma 2 (vgl. auch [12] und [19]). *Sei $g \geqq 2$. Sei $H(F_{2g})$ die Untergruppe von $\operatorname{Aut} F_{2g}$, die von den $\alpha \in \operatorname{Aut} F_{2g}$ mit*

$$\alpha\left(\prod_{i=1}^{g}[a_i, b_i]\right)=h\left(\prod_{i=1}^{g}[a_i, b_i]^\varepsilon\right) h^{-1},$$

$h \in F_{2g}$ und $\varepsilon=\pm 1$, erzeugt wird (es ist $H(F_{2g}) \neq \operatorname{Aut} F_{2g}$). Sei K_{2g} der von $\prod_{i=1}^{g}[a_i, b_i]$ erzeugte Normalteiler von F_{2g}. Sei $N(F_{2g}) \subset I(F_{2g})$ der von den Automorphismen $\kappa\colon c \mapsto t c t^{-1}$, $c \in F_{2g}$ und $t \in K_{2g}$, erzeugte Normalteiler von $H(F_{2g})$. Dann induziert $\bar{\varphi}_1$ einen Epimorphismus $\varphi_1\colon H(F_{2g}) \twoheadrightarrow \operatorname{Aut} F(1, g)$ durch $(\varphi_1(\delta))(D).=\bar{\varphi}_1(\delta(d))$ für $d \in F_{2g}$, $\delta \in H(F_{2g})$ und $\bar{\varphi}_1(d)=D$, und es gilt die kurze exakte Sequenz

$$1 \rightarrow N(F_{2g}) \rightarrow H(F_{2g}) \xrightarrow{\varphi_1} \operatorname{Aut} F(1, g) \rightarrow 1 .$$

Beweis. Aus $\bar{\varphi}_1(d)=\bar{\varphi}_1(d')$ folgt $\bar{\varphi}_1(\delta(d))=\bar{\varphi}_1(\delta(d'))$ für $d, d' \in F_{2g}$, $\delta \in H(F_{2g})$, denn es ist $\delta(K_{2g}) \subset K_{2g}$ für $\delta \in H(F_{2g})$.

Damit induziert $\bar{\varphi}_1$ einen Homomorphismus $\varphi_1\colon H(F_{2g}) \rightarrow \operatorname{Aut} F(1,g)$ durch $(\varphi_1(\delta))(D).=\bar{\varphi}_1(\delta(d))$ für $\bar{\varphi}_1(d)=D$ und $\delta \in H(F_{2g})$. Nach [18; p. 159] ist φ_1 sogar ein Epimorphismus. Natürlich gilt $N(F_{2g}) \subset \operatorname{Kern}(\varphi_1) \subset I(F_{2g})$. Um $\operatorname{Kern}(\varphi_1) \subset N(F_{2g})$ zu zeigen, brauchen wir nur zu zeigen: Sei $\alpha \in I(F_{2g})$ mit $\alpha \neq 1_{\operatorname{Aut} F_{2g}}$ und $\alpha\colon c \mapsto s c s^{-1}$ für $c, s \in F_{2g}$, aber $s \notin K_{2g}$. Dann ist $\varphi_1(\alpha) \neq 1_{\operatorname{Aut} F(1, g)}$. Das ist aber erfüllt, denn es gilt

$$(\varphi_1(\alpha))(A_1)=\bar{\varphi}_1(s a_1 s^{-1}) \neq A_1 . \quad \text{q.e.d.}$$

Satz 1. *Die Gruppen $F(n, 1)=\{A, B;\ [A, B]^n=1\}$ sind quasifrei.*

Beweis. Nach [8; p. 169] ist die Aussage für $n=1$ richtig. Sei nun $n \geqq 2$. Aus $\bar{\varphi}_n(d)=\bar{\varphi}_n(d')$ folgt $\bar{\varphi}_n(\delta(d))=\bar{\varphi}_n(\delta(d'))$ für $d, d' \in F_2$, $\delta \in \operatorname{Aut} F_2$. Damit induziert $\bar{\varphi}_n$ einen Homomorphismus $\varphi_n\colon \operatorname{Aut} F_2 \rightarrow \operatorname{Aut} F(n, 1)$ durch $(\varphi_n(\delta))(D).=\bar{\varphi}_n(\delta(d))$ für $\bar{\varphi}_n(d)=D$ und $\delta \in \operatorname{Aut} F_2$.

Sei nun $\tilde{\delta} \in \operatorname{Aut} F(n, 1)$. Nach [8; p. 269] ist dann $[\tilde{\delta}(A), \tilde{\delta}(B)]=(T[A, B]^\varepsilon T^{-1})^k$, $\varepsilon=\pm 1$, $k \in \mathbb{N}$ mit $1 \leqq k \leqq n-1$ und $T \in F(n, 1)$. Nach [14; Lemma 3] gibt es Erzeugende u, v von F_2 mit

$$\tilde{\delta}(A)=\bar{\varphi}_n(u)=\bar{\varphi}_n(\delta(a))=(\varphi_n(\delta))(A)$$

und

$$\tilde{\delta}(B)=\bar{\varphi}_n(v)=\bar{\varphi}_n(\delta(b))=(\varphi_n(\delta))(B)$$

für ein $\delta \in \operatorname{Aut} F_2$, d.h. φ_n ist ein Epimorphismus. q.e.d.

Satz 2. *Die Gruppen $F(n,g)$, $g\geqq 2$, sind fast quasifrei, aber nicht quasifrei.*

Beweis. Für $n=1$ sind wir fertig. Sei nun $n\geqq 2$. Sei $H(F_{2g})$ wie in Lemma 2 gegeben. Aus $\bar{\varphi}_n(d)=\bar{\varphi}_n(d')$ folgt $\bar{\varphi}_n(\delta(d))=\bar{\varphi}_n(\delta(d'))$ für $d, d'\in F_{2g}$, $\delta\in H\ (F_{2g})$. Damit induziert $\bar{\varphi}_n$ einen Homomorphismus φ_n: $H(F_{2g})\to \operatorname{Aut} F(n,g)$ durch $(\varphi_n(\delta))(D). = \bar{\varphi}_n(\delta(d))$ für $\bar{\varphi}_n(d)=D$ und $\delta\in H(F_{2g})$.

Aus $\bar{\psi}_n(C)=\bar{\psi}_n(C')$ folgt $\bar{\psi}_n(\tilde{\delta}(C))=\bar{\psi}_n(\tilde{\delta}(C'))$ für $C, C'\in F(n,g)$, $\tilde{\delta}\in \operatorname{Aut} F(n,g)$. Damit induziert $\bar{\psi}_n$ einen Homomorphismus ψ_n: $\operatorname{Aut} F(n,g)\to \operatorname{Aut} F(1,g)$ durch $(\psi_n(\tilde{\delta}))(\bar{\psi}_n(C)). = \bar{\psi}_n(\tilde{\delta}(C))$ für $\tilde{\delta}\in \operatorname{Aut} F(n,g)$. Wir erhalten folgendes kommutative Diagramm

$$\begin{array}{ccccccc} & & \operatorname{Aut} F_{2g} \neq H(F_{2g}) & \xrightarrow{\varphi_n} & \operatorname{Aut} F(n,g) & & \\ & & & \searrow^{\varphi_1} /\!/\!/ & \downarrow \psi_n & & \\ 1 \to N(F_{2g}) & \to & H(F_{2g}) & \xrightarrow{\varphi_1} & \operatorname{Aut} F(1,g) & \to 1, & \end{array}$$

wobei $N(F_{2g})$ wie in Lemma 2 gegeben.

Jedenfalls ist ψ_n ein Epimorphismus.

Weiter ist für $\tilde{\delta}\in \operatorname{Aut} F(n,g)$ nach [8; 269]

$$\tilde{\delta}\left(\prod_{i=1}^{g}[A_i,B_i]\right)=\left(T\prod_{i=1}^{g}[A_i,B_i]^{\varepsilon}\,T^{-1}\right)^{k}, \tag{$*$}$$

$\varepsilon=\pm 1$, $k\in\mathbb{N}$ mit $1\leqq k\leqq n-1$ und $T\in F(n,g)$.

Wir sind fertig, wenn wir zeigen können, daß der Kern von ψ_n nur aus inneren Automorphismen von $F(n,g)$ besteht, denn dann ist Kern (ψ_n) epimorphes Bild von $N(F_{2g})$ und φ_n ein Epimorphismus.

Sei also $\tilde{\gamma}\in$ Kern (ψ_n). Es ist

$$\bar{\psi}_n(\tilde{\gamma}(A_i))=(\psi_n(\tilde{\gamma}))(\bar{\psi}_n(A_i))=\bar{\psi}_n(A_i)$$

und

$$\bar{\psi}_n(\tilde{\gamma}(B_i))=\bar{\psi}_n(B_i).$$

Im Zusammenhang mit $(*)$ bedeutet dies aber, daß $\tilde{\gamma}$ ein innerer Automorphismus von $F(n,g)$ ist (vgl. [8; 271]). q.e.d.

Bemerkung. $F(n,g)$ läßt sich für $n\geqq 2$ auch präsentieren als Gruppe

$$G=\left\{S, A_1, B_1, \ldots, A_g, B_g;\ S^n=S\prod_{i=1}^{g}[A_i,B_i]=1\right\}.$$

In [18] hat Zieschang unter anderem gezeigt, daß jeder Automorphismus von G von einem Automorphismus von F_{2g+1} induziert wird. Da G aber schon von $2g$ Elementen erzeugt wird, wollen wir bei dieser Präsentierung nicht von einer fast quasifreien Gruppe sprechen.

Satz 3. *Sei F_q eine freie Gruppe von endlichem Rang q. Sei G eine von q, aber nicht von $q-1$, Elementen erzeugte Gruppe mit einer definierenden Relation* (Rang $(G)=q$).

Genau dann ist G eine quasifreie Gruppe, wenn entweder

(i) $q=1$ *und* $G=\{A; A^n=1\}$ *für* $n=1, 2, 3, 4$ *oder* 6 *oder*

(ii) $q=2$ *und* $G=\{A, B; [A, B]^n=1\}$ *für* $n \geqq 1$ *ist.*

Beweis. Für $q=1$ ist die Behauptung evident. Für $q=2$ ist die Behauptung nach Satz 1 hinreichend. Sei $G=\{A_1, \ldots, A_q; R(A_1, \ldots, A_q)=1\}$ quasifrei und $q \geqq 2$.

Nach [8; pp. 165, 172 und 261] ist dann $R(a_1, \ldots, a_q)$ in F_q konjugiert zu $R(a_2, a_1, a_3, \ldots, a_q)^{\varepsilon}$, zu $R(a_q, a_1, \ldots, a_{q-1})^{\varepsilon}$, zu $R(a_1^{-1}, a_2, \ldots, a_q)^{\varepsilon}$ und zu $R(a_1 a_2, a_2, \ldots, a_q)^{\varepsilon}$, $\varepsilon= \pm 1$ (vgl. auch [9] und [11]). Das ist aber nur möglich, wenn $q=2$ und $R(a_1, a_2)$ in F_2 konjugiert zu $([a_1, a_2]^{\varepsilon})^n$, $\varepsilon= \pm 1$, $n \geqq 1$, ist. q.e.d.

Bemerkung. Es scheint sehr schwierig zu sein, die quasifreien Gruppen zu charakterisieren.

Weitere Beispiele für quasifreie Gruppen sind:

1) $G=\{A, B; A^2=B^2=(AB)^2=1\}$,

2) $F_q/F_{q(2)}$ und $F_q/F_{q(3)}$, wobei $F_{q(n)}$ der n-te Term der unteren Zentralreihe von F_q ist [1], [3] und [8]).

3) F_2/F_2'' und $F_2/F_2''(F_2')^m$, wobei F_2'' die zweite Kommutatorgruppe von F_2 und m eine Primzahl ist ([2] und [4]).

Dagegen gilt:

a) $F_q/F_{q(n)}$ ist für $n \geqq 4$ im allgemeinen keine quasifreie Gruppe ([1] und [3]),

b) F_q/F_q'' ist für $q \geqq 3$ im allgemeinen keine quasifreie Gruppe ([5] und [6]).

3. Erzeugende von $F(n, g)$

Lemma 3. *Sei* $F(n, 1)=\{A, B; [A, B]^n=1\}$ *für* $n \geqq 2$. *Seien* U, V *beliebige Erzeugende von* $F(n, 1)$. *Dann ist* $[U, V]^r=1$ *für ein* $r \geqq 2$.

Beweis. $F(n, 1)$ sei treu dargestellt als Fuchssche Gruppe vom Geschlecht eins mit den kanonischen Erzeugenden A und B (vgl. [14]). Es seien also ohne Einschränkung $A, B \in PSL(2, \mathbb{R})$ mit $[A, B]^n=1$, $n \geqq 2$.

Seien U, V beliebige Erzeugende von $F(n, 1)$, und sei ohne Einschränkung $0 \leqq \mathrm{Sp}\, U$, $\mathrm{Sp}\, V$. Jedenfalls ist dann natürlich sogar $2 \leqq \mathrm{Sp}\, U$, $\mathrm{Sp}\, V$, $|\mathrm{Sp}\, UV|$, $|\mathrm{Sp}\, UV^{-1}|$.

Es bezeichne:

$E_F := \{(R, S) \mid$ Es gibt einen freien Übergang von (U, V) zu $(R, S)\}$. Insbesondere ist $F(n, 1)=\{R, S\}$ und $\mathrm{Sp}[R, S]=\mathrm{Sp}[U, V]$ für $(R, S) \in E_F$

(vgl. Lemma 1 und [16]). Setze:

$$L_F. = \{(\mathrm{Sp}\,R, \mathrm{Sp}\,S, \mathrm{Sp}\,RS) \mid (R,S) \in E_F \text{ und } 2 \leqq \mathrm{Sp}\,R \leqq \mathrm{Sp}\,S \leqq \mathrm{Sp}\,RS\}.$$

Wegen $2 \leqq \mathrm{Sp}\,U$, $\mathrm{Sp}\,V$ ist L_F nicht leer. Setze:

$$A_F. = \{\mathrm{Sp}\,R \mid (\mathrm{Sp}\,R, \mathrm{Sp}\,S, \mathrm{Sp}\,RS) \in L_F\},$$

$$B_F. = \{\mathrm{Sp}\,S \mid (\mathrm{Sp}\,R, \mathrm{Sp}\,S, \mathrm{Sp}\,RS) \in L_F\}$$

und

$$C_F. = \{\mathrm{Sp}\,RS \mid (\mathrm{Sp}\,R, \mathrm{Sp}\,S, \mathrm{Sp}\,RS) \in L_F\}.$$

Sei $x. = \inf A_F$, $y. = \inf B_F$ und $z. = \inf C_F$.

Dann ist:

a) $2 \leqq x \leqq y \leqq z$,

b) $x+y+z \leqq \mathrm{Sp}\,R + \mathrm{Sp}\,S + \mathrm{Sp}\,RS$ für alle $(\mathrm{Sp}\,R, \mathrm{Sp}\,S, \mathrm{Sp}\,RS) \in L_F$ und

c) $xy - z \geqq z$ oder $xy - z \leqq 0$ unter Beachtung der Gleichung $\mathrm{Sp}\,RS^{-1} = \mathrm{Sp}\,R \cdot \mathrm{Sp}\,S - \mathrm{Sp}\,RS$.

Der Fall $xy - z \leqq 0$ kann nicht eintreten, denn sonst gäbe es $(R,S) \in E_F$ mit $2 \leqq \mathrm{Sp}\,R$, $\mathrm{Sp}\,S$ und $\mathrm{Sp}\,RS^{-1} \leqq -2$, d.h. $F(n,1)$ wäre freie Gruppe ([13] und [16]). Es ist also $xy - z \geqq z$.

Angenommen

$$\mathrm{Sp}[U,V] = (\mathrm{Sp}\,U)^2 + (\mathrm{Sp}\,V)^2 + (\mathrm{Sp}\,UV)^2 - \mathrm{Sp}\,U \cdot \mathrm{Sp}\,V \cdot \mathrm{Sp}\,UV - 2 \geqq 2.$$

Wegen der Diskretheit von $F(n,1)$ ist sogar $\mathrm{Sp}[U,V] > 2$ (vgl. [16]). Dann ist $x^2+y^2+z^2-xyz > 4$ und sogar $x > 2$, denn für $x = 2$ wäre $x^2+y^2+z^2-xyz = 4$. Damit gilt: $2z^2 \leqq xyz < x^2+y^2+z^2-4$, d.h. $z-y < x-2$. Es folgt:

$$0 < x^2 - 4 + (z-y)^2 - (x-2)\,yz < (x-2)(2x - yz),$$

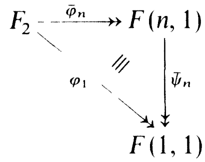
d.h. $yz < 2x$ im Widerspruch zu $2 \leqq x \leqq y \leqq z$.

Es ist daher $\mathrm{Sp}[U,V] < 2$. Es kann auch nicht $\mathrm{Sp}[U,V] \leqq -2$ sein, denn sonst wäre $F(n,1)$ freie Gruppe ([13] umd [16]). Damit ist insgesamt $-2 < \mathrm{Sp}[U,V] < 2$. Wegen der Diskretheit von $F(n,1)$ gibt es ein $r \in \mathbb{N}$ mit $[U,V]^r = 1$ ([7; p. 91]). Weiter ist $r \geqq 2$, da $F(n,1)$ sonst abelsch wäre. q.e.d.

Satz 4. *Sei* $F(n,1) = \{A, B; [A,B]^n = 1\}$ *für* $n \geqq 2$. *Zwei Elemente* U, V *von* $F(n,1)$ *sind genau dann Erzeugende von* $F(n,1)$, *wenn* $[U,V]$ *in* $F(n,1)$ *konjugiert ist zu* $[A,B]^\varepsilon$ *mit* $\varepsilon = \pm 1$.

Beweis. I) Seien U, V beliebige Erzeugende von $F(n,1)$. Dann ist jedenfalls $[U,V]^r = 1$ für ein $r \geqq 2$, wobei ohne Einschränkung $[U,V]^k \neq 1$ für $1 \leqq k \leqq r-1$ ist.

Wir haben folgendes kommutative Diagramm:



Dann sind $\overline{\psi}_n(U)$ und $\overline{\psi}_n(V)$ Erzeugende von $F(1,1)$, und es gibt Erzeugende u, v von F_2 mit $\overline{\psi}_n(U)=\varphi_1(u)=\overline{\psi}_n(\overline{\varphi}_n(u))$ und $\overline{\psi}_n(V)=\overline{\psi}_n(\overline{\varphi}_n(v))$.

Das bedeutet:

1) $\overline{\varphi}_n(u)$, $\overline{\varphi}_n(v)$ sind auch Erzeugende von $F(n,1)$,

2) $\overline{\varphi}_n(u)=UX$, $\overline{\varphi}_n(v)=VY$ für $X, Y\in \operatorname{Kern}(\overline{\psi}_n)$,

3) $[UX, VY]^n=1$ und $[UX, VY]^k\neq 1$ für $1\leqq k\leqq n-1$.

Damit ist $r=n$, $U=R\,\overline{\varphi}_n(u)\,R^{-1}$ und $V=R\,\overline{\varphi}_n(v)\,R^{-1}$ für ein $R\in F(n,1)$. Daher ist die Bedingung notwendig.

II) Seien U, V Elemente von $F(n,1)$ so, daß $[U, V]$ in $F(n,1)$ konjugiert ist zu $[A,B]^\varepsilon$ mit $\varepsilon=\pm 1$. Nach Lemma 1 gibt es Erzeugende u_1, v_1 von F_2 mit

$$1\neq [U, V]=\overline{\varphi}_n(h[a,b]^\varepsilon h^{-1})=\overline{\varphi}_n([u_1, v_1])=[U_1, V_1]$$

für $h\in F_2$, $U_1=\overline{\varphi}_n(u_1)$ und $V_1=\overline{\varphi}_n(v_1)$.

Dann sind nur folgende vier Fälle möglich:

1) $U=U_1 V_1^r$, $V=V_1$ $(\operatorname{mod} \operatorname{Kern}(\overline{\varphi}_n))$ für $n\geqq 2$,

2) $U=U_1$, $V=V_1 U_1^s$ $(\operatorname{mod} \operatorname{Kern}(\overline{\varphi}_n))$ für $n\geqq 2$,

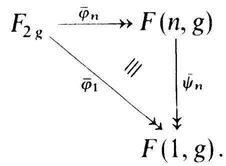
3) $U=V_1 U_1^r$, $V=U_1$ $(\operatorname{mod} \operatorname{Kern}(\overline{\varphi}_n))$ für $n=2$ und

4) $U=V_1$, $V=U_1 V_1^s$ $(\operatorname{mod} \operatorname{Kern}(\overline{\varphi}_n))$ für $n=2$.

Damit sind U und V Erzeugende von $F(n,1)$. q.e.d.

Satz 5. *Sei* $g\neq 3$. *Sei* $(U_1, \dots, U_{2g})$ *ein beliebiges Erzeugendensystem von* $F(n,g)$. *Dann gibt es einen freien Übergang von* $(A_1, B_1, \dots, A_g, B_g)$ *zu* $(U_1, \dots, U_{2g})$.

Beweis. Für $n=1$ ist dies Satz 6 von [19]. Für $g=1$ und $n\geqq 2$ ist dies Satz 4. Sei also nun $g=2$ oder $g\geqq 4$ und $n\geqq 2$. Wir haben folgendes kommutative Diagramm:



Sei $(U_1, V_1, \ldots, U_g, V_g)$ ein beliebiges Erzeugendensystem von $F(n,g)$. Dann gibt es nach Satz 6 von [19] ein Erzeugendensystem $(u_1, v_1, \ldots, u_g, v_g)$ von F_{2g} mit $\bar{\psi}_n(U_i) = \bar{\psi}_n(\bar{\varphi}_n(u_i))$ und $\bar{\psi}_n(V_i) = \bar{\psi}_n(\bar{\varphi}_n(v_i))$ für $i = 1, \ldots, g$, denn $(\bar{\psi}_n(U_1), \bar{\psi}_n(V_1), \ldots, \bar{\psi}_n(U_g), \bar{\psi}_n(V_g))$ ist ein Erzeugendensystem von $F(1,g)$.

Nun können wir (eventuell nach einem geeigneten freien Übergang) ohne Einschränkung $u_i = a_i$ und $v_i = b_i$, d.h. $\bar{\varphi}_n(u_i) = A_i$ und $\bar{\varphi}_n(v_i) = B_i$, annehmen.

Das bedeutet:

1) $U_i X_i = A_i$, $V_i Y_i = B_i$ für $X_i, Y_i \in \mathrm{Kern}\,(\bar{\psi}_n)$, $i = 1, \ldots, g$,

2) $\left(\prod_{i=1}^{g} [U_i X_i, V_i Y_i]\right)^n = 1$ und

3) $\left(\prod_{i=1}^{g} [U_i X_i, V_i Y_i]\right)^k \neq 1$ für $1 \leqq k \leqq n-1$.

Diese Aussagen sind analog zu den Aussagen im Fall $g = 1$, und wir können auf den Fall $g = 1$ zurückgreifen. Wegen $\bar{\psi}_n(U_i) = \bar{\psi}_n(U_i X_i)$ ist $U_i = R_i U_i X_i Z_i R_i^{-1}$ (mod Kern $(\bar{\varphi}_n)$) für $R_i \in \mathrm{Kern}\,(\bar{\psi}_n)$ und

$$Z_i = Z_i(U_1 X_1, V_1 Y_1, \ldots, U_g X_g, V_g Y_g).$$

Dann ist aber $Z_i = 1$, denn die Annahme $Z_i \neq 1$ führt nach Durchführung aller möglichen Kürzungen in $F(n,g)$ zu einer weiteren, von 2) unabhängigen definierenden Relation von $F(n,g)$. Weiter ist $U_i U_j = R_i U_i X_i R_i^{-1} R_j U_j X_j R_j^{-1}$ (mod Kern $(\bar{\varphi}_n)$) d.h. $R_i = R_j = .R$. Entsprechend ist $V_i = S V_i Y_i S^{-1}$ für ein $S \in \mathrm{Kern}\,(\bar{\psi}_n)$ und $U_i V_j = R U_i X_i R^{-1} S V_j Y_j S^{-1}$ (mod Kern $(\bar{\varphi}_n)$), d.h. $R = S$. Es ist also $U_i = R A_i R^{-1}$ und $V_i = R B_i R^{-1}$. Daher gibt es einen freien Übergang von

$$(A_1, B_1, \ldots, A_g, B_g) \quad \text{zu} \quad (U_1, V_1, \ldots, U_g, V_g). \quad \text{q.e.d.}$$

Bemerkungen. 1) Die Einschränkung $g \neq 3$ war nur durch den Beweis von Satz 6 aus [19] bedingt. Wahrscheinlich ist die Aussage von Satz 5 auch für $g = 3$ richtig.

2) In [15] und [17] wird unter Benutzung eines bestimmten Systems von Automorphismen von F_q (T-Transformationen) ein konstruktives Verfahren zur Lösung folgender Frage gegeben: Gibt es zu

$$a_1, \ldots, a_k, w_1, \ldots, w_k \in F_q, \qquad 1 < k \leqq q,$$

ein $\alpha \in \mathrm{Aut}\, F_q$ mit $w_i = \alpha(a_i)$, $i = 1, \ldots, k$? Auf Grund von Satz 5 kann man daher mit Hilfe dieses Verfahrens entscheiden, ob eine beliebige Gruppe mit einer definierenden Relation isomorph zu einer Gruppe $F(n,g)$, $g \neq 3$, ist oder nicht.

3) Wir formulieren folgendes allgemeine Problem: Sei $G=\{A_1, \ldots, A_q\}$ eine quasifreie Gruppe vom Rang q. Sei $(U_1, \ldots, U_q)$ ein beliebiges Erzeugendensystem von G. *Gibt es dann einen freien Übergang von* $(A_1, \ldots, A_q)$ *zu* $(U_1, \ldots, U_q)$?

Literatur

1. Andreadakis, S.: On the automorphisms of free groups and free nilpotent groups. Proc. London Math. Soc. (3) **15**, 239 – 268 (1965).
2. Bachmuth, S.: Automorphisms of free metabelian groups. Trans. Amer. Math. Soc. **118**, 93 – 268 (1965).
3. Bachmuth, S.: Induced automorphisms of free groups and free metabelian groups. Trans. Amer. Math. Soc. **122**, 1 – 17 (1966).
4. Bachmuth, S.: Automorphisms of a class of metabelian groups. Trans. Amer. Math. Soc. **127**, 284 – 293 (1967).
5. Chein, O.: *IA* automorphisms of free and free metabelian groups. Commun. Pure Appl. Math. **21**, 605 – 629 (1968).
6. Chein, O.: Induced automorphisms of free metabelian groups. (Preprint.)
7. Lehner, J.: Discontinous groups and automorphic forms. Math. Surveys N. VIII, Amer. Math. Soc. 1964.
8. Magnus, W., Karrass, A., Solitar, D.: Combinatorial group theory. New York: Interscience Publishers 1966.
9. Neumann, B. H.: Die Automorphismengruppe der freien Gruppen. Math. Ann. **107**, 367 – 386 (1932).
10. Nielsen, J.: Die Isomorphismen der allgemeinen, unendlichen Gruppe mit zwei Erzeugenden. Math. Ann. **78**, 385 – 397 (1918).
11. Nielsen, J.: Die Isomorphismengruppe der freien Gruppe. Math. Ann. **91**, 169 – 209 (1924).
12. Nielsen, J.: Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. Acta Math. **50**, 189 – 358 (1927).
13. Purzitsky, N.: Two generator discrete free products. Math. Z. **126**, 209 – 223 (1972).
14. Purzitsky, N., Rosenberger, G.: Two generator Fuchsian groups of genus one. Math. Z. **128**, 245 – 251 (1972).
15. Rapaport, E. S.: On free groups and their automorphisms. Acta Math. **99**, 139 – 163 (1958).
16. Rosenberger, G.: Fuchssche Gruppen, die freies Produkt zweier zyklischer Gruppen sind, und die Gleichung $x^2+y^2+z^2=xyz$. Math. Ann. **199**, 213 – 227 (1972).
17. Whitehead, J H. C.: On equivalent sets of elements in a free group. Ann. of Math. **37**, 782 – 800 (1936).
18. Zieschang, H.: Über Automorphismen ebener diskontinuierlicher Gruppen. Math. Ann. **166**, 148 – 167 (1966).
19. Zieschang, H.: Über die Nielsensche Kürzungsmethode in freien Produkten mit Amalgam. Inventiones math. **10**, 4 – 37 (1970).

Gerhard Rosenberger
D-2104 Hamburg 92
Haferacker 1
Bundesrepublik Deutschland

(Eingegangen am 18. Juli 1972)

Math. Z. 129, 269 – 277 (1972)
© by Springer-Verlag 1972

Stable Equivariant Bordism

Theodor Bröcker and Edward C. Hook

§ 0. Introduction

Given a compact Lie group G, there are at least two reasonable approaches one can use in defining a G-equivariant bordism theory. The first of these methods is the geometric approach initiated by Conner and Floyd in [3, 4], which has been extensively studied by Stong (cf., for exemple, [8]). The second approach is to construct an equivariant version of the appropriate Thom spectrum and then to define G-bordism theory to be a suitable homology theory with coefficients in this spectrum. This has been done by tom Dieck in [9, 10], where he considers the unitary case. We propose in this paper to investigate the relationship which exists between these two approaches; for simplicity, we restrict attention to the unoriented case, although similar results would hold in the unitary case. The main result is Theorem (4.1): Stabilized geometric equivariant bordism is isomorphic to homotopy theoretic equivariant bordism.

The second-named author would like to take this opportunity to express his gratitude to Prof. R. E. Stong for many helpful conversations and a great deal of encouragement.

§ 1. Homotopy-Theoretic Equivariant Bordism

Let G be a compact Lie Group. Let $RO(G)$ denote the set of real finite dimensional orthogonal representations of G. The set $RO(G)$ is partially ordered by $V<W$ if V is isomorphic to some G-submodule of W. If $V\in RO(G)$, let $|V|=$dimension of V, and $RO_k(G)=\{V\in RO(G) \mid |V|=k\}$.

If $V\in RO(G)$, let $D(V)$, $S(V)$ denote the unit disc and unit sphere in V, and $\Sigma(V)=D(V)/S(V)$ the quotient space, with base point $S(V)/S(V)$.

Let W be any real orthogonal representation of G (possibly infinite dimensional) and let $BO_n(W)$ be the Grassmannian of n-dimensional subspaces of W, with the G-action induced by the linear action on W. There is also the space

$$EO_n(W)=\{(V,x)\in BO_n(W)\times W \mid x\in V\}$$

which is the total space of the "tautological" n-plane bundle $\gamma^n(W)$ over $BO_n(W)$, and there is an action of G on $EO_n(W)$ by bundle maps covering the action on $BO_n(W)$ and such that the projection is equivariant; $\gamma^n(W)$ is a G-vector bundle. Moreover the unit disc bundle $D\gamma^n(W)$ and its boundary sphere bundle $S\gamma^n(W)$ are equivariant subspaces of $EO_n(W)$, such that we have a G-action on the Thom space

$$MO_n(W) := M\gamma^n(W) := D\gamma^n(W)/S\gamma^n(W).$$

If $|W| = n$, then $MO_n(W)$ is just the sphere $\Sigma(W)$.

We now define $\mathbb{R}^\infty(G)$ to be the orthogonal direct sum of countably many copies of each of the irreducible finite dimensional orthogonal representations of G with the obvious action of G. Then, for any non negative integer n, the above construction determines an object MO_n^G in the category $\mathrm{Top}_0(G)$ of pointed G-spaces. The bundle

$$\gamma^n\colon\ EO_n^G := EO_n(\mathbb{R}^\infty(G)) \to BO_n(\mathbb{R}^\infty(G)) =: BO_n^G$$

is known to be a universal equivariant n-plane bundle in the category of G-spaces.

If $W \in RO_k(G)$, one has a suspension map

$$m_{n,k}\colon\ \Sigma(W) \wedge MO_n^G \to MO_{n+k}^G$$

induced by the Whitney sum. Note that

$$MO_{n+k}^G = MO_{n+k}(\mathbb{R}^\infty(G)) = MO_{n+k}(W \oplus \mathbb{R}^\infty(G)).$$

The spaces MO_n^G together with the suspension maps $m_{n,k}$ constitute a G-spectrum, denoted by MO^G.

We may now define the *homotopy-theoretic G-bordism groups* of a space X in $\mathrm{Top}_0(G)$ to be the groups

$$\tilde{N}_n^G(X) := \varinjlim\, [\Sigma V, X \wedge MO_{|V|+n}^G]_0^G, \quad V \in RO(G),$$

where $[\ ,\ ]_0^G$ denotes equivariant homotopy classes of pointed maps, and the limit is taken over the direct system indexed by the partially ordered set $RO(G)$ and the maps induced by suspension.

These are the spectral homology groups of X with coefficients in MO^G, as introduced by tom Dieck [10]. They constitute an equivariant homology theory in the sense of Bredon [1], but in addition have *suspension isomorphisms* for all suspensions with linear G-action; i.e. if $V \in RO_k(G)$, there is a canonical isomorphism

(1.1) $$\sigma(V)\colon\ \tilde{N}_n^G(X) \to \tilde{N}_{n+k}^G(\Sigma(V) \wedge X)$$

(see [2, IV] for the completely analogous proofs in the non equivariant case). Moreover the graded group $N_*^G(X)$ is a natural N_*-module, and the suspensions are maps of N_*-modules (N_* is the unoriented bordism ring).

§ 2. Stable Equivariant Bordism

For any pair of G-spaces (X, A) one defines $\mathfrak{N}_*^G(X, A)$, the G-equivariant bordism of (X, A), as follows (see [8] for a more detailed account): A singular G-manifold of (X, A) is a pair (M, f) where M is a compact differentiable G-manifold with boundary, and $f\colon (M, \partial M) \to (X, A)$ is an equivariant map. Two singular G-manifolds (M, f), (M', f') are bordant, if there is a triple (V, V_0, F) where V is a compact differentiable G-manifold with boundary, and ∂V is the union of the invariant regularly imbedded G-manifolds M, V_0, M', with $M \cap V_0 = \partial M$; $M' \cap V_0 = \partial M'$; $M \cap M' = \emptyset$; $(M \cup M') \cap V_0 = \partial V_0$; and $F\colon (V, V_0) \to (X, A)$ is an equivariant map extending f on M and f' on M'. Bordism is an equivalence relation, and on the set $\mathfrak{N}_*^G(X, A)$ of equivalence classes one has a group structure, which is induced by disjoint union of manifolds.

G-equivariant bordism also defines an equivariant homology theory in the sense of Bredon [1]; this theory does not have suspension isomorphisms for suspension with non trivial G-action (e.g. for $G = \mathbb{Z}_2$). One does, however, have a *suspension homomorphism*

$$\sigma(V)\colon\ \tilde{\mathfrak{N}}_n^G(X) \to \tilde{\mathfrak{N}}_{n+|V|}^G(\Sigma(V) \wedge X) \tag{2.1}$$

assigning to $f\colon (M, \partial M) \to (X, *)$ the class of

$$(D(V) \times M, \partial(D(V) \times M)) \xrightarrow[f \times 1]{} (D(V) \times X, S(V) \times X \cup D(V) \times *)$$
$$\xrightarrow[\text{pr}]{} (\Sigma V \wedge X, *).$$

Obviously one has $\sigma(V \oplus W) = \sigma(W) \circ \sigma(V)$, hence for a pointed G-space X one has the direct system, indexed by $RO(G)$, of N_*-modules

$$\tilde{\mathfrak{N}}_*^G(\Sigma(V) \wedge X), \qquad V \in RO(G),$$

and suspension homomorphisms, and we define the *stable G-equivariant bordism group*

$$\tilde{\mathfrak{N}}_n^{G:S}(X) := \varinjlim \mathfrak{N}_{n+|V|}^G(\Sigma(V) \wedge X), \qquad V \in RO(G).$$

These groups also form an equivariant homology theory which, by construction, has suspension isomorphisms for all suspensions with linear G action. For pairs of G-spaces one has an analogous stabilization, and the suspension isomorphism take the form

$$\sigma(V)\colon\ \mathfrak{N}_*^{G:S}(X, A) \cong \mathfrak{N}_{*+|V|}^{G:S}(D(V) \times X, D(V) \times A \cup S(V) \times X).$$

Let $\xi: E \to X$ be a k-dimensional G-vector bundle and let $A \subset X$; then we have a natural *Thom homomorphism*

$$\tau(\xi): \mathfrak{N}_*^G(X, A) \to \mathfrak{N}_{*+k}^G(D(E), D(E|A) \cup S(E))$$

defined as follows: If $f: (M, \partial M) \to (X, A)$ represents a bordism element, then one has the induced bundle map

$$\begin{array}{ccc} f^*E & \xrightarrow{\bar{f}} & E \\ {\scriptstyle f^*\xi}\downarrow & & \downarrow{\scriptstyle \xi} \\ M & \xrightarrow{f} & X; \end{array}$$

$\tau(\xi)[M, f]$ is represented by

$$(D(f^*E), \partial(D(f^*E))) \xrightarrow{\bar{f}} (D(E), D(E|A) \cup S(E)).$$

If $\xi: E \to X$ and $\xi': E' \to X$ are G-vector bundles, then $\xi \oplus \xi'$ is the composite $\xi^* E' \to E \xrightarrow{\xi} X$ and

(2.2) $$\tau(\xi \oplus \xi') = \tau(\xi^* \xi') \circ \tau(\xi).$$

If $\pi: V \times X \to X$ is the trivial bundle with $V \in RO(G)$, then $\tau(\pi) = \sigma(V)$ is the suspension homomorphism. So by (2.2) in particular the Thom homomorphisms are compatible with suspensions, and passing to the limit we have a Thom homomorphism

(2.3) $$\tau(\xi): \mathfrak{N}_*^{G:S}(X, A) \to \mathfrak{N}_{*+k}^{G:S}(D(E), D(E|A) \cup S(E)).$$

(2.4) **Lemma.** *Let ξ be a G-vector bundle over X which is stably invertible. Then the Thom homomorphism* (2.3) *is an isomorphism.*

Proof. One has formula (2.2) also for the stable Thom homomorphism. If ξ' is an inverse bundle of ξ, then $\tau(\xi \oplus \xi')$ is an isomorphism, being a suspension. Thus $\tau(\xi)$ has a left inverse, and $\tau(\xi^* \xi')$ has a left and a right inverse, so $\tau(\xi^* \xi')$ is an isomorphism and $\tau(\xi)$ is an isomorphism. □

The bundle ξ is invertible in particular if X is compact, but in fact it is sufficient to assume that X is a limit of a sequence of closed subspaces X_i, such that $\xi | X_i$ is stably invertible, for stable bordism is compatible with direct limits. So we have a Thom isomorphism for every reasonable G-bundle, in particular for the universal bundle γ^n. In the absolute case one can pass to quotients to get a Thom isomorphism

(2.5) $$\tau(\xi): \mathfrak{N}_*^{G:S}(X) \to \mathfrak{N}_{*+k}^{G:S}(D(E), S(E)) \cong \tilde{\mathfrak{N}}_{*+k}^{G:S}(M(\xi)).$$

§ 3. The Pontrjagin-Thom Construction

As noted by tom Dieck [10], there is an equivariant *Pontrjagin-Thom construction*

(3.1) $$\Phi\colon \tilde{\mathfrak{N}}_*^G(-)\to \tilde{N}_*^G(-)$$

which will be shown to stabilize to a transformation

(3.2) $$\Phi^S\colon \tilde{\mathfrak{N}}_*^{G:S}(-)\to \tilde{N}_*^G(-).$$

The construction of Φ runs as follows:

Let $f\colon (M,\partial M)\to(X,*)$ represent an element of $\mathfrak{N}_*^G(X,*)$. Choose an equivariant imbedding $e\colon M\to D(V)$, for some $V\in RO(G)$ (see [6, 7]). Then M's tangent bundle $\tau(M)$ is a subbundle of $M\times V$, and the normal bundle ν of the imbedding e is defined to be the orthogonal complement of $\tau(M)$ in this bundle. Then a small disc bundle $D\nu$ of ν can be identified with a compact subset of $D(V)$, called a "tubular neighbourhood" although it is not a neighbourhood, and there is the usual collapsing map

$$k\colon \Sigma(V)=D(V)/S(V)\to D\nu/(D(\nu|\partial M)\cup S\nu).$$

On the other hand ν has a classifying map $u\colon M\to BO_{|V|-n}^G$ covered by $\bar{u}\colon D\nu\to D\gamma^{|V|-n}$, and the map $(f\circ\nu,\bar{u})\colon D\nu\to X\times D\gamma^{|V|-n}$ induces an equivariant map $D\nu/(D(\nu|\partial M)\cup S\nu)\to X\wedge MO_{|V|-n}^G$; if we compose this with the collapsing map k, we get a map which represents the element $\Phi[M,f]\in\tilde{N}_*^G(X,*)$.

One shows as in the classical case, that we have produced in this manner a well-defined natural transformation $\Phi\colon \tilde{\mathfrak{N}}_*^G(-)\to\tilde{N}_*^G(-)$, which preserves degrees and respects the N_*-module structures present.

In order to be able to stabilize this transformation we have to show:

(3.3) **Lemma.** *The natural transformation Φ is compatible with the suspension homomorphisms* (2.1) *and* (1.1).

Proof. We have to follow an element $[M,f]\in\tilde{\mathfrak{N}}_m^G(X)$ around the diagram

$$\begin{array}{ccc}\tilde{\mathfrak{N}}_m^G(X) & \xrightarrow[\sigma(W)]{} & \tilde{\mathfrak{N}}_{m+k}^G(\Sigma(W)\wedge X)\\ \downarrow{\scriptstyle\Phi} & & \downarrow{\scriptstyle\Phi}\\ \tilde{N}_m^G(X) & \xrightarrow[\sigma(W)]{} & \tilde{N}_{m+k}^G(\Sigma(W)\wedge X),\qquad W\in RO_k(G).\end{array}$$

If one constructs $\Phi[M,f]$ by means of an imbedding $e\colon M\to D(V)$, one may use the imbedding $1\times e\colon D(W)\times M\to D(W)\times D(V)=D(W\oplus V)$ for the construction of $\Phi\sigma(W)[M,f]$, and then the verification is straight forward. □

As a consequence we get by passing to the limit a natural transformation of N_*-modules

$$\Phi^S\colon \tilde{\mathfrak{N}}_*^{G:S}(-)\to \tilde{N}_*^G(-),$$

which preserves degrees and is compatible with suspension isomorphisms.

Remark. Our presentation of the Pontrjagin-Thom construction in this section differs from that in [5], chiefly in being more flexible, though both result in the same natural transformation.

§ 4. The Isomorphism Theorem

Call a pointed G-space *admissible*, if the inclusion $*\subset X$ of the basepoint is an equivariant cofibration. Then our main result is

(4.1) **Theorem.** *For any compact Lie group G and any admissible G-space X the natural transformation*

$$\Phi^S\colon \mathfrak{N}_*^{G:S}(X)\to \tilde{N}_*^G(X)$$

is an isomorphism.

Proof. One may suppose that X is compact because both theories are determined by their values on compact spaces, and one may suppose that the basepoint is disjoint, by looking at the exact sequences of the cofibration

$$\{*\}\cup\{+\}\subset X^+\to X.$$

We construct an inverse transformation Ψ for Φ^S using the following idea: By definition the homotopy-bordism of X^+ may be considered as stable equivariant homotopy groups of the spectrum $MO^G\wedge X^+$. If we denote such homotopy groups by $\pi_*^{G:S}(-)$ – without defining them – the transformation Ψ will be defined by the following diagram:

$$\begin{array}{ccc}
\pi_*^{G:S}(X^+\wedge MO^G) & \xrightarrow{\kappa} & \tilde{\mathfrak{N}}_*^{G:S}(X^+\wedge MO^G)\\
\Vert & & \downarrow{\scriptstyle \tau^{-1}}\\
N_*^G(X) & & \mathfrak{N}_*^{G:S}(X\times BO^G)\\
& {\scriptstyle\Psi}\searrow \quad \swarrow{\scriptstyle \mathrm{pr}_{1*}} & \\
& \mathfrak{N}_*^{G:S}(X) &
\end{array}$$

here κ is the canonical map which looks at a sphere as a particular manifold (which bounds when projected to the point), and τ is essentially the Thom isomorphism for the bundles $X\times\gamma^n$, where γ^n is the universal G-bundle.

