

Werk

Titel: Journal für die reine und angewandte Mathematik

Verlag: de Gruyter

Jahr: 1968

Kollektion: Mathematica

Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen

Werk Id: PPN243919689_0232

PURL: http://resolver.sub.uni-goettingen.de/purl?PPN243919689_0232

Übergeordnetes Werk

Werk Id: PPN243919689

PURL: <http://resolver.sub.uni-goettingen.de/purl?PPN243919689>

Terms and Conditions

The Goettingen State and University Library provides access to digitized documents strictly for noncommercial educational, research and private purposes and makes no warranty with regard to their use for other purposes. Some of our collections are protected by copyright. Publication and/or broadcast in any form (including electronic) requires prior written permission from the Goettingen State- and University Library.

Each copy of any part of this document must contain there Terms and Conditions. With the usage of the library's online system to access or download a digitized document you accept the Terms and Conditions.

Reproductions of material on the web site may not be made for or donated to other repositories, nor may be further reproduced without written permission from the Goettingen State- and University Library.

For reproduction requests and permissions, please contact us. If citing materials, please give proper attribution of the source.

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

Journal für die reine und angewandte Mathematik

gegründet von A. L. Crelle 1826

fortgeführt von

**C. W. Borchardt, K. Weierstrass, L. Kronecker,
L. Fuchs, K. Hensel, L. Schlesinger**

gegenwärtig herausgegeben von

Helmut Hasse und Hans Rohrbach

unter Mitwirkung von

**W. Brödel, M. Deuring, A. Grothendieck, P. R. Halmos, O. Haupt,
F. Hirzebruch, E. Hopf, M. Kneser, G. Köthe, K. Prachar, H. Reichardt,
P. Roquette, W. Schmeidler, L. Schmetterer, E. Stiefel**

Band 232



Walter de Gruyter & Co.

**vormals G. J. Göschen'sche Verlagshandlung / J. Guttentag, Verlags-
buchhandlung / Georg Reimer / Karl J. Trübner / Veit & Comp.**

Berlin 1968

1969.1993

Archiv-Nr. 18 03 68

Satz und Druck: Walter de Gruyter & Co., Berlin 30

Printed in Germany

Inhaltsverzeichnis des Bandes 232

	Seite
Banaschewski, B., and G. Bruns , Injective hulls in the category of distributive lattices	102
Benz, W. , Über die Grundlagen der Geometrie der Kreise in der pseudo-euklidischen (Minkowskischen) Geometrie	41
Bruns, G. , siehe Banaschewski, B.	
Douglas, R. G. , Structure theory for operators. I	180
Günzler, Hans , Means over countable semigroups and almost-periodicity in L^1	194
Gütting, Rainer , Zur Berechnung der Mahlerschen Funktionen w_n	122
Härtter, Erich , Differenzen von Mengen nichtnegativer ganzer Zahlen	112
Hildebrandt, Stefan , Über die alternierenden Verfahren von H. A. Schwarz und C. Neumann	136
Hofmeister, Gerd , Asymptotische Abschätzungen für dreielementige Extremalbasen in natürlichen Zahlen	77
Köthe, Gottfried , Die Bildräume abgeschlossener Operatoren	110
Müller, Bruno J. , Dominant dimension of semi-primary rings	173
Scheid, Harald , Über arithmetische Progressionen der Länge p	118
—, —, Arithmetische Funktionen über Halbordnungen. II	207
Schwarz, Wolfgang , Schwache asymptotische Eigenschaften von Partitionen	1
Schweiger, Fritz , Ein Kuzminscher Satz über den Jacobischen Algorithmus . .	35
Stade, Ulrich , Periodische Lösungen von nichtlinearen elliptischen Differentialgleichungen	156
Szinteniings, Dieter , Relativnullen in additiven abelschen Gruppen	17

Ausgabedatum des Bandes 232:

30. November 1968

Schwache asymptotische Eigenschaften von Partitionen

Von *Wolfgang Schwarz* in Freiburg

Herrn Professor Rohrbach zum 65. Geburtstag gewidmet

1. Einleitung

Gegeben sei eine Folge $\{\lambda_\nu\}$, $0 < \lambda_1 < \lambda_2 < \dots$ reeller Zahlen ohne endlichen Häufungspunkt; für die Anzahlfunktion

$$(1.1) \quad N(u) = \sum_{\lambda_\nu \leq u} 1$$

gelte für alle $\varepsilon > 0$ mit einer nur von ε abhängigen Konstanten $C(\varepsilon)$ die Abschätzung

$$(1.2) \quad N(u) \leq C(\varepsilon) \cdot \exp(\varepsilon u).$$

Das unendliche Produkt

$$(1.3) \quad g(s) = \prod_{\nu=1}^{\infty} (1 - \exp(-\lambda_\nu s))^{-1}$$

konvergiert wegen (1.2) für alle komplexen $s = \sigma + it$ mit $\operatorname{Re} s > 0$.

Durchläuft l die abzählbare Folge aller Linearkombinationen der λ_ν mit nicht-negativen ganzen Koeffizienten, so ist

$$(1.4) \quad g(s) = \sum_l p(l) \cdot e^{-ls};$$

die Reihe konvergiert absolut für $\operatorname{Re} s > 0$ ¹⁾. Die nichtnegativen Koeffizienten $p(l)$ sind gleich der Anzahl der Partitionen von l in Summanden aus der Folge $\{\lambda_\nu\}$ und auch gleich der Anzahl der Lösungen der diophantischen Gleichung

$$l = r_1 \lambda_1 + r_2 \lambda_2 + \dots$$

in ganzen Zahlen $r_j \geq 0$. Die summatorische Funktion

$$(1.5) \quad P(u) = \sum_{l \leq u} p(l)$$

wächst monoton mit u und genügt wegen (1.4) der zu (1.2) analogen Abschätzung

$$(1.6) \quad P(u) \leq g(\sigma) \exp(\sigma u)$$

¹⁾ Es ist $\sum_{l \leq \lambda_n} p(l) \cdot \exp(-l\sigma) \leq \prod_{\nu \leq n} (1 - \exp(-\lambda_\nu \sigma))^{-1}$.

für alle $\sigma > 0$. Wegen (1.3) ist somit für $\sigma > 0$

$$(1.7) \quad \log g(\sigma) = - \sum \log(1 - \exp(-\lambda_v \sigma)) \\ = - \int_{\lambda_1}^{\infty} \log(1 - e^{-u\sigma}) dN(u) = \sigma \int_{\lambda_1}^{\infty} N(u) \cdot (e^{u\sigma} - 1)^{-1} du,$$

und wegen (1.6) ist

$$(1.8) \quad g(\sigma) = \int_0^{\infty} e^{-u\sigma} dP(u) = \sigma \cdot \int_0^{\infty} P(u) e^{-u\sigma} du,$$

mit für $\sigma > 0$ konvergenten Integralen.

1958 zeigte Kohlbecker [9], womit er bekannte Ergebnisse von Hardy und Ramanujan [7] verallgemeinerte, daß aus der asymptotischen Beziehung

$$(1.9) \quad N(u) \sim u^{\alpha} L(u), \quad (u \rightarrow \infty),$$

wobei $\alpha > 0$ und L von langsamem Wachstum²⁾ ist, die asymptotische Beziehung

$$(1.10) \quad \log P(u) \sim u^{\beta} L^*(u), \quad (u \rightarrow \infty),$$

folgt, wobei $\beta = \alpha \cdot (\alpha + 1)^{-1}$ ist und L^* eine Funktion von langsamem Wachstum ist, die durch L asymptotisch eindeutig bestimmt ist³⁾.

Parameswaran [13] behandelte 1961 mit etwas abgewandelten Methoden den Fall $\alpha = 0$, der sich den Methoden Kohlbeckers entzogen hatte.

In der vorliegenden Arbeit soll aus möglichst wenigen Voraussetzungen über $N(u)$ eine asymptotische Formel für $\log P(u)$ (sogar mit Restglied) hergeleitet werden⁴⁾ (vgl. Satz 2 in § 4); das erzielte Ergebnis umfaßt und verbessert Ergebnisse von Hardy und Ramanujan [7], Brigham [2], Kohlbecker [9] und Parameswaran [13]. Damit können alle diese Ergebnisse mit einer einheitlichen Methode ohne allzu großen Rechenaufwand hergeleitet werden.

Das wichtigste Beweishilfsmittel ist ein reeller Taubersatz (Satz 1 aus § 3), der als bekannt gelten kann {vgl. [7], [1], [11], [9], [13] und [18], [19] und [20]}. Die für unsere Zwecke günstigste Fassung (Satz 1) findet sich jedoch in keiner dieser Arbeiten.

In § 5 und § 6 werden einige Anwendungen von Satz 2 gegeben; insbesondere wird gezeigt, wie man die Ergebnisse von Kohlbecker aus dem Hauptsatz (Satz 2) gewinnen kann.

2. Beispiele

Durch Angabe der Folge $\{\lambda_v\}$ beschreiben wir nun einige öfters untersuchte Partitionenfunktionen. Es wird sich zeigen, daß der Hauptsatz (Satz 2) auf alle diese Beispiele anwendbar ist⁵⁾.

²⁾ D. h. es ist $L(u) > 0$, L stetig und $\lim_{u \rightarrow \infty} L(cu) \cdot (L(u))^{-1} = 1$ für alle c mit $0 < c < \infty$.

³⁾ Kohlbecker hat auch bewiesen, daß umgekehrt aus (1.10) die Beziehung (1.9) gefolgert werden kann. Mit dieser Fragestellung beschäftigen wir uns hier nicht.

⁴⁾ In einer weiteren Arbeit soll aus schärferen Voraussetzungen über $N(u)$ auf das asymptotische Verhalten von $P(u)$ geschlossen werden.

⁵⁾ Für viele dieser Partitionenfunktionen sind bessere Ergebnisse bekannt (mindestens asymptotische Formeln für $P(u)$), nämlich für 2.1a (vgl. Ostmann [12], § 7), 2.1b ([16]), 2.2a ([8]), 2.2b ([16]; vgl. auch die Examensarbeit von K. H. Indlekofer, Freiburg 1967), 2.3a ([10], [3], [13], [17c]) und für 2.3c (W. Pohl, Examensarbeit Freiburg 1967). 2.2a, 2.3a, 2.3b und 2.3c werden auch in der in Fußnote 4) genannten Arbeit enthalten sein.

2. 1a. Uneingeschränkte Partitionen: $\lambda_\nu = \nu$; es ist $N(u) = u + O(1)$.
2. 1b. Partitionen in Primzahlen: $\lambda_\nu = p_\nu$, wobei p_ν die ν -te Primzahl bezeichnet; hier ist $N(u) \sim u (\log u)^{-1}$.
2. 2a. Partitionen in k -te Potenzen: $\lambda_\nu = \nu^k$; $N(u) = u^{\frac{1}{k}} + O(1)$.
2. 2b. $\lambda_\nu = f(\nu)$, wobei f ein ganzwertiges Polynom vom Grade k mit höchstem Koeffizienten 1 ist; hier ist $N(u) = u^{\frac{1}{k}} + O(1)$.
2. 3a. Mahlers Partitionenproblem: $\lambda_\nu = 2^{\nu-1}$; $N(u) = c \cdot \log u + O(1)$.
2. 3b. $\lambda_\nu = \nu^\nu$, oder $\lambda_\nu = \Gamma(\nu + 1)$; beide Male ist $N(u) \sim \log u \cdot (\log \log u)^{-1}$.
2. 3c. $\lambda_\nu = 2^{2^{\nu-1}}$; $N(u) \sim c \log \log u$.
2. 4a. $\lambda_\nu = \log \left(\prod_{p \leq p_\nu} p \right)$, wobei p_ν die ν -te Primzahl bezeichnet und p prim ist. Hier ist $N(u) \sim u (\log u)^{-1}$. In diesem Falle ist $P(u)$ die Anzahl der natürlichen Zahlen $q < u$ der Gestalt

$$q = 2^{\alpha_2} \cdot 3^{\alpha_3} \cdot 5^{\alpha_5} \cdots p^{\alpha_p} \text{ mit } \alpha_2 \geq \alpha_3 \geq \alpha_5 \geq \dots$$

Dieses Partitionenproblem stammt von Hardy und Ramanujan [7] und ist deshalb von Interesse, weil jede highly composite number (vgl. [15]) von der angegebenen Gestalt ist; allerdings ist nicht jede Zahl der angegebenen Gestalt eine highly composite number.

Außerdem haben Erdős und Mirsky [5] gezeigt, daß — wenn $D(u)$ die Anzahl der verschiedenen Werte von $d(n) = \sum_{d|n} 1$ für $1 \leq n < u$ bezeichnet — für $u \rightarrow \infty$

$$\log D(u) \sim 2^{\frac{1}{2}} (\log \log u)^{-\frac{1}{2}} \cdot \log P(u)$$

ist; $P(u)$ ist hierbei die in (2. 4a) behandelte Partitionenfunktion.

2. 4b. $\lambda_\nu = \log (\Gamma(\nu + 1))$, $N(u) \sim u (\log u)^{-1}$.
2. 4c. $\lambda_\nu = \log (\nu^\nu)$, $N(u) \sim u (\log u)^{-1}$.

3. Ein reeller Taubersatz

Hardy und Ramanujan [7] zeigten 1917:

Wenn $A(u)$ monoton mit u wächst, das Integral

$$(3. 1) \quad f(\sigma) = \sigma \int_0^\infty A(u) \exp(-u\sigma) du$$

für $\sigma > 0$ konvergiert und für $\sigma \rightarrow 0$

$$(3. 2) \quad \log f(\sigma) \sim A\left(\frac{1}{\sigma}\right)^\alpha \cdot \left(\log \frac{1}{\sigma}\right)^\beta, \quad \alpha > 0,$$

ist, dann ist für $u \rightarrow \infty$ mit einer explizit angebbaren Konstanten $\gamma = \gamma(\alpha, \beta, A)$

$$(3. 3) \quad \log A(u) \sim \gamma \cdot u^{\frac{\alpha}{\alpha+1}} \cdot (\log u)^{\frac{\beta}{\alpha+1}}.$$

Kohlbecker [9] verallgemeinerte diesen Satz 1958, indem er die Voraussetzung (3. 2) durch

$$(3. 2') \quad \log f(\sigma) \sim \left(\frac{1}{\sigma}\right)^\alpha \cdot L\left(\frac{1}{\sigma}\right), \quad \alpha > 0, \quad \sigma \rightarrow 0,$$

mit einer Funktion L von langsamem Wachstum ersetzte.

Den ausgeschlossenen Fall $\alpha = 0$ behandelte Parameswaran [13] im Jahre 1961 mit einer passenden Variante der Hardy-Ramanujanschen Beweismethode.

Diese Methode ist auch geeignet, in (3. 3) ein besseres Ergebnis zu erzielen, wenn man die Voraussetzung (3. 2) verschärft. Den ersten Satz dieser Art bewiesen Avakumović und Karamata [1] im Jahre 1936:

Aus

$$(3. 2'') \quad \log f(\sigma) = A \cdot \sigma^{-1} + O(1)$$

folgt

$$(3. 3'') \quad \log A(u) = 2(Au)^{\frac{1}{2}} + O(u \log u)^{\frac{1}{4}}.$$

Einen wesentlich allgemeineren Satz dieser Art folgerten Martin und Wiener [11] aus dem allgemeinen Wienerschen Taubersatz (vgl. etwa [21], chap. V). Vor kurzem zeigte E. Wagner [18], daß aus

$$(3. 2'') \quad \log f(\sigma) = \varphi(\sigma) + O(1),$$

wobei die Funktion φ gewisse weitere Bedingungen erfüllt, eine asymptotische Formel für $\log A(u)$ mit Restglied folgt (vgl. [18], Satz 1 und Satz 2). Die Bedingungen des Satzes von Wagner sind etwas zu allgemein und daher für unsere Zwecke ungeeignet; wir geben daher einen spezielleren Satz, dessen Voraussetzungen in § 4 besser nachzuprüfen sind; die gemachten Voraussetzungen ähneln den Voraussetzungen eines Inghamschen Taubersatzes ([8], Satz 1). Eine Skizze des Beweises (wiederum mit Hilfe der Hardy-Ramanujanschen Methode) geben wir der Vollständigkeit halber. Wir zeigen folgenden

Satz 1. Sei $A(u)$ eine mit u monoton wachsende reellwertige Funktion mit $A(0) = 0$; das Integral

$$(3. 1) \quad f(\sigma) = \sigma \int_0^\infty A(u) \exp(-u\sigma) du$$

sei für $\sigma > 0$ konvergent, und es gelte für $\sigma \rightarrow 0$

$$(3. 4) \quad \log f(\sigma) = \varphi(\sigma) + O(1),$$

wobei die zweimal stetig differenzierbare Funktion φ folgende Bedingungen erfüllt:

$$(3. 5a) \quad \varphi''(\sigma) > 0, \varphi''(\sigma) \nearrow \text{ für } \sigma \searrow 0;$$

$$(3. 5b) \quad -\sigma \varphi'(\sigma) \nearrow \infty \text{ für } \sigma \searrow 0;$$

es existieren Konstanten $\varrho \geq 0$ und C_1 , so daß die Ungleichung

$$(3. 6) \quad \frac{\sigma(\varphi''(\sigma))^{\varrho+1}}{|\varphi'(\sigma)|^{1\varrho+2}} \leq C_1$$

besteht. Dann ist für $T \rightarrow \infty$

$$(3. 7) \quad \log A(T) = \varphi(\sigma_T) + T\sigma_T + O\left\{T\sigma_T \cdot \left(\frac{\varphi''(\sigma_T)}{|\varphi'(\sigma_T)|^2} \cdot \log \frac{|\varphi'(\sigma_T)|^2}{\varphi''(\sigma_T)}\right)^{\frac{1}{2}}\right\},$$

wobei σ_T durch

$$(3. 8) \quad -\varphi'(\sigma_T) = T$$

für großes T eindeutig bestimmt ist⁶⁾.

⁶⁾ Wegen (3. 5b). $\varphi''(\sigma) \nearrow$ bedeutet, daß φ'' (schwach) monoton wächst; entsprechend ist $\searrow$ zu verstehen.

Bemerkungen. Aus (3. 5b) folgt

$$(3. 8') \quad \sigma_T \searrow 0 \text{ für } T \rightarrow \infty;$$

weiter folgt aus (3. 5b), wenn man φ'' mit Hilfe von (3. 6) abschätzt,

$$(3. 5c) \quad \varphi''(\sigma) \cdot |\varphi'(\sigma)|^{-2} \rightarrow 0 \text{ für } \sigma \rightarrow 0.$$

Das Fehlerglied ist somit von kleinerer Ordnung als das Hauptglied.

Schließlich folgt aus (3. 5b) durch Differentiation

$$(3. 5d) \quad \sigma \varphi''(\sigma) \geq -\varphi'(\sigma),$$

also gilt (mit (3. 5b))

$$\sigma^2 \varphi''(\sigma) \rightarrow \infty;$$

die im Fehlergliede stehende Funktion ist somit nicht beschränkt für $\sigma \rightarrow 0$.

Beweis von Satz 1:

a) Wegen (3. 1) und der Monotonie von $A(u)$ ist

$$f(\sigma) > \sigma \int_T^\infty A(u) \exp(-u\sigma) du \geq A(T) \exp(-T\sigma),$$

also folgt mit (3. 4) bei günstigster Wahl von σ :

$$(3. 7a) \quad A(T) \leq \gamma_1 \cdot \exp\{T\sigma_T + \varphi(\sigma_T)\} \text{ für } T \geq T_0$$

mit einer Konstanten $\gamma_1 > 0$.

b) Zur Abschätzung nach unten zerlegt man das Integral in (3. 1) in

$$\sigma \left\{ \int_0^{T_0} + \int_{T_0}^{(1-\varepsilon)T} + \int_{(1-\varepsilon)T}^R + \int_R^\infty \right\} = I_0 + I_1 + I_2 + I_3,$$

wobei $R = (1 + \varepsilon) \cdot T$ ist und $\varepsilon < \frac{1}{4}$ noch passend bestimmt wird, so daß I_1 und I_3 gegen I_2 vernachlässigbar klein werden. Offenbar ist $I_0 = O(1)$. Mit der Abkürzung

$$(3. 9) \quad \psi(u) = -u\sigma_T + u\sigma_u + \varphi(\sigma_u)$$

ist wegen (3. 7a)

$$I_1 \leq \gamma_1 \sigma_T \int_{T_0}^{(1-\varepsilon)T} \exp(\psi(u)) du;$$

entwickelt man den Exponenten des Integranden an der Stelle T , wobei man

$$(3. 9') \quad \begin{cases} \psi(T) = \varphi(\sigma_T), \\ \psi'(u) = -\sigma_T + \sigma_u \\ \psi''(u) = \frac{d}{du}(\sigma_u) = -(\varphi''(\sigma_u))^{-1} \end{cases}$$

beachtet, so erhält man

$$\psi(u) = \varphi(\sigma_T) + \frac{1}{2}(T - u)^2 \cdot \psi''(u^*)$$

mit einer Zwischenstelle u^* zwischen u und T . Aus (3. 8'), (3. 5a) und (3. 9') folgt

$$\psi''(u^*) \leq \psi''(T) = -(\varphi''(\sigma_T))^{-1},$$

also wird (mit $v = T - u$)

$$I_1 \leq \gamma_1 \sigma_T \exp \{ \varphi(\sigma_T) \} \cdot \int_{\varepsilon T}^{\infty} \exp \left\{ -\frac{1}{2} v^2 (\varphi''(\sigma_T))^{-1} \right\} dv,$$

und, nach kurzer Zwischenrechnung, wenn man nochmals (3.5a) und (3.8') beachtet,

$$(3.10a) \quad \begin{aligned} I_1 &\leq \gamma_1 \sigma_T \cdot \exp \left\{ \varphi(\sigma_T) - \frac{1}{2} \frac{1}{\varphi''(\sigma_T)} \cdot \varepsilon^2 T^2 \right\} \cdot \frac{\varphi''(\sigma_T)}{\varepsilon T} \\ &\leq \gamma_1 \sigma_T \cdot \exp \left\{ \varphi(\sigma_T) - \frac{1}{4} \frac{1}{\varphi''(\sigma_R)} \cdot \varepsilon^2 R^2 \right\} \cdot \frac{2\varphi''(\sigma_R)}{\varepsilon R}. \end{aligned}$$

c) Vor der Behandlung von I_3 soll die Differenz $\sigma_T - \sigma_R$ abgeschätzt werden. Wegen (3.5d) ist

$$(3.11) \quad 0 < \frac{-\varphi'(\sigma)}{\sigma \varphi''(\sigma)} \leq 1.$$

Aus $\varepsilon T = R - T = -\varphi'(\sigma_R) + \varphi'(\sigma_T) = (\sigma_T - \sigma_R) \varphi''(\sigma^*)$ mit einer Zwischenstelle σ^* , $\sigma_R \leq \sigma^* \leq \sigma_T$, erhält man unter Beachtung von (3.5a) und (3.11)

$$\sigma_T - \sigma_R \leq \frac{\varepsilon T}{\varphi''(\sigma_T)} = \frac{\varepsilon |\varphi'(\sigma_T)|}{\sigma_T \varphi''(\sigma_T)} \cdot \sigma_T \leq \varepsilon \cdot \sigma_T,$$

also

$$(3.12) \quad 0 \leq \sigma_T - \sigma_R \leq \varepsilon \sigma_T; \quad 0 \leq \sigma_T - \sigma_R \leq 2\varepsilon \sigma_R,$$

und, in der anderen Richtung,

$$(3.13) \quad \sigma_T - \sigma_R \geq \frac{\varepsilon T}{\varphi''(\sigma_R)} \geq \frac{1}{2} \frac{\varepsilon R}{\varphi''(\sigma_R)}.$$

d) Benützt man den Mittelwertsatz, (3.9') und (3.8'), so erhält man für $u \geq R$

$$\begin{aligned} \varphi(u) &\leq \varphi(R) - (u - R) |\varphi'(R)| \\ &\leq \left\{ \varphi(T) - \frac{1}{2} (R - T)^2 |\varphi''(R)| \right\} - (u - R) |\varphi'(R)|. \end{aligned}$$

Beachtet man (3.9'), so folgt hieraus

$$(3.10b) \quad I_3 \leq \gamma_1 \sigma_T \int_R^{\infty} \exp \{ \varphi(u) \} \cdot du \leq \frac{\gamma_1 \sigma_T}{|\varphi'(R)|} \cdot \exp \left\{ \varphi(\sigma_T) - \frac{1}{2} \varepsilon^2 T^2 \cdot \frac{1}{\varphi''(\sigma_R)} \right\}.$$

Aus (3.10a), (3.10b) mit (3.9') und (3.13) sowie (3.12) erhält man nun

$$(3.14) \quad I_1 + I_3 \leq 2\gamma_1 \cdot \exp \{ \varphi(\sigma_T) \} \cdot e^U$$

mit dem Exponenten

$$U = -\frac{1}{4} \varepsilon^2 R^2 \cdot \{\varphi''(\sigma_R)\}^{-1} + \log \frac{2\varphi''(\sigma_R)}{\varepsilon R} + \log 2\sigma_R.$$

Wir setzen nun mit der Konstanten $K = 2(4\rho + 1)^{\frac{1}{2}} > 0$

$$(3.15) \quad \varepsilon = \varepsilon(R) = K \left\{ \frac{\varphi''(\sigma_R)}{|\varphi'(\sigma_R)|^2} \cdot \log \frac{|\varphi'(\sigma_R)|^2}{\varphi''(\sigma_R)} \right\}^{\frac{1}{2}}.$$

Wegen (3. 5 c) und (3. 8') strebt ε gegen 0 für $R \rightarrow \infty$. Also ist offenbar mit einer geeigneten Konstanten K^*

$$\varepsilon \leq K^* \{\varphi''(\sigma_R) \cdot |\varphi'(\sigma_R)|^{-2}\}^{\frac{1}{4}},$$

also wird wegen (3. 6)

$$\frac{\sigma_R \varphi''(\sigma_R)}{|\varphi'(\sigma_R)|} \cdot \varepsilon^{4\varrho} \leq C_1 \cdot K^{*4\varrho},$$

somit ist (mit $|\varphi'(\sigma_R)| = R$)

$$(3. 16) \quad \log \frac{4\varphi''(\sigma_R) \cdot \sigma_R}{\varepsilon R} \leq \log 4 + (4\varrho + 1) \log \frac{1}{\varepsilon} + \log (C_1 K^{*4\varrho}) \\ = O(1) + (4\varrho + 1) \cdot \log \frac{1}{\varepsilon}.$$

Wegen der Wahl von K mit (3. 15), (3. 16) und wegen der Abschätzung

$$\log \frac{1}{\varepsilon} \leq \frac{1}{2} \log \frac{|\varphi'|^2}{\varphi''}$$

(für genügend kleines σ) folgt

$$U \leq -\frac{1}{4} K^2 \log \frac{|\varphi'|^2}{\varphi''} + O(1) + (4\varrho + 1) \frac{1}{2} \log \frac{|\varphi'|^2}{\varphi''} = O(1) - \frac{1}{8} K^2 \cdot \log \frac{|\varphi'(\sigma_R)|^2}{\varphi''(\sigma_R)}.$$

Wegen (3. 5 c) folgt

$$U \rightarrow -\infty \text{ für } R \rightarrow \infty,$$

und es wird, wenn ε nach (3. 15) gewählt ist,

$$(3. 17) \quad I_1 + I_3 = o\{\exp \varphi(\sigma_T)\} \text{ für } T \rightarrow \infty.$$

Aus der Monotonie von $A(u)$, aus (3. 4) und (3. 17) folgt nun mit einer Konstanten $\gamma_2 > 0$ für genügend großes T

$$A(R) \exp \{-\sigma_T(1 - \varepsilon)T\} \geq \sigma_T \int_{(1-\varepsilon)T}^R A(u) \exp(-u\sigma_T) du = I_2 \\ \geq \gamma_2 \exp \varphi(\sigma_T) - (I_0 + I_1 + I_3) > \frac{1}{2} \gamma_2 \cdot \exp \varphi(\sigma_T),$$

also wird wegen $T\sigma_T \leq R\sigma_R$ (vgl. (3. 5 b))

$$(3. 18) \quad \log A(R) \geq O(1) + \varphi(\sigma_T) + T\sigma_T - \varepsilon R\sigma_R.$$

e) Nun müssen noch $\varphi(\sigma_T)$ und $T\sigma_T$ durch $\varphi(\sigma_R)$ und $R\sigma_R$ ausgedrückt werden. Zunächst ist

$$0 \leq R\sigma_R - T\sigma_T \leq R \cdot |\sigma_R - \sigma_T| + \varepsilon T\sigma_T \leq 3\varepsilon R\sigma_R$$

nach (3. 12); aus dem Mittelwertsatz erhält man

$$0 \leq \varphi(\sigma_R) - \varphi(\sigma_T) \leq |\sigma_R - \sigma_T| \cdot |\varphi'(\sigma_R)| \leq R \cdot |\sigma_R - \sigma_T| \leq 2\varepsilon \cdot R\sigma_R;$$

dies zusammen mit (3. 18) gibt die benötigte Abschätzung nach unten:

$$(3. 7b) \quad \log A(R) \geq O(1) + \varphi(\sigma_R) + R\sigma_R - 6\varepsilon \cdot R\sigma_R.$$

Aus (3. 7a) und (3. 7b) folgt die Behauptung (3. 7).

Mit derselben Methode kann man, wenn man (3. 4) durch

$$(3. 4^*) \quad \log f(\sigma) \sim \varphi(\sigma) \text{ für } \sigma \rightarrow 0$$

ersetzt und die Voraussetzung

$$(3.6^*) \quad 0 < \gamma^* \leq \frac{(\varphi'(\sigma))^2}{\varphi(\sigma)\varphi''(\sigma)} \leq C_1^*$$

hinzunimmt, zeigen, daß für $T \rightarrow \infty$

$$(3.7^*) \quad \log A(T) \sim \varphi(\sigma_T) + T\sigma_T$$

gilt. Herr E. Wagner hat mir mitgeteilt, daß er (sogar unter schwächeren Voraussetzungen) einen derartigen Satz bewiesen hat. Seine Arbeit ([20]) wird in den Mathematischen Nachrichten erscheinen.

Es wäre sehr zu wünschen, einen solchen Satz ohne die linke Hälfte der Ungleichung (3.6*) zu beweisen (die etwa $\varphi(\sigma) = |\log \sigma|^2$ ausschließt).

4. Das Hauptergebnis

Wir verwenden die Bezeichnungen aus § 1. Der Taubersatz (Satz 1) aus § 3 soll angewandt werden, um aus schwachen Voraussetzungen über die Anzahlfunktion $N(u)$ der λ_v eine asymptotische Formel für $\log P(u)$ (mit Restglied) herzuleiten, nämlich das folgende Ergebnis⁷⁾:

Satz 2. *Es strebe*

$$(4.1) \quad N(u) = \sum_{\lambda_v < u} 1 \rightarrow \infty \text{ für } u \rightarrow \infty,$$

und es gelte mit einer Konstanten $C_2 > 0$ für $u > \lambda_1$ die Abschätzung

$$(4.2) \quad N(2u) \leq C_2 N(u).$$

Dann gilt für $u \rightarrow \infty$ die asymptotische Formel

$$(4.3) \quad \log P(u) = \varphi(\sigma_u) + u\sigma_u + O\left\{u\sigma_u \left(\frac{\varphi''(\sigma_u)}{|\varphi'(\sigma_u)|^2}\right) \cdot \log \frac{|\varphi'(\sigma_u)|^2}{\varphi''(\sigma_u)}\right\}^{\frac{1}{2}},$$

wobei

$$(4.4) \quad \varphi(\sigma) = - \sum_1^{\infty} \log(1 - \exp(-\lambda_v \sigma))$$

ist und σ_u durch $-\varphi'(\sigma_u) = u$ für große u eindeutig bestimmt ist.

Bemerkung 1. Satz 2 bleibt richtig, wenn die Voraussetzung (4.2) durch (4.7) ersetzt wird.

Bemerkung 2. Das Fehlerglied ist von geringerer Ordnung als das Hauptglied (vgl. (3.5c)).

Bemerkung 3. Für sämtliche in § 2 genannten Beispiele sind die Voraussetzungen von Satz 2 erfüllt, also gilt für alle diese Beispiele eine asymptotische Formel der Gestalt (4.3).

Beweis. Wir wenden den Taubersatz aus § 3 an; da über $N(u)$ zu wenig bekannt ist, können wir keine asymptotische Beziehung der Gestalt (3.4) für $\log g(\sigma)$ ⁸⁾ herleiten. Wir erzwingen diese Beziehung (3.4), indem wir

$$(4.4) \quad \varphi(\sigma) = \log g(\sigma) = - \sum_v \log(1 - e^{-\lambda_v \sigma}) = - \int_{\lambda_1}^{\infty} \log(1 - e^{-u\sigma}) dN(u)$$

⁷⁾ Satz 2 macht eine Aussage über das asymptotische Verhalten des Logarithmus der Partitionenfunktion $P(u)$, also eine „schwache asymptotische Aussage“ über $P(u)$.

⁸⁾ $g(\sigma)$ ist durch (1.3) erklärt.

(vgl. (1. 7)) setzen; unsere einzige Aufgabe ist, die in Satz 1 von φ verlangten Eigenschaften nachzuweisen.

Offenbar ist $\varphi(\sigma)$ mindestens dreimal stetig differenzierbar (für $\sigma > 0$), und es ist

$$(4. 5') \quad \varphi'(\sigma) = - \int_{\lambda_1}^{\infty} \frac{u}{e^{u\sigma} - 1} dN(u) = \int_{\lambda_1}^{\infty} N(u) \frac{d}{du} \left(\frac{u}{e^{u\sigma} - 1} \right) du,$$

$$(4. 5'') \quad \varphi''(\sigma) = \int_{\lambda_1}^{\infty} \frac{u^2 e^{u\sigma}}{(e^{u\sigma} - 1)^2} dN(u),$$

$$(4. 5''') \quad \varphi'''(\sigma) = - \int_{\lambda_1}^{\infty} \frac{u^3 e^{u\sigma} (e^{u\sigma} + 1)}{(e^{u\sigma} - 1)^3} dN(u),$$

somit ist $\varphi' \leq 0$, $\varphi'' > 0$, $\varphi''' \leq 0$, also $\varphi''(\sigma) \nearrow$ für $\sigma \searrow 0$, und (3. 5 a) ist erfüllt. Weiter ist

$$(4. 6) \quad -\sigma \varphi'(\sigma) = \int_{\lambda_1}^{\infty} \frac{u\sigma}{e^{u\sigma} - 1} dN(u) \nearrow \text{ für } \sigma \searrow 0,$$

weil $h(v) = v \cdot (e^v - 1)^{-1} \nearrow$ für $v \searrow 0$, wie eine einfache Überlegung zeigt⁹⁾.

Da $N(u)$ für $u \rightarrow \infty$ gegen Unendlich strebt, existiert zu jedem beliebig großen K ein $M = M(K)$ so, daß $N(u) \geq K$ für $u \geq M$ gilt; somit ist, wenn man

$$\frac{d}{du} (u(e^{u\sigma} - 1)^{-1}) \leq 0$$

beachtet,

$$\begin{aligned} -\varphi'(\sigma) &= - \int_{\lambda_1}^{\infty} N(u) \frac{d}{du} \left(\frac{u}{e^{u\sigma} - 1} \right) du \geq - \int_M^{\infty} \geq -K \int_M^{\infty} \frac{d}{du} \left(\frac{u}{e^{u\sigma} - 1} \right) du \\ &= KM \cdot (e^{\sigma M} - 1)^{-1} \geq \frac{1}{2} K \sigma^{-1} \end{aligned}$$

für genügend kleines σ , und (3. 5 b) ist erfüllt.

Bisher wurde die Voraussetzung (4. 2) noch nicht benötigt, und es steht nur noch der Beweis von (3. 6) aus. Zum Beweise von (3. 6) scheint man aber ohne (4. 2) oder andere Bedingungen nicht auszukommen. Aus (4. 5') erhält man mit der Substitution $u = 2v$ unter Beachtung von (4. 2)

$$\begin{aligned} -\varphi'(\sigma) &= - \int_{\lambda_1}^{\infty} N(u) \frac{d}{du} \left(\frac{u}{e^{u\sigma} - 1} \right) du = -2 \int_{\lambda_1}^{\infty} N(2v) \frac{d}{dv} \left(\frac{v}{e^{v2\sigma} - 1} \right) dv - 2 \int_{\frac{1}{2}\lambda_1}^{\lambda_1} \dots dv \\ &\leq 2C_2 |\varphi'(2\sigma)| + O\left(\frac{1}{\sigma}\right) \leq 3C_2 |\varphi'(2\sigma)| \end{aligned}$$

für genügend kleines σ (unter Benutzung von (3. 5 b)).

Wir benutzen jetzt nur noch die eben hergeleitete Beziehung

$$(4. 7) \quad |\varphi'(\sigma)| \leq 3C_2 |\varphi'(2\sigma)| \text{ (für genügend kleines } \sigma \text{)}.$$

⁹⁾ Es ist $h'(v) \leq 0$ für $v \geq 0$.

Die in $\operatorname{Re} s > 0$ erklärte Funktion

$$\varphi'(s) = - \sum_{\nu} \lambda_{\nu} \cdot (\exp(\lambda_{\nu} s) - 1)^{-1}$$

ist holomorph für $\operatorname{Re} s > 0$; wegen

$$|\exp(\lambda_{\nu} s) - 1|^{-1} \leq |\exp(\lambda_{\nu} \sigma) - 1|^{-1}$$

ist

$$(4.8) \quad |\varphi'(s)| \leq |\varphi'(\operatorname{Re} s)|.$$

Die Cauchysche Integralformel gibt

$$(4.9) \quad \varphi''(\sigma) = \frac{1}{2\pi i} \int_{\mathfrak{K}} \varphi'(z) (z - \sigma)^{-2} dz,$$

wobei über den Kreis $\mathfrak{K} = \left\{ z; |z - \sigma| = \frac{1}{2} \right\}$ integriert wird. Somit ist wegen (4.8), der Monotonie von $\varphi'(\sigma)$ und wegen (4.7)

$$0 \leq \varphi''(\sigma) \leq \frac{2}{\sigma} \cdot \sup_{z \in \mathfrak{K}} |\varphi'(z)| \leq \frac{2}{\sigma} \left| \varphi' \left(\frac{1}{2} \sigma \right) \right| \leq 6C_2 \cdot \frac{|\varphi'(\sigma)|}{\sigma},$$

also gilt

$$(4.10) \quad \frac{\sigma \varphi''(\sigma)}{|\varphi'(\sigma)|} \leq 6C_2,$$

und (3.6) ist mit $\varrho = 0$ und $C_1 = 6C_2$ nachgewiesen. Satz 1 gibt nun (unter Beachtung von (1.8) und (1.5)) die Behauptung.

5. Ein Taubersatz für Partitionen

Aus Satz 2 kann folgender Satz abgeleitet werden, der für Anwendungen unter Umständen günstiger ist.

Satz 3. Sei

$$(5.1) \quad g(\sigma) = \prod_1^{\infty} (1 - \exp(-\lambda_{\nu} \sigma))^{-1} = \sigma \int_0^{\infty} P(u) e^{-u\sigma} du$$

für $\sigma > 0$ konvergent; die Funktion $\Phi(\sigma)$ sei für $\sigma > 0$ stetig differenzierbar, und es gelte

$$(5.2) \quad -\Phi'(\sigma) \nearrow \infty \text{ für } \sigma \searrow 0.$$

Sei weiter eine der folgenden drei Bedingungen erfüllt:

$$(5. I) \quad a) \frac{d}{d\sigma} (\log g(\sigma)) = \frac{g'(\sigma)}{g(\sigma)} \sim \Phi'(\sigma) \text{ für } \sigma \rightarrow 0.$$

$$b) |\Phi'(\sigma)| \leq C_3 \cdot |\Phi'(2\sigma)|.$$

$$(5. II) \quad a) \log g(\sigma) \sim \Phi(\sigma).$$

$$b) \text{ Mit der Bezeichnung (1.1) sei } N(2u) \leq C_2 N(u) \text{ für } u > \lambda_1, \text{ und es gelte}$$

$$c') \sigma |\Phi'(\sigma)| \cdot (\Phi(\sigma))^{-1} \geq \gamma_4 \text{ für alle genügend kleinen } \sigma, \text{ oder}$$

$$c'') \sigma |\Phi'(\sigma)| \cdot (\Phi(\sigma))^{-1} = o(1) \text{ für } \sigma \rightarrow 0.$$

$$(5. III) \quad a) \log g(\sigma) \sim \Phi(\sigma) \text{ und}$$

$$b) \text{ für jedes reelle } c, 0 < c < \infty, \text{ ist } \lim_{u \rightarrow \infty} \frac{N(cu)}{N(u)} = 1.$$

Dann gilt die asymptotische Beziehung

$$(5.3) \quad \log P(u) \sim \Phi(t_u) + u t_u, \quad (u \rightarrow \infty),$$

wobei t_u durch $-\Phi'(t_u) = u$ bestimmt ist. Gilt (5. IIc'') oder (5. IIb), so erhält man schärfer

$$(5.3') \quad \log P(u) \sim \Phi(t_u).$$

Bemerkung 1. Für $u \rightarrow \infty$ geht t_u monoton gegen Null.

Bemerkung 2. Satz 3 (5. II) verallgemeinert den alten Taubersatz von Hardy und Ramanujan [7] (vgl. § 3) im Spezialfall, daß $f(\sigma)$ (vgl. (3.1)) erzeugende Funktion für eine gewisse Partitionenfunktion ist und dehnt ihn insbesondere auf den Fall $\alpha = 0$ aus. Man vergleiche hierzu auch das Ergebnis von Wagner [20].

Bemerkung 3. Ist $h(\sigma) > 0$, stetig differenzierbar für $\sigma > 0$, $-h'(\sigma) \nearrow \infty$ für $\sigma \searrow 0$ so gilt

$$\frac{-h'(\sigma) \cdot \sigma}{h(\sigma)} = o(1) \text{ für } \sigma \rightarrow 0$$

genau dann, wenn $h(\sigma)$ von langsamem Wachstum (für $\sigma \rightarrow 0$) ist, d. h. wenn $\lim_{\sigma \rightarrow 0+} \frac{h(c\sigma)}{h(\sigma)} = 1$ für alle c mit $0 < c < \infty$ ist.

