Werk

Label: Periodical issue

Jahr: 1968

PURL: https://resolver.sub.uni-goettingen.de/purl?266833020_0105|log16

Kontakt/Contact

Digizeitschriften e.V.
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

info@digizeitschriften.de

A 21267 D

Mathematische Zeitschrift

Herausgegeben von **H. Wielandt,** Tübingen

unter Mitwirkung von **B. Eckmann,** Zürich
E. Heinz, Göttingen
R. Nevanlinna, Helsinki
K. Zeller, Tübingen

Wissenschaftlicher Beirat H. Kneser, Tübingen
W. Magnus, New York
O. Perron, München
G. Pickert, Gießen

Band 105 · Heft 2 · 1968

Springer-Verlag · Berlin · Heidelberg · New York

Math. Z. 20. Mai 1968

Niedersächsische Staats- u.
Universitätsbibliothek
Göttingen

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

„**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 Zeitschrift erscheint, um eine rasche Publikation zu ermöglichen, in einzelnen Heften, die zu Bänden vereinigt werden. Ein Band besteht im allgemeinen aus 5 Heften. Der Preis eines Bandes beträgt DM 96,—.

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, sie 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 Vervielfältigung 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 0.30 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.

In the interest of speedy publications, this journal is issued at frequent intervals according to the material received. As a rule five numbers constitute one volume. The price is DM 96.— per volume.

"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 incumbent upon the author to submit no manuscript which has been or will be published elsewhere either at home or abroad. Unless special permission has been granted by the publishers, no photographic reproductions, microfilms, microphotos, or other reproductions of a similar nature may be made of the issues or 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:

B. Eckmann, Zürich (Schweiz), Eidgen. Technische Hochschule
E. Heinz, 34 Göttingen, Math. Inst. d. Universität
R. Nevanlinna, Helsinki (Finnland), Computing Center, University of Helsinki, Töölönkatu
H. Wielandt, 74 Tübingen, Math. Inst. d. Universität
K. Zeller, 74 Tübingen, Math. Inst. d. Universität

Springer-Verlag

69 Heidelberg 1	1 Berlin 33	Springer-Verlag
Postfach 1780	Heidelberger Platz 3	New York Inc.
Fernsprecher 49101	Fernsprecher 830301	175 Fifth Avenue
Fernschreibnummer 04-61723	Fernschreibnummer 01-83319	New York, N.Y. 10010

Math. Zeitschr. 105, 87—89 (1968)

Über das Nichtverschwinden der Funktionaldeterminante bei einer Klasse eineindeutiger Abbildungen

ERHARD HEINZ

Eingegangen am 24. Januar 1968

Es seien G und G^* zwei Gebiete im R^2, und $z\colon G\to G^*$ sei eine eineindeutige Abbildung der Klasse C^2 mit $z=z(w)=x+iy$ $(w=u+iv)$, wobei die reellwertigen Funktionen $x=x(u,v)$ und $y=y(u,v)$ in G dem quasilinearen elliptischen System

$$
\begin{aligned}
\Delta x &= h_1(x,y)(x_u^2+x_v^2)+h_2(x,y)(x_u y_u+x_v y_v)\\
&\quad +h_3(x,y)(y_u^2+y_v^2)+h_4(x,y)(x_u y_v-x_v y_u),\\
\Delta y &= \tilde{h}_1(x,y)(x_u^2+x_v^2)+\tilde{h}_2(x,y)(x_u y_u+x_v y_v)\\
&\quad +\tilde{h}_3(x,y)(y_u^2+y_v^2)+\tilde{h}_4(x,y)(x_u y_v-x_v y_u)
\end{aligned}
\tag{1}
$$

genügen und die Koeffizienten $h_1(x,y),\ldots,\tilde{h}_4(x,y)$ in G^* definiert sind. Von Bedeutung für differentialgeometrische Anwendungen ist die Frage, unter welchen zusätzlichen Voraussetzungen hinsichtlich $h_1,\ldots,\tilde{h}_4$ die Funktionaldeterminante $J(w)$ der Abbildung $w\to z(w)$ nicht verschwindet. Das erste Resultat in dieser Richtung verdankt man LEWY [6], der zeigte, daß stets $J(w)\neq 0$ gilt, falls die Funktionen $h_1,\ldots,\tilde{h}_4$ in G^* analytisch sind. Wie in [4] bewiesen wurde, läßt sich diese Bedingung durch die schwächere ersetzen, daß $h_1,\ldots,\tilde{h}_4$ in G^* einer Lipschitzbedingung genügen. In der vorliegenden Note beweisen wir eine weitergehende Verallgemeinerung, nämlich den folgenden

Satz. *Die Abbildung $z\colon G\to G^*$ sei eineindeutig und genüge dem System (1), wobei die Koeffizienten $h_1(x,y),\ldots,\tilde{h}_4(x,y)$ in G^* beschränkt sind und die Funktionen $\omega_1(x,y)=\tilde{h}_1(x,y)$, $\omega_2(x,y)=h_1(x,y)-\tilde{h}_2(x,y)$, $\omega_3(x,y)=h_2(x,y)-\tilde{h}_3(x,y)$ sowie $\omega_4(x,y)=h_3(x,y)$ in G^* eine Lipschitzbedingung erfüllen. Dann ist $J(w)\neq 0$ in G.*

Dieses Ergebnis stellt eine Ergänzung zu Satz 2 aus [5] dar. Die Voraussetzungen hinsichtlich $h_1,\ldots,\tilde{h}_4$ sind den differentialgeometrischen Anwendungen auf Einbettungs- und Verbiegungsprobleme in natürlicher Weise angepaßt. Der Beweis beruht auf einer einfachen Anwendung zweier analytischer Hilfssätze, die von HARTMAN-WINTNER [3] bzw. vom Verfasser [5] herrühren (Hilfssatz 1 und 2)[1], sowie eines von BERG [1] stammenden topologischen Lemmas (Hilfssatz 3).

[1] Die Beweisidee von Hilfssatz 1 geht auf CARLEMAN [2] und die von Hilfssatz 2 auf LEWY [6] zurück.

Hilfssatz 1. *Es sei* $\varphi=\varphi(w)=\varphi(u,v)$ *eine in* $|w|<\varepsilon$ *reellwertige Funktion der Klasse* C^2, *die der Differentialungleichung*

$$|\Delta\varphi|\leqq c\left(|\varphi_u|+|\varphi_v|+|\varphi|\right) \qquad (c<+\infty) \tag{2}$$

genügt. Außerdem sei

$$\varphi(0,0)=\varphi_u(0,0)=\varphi_v(0,0)=0 \tag{3}$$

und $\varphi\not\equiv 0$. *Dann gilt eine asymptotische Darstellung der Form*

$$\varphi(w)=\mathrm{Re}(A\,w^k)+o\left(|w|^k\right) \qquad (w\to 0) \tag{4}$$

mit $A\neq 0$ *und* $k>1$ *(k ganz).*

Beweis. Dies folgt unmittelbar aus [3], Theorem 1 und 2.

Hilfssatz 2. *Die Abbildung* $z\colon G\to G^*$ *erfülle die Voraussetzungen unseres Satzes. Außerdem sei* $0\in G, G^*$; $z(0)=0$; *und* a, b *seien zwei reelle Zahlen. Dann gibt es eine in* $|z|<\delta$ *reellwertige Funktion* $f=f(z)=f(x,y)$ *der Klasse* C^2 *mit*

$$f(0,0)=0; \quad f_x(0,0)=a, \quad f_y(0,0)=b, \tag{5}$$

so daß $\varphi(w)=f(z(w))$ *in* $|w|<\varepsilon$ *der Differentialungleichung (2) genügt.*

Beweis. Dies folgt unmittelbar aus [5], Hilfssatz 3.

Hilfssatz 3. $\Phi(w)$ *sei eine für* $|w|<\rho$ *komplexwertige, stetige Funktion. Außerdem sei die Abbildung* $w\to\Phi(w)$ *eineindeutig, und für* $\varphi(w)=\mathrm{Re}\,\Phi(w)$ *gelte die asymptotische Darstellung (4) mit* $A\neq 0$ *und k ganz-positiv. Dann hat man* $k=1$.

Beweis. Siehe [1], S. 314.

Beweis des Satzes. Ohne Beschränkung der Allgemeinheit dürfen wir $0\in G, G^*$ und $z(0)=0$ voraussetzen, und es genügt, $J(0)\neq 0$ zu beweisen. Angenommen, es sei $J(0)=0$. Dann gibt es zwei reelle Zahlen a, b mit $a^2+b^2=1$ und

$$\begin{aligned} a\,x_u(0,0)+b\,y_u(0,0)&=0,\\ a\,x_v(0,0)+b\,y_v(0,0)&=0. \end{aligned} \tag{6}$$

Nach Hilfssatz 2 existiert eine in $|z|<\delta$ reellwertige Funktion $f(z)=f(x,y)$ der Klasse C^2, so daß (5) gilt und $\varphi=\varphi(w)=f(z(w))$ für $|w|<\varepsilon$ der Differentialungleichung (2) genügt. Man setze

$$\Phi(w)=\varphi(w)+i\left(a\,y(u,v)-b\,x(u,v)\right). \tag{7}$$

Aus $a^2+b^2=1$ sowie (5) und der Eineindeutigkeit von $z\colon G\to G^*$ folgt, daß die Abbildung $w\to\Phi(w)$ $(|w|<\rho)$ für genügend kleines $\rho\leqq\varepsilon$ ebenfalls eineindeutig ist. Insbesondere ist $\varphi\not\equiv 0$, und wegen (5) und (6) hat man die Gl. (3).

Nach Hilfssatz 1 erhält man daher für $\varphi(w)=\mathrm{Re}\,\Phi(w)$ die asymptotische Darstellung (4) mit $A\neq 0$ und $k>1$ (k ganz). Hilfssatz 3 wird jetzt auf die Abbildung $w\to\Phi(w)$ anwendbar und liefert die Gleichung $k=1$, also Widerspruch.

Somit ist $J(0)\neq 0$, womit die Behauptung bewiesen ist.

Literatur

1. Berg, P.W.: On univalent mappings by solutions of linear elliptic partial differential equations. Trans. Amer. Math. Soc. **84**, 310—318 (1957).
2. Carleman, T.: Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables. Comptes Rendus Paris **197**, 471—474 (1933).
3. Hartman, P., and A. Wintner: On the local behavior of solutions of nonparabolic partial differential equations. Amer. J. Math. **75**, 449—476 (1953).
4. Heinz, E.: Über gewisse elliptische Systeme von Differentialgleichungen zweiter Ordnung mit Anwendung auf die Monge-Ampèresche Gleichung. Math. Annalen **131**, 411—428 (1956).
5. — Neue a-priori-Abschätzungen für den Ortsvektor einer Fläche positiver Gaußscher Krümmung durch ihr Linienelement. Math. Zeitschr. **74**, 129—157 (1960).
6. Lewy, H.: On the non-vanishing of the Jacobian in certain one-to-one mappings. Bull. Amer. Math. Soc. **42**, 689—692 (1936).

Prof. Dr. E. Heinz
Mathematisches Institut der Universität
Göttingen
3400 Göttingen, Bunsenstraße 3/5

Math. Zeitschr. 105, 90–109 (1968)

Unitary Groups Over Compact Valuation Rings of Characteristic 2

SO-FEI WEI

Received August 5, 1967

It is well-known that the unitary group is generated by quasi-symmetries and the proper unitary group is generated by unitary transvections in the field case [1–5]. In [6–8], KLINGENBERG initiated to extend the theory of linear groups, orthogonal groups and symplectic groups over a local ring of characteristic not 2 rather than a field. The unitary groups over a generalized valuation ring with 2 as a unit for the unramified case was studied by SCHNEIDER in his MIT thesis [11]. This paper will deal with the generation and the normal subgroups of unitary groups over a compact valuation ring of characteristic 2 (i.e. the formal power series ring $F[[\pi]]$ over a finite field F of characteristic 2 with $\pi^{\infty}=0$). Based on the structure of hermitian lattices due to SAH [10], we are able to establish theorems similar to the classical ones and further generalization.

The main results are contained in Theorems A, B, C for the unramified case and Theorems D, E, F for the ramified case. In view of Theorem C, every normal subgroup in $U^{+}(L)$ satisfies a ladden relation similar to the works of KLINGENBERG and SCHNEIDER. The result of normal subgroups on the ramified situation are more complicated and will be dealt with in a separate publication.

I wish to thank Professor CHIH-HAN SAH for suggesting the problems presented here, for his invaluable help, critism and encouragement.

§ 1. Notations

Throughout this paper K is a local field of characteristic 2 with a non-trivial involution $^{-}$. Let k be the fixed field of $^{-}$, hence $[K:k]=2$. We may then put $k=F\langle\langle\pi\rangle\rangle$, the formal power series field over a finite field F with uniformizer π such that $\pi^{\infty}=0$. $R=F[[\pi]]$ is the ring of integers of k and $S=$the integral closure of R in K. ord_K and ord_k are the ordinal functions in K and k respectively. For any a in K, $T(a)=a+\bar{a}$ and $N(a)=a\bar{a}$ are called the trace and the norm of a respectively. It is clear that $T(a)$ and $N(a)$ are in k for all a in K. We have $T(S)=\pi^{s}R$ for s a non-negative integer. If $s=0$, then K is an unramified extension of k; if $s>0$, the extension is ramified. We shall denote S^{*} and R^{*} to be the units of S and R respectively.

By a hermitian space (resp. lattice) L over K, we mean a vector space (resp. free module) of finite type over K (resp. S) together with a map $\langle\ ,\ \rangle$ from $L\times L$ to K such that $\langle\ ,\ \rangle$ is linear on the first variable and $\overline{\langle x,y\rangle}=\langle y,x\rangle$ holds for all x, y in L. A unitary transformation u of L is a one-to-one linear

map of L onto itself such that $\langle u(x), u(y)\rangle = \langle x, y\rangle$ for all x, y in L. The set of all unitary transformations form a group, $U(L)$, called the unitary group of L. The proper unitary group $U^+(L)$ of L consists of all unitary transformations of determinant 1. Thus, $U^+(L)$ is a normal subgroup of $U(L)$.

Let L be a hermitian lattice over S, then $KL = K \underset{S}{\otimes} L$ is a hermitian space over K. We shall denote $\dim L = \dim KL$ and index of L = index of KL. If L is the direct sum ot two sublattices M and N with $\langle M, N\rangle = 0$, we shall write $L = M \perp N$. If $L = Sx$ with $\langle x, x\rangle = a \neq 0$, we shall write $L = (a)$. If $L = Sx + Sy$ with $\langle x, x\rangle = a$, $\langle y, y\rangle = c$ and $\langle x, y\rangle = b \neq 0$, we shall denote $L = \begin{pmatrix} b \\ a & c \end{pmatrix}$. A hermitian lattice is called a hyperbolic plane, if it is isomorphic to $\begin{pmatrix} 1 \\ 0 & 0 \end{pmatrix}$. The orthogonal sum of hyperbolic planes will be called the hyperbolic lattice.

Suppose that the maximal ideal of S is I. A vector x in L is called maximal, if $x \neq 0$ modulo I. A maximal vector x in L is called isotropic if $\langle x, x\rangle = 0$ and anisotropic otherwise. Furthermore, if $\langle x, x\rangle$ is in R^*, we call it non-isotropic. By a unimodular lattice, we mean a hermitian lattice L satisfying $\langle x, L\rangle = S$ for all maximal vector x in L.

Definition 1.1. A linear map on L is called *a quasi-symmetry* with respect to a non-isotropic vector x in L and a scalar γ in $\Delta = \{a \in S \mid a\bar{a} = 1\}$ and is denoted by $\tau_{x,\gamma}$, if for all y in L,

$$\tau_{x,\gamma}(y) = y + (1+\gamma)\frac{\langle y, x\rangle}{\langle x, x\rangle} x .$$

It is clear that $\tau_{x,\gamma}(x) = \gamma x$ and $\tau_{x,\gamma}(y) = y$, if $\langle x, y\rangle = 0$. An easy calculation shows that $\tau_{x,\gamma} \in U(L)$ and $\tau_{x,\gamma}\tau_{x,\bar{\gamma}} = 1$. $D(L)$ will denote the subgroup of $U(L)$, generated by all the quasi-symmetries.

Definition 1.2. By *a unitary transvection* of L with respect to an isotropic vector x in L and c in R, denoted by $\sigma_{x,c}$, we mean an endomorphism of L such that for each y in L we have $\sigma_{x,c}(y) = y + c\langle y, x\rangle x$. It is clear that $\sigma_{x,c}(y) = y$ if $\langle y, x\rangle = 0$, and that $\sigma_{x,c}^2 = 1$. In the case $c = 1$ we shall write σ_x. Every unitary transvection is in $U^+(L)$. Let $T(L)$ be the subgroup generated by all the unitary transvections.

§ 2. Unramified Case

We suppose that K is an unramified extension of $k = F\langle\langle\pi\rangle\rangle$. Then K may be expressed as $E\langle\langle\pi\rangle\rangle$ with a quadratic extension E of F and $S = E[[\pi]]$. Then the involution $^-$ on K is exactly the involution on E with F as the fixed field and $\bar{\pi} = \pi$. We note that $N(S^*) = R^*$. Let $\Sigma = \{a \in E \mid a\bar{a} = 1\}$. It is easy to see that Σ is non-trivial and $E = F\Sigma$.

If L is a unimodular lattice over S, then it can be decomposed as

$$L = (1) \perp (1) \perp \cdots \perp (1)$$

(cf. [10]). *From now on we only consider the unimodular case.*

The natural projection p_r: $S \to S/\pi^r S$ induces a canonical surjective map p_r: $L \to L/\pi^r L$ for non-negative integer r. The induced hermitian form on $L/\pi^r L$ will be denoted by $\langle\ ,\ \rangle_r$. It is easy to see that $\langle p_r x, p_r y\rangle_r = p_r\langle x, y\rangle$. Therefore we obtain a homomorphism h_r: $U(L) \to U(L/\pi^r L)$ with $h_r \sigma p_r = p_r \sigma$ for each $\sigma \in U(L)$.

Lemma 2.1. *Let x and y be in L with $\langle x, x\rangle = \langle y, y\rangle \in R^*$ and $\langle x, y\rangle \in S^*$. Then there is a $\sigma \in D(L)$ such that $\sigma(x) = y$.*

Proof. Put $a = \langle x, y\rangle$ and $z = x + y$. It is easy to see that z is non-isotropic iff the leading coefficient of a, $\varphi(a)$, is not in F. If z is non-isotropic, then $\tau_{z,a}(x) = y$ where

$$\alpha = \frac{\overline{\langle x, z\rangle}}{\langle x, z\rangle} \in \Delta;$$

otherwise, there is a $\beta \in \Sigma$ such that $\beta\varphi(a) \notin F$, hence $\tau_{x,\beta}(x) + y$ is non-isotropic. We are through.

Lemma 2.2. *Except when $n = 2$ and $E = F_4$ (the field of 4 elements), $D(L)$ is transitive on the non-isotropic vectors of same length.*

Proof. It is well-known that $U(L/\pi L)$ is generated by quasi-symmetries except when $n = 2$ and $E = F_4$ (cf. [4]), since $L/\pi L$ is a hermitian space. Therefore h_1 is surjective on $D(L)$, so for arbitrary x and y in L with $\langle x, x\rangle = \langle y, y\rangle \in R^*$, we may then find a $\sigma \in D(L)$ such that $\sigma(x) + y \in \pi L$. Thus

$$\langle \sigma(x) + y, y\rangle = \langle \sigma(x), y\rangle + \langle y, y\rangle \in \pi S,$$

so $\langle \sigma(x), y\rangle \in S^*$. We can now apply Lemma 2.1 to conclude our assertion.

Lemma 2.3. *Suppose that $E = F_4$ and L is a unimodular lattice over $E[[\pi]]$ of dimension n.*

(a) *If $n = 2$, then $|U(L) : D(L)| = 2$.*

(b) *If $n = 3$, then $U(L) = D(L)$.*

Proof. (a) Let $L = (1) \perp (1) = Sx \perp Sy$. For arbitrary $\sigma \in U(L)$, put $\sigma(x) = ax + by$. Then at least one of a and b is in S^*. If $a \in S^*$, there is a $\tau \in D(L)$ such that $\tau\sigma(x) = x$ by Lemma 2.1, so $\tau\sigma(y) = \gamma y$ with $\gamma \in \Delta$, i.e. $\sigma \in D(L)$. If $b \in S^*$, similarly there is a $\rho \in D(L)$ such that $\rho\sigma(x) = y$, so $\rho\sigma = \eta$ where η is the transformation exchanging x and y. It is clear that $\eta \notin D(L)$ and $\eta^2 = 1$.

(b) Let $L = (1) \perp (1) \perp (1) = Sx \perp Sy \perp Sz$. According to (a) and Lemma 2.2, it suffices to show that the transformation σ which satisfies $\sigma(x) = x$, $\sigma(y) = z$ and $\sigma(z) = y$, is in $D(L)$. Take $\alpha \in E$, but $\alpha \notin F$, so $\alpha\bar{\alpha} = 1$ and $\alpha + \bar{\alpha} = 1$. Set $\tau = \tau_{x+y+z,\alpha}\sigma$, then $\tau(x) = \alpha x + \bar{\alpha}(y + z)$, $\tau(y) = \alpha z + \bar{\alpha}(x + y)$ and $\tau(z) = \alpha y + \bar{\alpha}(x + z)$. Let $\rho = \tau_{\tau(y)+y,\alpha}\tau$. Then by a straightforward computation, we have $\rho(x) = \bar{\alpha}x$, $\rho(y) = y$ and $\rho(z) = z$. Clearly, $\rho = \tau_{x,\bar{\alpha}} \in D(L)$; it follows that $\sigma \in D(L)$ as desired.

Therefore by the previous Lemmas and by induction, we have

Theorem A. *Let L be a unimodular lattice over $E[[\pi]]$. Then every unitary transformation of L is the product of quasi-symmetries, except when* $\dim L=2$ *and* $E=F_4$.

Lemma 2.4. *Let L be a binary unimodular lattice. Then $U^+(L)=T(L)$. Furthermore, it is exactly the special linear group over R, i.e. $SL_2(R)$.*

Proof. We know that

$$L=(1)\perp(1)\cong\begin{pmatrix}1 & \\ 0 & 0\end{pmatrix}=Sx+Sy.$$

Given an arbitrary $\sigma\in U^+(L)$, we have $\sigma\in SL_2(R)$ from the unitary properties and $\det(\sigma)=1$. It is well-known that $SL_2(R)$ is generated by $\sigma_{x,c}$ and $\sigma_{y,d}$ for all $c, d\in R$. Hence $U^+(L)=T(L)=SL_2(R)$.

Lemma 2.5. *Except when* $\dim L=n=3$ *and* $E=F_4$, *$T(L)$ is transitive on non-isotropic vectors in L of same length.*

Proof. We know that $U^+(L/\pi L)$ is generated by unitary transvections except when $n=3$ and $E=F_4$ (cf. [4]), so h_1 is a surjection from $T(L)$ to $U^+(L/\pi L)$. Let $x_1, x_2, \ldots, x_n$ be the orthonormal basis of L. Therefore from $N(S^*)=R^*$, it suffices to find a $\sigma\in T(L)$ which maps y to x_1 for arbitrary $y\in L$ with $\langle y, y\rangle=1$ and $x+y\in\pi L$. Put

$$y=\sum_i a_i x_i$$

with $a_1+1\in\pi S$ and $a_i\in\pi S$ for $i>1$. Note that γx_1+x_2 is isotropic for each $\gamma\in\Delta$. Then

$$\sigma_{\gamma x_1+x_2,c}(y)=[(1+c)a_1+ca_2\gamma]x_1+[a_2+c(\bar{\gamma}a_1+a_2)]x_2+\sum_{i>2}a_i x_i.$$

Put $a_1'=(1+c)a_1+ca_2\gamma$ and $a_2'=a_2+c(\bar{\gamma}a_1+a_2)$. Because of $\varphi(\bar{\gamma}a_1+a_2)=\varphi(\bar{\gamma})$ and $E=F\Sigma$, we may take suitable $\gamma\in\Sigma$ and $c\in\pi R$ such that $\mathrm{ord}_K(a_2')>\mathrm{ord}_K(a_2)$. If so, $\varphi(a_1')=\varphi(a_1)=1$. Therefore by repeated action of transvections $\sigma_{\gamma' x_1+x_2,c'}$ for suitable γ' and c', we may increase the order of a_2 to ∞, i.e. $a_2=0$. Now let M be the binary sublattice generated by x_1 and x_2. Then $L=M\perp N$. It is clear that $\sigma_{\gamma' x_1+x_2,c'}$ is a unitary transvection on M and leaves N fixed. Since $T(M)=SL_2(R)$ is a compact topological group, we may get $\rho\in T(M)$ that is the limit of the product of those $\sigma_{\gamma' x_1+x_2,c'}$. Let $\tau=\rho\perp 1_N$, then it is in $T(L)$. By calculation, we have $\sigma_{x_1+x_2}\sigma_{\tau(y)+x_2}\tau(y)=x_1$.

Remark. Here for the first time, we used the compactness of $U(L)$. Most of our results up to now could be stated and proved for general valuation rings and lattices.

Lemma 2.6. *If $E=F_4$ and $n=4$, then $U^+(L)=T(L)$.*

Proof. We have

$$L=(1)\perp(1)\perp(1)\perp(1)\cong\begin{pmatrix}1 & \\ 0 & 0\end{pmatrix}\perp\begin{pmatrix}1 & \\ 0 & 0\end{pmatrix}=H_1\perp H_2.$$

Let $H_i=Sx_i+Sy_i$. For any $\sigma\in U^+(L)$, there exists $\rho\in T(L)$ such that $\rho\sigma(x_1)+x_1\in\pi S$ by the surjection of h_1 on $T(L)$. Let $y=\rho\sigma(x_1)$, so $y=x_1+\sum a_i x_i+\sum b_i y_i$ with $a_i, b_i\in\pi S$. It is clear that $\sigma_{y+(1+\bar{a}_1)^{-1}y_1}(y)=(1+\bar{a}_1)^{-1}y_1$. We may find $z\in H_2$ such that $t=x_1+\bar{a}_1 y_1+z$ is isotropic. By a direct computation, we have $\sigma_t\sigma_{y+(1+\bar{a}_1)^{-1}y_1}(y)=(1+\bar{a}_1)^{-1}x_1+y_1+(1+\bar{a}_1)^{-1}z=v$ and $\sigma_{v+x_1}(v)=x_1$. Thus there is a $\tau\in T(L)$ such that $\tau\sigma(x_1)=x_1$. Let $\sigma_1=\tau\sigma$, then $\sigma_1(y)=\sum c_i x_i+\sum d_i y_i$ with $d_1=1$. Note that cx_1+dx_2 is isotropic if one of c and d is in S^*. We have $\sigma_{(\bar{c}_2+\bar{d}_2)x_1+x_2}\sigma_1(x_1)=x_1$ and $\sigma_{(\bar{c}_2+\bar{d}_2)x_1+x_2}\sigma_1(y_1)=[c_1+c_2(\bar{c}_2+\bar{d}_2)]x_1+y_1+d_2 y_2$. Similarly, after an action of $\sigma_{a x_1+y_2}$ for suitable $a\in S$, we may eliminate the coefficient of y_2, i.e. there exists $\tau_1\in T(L)$ such that $\tau_1\sigma_1=\rho_1\perp 1_{H_2}$ with $\rho_1\in U^+(H_1)$. By Lemma 2.4, we may conclude that $\sigma\in T(L)$.

Theorem B. *Suppose that L is a unimodular lattice over $E[[\pi]]$ of dimension $n>1$. Then every proper unitary transformation is the product of unitary transvections except when $n=3$ and $E=F_4$.*

Proof. It follows from Lemmas 2.4, 2.5, 2.6 and by induction.

Remark. $U(L)/U^+(L)\cong\Delta=\{a\in S\mid a\bar{a}=1\}$ where Δ is an abelian group under multiplication. This follows from the fact that every quasi-symmetry has determinant $\gamma\in\Delta$.

§ 3. Normal Subgroups

Assume that L is a unimodular hermitian lattice over $S=E[[\pi]]$ of dimension n, then so is $L/\pi^r L$ for all positive integers r. A vector x in $L/\pi^r L$ is called non-isotropic if $x\notin\pi L/\pi^r L$ and $\langle x,x\rangle_r\in(R/\pi^r R)^*$. Put $\Delta_r=\{a\in S/\pi^r S\mid a\bar{a}=1$ in $S/\pi^r S\}$ for $r>0$, it is clear that p_r is a surjection from Δ to Δ_r.

Lemma 3.1. (1) *$h_r: U(L)\to U(L/\pi^r L)$ is a surjection for all $r>0$.*

(2) *$h_r: U^+(L)\to U^+(L/\pi^r L)$ is a surjection for all $r>0$.*

Proof. Immitating all the proofs in § 2, we have $U(L/\pi^r L)=D(L/\pi^r L)$ except when $E=F_4$ and $n=2$. Furthermore, $|U(L/\pi^r L):D(L/\pi^r L)|=2$ for that exceptional case. Then it is clear to conclude all the assertions.

In the case $r=0$, $L/\pi^r L$ is a zero-space and $U(L/\pi^r L)=\{1\}$, so h_r is a trivial map.

Lemma 3.2. *The following are equivalent for a unimodular lattice L and $\sigma\in U(L/\pi^r L)$:*

(a) *$\sigma\varepsilon$ Center of $U(L/\pi^r L)$, $Z(L/\pi^r L)$.*

(b) *$\sigma(x)=\gamma x$ for some $\gamma\in\Delta_r$ and all $x\in L/\pi^r L$.*

(c) *$\sigma(x)=c_x x$ for all isotropic $x\in L/\pi^r L$, where c_x is a constant depending on x.*

Proof. (a)$\Rightarrow$(b): If $\sigma\in Z(L/\pi^r L)$, then $\sigma\tau_{x,\gamma}=\tau_{x,\gamma}\sigma$ for all quasi-symmetrices $\tau_{x,\gamma}$ on $L/\pi^r L$. It follows that $\sigma(x)=c_x x$ for all non-isotropic $x\in L/\pi^r L$, where c_x is a constant related to x. Let $x_1, x_2, \ldots, x_n$ be the orthonormal basis of $L/\pi^r L$ and let $\sigma(x_i)=c_i x_i$ for $c_i\in S$. We may take $\alpha\in S$ such that $x_i+\alpha x_j$ is non-isotropic then $\sigma(x_i+\alpha x_j)=\sigma(x_i)+\alpha\sigma(x_j)$, i.e. $d(x_i+\alpha x_j)=c_i x_i+\alpha c_j x_j$ so

$c_i=d=c_j$. Therefore $\sigma(x_i)=\gamma x_i$ for all i where $\gamma\in\Delta_r$. It is clear then that $\sigma(x)=\gamma x$ for all $x\in L/\pi^r L$. It is obvious (b)$\Rightarrow$(c).

(c)$\Rightarrow$(a): Let $\sigma(x)=c_x x$ for all isotropic $x\in L/\pi^r L$. We know that two vectors x and y are linearly dependent iff $x=ay$ for some $a\in S$. It is obvious that $c_{ax}=c_x$ for all isotropic x. Thus for two arbitrary linearly independent isotropic vectors x and y, there exists $\alpha\in S$ such that $x+\alpha y$ is isotropic, therefore $c_{x+\alpha y}=c_x=c_y$. Hence $\sigma(x)=cx$ for $c\in S$ and all isotropic $x\in L/\pi^r L$. Furthermore, since every vector is a linear combination of two isotropic ones, $\sigma(x)=cx$ for $c\in\Delta_r$ and all $x\in L/\pi^r L$; it is clear then $\sigma\in Z(L/\pi^r L)$.

Remark. If $\sigma\in$center of $U^+(L/\pi^r L)$, then $\sigma(x)=\gamma x$ for all $x\in L/\pi^r L$ such that $\gamma\in\Delta_r$ and $\gamma^n=1$. Thus $Z(L/\pi^r L)\cong\Delta_r$ and $Z(L/\pi^r L)\cap U^+(L/\pi^r L)\cong\{\gamma\in\Delta_r\mid\gamma^n=1\}$.

Definition 3.1. By *the r-th general congruence subgroup* $GU^+(n,r)$ *of* $U^+(L)$, we mean the group $GU^+(n,r)=h_r^{-1}(Z(L/\pi^r L))$. By the r-th special congruence subgroup $SU^+(n,r)$ of $U^+(L)$, we mean the normal subgroup of $U^+(L)$ generated by all the unitary transvections in $GU^+(n,r)$.

In particular, $GU^+(n,0)=U^+(L)$ and $SU^+(n,0)=T(L)$. We have the following inclusion relations:

$$GU^+(n,0)\supset GU^+(n,1)\supset GU^+(n,2)\supset\cdots$$

$$SU^+(n,0)\supset SU^+(n,1)\supset SU^+(n,2)\supset\cdots.$$

Then

$$\bigcap_{r=0}^{\infty} GU^+(n,r)=Z(L)$$

since $\pi^\infty=0$.

Let $\sigma\in U^+(L)$. The order $o(\sigma)$ of σ is the least integer r such that $\sigma\notin GU^+(n,r+1)$. A subgroup G of $U^+(L)$ is said to be of order r, i.e. $o(G)=r$, if r is the least integer such that $G\not\subset GU^+(n,r+1)$. We see that a quasi-symmetry $\tau_{x,\gamma}$ has order r iff $\mathrm{ord}_K(1+\gamma)=r$; whereas a unitary transvection $\sigma_{x,c}$ has order r iff $\mathrm{ord}_k(c)=r$.

Lemma 3.3. $SU^+(n,r)=$*kernel of* h_r, *for all* r *and* $n>1$. *Moreover,*

$$\bigcap_{r=0}^{\infty} SU^+(n,r)=\{1\}$$

and $GU^+(n,r)/SU^+(n,r)\cong\Delta_r$.

Proof. Let $K(n,r)=\ker h_r$. It is clear that $SU^+(n,r)\subset K(n,r)$. We will prove the opposite inclusion by induction on n.

$n=2$. We have $K(2,r)=SU^+(2,r)$ (cf. [6]).

Now we assume that $m\geqq 3$ and that the Lemma is true for all $n<m$. Take $\sigma\in K(m,r)$, then $\sigma(x_1)=x_1+\pi^r(a_1x_1+\cdots+a_mx_m)$ where $x_1,\ldots,x_m$ form an orthonormal basis of L. If $a_1\notin S^*$, we may take suitable $c\in R$ such that $\mathrm{ord}_k(c)=r$ and $\sigma_{x_1+x_2,c}\,\sigma(x_1)=x_1+\pi^r(b_1x_1+\cdots+b_mx_m)$ with $b_1\in S^*$. Note that $\sigma_{x_1+x_2,c}\in SU^+(m,r)$. Therefore, we may assume that a_1 is in S^*. Let $t=\pi^{-r}[\sigma(x_1)+x_1]$.

Case (1). If $a_1 \in R^*$, then t is isotropic. By a direct computation, we have $\sigma_{t,\pi^r a^{-1}}\sigma(x_1)=x_1$. Here $\sigma_{t,\pi^r a^{-1}} \in SU^+(m,r)$.