For a more explicite construction, suppose $\alpha\in\tilde{N}_n^G(X^+)$ is represented by $a\colon (D(V), S(V))\to(X^+\wedge M\gamma^{|V|-n}(Q), +)$ for some $V, Q\in RO(G)$; here we use the compactness of $D(V)$ to replace $\mathbb{R}^{\infty}(G)$ by the finite-dimensional representation Q (this in fact is unnecessary by the remark after (2.4)). The map a then also represents an element

$$\kappa(\alpha)=[D(V),a]\in\tilde{\mathfrak{N}}_{|V|}^{G:S}(X^+\wedge M\gamma^{|V|-n}(Q)).$$

The space in brackets is the Thom space of the bundle $\xi:=X\times\gamma^{|V|-n}$; thus we can apply to the element $\kappa(\alpha)$ the composite

$$\tilde{\mathfrak{N}}_{|V|}^{G:S}(M(\xi))\xrightarrow[\tau(\xi)^{-1}]{}\mathfrak{N}_n^{G:S}(X\times BO_{|V|-n}(Q))\xrightarrow[\mathrm{pr}_{1*}]{}\tilde{\mathfrak{N}}_n^{G:S}(X^+)$$

where $\tau(\xi)$ is the Thom isomorphism (2.5). Then it is easy to verify that the element $\Psi(\alpha):=\mathrm{pr}_{1*}\circ\tau(\xi)^{-1}\circ\kappa(\alpha)$ only depends on α and not on the choices involved in the definition, so that we have a well-defined degree-preserving function

$$\Psi\colon\ \tilde{N}_*^G(X^+)\to\tilde{\mathfrak{N}}_*^{G:S}(X^+)$$

which, it is claimed, inverts Φ^S.

It is easily seen that $\Psi\circ\Phi^S$ is the identity of $\mathfrak{N}_*^{G:S}(X)$. For the proof let $[M^n,f]\in\mathfrak{N}_n^{G:S}(X)$ where we may assume that the range of f is actually X, by passing to the appropriate suspension. To obtain $\Phi^S[M^n,f]$ we chose an imbedding $e\colon M\to D(V)$ with normal bundle v which has a bundle map $\bar{u}\colon Dv\to D\gamma^{|V|-n}(Q)$ for some $Q\in RO(G)$. These define the map of pairs

$$\begin{aligned}(f\circ v,\bar{u})\colon\ &(Dv, D(v\,|\,\partial M)\cup Sv)\\ &\to(X\times D\gamma^{|V|-n}(Q),\ \{*\}\times D\gamma^{|V|-n}(Q)\cup X\times S\gamma^{|V|-n}(Q)),\end{aligned}$$

and there is the canonical projection

$$\begin{aligned}p\colon\ &(X\times D\gamma^{|V|-n}(Q),\ \{*\}\times D\gamma^{|V|-n}(Q)\cup X\times S\gamma^{|V|-n}(Q))\\ &\qquad\to(X\wedge M\gamma^{|V|-n}(Q),*).\end{aligned}$$

Now in $\mathfrak{N}_n^{G:S}(M,\partial M)$ there is the fundamental class ι, represented by the identity of M, and $\tau(v)(\iota)\in\mathfrak{N}_{|V|}^{G:S}(Dv, D(v\,|\,\partial M)\cup Sv)$ is also the fundamental class. Therefore by an obvious bordism [3, p. 12f.] the element

$$p_*\circ(f\circ v,\bar{u})_*\,\tau(v)(\iota)\quad\text{represents }\kappa\Phi[M,f],$$

and by naturality of the Thom isomorphism

$$\tau(\xi)^{-1}(p_*\circ(f\circ v,\bar{u})_*\,\tau(v)(\iota))=\tau(\xi)^{-1}\,\kappa\Phi[M,f]$$

is represented by

$$(f,\bar{u}\,|\,M)\colon\ (M,\partial M)\to X\times BO_{|V|-n}(G).$$

This establishes that $\Psi\circ\Phi^S$ is the identity.

We now consider the composition $\Phi^S\circ\Psi$. Given $\alpha\in\tilde{N}_n^G(X^+)$ represented by $f\colon(DV,SV)\to(M\xi,*)$, we can find a G-manifold with boundary

$L \subset \operatorname{Int} D(V)$ of dimension $|V|$, such that $f^{-1}(B\xi) \subset L - \partial L$ and $f^{-1}(*) \cap L = \varnothing$ (see [3,(3.1)]). If we vary f within its G-homotopy class, we get a commutative diagram

$$
\text{(4.2)}\qquad
\begin{array}{ccccc}
 & & (D(V), S(V)) & & \\
 & & \downarrow & \searrow^{f} & \\
(\Sigma V, *) & \longrightarrow & (L/\partial L, *) & \longrightarrow & (M\xi, *) \\
 & & \uparrow & & \uparrow{\scriptstyle q} \\
 & & (L, \partial L) & \xrightarrow[f|L]{} & (D\xi, S\xi),
\end{array}
$$

therefore $\kappa[D(V), f] = \kappa[L, q \circ f|L] \in \mathfrak{N}_{|V|}^{G:S}(M\xi)$.

Now let $\xi \oplus \xi' = \pi: Q \times X \to X$ for $Q \in RO(G)$; then $\sigma(Q) = \tau(\xi^* \xi') \circ \tau(\xi)$ or equivalently

$$
\text{(4.3)}\qquad \sigma(Q) \circ \tau(\xi)^{-1} = \tau(\xi^* \xi').
$$

The right side of this equation describes $\tau(\xi)^{-1}$ up to suspension. We use this fact to define the manifold M to be $D((f|L)^* \xi^* \xi')$ which comes provided with an equivariant map into $D(Q) \times B(\xi)$; if we compose this map with the projection onto $D(Q) \times X$ and the canonical map into $\Sigma(Q) \wedge X^+$, we receive an equivariant map h: $M \to \Sigma(Q) \wedge X^+$, and by definition of Ψ and (4.3) $\sigma(Q) \circ \Psi(\alpha) = [M, h] \in \tilde{\mathfrak{N}}_{n+|Q|}^{G:S}(\Sigma(Q) \wedge X^+)$, so that to compute $\sigma(Q) \circ \Phi^S \circ \Psi(\alpha)$, we have to apply the Pontrjagin-Thom construction to the map h.

For this purpose, note first that M is equipped with an equivariant imbedding into

$$
D((f|L)^* \xi^* (\xi \oplus \xi')) = D(Q) \times L \subset D(Q) \times D(V) = D(Q \oplus V),
$$

so we may use this imbedding in performing the construction. Moreover, a "tubular neighbourhood" of this imbedding is just $D(Q) \times L$ and this is the "neighbourhood" we employ in the construction. Now consider the diagram

$$
\text{(4.4)}\qquad
\begin{array}{ccccccc}
D(Q) \times L & \xrightarrow[1 \times f|L]{} & D(Q) \times D(\xi) = D(Q) \times X \times D\gamma^{|V|-n}(Q) & & & & \\
\downarrow & & \downarrow \qquad\qquad \swarrow \qquad \searrow^{\mathrm{pr}_3} & & & & \\
M & \longrightarrow & D(\xi^* \xi') = D(Q) \times B\xi & & D\gamma^{|V|-n}(Q) & & \\
\downarrow & & \downarrow \qquad\qquad\qquad \downarrow & & \downarrow & & \\
L & \xrightarrow[f|L]{} & D\xi \quad \xrightarrow[\xi]{} \quad B\xi & \xrightarrow[\mathrm{pr}_2]{} & BO_{|V|-n}(Q) & &
\end{array}
$$

in which the unnamed maps are either bundle projections or induced by the pullback construction. This diagram is commutative and the map $M \to BO_{|V|-n}(Q)$ classifies the normal bundle of M in $D(Q) \times D(V)$. By an examination of diagrams (4.2) and (4.4), we eventually see that the Pontrjagin-Thom construction applied to the map h yields the "Q-fold" suspension of the map with which we began. Equivalently, we have shown that $\sigma(Q) \circ \Phi^S \circ \Psi(\alpha) = \sigma(Q)(\alpha)$; since $\sigma(Q)$ is an isomorphism this completes the proof. □

The way transversality was circumvented here may have many more applications in equivariant bordism theory.

Theorem (4.1) was proved in [5] for the special case of the group $\mathbb{Z}_2$, using very different methods. A proof of the present version was later discovered and shown to tom Dieck, among others. He and the first-named author raised objections to this proof, concerned with assumed stability properties of homotopy theory which led to false consequences in the case $G = \{1\}$. Fortunately, the first named author was able to circumvent the need for these assumptions.

References

1. Bredon, G.E.: Equivariant cohomology theories. Lecture Notes in Mathematics, vol. 34, Berlin-Heidelberg-New York: Springer 1967.
2. Bröcker, Th., tom Dieck, T.: Kobordismentheorie. Lecture Notes in Mathematics 178, Berlin-Heidelberg-New York: Springer 1970.
3. Conner, P.E., Floyd, E.E.: Differentiable periodic maps. Berlin-Göttingen-Heidelberg-New York: Springer 1964.
4. Conner, P.E., Floyd, E.E.: Maps of odd period. Ann. of Math. **84**, 132–156 (1966).
5. Hook, E.C.: Stable equivariant bordism. Dissertation, University of Virginia 1970.
6. Mostow, G.D.: Equivariant imbeddings in Euclidean space. Ann. of Math. **65**, 432–446 (1957).
7. Palais, R.S.: Imbedding of compact, differentiable transformation groups in orthogonal representations. J. Math. Mech. **6**, 673–678 (1957).
8. Stong, R.E.: Unoriented bordism and actions of finite groups. Amer. Math. Soc. Memoir No. 103, Providence, Rhode Island, 1970.
9. tom Dieck, T.: Kobordismentheorie und Transformationsgruppen. Aaarhus University Preprint Series 1968/69, No. 30.
10. tom Dieck, T.: Bordism of G-manifolds and integrality theorems. Topology **9**, 345–358 (1970).

Dr. Th. Bröcker
Mathematisches Institut
der Universität
D-8400 Regensburg
Federal Republic of Germany

Dr. E.C. Hook
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Mass. 02139
USA

(Received September 3, 1972)

Math. Z. 129, 279–286 (1972)
© by Springer-Verlag 1972

Fixelemente von Automorphismen nulldimensionaler abelscher Gruppen

Peter Plaumann

Reinhold Baer zum siebzigsten Geburtstag am 22. Juli 1972

Es ist bekannt, daß eine lokal kompakte Gruppe mit kompakter Zentrumsfaktorgruppe eine relativ kompakte Kommutatorgruppe hat (s. [3]). In [1] beweist Baer, daß für eine abelsche Gruppe G und eine Untergruppe A von $\operatorname{Aut} G$ die Endlichkeit von je zwei der Zahlen $|A|$, $|G\colon C_G(A)|$, $|G^{-1+A}|$ die Endlichkeit der dritten nach sich zieht. Daraus folgt sofort, daß eine Gruppe mit endlicher Kommutatorgruppe und einem abelschen Normalteiler von endlichem Index eine endliche Zentrumsfaktorgruppe hat. Beim Versuch von Endlichkeit auf Kompaktheit zu verallgemeinern, stößt man auf folgendes Gegenbeispiel: Sei I eine unendliche Menge, und für jedes $\alpha\in\mathsf{I}$ sei $G_\alpha = gp\{z_\alpha\}$ eine zyklische Gruppe der Ordnung 4. Ferner sei $H_\alpha = gp\{2z_\alpha\}$. Wir bilden $G = \prod_{\alpha\in\mathsf{I}} G_\alpha$ als abstrakte Gruppe, versehen mit der Topologie, die durch $H = \prod_{\alpha\in\mathsf{I}} H_\alpha$ mit der Produkttopologie als offene Untergruppe gegeben wird (s. [4], 6.15.c). Sei nun γ der Automorphismus $(g\mapsto -g)$ von G. Dann ist $G^{-1+\gamma} = H = C_G(\gamma)$ kompakt, aber G/H ist eine unendliche diskrete Gruppe.

Wir werden in der folgenden Note einen Satz beweisen, der die Natur des angegebenen Beispiels erläutert, und in einem Korollar dazu unter recht speziellen Voraussetzungen ein Kriterium für die Kompaktheit der Zentrumsfaktorgruppe herleiten.

Die auftretenden abelschen Gruppen schreiben wir additiv, ihre Automorphismengruppen multiplikativ. Mit $\prod$ bzw. $\sum$ bezeichnen wir die volle bzw. die eingeschränkte direkte Summe abelscher Gruppen.

1. Die Charaktergruppe $\mathsf{X}G$ einer kompakten, total unzusammenhängenden abelschen Gruppe G ist eine diskrete Torsionsgruppe ([4], Theorem 24.26) und zerfällt als solche in die direkte Summe ihrer Primärkomponenten $(\mathsf{X}G)_p$. Wenn wir in der üblichen Weise G mit $\mathsf{X}\mathsf{X}G$ identifizieren, so ist G zum direkten Produkt der Charaktergruppen der Primärkomponenten von $\mathsf{X}G$ isomorph ([4], Theorem 23.21).

Mit $(g,\eta)\mapsto g\eta$ bezeichnen wir die bilineare Abbildung von $G\times \mathsf{X}G$ in die Gruppe $\mathbb{K}$ der reellen Zahlen modulo 1 und wir setzen

$$\mathsf{A}(G,\mathsf{Y})=\{g\in G: g\mathsf{Y}=0\},$$

$$\mathsf{A}(H,\mathsf{X}G)=\{\eta\in\mathsf{X}G: H\eta=0\}.$$

Definition. Sei G eine kompakte, total unzusammenhängende abelsche Gruppe, und p sei eine Primzahl. Dann nennen wir

$$G_{(p)}=\mathsf{A}(G,\sum_{q\neq p}(\mathsf{X}G)_q)$$

die *topologische Primärkomponente von G zur Primzahl p.*

Eine interne Definition von $G_{(p)}$ liefert die folgende Bemerkung, von der wir allerdings keinen Gebrauch machen werden.

Bemerkung. Die topologische p-Komponente einer kompakten, total unzusammenhängenden abelschen Gruppe G besteht genau aus den Elementen von G, die p-Potenz-Ordnung haben, zusammen mit denjenigen Elementen von G, die in einer zu den ganzen p-adischen Zahlen isomorphen Untergruppe von G liegen.

Es liegt auf der Hand, wie sich die elementaren Eigenschaften der Primärkomponenten einer diskreten abelschen Gruppe ([2], Theorem 2.1) auf die topologischen Primärkomponenten übertragen.

Hilfssatz 1. *In einer kompakten, total unzusammenhängenden abelschen Gruppe G gelten folgende Aussagen:*

(1) *$G_{(p)}$ ist charakteristische Untergruppe von G,*

(2) $G=\prod_p G_{(p)}$,

(3) $G_{(p)}\cong\mathsf{X}((\mathsf{X}G)_p)$,

(4) *Für $q\neq p$ ist $G_{(p)}$ eindeutig durch q teilbar,*

(5) *G_p ist die Torsionsuntergruppe von $G_{(p)}$,*

(6) *$(G/G_{(p)})_{(p)}=0$ und $(G/G_{(p)})_{(q)}=G_{(q)}+G_{(p)}/G_{(p)}$.*

Beweis. Als Annihilator einer charakteristischen Untergruppe von $\mathsf{X}G$ ist $G_{(p)}$ charakteristisch in G.

Aus der Beziehung $\mathsf{X}G=\sum_p(\mathsf{X}G)_p$ folgt

$$G\cong\mathsf{X}\mathsf{X}G=\mathsf{X}(\sum_p(\mathsf{X}G)_p)=\prod_p\mathsf{A}(G,\sum_{q\neq p}(\mathsf{X}G)_q)=\prod_p G_{(p)}.$$

Die Aussage (3) folgt aus

$$\mathsf{X}((\mathsf{X}G)_p)\cong\mathsf{X}(\mathsf{X}G/\sum_{q\neq p}(\mathsf{X}G)_q)\cong\mathsf{A}(G,\sum_{q\neq p}(\mathsf{X}G)_q)=G_{(p)}.$$

Nach (3) gibt es also einen Isomorphismus σ von $G_{(p)}$ auf $\mathsf{X}((\mathsf{X}G)_p)$. Sei $x\in G_{(p)}$. Als p-Gruppe ist $(\mathsf{X}G)_p$ eindeutig durch $q \neq p$ teilbar. Also gibt es ein eindeutig bestimmtes Element $y^\sigma \in \mathsf{X}((\mathsf{X}G)_p)$ mit $y^\sigma \eta = x(q^{-1}\eta)$ für alle $\eta \in (\mathsf{X}G)_p$. Es folgt, daß $y = q^{-1}x$ ist.

Wegen (4) ist $(G_{(p)})_q = 0$ für alle $q \neq p$. Sei $x\in G$ ein p-Element. Wegen (2) ist $x = a + b$ mit $a \in G_{(p)}$ und $b \in \prod_{q\neq p} G_{(q)}$. Dann ist aber auch b ein p-Element, also ist $b = 0$ und damit $x \in G_{(p)}$.

Aussage (6) endlich folgt aus der Beziehung

$$\mathsf{X}(G/G_{(p)}) \cong \mathsf{X}\Big(\prod_{q\neq p} G_{(q)}\Big) \cong \sum_{q\neq p} \mathsf{X}G_{(q)} = \sum_{q\neq p} (\mathsf{X}G)_q .$$

Hilfssatz 2. *Sei G eine kompakte, total unzusammenhängende abelsche Gruppe, und sei p eine Primzahl. Dann sind die folgenden Aussagen äquivalent:*

(1) *$G_{(p)}$ ist noethersch [Maximalbedingung für abgeschlossene Untergruppen].*

(2) *$(\mathsf{X}G)_p$ ist artinsch [Minimalbedingung für (abgeschlossene) Untergruppen].*

(3) *$(\mathsf{X}G)_p \cong \mathbb{Z}(p^\infty)^n \oplus E$, n eine natürliche Zahl oder 0, E eine endliche Gruppe.*

(4) *$G_{(p)} \cong \Delta_p^n \oplus E$, n eine natürliche Zahl oder 0, E eine endliche Gruppe, Δ_p die Gruppe der ganzen p-adischen Zahlen.*

Beweis. Nach Hilfssatz 1 können wir $G_{(p)}$ mit $\mathsf{X}((\mathsf{X}G)_p)$ identifizieren. Die Äquivalenz von (1) und (2) folgt nun aus der Tatsache, daß die Abbildungen $H \mapsto \mathsf{A}(H, \mathsf{X}G)$ und $\mathsf{Y} \mapsto \mathsf{A}(G, \mathsf{Y})$ Dualitäten zwischen den Verbänden der abgeschlossenen Untergruppen von G und $\mathsf{X}G$ sind.

Die Gleichwertigkeit von (2) und (3) folgt nach Theorem 19.2 aus [2].

Schließlich sind (3) und (4) gleichwertig wegen

$$\mathsf{X}(\Delta_p^n \oplus E) \cong (\mathsf{X}\Delta_p)^n \oplus \mathsf{X}E \cong \mathbb{Z}(p^\infty)^n \oplus E .$$

(Siehe [4], Theorem 23.18, 25.2, 23.27(d).)

Für eine abelsche Gruppe G und eine natürliche Zahl n setzen wir $G(n) = \{g \in G : ng = 0\}$.

Hilfssatz 3. *Es sei G eine lokal kompakte abelsche Gruppe mit einer Untergruppe $H \cong \Delta_p^k$, so daß G/H eine beschränkte Torsionsgruppe ist. Ist G torsionsfrei, so ist G/H eine p-Gruppe vom Rang höchstens k.*

Beweis. Es sei $n(G/H) = 0$. Dann ist $\mu = (x \mapsto nx)$ ein Monomorphismus von G in H. Also ist G/H zu einem Torsionsfaktor von Δ_p^k isomorph, und für diese trifft die Behauptung zu.

2. Im folgenden betrachten wir eine diskrete Gruppe A, die auf einer topologischen abelschen Gruppe G als Gruppe von Automorphismen operiert. Wir setzen

$$D(G, A) = \overline{gp}\{g(1-\alpha): g \in G, \alpha \in A\},$$

$$C(G, A) = \{g \in G: g A = g\}.$$

Für $\eta \in \mathsf{X} G$ wird durch $x(\alpha\eta) = (x\alpha)\eta$ ein Element $\alpha\eta \in \mathsf{X} G$ definiert und die Abbildung $\alpha \mapsto (\eta \mapsto \alpha\eta)$ ist ein Antiisomorphismus von A in die Automorphismengruppe von $\mathsf{X} G$ (vgl. [4], Theorem 26.9). Wir betrachten im folgenden mit der Darstellung von A auf G zugleich immer die oben beschriebene Darstellung von A auf $\mathsf{X} G$.

Hilfssatz 4. *Sei G eine lokal kompakte abelsche Gruppe und sei A eine Gruppe von Automorphismen von G. Dann ist*

$$D(G, A) = \mathsf{A}(G, C(\mathsf{X} G, A)) \quad \textit{und} \quad C(G, A) = \mathsf{A}(G, D(\mathsf{X} G, A)).$$

Beweis. Wegen der Dualität genügt es, eine der beiden Beziehungen zu beweisen. Für ein $g \in G$ gilt aber

$$0 = g[(1-\alpha)\eta] = (g - g\alpha)\eta$$

für alle $\alpha \in A$ und $\eta \in \mathsf{X} G$ genau dann, wenn $g - g\alpha = 0$ für alle $\alpha \in A$ gilt.

Hilfssatz 5. *Sei G eine diskrete abelsche Gruppe und sei A eine Gruppe von Automorphismen von G mit $n = |A|$ endlich. Mit den Bezeichnungen*

$$\sigma = \sum_{\alpha \in A} \alpha \qquad \bar{\sigma} = n - \sigma$$

$$K = \{x \in G: x\sigma = 0\} \qquad \bar{K} = \{x \in G: x\bar{\sigma} = 0\}$$

$$D = D(G, A) \qquad C = C(G, A)$$

gelten dann die folgenden Aussagen:

(1) $nG \subseteq K + \bar{K}$,
(2) $K \cap \bar{K} \subseteq G(n)$,
(3) $nK \subseteq D \subseteq K$,
(4) $n\bar{K} \subseteq C \subseteq \bar{K}$,
(5) $C \cap D \subseteq G(n)$,
(6) $n^2 G \subseteq C + D$.

Beweis. Wir führen die wohl bekannten Rechnungen rasch durch. Sei $g \in G$. Dann ist

$$g\sigma^2 = g(\sum\alpha)(\sum\alpha) = g(n(\sum\alpha)) = g n\sigma.$$

Also ist
$$\bar{\sigma}\,\sigma=\sigma\,\bar{\sigma}=\sigma(n-\sigma)=-\sigma^2+n\,\sigma=-n\,\sigma+n\,\sigma=0.$$

Für alle $g\in G$ ist also
$$g\,\bar{\sigma}\in K \quad \text{und} \quad g\,\sigma\in K.$$
woraus
$$n\,g=g(n-\sigma+\sigma)=g\,\bar{\sigma}+g\,\sigma\in K+K$$
folgt. Also gilt (1).

Für $x\in K\cap K$ ist
$$x\,\sigma=0=x\,\bar{\sigma}=x(n-\sigma),$$
woraus $n\,x=x\,\sigma=0$ folgt. Also gilt (2). Sei nun $g\in G$ und $\alpha\in A$. Dann ist
$$g(1-\alpha)\,\sigma=g(\sigma-\alpha\,\sigma)=g(\sigma-\sigma)=0.$$
Also ist $D\subseteq K$. Für $x\in K$ ist wegen $x\,\sigma=0$
$$n\,x=n\,x+x(-\sigma)=x(n-\sigma)=x\Big[\sum_{\alpha\in A}(1-\alpha)\Big]\in D.$$
Also ist $n\,K\subseteq D$, und (3) ist bewiesen.

Für $g\in C$ ist $g\,\sigma=n\,g$. Daraus folgt
$$g\,\bar{\sigma}=g(n-\sigma)=0$$
und somit $C\subseteq\bar{K}$.

Für $x\in\bar{K}$ ist $x\,\sigma=n\,x$. Also ist für alle $\alpha\in A$
$$n\,x(1-\alpha)=x\,\sigma(1-\alpha)=x(\sigma-\sigma\,\alpha)=x(\sigma-\sigma)=0.$$
Es folgt, daß $n\,\bar{K}\subseteq C$ ist.

Schließlich ist (5) eine Konsequenz von (3) und (4) sowie (2) und (6) eine Konsequenz von (3) und (4) sowie (1).

Es sei angemerkt, daß ein Hilfssatz 5 entsprechender Sachverhalt stets dann gilt, wenn es ein invariantes Mittel über die Funktionen auf A mit Werten in G gibt (vgl. dazu [5]).

3. Wir beweisen nun das Hauptergebnis dieser Note.

Satz. *Sei G eine total unzusammenhängende, lokal kompakte abelsche Gruppe, und sei A eine endliche Gruppe von Automorphismen von G mit $|A|=n$. Ferner sei $D(G,A)$ kompakt und es sei $D(G,A)_{(p)}$ noethersch für alle Primteiler p von n. Dann is $G/C(G,A)$ kompakt.*

Beweis. Beim Beweis verwenden wir die folgenden beiden Reduktionsschlüsse:

(I) *Ist H eine A-zulässige Untergruppe von G und U irgendeine Untergruppe von G, so gilt genau dann $U/H \subseteq C(G/H, A)$, wenn $D(U, A) \subseteq H$ ist.*

Die Gültigkeit von (I) folgt aus

$$(g+H)\alpha = g+H \Leftrightarrow g(1-\alpha) = g - g\alpha \in H.$$

(II) *Ist H eine A-zulässige Untergruppe von G und ist $X/H = C(G/H, A)$ so folgt aus der Kompaktheit von $X/C(X, A)$ und $(G/H)/C(G/H, A)$ die Kompaktheit von $G/C(G, A)$.*

Dies gilt, weil unter den Voraussetzungen von (II) ja bereits $G/C(X, A)$ kompakt ist.

Zunächst sei nun $H = \sum_{p|n} D(G, A)_p$. Wegen Hilfssatz 1, (5) und Hilfssatz 2 ist H eine endliche Gruppe. Nach (I) ist für $X/H = C(G/H, A)$ dann $D(X, A)$ als Untergruppe von H gleichfalls endlich. Nach [1] folgt daraus die Endlichkeit von $X/C(X, A)$. Wegen (II) müssen wir also nur noch zeigen, daß $(G/H)/C(G/H, A)$ kompakt ist. Wir können also im folgenden o.B.d.A. stets annehmen, daß $D(G, A)_p = 0$ für alle Primteiler p von n gilt, da ja $D(G/H, A) = D(G, A)/H$ ist.

Unter der oben gemachten Annahme ist $C(G, A) \cap D(G, A) \subseteq D(G, A)(n) = 0$ wegen Hilfssatz 5, (5). Nach (I) folgt daraus $C(G/C(G, A), A) = 0$, so daß wir $C(G, A) = 0$ annehmen können und dann die Kompaktheit von G zu zeigen haben.

Wir bezeichnen mit $d(G)$ die Anzahl der Primteiler von n, für die $D_{(p)} \neq 0$ ist, und beweisen den Satz durch Induktion nach $d(G)$. Wir setzen $D = D(G, A)$.

Ist $d(G) = 0$, so besagt Hilfssatz 1, (4), daß D eindeutig durch n teilbar ist, während aus Hilfssatz 5, (6) folgt, daß $n^2 G \subseteq D$ ist. Also wird durch

$$\tau = (g \mapsto (n^2)^{-1}(n^2 g))$$

ein Epimorphismus von G auf D mit $\tau^2 = \tau$ definiert. Es folgt, daß $G = D \oplus G(n^2)$ ist. Nun ist aber $G(n^2)$ eine charakteristische Untergruppe von G, also sicher A-zulässig. Also ist

$$D(G(n^2), A) \subseteq D \cap G(n^2) \subseteq D(n^2) = 0,$$

woraus $G(n^2) \subseteq C(G, A) = 0$ und $G = D$ kompakt folgt.

Wir nehmen nun an, der Satz sei falsch. Dann gibt es ein Gegenbeispiel G mit minimalem $d(G) \geqq 1$. Wegen Hilfssatz 1, (6) und $D(G/D_{(p)}, A) = D/D_{(p)}$ ist $d(G/D_{(p)}) = d(G) - 1$. Um also (II) mit $H = D_{(p)}$ und $C(G/D_{(p)}, A) = X/D_{(p)}$ anwenden zu können, müssen wir noch die Kompaktheit von $X/C(X, A)$ nachweisen. Nach (I) ist $D(X, A) \subseteq D_{(p)}$, so daß es wegen Hilfssatz 2 eine natürliche Zahl k mit $D(X, A) \cong \Delta_p^k$ gibt. Wäre die Torsionsuntergruppe von X von 0 verschieden, so gäbe es eine

Primzahl q mit $S=X(q)\neq 0$. Da aber S eine abgeschlossene charakteristische Untergruppe von X ist, gilt $D(S,A)\subseteq D(X,A)\cap S=0$. woraus $S\subseteq C(X,A)=0$ folgte. Also ist X torsionsfrei und wegen Hilfssatz 5, (6) ist $X/D(X,A)$ eine beschränkte Torsionsgruppe. Nach Hilfssatz 3 ist dann $X/D(X,A)$ endlich und somit X kompakt. Wir können also (II) anwenden. und es folgt, daß G kein Gegenbeispiel ist.

Durch Dualisierung erhalten wir das folgende

Korollar 1. *Es sei G eine lokal kompakte abelsche Gruppe, für die jedes Element in einer kompakten Untergruppe liegt, und es sei A eine endliche Gruppe von Automorphismen von G mit $|A|=n$. Ferner sei $C(G,A)$ eine offene Untergruppe von G und es sei $(G/C(G,A))_p$ artinsch für alle Primteiler p von n. Dann ist $D(G,A)$ eine diskrete Untergruppe von G.*

Beweis. Nach [4], 24.18 ist $\mathsf{X}G$ total unzusammenhängend. Wegen Hilfssatz 4 und der bekannten Sätze aus der Dualitätstheorie der lokal kompakten abelschen Gruppen sind in $\mathsf{X}G$ die Voraussetzungen des Satzes erfüllt, so daß

$$\mathsf{X}D(G,A)\cong(\mathsf{X}G)/\mathsf{A}(\mathsf{X}G,D(G,A))=(\mathsf{X}G)/C(\mathsf{X}G,A)$$

kompakt ist.

Ist G irgendeine lokal kompakte Gruppe, so bezeichnen wir mit $\mathfrak{d}G$ den Abschluß der Kommutatorgruppe von G und mit $\mathfrak{z}G$ das Zentrum von G. Mit G_0 bezeichnen wir die zusammenhängende Komponente des Neutralelements von G.

Korollar 2. *Sei G eine lokal kompakte Gruppe mit endlich-dimensionaler, offener Zusammenhangskomponente. Ist $\mathfrak{d}G$ kompakt und besitzt G einen total unzusammenhängenden abelschen Normalteiler N mit kompaktem G/N, so ist $G/\mathfrak{z}G$ kompakt.*

Beweis. Da die abgeschlossene Hülle der Kommutatorgruppe von N und G_0 eine zusammenhängende Untergruppe von N ist, liegt NG_0 in $C_G(N)$, dem Zentralisator von N in G. Also ist die von G in N induzierte Gruppe $A=G/C_G(N)$ von Automorphismen endlich. Es sei $D=D(N,A)$. Dann ist $D\subseteq\mathfrak{d}G$ kompakt. Da G_0 in G offen ist, ist $D/(D\cap G_0)$ diskret, also endlich.

Nach den bekannten Sätzen über zusammenhängende Gruppen (vgl. [6]) gilt folgendes: Zunächst ist $D\cap G_0$ als kompakte Untergruppe von G_0 in einer maximalen kompakten Untergruppe K von G_0 enthalten, welche zusammenhängend ist. Weiterhin ist $K=RB$, wobei R zusammenhängender abelscher und B halbeinfacher Normalteiler von K ist. Wegen der endlichen Dimension von G_0 ist B das Produkt endlich vieler einfacher liescher Normalteiler, so daß $R\cap B\subseteq\mathfrak{z}B$ endlich ist. Es sei $\bar{K}=K/\mathfrak{z}B$ und es sei $\bar{F}=(D\cap G_0)\mathfrak{z}B/\mathfrak{z}B$. Als total unzusammenhängender

Normalteiler von $\bar{K}$ ist $\bar{F}$ in $\mathfrak{Z}\bar{K}=\bar{R}$ enthalten. Nach [4], 24.25 und 24.28 ist $\mathsf{X}\bar{R}$ eine diskrete torionsfreie abelsche Gruppe endlichen Ranges. Also hat $\mathsf{X}\bar{F}\cong\mathsf{X}\bar{R}/\mathsf{A}(\bar{F},\mathsf{X}\bar{R})$ endlichen p-Rang für alle Primzahlen p. Dasselbe gilt nun für $D\cap G_0$ wegen $\bar{F}=(D\cap G_0)/(D\cap G_0\cap\mathfrak{Z}B)$ mit endlichem $D\cap G_0\cap\mathfrak{Z}B$. Es folgt also, daß $(D\cap G_0)_{(p)}$ noethersch ist für alle Primzahlen p. Die bereits bewiesene Endlichkeit von $D/(D\cap G_0)$ sichert dann die Anwendbarkeit des Satzes auf N, A. Die Behauptung folgt nun aus $C(N,A)\subseteq\mathfrak{Z}G$.

Daß in den Voraussetzungen von Korollar 2 die Annahme, daß N total unzusammenhängend sei, auch unter sonst günstigen Bedingungen unentbehrlich ist, zeigt folgendes Beispiel: Es seien K_1 und K_2 isomorph zur Gruppe $\mathbb{K}$ der reellen Zahlen modulo 1, und es sei α der durch

$$\alpha=\big((u,v)\mapsto(u,u+v)\big):K_1\oplus K_2\to K_1\oplus K_2$$

gegebene Automorphismus. Wir bilden die semidirekte Erweiterung von $K_1\oplus K_2$ mit der von α erzeugten Gruppe A von Automorphismen. Dann ist $\mathfrak{d}G=K_2=\mathfrak{Z}G$, und es ist K_2A ein abelscher Normalteiler von G mit kompaktem G/K_2A.

Literatur

1. Baer, R.: Endlichkeitskriterien für Kommutatorgruppen. Math. Ann. **124**, 161 – 177 (1952).
2. Fuchs, L.: Abelian groups. Oxford: Pergamon Press 1960.
3. Grosser, S., Moskowitz, M.: Compactness conditions in topological groups. J. reine angew. Math. **246**, 1 – 40 (1971).
4. Hewitt, E., Ross, K. A.: Abstract harmonic analysis I. Berlin-Göttingen-Heidelberg: Springer 1963.
5. Hofmann, K. H.: Zerfällung topologischer Gruppen. Math. Z. **84**, 16 – 37 (1964).
6. Iwasawa, K.: On some types of topological groups. Ann. of Math. (2) **50**, 507 – 558 (1949).
7. Plaumann, P.: Erweiterungen kompakter Gruppen durch abelsche Gruppen. Math. Z. **99**, 123 – 140 (1967).

Prof. Peter Plaumann
FB Mathematik
Universität Trier-Kaiserslautern
D-675 Kaiserslautern, Postfach 1049
Bundesrepublik Deutschland

(Eingegangen am 20. April 1972)

Math. Z. 129, 287–298 (1972)
© by Springer-Verlag 1972

θ Maps between FK Spaces and Summability

Robert DeVos

1. Introduction

The problem of determining when one FK space is closed in another has been considered by several authors. Wilansky and Zeller in [17] considered the problem for conservative matrices and showed that c is closed in c_A if and only if A sums no bounded divergent sequences. This equivalence was extended by Berg, Crawford and Whitley in [3, 4, 13] to include the condition if and only if $A \in \theta(c, c)$ where $\theta(c, c) \subseteq B(c, c)$. For certain classes of FK spaces and matrices, we show for A a matrix in $B(X, Y)$, X is closed in Y_A if and only if $A \in \theta(X, Y)$. This gives us the Berg, Crawford, Whitley result as a special case.

This result, along with some results of Lacey, Whitley [9, 14] and Stiles [12] enable us to prove that ℓ^p for $0<p<\infty$ cannot equal, nor can it be a closed subspace of $(c_0)_A$ or ℓ^q_A for $p \neq q$, $0<q<\infty$ and A any matrix.

In [2], Bennett has shown that if $E \cap \ell^p$, $1<p<\infty$ is not closed in E, then E contains a sequence in $cs \setminus \ell^p$. In Section 4 we prove that if $E \cap \ell^p$, $0<p<\infty$ is not closed in E, then E contains a sequence in $\bigcap_{q>p} \ell^q$.

2. Preliminaries

An FK space X is a vector space of complex sequences which is also a Fréchet space, (complete linear metric) with continuous coordinates. We shall also consider that all FK spaces considered contain E^∞, the set of all sequences all but finitely many of whose terms are zero. A BK space is a normed FK space. The basic properties of FK spaces can be found in [15] and [20].

The following BK-spaces will be important in the sequel: m, the space of all bounded sequences; c, the space of all convergent sequences; c_0, the space of all null sequences. ℓ^p, $o<p<\infty$, the space of all absolutely p-summable sequences. As usual $\| \, \|_p$ denotes the norm on ℓ^p, for $1 \leqq p<\infty$ and $!!_p$ the paranorm of ℓ^p for $0<p<1$, and $\| \, \|_\infty$ the norm on m, c and c_0.

e denotes the sequence of "ones", $(1, 1, \ldots)$; $e^j, j=1, 2, \ldots$, the sequence $(0, \ldots, 0, 1, 0, \ldots)$ with the "one" in the j-th position. Let E be an FK

space, $x \in E$ and define $P_n(x) = \sum_{i=1}^{n} x_i e^i$ the n-th section of x. If $P_m(x) \to x$ then we say x has AK. E is said to have AK if and only if $P_m(x) \to x$ for all $x \in E$.

Let s be the space of all sequences. For any subset E of s and any matrix A, E_A denotes the set:

$$\{x \in s: Ax \text{ exists and } Ax \in E\}.$$

By [18], Theorem 4.10, E_A is an FK space if E is, and E_A will be called a summability domain. d_A is the existence domain of the matrix and is just E_A for $E = s$.

3.

Let X and Y be FK spaces, $B(X, Y)$ will denote the set of all continuous linear transformations from X to Y. $\theta(X, Y)$ will denote those continuous linear transformations which are range closed and have a finite dimensional null space.

Theorem 1. *Let X and Y be FK spaces and A a matrix mapping X to Y. If $A \in \theta(X, Y)$ then X is closed in Y_A.*

Proof. $A: X \to Y$ has closed range and finite dimensional kernel, hence there exists subspaces W and Z of X such that $Z = \ker A$ and $X = W \oplus Z$.

The range of A is closed, hence the $R[A]$ is an FK space where its topology is the restriction of the Y topology. So $A: W \to R[A]$ is a continuous, one to one, onto mapping between FK spaces, hence by the open mapping theorem A is a homeomorphism between W and $R[A]$.

To show W is closed in Y_A, let w and w^n, $n = 1, 2, \ldots$, be elements of W s.t. $w^n \to w$ in the topology of Y_A. Hence $Aw^n \to Aw$. W and $R[A]$ are homeomorphic, so $w^n \to w$ in the topology of W. This implies that the Y_A topology restricted to W is equivalent to the natural FK topology of W. Hence W is closed in Y_A.

Since X is the topological direct sum of W and the finite dimensional space Z, X is closed in Y_A.

We would like to thank John Hampson for his collaboration in the proof of the previous theorem.

Theorem 2. *Let X be a BK, AK space such that X' is a BK, AK space, Y an FK space and $A \in B(X, Y)$. If X is a closed subspace of Y_A then $A \in \theta(X, Y)$.*

Proof. Let $\| \ \|$ and $\| \ \|'$ be the norms of X and X' respectively and q be the paranorm of Y. By [14], p. 207, Theorem 1, we may assume that

X and X' have $\{e^k\}_{k=1}^{\infty}$ as a monotone basis. That is for $x \in X\,(z \in X')$ $\|P_n x\|$ $(\|P_n z\|')$ is a monotonic increasing function of n.

Suppose $A \notin 0(X, Y)$ then for each $\varepsilon > 0$ and each positive integer N, there exists a $z \in E'$ such that

$$\begin{aligned} &\text{a)} && \|z\| = 1, \\ &\text{b)} && P_N(z) = 0, \qquad (*) \\ &\text{c)} && q(A\,z) < \varepsilon. \end{aligned}$$

Let α^n be the n-th row of A, and let $r(n)$ be an increasing sequence of positive integers such that $\|\alpha^j - P_{r(n)}\,\alpha^j\|' < \frac{1}{2^n}$ for $j \leqq n$. This is possible since X' has AK.

Now, we inductively define a sequence $k(n)$ of positive integers and z^n a sequence of elements of E' of norm 1. Let $k(1) = 1$, and pick $z^1 \in E^{\infty}$ of norm 1 such that $q(A\,z^1) < \frac{1}{2}$. Next choose $k(2) \geqq r(2)$ such that $P_{k(2)} z^1 = z^1$. By $(*)$ there exists a $z^2 \in E'$ such that $\|z^2\| = 1$, $P_{k(2)+1}\, z^2 = 0$ and $q(A\,z^2) < \frac{1}{2^2}$. Continuing this process, we get an increasing sequence $k(n)$ of positive integers and z^n of elements of E' of norm 1 such that

a) $P_{k(n)}\; z^j = z^j$ for $j \leqq n-1$,

b) $P_{k(n)+1}\; z^n = 0$,

c) $q(A\,z^n) < \frac{1}{2^n}$.

Let $x^n = \sum_{i=1}^{n} z^i$ and x be the sequence such that

$$x_k = \begin{cases} 0 & \text{if } k \in \text{support of any } z^i \\ z_k^i & \text{if } k \in \text{support of } z^i. \end{cases}$$

Clearly $x^n \to x$ pointwise. $\{A(z^n)\}$ forms an absolutely convergent series in Y, Y is an FK space, hence there exists a $w \in Y$ such that

$$\sum_{n=1}^{\infty} A(z^n) = w.$$

Next, we would like to show that $x \in d_A$ and $A(x) = w$. $x \in d_A$ if and only if $\alpha^q \cdot x$ exists for $q = 1, 2, \ldots$ i.e. $\sum_{k=1}^{\infty} a_{qk}\, x_k$ is convergent for $q = 1, 2, \ldots$. Let $h > q$. By the definition of z^h we have

$$\left| \sum_{j=k(n)+1}^{k(n+1)} a_{qj}\, z_j^h \right| < \frac{1}{2^h},$$

$\|\alpha^q - P_{r(h)}\alpha^q\|' < \frac{1}{2^h}$ hence $\|P_m\alpha^q - P_{k(n)}\alpha^q\|' < \frac{1}{2^h}$ so

$$\left|\sum_{j=k(n)+1}^{m} a_{qj} z_j^h\right| < \frac{1}{2^h}.$$

Hence $\sum_{k=1}^{\infty} a_{qk} x_k$ converges. Since $\lim_{n\to\infty} \sum_{j=1}^{k(n)} a_{qj} x_j \to w_l$ we have $\alpha^q \cdot x = w_q$. Hence $Ax = w$.