Der ganz einfache Beweis darf unterdrückt werden. Wegen dieser Bemerkung ist Satz 3 unter den Bedingungen (5. II) mit c'' ein Spezialfall eines Ergebnisses von Parameswaran ([13], Theorem III).

Bemerkung 4. Satz 3 (5. I) kommt mit den wenigsten Voraussetzungen aus. Bei zahlentheoretischen Anwendungen ist es im allgemeinen nicht schwieriger, $\frac{d}{d\sigma}(\log g(\sigma))$ zu approximieren, als $\log g(\sigma)$ zu approximieren (vgl. (4.4) und (4.5')). Außerdem läßt sich (5. Ia) mit Hilfe von Lemma 1 folgern, wenn man $\Phi(\sigma) \sim \log g(\sigma)$,

$$\sigma |\Phi'(\sigma)| \cdot (\Phi(\sigma))^{-1} \geq \gamma_4$$

für alle genügend kleinen σ und $\Phi''(\sigma) \leq \gamma_3 \sigma^{-1} \cdot |\Phi'(\sigma)|$ (und natürlich die zweimalige stetige Differenzierbarkeit von Φ) voraussetzt.

Beweis von Satz 3. Wir zeigen zunächst, daß (5.3) aus den Voraussetzungen (5. I) gefolgert werden kann. Wir setzen $\varphi(\sigma) = \log g(\sigma)$. Wegen (5. Ia) und (5. Ib) ist (4.7) erfüllt, also gibt Satz 2

$$\log P(u) \sim \varphi(\sigma_u) + u \sigma_u,$$

wobei σ_u durch $-\varphi'(\sigma_u) = u$ bestimmt ist. Um (5.3) herzuleiten, muß ein Zusammenhang zwischen σ_u und t_u hergestellt werden. Hierzu wird die Bedingung (5. Ib) nicht mehr benötigt.

Aus $-\Phi'(t_u) = u$ folgt wegen $\varphi'(\sigma) \sim \Phi'(\sigma)$: Zu jedem $\varepsilon > 0$ existiert ein $u_0(\varepsilon)$, so daß für $u \geq u_0$

$$-\varphi'(t_u) = u(1 + \theta) \text{ mit } |\theta| < \varepsilon$$

ist. Somit ist $t_u = \sigma_{u(1+\theta)}$. Dieselbe Rechnung, die in § 3 zu (3.12) führte¹⁰⁾, gibt das Ergebnis

$$|t_u - \sigma_u| = |\sigma_{u(1+\theta)} - \sigma_u| \leq 2\varepsilon \cdot \text{Min}(\sigma_u, t_u).$$

¹⁰⁾ Die dort verwendete Monotonie von $-\sigma\varphi'(\sigma)$ ist nach § 4 erfüllt.

Somit ist für $u \rightarrow \infty$

$$(5.4) \quad t_u \sim \sigma_u.$$

Mit einer Zwischenstelle σ^* zwischen t_u und σ_u folgt hieraus

$$|\varphi(t_u) - \varphi(\sigma_u)| = |\varphi'(\sigma^*)| \cdot |t_u - \sigma_u|.$$

Wegen der Monotonie von φ' unter Beachtung von $|\varphi'(\sigma_u)| = u$ und $|\varphi'(t_u)| \sim |\Phi'(t_u)| = u$ und (5.4) folgt nun für $u \rightarrow \infty$

$$\varphi(t_u) - \varphi(\sigma_u) = o(u\sigma_u),$$

woraus man leicht

$$(5.5) \quad |\varphi(\sigma_u) + u\sigma_u - \Phi(t_u) - ut_u| = o(u\sigma_u + \varphi(\sigma_u))$$

erhält. Also gilt für $u \rightarrow \infty$

$$\log P(u) \sim \Phi(t_u) + ut_u.$$

Um aus (5. II) die Behauptung herzuleiten, benötigen wir die Hardy-Littlewood'sche Technik¹¹⁾ der Differentiation asymptotischer Formeln, die wir als Hilfssatz formulieren:

Lemma 1. Seien h bzw. H positive, zweimal bzw. einmal stetig differenzierbare Funktionen mit

$$(5.6') \quad h(\sigma) \sim H(\sigma) \text{ für } \sigma \rightarrow 0,$$

$$(5.6'') \quad -h'(\sigma) \nearrow \infty, \quad -H'(\sigma) \nearrow \infty \text{ für } \sigma \searrow 0,$$

$$(5.7) \quad 0 < h''(\sigma) \leq \gamma_3 \cdot \sigma^{-1} \cdot |h'(\sigma)|.$$

Dann gilt für $\sigma \rightarrow 0$

$$H'(\sigma) = h'(\sigma) + o(\sigma^{-1}h(\sigma)) + o(h'(\sigma)).$$

Zusatz. Gilt zusätzlich die Voraussetzung

$$-\sigma h'(\sigma) \{h(\sigma)\}^{-1} \geq \gamma_5 > 0,$$

so ist für $\sigma \rightarrow 0$

$$h'(\sigma) \sim H'(\sigma).$$

Beweis. Sei $\varepsilon > 0$ gegeben; wähle $\delta = 2\varepsilon \cdot \gamma_3^{-1}$. Wegen der Monotonie von $-H'$ ist

$$\begin{aligned} -\delta\sigma H'(\sigma) &\geq \int_{\sigma}^{\sigma+\delta\sigma} (-H'(\sigma)) d\sigma = H(\sigma) - H(\sigma + \delta\sigma) \\ &\geq h(\sigma) - h(\sigma + \delta\sigma) - 2\varepsilon\delta \cdot h(\sigma), \end{aligned}$$

falls $\sigma \leq \sigma_0(\varepsilon, \delta)$, wegen (5.6') und der nach (5.6'') für genügend kleines σ bestehenden Monotonie von h . Die Taylorsche Formel gibt mit einer Stelle σ^* zwischen σ und $\sigma + \delta\sigma$

$$\begin{aligned} -\delta\sigma H'(\sigma) &\geq \left(-\delta\sigma h'(\sigma) - \frac{1}{2} \delta^2 \sigma^2 h''(\sigma^*) \right) - 2\varepsilon\delta h(\sigma) \\ &\geq \delta\sigma \cdot \left\{ -h'(\sigma) - \frac{1}{2} \delta \gamma_3 |h'(\sigma)| - 2\varepsilon \cdot \frac{1}{\sigma} \cdot h(\sigma) \right\} \end{aligned}$$

wegen (5.7) und der Monotonie von h' . Somit folgt

$$-H'(\sigma) \geq -h'(\sigma) - \varepsilon \cdot |h'(\sigma)| - 2\varepsilon \cdot \sigma^{-1} \cdot h(\sigma).$$

Da eine entsprechende Abschätzung in der anderen Richtung gilt, ist Lemma 1 bewiesen. Der Beweis des Zusatzes ist offensichtlich.

¹¹⁾ [6], § 7. 12 oder [21], Ch. V, Theorem 4. 4.

Wir zeigen noch

Lemma 2. Sei $\varphi(\sigma) = \log g(\sigma)$. Aus

$$-\sigma\varphi'(\sigma) \cdot \{\varphi(\sigma)\}^{-1} = o(1) \text{ für } \sigma \rightarrow 0$$

folgt

$$u\sigma_u = o(\varphi(\sigma_u)).$$

Der Beweis ist wegen $u = -\varphi'(\sigma_u)$ trivial.

Seien nun die Voraussetzungen (5. II), a), b) erfüllt. Nach Satz 2 ist mit $\varphi(\sigma) = \log g(\sigma)$

$$\log P(u) \sim \varphi(\sigma_u) + u\sigma_u.$$

Wir wissen aus § 4, daß $-\varphi' \nearrow$ und daß $0 < \varphi''(\sigma) \leq \gamma_3 \sigma^{-1} |\varphi'(\sigma)|$ (vgl. (4. 10)) gilt. Somit ist Lemma 1 auf φ und Φ anwendbar und liefert für $\sigma \rightarrow 0$

$$(5. 8) \quad \Phi'(\sigma) = \varphi'(\sigma) + o(\sigma^{-1}\varphi(\sigma)) + o(\varphi'(\sigma)).$$

Gilt nun noch (5. II c'), so ist für $\sigma \rightarrow 0$

$$\frac{-\sigma\varphi'(\sigma)}{\varphi(\sigma)} \sim \frac{-\sigma\Phi'(\sigma)}{\Phi(\sigma)} + o(1) \geq \gamma_5^* > 0,$$

also gibt der Zusatz zu Lemma 1 die asymptotische Beziehung $\varphi'(\sigma) \sim \Phi'(\sigma)$, und aus (5. 5) folgt die Behauptung.

Gilt (5. II c''), so ist nach (5. 8)

$$-\sigma\varphi'(\sigma) (\varphi(\sigma))^{-1} \sim -\sigma\Phi'(\sigma) (\Phi(\sigma))^{-1} + o(1) = o(1),$$

und nach Lemma 2 ist für $u \rightarrow \infty$

$$u\sigma_u = o(\varphi(\sigma_u)), \quad ut_u = o(\Phi(t_u)).$$

Um (5. 3') zu zeigen, braucht nur noch $\varphi(\sigma_u) \sim \Phi(t_u)$ nachgewiesen zu werden.

Ist nun $\sigma_u \leq t_u$, so ist

$$|\varphi(\sigma_u) - \varphi(t_u)| = \left| \int_{\sigma_u}^{t_u} (-\varphi'(\sigma)) d\sigma \right| \leq (t_u - \sigma_u) |\varphi'(\sigma_u)| \leq ut_u + u\sigma_u,$$

also ist

$$(5. 9) \quad |\varphi(\sigma_u) - \Phi(t_u)| \leq |\varphi(\sigma_u) - \varphi(t_u)| + |\varphi(t_u) - \Phi(t_u)| \\ \leq ut_u + u\sigma_u + o(\Phi(t_u)) = o(\varphi(\sigma_u) + \Phi(t_u)).$$

Ist $\sigma_u > t_u$, so geht man von $\Phi(\sigma_u) - \Phi(t_u)$ aus, schließt genau wie eben und erhält ebenfalls (5. 9). Somit ist

$$\varphi(\sigma_u) \sim \Phi(t_u),$$

und (5. 3') ist gezeigt.

(5. III) ist ein Spezialfall von (5. II). Aus der Darstellung

$$(1. 7) \quad \varphi(\sigma) = \sigma \int_{\lambda_1}^{\infty} N(u) \cdot (e^{u\sigma} - 1)^{-1} du$$

folgt nämlich mit (5. III b) durch eine einfache Rechnung, daß φ von langsamem Wachstum ist, also gilt nach Bemerkung 3 zu Satz 3 die Beziehung (5. II c''); (5. II b) ist wegen (5. III b) offensichtlich erfüllt.

6. Herleitung eines Ergebnisses von Kohlbecker

Kohlbecker [9] ging von der Voraussetzung

$$(6.1) \quad N(u) \sim u^\alpha \cdot L(u), \quad \alpha > 0,$$

aus, wobei L eine Funktion von langsamem Wachstum ist. Nach Satz 2, dessen Voraussetzungen erfüllt sind, ist dann

$$(6.2) \quad \log P(u) \sim \varphi(\sigma_u) + u \sigma_u;$$

der rechtsstehende Ausdruck soll jetzt auf die in (6.1) gegebenen „Daten“ (α und L) umgerechnet werden¹²).

Nach [9], Theorem 1 (S. 350) ist für $\sigma \rightarrow 0$

$$(6.3) \quad \varphi(\sigma) \sim A \sigma^{-\alpha} L(\sigma^{-1})$$

mit der Konstanten $A = \Gamma(\alpha + 1) \cdot \zeta(\alpha + 1)$.

Lemma 3. Aus (6.1) folgt die Beziehung

$$(6.4) \quad \varphi'(\sigma) \sim -A\alpha \cdot \sigma^{-(\alpha+1)} \cdot L(\sigma^{-1}) \text{ für } \sigma \rightarrow 0.$$

Beweis. (6.4) könnte mit Hilfe von (4.5') ähnlich dem Beweise von Theorem 1 aus [9] gezeigt werden. Einfacher ist es, (6.4) aus (6.3) durch „Differentiation“ abzuleiten.

Sei $\varepsilon > 0$ gegeben. Wähle $\delta \leq \varepsilon \cdot (3C_2)^{-1}$ (vgl. (4.2)) so, daß

$$(6.5) \quad (1 + \delta)^{-\alpha} - 1 \geq -\alpha\delta(1 + \varepsilon)$$

ist. Wegen (6.3) und wegen des langsamen Wachstums von L ist für genügend kleines $\sigma \leq \sigma_0(\varepsilon, \delta)$

$$\begin{aligned} \varphi(\sigma + \delta\sigma) &\geq A(1 + \delta)^{-\alpha} \sigma^{-\alpha} \cdot L(\sigma^{-1}(1 + \delta)^{-1}) - \varepsilon \delta \sigma^{-\alpha} \cdot 2L(\sigma^{-1}) \\ &\geq A(1 + \delta)^{-\alpha} \sigma^{-\alpha} \cdot L(\sigma^{-1}) - 3\varepsilon \delta \sigma^{-\alpha} L(\sigma^{-1}) \end{aligned}$$

und

$$-\varphi(\sigma) \geq -A \sigma^{-\alpha} L(\sigma^{-1}) - \varepsilon \delta \sigma^{-\alpha} L(\sigma^{-1}).$$

Aufgrund der Taylorschen Formel ist mit einer Stelle σ^* zwischen σ und $\sigma + \delta\sigma$

$$\varphi'(\sigma) = \frac{1}{\delta\sigma} \cdot (\varphi(\sigma + \delta\sigma) - \varphi(\sigma)) - \frac{1}{2} \delta\sigma \varphi''(\sigma^*),$$

woraus wegen der Monotonie von φ'' , wegen (4.10), obigen Abschätzungen und (6.5)

$$\varphi'(\sigma) \geq A \sigma^{-(\alpha+1)} L(\sigma^{-1}) \{(1 + \delta)^{-\alpha} - 1\} \cdot \delta^{-1} - 4\varepsilon \sigma^{-(\alpha+1)} L(\sigma^{-1}) - \frac{1}{2} \delta \cdot 6C_2 \cdot |\varphi'(\sigma)|,$$

also

$$\varphi'(\sigma) \cdot (1 - \varepsilon) \geq -A\alpha \sigma^{-(\alpha+1)} L(\sigma^{-1}) - \varepsilon \cdot (4 + A\alpha) \cdot \sigma^{-(\alpha+1)} L(\sigma^{-1})$$

folgt. Da eine entsprechende Ungleichung in der anderen Richtung gilt, ist (6.4) bewiesen.

Nun ist nach de Bruijn (vgl. [9], S. 349) für $u \rightarrow \infty$

$$(6.6) \quad L(u) \sim J(u) := K \exp \left\{ \int_1^u \eta(t) t^{-1} dt \right\},$$

¹² Satz 3 kann hierzu nicht benützt werden, weil über L keine Differenzierbarkeitsvoraussetzungen gemacht sind.

wobei η stetig ist und $\lim_{t \rightarrow \infty} \eta(t) = 0$ gilt. $J(u)$ ist differenzierbar, und

$$(6.7) \quad A\alpha \cdot t_u^{-(\alpha+1)} \cdot J(t_u^{-1}) = u$$

ist eindeutig lösbar für große u mit der Lösung t_u .

Lemma 4. Wenn $-\varphi'(\sigma_u) = u$ ist und t_u durch (6.7) erklärt ist, so gilt für $u \rightarrow \infty$

$$(6.8) \quad \sigma_u \sim t_u.$$

Beweis. Man vgl. die Herleitung von (5.4).

Aus (6.2), (6.3) und (6.8) folgt nun

$$(6.9) \quad \log P(u) \sim t_u^{-\alpha} L(t_u^{-1}) \cdot A \cdot (\alpha + 1).$$

Benützt man noch [9], Lemma 3 (S. 356), welches für genügend große u die Existenz eines asymptotisch eindeutig bestimmten $\omega_u > 0$ aussagt, das

$$(6.10) \quad u = A\alpha\omega_u^{-(\alpha+1)}L(\omega_u^{-1})$$

und

$$\omega_u \sim t_u$$

erfüllt, so erhält man schließlich aus (6.8) ein Ergebnis von Kohlbecker [9]:

Satz 4. Wenn für $u \rightarrow \infty$

$$N(u) \sim u^\alpha L(u), \quad \alpha > 0,$$

gilt, dann gilt mit der Konstanten $A = \Gamma(\alpha + 1) \cdot \zeta(\alpha + 1)$ für $u \rightarrow \infty$

$$\log P(u) \sim \omega_u^{-\alpha} L(\omega_u^{-1}) \cdot A \cdot (\alpha + 1),$$

wobei ω_u eine Lösung von (6.10) ist.

Literaturverzeichnis

- [1] V. Avakumović und J. Karamata, Über einige Tauber'sche Sätze, deren Asymptotik von Exponentialcharakter ist. Math. Z. **41** (1936), 345—356.
- [2] N. Brigham, A general asymptotic formula for partition functions. Proc. Amer. Math. Soc. **1** (1950), 182—191.
- [3] N. G. de Bruijn, On Mahler's partition problem. Indag. Math. **10** (1948), 210—220.
- [4] P. Erdős, On an elementary proof of some asymptotic formulas in the theory of partitions. Annals of Math. (2) **43** (1942), 437—450.
- [5] P. Erdős and L. Mirsky, The distribution of the values of the divisor function $d(n)$. Proc. London Math. Soc. (3) **2** (1952), 257—271.
- [6] G. H. Hardy, Divergent Series, Oxford 1949.
- [7] G. H. Hardy and S. Ramanujan, Asymptotic formulae for the distribution of integers of various types. Proc. London Math. Soc. (2) **16** (1917), 112—132.
- [8] A. E. Ingham, A Tauberian theorem for partitions. Annals of Math. **42** (1941), 1075—1090.
- [9] E. E. Kohlbecker, Weak asymptotic properties of partitions. Transactions Amer. Math. Soc. **88** (1958), 346—365.
- [10] K. Mahler, On a special functional equation. J. London Math. Soc. **15** (1940), 115—123.
- [11] W. T. Martin and N. Wiener, Taylor's series of functions of smooth growth in the unit circle. Duke Math. J. **4** (1938), 384—392.

- [12] *H. H. Ostmann*, Additive Zahlentheorie I. Berlin 1956.
- [13] *S. Parameswaran*, Partition functions, whose logarithms are slowly oscillating. Transactions Amer. Math. Soc. **100** (1961), 217—240.
- [14] *W. B. Pennington*, On Mahler's partition problem. Annals of Math. (2) **57** (1953), 531—546.
- [15] *S. Ramanujan*, Highly composite numbers. Proc. London Math. Soc. (2) **14** (1915), 347—409.
- [16] *K. F. Roth* and *G. Szekeres*, Some asymptotic formulae in the theory of partitions. Quart. J. Math. Oxford (2) **5** (1954), 241—259.
- [17 a, b, c] *W. Schwarz*, Einige Anwendungen Tauberscher Sätze in der Zahlentheorie. A, B. J. reine angew. Math. **219** (1965), 67—96, 157—179. C. Mahler's Partitionsproblem, J. reine angew. Math. **228** (1967), 182—188.
- [18] *E. Wagner*, Tauber'sche Sätze reeller Art für die Laplace-Transformation. Math. Nachr. **31** (1966), 153—168.
- [19] *E. Wagner*, Tauber'sche Sätze für die Mellin-Transformation. Wiss. Z. Univ. Halle **15** (1966), 617—624.
- [20] *E. Wagner*, Ein reeller Taubersatz für die Laplace-Transformation. Math. Nachr. (im Druck).
- [21] *D. V. Widder*, The Laplace Transform, Princeton 1946.

Mathematisches Institut der Universität, 78 Freiburg

Eingegangen 13. März 1967

Relativnullen in additiven abelschen Gruppen¹⁾

Von *Dieter Szimtenings* in Mainz

Herrn Professor Dr. H. Rohrbach zum 65. Geburtstag

§ 1. Einleitung

Existiert zu einer nicht leeren Teilmenge A der Menge Z der nicht negativen ganzen Zahlen eine Menge $N(A) \subseteq Z$ mit mindestens einem von Null verschiedenen Element derart, daß $A + N(A) = A$ ist, so wird $N(A)$ nach H.-H. Ostmann ([2], [3]) eine Relativnull bezüglich A genannt. Es liegt nahe, diesen Begriff auf Mengen rationaler Zahlen zu übertragen und analoge Ergebnisse zu denen, die im Ostmannschen Bericht [2], S. 22–24 dargestellt sind, abzuleiten.

Dabei hat man es in der Hand, entweder von der Menge aller rationalen Zahlen $R(-\infty, +\infty)$ als Obermenge bzw. von solchen der Art $R(r, \infty)$, r reell, auszugehen oder sich auf Teilmengen der Menge der rationalen Zahlen zwischen 0 (einschließlich) und 1 (ausschließlich), $R[0, 1)$, bzw. allgemeiner $R[a, a+r)$, a reell, r ganz, zu beschränken. Die erstgenannte Menge ist bezüglich der Addition abgeschlossen; um dies auch für die zweite Menge $R[0, 1)$ zu erreichen, wird als Verknüpfung die Addition modulo 1 zugrunde gelegt.

Im Falle der Menge $R[0, 1)$ kann man die meisten Betrachtungen über Relativnullen gleich auf eine umfassende Klasse von Obermengen verallgemeinern. Denn bekanntlich ist die Menge $R[0, 1)$ bezüglich der Addition modulo 1 eine abelsche Gruppe; jedes der abzählbar vielen Elemente dieser Gruppe hat eine endliche Ordnung, d. h. $R[0, 1)$ ist im Sinne von M. Hall [1] eine periodische Gruppe. Da in der Hauptsache die eben genannten Eigenschaften von $R[0, 1)$ ausgenutzt werden, können die Definitionen und Sätze gleich auf beliebige periodische abelsche Gruppen G_0 übertragen werden.

Die Menge $R(-\infty, +\infty)$ ist ebenfalls eine abelsche Gruppe, wobei aber hier jedes der abzählbar vielen Elemente eine unendliche Ordnung hat; es handelt sich also um eine aperiodische Gruppe (vgl. M. Hall [1]). Statt dieser Obermenge können weiter beliebige aperiodische abelsche Gruppen G_∞ betrachtet und schließlich sogar die Unterscheidung zwischen periodischen und aperiodischen Gruppen fallen gelassen werden. Im folgenden werden daher in § 2 und § 3 die Begriffe „Summenmenge“ und „Relativnull“ für die Teilmengen beliebiger abelscher Gruppen G definiert²⁾ und dann hierfür die Ergebnisse über Mengen mit Relativnullen von H.-H. Ostmann soweit wie möglich auf Teilmengen von G übertragen.

¹⁾ D 65, Teil I.

²⁾ Bezeichnet man mit G_e die Menge aller Elemente aus G mit endlicher Ordnung, so ist G_e selbst eine periodische abelsche Gruppe, d. h. vom Typ G_0 .

Vorbereitend auf die eigentliche Problemstellung werden im § 2 Summenmengen in G untersucht, um dann anschließend Mengen mit Relativnullen in G zu betrachten. Besondere Beachtung verdient dabei die Frage: „Wie muß eine Menge $A \subseteq G$ beschaffen sein, damit für sie Relativnullen $N(A)$ existieren?“ Die Antwort darauf gibt der Satz 2 in § 3. In den Sätzen 3 und 4 sind einige bemerkenswerte Aussagen über Relativnullen zusammengestellt. Der § 4 handelt im ersten Teil von endlichen Mengen mit Relativnullen; der Satz 5b) ist in diesem Zusammenhang die bedeutendste Aussage. Im zweiten Teil des § 4 wird die Möglichkeit zur Gewinnung von Relativnullen mit Hilfe von Produktmengen (s. Definition 3) erörtert. Anschließend kommt das Problem zur Sprache: „Gibt es zu jeder gegebenen Menge $A \subseteq G$ mit Relativnullen eine maximale Relativnull $N^+(A)$?“ Mit dieser Fragestellung befaßt sich der § 5, wo auch die Struktur der maximalen Relativnull (vgl. Satz 7) angegeben wird. Der § 6 bringt eine Ausdehnung der zuvor erhaltenen Ergebnisse auf abelsche Halbgruppen. Die §§ 7 und 8 zeigen, wie sich die erhaltenen Ergebnisse in den Spezialfällen G_0 und G_∞ erweitern und z. T. verschärfen lassen.

In den Paragraphen 2 bis 5 sei mit G eine beliebige abelsche Gruppe mit additiver Verknüpfung bezeichnet.

§ 2. Summenmengen in G

Definition 1. Sind A und B zwei nicht leere Teilmengen (Komplexe) von G , dann heiße die Menge

$$C = A + B = \{a + b \mid a \in A, b \in B\}$$

die *Summenmenge* von A und B (vgl. [2], S. 1). A und B heißen *Summanden* von C . Als Summenmenge einer Menge $A \subseteq G$ und der leeren Menge $\emptyset$ wird die leere Menge $\emptyset$ definiert: $A + \emptyset = \emptyset$.

Offenbar ist die Summenmenge $A + B$ wieder eine Teilmenge der Gruppe G .

Satz 1. Seien A, B, C und A_i Teilmengen von G , $i \in I$, I eine endliche oder unendliche Indexmenge, dann gilt:

- a) $A + B = B + A$ (Kommutativgesetz).
- b) $A + (B + C) = (A + B) + C$ (Assoziativgesetz).
- c) Aus $A \subseteq B$ folgt $A + C \subseteq B + C$ für alle $C \subseteq G$.
- d) Ist $0 \in \bigcap_{i=1}^n A_i$, so gilt $A_i \subseteq A_1 + A_2 + \cdots + A_n$ für alle $i \in \{1, \dots, n\}$.
- e) $\bigcup_{i \in I} A_i + B = \bigcup_{i \in I} (A_i + B)$.
- f) $\bigcap_{i \in I} A_i + B \subseteq \bigcap_{i \in I} (A_i + B)$.
- g) Sind A und B Unter(halb)gruppen in G , dann ist auch $A + B$ Unter(halb)gruppe in G .

Beweis. a) und b) sind trivial, da G eine abelsche Gruppe ist.

c) Sei $d \in A + C$, dann ist d darstellbar in der Form $d = a + c$, $a \in A$, $c \in C$. Da nach Voraussetzung $a \in B$ ist, folgt $d = a + c \in B + C$.

d) Durch mehrfache Anwendung von c) folgt die Behauptung, da $\{0\} \subseteq A_i$ für $i = 1, \dots, n$ gilt.

e) $r \in \bigcup_{i \in I} A_i + B$, dann existiert (nach Definition der Summenmenge) mindestens eine Darstellung $r = a + b$ mit $b \in B$ und $a \in \bigcup_{i \in I} A_i$. Daraus folgt: Es gibt ein $t \in I$ so, daß $a \in A_t$ ist, und hieraus $r = (a + b) \in (A_t + B)$. Somit ist $r \in \bigcup_{i \in I} (A_i + B)$. Ist andererseits $s \in \bigcup_{i \in I} (A_i + B)$, dann gibt es ein $t \in I$ so, daß $s \in A_t + B$ ist. Daher läßt sich s in der Form $s = a_t + b$ mit $a_t \in A_t$ und $b \in B$ darstellen. Es folgt: $a_t \in \bigcup_{i \in I} A_i$ und damit $s = a_t + b \in \bigcup_{i \in I} A_i + B$.

f) Sei $r \in \bigcap_{i \in I} A_i + B$ beliebig. Dann muß r von der Gestalt $r = a + b$ sein mit $a \in \bigcap_{i \in I} A_i$ und $b \in B$, und es gilt $a \in A_i$ für alle $i \in I$. Somit folgt $r = a + b \in A_i + B$ für alle $i \in I$. Also ist $r \in \bigcap_{i \in I} (A_i + B)$.

g) Mit $a_1, a_2 \in A$ und $b_1, b_2 \in B$ folgt (wegen der Assoziativität und Kommutativität in G):

$$(a_1 + b_1) + (a_2 + b_2) = (a_1 + a_2) + (b_1 + b_2) \in A + B,$$

also ist $A + B$ Unterhalbgruppe. Falls A und B Untergruppen sind, ist $0 \in A \cap B$, so daß $0 \in A + B$ erfüllt ist. Da mit $a \in A$ und $b \in B$ auch $-a \in A$, $-b \in B$ gilt, folgt $-(a + b) = (-a) + (-b) \in A + B$.

Bemerkung 1. Die Umkehrung zu c) gilt im allgemeinen nicht, wie das Beispiel $A = \{1\}$, $B = \{0\}$, $C = \{0, 1, 2, \dots\}$ zeigt. Ebenso ist in der Eigenschaft f) das Gleichheitszeichen i. a. falsch; man wähle etwa $A_1 = \{0\}$, $A_2 = \{1\}$ und $B = \{0, 1\}$ und beachte Definition 1.

Bemerkung 2. Für Summenmengen aus n gleichen Summanden A wird nA geschrieben.

§ 3. Relativnullen in G

Definition 2. Existiert zu einer nicht leeren Menge $A \subseteq G$ eine Menge $N(A) \subseteq G$ mit mind)stens einem von Null verschiedenen Element derart, daß die Beziehung $A + N(A) = A$ erfüllt ist, dann wird $N(A)$ eine *Relativnull* bezüglich der Menge A genannt. A selbst heißt eine Menge mit Relativnullen.

Aufgrund dieser Definition muß eine Menge mit Relativnullen mindestens ein von Null verschiedenes Element enthalten. Als Beispiele für Mengen mit Relativnullen seien genannt:

Beispiel 1. Jede Unterhalbgruppe $U \subseteq G$ mit $0 \in U$ ist zu sich selbst Relativnull: $N(U) = U$. Denn für Unterhalbgruppen gilt allgemein $U + U \subseteq U$ und hier außerdem (mit Satz 1 c)) $U + U \supseteq U + 0 = U$. Also ist tatsächlich $U + U = U$.

Beispiel 2. Die dyadischen Brüche aus $R(-\infty, +\infty)$ bilden bezüglich der Addition eine abelsche Gruppe

$$G = \left\{ 0, \frac{n}{2^r} \mid n = \pm 1, \pm 2, \pm 3, \dots; \quad r = 1, 2, 3, \dots \right\}.$$

Die Teilmenge der positiven dyadischen Brüche

$$A = \left\{ \frac{n}{2^r} \mid n = 1, 2, 3, \dots; \quad r = 1, 2, 3, \dots \right\}$$

besitzt u. a. die Relativnullen

$$N_j(A) = \left\{ 0, \frac{n}{2^r} \mid n = 1, 3, 5, \dots; \quad 1 \leq r \leq j \right\}, \quad j = 1, 2, 3, \dots$$

und

$$N_k(A) = \left\{ 0, \frac{n}{2^r} \mid n = 1, 2, \dots; \quad 1 \leq r \leq k \right\}, \quad k = 1, 2, 3, \dots$$

Satz 2. a) Eine Menge $A \subseteq G$ besitzt dann und nur dann Relativnullen, wenn es eine eigentliche Unterhalbgruppe U (d. h. $U \neq \{0\}$) von G mit $0 \in U$ gibt, derart, daß

$$(1) \quad \{a\} + U \subseteq A \text{ gilt für alle } a \in A.$$

b) Notwendig und hinreichend dafür, daß eine Menge $A \subseteq G$ Relativnullen besitzt, ist die Existenz einer eigentlichen Unterhalbgruppe U mit $0 \in U$, so daß A eine Zerlegung

$$(2) \quad A = U + K, \quad K \subseteq A$$

mit U als Summand zuläßt. Diese Unterhalbgruppe U ist selbst Relativnull bezüglich A .

Beweis. a) $N(A)$ sei eine Relativnull bezüglich A ; dann gibt es ein $n \in N(A)$ mit $n \neq 0$. Für jedes $a \in A$ müssen auch die Elemente $a + rn$, $r \in \mathbb{Z}$, $\mathbb{Z} = \{0, 1, 2, \dots\}$ zu A gehören, d. h. es gilt $\{a\} + (n \times \mathbb{Z}) \subseteq A$. Die Menge $n \times \mathbb{Z} = \{0, n, 2n, \dots\}$ ist also eine Unterhalbgruppe U von G mit $0 \in U$, die (1) erfüllt. Es existiere nun eine solche Unterhalbgruppe $U \subseteq G$, derart, daß $\{a\} + U \subseteq A$, $0 \in U \neq \{0\}$ für alle $a \in A$ gilt. Aus $0 \in U$ folgert man $A + U \supseteq A + \{0\} = A$. Zum andern gilt mit Satz 1e) und (1)

$$A + U = \bigcup_{a \in A} \{a\} + U = \bigcup_{a \in A} (\{a\} + U) \subseteq A.$$

Also ist $U + A = A$, d. h. die Menge A besitzt die Relativnull $U \neq \{0\}$.

b) Zum Beweis braucht man nur zu zeigen, daß die Eigenschaft (2) äquivalent ist mit der Eigenschaft (1). Mit (2) und Beispiel 1 folgt

$$U + A = U + U + K = U + K = A,$$

womit (1) nachgewiesen ist.

Wie im Beweis von Teil a) gezeigt wurde, folgt $A + U = A$ aus der Eigenschaft (1), d. h. (2) ist mit $K = A$ erfüllt.

Bemerkung 3. Wird in Satz 2b) die Menge K „minimal“ in dem Sinne gewählt, daß $(K - \{r\}) + U \not\subseteq A$ für alle $r \in K$ folgt, dann heißt die Menge $K = K(U, A)$ ein U -Kern von A (vgl. [2], S. 22, wo die entsprechenden Mengen als o_A -Kern bezeichnet werden). $K(U, A)$ ist im allgemeinen nicht eindeutig bestimmt; in $R[0, 1)$ kann der U -Kern von A in eindeutiger Weise festgelegt werden (s. Satz 9).

Satz 3. Sei $A \subseteq G$ eine Menge mit Relativnullen $N_i(A)$, $i \in I$, I eine Indexmenge und sei B eine weitere Menge mit Relativnullen, dann gilt:

- a) Die Summenmenge von endlich vielen Relativnullen $N_i(A)$, $i = 1, 2, \dots, r$ bezüglich A ist wieder eine Relativnull bezüglich A .
- b) Die Vereinigung von Relativnullen bezüglich A ist auch eine Relativnull bezüglich A .
- c) $N(A) \cap N(B)$ ist Relativnull bezüglich $A \cup B$, falls $\{0\} \not\subseteq N(A) \cap N(B)$ ist.
- d) $N(A) \cup N(B)$ ist Relativnull bezüglich $A \cap B$, falls $\{0\} \not\subseteq N(A) \cup N(B)$ ist.
- e) Mit A besitzt auch die Menge $A^- = \{-a \mid a \in A\}$ Relativnullen.

^{a)} Ist $B \subseteq C$, dann werde die „Restmenge“ erklärt als $C - B = \{d \mid d \in C, d \notin B\}$.

Beweis. a) Das Assoziativgesetz (Satz 1b)) ergibt unmittelbar:

$$N_1(A) + \cdots + N_r(A) + A = N_1(A) + \cdots + N_{r-1}(A) + A = \cdots = N_1(A) + A = A.$$

b) Mit Satz 1e) findet man: $\bigcup_{i \in I} N_i(A) + A = \bigcup_{i \in I} (N_i(A) + A) = A.$

c) Durch Anwendung der Sätze 1e) und 1f) und Ausnutzung der Voraussetzung über die Relativnullen $N(A)$ und $N(B)$ findet man:

$$\begin{aligned} (A \cup B) + (N(A) \cap N(B)) &= (A + (N(A) \cap N(B))) \cup (B + (N(B) \cap N(A))) \\ &\subseteq (A \cap (A + N(B))) \cup (B \cap (B + N(A))) \subseteq A \cup B, \end{aligned}$$

beziehungsweise

$$(A \cup B) + (N(A) \cap N(B)) \supseteq (A \cup B) + \{0\} = A \cup B.$$

d) Ebenso wie im Teil c) schließt man hier:

$$\begin{aligned} (A \cap B) + (N(A) \cup N(B)) &= ((A \cap B) + N(A)) \cup ((A \cap B) + N(B)) \\ &\subseteq (A \cap (B + N(A))) \cup (B \cap (A + N(B))) \subseteq A \cup B, \end{aligned}$$

beziehungsweise

$$(A \cap B) + (N(A) \cup N(B)) \supseteq (A \cap B) + \{0\} = A \cap B.$$

e) Da G eine Gruppe ist, ist z. B. die Menge $N = \{-n \mid n \in N(A)\}$ eine Relativnull bezüglich A^- , wenn $N(A)$ Relativnull bezüglich A ist.

Folgerung 1. Die kleinste Unterhalbgruppe M , die alle gegebenen $N_i(A)$, $i \in I$, enthält, ist auch Relativnull bezüglich A .

Beweis. Sei $(i_1, \dots, i_r)$, $r \geq 1$, ein r -tupel mit $i_j \in I$; dann ist $N_{i_1}(A) + \cdots + N_{i_r}(A)$ nach Satz 3a) eine Relativnull bezüglich A . Bildet man für alle endlichen r -tupel $(i_1, \dots, i_r)$, $i_j \in I$, die zugehörigen $N_{i_1}(A) + \cdots + N_{i_r}(A)$, so ist

$$M = \bigcup_{\substack{(i_1, \dots, i_r) \\ r=1,2,\dots; i_j \in I}} (N_{i_1}(A) + N_{i_2}(A) + \cdots + N_{i_r}(A))$$

nach Satz 3b) Relativnull bezüglich A und nach Satz 3a) Halbgruppe, die nach Konstruktion die verlangte Minimaleigenschaft hat.

Satz 4. Sei $A \subseteq G$ eine Menge mit Relativnullen. Dann gilt:

a) Diejenigen Elemente $n \in G$, für die $\{n\} + A = A$ erfüllt ist, bilden eine Untergruppe $V(A)$ von G mit $V(A) + A = A$.

b) Diejenigen Elemente $m \in G$, für die $\{m\} + A \subseteq A$ ist, bilden eine Unterhalbgruppe $W(A)$ von G , und für alle $m \in W(A)$ ist $\{m\} + V(A) \subseteq W(A)$, d. h. $W(A)$ besteht aus vollen Restklassen von G modulo $V(A)$.

c) Sei $N(A)$ eine Relativnull bezüglich A und sei $B \neq \{0\}$ eine Teilmenge von $N(A)$. Existiert dann ein Element $b \in B \cap V(A)$, so ist B eine Relativnull bezüglich A .

Beweis. a) Seien $n, n' \in V(A)$, d. h. es ist $\{n\} + A = A$ und $\{n'\} + A = A$. Daraus folgert man: $\{n + n'\} + A = \{n\} + \{n'\} + A = \{n\} + A = A$. Also ist $n + n' \in V(A)$. Sodann gilt $\{0\} + A = A$. Also ist $0 \in V(A)$. Zum Beweis von $\{-n\} + A = A$ überlegt man: Sei $a \in A$, dann ist $a + n = d \in A$. Damit gilt $a = (-n) + d \in A$ und $a \in \{-n\} + A$. Sei umgekehrt $b \in \{-n\} + A$; daraus ergibt sich (mit Satz 1a) und b)):

$$b + n \in \{-n\} + A + \{n\} = A = A + \{n\},$$

somit $b \in A$. Also ist $-n \in V(A)$.

b) Aus $\{m_i\} + A \subseteq A$, $i = 1, 2$, folgt

$$\{m_1 + m_2\} + A = \{m_1\} + \{m_2\} + A \subseteq \{m_1\} + A \subseteq A.$$

Geht man von der Beziehung $\{m\} + V(A) + A = \{m\} + A \subseteq A$ aus, dann findet man weiter $\{m\} + V(A) \subseteq W(A)$. Ferner ist $V(A) \cap W(A) = \emptyset$, also ist $\{m\} + V(A)$ Restklasse von G modulo $V(A)$.

c) Wegen $b \in V(A)$ ist $\{b\} + A = A$, also $A = \{b\} + A \subseteq B + A \subseteq N(A) + A = A$. Damit ist c) und folglich Satz 4 vollständig bewiesen.

Als Ergänzung zu Satz 4 a) und b) seien zwei Beispiele genannt.

Beispiel 3. Zur Menge $A_1 = \left\{ r + \frac{1}{2} + s\sqrt{2} \mid r \text{ ganz; } s \geq 1, \text{ ganz} \right\}$ gehören $V(A_1) = \{r \mid r \text{ ganz}\}$ und $W(A_1) = \{n + s\sqrt{2} \mid n \in V(A_1); s \geq 1, \text{ ganz}\}$.

Beispiel 4. Für $A_2 = \left\{ r + \frac{1}{2} + s\sqrt{2} \mid r \geq 0, \text{ ganz; } s \geq 1, \text{ ganz} \right\}$ gilt $V(A_2) = \{0\}$ und $W(A_2) = \{r + s\sqrt{2} \mid r \geq 0, \text{ ganz; } s \geq 0, \text{ ganz; } s + r \geq 1\}$.

Folgerung 2. Für jede Teilmenge $T \subseteq N(A)$ mit mindestens einem von Null verschiedenen Element n gilt: $T \cup \{0\}$ ist ebenfalls eine Relativnull bezüglich A . Für $T = \{n\}$ ergibt sich insbesondere $\{0, n\} + A = A$.

Beweis. Da nach Satz 4a) $0 \in V(A)$ ist, folgt aus $(T \cup \{0\}) \cap V(A) \cong \{0\}$ mit Satz 4c) die Behauptung.

Folgerung 3. Enthält der Durchschnitt ($\neq \{0\}$) von Relativnullen $N_i(A)$, $i \in I$, I Indexmenge, mindestens ein Element aus $V(A)$ (oder spezieller die Null und mindestens ein von Null verschiedenes Element), dann ist auch dieser Durchschnitt eine Relativnull bezüglich A .