Case (2). Assume $a_1 \in S^*$ but $a_1 \notin R$. If $\langle t,t\rangle = \pi^{-r}(a_1+\bar{a}_1) \notin R^*$. We may take $\alpha = 1+\pi^r\delta \in \Delta$ with $\delta \in S^*$. It is clear that $\tau_{x_2,\bar{\alpha}}\tau_{x_1,\alpha}$ is in $K(2,r)=SU^+(2,r)$. Hence it is in $SU^+(m,r)$. Let $\tau=\tau_{x_2,\bar{\alpha}}\tau_{x_1,\alpha}\sigma$, we have $\tau(x_1)=(1+\pi^r a_1)\alpha x_1 + \pi^r(a_2\bar{\alpha}x_2+a_3x_3+\cdots+a_mx_m)$. Put $t'=\pi^{-r}[\tau(x_1)+x_1]$. Then $\langle t',t'\rangle = \delta\bar{\delta}+a_1\delta+\bar{a}_1\bar{\delta}+\pi^{-r}(a_1+\bar{a}_1)$, so $\varphi(\langle t',t'\rangle)=\varphi(\delta\bar{\delta})\neq 0$, i.e. t' is non-isotropic in L. Hence we may assume that $t=\pi^{-r}[\sigma(x_1)+x_1]$ is non-isotropic. By a straightforward computation, we have $\tau_{t,\gamma}\sigma(x_1)=x_1$, for $\gamma=\bar{a}_1/a_1$. Then

$$\mathrm{ord}_K(1+\gamma)=\mathrm{ord}_K\left(\frac{a_1+\bar{a}_1}{a_1}\right)\geqq r.$$

Since $m\geqq 3$, it is possible to find a non-isotropic y in L such that y is orthogonal to t and x_1, so $\tau_{y,\bar{\gamma}}\tau_{t,\gamma}\sigma(x_1)=x_1$.

In both cases, we have found $\rho\in SU^+(m,r)$ such that $\rho\sigma = 1\perp\sigma'$ with $\sigma'\in K(m-1,r)$. Thus, by induction we get $\sigma\in SU^+(m,r)$.

The second statement is a trivial consequence.

Lemma 3.4. *If $n>2$, then $U^+(L)$ is transitive on isotropic vectors in L.*

Proof. Take arbitrary two isotropic vectors x and y in L.

Case (1). If $\langle x\,y\rangle = a\in R^*$, then $\sigma_{x+y,a^{-1}}(x)=y$.

Case (2). If $\langle x,y\rangle = a\in S^*$, but $a\notin R$, then $\langle x,ay\rangle = a\bar{a}\in R^*$ and we can apply Case (1) to find $\tau\in T(L)$ such that $\tau(x)=ay$. We may decompose the lattice

$$L=\begin{pmatrix}1\\0\end{pmatrix}\perp(1)\perp\cdots\perp(1)$$

by taking y to be the first basis element, that is $L=(Sy+Sy_2)\perp Sy_3\perp\cdots\perp Sy_n$. There is a $\rho\in U^+(L)$ such that $\rho(y)=a^{-1}y$, $\rho(y_2)=\bar{a}y_2$, $\rho(y_3)=a\bar{a}^{-1}y_3$ and $\rho(y_i)=y_i$ for all $i>3$. Then it is clear that $\rho\tau(x)=y$ and $\rho\tau\in U^+(L)$.

Case (3). Assume $\langle x,y\rangle = a\in\pi S$, then with the same decomposition used in Case (2), $x=by+ay_2+c_3y_3+\cdots+c_ny_n$. If $b\in S^*$, σ_{y+y_2} will change the role of y and y_2. This reduces to Case (2). Therefore we may assume $b\notin S^*$ and $c_3\in S^*$. $z=dy+y_2+y_3$ is isotropic, if we take $d\in S^*$ with $d+\bar{d}=1$. Then $\sigma_z(x)=x+(a\bar{d}+b+c_3)(dy+y_2+y_3)$ so that $\langle\sigma_z(x),y\rangle = a+a\bar{d}+b+c_3\in S^*$. This again reduces to Case (2).

Lemma 3.5. *Suppose that $n>2$ and σ_1,σ_2 are two unitary transvections. Then $o(\sigma_1)=o(\sigma_2)$ iff σ_1 is conjugate to σ_2 in $U^+(L)$.*

Proof. Necessary. This is obvious, because $SU^+(n,r)$ is normal in $U^+(L)$ for all r.

Sufficient. Assume that σ_i is the unitary transvection with respect to isotropic vector z_i and scalar c_i in R for $i=1$ and 2, i.e. $\sigma_i=\sigma_{z_i,c_i}$. Thus $o(\sigma_1)=o(\sigma_2)$ implies $c_2c_1^{-1}\in R^*$, i.e. $c_2c_1^{-1}=a\bar{a}$ for some $a\in S^*$. By Lemma 3.4, there is a $\tau\in U^+(L)$ such that $\tau(z_1)=az_2$, then $\tau\sigma_1\tau^{-1}=\sigma_2$.

Lemma 3.6. *Suppose that $n>2$ and G is a normal subgroup of $U^+(L)$. If G contains a unitary transvection of order r, then $G\supset SU^+(n, r)$.*

Proof. It follows from Lemma 3.5, that $\sigma_{x,c}\sigma_{x,c+\pi^j}=\sigma_{x,\pi^j}\in G$ for all $j>r$, i.e. G contains all the unitary transvections of order $\geqq r$. Hence $G\supset SU^+(n, r)$

Let $R=F[[\pi]]$ with F a finite field of characteristic 2 and $\pi^\infty=0$. We write $\wp(R)=\{a^2+a\,|\,a\in R\}$ and $\wp(F)=\{a^2+a\,|\,a\in F\}$. It is known that $\wp(F)$ is of index 2 in F. If there is no confusion, we write $\mathrm{ord}(x)$ instead of $\mathrm{ord}_k(x)$ for $x\in R$.

Lemma 3.7. *$b\in\wp(R)$ iff $\varphi(b)\in\wp(F)$.*

(A special case of HENSEL's Lemma.)

Proof. Let

$$b=\sum_{i=0}^{\infty}b_i\pi^i \quad\text{and}\quad a=\sum_{i=0}^{\infty}a_i\pi^i.$$

Then $b=a^2+a$ iff $b_{2i}=a_{2i}+a_i^2$ and $b_{2i+1}=a_{2i+1}$ for all $i\geqq 0$. If $b\in\wp(R)$, then $b_0=a_0^2+a_0\in\wp(F)$. If $\varphi(b)\in\wp(F)$, then a_i is solvable in F for all $i>0$, so $b\in\wp(R)$.

Lemma 3.8. *For $a, b, c\in R$ with $\mathrm{ord}(a)\leqq\mathrm{ord}(b)$, $\mathrm{ord}(a)\leqq\mathrm{ord}(c)$ and*

$$\frac{ac}{b^2}\in R,$$

the equation $ax^2+bx+c=0$ is solvable in R iff

$$\frac{ac}{b^2}\in\wp(R).$$

Furthermore, if

$$\varphi\left(\frac{ac}{b^2}\right)=0,$$

then x can be found in πR.

Proof. Necessary. If

$$\frac{ac}{b^2}=d^2+d$$

for some $d\in R$, then

$$x=\frac{bd}{a}\in R,$$

is the solution.

Sufficient. Suppose $ax^2+bx+c=0$ for some $x\in R$, then

$$\left(\frac{ax}{b}\right)^2+\frac{ax}{b}+\frac{ac}{b^2}=0.$$

Since

$$\frac{ac}{b^2}\in R, \quad\text{we have}\quad \frac{ax}{b}\in R, \quad\text{i.e.}\quad \frac{ac}{b^2}\in\wp(R).$$

If

$$\varphi\left(\frac{ac}{b^2}\right)=0, \quad \text{i.e.} \quad \varphi\left(\frac{ac}{b^2}\right)\in\wp(F),$$

then it is easy to see that x is solvable in πR by Lemma 3.7. Q.e.d.

In case $n=\dim L=2$, we know that $U^+(L)$ is the special linear group over R, i.e. $SL_2(R)$. Thus if $\rho\in U^+(L)$ is of order r, then ρ may be expressed as

$$\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}$$

with respect to the hyperbolic pair basis such that

$$u\in R^*, \quad x,y,z\in R, \quad \min\{\mathrm{ord}(x),\ \mathrm{ord}(y),\ \mathrm{ord}(z)\}=r \quad \text{and} \quad u(u+z)+xy=1.$$

If $r>0$, we have $\varphi(u)=1$. The normal subgroup of $U^+(L)$ generated by ρ consists of the following transformations for arbitrary

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\in SL_2(R).$$

(*)
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}=\begin{pmatrix} u+acx+bdy+bcz & a^2x+b^2y+abz \\ c^2x+d^2y+cdz & u+acx+bdy+adz \end{pmatrix}$$
$$=\begin{pmatrix} u' & x' \\ y' & u'+z \end{pmatrix}=\rho'.$$

(**)
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}$$
$$=\begin{pmatrix} u^2+(acx+bdy+bcz)u+a^2xy+b^2y^2+abyz & [(1+a^2)x+b^2y+abz]u+acx^2+abz^2+bdxy+(a^2+bc)xz+b^2yz \\ [c^2x+(1+d^2)y+cdz]u+bdy^2+adyz+acxy & u^2+(acx+bdy+bcz)u+(ac+cd)xz+d^2xy+bdyz+c^2x^2+adz^2 \end{pmatrix}$$
$$=\begin{pmatrix} u'' & x'' \\ y'' & u''+z'' \end{pmatrix}=\rho''$$

with

$$z''=(a^2+d^2)xy+(ab+bd)yz+(ac+cd)xz+c^2x^2+b^2y^2+adz^2.$$

(***)
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}^{-1}$$
$$=\begin{pmatrix} u^2+(acx+bdy+adz)u+a^2xy+(ab+bd)yz+acxz+bcz^2+b^2y^2 & [(1+a^2)x+b^2y+abz]u+acx^2+bcxz+bdxy \\ [c^2x+(1+d^2)y+cdz]u+c^2xz+acxy+(ad+d^2)yz+bdy^2+cdz^2 & u^2+(acx+bdy+adz)u+c^2x^2+d^2xy+cdxz \end{pmatrix}$$
$$=\begin{pmatrix} u''' & x''' \\ y''' & u'''+z''' \end{pmatrix}=\rho'''$$

with

$$z''' = (a^2+d^2)\,x\,y+(a\,b+b\,d)\,y\,z+(a\,c+c\,d)\,x\,z+c^2\,x^2+b^2\,y^2+b\,c\,z^2.$$

Lemma 3.9. *Suppose that G is a normal subgroup of $SL_2(R)$, containing a transformation of order r, i.e.*

$$\rho=\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\in G,$$

when $r>0$ and $F\neq F_2$. Then there exist

$$\rho_1=\begin{pmatrix} u_1 & x_1 \\ y_1 & u_1+z_1 \end{pmatrix}\in G$$

with $\mathrm{ord}(x_1)=r$ *and* $\mathrm{ord}(y_1)=\mathrm{ord}(z_1)=2r$.

Proof. First of all we may assume $\mathrm{ord}(x)=r$. If not, either $\mathrm{ord}(y)=r$ or $\mathrm{ord}(z)=r$. If $\mathrm{ord}(y)=r$, let $a=d=0$, $b=c=1$ in (*); then we change the role of x and y. If $\mathrm{ord}(x)>r$ and $\mathrm{ord}(y)>r$, but $\mathrm{ord}(z)=r$, let a, $b\in R^*$ in (*), we have $\mathrm{ord}(x')=r$. Now we assume $\mathrm{ord}(x)=r$. Thus we can take $c\in R^*$ such that $\mathrm{ord}(cx+z)=\mathrm{ord}(x)$ since $F\neq F_2$. Then let $a=d=1$ and $b=0$ in (**); we have $\mathrm{ord}(x'')=2r$, $\mathrm{ord}(y'')=r$ and $\mathrm{ord}(z'')=2r$. After changing the role of x'' and y'', we get $\rho_1\in G$.

Lemma 3.10. *Under the same assumptions of Lemma 3.9, then there exists*

$$\rho_2=\begin{pmatrix} u_2 & x_2 \\ y_2 & u_2+z_2 \end{pmatrix}\in G$$

with $\mathrm{ord}(y_2)=r$ *and* $\mathrm{ord}(x_2)+\mathrm{ord}(y_2)>2\,\mathrm{ord}(z_2)$.

Proof. By Lemma 3.9, we may assume

$$\rho=\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\in G$$

with $\mathrm{ord}(x)=r$ and $\mathrm{ord}(y)=\mathrm{ord}(z)=2r$.

(1) In (**), if $a=1$, $c\in R^*$ and $b^2yu+bzu+cx^2=0$ for some $b\in R$, we have $\mathrm{ord}(x'')\geqq 3r$, $\mathrm{ord}(y'')=r$ and $\mathrm{ord}(z'')=2r$. From Lemmas 3.7 and 3.8 we get that $b^2yu+bzu+cx^2=0$ is solvable iff

$$\varphi\left(\frac{c\,x^2\,y}{z^2}\right)\in\wp(F).$$

Since

$$\frac{x^2\,y}{z^2}\in R^* \quad\text{and}\quad F\neq F_2,$$

we may find $c\in R^*$ such that

$$\varphi\left(\frac{c\,x^2\,y}{z^2}\right)\in\wp(F).$$

Thus b is solvable.

(2) If $\mathrm{ord}(x'')>3r$, then $\mathrm{ord}(x'')+\mathrm{ord}(y'')>2\,\mathrm{ord}(z'')$. Therefore we only have to consider the case that $\mathrm{ord}(x)=3r$, $\mathrm{ord}(y)=r$ and $\mathrm{ord}(z)=2r$. We know that $b^2(yu+yz)+bazu+(1+a^2)ux=0$ for some $b\in R$ iff

$$\varphi\left[\frac{(1+a^2)\,u\,x(y\,u+y\,z)}{a^2\,z^2\,u^2}\right]=\varphi\left(\frac{(1+a^{-2})\,x\,y}{z^2}\right)\in\not\wp(F).$$

Since

$$\frac{x\,y}{z^2}\in R^* \quad \text{and} \quad F\neq F_2,$$

we may take $a\in R$ such that

$$\varphi\left(\frac{(1+a^{-2})\,x\,y}{z^2}\right)\neq 0 \qquad \text{in } \not\wp(F).$$

Then b is solvable in R^* from Lemmas 3.7 and 3.8. In (**), take a and b as required above and $d\in\pi R$. Then c has to be in R^*, so $\mathrm{ord}(x'')\geqq 4r$, $\mathrm{ord}(y'')=r$, $\mathrm{ord}(z'')=2r$. Hence the conclusion follows.

Lemma 3.11. *Let G be a normal subgroup of $U^+(L)$ with $\dim L=2$. If a unitary transformation of order $r>0$ is in G, then G contains a unitary transvection of order r, when $F\neq F_2$.*

Proof. By Lemma 3.10, we may assume

$$\rho=\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\in G$$

with $\mathrm{ord}(y)=r$, $\mathrm{ord}(z)=2r$ and $\mathrm{ord}(x)+\mathrm{ord}(y)>2\,\mathrm{ord}(z)$.

(1) Since

$$\varphi\left(\frac{x\,y}{z^2}\right)=0,$$

$b^2y+abz+a^2x=0$ is solvable in πR by Lemma 3.8. Therefore in (*), take this $b\in\pi R$, and $a, d\in R^*$, then $x'=0$ and $\mathrm{ord}(y')=r$.

(2) We may assume then

$$\rho=\begin{pmatrix} u & 0 \\ y & u+z \end{pmatrix}\in G$$

with $\mathrm{ord}(y)=r$. In (***), let $b=c=0$ and $a\in R^*$ but $\varphi(a)\neq 1$ and $d=a^{-1}$, then $x'''=0$, $z'''=0$ and $y'''=(1+a^{-2})\,y(u+z)$. Therefore $\mathrm{ord}(y''')=r$, i.e. ρ''' is a unitary transvection of order r.

Lemma 3.12. *Suppose that $n=2$ and $F\neq F_2$. If a normal subgroup G of $U^+(L)$ contains a unitary transformation of order 0, then G contains a unitary transvection of order 0.*

Proof. Let

$$\rho=\begin{pmatrix} u & x \\ y & u+z \end{pmatrix}\in G$$

with $u(u+z)+xy=1$ and $u\in R^*$, $x, y, z\in R$. Since $o(\rho)=0$, at least one of $\mathrm{ord}(x)$, $\mathrm{ord}(y)$, $\mathrm{ord}(z)$ is 0. We may assume $\mathrm{ord}(z)=0$. If not, let $\mathrm{ord}(x)=0$; we may take $a=c=d=1$ and $b=0$ in (***), then $z'''=x^2$, i.e. $\mathrm{ord}(z''')=0$.

Case (1). If $\mathrm{ord}(z)=0$ and either $\mathrm{ord}(x)$ or $\mathrm{ord}(y)>0$, then

$$\varphi\left(\frac{xy}{z^2}\right)=0,$$

so $b^2y+baz+a^2x=0$ is solvable in πR. In (*), take the so determined $b\in\pi R$ and $a=d=1$, $c=0$. Then we have

$$\rho'=\begin{pmatrix} u+by & 0 \\ y & u+by+z \end{pmatrix}\in G.$$

Case (2). If $\mathrm{ord}(x)=\mathrm{ord}(y)=\mathrm{ord}(z)=0$, in (***), let $b=0$,

$$c=(a+a^{-1})\frac{u}{x}$$

and $a\in R^*$ but $\varphi(a)\neq 1$, $d=a^{-1}$. Then $\rho'''\in G$ with $x'''=0$, $z'''=a^2+a^{-2}\in R^*$.

From the above two cases, we may assume that

$$\rho=\begin{pmatrix} u & 0 \\ y & u+z \end{pmatrix}\in G$$

with $\mathrm{ord}(z)=0$. In (***), let $b=0$, $d=a^{-1}$ and $a\in R^*$ where $\varphi(a)\neq 1$. Then $u'''=1$, $x'''=0$, $z'''=0$ and $y'''=(u+z)[(1+a^{-2})y+ca^{-1}z]$. Therefore we take $c\in R$ such that $(1+a^{-2})y+ca^{-1}z\in R^*$. Then $\mathrm{ord}(y''')=0$.

Lemma 3.13. *Let G be a normal subgroup of $U^+(L)$ with $\dim L=n>2$ and $o(G)=r$. Then G contains a unitary transvection of order r when $F\neq F_2$.*

Proof. There exists $\rho\in G$ such that $o(\rho)=r$, i.e. $\rho\in GU^+(n, r)$. By Lemma 3.2, we may find isotropic $x\in L$ such that $\rho(x)=\gamma x+\pi^r y$ with $\gamma\in\Delta$ and y maximal in L.

(1) If $\langle x, y\rangle\in S^*$, then x and y form a basis of a plane P. Let

$$\tau=\sigma_{x,c}\rho\sigma_{x,c}^{-1}\rho^{-1},$$

then $\tau\in G$ and $\tau(z)=z$ for all z such that $\langle z, P\rangle=0$, i.e. τ is a unitary transformation on P. We have

$$\begin{aligned}\tau(x)&=[1+c\gamma\pi^r\langle x, y\rangle+c^2\pi^{2r}\langle x, y\rangle\langle y, x\rangle]x\\&\quad+c\pi^{2r}\langle x, y\rangle y=ax+by.\\ \tau(y)&=[c\gamma\pi^r\langle y, y\rangle+c^2\bar{\gamma}\pi^r\langle y, x\rangle^2+c^2\pi^{2r}\langle y, y\rangle\langle y, x\rangle]x\\&\quad+[1+c\bar{\gamma}\pi^r\langle y, x\rangle+c\pi^{2r}\langle y, y\rangle]y=a'x+b'y.\end{aligned}$$

If $\langle y, y\rangle\in\pi R$, let $c=1$, then $\mathrm{ord}_K(a')=r$.

If $\langle y, y\rangle\in R^*$, take $c\in R^*$, such that $\mathrm{ord}_K(a')=r$. It is possible since $F\neq F_2$. Therefore τ is a unitary transformation of order r on P. By Lemmas 3.11 and 3.12, we get our conclusion.

(2) It suffices to find an isotropic $x\in L$ such that $\mathrm{ord}_K(\langle\rho(x), x\rangle)=r$. If so, $\rho(x)=\gamma x+\pi^r y$ with $\langle x, y\rangle\in S^*$, thus the assertion is proven by (1). Now we assume that $\mathrm{ord}_K(\langle\rho(x), x\rangle)>r$ for all isotropic $x\in L$. Consider an arbitrary isotropic vector x, let $\rho(x)=\gamma x+\pi^r y$ with $y\in L$; then $\langle\rho(x), x\rangle=\pi^r\langle y, x\rangle\in \pi^{r+1}S$, so $\langle x, y\rangle\in\pi S$. There exists an isotropic $z\in L$ such that $\langle x, z\rangle=1$ and y is lying on the plane spanned by x and z. Let $y=ax+bz$ with $a, b\in S$; since $\langle x, y\rangle\in\pi S$, we have $b\in\pi S$. Then $\rho(x)=(\gamma+\pi^r a)x+\pi^r bz$, so $p_{r+1}\rho(x)=p_{r+1}(\gamma+\pi^r a)\,p_{r+1}(x)=h_{r+1}\rho\,p_{r+1}(x)$. Therefore $(h_{r+1}\rho)\,(p_{r+1}x)=c_x\,p_{r+1}(x)$, for all isotropic $x\in L$. It follows from Lemma 3.2, that we have $h_{r+1}\rho\in Z(L/\pi^{r+1}L)$ i.e. $o(\rho)\geqq r+1$. It is a contradiction.

Theorem C. *If L is a unimodular hermitian lattice over $S=E[[\pi]]$ of dimension $L=n>2$ and G is a normal subgroup of order r in $U^+(L)$, then $GU^+(n, r)\supset G\supset SU^+(n, r)$, when $E\neq F_4$.*

Proof. It follows from Lemmas 3.6 and 3.13.

Remark. In the case $n=2$ and $E\neq F_4$, it is easy to see that $\sigma_{x,c}$ and $\sigma_{x,d}$ are conjugate in $U(L)$ iff $cd^{-1}\in R^*$; whereas conjugate in $U^+(L)$ iff $cd^{-1}\in R^{*2}=\{a^2\mid a\in R^*\}$. Therefore if G is a normal subgroup of order r in $U^+(L)$ and also normal under $U(L)$, then we may use the same argument as above to show that $G\supset SU^+(2, r)$. However if G is only normal under $U^+(L)$, then we can not get such a conclusion.

§ 4. Ramified Case

From now on, K is a ramified extension of $k=F[[\pi]]$. Thus we may assume $K=F\langle\langle\Pi\rangle\rangle$ with $\pi=N(\Pi)$. $S=F[[\Pi]]$ and $R=F[[\pi]]$ are the integral rings of K and k respectively. We will use the same definitions and notations as in §2.

Let L be a hermitian lattice over S. $Q(L)=\{\langle x, x\rangle\mid x\in L\}$ is called the norm of L. $q(L)=\{R\langle x, x\rangle\mid x\in L\}$ is called the norm group of L. It is clear that $q(L)$ is an R-module of finite type in k.

In the ramified case, we have $T(\Pi)=\xi\pi^s$ with $s>0$ and $\xi\in R^*$. $R^*/N(S^*)$ is a cyclic group of order 2, thus we fix the choice of $\lambda\in F$ such that $1+\lambda\pi^{2s-1}$ determine the non-trivial coset of $R^*/N(S^*)$. We know that a unimodular hermitian lattice L over $S=F[[\Pi]]$ of dimension n has the following four distinct normal types (cf. [10]).

(i) $L=\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp\cdots\perp\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp(1)$.

(ii) $L=\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp\cdots\perp\begin{pmatrix}1\\0\ 0\end{pmatrix}$.

(iii) $L=\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp\cdots\perp\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp\begin{pmatrix}1\\\pi^j\ \lambda\pi^{2s-1-j}\end{pmatrix}$, $\quad 0<j<s$.

(iv) $L=\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp\cdots\perp\begin{pmatrix}1\\0\ 0\end{pmatrix}\perp(1)\perp(1+\lambda\pi^{2s-1})$.

Let $V=L/\Pi L$ be the residue space over F of L. We will use the same form $\langle\ ,\ \rangle$ on V. Let $G(V)=\{v\in\mathrm{Aut}(V)\mid\langle v(x), v(y)\rangle=\langle x, y\rangle$ for all $x, y\in V\}$. $\langle\ ,\ \rangle$ on types (ii) and (iii) is an alternating form, thus $G(V)=\mathrm{Sp}(V)$. On types (i) and (iv), $\langle\ ,\ \rangle$ is a symmetric bilinear form, then the structure of $G(V)$ is done by DIEUDONNÉ [5].

Let $L=H_1\perp\cdots\perp H_r\perp M$ where

$$M=(1),\ \begin{pmatrix}1&\\0&0\end{pmatrix},\ \begin{pmatrix}1&\\\pi^j&\lambda\pi^{2s-1-j}\end{pmatrix}\quad\text{or}\quad(1)\perp(1+\lambda\pi^{2s-1})$$

and H_i is the hyperbolic plane for all i. We will fix a basis so that

$$H_i=\begin{pmatrix}1&\\0&0\end{pmatrix}=Sx_i+Sy_i$$

for each i.

Lemma 4.1. *Suppose that x and y are isotropic in L and $a=\langle x, y\rangle\in R^*$. Then $\sigma_{x+y,a^{-1}}(x)=y$.*

Proof. Obvious.

Lemma 4.2. *If $L=H_1\perp\cdots\perp H_r$ is a hyperbolic lattice, then $T(L)$ is transitive on the isotropic basis elements.*

Proof. It is clear that $\sigma_{x_i+y_i}$ exchanges x_i and y_i. Therefore without loss of generality, it suffices to show that there is a $\sigma\in T(L)$ such that $\sigma(x_1)=x_2$. We note that $\sigma_{x_2+y_1+y_2}\,\sigma_{x_1+y_1+y_2}(x_1)=x_2$.

Lemma 4.3. *Suppose that $L=H_1\perp\cdots\perp H_r\perp M$ and $r>0$. If x is isotropic in L and $a\in S^*$, then there is a $\sigma\in T(L)$ such that $\sigma(ax)=x$.*

Proof. We may assume $x=x_1$. For the last three types, we use the fact $Q(M)=q(M)$ (cf. [10]) to find $z\in M$ such that $\langle z, z\rangle=a+\bar{a}$. In case $M=(1)$, there is $z\in M$ such that $\langle z, z\rangle=c(a+\bar{a})$ where $c=1$ or $1+\lambda\pi^{2s-1}$ according to $a+\bar{a}\in N(S)$ or not. Let $t=(\bar{a}+1)x_1+cy_1+z$ where $c=1+\lambda\pi^{2s-1}$ for $a+\bar{a}\notin N(S)$ in type (i), otherwise $c=1$. t is then isotropic and $\sigma_{t,c^{-1}}(ax_1)=a\bar{a}x_1+cay_1+az$ so that $\langle\sigma_{t,c^{-1}}(ax_1), y_1\rangle=a\bar{a}\in R^*$. Using Lemmas 4.1 and 4.2, we may find $\sigma\in T(L)$ transforming ax_1 to x_1.

Lemma 4.4. *Suppose that $L=H_1\perp\cdots\perp H_r\perp M$ and $r>0$. For arbitrary $\sigma\in U(L)$, there is a $\rho\in T(L)$ such that $\rho\sigma$ leaves the first basis element invariant.*

Proof. Put $\sigma(x_1)=\sum a_ix_i+\sum b_iy_i+z$ with $z\in M$.

Case (1). If one of a_i or b_i is in S^*, we may apply Lemma 4.2 and assume $b_1\in S^*$. $\langle\sigma(x_1), b_1x_1\rangle=b_1\bar{b}_1\in R^*$, so there is a $\tau\in T(L)$ such that $\tau\sigma(x_1)=b_1x_1$. Then the assertion follows from Lemma 4.3.

Case (2). If none of a_i and b_i are in S^*, then z is maximal in M, i.e. $z\notin\Pi M$. From $\langle\sigma(x_1),\sigma(x_1)\rangle=0=\sum(\bar{a}_ib_i+a_i\bar{b}_i)+\langle z, z\rangle$, we have $\mathrm{ord}_k(\langle z, z\rangle)\geqq s$. Now we consider each type separately.

(i) $M=(1)$. If z is maximal in M, then $\langle z, z\rangle \in R^*$, this contradicts the statement $\mathrm{ord}_k(\langle z, z\rangle)\geqq s>0$.

(ii) $M=\begin{pmatrix}1\\0\ 0\end{pmatrix}=Su+Sv$ with $z=cu+dv$.

One of c and d must be in S^*. Then Case (1) applies in this situation.

(iii) $M=\begin{pmatrix}1\\ \pi^j\ \ \lambda\pi^{2s-1-j}\end{pmatrix}=Su+Sv$ for $0<j<s$.

It is clear that $j<s\leqq 2s-1-j$. Put $z=cu+dv$, so $\langle z,z\rangle=c\bar{c}\pi^j+d\bar{d}\lambda\pi^{2s-1-j}+\bar{c}d+c\bar{d}$. We conclude that $c\in \Pi S$ and $d\in S^*$ from the fact that $\mathrm{ord}_k(\langle z,z\rangle)\geqq s$ and $z\notin \Pi M$. Now let $\sigma(y_1)=\sum a_i' x_i+\sum b_i' y_i+c'u+d'v$. If one of a_i' or b_i' is in S^*, by the symmetry of x_1 and y_1, we are done by Case (1). Therefore, we may assume that $a_i', b_i', c'\in \Pi S$ for all i and $d'\in S^*$. However

$$\langle\sigma(x_1),\ \sigma(y_1)\rangle=\sum a_i\bar{b}_i'+\sum\bar{a}_i' b_i+c\bar{c}'\pi^j+d\bar{d}'\lambda\pi^{2s-1-j}+c\bar{d}'+\bar{c}'\ d\in \Pi S$$

that contradicts to $\langle\sigma(x_1),\sigma(y_1)\rangle=\langle x_1, y_1\rangle=1$.

(iv) $M=(1)\perp(1+\lambda\pi^{2s-1})=Su+Sv$. Let $z=cu+dv$. Using the same techniques as in (iii), we either lead to a contradiction or reduce to Case (1).

Lemma 4.5. *Suppose that $r>1$ and $L=H_1\perp\cdots\perp H_r\perp N$ where N is*

$$(1),\ 0,\ \begin{pmatrix}1\\ \pi^j\ \ \lambda\pi^{2s-1-j}\end{pmatrix}\quad\text{or}\quad (1)\perp(1+\lambda\pi^{2s-1}).$$

Then for any $\sigma\in U(L)$, we can find a $\rho\in T(L)$ such that $\rho\sigma$ is the identity map on H_1.

Proof. Using Lemma 4.4, we may assume $\sigma(x_1)=x_1$, it follows that $\sigma(y_1)=\sum a_i x_i+\sum b_i y_i+z$ with $z\in N$ and $b_1=1$. We may also assume that the coefficient of x_1 in $\sigma(y_1)$ is in S^*, for otherwise, $\sigma_{x_1}\sigma(x_1)=x_1$ and $\sigma_{x_1}\sigma(y_1)$ will have the property asserted.

Case (1). If $a_1\in R^*$, then $\sigma_{\sigma(y_1)+y_1, a_1^{-1}}\sigma$ leaves x_1 and y_1 invariant.

Case (2). If $a_1\in S^*$, but $a_1\notin R$, we may assume that the coefficient of x_2 in $\sigma(y_1)$, i.e. a_2, is in S^*. Otherwise $\sigma_{x_1+x_2}\sigma(y_1)$ or $\sigma_{x_2+y_2}\sigma(y_1)$ will have that property; furthermore, $\sigma_{x_1+x_2}\sigma$ or $\sigma_{x_2+y_2}\sigma$ does not move x_1. Thus by computation, $\sigma_{x_1+a_1\bar{a}_2^{-1}y_2}\sigma(x_1)=x_1$ and the coefficient of x_1 in $\sigma_{x_1+a_1\bar{a}_2^{-1}y_2}\sigma(y_1)$ is $1+a_1+\bar{a}_1\in R^*$, so we come back to Case (1).

Having proved Lemma 4.5, we only need to investigate the structure of the unitary groups of the lower dimensional unimodular lattice.

Lemma 4.6. *Suppose that*

$$L=\begin{pmatrix}1\\0\ 0\end{pmatrix}=Sx+Sy.$$

Then

(1) $A_S=\{\rho_a \mid \rho_a(x)=ax$ *and* $\rho_a(y)=\bar{a}^{-1}y$ *with* $a\in S^*\}$ *is a subgroup of* $U(L)$. *Furthermore,* $T(L)\cap A_S=A_R=\{\rho_a \mid a\in R^*\}$.

(2) $U(L)=T(L)\cdot A_S$.

(3) $U^+(L)=T(L)$. $U(L)/T(L)\cong S^*/R^*$.

Proof. (1) Obviously, ρ_a is in $U(L)$ and A_S is a multiplicative group isomorphic to S^*. Because of $\det(\rho_a)=a/\bar{a}\neq 1$ for any $a\in S^*$ but $a\notin R$, we conclude that $\rho_a\notin T(L)$. If $a\neq 1$ in R^*, it is clear that $\rho_a=\sigma_{\rho_a(x)+y,\,\bar{a}^{-1}\sigma_{x+y}}\in T(L)$.

(2) Let an arbitrary $\sigma\in U(L)$ be given. Put $\sigma(x)=ax+by$. We may assume $b\in S^*$ by the symmetry of x and y, so $\langle\sigma(x), by\rangle=b\bar{b}\in R^*$. Then we use Lemma 4.1 and (1) to get the conclusion.

(3) The statements follow from (1) and (2).

Theorem D. *Suppose that* L *is a hyperbolic lattice. Then* $U(L)=T(L)\cdot A_S$ *where* A_S *acts on a fixed hyperbolic plane as in Lemma 4.6.* $U^+(L)=T(L)$. *Finally* $U(L)/T(L)=S^*/R^*$.

Proof. The assertions follow from Lemmas 4.5 and 4.6.

Lemma 4.7. *Let*

$$L=\begin{pmatrix}1&\\0&0\end{pmatrix}\perp(1).$$

Then

(1) $U(L)=T(L)\cdot(P,\Delta)$, *where* $P=\{c\in S \mid c\bar{c}\in T(S)\}=\Pi^s S$ *and* $\Delta=\{a\in S \mid a\bar{a}=1\}$.

(2) (P,Δ) *is the group under multiplication with* $(0,1)$ *as the identity and* $(c_1,\gamma_1)(c_2,\gamma_2)=(c_1+c_2\gamma_1,\gamma_1\gamma_2)$.

(3) $U^+(L)=T(L)\cdot(P,1)$. *Moreover, if* $F\neq F_2$, *then* $U^+(L)=T(L)$.

Proof. (1) Let

$$L=\begin{pmatrix}1&\\0&0\end{pmatrix}\perp(1)=(Sx+Sy)\perp Sz.$$

For any $\sigma\in U(L)$, we may assume that $\sigma(x)=x$ according to Lemma 4.4. Then $\sigma(y)=ax+y+cz$ and $\sigma(z)=\gamma(z+\bar{c}x)$ with $a+\bar{a}+c\bar{c}=0$ and $\gamma\bar{\gamma}=1$. Suppose that there is another $\sigma'\in U(L)$ such that $\sigma'(x)=x$, $\sigma'(y)=a'x+y+cz$ and $\sigma'(z)=\gamma(z+\bar{c}x)$. It is clear that $a+a'\in R$ and $\sigma_{x,\,a+a'}\sigma=\sigma'$. Therefore such a σ is determined by γ and c. We denote it by (c,γ).

(2) If $\sigma_i=(c_i,\gamma_i)$ for $i=1,2$, then $\sigma_i(x)=x$, $\sigma_i(y)=a_ix+y+c_iz$, $\sigma_i(z)=\gamma_i(z+\bar{c}_ix)$. By a straightforward calculation, we have $\sigma_1\sigma_2(x)=x$, $\sigma_1\sigma_2(y)=(a_1+a_2+\bar{c}_1c_2\gamma_1)x+y+(c_1+c_2\gamma_1)z$ and $\sigma_1\sigma_2(z)=\gamma_1\gamma_2[z+(\bar{c}_1+\bar{c}_2\bar{\gamma}_1)x]$. Thus $\sigma_1\sigma_2=(c_1+c_2\gamma_1,\gamma_1\gamma_2)\in(P,\Delta)$. Now it is easy to see that $(0,1)$ is the identity of (P,Δ) and $(c\bar{\gamma},\bar{\gamma})$ is the inverse of (c,γ).