The topology of Y_A is given by

$$x \to |x_j| \qquad j = 1, 2, \ldots$$

$$x \to \sup_n \left|\sum_{j=1}^{n} a_{qj} x_j\right| \qquad q = 1, 2, \ldots$$

$$x \to q(Ax).$$

In order to prove $x^n \to x$ in Y_A, it remains to show that

$$\sup_m \left|\sum_{j=1}^{m} a_{qj}(x^n - x)\right| \to 0 \quad \text{as } n \to \infty \text{ for } q = 1, 2, \ldots.$$

Let q be fixed and $\varepsilon > 0$ be given. $\alpha^q \in (Y_A)'$ so there exists a K such that for $k, m \geqq K$, $\left|\sum_{i=k}^{m} a_{qi} x_i\right| < \varepsilon$. Choose r such that there exists a d in the support of x^r and $d \geqq K$. For each $s \geqq r$

$$\left|\sum_{i=1}^{m} a_{qi}(x_i^s - x_u)\right| = \left|\sum_{i=k(s+1)}^{m} -a_{qi} x_i\right| < \varepsilon.$$

Hence $x^n \to x$ in Y_A and each $x^n \in E^{\infty}$ which implies that $x \in \overline{E^{\infty}}$. $x \notin X$ since $\{P_{k(n)} x\}$ is not a Cauchy sequence in X. Thus X is not a closed subspace of Y_A.

The spaces c_0 and ℓ^p, $1 < p < \infty$ are BK, AK spaces whose duals have AK hence they are choices for X in the previous theorem. The contribution of Berg Crawford and Whitley to the Theorem mentioned in the introduction is a special case of Theorems 1 and 2 where $X = c_0$ and $Y = c$.

In the proof of Theorem 2, the only place we used the AK property of the dual was to insure that the rows of the matrix had AK. Hence the following theorem follows with a similar proof.

Theorem 3. *Let X be BK, AK space, Y an FK space and $A \in B(X, Y)$ such that the rows of A as elements of X' have AK. If X is a closed subspace of Y_A then $A \in \theta(X, Y)$.*

Theorems 2 and 3 are partial converses of Theorem 1. We have not been able to prove the converse in general. For $X = \ell^p$, $1 < p < \infty$. Theorem 2 is the converse. The following theorem is a converse for $X=\ell^p$ for $0<p\leqq 1$.

Theorem 4. *Let $X=\ell^p$, $0<p\leqq 1$, Y an FK space and $A\in B(X, Y)$. If X is a closed subspace of Y_A then $A\in\theta(X, Y)$.*

Proof. Let !! be the paranorm of ℓ^p. Assume $A\notin\theta(X, Y)$. We inductively define a sequence $k(n)$ of positive integers and z^n a sequence of elements of E^∞ such that

a) $!z!=\frac{1}{n}$,

b) $P_{k(n)} z^j=z^j$ for $j\leqq n-1$,

c) $P_{k(n)+1} z^n=0$,

d) $q(Az^n)<\frac{1}{2} n$ where q is the paranorm of Y,

and proceed as in Theorem 2.

Mr. Glen Meyers has a proof for this Theorem in the case where X and Y are ℓ.

A transformation $A\in B(X, Y)$ is strictly singular if it is not a homeomorphism when restricted to any infinite dimensional subspace of X. Clearly if A is strictly singular it is not in $\theta(X, Y)$. In [14], Whitley proved that if $1<p<\infty$ and $1\leqq q<\infty$, $p\neq q$ then all maps in $B(\ell^p, \ell^q)$ are strictly singular. By Theorem 1.2 of [14] every mapping from ℓ to ℓ^q for $1<q<\infty$ is strictly singular. Hence for $p\neq q$, $1\leqq p, q<\infty$ every mapping is strictly singular hence not in $\theta(\ell^p, \ell^q)$.

In Corollary 2.2 of [12] Stiles has shown the space ℓ^p, $0<p<1$ contains no infinite dimensional, subspace isomorphic to a Banach space. Hence we have

Proposition. *Let $0<p<1$ and X any Banach space. All maps in $B(\ell^p, X)$ and $B(X, \ell^p)$ are strictly singular.*

So if p or q lies between 0 and 1 and the other is $\geqq 1$, $\theta(\ell^p, \ell^q)$ is empty. Hence we have the following.

Theorem 5. *Let X be a BK, AK space s.t. X' has AK, Y an infinite dimension BK space and $0<p<1$.*

a) *X cannot equal nor can it be a closed subspace of ℓ^p_A for any matrix A.*

b) *ℓ^p cannot equal nor can it be a closed subspace of Y_A.*

Now suppose $0<p, q<1, p\neq q$. In Theorem 5.1 of [12] Stiles has shown that $\dim_\ell \ell^p$ and $\dim_\ell \ell^q$ are incomparable. Hence ℓ^p is not isomorphic to a subspace of ℓ^q and vice versa which implies $\theta(\ell^p, \ell^q)$ is empty.

Combining the statements in the previous paragraphs we have. If $0<p, q<\infty$, $p\neq q$. $\theta(\ell^p, \ell^q)$ is empty. This statement along with Theorems 1, 2 and 4 prove the following.

Theorem 6. *Let* $A\in B(\ell^p, \ell^q)$, $0<p, q<\infty$

a) *If* $p=q$, ℓ^p *is closed in* ℓ^p_A *if and only if* $A\in\theta(\ell^p, \ell^p)$.

b) *If* $p\neq q$, ℓ^p *is not closed in* ℓ^q_A.

For $q=1$ and $p>1$, this result was proven by Bennett as Corollary 2, p. 104 of [2]. His techniques are completely different.

As Corollary 12, p. 3 of [9] Lacey and Whitley prove that every continuous map from a weakly complete space to c_0 is strictly singular. This statement along with Proposition 1 and Theorems 2 and 4 prove.

Theorem 7. ℓ^p *for* $0<p<\infty$ *cannot equal nor can it be a closed subspace of* $(c_0)_A$ *or* c_A *for any matrix* A.

A proof for $p>1$ was given by Bennett, using different techniques as Corollary 2, p. 109 of [2].

Lacey and Whitley in [9] have also shown that any mapping from c_0 to a weakly complete space is compact. This statement along with Proposition 1 and Theorems 2 and 4 prove.

Theorem 8. *Let* $0<p<\infty$. c_0 *cannot equal nor can it be a closed subspace of* ℓ^p_A *for any matrix* A.

We can replace c_0 in the previous theorem by any FK space X which contains c_0 as a closed subspace and obtain the same conclusion.

We will now restrict ourselves to the case where the first space is c_0. We will need the following lemma

Lemma 1. *Let* X *and* Y *be* BK, AK *spaces and* $A\colon X\to Y$ *then* $A'=A^t$.

Proof: $A' g(x)=\lim_{n\to\infty} A' g(P_n x)$ since x has AK.

$$=\lim_{n\to\infty} g\, A(P_n x)=\lim_{n\to\infty}\sum g_i [A(P_n x)]_i \text{ since } Y'=Y^\beta$$

$$=\lim_{n\to\infty}\sum_{i=1}^{\infty} g_i \sum_{k=1}^{n} a_{ik} x_k=\lim_{n\to\infty}\sum_{k=1}^{n}\left(\sum_{i=1}^{\infty} q_i a_{ik}\right) x_k$$

$$=\lim_{n\to\infty}\sum_{k=1}^{n}(A^t g)_k x_k=\sum_{k=1}^{\infty}(A^t g)_k x_k$$

$$=A^t g(x).$$

Theorem 9. *Let* Y *be a* BK, AK *space such that* Y' *has* AK *and* $A\colon c_0\to Y$. *The following are equivalent.*

a) $A\in\theta(c_0, Y)$

b) $Y_A\cap m=c_0$

c) c_0 *is a closed subspace of* Y_A.

Proof. The equivalence of a) and c) was shown in Theorems 1 and 2.

Hence, the proof will be complete when we show that a) $\Rightarrow$ b) and b) $\Rightarrow$ c). a) $\Rightarrow$ b) A: $c_0 \to Y$ where c_0 and Y are BK, AK spaces hence by Lemma 1. $A' = A^t$. A': $Y' \to \ell$ where Y' and ℓ are BK, AK spaces, hence applying the lemma again we have $A'' = A'^t = A^{tt} = A$. Theorem 2 of [7] proves that if $T \in B(X, Y)$ has closed range, then the null space of T is reflexive if and only if $T''^{-1}\, Y \subseteq X$. $A'' = A$ so $A^{-1}\, Y \subseteq X$ i.e. $Y_A \cap m = X$. b) $\Rightarrow$ c). Let p be the paranorm on Y_A and suppose c_0 is not a closed subspace of Y_A. Then for any $\varepsilon > 0$ and any positive integer N, there exists $x \in c_0$ s.t. $\|x\|_\infty = 1$, $p(x) < \varepsilon$ and $P_N x = 0$. Call this property (*).

Using (*), we inductively define an increasing sequence $\{k(n)\}$ of positive integers and a sequence $\{x^n\}$ of elements of c_0 such that for $n = 1, 2, \ldots$

(a) $\|x^n\|_\infty = 1$,

(b) $p(x^n) < \dfrac{1}{2^n}$,

(c) $P_{k(n)} x^n = 0$,

(d) $\|x^n - P_{k(n+1)} x^n\|_\infty \leqq \dfrac{1}{2^n}$.

Let $k(1) = 1$ and by (*) pick $x^1 \in c_0$ of norm 1 such that $P_{k(1)} x_1 = 0$ and $p(x^1) < \frac{1}{2}$. By (*) there exists $x^2 \in c_0$ of norm one such that $P_{k(2)} x^2 = 0$. Continuing this process, we get a sequence $\{k(n)\}$ of positive integers and $\{x^n\}$ of elements of c_0 with the required properties.

Let $x = \sum_n x^n$. $x \in Y_A$ and $x \in m \setminus c_0$ hence $Y_A \cap m \neq c_0$.

Theorem 10. *Let X be a BK space containing c_0 as a closed subspace, Y a BK, AK space such that Y' has AK and A a matrix, A: $X \to Y$. If $X \cap m \neq c_0$ then $A \notin \theta(X, Y)$.*

Proof. Suppose $A \in \theta(X, Y)$. Then A: $c_0 \to Y$ is in $\theta(c_0, Y)$ and the result follows as an application of the previous theorem.

Note that if X and Y satisfy the hypothesis of the previous theorem and $X \cap m \neq c_0$ then X and Y are not linearly homeomorphic via a matrix transformation. Wilansky in Corollary 1 of Theorem 3 of [16] proved the above result for $X = c$ and $Y = c_0$.

In the proof of their part of the Theorem mentioned in the introduction, Wilansky and Zeller used the "gliding hump" technique to show that if c is not a closed subset of c_A, then A sums a bounded divergent sequence. This result was extended by Meyer-König and Zeller in [10, 11] who replaced c_A by an arbitrary FK space E. In [18] Yurimae proved that if ℓ is not closed in ℓ_A then there exists an $x \in (bs \setminus \ell) \cap \ell_A$. Bennett in [2] extended both the results of Meyer-König and Zeller and those of Yurimae. He proved that if $E\cap\ell^p$, $1\leqq p<\infty$ is not closed in E then E contains a sequence in $cs\setminus\ell^p$. The following theorem is of a similar type.

Theorem 11. *Let E be an FK space with $E\cap\ell^p$, $o<p<\infty$ not closed in E, then E contains a sequence in $[\ell^p]=\bigcap_{q>p}\ell^p$.*

Proof. Assume first that $p\geqq 1$. A similar proof will follow for $0<p<1$. We may assume the topology of E is given by a family of seminorms. $\{q_n\}_{n=1}^{\infty}$ with the property that $\left(\sum_{i=1}^{m}|x_i|^p\right)^{1/p}\leqq q_m(x)\leqq q_{m+1}(x)$ for $x\in E$, $m=1,2,\dots$. $E\cap\ell^p$ is not closed in E, hence given $\varepsilon>0$, $M>0$ and any positive integer m, we can find an $x\in E\cap\ell^p$ such that

$$q_m(x)<\varepsilon \quad\text{and}\quad \|x\|_p=M. \tag{1}$$

Let $\varepsilon_n=\frac{1}{2^{n+1}}$ and $M_n=\frac{1}{n^{1/p}}$. We want to inductively define an increasing sequence $\{m_n\}$ of positive integers and $\{x^n\}$ of elements of $E\cap\ell^p$ such that

$$\begin{aligned}
&\text{(a)} && \|x^n\|_p=\frac{1}{n^{1/p}},\\
&\text{(b)} && \|x^i-P_{m_n}x^i\|_p\leqq\frac{1}{2^{n+1}} \quad \text{for } i=1,\dots,n-1,\\
&\text{(c)} && q_{m_n}(x^n)<\frac{1}{2^{n+1}}.
\end{aligned}\tag{2}$$

Let $m_1=1$, and choose x^1 as in (1). With $\varepsilon_2=\frac{1}{2^3}$ and $M_2=\frac{1}{2^{1/p}}$, choose m_2 such that $\|x^1-P_{m_2}x^1\|_p<\frac{1}{2^3}$. By (1) choose x^2 s.t. $q_{m_2}(x^2)<\varepsilon^2$ and $\|x^2\|_p=\frac{1}{2^{1/p}}$. Suppose now that $m_1,\dots,m_{n-1}$ and $x^1,\dots,x^{n-1}$ have been chosen. Pick $m_n>m_{n-1}$ s.t.

$$\|x^i-P_{m_n}x^i\|_p\leqq\frac{1}{2^{n+1}} \quad \text{for } i=1,\dots,n-1.$$

By (1) we can find an $x^n\in E\cap\ell^p$ s.t. $q_{m_n}(x^n)<\frac{1}{2^{n+1}}$ and $\|x^n\|_p=\frac{1}{n^{1/p}}$. Hence we have sequences $\{m_n\}$ of positive integers and $\{x^n\}$ of elements of $E\cap\ell^p$ which satisfy (2).

Note that if $j<n$, $q_j(x^n)\leqq q_n(x^n)\leqq q_{m_n}(x^n)<\frac{1}{2^{n+1}}$. Let $x=\sum_{n=1}^{\infty} x^n$ and we claim $x\in E$. Let j be a fixed positive integer and $y^k=\sum_{n=1}^{k} x^n$. $q_j(y^k)$ is a Cauchy sequence of real numbers for each j. E is an FK space hence y^k converges to some point of E. y^k converges pointwise to x, hence $x\in E$.

Now let $q>p$ be fixed. We want to show $x\in \ell^q\setminus \ell^p$. In the interval $m_j<k\leqq m_{j+1}$ $(j=1,2,\ldots)$ we have

$$
\begin{aligned}
\|(P_{m_{j+1}}-P_{m_j})\,x\|_p &= \left(\sum_{k=m_j+1}^{m_{j+1}} |x_p|^p\right)^{1/p}\\
&= \left(\sum_{k=m_j+1}^{m_{j+1}} \left|\sum_{i=1}^{\infty} x_k^i\right|\right)^{1/p}\\
&\leqq \sum_{i=1}^{\infty} \|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p\\
&\leqq \sum_{i=1}^{j-1} \|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p + \|(P_{m_{j+1}}-P_{m_j})\,x^j\|_p\\
&\quad + \sum_{i=j+1}^{\infty} \|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p\\
&\leqq A+B+C\,.
\end{aligned}
$$

Let $i<j$, $\|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p\leqq \|x^i-P_{m_j}x^i\|_p<\frac{1}{2^{j+1}}$ by (2)(b). Hence

$$\sum_{i=1}^{j-1} \|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p\leqq \frac{j-1}{2^{j+1}}\,.$$

So

$$A\leqq \frac{j-1}{2^{j+1}}\,\|(P_{m_{j+1}}-P_{m_j})\,x^j\|_p\leqq \frac{1}{j^{1/p}}\,.$$

$B\leqq \frac{1}{j^{1/p}}$. Let $i>j$,

$$\|(P_{m_{j+1}}-P_{m_j})\,x\|_p\leqq \|(P_{m_{j+1}}x^i)\|_p\leqq q_{m_{j+1}}(x^i)\leqq q_{m_i}(x^i)<\frac{1}{2^{i+1}}$$

by (2)(c).

Hence

$$\sum_{i=j+1}^{\infty} \|(P_{m_{j+1}}-P_{m_j})\,x^i\|_p\leqq \sum_{i=j+1}^{\infty} \frac{1}{2^{i+1}}=\frac{1}{2^{j+1}}\,.$$

So $C \leqq \frac{1}{2^{j+1}}$. Thus

$$\left(\sum_{k=m_j+1}^{m_{j+1}} |x_k|^p\right)^{1/p} < \frac{j-1}{2^{j+1}} + \frac{1}{j^{1/p}} + \frac{1}{2^{j+1}} < \frac{2}{j^{1/p}} \tag{3}$$

for j large. By the triangular inequality we have

$$\begin{aligned} \|(P_{m_{j+1}} - P_{m_j}) x^j\|_p &- \sum_{i=1}^{j-1} \|(P_{m_{j+1}} - P_{m_j}) x^i\|_p \\ &- \sum_{i=j+1}^{\infty} \|(P_{m_{j+1}} - P_{m_j}) x^i\|_p \\ &\leqq \|(P_{m_{j+1}} - P_{m_j}) x\|_p \end{aligned}$$

$$\|(P_{m_{j+1}} - P_{m_j}) x^j\|_p \geqq \|x^j\|_p - \|P_{m_j} x^j\|_p - \|x^j - P_{m_{j+1}} x^j\|_p.$$

By (2)(a), (b) and (c) respectively $\|x^j\|_p = \frac{1}{j^{1/p}}$,

$$\|x^j - P_{m_{j+1}} x^j\|_p \leqq \frac{1}{2^{j+2}} \quad \text{and} \quad \|P_{m_j} x^j\|_p < \frac{1}{2^{j+1}}.$$

Hence

$$\|(P_{m_{j+1}} - P_{m_j}) x^j\|_p \geqq \frac{1}{j^{1/p}} - \frac{1}{2^{j+1}} - \frac{1}{2^{j+2}}.$$

So

$$\frac{1}{j^{1/p}} - \frac{1}{2^{j+1}} - \frac{1}{j^{j+2}} - \frac{1}{2^{j+1}} - \frac{1}{2^{j+2}} - \frac{j-1}{2^{j+1}} - \frac{1}{2^{j+1}} \leqq \|(P_{m_{j+1}} - P_{m_j}) x\|_p.$$

Hence

$$\frac{1/2}{j^{1/p}} \leqq \|(P_{m_{j+1}} - P_{m_j}) x\|_p \qquad \text{for large } j. \tag{4}$$

Combining (3) and (4) we get

$$\frac{1/2}{j^{1/p}} \leqq \left(\sum_{k=m_j+1}^{m_{j+1}} |x_k|^p\right)^{1/p} \leqq \frac{2}{2^{1/p}} \qquad \text{for large } j.$$

Hence

$$\frac{(1/2)^p}{j} \leqq \sum_{k=m_j+1}^{m_{j+1}} |x_k|^p \leqq \frac{2^p}{j} \qquad \text{for large } j. \tag{5}$$

Since (5) is true for all large j, $x \notin \ell^p$. It is easily seen that $(a+b)^r \geqq a^r + b^r$ for a and b non negative and $r \geqq 1$. Let $r = \frac{q}{p}$

$$\left(\frac{2^p}{j}\right)^{q/p} \geqq \left(\sum_{k=m_j+1}^{m_{j+1}} |x_k|^p\right)^{q/p} \geqq \sum_{k=m_j+1}^{m_{j+1}} |x_k|^q.$$

Hence $x \in \ell^q$.

Note that for $r>q>p$, we cannot say there exists an $x\in(\ell^r\setminus\ell^q)\cap\ell^p$ for let $r=3$, $q=2$ and $p=1$. $\ell^1\cap\ell^2=\ell^1$ is not closed in ℓ^2 but $(\ell^3\setminus\ell^2)\cap\ell^2=\varnothing$.

Corollary 1. *Let $o<p<\infty$, Y and FK space and $A\in B(\ell^p, Y)$. If $A\notin\theta(\ell^p, Y)$ then Y_A contains a sequence in $[\ell^p]$.*

It is well known (see Bennett [2]) that the union of a countable family of nested increasing FK spaces is not an FK space.

Theorem 12. *Let $E\cap\ell^p$ not be closed in E for all p, $1<p<\infty$. Then E contains a sequence in $c_0\setminus\bigcap_{n=1}^{\infty}\ell^p$.*

Proof. $E\cap\left(\bigcup_{n=1}^{\infty}\ell^n\right)=\bigcup_{n=1}^{\infty}(E\cap\ell^n)$. $E\cap\ell^n$ is not closed in E, hence $\{E\cap\ell^n\}$ is a nested strictly increasing sequence of FK spaces. By the remark preceeding this theorem, the union is not an FK space. $E\cap c_0$ is an FK space and $E\cap c_0\supset\bigcup_{n=1}^{\infty}(E\cap\ell^n)$ hence we have strict containment.

Bennett in [2], p. 26 has shown that if $\{F^n\}$ is a decreasing sequence of FK spaces and E a BK space such that $E\subseteq\bigcap_{n=1}^{\infty}F^n$. Then E is closed in the intersection if and only if E is closed in F^m for some positive integer m. The following theorem is an application of Bennett's result and Theorem [11].

Theorem 13. *Let A^n be a sequence of matrix maps: $X\to\ell^q$ where $X=c_0$ $(\ell^p, 1\leqq p<\infty$ and $p\neq q)$ and $o<q<\infty$. Such that $\ell^q_{A^n}\supset\ell^q_{A^{n+1}}$. Then there exists $x\in m\setminus c_0([\ell^p])$ such that $x\in\bigcap_{n=1}^{\infty}\ell^q_{A^n}$.*

Proof. By Theorems 6 and 8, X is not a closed subspace of $\ell^q_{A^n}$ for any integer n. Hence by Bennett's Theorem $c_0(\ell^p)$ is not a closed subspace of $\bigcap_{n=1}^{\infty}\ell^q_{A^n}$. So the intersection contains a bounded divergent sequence by the result of Meyer-König and Zeller quoted before Theorem 11 (contains a sequence in $[\ell^p]$ by Theorem 11).

References

1. Bennett, G.: Some inclusion theorems for sequence spaces. To appear in Pacific J. Math.
2. Bennett, G.: Applications of functional analysis to summability theory. Ph.D. Thesis, Cambridge University (1970).
3. Berg, I.D.: A Banach algebra criterion for Tauberian theorems. Proc. Amer. Math. Soc. **15**, 648–652 (1964).

4. Crawford, J. P.: Transformations in Banach spaces with application to summability theory. Ph. D. Thesis, Lehigh University (1966).
5. DeVos, R.: Distinguished subsets and matrix maps between FK spaces. Ph. D. Thesis, Lehigh University (1972).
6. Dunford, N., Schwartz, J.T.: Linear operators, part I. New York: Interscience 1964.
7. Garling, D.J.H., Wilansky, A.: The Berg-Crawford-Whitley summability theorem. To appear in Proc. Cambridge Philos. Soc.
8. Goldberg, S.: Unbounded linear operators. New York: McGraw-Hill 1966.
9. Lacey, E., Whitley, R.J.: Conditions under which all the bounded linear maps are compact. Math. Ann. **158**, 1-5 (1965).
10. Meyer-König, W., Zeller, K.: Lückenumkehrsätze und Lückenperfektheit. Math. Z. **66**, 203-224 (1956).
11. Meyer-König, W., Zeller, K.: FK-Räume und Lückenperfektheit. Math. Z. **78**, 143-148 (1962).
12. Stiles, W.J.: On properties of subspaces of ℓ^p, $0<p<1$. Trans. Amer. Math. Soc. **149**, 405-415 (1970).
13. Whitley, R.: Conull and other matrices which sum a bounded divergent sequence. Amer. Math. Monthly **74**, 798-801 (1967).
14. Whitley, R.: Strictly singular operators and their conjugates. Trans. Amer. Math. Soc. **13**, 252-261 (1964).
15. Wilansky, A.: Functional analysis. New York: Blaisdell 1964.
16. Wilansky, A.: Topological divisors of zero and Tauberian theorems. Trans. Amer. Math. Soc. **113**, 240-251 (1964).
17. Wilansky, A., Zeller, K.: Summation of bounded divergent sequences, topological methods. Trans. Amer. Math. Soc. **98**, 501-509 (1955).
18. Yurimyae, E.: Topological properties of conull matrices. Tartu Rukl. Ul. Taimetised **177**, 43-61 (1965). [Russian with Estonian and German summaries.]
19. Zeller, K.: Faktorfolgen bei Limitierungsverfahren. Math. Z. **56**, 134-151 (1952).
20. Zeller, K., Beekmann, W.: Theorie der Limitierungsverfahren. Berlin-Heidelberg-New York: Springer 1970.

Dr. Robert DeVos
Mathematics Department
Villanova University
Villanova, Pennsylvania 19085
USA

(Received June 2, 1972)

Math. Z. 129, 299–319 (1972)
© by Springer-Verlag 1972

Non-Linear Potentials and Approximation in the Mean by Analytic Functions

Lars Inge Hedberg

1. Introduction

The first part of this paper is devoted to a variant of potential theory. We study the non-linear potential $V^\mu = K * (K * \mu)^{p-1}$ of a measure μ on $\mathbb{R}^d$, where K is a kernel depending only on $|x|$, $*$ denotes convolution, and p is a number greater than one. This kind of potential appears implicitly in the theory of potentials of functions in Lebesgue classes of Fuglede [18, 20], Meyers [34], and Rešetnjak [37]. It was introduced as an independent object of investigation and given its name in the work of Havin and Maz'ja [24, 25]. The classical Newton and M. Riesz potentials are obtained by setting $p=2$ and choosing suitable K.

Following Brelot [9, 10] we define thinness (effilément) of sets with respect to these potentials. We give a necessary condition of Wiener type for a set E in $\mathbb{R}^d$ to be thin at a point in $\mathbb{R}^d$ (Theorem 2). Under supplementary assumptions on p, d and K this condition is also sufficient and contains the classical Wiener criterion as a special case. Among the consequences we can mention a result on the local behavior (pseudo-continuity) of functions in general Sobolev spaces (Theorem 5). Finally we extend some earlier results on the "instability of capacities" to this general setting (Theorem 9).

In the second part of the paper we specialize to the Sobolev spaces W_1^q of functions whose first derivatives are in L^q. (Here and throughout the paper $\frac{1}{p} + \frac{1}{q} = 1$.) By duality this gives results on the possibility of approximation in L^p on planar sets by analytic functions, and in W_1^p by harmonic functions. We obtain conditions which improve those given earlier by the author [27], and while being slightly short of necessary and sufficient they are more explicit than the necessary and sufficient condition given by Bagby [6, 7].

The author is grateful to V. P. Havin for valuable comments.

2. Thin Sets for Non-Linear Potentials

We first recall some known results. We refer to the papers mentioned in the introduction for details.

Let $K(r)$, $r>0$, be a positive, decreasing, continuous function. For $x\in\mathbb{R}^d$, $x\neq 0$, we define $K(x)=K(|x|)$, and we assume

$$\int_{|x|<1} K(x)\,dm(x)<\infty,$$

where m is d-dimensional Lebesgue measure.

For $f\in L^q(\mathbb{R}^d)$, $1\leqq q\leqq\infty$, we define a potential

$$U^f(x)=U_K^f(x)=(K*f)(x)=\int K(x-y)\,f(y)\,dm(y),$$

and for a positive (Borel) measure μ we similarly define

$$U^\mu(x)=U_K^\mu(x)=(K*\mu)(x)=\int K(x-y)\,d\mu(y).$$

For an arbitrary set $E\subset\mathbb{R}^d$ the (outer) (K,q)-capacity $C_{K,q}(E)$ is defined by

$$C_{K,q}(E)=\inf_f\int|f(x)|^q\,dm(x)=\inf_f\|f\|_q^q,$$

where the infimum is taken over functions $f\in L^q$ such that $f\geqq 0$ and $U_K^f(x)\geqq 1$ on E.

A property which holds for all points outside a set E with $C_{K,q}(E)=0$ is said to hold (K,q)-quasi everywhere or (K,q)-q.e.

For an arbitrary Borel set $E\subset\mathbb{R}^d$ we also define a capacity $c_{K,q}(E)$ by

$$c_{K,q}(E)=\sup_\mu \mu(E),$$

where the supremum is taken over all positive Borel measures concentrated on E such that

$$\|U_K^\mu\|_p\leqq 1.$$

The two extremal problems are paired by the bilinear form

$$\Phi(f,\mu)=\int U_K^f\,d\mu=\int U_K^\mu f\,dm,$$

and this duality gives the following theorem. See the literature quoted above for proofs, especially Meyers [34]. The closure of E is denoted $\bar{E}$.

Theorem 1. *Let E be a Borel set, and let $1<q<\infty$. Then*

(a) $c_{K,q}(E)=C_{K,q}(E)^{1/q}$.

(b) *There is a measure μ supported by $\bar{E}$ such that $\mu(\mathbb{R}^d)=c_{K,q}(E)$, and $\|U_K^\mu\|_p=1$.*

(c) *There is a unique function $f\in L^q$ such that $\|f\|_q^q=C_{K,q}(E)$, and $U_K^f(x)\geqq 1$ (K,q)-q.e. on E.*

(d) *μ and f are related by*

$$f=c_{K,q}(E)\,(U_K^\mu)^{p-1},\qquad \text{i.e. } U_K^f=c_{K,q}(E)\big(K*(K*\mu)^{p-1}\big).$$

(e) μ *is concentrated on the set* $\bar{E} \cap \{U_K^f(x)=1\}=F$, *and* $C_{K,q}(F)=C_{K,q}(E)$.

We write

$$V^\nu = V_{K,q}^\nu = K * (K * \nu)^{p-1}.$$

This is called a non-linear potential of ν, and if $\nu = c_{K,q}(E)^{q-1}\mu$, where μ is the extremal measure above, we call $V_{K,q}^\nu$ the (K,q)-capacitary potential for E. Then $\nu(E)=c_{K,q}(E)^q = C_{K,q}(E)$, and we call ν the (K,q)-capacitary measure for E.

It is easy to see that $C_{K,q}(E)$ can also be defined by the following extremal property.

$$C_{K,q}(E)=\sup_\mu \mu(E),$$

where the supremum is taken over measures μ supported on E such that

$$V_{K,q}^\mu(x) \leqq 1, \qquad x \in \operatorname{Supp}\mu.$$

The space of functions $U_K^f, f \in L^q$, normed by $\|U_K^f\| = \|f\|_q$, is denoted by $\mathscr{L}_K^q$. If $K(x)$ is a Riesz kernel $R_\alpha(x)=|x|^{\alpha-d}$, $0<\alpha<d$, or equivalently the corresponding Bessel kernel J_α, defined as the Fourier transform of $(1+|t|^2)^{-\alpha/2}$, we write U_α^f, $\mathscr{L}_\alpha^q$, $C_{\alpha,q}$, etc.

It is a well-known consequence of the Calderón-Zygmund theory that for integral α $\mathscr{L}_\alpha^q$ equals the Sobolev space W_α^q of functions whose distribution derivatives of order up to α are functions in L^q, and that the norms are equivalent. See Calderón [11], or Stein [39; Ch. V:3.3].

In classical potential theory there are various equivalent definitions of thin (effilé) sets. See Brelot [9, 10]. We adopt the following one as suitable for our purposes.

Definition. Let $E \subset \mathbb{R}^d$ be an arbitrary set. E is said to be (K,q)-thin at a point x_0 if and only if there is a positive measure ν such that the non-linear potential $V_{K,q}^\nu$ is bounded and satisfies

$$V_{K,q}^\nu(x_0) < \liminf_{x \to x_0,\, x \in E \setminus \{x_0\}} V_{K,q}^\nu(x).$$

Remark. By the maximum principle of Havin and Maz'ja [24, 25] (see also Adams and Meyers [2, 3]) it is sufficient to assume that $V_{K,q}^\nu$ is bounded on the support on ν. In particular any capacitary potential is bounded.

Denote by $B_x(\delta)$ the ball $\{y; |y-x|<\delta\}$.

Theorem 2. *Let K be a kernel as above, and assume that K in addition satisfies*

$$K(\delta) \leqq A K(2\delta) \tag{1}$$

for some $A>0$ and all sufficiently small $\delta>0$. Let $E\subset\mathbb{R}^d$ be a Borel set. Then E is not (K,q)-thin at a point x if

$$\int_0 C_{K,q}(E\cap B_x(\delta))^{p-1}K(\delta)^p\,\delta^{d-1}\,d\delta=\infty. \tag{2}$$

If $K(x)$ is a Riesz or Bessel kernel the above theorem takes the following form.

Theorem 2′. *A Borel set $E\subset\mathbb{R}^d$ is not (α,q)-thin at x if*

$$\int_0\left(\frac{C_{\alpha,q}(E\cap B_x(\delta))}{\delta^{d-\alpha q}}\right)^{p-1}\frac{d\delta}{\delta}=\infty. \tag{3}$$

Remark 1. Theorem 2′ is interesting only for $\alpha q\leqq d$, since U_α^f is continuous for all $f\in L^q$ if $\alpha q>d$. Note that

$$C_{\alpha,q}(B_x(\delta))\approx\delta^{d-\alpha q}\quad\text{for } \alpha q<d, \quad\text{and}\quad C_{\alpha,q}(B_x(\delta))\approx\left(\log\frac{1}{\delta}\right)^{1-q}$$

for $\alpha q=d$. See Meyers [34; Lemmas 7, 8].

Remark 2. Maz'ja [33] has studied the Dirichlet problem for a class of quasi-linear elliptic equations including the equation

$$\operatorname{div}(\operatorname{grad} u\,|\operatorname{grad} u|^{q-2})=0$$

and proved that if x is a boundary point of a region D, then (3) with $\alpha=1$ and $E=\partial D$ is a sufficient condition for x to be regular. His proof is quite different, and the relation between the two problems is not clear.

Remark 3. For $p=2$ the potential $V_{K,2}^\nu$ becomes a classical, linear potential with respect to the kernel $K*K$. By well-known convolution properties of Riesz and Bessel kernels it follows that $C_{\alpha,2}$ equals the classical Riesz or Bessel capacity $C_{2\alpha}$. If we set $2\alpha=\beta$ (3) becomes

$$\int_0\frac{C_\beta(E\cap B_x(\delta))}{\delta^{d-\beta+1}}\,d\delta=\infty,\qquad 0<\beta\leqq d.$$

Together with Theorem 4 below this gives the classical Wiener criterion. Naturally the proof simplifies in this case. Previously the Wiener criterion seems to be recorded in the literature only for the case $0<\beta\leqq 2$, i.e. when the strict maximum principle holds. See e.g. Landkof [31, p. 356].

Remark 4. It is easily seen that (2) is equivalent to

$$\sum_{n=1}^{\infty}C_{K,q}(E\cap B_x(2^{-n}))^{p-1}K(2^{-n})^p\,2^{-nd}=\infty.$$

Remark 5. In a letter received just before this paper was first submitted for publication D.R. Adams kindly informed the author that he and N.G. Meyers have worked extensively on the problem of finding a Wiener criterion for non-linear potentials. In particular they have also found Theorem 2. (*Added in proof:* This work has appeared in Indiana Univ. Math. J. **22**, 169–197 (1972).) See also the recent papers [1–4, 21, 35].

Proof of Theorem 2. We let $x=0$, and we denote $E\cap B_0(\delta)$ by E_δ. The restriction of a measure ν to $B_0(\delta)=B_\delta$ is denoted ν_δ.

We assume that (2) holds, i.e.

$$\int_0 C_{K,q}(E_\delta)^{p-1} K(\delta)^p \delta^{d-1} d\delta = \infty .$$

We start by the observation that for any positive measure ν

$$\int_0^\infty \nu(B_\delta)^{p-1} K(\delta)^p \delta^{d-1} d\delta \leqq A V_{K,q}^{\nu}(0) \tag{4}$$

for some finite constant A, independent of ν. In fact, using (1) we find

$$\begin{aligned} V_{K,q}^{\nu}(0) &= \int K(y)\,dm(y)\{\int K(t-y)\,d\nu(t)\}^{p-1} \\ &\geqq \sum_{n=-\infty}^{\infty} \int_{2^{-n-1}<|y|\leqq 2^{-n}} K(y)\,dm(y)\{\int_{|t|\leqq 2^{-n}} K(t-y)\,d\nu(t)\}^{p-1} \\ &\geqq \sum_{n=-\infty}^{\infty} \int_{2^{-n-1}<|y|\leqq 2^{-n}} K(y)\,K(2^{-n+1})^{p-1}\,\nu(B_{2^{-n}})^{p-1}\,dm(y) \\ &\geqq A_1 \int K(y)^p\,\nu(B_{|y|})^{p-1}\,dm(y) \\ &= A \int_0^\infty K(\delta)^p\,\nu(B_\delta)^{p-1}\,\delta^{d-1}\,d\delta . \end{aligned}$$

We now assume that ν is a positive measure such that $V_{K,q}^{\nu}=V^{\nu}\geqq 1$ (K,q)-quasi everywhere on $E\cap G$, for some neighborhood G of 0. We also assume that V^{ν} is bounded, or somewhat weaker, that

$$\limsup_{\delta\to 0} \frac{1}{\nu(B_\delta)} \int_{B_\delta} V^{\nu_\delta}\,d\nu = \limsup_{\delta\to 0} \frac{1}{\nu(B_\delta)} \int (U^{\nu_\delta})^p\,dm < \infty . \tag{5}$$

We claim that under these assumptions $V^{\nu}(0)\geqq 1$. Since V^{ν} is lower semicontinuous it follows that $V^{\nu}(0)=\liminf_{x\to 0,\,x\in E\setminus\{0\}} V^{\nu}(x)$.

The proof is a refinement of a method used by Carleson in [12; Lemma 1] and [13; Theorem III:3]. It consists in approximating $V^{\nu}(0)$ with a suitable sequence of averages of V^{ν}. We choose measures σ_δ with (compact) support in E_δ, such that σ_δ has unit mass, and

$$\|U^{\sigma_\delta}\|_p \leqq 2\,C_{K,q}(E_\delta)^{-1/q}. \tag{6}$$

Then, if δ is so small that $E_\delta \subset G$,

$$\int V^\nu(x)\, d\sigma_\delta(x) \geqq 1,$$

or, by changing the order of integration,

$$\int U^{\sigma_\delta}(y)\{U^\nu(y)\}^{p-1}\, dm(y) \geqq 1.$$

Let $\varepsilon > 0$, and choose $\rho > 0$ so small that

$$\int_{|y|<\rho} K(y)\{U^\nu(y)\}^{p-1}\, dm(y) < \varepsilon,$$

which is possible if $V^\nu(0) < \infty$. Consider

$$I_1 = \int_{|y|\geqq\rho} U^{\sigma_\delta}(y)\{U^\nu(y)\}^{p-1}\, dm(y).$$

As $\delta \to 0$

$$U^{\sigma_\delta}(y) = \int K(x-y)\, d\sigma_\delta(x) \to K(y),$$

uniformly for $|y| \geqq \rho$, so

$$\lim_{\delta\to 0} I_1 = \int_{|y|\geqq\rho} K(y)\{U^\nu(y)\}^{p-1}\, dm(y),$$

which tends to $V^\nu(0)$ as $\rho \to 0$. If we can show that

$$\liminf_{\delta\to 0} \int_{|y|\leqq\rho} U^{\sigma_\delta}(y)\{U^\nu(y)\}^{p-1}\, dm(y) < A\,\varepsilon$$

it will follow that

$$V^\nu(0) = \liminf_{\delta\to 0} \int V^\nu(x)\, d\sigma_\delta(x) \geqq 1.$$

We first consider

$$J_\delta = \int U^{\sigma_\delta}\{U^{\nu_{4\delta}}\}^{p-1}\, dm.$$

By Hölder's inequality

$$J_\delta \leqq \|U^{\sigma_\delta}\|_p \,\|(U^{\nu_{4\delta}})^{p-1}\|_q = \|U^{\sigma_\delta}\|_p \,\|U^{\nu_{4\delta}}\|_p^{p-1}.$$

By (5) and (6)

$$J_\delta \leqq A\, C_{K,q}(E_\delta)^{-1/q}\, \nu(B_{4\delta})^{1/q}$$

for sufficiently small δ, and by (2), (4), and (1)

$$\liminf_{\delta\to 0} J_\delta = 0.$$

It remains to estimate J'_δ, where

$$J'_\delta = \int_{|y|<\rho} dm(y) \int K(x-y)\, d\sigma_\delta(x)\{\int_{|t|\geqq 4\delta} K(t-y)\, d\nu(t)\}^{p-1}$$
$$= \int d\sigma_\delta(x) \int_{|y|<\rho} K(x-y)\, dm(y)\{\int_{|t|\geqq 4\delta} K(t-y)\, d\nu(t)\}^{p-1}.$$

First consider $\frac{3}{2}\delta \leqq |y| < \rho$. (We can assume $4\delta < \rho$.) Then, for $|x| < \delta$, we have $|x-y| \geqq |y| - |x| > \frac{1}{3}|y|$, so by (1)

$$\int_{\frac{3}{2}\delta \leqq |y| < \rho} K(x-y)\,dm(y)\Big\{\int_{|t| \geqq 4\delta} K(t-y)\,dv(t)\Big\}^{p-1}$$
$$\leqq A \int_{|y|<\rho} K(y)\,dm(y)\Big\{\int K(t-y)\,dv(t)\Big\}^{p-1}, \qquad \text{for some } A.$$

Now let $|y| < \frac{3}{2}\delta$. For $|x| \leqq \delta$, and $|t| \geqq 4\delta$ we then have

$$|t-y| \geqq |t-(y-x)| - |x| \geqq \tfrac{1}{3}|t-(y-x)|,$$

since $|y-x| < \frac{5}{2}\delta$, and $|t-(y-x)| > \frac{3}{2}\delta \geqq \frac{3}{2}|x|$. Thus, for $|x| \leqq \delta$, and some A, (1) gives

$$\int_{|y|<\frac{3}{2}\delta} K(x-y)\,dm(y)\Big\{\int_{|t|\geqq 4\delta} K(t-y)\,dv(t)\Big\}^{p-1}$$
$$\leqq A \int_{|y|<\frac{3}{2}\delta} K(x-y)\,dm(y)\Big\{\int_{|t|\geqq 4\delta} K\big(t-(y-x)\big)\,dv(t)\Big\}^{p-1}$$
$$\leqq A \int_{|y-x|<\rho} K(x-y)\,dm(y)\Big\{\int K\big(t-(y-x)\big)\,dv(t)\Big\}^{p-1}$$
$$= A \int_{|y|<\rho} K(y)\,dm(y)\Big\{\int K(t-y)\,dv(t)\Big\}^{p-1},$$

where the last equality is simply a change of coordinates.

It follows from the finiteness of $V^{v}(0)$ that for any $\varepsilon > 0$

$$\limsup_{\delta \to 0} J'_\delta \leqq \varepsilon,$$

as soon as ρ is small enough. The theorem follows.

Remark. We have actually proved more, namely, that

$$V^{v}(0) = \liminf_{x \to 0,\, x \in E \setminus \{0\}} V^{v}(x)$$

for all sets E and measures v such that

$$\liminf_{\delta \to 0} \frac{1}{C_{K,q}(E_\delta)} \int (U^{v_{4\delta}})^p\,dm = 0. \tag{7}$$

Whether the converse to Theorem 2 or 2′ is true in general appears to be an open problem, but the following result can sometimes serve as a substitute. It is due to D. R. Adams (Univ. of Minnesota thesis, 1969), and, independently, to Havin and Maz'ja (see [24], where it is implicit in Theorem 4, and [25]). I am grateful to these mathematicians for communicating the theorem to me. A relatively simple proof, due to N. G. Meyers, is found in Adams [1].

Theorem 3. *Let μ be a positive measure, and let $\omega(\mu,\delta)=\sup_x \mu(B_x(\delta))$. Then, for $1<q<\infty$, $q\alpha\leqq d$, there is a constant A, depending only on q, α and d, such that for all x*

$$V^{\mu}_{\alpha,q}(x)\leqq A\int\left(\frac{\omega(\mu,\delta)}{\delta^{d-\alpha q}}\right)^{p-1}\frac{d\delta}{\delta}.$$

When $q>2-\frac{\alpha}{d}$, i.e. $(d-\alpha)(p-1)<d$, the situation is much simpler, and the work of the above authors gives the following (see [25, 1]).