Beweis. Aus der für alle $j \in I$ gültigen Beziehung $\bigcap_{i \in I} N_i(A) \subseteq N_j(A)$ ergibt sich mit Satz 4c) (bzw. mit der Folgerung 2) die behauptete Eigenschaft.

§ 4. Endliche Mengen und Produktmengen mit Relativnullen

Satz 5. a) Für endliche Mengen $A \subseteq G$ mit Relativnullen $N(A)$ gilt: Aus $n \in N(A)$ folgt $A + \{n\} = A$ (Ergänzung zu Folgerung 2).

b) Eine r -elementige Menge $A \subseteq G$ besitzt genau dann Relativnullen, wenn in G ein Element $n \neq 0$ mit $|n| \mid r$ existiert und wenn A in $\frac{r}{|n|}$ Teilmengen der Gestalt

$$A_i = \{a_i, a_i + n, \dots, a_i + (|n| - 1)n\} \text{ mit } a_i \in A$$

zerlegt werden kann. Dabei sei $|n|$ die Ordnung von n .

Beweis. a) Selbstverständlich wird von den Elementen der Menge $A = \{a_1, a_2, \dots, a_r\}$ vorausgesetzt, daß $a_i \neq a_j$ für $i \neq j$, $i, j \in \{1, 2, \dots, r\}$ gilt. Wären nicht sämtliche r Elemente $a_1 + n, a_2 + n, \dots, a_r + n$ voneinander verschieden, so würde daraus die Gleichheit zweier a_i, a_j mit $i \neq j$ folgen. Da A nur r Elemente enthält, stellt die Gesamtheit der $a_i + n$, $i = 1, 2, \dots, r$ schon die Menge A dar. Übrigens erkennt man unmittelbar $|n| \leq r$.

b) Nach Teil a) kann man sich darauf beschränken, Elemente $n \in G$ mit $n \neq 0$ und $A + \{n\} = A$ zu betrachten. Durch die Bildung der Summen $a_i + n$ entstehen alle Elemente aus A , und es ist nicht möglich, daß $a_i + n = a_i$ für irgendein $i \in \{1, \dots, r\}$ wird, da sofort $n = 0$ folgen würde. Daher muß ein j_1 mit $1 < j_1 \leq r$ so existieren, daß $a_1 + n = a_2, a_2 + n = a_3, \dots, a_j + n = a_1$ gilt (evtl. erst durch Umnúmerieren der a_i erreichbar!). Daraus ergeben sich für die Menge $A_1 = \{a_1, a_2, \dots, a_{j_1}\} \subseteq A$ folgende Eigenschaften:

$$(3) \quad |n| = j_1 \geq 2,$$

$$(4) \quad |a_{i+1} - a_i| = j_1 \text{ für } i = 1, 2, \dots, (j_1 - 1) \text{ und } |a_1 - a_{j_1}| = j_1$$

$$(5) \quad a_i = a_1 + (i - 1)n \text{ für } i = 1, \dots, j_1; \text{ also}$$

$$A_1 = \{a_1, a_1 + n, \dots, a_1 + (|n| - 1)n\},$$

(6) A_1 hat die maximale Relativnull

$$N^+(A_1) = \{0, n, \dots, (|n| - 1)n\}.$$

Ist $j_1 < r$, dann gibt es ein j_2 mit $1 < j_2 \leq r - j_1$, derart, daß

$$a_{j_1+1} + n = a_{j_1+2}, \dots, a_{j_1+j_2} + n = a_{j_1}$$

erfüllt ist, und wieder gelten die Eigenschaften (3) bis (6) für j_2 statt j_1 . (Dabei denke man sich a_1 durch a_{j_1+1} und A_1 durch A_2 ersetzt.) Es ist $N^+(A_1) = N^+(A_2)$; also enthalten A_1 und A_2 gleich viele Elemente.

Ist auch noch $j_1 + j_2 < r$, dann existiert ein j_3 mit $1 < j_3 \leq r - (j_1 + j_2)$ und genauso wie für j_1 und j_2 gelten die obigen Eigenschaften (3) bis (6) jetzt für $j_3, \dots$ usw. Das Verfahren bricht (spätestens nach r Schritten) ab. Da jedes $a_1 \in A$ bei diesem Prozeß erfaßt werden muß, wird allein durch die Annahme eines Elementes $n \in N(A)$, $n \neq 0$ die Menge A in $|n|/r$ Teilmengen von A der Gestalt

$$A_t = \{a_t, a_t + n, \dots, a_t + (|n| - 1)n\}, \quad t = 1, 2, \dots, \frac{r}{|n|}$$

zerlegt, und jede dieser Teilmengen besitzt (unabhängig von t) die Menge

$$N^+(A) = N^+(A_t) = \{0, n, \dots, (|n| - 1)n\}$$

als maximale Relativnull. Selbstverständlich ist eine Zerlegung dieser Art nur möglich, wenn in G ein Element $n \neq 0$ mit $|n| \mid r$ existiert.

Da A nichts anderes als die Vereinigung der $\frac{r}{|n|}$ Teilmengen $A_t \subseteq A$ ist, ergibt sich mit Teil a) sofort: $N^+(A)$ ist bezüglich $A = \bigcup_{t=1, \dots, \frac{r}{|n|}} A_t$ maximale Relativnull.

Beispiel 5. Enthält die Obermenge G Elemente endlicher Ordnung, dann existieren stets endliche Teilmengen $A \subseteq G$, die Relativnullen besitzen. Seien $a, n \in G$ mit $|n| < \infty$ und $|a| \leq \infty$, wobei $|a|$ die Ordnung des Elementes a bezeichnen soll. Dann hat $A = \{a, a + n, \dots, a + (|n| - 1)n\}$ Relativnullen, z. B. $N(A) = \{0, n, 2n, \dots, (|n| - 1)n\}$. Speziell sei A die abelsche Gruppe der reellen Zahlen aus $[0, 5)$ mit der Addition modulo 5 als Verknüpfung. Dann gilt $|\sqrt{2}| = \infty$, $|1| = 5$, und $N(A) = \{0, 1, 2, 3, 4\}$ ist eine Relativnull bezüglich $A = \{\sqrt{2}, \sqrt{2} + 1, \dots, \sqrt{2} + 4\}$.

Beispiel 6. Sei G die additive Gruppe der ganzen Zahlen $0, 1, \dots, 7 \pmod{8}$. Dann hat die Menge $A = \{0, 4; 1, 5; 2, 6\} \subseteq G$ die maximale Relativnull $N^+(A) = \{0, 4\} \subseteq G$. (Vgl. Satz 5b).)

Folgerung 4. a) Eine zweielementige Menge $A = \{a_1, a_2\} \subseteq G$ besitzt genau dann Relativnullen, wenn ein $n \in G$ mit $|n| = 2$ existiert und $|a_1 - a_2| = 2$ ist.

b) Eine dreielementige Menge $A = \{a_1, a_2, a_3\} \subseteq G$ besitzt genau dann Relativnullen, wenn ein $n \in G$ mit $|n| = 3$ existiert und $|a_i - a_j| = 3$ für $i \neq j$, $i, j \in \{1, 2, 3\}$ ist.

Beweis. a) Für $r = 2$ muß (nach den Eigenschaften (3) und (4) des Beweises von Satz 5b)) $|n| = 2$ und $|a_2 - a_1| = 2$ sein. Also gilt: $A = \{a, a + n\}$ und $N^+(A) = \{0, n\}$.

b) Für $r = 3$ folgt (aus $|n| \mid r$ und den Eigenschaften (3) und (4)) $|n| = 3$ und $|a_i - a_j| = 3$ für $i \neq j$, $i, j \in \{1, 2, 3\}$, und es ist $A = \{a, a + n, a + 2n\}$ und $N^+(A) = \{0, n, 2n\}$.

Bisher wurde der Begriff Relativnull allein im Zusammenhang mit Summenmengen verwendet. Eine andere Möglichkeit zur Gewinnung von Relativnullen liefert der Begriff der „Produktmenge“. Dazu definiert man:

Definition 3. Gegeben sei eine Menge $S = \{s, t, \dots\}$, deren Elemente mit denen der Menge G derart verknüpfbar seien, daß für alle $a, b \in G$ und für alle $s \in S$ das Produkt sa existiert und in G liegt und außerdem $sa + sb = s(a + b)$ gilt. Dann wird S *Operatorenbereich* für G genannt und für jede Teilmenge $A \subseteq G$ die Menge

$$s \times A = \{sa \mid a \in A, s \in S\}$$

als *Produktmenge* bezeichnet.

Zum Beispiel ist die Menge Z der nicht negativen ganzen Zahlen ein Operatorbereich für jede Teilmenge einer abelschen Gruppe G .

Satz 6. a) Ist S Operatorenbereich für G , so gilt: Für alle $s \in S$ und für alle $A, B \subseteq G$ ist $(s \times A) + (s \times B) = s \times (A + B)$.

b) Mit $N(A)$ ist auch $d \times N(A)$, d natürliche Zahl, Relativnull bezüglich A , falls $N(A) \cap V(A) \neq \emptyset$ ist.

c) Mit $A \subseteq G$ ist auch $t \times A$, t Element eines Operatorenbereiches, eine Menge mit Relativnullen, falls $t \times N(A)$ für mindestens eine Relativnull $N(A)$ ein von Null verschiedenes Element enthält.

Beweis. a) Aus $s \times A = \{sa \mid a \in A\}$, $s \times B = \{sb \mid b \in B\}$ folgt mit Definition 3: $(s \times A) + (s \times B) = \{sa + sb \mid a \in A, b \in B\} = \{s(a + b) \mid a + b \in A + B\} = s \times (A + B)$.

b) Man beachte, daß $d \times N(A) \subseteq dN(A)$ ist (vgl. Bemerkung 2) und wende zuerst Satz 3a) und dann Satz 4c) an.

c) Mit Teil a) ist $(t \times A) + (t \times N(A)) = t \times (A + N(A)) = t \times A$.

§ 5. Maximale Relativnullen in G

Definition 4. a) Eine Relativnull $N^+(A)$ bezüglich $A \subseteq G$ heißt *maximal*, wenn keine Relativnull bezüglich A existiert, die $N^+(A)$ als echte Teilmenge enthält.

b) Eine Unterhalbgruppe $U^+(A) \subseteq G$, die bezüglich A die Eigenschaft (1) (vgl. Satz 2a)) besitzt, werde *maximal* genannt, wenn keine weitere Unterhalbgruppe von G mit der Eigenschaft (1) existiert, die $U^+(A)$ als echte Teilmenge enthält.

Satz 7. a) Für jede Teilmenge $A \subseteq G$ gilt: Entweder es gibt keine Relativnull bezüglich A oder aber es existiert genau eine maximale Relativnull $N^+(A)$.

b) Für jede Relativnull $N(A)$ bezüglich A , $A \subseteq G$, gilt: $N(A) \subseteq N^+(A)$ und, falls $0 \in A$ ist, sogar $N(A) \subseteq N^+(A) \subseteq A^4$. Die Null ist stets ein Element der maximalen Relativnull.

c) Falls $A \subseteq G$ Relativnullen besitzt, existiert genau eine maximale Unterhalbgruppe $U^+(A) \subseteq G$ mit $0 \in U^+(A)$, die bezüglich A die Eigenschaft (1) besitzt.

d) Die maximale Relativnull $N^+(A)$ stimmt mit der maximalen Unterhalbgruppe $U^+(A) \subseteq G$ (mit $0 \in U^+(A)$) überein, die bezüglich A die Eigenschaft (1) besitzt.

Beweis. a) Angenommen man hätte zwei verschiedene maximale Relativnullen $N_1^+(A)$ und $N_2^+(A)$ gefunden. Wegen Satz 3b) ist auch $N_1^+(A) \cup N_2^+(A)$ Relativnull bezüglich A . Mit $N_1^+(A) \cup N_2^+(A) \supsetneq N_1^+(A)$ folgt ein Widerspruch dazu, daß $N_1^+(A)$ maximale Relativnull bezüglich A ist.

b) Angenommen es existiert ein $n \in N(A)$ mit $n \notin N^+(A)$, dann ist (nach Satz 1e)) wegen

$$(N^+(A) \cup \{n\}) + A = (N^+(A) + A) \cup (\{n\} + A) = A \cup (\{n\} + A) = A$$

auch $N^+(A) \cup \{n\}$ eine Relativnull bezüglich A . Mit $N^+(A) \cup \{n\} \supsetneq N^+(A)$ ergibt sich somit ein Widerspruch zur Maximaleigenschaft von $N^+(A)$. Falls $0 \in A$ ist, folgt aus der Beziehung für Relativnullen (mit Satz 1c)) noch:

$$A = A + N^+(A) \supseteq \{0\} + N^+(A) = N^+(A).$$

Die letzte Bemerkung ist klar, da stets $A + \{0\} = A$ gilt.

c) Seien $U_1^+(A)$ und $U_2^+(A)$ zwei maximale Unterhalbgruppen (im folgenden mit U_1^+ bzw. U_2^+ abgekürzt), so daß (1) erfüllt ist. Zuerst sieht man, daß wegen $0 \in U_i^+$, $i = 1, 2$, auch $0 \in U_1^+ + U_2^+$ ist. Weiter ist mit Satz 1g) auch $U_1^+ + U_2^+$ Unterhalbgruppe von G . Schließlich folgert man aus $A + U_2^+ = A$ und aus der Tatsache, daß $\{a\} + U_1^+ \subseteq A$ für alle $a \in A$ erfüllt ist,

$$\{a\} + U_1^+ + U_2^+ \subseteq A + U_2^+ = A \text{ für alle } a \in A.$$

Die drei Aussagen zusammen bedeuten: Auch $U_1^+ + U_2^+$ ist eine Unterhalbgruppe von G mit $0 \in U_1^+ + U_2^+$, die bezüglich A die Eigenschaft (1) besitzt. Da nun aber $U_1^+ + U_2^+ \supseteq U_i^+ + \{0\} = U_i^+$, $i = 1, 2$, ist, folgt aus der Maximaleigenschaft der U_i^+ , $i = 1, 2$ sofort: $U_1^+ = U_1^+ + U_2^+ = U_2^+$.

d) Die maximale Unterhalbgruppe $U^+(A)$ mit der Eigenschaft (1) ist nach Definition 4b) und Satz 2b) eine Relativnull bezüglich A , und mit Satz 7b) gilt: $U^+(A) \subseteq N^+(A)$. Für jedes von Null verschiedene Element $n \in N^+(A)$ folgt die Existenz einer Unterhalbgruppe $U = \{0, n, 2n, \dots\} = n \times Z$ mit den Eigenschaften $\{a\} + U \subseteq A$ für alle $a \in A$ und $U + A = A$ (vgl. den Beweis von Satz 2a)). Aufgrund der Bedeutung von $U^+(A)$ ist $U \subseteq U^+(A)$ (beachte die Schlußweise für die Menge $N^+(A)$ im Beweis von Teil b)); daher folgt mit $n \in U$ auch $n \in U^+(A)$. Somit ist $U^+(A) \supseteq N^+(A)$.

Folgerung 5. Dann und nur dann gilt $N^+(A) = A$, wenn $A \subseteq G$ Unterhalbgruppe mit $0 \in A$ ist. Insbesondere ist genau dann $N^+(A) = G$, wenn $A = G$ ist.

Beweis. Sei $N^+(A) = A$. Nach Satz 7d) ist $N^+(A)$ Unterhalbgruppe mit $0 \in N^+(A)$, womit die eine Richtung schon bewiesen ist. Umgekehrt folgt für eine Unterhalbgruppe

⁴⁾ Die letzte Ungleichung bleibt auch richtig, wenn man an Stelle von $0 \in A$ die schwächere Bedingung $A \cap V(N^+(A)) \neq \emptyset$ setzt.

A mit $0 \in A$ nach Beispiel 1) aus § 3 stets $A + A = A$ und daraus mit Satz 7b) die Behauptung.

Folgerung 6. Die maximale Relativnull $N^+(A)$ einer Menge $A \subseteq G$ ist eine Menge mit Relativnullen und zugleich ihre eigene maximale Relativnull. In Ergänzung zu Satz 4 gilt ferner die Beziehung $N^+(A) = V(A) \cup W(A)$.

Beweis. Da die maximale Relativnull $N^+(A)$ Unterhalbgruppe von G mit $0 \in N^+(A)$ ist, stellt die obige Aussage nur einen Spezialfall der Folgerung 5 dar.

Folgerung 7. Für beliebige Mengen $A, B \subseteq G$ mit Relativnullen gilt:

$$(7) \quad N^+(A) + N^+(B) \subseteq N^+(A + B).$$

Sind A und B Unterhalbgruppen von G mit $0 \in A \cap B$, dann folgt sogar:

$$N^+(A) + N^+(B) = N^+(A + B).$$

Beweis. Wegen Satz 1a) und b) gilt:

$$(A + B) + (N^+(A) + N^+(B)) = (A + N^+(A)) + (B + N^+(B)) = A + B,$$

d. h. $N^+(A) + N^+(B)$ ist Relativnull bezüglich $A + B$. Daraus folgt mit Satz 7b) die Behauptung für beliebige Mengen $A, B \subseteq G$. Im Spezialfall ist wegen Satz 1g) auch $A + B$ Unterhalbgruppe mit $0 \in A + B$, so daß sich mit Folgerung 5 ergibt:

$$N^+(A + B) = A + B = N^+(A) + N^+(B).$$

Bemerkung 4. Sind in Folgerung 7 die Mengen $A, B \subseteq G$ nicht Unterhalbgruppen, dann ist das Gleichheitszeichen in der Ungleichung (7) im allgemeinen falsch. So besitzen etwa die Mengen $A = B = R[0, 1) - \left\{0, \frac{1}{2}\right\}$ als maximale Relativnullen die Unterhalbgruppen $N^+(A) = N^+(B) = \left\{0, \frac{1}{2}\right\}$, so daß $N^+(A) + N^+(B) = \left\{0, \frac{1}{2}\right\}$ folgt. Da $A + B = R[0, 1)$ ist, gilt $N^+(A + B) = R[0, 1)$. Also in der Tat

$$N^+(A) + N^+(B) \subsetneq N^+(A + B).$$

Die Mengen $A = \{3, 5, 6; 8, 9, 10, \dots\}$ und $B = \{0, 2, 6, 7; 9, 10, 11, \dots\}$ sind zwei Unterhalbgruppen mit $0 \notin A + B$ in der (additiven) Gruppe der ganzen Zahlen. Aus $N^+(A) = \{0, 3; 5, 6, 7, \dots\}$ bzw. $N^+(B) = \{0, 7; 9, 10, 11, \dots\}$ ergibt sich

$$N^+(A) + N^+(B) = \{0, 3; 5, 6, 7, \dots\}.$$

Andererseits folgt aus $A + B = \{3; 5, 6, 7, \dots\}$ sofort $N^+(A + B) = \{0; 2, 3, 4, \dots\}$. Auch in diesem Beispiel ist $N^+(A) + N^+(B) \subsetneq N^+(A + B)$.

Folgerung 8. Ist $N(A)$ eine beliebige Relativnull bezüglich A , dann gilt: Für alle $n \in N(A)$, $n \neq 0$, ist

$$U_n = \{0, n, 2n, \dots\} \subseteq N^+(A).$$

Beweis. Man beachte, daß mit $a \in A$ auch alle Elemente $a + n, a + 2n, a + 3n, \dots$ in A enthalten sind.

Folgerung 9. Für jede echte Unterhalbgruppe U von $N^+(A)$, $0 \in U$, gibt es Mengen $C \subseteq A$, $D \subseteq N^+(A)$ mit den Eigenschaften $A = U + C$ und $N^+(A) = U + D$.

Beweis. Man wähle für die erste Aussage $C = A$ und beachte den Satz 2b). Zur zweiten Aussage: Wegen Folgerung 6 ist $N^+(A)$ Relativnull bezüglich $N^+(A)$, so daß

nach Folgerung 2 die Menge $N^+(A)$ auch die Unterhalbgruppe $U \subseteq N^+(A)$ als Relativnull besitzt. Also gilt: $N^+(A) + U = N^+(A)$ und damit ist die Richtigkeit der zweiten Behauptung mit $D = N^+(A)$ gezeigt.

Bemerkung 5. Die in den Beispielen 1) und 5) genannten Relativnullen sind maximal, während im Beispiel 2) die maximale Relativnull bezüglich A durch $N^+(A) = \lim_{k \rightarrow \infty} N_k(A)$ gegeben ist.

§ 6. Ausdehnung auf abelsche Halbgruppen

Die Sätze und Beweise der §§ 2—5 legen es nahe, anstatt von abelschen Gruppen G von abelschen Halbgruppen H als Obermengen auszugehen. Bei den meisten Beweisen wird nur die Halbgruppeneigenschaft der Gruppe G ausgenutzt, die Null wird dagegen fast überall benötigt. Bei einer solchen Erweiterung müßte der Satz 3e) gestrichen werden. Die in Satz 4a) für jedes $A \subseteq G$ definierte Gruppe $V(A)$ müßte durch die Menge $V(A)$ der Elemente $n \in G$ mit $\{n\} + A = A$ ersetzt werden. $V(A)$ ist dann nur Unterhalbgruppe mit $0 \in V(A)$, so daß die Bemerkung in Satz 4b) „ $W(A)$ besteht aus vollen Restklassen von G modulo $V(A)$ “ wegzulassen ist. Mit dieser Abschwächung bleiben alle bisher bewiesenen Sätze mit Ausnahme des Satzes 3e) richtig, wenn man statt einer beliebigen abelschen Gruppe G eine beliebige abelsche Halbgruppe H mit $0 \in H$ zugrunde legen würde. Die Beispiele 1), 4) und 5) der §§ 3 und 4 sind von Anfang an auf Halbgruppen als Obermengen zugeschnitten. Als Ergänzung das

Beispiel 7. Die Halbgruppe $H = \{r + s\sqrt{2} \pmod{5} \mid r, s \in \mathbb{Z}\}$ sei Obermenge. Dann ist auch hier $V(H) = \{r \pmod{5} \mid r \in \mathbb{Z}\}$ eine Untergruppe von H , und es gilt

$$W(H) = \{r + s\sqrt{2} \pmod{5} \mid r, s \in \mathbb{Z}, s \geq 1\}.$$

§ 7. Relativnullen in G_∞ und G_0

1. G_∞ sei eine aperiodische abelsche Gruppe mit additiver Verknüpfung. Speziell bildet die Menge der ganzen Zahlen bezüglich der Addition eine solche Gruppe. Hier ist $V(A) = \{0\}$ für alle $A \subseteq G_\infty$. Die von H.-H. Ostmann [2] betrachteten Teilmengen der Menge $\mathbb{Z}$ der nicht negativen ganzen Zahlen sind auch Teilmengen dieser Gruppe G_∞ . Somit sind, wie bereits in der Einleitung erwähnt, die nachstehend bewiesenen Sätze Verallgemeinerungen der Betrachtungen von H.-H. Ostmann.

Alle Sätze der §§ 2—6 haben für die Teilmengen von G_∞ Gültigkeit. Doch sind die Aussagen der Sätze 5a) und b) und der Folgerung 4 leer. Da auch die additive (abelsche) Gruppe der reellen Zahlen aperiodisch ist, gilt dies insbesondere für die Teilmengen der Menge der reellen Zahlen.

Beispiel 8. Die Menge $A = \{1 + \sqrt{2}, 1 + 3\sqrt{2}, 1 + 4\sqrt{2}, 1 + 5\sqrt{2}, \dots\}$ hat als maximale Relativnull die Unterhalbgruppe

$$N^+(A) = \{0, 2 \cdot \sqrt{2}, 3 \cdot \sqrt{2}, 4 \cdot \sqrt{2}, \dots\}.$$

Die Menge $N(A) = \{0, 2, 4, 6, \dots\}$ ist Relativnull bezüglich

$$A = [0, 1) \cup [2, 3) \cup [4, 5) \cup \dots \cup [2n, 2n+1) \cup \dots$$

Charakteristisch für aperiodische abelsche Gruppen ist der nachstehende Satz.

Satz 8. Keine endliche Menge $A \subseteq G_\infty$ besitzt Relativnullen.

Beweis. Es sei $A \subseteq G_\infty$ eine Menge mit Relativnullen und $N^+(A)$ ihre maximale Relativnull. Dann folgt für ein $n \in N^+(A)$, $n \neq 0$, und ein $a \in A$, daß die unendlich vielen verschiedenen (man beachte $|n| = \infty$!) Elemente $a, a + n, a + 2n, \dots$ zu A gehören. Also kann A nicht endlich sein.

Bemerkung 6. Die Aussage von Satz 8 gilt, wie man unmittelbar erkennt, im Sinne von § 6 auch für aperiodische abelsche Halbgruppen H_∞ .

2. G_0 sei eine beliebige periodische abelsche Gruppe mit additiver Verknüpfung. Das Beispiel 5) aus § 4 zeigt, daß hier, im Gegensatz zu Satz 8, endliche Teilmengen Relativnullen besitzen können. Die Aussagen der §§ 2—6 bleiben auch für G_0 richtig, lassen sich aber hinsichtlich der Voraussetzungen abschwächen bzw. hinsichtlich der Ergebnisse etwas verschärfen.

Ist $U \subseteq G_0$ eine Unterhalbgruppe, so gilt für jedes $n \in U$, $n \neq 0$, hier $2n \in U$, $3n \in U, \dots, (|n| - 1)n \in U$, $|n| \cdot n \in U$; somit ist $-n \in U$ und $0 \in U$. Also ist in G_0 jede Unterhalbgruppe Untergruppe. Zugleich ergibt sich, daß in G_0 die Voraussetzung $0 \in U$ von selbst erfüllt ist. Die Sätze 2a), b), 7c), d) und die Folgerungen 5 und 7 sind daher für die Obermenge G_0 so abzuändern, daß statt „Unterhalbgruppe U mit $0 \in U$ “ einfach „Untergruppe U “ und für die Eigenschaft (1) etwa „ A besteht aus vollen Restklassen nach der Untergruppe U “ geschrieben wird. Die Folgerung 2 läßt sich wie folgt verschärfen:

Folgerung 10. Jede Teilmenge T mit $T \neq \{0\}$, $T \neq \emptyset$, einer Relativnull $N(A)$ ist wieder eine Relativnull bezüglich A . Speziell gilt $\{n\} + A = A$ für alle $n \in N(A)$.

Beweis. Für jedes $a \in A$ gilt

$$a + n \in A, \dots, a + (|n| - 1)n = a - n \in A.$$

Daraus folgt: $a = (a - n) + n \in A + \{n\}$. Also gilt $A + \{n\} \supseteq A$, so daß

$$T + A = \bigcup_{n \in T} \{n\} + A = \bigcup_{n \in T} (A + \{n\}) \supseteq A$$

ist. Mit $T + A \subseteq N^+(A) + A = A$ folgt die Behauptung.

Die Sätze 4a), b) und 7b) kann man in G_0 ebenfalls ergänzen.

Folgerung 11. Für alle Mengen $A \subseteq G_0$ mit Relativnullen gilt:

$$(8) \quad V(A) = N^+(A) \text{ und } W(A) = \emptyset.$$

$$(9) \quad \text{Ist } 0 \notin A, \text{ so folgt } N^+(A) \cap A = \emptyset.$$

Beweis. Der erste Teil ist klar aufgrund der Folgerung 10. Zum zweiten Teil: Angenommen es gibt ein $a \in A$ mit $a \in N^+(A) = V(A)$. Dann ist auch $-a \in N^+(A)$ (wegen der Gruppeneigenschaft von $V(A)$). Also ist $0 = a - a \in A + N^+(A) = A$, im Widerspruch zur Voraussetzung.

Bemerkung 7. In beliebigen abelschen Gruppen G ist die Aussage (9) im allgemeinen falsch, wie etwa das nachstehende Beispiel zeigt:

In der Gruppe G der ganzen Zahlen hat die Menge

$$A = \{n_0 + r \mid r = 0, 1, 2, \dots; n_0 \in \mathbb{Z}, n_0 \geq 1, \text{ fest}\}$$

die maximale Relativnull $N^+(A) = \mathbb{Z}$, so daß $N^+(A) \cap A = A \neq \emptyset$ gilt.

§ 8. Relativnullen in $R[0, 1)$

Die Obermenge $R[0, 1)$ der rationalen Zahlen x mit $0 \leq x < 1$ ist gegenüber der Addition modulo 1 eine periodische abelsche Gruppe, also vom Typus G_0 . Schon in G_0 waren erhebliche Unterschiede gegenüber dem allgemeinen Fall festgestellt worden. Insbesondere weichen die Relativnullen $N(A)$ und die dazugehörigen Mengen A in der Obermenge Z der nicht negativen ganzen Zahlen (vgl. [2]) recht deutlich in ihrer Struktur von den entsprechenden Mengen in $R[0, 1)$ (bzw. G_0) ab. In G_0 (also auch in $R[0, 1)$) kann eine Menge aus endlich vielen Elementen Relativnullen besitzen (vgl. etwa das Beispiel 5) aus § 4 oder den Satz 5b), was nach Ostmann [2], S. 22, für Z nicht zutrifft. Ebenso muß in G_0 eine Relativnull nicht die Null enthalten (vgl. die Folgerung 10), was aber für Z (s. [2]) gefordert werden muß. Für den Spezialfall $R[0, 1)$ von G_0 läßt sich leicht ein Beispiel angeben, das sogar beide Abweichungen zeigt:

Beispiel 9. $A = \left\{ \frac{1}{4}, \frac{3}{4} \right\}$ mit $N(A) = \left\{ \frac{1}{2} \right\}$.

Im folgenden soll nun die Menge $R[0, 1)$ in bezug auf Mengen mit Relativnullen näher untersucht werden. Dabei zeigt sich, daß ein U -Kern einer Teilmenge A von $R[0, 1)$ (vgl. Bemerkung 3) und zwar als eindeutig bestimmte Menge explizit angebbar ist.

Zunächst seien einige für das folgende benötigte Mengen definiert. Es sei (vgl. Definition 3) für n ganz, $n \geq 2$, $F_n = \{0, 1, \dots, n-1\}$, $E_n = \frac{1}{n} \times F_n$.

Dann ist E_n eine zyklische Untergruppe endlicher Ordnung von $R[0, 1)$, erzeugt etwa vom Element $\frac{1}{n}$. Sei ferner A eine Menge mit Relativnullen und $\frac{i}{n} \in N^+(A)$, $(i, n) = 1$. Dann sei bei festem n und für beliebiges ganzes $q \geq 1$

$$(10) \quad K(nq, A) = \left\{ \frac{t_1}{nq}, \dots, \frac{t_{q'}}{nq} \left| \begin{array}{l} \frac{t_i}{nq} \in A, (t_i, nq) = 1, t_i \text{ minimal mod } q^5 \\ \text{für } i = 1, \dots, q'; t_i \not\equiv t_j \pmod{q} \text{ für } i \neq j \end{array} \right. \right\},$$

ferner $K(0, A) = \{0\} \cap A$ und

$$(11) \quad K(E_n, A) = \bigcup_{q=0,1,2,\dots} K(nq, A).$$

Dann gilt $E_n \subseteq N^+(A)$. Denn es ist $F_n = \{ji \pmod{n} \mid j \in F_n\}$ und daher mit $\frac{i}{n} \in N^+(A)$ auch $\frac{1}{n} \in N^+(A)$ und $\frac{1}{n} \times F_n \subseteq N^+(A)$. Ferner ist $K(nq, A) \subseteq A$, also auch $K(E_n, A) \subseteq A$. Außerdem gilt für $q_1 \neq q_2$ stets

$$(K(nq_1, A) + E_n) \cap (K(nq_2, A) + E_n) = \emptyset.$$

Denn ist eine der Mengen $K(nq_i, A)$, $i = 1, 2$, leer, so ist nichts zu zeigen. Angenommen $K(nq_i, A)$, $i = 1, 2$, enthielten ein Element gemeinsam, etwa

$$\frac{t_1}{q_1 n} + \frac{r_1}{n} = \frac{t_2}{q_2 n} + \frac{r_2}{n},$$

dann müßte

$$(q_2 t_1 - q_1 t_2) \equiv q_1 q_2 (r_2 - r_1) \pmod{q_1 q_2 n}$$

⁵⁾ „ t_i minimal mod q “ bedeutet, daß für alle $\frac{k_i}{nq} \in A$ mit $0 < k_i < nq$, $(k_i, nq) = 1$, $k_i \equiv t_i \pmod{q}$ stets $0 < t_i \leq k_i$ gilt.

gelten, woraus sich $q_i/(q_2 t_1 - q_1 t_2)$, $i = 1, 2$, ergäbe. Also wäre q_1/q_2 und q_2/q_1 , im Widerspruch zu $q_1 \neq q_2$.

Schließlich läßt sich zeigen, daß $q' \leq \varphi(q)$ sein muß. Da diese Aussage hier keine Rolle spielt, wird der Beweis im Anhang dieses Abschnittes gebracht.

Satz 9. *Unter den gemachten Voraussetzungen gilt:*

a) $K(E_n, A) + E_n = A$.

b) Für alle $t \in K(E_n, A)$ ist $(K(E_n, A) - \{t\}) + E_n \subsetneq A$.

Bemerkung 8. Die Eigenschaft a) liefert eine Zerlegung (2), wie in Satz 2b) gefordert wird. Die Eigenschaft b) bedeutet, daß $K(E_n, A)$ der (eindeutig bestimmte) E_n -Kern von A im Sinne von Bemerkung 3 ist.

Mit Rücksicht auf die in $R[0, 1)$ zugrunde gelegte Verknüpfung ist Gleichheit zwischen rationalen Zahlen stets als Kongruenz modulo 1 zu verstehen.

Beweis von Satz 9. a) Wegen $K(E_n, A) \leq A$ gilt mit Satz 1 c):

$$K(E_n, A) + E_n \leq A + E_n \leq A + N^+(A) = A.$$

Ist $\frac{i}{a}$ mit $(i, a) = 1$ ein beliebiges Element von A und etwa $\frac{1}{n} \in N^+(A)$, dann gilt: Für alle $r = 1, \dots, n$ ist

$$(12) \quad A \ni \frac{i}{a} + \frac{r}{n} = \frac{i \frac{n}{(a, n)} + r \frac{a}{(a, n)}}{\frac{a}{(a, n)} n}.$$

Da die Kongruenz $\frac{a}{(a, n)} r \equiv \left(\frac{a}{(a, n)}, n \right) \pmod{n}$ für mindestens ein r mit $0 < r < n$ lösbar ist, folgt für dieses r

$$\left(i \frac{n}{(a, n)} + r \frac{a}{(a, n)}, \frac{a}{(a, n)} n \right) = \left(i \frac{n}{(a, n)} + \left(\frac{a}{(a, n)}, n \right), n \right) = 1.$$

Also existiert ein r mit (s. (12)):

$$(13) \quad \frac{i}{a} + \frac{r}{n} = \frac{j}{nq} \in A, \quad q = \frac{a}{(a, n)}, \quad (j, nq) = 1.$$

Für jeden anderen Bruch $\frac{k}{nq} \in A$ mit $k \equiv j \pmod{q}$ folgt

$$\frac{k}{nq} = \frac{j + k - j}{nq} = \frac{j}{nq} + \frac{k - j}{q} \cdot \frac{1}{n} \in A + E_n = A.$$

Also muß wegen (13) in $K(nq, A)$ einer der Brüche $\frac{j'}{nq} \in A$ mit $(j', nq) = 1$, $j' \equiv j \pmod{q}$ vorkommen. Es ist daher für alle $s \in F_n$

$$\frac{j' + sq}{nq} = \frac{j'}{nq} + \frac{s}{n} \in K(E_n, A) + E_n.$$

Dies liefert für $s \equiv \frac{j - j'}{q} - r \pmod{n}$ wegen (13):

$$\frac{j'}{nq} + \frac{j - j'}{nq} - \frac{r}{n} = \frac{j}{nq} - \frac{r}{n} = \frac{i}{a} \in K(E_n, A) + E_n.$$

Also ist $A \leq K(E_n, A) + E_n$.

b) Das aus $K(E_n, A)$ gestrichene t kann in der Form $t = \frac{i}{nq}$ mit passendem q und $(i, nq) = 1$, $\frac{i}{nq} \in A$ angesetzt werden. Für dieses q enthält $K(nq, A) - \{t\}$ kein Element der Gestalt $\frac{i'}{nq}$, $(i', nq) = 1$, $i' \equiv i \pmod{q}$, da höchstens ein Element dieser Art in $K(nq, A)$ vorhanden ist. Würde nun in der Aussage von Satz 9b) das Gleichheitszeichen gelten, so gäbe es ein $\frac{k}{nr} \in A$ mit $(k, nr) = 1$ derart, daß gleichzeitig $\frac{k}{nr} \in K(E_n, A) - \{t\}$ ist. Daraus folgt die Existenz eines Elementes $\frac{j}{n} \in E_n$ mit

$$\frac{k}{nr} + \frac{j}{n} = t = \frac{i}{nq} \quad \text{oder} \quad \frac{k}{nr} = \frac{i}{nq} - \frac{j}{n} = \frac{i - jq}{nq} \quad \text{oder} \quad \frac{k}{r} = \frac{i - jq}{q}.$$

Aus $(i, nq) = 1$ und $(k, nr) = 1$ folgt $(i - jq, q) = (i, q) = 1$ bzw. $(k, r) = 1$. Daher ergibt sich $r = q$ und $k = i - jq \equiv i \pmod{q}$. Also wäre $\frac{k}{nq} \in K(nq, A) - \{t\}$ mit $(k, nq) = 1$ und $k \equiv i \pmod{q}$, im Widerspruch zu dem zuvor Bewiesenen.

Bemerkung 9. Für das bei vorgegebenem $N^+(A)$ kleinstmögliche $n \geq 2$ wird (nach Bemerkung 3) der E_n -Kern von A (aus Satz 9b)) zum Kern von A schlechthin, und E_n heißt dann eine erzeugende Relativnull (vgl. [2], S. 24).

Beispiel 10. Sei

$$A = \left\{ \frac{1, 1 + 6^r, 1 + 2 \cdot 6^r, \dots, 1 + 5 \cdot 6^r}{6^{r+1}} \mid r = 1, 2, 3, \dots \right\}.$$

Dann ist

$$N^+(A) = \left\{ 0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6} \right\}.$$

Zuerst sei $n = 6$ gewählt. Die einzigen nicht leeren $K(6q, A)$, $q \in \mathbb{Z}$, sind die Mengen

$$K(6 \cdot 6^r, A) = \left\{ \frac{1}{6^{r+1}} \right\} \quad \text{für } r \geq 1.$$

Der E_6 -Kern von A ist also die Menge

$$K(E_6, A) = \left\{ \frac{1}{6^r} \mid r = 2, 3, \dots \right\}.$$

Geht man statt von E_6 von der erzeugenden Relativnull E_2 (vgl. Bemerkung 9) aus, d. h. ist $n = 2$, dann haben die nicht leeren $K(2q, A)$, $q \in \mathbb{Z}$ die Gestalt

$$K(2 \cdot 3 \cdot 6^r, A) = \left\{ \frac{1, 1 + 6^r, 1 + 2 \cdot 6^r}{6^{r+1}} \right\} \quad \text{für } r \geq 1,$$

und der E_2 -Kern von A ist durch

$$K(E_2, A) = \left\{ \frac{1, 1 + 6^r, 1 + 2 \cdot 6^r}{6^{r+1}} \mid r = 1, 2, 3, \dots \right\}$$

gegeben; $K(E_2, A)$ ist der Kern von A schlechthin.

Der Satz 9 sei noch für $N^+(A)$ statt für A formuliert, denn nach Folgerung 6 ist $N^+(A)$ eine Menge mit Relativnullen, genauer $N^+(A)$ ist zu sich selbst maximale Relativnull.

Folgerung 12. Unter den Voraussetzungen und in den Bezeichnungen von Satz 9 gilt:

a) $K(E_n, N^+(A)) + E_n = N^+(A)$.

b) Für alle $t \in K(E_n, N^+(A))$ ist

$$(K(E_n, N^+(A)) - \{t\}) + E_n \leq N^+(A).$$

Abschließend werde an zwei einfachen Fällen gezeigt, wie einschränkend sich die Vorgabe einer bestimmten Menge N , die Relativnull bezüglich $A \leq R[0, 1)$ werden soll, auf die Menge A auswirkt.

Satz 10. a) Ist in einer Menge N als Teilmenge von $R[0, 1)$ jede natürliche Zahl als echter Nenner mindestens einmal vertreten, dann kann N nur noch Relativnull bezüglich der ganzen Menge $R[0, 1)$ sein.

b) Wenn $N = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \dots, \frac{1}{p}, \dots \right\}$ p Primzahl ist, dann kann N höchstens bezüglich solcher Mengen $A \leq R[0, 1)$ Relativnull sein, für die ein Element $b \in A$ existiert mit $b \neq 0$ und

$$\left\{ \frac{i}{a} \mid (i, a) = 1, \quad 0 < \frac{i}{a} < 1, \quad \mu(a) = \pm 1 \right\} \leq A \text{ für } 0 \in A,$$

$$\left\{ \frac{i}{a} \mid (i, a) = 1, \quad 0 < \frac{i}{a} < 1, \quad \mu(a) = 0 \right\} \geq A \text{ für } 0 \notin A.$$

Beweis. a) Mit der Menge N besitzt eine Menge A , die N zur Relativnull hat, nach Folgerung 8 auch $N^+(A) = R[0, 1)$ als maximale Relativnull.

b) Nach Folgerung 8 gehört mit $\frac{1}{p} \in N(A)$ gleich die ganze Untergruppe $E_p = \frac{1}{p} \times \{0, 1, \dots, p-1\}$ zur maximalen Relativnull $N^+(A)$. Weiterhin folgt aus

$$\frac{i_j}{p_j} \in N^+(A), \quad (i_j, p_j) = 1, \quad p_j \text{ Primzahl}, \quad j = 1, \dots, r$$

mit $p_s \neq p_t$ für $s \neq t$ und $P = \prod_{i=1, \dots, r} p_i$ sofort

$$\frac{i_1}{p_1} + \dots + \frac{i_r}{p_r} = \frac{1}{P} \left(i_1 \frac{P}{p_1} + \dots + i_r \frac{P}{p_r} \right) \leq N^+(A)$$

mit $\left(i_1 \frac{P}{p_1} + \dots + i_r \frac{P}{p_r}, P \right) = 1$, so daß die von N erzeugte Untergruppe

$$N^+(A) = \left\{ 0, \frac{i}{P} \mid (i, P) = 1 \right\}$$

als kleinste maximale Relativnull in Frage kommt. Ist $0 \in A$, dann folgt mit Satz 7b) ($N^+(A) \leq A$) die Behauptung. Ist dagegen $0 \notin A$, dann wende man die Aussage (9) der Folgerung 11 an.