(3) Because of $\det(c,\gamma)=\gamma$, we conclude that $U^+(L)=T(L)\cdot(P,1)$. Next we claim that $(P,1)\subset T(L)$ if $F\neq F_2$. According to Lemma 4.6(3), it suffices

to show that if $\sigma=(c, 1)$ with $c\neq 0$, we can find a $\rho\in T(L)$ such that $\rho\sigma(z)=z$. There is a non-zero $b\in F$ distinct from 1, so that $t=b^2ax+y+bcz$ is isotropic, where $a+\bar{a}=c\bar{c}$. We know $\sigma(z)=\bar{c}x+z$, then $\langle\sigma_t\sigma(z), z\rangle=1+b(1+b)c\bar{c}\in R^*$. Hence $v=[\sigma_t\sigma(z)+z]/\bar{c}$ is isotropic and $\sigma_{v,d}\sigma_t\sigma(z)=z$, if we set $d=b^{-1}(1+b)^{-1}$.

Lemma 4.8. *Suppose that $F=F_2$ and $L=H_1\perp H_2\perp Sz$ with $\langle z, z\rangle=1$. Then $U^+(L)=T(L)$.*

Proof. Using Lemmas 4.5 and 4.7 (3), we only need to show that for $c\neq 0$, $\sigma=(c,1)$ on $H_2\perp Sz$ is in $T(L)$. According to Theorem D, it suffices to find a $\rho\in T(L)$ such that $\rho\sigma(z)=z$. We may choose $a\in S$ satisfying $a+\bar{a}=c\bar{c}$. Then $t=ax_1+y_1+cz$ is isotropic. We know that $\sigma(z)=\bar{c}x_2+z$, so $\sigma_t\sigma(z)=\bar{c}x_2+(1+c\bar{c})z+\bar{c}(ax_1+y_1)$. Thus, $v=[\sigma_t\sigma(z)+z]/\bar{c}$ is isotropic and $\sigma_v\sigma_t\sigma(z)=z$ follows from a direct computation.

Therefore we have by induction

Theorem E. *Suppose that $r>0$ and $L=H_1\perp\cdots\perp H_r\perp M$ with $M=(1)$. Then $U(L)=T(L)\cdot(P,\Delta)$ where (P,Δ) acts on $H_r\perp M$ as in Lemma 4.7. Furthermore, $U^+(L)=T(L)$ holds except when $r=1$ and $F=F_2$.*

The known techniques will not work for the remaining two types, therefore we shall define a new unitary transformation.

Definition. By *a pseudo-symmetry* with respect to an anisotropic x and a scalar $\gamma\in\Delta$ with

$$\frac{1+\gamma}{\langle x, x\rangle}\in S,$$

denoted by $\rho_{x,\gamma}$, we mean a linear endomorphism on L such that for each $y\in L$, we have

$$\rho_{x,\gamma}(y)=y+(1+\gamma)\frac{\langle y, x\rangle}{\langle x, x\rangle}x.$$

Then $\rho_{x,\gamma}\in U(L)$. It is clear that $\rho_{x,\gamma}(x)=\gamma x$ and $\rho_{x,\gamma}(y)=y$ if $\langle y, x\rangle=0$, and $\rho_{x,\gamma}\rho_{x,\bar{\gamma}}=1$. The group generated by all the pseudo-symmetries is denoted by $F(L)$. Then $F(L)$ contains $D(L)$, the group generated by all the quasi-symmetries.

Lemma 4.9. Let

$$L=\begin{pmatrix}1 & \\ \pi^j & \lambda\pi^{2s-1-j}\end{pmatrix}=Su+Sv \quad \text{for} \quad 0<j<s.$$

Then,

(1) $U(L)=F(L)$ and

(2) $U^+(L)=B(L)$, the group generated by all $\rho_{z,\gamma}\rho_{t,\bar{\gamma}}$ for arbitrary anisotropic vectors $z, t\in L$ and $\gamma\in\Delta$.

Proof. (1) First we note that every maximal vector in L is anisotropic. For any $\sigma\neq 1\in U(L)$, put $\sigma(u)=au+bv$ so that $a\bar{a}\pi^j+b\bar{b}\lambda\pi^{2s-1-j}+a\bar{b}+\bar{a}b=\pi^j$. From $0<j<s\leqq 2s-1-j$, we obtain $a+1\equiv 0 \pmod{\Pi S}$.

Case (I). $\mathrm{ord}_K(a+1) \geqq \mathrm{ord}_K(b)$. Let $t=[\sigma(u)+u]/b$, then $\rho_{t,\gamma}(\sigma(u))=u$ if we set $\gamma=[(1+\bar{a})\pi^j+\bar{b}]/[(1+a)\pi^j+b]$. We note that

$$\frac{1+\gamma}{\langle t,t\rangle}=\bar{b}\Big/\left[1+\left(\frac{1+a}{b}\right)\pi^j\right]\in S,$$

namely $\rho_{t,\gamma}\in F(L)$. Put $\tau=\rho_{t,\gamma}\sigma$, then $\tau(u)=u$ and $\tau(v)=cu+(1+c\pi^j)v$ with $c+\bar{c}+c\bar{c}\pi^j=0$. It is easy to see that $\mathrm{ord}_K(c)>0$. If $c\neq 0$, let $z=[\tau(v)+v]/c$, then $\rho_{z,\alpha}\tau(v)=v$ where $\alpha=\bar{c}/c$ and

$$\frac{1+\alpha}{\langle z,z\rangle}=\frac{\bar{c}}{1+\lambda\pi^{2s-1}}\in S.$$

Moreover, $\rho_{z,\alpha}\tau(u)=u$, since $\langle u,z\rangle=0$. Thus $\rho_{z,\alpha}\rho_{t,\gamma}\sigma=1$ and $\sigma\in F(L)$.

Case (II). $\mathrm{ord}_K(a+1)<\mathrm{ord}_K(b)$. Consider

$$\rho_{v,\beta}\sigma(u)=au+bv+(1+\beta)\frac{a+b\lambda\pi^{2s-1-j}}{\lambda\pi^{2s-1-j}}v.$$

Then by Case (I), it suffices to show that we may choose a suitable $\beta\in\Delta$ to decrease the order of the coefficient of v in $\rho_{v,\beta}\sigma(u)$. Let $1+\beta=\lambda\pi^{2s-1-j}d$, so $d+\bar{d}+\lambda\pi^{2s-1-j}d\bar{d}=0$. It is possible to find a $\delta\in S^*$ such that $\delta+\bar{\delta}=\lambda\pi^{2s-1-j}$. Then putting $d=\delta^{-1}$, we are through.

(2) Since $\det(\rho_{z,\gamma})=\gamma$ and $\rho_{z,\gamma}\rho_{z,\alpha}=\rho_{z,\gamma\alpha}$, it is clear that $U^+(L)=B(L)$.

Lemma 4.10. *Let* $L=(1)\perp(1+\lambda\pi^{2s-1})=Su+Sv$. *Then*

(1) $U(L)=F(L)$ *and*

(2) $U^+(L)=B(L)$.

Proof. (1) We note that there is no isotropic vector in L. Put $\sigma(u)=au+bv$ for any $\sigma\in U(L)$. Then $\varphi(a+b)=1$.

Case (I). If $1+a\in S^*$, then $z=\sigma(u)+u$ is maximal. If we put

$$\gamma=\frac{1+\bar{a}}{1+a},\quad \text{then}\quad \frac{1+\gamma}{\langle z,z\rangle}=\frac{1}{1+a}\in S$$

and $\rho_{z,\gamma}\sigma(u)=u$, $\rho_{z,\gamma}\sigma(v)=\beta v$ for some $\beta\in\Delta$, so $\sigma\in F(L)$.

Case (II). If $1+a\in\Pi S$, then $b\in\Pi S$. Consider

$$\rho_{u+v,\alpha}\sigma(u)=\sigma(u)+(1+\alpha)\frac{a+b(1+\lambda\pi^{2s-1})}{\lambda\pi^{2s-1}}(u+v).$$

Itis clear that using the same techniques as in the proof of Lemma 4.9(2), we can get an $\alpha\in\Delta$ such that

$$\frac{1+\alpha}{\lambda\pi^{2s-1}}\in S^*.$$

Then we come back to Case (I).

(2) The same as in Lemma 4.9 (2).

Lemma 4.11. *Let $L=H\perp M$ where*

$$H=\begin{pmatrix}1\\0\ 0\end{pmatrix} \quad \text{and} \quad M=(1)\perp(1+\lambda\pi^{2s-1})$$

or

$$\begin{pmatrix}1\\ \pi^j\ \ \lambda\pi^{2s-1-j}\end{pmatrix} \quad \text{for} \quad 0<j<s.$$

Then $U(L)=T(L)\cdot F(L)$ and $U^+(L)=T(L)\cdot B(L)$.

Proof. Let $H=Sx+Sy$. From Lemma 4.4, for any $\sigma\in U(L)$ there exists a $\rho\in T(L)$ such that $\rho\sigma(x)=x$. Let $\tau=\rho\sigma$, then $\tau(y)=ax+y+z$ with $z\in M$. We may assume that $a\in S^*$, otherwise we consider $\sigma_x\tau(y)$.

Case (1). If $a\in R^*$, then $\sigma_{\tau(y)+y,a^{-1}}\tau(y)=y$ and $\sigma_{\tau(y)+y,a^{-1}}\tau(x)=x$.

Case (2). If $a\in S^*$ but $a\notin R$, then $t=\tau(y)+y$ is anisotropic. We have $\rho_{t,\alpha}\tau(x)=x$ and $\rho_{t,\alpha}\tau(y)=y$, if we set $\alpha=\bar{a}/a$. Note that

$$\frac{1+\alpha}{\langle t,t\rangle}=a^{-1}\in S.$$

Therefore in both cases, the assertions follow from Lemmas 4.9 and 4.10.

Theorem F. *Suppose that $L=H_1\perp\cdots\perp H_r\perp M$ with*

$$H_i=\begin{pmatrix}1\\0\ 0\end{pmatrix} \quad \text{for} \quad 1\leqq i\leqq r$$

and $M=(1)\perp(1+\lambda\pi^{2s-1})$ or

$$\begin{pmatrix}1\\ \pi^j\ \ \lambda\pi^{2s-1-j}\end{pmatrix} \quad \text{for} \quad 0<j<s.$$

Then $U(L)=T(L)\cdot F(H_r\perp M)$ and $U^+(L)=T(L)\cdot B(H_r\perp M)$. Furthermore, $U^+(L)=T(L)$ when $r\geqq 2$ and $M=(1)\perp(1+\lambda\pi^{2s-1})$.

Proof. All the statements except the last one follow from Lemmas 4.5, 4.9, 4.10 and 4.11. Now we only have to show that $U^+(L)=T(L)$ when $r\geqq 2$ and $M=(1)\perp(1+\lambda\pi^{2s-1})=Su+Sv$. Given arbitrary $\sigma\in U^+(L)$, we may assume by Lemma 4.5 that σ is the identity on H_i for all $i<r$. According to Theorem E, it suffices to find $\rho\in T(L)$ such that $\rho\sigma(v)=v$. Put $\sigma(v)=ax_r+by_r+cu+dv$, then $\varphi(c+d)=1$ from $\langle\sigma(v),\sigma(v)\rangle=1+\lambda\pi^{2s-1}$.

Case (1). If a or b is in S^*, we may assume that $a\in S^*$ by the symmetry of x_r and y_r. Let $t=x_1+\alpha y_1+\bar{a}^{-1}(\bar{c}+\bar{d}\lambda\pi^{2s-1})y_r+u+v$ for $\alpha+\bar{\alpha}=\lambda\pi^{2s-1}$, then t is isotropic and $\langle\sigma_t\sigma(v),v\rangle=0$. Thus $z=\sigma_t\sigma(v)+v$ is isotropic and $\sigma_{z,(1+\lambda\pi^{2s-1})^{-1}}\sigma_t\sigma(v)=v$.

Case (2). If both a and b are in ΠS, then we consider $\sigma_{x_r+\alpha y_r+u+v}\sigma(v)$ for $\alpha+\bar{\alpha}=\lambda\pi^{2s-1}$. It is clear that the coefficient of x_r is $a+c+d(1+\lambda\pi^{2s-1})+a\bar{\alpha}+b\in S^*$, because $\varphi(c+d)=1$ and $\varphi(a)=\varphi(b)=0$. Thus, this reduces to Case (1). Q.e.d.

References

1. BOURBAKI, N.: Algèbre, Chapter IX. Actual. Scient. et Ind., No. 1272. Paris 1959.
2. DIEUDONNÉ, J.: On the structure of unitary groups. Trans. Amer. Math. Soc. **72**, 367—385 (1952).
3. — On the structure of unitary groups (II). Amer. J. Math. **75**, 665—678 (1953).
4. — La géométrie des groupes classiques. Ergebnisse der Mathematik und ihrer Grenzgebiete (**2**), Heft 5 (1955).
5. — Sur les groupes classiques. Paris: Hermann 1958.
6. KLINGENBERG, W.: Lineare Gruppen über lokalen Ringen. Amer. J. Math. **83**, 137—153 (1961).
7. — Orthogonale Gruppen über lokalen Ringen. Amer. J. Math. **83**, 281—320 (1961).
8. — Symplectic groups over local rings. Amer. J. Math. **85**, 232—240 (1963).
9. O'MEARA, O.T.: Introduction to quadratic forms. Berlin-Göttingen-Heidelberg: Springer 1963.
10. SAH, C.H.: A note on hermitian forms over fields of characteristic 2. Amer. J. Math. **86**, 262—269 (1964).
11. SCHNEIDER, DAVID I.: The unitary group over generalized valuation rings. MIT thesis, 1964.

Dr. SO-FEI WEI
Department of Mathematics
University of Wisconsin-Milwaukee
U.S.A.

Math. Zeitschr. 105, 110—113 (1968)

A Simple Group of Order 44,352,000

DONALD G. HIGMAN and CHARLES C. SIMS

Received November 20, 1967

The group G of the title is obtained as a primitive permutation group of degree 100 in which the stabilizer of a point has orbits of lengths 1, 22 and 77 and is isomorphic to the Mathieu group M_{22}. Thus G has rank 3 in the sense of [1]. G is an automorphism group of a graph constructed from the Steiner system $\mathfrak{S}(3, 6, 22)$.

WITT [3] defined a Steiner system $\mathfrak{S}(d, m, n)$ to be a set S of n points together with a set B of subsets of S (referred to here as *blocks*) such that each block contains exactly m points and each set of d points is contained in exactly one block. WITT [4] showed that Steiner systems $\mathfrak{S}(3, 6, 22)$ exist and that they are unique up to isomorphism. The automorphism group $\overline{M}_{22}$ of an $\mathfrak{S}(3,6,22)$ contains the Mathieu group M_{22} as a subgroup of index 2 and is the normalizer of M_{22} in M_{24}.

Throughout the rest of the paper we shall use the following notation: S and B will denote the sets of points and blocks, respectively, of a fixed $\mathfrak{S}(3, 6, 22)$. Points will be denoted by Greek letters $\alpha, \beta, \ldots$ and blocks by Roman letters $u, v, \ldots$. For each $\alpha \in S$, $[\alpha]$ will denote the set of blocks containing α.

We shall use the following facts about $\mathfrak{S}(3, 6, 22)$ and $\overline{M}_{22}$:

(1) Each point α is contained in exactly 21 blocks. Thus $|[\alpha]| = 21$.

(2) Two distinct points are contained in exactly 5 blocks.

(3) Two distinct blocks have 0 or 2 points in common, 16 blocks being disjoint from a given block and 60 meeting it in 2 points.

(4) If u is a block not in $[\alpha]$, then exactly 6 blocks in $[\alpha]$ are disjoint from u.

(5) Given distinct points α and β and distinct blocks u and v in $[\alpha] \cap [\beta]$ there exist exactly 4 blocks disjoint from u and v.

(6) No 3 blocks are pairwise disjoint.

(7) $\overline{M}_{22}$ contains an involution fixing exactly 8 points and 21 blocks.

(1)—(6) are easily proved by counting arguments. (7) can be seen from an inspection of the character table of $\overline{M}_{22}$ given in [2].

We now construct an undirected graph $\mathscr{G}$ with vertex set

$$\{*\} \cup S \cup B,$$

where $*$ is a new symbol. In $\mathscr{G}$,

(a) $*$ is joined to each point in S.

(b) Each point $\alpha\in S$ is joined to the 21 blocks in $[\alpha]$.

(c) Two blocks are joined if and only if they are disjoint.

Let $\bar{G}$ denote the automorphism group of $\mathcal{G}$. It is clear that the stabilizer of $*$ in $\bar{G}$ is isomorphic to the automorphism group of $\mathfrak{S}(3,6,22)$, that is, $\overline{M}_{22}$. We shall show that $\bar{G}$ is transitive on the vertices of $\mathcal{G}$, from which it follows that $\bar{G}$ has order 88,704,000. Since by (7) $\bar{G}$ contains an odd permutation, $\bar{G}$ is not simple but contains a simple subgroup G of index 2.

Take $\alpha\in S$ and let $S(\alpha)$ and $B(\alpha)$ be the sets of vertices of $\mathcal{G}$ at distance 1 and 2 from α, respectively. $S(\alpha)=\{*\}\cup[\alpha]$. Thus $|S(\alpha)|=22$ and no two vertices of $S(\alpha)$ are joined. If $\beta\in S-\{\alpha\}$, then β is joined to $*$ and so $\beta\in B(\alpha)$. If $v\in B-[\alpha]$, then by (4) v is joined to some block in $[\alpha]$ and so $v\in B(\alpha)$. Hence

$$B(\alpha)=(S-\{\alpha\})\cup(B-[\alpha])$$

and $|B(\alpha)|=77$.

We shall prove that

(i) *Each vertex in* $B(\alpha)$ *is joined to exactly* 6 *vertices in* $S(\alpha)$.

(ii) *Three distinct vertices in* $S(\alpha)$ *are joined to exactly one vertex in* $B(\alpha)$.

(iii) *Two vertices in* $B(\alpha)$ *are joined if and only if they are not joined to a common vertex in* $S(\alpha)$.

From (i), (ii), (iii) *and the uniqueness of* $\mathfrak{S}(3,6,22)$ *it follows that the stabilizer of* α *in* $\bar{G}$ *is also isomorphic to* $\overline{M}_{22}$ *and this implies that* $\bar{G}$ *is transitive.*

Proof of (i). A vertex in $B(\alpha)$ is either a point $\beta\in S-\{\alpha\}$ or a block u in $B-[\alpha]$. If $\beta\in S-\{\alpha\}$, then by (2) β is joined to $*$ and to the 5 blocks containing α and β and to no other vertices in $S(\alpha)$. If $u\in B-[\alpha]$, then by (4) u is joined to the 6 blocks in $[\alpha]$ disjoint from u and to no other vertices in $S(\alpha)$.

Proof of (ii). We consider in turn each of the three types of sets of 3 distinct vertices in $S(\alpha)$. Since by (i) each vertex in $B(\alpha)$ is joined to 20 triples and there are $77\cdot 20$ triples altogether, it suffices to show that each triple is joined to at least one vertex in $B(\alpha)$.

Type I. $\{*, v, w\}$, $v, w\in[\alpha]$. In this case $*$, v and w are joined to β, where $v\cap w=\{\alpha,\beta\}$.

Type II. $\{u, v, w\}$, $u, v, w\in[\alpha]\cap[\beta]$, $\beta\in S-\{\alpha\}$. Here u, v and w are joined to β.

Type III. $\{u,v,w\}$, $u, v, w\in[\alpha]$, $u\cap v=\{\alpha,\beta\}$, $u\cap w=\{\alpha,\gamma\}$, $v\cap w=\{\alpha,\delta\}$, with β, γ and δ distinct points of $S-\{\alpha\}$. We must show the existence of a block disjoint from u, v and w. Let $\overline{w}=w-\{\alpha,\gamma,\delta\}$. By (5) there are 4 blocks disjoint from u and v, say z_1, z_2, z_3, z_4. Suppose all of the z_i intersect w non-trivially. Let $\bar{z}_i=z_i-w$. By (3) $|\bar{z}_i|=4$. Let $1\leqq i<j\leqq 4$. $w\cap z_i$ and $w\cap z_j$ are contained in $\overline{w}$ and each contain 2 points. Hence $w\cap z_i\cap z_j$ is non-empty. Since $|z_i\cap z_j|\leqq 2$, we have $|\bar{z}_i\cap\bar{z}_j|\leqq 1$. Therefore

$$\Big|\bigcup_i \bar{z}_i\Big|\geqq\sum_i|z_i|-\sum_{i<j}|\bar{z}_i\cap\bar{z}_j|\geqq 16-6=10.$$

However,

$$\bigcup_i \bar{z}_i \subseteq S - u \cup v \cup w$$

and $|u \cup v \cup w| = 13$. Thus

$$\left|\bigcup_i \bar{z}_i\right| \leqq 9,$$

a contradiction.

Proof of (iii). By (ii) each vertex in $B(\alpha)$ is joined to 16 other vertices in $B(\alpha)$. By (i) and (ii) we may consider $B(\alpha)$ to be the set of blocks of an $\mathfrak{S}(3,6,22)$ with point set $S(\alpha)$. By (3) it suffices to show that if two vertices in $B(\alpha)$ are joined, then they are not joined to a common vertex in $S(\alpha)$. There are three types of two-element subsets of $B(\alpha)$.

Type I. $\{\beta,\gamma\} \subseteq S - \{\alpha\}$. β and γ are not joined.

Type II. $\{\beta,u\}$, $\beta \in S - \{\alpha\}$, $u \in B - [\alpha]$. If β and u are joined, then $\beta \in u$. If β and u are joined to a common vertex in $S(\alpha)$, then that vertex must be a block $v \in [\alpha]$. But then $\beta \in v$ and so $u \cap v \neq \phi$. Therefore u and v are not joined.

Type III. $\{u,v\} \subseteq B - [\alpha]$. A vertex in $S(\alpha)$ joined to u and v must be a block w in $[\alpha]$. If u is also joined to v, then u, v and w are pairwise disjoint, contradicting (6).

We conclude by giving generating permutations for G. Numbering the vertex $*$ as 1, the points of S as 2, 3, ..., 23, and the blocks in B as 24, 25, ..., 100 in an appropriate manner, G is found to be generated by the permutations

a = (1) (2, 8, 13, 17, 20, 22, 7) (3, 9, 14, 18, 21, 6, 12)
(4, 10, 15, 19, 5, 11, 16) (23) (24, 77, 99, 72, 64, 82, 40)
(25, 92, 49, 88, 28, 65, 90) (26, 41, 70, 98, 91, 38, 75)
(27, 55, 43, 78, 86, 87, 45) (29, 69, 59, 79, 76, 35, 67)
(30, 39, 42, 81, 36, 57, 89) (31, 93, 62, 44, 73, 71, 50)
(32, 53, 85, 60, 51, 96, 83) (33, 37, 58, 46, 84, 100, 56)
(34, 94, 80, 61, 97, 48, 68) (47, 95, 66, 74, 52, 54, 63)

and

b = (1, 35) (2) (3, 81) (4, 92) (5) (6, 60) (7, 59) (8, 46)
(9, 70) (10, 91) (11, 18) (12, 66) (13, 55) (14, 85) (15, 90)
(16) (17, 53) (19, 45) (20, 68) (21, 69) (22) (23, 84)
(24, 34) (25, 31) (26, 32) (27) (28) (29) (30) (33) (36)
(37, 39) (38, 42) (40, 41) (43, 44) (47) (48) (49, 64)
(50, 63) (51, 52) (54, 95) (56, 96) (57, 100) (58, 97)
(61, 62) (65, 82) (67, 83) (71, 98) (72, 99) (73) (74, 77)
(75) (76, 78) (79) (80) (86) (87, 89) (88) (93) (94).

References

1. HIGMAN, D. G.: Finite permutation groups of rank 3. Math. Z. **86**, 145—156 (1964).

2. TODD, J. A.: A representation of the Mathieu group M_{24} as a collineation group. Ann. Mat. Pura App. (4) **71**, 199—238 (1966).

3. WITT, E.: Die 5-fach transitiven Gruppen von Mathieu. Abh. Math. Sem. Hamb. **12**, 256—264 (1937).

4. — Über Steinersche Systeme. Abh. Math. Sem. Hamb. **12**, 265—275 (1937).

Prof. D. G. HIGMAN
Department of Mathematics
University of Michigan
Ann Arbor, Michigan
U.S.A.

Prof. C. C. SIMS
Department of Mathematics
Rutgers, The State University
New Brunswick, N. J. 08903
U.S.A.

Math. Zeitschr. 105, 114—127 (1968)

Moufang Veblen-Wedderburn Systems

MICHAEL J. KALLAHER*

Received July 20, 1967

1. Introduction

In this paper we investigate Veblen-Wedderburn systems (and the corresponding projective planes) in which the multiplicative identity

$$(x\,y)(z\,x)=(x\cdot y\,z)\,x$$

holds. Such a system is called a Moufang Veblen-Wedderburn system. Examples of such systems include the alternative division rings and the nearfields. Alternative division rings were characterized by BRUCK and KLEINFELD [3] for characteristic not two and by KLEINFELD [6] in the characteristic two case.

ZASSENHAUS [10] completely determined all finite nearfields and ANDRÉ [1] investigated the projective planes coordinatized by nearfields. ANDRÉ proved the

Theorem. *A projective plane $\mathscr{P}$ which is not the exceptional plane*[1] *is Desarguesian if there exist in $\mathscr{P}$ points U, V, U', and V' such that $\mathscr{P}$ is coordinatized by a nearfield with respect to the points U and V and also with respect to the points U' and V'.*

A corollary to this theorem is the result that every collineation of a projective plane $\mathscr{P}$ coordinatized by a proper nearfield (i.e., not a division ring) with respect to the points U and V either fixes U and V or interchanges them, provided $\mathscr{P}$ is not the exceptional plane. With the aid of this result ANDRÉ describes completely the collineation groups of such planes.

We will extend these two results of ANDRÉ to the class of Moufang planes. In section 2 an algebraic investigation of Moufang Veblen-Wedderburn systems is conducted. Two necessary and sufficient conditions for such a system to be an alternative division ring are given.

In section 3 we define a Bol plane as a projective plane coordinatized by a Moufang Veblen-Wedderburn system. We restrict ourselves to a certain class of Moufang Veblen-Wedderburn systems, members of which are called special, and we define a special Bol plane as one coordinatized by such a system. We prove:

* The results in this paper form a part of the author's doctoral dissertation written at Syracuse University under the direction of Professor ERWIN KLEINFELD. This research was supported partially by the U. S. Army Research Office in Durham and partially by a National Aeronautics and Space Administration Fellowship at Syracuse University.

[1] This is the projective plane over the associative Hall Veblen-Wedderburn system of nine elements.

Theorem. *Let $\mathscr{P}$ be a special Bol plane with respect to two points U and V. If there exist points U' and V' with $U' \neq U, V$ such that $\mathscr{P}$ is also a Bol plane with respect to U' and V', then $\mathscr{P}$ is a Moufang plane.*

Any Moufang Veblen-Wedderburn system with characteristic $\neq 3$ and 5 is shown to be special. Also any nearfield, except the one with nine elements, is a special Moufang Veblen-Wedderburn system. Thus our theorem generalizes ANDRÉ's theorem to the larger class of Moufang planes.

Finally, the corollary to the theorem of ANDRÉ extends to planes coordinatized by proper special Moufang Veblen-Wedderburn systems (i.e., those which are not alternative division rings). In section 4, by means of this fact, a necessary and sufficient condition for a proper special Moufang Veblen-Wedderburn system to be a nearfield is proven.

For the principal definitions and concepts of projective planes we refer the reader to either HALL [5] or PICKERT [7]. In particular we shall assume a familiarity with the notions of coordinatizing planes by means of ternary rings. The notation (R, F) will be used to denote the ternary ring in which R is the set of symbols and F the ternary function defined on R.

2. Moufang Veblen-Wedderburn Systems

A Moufang Veblen-Wedderburn system (hereafter called an MVW system) is a ternary ring (R, F) which satisfies

(i) $F(a, b, c) = ab + c$ for all $a, b, c \in R$;

(ii) R is an abelian group under addition $(+)$;

(iii) $a(b+c) = ab + ac$ for all $a, b, c \in R$;

(iv) $(ab)(ca) = (a \cdot bc)a$ for all $a, b, c \in R$.

Thus an MVW system is a Veblen-Wedderburn system whose multiplicative loop is a Moufang loop. For loops satisfying (iv) we have the following theorem (BRUCK [2], pp. 115–117):

Theorem. *Let M be a loop satisfying* (iv). *The following identities hold for all $x, y, z \in M$:*

(a) $(xy \cdot z)y = x(y \cdot zy)$,

(b) $x(y \cdot xz) = (xy \cdot x)z$,

(c) $xx \cdot y = x \cdot xy$, $xy \cdot x = x \cdot yx$, *and* $yx \cdot x = y \cdot xx$.

Moreover, if x, y, z are elements of M such that $xy \cdot z = x \cdot yz$, then x, y, and z generate an associative subloop of M.

With respect to an MVW system the theorem tells us that every non-zero element has a unique multiplicative inverse and any two non-zero elements generate a group under multiplication. In what follows we shall use these two facts plus identities (a), (b), (c) freely without specific reference.

A *proper MVW system* is one in which the second distributive law, $(a+b)c = ac + bc$, does not hold. A *nearfield* is an MVW system in which $xy \cdot z = x \cdot yz$ holds for all x, y, z. The only proper MVW systems known are

the nearfields. However there do exist MVW systems which are not associative under multiplication – the Cayley-Dickson division algebras. Thus it appears likely that proper MVW systems which are not nearfields do exist. Our first task is therefore to discover under what conditions an MVW system is either associative or a Cayley division algebra.

Definition. Let (R, F) be an MVW system. The *nucleus* of R is the set $N(R)=\{n \mid n\in R,\ n\cdot xy=nx\cdot y$ for all $x, y\in R\}$; the *kernel* of R is $K(R)=\{k \mid k\in N(R), (x+y)k=xk+yk$ for all $x, y\in R\}$; and the *center* of R is $C(R)=\{c \mid c\in N(R), cx=xc$ for all $x\in R\}$.

Lemma 2.1. *Let (R, F) be a Moufang Veblen-Wedderburn system. If $n\in N(R)$, then $xn\cdot y=x\cdot ny$ and $xy\cdot n=x\cdot yn$ for all $x, y\in R$. Conversely, if n satisfies $xn\cdot y=x\cdot ny$ $(xy\cdot n=x\cdot yn)$ for all $x, y\in R$, then $n\in N(R)$.*

Proof. The first statement follows from the second statement of the theorem above. Assume $xn\cdot y=x\cdot ny$. Let $y=np$. Then $n\cdot xy=n(x\cdot np)=(nx\cdot n)p=nx\cdot np=nx\cdot y$. Thus $n\in N(R)$. A similar proof gives: $xy\cdot n=x\cdot yn$ for all $x, y\in R$ implies $n\in N(R)$.

Lemma 2.2. *In a Moufang Veblen-Wedderburn system (R, F)*

$$b(c\cdot ab^{-1})=(bc\cdot a)b^{-1}, \tag{2.1}$$

$$b^{-1}(bc\cdot a)=(c\cdot ab^{-1})b \tag{2.2}$$

for all $a, b, c\in R$.

Proof. We have

$$b(c\cdot ab^{-1})\cdot b=b\cdot(c\cdot ab^{-1})b=bc\cdot(ab^{-1}\cdot b)=bc\cdot a.$$

The lemma follows.

Lemma 2.3. *In a Moufang Veblen-Wedderburn system (R, F)*

$$a(-b)=(-a)b=-ab$$

for all $a, b\in R$, where $-b$ is defined by $b+(-b)=0$.

Proof. By the left distributive law we have $ab+a(-b)=0$. Hence $a(-b)=-ab$. In particular $(-1)(-1)=1$, and hence $(-1)[(-1)a+a]=a+(-1)a$. Thus either $-1=1$ or $(-1)a=-a$. If the second alternative holds, then $(-1)a=-a=a(-1)$. Let $b=-1$ in (2.1). Then

$$(-1)[c\cdot a(-1)]=[(-1)c\cdot a](-1)=(-1)[(-1)c\cdot a],$$

from which we obtain $-ca=c(-a)=(-c)a$. If $-1=1$, then $-a=a$ for all $a\in R$ and there is nothing to prove.

With the aid of Lemmas 2.1 and 2.3 it is easy to prove that $N(R)$ is a nearfield. $K(R)$ then forms an associative division ring and thus we can consider R as a (right) vector space over $K(R)$ (PICKERT [7], pp. 202–205). But $C(R)$, although an abelian group under multiplication, need not be closed under addition.

Theorem 2.1. *Let* (R, F) *be a Moufang Veblen-Wedderburn system. If*

$$(a+1)b=ab+b \tag{2.3}$$

for all $a, b \in R$, *then* R *is an alternative division ring.*

Proof. It is sufficient to show $(a+b)c=ac+bc$ for all $a, b, c \in R$. Without loss of generality we may assume $a \neq 0$, $c \neq 0$. Let $a=dc$, $d \neq 0$. Then

$$\begin{aligned}(a+b)c &= a(1+a^{-1}b)\cdot c=[dc\cdot(1+a^{-1}b)]c\\ &= d[c\cdot(1+a^{-1}b)c]=dc\cdot c+d[c(a^{-1}b\cdot c)].\end{aligned}$$

Using the fact that $a^{-1}=c^{-1}d^{-1}$ and applying (2.2) we have

$$\begin{aligned}d[c(a^{-1}b\cdot c)] &= d[c(c^{-1}d^{-1}\cdot b)\cdot c]=d(d^{-1}\cdot bc)\\ &= bc.\end{aligned}$$

Thus $(a+b)c=ac+bc$.

In [1] ANDRÉ showed that a nearfield is a division ring if the relation $(b+c+bc)a=ba+ca+bc\cdot a$ holds for all a, b, c. This is true for MVW systems of characteristic not two as the following theorem shows.

Theorem 2.2. *Let* (R, F) *be a Moufang Veblen-Wedderburn system in which* $1+1 \neq 0$. *If*

$$(b+c+bc)a=ba+ca+bc\cdot a \tag{2.4}$$

holds for all $a, b, c \in R$, *then* R *is an alternative division ring.*

Proof. Letting $b=c=1$ in (2.4) gives $(1+1+1)a=a+a+a$. If $c=1$, then (2.4) gives

$$(b(1+1)+1)a=b(a+a)+a \tag{2.5}$$

for all $a, b \in R$. Letting $b=(1+1)^{-1}$ and multiplying both sides of the resulting equation by $(1+1)$ on the left, we get

$$(1+1)^2\cdot a=a(1+1)+(1+1)a. \tag{2.6}$$

Let $b=(1+1+1)(1+1)^{-1}$ in (2.5). Then we have

$$(1+1+1+1)a=(1+1+1)(1+1)^{-1}\cdot(a+a)+a.$$

Applying (2.6) to this equation gives

$$\begin{aligned}a+(1+1)a &= (1+1)^{-1}\cdot a+(1+1)^{-1}\cdot a+(1+1)^{-1}\cdot a\\ &\quad +(1+1)^{-1}\cdot a+(1+1)^{-1}\cdot a+(1+1)^{-1}\cdot a.\end{aligned}$$

Multiplying both sides of this equation by $(1+1)$ gives

$$(1+1)a+(1+1)^2a=a+a+a+a+a+a.$$

Applying (2.6) again we get

$$(1+1)a\cdot(1+1)=(a+a)(1+1).$$

Thus $(1+1)a=a+a$ for all $a\in R$. Returning to (2.5) we have

$$[b(1+1)+1]a=b\cdot(1+1)a+a=b(1+1)\cdot a+a,$$

since $(1+1)\in N(R)$. If $c\in R$, pick $b=c(1+1)^{-1}$. Then

$$(c+1)a=[b(1+1)+1]a=b(1+1)\cdot a+a=ca+a.$$

Theorem 2.1 completes the proof.