Theorem 4. *Let $1<q<\infty$, $q\alpha\leqq d$, and $q>2-\frac{\alpha}{d}$. Then a Borel set E is (α,q)-thin at x if (and thus, by Theorem 2', if and only if)*

$$\int\left(\frac{C_{\alpha,q}(E\cap B_x(\delta))}{\delta^{d-\alpha q}}\right)^{p-1}\frac{d\delta}{\delta}<\infty. \tag{8}$$

It should be noted that Theorem 4 contains the linear case $p=2$.

We shall now draw some consequences of the definition of thinness.

Definition. (See Fuglede [18].) A function f is said to be *(K,q)-quasicontinuous* if it is defined (K,q)-q.e., and if for any $\varepsilon>0$ there exists an open set G with $C_{K,q}(G)<\varepsilon$ such that the restriction of f to $\complement G$ is continuous.

For classical capacities the following lemma is due to Deny [16].

Lemma 1. *Functions in $\mathscr{L}^q_K$ are (K,q)-quasicontinuous.*

Proof. Let $g=U^f_K$, $f\in L^q$, $f\geqq 0$. Then, by the definition of (K,q)-capacity

$$C_{K,q}\{U^f_K(x)>\lambda\}\leqq\frac{1}{\lambda^q}\|f\|^q_q.$$

The lemma follows as in [16].

Definition. (See Brelot [9].) A function g which is defined (K,q)-q.e. is said to be *(K,q)-pseudocontinuous* at x_0 if for all $\lambda>0$ the set $\{x;\,|g(x)-g(x_0)|\geqq\lambda\}$ is (K,q)-thin at x_0.

The following theorem is due to Fuglede [19, Theorem 3]. See also Deny [15, p. 171].

Theorem 5. *Every (K,q)-quasicontinuous function is (K,q)-pseudocontinuous (K,q)-quasi everywhere.*

Proof. For the reader's convenience we reproduce Fuglede's proof. First some notation.

For any set E the set of points where E is (K,q)-thin is denoted by $e(E)$. Thus $e(E)$ contains the exterior and part of the boundary of E.

The set $E \cup \complement(e(E))$ is denoted $\tilde{E}$. Clearly $C_{K,q}(E) = C_{K,q}(\tilde{E})$, since $E \subset \tilde{E}$ and the capacitary potential for E is $\geqq 1$ q.e. on $\tilde{E}$.

Consider all open intervals (α, β) with rational α and β, and let $\{A_n\}_1^\infty$ be an enumeration of their complements. Suppose g is (K, q)-quasi-continuous. Then, for any $\varepsilon > 0$ and each A_n there is an open set G with $C_{K,q}(G) < \varepsilon$ such that $g^{-1}(A_n) \setminus G$ is closed. Denote $g^{-1}(A_n)$ by B_n.

Then

$$\tilde{B}_n \subset ((B_n \setminus G) \cup G)^{\sim} \subset (B_n \setminus G)^{\sim} \cup \tilde{G} = (B_n \setminus G) \cup \tilde{G} \subset B_n \cup \tilde{G}.$$

Thus $\tilde{B}_n \setminus B_n \subset \tilde{G}$, and $C_{K,q}(\tilde{B}_n \setminus B_n) \leqq C_{K,q}(\tilde{G}) = C_{K,q}(G) < \varepsilon$. It follows that $C_{K,q}(\tilde{B}_n \setminus B_n) = 0$, and then also $C_{K,q}\left(\bigcup_{n=1}^{\infty} (\tilde{B}_n \setminus B_n)\right) = 0$.

But for all $x_0 \notin \bigcup_{n=1}^{\infty} (\tilde{B}_n \setminus B_n)$, every B_n such that $x_0 \notin B_n$ is thin at x_0. Thus for every $\varepsilon > 0$ the set $\{x; |g(x) - g(x_0)| \geqq \varepsilon\}$ is thin at x_0, which proves the theorem.

The following theorem is due to Choquet [14] in the classical case. Unfortunately we have not been able to extend it to all kernels K.

Theorem 6. *Let $E \subset \mathbb{R}^d$ be a Borel set, and let $q > 2 - \frac{\alpha}{d}$. Then for any $\varepsilon > 0$ there exists an open set G such that G contains the set $e(E)$ of points where E is (α, q)-thin, and $C_{\alpha,q}(E \cap G) < \varepsilon$.*

Proof. The proof follows Choquet [14]. Let $\{O_n\}_1^\infty$ be an enumeration of all balls $B_x(\delta)$ with rational x and δ. Let V_n be the capacitary potential for $E \cap O_n$, and let A_n be the subset of O_n where $V_n < 1$. By Theorem 4 it is clear that $\bigcup_1^\infty A_n = e(E)$.

By Lemma 1 and Theorem 1 there exists for given $\varepsilon > 0$ an open set ω_n such that ω_n contains the points of $E \cap O_n$ where $V_n < 1$, $C_{\alpha,q}(\omega_n) < \varepsilon 2^{-n}$, and V_n restricted to $\complement \omega_n$ is continuous. Let $F = E \setminus \bigcup_1^\infty \omega_n$, which implies that $\bar{F} \cap \left(\bigcup_1^\infty \omega_n\right) = \varnothing$. For all $x \in F$ and all n we have $V_n(x) \geqq 1$. Since V_n is continuous off ω_n it follows that $V_n(x) \geqq 1$ for all $x \in \bar{F}$. Thus $\bar{F} \cap e(E) = \varnothing$.

Let $G = \complement \bar{F}$. Then $e(E) \subset G$, and moreover,

$$G \cap E \subset \complement F \cap E \subset \bigcup_1^\infty \omega_n,$$

so $C_{\alpha,q}(G \cap E) \leqq \sum_1^\infty C_{\alpha,q}(\omega_n) \leqq \varepsilon$.

Theorem 6 has the following immediate corollaries [14].

Corollary 1. *The subset of E where E is* (α, q)*-thin,* $q>2-\frac{\alpha}{d}$*, has* (α, q)*-capacity zero, i.e.*

$$C_{\alpha,q}(E\cap e(E))=0.$$

Corollary 2. *A set which is* (α, q)*-thin,* $q>2-\frac{\alpha}{d}$*, at each of its points has* (α, q)*-capacity zero, and conversely.*

Theorem 7. *Let* $E\subset \mathbb{R}^d$ *be a Borel set. Suppose that the part of* $\bar{E}$ *where E is* (K, q)*-thin has* (K, q)*-capacity zero, i.e.* $C_{K,q}(\bar{E}\cap e(E))=0$*. Then* $C_{K,q}(E\cap G)=C_{K,q}(\bar{E}\cap G)$ *for all open G.*

If $q>2-\frac{\alpha}{d}$ *the following conditions are all equivalent.*

(a) $C_{\alpha,q}(E\cap G)=C_{\alpha,q}(\bar{E}\cap G)$ *for all open G.*

(b) *There is an* $\eta>0$ *such that* $C_{\alpha,q}(E\cap G)\geqq \eta\, C_{\alpha,q}(\bar{E}\cap G)$ *for all open G.*

(c) $C_{\alpha,q}(\bar{E}\cap e(E))=0$.

Proof. Suppose $C_{K,q}(\bar{E}\cap e(E))=0$. Then any bounded (K, q)-potential that is $\geqq 1$ (K, q)-quasi everywhere on $E\cap G$ is also $\geqq 1$ (K, q)-quasi everywhere on $\bar{E}\cap G$. This proves the first part of the theorem, and that (c) implies (a).

That (a) implies (b) is trivial. Now assume that $C_{\alpha,q}(\bar{E}\cap e(E))>0$, and let $\eta>0$ be arbitrary. By Theorem 6 there exists an open set G such that $e(E)\subset G$ and $C_{\alpha,q}(E\cap G)<\eta\, C_{\alpha,q}(\bar{E}\cap e(E))$. But $C_{\alpha,q}(\bar{E}\cap e(E)) \leqq C_{\alpha,q}(\bar{E}\cap G)$, which proves that (b) implies (c).

Theorems 2 and 4 immediately give quantitative corollaries to the above Theorems 5–7. The following consequences of Theorem 3 will be of use later in the absence of a converse to Theorem 2.

Theorem 8. *Let* $E\subset \mathbb{R}^d$ *be a Borel set,* $1<q<\infty$*, and* $q\alpha<d$*. Let h be an increasing function such that*

$$\int_0 \left(\frac{h(\delta)}{\delta^{d-\alpha q}}\right)^{p-1} \frac{d\delta}{\delta}<\infty.$$

Let F be the set of all points $x\in E$ *such that*

$$\limsup_{\delta\to 0} \frac{C_{\alpha,q}(E\cap B_x(\delta))}{h(\delta)}<\infty.$$

Then $C_{\alpha,q}(F)=0$.

Proof. Suppose $C_{\alpha,q}(F)>0$. Then there exist an $A<\infty$, a $\delta_0>0$, and an $F'\subset F$ such that $C_{\alpha,q}(F')>0$ and $C_{\alpha,q}(E\cap B_x(\delta))\leqq A\, h(\delta)$ for $\delta\leqq\delta_0$ and $x\in F'$. We can also assume that F' is compact and that its diameter is less than any given $r>0$.

Let $\varepsilon>0$ be given, and choose r so small that $r\leqq\delta_0$, and

$$\int_0^r \left(\frac{h(\delta)}{\delta^{d-\alpha q}}\right)^{p-1}\frac{d\delta}{\delta}+h(r)^{p-1}\int_r^{\prime}\frac{d\delta}{\delta^{(d-\alpha q)(p-1)+1}}<\varepsilon,$$

which is possible by assumption.

Let $V_{\alpha,q}^{\nu}$ be the capacitary potential for F', $\nu(F')=C_{\alpha,q}(F')$. Then

$$\nu(F'\cap B_x(\delta))\leqq C_{\alpha,q}(F'\cap B_x(\delta))\leqq C_{\alpha,q}(E\cap B_x(\delta))\leqq A\,h(2\delta)$$

for all x and $\delta<\delta_0/2$, and $\nu(F'\cap B_x(\delta))\leqq A\,h(r)$ otherwise. It follows from Theorem 3 that $V_{\alpha,q}^{\nu}(x)\leqq C\varepsilon$ for all x and some C. If $C\varepsilon<1$ this is a contradiction.

Corollary 1. *Let h be as in Theorem 8. Let E be a Borel set such that for all $x\in E$*

$$\limsup_{\delta\to 0}\frac{C_{\alpha,q}(E\cap B_x(\delta))}{h(\delta)}<\infty.$$

Then $C_{\alpha,q}(E)=0$.

Corollary 2. *Let h be as in Theorem 8. Let E be a Borel set and suppose that there is a set $F\subset\bar{E}$ such that $C_{\alpha,q}(F)>0$, and such that for all $x\in F$*

$$\limsup_{\delta\to 0}\frac{C_{\alpha,q}(E\cap B_x(\delta))}{h(\delta)}<\infty.$$

Then, for any $\eta>0$, there is a ball $B_x(\delta)$ such that

$$C_{\alpha,q}(E\cap B_x(\delta))<\eta\, C_{\alpha,q}(\bar{E}\cap B_x(\delta)).$$

Proof of Corollary 2. Suppose there is an $\eta>0$ so that

$$C_{\alpha,q}(E\cap B_x(\delta))\geqq\eta\, C_{\alpha,q}(\bar{E}\cap B_x(\delta))$$

for all x and δ. This leads to a contradiction in the same way as in Theorem 8.

For everywhere dense sets Theorem 7 can be sharpened.

Theorem 9. *Let K be as in Theorem 2, and let $E\subset\mathbb{R}^d$ be an everywhere dense Borel set. Then the following are equivalent.*

(a) $C_{K,q}(E\cap G)=C_{K,q}(G)$ *for every open G.*

(b) $C_{K,q}(E\cap B_x(\delta))=C_{K,q}(B_x(\delta))$ *for all x and δ.*

(See Remark 1 to Theorem 2.)

(c) *For almost all x (with respect to Lebesgue measure)*

$$\limsup_{\delta\to 0}\frac{C_{K,q}(E\cap B_x(\delta))}{\delta^d}>0. \tag{9}$$

Remark. A slightly weaker theorem for $\alpha=1$, $d=2$ was proved in Hedberg [27]. In that paper references to earlier work on the "instability of capacity" are also given. For $\alpha=2$ the above theorem was proved in the report [29]. The proofs given in [27] and [29] have recently been extended by Polking [36] to prove a result similar to the above for Bessel kernels.

Proof. It is clearly enough to prove that (c) implies (a). The basic idea of the proof is the same as in the proof of Theorem 2 above.

Let G be open and assume (9) holds almost everywhere in G. Let $U_K^f=U^f$ belong to $\mathscr{L}_K^q$, i.e. $f\in L^q$, and assume $U^f\geqq 1$ (K,q)-quasi everywhere on $E\cap G$. We claim that $U^f\geqq 1$ almost everywhere, and thus everywhere, in G. This will clearly prove the theorem.

Choose $x_0\in G$ so that (9) holds at x_0, so that

$$\int K(x_0-x)|f(x)|\,dm(x)<\infty, \tag{10}$$

and

$$\lim_{\delta\to 0}\delta^{-d}\int_{|x-x_0|<\delta}|f(x)-f(x_0)|^q\,dm(x)=0. \tag{11}$$

Here (11) is true for almost all $x_0\in G$ by a simple modification of the Lebesgue differentiation theorem (see e.g. Stein [39; Ch. I.5.7]).

Set $x_0=0$. For any $\delta>0$ we can find a probability measure ν with support in $E\cap B_0(\delta)$ such that

$$\|U_K^\nu\|_p\leqq 2\,C_{K,q}(E\cap B_0(\delta))^{-1/q}. \tag{12}$$

Then, if $B_0(\delta)\subset G$,

$$1\leqq\int U^f\,d\nu=\int f\,U^\nu\,dm.$$

We shall show that for a suitable sequence $\{\delta_i\}$

$$\lim_{\delta_i\to 0}\int f\,U^\nu\,dm=U^f(0).$$

Choose $\varepsilon>0$ and $\rho>0$ so that

$$\int_{|x|<\rho}K(x)|f(x)|\,dm(x)<\varepsilon. \tag{13}$$

Let δ be arbitrary, $0<\delta<\rho/2$. Then

$$\begin{aligned}\left|\int f\,U^\nu\,dm-U^f(0)\right|&=\left|\int f(x)\big(U^\nu(x)-K(x)\big)\,dm(x)\right|\\&\leqq\int_{|x|<\rho}K(x)|f(x)|\,dm(x)+\left|\int_{|x|<2\delta}f(x)\,U^\nu(x)\,dm(x)\right|\\&\quad+\int_{2\delta\leqq|x|<\rho}|f(x)|\,U^\nu(x)\,dm(x)\\&\quad+\int_{|x|\geqq\rho}|f(x)|\,|U^\nu(x)-K(x)|\,dm(x)\\&=I_1+|I_2|+I_3+I_4.\end{aligned}$$

By (13) $I_1 < \varepsilon$, and for $|x| \geqq \rho$ $\lim_{\delta \to 0} U^\nu(x) = K(x)$ uniformly, so $\lim_{\delta \to 0} I_4 = 0$.

$$I_2 = \int_{|x|<2\delta} (f(x) - f(0))\, U^\nu(x)\, dm(x) + f(0) \int_{|x|<2\delta} U^\nu(x)\, dm(x) = I_2' + I_2'',$$

$$|I_2'| \leqq \Big\{ \int_{|x|<2\delta} |f(x) - f(0)|^q\, dm(x) \Big\}^{1/q} \|U^\nu\|_p$$
$$\leqq 2 \Big\{ \delta^{-d} \int_{|x|<2\delta} |f(x) - f(0)|^q\, dm(x) \Big\}^{1/q} \{\delta^{-d}\, C_{K,q}(E \cap B_0(\delta))\}^{-1/q}$$

by (12). Thus $\lim_{\delta_i \to 0} I_2' = 0$ by (11) and (9) for a suitable sequence $\{\delta_i\}$. $\operatorname{Lim}_{\delta \to 0} I_2'' = 0$, since

$$|I_2''| \leqq |f(0)| \int d\nu(y) \int_{|x|<2\delta} K(x-y)\, dm(x) \leqq |f(0)| \int_{|x|<3\delta} K(x)\, dm(x).$$

Finally

$$I_3 = \int d\nu(y) \int_{2\delta \leqq |x| < \rho} K(x-y) |f(x)|\, dm(x)$$
$$\leqq \int d\nu(y) \int_{|x|<\rho} K(x/2) |f(x)|\, dm(x) \leqq A \int_{|x|<\rho} K(x) |f(x)|\, dm(x) < A\varepsilon$$

by (1) and (13). Since ε is arbitrary the theorem follows.

The proof also gives the following.

Corollary. *Let $g \in \mathscr{L}_K^q$, and suppose that g vanishes (K, q)-quasi everywhere on a set E satisfying either of the conditions in Theorem 9. Then $g \equiv 0$. I.e. $\complement E$ is a set of uniqueness for $\mathscr{L}_K^q$.*

3. Approximation in the Mean by Analytic Functions

In this section we specialize the results of the previous section to $\mathscr{L}_1^q$, or equivalently W_1^q. In this case more can be said, because these spaces are closed under truncation.

If G is open we denote by $\mathring{W}_1^q(G)$ the closure in $W_1^q(\mathbb{R}^d)$ of the C^∞ functions with compact support in G. The interior of a set E is denoted E^0. For a set $E \subset \mathbb{C}$ with positive Lebesgue measure we denote by $L_a^p(E)$ the subspace of $L^p(E)$ which consists of functions analytic in E^0, i.e. $f \in L_a^p(E)$ if $\int_E |f(z)|^p\, dm(z) < \infty$ and $\frac{\partial f}{\partial \bar z} = 0$ in E^0. The closure in $L^p(E)$ of the rational functions with poles off $\bar E$ (or equivalently the closure of the functions analytic in neighborhoods of $\bar E$) is denoted $R^p(E)$. We give several conditions for when functions in $W_1^q(\mathbb{R}^d)$, whose (compact) supports belong to a set E, actually belong to $\mathring{W}_1^q(E^0)$. By duality we get conditions for when $R^p(E) = L_a^p(E)$, and similarly for approximation in W_1^p by harmonic functions. For $1 < p \leqq 2$ $(p \leqq d/(d-1))$ the results are essentially due to Havin [22], who also pointed out the need for a study of the fine continuity properties of functions in W_1^q in order to solve the problem in general. Bagby [7] gave a necessary and sufficient condition for the above approximation property to be true. By applying the conditions for thinness given above we get more effective conditions which in the general case are slightly short of being necessary and sufficient, and improve those given earlier by the author [27]. We also show that for open sets G analytic functions of the form $\mu * z^{-1}$, where μ are complex-valued measures on $\complement G$, are always dense in $L_a^p(G)$. See the remarks below for a more detailed discussion of earlier work.

We write $\mu * z^{-1} = \hat{\mu}$, the Cauchy transform of μ. Similarly the Cauchy transform of a function f is $\hat{f} = f * z^{-1}$. From now on we will write C_q instead of $C_{1,q}$, q-thin instead of $(1, q)$-thin, etc.

The following lemma will be essential to what follows. A proof can be found e.g. in Deny-Lions [17; p. 316].

Lemma 2. *Let* $f \in W_1^q(\mathbb{R}^d)$, $1 \leqq q < \infty$. *Then*

$$f^+ = \operatorname{Max}(f, 0) \in W_1^q(\mathbb{R}^d),$$

and

$$\int |\operatorname{grad} f^+|^q \, dm = \int_{f>0} |\operatorname{grad} f|^q \, dm \leqq \int |\operatorname{grad} f|^q \, dm.$$

We take the opportunity to clarify the relationship between C_q and various other equivalent capacities. We will assume that $1 < q < d$, and omit the easy modifications needed for the case $q = d$. For $q > d$ all non-empty sets have positive q-capacity.

Definition. Let $E \subset \mathbb{R}^d$ be an arbitrary set. Set $\Gamma_q(E) = \inf_\omega \int |\operatorname{grad} \omega|^q \, dm$, where the infimum is taken over all $\omega \in W_1^q$ such that $\omega \geqq 1$ q-q.e. on E.

By Lemma 2 it is equivalent to consider ω such that $\omega = 1$ q-q.e. on E.

This capacity is well known. It was introduced by Maz'ja [32] and has been studied and used by several authors. See e.g. [6, 7, 23, 27, 28, 33, 38, 41, 42].

It is an immediate consequence of the equivalence of norms in $\mathscr{L}_1^q$ and W_1^q that Γ_q is equivalent to C_q, i.e. there are constants A_1 and A_2 independent of E such that

$$A_1 C_q(E) \leqq \Gamma_q(E) \leqq A_2 C_q(E).$$

In the complex plane $\mathbb{C}$ we also define the following capacities, in addition to c_q, C_q and Γ_q.

Definition. (a) Let $E \subset \mathbb{C}$ be arbitrary. Set

$$\tilde{\Gamma}_q(E) = \inf_\omega \int \left| \frac{\partial \omega}{\partial \bar{z}} \right|^q dm,$$

where the infimum is taken over all complex-valued $\omega \in W_1^q$ such that $\omega = 1$ q-q.e. on E.

(b) Let $E \subset \mathbb{C}$ be compact. Set $\gamma_q(E) = \sup_\mu \mu(E)$ where the supremum is taken over positive measures μ on E such that $\|\mu * z^{-1}\|_p \leqq 1$.

(c) Let $E \subset \mathbb{C}$ be compact. Set

$$\tilde{\gamma}_q(E) = \sup_f |f'(\infty)| = \sup_f \left|\lim_{z\to\infty} z f(z)\right|,$$

where the supremum is taken over all $f \in L_a^p(\complement E)$ with $\|f\|_p \leqq 1$. ($\tilde{\gamma}_q(E)$ is an "analytic q-capacity". See [27].)

For arbitrary sets F we define $\gamma_q(F) = \sup_{E \subset F} \gamma_q(E)$, E compact, and similarly for $\tilde{\gamma}_q$.

Lemma 3. *Let $E \subset \mathbb{C}$ be a Borel set. Then there is a constant A not depending on E such that*

$$c_q(E) \leqq \gamma_q(E) \leqq \tilde{\gamma}_q(E) = \frac{1}{\pi} \tilde{\Gamma}_q(E)^{1/q} \leqq \frac{1}{2\pi} \Gamma_q(E)^{1/q} \leqq A C_q(E)^{1/q} = A c_q(E).$$

The following corollary is known, see Havin-Maz'ja [24].

Corollary. *Let E be compact. Then $L_a^p(\complement E)$, $2 \leqq p < \infty$, is non-trivial if and only if $c_q(E) > 0$, i.e. if and only if there exists a non-zero measure μ on E such that $\|\mu * |z|^{-1}\|_p < \infty$.*

Proof of Lemma 3. It is enough to assume E is compact. That $c_q(E) = C_q(E)^{1/q}$ is part of Theorem 1. It is clear that $c_q(E) \leqq \gamma_q(E) \leqq \tilde{\gamma}_q(E)$, and that $\tilde{\Gamma}_q(E) \leqq 2^{-q} \Gamma_q(E) \leqq A C_q(E)$.

For any differentiable function ω with compact support, such that $\omega = 1$ in a neighborhood of E, and a suitable curve c, we find

$$\begin{aligned}|f'(\infty)| &= \frac{1}{2\pi}\left|\int_c f(z)\,dz\right| = \frac{1}{2\pi}\left|\int_c f(z)\,\omega(z)\,dz\right| \\ &= \frac{1}{2\pi}\left|\int \frac{\partial(f\omega)}{\partial \bar{z}}\, d\bar{z} \wedge dz\right| = \frac{1}{\pi}\left|\int f\,\frac{\partial \omega}{\partial \bar{z}}\, dm\right| \leqq \frac{1}{\pi} \|f\|_p \left\|\frac{\partial \omega}{\partial \bar{z}}\right\|_q.\end{aligned}$$

Thus $\tilde{\gamma}_q(E) \leqq \frac{1}{\pi} \tilde{\Gamma}_q(E)^{1/q}$. That equality holds was proved in [27, Lemma 1].

The following lemma is well-known for $q > d$, and due to Bagby [7] for $q < d$. We give a proof which differs somewhat from that of Bagby.

Lemma 4. *Let $\varphi \in W_1^q(\mathbb{R}^d)$, $1 \leqq q < \infty$. Suppose $\varphi = 0$ q-q.e. on a compact set F with $C_q(F) > 0$ ($F \neq \varnothing$ for $q > d$). Then $\varphi \in \mathring{W}_1^q(\complement F)$.*

Proof. It is no restriction to assume that φ has compact support. By Lemma 2 we can assume that $\varphi \geqq 0$, and it follows from the same lemma that we can assume that φ is bounded.

We first assume $q>d$. Then φ is continuous. Let $\varepsilon>0$ and define $\varphi_\varepsilon=(\varphi-\varepsilon)^+$. Then $\varphi_\varepsilon=0$ in a neighborhood of F, and

$$\int |\operatorname{grad}(\varphi-\varphi_\varepsilon)|^q\,dm=\int_{0<\varphi<\varepsilon} |\operatorname{grad}\varphi|^q\,dm,$$

which tends to zero with ε.

Now let $q\leqq d$. By Lemma 1 there is for every n an open set G_n such that $C_q(G_n)<\frac{1}{n}$, the restriction of φ to $\complement G_n$ is continuous, and $\varphi=0$ on $F\setminus G_n$. There exist $\omega_n\in W_1^q$ such that $\omega_n=1$ on G_n and $\int|\operatorname{grad}\omega_n|^q\,dm\leqq A/n$ for some A. Let $\psi_n=\left(\varphi-\frac{1}{n}\right)^+$, and set $\varphi_n=\psi_n(1-\omega_n)$. Then $\varphi_n\in W_1^q$, and $\varphi_n=0$ in a neighborhood of F. And

$$\|\operatorname{grad}(\varphi-\varphi_n)\|_q\leqq\|\operatorname{grad}(\varphi-\psi_n)\|_q+\|\psi_n\operatorname{grad}\omega_n\|_q+\|\omega_n\operatorname{grad}\psi_n\|_q.$$

As $n\to\infty$ the first term tends to zero as above, and the second term is $\leqq A\operatorname{Max}|\varphi|\,n^{-1/q}$. The third term is $\leqq\|\omega_n\operatorname{grad}\varphi\|_q$, which tends to zero since ω_n clearly tends to zero in measure. This proves the lemma.

Theorem 10. *Let $1\leqq p<\infty$. Let $G\subset\mathbb{C}$ be open with compact boundary and suppose $C_q(\complement G)>0$.* ($\complement G\neq\varnothing$ for $p<2$.) *Suppose G is bounded if $p\leqq 2$. Then the Cauchy transforms, $\hat\mu(z)=\mu*z^{-1}$, of complex-valued measures μ supported by $\complement G$ are dense in $L_a^p(G)$.*

Remark. For $1\leqq p<2$ the theorem is known and due to Bers [8] and Havin [22]. See also [27].

Proof. Suppose g is a complex-valued function in $L^q(G)$ such that $\int g\hat\mu\,dm=0$ for all μ supported by $\complement G$ such that $\|\mu*|z|^{-1}\|_p<\infty$. Then by Fubini's theorem $\int\hat g\,d\mu=0$ for all such μ. By Lemma 3 this implies that $\hat g=0$ q-q.e. on $\complement G$. But $\hat g\in W_1^q$, and by Lemma 4 the real and imaginary parts of $\hat g$ can be approximated in W_1^q by functions with compact support in G. For every such function φ_ε and every $f\in L_a^p(G)$ we have

$$\int\frac{\partial\varphi_\varepsilon}{\partial\bar z}f\,dm=\int\frac{\partial}{\partial\bar z}(\varphi_\varepsilon f)\,dm=0$$

by Green's theorem. Thus

$$\int gf\,dm=\pi\int\frac{\partial\hat g}{\partial\bar z}f\,dm=\int\left(\frac{\partial}{\partial\bar z}(\hat g-\varphi_\varepsilon)\right)f\,dm,$$

so

$$\left|\int gf\,dm\right|\leqq\|\operatorname{grad}(\hat g-\varphi_\varepsilon)\|_q\,\|f\|_p$$

which is arbitrarily small. This proves the theorem.

If $p<2$ the measures on the complement of G in Theorem 10 can always be assumed to be point masses. We shall now investigate for $p \geqq 2$ when the measures on $\complement G$ can be replaced by point masses on the exterior of G. We will assume that G is the interior of $\bar{G}$, and we are therefore going to consider compact sets E.

Theorem 11. *Let $E \subset \mathbb{C}$ be compact, and suppose $2 \leqq p < \infty$. Then the following are equivalent.*

(a) $R^p(E) = L_a^p(E)$.

(b) *If $\varphi \in W_1^q$ and $\varphi = 0$ q-q.e. on $\complement E$, then $\varphi \in \mathring{W}_1^q(E^0)$.*

(c) *If $\varphi \in W_1^q$ and $\varphi = 0$ q-q.e. on $\complement E$, then $\varphi = 0$ q-q.e. on ∂E.*

(d) *$C_q(G \smallsetminus E) = C_q(G \smallsetminus E^0)$ for all open G.*

(e) *For some $\eta > 0$ $C_q(G \smallsetminus E) \geqq \eta\, C_q(G \smallsetminus E^0)$ for all open G.*

Remark. The theorem is essentially known. The equivalence of (a) and (b) is well known. See e.g. Havin-Maz'ja [23]. The equivalence of (a), (c) and (d) was first proved by Bagby [6, 7]. See also [27], where further references are given. The theorem and the corollaries below should be compared to the results of Vituškin [40] for uniform rational approximation.

Proof. (c) implies (b) by Lemma 4, and that (b) implies (a) follows as in the proof of Theorem 10. We will prove that (e) implies (c), (a) implies (c), (b) implies (c), and that (c) implies (d).

To prove that (e) implies (c) we reproduce an argument used by Bagby [7] (see also Fuglede [19, Lemma 1]).

A set U is called quasi open if for every $\varepsilon > 0$ there is an open G with $C_q(G) < \varepsilon$, such that $U \smallsetminus G$ is open in $\complement G$. Then (e) implies that $C_q(U \smallsetminus E) \geqq \eta\, C_q(U \smallsetminus E^0)$ for all quasi open U. In fact, let G be open, $C_q(G) < \varepsilon$, and $U \smallsetminus G$ open in $\complement G$. Then

$$\begin{aligned} C_q(U \smallsetminus E^0) &\leqq C_q((U \cup G) \smallsetminus E^0) \\ &\leqq \eta^{-1} C_q((U \cup G) \smallsetminus E) \leqq \eta^{-1}(C_q(U \smallsetminus E) + C_q(G)). \end{aligned}$$

Now, $\varphi \in W_1^q$ implies that φ is quasi continuous, i.e. $\{|\varphi| > 0\}$ is quasi open. Let this set be U. Then

$$C_q(U \smallsetminus E^0) \leqq \eta^{-1} C_q(U \smallsetminus E) = 0,$$

which proves (c).

Suppose (c) does not hold. Then there exists a $\varphi \in W_1^q$ such that $\varphi = 0$ q-q.e. off E, but $\varphi > 0$ on a compact subset K of ∂E with $C_q(K) > 0$. There is by Lemma 3 a positive (non-zero) measure μ supported by K such that $\hat{\mu} \in L_a^p(E)$, and even $\|\mu * |z|^{-1}\|_p < \infty$. Then $\int \varphi\, d\mu > 0$, but $\int \psi\, d\mu = 0$ for all ψ with support in E^0, so $\varphi \notin \mathring{W}_1^q(E^0)$, and thus (b) implies (c). Moreover, $\varphi=\frac{1}{\pi}\frac{\partial\varphi}{\partial\bar z}*z^{-1}$, so $\int\varphi\,d\mu=-\frac{1}{\pi}\int\frac{\partial\varphi}{\partial\bar z}\hat\mu\,dm$. Thus $\frac{\partial\varphi}{\partial\bar z}$ does not annihilate all of $L^p_a(E)$ although it does annihilate all of $R^p(E)$ since $\varphi(z)=0$ off E. Thus (a) implies (c).

Now assume that (d) does not hold. Then there is an open set G such that $C_q(G\setminus E)<C_q(G\setminus E^0)$. Let φ be the q-capacitary potential for $G\setminus E$, and let $\psi=\min(\varphi,1)$. Then $\psi\in W^q_1$ by Lemma 2. There is a compact $K\subset\partial E\cap G$ with $C_q(K)>0$ where $\psi<1$, since otherwise $C_q(G\setminus E)=C_q(G\setminus E^0)$. Let ω be any function in $C^\infty_0(G)$ which is 1 in a neighborhood of K. Then $\omega(1-\psi)\in W^q_1$, $\omega(1-\psi)=0$ on $\complement E$, and $\omega(1-\psi)>0$ on K. Thus (c) implies (d). The theorem is proved.

Theorems 2′ and 9 now give the following corollaries to the above theorem.

Corollary 1. *Let $E\subset\mathbb{C}$ be compact without interior, and let $2\leqq p<\infty$. Then the following are equivalent.*

(a) $R^p(E)=L^p(E)$.

(b) $\varphi\in W^q_1$, *and* $\varphi=0$ *q-q.e. off E implies $\varphi\equiv 0$, i.e. E is a set of uniqueness for W^q_1.*

(c) $C_q(G\setminus E)=C_q(G)$ *for all open G.*

(d) $C_q(B_z(\delta)\setminus E)=C_q(B_z(\delta))$ *for all balls* $B_z(\delta)$. (Note that $C_q(B_z(\delta))\approx\delta^{2-q}$, $q<2$, $C_2(B_z(\delta))\approx 1/\log 1/\delta$.)

(e) $\limsup_{\delta\to 0}\frac{C_q(B_z(\delta)\setminus E)}{\delta^2}>0$ *for almost all z.*

Remark. For $d=1$ and $q=2$ the equivalence of (b) and (c) is due to Ahlfors and Beurling [5, p. 124]. A related problem is solved in Carleson [13; Theorem VI.2]. The sets of uniqueness for $\mathcal{L}^2_\alpha(\mathbb{R}^1)$, $\alpha<1$, have been similarly characterized by Carleson. See [26, p. 79]. A slightly weaker result than the equivalence of (a), (c), (d) and (e) was proved by the author in [27, Theorems 2 and 3]. See also Bagby [6] and Polking [36].

Corollary 2. *Let E be compact, $2\leqq p<\infty$. Then $R^p(E)=L^p_a(E)$ if (if and only if for $p<3$) the part of ∂E where $\complement E$ is q-thin has q-capacity zero. Thus $R^p(E)=L^p_a(E)$ if (if and only if for $p<3$)*

$$\int_0\left(\frac{C_q(B_z(\delta)\setminus E)}{\delta^{2-q}}\right)^{p-1}\frac{d\delta}{\delta}=\infty$$

for q-quasi all $z\in\partial E$.

Proof. To prove Corollary 2 one only has to combine Theorem 2′ and Theorem 5 or Theorem 7 with Theorem 11. It is also easy to see that on $\partial E\setminus\partial(E^0)$ it is enough to assume that

$$\limsup_{\delta\to 0} \frac{C_q(B_z(\delta)\setminus E)}{\delta^2} > 0$$

for almost all z.

Corollary 3. *Let E be compact, $2 \leqq p < \infty$. Then $R^p(E) = L^p_a(E)$ if the inner boundary $\partial' E$ (i.e. the points on ∂E which are not on the boundary of any component of $\complement E$) has $C_q(\partial' E) = 0$.*

Remark. For $p=2$ Corollary 2 is due to Havin [22]. Instead of applying Theorem 2′ we could also have applied the Wiener criterion of Maz'ja [33] referred to earlier. A weaker form of the corollary was proved in [27] by yet another method. In that paper Corollary 3 was also given.

Corollary 4. *Let E be compact, $2 \leqq p < \infty$. Then $R^p(E) \neq L^p_a(E)$ if there is an increasing function h such that*

$$\int_0 \left(\frac{h(\delta)}{\delta^{2-q}}\right)^{p-1} \frac{d\delta}{\delta} < \infty$$

and a set $K \subset \partial E$ with $C_q(K) > 0$ such that for all $z \in K$

$$\limsup_{\delta\to 0} \frac{C_q(B_z(\delta)\setminus E)}{h(\delta)} < \infty .$$

Remark. A weaker version of this corollary was given in [27].

Proof. Just apply Corollary 2 of Theorem 8.

Finally we note that the equivalence of (b)–(e) in Theorem 11 also gives approximation theorems in higher dimensions.

If $G \subset \mathbb{R}^d$ is open we denote by $W^p_{1,h}(G; \mathbb{R}^d)$ the subspace of $W^p_1(\mathbb{R}^d)$ which consists of functions harmonic in G. The following theorem is a counterpart to Theorem 10.

Theorem 12. *Let $1 \leqq p < \infty$. Let $G \subset \mathbb{R}^d$ be open with compact boundary, and suppose $C_q(\complement G) > 0$ if $p \geqq \frac{d}{d-1}$. Suppose G is bounded if $p \leqq \frac{d}{d-1}$. Then potentials $\mu * |x|^{2-d}$ of signed measures μ supported by $\complement G$ are dense in $W^p_{1,h}(G; \mathbb{R}^d)$.*

Now let E be compact. Denote by $W^p_{1,h}|_E$ the space of restrictions to E of functions in $W^p_{1,h}(E^0; \mathbb{R}^d)$, normed by the usual quotient norm, $\|f\|_E = \inf_g \|g\|$ for $g=f$ on E. $D^p_1|_E$ is the closure in this norm of functions harmonic in neighborhoods of E.

Theorem 13. *Let $E \subset \mathbb{R}^d$ be compact. Then $D^p_1|_E = W^p_{1,h}|_E$ if $1 \leqq p < \frac{d}{d-1}$. If $p \geqq \frac{d}{d-1}$, $D^p_1|_E = W^p_{1,h}|_E$ if and only if one of the equivalent conditions (b)–(e) in Theorem 11 holds.*

We omit formulating the corollaries corresponding to Corollaries 1–4 above.

The problem of approximation in L^p by harmonic functions and by solutions of other elliptic equations leads by duality to problems of approximation and continuity in the Sobolev spaces W_α^p, where α is an integer $\geqq 2$. Here only partial results are known, except in the case of sets without interior. See Polking [36] and the author [30].

References

1. Adams, D.R.: Traces of potentials. II. Indiana Univ. Math. J., to appear.
2. Adams, D.R., Meyers, N.G.: Bessel potentials. Inclusion relations among classes of exceptional sets. Bull. Amer. Math. Soc. **77**, 968-970 (1970).
3. Adams, D.R., Meyers, N.G.: Bessel potentials. Inclusion relations among classes of exceptional sets. Indiana Univ. Math. J., to appear.
4. Adams, D.R., Polking, J.C.: The equivalence of two definitions of capacity. Proc. Amer. Math. Soc., to appear.
5. Ahlfors, L., Beurling, A.: Conformal invariants and function-theoretic null-sets. Acta Math. **83**, 101–129 (1950).
6. Bagby, T.: L_p approximation by analytic functions. J. Approximation Theory **5**, 401-404 (1972).
7. Bagby, T.: Quasi topologies and rational approximation. J. Functional Analysis **10**, 259-268 (1972).
8. Bers, L.: An approximation theorem. J. Analyse Math. **14**, 1-4 (1965).
9. Brelot, M.: Sur les ensembles effilés. Bull. Sci. Math. **68**, 12-36 (1944).
10. Brelot, M.: Introduction axiomatique de l'effilément. Ann. Mat. Pura Appl. **59**, 77–95 (1962).
11. Calderón, A.P.: Lebesgue spaces of differentiable functions and distributions. Proc. Sympos. Pure Math. **4**, 33–49. Providence, R.I.: Amer. Math. Soc. 1961.
12. Carleson, L.: Mergelyan's theorem on uniform polynomial approximation. Math. Scandinav. **15**, 167–175 (1964).
13. Carleson, L.: Selected problems on exceptional sets. Princeton, N.J.: Van Nostrand 1967.
14. Choquet, G.: Sur les points d'effilément d'un ensemble. Application à l'étude de la capacité. Ann. Inst. Fourier (Grenoble) **9**, 91–101 (1959).
15. Deny, J.: Les potentiels d'energie finie. Acta Math. **82**, 107–183 (1950).
16. Deny, J.: Sur la convergence de certaines intégrales de la théorie du potentiel. Arch. der Math. **5**, 367–370 (1954).
17. Deny, J., Lions, J.L.: Les espaces du type de Beppo Levi. Ann. Inst. Fourier (Grenoble) **5**, 305–370 (1953–1954).
18. Fuglede, B.: Le théorème du minimax et la théorie fine du potentiel. Ann. Inst. Fourier (Grenoble) **15**, 65–88 (1965).
19. Fuglede, B.: Quasi topology and fine topology. Séminaire Brelot-Choquet-Deny, 10e année (1965–1966).
20. Fuglede, B.: Applications du théorème minimax à l'étude de diverses capacités. C.R. Acad. Sci. Paris, Sér. A **266**, 921–923 (1968).
21. Harvey, R., Polking, J.C.: A notion of capacity which characterizes removable singularities. Trans. Amer. Math. Soc. **169**, 183–195 (1972).
22. Havin, V.P.: Approximation in the mean by analytic functions. Doklady Akad. Nauk SSSR **178**, 1025-1028 (1968). (Soviet Math. Doklady **9**, 245-248 (1968).)

23. Havin, V. P., Maz'ja, V. G.: On approximation in the mean by analytic functions. Vestnik Leningrad. Univ. **23**, no. **13**, 62–74 (1968) [Russian].
24. Havin, V. P., Maz'ja, V. G.: A non-linear analogue of the Newtonian potential and metric properties of the (p, l)-capacity. Doklady Akad. Nauk SSSR **194**, 770–773 (1970). (Soviet Math. Doklady **11**, 1294–1298 (1970).)
25. Havin, V. P., Maz'ja, V. G.: Non-linear potential theory. [Russian.] Uspehi Mat. Nauk, **26** (6) 67–138 (1972). Applications of (p, l)-capacity to some problems in the theory of exceptional sets. [Russian.] Mat. Sb., to appear.
26. Hedberg, L. I.: The Stone-Weierstrass theorem in certain Banach algebras of Fourier type. Ark. Mat. **6**, 77–102 (1965).
27. Hedberg, L. I.: Approximation in the mean by analytic functions. Trans. Amer. Math. Soc. **163**, 157–171 (1972).
28. Hedberg, L. I.: Bounded point evaluations and capacity. J. Functional Analysis **10**, 269–280 (1972).
29. Hedberg, L. I.: Approximation in the mean by analytic and harmonic functions and capacities. Preliminary report. Uppsala Univ. Dept. of Math. Feb. 1971.
30. Hedberg, L. I.: Approximation in the mean by solutions of elliptic equations. Duke Math. J., to appear.
31. Landkof, N. S.: Foundations of modern potential theory. Moscow: Nauka 1966. (English translation, Springer 1972.)
32. Maz'ja, V. G.: p-conductance and theorems of imbedding certain function spaces into the space C. Doklady Akad. Nauk SSSR **140**, 299–302 (1961). (Soviet Math. Doklady **2**, 1200–1203 (1961).)
33. Maz'ja, V. G.: On the continuity at a boundary point of the solution of quasi-linear elliptic equations. [Russian.] Vestnik Leningrad. Univ. **25** no. **13**, 42–55 (1970). Correction, ibid. **27** no. **1**, 160 (1972).
34. Meyers, N. G.: A theory of capacities for potentials of functions in Lebesgue classes. Math. Scandinav. **26**, 255–292 (1970).
35. Meyers, N. G.: Continuity of Bessel potentials. Israel J. Math. **11**, 271–283 (1972).
36. Polking, J. C.: Approximation in L^p by solutions of elliptic partial differential equations. Amer. J. Math., to appear.
37. Rešetnjak, Ju. G.: On the concept of capacity in the theory of functions with generalized derivatives. Sibirsk. Mat. Ž. **10**, 1109–1138 (1969).
38. Serrin, J.: Local behavior of solutions of quasi-linear equations. Acta Math. **111**, 247–302 (1964).
39. Stein, E. M.: Singular integrals and differentiability properties of functions. Princeton Univ. Press 1970.
40. Vituškin, A. G.: Analytic capacity of sets in problems of approximation theory. Uspehi Mat. Nauk **22**, 141–199 (1967). (Russian Math. Surveys **22**, 139–200 (1967).)
41. Wallin, H.: A connection between α-capacity and L^p-classes of differentiable functions. Ark. Mat. **5**, 331–341 (1964).
42. Ziemer, W. P.: Extremal length as a capacity. Michigan Math. J. **17**, 117–128 (1970).