§ 9. Anhang. Beweis für die Aussage $q' \leq \varphi(q)$

1. Jede Restklasse $\left\{ \frac{t}{nq} \right\} + E_n$, $(t, nq) = 1$, enthält genau $\frac{\varphi(nq)}{\varphi(q)}$ gekürzte Brüche mit dem Nenner nq .

Alle Elemente einer solchen Restklasse haben die Gestalt $\frac{t + rq}{nq}$, $r \in F_n$. Daher braucht man zum Beweis der Zwischenbehauptung nur zu untersuchen, wie viele Zahlen

$r \in F_n$ mit der Eigenschaft $(t + rq, nq) = 1$ existieren. Sei $n = \prod p^i \cdot K$ und $q = \prod p^j \cdot L$ mit $(p, K) = (p, L) = (K, L) = 1$. Da $(t, q) = 1$ und $(t + rq, \prod p^i) = 1$ ist, gilt es, die Anzahl derjenigen $r \in F_n$ mit $(t + rq, K) = 1$ zu finden. Durchläuft r die Werte $0, 1, \dots, n-1$ dann ergeben die Zahlen $t + rq$ genau $\frac{n}{K}$ vollständige Restsysteme modulo K . Jedes dieser Restsysteme modulo K enthält $\varphi(K)$ zu K und wegen $(t + rq, \prod p^i) = 1$ auch zu n teilerfremde Zahlen. Also sind unter den Elementen $t + rq$, $r = 0, 1, \dots, n-1$, genau $\frac{n}{K} \cdot \varphi(K)$ zu n teilerfremde Zahlen enthalten.

Da einerseits $\frac{n}{K} \cdot \varphi(K) = \prod p^i \cdot \varphi(K)$ und andererseits auch

$$\frac{\varphi(nq)}{\varphi(q)} = \frac{\prod p^{i+j} \left(1 - \frac{1}{p}\right) \cdot \varphi(K) \cdot \varphi(L)}{\prod p^j \left(1 - \frac{1}{p}\right) \cdot \varphi(L)} = \prod p^i \cdot \varphi(K)$$

gilt, sind in jeder Restklasse $\left\{ \frac{t}{nq} \right\} + E_n$, $(t, nq) = 1$ genau $\frac{\varphi(nq)}{\varphi(q)}$ gekürzte Brüche mit Nenner nq enthalten.

2. Für alle Elemente einer Restklasse $\left\{ \frac{t}{nq} \right\} + E_n$ gilt wegen

$$\frac{j}{nq} = \frac{t}{nq} + \frac{r}{n} = \frac{t + rq}{nq}$$

stets $j \equiv t \pmod{q}$. Also ist in $K(nq, A)$ höchstens ein Element einer festen Restklasse $\left\{ \frac{t}{nq} \right\} + E_n$ enthalten.

3. In $R[0, 1)$ existieren genau $\varphi(nq)$ verschiedene Brüche $\frac{s}{nq}$ mit $(s, nq) = 1$, und nach 1. enthält jede Restklasse $\left\{ \frac{t}{nq} \right\} + E_n$ genau $\frac{\varphi(nq)}{\varphi(q)}$ solcher Brüche $\frac{s}{nq}$. Also gibt es $\varphi(q)$ voneinander verschiedene Restklassen der Gestalt $\left\{ \frac{t}{nq} \right\} + E_n$, $(t, nq) = 1$. Mit 2. folgt, daß $K(nq, A)$ höchstens $\varphi(q)$ Elemente $\frac{t_i}{nq}$ mit $(t_i, nq) = 1$, $t_i \equiv t_j \pmod{q}$ für $i \neq j$ enthält.

Bemerkung 10. Die Betrachtungen über Mengen mit Relativnullen bilden nur einen Bruchteil der gesamten gewöhnlichen additiven Zahlentheorie (bezüglich der Menge Z der nicht negativen ganzen Zahlen). Daher wäre es wünschenswert, auch die anderen Grundbegriffe der additiven Zahlentheorie (vgl. z. B. [4] oder [2]) (etwa finite, asymptotische und natürliche Dichte; Basis, Ordnung, wesentliche Komponente) auf geeignete abelsche Gruppen, insbesondere auf die Sonderfälle der Mengen rationaler Zahlen $R[0, 1)$ und $R(-\infty, +\infty)$, zu übertragen. H. B. Mann (An addition theorem for sets of elements of abelian groups, Proc. Amer. Math. Soc. 4 (1953), S. 423) hat für endliche abelsche Gruppen G eine Abschätzung der Elementanzahl einer Menge $C = AB = \{ab \mid a \in A, b \in B\}$ durch die Elementanzahlen der Ausgangsmengen $A \subseteq G$, $B \subseteq G$ angegeben.

Literaturverzeichnis

- [1] *M. Hall*, The theory of groups, New York 1959.
- [2] *H.-H. Ostmann*, Additive Zahlentheorie. I, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge Heft 7, 1956.
- [3] *H.-H. Ostmann*, Untersuchungen über den Summenbegriff in der additiven Zahlentheorie, Math. Annalen **120** (1947/49), 165—196.
- [4] *H. Rohrbach*, Einige neuere Untersuchungen über die Dichte in der additiven Zahlentheorie, Jahresbericht DMV **48** (1938), 199—236.

6501 Finthen, Am Hang 13

Eingegangen 20. März 1967

Ein Kuzminscher Satz über den Jacobischen Algorithmus

Von Fritz Schweiger in Wien

Herrn Professor Dr. Hans Rohrbach zum 65. Geburtstag

In einigen vorangegangenen Arbeiten ([2], [3] und [4]), die sich mit den ergodischen Eigenschaften der Transformation δ des Einheitswürfels in sich

$$\delta(x_1, \dots, x_n) = \begin{cases} \frac{x_2}{x_1} - \left[\frac{x_2}{x_1} \right], \dots, \frac{1}{x_1} - \left[\frac{1}{x_1} \right] & \text{sonst} \\ 0 & \text{für } (x_1, \dots, x_n) = (0, 0, \dots, 0) \end{cases}$$

beschäftigt haben, wurde unter anderem bewiesen: Es gibt ein bezüglich δ invariantes und bezüglich des Lebesgueschen Maßes L äquivalentes Maß μ , welches, da δ unzerlegbar in bezug auf L ist, 1-deutig bestimmt ist. — Leider ist es mir bis jetzt nicht gelungen, eine „explizite“ Gestalt dieses Maßes μ anzugeben. Es ist daher bemerkenswert, daß sich unter einer Zusatzvoraussetzung, deren Richtigkeit höchstwahrscheinlich ist, ein Analogon zu dem bekannten Satze von Gauß-Kuzmin beweisen läßt. Die Voraussetzung, die ich Kuzminsche Bedingung nennen will, lautet:

Die Dichte $\varrho = \varrho(x_1, \dots, x_n)$ des invarianten Maßes μ besitzt beschränkte partielle Ableitungen 1. Ordnung. d. h. es gilt:

$$(1) \quad \frac{\partial \varrho}{\partial x_i} < A.$$

Aus dem Satz von Radon-Nikodym wissen wir nur, daß $\varrho \in L_1$ ist. — Der Einfachheit halber beschränken wir uns nun auf den Fall $n = 2$. Es ist dann

$$\delta(x_1, x_2) = \left(\frac{x_2}{x_1} - \left[\frac{x_2}{x_1} \right], \frac{1}{x_1} - \left[\frac{1}{x_1} \right] \right) \text{ für } (x_1, x_2) \neq (0, 0)$$

und wenn wir

$$a_1(x) = \left[\frac{x_2}{x_1} \right], \quad a_2(x) = \left[\frac{1}{x_1} \right]$$

und induktiv

$$a_1^{(v)}(x) = a_1(\delta^{v-1}x), \quad a_2^{(v)}(x) = a_2(\delta^{v-1}x)$$

setzen, so sei wie immer

$$B^{(0)} = \{(x_1, x_2) \mid 0 < x_1 < 1, 0 < x_2 < 1\}$$

$$B^{(v)}(k_i^{(1)}, \dots, k_i^{(v)}) = \{x \in B^{(0)} \mid a_i^{(\mu)}(x) = k_i^{(\mu)}, \mu = 1, \dots, v\}.$$

Es ist dann

$$\delta^{-1}: B^{(0)} \rightarrow B^{(1)}$$

gegeben durch

$$\delta^{-1}(x_1, x_2) = \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right).$$

Wir teilen nun das Einheitsquadrat $B^{(0)}$ in zwei Gebiete:

$$I_1 = \{(x_1, x_2) \mid 0 < x_1 < 1, x_1 \leq x_2 < 1\}$$

$$I_2 = \{(x_1, x_2) \mid 0 < x_1 < 1, 0 < x_2 < x_1\}.$$

Man sieht nun, daß wenn $(x_1, x_2) \in I_1$ in $\delta^{-1}(x_1, x_2)$ alle zulässigen Werte für k_1 und k_2 auftreten können, nämlich

$$0 \leq k_1 \leq k_2, \quad 1 \leq k_2 < \infty.$$

Ist hingegen $(x_1, x_2) \in I_2$, so treten in $\delta^{-1}(x_1, x_2)$ nur folgende Werte auf:

$$0 \leq k_1 < k_2, \quad 1 \leq k_2 < \infty,$$

d. h. $k_1 = k_2$ kann nicht vorkommen. Denn da stets $(k_1 + x_1)/(k_2 + x_2) < 1$ sein muß, ist für $k_1 = k_2$ doch $x_1 < x_2$ erforderlich. Für den Fall $n > 2$ muß man den Einheitswürfel $B^{(0)}$ in mehr solche Gebiete zerlegen.

Ist nun $E \subset I_i$, $i = 1$ oder 2 , ein meßbarer Bereich, so gilt:

$$\mu(E) = \int_E \int_E \varrho(x_1, x_2) dx_1 dx_2 \Rightarrow \int_E \int_E \varrho(x_1, x_2) dx_1 dx_2 = \int_{\delta^{-1}} \int_E \varrho(x_1, x_2) dx_1 dx_2.$$

Ist nun $E \subset I_1$, so ist dies weiter gleich

$$\sum_{k_1 \geq 1} \sum_{0 \leq k_1 \leq k_2} \int_E \int_E \varrho \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right) \frac{dx_1 dx_2}{(k_2 + x_2)^3};$$

ist $E \subset I_2$, so ist dies weiter gleich:

$$\sum_{k_1 \geq 1} \sum_{0 \leq k_1 < k_2} \int_E \int_E \varrho \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right) \frac{dx_1 dx_2}{(k_2 + x_2)^3}.$$

Demnach gilt für die Dichte ϱ die Funktionalgleichung

$$(2a) \quad \varrho(x_1, x_2) = \sum_{k_1 \geq 1} \sum_{0 \leq k_1 \leq k_2} \varrho \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right) \frac{1}{(k_2 + x_2)^3}, \quad \text{für } (x_1, x_2) \in I_1,$$

$$(2b) \quad \varrho(x_1, x_2) = \sum_{k_1 \geq 1} \sum_{0 \leq k_1 < k_2} \varrho \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right) \frac{1}{(k_2 + x_2)^3}, \quad \text{für } (x_1, x_2) \in I_2.$$

Es sei nun $\sigma(x)$ eine monotone Funktion mit

- (3) 1. $\sigma(x) \rightarrow 0$ für $x \rightarrow \infty$
2. $\text{diam } B^{(\nu)}(x) \leq \sigma(\nu)$.

Da der Jacobische Algorithmus für alle $x \in B^{(0)}$ konvergent ist, geht der Durchmesser $\text{diam } B^{(\nu)}$, für jedes $x \in B^{(0)}$ nach Null, weswegen eine solche Funktion existiert. Es gilt nun folgender Satz:

Satz. Wenn $\varrho(x_1, x_2)$ auf $B^{(0)}$ die Kuzminsche Bedingung erfüllt und $f_0, f_1, f_2, \dots$ eine Folge reeller Funktionen auf $B^{(0)}$ ist, wobei die f_i den Rekursionsgleichungen

$$(4a) \quad f_{n+1}(x_1, x_2) = \sum_{k_2 \geq 1} \sum_{0 \leq k_1 \leq k_2} f_n \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_1}{k_2 + x_2} \right) \frac{1}{(x_2 + k_2)^3} \quad (x_1, x_2) \in I_1$$

$$(4b) \quad f_{n+1}(x_1, x_2) = \sum_{k_2 \geq 1} \sum_{0 \leq k_1 < k_2} f_n \left(\frac{1}{k_2 + x_2}, \frac{k_1 + x_2}{k_2 + x_2} \right) \frac{1}{(k_2 + x_2)^3} \quad (x_1, x_2) \in I_2$$

genügen, ferner die Bedingungen

$$0 < m \leq f_0 \leq M \text{ auf } B^{(0)} \quad \left| \frac{\partial f_0}{\partial x_1} \right| < N, \quad \left| \frac{\partial f_0}{\partial x_2} \right| < N$$

erfüllt sind, so gilt

$$(5) \quad |f_n(x_1, x_2) - a\varrho(x_1, x_2)| < b(e^{-\lambda\sqrt{n}} + \sigma(\sqrt{n}))$$

mit Konstanten a, b und $\lambda > 0$.

Beweis. Der Beweis verläuft wie in [1]. Daher werden nur diejenigen Schritte ausgeführt, wo eine leichte Abweichung auftritt. Zunächst gelten die folgenden Behauptungen:

$$(a) \quad f_n(x_1, x_2) = \sum^{(n)} f_0 \left(\frac{\omega_{i2}^{(n+1)} + \omega_{i1}^{(n)} x_1 + \omega_{i2}^{(n)} x_2}{\omega_{02}^{(n+1)} + \omega_{01}^{(n)} x_1 + \omega_{02}^{(n)} x_2} \right) \cdot \frac{1}{(\omega_{02}^{(n+1)} + \omega_{01}^{(n)} x_1 + \omega_{02}^{(n)} x_2)^3}.$$

Die Summation läuft dabei über die jeweils zulässigen $k^{(1)}, \dots, k^{(n)}$, je nachdem ob (x_1, x_2) aus I_1 oder I_2 ist. Die $\omega_{ij}^{(n)}$ sind rekursiv definierte Größen, deren Bedeutung in früheren Arbeiten auseinandergesetzt wurde. (a) wird durch vollständige Induktion bewiesen. Man beachte:

$$\begin{aligned} \omega_{j2}^{(n+1)} + \omega_{j1}^{(n)} \frac{1}{k_2 + x_2} + \omega_{j2}^{(n)} \frac{k_1 + x_1}{k_2 + x_2} \\ &= \frac{1}{k_2 + x_2} (\omega_{j2}^{(n+1)} k_2 + \omega_{j2}^{(n+1)} x_2 + \omega_{j1}^{(n)} + \omega_{j2}^{(n)} k_1 + \omega_{j2}^{(n)} x_1) \\ &= \frac{1}{k_2 + x_2} (\omega_{j2}^{(n+2)} + \omega_{j1}^{(n+1)} x_1 + \omega_{j2}^{(n+1)} x_2) \\ (b) \quad &\left| \frac{\partial f_n}{\partial x_j} \right| < c_1 N \sigma(n) + c_2 M. \end{aligned}$$

Sei $u_i = \frac{\omega_{i2}^{(n+1)} + \sum \omega_{ij}^{(n)} x_j}{\omega_{02}^{(n+1)} + \sum \omega_{0j}^{(n)} x_j}$, dann ist nach (a):

$$f_n = \sum^{(n)} f_0(u_1, u_2) \Delta_n,$$

wo $\Delta_n = (\omega_{02}^{(n+1)} + \sum \omega_{0j}^{(n)} x_j)^{-3}$ ist.

Da gliedweise Differentiation nach dem folgenden erlaubt ist, schreiben wir

$$\begin{aligned} \frac{\partial f_n}{\partial x_j} &= \sum^{(n)} \left\{ \frac{\partial f_0}{\partial u_1} \cdot \frac{\partial u_1}{\partial x_j} + \frac{\partial f_0}{\partial u_2} \cdot \frac{\partial u_2}{\partial x_j} \right\} \Delta_n + \sum^{(n)} f_0 \frac{\partial \Delta_n}{\partial x_j} \\ \frac{\partial u_i}{\partial x_j} &= \frac{\omega_{ij}^{(n)} - u_i \omega_{0j}^{(n)}}{\omega_{02}^{(n+1)} + \omega_{01}^{(n)} x_1 + \omega_{02}^{(n)} x_2} = \frac{\omega_{0j}^{(n)}}{\omega_{02}^{(n+1)} + \omega_{01}^{(n)} x_1 + \omega_{02}^{(n)} x_2} \left(\frac{\omega_{ij}^{(n)}}{\omega_{0j}^{(n)}} - u_i \right) \\ &\Rightarrow \left| \frac{\partial u_i}{\partial x_j} \right| < \sigma(n); \text{ da } \left| \frac{\partial \Delta_n}{\partial x_j} \right| < 3 |\Delta_n| < c_3 L(B^{(n)}) \text{ ist, denn es ist ja} \\ &\int \int_{\delta^n B^{(n)}} \Delta_n dx_1 dx_2 = L(B^{(n)}) \Rightarrow \left| \frac{\partial f_n}{\partial x_j} \right| < 2c_3 N \sigma(n) + c_3 M, \text{ da } \sum^{(n)} L(B^{(n)}) = 1 \text{ ist.} \end{aligned}$$

(c) Aus $t\varrho(x_1, x_2) < f_n < T\varrho(x_1, x_2)$ folgt $t\varrho(x_1, x_2) < f_{n+1} < T\varrho(x_1, x_2)$. Ist aus (2a), (2b) und (4a), (4b) klar!

$$(d) \quad \iint_{B^{(0)}} f_n(x_1, x_2) dx_1 dx_2 = \iint_{B^{(0)}} f_0(x_1, x_2) dx_1 dx_2$$

Folgt aus (a). —

Aus der Voraussetzung $0 < m \leq f_0 \leq M$ und der Tatsache $K_1 \leq \varrho(x_1, x_2) \leq K_2$ (μ ist zu L äquivalent!) folgt: $g\varrho(x_1, x_2) < f_0(x_1, x_2) < G\varrho(x_1, x_2)$ für konstante g, G . Wir bilden nun die Hilfsfunktionen

$$(6a) \quad \varphi_n = f_n - g\varrho(x_1, x_2)$$

$$(6b) \quad \psi_n = G\varrho(x_1, x_2) - f_n.$$

Es ist

$$\varphi_v(x_1, x_2) = \sum^{(v)} \varphi_0(u_1, u_2) \Delta > c_4 \sum^{(v)} \varphi_0(u_1, u_2) \cdot L(B^{(v)}),$$

denn es ist $c_4 L(B^{(v)}) < \Delta_v$ mit $0 < c_4 \leq 1$ (vgl. [4]).

Nach dem Zwischenwertsatz ist

$$\iint_{B^*} \varphi_0(z_1, z_2) dz_1 dz_2 = \sum^{(v)} \varphi_0(z'_1, z'_2) L(B^{(v)}),$$

wo B^* ein passender Teilbereich von $B^{(0)}$ sei, und demnach

$$\varphi_v - c_4 \iint \varphi_0 dz_1 dz_2 > c_4 \sum \{\varphi_0(z_1, z_2) - \varphi_0(z'_1, z'_2)\} L(B^{(v)}).$$

Nun ist

$$|\varphi_0(z_1, z_2) - \varphi_0(z'_1, z'_2)| < c_5(N + gA)\sigma(v) \text{ für } (z_1, z_2), (z'_1, z'_2) \in B^{(v)}$$

nach dem Mittelwertsatz der Differentialrechnung. An dieser Stelle und nur hier wird die Kuzminsche Voraussetzung verwendet. Man sieht, daß die Lipschitzbedingung

$$|\varrho(z_1, z_2) - \varrho(z'_1, z'_2)| < B \|z - z'\|$$

dasselbe leisten würde! Im ganzen ist daher

$$\varphi_v > c_4 \iint \varphi_0 dz_1 dz_2 - c_6 \sigma(v).$$

Setzt man $c_4 \iint \varphi_0 dz_1 dz_2 = l > 0$, so ist

$$(7a) \quad \varphi_v > l - c_6 \sigma(v).$$

Analog erhält man mit

$$c_4 \iint \psi_0 dz_1 dz_2 = l'$$

die Abschätzung

$$(7b) \quad \psi_v > l' - c_7 \sigma(v).$$

Daraus folgern wir:

$$f_n - g\varrho > l - c_6 \sigma(n),$$

$$f_n > g\varrho + l - c_6 \sigma(n) = \varrho \left(g + \frac{1}{\varrho(x_1, x_2)} - \frac{c_6 \sigma(n)}{\varrho(x_1, x_2)} \right) > \varrho(x_1, x_2) \left(g + \frac{1}{K_2} - \frac{c_6}{K_1} \sigma(n) \right).$$

Für $n \geq n_0$ ist daher $g_1 = g + lK_2^{-1} - c_6 K_1^{-1} \sigma(n) > g$.

Ebenso gilt

$$G\varrho - f_n > l' - c_7\sigma(n) - f_n > \varrho(x_1, x_2) \left(-G + \frac{l'}{\varrho} - \frac{c_7}{\varrho} \sigma(n) \right) f_n < \varrho \left(G - \frac{l'}{K_2} + \frac{c_7}{K_1} \sigma(n) \right)$$

und für $n \geq n_0$

$$G_1 = G - l' K_2^{-1} + c_7 K_1^{-1} \sigma(n) < G.$$

Es ist

$$G_1 - g_1 \leq G - g - (l + l') K_2^{-1} + (c_8 + c_7) K_1^{-1} \sigma(n).$$

Da

$$l + l' = c_4 \iint_{B(0)} (\varphi_0 + \psi_0) dz_1 dz_2 = c_4 \iint_{B(0)} (G - g) \varrho(z_1, z_2) dz_1 dz_2 = c_4 (G - g)$$

ist, so folgt

$$G_1 - g_1 \leq (G - g) (1 - K_2^{-1} c_4) + c_8 \sigma(n);$$

da $0 < c_4 < 1$, $K_2 \geq 1$ gilt, so ist $0 < q = 1 - K_2^{-1} c_4 < 1$.

Demnach ist

$$(8.1) \quad G_1 - g_1 \leq (G - g)q + c_8 \sigma(n).$$

Es war $g\varrho < f_0 < G\varrho$, wobei $\left| \frac{\partial f_0}{\partial x_j} \right| < N$ vorausgesetzt war. Für $n \geq n_0$ gilt $g_1\varrho < f_n < G_1\varrho$, wobei $\left| \frac{\partial f_n}{\partial x_j} \right| < c_1 N \sigma(n) + c_2 M < P$ etwa. Durch Iteration des Vorgangs sieht man (man beachte, daß P von n unabhängig ist!).

$$g_2\varrho < f_{2n} < G_2\varrho$$

mit

$$g_1 < g_2 < G_2 < G_1$$

und

$$(8.2) \quad G_2 - g_2 \leq (G_1 - g_1)q + c_9 \sigma(n),$$

wobei c_9 sich aus c_4 und P etc. analog zu vorher, nun aber für die weiteren Iterationsschritte fest, ergibt.

Durch Iteration gelangt man zu Folgen G_r und g_r , die strikt monoton fallen bzw. wachsen und für die gilt:

$$(8.r) \quad G_r - g_r \leq (G_{r-1} - g_{r-1})q + c_9 \sigma(n) \leq q^r (G - g) + c_9 \sigma(n) (q^{r-1} + q^{r-2} + \dots + 1) \\ \leq q^r (G - g) + \frac{c_9}{1 - q} \sigma(n),$$

das heißt

$$G_r - g_r \leq c_{10} e^{-\lambda r} + c_{11} \sigma(n),$$

woraus $\lim G_r = \lim g_r = a$ folgt.

Letztlich folgt daraus für $r = n$:

$$(9) \quad |f_n - a\varrho| < K_2 (c_{10} e^{-\lambda n} + c_{11} \sigma(n)).$$

Daher ist

$$a = \lim_{n \rightarrow \infty} \iint_{B(0)} f_n(z_1, z_2) dz_1, dz_2.$$

Für $n^2 \leq N < (n+1)^2$ folgt aus (9) mit Hilfe von (c):

$$|f_N - a\varrho| < b(e^{-\lambda/\sqrt{N}} + \sigma(\sqrt{N})) \text{ mit } b = \max(K_2 K_1^{-1} c_{10}, K_2 K_1^{-1} c_{11}).$$

Bemerkung. Da für $n=1$, d. h. für den Fall der gewöhnlichen Kettenbrüche $\sigma(x) = \exp(-(x-1)\log 2)$ brauchbar ist, kann der Term $\sigma(\sqrt{n})$ in den Exponentialterm gezogen werden weswegen im Fall $n=1$ die Aufspaltung in die zwei Terme (exponentieller Anteil und Anteil, der sich aus der Konvergenzgeschwindigkeit des Algorithmus ergibt) nicht zu bemerken ist. Welche Funktionen $\sigma(x)$ für $n \geq 2$ brauchbar sind, ist unbekannt.

Literatur

- [1] A. Ya. Khintchine, Continued Fractions, Groningen 1963 oder: Kettenbrüche, Leipzig 1956.
- [2] F. Schweiger, Ergodische Theorie des Jacobischen Algorithmus, Acta Arithmetica **11** (1966), 451—460.
- [3] F. Schweiger, Existenz eines invarianten Maßes beim Jacobischen Algorithmus, Acta Arithmetica **12** (1967), 263—268.
- [4] F. Schweiger, Mischungseigenschaften und Entropie beim Jacobischen Algorithmus, Journal für Reine und Angewandte Mathematik **229** (1968), 50—56.

Mathematisches Institut der Universität, 1090 Wien, Strudlhofgasse 4

Eingegangen 13. Juli 1967

Über die Grundlagen der Geometrie der Kreise in der pseudo-euklidischen (Minkowskischen) Geometrie

Von *Walter Benz* in Bochum

Hans Rohrbach zum 65. Geburtstag zugeeignet

Inhalt

Einleitung

§ 1. Die Geometrie $A(\mathbb{R})$

§ 2. Das Hyperbel-Modell

§ 3. Die ebenen Schnitte eines Hyperboloids

§ 4. Zwei $(8_3, 6_4)$ -Konfigurationen: Büschelsatz, Satz von Miquel

§ 5. Kongruente, äquiforme Abbildungen. Distanzen, Öffnungen. Pseudo-euklidische Winkel. Orthogonalität

§ 6. Die Automorphismengruppen der Geometrie $A(\mathbb{R})$

§ 7. Eine axiomatische Kennzeichnung der Geometrie $A(\mathbb{R})$.

Literaturverzeichnis

Einleitung

Der Körper $\mathbf{R}$ der reellen Zahlen besitzt drei kommutative Ringerweiterungen vom Rang 2, bei denen die reelle Zahl 1 Ringeins ist: Den Körper $\mathbf{C}$ der *komplexen Zahlen*, den Ring $\mathbf{D}$ der *dualen Zahlen*, den Ring $\mathbf{A}$ der *anormal-komplexen Zahlen*. Zu $\mathbf{C}$ gehört die ebene Möbius-Geometrie $M(\mathbf{R})$, zu $\mathbf{D}$ die ebene Laguerre-Geometrie $L(\mathbf{R})$ und zu $\mathbf{A}$ die zweidimensionale *pseudo-euklidische Kreisgeometrie* $A(\mathbf{R})$. Die Geometrien $M(\mathbf{R})$, $L(\mathbf{R})$, $A(\mathbf{R})$ können nacheinander interpretiert werden als Geometrie der ebenen Schnitte einer Kugel, eines Kegels, eines einschaligen Hyperboloids (s. U. Graf [10]). Dabei werden nur solche ebenen Schnitte zugelassen, die mehr als einen Punkt enthalten und die darüber hinaus im Falle des Kegels keine Gerade tragen und deren Trägerebenen im Falle des einschaligen Hyperboloids nicht selbstkonjugiert sind. Während polymorphe Typen zu den Geometrien M , L mehrfach untersucht und gekennzeichnet wurden (s. die zusammenfassenden Berichte Benz [4], Benz-Mäurer [8]), ist dies für den Fall der Geometrie A bisher unterblieben. Die vorliegende Arbeit beginnt damit, diese Lücke zu schließen.

In § 1 geben wir eine (auch für den klassischen Fall neue) Definition der ebenen pseudo-euklidischen Kreisgeometrie $A(\mathbb{R})$ über einem beliebigen kommutativen Körper $\mathbb{R}$ von Charakteristik $\neq 2$. Die unserer Definition zugrunde liegende homogene Beschrei-

bung der pseudo-euklidischen Geometrie mit Hilfe der projektiven Geraden über dem Ring $\mathfrak{Q}$ der anormal-komplexen Zahlen über $\mathfrak{K}$ ist ein Hauptresultat der vorliegenden Arbeit. Eine inhomogene Beschreibung der klassischen pseudo-euklidischen Geometrie mit Hilfe des Ringes $\mathcal{A}$ hat schon U. Graf [10] gegeben; natürlich erfaßt er hierbei unendlich viele Punkte des einschaligen Hyperboloids nicht. Während der Begriff der projektiven Geraden über einem lokalen Ring (wie er in der Laguerre-Geometrie gebraucht wird) nahe lag nach ringgeometrischen Untersuchungen von Hjelmslev [11] und Klingenberg [13], ergab sich der von uns benötigte Begriff der projektiven Geraden über einem (nicht notwendig lokalen) Ring aus ringgeometrischen Untersuchungen von D. Barbilian [1], [2]: Geordnete Paare (x_1, x_2) über dem kommutativen Ring $\mathfrak{Q}$ mit Einselement werden als Punkte angesprochen, bei denen das durch x_1, x_2 erzeugte Ideal der Ring $\mathfrak{Q}$ ist.

In § 1 werden neben dem bereits Geschilderten noch Transitivitäts- und Verknüpfungseigenschaften der Geometrie $\mathcal{A}(\mathfrak{K})$ bewiesen. Man hat hier zwei Parallelitätsrelationen auf der Menge der Punkte zu betrachten: Wir sprechen von Plus-Parallelität und Minus-Parallelität. H. Beck hat in [3] ein sehr schönes Modell der Geometrie der Kreise in der Minkowskischen Ebene angegeben; es handelt sich um die Geometrie der orientierten Geraden (Pfeile) und der Zykel (orientierte Kreise, Abstandskurven, Grenzkurven) der hyperbolischen nichteuklidischen Geometrie. Er spricht dabei von α -Parallelität und ω -Parallelität.

Da die in § 1 gegebene Definition der pseudo-euklidischen Geometrie, spezialisiert auf den Fall $\mathfrak{K} = \mathbf{R}$, nicht die klassische Definition ist, zeigen wir in § 2 die Isomorphie von $\mathcal{A}(\mathfrak{K})$ zum — geeignet abgeschlossenen — verallgemeinerten klassischen Modell. Interesse beansprucht der hier vorgeschlagene uneigentliche Abschluß der pseudo-euklidischen Ebene. Einen uneigentlichen Abschluß der klassischen pseudo-euklidischen Ebene gibt schon L. Tamassy [20] an.

In § 3 betrachten wir ein weiteres Modell der pseudo-euklidischen Geometrie über einem beliebigen Körper von Char $\mathfrak{K} \neq 2$, das der Punkte und ebenen Schnitte eines einschaligen Hyperboloids im dreidimensionalen projektiven Raum über dem Körper $\mathfrak{K}$.

In § 4 studieren wir zwei $(8_3, 6_4)$ -Konfigurationen, die Büschelsatz-Konfiguration und die Miquel-Konfiguration. Der Satz von Miquel ist eine spezielle Aussage über pseudo-euklidische Winkel. Diese werden allgemein in § 5 eingeführt und untersucht.

In § 6 geben wir die Automorphismengruppen der Geometrien $\mathcal{A}(\mathfrak{K})$ an. Es handelt sich also hier insbesondere um den Nachweis des von Staudtschen Theorems der pseudo-euklidischen Geometrie.

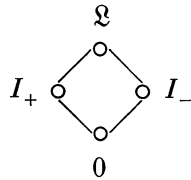
Der § 7 ist einer Axiomatik der ebenen pseudo-euklidischen Geometrie $\mathcal{A}(\mathfrak{K})$ vorbehalten. Mit dem axiomatisch gefaßten Begriff der pseudo-euklidischen Ebene wird gezeigt, daß die Klasse der Geometrien $\mathcal{A}(\mathfrak{K})$ identisch ist mit der Klasse der pseudo-euklidischen Ebenen. Merkwürdigerweise wird unter den einfachen geometrischen Eigenschaften, die die Algebraisierung erzwingen, nicht der Berührsatz (Eigenschaft (4) in Satz 1.3) benötigt. Er kann also erschlossen werden. Damit ist ein Studium von Nicht-Berührsatz-Geometrien (pseudo-euklidischer Natur) nur unter Verzicht auf natürliche geometrische Eigenschaften der pseudo-euklidischen Geometrie möglich. Insofern nimmt die pseudo-euklidische Geometrie eine Sonderstellung neben den Geometrien von Möbius und Laguerre ein (s. Benz [4], Benz-Mäurer [8]). Natürlich liegt der tiefere Grund für diese Sonderstellung darin, daß die beiden Parallelitätsrelationen der pseudo-euklidischen Geometrie den Rang 2-Charakter der Geometrie bedingen.

§ 1. Die Geometrie $A(\mathfrak{K})$

Sei $\mathfrak{K}$ ein beliebiger kommutativer Körper, sei $\mathfrak{L}$ der Ring der anormal-komplexen Zahlen über $\mathfrak{K}$, d. h. die Menge der Matrizen $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$, $a, b \in \mathfrak{K}$, unter Zugrundelegung von Matrizenaddition und -multiplikation. $\mathfrak{K}$ ist in der Form

$$\left\{ \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \mid k \in \mathfrak{K} \right\}^1) \quad \text{in } \mathfrak{L}$$

isomorph eingebettet. Identifizieren wir das Element $k \in \mathfrak{K}$ mit $\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \in \mathfrak{L}$ und benutzen wir für das Element $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathfrak{L}$ die Abkürzung j , so haben die anormal-komplexen Zahlen über $\mathfrak{K}$ die Gestalt $a + bj$, $a, b \in \mathfrak{K}$, wo j die anormal-komplexe Einheit mit $j^2 = 1$ ist. Der Idealverband von $\mathfrak{L}$ hat das folgende Aussehen



im Falle, daß die Charakteristik von $\mathfrak{K}$ von 2 verschieden ist. Dabei ist

$$I_+ = \mathfrak{K}(1 + j) = \{k(1 + j) \mid k \in \mathfrak{K}\}$$

und $I_- = \mathfrak{K}(1 - j)$ gesetzt. Ist $\text{Char } \mathfrak{K} = 2$, so ist $\mathfrak{L}$ ein lokaler Ring und isomorph zum Ring der Clifford-Zahlen über $\mathfrak{K}$. Die nachfolgend gegebene Definition der Geometrie $A(\mathfrak{K})$ reduziert sich in diesem Fall auf die der Laguerre-Geometrie über dem Körper $\mathfrak{K}$. Wir schließen deshalb ein für allemal für die vorliegenden Untersuchungen den Fall $\text{Char } \mathfrak{K} = 2$ aus. Ist $\mathfrak{K}$ die Einheitengruppe von $\mathfrak{L}$, so hat man also die Zerlegung

$$\mathfrak{L} = \mathfrak{K} \cup I_+ \cup I_- \quad \text{mit } \mathfrak{K} \cap (I_+ \cup I_-) = \emptyset \quad \text{und } I_+ \cap I_- = \{0\}.$$

Ferner stellt $I_+ \cdot I_-$ das Nullideal dar.

Wir nennen jedes geordnete Paar (u, v) , $u, v \in \mathfrak{L}$ einen *pseudo-euklidischen Punkt*, wenn nicht beide Elemente u, v im selben Ideal I_+ oder I_- liegen. Die Punkte (u, v) , (u', v') heißen genau dann *gleich*, wenn es eine Einheit $r \in \mathfrak{K}$ mit $u' = ru$, $v' = rv$ gibt. Die Punkte (u, v) , (u', v') heißen zueinander *p-parallel* („plus-parallel“), in Zeichen $(u, v) \parallel_+ (u', v')$, wenn

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix} \in I_+ \quad \text{ist,}$$

und sie heißen *m-parallel* („minus-parallel“), in Zeichen $(u, v) \parallel_- (u', v')$, wenn

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix} \in I_- \quad \text{gilt.}$$

Diese Parallelitätsdefinitionen hängen, da I_+ , I_- Ideale sind, nicht von dem gewählten Repräsentanten (ru, rv) , $r \in \mathfrak{K}$, des Punktes (u, v) und dem gewählten Repräsentanten $(r'u', r'v')$, $r' \in \mathfrak{K}$, des Punktes (u', v') ab.

¹⁾ Die Menge der Elemente g einer Grundmenge $\mathfrak{G}$ mit der Eigenschaft $\mathfrak{E}$ schreiben wir in der Gestalt $\{g \mid \mathfrak{E}\}$.

Hilfssatz 1. 1. Die Relationen $||_+$, $||_-$ sind Äquivalenzrelationen auf der Menge $\mathfrak{P}$ der pseudo-euklidischen Punkte.

Beweis. Dies zeigen wir für den Fall der p -Parallelität: Reflexivität und Symmetrie liegen auf der Hand. Zur Transitivität:

Sei $(u, v) ||_+ (u', v')$, $(u', v') ||_+ (u'', v'')$. Sei $D = \begin{vmatrix} u & v \\ u'' & v'' \end{vmatrix}$. Dann gilt jedenfalls

$$(1) \quad u'D = u' \begin{vmatrix} u & v \\ u'' & v'' \end{vmatrix} = u'' \begin{vmatrix} u & v \\ u' & v' \end{vmatrix} + u \begin{vmatrix} u' & v' \\ u'' & v'' \end{vmatrix} \in I_+$$

und

$$(2) \quad v'D = v' \begin{vmatrix} u & v \\ u'' & v'' \end{vmatrix} = v'' \begin{vmatrix} u & v \\ u' & v' \end{vmatrix} + v \begin{vmatrix} u' & v' \\ u'' & v'' \end{vmatrix} \in I_+.$$

1. Fall: u' regulär oder v' regulär. Aus (1) resp. (2) folgt dann $D \in I_+$. 2. Fall: u' irregulär und v' irregulär. Da (u', v') ein Punkt ist, liegen die Elemente u' , v' nicht im gleichen Ideal I_+ oder I_- und wegen $\mathfrak{L} = \mathfrak{R} \cup I_+ \cup I_-$ liegt also genau eines der Elemente u' , v' in I_+ und genau eines in I_- . Sei das in I_- liegende Element mit a bezeichnet. Wegen $a \notin I_+$ ist $a \neq 0$. Aus $a \in I_-$ folgt $aD \in I_-$. Mit (1) resp. (2) hat man also $aD = 0$, wenn man $I_+ \cap I_- = \{0\}$ beachtet. Aus $aD = 0$, $a \neq 0$, $a \in I_-$ folgt $(1 - j)D = 0$, d. h. $D \in I_+$.

Wir betrachten die Substitutionen

$$(3) \quad \begin{aligned} \varrho x' &= a_{11}x + a_{12}y \\ \varrho y' &= a_{21}x + a_{22}y \end{aligned}$$

von $\mathfrak{P}$, wobei ϱ , a_{ij} Elemente von $\mathfrak{L}$ sein sollen und ϱ , $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ als Einheiten vorausgesetzt werden.

Hilfssatz 1. 2. Die Substitutionen (3) sind 1,1-deutige Abbildungen von $\mathfrak{P}$ auf sich, die die Relationen $||_+$, $||_-$ in beiden Richtungen erhalten. Hintereinanderschaltung als Verknüpfung genommen, bildet die Gesamtheit der Abbildungen (3) eine Gruppe, die wir im folgenden mit Γ bezeichnen.

Beweis. Wir können uns auf die Angabe weniger Hinweise beschränken: Gegeben sei eine Substitution (3). Gehören x , y nicht beide demselben Ideal I_+ oder I_- an, so gilt dies auch für das Paar x' , y' : Da nämlich $|| a_{ij} ||$ regulär ist, kann man x und y linear und homogen in x' , y' ausdrücken; dies würde aber bedeuten, daß x und y ebenfalls dem gleichen Ideal I_+ oder I_- angehören müßten, wenn dies für das Paar x' , y' der Fall wäre. Verschiedene Repräsentanten desselben Punktes führen zum gleichen Bildpunkt. — Zur Invarianz der Relationen $||_+$, $||_-$ in beiden Richtungen: Da die Umkehrabbildung von (3) ebenfalls von der Gestalt (3) ist, genügt es zu zeigen, daß aus

$$(x_1, y_1) ||_+ (x_2, y_2) \text{ die Aussage } (x'_1, y'_1) ||_+ (x'_2, y'_2)$$

folgt. (Das Analoge gilt für $||_-$.)

Es ist aber

$$\varrho_1 \varrho_2 \begin{vmatrix} x'_1 & x'_2 \\ y'_1 & y'_2 \end{vmatrix} = \begin{vmatrix} \varrho_1 x'_1 & \varrho_2 x'_2 \\ \varrho_1 y'_1 & \varrho_2 y'_2 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix},$$

was sofort in beiden Fällen $||_+$, $||_-$ die Behauptung ergibt. Der Punkt P heiße *nicht-parallel* zum Punkt Q , in Zeichen $P \# Q$, wenn weder $P ||_+ Q$ noch $P ||_- Q$ gilt. Umgekehrt sollen die Punkte P , Q *parallel* heißen, wenn sie p -parallel oder m -parallel sind. Dann gilt:

Satz 1.1. Sind A, B, C paarweise nicht parallele Punkte und ebenso A', B', C' , so gibt es genau ein $\tau \in T$ mit $A^\tau = A', B^\tau = B', C^\tau = C'$.