Of interest also are sufficient conditions under which an MVW system is a nearfield. We have the following isolated result:

Theorem 2.3. *Let (R, F) be a Moufang Veblen-Wedderburn system of dimension two over $K(R)$. If $K(R)=C(R)$, then (R, F) is a nearfield.*

Proof. Let 1 and u be a basis for R over $K(R)$. Define α, σ: $K(R)\times K(R)\to K(R)$ as follows: If $a, b\in K(R)$, then

$$(a+ub)u=\alpha(a, b)+u\sigma(a, b).$$

α and σ are well-defined and onto. Also

$$R=\{a+ub\mid a, b\in K(R)\} \quad \text{implies} \quad R=\{\alpha(a, b)+u\sigma(a, b)\mid a, b\in K(R)\}.$$

Let $a+ub$, $c+ud$, $e+uf$ be three elements of R. Then

$$(a+ub)(c+ud)\cdot(e+uf)=(a+ub)(c+ud)\cdot e+(a+ub)(c+ud)\cdot uf$$

and

$$(a+ub)\cdot(c+ud)(e+uf)=(a+ub)\cdot(c+ud)e+(a+ub)\{(c+ud)\cdot uf\}.$$

Since e and f are in $K(R)$ we need only prove

$$(a+ub)(c+ud)\cdot u=(a+ub)\cdot(c+ud)u. \tag{2.7}$$

The left-hand side of (2.7) equals

$$\alpha(ac+\alpha(a, b)d, bc+\sigma(a, b)d)+u\sigma(ac+\alpha(a, b)d, bc+\sigma(a, b)d), \tag{2.8}$$

while the right-hand side is

$$a\alpha(c, d)+ub\alpha(c, d)+\alpha(a, b)\sigma(c, d)+u\sigma(a, b)\sigma(c, d). \tag{2.9}$$

The proof that (2.8) equals (2.9) is contained in the next two lemmas.

Lemma 2.4. *In the (R, F) of Theorem 2.3*

$$\alpha(ae, be)=\alpha(a, b)e, \qquad \sigma(ae, be)=\sigma(a, b)e$$

for all $a, b, e\in K(R)$.

Proof. These two identities are obtained by performing the multiplications in the identity

$$(a+ub)\cdot ue=(ae+ube)u,$$

and using the fact that 1 and u are linearly independent.

Lemma 2.5. *In the MVW system of Theorem 2.3,*

$$\alpha(a\,c+\alpha(a,b)\,d,\, b\,c+\sigma(a,b)\,d)=a\,\alpha(c,d)+\alpha\big(a\,\sigma(c,d),\, b\,\sigma(c,d)\big),$$

$$\sigma(a\,c+\alpha(a,b)\,d,\, b\,c+\sigma(a,b)\,d)=b\,\alpha(c,d)+\sigma\big(a\,\sigma(c,d),\, b\,\sigma(c,d)\big)$$

for all $a, b, c, d\in K(R)$.

Proof. Consider

$$(a+u\,b)(c+u\,d)\cdot(c+u\,d)=(a+u\,b)\cdot(c+u\,d)^2,$$

true for all $a, b, c, d\in K(R)$. The left-hand side of this identity is

$$a\,c^2+\alpha(a\,d, b\,d)\,c+\alpha(a\,c\,d+\alpha(a\,d, b\,d)\,d, b\,c\,d+\sigma(a\,d, b\,d)\,d)$$
$$+u[b\,c^2+\sigma(a\,d, b\,d)\,c+\sigma(a\,c\,d+\alpha(a\,d, b\,d)\,d, b\,c\,d+\sigma(a\,d, b\,d)\,d)],$$

while the right-hand side equals

$$a\,c^2+a\,\alpha(c\,d, d^2)+\alpha(a\,d\,c, b\,d\,c)+\alpha\big(a\,\sigma(c\,d, d^2), b\,\sigma(c\,d, d^2)\big)$$
$$+u[b\,c^2+b\,\alpha(c\,d, d^2)+\sigma(a\,d\,c, b\,d\,c)+\sigma(a\,\sigma(c\,d, d^2), b\,\sigma(c\,d, d^2))].$$

The lemma follows by setting like components of these two expressions equal and applying Lemma 2.4.

Returning to the theorem, a direct application of Lemmas 2.4 and 2.5 to the expression (2.8) gives (2.9). Therefore (R, F) is a nearfield.

Theorem 2.3 has three interesting corollaries. The first one is

Corollary 2.3.1. *If* (R, F) *is a Moufang Veblen-Wedderburn system of dimension two over* $K(R)$ *and* $N(R)=C(R)$, *then* R *is a field.*

Proof. Since $C(R)\subset K(R)\subset N(R)$ we have $K(R)=C(R)$ and hence (R, F) is a nearfield. But then $N(R)=R=C(R)$ and thus R is a field.

Corollary 2.3.2. *Let* (R, F) *be a Hall Veblen-Wedderburn system. If* (R, F) *is not the associative system consisting of nine elements, then the multiplicative loop of* (R, F) *is not Moufang.*

Proof. See HALL [4] for the definition of a Hall Veblen-Wedderburn system. All that is needed here are the facts that, if it is Moufang, then 1) such a system has dimension two over its kernel and 2) $N(R)=C(R)$, except when R has nine elements. Therefore the corollary follows from Corollary 2.3.1.

The last corollary gives a characterization of the associative Hall Veblen-Wedderburn system consisting of nine elements:

Corollary 2.3.3. *If* (R, F) *is a proper Moufang Veblen-Wedderburn system satisfying*

$$a^2=-1 \qquad \text{for } a\neq 0, \pm 1, a\in R, \tag{2.10}$$

then (R, F) *is the associative Hall Veblen-Wedderburn system of nine elements.*

Proof. ANDRÉ ([1], pp. 141 – 142) considered a proper nearfield (N, G) which satisfied (2.10). He proved that (N, G) must be the associative Hall Veblen-Wedderburn system of nine elements. In his proof he derived the following facts about (N, G):

(1) $a^{-1} = -a$ for all $a \in N$, $a \neq 0, \pm 1$;

(2) $ba = -ab$ for all $a, b \in N$, $a \neq 0, \pm 1$, $b \neq 0, \pm 1, \pm a$;

(3) N has characteristic three;

(4) N has at most two linearly independent elements over the prime field of its kernel.

(1) to (4) hold also for (R, F) of this corollary and ANDRÉ's proofs go over directly. Thus, by (3) the prime field of $K(R)$ is $GF(3)$. $R \neq GF(3)$ since (R, F) is proper. Hence R has dimension two over $GF(3)$. If $K(R) \neq GF(3)$, then $K(R) = R$ and R is a field, again contradicting the assumption that (R, F) is proper. Hence $K(R) = GF(3)$. Then $K(R) = C(R)$ since $GF(3) \subset C(R)$. Therefore by Theorem 2.3 (R, F) is a nearfield. The result of ANDRÉ implies that (R, F) is the associative Hall Veblen-Wedderburn system of nine elements.

3. Bol Planes

We first define the Bol configuration which will play an important part in this section. Let $\mathscr{P}$ be a projective plane and A, B, C three distinct points of $\mathscr{P}$ with $A \notin BC$, the line determined by B and C. The Bol configuration with basis points A, B, C is defined as follows: From

$$E_1 C = E_2 C \qquad E_2 B = Q_2 B \qquad E_i A = P_i A \qquad (i = 1, 2)$$

$$P_1 C = P_2 C \qquad P_1 B = R_1 B \qquad Q_i A = R_i A \qquad (i = 1, 2)$$

$$Q_1 C = Q_2 C \qquad E_1 B = Q_1 B = P_2 B = R_2 B$$

$$E_i, \quad P_i, \quad Q_i \notin AB, \quad BC, \quad AC \qquad (i = 1, 2)$$

follows $R_1 C = R_2 C$ (Fig. 1).

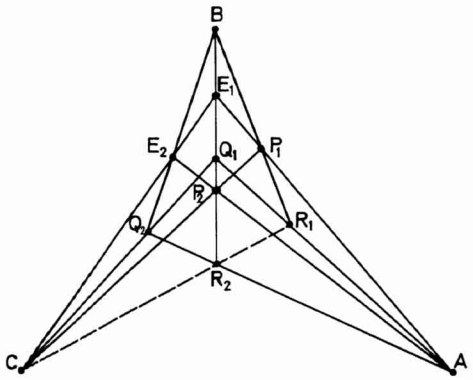


Fig. 1

Secondly we shall follow HALL's method of coordination. If $\mathscr{P}$ is a projective plane coordinatized by a ternary ring (R, F), then with (R, F) can be associated four points U, V, O, I of $\mathscr{P}$. U is the point (0) or the point at infinity of the x-axis, V is the point (∞) or the point at infinity of the y-axis, O is the point (0, 0), and I is the point (1, 1). U, V, O, I are called the basis points of (R, F); U and V are called the points at infinity of (R, F). Note that in this terminology order is important: the first point will always be the point (0), the second the point (∞), etc.

Let $\mathscr{P}$ be a projective plane over an MVW system (R, F) and assume the basis points of (R, F) in $\mathscr{P}$ are U, V, O, I. Geometrically $\mathscr{P}$ is (UV, UV)-transitive since (R, F) is a Veblen-Wedderburn system ([7], p. 101) and the Moufang property of (R, F) is expressed in the fact that the Bol configuration holds in $\mathscr{P}$ for the following cases: (1) $A=O$, $B=V$, $C=U$; (2) $A=V$, $B=O$, $C=U$; (3) $A=V$, $B=U$, $C=O$ ([7], pp. 50–58). Note that because of the (UV, UV)-transitivity every ternary ring of $\mathscr{P}$ which has the points U and V as its points at infinity will be an MVW system.

Definition. A projective plane $\mathscr{P}$ is a *Bol plane* (with respect to the points U and V) if one of the ternary rings of $\mathscr{P}$ (with U and V as its points at infinity) is an MVW system.

Theorem 3.1. *A projective plane $\mathscr{P}$ is a Moufang plane if there exist in $\mathscr{P}$ distinct points U, V, V' with $UV=UV'$ such that $\mathscr{P}$ is a Bol plane with respect to U and V and also with respect to U and V'.*

Proof. Choose an MVW system (R, F) of $\mathscr{P}$ such that U and V are its points at infinity and $V'=(1)$. This we may do since $\mathscr{P}$ is (UV, UV)-transitive. In (R, F) we have by Lemma 2.3 $a(-b)=(-a)b=-ab$.

Pick $a, b, c \in R$, all non-zero, $a \neq b$, and $c \neq 1$. Consider the points $E_1=(0, ca)$, $E_2=(a, ca)$, $Q_1=(0, cb)$, $Q_2=(b, cb)$, $P_1=(-a, -a+ca)$, $P_2=(0, -a+ca)$, $R_2=(0, -b+cb)$, $R_1=(f, f+cb)$, where f is defined by $(-a+ca)(-a)^{-1} \cdot f = -b+cb$ (Fig. 2).

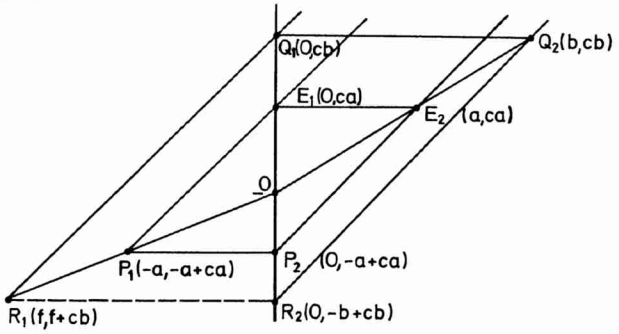


Fig. 2

Since $\mathscr{P}$ has a ternary ring with U and V' as its points at infinity which is an MVW system, BOL's configuration holds for $A=V'$, $B=O$ (the origin for

(R, F)), and $C=U$. The points E_i, P_i, Q_i, R_i $(i=1, 2)$ fulfill the hypothesis of BOL's configuration and hence $R_1 U=R_2 U$. Thus $-b+cb=f+cb$, or $f=-b$. Hence

$$(-a+ca)(-a)^{-1}\cdot(-b)=-b+cb.$$

If $a=-1$, then

$$(1-c)b=b-cb \tag{3.1}$$

for all b, c with $b\neq 0, -1$, $c\neq 0, 1$. If $b=0, -1$ or $c=0, 1$ (3.1) still holds. Thus given any $b, d\in R$ we have

$$(1+d)b=[1-(-d)]b=b-(-d)b=b+db.$$

By Theorem 2.1 (R, F) is an alternative division ring and therefore $\mathscr{P}$ is a Moufang plane.

Definition. A Moufang Veblen-Wedderburn system is *special* if $N(R)$ is neither the associative Hall Veblen-Wedderburn system with nine elements, $GF(3)$, nor $GF(5)$ [2]. A projective plane $\mathscr{P}$ is a *special* Bol plane (with respect to the points U and V) if one of the ternary rings of $\mathscr{P}$ (with U and V as its points at infinity) is a special MVW system.

Lemma 3.1. *If (R, F) is a special Moufang Veblen-Wedderburn system with characteristic $\neq 2$, there exists $b\in N(R)$ with $b\neq 0, 1, -1$, and $b^2\neq -1$.*

Proof. If $N(R)$ is a proper nearfield, the assertion follows from the results of ANDRÉ given in the proof of Corollary 2.3.3. Obviously $GF(3)$ and $GF(5)$ are the only fields which do not fulfill the assertion. In the case of a non-commutative division ring the assertion follows from the fact that to every subfield there exists a pure extension field.

Theorem 3.2. *Let $\mathscr{P}$ be a special Bol plane and let (R, F) be the ternary ring of the definition with basis points U, V, O, I. $\mathscr{P}$ is a Moufang plane if there exist distinct points U', V' satisfying $U'V'=UV$, $U'\neq U, V$, such that $\mathscr{P}$ is a Bol plane with respect to U' and V'.*

Proof. Note that if $\mathscr{P}$ is a Bol plane with respect to U and V, then $\mathscr{P}$ is a Bol plane with respect to the points V and U. Thus if $V'=U$ or $V'=V$ we can apply Theorem 3.1 and obtain the theorem. We assume therefore that $V'\neq U, V$.

We first prove that we may assume $U'=(1)$ with respect to (R, F). If $U'\neq(1)$, then $U'=(g)$ for some $g\in R$, $g\neq 0$. We define a new ternary ring (R, F') with basis points U, V, O, $I'=(1, g)$, giving the new coordinate pair $[a, a]$ to the point (a, ga) of OI' and thus the new coordinate pair $[a, b]$ to the point (a, gb), $a, b\in R$. Then $F'(a, b, c)=g^{-1}(ga\cdot b+gc)$ and, if $\oplus$ and $\otimes$ denote the new addition and multiplication respectively, then $a\otimes b=g^{-1}(ga\cdot b)$ and $a\oplus b=g^{-1}(ga+gb)=a+b$.

Because $\mathscr{P}$ is (UV, UV)-transitive and the multiplicative loops (R, F), (R, F') are isotopic, (R, F') is an MVW system. Furthermore, if $b\in N(R)$ in the

[2] Professor GÜNTER PICKERT has pointed out that "nonexceptional" is a better name than "special". The author wishes to thank Professor PICKERT for many helpful suggestions for improving this section. In particular, Lemma 3.1 and its proof are due to him.

old ternary ring, then $b \in N(R)$ in the new one. For, using Lemma 2.2, $(b \otimes c) \otimes a = g^{-1} \cdot (gb \cdot c)\, a = b(c \cdot ag^{-1}) \cdot g$, and $b \otimes (c \otimes a) = g^{-1}[gb \cdot g^{-1}(gc \cdot a)] = b(c \cdot ag^{-1}) \cdot g$ for all $a, c \in R$. Also if $b^{-1} \neq -b$ in (R, F), then $b \otimes (-b) = g^{-1}(gb \cdot (-b)) = -b^2 \neq 1$, and hence $b^{-1} \neq -b$ in (R, F'). Thus we may assume $U' = (1)$.

We may also assume further that $V' \notin (IU \cap OV)(IV \cap OU)$. Suppose otherwise. Then $V' = (-1)$, $VI \cap OV' = (1, -1)$. Since $U' \neq V'$ we must have that the characteristic of (R, F) is not two. $\mathcal{P}$ has the (V, OU)-collineation σ with $\sigma(x, y) = (x, by)$, $\sigma(y) = (by)$, $\sigma(U) = U$, $\sigma(V) = V$, where $b \in N(R)$, $b \neq 0, \pm 1$, $b^{-1} \neq -b$. Replace U, V by $\overline{U} = U'$, $\overline{V} = V'$, and U', V' by $\overline{U}' = (b) = \sigma(U')$, $\overline{V}' = (-b) = \sigma(V')$. $\mathcal{P}$ is a Bol plane with respect to $\overline{U}$, $\overline{V}$, and also with respect to $\overline{U}'$, $\overline{V}'$. Let $\bar{I} = (1, b)$. Since in (R, F) $(1+1)^{-1}$ distributes on the right, we obtain $(\bar{I}\overline{U} \cap \bar{O}\overline{V})(\bar{I}\overline{V} \cap \bar{O}\overline{U}) \cap \overline{U}\,\overline{V} = (b^{-1})$. Hence $\overline{V}' \notin (\bar{I}\overline{U} \cap \bar{O}\overline{V}) \cdot (\bar{I}\overline{V} \cap \bar{O}\overline{U})$. Note that $\overline{U}' = (1)$ with respect to a ternary ring (R, F) (which will also be an MVW system) having basis points $\overline{U}$, $\overline{V}$, $\bar{O}$, $\bar{I}$.

We have reduced the proof to the following case: $\mathcal{P}$ is a Bol plane with respect to U and V and also with respect to U' and V', where $U' = (1)$ and $V' \neq U, V$, $V' \notin (IU \cap OV)$ $(IV \cap OU)$ (i.e., $V' \neq (-1)$). Representing the points of $\mathcal{P}$ by their coordinates with respect to (R, F), the mapping π with

$$\pi(x, y) = (y, x), \quad \pi(z) = (z^{-1}), \quad \pi(U) = V, \quad \pi(V) = U$$

is a collineation of $\mathcal{P}$. Note that $\pi(P) = P$ for all $P \in OI$. If $Q = (-1)$, then $(x, y) \in (0, c)\, Q$ is equivalent to $x + y = c$. Hence π fixes every line through Q and therefore π is a (Q, OI)-collineation. $V' \neq (-1)$ and $V' \notin OI$ implies $\pi(V') = V'' \neq V'$. Hence $\mathcal{P}$ is a Bol plane with respect to U' and V' and also with respect to U' and V''. Thus, by Theorem 3.1, $\mathcal{P}$ is a Moufang plane.

Corollary 3.2.1. *Let $\mathcal{P}$ be a special Bol plane with respect to the distinct points U and V. If there exist distinct points U' and V' with $U' \neq U, V$ such that $\mathcal{P}$ is also a Bol plane with respect to U' and V', then $\mathcal{P}$ is a Moufang plane.*

Proof. If $U'V' \neq UV$, then $\mathcal{P}$ is a translation plane with respect to the lines UV and $U'V'$. Thus $\mathcal{P}$ is a Moufang plane (HALL [5], p. 372). If $U'V' = UV$ then Theorem 3.2 applies and $\mathcal{P}$ is Moufang.

Corollary 3.2.2. *Let $\mathcal{P}$ be a special Bol plane with respect to the distinct points U and V. $\mathcal{P}$ is a Moufang plane if and only if $\mathcal{P}$ has a collineation which maps U (or V) into a point U' with $U' \neq U, V$.*

An application of Corollary 3.2.2 gives the following characterization of Moufang planes:

Theorem 3.3. *Let $\mathcal{P}$ be a projective plane which with respect to distinct points U and V is coordinatized by a division ring and with respect to distinct points U' and V' is a special Bol plane. Then $\mathcal{P}$ is a Moufang plane.*

Proof. If $UV \neq U'V'$, then $\mathcal{P}$ is a translation plane with respect to two lines and is thus a Moufang plane. If $UV = U'V'$ and either $V = U'$ or $V = V'$, then

$\mathscr{P}$ has a ternary ring which is a division ring and whose multiplicative loop is Moufang. Hence it is an alternative division ring and thus $\mathscr{P}$ is a Moufang plane.

We have reduced the proof to the case where $UV=U'V'$ and $V\neq U'$, $V\neq V'$. $\mathscr{P}$ coordinatized by a division ring with respect to the points U and V implies that $\mathscr{P}$ is (V, γ)-transitive for some line $\gamma\neq UV$. Hence $\mathscr{P}$ has a collineation which fixes UV and maps U' into a point $U''\neq U', V'$. By Corollary 3.2.2 $\mathscr{P}$ is a Moufang plane.

We close this section with one more characterization of Moufang planes. First we state a lemma whose proof parallels that of Lemma 2.3 and is therefore omitted.

Lemma 3.2. *Let (R, F) be a Veblen-Wedderburn system which satisfies*

$$(b\cdot cb)a=b(c\cdot ba)$$

for all $a, b, c\in R$. Then

$$(-1)a=a(-1)=-a, \tag{3.2}$$

$$b(-a)=-ba \tag{3.3}$$

for all $a, b\in R$.

Theorem 3.4. *Let $\mathscr{P}$ be a projective plane with a line γ having the property that every ternary ring of $\mathscr{P}$ with γ as the line at infinity*[3] *is a Veblen-Wedderburn system satisfying*

$$(b\cdot cb)a=b(c\cdot ba). \tag{3.4}$$

Then $\mathscr{P}$ is a Moufang plane.

Proof. (3.4) holding in all ternary rings having γ as the line at infinity says that the Bol configuration holds if $B\notin\gamma$, $A, C\in\gamma$ ([7], p. 57). Pick points U, V, O, I with $UV=\gamma$, $O\notin\gamma$, $I\notin UV$, OV, OU. Let (R, F) be a ternary ring of $\mathscr{P}$ having U, V, O, I as basis points. Then (3.4) holds in (R, F) and it is a Veblen-Wedderburn system. Also by the lemma (3.2) and (3.3) hold in (R, F).

Let a, b, c, d be non-zero elements of R with $a\neq b$, $c\neq d$. Let $A=(d)$, $C=U$, $B=0$. Pick $E_1=(0, ca)$, $E_2=(a, ca)$, $P_1=(-a, -da+ca)$, $P_2=(0, -da+ca)$, $Q_1=(0, cb)$, $Q_2=(b, cb)$, $R_1=(f, df+cb)$, $R_2=(0, -db+cb)$, where f is defined by the equations $x(-a)=-da+ca$, $xf=df+cb$.

These eight points satisfy BOL's configuration for A, B, C. Hence $R_1C=R_2C$ or $df+cb=-db+cb$. Therefore $f=-b$ and this gives

$$x(-b)=-db+cb.$$

Let $a=1$. Then $x=d-c$ and

$$(d-c)\cdot(-b)=-db+cb \tag{3.5}$$

[3] By this is meant the basis points (0) and (∞) lie on the line γ.

for all non-zero b, c, d with $b \neq 1$, $c \neq d$. But if $b=1$, if $c=d$, or if b, c, or d is 0, (3.5) still is true. Thus (3.5) holds in general. Let $e=-b$, $h=d-c$ in (3.5). Then

$$h e+c e=(h+c) e$$

for all h, c, $e \in R$. Thus (R, F) is a division ring satisfying (3.4). SKORNYAKOV [9] and SAN SOUCIE [8] have shown that all such division rings are alternative. Hence $\mathscr{P}$ is a Moufang plane.

4. A Characterization of Nearfields

In this section we determine a necessary and sufficient condition for a proper special MVW system to be a nearfield. For this purpose we need to determine the central collineations of a plane coordinatized by a proper special MVW system. Such a plane will be called a *super Bol plane*.

The reader is referred to either [5] or [7] for the definition and properties of central collineations. If $\mathscr{P}$ is a projective plane, P a point, and γ a line of $\mathscr{P}$, then $G(P, \gamma)$ denotes the set of central collineations of $\mathscr{P}$ with center P and axis γ.

Our first theorem is an immediate corollary of Theorem 3.2:

Theorem 4.1. *If $\mathscr{P}$ is a super Bol plane with respect to the points U and V, then every collineation of $\mathscr{P}$ either fixes U and V or interchanges them.*

Theorem 4.2. *Let $\mathscr{P}$ be a super Bol plane, (R, F) a proper special Moufang Veblen-Wedderburn system coordinatizing it, and U, V, O, I the basis points of (R, F). Representing the points of $\mathscr{P}$ by their coordinates with respect to (R, F), $\mathscr{P}$ has four types of central collineations:*

I. $(x, y) \to (xs+c, ys+d)$, $(m) \to (m)$, $U \to U$, $V \to V$, where $c, d \in R$, $s \in K(R)$, $s \neq 0$.

II. $(x, y) \to (x, ay+d)$, $(m) \to (am)$, $U \to U$, $V \to V$, with $d \in R$, $a \in N(R)$, $a \neq 0$.

III. $(x, y) \to (ax+c, y)$, $(m) \to (ma^{-1})$, $U \to U$, $V \to V$, with $c \in R$, $a \in N(R)$, $a \neq 0$.

IV. $(x, y) \to (a^{-1}(y-k), ax+k)$, $(m) \to (am^{-1}a)$, $U \to V$, $V \to U$, where $a, k \in R$, $a \neq 0$.

Proof. Note first that Theorem 4.1 says that every central collineation either fixes U and V or interchanges them. Let ξ be a (C, γ)-collineation of $\mathscr{P}$. If $\gamma = UV$, then γ is of the form I ([7], pp. 199–203).

The (C, γ)-collineations with $\gamma \neq UV$ must fulfill $C \in UV$. If $U \in \gamma$ a (C, γ)-collineation ν must also fix V; hence $C=V$ and ν is a (V, γ)-collineation with $U \in \gamma$. Assume first that $\gamma = OU$. It is well known that ν must then have the form II with $d=0$ ([7], p. 103, proof of th. 45). Conversely, any collineation of the form II with $d=0$ is a (V, OU)-collineation.

Assume now that γ is a line through U which is not UO. Let $(0, b) \in \gamma$. If τ is the (V, UV)-collineation which maps $(0, 0)$ into $(0, b)$, then $\tau^{-1} \nu \tau$ is a collineation of $\mathscr{P}$ which fixes V and the line OU pointwise. Hence $\tau^{-1} \nu \tau = \sigma$,

where σ is a collineation of the form II with $d=0$. Thus $\nu=\tau\,\sigma\,\tau^{-1}$ has the form II. Conversely, any collineation of the form II is a (V, γ)-collineation, where $\gamma=(0, d)\,U$.

Similarly any central collineation of $\mathscr{P}$ with center U and axis $\gamma \neq UV$ going through V is of the form III and conversely.

Let γ be a line of $\mathscr{P}$ going through O but through neither U nor V, and assume $(c)\in\gamma$. Then the mapping of form IV with $a=c$ and $k=0$ is a central collineation δ_c of $\mathscr{P}$ with axis γ and center $(-c)$. If σ is a (C, γ)-collineation for some $C\in UV$, then σ must interchange U and V. Thus $\sigma\delta_c^{-1}$ fixes U and V and also γ pointwise. Therefore $\sigma\delta_c^{-1}=1$ or $\sigma=\delta_c$. Hence if γ is a line of $\mathscr{P}$ going through O but through neither U nor V, there exists one and only one (non-identical) central collineation (given by IV with $k=0$) of $\mathscr{P}$ which has γ as its axis.

Finally, let γ be a line of $\mathscr{P}$ not going through O, U, or V. Assume $(0, b)$ and (c) lie on γ. Consider the mapping of the form IV with $a=c$ and $k=b$. This is a collineation of $\mathscr{P}$ which fixes every point on γ and also the point $(-c)$. Proceeding as in the last paragraph we can show that this is the only central collineation of $\mathscr{P}$ which has γ as its axis. This completes the proof of Theorem 4.2.

Definition. The *projective group* of a projective plane is the group of collineations generated by its central collineations. A *projective collineation* is an element of this group.

Theorem 4.2 tells us that an element of the projective group of a super Bol plane must be a product of at most four types of central collineations. If in the ternary ring (R, F) of Theorem 4.2 multiplication is associative then we can actually give the form of any projective collineation of $\mathscr{P}$. For then $N(R)=R$ and hence any product of collineations of types I to IV must have the form

$$(x, y)\to(a\,x\,s+c,\ b\,y\,s+d)$$

or the form

$$(x, y)\to(a\,y\,s+c,\ b\,x\,s+d).$$

Here we have restrict ourselves to describing the action on the points of $\mathscr{P}-UV$. Thus we have proved the following theorem due to ANDRÉ [1]:

Corollary 4.2.1 (André). *Let $\mathscr{P}$ be a projective plane over a proper nearfield (R, F) having basis points U, V, O, I, and assume $\mathscr{P}$ is not the exceptional plane. Then any projective collineation of $\mathscr{P}$ can be written as a product of at most four central collineations and, describing its action on the points of $\mathscr{P}-UV$, it has either the form*

$$(x, y)\to(a\,x\,s+c,\ b\,y\,s+d) \tag{4.1}$$

or the form

$$(x, y)\to(a\,y\,s+c,\ b\,x\,s+d), \tag{4.2}$$

where $a, b, c, d\in R$ and $s\in K(R)$ with a, b, and s non-zero.

For proper special Moufang Veblen-Wedderburn systems (R, F) with $K(R)=C(R)$ the converse of Corollary 4.2.1 is true. Specifically we have:

Theorem 4.3. *Let (R, F) be a proper special Moufang Veblen-Wedderburn system with $K(R)=C(R)$, let $\mathscr{P}$ be the super Bol plane coordinatized by (R, F), and assume U and V are the points at infinity of (R, F). (R, F) is a nearfield if and only if every projective collineation of $\mathscr{P}$ has either the form* (4.1) *or the form* (4.2) *(again restricting ourselves to $\mathscr{P}-UV$).*

Proof. The necessity is simply Corollary 4.2.1. For the sufficiency let e and f be any two non-zero elements of R and consider the collineations δ_e and δ_f of form IV with $k=0$, $a=e$, and $k=0$, $a=f$ respectively. Then the collineation $\delta_f\delta_e$ maps the point (x, y) into $(f^{-1}\cdot ex, f\cdot e^{-1}y)$. Since $\delta_f\delta_e$ is a projective collineation it must be of the form (4.1) for some $a, b, c, d\in R$, $s\in K(R)$. Clearly $c=0=d$. If (x, y) is the point (1.1) then we obtain $f^{-1}e=as$. Since $s\in C(R)$ we have

$$f^{-1}\cdot ec=acs=as\cdot c=f^{-1}e\cdot c$$

for all $c\in R$. Hence R is associative.

Bibliography

1. André, J.: Projektive Ebenen über Fastkörpern. Math. Z. **62**, 137–160 (1955).
2. Bruck, R. H.: A survey of binary systems. Berlin-Göttingen-Heidelberg: Springer 1958.
3. –, and E. Kleinfeld: The structure of alternative division rings. Proc. Amer. Math. Soc. **2**, 878–890 (1951).
4. Hall, M.: Projective planes. Trans. Amer. Math. Soc. **54**, 229–277 (1943).
5. – Theory of groups. New York: Macmillan 1959.
6. Kleinfeld, E.: Alternative divisoin rings of characteristic 2. Proc. Nat. Acad. Sci. U.S.A. **37**, 818–820 (1951)
7. Pickert, G.: Projektive Ebenen. Berlin-Göttingen-Heidelberg: Springer 1955.
8. San Soucie, R. L.: Right alternative division rings of characteristic 2. Proc. Amer. Math. Soc. **6**, 291–296 (1955).
9. Skornyakov, L. A.: Right-alternative fields. Izv. Akad. Nauk. SSSR Ser. Mat. **15**, 177–184 (1951) [Russian].
10. Zassenhaus, H.: Über endliche Fastkörper. Abh. Math. Seminar Univ. Hamburg **11**, 187–220 (1936).

Dr. M. J. Kallaher
Department of Mathematics
University of Manitoba
Winnipeg, Canada

Math. Zeitschr. 105, 128—140 (1968)

A Cohomology Theory for Commutative Galois Extensions

MORRIS ORZECH*

Received September 25, 1967

Introduction

Let R be a commutative ring, G a finite abelian group. Let $E(R, G)$ denote the set of isomorphism classes of Galois extensions of R with Galois group G. Let $A(R, G)$ denote the subset of $E(R, G)$ consisting of the classes of extensions having a normal base. Let $T(R, G)$ denote the subset of $E(R, G)$ of classes of commutative extensions. Define

$$N(R, G) = A(R, G) \cap T(R, G).$$

In [5] and [10] the authors obtain cohomological descriptions of $E(R, G)$ and $A(R, G)$. In this paper we use these descriptions and a cohomology theory for functor categories, developed by CHASE, to describe $T(R, G)$ and $N(R, G)$ cohomologically.

In Section 1 we set forth the definitions and properties of an additivizing functor which forms the basis for the cohomology theory for functor categories referred to above. We also review the definition of a related cohomology theory based on Harrison cohomology [7].

In Section 2 we introduce a more general setting for the theory introduced in Section 1. It is related to the Harrison-Amitsur bi-cohomology presented in [10] in the same way as the cohomology theories in Section 1 are related.

In Section 3 we review briefly the definitions of the various sets of Galois extensions which we wish to classify, and their properties as abelian groups and as functors. We also recall the isomorphisms $H^2 UR(G) \cong A(R, G)$, $\underline{H}^1 U(R, G) \cong E(R, G)$ which permits us to describe $E(R, G)$ and $A(R, G)$ cohomologically.

In Section 4 we establish the isomorphisms

$$h^1 U R(G) \cong N(R, G), \quad h^1 U(R, G) \cong E(R, G).$$

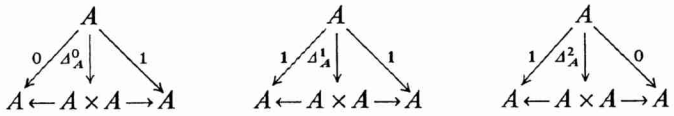
This paper constitutes a portion of the author's doctoral dissertation at Cornell University. The author wishes to thank S. U. CHASE for his helpful advice and encouragement.

§ 1. An Additivizing Functor

Let $\mathscr{C}$ be an abelian category, and $F: \mathscr{C} \to Ab$ a covariant functor from $\mathscr{C}$ to the category of abelian groups. We begin by showing how to associate to F an additive functor $hF: \mathscr{C} \to Ab$ such that if F is additive, hF is naturally equivalent to F.

* This work was done while the author was a National Foundation fellowship holder.

For A an object of $\mathscr{C}$, let maps $\Delta_A^i: A \to A \times A$, $i=0, 1, 2$, be defined as the unique maps rendering the diagrams below commutative; the unlabeled maps in these diagrams are the canonical projections.