Dr. L. I. Hedberg
Uppsala University
Sysslomansgatan 8
S-75223 Uppsala
Sweden

(Received December 17, 1971)

Math. Z. 129, 321 – 329 (1972)
© by Springer-Verlag 1972

Gleason Parts and a Problem in Prediction Theory

Samuel Merrill, III*

0. Introduction

Two linear manifolds in a Hilbert space are said to be at *positive angle* if

$$\sup |(f, g)| < 1$$

where f and g range over the elements of the manifolds, respectively, of norm 1. Let μ be a finite positive regular Borel measure on

$$T = \{Z \in \mathbb{C}: |Z| = 1\}.$$

Denote by P the linear manifold spanned by the functions $e^{in\theta}$, $n = 0, 1, \ldots$, and P_0 the submanifold of P consisting of polynomials with constant term 0. For any set S of complex-valued functions, write $\bar{S} = \{\bar{f}: f \in S\}$.

In [4] Helson and Szegö obtain the following result.

Proposition. *The manifolds P_0 and $\bar{P}$ are at positive angle in $L^2(d\mu)$ if and only if μ is absolutely continuous with respect to Lebesgue measure dm on T, so that $d\mu = w\,dm$, $w \in L^1(dm)$, and $w = e^u |V|$ where u is real and bounded, $V \in H^1$, and $|\arg V| \leqq (\pi/2) - \varepsilon$ for some $\varepsilon > 0$.*

Here H^1 is the closure of P in $L^1(dm)$. Helson and Szegö also show that V is outer, i.e., the functions $\{Vf: f \in P\}$ are dense in H^1. As they indicate, this result solves a prediction theory problem concerning stationary stochastic processes and gives a necessary and sufficient condition that the projection of $P_0 + \bar{P}$ onto P_0 be bounded in $L^2(d\mu)$.

Devinatz [2] and Ohno [10] extend this result to a setting in which P is replaced by a uniform algebra $A(X)$ and P_0 by $A_m = \{f \in A(X): \int f\,dm = 0\}$ where dm is a (unique) representing measure on X for a complex homomorphism of $A(X)$. Our result is in this setting except that we consider different linear manifolds defined in terms of the Gleason part of m, which we assume non-trivial.

In Section 1 we prove the main theorem and derive from it a necessary and sufficient condition that the projection in $L^2(d\mu)$ onto the closure of the ideal of functions in $H^\infty(dm)$ vanishing on the Gleason part of m be bounded in the norm of $L^2(d\mu)$. In Section 2 we strengthen the theorem

---

* Partially supported by NSF Grant.

for the case in which $A(X)$ is an algebra associated with an ordered group, and in Section 3, we interpret the latter result *to solve a problem of prediction theory type* concerning doubly stationary stochastic processes.

1. The Main Theorem

Let $A=A(X)$ be a uniform algebra, that is a uniformly closed, separating algebra of continuous, complex-valued functions containing the constants on a compact Hausdorff space. Assume that m is a complex homomorphism of A such that m has a unique representing measure dm on X and the Gleason part $P(m)=\{\sigma\in\text{spectrum}(A)\colon \|m-\sigma\|<2\}\neq\{m\}$. Define $\mathbf{H}^p$ as the closure of A in $L^p(dm)$ (norm closure for $1\leqq p<\infty$; w^* closure for $p=\infty$). For $1\leqq p\leqq\infty$, define $\mathbf{I}^p=\{f\in\mathbf{H}^p\colon \int f\,d\sigma=0$ for all $\sigma\in P(m)\}$, $\mathbf{L}^p=\{f\in L^p(dm)\colon \int fh\,dm=0$ for all $h\in\mathbf{I}^\infty\cup\bar{\mathbf{I}}^\infty\}$, and finally $\mathbf{J}^p=\mathbf{L}^p\oplus\mathbf{I}^p$. By [9, Lemma 5, p. 467], $\mathbf{H}^p\subseteq\mathbf{J}^p$. If $f\in L^1(dm)$, we define the *support set* E_f of f as the complement of a set of maximal m-measure on which f is null. Clearly E_f is defined only up to sets of measure zero. Note that if $f\in\mathbf{L}^2$, then the characteristic function of its support set is also in $\mathbf{L}^2$ [9, Lemma 6, p. 468].

Let μ be a finite positive regular Borel measure on X which is absolutely continuous with respect to dm, so that $d\mu=w\,dm$, $w\in L^1(dm)$. Write $E=E_w$ for the support set of w. Denote the norm in $L^2(d\mu)$ by $\|\ \|$ and let $S_I=\{f\in\mathbf{I}^\infty\colon \|f\|=1\}$, $S_J=\{f\in\mathbf{J}^\infty\colon \|f\|=1\}$.

Theorem 1. *The manifolds* $\mathbf{I}^\infty$ *and* $\bar{\mathbf{J}}^\infty$ *are at positive angle in* $L^2(d\mu)$, *i.e.,*

$$\rho=\sup\left|\int fg\,d\mu\right|<1 \qquad (f\in S_I,\, g\in S_J), \tag{1}$$

if and only if

(A) $\chi_E\in\mathbf{L}^\infty$ *and* $w=e^u|V|$ *where* u *is real and bounded,* $V\in\mathbf{J}^1$, *and* $|\arg V|\leqq\pi/2-\varepsilon \pmod{2\pi}$ *for some* $\varepsilon>0$.

Corollary 1. *Let* T *be the projection of* $\mathbf{I}^\infty\oplus\bar{\mathbf{J}}^\infty$ *onto* $\mathbf{I}^\infty$. *Assume that* $\int|Tf|^2\,d\mu\neq0$ *for some* $f\in\mathbf{I}^\infty$. *Then the operator* T *is bounded in the norm of* $L^2(d\mu)$ *if and only if* (A) *holds. Moreover,* $\mathbf{I}^\infty\oplus\bar{\mathbf{J}}^\infty$ *is dense in* $L^2(d\mu)$, *so that if* T *is bounded it has a unique bounded extension to* $L^2(d\mu)$.

Proof of Corollary 1. First note that $\rho<1$ if and only if $\psi>0$, where

$$\psi=2-2\rho=\inf\int|f+\bar{g}|^2\,d\mu \qquad (f\in S_I,\, g\in S_J). \tag{2}$$

If T is bounded with norm $K>0$, then $\|f+\bar{g}\|^2\geqq K^{-1}\|f\|^2=K^{-1}$ for $f\in S_I$ and $g\in S_J$. Conversely, if $\psi>0$, $|\mathrm{Re}(f,\bar{g})|=|\lambda|\leqq\alpha<1$ if $f\in S_I$, $g\in S_J$. To see that T is bounded it suffices to obtain $\delta>0$ such that for each fixed $f\in S_I$ and $g\in S_J$, $\|f+c\,\bar{g}\|^2\geqq\delta$ for all real c. But $p(c)=\|f+c\,\bar{g}\|^2=1+2\lambda\,c+c^2\geqq p(-\lambda)=1-\lambda^2\geqq1-\alpha^2$. Let $\delta=1-\alpha^2$.

To prove density, suppose $g \in L^2(d\mu)$ with $\int f g\, w\, dm = 0$ $(f \in \mathbf{I}^\infty \oplus \bar{\mathbf{J}}^\infty)$. Since $g w \in L^1(dm)$ and $\mathbf{I}^\infty \oplus \bar{\mathbf{J}}^\infty$ is w^* dense in L' (dm) by [9, Lemma 5, p. 467], $g \equiv 0(d\mu)$, so $\mathbf{I}^\infty \oplus \bar{\mathbf{J}}^\infty$ is dense in $L^2(d\mu)$.

Proof of Theorem 1. First assume that $\rho < 1$ so that $\psi > 0$. In particular

$$\inf \int |1-f|^2\, d\mu > 0 \qquad (f \in \mathbf{I}^\infty). \tag{3}$$

Hence the projection G of 1 into the closure in $L^2(d\mu)$ of $\mathbf{I}^\infty$ is not identically 1. If $h \in \mathbf{I}^\infty$, and $f_n \in \mathbf{I}^\infty$ such that $\|f_n - G\| \to 0$, then $\|(1-f_n)h - (1-G)h\| \to 0$ so that $(1-G)$ is orthogonal to $(1-G)h$ in $L^2(d\mu)$. Thus

$$\int |1-G|^2\, h\, d\mu = 0 \qquad (h \in \mathbf{I}^\infty \cup \bar{\mathbf{I}}^\infty).$$

It follows that

$$|1-G|^2\, w = k^2, \tag{4}$$

where $k \in \mathbf{L}^2$, $k \geqq 0$ (see the proof of [9, Lemma 8, p. 469]).

We define $G \equiv 1$ off E, so the support set F of $1-G$ is contained in E. We show $F = E$.

Let $\mathscr{A}$ be the σ-algebra of Borel subsets A of X for which the characteristic function $\chi_A \in \mathbf{L}^\infty$. For $A \in \mathscr{A}$ and $\mu(A) > 0$, define

$$\psi(A) = \inf \int_A |f + \bar{g}|^2\, d\mu \tag{5}$$

where $f \in \mathbf{I}^\infty$, $g \in \mathbf{J}^\infty$, and $\int_A |f|^2\, d\mu = \int_A |g|^2\, d\mu = 1$. If no such f exist, define $\psi(A) = 2$. Since F (the support set of $1-G$) is also the support set of k, $F \in \mathscr{A}$ by [9, Lemma 6, p. 468] and so is $F' = X \setminus F$. By definition of G, there exists $f_n \in \mathbf{I}^\infty$ such that $\int |G - f_n|^2\, d\mu \to 0$. Since $G \equiv 1$ on F',

$$\inf_{f \in \mathbf{I}^\infty} \int_{F'} |1-f|^2\, d\mu = 0. \tag{6}$$

We may assume that the functions f in (6) have supports in F' (by multiplying by $\chi_{F'}$ if necessary). If $\mu(F') > 0$, then $\psi(F') = 0$, and it is then easy to check that $\psi(X) = 0$. This contradiction implies that $\mu(F') = 0$, i.e., $E = F$ and in particular $E \in A$.

Thus we may define

$$D = k/(1-G) \tag{7}$$

on E and $D = 0$ off E. Thus $w = |D|^2$. We show that $[D\,\mathbf{I}^\infty] = \chi_E I^2$, where $[\ ]$ denotes the closure in $L^2(dm)$.

Lemma 1. *If h lies in the closure of $\mathbf{I}^\infty$ in $L^2(d\mu)$, then*

$$\int \frac{(1-\bar{G})h}{k}\, d\mu = 0.$$

Proof. The equality holds for k replaced by $k+\varepsilon$ and continues to hold as $\varepsilon\to 0$, since convergence in $L^2(dm)$ is dominated by $|1-\bar{G}|h|wk^{-1} = |h|w^{\frac{1}{2}}\in L^2(dm)$.

Lemma 2. *The function D defined in (7) is an element of $\mathbf{J}^2$, $k=D_1$ where D_1 is the projection of D onto $\mathbf{L}^2$ in $L^2(dm)$, and*

$$\int |D_1|^2\,dm = \inf_{f\in\mathbf{I}^\infty}\int |1-f|^2\,d\mu > 0. \tag{8}$$

Proof. If $h\in\mathbf{I}^\infty$,

$$\int_E h D\,dm = \int_E h\cdot k/(1-G)\,dm = \int_E \frac{h(1-\bar{G})k^2}{k|1-G|^2}\,dm = \int_E \frac{h(1-\bar{G})}{k}\,d\mu = 0$$

by Lemma 1.

To show that $D_1=k$, it suffices to show that $\int hD\,dm=\int hk\,dm$ for all $h\in\mathbf{L}^\infty$. Using Lemma 1 for the second equality,

$$\begin{aligned}\int hD\,dm &= \int h(1-\bar{G})k^{-1}\,d\mu = \int h(1-\bar{G})k^{-1}\,d\mu - \int h(1-\bar{G})Gk^{-1}\,d\mu\\ &= \int h|1-G|^2k^{-1}\,d\mu = \int hk\,dm.\end{aligned}$$

Finally

$$\int |D_1|^2\,dm = \int k^2\,dm = \int |1-G|^2\,d\mu = \inf_{f\in\mathbf{I}^\infty}\int |1-f|^2\,d\mu > 0$$

by (3).

Lemma 3. *For $A\in\mathscr{A}$, define*

$$\nu(A) = \inf\int_A |1-f|^2\,d\mu \qquad (f\in\mathbf{I}^\infty). \tag{9}$$

Then ν is a countable additive measure on $(X,\mathscr{A})$ and

$$\nu(A) = \int_A |1-G|^2\,d\mu. \tag{10}$$

Proof. Suppose A, $B\in\mathscr{A}$, $A\cap B=\varnothing$. By standard properties of infima, $\nu(A\cup B)\geqq \nu(A)+\nu(B)$. Also there exist f, $g\in\mathbf{I}^\infty$ with support sets contained in A and B, respectively, and

$$\begin{aligned}\nu(A)+\nu(B)+\varepsilon &\geqq \int_A |1-f|^2\,d\mu + \int_B |1-g|^2\,d\mu\\ &= \int_{A\cup B} [|\chi_A-f|^2+|\chi_B-g|^2]\,d\mu = \int_{A\cup B} |1-(f+g)|^2\,d\mu \geqq \nu(A\cup B).\end{aligned}$$

Countable additivity is proved similarly, using the dominated convergence theorem. Details are omitted as we will need only finite additivity. To prove (10) note that if $f_n\in\mathbf{I}^\infty$ and $\|f_n-G\|\to 0$, then $\chi_A f_n\in\mathbf{I}^\infty$ and $\|\chi_A f_n-\chi_A G\|\to 0$. Hence $\nu(A)\leqq \int_A |1-G|^2\,d\mu$. The same inequality holds for $X\setminus A$. Since $\nu(X)=\nu(A)+\nu(X\setminus A)$, equality follows.

Lemma 4. *If* $D \in \mathbf{J}^2$ *is any function such that* $|D|^2 = w$ *and* (8) *holds, then* $[D\mathbf{I}^\infty] = \chi_E \cdot \mathbf{I}^2$ *and* $[D\mathbf{J}^\infty] = \chi_E \cdot \mathbf{J}^2$.

Proof. Assume $\int |D_1|^2\, dm = 1$. By (8) there exist $f_n \in \mathbf{I}^\infty$ such that

$$\int |1 - f_n|^2 \, |D|^2\, dm \to 1 .$$

Thus

$$\int |D_1 - (1 - f_n) D|^2\, dm = \int |D_1|^2\, dm - 2\,\mathrm{Re} \int \bar{D}_1 (1 - f_n)(D_1 + D_2)\, dm + \int |1 - f_n|^2 \, |D|^2\, dm \to 1 - 2 + 1 = 0 .$$

Thus $D_1 \in [D\mathbf{J}^\infty]$. Since $D_1 \in \mathbf{L}^2$, $[D_1 \mathbf{J}^\infty] = [D_1 \mathbf{L}^\infty] + [D_1 \mathbf{I}^\infty]$. Thus to complete the proof it suffices to show that $\chi_E \in [D_1 \mathbf{L}^\infty]$. Since $[D_1 \mathbf{L}^\infty]$ corresponds to a doubly invariant subspace of $L^2(T)$ in the sense of [9, Lemma 4, p. 466], $[D_1 \mathbf{L}^\infty] = \chi_{E_1} \cdot \mathbf{L}^2$ where E_1 is the support set of D_1. Note that $\chi_{E_1} \in \mathbf{L}^\infty$. We must show that $E_1 = E$.

Clearly $E_1 \subseteq E$, for $D_2 = -D_1$ off E, so $(1 - \chi_E) D_2 \in \mathbf{L}^2 \cap \mathbf{I}^2$, so $m(E_1 \setminus E) = 0$. Observe that

$$\int |D_1|^2\, dm = v(X) \geqq v(E_1) = \inf_{f \in \mathbf{I}^\infty} \int |\chi_{E_1} \cdot D - f \chi_{E_1} \cdot D|^2\, dm \geqq \int |D_1|^2\, dm$$

since D_1 is the projection of $\chi_{E_1} \cdot D$ onto $\mathbf{L}^2$. Hence

$$0 = v(E \setminus E_1) = \int_{E \setminus E_1} k^2\, dm$$

by (10) and (4). Since k does not vanish on E, $m(E \setminus E_1) = 0$, so we may assume $E_1 = E$.

Lemma 5. *The set of products* fg *with* $f \in \mathbf{J}^2$, $g \in \mathbf{I}^2$ *and* $\int |f|^2\, dm \leqq 1$ *and* $\int |g|^2\, dm \leqq 1$ *is dense in the set of functions* $h \in \mathbf{I}^1$ *with* $\int |h|\, dm \leqq 1$.

Proof. Suppose $h \in \mathbf{I}^1$, $\int |h|\, dm = 1$ and $0 < \inf \int |1 - \alpha|^2 \, |h|\, dm\, (\alpha \in \mathbf{I}^\infty)$. Let H be the projection of 1 into the closure of $\mathbf{I}^\infty$ in $L^2(|h|\, dm)$. Then by previous argument the support set F of $1 - H$ is the same as that of

$$p^2 = |1 - H|^2 \, |h|\, dm, \tag{11}$$

and $\chi_F \in \mathbf{L}^\infty$. Then $h = fg$ where $f = p/(1 - H)$ and $g = (1 - H) h/p$ on F and zero off F. By direct computation using (11) we find $\int |f|^2\, dm = \int |g|^2\, dm = 1$. We have $f \in \mathbf{J}^2$ by the proof of Lemma 2 and $g \in \mathbf{I}^2$ by the definition of H.

Now suppose

$$0 = \inf \int |1 - \alpha|^2 \, |h|\, dm \qquad (\alpha \in \mathbf{I}^\infty) \tag{12}$$

and let $h_\delta = h + \delta$. For fixed $\alpha \in \mathbf{I}^\infty$, $\lim_\delta \int |1 - \alpha|^2 \, |h_\delta|\, dm = \int |1 - \alpha|^2 \, |h|\, dm$. Given $\varepsilon > 0$, we can choose $0 < \delta < \varepsilon$ by (12) such that

$$\inf \int |1 - \alpha|^2 \, |h_\delta|\, dm < \varepsilon \qquad (\alpha \in \mathbf{I}^\infty). \tag{13}$$

Define $A_m = \{\alpha \in A : \int \alpha\, dm = 0\}$. Since $h_\delta \in \mathbf{H}^1$ the generalized Szegö theorem and Jensen inequality [6] imply that

$$\inf_{\alpha \in \mathbf{I}^\infty} \int |1-\alpha|^2 |h_\delta|\, dm \geqq \inf_{\alpha \in A_m} \int |1-\alpha|^2 |h_\delta|\, dm \geqq \int \delta\, dm > 0.$$

Applying the argument above to h_δ, we have $h_\delta = fg$ where here f and g are in $\mathbf{J}^2$. Writing $f = f_1 + f_2$, $f_1 \in \mathbf{L}^2$, $f_2 \in \mathbf{I}^2$, we conclude from Lemma 2 and (13) that $\int |f_1|^2\, dm < \varepsilon$. Thus

$$\|f_2 g - h_\delta\|_1^2 = \|f_1 g\|_1^2 \leqq \|f_1\|_2^2 \|g\|_2^2 < \varepsilon(1+\delta) < \varepsilon(1+\varepsilon).$$

The conclusion of the lemma is now straightforward.

To complete the proof of Theorem 1, write $w = e^{-i\varphi} D^2$ on E. Then

$$\rho = \sup \left|\int (fD)(gD)\, e^{-i\varphi}\, dm\right| \qquad (f \in S_I, g \in S_J).$$

By Lemmas 4 and 5

$$\rho = \sup \left|\int_E h\, e^{-i\varphi}\, dm\right| \qquad \left(h \in \mathbf{I}^1;\ \int_E |h|\, dm = 1\right).$$

Thus ρ is the norm of the linear functional on $\mathbf{I}^1$ defined by

$$\psi(h) = \int_E h\, e^{-i\varphi}\, dm.$$

Arguing as in [4, p. 125], $\rho = \inf \|\chi_E e^{-i\varphi} - A\|_\infty$ $(A \in \mathbf{J}^\infty)$, and $\rho < 1$ if and only if there exist $\varepsilon > 0$ and $A \in \mathbf{J}^\infty$ such that $|A(x)| \geqq \varepsilon$ and $|\varphi(x) + \arg A(x)| \leqq \pi/2 - \varepsilon \pmod{2\pi}$ for $x \in E$. Thus $w = |D|^2 = |AD^2|(1/|A|) = |V|\, e^u$ where u is real and bounded and $|\arg V| \leqq \pi/2 - \varepsilon \pmod{2\pi}$.

Conversely if $w = |V|\, e^u$ where u is real and bounded, $V \in \mathbf{J}^1$, and $|\arg V| \leqq \pi/2 - \varepsilon \pmod{2\pi}$, then $\rho < 1$. For we may assume $w = |V|$ $(u \equiv 0)$. Let $A \equiv 1$, $w = V e^{-i\varphi}$, so $\arg V = \varphi$, and $|\varphi + \arg A| = |\arg V| \leqq \pi/2 - \varepsilon \pmod{2\pi}$.

2. Algebras Associated with Ordered Groups

Let G be a compact abelian group with ordered dual Γ determined by a positive semigroup Γ_+ (see [11, Chapter 8]), $A(G) = \{f \in C(G): \hat{f}(\gamma) = 0$, for $\gamma < 0\}$, and dm be normalized Haar measure. Clearly $A(G)$ is a uniform algebra and dm is the unique representing measure for a complex homomorphism m. In this section we show that if m lies in a non-trivial Gleason part, then the circle group is a direct factor of G, and the function V of Theorem 1 has a property analogous to the generative property of outer functions.

Lemma 6. *If the Gleason part of m is non-trivial, there exists a least positive character γ_0 which considered as a function on G is the Wermer embedding function $\mathbf{Z}$, except for a constant factor of modules* 1 (see [9, p. 464]). *If $H = \{x \in G: \gamma_0(x) = 1\}$, then $G = H \times (G/H)$ and $G/H \cong T$.*

Proof. Choose $\sigma\in P(m)\setminus\{m\}$. By [1, Theorem 3.1 and Theorem 4.1, p. 381–382], there exist $x\in G$ and $\rho\colon \Gamma_+\to[0,1]$ with $\rho(\gamma+\delta)=\rho(\gamma)\,\rho(\delta)$ $(\gamma,\delta\in\Gamma_+)$ such that $\sigma(\gamma)=\rho(\gamma)\,(x,\gamma)$ $(\gamma\geqq 0)$. Since $\sigma\in P(m)$ and $m(\gamma)=0$ for $\gamma>0$,

$$\sup_{\gamma>0}\rho(\gamma)=\sup_{\gamma>0}|\sigma(\gamma)|=c<1. \tag{14}$$

If there is no least positive character γ_0 in Γ, then there exist

$$\gamma_1>\gamma_2>\cdots>0$$

with $\rho(\gamma_n)/c>c$ for large n. But then

$$\rho(\gamma_{n+1})=\rho(\gamma_n)/\rho(\gamma_n-\gamma_{n+1})\geqq\rho(\gamma_n)/c>c,$$

which contradicts (14). Since the function γ_0 has constant modulus one, we can conclude that $\gamma_0=\mathbf{Z}$ by showing that $\bar\lambda\gamma_0$ is the projection of 1 into $H^2_m(d\sigma)$ for an appropriate constant λ. Since $\gamma_0\in H^2_m(d\sigma)$, it suffices to show $1-\bar\lambda\gamma_0\perp H^2_m(d\sigma)$. For $\gamma>0$, $\int(1-\lambda\bar\gamma_0)\gamma\,d\sigma=\rho(\gamma)(x,\gamma)\{1-\lambda\bar\gamma_0(x)/\rho(\gamma_0)\}=0$ if we choose $\lambda=\rho(\gamma_0)\,\gamma_0(x)$.

Clearly H is a subgroup and $G/H\cong T$. If $\Lambda=\{\gamma\in\Gamma\colon\gamma(x)=1 \text{ for } x\in H\}$ is the annihilator of H, $\Lambda\cong$ the integers, so that $\Lambda\cong\Gamma_0=\{n\gamma_0\colon n \text{ an integer}\}$. This coupled with the fact that if $\gamma\in\Gamma_+\setminus\Gamma_0$ then $\gamma>n\gamma_0$ for all integers n implies that the order on Γ induces an order on Γ/Λ. Thus Γ/Λ is torsion-free [11, Theorem 8.1.2.(a), p. 194] so that its dual H is divisible [5, 24.25, p. 385]. In turn $G\cong H\times(G/H)$ [5, A. 8, p. 441].

We thank our colleague T. Wilcox for discussions about topological groups.

The following result (for the special case $H=T$) appears in the thesis of Lal [7]. We give his proof.

Lemma 7. *Suppose $G=TH$, $T\cap H=\{e\}$, where T and H are subgroups. Then*

$$\inf_{f\in\mathbf{I}^\infty}\int|1-f|^2\,w\,dm=\int_T\Big[\exp\int_H\log w(t,h)\,dh\Big]\,dt$$

where dt and dh denote the normalized Haar measures on T and H.

Proof. If G is defined as in the proof of Theorem 1, then there exist $g_n\in\mathbf{I}^\infty$ with $\int|g_n-G|^2\,w\,dm\to 0$. Replacing g_n by a subsequence if necessary, this implies that

$$\int_H|g_n(t\,h)-G(t\,h)|^2\,w(t\,h)\,dh\to 0$$

for almost all t. Define $G_t(h)=G(t\,h)$. It follows by the generalized Szegö theorem [6, p. 289] that

$$\int_H|1-G_t(h)|^2\,w_t(h)\,dh=\exp\int_H\log w_t(h)\,dh.$$

Integrating with respect to t, we obtain the result.

Theorem 2. *If $A=A(G)$, dm is normalized Haar measure, and m lies in a non-trivial Gleason part, then the function V of Theorem 1 has the property $[V\cdot \mathbf{I}^\infty]_1=\chi_E\cdot \mathbf{I}^1$ where $[\;]_1$ denotes closure in $L^1(dm)$.*

Proof. By Lemma 6, we may assume $G=H\times T$. By Theorem 1, $\operatorname{Re} V\geqq 0$, so $\operatorname{Re} V_t\geqq 0$. It is well known that this implies that V_t is outer for almost all t with $\{t\}\times H\subseteq E$ (see, e.g., [8, II.2.2, p. 53]), so that

$$\exp\int_H \log|V_t(h)|\,dh=\left|\int_H V_t(h)\,dh\right|=|V_1(t)|>0$$

for such t (V_1 denotes the projection of V into $\mathbf{L}^1$). The second equality follows from the facts that $Z=\gamma_0$ (Lemma 6) and L^1 is the closure of the characters in Γ_0. Integrating with respect to t we obtain

$$\int_T\left[\exp\int_H \log|V|\,dh\right]dt=\int|V_1|\,dt>0. \tag{15}$$

But $V=(AD)D$ is the product of functions in $\mathbf{J}^2$, each of which must also satisfy (15). Thus using Lemmas 4 and 7 and an argument similar to that of [6, Theorem 6.3, p. 299] we conclude that

$$[V\cdot \mathbf{I}^\infty]_1=\chi_E\cdot \mathbf{I}^1.$$

In fact $V=e^{\alpha+i\beta}$ where the restriction of $\alpha+i\beta$ to $\{t\}\times H$ is in $L^1(H)$ and α_t is the conjugate of $-\beta_t$ in the sense of [3, II, Theorem 6, p. 192] for almost all t with $\{t\}\times H\subseteq E$ (cf. [4, Theorem 1, p. 123]).

3. Application to Doubly Stationary Stochastic Processes

Let (Ω, Σ, P) be a probability measure space. A stochastic process $\chi_{nm}(n, m\in\mathbb{Z})$ of complex-valued random variables in $L^2(dP)$ is called *doubly stationary* [3, I, p. 168] if the covariance

$$(\chi_{nm}, \chi_{n'm'})=(\chi_{n-n',m-m'}, \chi_{00})=\rho(n-n', m-m')$$

is a function of $(n-n')$ and $(m-m')$. In this case ρ is a positive definite function on $\mathbb{Z}\times\mathbb{Z}$, so by a theorem of Bochner [11, p. 19], there exist a finite positive Borel measure μ on $G=T\times T$ (the spectral measure of the process) such that $\hat{\mu}=\rho$. The correspondence of the functions χ_{nm} to the characters $e^{-in\theta}e^{-im\varphi}$ on G induces an isometry of the closed linear span of χ_{nm} onto $L^2(d\mu)$.

Let $A(G)$ be the uniform algebra determined by the semigroup $\{(n, m)\colon n>0\}\cup\{(0, m)\colon m\geqq 0\}$ (see Section 2). The maximal ideal space of $A(G)$ can be identified as

$$\{(\theta, \varphi)\colon |\theta|\leqq 1, |\varphi|=1\}\cup\{(0, \varphi)\colon |\varphi|<1\}.$$

Haar measure on G represents the center of the non-trivial Gleason part $P=\{(0, \varphi)\colon |\varphi|<1\}$. $\mathbf{I}^\infty$ and $\mathbf{J}^\infty$ are defined in terms of this part.

We assume that the spectral measure μ is absolutely continuous with respect to Haar measure and define the *past* of the stochastic process by $I = \text{span}\{\chi_{nm}: n<0, m\in\mathbb{Z}\}$ and the *future* by $J = \text{span}\{\chi_{nm}: n\geqq 0, m\in\mathbb{Z}\}$. Then

$$\sup|(x, y)| = \sup|\int fg\, d\mu| \tag{16}$$

where $x\in I$, $y\in J$, $\|x\| = \|y\| = 1$; and $f\in \mathbf{I}^{\alpha}$, $g\in \mathbf{J}^{\alpha}$, $\int|f|^2\, d\mu = \int|g|^2\, d\mu = 1$. Theorem 2 applies and gives a necessary and sufficient condition that the suprema in (16) be strictly less than 1. This is a result of prediction theory type, in the extended sense of [3] and [4].

References

1. Arens, R., Singer, I.: Generalized analytic functions. Trans. Amer. Math. Soc. **81**, 379–393 (1956).
2. Devinatz, A.: Toeplitz operators on H^2 spaces. Trans. Amer. Math. Soc. **112**, 304–317 (1964).
3. Helson, H., Lowdenslager, D.: Prediction theory and Fourier series in several variables, I and II. Acta Math. **99**, 165–202 (1958) and **106**, 173–213 (1961).
4. Helson, H., Szegö, G.: A problem in prediction theory. Ann. Mat. Pura Appl. **51**, 107–138 (1960).
9. Hewitt, E., Ross, K.: Abstract harmonic analysis. Berlin-Göttingen-Heidelberg: Springer 1963.
6. Hoffman, K.: Analytic functions and logmodular Banach algebras. Acta Math. **108**, 271–317 (1962).
7. Lal, N.: Isometries of H^p spaces derived from logmodular algebras. Thesis, University of Rochester 1969.
8. Merrill, S.: H^p spaces derived from function algebras. Thesis, Yale University 1965.
9. Merrill, S., Lal, N.: Characterization of certain invariant subspaces of H^p and L^p spaces derived from logmodular algebras. Pacific J. Math. **30**, 463–474 (1969).
10. Ohno, Y.: Remarks on Helson-Szegö problems. Tôhoku Math. J. **18**, 54–59 (1966).
11. Rudin, W.: Fourier analysis on groups. New York: Interscience 1962.

Prof. S. Merrill, III
Department of Mathematics
The University of Rochester
Rochester
New York 14627 (USA)

(Received January 31, 1972)

Math. Z. 129, 331–334 (1972)
© by Springer-Verlag 1972

Flat Modules and Torsion Theories

William Schelter and Paul Roberts

In Morita's paper [4] he investigates the torsion theory and the quotient ring arising from a flat module. If ${}_R U$ is a flat module then one can consider as torsion all modules M_R such that $M \otimes_R U = 0$, or one may look at the idempotent Gabriel filter $\{D \leqq_r R: DU = U\}$, or at the torsion theory coming from the injective right R-module $\mathrm{Hom}_{\mathbb{Z}}[{}_R U, Q/\mathbb{Z}]$. All three coincide. But which torsion theories come from flat modules, since we know, for instance, that every torsion theory comes from an injective? We give an example of a torsion theory on a commutative ring, which does not come from a flat module. We also show that if the Gabriel filter has a countable base then it must arise from a flat module.

Theorem. *If $\mathscr{D}$ is a right Gabriel filter for the ring R, and if $\mathscr{D}$ has a countable base, then there is a flat left R-module ${}_R U$ such that $\mathscr{D} = \{D \leqq_r R: DU = U\}$.*

Proof. Let $(D_n)_{n \in \mathbb{N}}$ be a base for $\mathscr{D}$, such that $D_{n+1} \subseteq D_n$. Let

$$I = \{(d_i)_{i \in \mathbb{N}} \in \prod_{i \in \mathbb{N}} D_i : \forall_{n \in \mathbb{N}} \forall_{j \leqq n}\, d_j d_{j+1} \ldots d_n \in D_n\}$$

and let I_n be the restriction to $\prod_{i \geqq n} D_i$, for each $n \in \mathbb{N}$.

If $d = (d_i)_{i \in \mathbb{N}} \in I$, we denote $(d_i)_{i \geqq n}$ by d^n. Let F_n be a free R-module with $|I_n|$ basis elements, denoted $\varepsilon(d^n)$. We define

$$\alpha_n : F_n \to F_{n+1} = \varepsilon(d^n) \mapsto d_n\, \varepsilon(d^{n+1})$$

to be a left R-homomorphism. Let ${}_R U = \varinjlim F_n$, which is a flat left R-module.

We show that, if M is $\mathscr{D}$-torsion, then $M \otimes_R U = 0$. Take $m \in M$ and $\varepsilon(d^n) \in F_n$. Since $M \otimes U = \varinjlim (M \otimes F_n)$, we must find a $j > n$ such that $m \otimes \alpha_j \alpha_{j-1} \ldots \alpha_n\, \varepsilon(d^n) = 0$. We select $j > n$ such that $m D_j = 0$, and observe that

$$\alpha_j \alpha_{j-1} \ldots \alpha_n\, \varepsilon(d^n) = d_n d_{n+1} \ldots d_j\, \varepsilon(d^j)$$

and

$$d_n d_{n+1} \ldots d_j \in D_j.$$

Next we show that if $M \neq 0$ is $\mathcal{D}$-torsionfree then $M \otimes_R U \neq 0$. To do this we build a sequence $d=(d_i)_{i\in\mathbb{N}} \in I$ such that

$$m \otimes \alpha_n \dots \alpha_1 \varepsilon(d) \neq 0, \qquad \text{all } n \in \mathbb{N},$$

for some fixed $m \in M$. Take any $m \in M$. We know $mD_1 \neq 0$, so take $d_1 \in D_1$ such that $m d_1 \neq 0$. Now $m d_1 \in M$, which is $\mathcal{D}$-torsionfree, so take

$$d_2 \in D_2 \cap d_1^{-1}(D_2)$$

such that $m d_1 d_2 \neq 0$. In general take

$$d_n \in D_n \cap (d_1 \dots d_{n-1})^{-1} D_n \cap \dots \cap d_{n-1}^{-1}(D_n)$$

such that $m d_1 \dots d_n \neq 0$. Clearly $d=(d_n)_{n\in\mathbb{N}}$ does what is required.

It remains to show that if $M \otimes U = 0$ then M is $\mathcal{D}$-torsion. If not, $M/T(M) \neq 0$ is $\mathcal{D}$-torsionfree but

$$0 = M \otimes U \to M/T(M) \otimes U \to 0$$

is exact, a contradiction. //

We now give an example of a commutative ring, such that the Utumi filter, i.e. the largest Gabriel filter such that R_R is torsion free, does not come from a flat module. Let $R=C(X)$, where X is the real unit interval [0, 1], and $C(X)$ is the ring of real valued continuous functions on it. An ideal D of R is in the Utumi filter $\mathcal{I}$ iff the cozero set of D (i.e. $\{x \in X\colon f \in D,\ f(x) \neq 0\}$) is an open dense subset of X. Let $D_x = \{f \in R\colon f^{-1}(0)$ is a neighborhood of $x\}$ for each x in X. Suppose $\mathcal{I}$ comes from a flat module U. By Lazard's result (see [3]) we can write $U = \varinjlim_{i \in J} F_i$, F_i finitely generated free, where J is directed and we have $p_{ij}\colon F_i \to F_j$, where $i<j$. If $U \neq 0$ then there is a $z \in F_i$, for some i, such that $p_{ij}(z) \neq 0$, for all $i<j$.

We know $0 = R/D_x \otimes U = \varinjlim (R/D_x \otimes F_i)$. Thus there is a j_x such that $(1+D_x) \otimes p_{ij_x}(z) = 0$. This means that each component of $p_{ij_x}(z)$ lies in D_x. Let Z_x be the intersection of the zero sets of each component, thus Z_x is a neighborhood of x. Let $\bigcup_{k=1}^{n} Z_{x_k}$ be a finite subcovering of X. Choose $s > j_{x_k}$, for all $k \leqq n$. It is now clear that each component of $p_{is}(z)$ must be zero, since it must be zero on every Z_{x_k}. Thus $U=0$, and so could not have given us the Utumi filter.

Proposition. *If R is a commutative noetherian ring every Gabriel filter arises from a flat module.*

Proof. Let E be an injective which gives rise to the Gabriel filter in question. $E = \bigoplus_{i \in I} E(R/P_i)$, where the P_i are prime ideals. The torsion theory determined by $E(R/P_i)$ is the same as the one coming from the flat R-module R_{P_i} (i.e. R localized at P_i).

$$\operatorname{Hom}_{\mathbb{Z}}[\oplus R_{P_i}, Q/\mathbb{Z}] \simeq \prod \operatorname{Hom}_{\mathbb{Z}}[R_{P_i}, Q/\mathbb{Z}],$$

and the latter injective gives the same torsion theory as $\prod E(R/P_i)$, which gives the same torsion theory as $\oplus E(R/P_i) = E$. //

If U_1 and U_2 are two flat R-modules, then it is clear that the intersection of the corresponding torsion theories is given by $U_1 \oplus U_2$.

We shall call U_1 equivalent to U_2 if U_1 and U_2 give the same torsion theory. Denote this by $U_1 \sim U_2$.

Corollary. *If R is commutative and noetherian, and U_1 and U_2 are flat R-modules, then $U_1 \otimes U_2$ gives rise to the smallest torsion theory containing those of U_1 and U_2.*

Proof. First observe that if $U_1 \sim U_1'$ and $U_2 \sim U_2'$ then $U_1 \otimes_R U_2 \sim U_1' \otimes_R U_2'$, since if $M \in \operatorname{Mod} R$

$$\begin{aligned} M \otimes (U_1 \otimes U_2) = 0 &\Leftrightarrow (M \otimes U_1) \otimes U_2' = 0 \\ &\Leftrightarrow (M \otimes U_2') \otimes U_1' = 0. \end{aligned}$$

Thus, by the proof of the Proposition, it suffices to check the corollary where U_1 and U_2 are sums of localizations. If $U_1 = U_2 \sim \bigoplus_{i \in I} R_{P_i}$, then we must show $M \otimes U_1 \otimes U_1 = 0 \Rightarrow M \otimes U_1 = 0$. But

$$\begin{aligned} 0 = M \otimes (\bigoplus_{i \in I} R_{P_i}) \otimes (\bigoplus_{i \in I} R_{P_i}) &= M \otimes \bigoplus_{i, j \in J} (R_{P_i} \otimes R_{P_j}) \\ &= (M \otimes (\oplus R_{P_i})) \oplus (M \otimes (\bigoplus_{i \neq j} R_{P_i} \otimes R_{P_j})) \\ &\Rightarrow M \otimes (\oplus (R_{P_i})) = 0. \end{aligned}$$

If $U_1 \neq U_2$, suppose U is a flat module representing the join of the torsion theories of U_1 and U_2. Suppose further there is an $M \in \operatorname{Mod} R$ such that

$$M \otimes U \neq 0 \quad \text{but } M \otimes U_1 \otimes U_2 = 0.$$

Thus $M \otimes U_1$ is U_2-torsion and hence $M \otimes U_1 \otimes U = 0$. Thus $M \otimes U$ is U_1-torsion so that $M \otimes U \otimes U = 0$. This says, by the first case, that $M \otimes U = 0$. //

The reader should note that for non-noetherian rings the corollary does not hold. Indeed, if we let $R = \mathbb{Q}[x_i : i \in \mathbb{N}]/(x_i^2 : i \in \mathbb{N})$, and let

$$U = \varinjlim (R \xrightarrow{x_1} R \xrightarrow{x_2} R \rightarrow \cdots)$$

where the maps are multiplications by x_i, then $U \otimes_R U = 0$, but U is not equivalent to zero, since only zero is equivalent to zero.

Bibliography

1. Fine, N., Gillman, L., Lambek, J.: Rings of quotients of rings of functions, p. 11. McGill University Press 1961.
2. Lambek, J.: Torsion theories, additive semantics and rings of quotients. Lecture Notes in Mathematics 177. Berlin-Heidelberg-New York: Springer 1971.
3. Lazard, D.: Sur les modules plats. C.R. Acad. Sci. Paris **258**, 6313–6316 (1964).
4. Morita, K.: Flat modules, injective modules and quotient rings. Math. Z. **120**, 25–40 (1971).

Dr. W. Schelter and Dr. P. Roberts
Department of Mathematics
McGill University
Montreal, Canada

(Received May 27, 1972)

Math. Z. 129, 335–349 (1972)
© by Springer-Verlag 1972

Distinguished Selfadjoint Extensions of Dirac Operators

Upke-Walther Schmincke

§ 1. Introduction

The Hamiltonians in quantum mechanics are given mostly as formal differential expressions. To realize them, i.e. to get them as operators, one has to choose suitable domains of definition in a Hilbert function space. The point which decides on the physical relevancy of such a formal Hamiltonian is the question whether it generates a selfadjoint operator in a certain unique manner, or whether there exists a domain on which it defines a selfadjoint operator distinguished among other selfadjoint operators with the same formal part (if there exist any at all).

In nonrelativistic quantum mechanics there is no difficulty to make the notion of the "distinguished selfadjoint realization" precise. For it is a fact that, for instance, all physically important Schrödinger operators, defined as minimal operators on suitable sets of good-natured functions, are essentially selfadjoint or semibounded. Then unless the closure of such a minimal Schrödinger operator is already selfadjoint one can take its Friedrichs extension as the distinguished selfadjoint realization. It is this realization which in the respective case is employed in physics; usually it is obtained by restricting the adjoint of the minimal operator in terms of a "distinguished boundary condition" (Friedrichs [2], Rellich [9]; for a discussion, cf. Kalf [3]).