Beweis. Seien $A, A', \dots$ sukzessive gegeben durch $(a_1, a_2), (a'_1, a'_2), \dots$. Kürzen wir $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ durch $[a, b]$ ab usf., so gilt also

$$[a, b], [b, c], [c, a], [a', b'], [b', c'], [c', a'] \notin I_+ \cup I_-.$$

Sei τ_0 die Transformation

$$\mathfrak{x}' = \mathfrak{A}\mathfrak{x}, \quad \mathfrak{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathfrak{x}' = \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}, \quad \mathfrak{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

mit

$$\mathfrak{A} = \begin{pmatrix} a'_1 & b'_1 \\ a'_2 & b'_2 \end{pmatrix} \begin{pmatrix} \frac{[b', c']}{[b, c]} & 0 \\ 0 & -\frac{[c', a']}{[c, a]} \end{pmatrix} \begin{pmatrix} b_2 - b_1 \\ a_2 - a_1 \end{pmatrix},$$

die wegen

$$|\mathfrak{A}| = [a', b'] \frac{[b', c'] [c', a']}{[b, c] [c, a]} [a, b] \in \mathfrak{R}$$

zu T gehört. Es gilt $A^{\tau_0} = A', B^{\tau_0} = B', C^{\tau_0} = C'$. Wir kommen zur Eindeutigkeit der Abbildung: Seien τ, σ Elemente aus T mit $A^\tau = A' = A^\sigma, B^\tau = B' = B^\sigma, C^\tau = C' = C^\sigma$. Sei $\alpha \in T$ eine — aufgrund des schon bewiesenen Teiles des Satzes 1.1 vorhandene — Abbildung mit $(1, 0)^\alpha = A, (0, 1)^\alpha = B, (1, 1)^\alpha = C$. Dann sind $(1, 0), (0, 1), (1, 1)$ Fixpunkte der Abbildung $\gamma = \alpha\tau\sigma^{-1}\alpha^{-1}$. Ist γ gegeben durch

$$\varrho x' = c_{11}x + c_{12}y,$$

$$\varrho y' = c_{21}x + c_{22}y,$$

so hat man also $c_{21} = c_{12} = 0, c_{11} = c_{22}$. Es ist demnach γ die Identität, was $\tau = \sigma$ ergibt.

Die Menge $\mathfrak{p}$ der Punkte der projektiven Geraden über dem Körper $\mathfrak{R}$ ist in $\mathfrak{P}$ enthalten.

Die Bilder von $\mathfrak{p}$ unter Abbildungen aus T nennen wir *pseudo-euklidische Kreise*, kurz *Kreise*.

Unter einer *Kreisverwandtschaft* wird eine 1,1-deutige Abbildung von $\mathfrak{P}$ auf sich verstanden, die in beiden Richtungen Kreise in Kreise überführt.

Satz 1.2. Die Abbildungen aus T sind Kreisverwandtschaften. Sie erhalten darüber hinaus in beiden Richtungen die Relationen $||_+$ und $||_-$.

Beweis. Wegen Hilfssatz 1.2 haben wir nur noch zu zeigen, daß die τ aus T in beiden Richtungen Kreise in Kreise überführen. Da T natürlich gegen Inversenbildung abgeschlossen ist, haben wir nur zu zeigen, daß die τ aus T Kreise in Kreise überführen: Dies liegt aber auf der Hand, da aus $\mathfrak{p}^\sigma, \sigma \in T$, unter τ der Kreis $\mathfrak{p}^{(\sigma\tau)}$ entsteht.

Ist P ein Punkt, ist k ein Kreis mit $P \in k$, so verwenden wir die üblichen Redeweisen: k geht durch P , P liegt auf k usf. Die Punktmenge $\{P, Q, \dots\}$ heißt *konzyklisch*, wenn es einen pseudo-euklidischen Kreis gibt, der alle Punkte der Menge enthält.

Es gilt der

- Satz 1. 3.** (1) *Durch drei paarweise nicht parallele Punkte geht genau ein Kreis²⁾.*
 (2) *Zwei verschiedene Punkte eines Kreises sind nicht parallel.*
 (2') *Sind S, T nicht parallele Punkte, so gibt es wenigstens einen Kreis durch S, T .*
 (3) *Jeder Kreis enthält wenigstens 4 verschiedene Punkte. Nicht alle Punkte liegen auf einem einzigen Kreis. Es gibt wenigstens einen Kreis.*
 (4) *Ist k ein Kreis, sind P, Q nicht parallele Punkte mit $P \in k, Q \notin k$, so gibt es genau einen Kreis k' durch P, Q , der mit k nur den Punkt P gemeinsam hat.*
 (5) *Ist k ein Kreis, ist P ein Punkt, so gibt es genau einen Punkt P' auf k mit $P' \parallel_+ P$ und genau einen Punkt P'' auf k mit $P'' \parallel_- P$.*

Beweis. Zu (1): Seien A, B, C paarweise nicht parallele Punkte. Sei $\tau \in T$ die Abbildung (s. Satz 1. 1.) mit $(1, 0)^\tau = A, (0, 1)^\tau = B, (1, 1)^\tau = C$. Dann geht offenbar durch A, B, C der Kreis p^τ .

Sei $p^\sigma, \sigma \in T$, ein weiterer Kreis durch A, B, C . Die Punkte $A^{\sigma^{-1}}, B^{\sigma^{-1}}, C^{\sigma^{-1}}$ auf p sind nach Hilfssatz 1. 2 paarweise nicht parallel.

Sei $\alpha \in T$ die Abbildung mit $(1, 0)^\alpha = A^{\sigma^{-1}}, (0, 1)^\alpha = B^{\sigma^{-1}}, (1, 1)^\alpha = C^{\sigma^{-1}}$. Dies leistet aber schon eine Abbildung α' aus $PGL(2, \mathbb{R})$, die natürlich — fortgesetzt — zu T gehört und also nach Satz 1. 1. mit α übereinstimmen muß. Für α gilt dann $p^\alpha = p$. Weiterhin folgt $\alpha\sigma = \tau$. Also ist $p^\tau = p^{(\alpha\sigma)} = (p^\alpha)^\sigma = p^\sigma$.

Zu (2): Die Behauptung folgt unmittelbar aus Hilfssatz 1. 2, wenn man beachtet, daß natürlich je zwei verschiedene Punkte von p nicht parallel sind.

Zu (2'): Sind $S = (s_1, s_2), T = (t_1, t_2)$ nicht parallel, so stellt

$$\sigma: \begin{aligned} x'_1 &= s_1x_1 + t_1x_2 \\ x'_2 &= s_2x_1 + t_2x_2 \end{aligned}$$

eine Abbildung aus T dar. Damit geht durch S, T z. B. der Kreis p^σ .

Zu (3): Wegen $\text{Char } \mathbb{R} \neq 2$ enthält p wenigstens die verschiedenen Punkte $(1, 0), (0, 1), (1, 1), (1 + 1, 1)$. Also enthält auch jedes $p^\tau, \tau \in T$, wenigstens vier verschiedene Punkte. Würde der Kreis $p^\tau, \tau \in T$, alle Punkte enthalten, also insbesondere die Punkte $(1, 0), (0, 1), (1, 1)$, so wäre nach (1) $p^\tau = p$. Der Kreis p enthält aber nicht den Punkt $(j, 1)$.

Zu (4): Wegen des schon bewiesenen Teils (1), (2), (3) und der Sätze 1. 1., 1. 2. können wir folgende spezielle Situation annehmen: P besitzt die Koordinaten $(1, 0)$, k ist der Kreis p . Es habe Q die Koordinaten (q_1, q_2) . Wegen $Q \nparallel_+ P, Q \nparallel_- P$ gilt $q_2 \in \mathbb{R}$.

Dann betrachten wir für Q die Koordinaten $(q, 1), q = \frac{q_1}{q_2}$. Wegen $Q \notin p$ ist $q \notin \mathbb{R}$; sei $q = \alpha + \beta j, \alpha, \beta \in \mathbb{R}$ mit also $\beta \neq 0$. Wir betrachten den nach (1) 1-deutig bestimmten Kreis k'_0 durch die paarweise nicht parallelen Punkte $(1, 0), (q, 1), (q + 1, 1)$. Da für das folgende τ aus T

$$\tau: \begin{aligned} x'_1 &= x_1 + qx_2 \\ x'_2 &= x_2 \end{aligned}$$

ersichtlich $(1, 0)^\tau = (1, 0), (0, 1)^\tau = (q, 1), (1, 1)^\tau = (q + 1, 1)$ gilt, so ist

$$k'_0 = p^\tau = \{(\lambda_1 + q\lambda_2, \lambda_2) \mid \lambda_1, \lambda_2 \in \mathbb{R}, \text{ nicht beide } 0\}.$$

²⁾ Sind A, B, C paarweise nicht parallele Punkte, so bezeichnen wir den eindeutig bestimmten Kreis durch A, B, C mit (ABC) .

Damit ist $k'_0 - \{P\} = \{(\lambda + q, 1) \mid \lambda \in \mathfrak{R}\}$. Es gilt $k'_0 - \{P\} \cap k = \emptyset$: Wegen $q \notin \mathfrak{R}$, $\lambda \in \mathfrak{R}$ ist nämlich stets $\lambda + q \notin \mathfrak{R}$. — Angenommen, es gäbe einen weiteren Kreis $k'_1 \neq k'_0$ durch P, Q , der mit k nur den Punkt P gemeinsam hat. Sei A ein nach (3) vorhandener Punkt $\neq P, Q$ auf k'_1 . Nach (2) sind dann die Punkte P, Q, A paarweise nicht parallel. Wir betrachten die Abbildung $\sigma \in \mathbb{T}$ mit $(1, 0)^\sigma = (1, 0)$, $(0, 1)^\sigma = (q, 1)$, $(1, 1)^\sigma = (a, 1)$. Es seien hierbei $(a, 1)$ die Koordinaten von A : Wegen $A \#_+ P$, $A \#_- P$ ist nämlich die 2. Komponente von A regulär. Da A nicht parallel zu Q ist, folgt noch $a - q \in \mathfrak{R}$. Offenbar ist $\mathfrak{p}^\sigma = k'_1$. Es hat σ die Form:

$$\sigma: \begin{aligned} x'_1 &= (a - q)x_1 + qx_2 \\ x'_2 &= x_2. \end{aligned}$$

Damit ist $k'_1 - \{P\} = \{((a - q)\lambda + q, 1) \mid \lambda \in \mathfrak{R}\}$.

Wegen $k'_1 - \{P\} \cap k = \emptyset$ muß $(a - q)\lambda + q \notin \mathfrak{R}$ sein für alle λ aus $\mathfrak{R}$. Setzen wir noch $a = \gamma + \delta j$, $\gamma, \delta \in \mathfrak{R}$, so haben wir also

$$(\delta - \beta)\lambda + \beta \neq 0 \text{ für alle } \lambda \text{ aus } \mathfrak{R},$$

was auf $\delta = \beta$, d. h. $a - q \in \mathfrak{R}$, führt und wegen $a - q \in \mathfrak{R}$ dann $0 \neq a - q \in \mathfrak{R}$ ergibt. Jetzt ist aber $\left(\frac{1}{a - q}, 1\right)^\sigma = (1 + q, 1)$. Also ist $(1 + q, 1) \in k'_1$, was doch $k'_1 = k'_0$ ergäbe.

Zu (5): Wir können die spezielle Situation $k = \mathfrak{p}$ annehmen. Im Falle $P \in k$ ist nach (2) nichts zu zeigen. Sei $P \notin k$. Sei $P = (p_1, p_2)$. Ist $p_i = \alpha_i + \beta_i j$, $\alpha_i, \beta_i \in \mathfrak{R}$, $i = 1, 2$, so folgt aus $P \notin k$ jedenfalls $\gamma = \begin{vmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{vmatrix} \neq 0$. Die Punkte $P'_0 = (\alpha_1 - \beta_1, \alpha_2 - \beta_2)$ (wegen $\gamma \neq 0$ kann nicht gleichzeitig $\alpha_1 - \beta_1 = 0$ und $\alpha_2 - \beta_2 = 0$ sein),

$$P''_0 = (\alpha_1 + \beta_1, \alpha_2 + \beta_2)$$

auf k haben die Eigenschaft $P'_0 \parallel_+ P$, $P''_0 \parallel_- P$. Gäbe es noch einen Punkt P'_1 auf k mit $P'_1 \parallel_+ P$, so würde Hilfssatz 1. 1. auf $P'_0 \parallel_+ P'_1$ führen, was wegen (2) $P'_0 = P'_1$ ergäbe.

Mit Hilfe von Satz 1. 3 erhalten wir den

Satz 1. 4. *Eine Kreisverwandtschaft κ überführt in beiden Richtungen parallele Punkte in parallele Punkte.*

Beweis. Sind A, B verschiedene und parallele Punkte, so müssen A^*, B^* ebenfalls parallel sein: Im anderen Falle ginge nach Satz 1. 3. (2') durch A^*, B^* ein Kreis k . Dann würde $k^{\kappa^{-1}}$ die Punkte A, B enthalten, was Satz 1. 3. (2) widerspricht. Daß ebenfalls $A^{\kappa^{-1}}, B^{\kappa^{-1}}$ parallel sein müssen, folgt aus dem gerade Bewiesenen und der Tatsache, daß auch κ^{-1} eine Kreisverwandtschaft ist.

§ 2. Das Hyperbel-Modell

Sei $\mathfrak{R}$ ein kommutativer Körper ($\text{Char } \mathfrak{R} \neq 2$). Jedes geordnete Paar (k_1, k_2) von Elementen k_1, k_2 aus $\mathfrak{R}$ heiße ein *eigentlicher pseudo-euklidischer Punkt* (oder auch nur kurz ein *eigentlicher Punkt*). Verschiedene Paare (a_1, a_2) , (b_1, b_2) sollen verschiedene eigentliche Punkte sein. Wir kommen zur Definition der *uneigentlichen pseudo-euklidischen Punkte*: Das Symbol ∞ heiße ein *uneigentlicher Punkt*. Die weiteren (von ∞ und untereinander verschiedenen) *uneigentlichen pseudo-euklidischen Punkte* sollen die Geraden $\{(x, y) \mid y = x + u\}$, $\{(x, y) \mid y = -x + u\}$ für alle u aus $\mathfrak{R}$ sein. Sind a, b, c, d Elemente aus $\mathfrak{R}$ mit $b^2 - c^2 \neq 4ad$, so heiße

$$(*) \quad \{(x, y) \mid a(x^2 - y^2) + bx + cy + d = 0\}$$

ein *pseudo-euklidischer Kreis*. Dabei soll $(*)$ im Falle $a = 0$ noch den uneigentlichen Punkt ∞ enthalten und im Falle $a \neq 0$ die beiden uneigentlichen Punkte $y + x = \frac{c-b}{2a}$, $y - x = \frac{c+b}{2a}$, die im klassischen Fall die Asymptoten an $(*)$ darstellen³⁾.

Es werden also alle Geraden (als Punktmengen aufgefaßt) pseudo-euklidische Kreise genannt, sofern sie nicht zur Geraden $y = x$ oder zur Geraden $y = -x$ parallel sind; dabei ist stets noch der uneigentliche Punkt ∞ hinzuzufügen. Weiterhin sind alle Hyperbeln (als Punktmengen aufgefaßt) der Gleichung $(x + \alpha)^2 - (y + \beta)^2 = \gamma \neq 0$ pseudo-euklidische Kreise; dabei sind noch die beiden Asymptoten als uneigentliche Punkte hinzuzufügen.

Um den Anschluß an § 1 zu bekommen, ordnen wir den pseudo-euklidischen Punkten homogene Koordinaten zu⁴⁾: Dem eigentlichen Punkt (a, b) , $a, b \in \mathbb{R}$, werden die Paare $(\varrho(a + bj), \varrho)$, $\varrho \in \mathbb{R}$, zugeordnet und dem uneigentlichen Punkt ∞ die Paare $(\varrho, 0)$, $\varrho \in \mathbb{R}$. Der uneigentliche Punkt, repräsentiert durch die Gerade $y + x = u$, erhält die homogenen Koordinaten $\varrho((u + 1) + (u - 1)j, 1 + j)$, $\varrho \in \mathbb{R}$, und der uneigentliche Punkt $y - x = u$ die Koordinaten $\varrho((1 - u) + (1 + u)j, 1 - j)$, $\varrho \in \mathbb{R}$. Die beiden letzten Zuordnungen lassen sich in der folgenden (geringfügig abgeänderten) Form zusammenfassen: Der uneigentliche Punkt $x \pm y = v$ erhält die Koordinaten $\varrho((v + 1) \pm (v - 1)j, 1 \pm j)$, wobei überall, wo $\pm$ steht, dasselbe Zeichen zu nehmen ist.

Damit gilt

Hilfssatz 2.1. *Die angegebene Zuordnung σ stellt eine 1,1-deutige Abbildung der Menge $\mathfrak{E}$ der pseudo-euklidischen Punkte auf die Menge $\mathfrak{P}$ des § 1 dar.*

Beweis. Da für die aufgeschriebenen Paare $(a + bj, 1)$, $(1, 0)$, $(u + 1 + (u - 1)j, 1 + j)$, $(1 - u + (u + 1)j, 1 - j)$ die Komponenten nicht im selben Ideal I_+ oder I_- liegen, stellt σ eine 1-deutige Abbildung von $\mathfrak{E}$ in $\mathfrak{P}$ dar. Jedes Element aus $\mathfrak{P}$ tritt als Bild genau eines Elementes aus $\mathfrak{E}$ auf: Sei $(s_1, s_2) \in \mathfrak{P}$.

1. Fall: $s_2 \in \mathbb{R}$. Hier hat man als einziges Urbild den Punkt $\left(Re \frac{s_1}{s_2}, An \frac{s_1}{s_2}\right)$ ⁵⁾.
2. Fall: $s_2 = 0$. Hier ist $s_1 \in \mathbb{R}$. Als Urbild hat man den Punkt ∞ .
3. Fall: $s_2 \in I_+ - \{0\}$. Für (s_1, s_2) können wir dann zu Koordinaten $(a + bj, 1 + j)$, $a, b \in \mathbb{R}$, $a \neq b$, übergehen. In $x + y = \frac{a+b}{2}$ haben wir ein Urbild: Denn es ist

$$(a + bj, 1 + j) = \left(\left(\frac{1}{2} + \frac{a-b}{4} \right) + \left(\frac{1}{2} - \frac{a-b}{4} \right) j \right) \cdot \left(\left(\frac{a+b}{2} + 1 \right) + \left(\frac{a+b}{2} - 1 \right) j, 1 + j \right),$$

wobei wir beachten, daß

$$\left(\frac{1}{2} + \frac{a-b}{4} \right) + \left(\frac{1}{2} - \frac{a-b}{4} \right) j$$

regulär ist wegen $a \neq b$. $x + y = \frac{a+b}{2}$ ist das einzige Urbild: Jedenfalls können ein eigentlicher Punkt und auch ∞ kein Urbild sein. Ist $x + y = u$, $u \in \mathbb{R}$, ein Urbild, so muß

$$\varrho((u + 1) + (u - 1)j, 1 + j) = (a + bj, 1 + j), \varrho \in \mathbb{R},$$

³⁾ Auch im allgemeinen Fall wollen wir diese beiden uneigentlichen Punkte die *Asymptoten* von $(*)$ nennen.

⁴⁾ Unter $\mathfrak{L}$ wird hier wieder der Ring der anormal-komplexen Zahlen über $\mathbb{R}$ verstanden. Wir übernehmen auch die anderen Bezeichnungen wie $\mathbb{R}$, $\mathfrak{P}$, I_+ usw.

⁵⁾ Ist $a, b \in \mathbb{R}$, so schreiben wir $Re(a + bj) = a$ und $An(a + bj) = b$.

gelten. Also ist ϱ von der Form $\alpha + (1 - \alpha)j$ mit $\alpha \in \mathfrak{R}$ und $\alpha \neq \frac{1}{2}$. Dies ergibt $(\alpha + (1 - \alpha)j) \cdot ((u + 1) + (u - 1)j) = a + bj$, was $u = \frac{a+b}{2}$ und $\alpha \neq \frac{1}{2}$ wegen $a \neq b$ zur Folge hat. Wäre $y - x = u$ ein Urbild, so müßte

$$\varrho((1 - u) + (1 + u)j, 1 - j) = (a + bj, 1 + j), \varrho \in \mathfrak{R},$$

sein, was wegen $\varrho(1 - j) \in I_- - \{0\}$ und $1 + j \in I_+$ nicht geht.

4. Fall: $s_2 \in I_- - \{0\}$. Für (s_1, s_2) können wir zu Koordinaten $(a + bj, 1 - j)$, $a, b \in \mathfrak{R}$, $a \neq -b$, übergehen. In $y - x = \frac{b-a}{2}$ haben wir das einzige Urbild: Es ist $(a + bj, 1 - j)$

$$= \left(\left(\frac{1}{2} + \frac{a+b}{4} \right) - \left(\frac{1}{2} - \frac{a+b}{4} \right) j \right) \left(\left(1 - \frac{b-a}{2} \right) + \left(1 + \frac{b-a}{2} \right) j, 1 - j \right).$$

Bemerkung. Die Zuordnung σ kann auch in der folgenden Form (unter Benutzung von Normalformen) wiedergegeben werden:

$$(a, b) \rightarrow (a + bj, 1)$$

$$\infty \rightarrow (1, 0)$$

$$x + \varepsilon y = 0 \rightarrow (1 - \varepsilon j, 1 + \varepsilon j)$$

$$x + \varepsilon y = u \neq 0 \rightarrow (2u, 1 + \varepsilon j), \varepsilon \in \{+1, -1\}.$$

Mit $z \rightarrow \bar{z}$ bezeichnen wir den involutorischen Automorphismus $a + bj \rightarrow a - bj$, $a, b \in \mathfrak{R}$, von $\mathfrak{L}$. Ist $\mathfrak{A} = (a_{\mu\nu})$ eine Matrix über $\mathfrak{L}$, so bezeichnet $\mathfrak{A}^T$ die zu $\mathfrak{A}$ transponierte Matrix und $\bar{\mathfrak{A}}$ die Matrix $(\bar{a}_{\mu\nu})$.

Der folgende Satz 2.1 gibt eine Darstellung der pseudo-euklidischen Kreise, der auch die uneigentlichen Punkte des betreffenden Kreises genügen.

Satz 2.1. Sind a, b, c, d Elemente aus $\mathfrak{R}$ mit $ad - \frac{b^2 - c^2}{4} \neq 0$, so ist die Menge der Punkte (z_1, z_2) ⁶⁾, die der Gleichung

$$\mathfrak{z}^T \begin{pmatrix} a & \frac{b+cj}{2} \\ \frac{b-cj}{2} & d \end{pmatrix} \bar{\mathfrak{z}} = 0, \quad \mathfrak{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

genügen, genau die Menge der Punkte (mit Einschluß der uneigentlichen) des Kreises

$$a(x^2 - y^2) + bx + cy + d = 0.$$

Beweis. Wir betrachten das Lösungsgebilde Λ von

$$(*) \quad az_1 \bar{z}_1 + \frac{b+cj}{2} z_1 \bar{z}_2 + \frac{b-cj}{2} \bar{z}_1 z_2 + dz_2 \bar{z}_2 = 0.$$

An eigentlichen Punkten $(x + yj, 1)$ gehören dem Lösungsgebilde genau an die, für die gilt $a(x^2 - y^2) + bx + cy + d = 0$. Wir fragen nach den uneigentlichen Punkten, die

⁶⁾ Hier sind homogene Koordinaten gemeint.

dem Lösungsgebilde Λ angehören. ∞ gehört hierzu genau dann, wenn $a = 0$ ist. Im Falle $a = 0$ gehört kein weiterer uneigentlicher Punkt zum Lösungsgebilde Λ : Sei $P = (z_1 z_2)$ eine Lösung mit $z_2 \in I_+ \cup I_-$. Wegen $z_2 \bar{z}_2 \in I_+ \cdot I_- = \{0\}$ reduziert sich $(*)$ zu

$$(**) \quad \frac{b + cj}{2} z_1 \bar{z}_2 + \frac{b - cj}{2} \bar{z}_1 z_2 = 0.$$

Setzt man $\frac{b - cj}{2} \bar{z}_1 z_2 = w = (\alpha + \beta j)$, so nimmt $(**)$ die Gestalt $\bar{w} + w = 2\alpha = 0$ an, woraus $w = 0$ folgt, wenn wir noch $w \in I_+ \cup I_-$ beachten. Wegen

$$-ad + \frac{b^2 - c^2}{4} = \frac{(b + c)(b - c)}{4} \neq 0$$

ist $\frac{b - cj}{2} \in \mathfrak{R}$ und wir erhalten $\bar{z}_1 z_2 = 0$. Die Bemerkung zu Hilfssatz 2.1 zeigt, daß ∞ der einzige uneigentliche Punkt ist, für den $\bar{z}_1 z_2 = 0$ ist.

Sei $a \neq 0$. Also kommen hier nur uneigentliche Punkte der Form $(2u, 1 + \varepsilon j)$, $(1 - \varepsilon j, 1 + \varepsilon j)$, $\varepsilon \in \{+1, -1\}$, $u \neq 0$, als zu Λ gehörig in Frage. Da wegen $(*)$

$$az_1 \bar{z}_1 + 2 \operatorname{Re} \left(\frac{b + cj}{2} z_1 \bar{z}_2 \right) = 0$$

gelten muß, hat man genau die uneigentlichen Punkte

$$x + y = \frac{c - b}{2a}, \quad x - y = -\frac{c + b}{2a},$$

d. h. die Asymptoten als Lösungen.

Aufgrund von Satz 2.1 hat also jeder pseudo-euklidische Kreis eine Darstellung $\mathfrak{z}^T \mathfrak{M} \bar{\mathfrak{z}} = 0$, wo $\mathfrak{M}$ eine nichtsinguläre Matrix vom Typ $(2, 2)$ über $\mathfrak{L}$ ist mit $\mathfrak{M} = \bar{\mathfrak{M}}^T$. Umgekehrt gilt

Satz 2.2. *Sei $\mathfrak{M}$ eine nichtsinguläre Matrix vom Typ $(2, 2)$ über $\mathfrak{L}$ mit $\mathfrak{M} = \bar{\mathfrak{M}}^T$. Dann stellt das Lösungsgebilde von $\mathfrak{z}^T \mathfrak{M} \bar{\mathfrak{z}} = 0$ einen pseudo-euklidischen Kreis dar.*

Beweis. Sei $\mathfrak{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$. Aus $\mathfrak{M} = \bar{\mathfrak{M}}^T$ folgt $m_{11} = \overline{m_{11}}$, $m_{22} = \overline{m_{22}}$, d. h. $a \equiv m_{11} \in \mathfrak{R}$, $d \equiv m_{22} \in \mathfrak{R}$. Setzen wir $m_{12} = \frac{b + cj}{2}$ mit $b, c \in \mathfrak{R}$, so ist also

$$m_{21} = \overline{m_{12}} = \frac{b - cj}{2}.$$

Wegen $|\mathfrak{M}| \neq 0$ folgt weiterhin $ad - \frac{b^2 - c^2}{4} \neq 0$. Damit hat man aber tatsächlich einen pseudo-euklidischen Kreis vor sich.

Damit kommen wir zu

Satz 2.3. *Bei der (aufgrund von Hilfssatz 2.1) 1,1-deutigen Abbildung σ von $\mathfrak{E}$ auf $\mathfrak{P}$ sind — bei Identifikation — die pseudo-euklidischen Kreise genau die pseudo-euklidischen Kreise des § 1.*

Beweis. Die Kreise des § i wollen wir für diesen Beweis i-Kreise nennen. Durch Abbildungen von $\mathbf{T}$ werden 2-Kreise in 2-Kreise überführt: Ist $\mathfrak{z}^T \mathfrak{M} \bar{\mathfrak{z}} = 0$ ein 2-Kreis,

ist $\mathfrak{z}' = \mathfrak{A}\mathfrak{z}$ eine Abbildung aus $\mathbb{T}$, die wir in der Form $\mathfrak{z} = \mathfrak{B}\mathfrak{z}'$ schreiben wollen, so hat man als Bild $\mathfrak{z}'^T \mathfrak{B}^T \mathfrak{M} \mathfrak{B} \bar{\mathfrak{z}}' = 0$; dies ist wieder ein 2-Kreis nach Satz 2.2 wegen $(\mathfrak{B}^T \mathfrak{M} \mathfrak{B})^T = \mathfrak{B}^T \mathfrak{M}^T \mathfrak{B} = \mathfrak{B}^T \mathfrak{M} \mathfrak{B}$ und $|\mathfrak{B}^T \mathfrak{M} \mathfrak{B}| = |\mathfrak{B}| |\mathfrak{M}| |\mathfrak{B}| \neq 0$. Der 1-Kreis $\mathfrak{p}$ ist ein 2-Kreis, da er in der Form $\mathfrak{z}^T \begin{pmatrix} 0 & \frac{j}{2} \\ -\frac{j}{2} & 0 \end{pmatrix} \bar{\mathfrak{z}} = 0$ geschrieben werden kann. Also sind nach dem vorhin Bewiesenen alle 1-Kreise $\mathfrak{p}^T, \tau \in \mathbb{T}$, 2-Kreise. Umgekehrt ist jeder 2-Kreis auch ein 1-Kreis:

Ist nämlich

$$\mathfrak{z}^T \mathfrak{M} \bar{\mathfrak{z}} = 0, \quad \mathfrak{M} = \begin{pmatrix} a & \frac{b+cj}{2} \\ \frac{b-cj}{2} & d \end{pmatrix},$$

ein 2-Kreis, so geht er durch die T -Abbildung $\mathfrak{x} = \mathfrak{B}^{-1}\mathfrak{z}$ aus

$$\mathfrak{x}^T \begin{pmatrix} 0 & \frac{j}{2} \\ -\frac{j}{2} & 0 \end{pmatrix} \bar{\mathfrak{x}} = 0$$

hervor, wenn $|\mathfrak{B}| \in \mathfrak{R}$ und

$$\mathfrak{B}^T \mathfrak{M} \mathfrak{B} = \varrho \mathfrak{M}_0, \quad \mathfrak{M}_0 = \begin{pmatrix} 0 & \frac{j}{2} \\ -\frac{j}{2} & 0 \end{pmatrix}, \quad \varrho \in \mathfrak{R},$$

gilt.

1. Fall: $a = 0$. Seien $(\xi_1, \eta_1), (\xi_2, \eta_2)$ verschiedene Punkte der Geraden $bx + cy + d = 0$. Dann gilt $(\xi_2 - \xi_1) + (\eta_2 - \eta_1)j \in \mathfrak{R}$, da sonst $(\xi_2 - \xi_1)^2 = (\eta_2 - \eta_1)^2$ wäre, was mit $b(\xi_2 - \xi_1) + c(\eta_2 - \eta_1) = 0$ auf $b^2 - c^2 = 0$ führte, was wegen $ad - \frac{b^2 - c^2}{4} \neq 0$ nicht geht.

Wir betrachten die Matrix $\mathfrak{B} = \begin{pmatrix} (\xi_2 - \xi_1) + j(\eta_2 - \eta_1) & \xi_1 + \eta_1 j \\ 0 & 1 \end{pmatrix}$, die das Gewünschte leistet.

2. Fall: $a \neq 0$. Die T -Abbildung $\varphi: \mathfrak{z} \rightarrow \mathfrak{y}$ mit

$$\mathfrak{z} = \begin{pmatrix} \left(\frac{c-b}{2a} + 1\right) + \left(\frac{c-b}{2a} - 1\right)j & \alpha \\ 1 + j & 1 \end{pmatrix} \mathfrak{y}, \quad \text{wobei } \alpha = \begin{cases} 1 - j & \text{falls } b \neq c \\ 1 + j & \text{falls } b = c, \end{cases}$$

überführt den 2-Kreis $\xi: a(x^2 - y^2) + bx + cy + d = 0$ in einen 2-Kreis

$$\eta: a'(x^2 - y^2) + b'x + c'y + d' = 0$$

mit $a' = 0$, da $\left(\left(\frac{c-b}{2a} + 1\right) + \left(\frac{c-b}{2a} - 1\right)j, 1 + j\right) \in \xi$ in $(1, 0)$ übergeht.

Aufgrund des Falles 1 geht η durch eine T -Abbildung ψ aus $\mathfrak{p}$ hervor. Also ist $\xi^\psi = \eta, \eta = \mathfrak{p}^\psi$, was $\xi = \mathfrak{p}^{\psi\varphi^{-1}}$ mit $\psi\varphi^{-1} \in \mathbb{T}$ ergibt.

Also auch im Falle 2 ist ξ ein 1-Kreis.

Zum Abschluß dieses Paragraphen gehen wir noch auf die Deutung des Parallelitätsbegriffes im Hyperbelmodell ein. Wir nennen dabei die uneigentlichen Punkte der Form $y = x + k$, $k \in \mathbb{R}$, p -Punkte, und die uneigentlichen Punkte der Form $y = -x + k$, $k \in \mathbb{R}$, m -Punkte. Dann gilt:

Satz 2.4. a) Der Punkt ∞ ist zu keinem eigentlichen Punkt parallel. Er ist p -parallel genau zu den m -Punkten und zu ∞ und m -parallel genau zu den p -Punkten und zu ∞ .

b) Seien P, Q zwei untereinander und von ∞ verschiedene Punkte. Sind P, Q eigentlich, so sind sie p -parallel (bzw. m -parallel) genau dann, wenn ihre Verbindungsgerade ein p -Punkt (bzw. ein m -Punkt) ist. Sind P, Q uneigentlich, so sind sie p -parallel (bzw. m -parallel) genau dann, wenn sie beide m -Punkte (bzw. beide p -Punkte) sind. Ist P eigentlich und ist Q uneigentlich, so ist $P \parallel_+ Q$ (bzw. $P \parallel_- Q$) genau dann, wenn Q ein p -Punkt (bzw. ein m -Punkt) ist, der als Gerade betrachtet mit P inzidiert.

Beweis. Zu a). $\begin{vmatrix} 1 & 0 \\ a + bj & 1 \end{vmatrix} = 1 \in \mathbb{R}$ zeigt, daß ∞ zu keinem eigentlichen Punkt parallel ist. Ein p -Punkt ist allgemein von der Form (hier liegt $y = x + u$ zugrunde)

$$(p) \quad ((1 - u) + (1 + u)j, 1 - j), \quad u \in \mathbb{R},$$

und ein m -Punkt allgemein von der Form (hier liegt $y = -x + u$ zugrunde)

$$(m) \quad ((u + 1) + (u - 1)j, 1 + j), \quad u \in \mathbb{R}.$$

Hieraus folgt der verbleibende Teil der Behauptung a).

Zu b). Seien die eigentlichen Punkte $P = (a + bj, 1)$, $Q = (c + dj, 1)$ p -parallel:

$\begin{vmatrix} a + bj & 1 \\ c + dj & 1 \end{vmatrix} \in I_+$. Dies ist offenbar genau dann der Fall, wenn die Verbindungsgerade ein p -Punkt ist. Entsprechendes gilt für m -Parallelität. Der Fall P, Q uneigentlich erledigt sich sofort mit (p) und (m). Sei nun noch $P = (a + bj, 1)$ ein eigentlicher Punkt und Q ein uneigentlicher Punkt $\neq \infty$. Ist Q ein p -Punkt, so ist nach (p) gewiß P nicht m -parallel zu Q , und es ist P p -parallel zu Q genau dann, wenn $b - a = u$ ist, wenn $Q = ((1 - u) + (1 + u)j, 1 - j)$ gesetzt war. Dies zeigt die Behauptung. Entsprechendes gilt für den verbleibenden Teil, daß nämlich Q ein m -Punkt ist.

§ 3. Die ebenen Schnitte eines Hyperboloids

Im dreidimensionalen projektiven Raum über $\mathbb{R}$, $\text{Char } \mathbb{R} \neq 2$, betrachten wir — bezogen auf ein projektives Koordinatensystem (x, y, z, t) — das „Hyperboloid“ $\mathfrak{Q} = \{(x, y, z, t) \mid x^2 - y^2 + z^2 = t^2\}$.

Hilfssatz 3.1. Genau dann ist die Ebene $E: ax + by + cz + dt = 0$ Tangentialebene an $\mathfrak{Q}$ (d. h. Polare eines Quadrikkpunktes), wenn $a^2 - b^2 + c^2 = d^2$ ist.

Beweis. Sei $a^2 - b^2 + c^2 = d^2$. Dann ist $(a, -b, c, -d) \in \mathfrak{Q} \cap E$. Die Tangentialebene im Punkt $(a, -b, c, -d)$ an $\mathfrak{Q}$ ist aber gerade die Ebene E . — Sei umgekehrt E eine Tangentialebene an $\mathfrak{Q}$; sei E die Polare von (p, q, r, s) . Dann hat E die Darstellung: $px - qy + rz = st$. Es gilt $(p, -q, r, -s) = \lambda(a, b, c, d)$ mit einem $\lambda \neq 0$ aus $\mathbb{R}$. Wegen $(p, q, r, s) \in \mathfrak{Q}$, d. h. wegen $p^2 - q^2 + r^2 = s^2$, folgt dann aber $a^2 - b^2 + c^2 = d^2$.

Es sei nun p der Punkt $(0, 0, 1, 1)$ und Π die Polare von p . Wegen $\Pi: z = t$ ist $\Pi \cap \mathfrak{Q}$ die Menge aller Punkte (x, y, z, t) mit $x^2 - y^2 = 0$ und $z = t$. Es besteht also $\alpha = \Pi \cap \mathfrak{Q}$

aus den beiden Geraden

$$\alpha_+ = \{(x, y, z, t) \neq 0 \mid x = y, z = t\},$$

$$\alpha_- = \{(x, y, z, t) \neq 0 \mid x = -y, z = t\}.$$

Wir führen eine *stereographische Projektion* σ durch: Dem Punkt $\mathfrak{x} \in \mathfrak{Q} - \alpha$ ordnen wir den Schnittpunkt der Geraden durch $\mathfrak{p}$, $\mathfrak{x}$ mit der Ebene $\Gamma: z = 0$ zu. Bezeichnet ω die uneigentliche Gerade von Γ , so gilt der

Hilfssatz 3.2. *Es ist σ eine 1,1-deutige Abbildung von $\mathfrak{Q} - \alpha$ auf $\Gamma - \omega$.*

Beweis. Wegen $\mathfrak{p} \notin \Gamma$ ist σ eine eindeutige Abbildung von $\mathfrak{Q} - \alpha$ in Γ . Wegen $\mathfrak{x} \in \mathfrak{Q}$, $\mathfrak{x} \notin \Pi$ und $\Pi \cap \Gamma = \omega$ ist σ eine eindeutige Abbildung von $\mathfrak{Q} - \alpha$ in $\Gamma - \omega$. Jeder Punkt aus $\Gamma - \omega$ tritt genau einmal als Bild auf: Sei dieser Punkt durch $(\xi, \eta, 0, 1)$ gegeben. Für ein eventuell vorhandenes Urbild (x, y, z, t) in $\mathfrak{Q} - \alpha$ muß gelten: $(x, y, z, t) = \lambda(\xi, \eta, 0, 1) + \mu(0, 0, 1, 1)$, $\lambda, \mu \in \mathbb{R}$, $\lambda \neq 0$ und $x^2 - y^2 + z^2 = t^2$.

Dies führt auf: $\lambda^2(\xi^2 - \eta^2) + \mu^2 = (\lambda + \mu)^2$, d. h. auf $\frac{\mu}{\lambda} = \frac{\xi^2 - \eta^2 - 1}{2}$. Es gibt also höchstens ein Urbild, nämlich $(\xi, \eta, 0, 1) + \frac{\xi^2 - \eta^2 - 1}{2}(0, 0, 1, 1)$; dieser Punkt ist aber offenbar auch ein Urbild aus $\mathfrak{Q} - \alpha$.

Wir wollen jetzt die Abbildung σ fortsetzen zu einer 1,1-deutigen Abbildung $\bar{\sigma}$ von $\mathfrak{Q}$ auf die Menge der pseudo-euklidischen Punkte (hier mit Einschluß der uneigentlichen): Der Punkt $\mathfrak{p}$ soll dem uneigentlichen Punkt ∞ entsprechen. Abgesehen von $\mathfrak{p}$ gehören zu α noch die Punkte $(1, 1, k, k)$, $k \in \mathbb{R}$, und $(1, -1, k, k)$, $k \in \mathbb{R}$. Wir ordnen zu

$$\bar{\sigma}: (1, 1, k, k) \rightarrow \xi - \eta = k,$$

$$\bar{\sigma}: (1, -1, k, k) \rightarrow \xi + \eta = k.$$

Man verifiziert in der Tat leicht den

Hilfssatz 3.3. *Es ist $\bar{\sigma}$ eine 1,1-deutige Abbildung von $\mathfrak{Q}$ auf die Menge der pseudo-euklidischen Punkte (mit Einschluß der uneigentlichen).*

Wir nennen die Punktmenge $E \cap \mathfrak{Q}$ einen *zulässigen ebenen Schnitt*, wenn die sonst beliebige Ebene E keine Tangentialebene an $\mathfrak{Q}$ ist. Dann gilt

Hilfssatz 3.4. *Das $\bar{\sigma}$ -Bild eines jeden zulässigen ebenen Schnittes ist ein pseudo-euklidischer Kreis. Umgekehrt ist das Urbild eines pseudo-euklidischen Kreises ein zulässiger ebener Schnitt.*

Beweis. Sei $E \cap \mathfrak{Q}$ ein zulässiger ebener Schnitt. Ist E durch $ax + by + cz + dt = 0$ gegeben, so gilt nach Hilfssatz 3.1: $a^2 - b^2 + c^2 \neq d^2$.