Now define $hF(A)=\text{kernel}\,(F(\Delta_A^0)-F(\Delta_A^1)+F(\Delta_A^2))$. It is not difficult to verify that hF defines a functor from $\mathscr{C}$ to Ab: Writing

$$\delta F(A)=F(\Delta_A^0)-F(\Delta_a^1)+F(\Delta_A^2)$$

and $\imath F(A)$ for the inclusion map from $hF(A)$ to $F(A)$, we define $hF(\alpha)$, for $\alpha: A \to B$ a map in $\mathscr{C}$, to be the unique map such that the following diagram is commutative

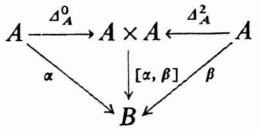
$$\begin{array}{ccccccc} 0 \to & hF(A) & \xrightarrow{\imath F(A)} & F(A) & \xrightarrow{\delta F(A)} & F(A \times A) \\ & \downarrow{\scriptstyle hF(\alpha)} & & \downarrow{\scriptstyle F(\alpha)} & & \downarrow{\scriptstyle F(\alpha \times \alpha)} \\ 0 \to & hF(B) & \xrightarrow[\imath F(B)]{} & F(B) & \xrightarrow[\delta F(B)]{} & F(B \times B). \end{array}$$

It is easy to verify that $\imath F, \delta F$ define natural transformations of functors.

Theorem 1. (a) *$\imath F$ is a natural equivalence if and only if F is additive.*

(b) *For any functor $F: \mathscr{C} \to Ab$, $\imath hF$ is an equivalence.*

Proof. We first observe that Δ_A^0, Δ_A^2 are the canonical injections of A into the second and first factors of $A \times A$, the latter being viewed as a direct sum. It follows easily that $\Delta_A^1=\Delta_A^0+\Delta_A^2$, so that if F is additive, $hF(A)=F(A)$ for all A and $\imath F$ is an equivalence. Conversely, for $\alpha, \beta: A \to B$ in $\mathscr{C}$ we have $\alpha+\beta=\Delta_A^1[\alpha, \beta]$, where $[\alpha, \beta]$ renders the following diagram commutative



So if $\imath F$ is a natural equivalence,

$$\begin{aligned} F(\alpha+\beta)=F(\Delta_A^1[\alpha, \beta])&=F(\Delta_A^1)\,F([\alpha, \beta])=(F(\Delta_A^0)+F(\Delta_A^2))\,F([\alpha, \beta]) \\ &=F(\Delta_A^0[\alpha, \beta])+F(\Delta_A^1[\alpha, \beta])=F(\alpha)+F(\beta). \end{aligned}$$

This proves (a).

We now wish to show that $\delta hF=0$. It clearly suffices to show that for each object A of $\mathscr{C}$ the diagram below is commutative, since $\delta F(A)\,\imath F(A)=0$ by

definition, and $\iota F(A\times A)$ is monic by definition

$$\begin{array}{ccc} hF(A) & \xrightarrow{\iota F(A)} & F(A) \\ {\scriptstyle \delta hF(A)}\downarrow & & \downarrow{\scriptstyle \delta F(A)} \\ hF(A\times A) & \xrightarrow[\iota F(A\times A)]{} & F(A\times A). \end{array}$$

To show that this diagram is commutative it suffices to show that

$$\iota F(A\times A)\, hF(\Delta_A^i)=F(\Delta_A^i)\,\iota F(A) \quad \text{for} \quad i=0,1,2.$$

Now $hF(\Delta_A^i)$ is defined to be the unique map which makes the following diagram commutative

$$\begin{array}{ccccccc} 0 & \longrightarrow & hF(A) & \xrightarrow{\iota F(A)} & F(A) & \xrightarrow{\delta F(A)} & F(A\times A) \\ & & {\scriptstyle hF(\Delta_A^i)}\downarrow & & \downarrow{\scriptstyle F(\Delta_A^i)} & & \downarrow{\scriptstyle F(\Delta_A^i\times\Delta_A^i)} \\ 0 & \to & hF(A\times A) & \xrightarrow[\iota F(A\times A)]{} & F(A\times A) & \xrightarrow[\delta F(A\times A)]{} & F(A\times A\times A\times A). \end{array}$$

This implies the desired result, proving (b).

The correspondence $F\to hF$ is natural in a sense we shall now make explicit. Assume $\mathscr{C}$ has a small skeleton, and let $\mathscr{P}(\mathscr{C})$, or more briefly $\mathscr{P}$, denote the category of functors from $\mathscr{C}$ to Ab. Let $\mathscr{AP}(\mathscr{C})$, or $\mathscr{AP}$, denote the full subcategory of $\mathscr{P}(\mathscr{C})$ the objects of which are additive functors; the morphisms in $\mathscr{P}$ and $\mathscr{AP}$ are the natural transformations of functors. Let $i\colon \mathscr{AP}\to\mathscr{P}$ denote the embedding, or forgetful functor.

We now remark that $\mathscr{P}$ and $\mathscr{AP}$ are abelian categories possessing arbitrary sums and products; these are computed object-wise. A sequence of objects and maps in $\mathscr{P}$ or $\mathscr{AP}$, say $F'\to F\to F''$, is exact if for each object A of $\mathscr{C}$ the sequence $F'(A)\to F(A)\to F''(A)$ is exact [6, Chapter 6].

Theorem 2. *The correspondence $F\to hF$ defines an additive functor $h\colon \mathscr{P}\to\mathscr{AP}$. The functor h is a right adjoint to the functor $i\colon \mathscr{AP}\to\mathscr{P}$. In particular, h is left-exact.*

Proof. Let $\theta\colon F\to G$ be a natural transformation between two objects of $\mathscr{P}$. Define a natural transformation $h\theta\colon hF\to hG$ by defining $h\theta(A)$, for A an object of $\mathscr{C}$, to be the unique map rendering the following diagram commutative

$$\begin{array}{ccccccc} 0 & \to & hF(A) & \xrightarrow{\iota F(A)} & F(A) & \xrightarrow{\delta F(A)} & F(A^2) \\ & & {\scriptstyle h\theta(A)}\downarrow & & \downarrow{\scriptstyle \theta(A)} & & \downarrow{\scriptstyle \theta(A^2)} \\ 0 & \to & hG(A) & \xrightarrow[\iota G(A)]{} & G(A) & \xrightarrow[\delta G(A)]{} & G(A^2). \end{array}$$

That $h\theta$ is a natural transformation follows by an easy diagram chase around a rather obvious three-dimensional diagram, using the fact that $\iota F(A)$ is monic for each F and for each A.

That h is additive i.e. that $h(\theta+\theta')=h\theta+h\theta'$, follows from uniqueness of $h\theta$ in the diagram above.

The adjointness relation between i and h follows from Theorem 1. More explicitly, let F be an object of $\mathscr{P}$, G an object of $\mathscr{AP}$. We wish to set up an isomorphism between the morphism classes $\mathscr{AP}(G, hF)$ and $\mathscr{P}(iG, F)$. Given $\theta\colon G \to hF$, we associate to θ the composite $(\imath F)\theta\colon G \to hF \to F$. Conversely, given $\mu\colon G \to F$ in $\mathscr{P}$, $h\mu\colon hG \to hF$ in $\mathscr{AP}$, so to μ we associate $(h\mu)(\imath G)^{-1}$. It is straight-forward to verify that h, i, are adjoints. The left-exactness of h follows from the fact that any right adjoint is left-exact, or it can easily be shown directly via a diagram-chasing argument.

Having constructed $h\colon \mathscr{P} \to \mathscr{AP}$, we remark that it can be shown that each object F of $\mathscr{P}$ admits a monomorphism $F \to E$ into an injective object of $\mathscr{P}$ [3, Cor. 2.7, p. 17]. We may thus consider the right-derived functors h^n of h, $n \geqq 0$. Since h is left-exact, $h^0 = h$.

Theorem 3. (a) *Given a short exact sequence*

$$(\xi) = 0 \to F' \xrightarrow{\alpha} F \xrightarrow{\beta} F'' \to 0$$

in $\mathscr{P}$, *there exist natural transformations* $V_n(\xi)\colon h^n F'' \to h^{n+1} F'$ *in* $\mathscr{AP}$, $n \geqq 0$, *such that the following sequence is exact:*

$$0 \to h^0 F' \xrightarrow{h^0\alpha} h^0 F \xrightarrow{h^0\beta} h^0 F'' \xrightarrow{V_1(\xi)} h^1 F' \xrightarrow{h_1\alpha} \cdots \xrightarrow{h^n\beta} h^n F'' \xrightarrow{V_n(\xi)} h^{n+1} F' \xrightarrow{h^{n+1}\alpha} \cdots .$$

(b) *Given a map of exact sequences over* $\mathscr{P}$, *i.e. a commutative diagram with exact rows:*

$$\begin{array}{lccccccc} (\xi)\colon & 0 \to & F' & \xrightarrow{\alpha} & F & \xrightarrow{\beta} & F'' & \to 0 \\ & & \downarrow{\scriptstyle \theta'} & & \downarrow{\scriptstyle \theta} & & \downarrow{\scriptstyle \theta''} & \\ (\eta)\colon & 0 \to & G' & \xrightarrow[\gamma]{} & G & \xrightarrow[\lambda]{} & G'' & \to 0, \end{array}$$

the following diagram commutes for all $n \geqq 0$

$$\begin{array}{ccccccc} h^n F' & \xrightarrow{h^n\alpha} & h^n F & \xrightarrow{h^n\beta} & h^n F'' & \xrightarrow{V_n(\xi)} & h^{n+1} F' \\ \downarrow{\scriptstyle h^n\theta'} & & \downarrow{\scriptstyle h^n\theta} & & \downarrow{\scriptstyle h^n\theta''} & & \downarrow{\scriptstyle h^{n+1}\theta'} \\ h^n G' & \xrightarrow[h^n\gamma]{} & h^n G & \xrightarrow[h^n\lambda]{} & h^n G'' & \xrightarrow[V_n(\eta)]{} & h^{n+1} G' . \end{array}$$

(c) *If* E *is an injective object in* $\mathscr{P}$, $h^n E = 0$ *for* $n > 0$.

Proof. This is a well-known homological result.

Definition. If $\{k^n\}$ is a sequence of functors satisfying (a) and (b) of the above theorem (with k^n replacing h^n), we shall say that $\{k^n, V_n\}$ is a *connected sequence of functors* [9].

We now review the construction of the generalized Harrison cohomology groups [10, §2]. For A an object of $\mathscr{C}$ and $n \geqq 0$ an integer, define maps Δ_A^i: $A^n \to A^{n+1}$, $i = 0, \ldots, n+1$, by the formulae

$$\Delta_A^i = \begin{cases} (0, 1_A n) & i = 0 \\ (1_A i - 1, 1_A, 1_A, 1_A n - i) & 0 < i < n+1 \\ (1_A n, 0) & i = n+1 . \end{cases}$$

Now define

$$\delta^n F(A)\colon\ F(A^n)\to F(A^{n+1}) \quad \text{by} \quad \delta^n F(A)=\sum_{i=0}^{n+1}(-1)^i F(\Delta_A^i).$$

Then $\delta^{n+1}\delta^n=0$, so we obtain a cochain complex

$$CF(A)=\{C^n F(A),\delta^n F(A)\}\colon\ 0\to F(A)\xrightarrow{\delta^1 F(A)} F(A^2)\xrightarrow{\delta^2 F(A)} F(A^3)\to\cdots.$$

We denote the cohomology groups of this complex by $H^n F(A)$. In particular, $\delta^1 F(A)=\delta F(A)$ and $H^1 F(A)=hF(A)$.

Theorem 4. (a) *The correspondence $F\to CF$ defines an exact covariant functor from $\mathscr{P}$ to the category of cochain complexes over $\mathscr{P}$.*

(b) *For each $n\geqq 0$, H^n defines a functor from $\mathscr{P}$ to $\mathscr{P}$.*

(c) *Given an exact sequence $(\xi)=0\to F'\xrightarrow{\alpha}F\xrightarrow{\beta}F''\to 0$ in $\mathscr{P}$, there is associated with it a sequence of natural transformations $\nabla^n\xi\colon H^nF''\to H^{n+1}F'$ in $\mathscr{P}$ such that $\{H^n,\nabla^n\}$ is a connected sequence of functors.*

(d) *There exist natural transformations $\phi_n\colon h^n\to H^{n+1}$, of functors from $\mathscr{P}$ to $\mathscr{P}$ such that* (i) *$\phi_0=$ identity,* (ii) *ϕ_1 is monic, and* (iii) *given an exact sequence*

$$(\xi)=0\to F'\to F\to F''\to 0$$

in $\mathscr{P}$, the diagram below commutes for $n\geqq 0$.

$$\begin{array}{ccccccc} h^nF' & \longrightarrow & h^nF & \longrightarrow & h^nF'' & \xrightarrow{\nabla^n\xi} & h^{n+1}F' \\ \downarrow{\scriptstyle \phi_nF'} & & \downarrow{\scriptstyle \phi_nF} & & \downarrow{\scriptstyle \phi_nF''} & & \downarrow{\scriptstyle \phi_{n+1}F'} \\ H^{n+1}F' & \to & H^{n+1}F & \to & H^{n+1}F'' & \xrightarrow[\nabla^n\xi]{} & H^{n+2}F'. \end{array}$$

Proof. All these assertions follow by straightforward arguments and well-known homological results. We sketch here some of the constructions needed to obtain the theorem. First we establish the functorial nature of the correspondences $F\to CF$, $F\to H^nF$.

Let $\theta\colon F\to G$ be a natural transformation of functors. Then for each $n\geqq 0$ and each $i=0,1,\ldots,n+1$ we have a commutative diagram

$$\begin{array}{ccc} F(A^n) & \xrightarrow{F(\Delta_A^i)} & F(A^{n+1}) \\ \downarrow{\scriptstyle \theta(A^n)} & & \downarrow{\scriptstyle \theta(A^{n+1})} \\ G(A^n) & \xrightarrow[G(\Delta_A^i)]{} & G(A^{n+1}). \end{array}$$

This establishes the relations $\theta(A^{n+1})\delta^nF(A)=\delta^nG(A)\theta(A^n)$, so θ is seen to induce a map of complexes $C\theta\colon CF\to CG$. Thus θ also induces maps $H^n\theta\colon H^nF\to H^nG$. The existence of a connecting homomorphism $\nabla^n\xi$ as claimed in (c) is a standard result.

We now outline the construction of ϕ_n. Let F be an object of $\mathscr{P}$. Let E be an injective object of $\mathscr{P}$ such that we have an exact sequence $(\xi)=0\to F\to E\to G\to 0$ in $\mathscr{P}$. We have remarked that $h=H^1$, so $h^0=H^1$. We take $\phi_0F\colon h^0F\to H^1F$ to be the identity map. Suppose ϕ_nF has been defined for each object F of $\mathscr{P}$.

We have a commutative diagram with exact rows

$$\begin{array}{ccccccccc} h^n F & \longrightarrow & h^n E & \longrightarrow & h^n G & \xrightarrow{\nabla_n \xi} & h^{n+1} F & \to & h^{n+1} E = 0 \\ \downarrow{\scriptstyle \phi_n F} & & \downarrow{\scriptstyle \phi_n E} & & \downarrow{\scriptstyle \phi_n G} & & \downarrow{\scriptstyle \phi_{n+1} F} & & \\ H^{n+1} F & \to & H^{n+1} E & \to & H^{n+1} G & \xrightarrow[\nabla_n \xi]{} & H^{n+2} F & \to & H^{n+2} E . \end{array}$$

$h^{n+1} E = 0$, and $\phi_{n+1} F$ is defined as the unique map making the above diagram commutative. Of course, ϕ_n does not depend on the particular choice of the sequence (ξ). That ϕ_0 is the identity map holds by definition. It is then obvious from the diagram below that $\phi_1 F$ is monic for each F (where E, G are as above)

$$\begin{array}{ccccccc} h^0 E & \to & h^0 G & \xrightarrow{\nabla_1 \xi} & h^1 F & \to & 0 \\ \| {\scriptstyle \phi_0 E} & & \| {\scriptstyle \phi_0 G} & & \downarrow{\scriptstyle \phi_1 F} & & \\ H^1 E & \to & H^1 G & \longrightarrow & H^2 F . & & \end{array}$$

This completes the proof of the theorem.

§ 2. The Bifunctor Cohomologies

We review briefly the definition of Amitsur cohomology, in a setting suited for our later use. Let Ω be a subcategory of the category of commutative rings which satisfies the following properties:

(i) Ω is small.

(ii) Ω has an initial object ω.

(iii) Ω possesses finite direct sums.

A typical setting will be when Ω consists of certain R-algebras, where R is a commutative ring. In this case tensor products will yield direct sums; for this reason, if S, T are objects of Ω, we will write $S \otimes T$ for the direct sum of S and T in Ω. The injections $S \to S \otimes T$, $T \to S \otimes T$ must be interpreted as the injections into the direct sum. But the definition of Amitsur cohomology given below is in fact valid whenever we have a category with finite direct sums.

For S an object of Ω, let $S^n = S \otimes \cdots \otimes S$. Define maps ε_S^i: $S^n \to S^{n+1}$, $i = 1, 2, \ldots, n+1$ by

$$\varepsilon_S^i(x_1 \otimes \cdots \otimes x_n) = x_1 \otimes \cdots \otimes x_{i-1} \otimes 1 \otimes x_i \otimes \cdots \otimes x_n .$$

These maps satisfy the co-semisimplicial identity $\varepsilon_i \varepsilon_j = \varepsilon_{j+1} \varepsilon_i$ for $i \leqq j$; thus if F: $\Omega \to Ab$ is a covariant functor, we can construct the cochain complex

$$F(S) \xrightarrow{d^0 F(S)} F(S^2) \xrightarrow{d^1 F(S)} F(S^3) \to \cdots$$

where

$$d^n F(S) = \sum_{i=1}^{n+1} (-1)^i F(\varepsilon_s^i) .$$

We now consider the category $\mathscr{P}(\Omega)$ of functors from Ω to $\mathscr{P}$, the morphisms being natural transformations of functors. We remark, as we did in the case of $\mathscr{P}$, that $\mathscr{P}(\Omega)$ is an abelian category with arbitrary limits, that the

operations in $\mathscr{P}(\Omega)$ are performed object-wise, and that a sequence of objects in $\mathscr{P}(\Omega)$ is exact if it yields an exact sequence in $\mathscr{P}$ when evaluated at each object S of Ω. The same remarks hold when $\mathscr{P}(\Omega)$ is replaced by the full subcategory $\mathscr{A}\mathscr{P}(\Omega)$ consisting of functors from Ω to $\mathscr{A}\mathscr{P}$.

Given an object F of $\mathscr{P}(\Omega)$ we may regard F as a bifunctor $F\colon \Omega\times\mathscr{C}\to Ab$, namely by setting $F(S, A)=F(S)(A)$ for each object S of Ω and each object A of $\mathscr{C}$. We shall adopt this point of view in the following discussion, introducing the following bit of notational redundancy: for $F\colon \Omega\times\mathscr{C}\to Ab$ and S an object of Ω we will write $FS=F(S, -)$ for the functor from $\mathscr{C}$ to Ab given by $FS(A)=F(S, A)$; likewise, if A is an object of $\mathscr{C}$, F and A determine a functor $FA=F(-, A)\colon \Omega\to Ab$.

Let $F\colon \Omega\times\mathscr{C}\to Ab$ be an object in $\mathscr{P}(\Omega)$. For p, q integers, define an object $C^{p,q}F$ in $\mathscr{P}(\Omega)$ by

$$C^{p,q}F(S,A)=\begin{cases}F(S^{q+1},A^{p+1}) & \text{if } p,q\geqq 0.\\ 0 & \text{if } p<0 \text{ or } q<0.\end{cases}$$

We make $C^{p,q}F(S, A)$ into a bicomplex of abelian groups by defining coboundary operators

$$d_1^{p,q}\colon\ C^{p,q}\to C^{p+1,q},\qquad d_2^{p,q}\colon\ C^{p,q}\to C^{p,q+1}\qquad \text{for } p,q\geqq 0,$$

by the formulae

$$d_1^{p,q}F(S,A)=(-1)^p\,\delta^{p+1}FS^{q+1}(A),$$
$$d_2^{p,q}F(S,A)=d^qFA^{p+1}(S).$$

That $C^{p,q}F(S, A)$ satisfies the conditions for being a bicomplex of abelian groups [6, p. 60] is easily verified using the fact that the ε^i's and Δ^j's commute with one another [10, §3]. The bicomplex gives rise to a total complex [6, p. 60] $CF(S, A)$, and to cohomology groups $\underline{H}^n_\Omega F(S, A)$.

Theorem 5. (a) *The correspondences $F\to C^{p,q}F$, $F\to \underline{H}^n_\Omega F$ give rise to functors from $\mathscr{P}(\Omega)$ to $\mathscr{P}(\Omega)$. The correspondence $F\to CF$ yields a functor from $\mathscr{P}(\Omega)$ to the category of cochain complexes over $\mathscr{P}(\Omega)$. An exact sequence $F'\to F\to F''$ in $\mathscr{P}(\Omega)$ is transformed to an exact sequence $CF'\to CF\to CF''$ of complexes.*

(b) *Given an exact sequence $(\xi)=0\to F'\xrightarrow{\alpha}F\xrightarrow{\beta}F''\to 0$ in $\mathscr{P}(\Omega)$, we obtain natural transformations $\Delta^n\xi\colon \underline{H}^n_\Omega F''\to \underline{H}^{n+1}_\Omega F'$ such that $\{\underline{H}^n_\Omega, \Delta^n\}$ is a connected sequence of functors.*

(c) *Let $f, g\colon S\to T$ be maps in Ω, and let A be an object of $\mathscr{C}$. Then f, g induce maps $CF(f, A)$, $CF(g, A)\colon CF(S, A)\to CF(T, A)$ which are chain homotopic. In particular,*

$$\underline{H}^n_\Omega F(f,-)=\underline{H}^n_\Omega F(g,-)\colon\ \underline{H}^n_\Omega F(S,-)\to H^n_\Omega F(T,-).$$

Proof. (a) and (b) follow by straightforward arguments and obvious constructions. (c) is proved using AMITSUR's proof of a corresponding fact

about Amitsur cohomology [2, Th. 2.7] – the introduction of the additional differentiation operators δ^n does not interfere with the proof since the ε^i's and Δ^j's commute. An explicit proof is contained in [10, Prop. 3.10].

Let ω be the initial object of Ω, so that $\omega^n=\omega$. Then for G a functor from Ω to Ab, $d^nG(\omega)$ is either the zero map or the identity map, depending on whether n is odd or even. It is not difficult to see that $\underline{H}^n_\Omega G(\omega, A)\cong H^{n+1}G\omega(A)$ i.e. $\underline{H}^n_\Omega G(\omega, -)\sim H^{n+1}(G(\omega, -))$.

Let F, S, A, be objects of $\mathscr{P}(\Omega)$, Ω, $\mathscr{C}$, respectively. Let $kF(S, A)=$kernel $\{d^0FA(S)\colon F(S, A)\to F(S^2, A)\}$. It is straightforward to verify that the correspondence $F\to kF$ induces a functor $k\colon \mathscr{P}(\Omega)\to\mathscr{P}(\Omega)$, and we obtain a natural transformation $\lambda FA\colon kF(-, A)\to F(-, A)$ given by $\lambda FA(S)=$inclusion map of $kF(S, A)$ into $F(S, A)$. Moreover, it is not difficult to verify that.

$$0\to kF(S, -)\xrightarrow{\lambda F(-)(S)} F(S, -)\xrightarrow{d^0F(-)(S)} F(S^2, -)$$

is an exact sequence of objects and maps in $\mathscr{P}$. We know from Theorem 2 that h is left exact. Applying h to the sequence of functors just described, and using the definition of h, we obtain a commutative diagram, as below, with exact rows and columns (we write $hF(S, A)$ for $hF(S, -)(A)$)

$$\begin{array}{ccccccc}
 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & hkF(S, A) & \xrightarrow{h\lambda FA(S)} & hF(S, A) & \xrightarrow{hd^0FA(S)} & hF(S^2, A) \\
 & & \downarrow{\scriptstyle \iota kFS(A)} & & \downarrow{\scriptstyle \iota FS(A)} & & \downarrow{\scriptstyle \iota FS^2(A)} \\
0 & \to & kF(S, A) & \xrightarrow{\lambda FA(S)} & F(S, A) & \xrightarrow{d^0FA(S)} & F(S^2, A) \\
 & & \downarrow{\scriptstyle \delta kFS(A)} & & \downarrow{\scriptstyle \delta FS(A)} & & \downarrow{\scriptstyle \delta FS^2(A)} \\
0 & \to & kF(S, A^2) & \xrightarrow[\lambda FA^2(S)]{} & F(S, A^2) & \xrightarrow[d^0FA^2(S)]{} & F(S^2, A^2).
\end{array}$$

Since $\lambda FA(S)$, $\iota FS(A)$, etc. are inclusion maps, an easy diagram chase shows that $hkF(S, A)=$kernel $(h^0FA(S))\cap$kernel $(\delta FS(A))=\underline{H}^0_\Omega F(S, A)$. We know from Theorem 1 that for G any object of $\mathscr{P}$, hG is an additive functor. We have thus shown

Theorem 6. *$\underline{H}^0_\Omega F(S, -)$ is an additive functor. In particular, the correspondence $F\to\underline{H}^0_\Omega F$ defines a functor $\mathscr{P}(\Omega)\to\mathscr{AP}(\Omega)$, where $\mathscr{AP}(\Omega)$ is the category of functors from Ω to $\mathscr{AP}$.*

We now consider the right derived functors of H^0, which we denote by h^n_Ω, $n\geqq 0$ and $h^n_\Omega\colon \mathscr{P}(\Omega)\to\mathscr{AP}(\Omega)$.

Proposition 7. (a) *H^0 is left exact. Thus we may identify $\underline{H}^0_\Omega$ with h^0_Ω.*

(b) *Let S, T be objects of Ω, and f, $g\colon S\to T$ morphisms in Ω. Then for each $n\geqq 0$ and each object F of $\mathscr{P}(\Omega)$, $h^n_\Omega F(f, -)=h^n_\Omega F(g, -)$.*

Proof. Let $0\to F'\to F\to F''$ be an exact sequence in $\mathscr{P}(\Omega)$. By (a) of Theorem 5, it follows that this sequence induces an exact sequence $0\to\underline{H}^0_\Omega F'\to\underline{H}^0_\Omega F\to\underline{H}^0_\Omega F''$. This proves (a). To prove (b) we remark that by (c) of Theo-

rem 5, $\underline{H}^0_\Omega F(f,-)=\underline{H}^0_\Omega F(g,-)$. But $h^n_\Omega F(\mathrm{f},-)$ is induced by $\underline{H}^0_\Omega E_n(F,-)$, where $0\to F\to E_0\to E_1\to\cdots$ is an injective resolution of F. This proves (b).

We remark that by (a) we can identify $h^0_\Omega=\underline{H}^0_\Omega$.

For S, T objects of Ω, write $S<T$ if there exists a morphism $S\to T$ in Ω. We can define an equivalence relation on $|\Omega|$, the set of objects of Ω, by setting $S\sim T$ if $S<T$ and $T<S$. Let $\underset{\sim}{\Omega}$ denote the set of equivalence classes of object of $|\Omega|$. For S in $|\Omega|$, let $\underset{\sim}{S}$ denote the equivalence class of S in $\underset{\sim}{\Omega}$. We define a partial ordering on $\underset{\sim}{\Omega}$ by setting $\underset{\sim}{S}<\underset{\sim}{T}$ if $S<T-$ this is clearly well-defined. $\underset{\sim}{\Omega}$ is a directed set under this ordering, since $\underset{\sim}{S}$, $\underset{\sim}{T}<\underset{\sim}{U}$ where $U=S\otimes T$.

Let F be a functor in $\mathscr{P}(\Omega)$ which has the following property:

(A) Given S, T objects of Ω with f, g: $S\to T$ in Ω, then $F(f,-)=F(g,-)$.

We define a functor $\hat{F}$ in $\mathscr{P}(\Omega)$ by

$$\hat{F}(A)=\varinjlim_{\underset{\sim}{S}\in\underset{\sim}{\Omega}} F(S,A).$$

(It is clear that if $S\sim S'$, with f: $S\to S'$, g: $S'\to S$ in Ω, $F(S,A)=F(S',A)$ since $G(ff',-)=$identity, $G(f'f,-)=$identity.) For F an object of $\mathscr{A}\mathscr{P}(\Omega)$, $\hat{F}$ is an object of $\mathscr{A}\mathscr{P}$. For F an object of $\mathscr{P}(\Omega)$, we shall write $\underline{H}^iF$, $\underline{h}^iF$ for $\hat{H^i_\Omega F}$ and $\hat{h^i_\Omega F}$ respectively, $i\geqq 0$. Thus $\underline{H}^i$, $\underline{h}^i$ are functors from $\mathscr{P}(\Omega)$ to $\mathscr{P}-$ actually $\underline{h}^i$ is a functor to $\mathscr{A}\mathscr{P}$ as well. Note $\underline{h}^0=\underline{H}^0$.

Theorem 7. (a) *Let* $(\xi)=0\to F\xrightarrow{\alpha}F\xrightarrow{\beta}F''\to 0$ *be an exact sequence in* $\mathscr{P}(\Omega)$. *Then there exist natural transformations* $\Delta_n\xi$: $h^n_\Omega F''\to h^{n+1}_\Omega F'$ *such that* $\{h^n_\Omega,\Delta_n\}$ *is a connected sequence of functors. If* E *is injective in* $\mathscr{P}(\Omega)$, $h^n_\Omega E=0$ *for* $n>0$.

(b) *There exist natural transformations of functors,* Ψ_n: $h^n_\Omega\to\underline{H}^n_\Omega$ *such that*

(i) $\Psi_0=$*identity*, (ii) Ψ_1 *is monic.*

(c) *Given an exact sequence* (ξ) *as in* (a), *the following diagram commutes for each* $n\geqq 0$

$$\begin{array}{ccccccc}
h^n_\Omega F' & \xrightarrow{h^n_\Omega\alpha} & h^n_\Omega F & \xrightarrow{h^n_\Omega\beta} & h^n_\Omega F'' & \xrightarrow{\Delta_n\xi} & h^{n+1}_\Omega F' \\
\downarrow{\scriptstyle \Psi_n F'} & & \downarrow{\scriptstyle \Psi_n F} & & \downarrow{\scriptstyle \Psi_n F''} & & \downarrow{\scriptstyle \Psi_{n+1}F'} \\
\underline{H}^n_\Omega F' & \xrightarrow[\underline{H}^n_\Omega\alpha]{} & \underline{H}^n_\Omega F & \xrightarrow[\underline{H}^n_\Omega\beta]{} & \underline{H}^n_\Omega F'' & \xrightarrow[\Delta^n\xi]{} & \underline{H}^{n+1}_\Omega F'.
\end{array}$$

Proof. The proof is as in Theorems 3 and 4.

Let $0\to F'\to F\to F''\to 0$ be an exact sequence in $\mathscr{P}(\Omega)$. For each object S of Ω and for each object A of $\mathscr{C}$, $0\to F'(S,A)\to F(S,A)\to F''(S,A)\to 0$ is an exact sequence of abelian groups. Since taking direct limits over a given directed set $\underset{\sim}{\Omega}$ yields an exact functor, we obtain an exact sequence $0\to\hat{F}'(A)\to\hat{F}(A)\to\hat{F}''(A)\to 0$. Thus $\hat{}$ is an exact functor. Relating this to the last theorem we obtain

Theorem 8. (a) *Let* $(\xi)=0\to F'\to F\to F''\to 0$ *be an exact sequence in* $\mathscr{P}(\Omega)$. *Then there exist natural transformations* $\delta_n\xi$: $\underline{h}^nF''\to\underline{h}^{n+1}F'$ *such that* $\{\underline{h}^n, \delta_n\}$ *is a connected sequence of functors.*

(b) *Let* (ξ) *be as above. Then there exist natural transformations* $\delta^n\xi$: $\underline{H}^nF''\to\underline{H}^{n+1}F'$ *such that* $\{\underline{H}^n, \delta^n\}$ *is a connected sequence of functors.*

(c) *There exist natural transformations* χ_n: $\underline{h}^n\to\underline{H}^n$ *such that for* (ξ) *a sequence as above, the following commutes for each* $n\geqq 0$:

$$\begin{array}{ccccccc}
\underline{h}^nF' & \to & \underline{h}^nF & \to & \underline{h}^nF'' & \to & \underline{h}^{n+1}F' \\
\downarrow{\scriptstyle \chi_nF'} & & \downarrow{\scriptstyle \chi_nF} & & \downarrow{\scriptstyle \chi_nF''} & & \downarrow{\scriptstyle \chi_{n+1}F'} \\
\underline{H}^nF' & \to & \underline{H}^nF & \to & \underline{H}^nF'' & \to & \underline{H}^{n+1}F'.
\end{array}$$

Moreover $\chi_0=$*identity*, χ_1 *is monic.*

§ 3. Groups of Galois Extensions

Let R be a commutative ring. Let A be an R-algebra and G a finite group of R-algebra automorphisms of A. Let $e_G(A)$ be the set of all maps from G to A. We say that A *is a Galois extension of* R *with Galois group* G if $A^G=R$, where $A^G=\{x\in A\mid \sigma(x)=x \text{ for all } \sigma \text{ in } G\}$, and if the map h defined below is a bijection

h: $A\otimes A\to e_G(A)$ by $h(x\otimes y)(\sigma)=x\sigma(y)$ for x, y in A, σ in G.

A Galois extension A of R with Galois group G is said to be a Galois (R, G)-algebra, or to have a normal base, if $A\cong R(G)$ as an $R(G)$-module.

If A and B are Galois extensions of R with Galois group G, we shall say that A and B are equivalent if there exists an isomorphism j: $A\to B$ which is an R-algebra and an $R(G)$-module map. We will use [4] and [8] as our reference for facts about Galois extensions. We will outline here some properties and constructions which we will need later.

We now assume G is a finite abelian group. We let $E(R, G)$ denote the set of equivalence classes of Galois extensions of R with Galois group G, the equivalence being that of isomorphic objects, as above. $A(R, G)$ denotes the subset of $E(R, G)$ of extensions with normal bases, $T(R, G)$ denotes the set of extensions which are commutative R-algebras, and $N(R, G)=A(R, G)\cap T(R, G)$.