In relativistic quantum mechanics the situation is more complicated. Dirac operators, for instance, as all relativistic Hamiltonians for particles with a half-integral spin, will never be semibounded. So the problem arises to characterize those minimal operators which are not essentially selfadjoint but nevertheless have distinguished selfadjoint extensions in a familiar sense. As the most important example we consider the minimal Dirac operator T of one-electron ions which has the formal part

$$\mathscr{T}=\vec{\alpha}\cdot\vec{p}+\beta-\frac{Z\alpha}{|x|}$$

($\alpha=1/137{,}038\ldots$ fine structure constant, Z atomic number; for its domain and the other notations see § 3). Applying the Weyl-Stone theory (for systems) to the separated radial part of $\mathscr{T}$ one obtains (cf. Weidmann [15])

a) limit-point case at ∞ for $0 \leqq Z\alpha < \infty$,

b) limit-point case at 0 (hence essential selfadjointness of T) for $0 \leqq Z\alpha \leqq \frac{1}{2}\sqrt{3}$ (which corresponds to $Z \leqq 118$),

c) limit-circle case at 0 (hence non essential selfadjointness of T) for $Z\alpha > \frac{1}{2}\sqrt{3}$ ($Z \geqq 119$).

However, in the *transition interval* $\frac{1}{2}\sqrt{3} < Z\alpha < 1$ ($119 \leqq Z \leqq 137$) we have only a "weak" limit-circle case at 0 because of the nonoscillatory fundamental system of the radial part of $\mathcal{T}$. Hence a suitable boundary condition eliminating the faster divergent solution defines a distinguished selfadjoint realization of $\mathcal{T}$ (which is actually used in physics; there it is only the range $Z\alpha > 1$ ($Z > 137$) that is in general excluded from consideration, cf. Landau-Lifschitz [7]).

It is the aim of this paper to give a criterion for formal Dirac operators to possess such a distinguished selfadjoint realization. The criterion attaches to the possibility to construct this realization as an extension of the minimal operator. The method applied can be considered as a generalization of the closing procedure: we introduce a suitable operator G (the "intercalary multiplicator") and receive the desired selfadjoint extension of the given operator A as the closure of the aggregate $\overline{G^{-1}}\,\overline{GA}$. The abstract construction is carried out in §2 (Theorem 1). Having established some relations for concrete vector operators (§3) we apply the abstract result to the Dirac operator (§4). We obtain a large class of potentials for which this extension exists (Theorem 2). Finally we give the characterization of this extension by boundary conditions (Theorem 3). So we are able to show that in the special case of the Dirac operator for one-electron ions with $Z\alpha < 1$ ($Z \leqq 137$) the constructed selfadjoint extension coincides with the usually employed one. In this case the boundary conditions (as in the analogous case for the Schrödinger operator $-\Delta + v/|x|^2$ in $\mathbb{R}^3 \setminus \{o\}$ where $-\frac{1}{4} < v < \frac{3}{4}$ is the transition interval) can be interpreted physically as the requirement for the potential energy to be finite.

§2. The Intercalary Multiplicator Extension

The following is based upon a Hilbert space H. All operators are assumed to be linear. The restriction of an operator A to a subspace $D \subseteq D(A)$ is denoted by $A|D$. Bounded operators are necessarily densely defined for us; the notion of relative boundedness is the same as in Kato [5], though.

Lemma 1. *Let A, B be densely defined operators with $D(B) \supseteq D(A)$ and*

$$\|Bu\| \leqq a\|u\| + b\|Au\|, \qquad u \in D(A),$$

where $a \geqq 0$, $b \in [0, 1)$ *are suitable constants. Moreover, let* A *be closable. Then the following statements hold:*

(i) *The operator* $A+B$ *is closable, and* $D(\overline{A+B}) = D(\bar{A})$.

(ii) *If* B *is closable, then* $\overline{A+B} = \bar{A} + \bar{B}$.

(iii) *If* A *has a bounded inverse* A^{-1} *with* $a\|A^{-1}\| + b < 1$, *then the operators* $A+B$, $\overline{A+B}$ *have bounded inverses, too; in particular*

$$\overline{R(A+B)} = R(\overline{A+B}) = R(\bar{A}) = H.$$

Proof. See Kato [5], p. 190 and 196. ∎

The proofs of the following two lemmata are not difficult and may be omitted.

Lemma 2. *Let* A, B *be closable operators; furthermore, let the product* AB *be closable.*

(i) *If* B *is relatively bounded with respect to* AB, *then* $\overline{AB} \subseteq \bar{A}\bar{B}$.

(ii) *If* B *is invertible and the operator* B^{-1} *closable and relatively bounded with respect to* A, *then* $\bar{A}\bar{B} \subseteq \overline{AB}$.

Lemma 3. *Let* K, L *be operators with* $R(K) = H$, $R(L) \supseteq D(K)$, *where* K *is densely defined. Then* $R(LKL)$ *is dense in* H.

Theorem 1. *Let* A_0, B, C, G *be operators, where* A_0 *is essentially selfadjoint and* B *and* G *symmetric. Assume that the following hypotheses are fulfilled:*

(i) $D_0 := D(A_0) = D(B) = D(C) \subseteq D(G) =: D_1$;

(ii) $R(A_0) \cup R(B) \cup R(C) \subseteq D_1$;

(iii) $R(G) = D_1$, $R(G|D_0) = D_0$;

(iv) *there are numbers* $c_0 > 0$, $c_1 > 0$, $c_2 > 1$, $c_3 > 1$, $c_4 > 1$ *such that all* $u \in D_0$ *satisfy the following inequalities:*

$$\|A_0 u\| \geqq c_0 \|u\|, \qquad \|A_0 G^2 u\| \geqq c_1 \|u\|, \tag{1), (2}$$

$$\|A_0 G u\| \geqq c_2 \|(A_0 G - G A_0) u\|, \tag{3}$$

$$\|G A_0 u\| \geqq c_3 \|G C u\|, \tag{4}$$

$$\|G(A_0 + C) u\| \geqq c_4 \|G(B + C) u\|. \tag{5}$$

Then G *is invertible, the operator* GA *with* $A := A_0 - B$ *is closable, and the operator*

$$A_G := \overline{G^{-1}}\, \overline{GA}$$

is an essentially selfadjoint extension of A.

Remark. As our example in §4 will show, the assumptions of the theorem do not imply the essential selfadjointness of A.

Proof. We shall prove the theorem in three steps.

Step 1. First we deduce the following statements:

(I) (i) The operator $A_0 G^2$ is closable and has a bounded inverse;

(ii) the operator $\hat{A}_0 := G A_0 G$ is closable, and $D(\hat{A}_0) = D_0$, $R(\overline{\hat{A}_0}) = H$;

(iii) there are positive numbers c_5, c_6, c_7 such that for all $u \in D_0$ the inequalities

$$\|\hat{A}_0 u\| \geqq c_5 \|u\|, \quad \|\hat{A}_0 u\| \geqq c_6 \|G^2 u\|, \quad \|\hat{A}_0 u\| \geqq c_7 \|A_0 G^2 u\| \qquad (6), (7), (8)$$

hold.

Ad (I.i). Obviously $D(A_0 G^2) = D_0$. The closability of $A_0 G^2$ is then a simple consequence of the symmetry of the operators A_0 and G^2. In view of (2) it remains to be shown that $R(A_0 G^2)$ is dense in H. Now $\overline{A_0}$ is selfadjoint; according to (1) the point 0 belongs to the resolvent set of $\overline{A_0}$ so that $\overline{A_0}$ has a bounded inverse with domain H. Because of $R(A_0 G^2) = R(A_0)$ this gives

$$\overline{R(A_0 G^2)} = \overline{R(A_0)} \supseteq R(\overline{A_0}) = H.$$

Ad (I.ii). The symmetry and so the closability of $\hat{A}_0$ are evident. It is also clear that the operators $\hat{A}_0$ and $P := \hat{A}_0 - A_0 G^2$ have domain D_0. From (3) the inequality

$$\|P u\| \leqq \frac{1}{c_2} \|A_0 G^2 u\|, \qquad u \in D_0, \tag{9}$$

follows, hence regarding (I.i) and applying Lemma 1 to the operators $A_0 G^2$, P we get our assertion.

Ad (I.iii). These inequalities are trivial consequences of (9), (1), and (2).

Step 2. Now we prove:

(II) The operator $\hat{A} := GAG = G(A_0 - B) G$ is symmetric and has the properties:

(i) $D(\overline{\hat{A}}) = D(\overline{\hat{A}_0})$, $R(\overline{\hat{A}}) = H$;

(ii) there is a positive number c_8 such that all $u \in D_0$ satisfy the inequality

$$\|\hat{A} u\| \geqq c_8 \|\hat{A}_0 u\|. \tag{10}$$

The symmetry of $\hat{A}$ is trivial. We introduce the operators $\hat{B} := GBG$ and $\hat{C} := GCG$. Obviously $D(\hat{B}) = D(\hat{C}) = D_0$ and for all $u \in D_0$

$$\|\hat{C} u\| \leqq \frac{1}{c_3} \|\hat{A}_0 u\|, \tag{11}$$

$$\|(\hat{B} + \hat{C}) u\| \leqq \frac{1}{c_4} \|(\hat{A}_0 + \hat{C}) u\|, \tag{12}$$

according to (4) and (5). Taking account of (I.ii) and (11) and applying Lemma 1 to the operators $\hat{A}_0$, $\hat{C}$ we obtain the closability of $\hat{A}_0+\hat{C}$ and the relations

$$D(\overline{\hat{A}_0+\hat{C}})=D(\overline{\hat{A}_0}), \quad R(\overline{\hat{A}_0+\hat{C}})=R(\overline{\hat{A}_0})=H.$$

Application of Lemma 1 to the operators $\hat{A}_0+\hat{C}$, $\hat{B}+\hat{C}$ yields

$$D(\bar{\hat{A}})=D(\overline{\hat{A}_0+\hat{C}-(\hat{B}+\hat{C})})=D(\overline{\hat{A}_0+\hat{C}})=D(\overline{\hat{A}_0}), \; R(\bar{\hat{A}})=R(\overline{\hat{A}_0+\hat{C}})=H.$$

The inequality (10) is trivial.

Step 3. Finally we show:

(III) (i) The operator GA is closable, and

$$A_G := \overline{G^{-1}}\;\overline{GA}=\overline{G^{-1}}\;\bar{\hat{A}}\;\overline{G^{-1}}. \tag{13}$$

(ii) Let κ be a real number with $0<|\kappa|<c_6\,c_8$ and put $K := \overline{\hat{A}+i\kappa G^2}$, $L := \overline{G^{-1}}$. Then

$$K=\bar{\hat{A}}+i\kappa\overline{G^2}, \tag{14}$$

$$D(\bar{\hat{A}})=D(K)\subseteq R(L), \quad R(K)=H, \tag{15}$$

$$LKL=A_G+i\kappa E. \tag{16}$$

(iii) A_G is essentially selfadjoint.

Ad (III.i). The closability of GA can easily be inferred from the symmetry of the operators G and A. The existence of G^{-1} is trivial, too. From (7) and (10) we get

$$\|\hat{A}u\|\geqq c_6\,c_8\,\|G^2 u\|, \quad u\in D_0, \tag{17}$$

and as a consequence

$$\|Gu\|^2=(G^2u,u)\leqq\frac{1}{c_6\,c_8}\,\|\hat{A}u\|\cdot\|u\|, \quad u\in D_0. \tag{18}$$

Therefore $(G^{-1})^{-1}$ is relatively bounded with respect to $\hat{A}$. Noting the equality $\hat{A}G^{-1}=GA$ and applying Lemma 2, (ii) to the operators $\hat{A}$, G^{-1} we obtain $\bar{\hat{A}}\,\overline{G^{-1}}\subseteq\overline{GA}$. The opposite inclusion follows by Lemma 2, (i) from the relative boundedness of G^{-1} with respect to $GA=\hat{A}G^{-1}$ which evidently results from (10) and (6):

$$\|\hat{A}G^{-1}u\|\geqq c_8\,\|\hat{A}_0\,G^{-1}u\|\geqq c_5\,c_8\,\|G^{-1}u\|, \quad u\in D_0.$$

Thus the equality $\bar{\hat{A}}\,\overline{G^{-1}}=\overline{GA}$ is deduced; left-sided multiplication with $\overline{G^{-1}}$ gives (13).

Ad (III.ii). According to (17) the operator $i\kappa G^2$ is relatively bounded with respect to $\hat{A}$ with an $\hat{A}$-bound smaller than 1; hence by Lemma 1 the relations (14) and $D(K)=D(\bar{\hat{A}})$ are obvious. Also we have $R(K)=H$ in virtue of (II.i). In addition we get regarding (18)

$$D(K)=D(\bar{\hat{A}})\subseteq D(\bar{G})=R(\overline{G^{-1}})=R(L);$$

so (15) is proved. Now the symmetry of G implies the relative boundedness of $G=(G^{-1})^{-1}$ with respect to G^2. Hence according to Lemma 2, (ii)

$$\overline{G^2}\,\overline{G^{-1}}\subseteq\overline{G^2\,\overline{G^{-1}}}=\bar{G}.$$

Therefore we get with (14) and (13)

$$LKL=\overline{G^{-1}}(\bar{\hat{A}}+i\kappa\,\overline{G^2})\,\overline{G^{-1}}\subseteq\overline{G^{-1}}(\bar{\hat{A}}\,\overline{G^{-1}}+i\kappa\bar{G})=A_G+i\kappa E.$$

To prove the equality in (16) the inclusion

$$D(A_G+i\kappa E)=D(A_G)\subseteq D(LKL) \tag{19}$$

remains to be verified. Let u be an element with

$$u\in D(A_G)=\{v\,|\,v\in D(\overline{G^{-1}}),\ \overline{G^{-1}}\,v\in D(\bar{\hat{A}}),\ \bar{\hat{A}}\,\overline{G^{-1}}\,v\in D(\overline{G^{-1}})\}.$$

Since $D(\bar{\hat{A}})=D(K)\subseteq D(\overline{G^2})$, we have $Lu=\overline{G^{-1}}\,u\in D(K)$ and

$$KLu=(\bar{\hat{A}}+i\kappa\,\overline{G^2})\,\overline{G^{-1}}\,u=\bar{\hat{A}}\,\overline{G^{-1}}\,u+i\kappa\,\bar{G}\,u\in D(L);$$

thus $u\in D(LKL)$. This proves (19).

Ad (III.iii). Applying Lemma 3 to the operators K, L owing to (III.ii) we obtain

$$\overline{R(A_G+i\kappa E)}=\overline{R(LKL)}=H.$$

On account of the symmetry of A_G (which is easily seen from (13)) we thus have deduced the essential selfadjointness of A_G. – As A_G is an extension of A the proof of the theorem is complete. ▮

§ 3. Relations for Some Vector Operators

The Hilbert space we shall use in the subsequent sections is $H=[L^2(\mathbb{R}^3)]^4$, the space of all functions u: $\mathbb{R}^3\to\mathbb{C}^4$ with components $u_j\in L^2(\mathbb{R}^3)$ (these functions just as the elements of $\mathbb{C}^4$ are written as column vectors). The underlying inner product is

$$(u,v)=\int_{\mathbb{R}^3}[u(x)]^t\,\overline{v(x)}\,dx;$$

here "t" denotes the transposition of the matrix $u(x)$. Writing $\mathbb{R}^3_+:=\mathbb{R}^3\setminus\{o\}$ we also use the spaces

$$D_0:=\{u\,|\,u\in H,\ u_j\in C_0^\infty(\mathbb{R}^3_+)\ \ (j=1,\dots,4)\},$$

$$D_1:=\{u\,|\,u\in H,\ u_j\in C_0^0(\mathbb{R}^3_+)\ \ (j=1,\dots,4)\}.$$

Moreover, let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be Hermitian 4×4 matrices which satisfy the commutation relations

$$\alpha_j \alpha_k + \alpha_k \alpha_j = 2\delta_{jk} E \qquad (j, k = 1, 2, 3, 4),$$

where E is the unit 4×4 matrix (also being identified with the identity operator in H).

We then introduce the following vector operators (cf. [11]) the components of which are assigned to have the domain D_0:

a) $\vec{x} := (x_1, x_2, x_3)$ (x_j denotes the operator of multiplication by the position coordinate x_j),

b) $\vec{p} = (p_1, p_2, p_3) := -i\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right)$,

c) $\vec{L} = (L_1, L_2, L_3) := [\vec{x}, \vec{p}]$,

d) $\vec{\alpha} := (\alpha_1, \alpha_2, \alpha_3)$,

e) $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3) := -\frac{i}{2}[\vec{\alpha}, \vec{\alpha}]$.

If f is a real-valued function $\in C^0((0, \infty))$ the operator of multiplication by $f(|x|)$ defined on D_1 is denoted by $f(r)$ (or sometimes only by f). With the help of the above vector operators we then define the following "scalar" operators:

f) $p_r := r^{-1}(\vec{x} \cdot \vec{p} - iE)$, g) $L_\sigma := \vec{\sigma} \cdot \vec{L}$, h) $\alpha_r := r^{-1}(\vec{\alpha} \cdot \vec{x})$.

Lemma 4. *The operators* f), g), h) *and the components of the vector operators* a), ..., e) *are symmetric. Let $f \in C^1((0, \infty))$. Then the following identities hold* ($u \in D_0$):

$$(\vec{p} \cdot \vec{x} - \vec{x} \cdot \vec{p})\, u = -3iu, \tag{20}$$

$$(\vec{p} f(r) - f(r)\vec{p})\, u = -i\vec{x} f'(r) r^{-1} u, \tag{21}$$

$$(p_r f(r) - f(r) p_r)\, u = -i f'(r)\, u, \tag{22}$$

$$\vec{p}^2 u = p_r^2 u + r^{-2} \vec{L}^2 u, \tag{23}$$

$$(\vec{\alpha} \cdot \vec{x})^2 u = r^2 u, \qquad (\vec{\alpha} \cdot \vec{p})^2 u = \vec{p}^2 u, \tag{24}$$

$$\begin{aligned} (\vec{\alpha} \cdot \vec{x})(\vec{\alpha} \cdot \vec{p})\, u &= (\vec{x} \cdot \vec{p})\, u + i L_\sigma u, \\ (\vec{\alpha} \cdot \vec{p})(\vec{\alpha} \cdot \vec{x})\, u &= (\vec{p} \cdot \vec{x})\, u - i L_\sigma u, \end{aligned} \tag{25}$$

$$(\vec{\alpha} \cdot \vec{p}) f(r) (\vec{\alpha} \cdot \vec{p})\, u = p_r f(r) p_r u + f(r) r^{-2} \vec{L}^2 u + r^{-1} f'(r)(L_\sigma + E)\, u, \tag{26}$$

$$[(\vec{\alpha} \cdot \vec{p}) f(r) - f(r)(\vec{\alpha} \cdot \vec{p})]\, u = -i\alpha_r f'(r)\, u, \tag{27}$$

$$[(\vec{\alpha} \cdot \vec{p}) f(r)\alpha_r - f(r)\alpha_r(\vec{\alpha} \cdot \vec{p})]\, u = -i f'(r)\, u - 2i r^{-1} f(r)(L_\sigma + E)\, u, \tag{28}$$

$$[(\vec{\alpha} \cdot \vec{p}) f(r)\alpha_4 + \alpha_4 f(r)(\vec{\alpha} \cdot \vec{p})]\, u = i\alpha_4 \alpha_r f'(r)\, u. \tag{29}$$

Proof. Most of these identities have been proved in [11]. The rest can easily be verified by calculation. ∎

Lemma 5. *Let* $f \in C^1((0, \infty))$ *be a positive-valued function. Then* $f(r)$ *commutes with all components of* $\vec{L}$ *and with* L_σ, *and the following inequalities hold* $(u \in D_0)$:

$$(f(r)\vec{L}^2 u, u) \geqq 0, \tag{30}$$

$$(f(r) L_\sigma u, u) \geqq -(f(r)\vec{L}^2 u, u). \tag{31}$$

Proof. See [11]. ▮

Lemma 6. *(An inequality of Hardy's type.) Let m be a real number and let* $u \in D_0$. *Then*

$$(r^{2m} p_r u, p_r u) \geqq (m - \tfrac{1}{2})^2 (r^{2m-2} u, u). \tag{32}$$

Proof. We start from the commutation relation (see (22))

$$p_r r^{2m-1} - r^{2m-1} p_r = -i(2m-1) r^{2m-2} | D_0 .$$

Right- and left-sided multiplication by r^{1-m} gives

$$r^{1-m} p_r r^m - r^m p_r r^{1-m} = -i(2m-1) E | D_0 .$$

Now, let $u \in D_0$ and (excluding the trivial case) $u \neq \Theta$, $m \neq \frac{1}{2}$. Then $v := r^{m-1} u \neq \Theta$, and we obtain:

$$\begin{aligned} |2m-1| \, \|v\|^2 &= |(-i(2m-1) v, v)| = |(r^{1-m} p_r r^m v - r^m p_r r^{1-m} v, v)| \\ &= |(v, r^m p_r r^{1-m} v) - (r^m p_r r^{1-m} v, v)| \leqq 2 \, \|v\| \, \|r^m p_r r^{1-m} v\| , \end{aligned}$$

hence after division by $2\|v\|$ and squaring

$$(m - \tfrac{1}{2})^2 \, \|v\|^2 \leqq \|r^m p_r r^{1-m} v\|^2 .$$

Revoking the substitution we get our assertion. ▮

§ 4. Selfadjoint Extensions of Dirac Operators

Let $q \in L^2_{\text{loc}}(\mathbb{R}^3_+)$ be a real-valued function. We consider the formal Dirac operator

$$\mathcal{T} := \vec{\alpha} \cdot \vec{p} + \beta + q(x)$$

(here we write conventionally $\beta := \alpha_4$ and understand the vector operators only in a formal sense) and associate with it the *minimal Dirac operator* T given by

$$Tu = \mathcal{T} u, \qquad D(T) = D_0 .$$

As an application of Theorem 1 we now state our main theorem.

Theorem 2. *Let* $s \in [0, 1)$, $k > 1$, $c > 1$. *Assume that q can be expressed as* $q = q_1 + q_2$ *where* $q_1 \in C^0(\mathbb{R}^3_+)$, $q_2 \in L^\infty(\mathbb{R}^3)$ *are real-valued functions* *with* $(r:=|x|)$

$$\frac{1}{r^2}\left[f(r)+\frac{s}{2}\right]^2 \leqq k\left[|q_1(x)|^2+\frac{1}{r^2}f^2(r)\right]$$

$$\leqq \frac{1}{r^2}\left[f(r)+\frac{s+1}{2}\right]^2+\frac{1}{r}f'(r), \qquad x\in \mathbb{R}^3_+,$$

$f\in C^1((0,\infty))$ *being a positive-valued function which is bounded from above by* $(1-s)/2c$. *Then there exists a bounded symmetric operator* S *such that the operator*

$$T_G:=\overline{r^{-s/2}\,\overline{r^{s/2}(T-S)}}+\bar{S}$$

is an essentially selfadjoint extension of T.

Remark. For $s=0$ the theorem reduces to an assertion we have essentially proved in [11].

Proof. We shall apply Theorem 1. Therefore we identify the Hilbert spaces equally denoted, in like manner the spaces denoted by D_0, D_1, respectively. Furthermore, setting

$$g(t):=t^{-1}f(t) \qquad (0<t<\infty),$$

$$S_0:=\frac{1}{2k}r^{-1}(g^2+q_1^2)^{-1}(s+2rg)[g\beta-i\,q_1\,\beta\,\alpha_r],$$

$$S:=S_0+q_2, \qquad D(S)=D(S_0)=D_0,$$

we make the following identifications:

$$A_0=\vec{\alpha}\cdot\vec{p}+\beta, \qquad C=i\,g\,\alpha_r,$$

$$B=-q_1+S_0 \quad (D(B)=D_0), \quad G=r^{s/2}.$$

The symmetry and the boundedness of S are easy to verify. Since A_0 is essentially selfadjoint (see [11]) and the inclusions (i), (ii), (iii) involved in Theorem 1 are obvious it remains to show that the inequalities in (iv) hold. For then we have proved that the operator $\overline{G^{-1}\,\overline{G(A_0-B)}}$ is an essentially selfadjoint extension of

$$A:=A_0-B=T-q_1-q_2+q_1-S_0=T-S$$

from which the proposition of the theorem is evident.

Let $u\in D_0$. Then we have

$$\|A_0\,u\|^2=\|(\vec{\alpha}\cdot\vec{p}+\beta)\,u\|^2=(\vec{p}^2\,u,u)+\|u\|^2\geqq\|u\|^2$$

so that (1) is satisfied with $c_0=1$.

Next, putting $v:=G^2 u$ we get in virtue of Lemma 4, (23), Lemma 5, (30), and Hardy's inequality (Lemma 6 with $m=0$):

$$\|A_0 v\|^2 \geqq \tfrac{1}{4}(r^{-2} v, v) + \|v\|^2 = (\{\tfrac{1}{4} r^{-2(1-s)} + r^{2s}\} u, u).$$

It is clear that the right-hand side is greater than const $\|u\|^2$ with a positive constant which proves (2).

Now put $w:=Gu$. Then according to Lemma 4, (27), and Hardy's inequality we get

$$\begin{aligned}\|(A_0 G - G A_0) G^{-1} w\|^2 &= \|[(\vec{\alpha}\cdot\vec{p}) r^{s/2} - r^{s/2}(\vec{\alpha}\cdot\vec{p})] r^{-s/2} w\|^2\\ &= \left\|\left(-i\alpha_r \frac{s}{2} r^{-1}\right) w\right\|^2\\ &= \frac{s^2}{4}(r^{-2} w, w) \leqq s^2 (p_r^2 w, w) \leqq s^2 \|A_0 w\|^2,\end{aligned}$$

so that (3) is fulfilled with any $c_2 \in (1, 1/s)$.

As to the inequality (4), we state that it is satisfied with $c_3 = c$. First we obtain by means of the identities of Lemma 4:

$$\begin{aligned}&A_0 G^2 A_0 u - c^2 G^2 C^* C u\\ &\quad = (\vec{\alpha}\cdot\vec{p}) r^s (\vec{\alpha}\cdot\vec{p}) u + \{(\vec{\alpha}\cdot\vec{p}) r^s \beta + \beta r^s (\vec{\alpha}\cdot\vec{p})\} u + r^s u - c^2 r^s g^2 u\\ &\quad = p_r r^s p_r u + r^{s-2} \vec{L}^2 u + s r^{s-2}(L_\sigma + E) u + i\beta\alpha_r s r^{s-1} u + r^s u - c^2 r^s g^2 u.\end{aligned}$$

Using the inequalities of the Lemmata 5 and 6 and the assumptions of the theorem we find

$$\begin{aligned}&\|G A_0 u\|^2 - c^2 \|G C u\|^2 = (A_0 G^2 A_0 u, u) - c^2 (G^2 C^* C u, u)\\ &\quad \geqq \left(\frac{s}{2} - \frac{1}{2}\right)^2 (r^{s-2} u, u) + (r^{s-2} \vec{L}^2 u, u) - s(r^{s-2} \vec{L}^2 u, u) + s(r^{s-2} u, u)\\ &\qquad - s(r^{s-1} u, u) + (r^s u, u) - c^2 (f^2 r^{s-2} u, u) \geqq 0.\end{aligned}$$

To verify the last inequality—with $c_4 = \sqrt{k}$—we start from the expression

$$\begin{aligned}\delta &:= \|G(A_0 + C) u\|^2 - k \|G(B + C) u\|^2\\ &= (A_0 G^2 A_0 u + \{A_0 G^2 C + C^* G^2 A_0\} u + G^2 C^* C u, u)\\ &\quad - k\big(G^2 [(-q_1 + C^*) + S_0][(-q_1 + C) + S_0] u, u\big).\end{aligned}$$

Some calculations yield:

$$\{A_0 G^2 C + C^* G^2 A_0\} u = (r^{s-1} f)' u + 2 r^{s-2} f (L_\sigma + E) u + 2 i r^{s-1} f \beta \alpha_r u,$$

$$i \beta \alpha_r (s + 2f) r^{s-1} u = k G^2 [(-q_1 + C^*) S_0 + S_0(-q_1 + C)] u,$$

$$S_0^2 = \frac{1}{4k^2} r^{-2} (g^2 + q_1^2)^{-1} (s + 2f)^2.$$

Regarding the assumption of the theorem we get with the help of the relations of §3:

$$\begin{aligned}
\delta = {} & (r^s p_r u, p_r u) + (r^{s-2} \vec{L}^2 u, u) + s(r^{s-2} L_\sigma u, u) + s(r^{s-2} u, u) \\
& + i(\beta \alpha_r s r^{s-1} u, u) + (r^s u, u) + (f^2 r^{s-2} u, u) + (s-1)(r^{s-2} f u, u) \\
& + (r^{s-1} f' u, u) + 2(r^{s-2} f L_\sigma u, u) + 2(r^{s-2} f u, u) + (i r^{s-1} 2 f \beta \alpha_r u, u) \\
& - k(r^s (q_1^2 + g^2) u, u) - k(G^2 [(-q_1 + C^*) S_0 + S_0(-q_1 + C)] u, u) \\
& - k\left(r^s \frac{1}{4k^2} r^{-2} (g^2 + q_1^2)^{-1} (s + 2f)^2 u, u\right) \\
\geqq {} & \left(\left\{\frac{s^2}{4} - \frac{s}{2} + \frac{1}{4} + s + f^2 - k(q_1^2 + g^2) r^2 + (s+1) f + r f'\right\} r^{s-2} u, u\right) \\
& + (\{1 - s - 2f\} r^{s-2} \vec{L}^2 u, u) + \left(r^s \left\{1 - \frac{1}{4k} r^{-2} (g^2 + q_1^2)^{-1} (s + 2f)^2\right\} u, u\right) \\
\geqq {} & 0.
\end{aligned}$$

Thus (iv) is shown to be satisfied and the theorem is proved. ∎

A more explicit characterization of the constructed selfadjoint extension $\bar{T}_G$ of T is given in the following theorem. We point out here that under the above assumptions on q the adjoint of T can be characterized by

$$T^* u = \mathcal{T} u,$$

$$D(T^*) = \{u \mid u \in H,\ u_j \in H^1_{\text{loc}}(\mathbb{R}^3_+)(j = 1, \ldots, 4),\ \mathcal{T} u \in H\},$$

$H^1_{\text{loc}}(\mathbb{R}^3_+)$ being the usual Sobolev space of L^2_{loc}-functions which possess local derivatives of first order in the distributional sense (cf. Evans [1]).

Theorem 3. *Assume q to satisfy the hypotheses of Theorem 2. Let T^*_σ denote the restriction of T^* to the subspace $D^*_\sigma := D(T^*) \cap D(r^{-\sigma})$, σ being a real number. Then*

$$\bar{T}_G = T^*_\sigma \quad \left(\frac{1}{2} \leqq \sigma \leqq 1 - \frac{s}{2}\right). \tag{33}$$

Especially $T^*_{\frac{1}{2}}$, the restriction of T^* to the subspace

$$D^*_{\frac{1}{2}} = \left\{ u \,|\, u \in H,\ u_j \in H^1_{\mathrm{loc}}(\mathbb{R}^3_+)\,(j=1,\ldots,4),\ \mathcal{T}u \in H,\ \int_{\mathbb{R}^3} \frac{|u(x)|^2}{|x|}\,dx < \infty \right\},$$

is an essentially selfadjoint extension of T (its closure being equal to $\overline{T_G}$).

Proof. Let $\sigma \in \left[\frac{1}{2}, 1 - \frac{s}{2}\right]$. We proceed in two steps.

Step 1. $\overline{T_G} \subseteq \overline{T^*_\sigma}$.

Since $T \subseteq T_G$ and therefore $T_G \subseteq \overline{T_G} = T^*_G \subseteq T^*$, it is sufficient to prove the inclusion

$$D(T_G) \subseteq D(\overline{r^{-\sigma}}). \tag{34}$$

We start from the elementary fact that there are two positive numbers a_1, a_2 such that the inequality

$$1 \leqq a_1\, t^{2\sigma - s} + a_2\, t^{-(2-s-2\sigma)}$$

holds for all $t \in (0, \infty)$. Multiplying by $t^{-2\sigma}$ we obtain an inequality from which the relation

$$\|r^{-\sigma} u\|^2 \leqq a_1\, \|r^{-s/2} u\|^2 + a_2\, \|r^{(s/2)-1} u\|^2, \qquad u \in D_0, \tag{35}$$

is obvious. On the other hand, connecting the inequalities (10), (6), and (8) in the proof of Theorem 1 and replacing u by $G^{-1} u$ we get $(u \in D_0)$:

$$\begin{aligned} \|GAu\| &\geqq c_5\, c_8\, \|G^{-1} u\| = c_5\, c_8\, \|r^{-(s/2)} u\|, \\ \|GAu\| &\geqq c_7\, c_8\, \|A_0\, G u\| \geqq \tfrac{1}{2}\, c_7\, c_8\, \|r^{(s/2)-1} u\|, \end{aligned} \tag{36}$$

where we have used Hardy's inequality again. From the inequalities (35) and (36) we easily infer

$$D(\overline{GA}) \subseteq D(\overline{r^{-\sigma}}).$$

Since $T_G = A_G + \bar{S}$ and $\bar{S}$ is bounded we get

$$D(T_G) = D(A_G) = D(\overline{G^{-1}}\,\overline{GA}) \subseteq D(\overline{GA}) \subseteq D(\overline{r^{-\sigma}}),$$

which proves (34).

Step 2. $\overline{T^*_\sigma} \subseteq \overline{T_G}$.

It is sufficient to prove the symmetry of T^*_σ since in that case we obtain

$$\overline{T_G} \subseteq \overline{T^*_\sigma} \subseteq (T^*_\sigma)^*,$$

hence

$$\overline{T^*_\sigma} = (T^*_\sigma)^{**} \subseteq (\overline{T_G})^* = \overline{T_G}.$$

Thus we have to show

$$\operatorname{Im}(T^* u, u) = 0, \qquad u \in D^*_\sigma. \tag{37}$$

Let $u \in D_\sigma^*$, $u \neq \Theta$. Let ρ, R be numbers with $0<\rho<R$ and set

$$I_{\rho,R} := i \int_{G_{\rho,R}} \{[\mathcal{T}u(x)]^t \overline{u(x)} - [u(x)]^t \overline{\mathcal{T}u(x)}\}\, dx$$

where $G_{\rho,R} := \{x \mid x \in \mathbb{R}^3, \rho < |x| < R\}$. Then obviously

$$\eta := i\{(T^* u, u) - (u, T^* u)\} = \lim_{\substack{\rho \to 0 \\ R \to \infty}} I_{\rho,R}.$$

Now it is our aim to apply Gauss' theorem and then to estimate the boundary terms. In general this is not possible since the functions in the class (!) u will not be sufficiently regular. However, according to a theorem of Sobolev every class $v \in H^1_{\mathrm{loc}}(\mathbb{R}^3_+)$ possess a distinguished representative for which the restriction to the twodimensional spheres $|x| = \tau$ $(0 < \tau < \infty)$ can be defined as a square integrable function; moreover, this representative allows the application of Gauss' theorem (Ladyženskaya-Ural'ceva [6], Smirnov [13]; see also Kalf [3]). We assume the representative of our $u \in D_\sigma^*$ to be chosen such that his four components have the mentioned distinguished property. Then an easy calculation yields:

$$I_{\rho,R} = \int_{G_{\rho,R}} \left\{ \sum_{j=1}^{3} \left[\alpha_j \frac{\partial u}{\partial x_j}\right]^t \bar{u} + u^t \left[\sum_{j=1}^{3} \bar{\alpha}_j \frac{\partial \bar{u}}{\partial x_j}\right] \right\} dx$$

$$= \int_{\partial G_{\rho,R}} \sum_{j=1}^{3} [\alpha_j u(x)]^t \overline{u(x)}\, \nu_j\, do = I_\rho + I_R,$$

where we put

$$I_\rho := -\frac{1}{\rho} \int_{|x|=\rho} \sum_{j=1}^{3} [\alpha_j u]^t \bar{u}\, x_j\, do, \qquad I_R := \frac{1}{R} \int_{|x|=R} (.\ .)\, do.$$

Since the limits exist separately we have

$$\eta = \lim_{\rho \to 0} I_{\rho,R} = \lim_{\rho \to 0} I_\rho + \lim_{R \to \infty} I_R \equiv \gamma_0 + \gamma_1.$$

Now we show $\gamma_1 = \gamma_0 = 0$. If γ_1 would not be equal to zero, then for a certain $R_0 > 0$

$$0 < \tfrac{1}{2}|\gamma_1| \leqq |I_R| \leqq \mathrm{const} \int_{|x|=R} |u|^2\, do \qquad (R \geqq R_0).$$

Integration from R_0 to ∞ leads to $\infty \leqq \|u\|^2$ which is a contradiction to $u \in H$. In order to prove $\gamma_0 = 0$ we need the boundary condition $u \in \overline{D(r^{-\sigma})}$. Assume $\gamma_0 \neq 0$. Then for a certain $\rho_0 > 0$ we have $|I_\rho| \geqq \frac{1}{2}|\gamma_0|$ $(0 < \rho \leqq \rho_0)$, hence with a positive constant

$$\infty > \|\overline{r^{-\sigma}} u\|^2 \geqq \int_{|x| < \rho_0} |u(x)|^2 |x|^{-2\sigma}\, dx \geqq \mathrm{const} \int_0^{\rho_0} \rho^{-2\sigma}\, d\rho = \infty$$

which is absurd. This proves $\eta = 0$ and so (37). ∎

Example. We consider the Dirac operator T with the Coulomb potential

$$q(x) = -\frac{\mu}{|x|}, \qquad x \in \mathbb{R}^3_+,$$

μ being a positive number. One easily verifies that the assumptions of Theorem 2 are satisfied if the inequality

$$1 - 4\mu^2 \leqq (1-s)^2 < 4(1-\mu^2)$$

holds ($s \in [0, 1)$ a suitable number) so that they can be fulfilled for any $\mu < 1$.

In the case $\mu < \frac{1}{2}\sqrt{3}$ it is possible to choose $s = 0$ so that T is seen to be essentially selfadjoint (cf. Rellich [8, 10], Evans [1], Weidmann [15], and [11]).

In the case $\frac{1}{2}\sqrt{3} < \mu < 1$ one must choose $s > 0$ (because T is no longer essentially selfadjoint as we know!). Then according to Theorem 3 for $\frac{1}{2} \leqq \sigma < \frac{1}{2} + \sqrt{1-\mu^2}$ the closure of T_σ^* is constant and equals the distinguished selfadjoint extension $\overline{T_G}$. The special boundary condition $u \in D(r^{-\frac{1}{2}})$ can obviously be regarded as an expression for the finiteness of the potential energy, just as in analogous cases for the Schrödinger operator (cf. Kalf [3]), so that the distinction of $\overline{T_G}$ becomes physically conspicuous.

References

1. Evans, W. D.: On the unique self-adjoint extension of the Dirac operator and the existence of the Green matrix. Proc. London Math. Soc. (3) **20**, 537–557 (1970).
2. Friedrichs, K. O.: Über die ausgezeichnete Randbedingung in der Spektraltheorie der halbbeschränkten gewöhnlichen Differentialoperatoren zweiter Ordnung. Math. Ann. **112**, 1–23 (1935/36).
3. Kalf, H.: On the characterization of the Friedrichs extension of ordinary or elliptic differential operators with a strongly singular potential. J. Functional Analysis **10**, 230–250 (1972).
4. Kato, T.: Fundamental properties of Hamiltonian operators of Schrödinger type. Trans. Amer. Math. Soc. **70**, 195–211 (1951).
5. Kato, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966.
6. Ladyženskaya, O. A., Ural'ceva, N. N.: Linear and quasi-linear elliptic equations. New York and London: Academic Press 1968.
7. Landau, L. D., Lifschitz, E. M. (W. B. Berestetzki, L. P. Pitajewski): Relativistische Quantentheorie. Berlin: Akademie-Verlag 1970.
8. Rellich, F.: Die zulässigen Randbedingungen bei den singulären Eigenwertproblemen der mathematischen Physik. Math. Z. **49**, 702–723 (1943/44).
9. Rellich, F.: Halbbeschränkte gewöhnliche Differentialoperatoren zweiter Ordnung. Math. Ann. **122**, 343–368 (1951).
10. Rellich, F.: Eigenwerttheorie partieller Differentialgleichungen II. Lecture notes: Göttingen 1953.
11. Schmincke, U.-W.: Essential selfadjointness of Dirac operators with a strongly singular potential. Math. Z. **126**, 71–81 (1972).

12. Schmincke, U.-W.: Ausgezeichnete selbstadjungierte Erweiterungen von Dirac-Operatoren mit stark singulärem (insbesondere Coulomb-) Potential. Habilitationsschrift RWTH Aachen 1972.
13. Smirnov, V. I.: Functional analysis, a course of higher mathematics, V, Moscow: 1959 [in Russian]. English translation: Reading, Mass. 1964.
14. Sobolev, S. L.: Applications of functional analysis in mathematical physics. Providence, RI: Amer. Math. Soc. 1963.
15. Weidmann, J.: Oszillationsmethoden für Systeme gewöhnlicher Differentialgleichungen. Math. Z. **119**, 349–373 (1971).

Dr. Upke-Walther Schmincke
Institut für Mathematik
Technische Hochschule Aachen
D-5100 Aachen, Templergraben 55
Federal Republic of Germany

(Received September 4, 1972)

Math. Z. 129, 351–358 (1972)
© by Springer-Verlag 1972

Vanishing Tensor Powers of Modules

Roger Wiegand and Sylvia Wiegand

A module M over a commutative ring R is said to *vanish* if the tensor product of some finite number of copies of M is 0. The smallest integer d such that $\otimes_R^{d+1} M=0$ is denoted by $\delta_R(M)$, and we set $\Delta(R)=\sup \delta_R(M)$, the supremum taken over all vanishing modules. In § 1 we show that if R is noetherian then $\Delta(R)$ is less than or equal to the Krull dimension of R. The second section is devoted to a $\mathbb{Z}/(p^\infty)$-like construction over general commutative rings, and some applications are given in § 3. In particular, it is shown that if some proper ideal of R contains an R-sequence of length d, then $\Delta(R)\geqq d$. In the last two sections we show that $\Delta(R/(\mathrm{soc}(R))^2)=\Delta(R)$ for every ring R, and that the vanishing R-modules form a (non-hereditary) torsion class provided $\Delta(R)$ is finite.

We make the standing assumption that all rings are commutative and all modules and ring homomorphisms are unitary. When there is no danger of confusion, we write $M\otimes N$ for $M\otimes_R N$, and $\delta(M)$ for $\delta_R(M)$.

1. Noetherian Rings

We begin with a simple necessary condition for a module to vanish, which implies, in particular, that a non-zero vanishing module cannot be finitely generated.

1.1. Proposition. *If $_RM$ is a vanishing module, then $M_P=PM_P$ for each prime P.*

Proof. Let $K=R_P/PR_P$, and let $V=M_P/PM_P$. Then, if $\otimes_R^n M=0$, we have $0=(\otimes_R^n M)\otimes_R K=\otimes_K^n(M\otimes_R K)=\otimes_K^n V$. Since K is a field, this clearly implies $V=0$, as desired.

If M is a vanishing abelian group, (1.1) says that $pM=M$ for each prime p, and $M\otimes Q=0$. In other words, M is a divisible torsion group. Conversely, if M is divisible and torsion, clearly $M\otimes M=0$.

The converse of (1.1) is false in general. For example, let R be a ring with exactly one prime ideal P, and suppose $0\neq P=P^2$. (A suitable homomorphic image of a non-discrete, rank 1 valuation ring will do.) Then $\otimes_R^n P$ maps onto $P^n=P$, so P does not vanish, but $P_P=PP_P$. In §3 we shall see that there are noetherian examples. On the other hand, if R is noetherian and has finite Krull dimension, we have the following:

1.2. Theorem. *Assume R is noetherian with finite Krull dimension d. Then the following conditions on a module M are equivalent:*

(i) *M vanishes.*

(ii) $\otimes^{d+1} M = 0$.