Es ist $(E \cap (\mathfrak{Q} - \alpha))^{\bar{\sigma}} = (E \cap (\mathfrak{Q} - \alpha))^{\sigma}$ gegeben genau durch die Menge der Punkte $(\xi, \eta, 0, 1)$ mit

$$(*) \quad (c + d)(\xi^2 - \eta^2) + 2a\xi + 2b\eta + (-c + d) = 0,$$

wenn man beachtet

$$\varrho x = \xi,$$

$$\varrho y = \eta,$$

$$\varrho z = \frac{\xi^2 - \eta^2 - 1}{2},$$

$$\varrho t = \frac{\xi^2 - \eta^2 + 1}{2} \text{ mit einem } \varrho \neq 0 \text{ aus } \mathbb{R}.$$

Wegen $a^2 - b^2 + c^2 \neq d^2$, d. h. wegen

$$(c + d)(-c + d) - \frac{(2a)^2 - (2b)^2}{4} \neq 0,$$

hat man in $(E \cap (\mathfrak{D} - \alpha))^{\bar{\sigma}}$ den eigentlichen Teil eines pseudo-euklidischen Kreises.

Wir bestimmen noch $(E \cap \alpha)^{\sigma}$; es wird sich herausstellen, daß es sich gerade um den uneigentlichen Teil des pseudo-euklidischen Kreises $(*)$ handelt: Sei $c + d = 0$. Dann ist $E \cap \alpha = \mathfrak{p}$. (Läge noch ein Punkt $(1, 1, k, k)$ oder ein Punkt $(1, -1, k, k)$ auf E , so erhielte man in Verbindung mit $c + d = 0$ die Gleichung $a + b = 0$ bzw. $a - b = 0$, woraus $0 = a^2 - b^2 = a^2 - b^2 + c^2 - d^2$ folgen würde.) Im Falle $c + d = 0$ besitzt also $(*)$ den einzigen uneigentlichen Punkt ∞ , der gleich $\mathfrak{p}^{\bar{\sigma}} = (E \cap \alpha)^{\bar{\sigma}}$ ist.

Sei $c + d \neq 0$. Dann ist $\mathfrak{p} \notin E \cap \alpha$ und

$$\begin{aligned} E \cap \alpha_+ &= \left(1, 1, -\frac{a+b}{c+d}, -\frac{a+b}{c+d}\right), \\ E \cap \alpha_- &= \left(1, -1, -\frac{a-b}{c+d}, -\frac{a-b}{c+d}\right). \end{aligned}$$

Es gilt $(E \cap \alpha_+)^{\bar{\sigma}} = \left\{\eta - \xi = \frac{a+b}{c+d}\right\}$, $(E \cap \alpha_-)^{\bar{\sigma}} = \left\{\eta + \xi = \frac{b-a}{c+d}\right\}$; dies sind aber nach Beginn von § 2 die beiden uneigentlichen Punkte von $(*)$.

Ist umgekehrt der pseudo-euklidische Kreis

$$(**) \quad \alpha(\xi^2 - \eta^2) + \beta\xi + \gamma\eta + \delta = 0, \quad \alpha\delta - \frac{\beta^2 - \gamma^2}{4} \neq 0,$$

gegeben, so ist mit dem Vorstehenden sofort zu sehen, daß das Urbild durch den zulässigen ebenen Schnitt

$$E \cap \mathfrak{D}, \quad E: \beta x + \gamma y + (\alpha - \delta)z + (\alpha + \delta)t = 0,$$

gegeben ist.

Wir fassen zusammen:

Satz 3.1. *Zugrunde gelegt sei ein beliebiger kommutativer Körper $\mathfrak{K}$ von Char $\neq 2$. Dann bildet die Gesamtheit der Punkte und der zulässigen ebenen Schnitte von $\mathfrak{D}$, die projektive Inzidenzrelation übernommen, ein Modell der pseudo-euklidischen Geometrie über $\mathfrak{K}$.*

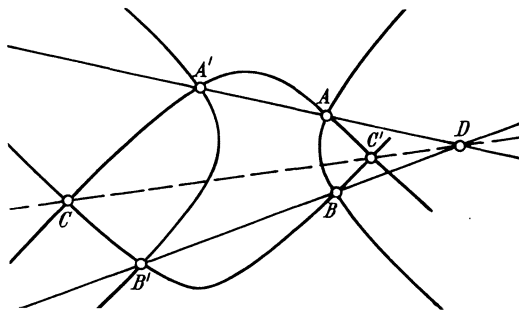
Mit Hilfe dieses Satzes läßt sich der Satz 1.3 geometrisch beweisen. Allerdings müssen dann Hilfsbetrachtungen vorausgeschickt werden, insbesondere Reichhaltigkeitsaussagen und weitergehende Betrachtungen, die den Parallelitätsbegriff betreffen. Eine Deutung des Parallelitätsbegriffes im Hyperboloid-Modell wird in § 4 gegeben (s. Hilfsatz 4.2).

§ 4. Zwei $(8_3, 6_4)$ -Konfigurationen

Wir beginnen mit der *Büschelsatz-Konfiguration*. Zugrunde gelegt sei die pseudo-euklidische Geometrie über dem kommutativen Körper $\mathfrak{K}$, Char $\mathfrak{K} \neq 2$.

Satz 4.1 (Büschelsatz). *Seien $A, A', B, B', C, C', D, D'$ paarweise nicht parallele pseudo-euklidische Punkte derart, daß die Quadrupel A, A', B, B' ; A, A', C, C' ; $A, A',$*

D, D' ; B, B', C, C' ; B, B', D, D' je konzyklisch liegen auf wenigstens vier verschiedenen Kreisen. Dann liegt auch das Quadrupel C, C', D, D' konzyklisch.



Figur 1. Büschelsatz

Wir führen den Beweis im Modell des § 3. Zunächst drei Hilfssätze.

Hilfssatz 4. 1. Sei S ein Punkt auf $\mathfrak{Q}$, sei Σ die Polare von S . Ist dann g eine Gerade auf $\Sigma \cap \mathfrak{Q}$, so gilt $S \in g$. Weiterhin sind je zwei Punkte von g zueinander konjugiert.

Beweis. Seien U, V verschiedene Punkte auf g . Seien

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}$$

Koordinatenmatrizen von U, V . Ist

$$\mathfrak{A} = \begin{pmatrix} 1 & & & 0 \\ & -1 & & \\ & & 1 & \\ 0 & & & -1 \end{pmatrix},$$

so gilt also $(\alpha u + \beta v)^T \mathfrak{A} (\alpha u + \beta v) = 0$ für alle $\alpha, \beta \in \mathbb{R}$. Hieraus folgt, daß U zu V konjugiert sein muß. Die Polare Φ von V enthält also U ; sie enthält weiterhin V wegen $V \in \mathfrak{Q}$ und S , da V zu S konjugiert ist wegen $V \in \Sigma$. Wären nun U, V, S nicht kollinear, so müßte $\Phi = U + V + S \equiv \Sigma$ sein, was nicht geht, da zu verschiedenen Punkten (hier V, S) auch verschiedene Polaren gehören.

Hilfssatz 4. 2. Es seien P, Q verschiedene Punkte auf $\mathfrak{Q}$. Dann sind die folgenden Aussagen gleichwertig:

- (a) P ist zu Q parallel.
- (b) Die Punkte P, Q sind zueinander konjugiert.
- (c) $P + Q \subseteq \mathfrak{Q}$.

Beweis. Aus (a) folgt (b): Da die Punkte P, Q zueinander parallel sind, gibt es nach Satz 1. 3 (2) keinen Kreis durch P, Q . Damit ist für jede Ebene E durch P, Q der Schnitt $E \cap \mathfrak{Q}$ nicht zulässig. Nach Hilfssatz 3. 1 gilt also für jede Ebene

$$E: ax + by + cz + dt = 0$$

durch P, Q die Relation

$$(*) \quad a^2 - b^2 + c^2 = d^2.$$

Setzen wir

$$\mathfrak{a} = \begin{pmatrix} a \\ -b \\ c \\ -d \end{pmatrix},$$

so haben E bzw. $(*)$ die Form $\mathfrak{a}^T \mathfrak{A} \mathfrak{x} = 0$ bzw. $\mathfrak{a}^T \mathfrak{A} \mathfrak{a} = 0$ mit der Matrix $\mathfrak{A}$ von Hilfssatz 4. 1. Wir können also sagen: Ist A ein beliebiger Punkt des Raumes, der zu P und zu Q konjugiert ist, so ist $A \in \mathfrak{Q}$. Bezeichnet g die Schnittgerade der Polaren von P und Q , so gilt also $g \subseteq \mathfrak{Q}$. Hilfssatz 4. 1 ergibt $g \ni P, Q$ und weiterhin, daß P, Q zueinander konjugiert sind.

Aus (b) folgt (c): Da P, Q verschiedene und zueinander konjugierte Punkte auf $\mathfrak{Q}$ sind, so gilt für jedes $R \in P + Q$ offenbar (es bezeichne hier T^Π die Polare des Punktes T) $R^\Pi \ni P, Q$ wegen $R \in P + Q \subseteq P^\Pi \cap Q^\Pi$. Also ist $R^\Pi \supseteq P + Q \ni R$.

Aus (c) folgt (a): Angenommen, die Punkte P, Q wären nicht parallel. Dann gibt es wegen Satz 1. 3, (2') eine Ebene

$$E: ax + by + cz + dt = 0$$

durch P, Q derart, daß $E \cap \mathfrak{Q}$ zulässig ist. Setzen wir wieder

$$\mathfrak{a} = \begin{pmatrix} a \\ -b \\ c \\ -d \end{pmatrix},$$

so haben wir also $\mathfrak{a}^T \mathfrak{A} \mathfrak{p} = 0$, $\mathfrak{a}^T \mathfrak{A} \mathfrak{q} = 0$, $\mathfrak{a}^T \mathfrak{A} \mathfrak{a} \neq 0$, wenn $\mathfrak{p}, \mathfrak{q}$ Koordinatenmatrizen von resp. P, Q sind. Ist A der Punkt mit der Koordinatenmatrix $\mathfrak{a}$, so gilt also $A \in P^\Pi \cap Q^\Pi$, $A \notin \mathfrak{Q}$. Damit ist $P + Q \subseteq A^\Pi$. Aus $P + Q \subseteq \mathfrak{Q}$, d. h. aus

$$(\alpha \mathfrak{p} + \beta \mathfrak{q})^T \mathfrak{A} (\alpha \mathfrak{p} + \beta \mathfrak{q}) = 0$$

für alle $\alpha, \beta \in \mathfrak{R}$ folgt $P \in Q^\Pi$. Also ist $\mathfrak{Q} \supseteq P + Q = P^\Pi \cap Q^\Pi \ni A$, was $A \notin \mathfrak{Q}$ widerspricht.

Hilfssatz 4. 3. Ist g eine Gerade, so gilt entweder $g \subseteq \mathfrak{Q}$ oder aber $|g \cap \mathfrak{Q}| \leq 2$.

Beweis. Sind nämlich P, Q, R drei verschiedene Punkte auf g , die auch zu $\mathfrak{Q}$ gehören, so seien $\mathfrak{p}, \mathfrak{q}, \mathfrak{p} + \mathfrak{q}$ Koordinatenmatrizen von resp. P, Q, R . Aus

$$0 = (\mathfrak{p} + \mathfrak{q})^T \mathfrak{A} (\mathfrak{p} + \mathfrak{q}) = 2\mathfrak{p}^T \mathfrak{A} \mathfrak{q}$$

folgt dann $(\alpha \mathfrak{p} + \beta \mathfrak{q})^T \mathfrak{A} (\alpha \mathfrak{p} + \beta \mathfrak{q}) = 2\alpha\beta \mathfrak{p}^T \mathfrak{A} \mathfrak{q} = 0$ für alle α, β aus $\mathfrak{R}$.

Nun zum Beweis von Satz 4. 1:

Das Quadrupel A, A', B, B' ist koplanar. Die Geraden $A + A', B + B'$ sind verschieden. Im anderen Falle müßte nach Hilfssatz 4. 3 die Gerade $A + A' = B + B'$ ganz zu $\mathfrak{Q}$ gehören. Dann wären aber nach Hilfssatz 4. 2, (c) $\Rightarrow$ (a), die Punkte A, A', B, B' paarweise parallel. Sei $T = (A + A') \cap (B + B')$. Wir zeigen, daß T auf $C + C'$ liegt (mutatis mutandis ergibt sich $T \in D + D'$): Wie oben ergibt sich die Existenz von Punkten U, V mit $U = (A + A') \cap (C + C')$, $V = (B + B') \cap (C + C')$. Im Falle $U = V$ ist $U = V = T$. Im Falle $U \neq V$ wäre $C + C' = U + V \subseteq (A + A') + (B + B')$; hier stimmen also die Kreise durch $A, A', B, B'; A, A', C, C'; B, B', C, C'$ überein, was nicht geht, da sonst die Quadrupel $A, A', B, B'; A, A', C, C'; A, A', D, D'; B, B', C, C'; B, B', D, D'$ schon auf drei Kreise verteilt wären.

Aus $T \in C + C', D + D'$ folgt, daß C, C', D, D' koplanar liegen. Kollinear können keine drei dieser Punkte liegen, da sonst nach Hilfssatz 4.3 die betreffende Gerade ganz zu $\mathfrak{Q}$ gehörte und also unter den Punkten C, C', D, D' wenigstens zwei zueinander parallel wären nach Hilfssatz 4.2 Die Ebene E durch C, C', D, D' hat die Eigenschaft, daß $E \cap \mathfrak{Q}$ zulässig ist: Denn es gilt schon $E = C + C' + D$ und durch die drei paarweise nicht parallelen Punkte C, C', D geht nach Satz 1.3, (1) ein Kreis.

Wir arbeiten jetzt im Modell des § 1.

Sind $P = (p_1, p_2), Q = (q_1, q_2), R = (r_1, r_2), S = (s_1, s_2)$ verschiedene Punkte so, daß P nicht zu S und Q nicht zu R parallel ist, so erklären wir das dem geordneten Punktequadrupel P, Q, R, S zugeordnete *Doppelverhältnis*

$$\begin{bmatrix} P & Q \\ S & R \end{bmatrix} = \frac{\begin{vmatrix} p_1 & p_2 \\ r_1 & r_2 \end{vmatrix}}{\begin{vmatrix} p_1 & p_2 \\ s_1 & s_2 \end{vmatrix}} \cdot \frac{\begin{vmatrix} q_1 & q_2 \\ s_1 & s_2 \end{vmatrix}}{\begin{vmatrix} q_1 & q_2 \\ r_1 & r_2 \end{vmatrix}}.$$

Wir beachten dabei, daß aufgrund der gemachten Voraussetzung die Nenner regulär sind. Dieser Wert aus $\mathfrak{Q}$ hängt offenbar nicht von den gewählten Repräsentanten $(p_1, p_2), \dots$ der Punkte $P, \dots$ ab.

Hilfssatz 4.4. Sind P, Q, R, S verschiedene Punkte mit $P \nparallel S$ und $Q \nparallel R$ und ist τ eine Abbildung aus T , so gilt

$$\begin{bmatrix} P^\tau & Q^\tau \\ S^\tau & R^\tau \end{bmatrix} = \begin{bmatrix} P & Q \\ S & R \end{bmatrix}.$$

Beweis. Zunächst stellen wir fest, daß $P^\tau, Q^\tau, R^\tau, S^\tau$ paarweise verschiedene Punkte sind und daß, aufgrund von Hilfssatz 1.2, $P^\tau \nparallel S^\tau, Q^\tau \nparallel R^\tau$ gilt. Hat τ die Gestalt

$$\varrho_x \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \mathfrak{U} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathfrak{U} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

so ist also (wenn (p'_1, p'_2) ein Repräsentant von P^τ ist usw.),

$$\begin{pmatrix} \varrho_p p'_1 & \varrho_r r'_1 \\ \varrho_p p'_2 & \varrho_r r'_2 \end{pmatrix} = \mathfrak{U} \begin{pmatrix} p_1 & r_1 \\ p_2 & r_2 \end{pmatrix},$$

was auf

$$\varrho_p \varrho_r \begin{vmatrix} p'_1 & r'_1 \\ p'_2 & r'_2 \end{vmatrix} = |\mathfrak{U}| \begin{vmatrix} p_1 & r_1 \\ p_2 & r_2 \end{vmatrix}$$

führt. Entsprechendes für die anderen benötigten Determinanten aufgeschrieben, führt sofort auf die Behauptung.

Hilfssatz 4.5. Sind P, Q, R, S paarweise nicht parallele Punkte, so gilt $\begin{bmatrix} P & Q \\ S & R \end{bmatrix} \in \mathfrak{R} - \{0\}$ genau dann, wenn die Punkte P, Q, R, S konzyklisch liegen.

Beweis. Sei α die nach Satz 1.1 eindeutig existierende Abbildung aus T mit $P^\alpha = (1, 0), Q^\alpha = (0, 1), R^\alpha = (1, 1)$. Wegen $S^\alpha \nparallel P^\alpha = (1, 0)$ ist — wenn $S^\alpha = (u, v)$ gesetzt wird — $\begin{vmatrix} u & v \\ 1 & 0 \end{vmatrix} \in \mathfrak{R}$. Da also v regulär ist, können wir S^α durch $(k, 1)$ wiedergeben; hier muß $k \neq 0$ sein wegen $S^\alpha \nparallel Q^\alpha$. Nach Hilfssatz 4.4 gilt

$$k = \begin{bmatrix} P^\alpha & Q^\alpha \\ S^\alpha & R^\alpha \end{bmatrix} = \begin{bmatrix} P & Q \\ S & R \end{bmatrix}.$$

Liegen nun P, Q, R, S konzyklisch, so also auch $P^\alpha, Q^\alpha, R^\alpha, S^\alpha$, was $k \in \mathfrak{K}$ ergibt, da $\mathfrak{p}$ der Kreis durch $(1, 0), (0, 1), (1, 1)$ ist. Gilt umgekehrt $k \in \mathfrak{K}$, so ist $S^\alpha \in \mathfrak{p}$ und $P^\alpha, Q^\alpha, R^\alpha, S^\alpha$ liegen damit konzyklisch, was sich auf das Quadrupel P, Q, R, S überträgt.

Hilfssatz 4. 6. Sind A, B, C paarweise nicht parallele Punkte, ist $Q \neq A$ ein Punkt mit $Q \neq B, C$ und $\begin{bmatrix} A & Q \\ B & C \end{bmatrix} \in \mathfrak{K}$, so liegt das Quadrupel A, B, C, Q konzyklisch und insbesondere gilt $Q \neq A$.

Beweis. Das Doppelverhältnis $\begin{bmatrix} A & Q \\ B & C \end{bmatrix}$ ist erklärt wegen $Q \neq C$ und $A \neq B$. Sei α die Abbildung aus T mit $A^\alpha = (1, 0), B^\alpha = (1, 1), C^\alpha = (0, 1)$. Dann ist

$$\mathfrak{K} \ni \begin{bmatrix} A & Q \\ B & C \end{bmatrix} = \begin{bmatrix} A^\alpha & Q^\alpha \\ B^\alpha & C^\alpha \end{bmatrix} = 1 - \frac{q_2}{q_1},$$

wenn Q die Form (q_1, q_2) hat; wir beachten dabei

$$q_1 = \begin{vmatrix} q_1 & q_2 \\ 0 & 1 \end{vmatrix} \in \mathfrak{K} \text{ wegen } Q^\alpha \neq C^\alpha.$$

Also ist $\frac{q_2}{q_1} \in \mathfrak{K}$, d. h. der Punkt Q — in der Form $\left(1, \frac{q_2}{q_1}\right)$ aufgeschrieben — gehört zu $\mathfrak{p}$. Damit ist das Quadrupel $A^\alpha, B^\alpha, C^\alpha, Q^\alpha$ konzyklisch und also auch das Quadrupel A, B, C, Q . Da es sich bei A, B, C, Q um paarweise verschiedene Punkte handelt, so gilt $Q \neq A$ nach Satz 1. 3, (2).

Wir sagen, daß der Kreis a den Kreis b im Punkt P berührt, wenn entweder $a = b \ni P$ gilt oder aber $a \cap b = \{P\}$ ist. Es gilt

Satz 4. 2. Es seien a, b Kreise durch den Punkt P mit $a \cap b = \{P\}$. Weiterhin seien A_1, A_2 verschiedene Punkte auf $a - b$ und B_1, B_2 verschiedene Punkte auf $b - a$. Dann gilt

$$\begin{bmatrix} P & A_1 \\ B_2 & A_2 \end{bmatrix} - \begin{bmatrix} P & A_1 \\ B_1 & A_2 \end{bmatrix} \in \mathfrak{K}^x \text{).}$$

Beweis. Wir können die spezielle Situation $P = (1, 0), A_1 = (0, 1), A_2 = (1, 1)$ annehmen. Ist $B_i = (b_{i1}, b_{i2}), i = 1, 2$, so ist b_{i2} regulär wegen $B_i \neq P$ (es sind ja doch B_i, P verschiedene Punkte auf einem Kreis). Wir können damit B_i in der Form $(b_i, 1)$ angeben.

Der Kreis b hat mit a nur den Punkt P gemeinsam. Der Kreis b' durch $P, B_1 (b_1 + 1, 1)$ hat aber mit a auch nur den Punkt P gemeinsam. Wegen Satz 1. 3, (4) — das dortige Q ersetze man durch B_1 — ist $b = b'$. Es gilt $b_2 - b_1 \in \mathfrak{K}^x$: Dies liegt auf der Hand für $B_2 = (b_1 + 1, 1)$. Ist $B_2 \neq (b_1 + 1, 1)$, so sind also $P, B_1, (b_1 + 1, 1), B_2$ paarweise verschieden und damit paarweise nicht parallel. Mit Hilfssatz 4. 5 folgt nun

$$\mathfrak{K}^x \ni \begin{bmatrix} P & B_1 \\ B_2 & (b_1 + 1, 1) \end{bmatrix} = b_2 - b_1 = \begin{bmatrix} P & A_1 \\ B_2 & A_2 \end{bmatrix} - \begin{bmatrix} P & A_1 \\ B_1 & A_2 \end{bmatrix}.$$

Satz 4. 3. Es seien $W, A_1, A_2, A_3, P_1, P_2, P_3$ verschiedene Punkte mit

- 1) $W \neq A_1, A_2, A_3, P_1, P_2, P_3$,
- 2) A_1, A_2, A_3 sind paarweise nicht parallel,
- 3) P_1, P_2, P_3 sind paarweise nicht parallel,
- 4) $A_i \neq P_j$ für $i \neq j$ und $i, j \in \{1, 2, 3\}$.

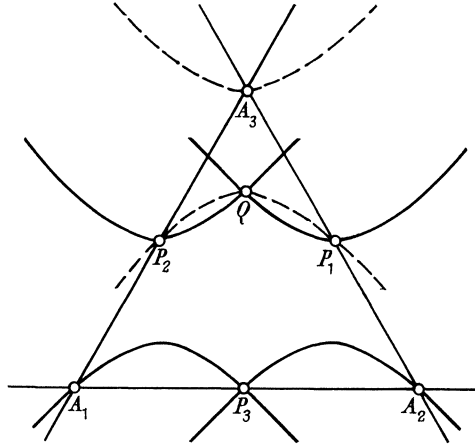
7) $\mathfrak{K}^x$ bezeichnet die multiplikative Gruppe des Körpers $\mathfrak{K}$.

Es seien $a_1, a_2, a_3, p_1, p_2, p_3$ pseudo-euklidische Kreise mit

$$P_i \in a_i \equiv (WA_{i+1}A_{i+2})$$

$$p_i \equiv (A_i P_{i+1} P_{i+2}),$$

wobei die Indizes — modulo 3 genommen — über 1, 2, 3 laufen. Besitzen dann die pseudo-euklidischen Kreise p_2 und p_3 den Durchschnitt $\{P_1, Q\}$, so folgt $Q \in p_1$ ⁸).



Figur 2. Satz von Miquel

Beweis. 1. Fall: $Q \neq P_1$. Wegen $P_1, Q \in p_2$ folgt nach Satz 1.3, (2) $P_1 \neq Q$. Die pseudo-euklidischen Kreise a_1, a_2, a_3 sind paarweise verschieden, da $p_2 \cap p_3 = \{P_1, Q\}$ ist: Im Falle $a_1 = a_2$ wäre $\{W, P_1, P_2, A_1, A_2, A_3\}$ konzyklisch, was auf $a_1 = a_2 = a_3 = p_3$ führte und mit $P_3 \in a_3$ dann $p_2 = p_3$ ergäbe. Mutatis mutandis ergeben die Fälle $a_2 = a_3$, $a_3 = a_1$ Widersprüche. Der Satz 4.3 liegt auf der Hand für $Q \in \{A_1, P_2, P_3\}$, da $p_1 = (A_1 P_2 P_3)$ ist. Sei $Q \notin \{A_1, P_2, P_3\}$. Wegen $Q, P_2 \in p_3$ und $Q, P_3 \in p_2$ ist dann $Q \neq P_2, P_3$. Es ist $Q \neq A_2$ und $Q \neq A_3$; im Falle $Q = A_2$ bzw. $Q = A_3$ hätte man $(A_3 P_1 P_2) = p_3 = (A_3 P_1 A_2) = a_1$ bzw. $(A_2 P_3 P_1) = p_2 = (A_2 A_3 P_1) = a_1$, was $P_2 \in a_1$ bzw. $P_3 \in a_1$ ergäbe, wo aber $a_2 \cap a_1 = \{W, A_3\}$ bzw. $a_3 \cap a_1 = \{W, A_2\}$ ist. — Wegen $Q, A_2 \in p_2$ folgt $Q \neq A_2$, und wegen $Q, A_3 \in p_3$ hat man $Q \neq A_3$. Nun ist

$$\begin{bmatrix} A_1 & Q \\ P_3 & P_2 \end{bmatrix} \equiv \begin{bmatrix} P_1 & P_3 \\ Q & A_2 \end{bmatrix} \begin{bmatrix} P_1 & P_2 \\ A_3 & Q \end{bmatrix} \begin{bmatrix} W & P_1 \\ A_3 & A_2 \end{bmatrix} \begin{bmatrix} W & P_2 \\ A_1 & A_3 \end{bmatrix} \begin{bmatrix} W & P_3 \\ A_2 & A_1 \end{bmatrix}.$$

Da jedes der Quadrupel $P_1 P_3 A_2 Q$, $P_1 P_2 Q A_3$, $W P_1 A_2 A_3$, $W P_2 A_3 A_1$, $W P_3 A_1 A_2$ aus vier paarweise nicht parallelen Punkten besteht (s. die Voraussetzungen und das ergänzend Gezeigte) und da außerdem jedes dieser Quadrupel konzyklisch ist, so ergibt die aufgeschriebene Identität zusammen mit Hilfssatz 4.5 offenbar $\begin{bmatrix} A_1 & Q \\ P_3 & P_2 \end{bmatrix} \in \mathfrak{K}$. Da A_1, P_2, P_3 paarweise nicht parallel sind und außerdem auch Q, P_2, P_3 , so folgt jetzt mit Hilfssatz 4.6 — wenn wir aus $Q \notin \{A_1, P_2, P_3\}$ noch $Q \neq A_1$ entnehmen —, daß $A_1 P_3 P_2 Q$ konzyklisch ist, d. h. daß $Q \in p_1$ gilt.

2. Fall: $Q = P_1$. Es gilt also $p_2 \cap p_3 = \{P_1\}$. Da P_3, A_2 verschiedene Punkte auf $p_2 - \{P_1\} = p_2 - p_3$ sind, und P_2, A_3 verschiedene Punkte auf $p_3 - \{P_1\} = p_3 - p_2$

⁸) Hier ist natürlich der Fall $P_1 = Q$ zugelassen. Er stellt einen wichtigen Entartungsfall dar.

sind, so ergibt Satz 4. 2

$$\begin{bmatrix} P_1 & P_3 \\ A_3 & A_2 \end{bmatrix} - \begin{bmatrix} P_1 & P_3 \\ P_2 & A_2 \end{bmatrix} \in \mathfrak{R}.$$

Die Identität

$$\begin{bmatrix} A_1 & P_1 \\ P_3 & P_2 \end{bmatrix} \equiv \left(\begin{bmatrix} P_1 & P_3 \\ A_3 & A_2 \end{bmatrix} - \begin{bmatrix} P_1 & P_3 \\ P_2 & A_2 \end{bmatrix} \right) \begin{bmatrix} W & P_1 \\ A_3 & A_2 \end{bmatrix} \begin{bmatrix} W & P_2 \\ A_1 & A_3 \end{bmatrix} \begin{bmatrix} W & P_3 \\ A_2 & A_1 \end{bmatrix}$$

führt damit und mit Hilfssatz 4. 5 auf

$$\begin{bmatrix} A_1 & P_1 \\ P_3 & P_2 \end{bmatrix} \in \mathfrak{R}.$$

Hilfssatz 4. 6 ergibt dann $Q = P_1 \in p_1$, wenn wir noch $P_1 \neq A_1$ beachten.

§ 5. Kongruente, äquiforme Abbildungen, Distanzen, Öffnungen.

Pseudo-euklidische Winkel. Orthogonalität

Im ersten Teil dieses Paragraphen betrachten wir nur eigentliche *pseudo-euklidische Punkte* und *pseudo-euklidische Geraden*; eine pseudo-euklidische Gerade ist dabei ein pseudo-euklidischer Kreis durch den uneigentlichen pseudo-euklidischen Punkt ∞ , wobei wir allerdings bei unseren Betrachtungen jetzt den Punkt ∞ nicht mit zu den pseudo-euklidischen Geraden rechnen wollen.

Wir stellen erweiternd zusammen: Sei $\mathfrak{R}$ ein kommutativer Körper von Char $\neq 2$, sei $\mathfrak{L}$ der Ring der anormal-komplexen Zahlen über $\mathfrak{R}$. Genau die Elemente von $\mathfrak{L}$ nennen wir *Punkte*. Der Punkt P heißt genau dann *parallel* zum Punkt Q , in Zeichen $P \parallel Q$, wenn $Q - P$ kein Inverses im Ring $\mathfrak{L}$ besitzt. Die Parallelitätsrelation ist reflexiv und symmetrisch. Sind P, Q nicht parallele Punkte, so heiße die Punktmenge

$$(PQ) = \{P + \lambda(Q - P) \mid \lambda \in \mathfrak{R}\}$$

eine *Gerade*. Die Gerade (PQ) heiße *parallel* zur Geraden (RS) genau dann, wenn $S - R = k(Q - P)$ ist mit $k \in \mathfrak{R}$. Sind P, Q Punkte, so werde unter der *Distanz* $d(P, Q)$ der Punkte P, Q das folgende Element aus $\mathfrak{R}$ verstanden

$$d(P, Q) = (Q - P) (\overline{Q - P}) \equiv (q_1 - p_1)^2 - (q_2 - p_2)^2;$$

dabei sei $P = p_1 + p_2j$, $Q = q_1 + q_2j$ mit $p_1, p_2, q_1, q_2 \in \mathfrak{R}$.

Hilfssatz 5. 1. Genau dann gilt $d(P, Q) = 0$, wenn $P \parallel Q$ ist.

Beweis. Aus $(Q - P) (\overline{Q - P}) = 0$ folgt $Q - P \notin \mathfrak{R}$, d. h. $P \parallel Q$. Vice versa bedeutet $d(P, Q) \neq 0$ wegen $d(P, Q) \in \mathfrak{R}$ offenbar $(Q - P) (\overline{Q - P}) \in \mathfrak{R}^\times$, d. h. $Q - P \in \mathfrak{R}$, d. h. $P \neq Q$.

Unter einer *kongruenten Abbildung* unserer über $\mathfrak{R}$ definierten Geometrie verstehen wir eine eindeutige Abbildung α der Menge der Punkte in die Menge der Punkte mit $d(P^\alpha, Q^\alpha) = d(P, Q)$ für alle Paare P, Q .

Satz 5. 1. Genau die Abbildungen

$$w = az + b$$

$$w = a\bar{z} + b$$

mit $a, b \in \mathfrak{L}$, $a\bar{a} = 1$, sind die kongruenten Abbildungen. Kongruente Abbildungen sind also bijektiv, geradentreu und parallelentreu (sowohl auf der Menge der Punkte als auch auf der Menge der Geraden).

Beweis. a) Gewiß sind die Abbildungen $w = az + b$, $w = a\bar{z} + b$ mit $a, b \in \mathfrak{L}$, $a\bar{a} = 1$, kongruente Abbildungen, die überdies bijektiv sind.

b) Sei α eine beliebige kongruente Abbildung. Aus $d(0^\alpha, 1^\alpha) = d(0, 1) = 1$ folgt $a\bar{a} = 1$, wenn $a = 1^\alpha - 0^\alpha$ gesetzt wird. Die bijektive kongruente Abbildung $w = az + 0^\alpha$ werde β genannt. Damit ist auch $\gamma = \alpha\beta^{-1}$ eine kongruente Abbildung. Es gilt $0^\gamma = 0$, $1^\gamma = 1$.

c) Aus $d(0, z) = d(0, z^\gamma)$ folgt $z\bar{z} = z^\gamma\bar{z}^\gamma$ und aus $d(1, z) = d(1, z^\gamma)$ folgt

$$(z - 1)(\bar{z} - 1) = (z^\gamma - 1)(\bar{z}^\gamma - 1).$$

Also hat man auch $z + \bar{z} = z^\gamma + \bar{z}^\gamma$; dies bedeutet $x = u$, wenn wir $z = x + jy$, $w = u + jv$ mit $x, y, u, v \in \mathfrak{R}$ setzen. Mit $x^2 - y^2 = u^2 - v^2$ gilt also auch $y^2 = v^2$. Damit ist $z^\gamma \in \{z, \bar{z}\}$. Insbesondere gilt $j^\gamma \in \{j, \bar{j}\}$.

1. Fall: $j^\gamma = j$. Wir zeigen, daß in diesem Falle stets $z^\gamma = z$ gilt für alle $z \in \mathfrak{L}$: Aus $d(j, z) = d(j, z^\gamma)$ folgt $(z - j)(\bar{z} + j) = (z^\gamma - j)(\bar{z}^\gamma + j)$, d. h. $(z - \bar{z})j = (z^\gamma - \bar{z}^\gamma)j$, d. h. $z - \bar{z} = z^\gamma - \bar{z}^\gamma$. Also gilt $y = v$ und damit $z^\gamma = z$.

Im ersten Fall hat somit α die Gestalt

$$w = (1^\alpha - 0^\alpha)z + 0^\alpha.$$

2. Fall: $j^\gamma = -j$. In diesem Falle ist stets $z^\gamma = \bar{z}$. Aus $d(j, z) = d(-j, z^\gamma)$ folgt nämlich $y = -v$, d. h. $z^\gamma = \bar{z}$. Im zweiten Fall hat damit α die Gestalt

$$w = (1^\alpha - 0^\alpha)\bar{z} + 0^\alpha.$$

d) Also sind genau die Abbildungen $w = az + b$ und $w = a\bar{z} + b$ mit $a, b \in \mathfrak{L}$, $a\bar{a} = 1$, die kongruenten Abbildungen. Die weiteren Aussagen von Satz 5. 1 sind leicht zu bestätigen. Teile hiervon liegen übrigens schon aufgrund früherer Überlegungen auf der Hand.

Gegeben sei ein geordnetes Punktetripel S, A, B . Ein solches geordnetes Punktetripel S, A, B nennen wir eine *Ecke*. Die Ecke S, A, B heiße *regulär*, wenn $S \# A, B$ ist. Unter der *Öffnung* $o(S, A, B)$ der Ecke S, A, B verstehen wir die Elementklasse $(B - S)(\overline{A - S})\mathfrak{R}^x$.

Eine eindeutige Abbildung α der Menge der Punkte in sich heiße eine *äquiiforme Abbildung 1. Art*, wenn sie öffnungstreu ist, und sie heiße eine *äquiiforme Abbildung 2. Art*, wenn sie Öffnungen konjugiert.

Satz 5. 2. *Genau die Abbildungen*

$$w = az + b$$

mit $a, b \in \mathfrak{L}$, $a \in \mathfrak{R}$, sind die äquiiformen Abbildungen 1. Art. Genau die Abbildungen

$$w = a\bar{z} + b$$

mit $a, b \in \mathfrak{L}$, $a \in \mathfrak{R}$, sind die äquiformen Abbildungen 2. Art. Äquiforme Abbildungen sind also bijektiv, geradentreu und paralleltreu (sowohl auf der Menge der Punkte als auch auf der Menge der Geraden).

Beweis. a) Gewiß sind die Abbildungen $w = az + b$ mit $a, b \in \mathfrak{L}$, $a \in \mathfrak{R}$, äquiforme Abbildungen 1. Art, und gewiß sind die Abbildungen $w = a\bar{z} + b$ mit $a, b \in \mathfrak{L}$, $a \in \mathfrak{R}$, äquiforme Abbildungen 2. Art. Die genannten Abbildungen sind darüber hinaus alle bijektiv.

b) Sei α eine beliebige äquiforme Abbildung 1. Art (bzw. 2. Art). Aus

$$o(0^\alpha, 1^\alpha, j^\alpha) = o(0, 1, j) = j \mathfrak{R}^x$$

(bzw. $o(0^\alpha, 1^\alpha, j^\alpha) = \overline{o(0, 1, j)} = \overline{j \mathfrak{R}^x} = +j \mathfrak{R}^x$) folgt $1^\alpha - 0^\alpha \in \mathfrak{R}$. Die bijektive äquiforme Abbildung $w = (1^\alpha - 0^\alpha)z + 0^\alpha$ werde β genannt. Damit ist auch $\gamma = \alpha\beta^{-1}$ eine äquiforme Abbildung 1. Art (bzw. 2. Art). Es gilt $0^\gamma = 0$, $1^\gamma = 1$.

c) Aus $o(0, 1, z) = o(0, 1, z')$ (bzw. $\overline{o(0, 1, z)} = \overline{o(0, 1, z')}$) folgt $z' = k(z) \cdot z$ mit $k(z) \in \mathfrak{R}^x$ (bzw. $z' = k'(z) \cdot \bar{z}$ mit $k'(z) \in \mathfrak{R}^x$). Aus $o(1, 0, z) = o(1, 0, z')$ (bzw. $\overline{o(1, 0, z)} = \overline{o(1, 0, z')}$) folgt $z' - 1 = k_1(z)(z - 1)$ mit $k_1(z) \in \mathfrak{R}^x$ (bzw. $z' - 1 = k'_1(z)(\bar{z} - 1)$). Zusammengefaßt hat man $[k(z) - k_1(z)]z = 1 - k_1(z)$ (bzw. $[k'(z) - k'_1(z)]\bar{z} = 1 - k'_1(z)$). Sei nun $z \notin \mathfrak{R}$. Dann folgt hieraus $k(z) = k_1(z) = 1$ (bzw. $k'(z) = k'_1(z) = 1$). Also gilt $z' = z$ (bzw. $z' = \bar{z}$) für alle $z \in \mathfrak{L} - \mathfrak{R}$.

Wir zeigen nun, daß auch $z' = z$ (bzw. $z' = \bar{z}$) für alle $z \in \mathfrak{R}$ gilt. Wir haben also $z' = z$ für alle $z \in \mathfrak{R} - \{0, 1\}$ zu zeigen. Wegen $j \in \mathfrak{L} - \mathfrak{R}$ gilt $j' = j$ (bzw. $j' = -j$). Aus $o(j, 0, z) = o(j, 0, z')$ (bzw. $\overline{o(j, 0, z)} = \overline{o(-j, 0, z')}$) folgt $z' - j = \hat{k}(z)(z - j)$ mit $\hat{k}(z) \in \mathfrak{R}^x$ (bzw. $z' + j = \hat{k}'(z)(\bar{z} + j)$ mit $\hat{k}'(z) \in \mathfrak{R}^x$). Mit $z' = k(z) \cdot z$ (bzw. $z' = k'(z) \cdot \bar{z}$) zusammengefaßt hat man

$$[k(z) - \hat{k}(z)]z = j(1 - \hat{k}(z)) \quad (\text{bzw. } [k'(z) - \hat{k}'(z)]\bar{z} = j(-1 + \hat{k}'(z))).$$

Mit $z \in \mathfrak{R}^x$, $j \notin \mathfrak{R}$ folgt hieraus $1 = \hat{k}(z) = k(z)$ (bzw. $+1 = \hat{k}'(z) = k'(z)$). Damit gilt auch für $z \in \mathfrak{R}$ offenbar $z' = z$ (bzw. $z' = \bar{z}$).

d) Also sind genau die Abbildungen $w = az + b$ (bzw. $w = a\bar{z} + b$) mit $a, b \in \mathfrak{L}$, $a \in \mathfrak{R}$, die äquiformen Abbildungen 1. Art (bzw. 2. Art). Die weiteren Aussagen von Satz 5. 2 sind leicht zu bestätigen.

Für die Zwecke der pseudo-euklidischen Kreisgeometrie werden wir jetzt einen *Winkelbegriff* einführen, der an die Seite zu stellen ist dem Begriff des orientierten Kreiswinkels in der Möbius-Geometrie. Dort versieht man im klassischen Falle die Kugel mit einer Indikatrix und mißt die Winkel mod π längs der durch die Indikatrix gegebenen Orientierung; als zugehörige Gruppe im Sinne des Erlanger Programms hat man die Gruppe der orientierungstreuen äquiformen Abbildungen. An die Stelle dieser Gruppe hat bei unserem allgemeinen Aufbau die Gruppe der äquiformen Abbildungen 1. Art zu treten.

Unter einem *Winkel* verstehen wir ein geordnetes Geradenpaar g, h zweier sich schneidender pseudo-euklidischer Geraden, mitsamt einem Schnittpunkt S dieser Geraden, in Zeichen $(g h; S)$. Die Gruppe Γ sei die Gruppe der äquiformen Abbildungen 1. Art. Die Winkel $(g h; S)$, $(g' h'; S')$ heißen genau dann gleich, wenn es ein $\gamma \in \Gamma$ mit $g^\gamma = g'$, $h^\gamma = h'$, $S^\gamma = S'$ gibt.

Satz 5.3. Γ ist eine Süß'sche Gruppe, die überdies maximal und damit involutorisch ist⁹⁾. Dabei sei $|\mathfrak{K}| > 3$.