As shown in [10, §1], $E(R, G)$, $A(R, G)$, etc. are abelian groups. The zero element of these groups is the class of $e_G(R)=\{\text{Set maps from } G \text{ to } R\}$. Given a map ϕ: $G\to H$ of finite abelian groups, it induces group homomorphisms $E(R, \phi)$: $E(R, G)\to E(R, H)$, $A(R, \phi)$: $A(R, G)\to A(R, H)$, etc. Given a ring homomorphism f: $R\to S$, it induces group homomorphisms $E(f, G)$: $E(R, G)\to E(S, G)$, etc. Explicitly, $E(f, G)$ is given by $E(f, G)(A)=(S\otimes A)$, where $(S\otimes A)$ denotes the equivalence class of $S\otimes A$. We have inclusion homomorphisms such as $N(R, G)\to T(R, G)$, $A(R, G)\to E(R, G)$, etc. The following facts hold [10, §1]

Theorem 9. (a) *With $E(R, G)$, $E(R, \phi)$, etc. as defined above, we obtain bifunctors E, A, T, N: $\Omega \times fAb \to Ab$, where Ω is any subcategory of the category of commutative rings, and fAb is the category of finite abelian groups.*

(b) *The various inclusions $N(R, G) \to T(R, G)$, etc. induce natural transformations of bifunctors.*

(c) *If R is semi-local, the natural transformation $A(R, -) \to E(R, -)$ is a natural equivalence i.e. every Galois extension of R has a normal base.*

(d) *If G is cyclic, the transformation $T(-, G) \to E(-, G)$ is an equivalence.*

For Ω as in (a) of Theorem 9, we define a functor U: $\Omega \times fAb \to Ab$ by $U(R, G) = U(R(G))$, the (multiplicative) group of units of the group ring $R(G)$. If f: $R \to S$ is a ring homomorphism, $U(f, G)$: $U(R, G) \to U(S, G)$ is defined by

$$U(f, G)\Big(\sum_{\sigma \in G} a_\sigma \sigma\Big) = \sum_\sigma f(a_\sigma)\sigma .$$

For ϕ: $G \to H$ a map of finite abelian groups, $U(R, \phi)$: $U(R, G) \to U(R, H)$ is given by

$$U(R, \phi)\Big(\sum_\sigma a_\sigma \sigma\Big) = \sum_{\tau \in H} \sum_{\phi(\sigma) = \tau} a_\sigma \tau .$$

From [5, Cor. 4.8] and [10, Th. 2.2] we obtain:

Theorem 10. *For R a commutative ring and G a finite abelian group there is an isomorphism of abelian groups $H^2 UR(G) \cong A(R, G)$ which determines a natural transformation of functors $H^2 U \sim A$.*

Now let R be a commutative ring. Let $\Omega(R)$ be the category whose objects are R-algebras of the form

$$\sum_{i=1}^{n} \oplus R_{x_i},$$

where x_i is in R and

$$\sum_{i=1}^{n} x_i = 1,$$

n is some integer, and R_x denotes the localization of R at the set $\{1, x, x^2, \ldots\}$. The maps in $\Omega(R)$ are taken to be the R-algebra homomorphism. $\Omega(R)$ is a small category with finite direct sums (tensor products) and initial object R. We may apply the constructions of Section 2 to $\Omega(R)$. In particular, let U be the units functor defined above, U: $\Omega(R) \times fAb \to Ab$. We have from [10, Cor. 3.12]:

Theorem 11. *There is a natural isomorphism of functors $\underline{H}^1_{\Omega(R)} U(R, G) \cong E(R, G)$.*

§ 4. Commutative Galois Extensions

Putting together the results of Theorems 4 and 10 we obtain a natural monomorphism $h^1 UR(G) \to A(R, G)$. If G is cyclic, by (d) of Theorem 9, we have a monomorphism $h^1 UR(G) \to N(R, G)$. Likewise, by Theorems 7 and 11 we obtain a monomorphism $\underline{h}^1 U(R, G) \to T(R, G)$ for G cyclic. Our aim is to

show that these monomorphisms are isomorphisms for G cyclic, and that they induce isomorphisms for all finite abelian groups G.

Lemma 12. *Let F, L be functors from fAb to Ab, with F additive. Suppose we have a natural transformation $\alpha = F \to L$ such that αG is an isomorphism for G a finite cyclic group. Then there is a natural equivalence $F \sim hL$.*

Proof. By Theorem 1, $F \sim hF$. α induces $h\alpha$: $hF \to hL$ and thus a natural transformation β: $F \to hL$ such that the diagram below commutes

$$\begin{array}{ccc} hF & \xrightarrow{h\alpha} & hL \\ & \overset{\beta}{\nearrow} & \downarrow \iota L \\ F & \xrightarrow{\alpha} & L \end{array}$$

ιL is a monomorphism, and from the assumptions it follows easily that βG is an isomorphism for G cyclic. Since F and hL are additive, β is an isomorphism by the fundamental theorem of abelian groups.

Corollary. *There exist natural equivalences $hE(R, -) \sim T(R, -)$, $hA(R, -) \sim N(R, -)$.*

Proof. $T(R, -)$, $N(R, -)$ are additive functors from fAb to Ab [8, § 1], [10, § 1]. The result now follows from Theorem 9.

$h^1 UR$ is an additive functor from fAb to Ab, so $hh^1 UR \sim h^1 UR$. Applying h to the composite transformation $h^1 UR \to H^2 UR \sim A(R, -)$ and using the above corollary, we obtain a natural transformation α: $h^1 UR \to N(R, -)$. α is monic since the map $h^1 UR \to H^2 UR$ is monic and since h is left exact. To show that α is an isomorphism, it suffices to show that αG is an isomorphism for G a finite cyclic group. Thus it suffices to show that αG is onto for G a finite cyclic group. But clearly, to show this it is sufficient to conclude that the map $h^1 UR(G) \to H^2 UR(G)$ is onto for G a finite cyclic group.

Lemma 13. *Let x be an element of $H^2 UR(G)$, where G is a finite cyclic group. Then there exists a functor F: $fAb \to Ab$ and an embedding $UR \to F$ such that $x \to 0$ under the map $H^2 UR(G) \to H^2 F$.*

Proof. We have an isomorphism $H^2 UR(G) \cong A(R, G)$. There is then a Galois extension S of R with Galois group G such that $x \to (S)$ under the isomorphism. S is commutative since G is cyclic. Also $S \otimes S$ is by the definition of Galois extension simply the trivial element of $A(S, G)$ i.e. under the map $A(R, G) \to A(S, G)$, $(S) \to 0$. Let F be the functor US i.e. $F(J) = U(S(J))$. The desired result follows from the commutativity of the diagram

$$\begin{array}{ccc} H^2 UR(G) & \to & H^2 US(G) \\ \cong \downarrow & & \downarrow \\ A(R, G) & \to & A(S, G). \end{array}$$

Theorem 14. *There exists a natural isomorphism $h^1 UR(G) \cong N(R, G)$ for G any finite abelian group.*

Proof. By the discussion preceding Lemma 13, it suffices to show that the map $h^1 UR(G) \to H^2 UR(G)$ is onto for G cyclic. Let x be in $H^2 UR(G)$, and let F in $\mathscr{P}(fAb)$ be chosen as in the statement of Lemma 13. We have an exact sequence $(\xi) = 0 \to UR \to F \to Q \to 0$ in $\mathscr{P}(fAb)$. From (c) and (d) of Theorem 4 we obtain a commutative diagram with exact rows

$$\begin{array}{ccccc} h^0 Q(G) & \xrightarrow{\nabla_0 \xi} & h^1 UR(G) & \longrightarrow & h^1 F(G) \\ \| & & \downarrow{\scriptstyle \phi_1 UR(G)} & & \downarrow{\scriptstyle \phi_1 F(G)} \\ H^1 Q(G) & \xrightarrow[\nabla^0 \xi]{} & H^2 UR(G) & \longrightarrow & H^2 F(G). \end{array}$$

Since the map $h^1 UR(G) \to H^2 UR(G)$ is monic, and $x \to 0$ under the map $H^2 UR(G) \to H^2 F(G)$, it is easy to see that there is an element y in $h^1 UR(G)$ such that $(\phi_1 UR(G))(y) = x$. Thus we are done.

If we now use the isomorphism $\underline{H}^1_{\Omega(R)} U(R, G) \cong E(R, G)$ of Theorem 11 we can apply exactly the arguments above (but with use of Theorem 7 replacing that of Theorem 4, for example) to obtain

Theorem 15. *There exists a natural isomorphism $h^1_{\Omega(R)} U(R, G) \cong T(R, G)$ for G any finite abelian group.*

Bibliography

1. AMITSUR, S. A.: Homology groups and double complexes for arbitrary fields. J. Math. Soc. Japan **14**, 1–25 (1962).
2. ARTIN, M.: Grothendieck topologies and cohomology. Lecture notes, Harvard Univ. 1962.
3. CARTAN, H., and S. EILENBERG: Homological algebra. Princeton: Princeton Univ. Press 1956.
4. CHASE, S. U., D. K. HARRISON, and ALEX ROSENBERG: Galois theory and Galois cohomology of commutative rings. Memoirs Amer. Math. Soc. **52**, 15–33 (1965).
5. –, and ALEX ROSENBERG: A theorem of Harrison, Kummer theory, and Galois algebras. Nagoya Math. J. **27**–2, 663–685 (1966).
6. FREYD, P.: Abelian categories. New York: Harper and Row 1964.
7. HARRISON, D. K.: Abelian extensions of arbitrary fields. Trans. Amer. Math. Soc. **106**, 230–235 (1963).
8. – Abelian extensions of commutative rings. Memoirs Amer. Math. Soc. **52**, 1–14 (1965).
9. NORTHCOTT, D. G.: An introduction to homological algebra. Cambridge: Cambridge Univ. Press 1960.
10. ORZECH, M.: A cohomological description of abelian Galois extensions. Trans. Amer. Math. Soc., (to appear).

Dr. M. ORZECH
Department of Mathematics
University of Illinois
Urbana, Illinois 61801, USA

Math. Zeitschr. 105, 141—149 (1968)

Über Strömungen in Gruppen, speziell in symmetrischen Gruppen

HORST MICHEL

Eingegangen am 7. September 1965, in veränderter Fassung am 22. Februar 1968

In linear geordneten Gruppen G, die nach NEUMANN [12] als O-Gruppen bezeichnet werden, gilt ein Satz von LEVI [9], demzufolge jede Gleichung $x^n=a$ ($a\in G$, n natürliche Zahl) höchstens eine Lösung in G, also a höchstens eine n-te Wurzel in G hat. Aus diesem Satz folgt, daß jedes Element a einer O-Gruppe G in eine eindeutig bestimmte Untergruppe S von G eingebettet werden kann, die zu einer Untergruppe H der additiven Gruppe Q_+ der rationalen Zahlen isomorph ist und alle in G existierenden rationalen Potenzen von a enthält, für die also $N(a, S)=N(a, G)$ gilt, wenn allgemein $N(a, G)$ die Menge der natürlichen Zahlen bezeichnet, für die $x^n=a$ eine Lösung in G hat (G Gruppe, $a\in G$).

In einer beliebigen Gruppe G läßt es sich nicht immer erreichen, daß zu jedem $a\in G$ eine zu einer Untergruppe H von Q_+ epimorphe (viel weniger isomorphe) Gruppe $S\subseteq G$ mit $N(a, S)=N(a, G)$ existiert. Es ist das Ziel dieser Arbeit,

a) notwendige und hinreichende Bedingungen dafür herzuleiten, daß zu $a\in G$ eine G-vollständige Strömung durch a, d. h. eine einparametrige abelsche Gruppe

$$S=\{a^{[\gamma]}\in G\,|\,\gamma\in H\}$$

existiert, so daß $a^{[1]}=a$ und $a^{[\gamma_1+\gamma_2]}=a^{[\gamma_1]}a^{[\gamma_2]}$ gilt, wobei als Parameterbereich H eine 1 enthaltende Untergruppe von Q_+ benutzt wird und die Zahlen $1/k$ ($k\in N(a, G)$) ein Erzeugendensystem in H bilden,

b) zu zeigen, daß in den symmetrischen Gruppen $\mathfrak{S}^\Omega$ (mit Ω von beliebiger Mächtigkeit) jedes Element P in eine $\mathfrak{S}^\Omega$-vollständige Strömung durch P einbettbar ist,

c) zu jeder 1 enthaltenden Untergruppe H von Q_+ eine Permutation P in einer „genügend großen" symmetrischen Gruppe zu konstruieren, so daß die $\mathfrak{S}^\Omega$-vollständige Strömung durch P gerade H zum Parameterbereich hat.

Mit der Lösungstheorie der Translationsgleichung

$$F[F(a, \gamma_1), \gamma_2]=F(a, \gamma_1+\gamma_2) \tag{1}$$

(vgl. [1], S. 170ff.) und der Theorie der Iterationsgruppen (vgl. z. B. [7], S. 37ff. und die dort angegebene Literatur) stehen die hier behandelten Fragen ersichtlich in engem Zusammenhang, wenn in (1) $a\in G$ und $\gamma_1, \gamma_2\in Q_+$ sind.

Analoge Berührungspunkte mit Strömungsproblemen der Ergodentheorie (z. B. Existenz einer n-ten Wurzel einer maßtreuen Transformation und ihre Einbettung in eine einparametrige Gruppe; [5], S. 97) sind offenbar. Im Hinblick auf diese wichtige Anwendung wurde auch die Bezeichnung Strömung für solche einparametrigen Untergruppen gewählt (vgl. auch [6], S. 1).

Bezüglich anderer Untersuchungen über Wurzeln in Gruppen sei auf die Arbeiten von NEUMANN [13] und BAUMSLAG [2–4] verwiesen. Den Herren Prof. Dr. H. WIELANDT und Prof. Dr. B. H. NEUMANN dankt der Verfasser für freundliche Ratschläge und Hinweise bei der Abfassung dieser Arbeit.

Bezeichnungen

Q_+	additive Gruppe der rationalen Zahlen
Z	additive Gruppe der ganzen Zahlen
N	Menge der natürlichen Zahlen
N^0	Menge der nichtnegativen ganzen Zahlen
$\{a_1, \ldots, a_k\}$	die von den Elementen $a_1, \ldots, a_k$ erzeugte Gruppe
$(k_1, \ldots, k_n)$	größter gemeinsamer Teiler der $k_1, \ldots, k_n \in N$
$[k_1, \ldots, k_n]$	größtes gemeinsames Vielfaches der $k_1, \ldots, k_n \in N$
$\{p_i\}, (i \in N)$	Menge aller Primzahlen in natürlicher Reihenfolge.

1. Strömungen und G-vollständige Strömungen durch ein Gruppenelement

Definition 1.1. *Es sei G eine Gruppe und $a \in G$. Dann sei*

$$N(a, G) = \{n \in N \mid \exists\, x \in G \text{ mit } x^n = a\}$$

und $[N(a, G)]$ bezeichne die von den Zahlen $1/n$ $(n \in N(a, G))$ erzeugte, 1 enthaltende Untergruppe von Q_+.

Definition 1.2. *Es sei G eine Gruppe, $a \in G$ und H eine 1 enthaltende Untergruppe von Q_+. Dann heiße eine einparametrige Untergruppe*

$$S = \{a^{[\gamma]} \in G \mid \gamma \in H\}$$

von G eine (H-) Strömung durch a, falls

$$a^{[1]} = a\,, \tag{2}$$

$$a^{[\gamma_1 + \gamma_2]} = a^{[\gamma_1]} a^{[\gamma_2]} \tag{3}$$

gilt.

Da G nicht torsionsfrei zu sein braucht, kann in S aus $\gamma_1 \neq \gamma_2$ sehr wohl $a^{[\gamma_1]} = a^{[\gamma_2]}$ folgen. Jede Strömung S ist wegen (3) abelsch. Jede a enthaltende Untergruppe S' von S ist auch eine Strömung durch a. Zu jedem a in einer

beliebigen Gruppe existiert eine Z-Strömung durch a: die von a erzeugte zyklische Untergruppe liefert ihre Elemente.

Für den in Definition 1.2 vorkommenden Parameterbereich H hat man folgende Inklusion:

Satz 1.3. *Ist S eine H-Strömung durch a in G, so gilt $N(a, S) \subseteq N(a, G)$ und $H \subseteq [N(a, G)]$.*

Beweis. Aus $n \in N(a, S)$ folgt $x^n = a$ mit $x \in S \subseteq G$, also auch $n \in N(a, G)$. Die zweite Inklusion ergibt sich wegen $H = [N(a, S)]$ aus der ersten.

Mit diesem Satz wird folgende Definition sinnvoll:

Definition 1.4. Eine Strömung S durch a in G heiße *eine G-vollständige Strömung* durch a, falls

$$N(a, S) = N(a, G)$$

gilt.

G-vollständige Strömungen durch Gruppenelemente a sind nicht immer möglich. Wählt man als G die von den zwei Elementen b, c erzeugte Gruppe mit der einzigen Relation $b^2 = c^3$, so ist G das verallgemeinerte freie Produkt der zyklischen Gruppen $\{b\}$, $\{c\}$ mit der durch $b^{2k} = c^{3k}$ $(k \in Z)$ definierten vereinigten Untergruppe. Mittels Normalwortdarstellung kann man zeigen, daß die Gleichung $x^n = a$ $(n \in N)$ zwar für $n = 2$, $n = 3$, nicht aber für $n = 6$ lösbar ist. Letzteres ist aber mit (3) nicht vereinbar, denn mit $a^{[1/2]}$ und $a^{[-1/3]}$ wäre auch ein Element $a^{[1/6]}$ in S.

Durch Verallgemeinerung eines in [11] benutzten Verfahrens ergibt sich aber folgendes Kriterium:

Satz 1.5. *Zu einem Element a einer Gruppe G existiert genau dann eine G-vollständige Strömung durch a, wenn eine (evtl. endliche) Folge $a = a_1, a_2, a_3, \ldots$ von Elementen aus G und eine (evtl. endliche) Folge $k_2, k_3, \ldots$ natürlicher Zahlen mit den Eigenschaften*

(4) $a_{i+1}^{k_{i+1}} = a_i$ $(i \in N)$,

(5) *zu jedem $k \in N(a, G)$ existiert ein i_0 mit $k \mid \prod_{j=2}^{i_0} k_i$*

existiert, wobei die Folge der a_i genau dann mit dem Index s abbricht, wenn die Folge der k_i mit dem Index s abbricht.

Beweis. 1. Gibt es eine G-vollständige Strömung durch a, so wähle man als Folge $k_2, k_3, \ldots$ eine Folge von Primzahlen, in der jedes p_i genau λ mal vorkommt, wobei λ durch $p_i^{\lambda} \in N(a, G)$ und $p_i^{\lambda+1} \notin N(a, G)$ festgelegt ist. Ist $p_i^{\lambda} \in N(a, G)$ für alle $\lambda \in N$, so richte man die Folge $k_2, k_3, \ldots$ derart ein, daß p_i abzählbar unendlich oft in ihr vorkommt. Diese Folge hat die Eigenschaft (5), da mit k auch alle Teiler von k zu $N(a, G)$ gehören. Für die Folge $a_2, a_3, \ldots$ wähle man $a_i = a^{[1/\gamma_i]}$ mit

$$\gamma_i = \prod_{j=2}^{i} k_j,$$

denn $1/\gamma_i$ liegt nach Konstruktion der k_i sicher in $[N(a, G)]$, also die a_i $(i=2, 3, \ldots)$ in G. Dann gilt

$$a_{i+1}^{k_{i+1}}=(a^{[1/\gamma_{i+1}]})^{k_{i+1}}=(a^{[1/\gamma_i k_{i+1}]})^{k_{i+1}}=a^{[1/\gamma_i]}=a_i,$$

also ist (4) auch erfüllt.

2. Aus (4) und (5) folgt, daß jedes $\gamma\in[N(a, G)]$ eine Darstellung $\gamma=m'/k$ mit $m'\in Z$, $k\in N(a, G)$ hat, für die nach (5) auch

$$\gamma=m\Big/\prod_{j=2}^{i_0}k_i \tag{6}$$

geschrieben werden kann. Setzt man nun

$$a^{[\gamma]}=(a_{i_0})^m, \tag{7}$$

so ist diese Definition wegen (4) von der Darstellung (6) von γ (die ja nicht eindeutig zu sein braucht!) nicht abhängig, obwohl ihre Parameter m, i_0 in (7) eingehen. Zieht man (4) heran, so folgen sofort die Beziehungen (2) und (3).

Für eine G-vollständige Strömung durch a ist offenbar folgende Bedingung notwendig: Aus $m, n\in N(a, G)$ folgt $[m, n]\in N(a, G)$. Diese Bedingung ist aber — wie das folgende Beispiel zeigt — nicht hinreichend.

Es sei $G=\{b_1, b_2, \ldots\}$ die von den abzählbar vielen Elementen b_i erzeugte Gruppe mit den alleinigen Relationen

$$b_1^2=b_2^{2^2}=b_3^{2^3}=\cdots.$$

Dann ist G das verallgemeinerte freie Produkt der zyklischen Untergruppen $\{b_i\}$ $(i\in N)$ mit der durch

$$b_1^{2k}=b_2^{2^2k}=b_3^{2^3k}=\cdots$$

definierten vereinigten Untergruppe. Setzt man in G $a=b_1^2=b_2^{2^2}=b_3^{2^3}=\cdots$, so ist offenbar $2^\lambda\in N(a, G)$ für alle $\lambda\in N$ und $p_i^\lambda\notin N(a, G)$ $(i=2, 3, \ldots; \lambda\in N)$, wie man mit der Normalwortdarstellung zeigen kann. $[N(a, G)]$ ist also die Gruppe aller rationalen Zahlen der Form $m\cdot 2^{-k}$ $(m\in Z, k\in N)$ und a hat mit jeder m-ten und jeder n-ten Wurzel auch eine $[m, n]$-te Wurzel. Existierte eine G-vollständige Strömung durch a, so müßte (z.B.) eine 2-te Wurzel b von a in dieser Gruppe existieren, die selbst beliebige 2^λ-te Wurzeln $(\lambda=1, 2, \ldots)$ hat. Die einzigen 2-ten Wurzeln von a in G sind aber $b_1, b_2^2, b_3^{2^2}, \ldots$, wie man mittels Normalwortdarstellung zeigen kann. Von diesen hat aber nicht eine 2^λ-te Wurzeln mit unbeschränkt großem $\lambda\in N$.

2. Hinreichende Bedingung für die Existenz G-vollständiger Strömungen

Die Bedingungen (4) und (5) von Satz 1.5 sind etwas kompliziert in der Handhabung. Es ist daher wünschenswert, eine auch nur hinreichende Bedingung für die Existenz G-vollständiger Strömungen zu haben. Zu ihrer Formulierung bedient man sich zweckmäßig gewisser Erzeugendensysteme für die Untergruppen von Q_+.

Jeder additiven, 1 enthaltenden Untergruppe H von Q_+ kann umkehrbar eindeutig eine auf der Menge aller Primzahlen p_i, die in natürlichen Reihenfolge indiziert seien, definierte Funktion $\varepsilon(p_i)$ mit

$$\varepsilon(p_i)=\begin{cases}0, & \text{falls } p_i^{-1}\notin H\\ l, & \text{falls } p_i^{-l}\in H \text{ und } p_i^{-l-1}\notin H\\ \infty, & \text{falls } p_i^{-l}\in H \text{ für alle } l\in N\end{cases}\qquad (i\in N) \tag{8}$$

zugeordnet werden. Dann ist H erzeugbar durch alle Zahlen $1/k$ $(k\in N)$, für welche

$$k=\prod_{i=1}^{\infty} p_i^{\kappa_i} \quad \text{und} \quad 0\leqq\kappa_i\leqq\varepsilon(p_i), \qquad (i\in N) \tag{9}$$

gilt (vgl. [8], S. 224ff.).

Definition 2.1. *Ist G eine Gruppe und $a\in G$, so bezeichne $\varepsilon_a(p_i)$ das System der Größen* (8), *die man erhält, wenn $H=[N(a, G)]$ gesetzt wird.*

Satz 2.2. *Sind a, b Elemente einer Gruppe G, ist $k\in N$, $b^k=a$ und*

$$k=\prod_{i=1}^{\infty} p_i^{\kappa_i},$$

so gilt $\varepsilon_b(p_i)\leqq\varepsilon_a(p_i)-\kappa_i$, $(i\in N)$.

Beweis. Aus der Annahme $\varepsilon_b(p_j)>\varepsilon_a(p_j)-\kappa_j$ für ein gewisses j würde $\varepsilon_b(p_j)<\infty$ und die Existenz einer $p_j^{\varepsilon_b(p_j)}$-ten Wurzel c von b folgen. c wäre aber $k\cdot p_j^{\varepsilon_b(p_j)}$-te Wurzel von a, woraus sich $\varepsilon_a(p_j)\geqq\varepsilon_b(p_j)+\kappa_j$ ergeben würde.

Definition 2.3. *Ein Element b einer Gruppe G heiße genau dann eine volle k-te $(k\in N)$ Wurzel von $a\in G$, wenn $b^k=a$ und*

$$\varepsilon_b(p_i)=\varepsilon_a(p_i)-\kappa_i \qquad \left(i\in N;\ k=\prod_{i=1}^{\infty} p_i^{\kappa_i}\right)$$

gilt.

Es kann sein, daß keine der zu einem $a\in G$ existierenden k-ten Wurzeln volle k-te Wurzel ist. So hat z. B. das in dem in Anschluß an Definition 1.4 konstruierten Beispiel mit a bezeichnete Element keine volle 2-te Wurzel.

Existiert eine volle k-te Wurzel, so brauchen nicht alle existierenden k-ten Wurzeln voll zu sein: Ist $H=\{a, b\}$, $ab=ba$, $H_1=\{a^2b^{-4}\}$, so ist in $G=H/H_1$ offenbar $(aH_1)^2=(b^2H_1)^2=a^2H_1$, aber aH_1 hat keine 2-te Wurzel wie b^2H_1.

Satz 2.4. *Besitzt in einer Gruppe G jedes Element b mit einer n-ten Wurzel c $(n\in N)$ auch stets eine volle n-te Wurzel $\bar{c}$ in G, so existiert zu jedem Element $a\in G$ eine G-vollständige Strömung durch a.*

Beweis. Sei $a\in G$ beliebig gewählt und $k_2, k_3, \ldots$ (wie im Beweis von Satz 1.5) eine Folge von Primzahlen, in der jedes p_i genau $\varepsilon_a(p_i)$ mal vorkommt. Dann ist (5) erfüllt. $a_1=a$ hat nach Voraussetzung eine volle k_2-te Wurzel a_2. Bezeichnen die κ_{ij} die Primfaktorexponenten von k_i, so ist insbesondere $\varepsilon_{a_2}(p_i)=\varepsilon_{a_1}(p_i)-\kappa_{2i}$ $(i\in N)$. Daher enthält die Folge $k_3, k_4, \ldots$ jede Primzahl

p_i so oft, wie $\varepsilon_{a_2}(p_i)$ angibt und a_2 hat nach Voraussetzung eine volle k_3-te Wurzel a_3, wobei $\varepsilon_{a_3}(p_i)=\varepsilon_{a_2}(p_i)-\kappa_{3\,i}$ $(i\in N)$ gilt. Durch Fortsetzung dieses Verfahrens erhält man eine Folge $a_1, a_2, \ldots$ von Elementen aus G mit der Eigenschaft (4). Mit Satz 1.5 hat man dann die Behauptung.

Die in Satz 2.4 gemachte Voraussetzung läßt sich wesentlich abschwächen für den Fall, daß nur Einbettbarkeit eines festen $a\in G$ in eine G-vollständige Strömung verlangt wird. Da die im Abschnitt 3 betrachteten symmetrischen Gruppen jedoch die Voraussetzung von Satz 2.4 erfüllen, gehen wir nicht weiter darauf ein.

3. *G*-vollständige Strömungen in symmetrischen Gruppen

Ist Ω eine Grundmenge beliebiger Mächtigkeit $|\Omega|$, so bildet die Menge aller Permutationen $P, Q, \ldots$ von Ω die symmetrische Gruppe $\mathfrak{S}^{\Omega}$. Jedes $P\in\mathfrak{S}^{\Omega}$ besitzt eine bis auf Anordnung eindeutige Zyklendarstellung

$$P=\prod_{\iota\in I} Z^{\iota}_{m_\iota},$$

wobei $Z^{\iota}_{m_\iota}$ einen Zyklus der Länge m_ι $(1\leqq m_\iota\leqq\aleph_0)$ bezeichne und $I=\{\iota\}$ eine der Menge der Zyklen von P gleichmächtige Menge ist. Die Menge der Zyklen der Länge m in P werde mit

$$\mathfrak{Z}_m(P)=\{Z^{\iota}_{m_\iota}\,|\,m_\iota=m\},\qquad 1\leqq m\leqq\aleph_0$$

bezeichnet, $|\mathfrak{Z}_m(P)|$ sei die Mächtigkeit dieser Menge.

Sind $k, m\in N$, $\{p_i\}$ $(i\in N)$ die Folge der Primzahlen in natürlicher Reihenfolge, κ_i bzw. μ_i $(i\in N)$ die Primfaktorexponenten von k bzw. m, so ist durch

$$d_{k,m}=\prod_{i=1}^{\infty} p_i^{\delta_i}\quad\text{und}\quad \delta_i=\begin{cases}\kappa_i & \text{für } \mu_i>0\\ 0 & \text{für } \mu_i=0\end{cases}$$

eine von k und m abhängende natürliche Zahl $d_{k,m}$ definiert, die insbesondere die Eigenschaft

$$[d_{k_1,m}, d_{k_2,m}, \ldots, d_{k_r,m}]=d_{[k_1,k_2,\ldots,k_r],m} \tag{10}$$

besitzt. Ergänzend werde $d_{k,\aleph_0}=k$ definiert. Mit diesen Bezeichnungen lautet ein von ŁOJASIEWICZ [10] bewiesener Satz über die Existenz von k-ten Wurzeln von Permutationen so:

Satz 3.1. (ŁOJASIEWICZ). *$P\in\mathfrak{S}^{\Omega}$ besitzt genau dann eine k-te Wurzel $Q\in\mathfrak{S}^{\Omega}$ $(Q^k=P)$, wenn für alle m mit $2\leqq m\leqq\aleph_0$ aus $|\mathfrak{Z}_m(P)|<\aleph_0$ stets $|\mathfrak{Z}_m(P)|\equiv 0$ $(d_{k,m})$ folgt.*

Formuliert man den Begriff der „G-vollständigen Strömung durch a" speziell in symmetrischen Gruppen, so ist die Einbettbarkeit einer beliebigen Permutation $P\in\mathfrak{S}^{\Omega}$ in eine $\mathfrak{S}^{\Omega}$-vollständige Strömung durch P schon gezeigt, wenn man die Voraussetzung von Satz 2.4 beweisen kann. Das soll im folgenden Satz geschehen.

Satz 3.2. *Jede Permutation $Q\in\mathfrak{S}^{\Omega}$, die eine k-te Wurzel Q' besitzt ($k\in N$), besitzt auch eine volle k-te Wurzel $\bar{Q}$.*

Beweis. Ist

$$k=\prod_{i=1}^{\infty} p_i^{\kappa_i},$$

so ist (vgl. Def. 2.1) $\varepsilon_Q(p_i)\geqq\kappa_i$ $(i\in N)$. Es ist zu zeigen, daß $\bar{Q}$ so gewählt werden kann, daß $\varepsilon_{\bar{Q}}(p_i)=\varepsilon_Q(p_i)-\kappa_i$ $(i\in N)$ gilt. Abkürzend werde im folgenden $\varepsilon_{\bar{Q}}(p_i)=\bar{\varepsilon}_i$, $\varepsilon_Q(p_i)=\varepsilon_i$ gesetzt. Sei nun m so gewählt, daß $|\mathfrak{Z}_m(Q)|<\aleph_0$ gilt. Dann ist es nach dem Beweis von Satz 3.1 in [10] möglich, eine solche k-te Wurzel $\bar{Q}$ zu konstruieren, daß dabei aus den $|\mathfrak{Z}_m(Q)|$ m-Zyklen von Q $|\mathfrak{Z}_m(Q)|/d_{k,m}$ $md_{k,m}$-Zyklen von $\bar{Q}$ werden. Es gilt bei dieser speziellen Konstruktion

$$|\mathfrak{Z}_{md_{k,m}}(\bar{Q})|=\frac{|\mathfrak{Z}_m(Q)|}{d_{k,m}}, \tag{11}$$

denn aus $m'd_{k,m'}=md_{k,m}$ folgt $m=m'$. Zur Berechnung von $\bar{\varepsilon}_j$ hat man $\bar{Q}$ auf p_j^{λ}-te Wurzeln zu untersuchen. Nach Satz 3.1 und (11) muß

$$|\mathfrak{Z}_m(Q)|\equiv 0 \quad (d_{p_j^{\lambda},md_{k,m}}\cdot d_{k,m})$$

gelten, woraus wegen $d_{p_j^{\lambda},md_{k,m}}\cdot d_{k,m}=d_{kp_j^{\lambda},m}$ die Kongruenz

$$|\mathfrak{Z}_m(Q)|\equiv 0 \quad (d_{kp_j^{\lambda},m}) \tag{12}$$

folgt. Für alle $\varepsilon_j<\infty$ gilt aber nach Voraussetzung $|\mathfrak{Z}_m(Q)|\equiv 0$ $(d_{p_j^{\varepsilon_j},m})$ sowie $|\mathfrak{Z}_m(Q)|\equiv 0$ $(d_{k,m})$. Daraus folgt mit (10)

$$|\mathfrak{Z}_m(Q)|\equiv 0 \quad (d_{kp_j^{\varepsilon_j-\kappa_j},m}).$$

In (12) darf also $\lambda=\varepsilon_j-\kappa_j$ gewählt werden. Durch einen analogen Schluß überzeugt man sich im Falle $\varepsilon_j=\infty$, daß λ beliebig groß gewählt werden darf. Damit ist $\bar{\varepsilon}_j\geqq\varepsilon_j-\kappa_j$ gezeigt. Nach Satz 2.2 folgt daraus $\bar{\varepsilon}_j=\varepsilon_j-\kappa_j$.

Mit Satz 2.4 gilt also der

Satz 3.3. *Zu jeder Permutation $P\in\mathfrak{S}^{\Omega}$ existiert eine $\mathfrak{S}^{\Omega}$-vollständige, d.h. eine $[N(P,\mathfrak{S}^{\Omega})]$-Strömung durch P.*

4. Typen $\mathfrak{S}^{\Omega}$-vollständiger Strömungen

Aus dem Einbettungssatz 3.3 ist nicht ersichtlich, ob in einer vorgegebenen Permutationsgruppe $\mathfrak{S}^{\Omega}$ zu einer vorgegebenen Gruppe $H\subseteq Q_+$ eine Permutation $P\in\mathfrak{S}^{\Omega}$ mit der Eigenschaft $[N(P,\mathfrak{S}^{\Omega})]=H$ existiert. Es wird sich zeigen, daß dies von $|\Omega|$ abhängt.

Satz 4.1. *Ist $P\in\mathfrak{S}^{\Omega}$ und $|\Omega|\leqq|\bar{\Omega}|$, so existiert ein $\bar{P}\in\mathfrak{S}^{\bar{\Omega}}$ mit $N(\bar{P},\mathfrak{S}^{\bar{\Omega}})=N(P,\mathfrak{S}^{\Omega})$.*

Beweis. In $\bar{\Omega}$ existiert eine zu Ω gleichmächtige Teilmenge Ω', auf der $\bar{P}$ „wie auf Ω" definiert werde. Die (evtl. verbleibende) Restmenge $\bar{\Omega}-\Omega'$ werde identisch auf sich abgebildet. Dann hat $\bar{P}$ (abgesehen von 1-Zyklen) dieselbe Zyklenstruktur wie P und da 1-Zyklen für die Existenz oder Nichtexistenz von k-ten Wurzeln (Satz 3.1) keine Rolle spielen, folgt die Behauptung.

Satz 4.2. *Zu jeder additiven, 1 enthaltenden Untergruppe H von Q_+ existiert in $\mathfrak{S}^\Omega$ mit $|\Omega|=\aleph_0$ eine Permutation P mit $H=[N(P, \mathfrak{S}^\Omega)]$.*

Beweis. Zu H existieren nach (8) Zahlen $\varepsilon(p_i)$ $(i\in N)$, die gemäß (9) ein Erzeugendensystem von H definieren. Man konstruiere $P\in\mathfrak{S}^\Omega$ so, daß für $m>1$

$$|\mathfrak{Z}_m(P)|=\begin{cases} p_j^{\varepsilon(p_j)}, & \text{falls } m=p_j \text{ und } \varepsilon(p_j)<\infty, \\ \aleph_0, & \text{falls } m=p_j \text{ und } \varepsilon(p_j)=\infty, \\ 0 & \text{sonst} \end{cases}$$

gilt. Dann ist

$$\sum_{n=2}^{\aleph_0} m\,|\mathfrak{Z}_m(P)|\leqq\aleph_0 .$$

Ergänzt man die Definition von P auf der evtl. verbleibenden Restmenge von Ω durch 1-Zyklen, so erhält man mit Def. 2.1 und Satz 3.1 die Gleichungen $\varepsilon_P(p_i)=\varepsilon(p_i)$ $(i\in N)$, die zur Behauptung äquivalent sind.

Man kann sich also auf $\mathfrak{S}^\Omega$ mit $|\Omega|=\aleph_0$ beschränken, wenn gewährleistet sein soll, daß in $\mathfrak{S}^\Omega$ zu jeder Gruppe $H\subseteq Q_+$ ein P mit $[N(P, \mathfrak{S}^\Omega)]=H$ existiert. Von Interesse ist der extreme Fall $H=Q_+$, der abschließend betrachtet werden soll.