(iii) *$M_P = PM_P$ for each prime ideal P.*

Proof. (ii) $\Rightarrow$ (i) trivially, and (i) $\Rightarrow$ (iii) by (1.1). To prove (iii) $\Rightarrow$ (ii) we may assume R is a local ring with maximal ideal $\mathcal{M}$. We proceed by induction on d. If $d=0$, then $\mathcal{M}$ is nilpotent, and we have $M = \mathcal{M} M = \dots = \mathcal{M}^n M = 0$. Now assume $d>0$, and set $N = \otimes^d M$. Then, for each prime $P \neq \mathcal{M}$, we have $N_P = \otimes^d_{R_P} M_P = 0$, since R_P has Krull dimension at most $d-1$.

Let F be an arbitrary finitely generated R-submodule of N, and let I be the annihilator of F in R. Let P be an arbitrary non-maximal prime. Then since $N_P = 0$, we have $F_P = 0$, and since F is finitely generated, $I \nsubseteq P$. Therefore either $I = R$ or $\mathcal{M}/I$ is the unique prime ideal of R/I, and in either case $\mathcal{M}^n \subseteq I$ for some n. Therefore $F \otimes M = F \otimes \mathcal{M}^n M = \mathcal{M}^n F \otimes M = 0$. Finally, $\otimes^{d+1} M = N \otimes M = \varinjlim_F (F \otimes M) = 0$.

If we let $\dim(R)$ denote the Krull dimension of R, then (1.2) implies that $\Delta(R) \leqq \dim(R)$ for noetherian rings R. In § 3 we shall see that if K is a field then $\Delta(K[x_1, \dots, x_d]) = d$, so that the estimate given above cannot be improved in general.

2. M-Sequences

Let M be an R-module and let $x_1, \dots, x_d$ be elements of R. Then $(x_1, \dots, x_d)$ is said to be an *M-sequence* provided

(i) For each j $(1 \leqq j \leqq d)$, x_j is a non-zero divisor on $M/M(x_1, \dots, x_{j-1})$, and

(ii) $M(x_1, \dots, x_d) \neq M$.

Remark. Let k be an integer $(1 \leqq k \leqq d)$. Then $(x_1, \dots, x_d)$ is an M-sequence if and only if $(x_1, \dots, x_k)$ is an M-sequence and $(x_{k+1}, \dots, x_d)$ is an $M/M(x_1, \dots, x_k)$-sequence.

Given an *arbitrary* sequence $x_1, \dots, x_d$ in R and an R-module M, we let $[M; x_1, \dots, x_d]$ denote the direct limit $\varinjlim_n M/M(x_1^n, \dots, x_d^n)$, with respect to the maps $\varphi_{n,m}$: $M/M(x_1^m, \dots, x_d^m) \to M/M(x_1^n, \dots, x_d^n)$ induced by multiplication by $(x_1 \dots x_d)^{n-m}$. (The notation $M/M(x_1^\infty, \dots, x_d^\infty)$ would be more suggestive, but is more cumbersome.) The next lemma will not be used until the next section, but this seems to be an appropriate place to record it.

2.1. Lemma. *Let M be an R-module and let $x_1, \ldots, x_d \in R$.*

(a) *For each permutation π on $\{1, \ldots, d\}$,*

$$[M; x_{\pi(1)}, \ldots, x_{\pi(d)}] \simeq [M; x_1, \ldots, x_d].$$

(b) *For each R-module N,*

$$M \otimes [N; x_1, \ldots, x_d] \simeq [M \otimes N; x_1, \ldots, x_d].$$

(c) *For each j $(1 < j < d)$,*

$$[[M; x_1, \ldots, x_j]; x_{j+1}, \ldots, x_d] \simeq [M; x_1, \ldots, x_d].$$

(d) *For each $x \in R$, $[M; x, x] = 0$.*

Proof. The proofs of (a), (b), and (c) are straightforward. To prove (d), simply note that $\varphi_{n,m}$ is induced by multiplication by $x^{2(n-m)}$; hence $\varphi_{2n,n}(M/M x^n) = 0$.

We will need the following technical lemma, part of which occurs as an exercise in [2]:

2.2. Lemma. *Let $(x_1, \ldots, x_d)$ be an M-sequence, and let n_j be positive integers $(1 \leqq j \leqq d)$. Then the following are true:*

(A) *$(x_1^{n_1}, \ldots, x_d^{n_d})$ is an M-sequence.*

(B) *If $a \in M$ and $a x_i \in M(x_1^{n_1}, \ldots, x_d^{n_d})$ for some i, then*

$$a \in M(x_1^{n_1}, \ldots, x_{i-1}^{n_{i-1}}, x_i^{n_i - 1}, x_{i+1}^{n_{i+1}}, \ldots, x_d^{n_d}).$$

Proof. We show first of all that

$$M x_1 \cap M(x_2, \ldots, x_d) = M x_1(x_2, \ldots, x_d).$$

Obviously

$$M x_1(x_2, \ldots, x_d) \subseteq M x_1 \cap M(x_2, \ldots, x_d).$$

For the reverse inclusion, let

$$a x_1 = a_2 x_2 + \cdots + a_d x_d.$$

Then $a_d \in M(x_1, \ldots, x_{d-1})$, and we may write

$$a x_1 = b_2 x_2 + \cdots + b_{d-1} x_{d-1} + c x_1 x_d.$$

Then $b_{d-1} \in M(x_1, \ldots, x_{d-2})$, and we can repeat the process, eventually getting $a x_1 \in M x_1(x_2, \ldots, x_d)$, as desired.

We can now verify the following special case of (B):

(B′) If $a \in M$, $n \geqq 1$, and $a x_1 \in M(x_1^n, x_2, \ldots, x_d)$, then $a \in M(x_1^{n-1}, x_2, \ldots, x_d)$. To see this, write $a x_1 = b x_1^n + c$, where $c \in M(x_2, \ldots, x_d)$. But then $c \in M x_1$, so $c \in M x_1(x_2, \ldots, x_d)$. Since x_1 is a non-zero divisor, (B′) follows.

Next, we prove a special case of (A), namely:

(A′) For each $n \geqq 1$, $(x_1^n, x_2, \ldots, x_d)$ is an M-sequence. Since (A′) is obvious if either $n=1$ or $d=1$, we may assume inductively that $n>1$, $d>1$, and $(x_1^{n-1}, x_2, \ldots, x_d)$ and $(x_1^n, x_2, \ldots, x_{d-1})$ are both M-sequences. Suppose $a x_d \in M(x_1^n, x_2, \ldots, x_{d-1})$. Then $a \in M(x_1, x_2, \ldots, x_{d-1})$, and we can write $a = b x_1 + c$, with $c \in M(x_2, \ldots, x_{d-1})$. Then $b x_1 x_d \in M(x_1^n, x_2, \ldots, x_{d-1})$, and, by (B′) $b x_d \in M(x_1^{n-1}, x_2, \ldots, x_{d-1})$. But then $b \in M(x_1^{n-1}, x_2, \ldots, x_{d-1})$. Since $a = b x_1 + c$, we have $a \in M(x_1^n, x_2, \ldots, x_{d-1})$, and (A′) is verified. Statement (A) follows immediately, by induction and the remark at the beginning of this section.

Finally, we prove (B). Set $y_j = x_j^{n_j}$ $(1 \leqq j \leqq d)$, and let $N = M(y_1, \ldots, y_{i-1})$. By (A), $(x_i, y_{i+1}, \ldots, y_d)$ is an (M/N)-sequence, and, letting $\bar{a} = a + N$, we have

$$\bar{a}\, x_i \in (M/N)(x_i^{n_i}, y_{i+1}, \ldots, y_d).$$

By (B′),

$$\bar{a} \in (M/N)(x_i^{n_i - 1}, y_{i+1}, \ldots, y_d),$$

and

$$a \in M(y_1, \ldots, y_{i-1}, x_i^{n_i - 1}, y_{i+1}, \ldots, y_d).$$

2.3. Theorem. *Let $(x_1, \ldots, x_d)$ be an M-sequence. Then the induced maps $\varphi_{n,m}$: $M/M(x_1^m, \ldots, x_d^m) \to M/M(x_1^n, \ldots, x_d^n)$ are monomorphisms. In particular, $[M; x_1, \ldots, x_d] \neq 0$.*

Proof. By 2.2(A) it will suffice to show that $\varphi_{2,1}$ is a monomorphism. Suppose $a \in M$ and $a x_1 \ldots x_d \in M(x_1^2, \ldots, x_d^2)$. Then

$$a x_1 \ldots x_{d-1} \in M(x_1^2, \ldots, x_{d-1}^2, x_d),$$

by 2.2(B). Repeating the process, we eventually get $a \in M(x_1, \ldots, x_d)$, as desired.

The construction of $[M; x_1, \ldots, x_d]$ was inspired by a conversation with David Eisenbud, who has shown (unpublished) that if $(x_1, \ldots, x_d)$ is an R-sequence generating the maximal ideal P of a regular local ring R, then $[R; x_1, \ldots, x_d]$ is the injective hull of the R-module R/P. Using (2.2) one can prove in general that $[M; x_1, \ldots, x_d]$ is an essential extension of $M/M(x_1, \ldots, x_d)$, provided $(x_1, \ldots, x_d)$ is an M-sequence. Of course, $[M; x_1, \ldots, x_d]$ may fail to be injective.

3. Applications

3.1. Proposition. *Let $S = R[x]$ be the polynomial ring in one indeterminate over R. If $\Delta(R) = \infty$ then $\Delta(S) = \infty$. If $\Delta(R) < \infty$, then $\Delta(S) \geqq \Delta(R) + 1$.*

Proof. It will suffice to show that if $\delta_R(M) = d$ then

$$\delta_S((M \otimes_R S) \oplus [S; x]) = d + 1.$$

We first observe that if A is any non-zero R module then (x) is an $A \otimes_R S$-sequence. (This is best seen by thinking of $A \otimes_R S$ as the module of polynomials with coefficients from A.)

Now if $\delta_R(M)=d$, we have $\otimes_S^{d+1}(M \otimes_R S)=0$, and since

$$[S; x] \otimes_S [S; x]=[S; x, x]=0,$$

it follows that $\otimes_S^{d+2}((M \otimes_R S) \oplus [S; x])=0$. On the other hand,

$$\begin{aligned}\otimes_S^{d+1}((M \otimes_R S) \oplus [S; x]) &= (\otimes_S^d(M \otimes_R S)) \otimes_S [S; x] \\ &= ((\otimes_R^d M) \otimes_R S) \otimes_S [S; x]=[(\otimes_R^d M) \otimes_R S; x] \neq 0,\end{aligned}$$

by Theorem 2.3.

Combining this result with Theorem 1.2 we have the following fact:

3.2. Corollary. *If K is a field then $\Delta(K[x_1, \ldots, x_n])=n$.*

Actually, Corollary 3.2 is a special case of a much more general result.

3.3. Theorem. *If some R-module M admits an M-sequence of length d, then $\Delta(R) \geqq d$.*

Proof. Let $(x_1, \ldots, x_d)$ be an M-sequence, and let

$$A=[R; x_1] \oplus \cdots \oplus [R; x_d].$$

From Lemma 2.1 we see that $\otimes^{d+1} A=0$, and $\otimes^d A=[R; x_1, \ldots, x_d]$. But then $M \otimes (\otimes^d A)=[M; x_1, \ldots, x_d] \neq 0$ by Theorem 2.3. Therefore $\otimes^d A \neq 0$.

The problem of determining $\Delta(R)$ is really a local one. The fact that $\Delta(R) \leqq \sup \Delta(R_P)$, P ranging over all maximal ideals, is easy, and has already been used (in the proof of Theorem 1.2). We prove the reverse inequality in a somewhat more general setting.

3.4. Lemma. *Let $R \to S$ be a ring homomorphism, and assume the multiplication map $S \otimes_R S \to S$ is an isomorphism. Then, for each S-module M, $\otimes_S^n M \simeq \otimes_R^n M$. Therefore $\Delta(S) \leqq \Delta(R)$.*

The proof is a straightforward juggling of associativity formulas. We record the following immediate consequence:

3.5. Proposition. *For each ring R,*

$$\Delta(R)=\sup\{\Delta(R_P) \mid P \text{ maximal}\}=\sup\{\Delta(R_P) \mid P \text{ prime}\}.$$

Recall that a Cohen-Macaulay ring is a noetherian ring R such that $R_{\mathcal{M}}$ contains an $R_{\mathcal{M}}$-sequence of length $\dim(R_{\mathcal{M}})$ for each maximal ideal $\mathcal{M}$. Combining (1.2), (3.3), and (3.5), we have the following observation:

3.6. Corollary. *If R is a Cohen-Macaulay ring then $\Delta(R)=\dim(R)$.*

3.7. Proposition. *If $\Delta(R)=0$ then $\dim(R)=0$, that is, prime ideals are maximal.*

Proof. If $\dim(R)>0$, then R maps onto a domain S that is not a field. If $0\neq x$ is a non-unit of S, then $\delta_S[S;x]=1$. Hence, by (3.4), $\Delta(R)\geqq\Delta(S)\geqq 1$.

G. Sabbagh has pointed out that the converse of (3.7) is false in general. For, suppose $f: R\to S$ is a non-surjective ring homomorphism satisfying the hypotheses of (3.4). Then it is readily verfied that the cokernel of f is a vanishing module, and hence $\Delta(R)>0$. Lazard [4, V, Proposition 3.2] has constructed a ring R of Krull dimension 0 that admits such a homomorphism.

We conclude this section with an example of a noetherian ring R for which $\Delta(R)=\infty$.

3.8. *Example* [5, p. 203]. Let K be a field and let P_i be the prime ideal of the polynomial ring $K[x_1, x_2, \ldots]$ generated by

$$\left\{x_j \,\middle|\, \frac{i(i-1)}{2}\leqq j\leqq\frac{i(i+1)}{2}\right\}.$$

Thus $P_1=(x_1)$, $P_2=(x_2, x_3), \ldots$. Let S be the intersection of the complements of the P_i, and let $R=S^{-1}K[x_1, x_2, \ldots]$. Then R_{P_iR} contains an R_{P_iR}-sequence of length i, and it follows that $\Delta(R)=\infty$. It is known [5, p. 203] that R is noetherian.

This example also shows that the converse of (1.1) can fail, even if R is noetherian. In fact, if R is an arbitrary ring such that $\Delta(R)=\infty$ then there exists a non-vanishing R-module M such that $M_P=PM_P$. To see this, choose, for each i, a module M_i such that $i\leqq\delta(M_i)<\infty$. Clearly, $M=\oplus\sum_i M_i$ is not a vanishing module.

4. Semiprime Rings

If M is an R-module we denote by $z(M)$ the singular submodule of M, that is, the set of elements x such that $(0:x)$ is an essential ideal of R. It is well-known and easily checked that R is semiprime (has no non-zero nilpotent elements) if and only if $z(R)=0$.

4.1. Lemma. *If R is a (von Neumann) regular ring then $\Delta(R)=0$.*

Proof. Apply (3.5), noting that R_P is a field for each P.

4.2. Theorem. *If R is semiprime and M is a vanishing R-module, then $z(M)=M$.*

Proof. Let Q be the maximal quotient ring of R. By [3, §2.4] Q is a commutative regular ring. Also, by [3, §§4.3, 4.5], Q is injective as an R-module.

Now suppose M is a vanishing R-module, say $\otimes_R^n M=0$. Then $0=(\otimes_R^n M)\otimes_R Q=\otimes_Q^n(M\otimes_R Q)$, and by (4.1), $M\otimes_R Q=0$. Now ${}_RQ\subseteq {}_R\mathrm{Hom}_R(Q,Q)$ via multiplications; hence

$$\mathrm{Hom}_R(M,Q)\subseteq \mathrm{Hom}_R(M,\mathrm{Hom}_R(Q,Q))=\mathrm{Hom}_R(M\otimes_R Q,Q)=0.$$

Since ${}_RQ$ is injective, it follows that $\mathrm{Hom}_R(Rx,Q)=0$ for each $x\in R$. If $(0:x)$ is not essential, choose $0\neq y\in R$ such that $(0:x)\cap Ry=0$. Then $x\to y$ determines a non-zero element of $\mathrm{Hom}_R(Rx,Q)$, a contradition.

4.3. Corollary. *Let I be an ideal contained in the socle of R. Then $\Delta(R/I^2)=\Delta(R)$.*

Proof. By Lemma 3.4, $\Delta(R/I^2)\leqq\Delta(R)$. To prove the reverse inequality, it will suffice to show that $I^2M=0$ for every vanishing R-module M. Let $x\in M$ and let J be the prime radical of R. Since M/JM is a vanishing R/J-module, (4.2) implies that $(JM:x)$ is an essential ideal of R. Therefore $I\subseteq(JM:x)$, and $I^2x\subseteq IJM$. But clearly $IJ=0$, so $I^2x=0$, as desired.

5. A Torsion Theory

By a *torsion class*, we mean a collection of R-modules closed under homomorphic images, arbitrary direct sums, and group extensions [1].

5.1. Theorem. *The vanishing R-modules form a torsion class if and only if $\Delta(R)<\infty$.*

Proof. If $\Delta(R)=\infty$, we have already seen (at the end of §3) that one can find a collection of vanishing modules whose direct sum is not vanishing. Conversely, assume $\Delta(R)=d$.

Claim. If $M_1,\dots,M_{d+1}$ are vanishing modules, then

$$M_1\otimes\cdots\otimes M_{d+1}=0.$$

To see this, set $M=M_1\oplus\cdots\oplus M_{d+1}$. Then M is clearly vanishing, and if $M_1\otimes\cdots\otimes M_{d+1}\neq 0$, we would have $\otimes^{d+1}M\neq 0$, contradicting $\Delta(R)=d$.

It follows from the claim that an arbitrary direct sum of vanishing modules is vanishing. Clearly, a homomorphic image of a vanishing module is vanishing. Finally, let $0\to A\to B\to C\to 0$ be an exact sequence, with A and C vanishing. For each $i,j,k\geqq 0$, let $D_{i,j,k}=(\otimes^i A)\otimes(\otimes^j B)\otimes(\otimes^k C)$. We show by induction on j, that $D_{i,j,k}=0$ whenever $i+j+k=d+1$. If $j=0$, this follows from the claim above. Now if $i+(j+1)+k=d+1$, we tensor the original exact sequence with

$D_{i,j,k}$ to get $D_{i+1,j,k} \to D_{i,j+1,k} \to D_{i,j,k+1} \to 0$. It follows that $D_{i,j+1,k}=0$, and the induction is complete. Setting $j=d+1$, we have $\otimes^{d+1} B=0$, as desired.

Note added in proof. M. Hochster has recently shown us that by interpreting the module $[R; x_1, \ldots, x_n]$ in terms of local cohomology, one can prove that $\Delta(R)=\dim(R)$ for *every* noetherian ring R.

References

1. Dickson, S.E.: A torsion theory for abelian categories. Trans. Amer. Math. Soc. **121**, 223–235 (1966).
2. Kaplansky, I.: Commutative rings. Boston: Allyn and Bacon 1970.
3. Lambek, J.: Lectures on rings and modules. Waltham, Mass.: Blaisdell 1966.
4. Lazard, D.: Autour de la platitude. Bull. Soc. Math. France **97**, 82–128 (1969).
5. Nagata, M.: Local rings. New York: Interscience Publishers 1962.

Prof. Roger A. Wiegand
Dr. Sylvia Wiegand
Department of Mathematics
The University of Nebraska
Lincoln, Nebraska 68508
USA

(Received November 22, 1971)

Math. Z. 129, 359–364 (1972)
© by Springer-Verlag 1972

Some Remarks on the Böhme-Berger Bifurcation Theorem

Floris Takens

1. Introduction and Statement of the Results

This paper originated from an attempt to give a rigorous proof of Berger's bifurcation theorem [1; 10.8]. In the mean time Böhme [2] gave a good proof of this theorem in an even greater generality than I intended to. In this paper we give some related results which, I hope, will give a better understanding of the "topological meaning" of this theorem.

The theorem, mentioned in the title, concerns eigenvalue problems of certain non-linear operators on real Hilbert space which are invariant under the involution "$-$id" and which are of "gradient type". More precisely, it concerns the solutions of the equation

$$L(u)+\phi_1(u)=\lambda(u+\phi_2(u)) \tag{1}$$

depending on $\lambda\in\mathbb{R}$ for u in some real Hilbert space H. In this equation L is supposed to be a self adjoint (linear) operator on H and ϕ_1 and ϕ_2 are supposed to be of *gradient type*, i.e. for certain functions F_1 and F_2 on H we have $\phi_i(u)=(\operatorname{grad} F_i)(u)$. Moreover we assume that (1) is *invariant under* $-id$, i.e. $L(-u)=-L(u)$ and $\phi_i(-u)=-\phi_i(u)$ for $i=1,2$. We also assume that $d\phi_i(0)=0$ for $i=1,2$.

Before we state the results, we want to point out that finding solutions of (1) is equivalent with finding critical points of a certain function depending on λ, namely:

Let F_1 and F_2, as above, be such that $\phi_i(u)=(\operatorname{grad} F_i)(u)$ and let the quadratic function Q be defined by $Q(u)=\frac{1}{2}\langle L(u),u\rangle$. Then it is clear that u_0 is a solution of (1) for $\lambda=\lambda_0$ if and only if u_0 is a critical point of K_{λ_0}: $H\to\mathbb{R}$; K_{λ_0} is defined by

$$K_{\lambda_0}(u)=Q(u)+F_1(u)-\lambda_0\left(\tfrac{1}{2}\|u\|^2+F_2(u)\right). \tag{2}$$

As the results, we want to prove, are essentially about the finite dimensional aspects of (1) (or (2)) we shall assume from now on that H is finite dimensional.

We first restate the theorem of Böhme and Berger for our (finite dimensional) case.

Theorem 1 [2]. *Suppose, in Eq.* (1) *above,* λ_0 *is an eigenvalue of* L *with multiplicity* r *and suppose also that* ϕ_1 *and* ϕ_2 *are at least* C^1 [*it is of course also assumed that* ϕ_1 *and* ϕ_2 *are of gradient type, that* (1) *is invariant under* "$-id$" *and that* $d\phi_i(0)=0$].

Then there is for every $\delta>0$ *an* $\varepsilon>0$ *such that, for each* $c\in(0,\varepsilon)$, *there are at least* $2r$ *points* $p_1,\ldots,p_{2r}\in H$ *with* $\frac{1}{2}\|p_i\|^2+F_2(p_i)=c$ (F_2 *as in* (2)) *and with, for each* p_i, *a real number* λ_i, $|\lambda_i-\lambda_0|<\delta$, *such that* p_i *is a solution of*

$$L(u)+\phi_1(u)=\lambda_i(u+\phi_2(u))$$

[*or* p_i *is a critical point of* K_{λ_i}].

This theorem suggests the following

Problem 1. *Does Theorem* 1 *remain true if we replace the condition* "$\frac{1}{2}\|p_i\|^2+F_2(p_i)=c$" *by* "$R(p_i)=c$ *for some fixed* C^2*-function* $R\colon H\to\mathbb{R}$ *with* $R(0)=0$, $dR(0)=0$ *and* $d^2R(0)$ *positive definite* (*i.e. with*

$$\sum_{i,j=1}^{n}\left(\frac{\partial^2 R}{\partial u_i\,\partial u_j}\right)(0)\,u_i\,u_j$$

a positive definite quadratic form)"?

or even the following

Problem 2. *Is it true that, under the hypothesis of Theorem* 1, *there is an* $\varepsilon>0$ *and at least* $2r$ *pairs* (p_i,λ_i) *with* $\|p_i\|^2=\varepsilon$, p_i *a solution of* (1) *for* $\lambda=\lambda_i$ *and such that each* (p_i,λ_i) *is in the same arcwise connected component of* S *as* $(0,0)$; $S=\{(u_0,\lambda_0)\in H\times\mathbb{R}\,|\,u_0$ *is a solution of* (1) *with* $\lambda=\lambda_0\}$?

We shall show that the answer to these questions is in general no.

Theorem 2[1]. *There is a* C^∞*-map* $\phi_1\colon\mathbb{R}^2\to\mathbb{R}^2$, *gradient of a* C^∞*-function* F_1, *with* $\phi_1(0)=0$, $d\phi_1(0)=0$ *and* $\phi_1(-u)=-\phi_1(u)$, *and there is a* C^∞*-function* $R\colon\mathbb{R}^2\to\mathbb{R}$ *with* $R(0)=0$, $dR(0)=0$ *and* $d^2R(0)$ *positive definite such that, for some sequence* $\{c_i\}_{i=0}^\infty$ *of positive real numbers with* $\lim_{i\to\infty}c_i=0$, *no solution of*

$$\phi_1(u)=\lambda\cdot u \tag{1'}$$

(*for any* $\lambda\in\mathbb{R}$) *lies in* $R^{-1}(c_i)$, $i=1,2,3,\ldots$.

Remark. In Theorem 2 we wrote $\mathbb{R}^2$ instead of H. This was not only done in order to make the explicite construction of ϕ_1, R and $\{c_i\}_{i=0}^\infty$ easier but also because the theorem is wrong in case we replace $\mathbb{R}^2$ by an odd dimensional Hilbert space:

[1] After this paper was submitted, a paper of R. Böhme (Math. Z. **127**, 105–126 (1972)) appeared, containing also a proof of this theorem.

Theorem 3 (see also [2; Satz 1] and [4, Ch. IV, §2, Theorem 2.1]). *Let H be an odd (and finite) dimensional Hilbert space. Let ϕ_1, ϕ_2: $H \times \mathbb{R} \to H$ be C^1 maps with, for each $\lambda \in \mathbb{R}$, $\phi_i(0, \lambda) = 0$ and $d\phi_i(0, \lambda) = 0$ (we do not require this time that ϕ_i is of gradient type or that $\phi_i(-u, \lambda) = -\phi_i(u, \lambda)$). Let R: $H \times \mathbb{R} \to \mathbb{R}$ be a C^2 function with, for each $\lambda \in \mathbb{R}$, $R(0, \lambda) = 0$, $dR(0, \lambda) = 0$ and $d^2 R_\lambda(0)$ positive definite (R_λ is defined by $R_\lambda(u) = R(u, \lambda)$).*

Then there is for each $\delta > 0$ an $\varepsilon > 0$ such that, for each $c \in (0, \varepsilon)$, there is at least one point $(p, \lambda) \in H \times \mathbb{R}$ with $R(p, \lambda) = c$, $|\lambda| < \delta$ and

$$\phi_1(p, \lambda) = \lambda \cdot (u + \phi_2(p, \lambda)). \tag{1''}$$

Remark. Using Theorem 3 one can show that if in Theorem 1 "$\frac{1}{2} \|p_i\|^2 + F_2(p_i) = c$" is replaced by "$R(p_i) = c$ for some fixed C^2-function R: $H \to \mathbb{R}$ with $R(0) = 0$, $dR(0) = 0$ and $d^2 R(0)$ positive definite" and if r (the multiplicity of λ_0) is odd, then there is at least one (instead of $2r$) solution (p, λ) for each $c \in (0, \varepsilon)$; hence under the assumptions of Theorem 3, the answer to Problem 1 is affirmative. I conjecture that under these assumptions also the answer to Problem 2 is affirmative.

2. Proofs of the Theorems

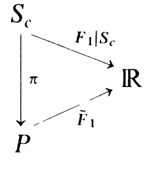
For a complete proof of Theorem 1 we refer to [2]. Here we give a proof for the special case that $r = \dim(H)$ because this proof has a strong relation with the proofs of the Theorems 2 and 3. Our proof is essentially the same as in [2].

2a. *The Proof of Theorem 1 with $r = \dim(H)$*

Without loss of generality we may assume that $\lambda_0 = 0$; this means that $L \equiv 0$. We take $\varepsilon_0 > 0$ so small that for each $c \in (0, \varepsilon_0]$,

$$S_c = \{p \in H \mid \tfrac{1}{2} \|p\|^2 + F_2(p) = c\}$$

is a sphere (of dimension $(r-1)$) and such that the map $S_c \to S^{r-1}$, which assigns to each $u \in S_c$ the unit vector $\dfrac{u}{\|u\|}$, is a diffeomorphism. Because of symmetry ($F_2(-u) = F_2(u)$), S_c is invariant under the involution "$-\mathrm{id}$". We now consider, for arbitrary $c \in (0, \varepsilon_0]$, the function $F_1 | S_c$ (F_1 was the function with $\phi_1(u) = (\operatorname{grad} F_1)(u)$). Also F_1 is invariant under "$-\mathrm{id}$", so we can make the following commutative diagram



where π is the projection map which identifies $u \in S_c$ with $-u \in S_c$; $\tilde{F}_1$ is determined by the requirement that the diagram commutes. Note that P is diffeomorphic with the $(r-1)$-dimensional projective space; from this it follows, using Lusternik-Schnirelman theory [5], that $\tilde{F}_1$ has at least r critical points on P. Hence $F_1 | S_c$ has at least $2r$ critical points (π is a double covering). Let $p \in S_c$ be such a critical point of $F_1 | S_c$; we then have, for some μ, $dF_1(p) = \mu \cdot dF_3(p)$, where $F_3(p)$ stands for $\frac{1}{2}\|p\|^2 + F_2(p)$. Hence it follows that p is a critical point of $K_\mu = F_1 - \mu F_3$.

Finally we have to show that if we take c small, then the corresponding values μ for $p \in S_c$, and p a critical point of $F_1 | S_c$, also become small. This is proved as follows:

For some constant $A > 0$ we have that for each $p \in S_c$, $c \in (0, \varepsilon_0]$, $\|dF_3(p)\| \geqq A \cdot \sqrt{c}$. On the other hand, because $dF_1(0) = 0$ and $d^2 F_1(0) = 0$, there is some continuous function α: $\mathbb{R}_+ \to \mathbb{R}_+$, with $\alpha(0) = 0$ ($\mathbb{R}_+ = \{t \in \mathbb{R} | t \geqq 0\}$), such that, for each $p \in S_c$, $c \in (0, \varepsilon_0]$, $\|dF_1(p)\| \leqq \sqrt{c} \cdot \alpha(\sqrt{c})$.

From this it follows that if $p \in S_c$ is a critical point of $F_1 | S_c$, then the corresponding μ satisfies

$$|\mu| \leqq \frac{\sqrt{c} \cdot \alpha(\sqrt{c})}{A \cdot \sqrt{c}} = A^{-1} \cdot \alpha(\sqrt{c});$$

the righthand side of this expression goes to zero for $c \to 0$. This completes the proof.

2b. The Proof of Theorem 2

We consider the annulus $A = \{(x_1, x_2) \in \mathbb{R} | x_1^2 + x_2^2 \in [1, 4]\}$. $C \subset A$ is the curve given by $\frac{1}{2}x_1^2 + \frac{1}{3}x_2^2 = 1$. First we want to construct a C^∞ function f: $\mathbb{R}^2 \to \mathbb{R}$ with

(i) $f(x) = f(-x)$,

(ii) $\sup(f) \subset A$,

(iii) for each $c \in C$, $\left(x_1 \frac{\partial f}{\partial x_2} - x_2 \frac{\partial f}{\partial x_1}\right)(c) = 1$.

To construct such f we first define a smooth function f on a small closed neighbourhood U of the four points $(0, \pm\sqrt{3})$ and $(\pm\sqrt{2}, 0)$ in $\mathbb{R}^2$ such that f_1 satisfies (i) and (iii) (one has to take U so that $x \in U$ implies $-x \in U$). If we take the four components of U (one for each of the four points) small enough, such f_1 can clearly be constructed. Next we extend f_1 differentiably over all of $C \cup U$ such that (i) still holds. In each point $c \in C \setminus U$ we then have the following situation: $x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1}$ is not tangent to C; the directional derivative of f_1

along C is given (because f_1 is defined on C) and by (iii) the derivative of f_1 (or rather of an extension of f_1 over $\mathbb{R}^2$ satisfying (iii)) in the direction $x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1}$ is given. This means that the 1-jet of any extension f_2 of $f_1 | C$, satisfying (iii), in each point of $c \in C$ is determined and depends C^∞ on c. Let f_2: $\mathbb{R}^2 \to \mathbb{R}$ be such that $f_2 | C = f_1 | C$ and such that in each point $c \in C$ the 1-jet of f_2 is such that (iii) holds. It is clear that f_2 and $f_2 \cdot (-\mathrm{id})$ have, in each point $c \in C$, the same 1-jet (this is because f_1 satisfied (i)). Hence f_3: $\mathbb{R}^2 \to \mathbb{R}$, defined by $f_3(x) = \frac{1}{2} f_2(x) + \frac{1}{2} f_2(-x)$, satisfies (i) and (iii). Finally we take $f(x) = \lambda(\|x\|) \cdot f_3(x)$ where λ is some C^∞-function with $\lambda([\sqrt{2}, \sqrt{3}]) = 1$ and $\lambda(\mathbb{R} \setminus [1, 2]) = 0$.

Using the above function f we write $F_1(x_1, x_2) = \sum_{i=0}^{\infty} \alpha_i f(2^i x_1, 2^i x_2)$, where the sequence $\{\alpha_i\}_{i=0}^\infty$, $\alpha_i \in \mathbb{R}$, $\alpha_i > 0$ is so that the C^i norm of $\alpha_i \cdot f(2^i x_1, 2^i x_2)$ is smaller than 2^{-i}. From this condition it is clear that F_1 is well defined and C^∞.

For R we take $R(x_1, x_2) = \frac{1}{2} x_1^2 + \frac{1}{3} x_2^2$. We now want to show that the sequence $c_i = 2^{-2i}$, $i = 0, 1, \ldots$, satisfies the conditions in the conclusion of Theorem 2 with $\phi_1(x) = (\operatorname{grad} F_1)(x)$.

Let $x = (x_1, x_2)$ be some point of $R^{-1}(c_i)$ for some i. Then, by definition, $2^i x = (2^i x_1, 2^i x_2)$ is a point of $C = R^{-1}(1)$. From the definition of F_1 it then follows that $\left(x_1 \frac{\partial F_1}{\partial x_2} - x_2 \frac{\partial F_1}{\partial x_1}\right)(c) \neq 0$. Hence, for no $\lambda \in \mathbb{R}$, x is a critical point of K_λ (defined by $K_\lambda(x) = F_1(x) - \frac{\lambda}{2} \|x\|^2$). Hence there is no λ for which x is a solution of $\phi_1(u) = \lambda \cdot u$; this proves the theorem.

2c. The Proof of Theorem 3

We first fix a $\delta > 0$. Then we take an $\varepsilon_1 > 0$ such that for each $c \in (0, \varepsilon_1)$, $R^{-1}(c) \cap \{(u, \lambda) \in H \times \mathbb{R} \,|\, |\lambda| \leqq \delta\}$ is diffeomorphic with $S^{n-1} \times \mathbb{R}$; $n = \dim(H)$. Next we take an $\varepsilon_2 \in (0, \varepsilon_1)$ such that, for each $c \in (0, \varepsilon_2)$, the map Ψ_c: $\{(u, \lambda) \in H \times \mathbb{R} \,|\, R(u, \lambda) = c$ and $|\lambda| \leqq \delta\} \to S^{n-1} \times [-\delta, +\delta]$, defined by $\Psi_c(u, \lambda) = \left(\frac{u}{\|u\|}, \lambda\right)$, is a diffeomorphism. Finally we take $\varepsilon \in (0, \varepsilon_2)$ such that in each point $(u, \lambda) \in R^{-1}(0, \varepsilon]$ with $\lambda = +\delta$, resp. $\lambda = -\delta$, $\langle [\lambda \cdot u + \lambda \cdot \phi_2(u, \lambda) - \phi_1(u, \lambda)], u \rangle > 0$, resp. < 0. Such ε exists because for each λ, $\phi_i(0, \lambda) = 0$ and $d\phi_i(0, \lambda) = 0$.

To prove the theorem we have to show that, for any $c \in (0, \varepsilon)$, $\{(u, \lambda) \,|\, R(u, \lambda) = c,\ |\lambda| \leqq \delta\}$ contains a solution of (1″). Suppose that, for some fixed $c \in (0, \varepsilon)$, $\{(u, \lambda) \,|\, R(u, \lambda) = c$ and $|\lambda| \leqq \delta\}$ does not contain a solution of (1″); we shall derive a contradiction from this assumption.

Consider the map $\Phi\Psi_c^{-1}$: $S^{n-1}\times[-\delta,+\delta]\to S^{n-1}$ where Ψ_c is as above and

$$\Phi(u,\lambda)=\frac{\lambda\cdot u+\lambda\cdot\phi_2(u,\lambda)-\phi_1(u,\lambda)}{\|\lambda\cdot u+\lambda\cdot\phi_2(u,\lambda)-\phi_1(u,\lambda)\|}\cdot\Phi\,|\,\mathrm{Im}(\Psi_c^{-1})$$

is defined if and only if $\mathrm{Im}(\Psi_c^{-1})=\{(u,\lambda)\,|\,R(u,\lambda)=c$ and $|\lambda|\leqq\delta\}$ does not contain any solution of (1″).

Because in each point of $R^{-1}(c)$ with $\lambda=+\delta$, resp. $\lambda=-\delta$, we have $\langle[\lambda\cdot u+\lambda\cdot\phi_2(u,\lambda)-\phi_1(u,\lambda)],u\rangle>0$, resp. <0,

$$\Phi\Psi_c^{-1}\,|\,S^{n-1}\times\{-\delta\}:\ S^{n-1}\times\{-\delta\}\to S^{n-1}$$

is homotopic with the "antipodal map" and

$$\Phi\Psi_c^{-1}\,|\,S^{n-1}\times\{+\delta\}:\ S^{n-1}\times\{+\delta\}\to S^{n-1}$$

is homotopic with the "identity". As in our case $n=\dim(H)$ is odd and hence $n-1=\dim(S^{n-1})$ is even, the identity and the antipodal map of S^{n-1} to itself are not homotopic [3]. On the other hand, $\Phi\Psi_c^{-1}$ realizes a homotopy between $\Phi\Psi_c^{-1}|S^{n-1}\times\{-\delta\}$ and $\Phi\Psi_c^{-1}|S^{n-1}\times\{+\delta\}$; this is the required contradiction.

References

1. Berger, M.S.: A bifurcation theory for nonlinear elliptic partial differential equations and related systems, in: Bifurcation theory and nonlinear eigenvalue problems, ed. by J.B. Keller and S. Antman. New York: Benjamin 1969.
2. Böhme, R.: Nichtlineare Störung der isolierten Eigenwerte selbstadjungierter Operatoren. Math. Z. **123**, 61-92 (1971).
3. tom Dieck, T., Kamps, K.H., Puppe, D.: Homotopietheorie. Lecture Notes in Mathematics **157**, Berlin-Heidelberg-New York: Springer 1970.
4. Krasnosel'skii: Topological methos in the theory of nonlinear integral equations. Oxford: Pergamon Press 1964.
5. Lusternik, L., Schnirelman, L.: Méthodes topologiques dans les problèmes variationnels. Paris: Hermann 1934.

Dr. Floris Takens
Mathematisch Instituut
Postbus 800
Groningen
The Netherlands

(*Received March 24, 1972*)

Math. Z. 129, 365–379 (1972)
© by Springer-Verlag 1972

On the Second Cohomology Groups of Semidirect Products

Ken-ichi Tahara

Introduction

Let G be a finite group, then $H^3(G, \mathbb{Z})$ is canonically isomorphic to $H^2(G, \mathbb{Q}/\mathbb{Z})$, which is called the Schur multiplicator of the group G. In general we study $H^2(G, A)$, where A is a G-module with trivial G-action. The second cohomology group $H^2(G, A)$ is the additive group of 2-cocycles taken modulo the subgroup of 2-coboundaries, where a map $f\colon G\times G \to A$ is a 2-cocycle on G if and only if f satisfies the following:

$$f(\sigma,\tau)-f(\rho\sigma,\tau)+f(\rho,\sigma\tau)-f(\rho,\sigma)=0 \qquad (\rho,\sigma,\tau\in G),$$

and f is a 2-coboundary on G if and only if there exists a map $g\colon G\to A$ such that

$$f(\sigma,\tau)=g(\tau)-g(\sigma\tau)+g(\sigma) \qquad (\sigma,\tau\in G).$$

Lyndon [5] proved that if a finite group G is the direct product of its normal subgroups N and T, then the cohomology group $H^n(G, A)$ of G with a coefficient module A with trivial G-action is related to the compound cohomology groups $H^k(T, H^{n-k}(N, A))$ $(0\leqq k\leqq n)$.

In this paper we show that, when G is the semidirect product of a normal subgroup N and a subgroup T, then

$$H^2(G, A)\cong H^2(T, A)\oplus \tilde{H}^2(G, A),$$

and we have the following canonical exact sequence

$$0\to H^1(T, H^1(N, A))\to \tilde{H}^2(G, A)\xrightarrow{\text{res}} H^2(N, A)^T$$
$$\xrightarrow{d_2} H^2(T, H^1(N, A))\to \tilde{H}^3(G, A),$$

where $\tilde{H}^i(G, A)$ is the kernel of the restriction homomorphism $H^i(G, A)\to H^i(T, A)$. When both N and T are finite cyclic subgroups, the homomorphism d_2, and $\tilde{H}^2(G, A)$ hence $H^2(G, A)$ can be determined completely for an arbitrary G-module A with trivial G-action. As its application, the Schur multiplicator of a semidirect product of cyclic subgroups is concretely computed.

The author would like to express his appreciation to Professor Y. Akagawa who pointed out to him an error of the first draft of this paper, and also to Professor T. Oda for his helpful suggestions and encouragement during the preparation of this paper and for his critical reading of the manuscript.

§ 1. Theorems

Let G be a finite group, and let A be a G-module with trivial G-action. Then it was shown by Eilenberg and MacLane [3] that the cohomology group $H^n(G, A)$ is unaffected even if we restrict ourselves only to those n-cochains f which satisfy $f(g_1, g_2, \dots, g_n)=0$ whenever one of the arguments g_i is 1 (= the identity element of G). Since this condition is very useful in computation, we henceforth assume that all cochains satisfy the condition.

If $G=N\cdot T$ is the semidirect product of a normal subgroup N and a subgroup T, then every element g of G is uniquely represented in the form $g=nt$ ($n\in N$, $t\in T$), and for any $n\in N$, $t\in T$,

$$tn={}^tnt \qquad \text{with } {}^tn=tnt^{-1}.$$

Proposition 1. *Let G be the semidirect product of a normal subgroup N and a subgroup T, and let A be a G-module with trivial G-action.*

(I) *Let a map $f\colon G\times G\to A$ be a 2-cocycle on G. Then f can always be normalized up to coboundaries as follows:*

$$(*) \qquad f(N, T)=0,$$

and hence

(a) $f(nt, n't')=f(t,t')+f(t,n')+f(n', {}^tn')$ $\quad (n, n'\in N,\ t, t'\in T)$.

We call such a 2-cocycle f a normal 2-cocycle. Thus a normal 2-cocycle f on G is determined uniquely by $f|_{N\times N}$, $f|_{T\times T}$ and $f|_{T\times N}$.

(II) *The data $f|_{N\times N}$, $f|_{T\times T}$ and $f|_{T\times N}$ determine a normal 2-cocycle on G if and only if they satisfy the following:*

(b) *f is a 2-cocycle on N,*

(c) *f is a 2-cocycle on T,*

(d) $f(tt', n)=f(t', n)+f(t, {}^{t'}n)$ $\quad (n\in N,\ t, t'\in T)$,

(e) $f(n, n')-f({}^tn, {}^tn')=f(t, n)-f(t, nn')+f(t, n')$ $\quad (n, n'\in N,\ t\in T)$.