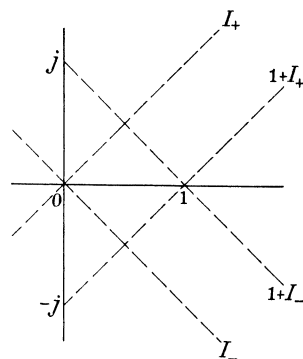
Beweis. Aufgrund der Sätze 2.1 und 3.14 in Benz [5] und des Satzes 5.2 der vorliegenden Arbeit verbleibt noch der Nachweis der Maximalität von Γ . Angenommen, es gäbe eine Automorphismengruppe Δ unserer Geometrie über $\mathfrak{K}$, die Γ echt umfaßt und die Süß'sche Gruppe ist. Sei $\delta \in \Delta - \Gamma$ und sei $0^\delta = A$, $1^\delta = B$. Wegen $0 \neq 1$ ist $A \neq B$. Sei $\gamma \in \Gamma$ die eindeutig bestimmte Abbildung aus Γ mit $A^\gamma = 0$, $B^\gamma = 1$. Dann ist $\delta' \equiv \delta\gamma \in \Delta - \Gamma$ eine Abbildung, die 0, 1 als Fixpunkte besitzt. Wenn wir zeigen können, daß δ' die Identität sein muß, so haben wir einen Widerspruch. Sei z ein Punkt (z. B. $z = 2j$) mit

$$z \notin \mathfrak{K} \cup I_+ \cup I_- \cup (1 + I_+) \cup (1 + I_-).$$

Also ist $0 \neq z$ und $1 \neq z$. Damit existieren die Geraden $g = (0z)$ und $h = (1z)$. Es gilt $g \cap h = \{z\}$. Da δ' die Punkte 0, 1 festläßt und als Ganzes auch die Gerade (01) , so gilt $g^{\delta'} = g$ und $h^{\delta'} = h$, da $\Delta \ni \delta'$ Süß'sche Gruppe ist. Also gilt, da Δ nur aus Automorphismen unserer Geometrie besteht,

$$\{z^{\delta'}\} = g^{\delta'} \cap h^{\delta'} = g \cap h = \{z\}, \text{ d. h. } z^{\delta'} = z.$$

Die Betrachtung, fortgeführt unter Benutzung weiterer Fixpunkte, führt zum Ziel für $|\mathfrak{K}| > 3$. Im Falle $|\mathfrak{K}| = 3$ liegt keine Maximalität vor, wie Herr W. Leißner feststellt.



Figur 3

Bemerkung zu Satz 5.2. Nach in Benz [5] durchgeführten Betrachtungen ist also die größte Automorphismengruppe der Geometrie über $|\mathfrak{K}| > 3$, die die eingeführte Winkelvergleichung invariant läßt, genau die Gruppe Γ .

Die vorliegende Geometrie ist eine affine Parallelstruktur mit Nachbarelementen besonderer Art: In ihr gibt es nur triviale Schmiegegeradenpaare¹⁰⁾, d. h. die Geraden g, h sind dann und nur dann Schmiegegeraden, wenn $g = h$ ist. Zwei sich schneidende Geraden sind also identisch oder aber Kreuzgeraden.

Da Γ Süß'sche Gruppe ist, geht durch Spezialisierung aus Benz [5] hervor

Satz 5.4. (1) Die über Γ definierte Gleichheitsrelation auf der Menge der Winkel ist eine Äquivalenzrelation. (Maximale Klassen gleicher Winkel nennen wir freie Winkel.)

(2) Gegeben seien ein Winkel $(a, b; P)$, ein Punkt Q , eine Gerade g durch Q , Dann gibt es eine und nur eine Gerade h durch Q mit $(ab; P) = (gh; Q)$.

(3) Sei g eine Gerade, seien P, Q Punkte auf g , seien p eine Gerade durch P und q eine Gerade durch Q . Dann folgt $(gp; P) = (gq; Q)$ genau dann, wenn p zu q parallel, in Zeichen $p \parallel q$, ist.

(4) Die Menge $\mathfrak{B}$ der freien Winkel bildet bezüglich der (natürlichen) Addition¹¹⁾ von Winkeln eine Gruppe, die isomorph zu $\mathfrak{K}/\mathfrak{K}^x$ ist.

⁹⁾ Für die Begriffe Süß'sche Gruppe, maximale Süß'sche Gruppe, involutorische Gruppe s. Benz [5]. Überhaupt liegt in unserer Punkt-Geraden-Geometrie eine affine Parallelstruktur mit Nachbarelementen (Nachbarrelation = Parallelitätsrelation auf der Menge der Punkte) vor, s. Benz [5], Satz 2.1. Die zwei Postulate an eine Süß'sche Gruppe sind (abgeschwächte) Abwandlungen auf den affinen Fall mit Nachbarelementen von Forderungen, die W. Süß [19] in möbiusgeometrischem Zusammenhang erhebt.

¹⁰⁾ Die Begriffe Schmiegegeraden, Kreuzgeraden gehen auf Hjelmslev [11 d], S. 5, zurück.

¹¹⁾ S. Benz [5].

(5) In einem Dreieck ABC ; abc^{12} seien α, β, γ die freien Winkel mit $\alpha \ni (bc; A)$, $\beta \ni (ca; B)$, $\gamma \ni (ab; C)$. Dann gilt $\alpha + \beta + \gamma = 0^{13}$.

Ist $(ab; P)$ ein Winkel, so gibt es, da eine affine Parallelstruktur m. N. zugrunde liegt, eine reguläre Ecke P, A, B mit $A \in a, B \in b$. Wir wollen $o(P, A, B)$ die Größe des Winkels $(ab; P)$ nennen. Man überzeugt sich leicht davon, daß $o(P, A, B) = o(P, A', B')$ ist, wenn A', B' beliebige Punkte sind mit $A' \in a, B' \in b, P, A', B'$ ist reguläre Ecke.

Hilfssatz 5.2. Die Winkel $(ab; P), (gh; S)$ sind genau dann gleich, wenn sie von gleicher Größe sind.

Beweis. Sind $(ab; P), (gh; S)$ gleiche Winkel, so sind sie auch von gleicher Größe, da die Abbildungen aus Γ öffnungstreu sind. Seien umgekehrt $(ab; P), (gh; S)$ Winkel gleicher Größe, $o(P, A, B) = o(S, G, H)$, P, A, B und S, G, H reguläre Ecken mit $A \in a, B \in b, G \in g, H \in h$. Sei $\gamma \in \Gamma$ die Abbildung mit $P^\gamma = S, A^\gamma = G$, d. h. die Abbildung

$$w = (S - G)(P - A)^{-1}(z - P) + S.$$

Jedenfalls gilt dann $(ab; P) = (g(SB^\gamma); S)$. Wir haben also noch $h = (SB^\gamma)$ nachzuweisen. Aus $o(P, A, B) = o(S, G, H)$ und $o(P, A, B) = o(S, G, B^\gamma)$ folgt aber $B^\gamma - S = \lambda(H - S)$ mit $\lambda \in \mathbb{R}^x$, d. h. $B^\gamma \in h$.

Ein Winkel heißt ein rechter Winkel, wenn der zugehörige freie Winkel φ eine Involution in $\mathfrak{X}$ ist, d. h. wenn $\varphi \neq 0$ und $\varphi + \varphi = 0$ ist. Wieviel Involutionen φ besitzt $\mathfrak{X} \cong \mathfrak{R}/\mathfrak{R}^x$? Offenbar genau eine, nämlich $j\mathfrak{R}^x$. Damit gibt es genau einen freien rechten Winkel, in Zeichen R . Indem wir den Begriff des Lotes und den Begriff des Senkrechstehens einführen, haben wir

Satz 5.5. (1) Ist g eine Gerade, ist P ein Punkt, so gibt es genau ein Lot durch P auf g , d. h. genau eine Gerade h durch P , die h in einem einzigen Punkt, wir nennen ihn S , schneidet mit $(gh; S) \in R$.

(2) Ist $(gh; S)$ ein rechter Winkel, wir schreiben dann auch $g \perp h$, so ist der zugehörige „Nebenwinkel“ $(hg; S)$ ebenfalls ein rechter Winkel. Aus $g \perp h$ folgt also $h \perp g$.

(3) Ist $g \perp h, h \parallel h'$, so ist auch $g \perp h'$.

Beweis zu (1). Da die äquiformen Abbildungen 1. Art winkeltreu sind, so können wir die spezielle Situation $g = (0 \ 1)$ annehmen. Sei $P = a + bj$, $a, b \in \mathfrak{R}$. Dann gilt für die Gerade h durch $S = a$ und $a + j$ offenbar $(gh; S) \in R$ wegen $o(a, a + 1, a + j) = j\mathfrak{R}^x$. Ist nun h' ein weiteres Lot durch P auf g , so sei S' der Schnittpunkt von g und h' . Es ist also $S' \in \mathfrak{R}$ und $o(S', S' + 1, H') = j\mathfrak{R}^x$ mit einem $H' \in h, H' \neq S'$. Demnach gilt $H' - S' = \lambda j$ mit einem $\lambda \in \mathfrak{R}^x$. Wegen $P \in h'$ und $h' = (S'H')$ ist also

$$P = S' + \mu(H' - S') = S' + \mu\lambda j;$$

$S' \in \mathfrak{R}$ bedeutet also $S' = a = S$, was zusammen mit

$$a + j = a + \frac{1}{\lambda} \cdot \lambda j = S' + \frac{1}{\lambda} (H' - S') \in h'$$

auf $h' = h$ führt. Auf die Angabe der einfachen Beweise zu (2) und (3) verzichten wir.

¹² Hier seien A, B, C paarweise nicht parallele Punkte und a, b, c resp. die Verbindungsgeraden $(BC), (CA), (AB)$.

¹³ 0 bezeichne das neutrale Element von $\mathfrak{R}/\mathfrak{R}^x$.

Im Hinblick auf die äquiformen Abbildungen 2. Art geben wir noch an

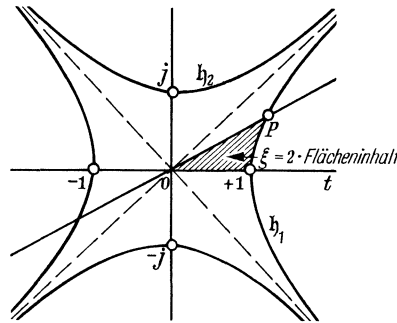
Satz 5. 6. *Die äquiformen Abbildungen 2. Art sind ebenfalls rechtwinkeltreu, d. h. sie überführen rechte Winkel in rechte Winkel. Weiterhin gilt für jede solche Abbildung α und jeden Winkel $(ab; P)$ die Aussage $(a^\alpha b^\alpha; P^\alpha) = (ba; P)$, d. h. die Orientierung der Winkel wird geändert.*

Der Beweis liegt auf der Hand, wenn man bedenkt, daß äquiforme Abbildungen 2. Art Öffnungen konjugieren.

Bevor wir auch Winkel im Bereiche der pseudo-euklidischen Kreise betrachten, wollen wir den Fall $\mathfrak{K} = \mathbf{R}$ näher ansehen. Dazu sei

$$\begin{aligned}\mathfrak{h}_1 &= \{x + yj \mid x, y \in \mathbf{R} \text{ und } x^2 - y^2 = 1 \text{ und } x > 0\}, \\ \mathfrak{h}_2 &= \{x + yj \mid x, y \in \mathbf{R} \text{ und } y^2 - x^2 = 1 \text{ und } y > 0\}, \\ \mathfrak{h} &= \mathfrak{h}_1 \cup \mathfrak{h}_2.\end{aligned}$$

Ist nun t die Gerade $(0 \ 1)$ und ist g eine pseudo-euklidische Gerade durch 0, so ist die Größe des Winkels $(tg; 0)$ durch PR^x gegeben; dabei wollen wir unter P den eindeutig bestimmten Schnittpunkt von g und $\mathfrak{h}$ verstehen.



Figur 4

Sei $P = x + yj$, $x, y \in \mathbf{R}$. Wir bestimmen jetzt die eindeutige Lösung ξ von (a) im Falle (b); dabei sei

- (a) $x = \mathfrak{Cof} \xi$, $\xi \geq 0$ für (b) $P \in \mathfrak{h}_1$, $y \geq 0$,
- (a) $x = \mathfrak{Cof} \xi$, $\xi < 0$ für (b) $P \in \mathfrak{h}_1$, $y < 0$,
- (a) $y = \mathfrak{Cof} \xi$, $\xi \geq 0$ für (b) $P \in \mathfrak{h}_2$, $x \geq 0$,
- (a) $y = \mathfrak{Cof} \xi$, $\xi < 0$ für (b) $P \in \mathfrak{h}_2$, $x < 0$.

Im Falle $P \in \mathfrak{h}_1$ haben wir $P = \mathfrak{Cof} \xi + j \mathfrak{Sin} \xi = e^{j\xi}$; im Falle $P \in \mathfrak{h}_2$ gilt

$$P = x + yj = j(y + xj) = je^{j\xi}.$$

Damit ist im Falle $P \in \mathfrak{h}_1$ die Größe des Winkels $(tg; 0)$ offenbar $e^{j\xi} R^x$; entsprechend ist im Falle $P \in \mathfrak{h}_2$ die Größe von $(tg; 0)$ durch $je^{j\xi} R^x$ gegeben. Bis auf Isomorphie ist also nach Satz 5. 4, (2) die Winkelgruppe $\mathfrak{W}$ gegeben durch

$$\mathfrak{W} \cong \{e^{j\xi} R^x \mid \xi \in \mathbf{R}\} \cup \{je^{j\xi} R^x \mid \xi \in \mathbf{R}\} \cong \{e^{j\xi} \mid \xi \in \mathbf{R}\} \cup \{je^{j\xi} \mid \xi \in \mathbf{R}\} \cong \mathbf{R}^+ \otimes \mathfrak{S}_2,$$

wobei $\mathbf{R}^+$ die additive Gruppe des Körpers der reellen Zahlen ist, und wo $\mathfrak{S}_2$ die symmetrische Gruppe über 2 Gegenständen bezeichnet. — Eine anschauliche Deutung von ξ entnehmen wir der elementaren Theorie (s. auch Figur 4).

Gegenüber der Multiplikation bildet also die Punktmenge $\mathfrak{h} = \mathfrak{h}_1 \cup \mathfrak{h}_2$ eine Gruppe, die isomorph zur Winkelgruppe $\mathfrak{W}$ ist. $\mathfrak{h}_1$ ist eine Untergruppe, die zur homogenen Lorentzgruppe¹⁴⁾ der zweidimensionalen Raum-Zeit-Welt isomorph ist: Ist c die Lichtgeschwindigkeit, so können wir das offene Intervall $(-c, c)$ eindeutig auf $\mathbf{R}$ beziehen vermöge

$$\xi \in \mathbf{R} \text{ entspreche}^{15)} -c \Im \xi \in (-c, +c).$$

Ist nun v eine konstante Geschwindigkeit aus $(-c, +c)$, so läßt sich die homogene Lorentztransformation (geschrieben in sonst üblichen Bezeichnungen)

$$x' = \frac{x - vt}{\sqrt[>0]{1 - \frac{v^2}{c^2}}}, \quad t' = \frac{t - \frac{v}{c^2}x}{\sqrt[>0]{1 - \frac{v^2}{c^2}}}$$

vermöge $z = x + ctj$, $w = x' + ct'j$ schreiben in der Gestalt

$$w = e^{\xi j} z.$$

Die Gruppe der homogenen Lorentztransformationen ist also außerdem noch eine Untergruppe der Gruppe der kongruenten Abbildungen. Unter Benutzung des Begriffs der Drehung liegt also eine Gruppe von Drehungen — nicht aller Drehungen — um den Punkt 0 vor.

Wir kommen jetzt zum Winkelbegriff im Bereich der pseudo-euklidischen Kreise. Die uneigentlichen Punkte sind bei den folgenden Betrachtungen wieder zugelassen. Seien a, b pseudo-euklidische Kreise, die wenigstens 2 verschiedene Punkte gemeinsam haben. Ist $P \in a \cap b$, so heiße das geordnete Tripel $(ab; P)$ ein *Winkel*. Wir ordnen dem Winkel $(ab; P)$ eine Größe zu:

1. Fall: P eigentlich.

a' bzw. b' seien die Tangenten in P an a bzw. b . Wir haben dabei den hier gebrauchten Begriff der Tangente zu erklären, etwa im Falle a' : Geht a durch den Punkt ∞ , so sei $a' = a$; ist $\infty \notin a$, so gilt $\infty \neq P$ nach Satz 2. 4, a), was nach Satz 1. 3, (4) auf einen eindeutig bestimmten Kreis $a' \ni P, \infty$ mit $a' \cap a = \{P\}$ führt. — Nun sei als die Größe von $(ab; P)$ die Größe von $(a'b'; P)$ verstanden¹⁶⁾.

2. Fall: P uneigentlich.

In den Fällen $P = \infty$, $P = (k, 1 \pm j)$ mit $k \in \mathbb{R}^x$ sei α die Abbildung $(x_1, x_2) \rightarrow (x_2, x_1)$ aus T. In den verbleibenden Fällen $P = (1 + j, 1 - j)$, $P = (1 - j, 1 + j)$ sei α die Abbildung $(x_1, x_2) \rightarrow (x_2, x_1 + x_2)$ aus T. In jedem dieser Fälle ist dann P^α ein eigentlicher Punkt; unter der Größe von $(ab; P)$ sei nun die schon definierte Größe von $(a^\alpha b^\alpha; P^\alpha)$ verstanden.

Zwei Winkel sollen *gleich* heißen, wenn sie von gleicher Größe sind. Es gilt

¹⁴⁾ S. z. B. Strubecker [15].

¹⁵⁾ S. Strubecker [15].

¹⁶⁾ Ist k ein pseudo-euklidischer Kreis durch ∞ , so bezeichnen wir die pseudo-euklidische Gerade $k - \{\infty\}$ auch durch $\underline{k}$.

Satz 5. 7. (1) Die Winkel $(ab; P)$, $(cd; Q)$ sind genau dann gleich, wenn es eine Abbildung $\tau \in T$ gibt mit $a^\tau = c$, $b^\tau = d$, $P^\tau = Q$.

(2) Sei $(ab; P)$ ein Winkel. Da $|a \cap b| \geq 2$ gilt, hat man die Fälle $a = b$ und $|a \cap b| = 2$. Im Falle $a = b$ ist $\mathbb{R}^x$ die Größe von $(ab; P)$. Im Falle $|a \cap b| = 2$ ist

$$\begin{bmatrix} Q & P \\ B & A \end{bmatrix} \mathbb{R}^x$$

die Größe von $(ab; P)$, wenn $\{P, Q\} = a \cap b$ ist und A, B irgendwelche Punkte mit $A \in a - b$, $B \in b - a$ sind.

(3) Sei $(ab; P)$ ein Winkel, $a \neq b$, berühre der pseudo-euklidische Kreis r den Kreis b im Punkte P . Dann ist auch $(ar; P)$ ein Winkel und es gilt $(ab; P) = (ar; P)$. Ist umgekehrt $(ar; P)$ ein Winkel mit $(ab; P) = (ar; P)$, so gilt $b \cap r = \{P\}$ oder $b = r$.

Wir fassen weitere Winkleigenschaften¹⁷⁾ zusammen:

Satz 5. 8. (1) (Eigenschaft der Abtragbarkeit). Es seien vorgegeben ein Winkel $(xy; S)$, ein Kreis x' , verschiedene Punkte S', T' auf x' . Dann gibt es genau einen Kreis $y' \ni S', T'$ mit $(xy; S) = (x'y'; S')$.

(2) Sind x, y Kreise mit $|x \cap y| = 2$, $x \cap y = \{S, T\}$, so gilt

$$(xy; S) = (yx; T).$$

(3) (Gegenstück zum Satz über den Sehntangentenwinkel der Möbius-Geometrie). Sind A, B, C, D paarweise nicht parallele Punkte, so gilt

$$((ABC) (ABD); B) = ((DCB) (DCA); C).$$

(4) (Gegenstück zum Peripheriewinkelsatz der Möbius-Geometrie). Es seien A, B, P, Q verschiedene Punkte auf einem Kreis z . Es sei $W \notin z$ ein Punkt, der zu keinem der Punkte A, B, P, Q parallel ist. Dann gilt

$$((WPA) (WPB); P) = ((WQA) (WQB); Q).$$

(5) (Gegenstück zum Satz über die Winkelsumme im Kreisdreieck). Es seien W, A, B, C paarweise nicht parallele Punkte, nicht gemeinsam auf einem Kreis. Dann gilt $\alpha + \beta + \gamma = 0$, wenn α, β, γ die freien Winkel sind mit

$$\alpha \ni ((WAB) (WAC); A), \quad \beta \ni ((WBC) (WBA); B), \quad \gamma \ni ((WCA) (WCB); C).$$

(6) (Verallgemeinerung des Satzes von Miquel¹⁸⁾). Es seien A, B, C, D, E, F, G, H paarweise nicht parallele Punkte derart, daß die Quadrupel $AEHD, FBCE, AFCH, EBGD$ konzyklisch sind. Dann gilt

$$((ABC) (ABD); B) = ((EFH) (EFG); F).$$

Ins Einzelne gehende Angaben von Orthogonalitätseigenschaften übergehen wir. Wir merken aber an den

¹⁷⁾ Die Beweise sind mutatis mutandis die der Laguerre-Geometrie (s. Benz [6], [7]).

¹⁸⁾ Satz 4. 3 im Falle acht verschiedener Punkte.

Satz 5. 9. Es seien a, b Kreise mit $|a \cap b| = 2$. Dann sind die folgenden Aussagen gleichwertig:

(a) $a \perp b$.

(b) $\begin{bmatrix} P & Q \\ B & A \end{bmatrix}^2 \in \mathbb{R}$,

wo $a \cap b = \{P, Q\}$ ist und A, B Punkte sind mit $A \in a - b$, $B \in b - a$.

(c) Die zu a, b gehörenden Ebenen im Modell des § 3 sind bezüglich des Hyperboloids $\mathfrak{Q}$ konjugiert.

Unter Benutzung des Distanzbegriffes kann man den eigentlichen Teil der pseudo-euklidischen Kreise, die nicht durch ∞ gehen, mit Hilfe eines Mittelpunktbegriffes in der Form

$$\{X \mid d(X, M) = \gamma \neq 0\}$$

schreiben.

Man erhält leicht Eigenschaften (natürlich nicht zur vollen Automorphismengruppe gehörend), in die der Mittelpunktsgriff eingeht.

§ 6. Die Automorphismengruppen der Geometrie $A(\mathbb{R})$

Gegeben sei eine Geometrie $A(\mathbb{R})$. Es bezeichne $\mathfrak{P}$ die Menge der Punkte, $\mathfrak{K}$ die Menge der Kreise, $\parallel_+, \parallel_-$ die beiden Parallelitätsrelationen auf $\mathfrak{P}$ und Γ wiederum die Gruppe $PGL(2, \mathbb{R})$.

Wir stellen die drei Fragen:

1. Wie lautet die Automorphismengruppe der Struktur $(\mathfrak{P}, \mathfrak{K})$, d. h. die Gruppe der Kreisverwandtschaften?
2. Wie lautet die Automorphismengruppe der Struktur $(\mathfrak{P}, \parallel_+, \parallel_-; \mathfrak{K})$, d. h. die Gruppe derjenigen Kreisverwandtschaften, die in beiden Richtungen die Relationen $\parallel_+, \parallel_-$ je in sich überführen?
3. Wie lautet die Gruppe der winkeltreuen Kreisverwandtschaften?

Hilfssatz 6. 1. Sind A, B, C verschiedene, aber paarweise parallele Punkte, so gilt entweder

$$\text{a) } A \parallel_+ B \parallel_+ C \text{ oder b) } A \parallel_- B \parallel_- C.$$

Beweis. Angenommen, falsch. O. B. d. A. können wir dann annehmen, es gilt $A \parallel_+ B \parallel_- C$. Weiterhin haben wir noch $C \parallel_+ A$ oder $C \parallel_- A$ ¹⁹⁾. Aus $A \parallel_+ B \parallel_- C \parallel_+ A$ folgt $C \parallel_+ A \parallel_+ B$, d. h. $C \parallel_+ B$, d. h. $C = B$ zusammen mit $C \parallel_- B$. Wdspr. Aus $A \parallel_+ B \parallel_- C \parallel_- A$ folgt $A \parallel_- B$, d. h. $A = B$ mit $A \parallel_+ B$. Wdspr.

Satz 6. 1. Ist α eine Kreisverwandtschaft, so ist sie von der Gestalt

$$\varrho x'_1 = a_{11}x_1^\varrho + a_{12}x_2^\varrho$$

$$\varrho x'_2 = a_{21}x_1^\varrho + a_{22}x_2^\varrho$$

¹⁹⁾ Übrigens bedeutet $P \parallel_+ Q$ und gleichzeitig $P \parallel_- Q$ natürlich $P = Q$ wegen $I_+ \cap I_- = \{0\}$. Allerdings muß $\begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix} = 0$ im Ringfall durchdiskutiert werden. Siehe aber Satz 2. 4.

wobei a_{ij} , ϱ Elemente aus $\mathfrak{L}$ sind mit $\|a_{ij}\|$, $\varrho \in \mathfrak{K}$ und wo φ ein Automorphismus von $\mathfrak{L}$ ist, der $\mathfrak{K}$ als Ganzes festläßt. Umgekehrt stellen diese Abbildungen Kreisverwandtschaften dar.

Beweis. Wir gehen von einer beliebigen Kreisverwandtschaft α aus. Sei $\infty^\alpha = A$, $0^\alpha = B$, $1^\alpha = C$. Die Punkte $\infty, 0, 1$ sind paarweise verschieden, also gilt dies auch für A, B, C . Durch $\infty, 0, 1$ geht ein Kreis; damit geht auch, da α ein Automorphismus von $(\mathfrak{P}; \mathfrak{k})$ ist, durch A, B, C ein Kreis. Nach Satz 1. 3. (2) sind also A, B, C paarweise nicht parallel. Sei $\tau \in \mathcal{T}$ die Abbildung mit $A^\tau = \infty$, $B^\tau = 0$, $C^\tau = 1$. Damit sind $\infty, 0, 1$ Fixpunkte von $\alpha\tau$. Wir arbeiten nun im Hyperbelmodell. Die Beschränkung von $\mu = \alpha\tau$ auf die Menge $\mathfrak{S}$ der eigentlichen Punkte ist eine Bijektivität von $\mathfrak{S}$, da μ doch ∞ festläßt und μ doch parallele Punkte in parallele Punkte überführt²⁰⁾. μ ist eine Affinität der affinen Ebene über $\mathfrak{K}$: Da μ eine Kreisverwandtschaft ist, werden jedenfalls wegen $\infty^\mu = \infty$ pseudo-euklidische Geraden in pseudo-euklidische Geraden überführt; affine Geraden der Gleichung $y = \pm x + u$ gehen in ebensolche über nach Hilfssatz 6. 1. Wir interessieren uns also jetzt für diejenigen Affinitäten der affinen Ebene E über $\mathfrak{K}$, die Geraden der Steigung ± 1 in Geraden der Steigung ± 1 überführen und die die Punkte 0 und 1 festlassen. Das Koordinatensystem von E sei so gewählt, daß 0 die affinen Koordinaten (0, 0) besitzt, 1 die Koordinaten (1, 0), j die Koordinaten (0, 1). Die in Frage stehenden Affinitäten haben dann das Aussehen

$$\begin{array}{ccc} x' = x^\sigma + (a - 1)y^\sigma & & x' = x^\sigma + (a - 1)y^\sigma \\ y' = & ay^\sigma & \text{bzw.} \quad y' = -ay^\sigma \end{array}$$

mit $a \in \mathfrak{K}^\times$, σ ein Automorphismus von $\mathfrak{K}$. Es gehört also $\mu/\mathfrak{S}$ zu diesen Abbildungen. Da μ als Kreisverwandtschaft den pseudo-euklidischen Kreis $x^2 - y^2 = 1$ in einen pseudo-euklidischen Kreis überführt, so folgt $a = 1$. Also ist $\mu/\mathfrak{S}$ von der Gestalt

$$\begin{array}{ccc} (*) \quad x' = x^\sigma & \text{oder} & (**) \quad x' = x^\sigma \\ y' = y^\sigma & & y' = -y^\sigma. \end{array}$$

Da es nur jeweils eine Fortsetzung dieser Abbildungen zu einer Kreisverwandtschaft gibt, folgt, daß μ von der Gestalt (jetzt liegen homogene Koordinaten zugrunde)

$$(x_1, x_2) \rightarrow (x_1^\sigma, x_2^\sigma)$$

ist, wo φ ein Automorphismus von $\mathfrak{L}$ der Form $a + bj \rightarrow a^\sigma + b^\sigma j$ oder $a + bj \rightarrow a^\sigma - b^\sigma j$ darstellt, also ein Automorphismus von $\mathfrak{L}$, der $\mathfrak{K}$ als Ganzes festläßt.

Ist umgekehrt φ ein Automorphismus von $\mathfrak{L}$, der $\mathfrak{K}$ als Ganzes festhält, $\mathfrak{K}^\varphi = \mathfrak{K}$, so nennen wir die Restriktion $\varphi/\mathfrak{K} = \sigma$. Aus $j^2 = 1$, $j \in \mathfrak{K}$, folgt über $(j^\varphi)^2 = 1$ offenbar $j^\varphi \in \{j, -j\}$. Damit setzt sich die Abbildung

$$\begin{array}{l} \varrho x'_1 = a_{11}x_1^\sigma + a_{12}x_2^\sigma \\ \varrho x'_2 = a_{21}x_1^\sigma + a_{22}x_2^\sigma, \quad \|a_{ij}\|, \varrho \in \mathfrak{L}, \end{array}$$

zusammen aus Abbildungen der Gestalt $(*)$ oder $(**)$ und einer T-Abbildung, ist also Kreisverwandtschaft.

Korollare zu Satz 6. 1.:

²⁰⁾ $P \parallel Q$ und $P \neq Q$ ist gleichwertig damit, daß es keinen Kreis gibt, der sowohl P als auch Q enthält (Satz 1. 3, (2), (2')).

Satz 6. 2. Ist α ein Automorphismus der Struktur $(\mathfrak{P}; \mathfrak{k})$, so gilt entweder

a) α ist ein Automorphismus von $(\mathfrak{P}, ||_+, ||_-; \mathfrak{k})$ oder

b) für jedes Punktepaar P, Q mit $P ||_+ Q$ gilt $P^\alpha ||_- Q^\alpha$ und für jedes Punktepaar P, Q mit $P ||_- Q$ gilt $P^\alpha ||_+ Q^\alpha$.

Satz 6. 3. Die Automorphismen der Struktur $(\mathfrak{P}, ||_+, ||_-; \mathfrak{k})$ sind genau die Abbildungen

$$\varrho x'_1 = a_{11}x_1^\varphi + a_{12}x_2^\varphi$$

$$\varrho x'_2 = a_{21}x_1^\varphi + a_{22}x_2^\varphi, \quad a_{ij}, \varrho \in \mathfrak{L} \text{ mit } ||a_{ij}||, \varrho \in \mathfrak{R},$$

wo φ ein Automorphismus von $\mathfrak{L}$ ist, der $\mathfrak{R}$ als Ganzes festläßt, und der j festhält.

Zusammen mit früheren Überlegungen (insbesondere Satz 5. 2) folgt noch

Satz 6. 4. Die winkeltreuen Automorphismen von $(\mathfrak{P}; \mathfrak{k})$ sind genau die Elemente von $\mathcal{T}$.

§ 7. Eine axiomatische Kennzeichnung der Geometrie $A(\mathfrak{R})$

Wir gehen aus von einer abstrakten Menge $\mathfrak{P}$, deren Elemente $P, Q, \dots$ wir *pseudo-euklidische Punkte*, kurz *Punkte*, nennen. Auf $\mathfrak{P}$ seien zwei Äquivalenzrelationen $||_+, ||_-$ gegeben. Gilt für das Punktepaar P, Q eine der Relationen $P ||_+ Q, P ||_- Q$, so nennen wir das Paar P, Q *parallel*, im anderen Falle nicht parallel. Es heißen P, Q *p-parallel* (*plus-parallel*) bzw. *m-parallel* (*minus-parallel*), wenn $P ||_+ Q$ bzw. $P ||_- Q$ gilt. Weiterhin sollen Teilmengen $a, b, \dots$ von $\mathfrak{P}$ gegeben sein, die *pseudo-euklidische Kreise*, kurz *Kreise*, genannt werden. Mit $\mathfrak{k}$ werde die Menge der Kreise bezeichnet. Für $P \in a$ bzw. $P \notin a$ übernehmen wir übliche Redewendungen wie „ P liegt auf a “ bzw. „ P liegt nicht auf a “ usf. Schließlich sei noch eine Gruppe $\mathcal{T}$ von Automorphismen der Struktur $(\mathfrak{P}, ||_+, ||_-; \mathfrak{k})$ gegeben. Wir nennen $\Sigma = (\mathfrak{P}, ||_+, ||_-; \mathfrak{k}; \mathcal{T})$ eine *pseudo-euklidische Ebene*, wenn die folgenden Eigenschaften erfüllt sind:

(A 1) Durch drei paarweise nicht parallele Punkte geht genau ein Kreis.

(A 2) Ist k ein Kreis, ist P ein Punkt, so gibt es eindeutig bestimmte Punkte P', P'' auf k mit $P' ||_+ P, P'' ||_- P$, die im Falle $P \notin k$ verschieden sind.

(A 3) Es gibt vier paarweise nicht parallele Punkte, die nicht gemeinsam einem Kreis angehören.

(T 1) Es seien A_1, A_2, A_3 paarweise nicht parallele Punkte und ebenso A'_1, A'_2, A'_3 . Dann gibt es genau ein $\tau \in \mathcal{T}$ mit $A'_i = A_i^\tau, i = 1, 2, 3$. Ist darüber hinaus $A_i ||_+ A'_i$ (bzw. $A_i ||_- A'_i$) für $i = 1, 2, 3$, so gilt $P ||_+ P^\tau$ (bzw. $P ||_- P^\tau$) für alle $P \in \mathfrak{P}$.

(T 2) Sind A, B, C, D paarweise nicht parallele Punkte auf einem Kreis, so gilt für die Abbildung $\tau \in \mathcal{T}$ mit $A^\tau = B, B^\tau = A, C^\tau = D$ auch $D^\tau = C$. — Es gibt paarweise nicht parallele Punkte P, Q, R, S , die gemeinsam auf einem Kreis liegen, so, daß für die Abbildung $\tau \in \mathcal{T}$ mit $P^\tau = P, Q^\tau = Q, R^\tau = S$ auch $S^\tau = R$ gilt.

Die über dem kommutativen Körper $\mathfrak{R}$ mit $\text{Char } \mathfrak{R} \neq 2$ erklärte pseudo-euklidische Geometrie bezeichnen wir mit $A(\mathfrak{R})$, wie schon in der Einleitung fixiert.

Dann wurde schon in § 1 (bis auf eine geringfügige Ergänzung) gezeigt der

Satz 7. 1. $A(\mathfrak{K})$ ist eine pseudo-euklidische Ebene.

Der vorliegende § 7 ist dem Nachweis des folgenden Satzes gewidmet:

Satz 7. 2. Ist $\Sigma = (\mathfrak{P}, ||_+, ||_-; \mathfrak{k}; T)$ eine pseudo-euklidische Ebene, so ist sie isomorph zu einer pseudo-euklidischen Ebene $A(\mathfrak{K}) = (\mathfrak{P}', ||'_+, ||'_-; \mathfrak{k}'; T')$.

Der längere Beweis dieses kennzeichnenden Satzes der Geometrien $A(\mathfrak{K})$ erfolgt in mehreren Schritten. Allen folgenden Erörterungen liege eine feste pseudo-euklidische Ebene Σ zugrunde.

1. Je zwei verschiedene Punkte P, Q eines Kreises k sind nicht parallel.

Beweis. Sei etwa $P ||_+ Q$. Da auch $P ||_+ P$ gilt, so folgt aus (A 2) wegen $P, Q \in k$ offenbar $Q = P$.

2. Sind P, Q Punkte mit $P ||_+ Q$ und $P ||_- Q$, so gilt $P = Q$.

Beweis. Angenommen, $P \neq Q$. Nach 1 geht dann jedenfalls durch P, Q kein Kreis. Wir zeigen, daß durch Q ein Kreis geht, der also nach dem Vorhergehenden nicht den Punkt P enthält. Seien A, B, C, D vier paarweise nicht parallele Punkte; solche existieren nach (A 3). Zu wenigstens zweien dieser Punkte kann Q nicht parallel sein, da die Relationen $||_+, ||_-$ Äquivalenzrelationen sind. Damit gibt es drei paarweise nicht parallele Punkte, zu denen Q gehört. Also haben wir nach (A 1) einen Kreis k durch Q . Wegen $P \notin k$ und (A 2) gilt $P' \neq P''$, was der Annahme $P ||_+ Q, P ||_- Q, Q \in k$ widerspricht.

3. Jeder Kreis enthält wenigstens vier verschiedene Punkte.

Beweis. Nach (A 3) gibt es paarweise nicht parallele Punkte A, B, C, D . Sei k ein Kreis und seien A', B', C', D' die gemäß (A 2) bestimmten Punkte bezüglich k . Diese Punkte auf k sind verschieden, da $||_+$ eine Äquivalenzrelation ist.

Sei $\mathfrak{p}$ ein fester Kreis, auf dem Punkte P, Q, R, S der Art (T 2) liegen. $\infty, 0, 1$ seien drei verschiedene feste Punkte auf $\mathfrak{p}$. Sie sind nach 1 paarweise nicht parallel. Die Standuntergruppe

$$T_{\mathfrak{p}} = \{\tau \in T \mid \mathfrak{p}^\tau = \mathfrak{p}\}$$

von $\mathfrak{p}$ in T betrachten wir nun zunächst nur in ihrer Wirkung auf $\mathfrak{p}$. Wir beachten dabei, daß im Falle der Gleichheit der Restriktionen $\tau_1 \mid \mathfrak{p}, \tau_2 \mid \mathfrak{p}, \tau_1, \tau_2 \in T_{\mathfrak{p}}$, auch $\tau_1 = \tau_2$ ist: Dies folgt aus der einfachen Transitivität (s. (T 1)) von T auf der Menge der geordneten Tripel paarweise nicht paralleler Punkte in Verbindung mit 1 und 3.

4. $T_{\mathfrak{p}} \mid \mathfrak{p}$ ist isomorph zu einer Gruppe $PGL(2, \mathfrak{K})$ mit einem kommutativen Körper $\mathfrak{K}$, der als Menge mit $\mathfrak{p} - \{\infty\}$ identifiziert werden kann. Es ist $\text{Char } \mathfrak{K} \neq 2$.

Beweis. $\Gamma \equiv T_{\mathfrak{p}} \mid \mathfrak{p}$ ist minimal dreifach transitiv auf $\mathfrak{p}$ nach (T 1). Da noch (T 2) gilt, folgt 4 aus bekannten Resultaten²¹⁾. Ist P ein Punkt, so ordnen wir ihm das geordnete Paar (P', P'') zu, wobei P', P'' die Punkte gemäß (A 2) in bezug auf den Kreis $\mathfrak{p}$ sind.

²¹⁾ Solche Resultate legte B. Petkantschin [14] vor. Sie haben ihre Wurzel in der v. Staudt'schen Wurfberechnung.

Es gilt

5. Die Abbildung $\lambda: P \rightarrow (P', P'')$ ist eine eindeutige Abbildung von $\mathfrak{P}$ auf das Mengenprodukt $\mathfrak{p} \times \mathfrak{p}$.

Beweis. Gemäß (A 2) liegt eine eindeutige Abbildung von $\mathfrak{P}$ in $\mathfrak{p} \times \mathfrak{p}$ vor. Gegeben sei nun ein geordnetes Punktepaar $(A, B) \in \mathfrak{p} \times \mathfrak{p}$. Wir wollen zeigen, daß ein Urbild existiert und daß dieses in eindeutiger Weise bestimmt ist. Im Falle $A = B$ gibt es nur das Urbild A . Sei $A \neq B$. Für den Punkt $U \notin \mathfrak{p}$ bestimmen wir U', U'' . Seien A, B, C verschiedene Punkte auf $\mathfrak{p}$ und ebenso U', U'', V . Sei τ die nach (T 1) existierende Abbildung aus $\mathcal{T}$ mit $U'^\tau = A, U''^\tau = B, V^\tau = C$. Sei $P = U^\tau$. Dann gilt $P = U^\tau \parallel_+ U'^\tau = A$ und $P = U^\tau \parallel_- U''^\tau = B$. Also ist P ein Urbild von (A, B) . Sei Q ein weiteres Urbild. Dann gilt $Q \parallel_+ A \parallel_+ P$ und $Q \parallel_- B \parallel_- P$, was $Q = P$ bedeutet nach 2.

Mit Hilfe der Abbildung λ wollen wir jetzt die Abbildungen aus $PGL(2, \mathfrak{K})$ als Abbildungen aus $\mathcal{T}_p$ schreiben. Dazu führen wir homogene Koordinaten für die Punkte P von $\mathfrak{P}$ ein. Sei $\mathfrak{L}$ der Ring der anormal-komplexen Zahlen über $\mathfrak{K}$.

1. Fall: P ist nicht parallel zu ∞ . Dann ist also $\infty \neq P', P''$. Wir setzen

$$P \rightarrow \frac{P'' + P'}{2} + \frac{P'' - P'}{2} j.$$

Homogene Koordinaten von P seien

$$\left(\varrho \left[\frac{P'' + P'}{2} + \frac{P'' - P'}{2} j \right], \varrho \right),$$

$\varrho \in \mathfrak{K}, \mathfrak{K}$ die Einheitengruppe von $\mathfrak{L}$.

2. Fall: $P = \infty$. Homogene Koordinaten seien $(\varrho, 0), \varrho \in \mathfrak{K}$.

3. Fall: $P \parallel_+ \infty, P \neq \infty$. Homogene Koordinaten $\varrho(1 - j, 1 + j), \varrho \in \mathfrak{K}$, für $P'' = 0$ und $\varrho(2P'', 1 + j)$ für $P'' \neq 0$.

4. Fall: $P \parallel_- \infty, P \neq \infty$. Homogene Koordinaten $\varrho(1 + j, 1 - j), \varrho \in \mathfrak{K}$, für $P' = 0$ und $\varrho(+2P', 1 - j)$ für $P' \neq 0$.