Satz 4.3. *Zu $P\in\mathfrak{S}^\Omega$ existiert genau dann eine (notwendigerweise $\mathfrak{S}^\Omega$-vollständige) Q_+-Strömung durch P, wenn aus $2\leqq m\leqq\aleph_0$ und $|\mathfrak{Z}_m(P)|>0$ stets $|\mathfrak{Z}_m(P)|\geqq\aleph_0$ folgt.*

Beweis. 1. Es sei $[N(P, \mathfrak{S}^\Omega)]=Q_+$. Jede m teilende Primzahl p hat die Eigenschaft $p^{-\lambda}\in[N(P, \mathfrak{S}^\Omega)]$. Ist $|\mathfrak{Z}_m(P)|>0$, so muß nach Satz 3.1

$$|\mathfrak{Z}_m(P)|\equiv 0 \quad (d_{p^\lambda, m}), \qquad \lambda=1, 2, \ldots$$

gelten, was wegen $d_{p^\lambda, m}=p^\lambda$ mit der Endlichkeit von $|\mathfrak{Z}_m(P)|$ nicht vereinbar ist. Den gleichen Widerspruch erhält man wegen $d_{p^\lambda, \aleph_0}=p^\lambda$ auch im Falle $m=\aleph_0$.

2. Folgt aus $2\leqq m\leqq\aleph_0$ und $|\mathfrak{Z}_m(P)|>0$ stets $|\mathfrak{Z}_m(P)|\geqq\aleph_0$, so erhält man zu beliebigem $k\in N$ eine k-te Wurzel Q von P, wenn jeweils k m-Zyklen von P zu einem km-Zyklus von Q zusammengesetzt werden. Ergänzend werde auf den 1-Zyklen $Q=P$ gesetzt. Damit ist $N(P, \mathfrak{S}^\Omega)=N$, also $[N(P, \mathfrak{S}^\Omega)]=Q_+$.

Unmittelbar ergibt sich hieraus die

Folgerung 4.4. *Ist $[N(P, \mathfrak{S}^\Omega)]=Q_+$, so ist entweder $P=I$ oder $|\Omega|\geqq\aleph_0$.*

Literatur

1. ACZÉL, J.: Vorlesungen über Funktionalgleichungen und ihre Anwendungen. Berlin 1961.
2. BAUMSLAG, G.: Wreath products and p-groups. Proc. Cambr. Phil. Soc. **55**, 224—231 (1959).
3. — Roots and wreath products. Proc. Cambr. Phil. Soc. **56**, 109—117 (1960).
4. — Some aspects of groups with unique roots. Acta Math. **104**, 217—303 (1960).
5. HALMOS, P.R.: Lectures on ergodic theory. Tokyo 1956.
6. JACOBS, K.: Neuere Methoden und Ergebnisse der Ergodentheorie. Berlin 1960.
7. KUCZMA, M.: A survey of the theory of functional equations. Publ. El. Fak. Univ. Beograd, Ser. Mat. i Fiz., No. **130** (1964).
8. KUROSCH, A.G.: Gruppentheorie. Berlin 1953.
9. LEVI, F.W.: Ordered groups. Proc. Indian Acad. Sci., Sect. A **16**, 256—263 (1942).
10. ŁOJASIEWICZ, S.: Solution générale de l'équation fonctionelle $f(f(\ldots f(x)\ldots))=g(x)$. Ann. Soc. Polon. Math. **24**, 88—91 (1951).
11. MICHEL, H.: Über die Fixpunktstruktur von Iterationsgruppen reeller Funktionen. Arch. Math. **16**, 47—59 (1965).
12. NEUMANN, B.H.: On ordered groups. Amer. Journ. Math. **71**, 1—18 (1949).
13. — Adjunction of elements to groups. J. Lond. Math. Soc. **18**, 12—20 (1943).

Dr. HORST MICHEL
II. Mathematisches Institut der
Martin-Luther-Universität
X 402 Halle/Saale, Reichardtstr. 9

Math. Zeitschr. 105, 150—152 (1968)

Zur Verteilung quadratischer Reste

KLAUS BURDE

Eingegangen am 14. Oktober 1967

PERRON bewies in einer 1952 erschienenen Arbeit[1] die folgende Anordnungseigenschaft quadratischer Reste.

Satz A. *Sei* $p=4k-1$ *eine rationale Primzahl und*

$$r_1, r_2, \ldots, r_{2k}$$

ein vollständiges System quadratischer Reste modulo p (inklusive 0*). Für jede zu p prime Zahl a sind dann unter den Zahlen*

$$r_1+a, r_2+a, \ldots, r_{2k}+a$$

stets k Reste (eventuell 0 *dabei) und k Nichtreste modulo p.*

Satz B. *Sei* $p=4k+1$ *eine rationale Primzahl und*

$$r_1, r_2, \ldots, r_{2k+1}$$

ein vollständiges System quadratischer Reste modulo p. Ist dann a eine zu p prime Zahl, so enthält die Folge der $(2k+1)$ *Zahlen*

$$r_1+a, r_2+a, \ldots, r_{2k+1}+a,$$

falls a selbst Rest modulo p ist $(k+1)$ *Reste und k Nichtreste modulo p. Ist jedoch a Nichtrest modulo p, so sind es umgekehrt k Reste und* $(k+1)$ *Nichtreste modulo p.*

Die Null wird dabei zu den Resten gezählt.

Diese Sätze lassen sich als Orthogonalitätsrelation im $(p+1)$-dimensionalen Vektorraum V_{p+1} formulieren. Sie sind, wie man leicht nachprüft, äquivalent mit

Satz 1. *Mit den Vektoren* $\mathfrak{e}_i$; $i=0, \ldots, p$ *bilden auch die* $(p+1)$ *Vektoren*

$$\mathfrak{q}_i=\frac{1}{\sqrt{p}}\left\{\left(\frac{i}{p}\right)\mathfrak{e}_0+\left(\frac{i+1}{p}\right)\mathfrak{e}_1+\cdots+\left(\frac{i+p-1}{p}\right)\mathfrak{e}_{p-1}+\mathfrak{e}_p\right\}; \quad i=0, \ldots, p-1,$$

$$\mathfrak{q}_p=\frac{1}{\sqrt{p}}\{\mathfrak{e}_0+\mathfrak{e}_1+\cdots+\mathfrak{e}_{p-1}\}$$

eine orthonormierte Basis des V_{p+1}.

[1] PERRON, O.: Bemerkungen über die Verteilung quadratischer Reste. Mathematische Zeitschrift **56**, 122—130 (1952).

Hierin ist $(0/p)=0$ gesetzt, da nur dann keine Fallunterscheidungen auftreten. Dementsprechend soll im folgenden die Null weder zu den (quadratischen) Resten noch zu den Nichtresten gezählt werden.

Aus der Betrachtung der Komponentenzerlegung gewisser Vektoren $\mathfrak{a}\in V_{p+1}$ bezüglich der orthonormierten Basis der $\mathfrak{q}_i$ aus Satz 1 erhält man auf einfache Weise Verteilungssätze über Legendresymbole. So lassen sich z.B. aus den Komponentenzerlegungen von Vektoren der Form

$$V_{p+1}\ni\mathfrak{a}_k=\sum_{i=0}^{p}a_i\mathfrak{e}_i \quad \text{mit} \quad a_i=0 \qquad \text{für } i\geqq k<p$$

Aussagen über k-gliedrige Sequenzen von Legendresymbolen gewinnen. Die umfangreicheren Untersuchungen darüber werden Gegenstand einer späteren Arbeit sein. Hier soll der

Satz 2. *Zu jeder rat. Primzahl p gibt es eine Folge von $\frac{p-1}{2}$ aufeinander folgenden*[2] *primen Restklassen modulo p, unter denen die Restklassen, die aus quadratischen Resten bestehen, um mindestens $\{\frac{1}{2}\sqrt{p}\}$ überwiegen.*

mit der eben skizzierten Methode hergeleitet werden. Er ergibt sich aus folgendem elementaren Sachverhalt.

Hilfssatz. *Bilden die Vektoren $\mathfrak{e}_i$; $i=1,\ldots,n$ eine orthonormierte Basis des V_n, dann existiert zu jedem Vektor $\mathfrak{o}\neq\mathfrak{a}\in V_n$ unter den $\mathfrak{e}_i$ mindestens einer, dessen Komponente in Richtung von $\mathfrak{a}$ absolut genommen $\geqq 1/\sqrt{n}$ ist.*

Beweis. $\mathfrak{a}_0$ sei der Einheitsvektor in Richtung von $\mathfrak{a}$. Es gilt

$$\mathfrak{a}_0=a_1\mathfrak{e}_1+a_2\mathfrak{e}_2+\cdots+a_n\mathfrak{e}_n,$$

wobei die a_i die Komponenten der Vektoren $\mathfrak{e}_i$ in Richtung von $\mathfrak{a}$ sind. Aus

$$\mathfrak{a}_0^2=a_1^2+a_2^2+\cdots+a_n^2=1$$

folgt daher sofort die Behauptung des Hilfssatzes.

Zum Beweis von Satz 2 wenden wir den Hilfssatz auf die von den Vektoren $\mathfrak{q}_i$ aus Satz 1 gebildete orthonormierte Basis des V_{p+1} und den Vektor

$$V_{p+1}\ni\mathfrak{a}=\mathfrak{e}_0+\mathfrak{e}_1+\cdots+\mathfrak{e}_{\frac{p-1}{2}}-\mathfrak{e}_{\frac{p+1}{2}}-\cdots-\mathfrak{e}_{p-1}+\mathfrak{e}_p$$

an. Es gibt unter den Vektoren $\mathfrak{q}_i$; $i=0,\ldots,p$ mindestens einen, etwa $\mathfrak{q}_j$ mit[3]

$$\pm\mathfrak{a}\,\mathfrak{q}_j\geqq\frac{\sqrt{p+1}}{\sqrt{p+1}}=1.$$

Dieser Vektor ist wegen

$$\pm\mathfrak{a}\,\mathfrak{q}_p=\pm\frac{1}{\sqrt{p}}<1$$

[2] Die Folge der primen Restklassen modulo p ist dabei als zyklisch geschlossen aufgefaßt. Auf die Restklasse $[p-1]$ folgt also die Restklasse $[1]$.

Mit $\{\frac{1}{2}\sqrt{p}\}$ ist die am nächsten bei $\frac{1}{2}\sqrt{p}$ liegende ganze Zahl gemeint.

[3] Die folgenden Ungleichungen gelten für eines der beiden Vorzeichen.

sicher nicht q_p. Es folgt

$$\pm\left(\sum_{\nu=0}^{\frac{p-1}{2}}\left(\frac{j+\nu}{p}\right)-\sum_{\nu=\frac{p+1}{2}}^{p-1}\left(\frac{j+\nu}{p}\right)+1\right)>\sqrt{p};\qquad 0\leqq j<p,$$

$$\pm(P-N-(P'-N')+1)>\sqrt{p},$$

wobei P die Anzahl der positiven und N die Anzahl der negativen Legendresymbole in der ersten Summe, P' und N' die entsprechenden Anzahlen für die zweite Summe sind. Mit

$$P+P'=N+N'=\frac{p-1}{2}$$

erhalten wir

$$\pm(2(P-N)+1)>\sqrt{p}.$$

Es gilt also eine der beiden Ungleichungen

$$P-N>\tfrac{1}{2}(\sqrt{p}-1),\qquad N-P>\tfrac{1}{2}(\sqrt{p}+1).$$

In beiden Fällen läßt sich nun eine Folge von $\frac{p+1}{2}$ aufeinander folgenden Zahlen angeben, die mindestens $[\frac{1}{2}(\sqrt{p}+1)]=\{\frac{1}{2}\sqrt{p}\}$ mehr Reste als Nichtreste modulo p enthält, woraus dann Satz 2 folgt. Gilt die erste Ungleichung, so sind es die Zahlen

$$j, j+1, \ldots, j+\frac{p-1}{2},$$

gilt die zweite Ungleichung, so enthält diese Folge mindestens $\{\frac{1}{2}p\}+1$ mehr Nichtreste als Reste modulo p, die $\frac{p+1}{2}$ aufeinander folgenden Zahlen

$$j+\frac{p+1}{2}, \ldots, j+p$$

daher mindestens $\{\frac{1}{2}p\}$ mehr Reste als Nichtreste modulo p. Durch Anwendung des Hilfssatzes auf andere Vektoren $a\in V_{p+1}$ ergeben sich weitere Sätze dieser Art.

KLAUS BURDE
Institut für Math. D der TH
33 Braunschweig, Pockelsstraße 14

Math. Zeitschr. 105, 153–156 (1968)

On Normaloid Operators

V. ISTRĂŢESCU and I. ISTRĂŢESCU

Received September 29, 1967

1.

Let T be a bounded operator T on complex Hilbert space H. The numerical range $W(T)$ is the set $\{\langle Tx,x\rangle: \|x\|=1\}$ and the spectrum of T is $\sigma(T)$. By con E we denote the convex hull of the set E. An attempt to characterize those operators T for which con $\sigma(T)=\overline{W(T)}$ (the bar denotes the closure) was made by WINTER [10].

For this he introduces a class of operators, called normaloid operators. But an example due to HALMOS shows that this is not possible [8].

The aim of this Note is to obtain some structure theorems for a subclass of normaloid operators – the class of invariant-normaloid operators – which contains the class of hyponormal operators and operators of class (N).

2.

Definition 1. [10]. An operator T *is normaloid* if

$$\|T\|=\sup\{|\langle Tx,x\rangle|;\ \ \|x\|=1\}.$$

Definition 2. An operator T is *invariant-normaloid* if the restriction to every invariant subspace is also normaloid.

Now, we give some examples of invariant-normaloid operators.

a) *Hyponormal operators.* An operator T is hyponormal if, $\|T^*x\|\leqq\|Tx\|$ for every $x\in H$.

It is known ([9] Lemma 4), [1] that the restriction of T onto invariant subspace M is also hyponormal.

b) *Operators of class* (N) [4]. An operator T is of class (N) if

$$\|Tx\|^2\leqq\|T^2x\|$$

for every $x\in H$, $\|x\|=1$.

It is easy to see that every hyponormal operator is of class (N). Also it is obvious that the restriction of T onto an invariant subspace is of class (N).

c) *Operators of class* (N,k). An operator T is of class (N,k) if

$$\|Tx\|^k\leqq\|T^kx\|$$

for every $x\in H$, $\|x\|=1$.

All operators in a), b), c) are normaloid [1, 9, 4, 5].

3.

Let T be an operator wich is invariant-normaloid.

Theorem 1. *If T is compact then T is normal.*

Proof. Since T is normaloid, there exists $\lambda\in\sigma(T)$ such that $|\lambda|=\|T\|$. But T is compact. Then the subspace

$$m_T(\lambda)=\{x,\ x\in H,\ Tx=\lambda x\}$$

which is invariant for T and T^* is not equal to $\{0\}$. Consider the family of subspaces of the following form

$$m_T(\lambda)\cap m_{T^*}(\bar{\lambda})=M_T(\lambda)$$

and it is easy to see that $M_T(\lambda)\perp M_T(\lambda')$ for $\lambda\neq\lambda'$. Let H_0 be the subspace

$$\sum_{\lambda\in\sigma_p(T)} M_T(\lambda)$$

which reduces T and denote $H_1=H_0^{\perp}$. The restriction of T onto H_0 is clearly normal. For the proof of the theorem it is sufficient to show that $T/H_1=0$.

But $T/H_1=T_1$ is normaloid. Then there exists $\mu\in\sigma(T_1)$ such that

$$|\mu|=\|T_1\|$$

and $T_1 y=\mu y$ for some non-zero element y in H_1. But $T_1^* y=\bar{\mu}y$. It is clear, now that

$$m_T(\mu)\cap m_{T^*}(\bar{\mu})=M_T(\mu)$$

is not equal to $\{0\}$ and is also orthogonal to H_0. We obtain in this way a contradiction and the theorem is proved.

In a similar way we can prove the following theorems:

Theorem 2. *If $p(\lambda)$ is a polynomial non-vanishing on $\sigma(T)-\{0\}$ and $p(T)$ is a compact operator, then T is normal.*

Theorem 3. *If there exists a system of integers $p_1, q_1, \ldots, p_n, q_n$ such that*

$$T^{*p_1}T^{q_1}\ldots T^{*p_n}T^{q_n}$$

is a compact operator, then T is normal.

These theorems represent a generalization to our case of a result known for operators in the class a), b), c) [4, 6, 5].

4. Tensor Product of Normaloid Operators

If T_1 and T_2 are normaloid operators, appears naturally the problem of the structure of operator $T_1\otimes T_2$. We prove the following:

Theorem 4. *If T_1 and T_2 are normaloid operators, the operator $T_1 \otimes T_2$ is normaloid.*

Proof. We use the following theorem of BROWN and PEARCY [2]: $\sigma(T_1 \otimes T_2) = \sigma(T_1) \cdot \sigma(T_2)$ (which is true for arbitrary operators T_1 and T_2). Thus it is clear that

$$\sup_{\mu \in \sigma(T_1 \otimes T_2)} |\mu| = \|T_1\| \, \|T_2\|$$

But the operator $T_1 \otimes T_2$ has the property

$$\sup_{\mu \in \sigma(T_1 \otimes T_2)} |\mu| \leqq \|T_1 \otimes T_2\| \leqq \|T_1\| \, \|T_2\|$$

and thus $T_1 \otimes T_2$ is normaloid.

In [7] was remarked that if T_1 and T_2 are hyponormal operators then $T_1 \otimes T_2$ is also hyponormal and thus normaloid.

Another proof, more elementary than this may be based on the following:

Lemma. *If T_1, T_2 are any operators, then the approximate point spectra satisfy*

$$\Pi_{\mathrm{ap}}(T_1) \, \Pi_{\mathrm{ap}}(T_2) \subset \Pi_{\mathrm{ap}}(T_1 \otimes T_2).$$

Proof. Let $\mu_i \in \Pi_{ap}(T_i)$, $i = 1, 2$ and choose sequences of unit vectors $\{u_n\}, \{v_n\}$ with the property

$$\|(T_1 - \mu_1) u_n\| \to 0, \quad \|(T_2 - \mu_2) v_n\| \to 0.$$

Then $\{u_n \otimes v_n\}$ is a sequence of unit vectors, and applying to it the operator

$$T_1 \otimes T_2 - \mu_1 \mu_2 I = (T_1 - \mu_1 T) \otimes T_2 + \mu_1 I \otimes (T_2 - \mu_2 I)$$

ones sees that

$$\mu_1 \mu_2 \in \Pi_{\mathrm{ap}}(T_1 \otimes T_2).$$

It in particular this furnishes a proof of Theorem 4 without using the theorem of BROWN and PEARCY [2].

Indeed if the T_i are normaloid and $|\mu_i| = \|T_i\|$ then

$$|\mu_1| \, |\mu_2| = \|T_1\| \, \|T_2\| = \|T_1 \otimes T_2\|$$

shows that the conclusion of the Theorem 4 is true.

We are indebted to Professor S.K. BERBERIAN which kindly communicated these to the authors.

References

1. BERBERIAN, S.K.: A note on hyponormal operators. Pacif. J. of Math. **12**(4), 1171–1175 (1962).
2. BROWN, A., and C. PEARCY: Spectra of tensor products of operators. Prof. Amer. Math. Soc. **17**, 162–166 (1966).
3. FURUTA, T.: On the class of Paranormal operators (to appear).
4. ISTRĂŢESCU, V.: On some hyponormal operators. Pacif. J. of Math. **22**, 413–417 (1967).

5. ISTRĂŢESCU, V., and I. ISTRĂŢESCU: On some classes of operators, I, II. (To appear.)
6. —, T. SAITO, and T. YOSHINO: On a class of operators. Tahoku Math. J. **18**(4), 410—414 (1966).
7. SAITO, T.: Numerical range of tensor products of operators. Tohoku Math. J. **19**(1), 98—100 (1966).
8. SCHREIBER, M.: Numerical range and spectral sets. Mich. Math. J. **10**, N. 3, 283—288 (1963).
9. STAMPFLI, J.G.: Hyponormal operators. Pacific. J. Math. **12**(4), 1453—1458 (1962).
10. WINTER, A.: Zur Theorie der beschränkten Bilinearformen. Math. Z. **30**, 228—282 (1929).

V. ISTRĂŢESCU
Institute of Mathematics
Romanian Academy of Sciences
str. M. Eminescu 47
Bucharest 9, Romania

I. ISTRĂŢESCU
Polytechnic Institute of Timişoara
Timişoara, Romania

Math. Zeitschr. 105, 157—159 (1968)

On a Conjecture of Samuel

BALWANT SINGH

Received July 25, 1967

§ 1. Introduction

In ([2], § 4) SAMUEL has made the following conjecture:

Let A be an integral domain of characteristic $p \neq 0$ and D a derivation of A. If a logarithmic derivative t belongs to the ideal $\mathfrak{q}$ generated by $D(A)$ then t is the logarithmic derivative of a unit.

He has proved there that the conjecture is true under the following hypotheses: i) $p=2$, 3 or 5, ii) A is local, iii) $D^p=D$, iv) $\mathfrak{q}$ is contained in the maximal ideal of A. We show in this note that the conjecture is true under the following hypotheses: i) $D^p=aD$ where a is a unit of A with $Da=0$, ii) $\mathfrak{q}$ is contained in the radical of A. We also give a trivial example to show that, in all its generality, the conjecture is false.

The author wishes to express his thanks to Professor R. SRIDHARAN for his guidance in the preparation of this note.

§ 2

Let A be an integral domain of characteristic $p \neq 0$ and K its quotient field. Let D be a derivation of K such that $D(A) \subset A$. Let $\mathfrak{q}$ be the ideal of A generated by $D(A)$. We shall denote by $\mathfrak{R}(A)$ the radical of A, i.e. the intersection of all maximal ideals of A. We shall use in the sequel the following formula due to HOCHSCHILD [1]: For any $\lambda \in K$, we have

$$(\lambda D)^p = \lambda^p D^p + (\lambda D)^{p-1}(\lambda) \cdot D$$

where λD is the derivation defined by $(\lambda D)(\mu) = \lambda(D\mu)$ for $\mu \in K$.

Lemma 1. *If $t \in K^*$ then*

$$\left(\frac{1}{t} D\right)^{p-1}(t^{p-1}) + \frac{1}{t} D^{p-1}(t) = 0 .$$

Proof. Let

$$\alpha = \left(\frac{1}{t} D\right)^{p-1}(t^{p-1}) + \frac{1}{t} D^{p-1}(t) .$$

If $Dt=0$, we have nothing to prove. We may therefore assume that $Dt \neq 0$. It is then sufficient to prove that $D\alpha=0$ and $D(t\alpha)=0$. If we write

$$D_1 = \frac{1}{t} D ,$$

an easy computation shows that

$$tD\alpha=(tD_1)^p(t)-t^pD_1^p(t)-(tD_1)^{p-1}(t)\cdot D_1 t$$

which is zero by Hochschild's formula. Therefore $D\alpha=0$. Further, we have

$$\begin{aligned}t\alpha&=t\left(\frac{1}{t}D\right)^{p-1}(t^{p-1})+D^{p-1}(t)\\&=t^{p+1}\left(\frac{1}{t}D\right)^{p-1}\left(\frac{1}{t}\right)+D^{p-1}(t)\end{aligned}$$

since t^p is a D-constant. Therefore

$$\begin{aligned}D(t\alpha)&=t^p\left[\frac{1}{t^p}D^p(t)+\left(\frac{1}{t}D\right)^{p-1}\left(\frac{1}{t}\right)\cdot Dt-\left(\frac{1}{t}D\right)^p(t)\right]\\&=0 \text{ by Hochschild's formula.}\end{aligned}$$

Proposition. *Let $D^p=aD$, where a is a unit of A with $Da=0$, and assume that the ideal $\mathfrak{q}$ generated by $D(A)$ is contained in $\mathfrak{R}(A)$. If t is a logarithmic derivative belonging to $\mathfrak{R}(A)$, then t is the logarithmic derivative of a unit.*

Proof. We may assume that $t\neq 0$. Since t is a logarithmic derivative, we have $D^{p-1}t=at-t^p$. (See [2], § 4, Lemma 2).

Let us write

$$x=\sum_{r=0}^{p-2}\left(\frac{1}{t}D\right)^r(t^{p-1})$$

and $u=a-x$. Then

$$\begin{aligned}\left(\frac{1}{t}D\right)(u)&=-\sum_{r=1}^{p-2}\left(\frac{1}{t}D\right)^r(t^{p-1})-\left(\frac{1}{t}D\right)^{p-1}(t^{p-1})\\&=-(x-t^{p-1})+\frac{1}{t}D^{p-1}(t)\quad\text{(Lemma 1)}\\&=-x+t^{p-1}+\frac{1}{t}(at-t^p)\\&=u\end{aligned}$$

so that $t=Du/u$.

Assuming now that $t\in\mathfrak{R}(A)$, we shall prove that $x\in\mathfrak{R}(A)$, which would imply that u is a unit of A. To this end we first prove

Lemma 2. *Let $B=k[X_1, X_2, \ldots, X_n, \ldots]$ where k is the prime field of p elements and let L be the quotient field of B. Let Δ be the derivation of L defined by $\Delta X_i=X_{i+1}$ and let*

$$f=\sum_{r=0}^{p-2}\left(\frac{1}{X_1}\Delta\right)^r(X_1^{p-1}).$$

Then f belongs to the ideal $\mathfrak{m}$ of B generated by $X_1, X_2, \ldots$.

Proof. We can write $f=g/X_1^n$ where $g\in B$ and n is a non-negative integer. Let $g\in B$ be such that $X_1\nmid g$. Since $X_1^{p-3}f\in B$, we have $n\leqq p-3$ and it is clear that in fact g belongs to $\mathfrak{m}$. If $f\notin\mathfrak{m}$, then $n\geqq 1$ and n is coprime to p. Now we have

$$\begin{aligned}\Delta f&=X_1\left[\sum_{r=1}^{p-2}\left(\frac{1}{X_1}\Delta\right)^r(X_1^{p-1})+\left(\frac{1}{X_1}\Delta\right)^{p-1}(X_1^{p-1})\right]\\&=X_1\left[f-X_1^{p-1}-\frac{1}{X_1}\Delta^{p-1}(X_1)\right]\qquad\text{(Lemma 1)}\\&=X_1 f-X_1^p-X_p.\end{aligned}$$

Therefore

$$\begin{aligned}\Delta g&=n X^{n-1}X_2 f+X_1^n\Delta f\\&=n X_1^{n-1}X_2 f+X_1 g-X_1^{n+p}-X_1^n X_p.\end{aligned}$$

Since $\Delta(B)\subset\mathfrak{m}$, $\Delta g\in\mathfrak{m}$. It follows that $nX_1^{n-1}X_2 f\in\mathfrak{m}$. Since n is coprime to p, $X_1^{n-1}X_2 f\in\mathfrak{m}$. Writing $h=X_1^{n-1}X_2 f$, we have $X_1 h=X_2 g$. This is a contradiction since $X_1\nmid g$. Therefore $f\in\mathfrak{m}$ and the lemma is proved.

Now in order to prove that $x\in\mathfrak{R}(A)$, it is sufficient to prove that x is in the ideal $\mathfrak{a}$ of A generated by $t, Dt, D^2t, \ldots$. We define a homomorphism $\varphi: L\to K$ of fields by setting $\varphi(X_i)=D^{i-1}(t)$. Then $\varphi(\mathfrak{m})\subset\mathfrak{a}$ and $\varphi\Delta=D\varphi$. Since $f\in\mathfrak{m}$ and $x=\varphi(f)$, we have $x\in\mathfrak{a}$ and the proof of the Proposition is complete.

Remark. Let $A=k[X, Y]$ be the polynomial ring in two variables over a field k. Define a k-derivation D of A by $DX=X$ and $DY=XY$. Then $DY/Y=X\in D(A)$. But X is not the logarithmic derivative of a unit, since the only units of A are the elements of k^* and we have $D(k^*)=0$. Thus, in its general form, the conjecture is false.

References

1. Hochschild, G.P.: Simple algebras with purely inseparable splitting fields. Trans. Amer. Math. Soc. **79**, 477–489 (1955).
2. Samuel, P.: Classes de diviseurs et derivees logarithmiques. Topology **3**, Suppl. 1, 81–96 (1964).

Balwant Singh
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 5

Math. Zeitschr. 105, 160–172 (1968)

Positive (Noncommuting) Operator Products and Semigroups*

KARL GUSTAFSON

Received July 17, 1967

Introduction

Let $A \geqq 0$ and $B \geqq 0$ be two positive bounded self-adjoint operators. The two algebraic questions which immediately arise are the following: (1) is $A+B \geqq 0$; and (2) is $BA \geqq 0$? The first question has (trivially) an affirmative answer; the second question also has an affirmative answer, *under the condition that A and B commute.* However, apparently the question (2), which due to the nonself-adjointness of BA in general must be reformulated as (2′) $\mathrm{Re}(BAx, x) \geqq 0$ for all x, has remained unanswered for general (i.e., *noncommuting*) operators A and B. This can obviously be an important question in matrix work, general mathematical operator theory, and especially in theoretical physics where one can say that having a zero commutator is the exceptional and less interesting case.

The purpose of this paper is to give some general sufficient conditions for (2″) $\mathrm{Re}[BAx, x] \geqq 0$ for all x; here we have written $[x, y]$ rather than (x, y) because our results will be obtained more generally for $[x, y]$ a semi-inner product on any Banach space (see LUMER [10]), where in the special case of a Hilbert space, $[x, y]=(x, y)$ necessarily. Our conditions may be described roughly as follows: letting $\|A\| = \|B\| = 1$, if A has a small (positive) numerical range, then B with a relatively large (positive) numerical range can be tolerated in the sense that the numerical range of BA will remain positive. Our procedure may be described relatively simply as follows: we write the problem as a perturbation $\varepsilon BA = (\varepsilon B - I)A + A$, and then pick the right ε so that the size of $(\varepsilon B - I)A$, considered as the perturbing operator, is minimized; this then permits us to show that the upper bound of $-\varepsilon BA$ is nonpositive.

It should be mentioned that some results for $\sigma(BA)$ in place of (BAx, x) in (2) are known. For example, the recent result of WILLIAMS [15] states that $B > 0$ and $A \geqq 0$ imply that $\sigma(BA) \geqq 0$; however, this is asking less in the sense that the spectrum is smaller than the (closed) numerical range.

In Section 1 we derive some basic lemmas concerning the behavior of the function $\min \|\varepsilon B - I\|$, $\varepsilon \geqq 0$, for a given B. In this connection we will refer to results of BOHNENBLUST and KARLIN [1], HILDEBRANDT [7], LUMER [10], and LUMER and PHILLIPS [11]. In Section 2 we give some results for commuting operators, for completeness and for use in later sections, although our primary

* Partially supported by NSF GP 7041X.

concern is noncommuting operators. In section 3 we apply the results of section 1 to the question (2''). In section 4 we give lower bounds for the positive operator products. In section 5 we apply these results to the question of when left multiplication by B of the generator A of a contraction semigroup yields BA as a generator; it was this question which motivated the present paper, since it is essential that BA have a negative sign.

It is expected that results for such basic questions as (2''), (2'), and (2) will have many applications, e.g., to products of elliptic systems in partial differential equations, to expectation values in physics, etc; further applications to semigroup theory and extensions to unbounded operators will be pursued in a later paper. The scope of the present paper is bounded operators, and whenever an operator appears, it may be assumed to be bounded and everywhere defined unless stated otherwise; if stated to be (unbounded), it is assumed to have a dense domain, and of course includes the bounded operators.

The author appreciates discussions with S. HILDEBRANDT concerning this paper.

1. Definitions and Lemmas

Consider the following quantitative functions defined on the Banach Algebra of bounded linear operators B: $X \to X$, X a complex Banach space; for unbounded operators, we will also use those which remain meaningful. Let $\|x\| = 1$:

$$\|B\| = \sup \|Bx\| = \text{norm of } B;$$

$$|W(B)| = \sup |[Bx, x]| = \text{numerical radius of } B;$$

$$\theta(B) = \sup \operatorname{Re}[Bx, x] = \text{upper semibound of } B;$$

$$m(B) = \inf \operatorname{Re}[Bx, x] = \text{lower semibound of } B;$$

$$\gamma(B) = \inf \|Bx\| = \text{minimum modulus of } B.$$

Here $[x, y]$ is the semi-inner product introduced by LUMER [10]; on a Hilbert space it is necessarily the regular inner product (x, y). By analogy with the self-adjoint case, $\theta(B)$ could be denoted as $M(B)$, but due to the conventions in the references, we use $\theta(B)$; in [7] it is called $a(B)$, and in [1] it is called $\phi(B)$. The above $\gamma(B)$ is slightly different from the usual

$$\gamma(B) = \inf\big(\|Bx\|/d(x, N(B))\big),$$

$x \notin N(B)$, $N(B)$ the kernel of B, but we will be chiefly interested in the case B^{-1} bounded, and then they are the same, and $\gamma(B) = \|B^{-1}\|^{-1}$. Always we have the relationship

$$m(B) \leqq \theta(B) \leqq \max\{|m(B)|, |\theta(B)|\} \leqq |W(B)| \leqq \|B\| \leqq \alpha |W(B)|,$$

where $\alpha = 2$ in Hilbert space, e in Banach space. We also note the relation $-\theta(-B) = m(B) \leqq \gamma(B)$. For strongly positive self-adjoint operators one has $0 < m(B) = \gamma(B) \leqq \theta(B) = |W(B)| = \|B\|$.

As generalizations of negative and positive self-adjoint operators, B is called *dissipative* if $\theta(B)\leqq 0$, *strongly dissipative* if $\theta(B)<0$, *accretive* if $\theta(-B)\leqq 0$, *strongly accretive* if $\theta(-B)<0$. We also consider the two functions $g(\varepsilon, B)=\|\varepsilon B+I\|$, $\varepsilon\geqq 0$, and $f(\varepsilon, B)=\varepsilon^{-1}[g(\varepsilon, B)-1]$, $\varepsilon>0$; we define $f(0, B)=\lim f(\varepsilon, B)$, $\varepsilon\to 0$. It is easy to see that both $f(\varepsilon, B)$ and $g(\varepsilon, B)$ are continuous in ε, $g(\varepsilon, B)$ is convex in ε, $f(\varepsilon, B)$ is monotone nondecreasing in ε, for $\varepsilon\geqq 0$. The important relation $\theta(B)=f(0, B)$ was shown by LUMER [10], see also HILDEBRANDT [7]; also $\theta(B)=\omega(B)$, the local type of the strongly continuous semigroup generated by B. Note that $\theta(B)$ is the Gateau derivative of the norm at the identity operator in the direction B, i.e., $g'(0, B)=\theta(B)=\delta_I^B\|\cdot\|$.

It is well-known (see [10–12]) that if B is dissipative, then $\|(\lambda-B)^{-1}\|\leqq\lambda^{-1}$ for all $\lambda>0$; we will need the following sharpened result.

Lemma 1.1. *Let B (unbounded) be dissipative. Then*

$$\gamma(\varepsilon B-I)\geqq 1-\varepsilon\theta(B), \qquad \text{for all } \varepsilon\geqq 0.$$

Proof. Letting $\|x\|=1$, we have

$$\|(\varepsilon B-I)x\|\geqq\left|\operatorname{Re}[(\varepsilon B-I)x, x]\right|\geqq 1-\varepsilon\operatorname{Re}[Bx, x]\geqq 1-\varepsilon\theta(B),$$

and thus the desired result; equivalently, that $\|(\lambda-B)^{-1}\|\leqq(\lambda-\theta(B))^{-1}$ for $\lambda>0$.