Proof. For the discussion in this proof, it is more convenient to treat the law of composition of A as multiplication. There is a one-to-one correspondence between the elements of $H^2(G, A)$ and the classes of equivalent central extensions

$$1\to A\to \Pi\to G\to 1,$$

and giving a 2-cocycle f representing a cohomology class corresponds to giving a section $\pi: G \to \Pi$, via the relation

$$f(g, g')\,\pi(gg') = \pi(g)\,\pi(g') \qquad (g, g' \in G).$$

It is clear that a 2-cochain f satisfies $(*)$ if and only if the corresponding section satisfies

$$(**) \qquad \pi(nt) = \pi(n)\,\pi(t) \qquad (n \in N, t \in T).$$

Hence the normalization $(*)$ of 2-cocycles on G is obvious, and since

$$\pi(nt)\,\pi(n't') = f(nt, n't')\,\pi(ntn't') = f(nt, n't')\,\pi(n\,{}^t n' tt') \quad (n, n' \in N, t, t' \in T),$$

and the left hand side is equal to

$$\begin{aligned}\pi(n)\,\pi(t)\,\pi(n')\,\pi(t') &= \pi(n)\,f(t, n')\,\pi({}^t n')\,\pi(t)\,\pi(t') \\ &= f(t, n')\,f(n, {}^t n')\,f(t, t')\,\pi(n\,{}^t n' tt'),\end{aligned}$$

we have (a). To prove (II), suppose f satisfies $(*)$, and let π be the corresponding section satisfying $(**)$. Since

$$\pi(t)\,\pi(n) = f(t, n)\,\pi(tn) = f(t, n)\,\pi({}^t n t) = f(t, n)\,\pi({}^t n)\,\pi(t) \qquad (n \in N, t \in T),$$

it follows that

$$\pi(t)\,\pi(n)\,\pi(t)^{-1} = f(t, n)\,\pi({}^t n) \qquad (n \in N, t \in T).$$

By replacing t by tt' or n by nn' in the above relation, we easily get (d) and (e), respectively.

Conversely it is easy to show that a map $f: G \times G \to A$ satisfying the relations (a)$\sim$(e) is a 2-cocycle on G. Q.E.D.

Let $\tilde{H}^i(G, A)$ be the kernel of the restriction homomorphism

$$\text{res}: H^i(G, A) \to H^i(T, A).$$

Then by the splitting of the exact sequence

$$\begin{array}{ccccccc} 1 \to N \to & G & \to & G/N & \to 1, \\ & \uparrow & \nearrow & & \\ & T & & & \end{array}$$

we see that

$$H^i(G, A) \cong H^i(T, A) \oplus \tilde{H}^i(G, A) \qquad (i \geqq 1),$$

canonically.

Theorem 2. *Let G be the semidirect product of a normal subgroup N and a subgroup T, and let A be a G-module with trivial G-action. Then*

(I) $H^2(G, A) \cong H^2(T, A) \oplus \tilde{H}^2(G, A)$ *canonically.*

(II) *We have a canonical exact sequence*

$$0 \to H^1(T, H^1(N, A)) \to \tilde{H}^2(G, A) \xrightarrow{\text{res}} H^2(N, A)^T \xrightarrow{d_2} H^2(T, H^1(N, A)) \to \tilde{H}^3(G, A),$$

where T acts on $H^1(N, A) = \mathrm{Hom}(N, A)$ and $H^2(N, A)$ via the canonical action induced by conjugation of T on N, and $H^2(N, A)^T$ is the subgroup of T-invarients. The homomorphism res: $\tilde{H}^2(G, A) \to H^2(N, A)^T$ *is induced by the restriction map* res: $H^2(G, A) \to H^2(N, A)$.

Proof. Let $Z_*^2 = Z_*^2(G, A)$ and $B_*^2 = B_*^2(G, A)$ be the groups of normal 2-cocycles and 2-coboundaries, respectively. By Proposition 1, we have $H^2(G, A) = Z_*^2/B_*^2$. Let $\tilde{Z}_*^2$ be the subgroup of elements f of Z_*^2 for which $f(T, T) = 0$. Put $\tilde{B}_*^2 = \tilde{Z}_*^2 \cap B_*^2$. Then by Proposition 1, it is easy to see that $\tilde{H}^2(G, A) \cong \tilde{Z}_*^2/\tilde{B}_*^2$, because f is an element of $\tilde{B}_*^2$ if and only if there exists a map $v: N \to A$ with $v(1) = 0$ such that

$$f(t, n) = v(n) - v({}^t n) \qquad (n \in N,\ t \in T),$$

and

$$f(n, n') = v(n) - v(nn') + v(n') \qquad (n, n' \in N).$$

By Proposition 1, the homomorphism sending f to $(f|_{T \times N}, f|_{N \times N})$ is an injection

$$\tilde{Z}_*^2 \to Z^1(T, C^1(N, A)) \oplus Z^2(N, A).$$

Its image consists of elements (u, h) satisfying

$$h(n, n') - h({}^t n, {}^t n') = u(t, n) - u(t, nn') + u(t, n') \qquad (n, n' \in N,\ t \in T). \quad (1)$$

We identify $\tilde{Z}_*^2$ with its image in the right hand side. The subgroup consisting of elements $(u, 0)$ with $u \in Z^1(T, H^1(N, A))$ is obviously contained in $\tilde{Z}_*^2$. We identify $Z^1(T, H^1(N, A))$ with its image in

$$Z^1(T, C^1(N, A)) \oplus Z^2(N, A).$$

It is straightforward to see that

$$\tilde{B}_*^2 \cap Z^1(T, H^1(N, A)) = B^1(T, H^1(N, A)).$$

Thus we have a canonical injection

$$H^1(T, H^1(N, A)) \to \tilde{H}^2(G, A).$$

Consider the homomorphism $p: \tilde{Z}_*^2 \to Z^2(N, A)$ sending (u, h) to h.

Lemma 3.
$$p^{-1}(B^2(N, A)) = Z^1(T, H^1(N, A)) + \tilde{B}_*^2.$$

Proof. Let (u, h) be an element of $\tilde{Z}_*^2$ with $h \in B^2(N, A)$. Then there exists a map $v: N \to A$ with $v(1) = 0$ such that

$$h(n, n') = v(n) - v(nn') + v(n') \qquad (n, n' \in N).$$

By (1), we get

$$\{u(t, n)-v(n)+v({}^t n)\}+\{u(t, n')-v(n')+v({}^t n')\}$$
$$=\{u(t, nn')-v(nn')+v({}^t(nn'))\} \qquad (n, n'\in N, t\in T).$$

Thus $u'(t, n)=u(t, n)-v(n)+v({}^t n)$ is in $Z^1(T, H^1(N, A))$ and $(u', 0)$ is congruent to (u, h) modulo $\tilde{B}^2_*$. Q.E.D.

Proof of Theorem 2 (*continued*). So far we obtained a canonical exact sequence

$$0\to H^1(T, H^1(N, A))\to \tilde{H}^2(G, A)\xrightarrow{p} H^2(N, A).$$

Moreover p is obviously induced by the restriction map res: $H^2(G, A)\to H^2(N, A)$. By (1), the image of p is contained in $H^2(N, A)^T$.

Next we define a canonical homomorphism

$$d_2\colon H^2(N, A)^T\to H^2(T, H^1(N, A)).$$

Let $h\colon N\times N\to A$ be a 2-cocycle on N representing a cohomology class of $H^2(N, A)^T$. Then there is a map $u\colon T\times N\to A$ with $u(1, n)=u(t, 1)=0$ $(n\in N, t\in T)$ such that

$$h(n, n')-h({}^t n, {}^t n')=u(t, n)-u(t, nn')+u(t, n') \qquad (n, n'\in N, t\in T).$$

Consider

$$(d_2 h)(t, t', n)=u(t', n)-u(tt', n)+u(t, {}^{t'}n).$$

$d_2 h$ is easily seen to be contained in $Z^2(T, H^1(N, A))$. We now show that this d_2 gives rise to the sought-for homomorphism. Suppose $h'\colon N\times N\to A$ is another 2-cocycle cohomologous to h. Then there exists a map $v\colon N\to A$ with $v(1)=0$ such that

$$h'(n, n')=h(n, n')+v(n)-v(nn')+v(n') \qquad (n, n'\in N).$$

Let $u'\colon T\times N\to A$ be a map with $u'(1, n)=u'(t, 1)=0$ $(n\in N, t\in T)$ such that

$$h'(n, n')-h'({}^t n, {}^t n')=u'(t, n)-u'(t, nn')+u'(t, n') \qquad (n, n'\in N, t\in T).$$

Then $w(t, n)=u'(t, n)-u(t, n)-v(n)+v({}^t n)$ is easily seen to be contained in $C^1(T, H^1(N, A))$. It is straightforward to see that

$$(d_2 h')(t, t', n)=(d_2 h)(t, t', n)+w(t', n)-w(tt', n)+w(t, {}^{t'}n) \quad (n\in N, t, t'\in T).$$

Thus $d_2\colon H^2(N, A)^T\to H^2(T, H^1(N, A))$ is well defined.

To show the exactness of the sequence, suppose $d_2 h$ is cohomologous to 0. Then there exists $w\in C^1(T, H^1(N, A))$ with

$$(d_2 h)(t, t', n)=w(t', n)-w(tt', n)+w(t, {}^{t'}n) \qquad (n\in N, t, t'\in T).$$

Then

$$u(t,n)=w(t,n)+z(t,n) \qquad (n\in N,\ t\in T),$$

with $z(t',n)-z(tt',n)+z(t,{}^{t'}n)=0$ $(n\in N,\ t,t'\in T)$, and

$$h(n,n')-h({}^tn,{}^tn')=z(t,n)-z(t,nn')+z(t,n') \qquad (n,n'\in N,\ t\in T).$$

Thus (z,h) is in $\tilde{Z}^2_*$.

Finally we consider a canonical homomorphism $H^2(T,H^1(N,A))\to\tilde{H}^3(G,A)$. Let $f\colon T\times T\to H^1(N,A)$ be a 2-cocycle on T representing a cohomology class of $H^2(T,H^1(N,A))$. Then f is considered as a map from $T\times T\times N$ to A, and it is easy to see that f is a 3-cocycle on G with

$$f(nt,n't',n''t'')=f(t,t',n'') \qquad (n,n',n''\in N,\ t,t',t''\in T).$$

Suppose f is cohomologous to zero. Then there exists a map $v\colon T\times N\to A$ with $v(1,n)=v(t,1)=0$ $(n\in N,\ t\in T)$ and

$$v(t,nn')=v(t,n)+v(t,n') \qquad (t\in T,\ n,n'\in N)$$

such that

$$f(t,t',n'')=v(t',n'')-v(tt',n'')+v(t,{}^{t'}n'') \qquad (t,t'\in T,\ n''\in N).$$

Putting $v(nt,n't')=v(t,n')$ $(n,n'\in N,\ t,t'\in T)$, we have easily

$$f(g,g',g'')=v(g',g'')-v(gg',g'')+v(g,g'g'')-v(g,g') \qquad (g,g',g''\in G).$$

Thus the canonical homomorphism $H^2(T,H^1(N,A))\to\tilde{H}^3(G,A)$ is well defined.

To show the last exactness of the sequence, let $h\colon N\times N\to A$ be a 2-cocycle on N representing a cohomology class of $H^2(N,A)^T$. Then there is a map $u\colon T\times N\to A$ with $u(1,n)=u(t,1)=0$ $(n\in N,\ t\in T)$ such that

$$h(n,n')-h({}^tn,{}^tn')=u(t,n)-u(t,nn')+u(t,n') \qquad (n,n'\in N,\ t\in T).$$

Putting $v(nt,n't')=u(t,n')+h(n,{}^tn')$ $(n,n'\in N,\ t,t'\in T)$, we have easily

$$(d_2h)(g,g',g'')=v(g',g'')-v(gg',g'')+v(g,g'g'')-v(g,g') \qquad (g,g',g''\in G).$$

Thus its image in $\tilde{H}^3(G,A)$ is zero. Conversely let $f\colon T\times T\to H^1(N,A)$ be a 2-cocycle on T representing a cohomology class of $H^2(T,H^1(N,A))$, and let its image in $\tilde{H}^3(G,A)$ be zero. Then there exists a map $v\colon G\times G\to A$ with $v(1,g)=v(g,1)$ $(g\in G)$ such that

$$f(g,g',g'')=v(g',g'')-v(gg',g'')+v(g,g'g'')-v(g,g') \qquad (g,g',g''\in G).$$

Putting $u(t,n)=v(t,n)-v({}^tn,t)$ $(t\in T,\ n\in N)$ and

$$h(n,n')=v(n,n') \qquad (n,n'\in N),$$

we have easily

$$h(n', n'') - h(nn', n'') + h(n, n'n'') - h(n, n') = 0 \qquad (n, n', n'' \in N),$$

$$h(n, n') - h({}^t n, {}^t n') = u(t, n') - u(t, nn') + u(t, n) \qquad (n, n' \in N, t \in T),$$

and

$$f(t, t', n'') = u(t, n'') - u(tt', n'') + u(t, {}^{t'}n'') \qquad (t, t' \in T, n'' \in N).$$

Thus h determine an element of $H^2(N, A)^T$, and the cohomology class of $d_2 h$ in $H^2(T, H^1(N, A))$ is equal to that of f. Q.E.D.

Remark 1. In general there is the Hochschild-Serre spectral sequence

$$E_2^{p,q} = H^p(G/N, H^q(N, A)) \Rightarrow H^{p+q}(G, A).$$

When $G \to G/N$ splits, $E_2^{p,q}$ with $q>0$ converges to $\tilde{H}^{p+q}(G, A)$. The exact sequence we obtained in Theorem 2 is nothing but the exact sequence of terms of low degree in this latter spectral sequence. Our proof gives us concrete descriptions of $\tilde{H}^2(G, A)$ hence $H^2(G, A)$, and d_2 (cf. Theorem 7, 8 below).

We denote by $H^2(N, A)^*$ the image of res: $\tilde{H}^2(G, A) \to H^2(N, A)^T$. By Theorem 2, $H^2(N, A)^* = \operatorname{Ker} d_2$.

Corollary. *If $G = N \times T$ is the direct product of N and T, then canonically*

$$H^2(G, A) \cong H^2(T, A) \oplus H^1(T, H^1(N, A)) \oplus H^2(N, A).$$

Proof. T acts trivially on N, hence $H^2(N, A)^T = H^2(N, A)$. Thus we have an exact sequence by Theorem 2,

$$0 \to H^2(T, A) \oplus H^1(T, H^1(N, A)) \to H^2(G, A) \to H^2(N, A).$$

By the splitting of the exact sequence

$$\begin{array}{ccccccccc} 1 & \to & T & \to & G & \to & G/T & \to & 1 \\ & & & & \uparrow & \nearrow & & & \\ & & & & N & & & & \end{array}$$

we have

$$H^2(G, A) \cong H^2(T, A) \oplus H^1(T, H^1(N, A)) \oplus H^2(N, A). \qquad \text{Q.E.D.}$$

Remark 2 (Akagawa). In general $H^2(N, A)^*$ is not equal to $H^2(N, A)^T$. Here is an example. Let G be the group generated by elements n, t with defining relations:

$$n^9 = t^3 = 1, \qquad tnt^{-1} = n^4.$$

Then G is the semidirect product of the normal subgroup $N = \{n\}$ and the subgroup $T = \{t\}$. Let $A = \mathbb{Z}/3\mathbb{Z}$ be the G-module with trivial G-action.

Then easily

$$H^2(T, A) = \mathbb{Z}/3\mathbb{Z}, \quad H^1(T, H^1(N, A)) = \mathbb{Z}/3\mathbb{Z},$$

and

$$H^2(N, A)^T = \mathbb{Z}/3\mathbb{Z}, \quad H^2(T, H^1(N, A)) = \mathbb{Z}/3\mathbb{Z}.$$

By Theorem 8 below, we have $d_2 = \text{identity}$ and hence $H^2(N, A)^* = 0$ and

$$H^2(G, A) \cong H^2(T, A) \oplus H^1(T, H^1(N, A)) = \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}.$$

Remark 3. In general $H^2(G, A)$ is not equal to

$$H^2(T, A) \oplus H^1(T, H^1(N, A)) \oplus H^2(N, A)^*,$$

namely the canonical exact sequence

$$0 \to H^1(T, H^1(N, A)) \to \tilde{H}^2(G, A) \to H^2(N, A)^* \to 0$$

is not split. Here is an example. Let G be the group generated by elements n, t with defining relations:

$$n^9 = t^9 = 1, \quad tnt^{-1} = n^7.$$

Then G is the semidirect product of the normal subgroup $N = \{n\}$ and the subgroup $T = \{t\}$. Let $A = \mathbb{Z}/27\mathbb{Z}$ be the G-module with trivial G-action. Then easily

$$H^2(T, A) = \mathbb{Z}/3\mathbb{Z}, \quad H^1(T, H^1(N, A)) = \mathbb{Z}/3\mathbb{Z},$$

and

$$H^2(N, A)^T = \mathbb{Z}/3\mathbb{Z}, \quad H^2(T, H^1(N, A)) = \mathbb{Z}/3\mathbb{Z}.$$

By Theorem 8 below, we have $d_2 = 3 \cdot \text{identity} = 0$, and hence

$$H^2(N, A)^* = H^2(N, A)^T = \mathbb{Z}/3\mathbb{Z}.$$

On the other hand we have by Theorem 7 below,

$$\tilde{H}^2(G, A) = \mathbb{Z}/9\mathbb{Z} \neq H^1(T, H^1(N, A)) \oplus H^2(N, A)^* = \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}.$$

We now determine completely $\tilde{H}^2(G, A)$ hence $H^2(G, A)$, and identify the homomorphism d_2, when N and T are both finite cyclic groups and G is their semidirect product.

The following is well known, and may be proved in the same way as in Proposition 1.

Lemma 4. *Let S be a cyclic group of order s generated by an element σ, and let A be a left S-module. Then a 2-cocycle $f\colon S \times S \to A$ can always be normalized as*

$$f(\sigma^i, \sigma^j) = \left(\left[\frac{i+j}{s} \right] - \left[\frac{i}{s} \right] - \left[\frac{j}{s} \right] \right) f(\sigma, \sigma^{-1}) \quad (i, j \in \mathbb{Z}),$$

where [ ] *is the Gauss symbol. Moreover the homomorphism sending* f *to* $f(\sigma,\sigma^{-1})\left(\mathrm{mod}\left(\sum_{l=0}^{s-1}\sigma^l\right)A\right)$ *gives a canonical isomorphism*

$$H^2(S,A)\cong A^S\Big/\left(\sum_{l=0}^{s-1}\sigma^l\right)A.$$

Lemma 5. *Let* $G=N\cdot T$ *be the semidirect product of a normal subgroup* N *and a finite cyclic subgroup* $T=\{t\}\cong\mathbb{Z}/m\mathbb{Z}$, *and let* A *be a G-module with trivial G-action. Let a map* $f\colon N\times N\to A$ *be a 2-cocycle on* N *representing an element of* $H^2(N,A)^T$. *Then there exists a map* $v\colon T\times N\to A$ *with* $v(1,n)=v(t',1)=0$ $(n\in N,\ t'\in T)$ *which satisfies the following:*

$$\text{(i)}\qquad v(t^i,n)=\sum_{l=0}^{i-1}v(t,{}^{t^l}n)\qquad(0\leqq i<m,\ n\in N),$$

and

$$\text{(ii)}\quad f(n,n')-f({}^{t'}n,{}^{t'}n')=v(t',n)-v(t',nn')+v(t',n')\quad(n,n'\in N,\ t'\in T).$$

Moreover $d_2\colon H^2(N,A)^T\to H^2(T,H^1(N,A))$ *sends the cohomology class of* f *to that of a 2-cocycle on* T *with coefficients in* $H^1(N,A)$:

$$\left(\left[\frac{i+j}{m}\right]-\left[\frac{i}{m}\right]-\left[\frac{j}{m}\right]\right)\sum_{l=0}^{m-1}v(t,{}^{t^l}n)\qquad(i,j\in\mathbb{Z},\ n\in N).$$

Proof. Let a map $f\colon N\times N\to A$ be a 2-cocycle representing an element of $H^2(N,A)^T$, then there exists a map $u\colon T\times N\to A$ with $u(1,n)=u(t',1)=0$ $(n\in N,\ t'\in T)$ such that

$$f(n,n')-f({}^{t'}n,{}^{t'}n')=u(t',n)-u(t',nn')+u(t',n')\qquad(n,n'\in N,\ t'\in T).$$

Consider

$$(d_2f)(t',t'',n)=u(t'',n)-u(t't'',n)+u(t',{}^{t''}n)\qquad(t',t''\in T,\ n\in N),$$

which represents an element of $H^2(T,H^1(N,A))$. Since T is cyclic, there exists a map $w\colon T\times N\to A$ with $w(1,n)=w(t',1)=0$ $(n\in N,\ t'\in T)$ and $w(t',nn')=w(t',n)+w(t',n')$ $(t'\in T,\ n,n'\in N)$, which normalizes d_2f as in Lemma 4, i.e.

$$\begin{aligned}&(d_2f)(t^i,t^j,n)+\{w(t^j,n)-w(t^{i+j},n)+w(t^i,{}^{t^j}n)\}\\&\quad=\{u(t^j,n)-u(t^{i+j},n)+u(t^i,{}^{t^j}n)\}+\{w(t^j,n)-w(t^{i+j},n)+w(t^i,{}^{t^j}n)\}\\&\quad=\left(\left[\frac{i+j}{m}\right]-\left[\frac{i}{m}\right]-\left[\frac{j}{m}\right]\right)\{(d_2f)(t,t^{-1},n)+w(t^{-1},n)+w(t,{}^{t^{-1}}n)\}\\&\hspace{28em}(i,j\in\mathbb{Z}).\end{aligned}\tag{2}$$

Then the map $v=u+w\colon T\times N\to A$ satisfies the relations (i) and (ii). The rest follows immediately. Q.E.D.

Lemma 6. *Let $G=N\cdot T$ be the semidirect product of a cyclic normal subgroup $N=\{n\}\cong\mathbb{Z}/k\mathbb{Z}$ and a subgroup T, and let A be a G-module with trivial G-action. Let a map f: $N\times N\to A$ be a 2-cocycle on N of the form $f(n^i, n^j)=\left(\left[\frac{i+j}{k}\right]-\left[\frac{i}{k}\right]-\left[\frac{j}{k}\right]\right)f(n, n^{-1})$ $(i, j\in\mathbb{Z})$ such that*

(iii) $f(n', n'')-f({}^tn', {}^tn'')=u(t, n')-u(t, n'n'')+u(t, n'')$ $(n', n''\in N, t\in T)$

for a map u: $T\times N\to A$ with $u(1, n')=u(t, 1)=0$ $(n'\in N, t\in T)$. Then

$$\text{(iv)}\quad u(t, n^i)=\left(\left[\frac{ir(t)}{k}\right]-\left[\frac{i}{k}\right]\right)f(n, n^{-1})+iu(t, n)\qquad (i\in\mathbb{Z}, t\in T),$$

where $r(t)$ is an integer with ${}^tn=n^{r(t)}$ and $[\ ]$ is the Gauss symbol.

Proof. Obvious.

Let $G_{k,m}$ be the group of order km generated by elements n, t with defining relations:

$$n^k=1,\qquad t^m=1,\qquad tnt^{-1}=n^r\quad\text{with } r^m\equiv 1 \pmod{k}.$$

Then $G_{k,m}$ is the semidirect product of the cyclic subgroups $N=\{n\}$ and $T=\{t\}$.

Theorem 7. *Let $G=G_{k,m}$ be as above, and let A be a G-module with trivial G-action. Then $\tilde{H}^2(G, A)$ is isomorphic to the additive group consisting of elements (a, b) in $A\times A$ with the relation:*

$$\text{(v)}\quad \left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m-1}{k}\right)\left\{\frac{r-1}{(k, r-1)}a+\frac{k}{(k, r-1)}b\right\}=0,$$

taken modulo the subgroup consisting of elements

$$\text{(vi)}\qquad (kc, (1-r)c)\quad\text{with } c\in A,$$

via the map sending a cohomology class of f to the element (a, b) with $a=f(n, n^{-1})$ and $b=f(t, n)$, provided f is normalized as

$$f(n^i, n^j)=\left(\left[\frac{i+j}{k}\right]-\left[\frac{i}{k}\right]-\left[\frac{j}{k}\right]\right)f(n, n^{-1})\qquad (i, j\in\mathbb{Z}).$$

Here $\left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m-1}{k}\right)$ and $(k, r-1)$ are the greatest common divisors of k, $r-1$, $\sum_{l=0}^{m-1} r^l$ and $\frac{r^m-1}{k}$; and k and $r-1$, respectively.

Proof. Let f be a normal 2-cocycle on G representing a cohomology class of $\tilde{H}^2(G, A)$. By Proposition 1, f is a 2-cocycle on N, hence by

Lemma 4, f can be normalized as follows:

$$f(n^i, n^j) = \left(\left[\frac{i+j}{k}\right] - \left[\frac{i}{k}\right] - \left[\frac{j}{k}\right]\right) f(n, n^{-1}) \qquad (i, j \in \mathbb{Z}). \tag{3}$$

But by Proposition 1,

$$f(n', n'') - f({}^{t'}n', {}^{t'}n'') = f(t', n') - f(t', n'n'') + f(t', n'') \qquad (n', n'' \in N, t' \in T),$$

hence by Lemma 6,

$$f(t, n^j) = \left(\left[\frac{jr}{k}\right] - \left[\frac{j}{k}\right]\right) f(n, n^{-1}) + j f(t, n) \qquad (j \in \mathbb{Z}). \tag{4}$$

On the other hand by Proposition 1,

$$f(t't'', n') = f(t'', n') + f(t', {}^{t''}n') \qquad (t', t'' \in T, n' \in N).$$

Then

$$f(t^i, n') = \sum_{l=0}^{i-1} f(t, {}^{t^l}n') \qquad (i \in \mathbb{Z}),$$

hence by (4),

$$f(t^i, n^j) = \left(\left[\frac{jr^i}{k}\right] - \left[\frac{j}{k}\right]\right) f(n, n^{-1}) + j \sum_{l=0}^{i-1} r^l f(t, n) \qquad (i, j \in \mathbb{Z}). \tag{5}$$

Since $f(1, n') = f(t', 1) = 0$ ($n' \in N, t' \in T$), we have

$$(r-1) f(n, n^{-1}) + k f(t, n) = 0 \tag{6}$$

and

$$\frac{r^m - 1}{k} f(n, n^{-1}) + \sum_{l=0}^{m-1} r^l f(t, n) = 0, \tag{7}$$

by setting $i = 1, j = k$, and $i = m, j = 1$ in (5), respectively. Moreover (6) and (7) can be unified to be a single equality

$$\left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m - 1}{k}\right) \left\{\frac{r-1}{(k, r-1)} f(n, n^{-1}) + \frac{k}{(k, r-1)} f(t, n)\right\} = 0, \tag{8}$$

which is (v) with $a = f(n, n^{-1})$ and $b = f(t, n)$.

Conversely if $f(n, n^{-1})$ and $f(t, n)$ satisfy (8), then $f(n^i, n^j)$ and $f(t^i, n^j)$ are well-defined by (3) and (5), respectively. Moreover they satisfy (b) ~ (e) in Proposition 1, hence determine a normal 2-cocycle on G with $f(T, T) = 0$.

Let f be a normal 2-coboundary on G with $f(T, T) = 0$, then there exists a map $v: G \to A$ with $v(1) = 0$ such that

$$f(g, g') = v(g) - v(gg') + v(g') \qquad (g, g' \in G).$$

Then it is straightforward to see that

$$f(n, n^{-1}) = k v(n) \quad \text{and} \quad f(t, n) = (1-r) v(n),$$

which is (vi) with $a = f(n, n^{-1})$, $b = f(t, n)$ and $c = v(n)$.

Conversely if $f(n, n^{-1}) = kc$ and $f(t, n) = (1-r)c$ with $c \in A$, then f is a normal 2-cobundary on G with $f(T, T) = 0$. Q.E.D.

Let A be a $G_{k,m}$-module with trivial $G_{k,m}$-action. Then by Lemma 4, the canonical isomorphism

$$H^2(N, A)^T \cong {}_{(1-r)}(A/kA)$$

sends a cohomology class of f to $f(n, n^{-1}) \pmod{{}_kA}$ with $(1-r) f(n, n^{-1}) + k u(t, n) = 0$, provided f is normalized as follows:

$$f(n^i, n^j) = \left(\left[\frac{i+j}{k} \right] - \left[\frac{i}{k} \right] - \left[\frac{j}{k} \right] \right) f(n, n^{-1}) \qquad (i, j \in \mathbb{Z}),$$

and

$$f(n', n'') - f({}^{t'}n', {}^{t'}n'') = u(t', n') - u(t', n'n'') + u(t', n'') \qquad (n', n'' \in N,\ t' \in T),$$

for a map u: $T \times N \to A$ with $u(1, n') = u(t', 1) = 0$ $(n' \in N,\ t' \in T)$.

Moreover by the same lemma, the canonical isomorphism

$$H^2(T, H^1(N, A)) \cong {}_{(r-1)}({}_kA) \Big/ \left(\sum_{l=0}^{m-1} r^l \right) {}_kA = {}_{(k, r-1)}A \Big/ \left(\sum_{l=0}^{m-1} r^l \right) {}_kA$$

sends a cohomology class of h to $h(t, t^{-1}, n)$ $\left(\mathrm{mod} \left(\sum_{l=0}^{m-1} r^l \right) {}_kA \right)$, provided h is normalized as

$$h(t^i, t^j, n') = \left(\left[\frac{i+j}{m} \right] - \left[\frac{i}{m} \right] - \left[\frac{j}{m} \right] \right) h(t, t^{-1}, n') \qquad (i, j \in \mathbb{Z},\ n' \in N).$$

Theorem 8. *Let $G = G_{k,m}$ be as above, and let A be a G-module with trivial G-action. Then the homomorphism*

$$d_2 \colon {}_{(1-r)}(A/kA) \to {}_{(k, r-1)}A \Big/ \left(\sum_{l=0}^{m-1} r^l \right) {}_kA$$

sends $a \pmod{kA}$ with $(1-r)a = kb$ to

$$\frac{r^m - 1}{k} a + \sum_{l=0}^{m-1} r^l b \left(\mathrm{mod} \left(\sum_{l=0}^{m-1} r^l \right) {}_kA \right).$$

Proof. Let f, with $f(n^i, n^j) = \left(\left[\frac{i+j}{k} \right] - \left[\frac{i}{k} \right] - \left[\frac{j}{k} \right] \right) f(n, n^{-1})$ $(i, j \in \mathbb{Z})$, be a 2-cocycle on N representing a cohomology class of $H^2(N, A)^T$, then by Lemma 5, there exists a map v: $T \times N \to A$ with $v(1, n') = v(t', 1) = 0$

$(n' \in N,\ t' \in T)$ such that

$$v(t^i, n') = \sum_{l=0}^{i-1} v(t, {}^{t^l}n') \qquad (0 \leqq i < m,\ n' \in N),$$

$$f(n^i, n^j) - f({}^t n^i, {}^t n^j) = v(t, n^i) - v(t, n^{i+j}) + v(t, n^j) \qquad (i, j \in \mathbb{Z}),$$

and that the cohomology class of $d_2 f$ is

$$\sum_{l=0}^{m-1} v(t, {}^{t^l}n) \left(\operatorname{mod} \left(\sum_{l=0}^{m-1} r^l \right) {}_k A \right).$$

On the other hand by Lemma 6,

$$v(t, n^j) = \left(\left[\frac{jr}{k} \right] - \left[\frac{j}{k} \right] \right) f(n, n^{-1}) + j\, v(t, n) \qquad (j \in \mathbb{Z}).$$

Hence

$$(1-r) f(n, n^{-1}) + k\, v(t, n) = 0,$$

and

$$\sum_{l=0}^{m-1} v(t, {}^{t^l}n) = \frac{r^m - 1}{k} f(n, n^{-1}) + \sum_{l=0}^{m-1} r^l v(t, n). \quad \text{Q.E.D.}$$

§ 2. An Application: Schur Multiplicator

As usual $\mathbb{Z}, \mathbb{Q}$ are the ring of integers, the field of rational numbers, respectively. Then

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0 \quad \text{(exact)}.$$

Let G be a finite group. Then

$$H^3(G, \mathbb{Z}) \cong H^2(G, \mathbb{Q}/\mathbb{Z}).$$

This group is called the Schur multiplicator of the group G.

Let G be a finite abelian group, then $H^3(G, \mathbb{Z})$ is trivial if and only if G is cyclic. We now compute $H^3(G, \mathbb{Z})$ when G is a semidirect product of its finite cyclic subgroups. Let $G_{k,m}$ be the group of order km generated by elements n, t with defining relations:

$$n^k = 1, \qquad t^m = 1, \qquad t n t^{-1} = n^r \quad \text{with } r^m \equiv 1 \pmod{k}.$$

Then $G_{k,m}$ is the semidirect product of the cyclic subgroups $N = \{n\}$ and $T = \{t\}$.

Proposition 9.

$$H^3(G_{k,m}, \mathbb{Z}) = \mathbb{Z} \Big/ \left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m - 1}{k} \right) \mathbb{Z},$$

where $\left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m-1}{k} \right)$ is the greatest common divisor of $k, r-1$, $\sum_{l=0}^{m-1} r^l$ and $\frac{r^m-1}{k}$.

Proof. Since the Schur multiplicator of a cyclic group is trivial, we have by Theorem 7

$$H^3(G_{k,m}, \mathbb{Z}) = H^1(T, H^1(N, \mathbb{Q}/\mathbb{Z})) = \mathbb{Z}\Big/\left(k, r-1, \sum_{l=0}^{m-1} r^l, \frac{r^m-1}{k}\right)\mathbb{Z}. \text{ Q.E.D.}$$

We list below the Schur multiplicators of all non-abelian groups of order $\leqq 30$, but in this table there are some groups whose ones are computed by the well-known methods, e.g. the restriction map and the transfer map. We use the notation of Coxeter-Moser [2] (p. 134).

Order	Symbol	$H^3(G, \mathbb{Z})$	Order	Symbol	$H^3(G, \mathbb{Z})$
6	D_3	0	21	R''	0
8	D_4	$\mathbb{Z}/2\mathbb{Z}$	22	D_{11}	0
	$Q = \langle 2, 2, 2\rangle$	0	24	$\mathbb{Z}_2 \times A_4$	$(\mathbb{Z}/2\mathbb{Z})^2$
10	D_5	0		$\mathbb{Z}_2 \times D_6$	$(\mathbb{Z}/2\mathbb{Z})^3$
12	D_6	$\mathbb{Z}/2\mathbb{Z}$		$\mathbb{Z}_3 \times D_4$	$\mathbb{Z}/2\mathbb{Z}$
	A_4	$\mathbb{Z}/2\mathbb{Z}$		$\mathbb{Z}_3 \times Q$	0
	$\langle 2, 2, 3\rangle$	0		$\mathbb{Z}_4 \times D_3$	$\mathbb{Z}/2\mathbb{Z}$
14	D_7	0		$\mathbb{Z}_2 \times \langle 2, 2, 3\rangle$	$\mathbb{Z}/2\mathbb{Z}$
16	$\mathbb{Z}_2 \times D_4$	$(\mathbb{Z}/2\mathbb{Z})^3$		D_{12}	$\mathbb{Z}/2\mathbb{Z}$
	$\mathbb{Z}_2 \times Q$	$(\mathbb{Z}/2\mathbb{Z})^2$		S_4	$\mathbb{Z}/2\mathbb{Z}$
	D_8	$\mathbb{Z}/2\mathbb{Z}$		$\langle 2, 3, 3\rangle$	0
	$\langle -2, 4 \mid 2\rangle$	0		$(4, 6 \mid 2, 2)$	$\mathbb{Z}/2\mathbb{Z}$
	$\langle 2, 2 \mid 2\rangle$	0		$\langle -2, 2, 3\rangle$	0
	$\langle 2, 2 \mid 4, 2\rangle$	$\mathbb{Z}/2\mathbb{Z}$		$\langle 2, 2, 6\rangle$	0
	$(4, 4 \mid 2, 2)$	$(\mathbb{Z}/2\mathbb{Z})^2$	26	D_{13}	0
	R	$(\mathbb{Z}/2\mathbb{Z})^2$	27	$(3, 3 \mid 3, 3)$	$(\mathbb{Z}/3\mathbb{Z})^2$
	$\langle 2, 2, 4\rangle$	0		R'''	0
18	$\mathbb{Z}_3 \times D_3$	0	28	D_{14}	$\mathbb{Z}/2\mathbb{Z}$
	D_9	0		$\langle 2, 2, 7\rangle$	0
	$(3, 3, 3 : 2)$	$\mathbb{Z}/3\mathbb{Z}$			
20	D_{10}	$\mathbb{Z}/2\mathbb{Z}$	30	$\mathbb{Z}_3 \times D_5$	0
	R'	0		$\mathbb{Z}_5 \times D_3$	0
	$\langle 2, 2, 5\rangle$	0		D_{15}	0

References

1. Cartan, H., Eilenberg, S.: Homological algebra. Princeton: University Press 1956.
2. Coxeter, H.S.M., Moser, W.O.J.: Generators and relations for discrete groups. Berlin-Heidelberg-New York: Springer 1957.
3. Eilenberg, S., MacLane, S.: Cohomology theory in abstract groups, I. Ann. of Math. **48**, 51–78 (1947).
4. Huppert, B.: Endliche Gruppen I. Berlin-Heidelberg-New York: Springer 1967.

5. Lyndon, R.C.: The cohomology theory of group extensions. Duke Math. J. **15**, 271-292 (1948).
6. Schur, I.: Über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen. J. reine angew. Math. **127**, 20-50 (1904).
7. Schur, I.: Untersuchungen über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen. J. reine angew. Math. **132**, 85-137 (1907).
8. Schur, I.: Über die Darstellungen der symmetrischen und alternierenden Gruppen durch gebrochene lineare Substitutionen. J. reine angew. Math. **139**, 155-250 (1911).
9. Wiegold, J.: Some groups with non-trivial multiplicators. Math. Z. **120**, 307-308 (1971).

Dr. Ken-ichi Tahara
Department of Mathematics
Aichi University of Education
Igaya-cho, Kariya-shi
Japan 448

(Received April 12, 1972)

INSTRUCTIONS TO AUTHORS

A. General

Papers submitted for publication may be written in English, French or German. A **summary** for Zentralblatt für Mathematik (including key words for classification) should be attached.

To speed up publication, authors will receive **only one set of proofs:** provisionally numbered page proofs. Authors are requested to **correct typographical errors only;** they will be charged for corrections involving changes, additions or deletions to the original manuscript.

Manuscripts must be in their **final form,** typed on one side of each sheet only, with double spacing and wide margins. Formulae should be typewritten whenever possible. Mimeographed copies are not acceptable unless clearly legible.

Please include a "Note for the Printer" explaining markings used. See suggestions overleaf.

Diagrams should be submitted on separate sheets, not included in the text. They should be drawn in Indian ink in clean and clear uniform lines, the whole about twice the size of the finished illustration. Inscriptions should allow for the figure 1, for example, to be about 2 mm high in the final version (i.e. 4 mm for reduction $\times \frac{1}{2}$). The author should mark in the margin of the manuscript where diagrams may be inserted.

Footnotes, other than those which refer to the title heading, should be numbered consecutively and placed at the foot of the page to which they refer (not at the end of the article).

Please give on the first page of the manuscript a **running head** (condensed title), which should not exceed 70 letters including spaces.

References to the literature should be listed at the end of the manuscript. The following information should be provided for **journal articles:** names and initials of all authors, full title of the paper, name of the journal, volume, first and last page numbers and year of publication. References to **books** should include name(s) of author(s), full title, edition, place of publication, publisher and year of publication.

Examples

Bombieri, E., Giusti, E.: Harnack's inequality for elliptic differential equations on minimal surfaces. Inventiones math. **15**, 24–46 (1971).

Tate, J.T.: *p*-Divisible Groups. In: Proceedings of a Conference on Local Fields, pp. 158–183. Berlin-Heidelberg-New York: Springer 1967.

B. Marking

1. Text

Words or sentences to be set in *italics* should be marked by single underlining.

The words **"Theorem"**, **"Proposition"**, **"Lemma"**, **"Corollary"** etc. are normally printed in **boldface**, followed by the formulation in italics (to be underlined in the manuscript).
The words "*Proof*", "*Remark*", "*Example*", "*Case*" etc. are printed in italics, and "Definition" in boldface with the formulation in ordinary typeface.

2. Formulae

Letters in formulae are normally printed in italics, figures in ordinary typeface.
It will help the printer if in doubtful cases the position of indices and exponents is marked thus: $b_{\wedge j}$, $a^{\vee}$. Spacing of indices and exponents must be specially indicated $(A_m{}^n{}_n{}^m)$, otherwise they will be set (A^{nm}_{mn}).
Underlining for special alphabets and typefaces should be done according to the following code:

single underlining:	small letter
double underlining:	capital letter
brown:	boldface (headings, boldface letters in formulae)
yellow:	upright (abbreviations e.g. Re, Im, log, sin, ord, id, lim, sup, etc.)
red:	Greek
blue:	Gothic
green:	Script
violet:	the numerals 1 and 0 (to distinguish them from the small letter *l* and the capital letter *O*)

The following are frequently confused:

$\circ, o, O, 0$; $\cup, \mathrm{u}, \bigcup, U$; $\times, x, X, \kappa$; $\vee, \nu, v$; $\theta, \Theta, \phi, \varphi, \Phi, \varnothing$; ψ, Ψ; $\iota, \in$;

a', a^1; the symbol a and the indefinite article a;

also the handwritten Roman letters:

c, C; e, l; j, J; k, K; o, O; p, P; s, S; u, U; v, V; w, W; x, X; z, Z;

Please take care to distinguish them in some way.

C. Examples

1. Special alphabets or typefaces

Script	$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{G}, \mathcal{H}, \mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{L}, \mathcal{M}, \mathcal{N}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$
	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z
Sanserif	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z
	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z
Gothic	$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}, \mathfrak{E}, \mathfrak{F}, \mathfrak{G}, \mathfrak{H}, \mathfrak{I}, \mathfrak{J}, \mathfrak{K}, \mathfrak{L}, \mathfrak{M}, \mathfrak{N}, \mathfrak{O}, \mathfrak{P}, \mathfrak{Q}, \mathfrak{R}, \mathfrak{S}, \mathfrak{T}, \mathfrak{U}, \mathfrak{V}, \mathfrak{W}, \mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$
	$\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{d}, \mathfrak{e}, \mathfrak{f}, \mathfrak{g}, \mathfrak{h}, \mathfrak{i}, \mathfrak{j}, \mathfrak{k}, \mathfrak{l}, \mathfrak{m}, \mathfrak{n}, \mathfrak{o}, \mathfrak{p}, \mathfrak{q}, \mathfrak{r}, \mathfrak{s}, \mathfrak{t}, \mathfrak{u}, \mathfrak{v}, \mathfrak{w}, \mathfrak{x}, \mathfrak{y}, \mathfrak{z}$
Boldface	**A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z**
	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z
Special Roman	$\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{I}, \mathbb{J}, \mathbb{K}, \mathbb{L}, \mathbb{M}, \mathbb{N}, \mathbb{O}, \mathbb{P}, \mathbb{Q}, \mathbb{R}, \mathbb{S}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{W}, \mathbb{X}, \mathbb{Y}, \mathbb{Z}, \mathbb{1}$
Greek	$\Gamma, \Delta, \Theta, \Lambda, \Xi, \Pi, \Sigma, \Phi, \Psi, \Omega$
	$\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta, \theta, \vartheta, \iota, \kappa, \lambda, \mu, \nu, \xi, o, \pi, \rho, \sigma, \tau, \upsilon, \varphi, \phi, \chi, \psi, \omega$

2. Notations

preferred form	instead of	preferred form	instead of
$A^*, \tilde{b}, \gamma', \mathfrak{v}, \mathbf{v}$	$\hat{A}, \hat{b}, \check{\gamma}, \vec{v}$	$\exp(-(x^2+y^2)/a^2)$	$e^{-\frac{x^2+y^2}{a^2}}$
lim sup, lim inf	$\overline{\lim}, \underline{\lim}$	$\cos(1/x)$	$\cos\frac{1}{x}$
inj lim, proj lim	$\underset{\longrightarrow}{\lim}, \underset{\longleftarrow}{\lim}$	$(a+b/x)^{1/2}$	$\sqrt{a+\frac{b}{x}}$
$f: A \to B$	$A \xrightarrow{f} B$		
f^{-1}	$\overset{-1}{f}$		