In the papers [1, 7, 10, 11] the behavior of $g(\varepsilon, B)$ near $g(0, B)=1$ was investigated (e.g., yielding $g'(0, B)=\theta(B)$); by this and the convexity of $g(\varepsilon, B)$, we may therefore consider three disjoint classes of $g(\varepsilon, B)$ curves: 1) $g(\varepsilon, B)$ strictly increasing, with the two subclasses 1.1) $\theta(B)>0$ and 1.2) $\theta(B)=0$; 2) $g(\varepsilon, B)$ initially flat, $\theta(B)=0$, with the two subclasses 2.1) $B\equiv 0$ or 2.2) $g(\varepsilon, B)$ strictly increasing for all sufficiently large ε; 3) $g(\varepsilon, B)$ initially decreasing, $\theta(B)<0$, and $g(\varepsilon, B)$ later strictly increasing. For our purposes we need to know more about the behavior of $g(\varepsilon, B)$ away from $\varepsilon=0$, especially for strongly dissipative B, especially for $0<\varepsilon<2\|B\|^{-1}$; the following Lemma gives us some information.

Lemma 1.2. (i)

$$\max\{\theta(B), \|B\|-2\varepsilon^{-1}\}\leqq f(\varepsilon, B)\leqq\min\{\|B\|, \theta(B)+\varepsilon[\|B\|+|\theta(B)|]^2\};$$

$$\max\{\varepsilon\|B\|-1, 1+\varepsilon\theta\}\leqq g(\varepsilon, B)\leqq\min\{1+\varepsilon\|B\|, 1+\varepsilon\theta+\varepsilon^2[\|B\|+|\theta|]^2;$$

$$\lim_{\varepsilon\to\infty} f(\varepsilon, B)=\|B\|, \qquad \lim_{\varepsilon\to\infty} g(\varepsilon, B)=\infty;$$

$\theta\leqq g'(\varepsilon, B)\leqq\|B\|$ *for all, except at most a countable number of values of,* ε.

(ii) *When B is dissipative,*

$$f(\varepsilon, B)\leqq\frac{\theta(B)+\varepsilon\|B\|^2}{1-\varepsilon\theta(B)};$$

$$\frac{\|B\|+\theta(B)}{\|B\|-\theta(B)}\leqq g(\varepsilon, B)\leqq\frac{1+\varepsilon^2\|B\|^2}{1-\varepsilon\theta(B)};$$

(iii) *When B is strongly dissipative, $g(\varepsilon, B)$ attains a minimum*

$$g_m(B)=\min_{\varepsilon\geqq 0} g(\varepsilon, B)$$

for at least one $\varepsilon_m>0$, where $0\leqq g_m(B)<1$ and $g_m(B)=0$ if and only if $B=-cI$, c a positive constant. Also, $g_m(B)$ is invariant under positive expansions or contractions of B.

Proof. All of the $f(\varepsilon, B)$ part of (i) follows directly from the monotonicity of f and triangle and reverse triangle inequalities, except the second upper bound, which will follow from part (ii). Similarly the $g(\varepsilon, B)$ part of (i) follows easily from the triangle inequalities, (i) for $f(\varepsilon, B)$, the relations $g=\varepsilon f+1$, and $g'(0)=\theta\leqq f=g'-\varepsilon f'\leqq g'=f+\varepsilon f'\leqq\|B\|$ (where the prime denotes the derivative with respect to ε, the last relation following by taking the difference quotient for g' where it exists), and the fact that g is convex. The second upper bounds for f and g are better than the first ones for small ε, but probably are not sharp; the upper bounds in (ii) will be more important. Letting $\|B\|=1$ and considering small ε, the upper bound for $f(\varepsilon, B)$ in (i) is a line extending upward from $\theta(B)$ with slope between 1 and 4, depending on $\theta(B)$; in [7, Lemma 5], for B on a Hilbert space, a line of slope 2 was obtained; in part (ii), the upper bound for $f(\varepsilon, B)$ will be a curve, emanating from $\theta(B)$, of slope between 0 and 2 and decreasing from initial slope $1+\theta^2(B)$. The lower bounds in (i) are sharp in the sense that examples exist in which $f(\varepsilon, B)$ and $g(\varepsilon, B)$ are exactly these bounding functions (see Lemma 1.3).

For part (ii), we first consider the upper bound for $g(\varepsilon, B)$. Because B is dissipative, B generates a contraction semigroup and $R(\lambda-B)=X$ for all $\lambda>0$ by the Hille-Yosida Theorem (e.g., see [8, 11]). Thus we can write $\varepsilon B+I=(\varepsilon B+I)(\varepsilon B-I)(\varepsilon B-I)^{-1}$ and thus

$$\|\varepsilon B+I\|\leqq\|I-\varepsilon^2 B^2\|\cdot\|\varepsilon B-I\|^{-1}\leqq(1+\varepsilon^2\|B\|^2)\cdot(1-\varepsilon\theta(B))^{-1}$$

by Lemma 1.1. The upper bound for $f(\varepsilon, B)$ in (ii) then follows from this bound for $g(\varepsilon, B)$. By the substitution $B_1=B-\theta(B)$ when $\theta(B)>0$, and by observing that then $\theta(B_1)=0$ and that $f(\varepsilon, B)\leqq f(\varepsilon, B_1)+\theta(B)$, one arrives at the upper bounds of (i).

In Hilbert space it is easy to see (whether or not B is dissipative) that $g^2(\varepsilon, B)\leqq\varepsilon^2\|B\|^2+1+2\varepsilon\theta(B)$. Also one can see from the identity

$$\begin{aligned}(\varepsilon^2+1+2\varepsilon\theta)\cdot[1+\varepsilon^2(1-\varepsilon\theta)^{-2}(1-\theta^2)]\\ =[(1+\varepsilon^2)(1-\varepsilon\theta)^{-1}]^2-4\theta^2\varepsilon^2(1-\varepsilon\theta)^{-2},\end{aligned}$$

where $\|B\|=1$ and $\theta=\theta(B)$, the comparison for strongly dissipative B

$$g(\varepsilon, B)\leqq(\varepsilon^2+1+2\varepsilon\theta)^{\frac{1}{2}}\leqq(1+\varepsilon^2)(1-\varepsilon\theta)^{-1}, \tag{1.1}$$

with respective minimums attained at $\varepsilon_m=-\theta$ and $((1+\theta^2)^{\frac{1}{2}}-1)\cdot(-\theta)^{-1}$, with corresponding minimum values $(1-\theta^2)^{\frac{1}{2}}$ and $2\theta^{-2}((1+\theta^2)^{\frac{1}{2}}-1)$. Thus

in Hilbert space one can somewhat improve the bounding functions of Lemma 1.2; however, the two upper bounding functions in (1.1) do not differ qualitatively in an important manner concerning $g_m(B)$. That is, our main interest will be to bound $g_m(B)$ itself when $\theta(B)<0$; it will be seen that the above two minimum values, although less than 1, are rather poor for self-adjoint operators, for example, because the bound minimums of both occur too early and are too close to the value 1. For a specific example, for self-adjoint B with $\|B\|=1$ and $\theta(B)=-\frac{1}{2}$, the minimums of the bounding functions are approximately 0.87 and 0.94, whereas $g_m(B)=0.33$.

On the other hand, the bounds in (1.1) are not improvable in general, as can be seen from the following example. Let B be the normal operator on l_2 constructed with eigenvalues a countable set dense in the subset $\{|z|\leqq 1\}\cap\{\text{Re } z\leqq\theta\}$ for some given $-1<\theta<0$, and containing also the point λ, $|\lambda|=1$, Re $\lambda=\theta$. Then the above compact subset is $\sigma(B)=\overline{W(B)}$, where $W(B)$ denotes the numerical range of B. Thus $\exists\lambda_n=(Bx_n, x_n)\varepsilon W(B)$ such that $\lambda_n\to\lambda$; also $1=|\lambda|\leftarrow|\lambda_n|=|(Bx_n, x_n)|\leqq\|Bx_n\|\leqq\|B\|=1$, from which it can be concluded that $g^2(\varepsilon, B)=\varepsilon^2+1+2\varepsilon\theta$. Therefore we can conclude that for the purpose of bounding $g(\varepsilon, B)$ and $g_m(B)$ from above, we cannot in general appreciably improve on Lemma 1.2 (ii) (although a slight improvement might be possible for the general Banach space bound $(1+\varepsilon^2\|B\|^2)(1-\varepsilon\theta(B))^{-1}$ by bounding $\|I-\varepsilon^2B^2\|$ closer, especially for special classes of operators).

This brings us to the lower bound in (ii); it can be obtained as the intersection of the lower bounding functions $1+\varepsilon\theta$ and $\varepsilon\|B\|-1$ from (i), yielding $\varepsilon_m=2(\|B\|-\theta(B))^{-1}$. Two symmetric but poorer lower bounds

$$(\|B\|-\|B^{-1}\|^{-1})\cdot(\|B\|+\|B^{-1}\|^{-1})^{-1}$$

and

$$[\theta(-B)+\theta(B)]\cdot[\theta(-B)-\theta(B)]^{-1}$$

are also obtainable; for strongly negative self-adjoint operators both of these lower bounds, the lower bound of Lemma 1.2 (ii), and $g_m(B)$ will be seen (Lemma 1.3) to all agree. If $g_m(B)$ attains the lower bound given in (ii), then clearly $g(\varepsilon, B)$ behaves as in the self-adjoint case, i.e., $g(\varepsilon, B)$ consists of exactly the two functions $1+\varepsilon\theta$ and $\varepsilon\|B\|-1$.

For part (iii), by the convexity of $g(\varepsilon, B)$ and the fact that $g'(0)=\theta(B)<0$, $g(\varepsilon, B)$ clearly lies properly between the values 1 and 0 for some interval $(0, \varepsilon_1)$, $0<\varepsilon_1<2\|B\|^{-1}$ by part (i), with the possible exception that $g(B, \varepsilon_0)=0$ for some $\varepsilon_0\in(0, \varepsilon_1)$; this can happen only for one ε_0 and only when B is a negative multiple of the identity, since $g(\varepsilon, -cI)=0$ clearly implies $\varepsilon_0=c^{-1}$, and $\|\varepsilon B+I\|=0$ implies $B=-\varepsilon^{-1}I$. Since $g(\varepsilon, B)$ is continuous in ε it attains a minimum; by elementary considerations it can be seen that $f(\varepsilon, B)$ is strictly increasing immediately to the right of any $\varepsilon_m(B)$. By the normal operator example given above and the following self-adjoint examples, $g_m(B)$ may be attained either smoothly or at a discontinuity of $g'(\varepsilon, B)$. The fact that $g_m(B)$ is invariant relative to the size of B is clear from $g(c^{-1}\varepsilon_m, cB)=g(\varepsilon_m, B)$, $c>0$.

This observation allows us the convenient assumption of $\|B\|=1$ when we wish it.

For self-adjoint operators, the behavior of $f(\varepsilon, B)$ and $g(\varepsilon, B)$ is completely determined, and in fact they assume their lower bounding values of Lemma 1.2 exactly. Because of the importance of this class of operators, we state the following:

Lemma 1.3. *Let B be self-adjoint. Then* $g(\varepsilon, B)=\max\{1+\varepsilon\theta(B), \varepsilon\|B\|-1\}$, $f(\varepsilon, B)=\max\{\theta(B), \|B\|-2\varepsilon^{-1}\}$. *If g' has a discontinuity, it occurs at* $\varepsilon_c=2(\|B\|-\theta(B))^{-1}$; *when* $\theta(B)<0$, $\varepsilon_m=\varepsilon_c$ *and*

$$g_m(B)=\big(\|B\|+\theta(B)\big)\cdot\big(\|B\|-\theta(B)\big)^{-1}.$$

Proof. We recall that B satisfies $m=m(B)\leqq(Bx, x)\leqq\theta(B)=M$ and thus $\varepsilon m+1=m(\varepsilon B+I)\leqq((\varepsilon B+I)x, x)\leqq\theta(\varepsilon B+I)=\varepsilon M+1$; Lemma 1.3 then follows in a straight forward manner by evaluating

$$\|\varepsilon B+I\|=\max\{|\varepsilon m+1|, |\varepsilon M+1|\},$$

recalling that $\|B\|=\max\{|m|, |M|\}$. We now list these different cases using the classification previously introduced for the $g(\varepsilon, B)$ graphs; we observe that for self-adjoint operators, the class 1.2) cannot occur, while the class 1.1) decomposes into the three classes of positive, positively dominated "mixed", and negatively dominated "mixed" operators.

1.1 a, b) $0\leqq|m|\leqq M$, $M>0$, $g(\varepsilon, B)=1+\varepsilon M$.

1.1 c) $0<M<-m$, $\varepsilon_c=-2(m+M)^{-1}$, $g(\varepsilon, B)=1+\varepsilon M$ on $[0, \varepsilon_c]$, $g(\varepsilon, B)=-m\varepsilon-1$ thereafter.

2.1) $m=M=0=B$, $g(\varepsilon, B)=1$.

2.2) $m<M=0$, $\max\varepsilon_m=\varepsilon_c=-2m^{-1}$, $g(\varepsilon, B)=1$ on $[0, \varepsilon_c]$, $g(\varepsilon, B)=-m\varepsilon-1$ thereafter.

3) $m\leqq M<0$, $\varepsilon_m=\varepsilon_c=-2(m+M)^{-1}$, $g(\varepsilon, B)=1+\varepsilon M$ on $[0, \varepsilon_m]$, $g(\varepsilon, B)=-m\varepsilon-1$ thereafter, $g(\varepsilon_m)=(m-M)\cdot(m+M)^{-1}$.

In case 3) by the spectral radius theorem we may write

$$\begin{aligned} g_m(B)&=\big(\|B\|-\|B^{-1}\|^{-1}\big)\cdot\big(\|B\|+\|B^{-1}\|^{-1}\big)^{-1}\\ &=\big(\theta(-B)+\theta(B)\big)\cdot\big(\theta(-B)-\theta(B)\big)^{-1}=\big(\|B\|+\theta(B)\big)\cdot\big(\|B\|-\theta(B)\big)^{-1}. \end{aligned}$$

2. Commuting Operator Products

We include this section only for completeness and for use later; the results no doubt can be extended. It is well-known that if A and B are positive commuting everywhere defined self-adjoint operators, then BA is self-adjoint and positive. We first extend this to *unbounded* A; unbounded B could also be

considered with the addition of conditions such as $D(BA)$ dense, etc. We recall that unbounded A commutes with bounded (everywhere defined) B iff $BA \subset AB$.

Lemma 2.1. *Let A (unbounded) and B be self-adjoint. Then BA is symmetric iff A commutes with B.*

Proof. As in the bounded case, from $(BA)^* = AB$.

Lemma 2.2. *Let A (unbounded) be self-adjoint and positive, B accretive, A commuting with B; then BA is accretive. In addition, let B be self-adjoint and strongly positive; then $BA = AB$ is self-adjoint and strongly positive.*

Proof. A possesses a positive self-adjoint square root $A^{\frac{1}{2}}$ such that $BA^{\frac{1}{2}} \subset A^{\frac{1}{2}}B$, so that $(BAx, x) = (BA^{\frac{1}{2}}x, A^{\frac{1}{2}}x)$ and thus $\mathrm{Re}(BAx, x) \geqq 0$. When B is self-adjoint, by Lemma 2.1 BA is symmetric, and then BA is self-adjoint by the following decomposition. We write $\varepsilon BA = (\varepsilon B - I)A + A$ and note that $\|\varepsilon B - I\| < 1$ for ε near $\varepsilon_m(-B)$; thus the (symmetric) perturbation $(\varepsilon B - I)A$ is relatively bounded relative to A with relative bound less than 1, sufficient for the presentation of self-adjointness (e.g., see [9]); clearly the lower bound is positive.

3. Positive (Noncommuting) Operator Products

First it should be noted that $A \geqq 0$ and $B \geqq 0$ does not necessarily imply that $\mathrm{Re}(BAx, x) \geqq 0$, even if $B > 0$ (in contrast with Lemma 2.2). Let:

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & \alpha \\ 0 & 1 \end{bmatrix}$$

on complex two space, with $0 < \alpha < 1$. Then by [3, 13], the numerical range of BA is a proper ellipse with focal points 0 and 1. This example shows that one cannot even rotate BA, i.e., consider $e^{i\phi}BA$ for some ϕ, to obtain a positive real part of the numerical range. On the other hand, $\sigma(BA)$ is the two points $\{0\}$ and $\{1\}$ and is real and non-negative in accordance with the results of [15].

The main result of this paper is the following Theorem.

Theorem 3.1. *Let A and B be strongly accretive bounded operators on a Banach space. If either $g_m(-B) \leqq m(A) \cdot \|A\|^{-1}$ or $g_m(-A) \leqq m(B) \cdot \|B\|^{-1}$, then both BA and AB are accretive.*

Corollary 3.2. *Let A and B be self-adjoint operators on a Hilbert space with $m_A > 0$ and $m_B > 0$. If $(M_B - m_B) \cdot (M_B + m_B)^{-1} \leqq m_A M_A^{-1}$, or (equivalently) vice versa, then both BA and AB are accretive.*

Proof. We treat the problem as a perturbation. We may write $\varepsilon BA = (\varepsilon B - I)A + A$, and then

$$\begin{aligned} \theta(-\varepsilon BA) &\leqq \theta\big(-(\varepsilon B - I)A\big) + \theta(-A) \\ &\leqq \|(\varepsilon B - I)A\| + \theta(-A) \\ &\leqq g(\varepsilon, -B) \cdot \|A\| + \theta(-A). \end{aligned}$$

Then letting $\varepsilon=\varepsilon_m(-B)$, or for that matter, any other ε such that $g(\varepsilon, -B)\leqq -\theta(-A)\cdot\|A\|^{-1}$, it follows that $-\varepsilon BA$ is dissipative, and hence BA is accretive. The proof for the other condition $g_m(-A)\leqq-\theta(-B)\cdot\|B\|^{-1}$ follows similarly from the decomposition $-\varepsilon BA=B(I-\varepsilon A)-B$. That AB is also accretive under the first condition follows from this second decomposition applied to A, namely, $-\varepsilon AB=A(I-\varepsilon B)-A$; the second condition for AB uses the first decomposition. That in fact the two conditions are equivalent for self-adjoint operators (Corollary 3.2) follows from the lower bounds for g_m in Lemma 1.3; letting $\|B\|=\|A\|=1$ for simplicity, it follows directly that

$$g_m(-B)\leqq-\theta(-A)\Rightarrow\big(1+\theta(-B)\big)\cdot\big(1-\theta(-B)\big)^{-1}$$
$$\leqq-\theta(-A)\Rightarrow g_m(-A)=\big(1+\theta(-A)\big)\cdot\big(1-\theta(-A)\big)^{-1}\leqq-\theta(-B),$$

a situation which does not hold in the general (non-self-adjoint) case. For self-adjoint $\|A\|=\|B\|=1$, the condition $m_A+m_B\geqq 1$, although more restrictive than that of Corollary 3.2, shows roughly the sufficient relationship of the operators; but the corresponding generalization $-\theta(-A)-\theta(-B)\geqq 1$ does not carry over, as seen from the normal operator example given above, since what is needed here is $g_m(-B)\leqq 1+\theta(-B)$, i.e., $f(\varepsilon_3, -B)\leqq\theta(-B)$ for some $\varepsilon_3\geqq 1$, a situation occurring for self-adjoint operators but not in general.

We should remark that Theorem 3.1 can be shown without the θ machinery, using just $g(\varepsilon, T)$ instead, from $\|-\varepsilon(\eta B)A+I\|\leqq\eta g(\varepsilon, -A)+g(\eta, -B)$, which is minimized by taking $\varepsilon=\varepsilon_m(-A)$, and then made less than 1 by requiring $g_m(-A)\leqq-\theta(-B)$, recalling the relation $g(\eta)=\eta f(\eta)+1$.

Let us observe that in the numerical analysis of matrices (e.g., see [14]), one encounters the so-called spectral condition number $\kappa(H)=\|H\|\cdot\|H^{-1}\|^{-1}$, where H is a similarity transformation converting a given matrix to diagonal form; thus for strongly positive self-adjoint A, the condition in Theorem 3.1 is $g_m(-B)\leqq[\kappa(A)]^{-1}$, a type of operator "conditioning" requirement.

Concerning the fact that both BA and AB are accretive in Theorem 3.1, we note that since they are similar, we have $\sigma(AB)=\sigma(BA)$ if they are on a Hilbert space; on the other hand, similarity does not always preserve the numerical range, as may be seen (for example) from the two by two matrix example of this section by replacing one of the α elements with $\beta\neq\alpha$.

In Theorem 3.1 (and in Lemma 2.2) we considered only positive B; examples of A positive, B not positive, but BA positive are easily constructed. However, the limitation on considering non-positive B can be seen from the following example. Let A and B commute, A self-adjoint and strongly positive. Then clearly $\|A^{\frac{1}{2}}x\|^2\geqq m(A)$ for $\|x\|=1$, i.e., $R(A^{\frac{1}{2}})=X$. Now suppose B is not accretive; then there exists $y=A_jx$ such that $\mathrm{Re}(BAx, x)=\mathrm{Re}(By, y)<0$.

Theorem 3.1 is sharp. Let $-A$, $^{\frac{1}{2}}j=1, 2$, be normal operators on l_2 as constructed in section 1, i.e., A_j has eigenvalues dense in that part of the unit ball to the right of the lines $\mathrm{Re}\,z=m(A_j)$, where $0<m(A_j)<\|A_j\|=1$; also A_j has the eigenvalue $\lambda_j=m(A_j)+i(1-m(A_j)^2)^{\frac{1}{2}}$, with common $(j=1, 2)$ nor-

malized eigenvector x_0. Then

$$\begin{aligned}\mathrm{Re}(A_1 A_2 x_0, x_0) = \mathrm{Re}\,\lambda_1\lambda_2 &= m(A_1)\cdot m(A_2) - (1-m(A_1)^2)^{\frac{1}{2}}\cdot(1-m(A_2)^2)^{\frac{1}{2}} \\ &= m(A_1)\cdot m(A_2) - g_m(-A_1)\cdot g_m(-A_2) \\ &< 0,\end{aligned}$$

if both $g_m(A_1) > m(A_2)$ and $g_m(A_2) > m(A_1)$. For a specific example, let $m(A_1) = 2^{-1}$; then $\arg(\lambda_1) = \pi/3$, and the limitation on A_2 is $\arg(\lambda_2) \leqq \pi/6$, i.e., $m(A_2) \geqq 3^{\frac{1}{2}}/2 = g_m(-A_1)$. Note that both of these operators A_j are strongly accretive and can none the less yield a non-accretive product.

4. Lower Bounds

Theorem 4.1. *The following are lower bounds for* $m_{BA} \equiv m(BA)$.

$b_1 = (m_A m_B^2 \|A\| + m_B m_A^2 \|B\|)\cdot(m_A m_B + \|A\|\cdot\|B\|)^{-1}$; *A and B strongly positive self-adjoint operators,* $AB = BA$.

$b_2 = m_A m_B$; *A (unbounded) positive self-adjoint, B strongly positive self-adjoint,* $BA \subset AB$.

$b_3 = m_A m_B$; *A (unbounded) positive self-adjoint, B accretive,* $BA \subset AB$.

$b_4 = (m_A - g_m(-B)\cdot\|A\|)\cdot(\varepsilon_m(-B))^{-1}$; *A and B strongly accretive on a Banach space,* $g_m(-B) \leqq m_A\cdot\|A\|^{-1}$.

Proof. The bounds are ordered in increasing generality of the operator classes involved. The bound b_1 is obtained as the uniform lower bound of an inequality of recent interest (see DIAZ and METCALF [2]), namely

$$(Ax, Bx)\cdot(M_A M_B + m_A m_B) \geqq m_A M_A (Bx, Bx) + m_B M_B (Ax, Ax),$$

valid for strongly positive bounded commuting self-adjoint operators; this inequality and others equivalent or very similar to it have been given by DIAZ and METCALF, GREUB and RHEINBOLDT, PETRYSHYN, MAKAI, KRASNOSELSKII and KREIN, KANTOROVICH, etc.; see [2].

The b_2 bound will be obtained by the perturbation techniques of this paper, while the b_3 bound will be obtained from the proof of Lemma 2.2. The fact that they are equal is evidence that the b_4 bound, obtained with procedures similar to those for the b_2 bound, is rather good as a general bound.

If one has a self-adjoint (unbounded) T_1, possessing a lower bound, which is perturbed by a symmetric T_2 which is relatively bounded with respect to T_1, relative bound $b < 1$, then (see [9, p. 291]) one has the lower bound

$$m(T_1 + T_2) \geqq m(T_1) - \max\{a(1-b)^{-1}, a + b\,|m(T_1)|\}.$$

Letting $T_1 = A$, $T_2 = (\varepsilon_m B - I)A$, $b = g_m(-B)$, $a = 0$, by the b_2 assumptions on A and B (and using Lemmas 2.1 and 1.3) we arrive at

$$m(BA) \geqq (m_A - g_m(-B)\cdot m_A)\cdot(\varepsilon_m(-B))^{-1} = b_2.$$

For b_3, we have from the proof of Lemma 2.2 that

$$\mathrm{Re}(BAx, x)=\mathrm{Re}(BA^{\frac{1}{2}}x, A^{\frac{1}{2}}x)\geqq m_B(Ax, x)\geqq m_B m_A.$$

The b_4 bound can be extracted from the proof of Theorem 3.1.

Letting A and B be strongly positive commuting bounded self-adjoint operators, so that we can make a comparison, one has by elementary calculations (and not using the condition $g_m(-B)\leqq m_A\cdot\|A\|^{-1}$ for b_4) that

$$b_4\leqq b_1\leqq b_2\equiv b_3,$$

and as $m_A\to\|A\|$, the case including the most B for b_4, all bounds approach m_Am_B. As two representative examples, letting $\|A\|=\|B\|=1$, we list:

$$m_A=m_B=2^{-1},\quad b_2=4^{-1},\quad b_1=5^{-1},\quad b_4=8^{-1};$$

$$m_A=4^{-1},\quad m_B=3\cdot 4^{-1},\quad b_2=3(16)^{-1},\quad b_1=3(19)^{-1},\quad b_4=3(32)^{-1}.$$

For our purposes, we conclude that b_4, *valid for noncommuting operators, on a Banach space*, is satisfactory.

5. Multiplication of Semigroup Generators

Much is known about the perturbation of semigroup generators, e.g., see [5, 8, 9, 12], but apparently very little is known about products of generators; more relevantly, about when left multiplication by B of a generator A produces a generator BA, more relevantly because often if B is also a generator, BA will not be a generator, having the wrong sign (Corollary 5.3). We restrict ourselves here to A (unbounded) any infinitesimal generator of a strongly continuous contraction semigroup on a Banach space X, letting $G(1, 0)$ denote all such generators, and to left multiplication by bounded (everywhere defined) B. The contraction semigroups are the most important type of semigroup for applications (e.g., see [9]).

In the recent paper DORROH [4] showed that in the special case when X is a space of bounded functions and $A\in G(1, 0)$, then $pA\in G(1, 0)$, where p was the operator given by multiplication by a positive bounded function, bounded below away from zero; in GUSTAFSON [6] part (i) of the following was shown.

Theorem 5.1. *Let X be any Banach space, A (unbounded) $\in G(1, 0)$, either of the following (disjoint) conditions holding for B;*

(i) *B is strongly accretive;*

(ii) $\exists\varepsilon>0\ni\varepsilon\|B^{-1}\|^{-1}>\|\varepsilon B-I\|\geqq 1$, $\theta(-B)\geqq 0$.

Then $BA\in G(1, 0)$ if and only if BA is dissipative.

Proof. We omit the proof of (i), which is given in [6], where (i) was stated as: $\exists\varepsilon>0\ni\|\varepsilon B-I\|<1$. Condition (ii), although somewhat peculiar, will be proved below, especially since the proof of (i) is similar. It is equivalent to

asking for an operator B_1 satisfying the condition $1 \leqq 1+\theta(B_1) \leqq \|B_1+I\| < \|B_1^{-1}\|^{-1}$, which excludes self-adjoint operators by the spectral radius theorem and the fact that the extremes of the numerical range are in the spectrum. That is, many strongly accretive operators satisfy the norm relations of (ii), but we have excluded them by $\theta(-B) \geqq 0$; by the graphs of Lemma 1.2, (ii) requires a small $\theta(B_1)$, large $\gamma(B_1)$, and a $g(\varepsilon, B_1)$ curve which ascends slowly at the beginning.

To prove (ii), we need only (e.g., see [11]) to prove that $R(I-\varepsilon BA)=X$; in the following we use some basic index theory (e.g., see [5, 9]). Writing $I-\varepsilon BA=\varepsilon B(I-A)+(I-\varepsilon B)$ and considering $I-\varepsilon B$ as perturbing $\varepsilon B(I-A)$, we have

$$\|I-\varepsilon B\| < \gamma(\varepsilon B) \leqq \gamma(I-A)\cdot\gamma(\varepsilon B) \leqq \gamma\big(\varepsilon B(I-A)\big) \tag{5.1}$$

by the hypothesis of (ii) and Lemma 1.1 applied to A. By GUSTAFSON [5] this yields $\beta(I-\varepsilon BA)=\beta(\varepsilon B(I-A))$, where β is the deficiency index (codimension of the closure of the range of the operator). Since $A \in G(1, 0)$, we have $R(I-A)=X=R(B)$, the latter by (ii) considered as $I-\varepsilon B$ perturbing εB. Thus

$$\beta(I-\varepsilon BA)=0$$

and the proof is concluded by noting that (5.1) implies that both $I-\varepsilon BA$ and $R(I-\varepsilon BA)$ are closed.

Remark. Theorem 5.1 motivates the present paper, because after the application of index theory, one sees that the essential, and more difficult for noncommuting operators, part of whether $BA \in G(1, 0)$ is the question of *when BA is dissipative.* Of course, if one can find a semi-inner product $[x, y]$ for which BA is dissipative, e.g., if one considers strongly positive function operators $B=p$ as in [4], this question can be disposed of. For a general criterion, restricting ourselves to bounded $A \in G(1, 0)$, we can now use Theorem 3.1.

Theorem 5.2. *Let $A \in G(1, 0)$, B accretive. If either $g_m(-B) \leqq m(-A)\cdot\|A\|^{-1}$, or $g_m(A) \leqq m(B)\cdot\|B\|^{-1}$, then both $BA \in G(1, 0)$ and $AB \in G(1, 0)$.*

Proof. We need only show that the products are dissipative, (e.g., see [11]); if $\theta(A)<0$ or $\theta(-B)<0$, the result follows from Theorem 3.1. If $\theta(A)=0$, then the hypothesis permits only B of the form cI, $c>0$, from the first condition, or $B=\|B\|\cdot I$ from the second condition, for then $m(B)=\|B\|$, which implies $[B-\|B\|\cdot Ix, x]=0$ for all $[x, x]=1$; similarly for $\theta(-B)=0$.

Corollary 5.3. *Let $A \in G(1, 0)$, $B \in G(1, 0)$, either $g_m(B)<m(-A)\cdot\|A\|^{-1}$ or $g_m(A)<m(-B)\cdot\|B\|^{-1}$ then neither BA nor AB is in $G(1, 0)$.*

Proof. Similar to the above: if A and B are both strongly dissipative, from the lower bound b_4 of Theorem 4.1; the $\theta(A)=0$ or $\theta(B)=0$ cases are not allowed by the hypotheses.

Corollary 5.4. *Let* $A \in G(1, 0)$, *B positive, both self-adjoint, either* $(M_B - m_B) \cdot (M_B + m_B)^{-1} \leqq m_A M_A^{-1}$ *or vice versa; then* $BA \in G(1, 0)$ *and* $AB \in G(1, 0)$.

Proof. Theorem 5.2 and Lemma 1.3.

For completeness we include the following commutative result, noting that here quantitative relations between the numerical ranges are not important.

Corollary 5.5. *Let A (unbounded)* $\in G(1, 0)$, *A self-adjoint, B strongly accretive,* $BA \subset AB$. *Then* $BA \in G(1, 0)$.

Proof. Lemma 2.2 and A dissipative (necessarily, see [9]) imply that BA is dissipative, which is sufficient by Theorem 5.1.

6. Additional Remarks

That ε_m in Lemma 1.2 (iii) is unique for Hilbert space is easily seen; that ε_m there is not unique in general can be seen from examples.

Although the criteria of section 3 is sharp, a significant improvement can be made if one knows the additional parameter $\cos A$; see GUSTAFSON, *The angle of an operator and positive operator products*, Bull. Amer. Math. Soc., to appear. For example, $g_m(-B) \leqq \cos A$ yields BA accretive, and $m(A) \cdot \|A\|^{-1}$ is a lower bound for $\cos A$, although in the normal operator examples of section 3 one has $\cos A = m(A) \cdot \|A\|^{-1}$. Moreover, using $\cos A$ in the lower bound b_4 of section 4 yields better values; e.g., in the first example there, $b_4' \cong 0.2424$.

The results of section 5 can be extended to unbounded B; see GUSTAFSON and LUMER, *Multiplicative perturbation of semigroups*, to appear.

References

1. BOHNENBLUST, H.F., and S. KARLIN: Geometrical properties of the unit sphere of Banach Algebras. Ann. of Math. **62**, 217—229 (1955).
2. DIAZ, J.B., and F.T. METCALF: Complementary inequalities III: inequalities complementary to SCHWARZ's inequality in Hilbert space. Math. Annalen **162**, 120—139 (1965).
3. DONOGHUE, W.F.: On the numerical range of a bounded operator. Michigan Math. J. **4**, 261—263 (1957).
4. DORROH, J.R.: Contraction semigroups in a function space. Pacific. J. Math. **19**, 35—38 (1966).
5. GUSTAFSON, K.E.: A perturbation lemma. Bull. Amer. Math. Soc. **72**, 334—338 (1966).
6. — A note on left multiplication of semigroup generators. Pacific J. Math. (to appear).
7. HILDEBRANDT, S.: Über den numerischen Wertebereich eines Operators. Math. Annalen **163**, 230—247 (1966).
8. HILLE, E., and R.S. PHILLIPS: Functional analysis and semigroups. Amer. Math. Soc. Colloq. Publ. 31 Providence: R.I. 1957.
9. KATO, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966.
10. LUMER, G.: Semi inner-product spaces. Trans. Amer. Math. Soc. **100**, 29—43 (1961).

11. —, and R.S. PHILLIPS: Dissipative operators in a Banach space. Pacific. J. Math. **11**, 679—698 (1961).
12. NELSON, EDWARD: Feynman integrals and the Schrödinger equation. J. Mathematical Physics **5**, 332—343 (1964).
13. TOEPLITZ, O.: Das algebraische Analogen zu einem Satze von FEYÉR. Math. Z. **2**, 187—197 (1918).
14. WILKINSON, J.H.: The algebraic eigenvalue problem. Oxford: Clarendon Press 1965.
15. WILLIAMS, J.P.: Spectra of products and numerical ranges. J. Math. Anal. and Applic. **17**, 214—220 (1967).

Dr. KARL GUSTAFSON
Institut Battelle
CH-1227 Carouge-Genève, Schweiz
and University of Minnesota
Minneapolis, Minnesota 55455, USA

Correction to my note "Degrees of Modular Irreducible Representations of p-Solvable Groups"

Math. Zeitschr. **104**, 141—143 (1968)

EVERETT C. DADE

Received May 13, 1968

The results of my note are all trivial consequences of a minor and obvious generalization of Theorem 6 of R. SWAN's paper "The Grothendieck Ring of a Finite Group" in Topology **2**, 85—110 (1963). I wish to apologize to SWAN for not being aware of his result.

Prof. EVERETT C. DADE
Dept. of Math., University of Illinois
Urbana, Illinois 61801, USA