Zunächst hat man

6. Es gilt $P \parallel_+ Q$ (bzw. $P \parallel_- Q$) genau dann, wenn $\begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix} \in I_+$ (bzw. $\in I_-$) ist, wobei $P = (p_1, p_2), Q = (q_1, q_2)$ sei und $I_+ = \mathfrak{K}(1 + j), I_- = \mathfrak{K}(1 - j)$.

*Beweis*²²⁾. $P \parallel_+ Q$ (bzw. $P \parallel_- Q$) ist gleichwertig mit $P' = Q'$ (bzw. $P'' = Q''$).

I. Fall: P ist nicht parallel zu ∞ .

I.1: Q ist nicht zu ∞ parallel. Mit

$$p_1 = \frac{P'' + P'}{2} + \frac{P'' - P'}{2} j, \quad p_2 = 1,$$

$$q_1 = \frac{Q'' + Q'}{2} + \frac{Q'' - Q'}{2} j, \quad q_2 = 1,$$

liegt 6 für I.1 auf der Hand.

²²⁾ Im Verlauf des Beweises kürzen wir $\begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix}$ durch D ab.

I.2: $Q = \infty$.

Hier gilt weder $D \in I_+$ (bzw. $\in I_-$), noch $P' = Q'$ (bzw. $P'' = Q''$).

I.3: $Q \neq \infty, Q \parallel_+ \infty$.

Also ist hier $Q' = \infty, Q'' \neq \infty$.

a) $Q'' = 0$. Hier ist $q_1 = 1 - j, q_2 = 1 + j$. Offenbar ist $P' \neq Q'$ und $D \notin I_+$. Weiterhin gilt offenbar $P'' = Q''$ genau dann, wenn $D \in I_-$ ist.

b) $Q'' \neq 0$. Hier ist $q_1 = 2Q'', q_2 = 1 + j$. Offenbar ist $P' \neq Q'$ und $D \notin I_+$. Weiterhin ist $P'' = Q''$ mit $D \in I_-$ gleichwertig.

I.4: $Q \neq \infty, Q \parallel_- \infty$.

Also ist hier $Q' \neq \infty, Q'' = \infty$.

a) $Q' = 0$. Es ist $q_1 = 1 + j, q_2 = 1 - j$. Es ist $P' = Q'$ mit $D \in I_+$ gleichwertig. Weiterhin gilt $P'' \neq Q''$ und $D \notin I_-$.

b) $Q' \neq 0$. Es ist $q_1 = 2Q', q_2 = 1 - j$. Es ist $P' = Q'$ mit $D \in I_+$ gleichwertig. Offenbar gilt weiterhin $P'' \neq Q''$ und $D \notin I_-$.

Da die Aussage 6 symmetrisch in P, Q ist, reduziert sich die weitere Falluntersuchung erheblich. Es verbleiben die nachstehenden weiteren Fälle:

II. $P = \infty$.

II.1. $Q = \infty$.

II.2. $Q \neq \infty, Q \parallel_+ \infty$.

II.3. $Q \neq \infty, Q \parallel_- \infty$.

III. $P \neq \infty, P \parallel_+ \infty$.

III.1. $Q \neq \infty, Q \parallel_+ \infty$.

III.2. $Q \neq \infty, Q \parallel_- \infty$.

IV. $P \neq \infty, P \parallel_- \infty$.

IV.1. $Q \neq \infty, Q \parallel_- \infty$.

Die Diskussion dieser Fälle, die immer wieder darauf hinausläuft, die Gleichwertigkeit von $P' = Q'$ (bzw. $P'' = Q''$) mit $D \in I_+$ (bzw. $\in I_-$) nachzuweisen, vereinfachen wir nochmals, indem wir die lästigen Fälle a), b) vermeiden. Man kann die folgende Zusammenfassung benutzen für die homogenen Koordinaten:

für $P \parallel_+ \infty, P \neq \infty$ die Koordinaten $((P'' + 1) + (P'' - 1)j, 1 + j)$;

für $P \parallel_- \infty, P \neq \infty$ die Koordinaten $((P' + 1) - (P' - 1)j, 1 - j)$.

Im Falle II.1 ist nichts zu zeigen.

Zu II.2.: Hier ist $Q' = \infty, Q'' \neq \infty$.

Es ist $q_1 = (Q'' + 1) + (Q'' - 1)j, q_2 = 1 + j$. Nun ist $P' = Q'$ und $D \in I_+$. Weiterhin ist $P'' \neq Q''$ und $D \notin I_-$.

Zu II.3.: Hier ist $Q' \neq \infty, Q'' = \infty$.

Es ist $q_1 = (Q' + 1) - (Q' - 1)j, q_2 = 1 - j$. Es gilt $P' \neq Q'$ und $D \notin I_+$. Weiterhin ist $P'' = Q''$ und $D \in I_-$.

Für den Fall III. haben wir $p_1 = (P'' + 1) + (P'' - 1)j, p_2 = 1 + j$.

Im Falle III.1. ist $q_1 = (Q'' + 1) + (Q'' - 1)j, q_2 = 1 + j$. Es gilt $P' = \infty = Q'$ und $D \in I_+$. Weiterhin ist $P'' = Q''$ gleichwertig mit $D \in I_-$.

Im Falle III.2. ist $q_1 = (Q' + 1) - (Q' - 1)j$, $q_2 = 1 - j$. Es gilt $P' = \infty \neq Q'$ und $D \notin I_+$. Weiterhin ist $P'' \neq \infty = Q''$ und $D \notin I_-$.

Es verbleibt der Fall IV.1.: Wegen $P \neq \infty$, $P \parallel_- \infty$ ist $P' \neq \infty$, $P'' = \infty$ und $p_1 = (P' + 1) - (P' - 1)j$, $p_2 = 1 - j$. Wegen $Q \neq \infty$, $Q \parallel_- \infty$ ist $Q' \neq \infty$, $Q'' = \infty$ und $q_1 = (Q' + 1) - (Q' - 1)j$, $q_2 = 1 - j$. Offenbar ist $P' = Q'$ gleichwertig mit $D \in I_+$. Weiterhin gilt $P'' = Q''$ und $D \in I_-$.

Damit ist 6 vollständig bewiesen.

7. Die Fortsetzung nach T_p der Abbildung

$$(*) \quad \begin{aligned} \varrho x'_1 &= a_{11}x_1 + a_{12}x_2 \\ \varrho x'_2 &= a_{21}x_1 + a_{22}x_2, \varrho \in \mathbb{R}^x, \|a_{ii}\| \in \mathbb{R}^x, \end{aligned}$$

$a_{ij} \in \mathbb{R}$, aus $PGL(2, \mathbb{R})$, ist die Abbildung gleicher Gestalt mit $\varrho \in \mathbb{R}$.

*Beweis*²³⁾. Mit Hilfe einer Restriktion $\tau|_p$ ist also jetzt auf τ zu schließen. Sei $P \in \mathfrak{P} - p$. Wir unterscheiden die Punkte P^τ und P^* , wo P^* das formal nach $(*)$ errechnete Bild von P sei. Jedenfalls ist die Abb. $(*)$ wegen Hilfssatz 1. 2 eine eindeutige Abbildung von $\mathfrak{P}$ auf sich, die wegen 6 nach Hilfssatz 1. 2 p -parallele Punkte in p -parallele Punkte überführt und m -parallele in m -parallele. Da τ auf p mit $(*)$ auf p übereinstimmt, gilt $P^\tau \parallel_+ P'^* \parallel_+ P^*$ und $P^\tau \parallel_- P'^* \parallel_- P^*$, was auf $P^\tau \parallel_+ P^*$ und $P^\tau \parallel_- P^*$ führt und wegen 2 somit $P^\tau = P^*$ ergibt.

An dieser Stelle fassen wir zusammen, wie weit wir beim Beweis von Satz 7. 2 schon gekommen sind! Mit dem in 4 erhaltenen Körper $\mathbb{R}$ und der hierüber konstruierten pseudo-euklidischen Ebene $\mathcal{A}(\mathbb{R}) = (\mathfrak{P}', \parallel_+, \parallel_-; \mathfrak{f}'; T')$ haben wir die Übereinstimmung: Eine über die Abbildung λ konstruierte bijektive Abbildung von $\mathfrak{P}$ auf $\mathfrak{P}'$, bei der nach 6 die Relationen $\parallel_+$, $\parallel'_+$ und $\parallel_-$, $\parallel'_-$ je ineinander übergehen. Weiterhin haben wir nach 7 die Übereinstimmung der Gruppen T'_p und T_p . Wir zeigen jetzt:

8. Unter Zugrundelegung der Identifikation $\mathfrak{P} \rightarrow \mathfrak{P}'$ können die Gruppen T, T' als Substitutionsgruppen auf derselben Menge $\mathfrak{P}'$ betrachtet werden. Gegenüber dieser Interpretation ist jede Substitution aus T auch eine Substitution aus T' und umgekehrt.

Vorweg zeigen wir

9. Sind P, Q Punkte, so gibt es genau einen Punkt R mit $R \parallel_+ P$ und $R \parallel_- Q$.

Den Beweis von 9 müssen wir in der axiomatischen Ebene erbringen. Wir können dabei nicht die Gruppe T' benutzen, um eine einfachere Situation herbeizuführen. Zunächst nehmen wir die Existenz eines solchen Punktes an; dabei werden wir sehen, daß R eindeutig bestimmt ist, und wir werden ein Konstruktionsverfahren erhalten, um einen gesuchten Punkt R zu finden. Für die nach (A 2) in bezug auf p bestimmten Punkte $P', P'', Q', Q'', R', R''$ gilt $R' = P', R'' = Q''$. Die Bijektivität der früher konstruierten Abbildung λ führt auf genau ein Urbild von (P', Q'') . Der gesuchte Punkt R ergibt sich als λ -Urbild von (P', Q'') .

Nun zum Beweis von 8: Sei $\tau \in T$.

Sei $\infty^\tau = A, 0^\tau = B, 1^\tau = C$. Mit den nach (A 2) in bezug auf p bestimmten Punkten $A', B', C', A'', B'', C''$ seien $\alpha, \beta \in T$ die Abbildungen mit

$$(A')^\alpha = A, (B')^\alpha = B, (C')^\alpha = C, (A'')^\beta = A, (B'')^\beta = B, (C'')^\beta = C.$$

²³⁾ Meinen ursprünglich etwas längeren Beweis von 7 ersetzte W. Leißner durch den vorliegenden.

Wir beachten dabei, daß A, B, C paarweise nicht parallel sind und daß deshalb A', B', C' und A'', B'', C'' je drei verschiedene Punkte von $\mathfrak{p}$ sind.

Offenbar gilt $\tau\alpha^{-1}, \tau\beta^{-1} \in T_{\mathfrak{p}} = T'_{\mathfrak{p}}$ mit 7. Wegen $A \parallel_+ A', B \parallel_+ B', C \parallel_+ C'$ gilt nach (T 1) $P \parallel_+ P^\alpha$ für alle $P \in \mathfrak{P}$. Genauso hat man $P \parallel_- P^\beta$ für alle $P \in \mathfrak{P}$. Damit ist $P^\tau \parallel_+ P^{\tau\alpha^{-1}}, P^\tau \parallel_- P^{\tau\beta^{-1}}$. Aufgrund von 9 ist hiermit P^τ eindeutig bestimmt. Wir können also P^τ konstruieren über die bekannten Abbildungen $\tau\alpha^{-1}, \tau\beta^{-1} \in T'_{\mathfrak{p}}$ und über die Relationen $\parallel_+, \parallel_-$. Sei $\varphi = \tau\alpha^{-1}, \psi = \tau\beta^{-1}$. Wegen $\varphi, \psi \in T'_{\mathfrak{p}}$ sind φ, ψ auch die Abbildungen aus T' mit $\infty^\varphi = A', 0^\varphi = B', 1^\varphi = C', \infty^\psi = A'', 0^\psi = B'', 1^\psi = C''$. Sind $\hat{\tau}, \hat{\alpha}, \hat{\beta}$ die Abbildungen aus T' mit $\infty^{\hat{\tau}} = A, 0^{\hat{\tau}} = B, 1^{\hat{\tau}} = C; (A')^{\hat{\alpha}} = A, (B')^{\hat{\alpha}} = B, (C')^{\hat{\alpha}} = C; (A'')^{\hat{\beta}} = A, (B'')^{\hat{\beta}} = B, (C'')^{\hat{\beta}} = C$, so gilt also $\varphi = \hat{\tau}\hat{\alpha}^{-1}, \psi = \hat{\tau}\hat{\beta}^{-1}$. Da nun auch $P^{\hat{\tau}}$ eindeutig bestimmt ist als der Punkt X mit $X \parallel_+ P^{\hat{\tau}\hat{\alpha}^{-1}}, X \parallel_- P^{\hat{\tau}\hat{\beta}^{-1}}$, so gilt $P^\tau = P^{\hat{\tau}}$, d. h. $\tau = \hat{\tau}$.

Ist nun umgekehrt $\gamma \in T'$, so sei $\infty^\gamma = A, 0^\gamma = B, 1^\gamma = C$. Sei $\tau \in T$ die Abbildung mit $\infty^\tau = A, 0^\tau = B, 1^\tau = C$. Also ist nach dem Vorhergehenden $\tau = \hat{\tau} \equiv \gamma$.

10. Bei der Abbildung $\mathfrak{P} \leftrightarrow \mathfrak{P}'$ entsprechen die axiomatischen Kreise den $A(\mathfrak{K})$ -Kreisen.

Beweis. Folgt aus $T = T'$ (vgl. 8) in Verbindung mit der Tatsache, daß T transitiv auf $\mathfrak{k}$ ist, also jeder Kreis als Bild von $\mathfrak{p}$ gewonnen werden kann, und das Gleiche für $T', \mathfrak{P}'$ gilt.

Damit ist Satz 7. 2 vollends bewiesen.

Wir bemerken, daß bei einer pseudo-euklidischen Ebene die Gültigkeit des Berührsatzes (Eigenschaft (4) in Satz 1. 3) nicht gefordert wurde, daß sie aber erschlossen werden kann, da ja nach Satz 7. 2 jede pseudo-euklidische Ebene eine Ebene $A(\mathfrak{K})$ ist.

Zusatz bei der Korrektur

In der Einleitung wiesen wir auf die Arbeit von U. Graf [10] hin, in der eine inhomogene Beschreibung der klassischen pseudo-euklidischen Geometrie gegeben wird. In dem mir jetzt zugänglich gewordenen Buch I. M. Yaglom, Complex Numbers in Geometry, New York-London 1968 (englische Übersetzung der 1963 in Moskau erschienenen Originalausgabe) wird ein Kalkül unendlich großer Zahlen für die pseudo-euklidische Geometrie (und übrigens auch für die Laguerre-Geometrie) entwickelt, mit dessen Hilfe die bei der inhomogenen Beschreibung auftretenden Ausnahmepunkte beseitigt werden können. Einen Kalkül unendlich großer Zahlen für den Fall der Laguerre-Geometrie hat schon Grünwald angegeben. Die Arbeiten von Grünwald und Graf und anderen werden allerdings bei I. M. Yaglom loc. cit. nicht zitiert — möglicherweise, weil es sich um ein einführendes Werk handelt.

Literaturverzeichnis

- [1] D. Barbilian, Zur Axiomatik der projektiven ebenen Ringgeometrien. I, J. ber. Deutsch. Math. Ver. **50** (1940), 179—229.
- [2] D. Barbilian, II, J. ber. Deutsch. Math. Ver. **51** (1941), 34—76.
- [3] H. Beck, Ein Seitenstück zur Möbius'schen Geometrie der Kreisverwandtschaften, Trans. Americ. Math. Soc. **11** (1910), 414—448.
- [4] W. Benz, Über Möbiusebenen, Ein Bericht, J. ber. Deutsch. Math. Ver. **63** (1960), 1—27.

- [5] W. Benz, Süß'sche Gruppen in affinen Ebenen mit Nachbarelementen und allgemeineren Strukturen, Abh. math. Sem. Hamburg **26** (1963), 83—101.
- [6] W. Benz, Laguerregeometrie über einem lokalen Ring, Math. Z. **87** (1965), 137—145.
- [7] W. Benz, Über eine Verallgemeinerung des Satzes von Miquel, Publ. Math. Debrecen **9** (1962), 227—230.
- [8] W. Benz und H. Maurer, Über die Grundlagen der Laguerre-Geometrie, Ein Bericht, J. ber. Deutsch. Math. Ver. **67** (1964), 14—42.
- [9] G. Ewald, Begründung der Geometrie der ebenen Schnitte einer Semiquadrik, Arch. Math. **8** (1957), 203—208.
- [10] U. Graf, Über komplexe Zahlssysteme und ihren Zusammenhang mit den äquidistanten Transformationen in Ebenen mit nichteuklidischer Maßbestimmung, Sitz.ber. Berliner math. Ges., 32. Jhrg. (1933).
- [11] J. Hjelmslev, Einleitung in die allgemeine Kongruenzlehre, Danske Vid. Selsk. math.-fys. Medd. a) 1. Mitt. **8** (1929); b) 2. Mitt. **10** (1929); c) 3. Mitt. **19** (1942); d) 4. Mitt. **22** (1945); e) 5. Mitt. **22** (1945); f) 6. Mitt. **25** (1949).
- [12] F. Klein, Vorlesungen über höhere Geometrie. Berlin 1926.
- [13] W. Klingenberg, Projektive Geometrien mit Homomorphismus. Math. Ann. **132** (1956), 180—200.
- [14] B. Petkantschin, Axiomatischer Aufbau der zweidimensionalen Möbiusschen Geometrie. Annuaire Univ. Sofia, Fac. Phys.-Math. **36** (1940), 219—325.
- [15] K. Strubecker, Geometrie in einer isotropen Ebene. I, MNU **15** (1962/63), 297—306.
- [16] K. Strubecker, Geometrie in einer isotropen Ebene. II, MNU **15** (1962/63), 343—351.
- [17] K. Strubecker, Geometrie in einer isotropen Ebene. III, MNU **15** (1962/63), 385—394.
- [18] W. Süß, Beiträge zur gruppentheoretischen Begründung der Geometrie. II, Tôhoku math. J. **27** (1926), 213—242.
- [19] W. Süß, Beiträge zur gruppentheoretischen Begründung der Geometrie. III, Tôhoku math. J. **28** (1927), 228—241.
- [20] L. Tamassy, A Möbius-Féle Körgeometria Egy Általánosításáról. Acta Univ. Debrecen. **2** (1955), 137—143 (ungarisch mit deutscher Zusammenfassung).

Mathematisches Institut der Universität, 463 Bochum

Eingegangen 24. Juli 1967

Asymptotische Abschätzungen für dreielementige Extremalbasen in natürlichen Zahlen

Von *Gerd Hofmeister* z. Z. in Princeton (N. J.)

Herrn Professor Dr. Hans Rohrbach zum 65. Geburtstag

Inhalt

- § 1. Bezeichnungen und Vorbemerkungen
- § 2. Abschätzungen von $n_h(3)$ nach unten
- § 3. Zwischenbemerkungen
- § 4. Abschätzungen von $n_h(3)$ nach oben
- § 5. Die Extremalbasen zu $n_h(3)$ für $h > h_0$
- § 6. Numerische Resultate für $h \leq 34$
- § 7. Vorbemerkungen zu h -Mengen
- § 8. Abschätzungen von $l_h(3)$ für gerade h
- § 9. Abschätzungen von $l_h(3)$ für ungerade h
- § 10. Die Extremalbasen zu $g_h(3)$ für $h > h_0$
- Literatur

§ 1. Bezeichnungen und Vorbemerkungen

Kleine lateinische und griechische Buchstaben bezeichnen — wenn nicht ausdrücklich anders gesagt — nichtnegative ganze Zahlen.

Ist x reell, so bezeichnet $[x]$ die größte ganze Zahl $\leq x$, $\langle x \rangle$ die kleinste ganze Zahl $\geq x$; ferner setzen wir $\{x\} = [x]$ für $x - [x] < \frac{1}{2}$ und $\{x\} = \langle x \rangle$ für $x - [x] \geq \frac{1}{2}$.

$[a, b]$ ist die Menge aller reellen x mit $a \leq x \leq b$.

o. B. d. A. = ohne Beschränkung der Allgemeinheit.

Ist $\mathfrak{A} = \{a_1, a_2, a_3\}$ eine gegebene Zahlenmenge mit $a_1 = 1 < a_2 < a_3$, dann heißt der Ausdruck

$$(1) \quad \lambda_3 a_3 + \lambda_2 a_2 + \lambda_1 a_1$$

eine Darstellung von m bezüglich $\mathfrak{A}$, wenn

$$(2) \quad m = \lambda_3 a_3 + \lambda_2 a_2 + \lambda_1 a_1$$

gilt. Wo keine Mißverständnisse zu befürchten sind, wird auch (2) als Darstellung von m bezüglich $\mathfrak{A}$ bezeichnet. Ist $h \geq 1$ gegeben und gilt

$$(3) \quad \lambda_3 + \lambda_2 + \lambda_1 \leq h,$$

so heißt (1) bzw. (2) eine zulässige Darstellung von m bezüglich $\mathfrak{A}$ und h .

Gelten in (1) die Ungleichungen

$$(4) \quad \lambda_1 a_1 \leq a_2 - 1 \text{ und } \lambda_2 a_2 + \lambda_1 a_1 \leq a_3 - 1,$$

so heißt (1) bzw. (2) die Euklidische Darstellung von m bezüglich $\mathfrak{A}$ ([2]). Setzt man für jedes s

$$(5) \quad s a_3 = \gamma_s a_2 - \beta_s \text{ mit } 0 \leq \beta_s \leq a_2 - 1,$$

so sagen wir, m hat eine s -Stelle, falls (4) gilt und zu gegebenem $h \geq 1$

$$(6) \quad m = (\lambda_3 - s) a_3 + (\lambda_2 + \gamma_s) a_2 + (\lambda_1 - \beta_s) a_1$$

eine zulässige Darstellung von m bezüglich $\mathfrak{A}$ und h ist.

Ist $\mathfrak{A} = \{a_1, a_2, a_3\}$ eine Zahlenmenge mit $a_1 = 1 < a_2 < a_3$ und ist $h \geq 1$, so werden drei Reichweiten betrachtet.

1. Bezeichnet $n_h(\mathfrak{A})$ die größte Zahl n , so daß jedes m in $[1, n]$ bezüglich $\mathfrak{A}$ und h eine zulässige Darstellung besitzt, so bezeichnet $n_h(3)$ die größte Zahl $n_h(\mathfrak{A})$, zu der ein $\mathfrak{A}$ existiert; ein solches $\mathfrak{A}$ heißt auch Extremalbasis zu $n_h(3)$ (Rohrbach [3]).

2. Bezeichnet man genau die Mengen $\mathfrak{A}$ als h -Mengen, für die h die kleinste Zahl mit $n_h(\mathfrak{A}) \geq a_3$ ist, so bezeichnet $l_h(3)$ die größte Zahl $n_h(\mathfrak{A})$, zu der eine h -Menge $\mathfrak{A}$ existiert; ein solches $\mathfrak{A}$ heißt auch Extremalbasis zu $l_h(3)$ (Salié [4]).

3. Bezeichnet man genau die Mengen $\mathfrak{A}$ als reguläre Mengen, für die jede Zahl m in $[1, n_h(\mathfrak{A})]$ eine 0-Stelle hat, so bezeichnet $g_h(3)$ die größte Zahl $n_h(\mathfrak{A})$, zu der eine reguläre Menge $\mathfrak{A}$ existiert; ein solches $\mathfrak{A}$ heißt auch Extremalbasis zu $g_h(3)$ ([4]).

Allgemeiner lassen sich diese Reichweiten für k -elementige Zahlenmengen definieren.

Nach Rohrbachs Arbeit [3] ist insbesondere $n_h(k)$ zum Gegenstand zahlreicher Betrachtungen geworden; $l_h(k)$ und $g_h(k)$ dienen dabei als Hilfsfunktionen zur Abschätzung von $n_h(k)$ nach unten oder überhaupt als etwas durchsichtigere, wohl auch in sich interessante Ausgangsbetrachtungen, von denen man in irgendeiner Hinsicht wenigstens eine gewisse Ähnlichkeit bezüglich $n_h(k)$ erhofft.

Abgesehen von einigen tabulierten Werten (vgl. Stöhr [5]) und abgesehen von den Fällen $n_1(k) = k$, $n_h(1) = h$ sowie

$$(7) \quad n_h(2) = \left\lfloor \frac{h^2 + 6h + 1}{4} \right\rfloor \quad (\text{Stöhr [5]})$$

klaffen zwischen den unteren und oberen Abschätzungen für $n_h(k)$ beträchtliche Lücken. In dem uns beschäftigenden Fall $k = 3$ ist in Salié [4]

$$(8) \quad l_h(3) > \frac{h^3}{27} + \frac{31}{36} h^2 \quad \text{für } h \geq 13$$

gezeigt, und dies war bislang für diese h auch zugleich die beste untere Abschätzung von $n_h(3)$.

Nach oben ist anscheinend nicht mehr als

$$(9) \quad n_h(3) < \binom{h+3}{3} = \frac{h^3}{6} + h^2 + O(h)$$

gezeigt (Rohrbach [3]). Bezüglich $g_h(3)$ ist in [1]

$$(10) \quad g_h(3) = \frac{h^3}{27} + \frac{5}{9}h^2 + O(h)$$

bewiesen.

Im folgenden wird gezeigt:

$$(11) \quad n_h(3) = \frac{4}{81}h^3 + \frac{2}{3}h^2 + O(h),$$

$$(12) \quad l_h(3) = \frac{\sqrt{3}}{36}h^3 + \frac{5 + 2\sqrt{3}}{12}h^2 + O(h) \quad \text{für gerade } h,$$

$$(13) \quad l_h(3) = \frac{\sqrt{3}}{36}h^3 + \frac{2 + \sqrt{3}}{6}h^2 + O(h) \quad \text{für ungerade } h.$$

Darüber hinaus werden die Extremalbasen zu allen drei Reichweiten im Fall $h > h_0$ angegeben. Bei den Abschätzungen nach oben werden jedoch keine Aussagen über h_0 gemacht. Der Vergleich von (11), (12) und (13) mit (10) zeigt übrigens, daß $n_h(3)$ und $l_h(3)$ beträchtlich größer als $g_h(3)$ werden. Andererseits erweist sich, daß die Extremalbasen zu $n_h(3)$ und $l_h(3)$ für $h > h_0$ in $[1, n_h(\mathfrak{A})]$ stets 0- oder 1-Stellen haben.

In den folgenden Betrachtungen spielt der Begriff der Euklidischen Darstellung von m bezüglich $\mathfrak{A}$ eine erhebliche Rolle. Offenbar besitzt jede Zahl $m \geq 1$ genau eine Euklidische Darstellung bezüglich $\mathfrak{A}$ (abgesehen von der Reihenfolge der Summanden). Um sie zu erhalten, hat man das größte Element a_3 so oft als möglich zu verwenden, dann mit a_2 bezüglich des verbleibenden Restes ebenso zu verfahren, entsprechend mit $a_1 = 1$. Klar ist noch die folgende Eigenschaft Euklidischer Darstellungen: Sind

$$y = y_3a_3 + y_2a_2 + y_1a_1$$

und

$$z = z_3a_3 + z_2a_2 + z_1a_1$$

die Euklidischen Darstellungen von y und z bezüglich $\mathfrak{A}$, so gilt $y < z$ dann und nur dann, wenn für den größten Index j mit $y_j \neq z_j$ auch $y_j < z_j$ gilt. Ferner ist zu beachten, daß die Euklidische Darstellung einer Zahl m bezüglich $\mathfrak{A}$ nicht notwendig minimale Summandenzahl aufweist, wie das einfachste Gegenbeispiel $\mathfrak{A} = \{1, 3, 4\}$ und $m = 6$ zeigt: die Euklidische Darstellung benötigt drei Summanden, $6 = 4 + 1 + 1$, obwohl sich 6 bereits mit zwei Summanden aus $\mathfrak{A}$ darstellen läßt. (Vergleiche auch [2].)

Wo keine Mißverständnisse zu befürchten sind, werden im folgenden bei Darstellungen die Zusätze „bezüglich $\mathfrak{A}$ “ oder „bezüglich $\mathfrak{A}$ und h “ weggelassen.

§ 2. Abschätzung von $n_h(3)$ nach unten

Wie üblich geschieht die Abschätzung von $n_h(3)$ nach unten durch Angabe einer geeigneten Zahlenmenge. Eine Motivation der Voraussetzungen des folgenden Hilfssatzes 1 kann erst nach den Abschätzungen von $n_h(3)$ nach oben gegeben werden, für die ihrerseits einigermaßen gute Abschätzungen nach unten benötigt werden.

Hilfssatz 1. *Es sei $h \geq 2$. Ferner seien β und γ beliebig mit*

$$(1) \quad 2 \leq \gamma \leq \beta \text{ und } 2\beta \leq h + 2$$

gegeben. Setzt man dann

$$(2) \quad a_1 = 1, \quad a_2 = 2\beta - \gamma + 1 \quad \text{und} \quad a_3 = \gamma a_2 - \beta,$$

so folgt

$$(3) \quad n_h(\mathfrak{A}) \geq (h + 4 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (\beta - 2)a_1.$$

Beweis. Wir bezeichnen die rechte Seite von (3) mit n und bemerken, daß diese Darstellung die Euklidische von n ist; denn die Koeffizienten der a_i sind nichtnegativ, weiter ist

$$(\beta - 2)a_1 < a_2 - 1$$

und

$$(\gamma - 2)a_2 + (\beta - 2)a_1 = \gamma a_2 - \beta - 2 - 2(a_2 - \beta) < \gamma a_2 - \beta - 2 = a_3 - 2.$$

Es sei nun m beliebig aus $[1, n]$ mit der Euklidischen Darstellung

$$m = e_3 a_3 + e_2 a_2 + e_1 a_1$$

gegeben.

Dann folgt unmittelbar

$$(4) \quad e_3 \leq h + 4 - \beta - \gamma,$$

$$(5) \quad e_2 \leq \left\lfloor \frac{a_3 - 1}{a_2} \right\rfloor = \gamma - 1,$$

$$e_1 \leq a_2 - 1$$

sowie

$$(6) \quad e_2 a_2 + e_1 a_1 \leq a_3 - 1.$$

a) Für $1 \leq m < a_2$ ist die Euklidische Darstellung $m = e_1 a_1$ bereits eine zulässige, denn es gilt $e_1 \leq a_2 - 1 = 2\beta - \gamma \leq 2\beta - 2 \leq h$.

b) Für $a_2 \leq m < a_3$ ist die Euklidische Darstellung $m = e_2 a_2 + e_1 a_1$ bereits eine zulässige, denn der Fall $e_2 = \gamma - 1$ und $e_1 = a_2 - 1$ führt wegen

$$e_2 a_2 + e_1 a_1 = (\gamma - 1)a_2 + (a_2 - 1)a_1 = \gamma a_2 - 1 > a_3$$

auf einen Widerspruch zu (6). Es ist also

$$e_2 + e_1 \leq \gamma + a_2 - 3 = 2\beta - 2 \leq h.$$

c) Sei daher $a_3 \leq m \leq n$, also $e_3 \geq 1$. Ist $e_3 + e_2 + e_1 \leq h$, so ist die Euklidische Darstellung bereits eine zulässige. Daher setzen wir im folgenden

$$(7) \quad e_3 + e_2 + e_1 \geq h + 1$$

voraus. Nach (4) und (5) folgt $e_3 + e_2 \leq h + 3 - \beta$. Der Fall $e_3 + e_2 = h + 3 - \beta$ führt auf $e_3 = h + 4 - \beta - \gamma$ und $e_2 = \gamma - 1$, also auf $m > n$, im Widerspruch zur Wahl von m . Desgleichen führt der Fall $e_3 = h + 4 - \beta - \gamma$ und $e_2 = \gamma - 2$ wegen (7) auf $e_1 \geq \beta - 1$, also auf $m > n$. Da der Fall $e_3 = h + 3 - \beta - \gamma$ und $e_2 = \gamma - 1$ wegen (7) auf $e_1 \geq \beta - 1$ führt, also wegen

$$\begin{aligned} e_2 a_2 + e_1 a_1 &\geq (\gamma - 1)a_2 + (\beta - 1)a_1 = \gamma a_2 - \beta + 2\beta - a_2 - 1 \\ &= \gamma a_2 - \beta + \gamma - 2 \geq \gamma a_2 - \beta = a_3 \end{aligned}$$

auf einen Widerspruch zu (6), können wir im folgenden

$$e_3 + e_2 \leq h + 1 - \beta$$

voraussetzen und daher wegen (7) noch

$$e_1 \geq \beta.$$

Dann sind aber in der folgenden neuen Darstellung von m ,

$$m = (e_3 - 1)a_3 + (e_2 + \gamma)a_2 + (e_1 - \beta)a_1,$$

die Koeffizienten der a_i nichtnegativ, und für die Koeffizientensumme folgt

$$\begin{aligned} e_3 - 1 + e_2 + \gamma + e_1 - \beta &\leq h + 1 - \beta - 1 + \gamma + a_2 - 1 - \beta \\ &= h + a_2 - 2\beta + \gamma - 1 = h. \end{aligned}$$

Also besitzt jedes m in $[1, n]$ mindestens eine zulässige Darstellung, überdies sogar eine 0- oder 1-Stelle.

Um eine untere Schranke von $n_h(3)$ für alle $h \geq 2$ zu erhalten, betrachten wir *Wahl 1*.

$$\beta = \left\lfloor \frac{5}{9}h \right\rfloor - \left\lfloor \frac{h}{9} \right\rfloor + 1$$

und

$$\gamma = \left\lfloor \frac{2}{9}h \right\rfloor + 2.$$

Dann gilt (1) für alle $h \geq 2$. Zieht man sodann bezüglich h eine Aufspaltung nach Restklassen mod 9 heran, so folgt mit (2) nach (3)

$$\textbf{Folgerung 1. } n_h(3) > \frac{4}{81}h^3 + \frac{2}{3}h^2 + \frac{47}{27}h - 2 \text{ für } h \geq 2.$$

Für $h \geq 18$ benutzt man vorteilhafter

Wahl 2. $\beta = \left\lfloor \frac{5}{9}h \right\rfloor - \left\lfloor \frac{h}{9} \right\rfloor + 2$ und $\gamma = \left\lfloor \frac{2}{9}h \right\rfloor + 2$. Dann ist (1) jedenfalls für alle $h \geq 18$ erfüllt. Eine Aufspaltung nach Restklassen mod 9 bezüglich h zeigt, daß für $h \geq 27$ die rechte Seite von (3) für $h \equiv 0 \pmod{9}$ minimal ist. Wir erhalten daher

$$\textbf{Folgerung 2. } n_h(3) \geq \frac{4}{81}h^3 + \frac{2}{3}h^2 + \frac{22}{9}h \text{ für } h \geq 27.$$

Es sei hier vorweggenommen, daß in Folgerung 2 für alle $h > h_0$ mit $h \equiv 0 \pmod{9}$ das Gleichheitszeichen gilt (siehe weiter unten).

Zur Abschätzung von $n_h(3)$ nach oben benötigen wir einige Hilfsbetrachtungen.

§ 3. Zwischenbemerkungen

Es sei $h \geq 2$ und $\mathfrak{A} = \{a_1, a_2, a_3\}$ mit $a_1 = 1 < a_2 < a_3$ gegeben. Weiter sei n eine Zahl aus $[1, n_h(\mathfrak{A})]$. Dann besitzt n eine zulässige Darstellung, etwa

$$n = x_3a_3 + x_2a_2 + x_1a_1,$$

wo $x_i \geq 0$, $x_3 + x_2 + x_1 \leq h$ und $x_1 \leq a_2 - 1$ vorausgesetzt werden kann. (Im Fall $x_1 \geq a_2$ liefert

$$n = x_3a_3 + \left(x_2 + \left\lfloor \frac{x_1}{a_2} \right\rfloor\right)a_2 + \left(x_1 - \left\lfloor \frac{x_1}{a_2} \right\rfloor a_2\right)a_1$$

eine Darstellung vom verlangten Typ.)

Bezeichne ferner

$$n = e_3 a_3 + e_2 a_2 + e_1 a_1$$

die Euklidische Darstellung von n . Ist nun $e_3 + e_2 + e_1 \geq h + 1$, so folgt unmittelbar $e_3 - x_3 \geq 0$. Weiter ist $e_2 - x_2 \leq 0$; denn die Annahme $e_2 - x_2 \geq 1$ führt auf den Widerspruch

$$(e_3 - x_3) a_3 + (e_2 - x_2) a_2 + (e_1 - x_1) a_1 \geq a_2 - (a_2 - 1) = 1.$$

Die Annahme $e_1 - x_1 \leq 0$ führt auf

$$(e_3 - x_3) a_3 = (x_2 - e_2) a_2 + (x_1 - e_1) a_1,$$

wo die Koeffizienten der a_i nichtnegativ sind, also auf

$$e_3 - x_3 \leq x_2 - e_2 + x_1 - e_1$$

oder

$$e_3 + e_2 + e_1 \leq x_3 + x_2 + x_1 \leq h,$$

im Widerspruch zur Voraussetzung.

Insgesamt folgt: Ist $n = e_3 a_3 + e_2 a_2 + e_1 a_1$ die Euklidische Darstellung von n , so besitzt n dann und nur dann zulässige Darstellungen, wenn es zulässige Darstellungen der Form

$$n = (e_3 - s) a_3 + (e_2 + \gamma_s) a_2 + (e_1 - \beta_s) a_1$$

gibt, wo s , γ_s und β_s alle nichtnegativ sind und überdies $\beta_s \leq a_2 - 1$ ist. Es gilt also

Folgerung 3. Eine Zahl hat genau dann mindestens eine zulässige Darstellung, wenn sie eine s -Stelle mit einem $s \geq 0$ hat.

Aufgrund dieser Folgerung genügt es, in $[1, n_h(\mathfrak{A})]$ statt zulässiger Darstellungen schlechthin die spezielleren s -Stellen zu betrachten.

Folgerung 4. Es sei $h \geq 2$ und $\mathfrak{A} = \{a_1, a_2, a_3\}$ eine Zahlenmenge mit $a_1 = 1 < a_2 < a_3$ und $n_h(\mathfrak{A}) \geq \frac{h^3}{21}$. Dann besitzt jede Zahl n aus $[1, n_h(\mathfrak{A})]$ mindestens eine s -Stelle mit einem $s \leq 21$.

Beweis. Annahme: Es gibt ein m in $[1, n_h(\mathfrak{A})]$, so daß m nur s -Stellen mit $s \geq 22$ besitzt. Dann sei r die kleinste dieser s -Stellen.

Es folgt $ra_3 = \gamma_r a_2 - \beta_r a_1$ mit $22 \leq r \leq \gamma_r - 1 \leq h - 1$ und $0 \leq \beta_r \leq a_2 - 1$ sowie $r + \beta_r \geq \gamma_r + 1$ (andernfalls hätte m bereits eine 0-Stelle).

Es folgt $r + \gamma_r a_2 - ra_3 \geq \gamma_r + 1$, also

$$r(a_3 - 1) \leq \gamma_r(a_2 - 1) - 1 < h(a_2 - 1)$$

oder

$$r < h \frac{a_2 - 1}{a_3 - 1}.$$

Nun ist offenbar $a_2 \leq h + 1$ und $a_3 \geq \frac{h^2}{21}$. Wegen $h \geq 23$ folgt daher

$$r < \frac{21h^2}{h^2 - 21} < 22,$$

im Widerspruch zur Annahme.

Zur Verschärfung von Folgerung 4 für $h > h_0$ benötigen wir einige Hilfssätze.

Hilfssatz 2 (Stöhr [5]). *Es sei $h \geq 1$ und $\mathfrak{B} = \{a_1, a_2\}$ eine Zahlenmenge mit $a_1 = 1 < a_2 \leq h + 1$. Dann gilt*

$$n_h(\mathfrak{B}) = (h + 3)a_2 - a_2^2 - 2.$$

Beweis. Jede Zahl z mit $1 \leq z \leq (h - a_2 + 3)a_2 - 2$ ist als Summe von höchstens h Summanden aus $\mathfrak{B}$ darstellbar (man wähle $\left[\frac{z}{a_2}\right]$ Summanden gleich a_2 , die übrigen gleich a_1), die Zahl $(h - a_2 + 3)a_2 - 1$ dagegen nicht; denn da sie $\equiv -1 \pmod{a_2}$ ist, müßten mindestens $a_2 - 1$ Summanden gleich a_1 sein, also höchstens $h - a_2 + 1$ der Summanden gleich a_2 ; es ist aber

$$(h - a_2 + 1)a_2 + (a_2 - 1)a_1 < (h - a_2 + 3)a_2 - 1.$$

Hilfssatz 3. *Es sei $h \geq 2$ und $\mathfrak{A} = \{a_1, a_2, a_3\}$ eine Zahlenmenge mit $a_1 = 1 < a_2 < a_3$. Weiter seien n und n' Zahlen aus $[1, n_h(\mathfrak{A})]$ mit den Euklidischen Darstellungen*

$$\begin{aligned} n &= e_3 a_3 + e_2 a_2 + e_1 a_1, \\ n' &= e'_3 a_3 + e'_2 a_2 + e'_1 a_1. \end{aligned}$$

Ist ferner s die kleinste Zahl, so daß n eine s -Stelle hat und gelten die Aussagen $e_3 + e_2 \leq e'_3 + e'_2$, $e_1 = e'_1$ und $e'_3 \geq s$, so hat auch n' keine s' -Stellen mit $s' < s$.

Beweis. Annahme: n' hat eine s' -Stelle mit $s' < s$. Ist $e_3 \leq s' - 1$, so ist $e_3 - s < 0$, und n hat keine s -Stelle. Sei daher $e_3 \geq s'$. Dann sind in

$$n = (e_3 - s')a_3 + (e_2 + \gamma_{s'})a_2 + (e_1 - \beta_{s'})a_1$$

die Koeffizienten der a_i nichtnegativ (speziell ist $e_1 - \beta_{s'} = e'_1 - \beta_{s'} \geq 0$, da n' eine s' -Stelle hat). Weiter gilt nach Voraussetzung

$$e_3 - s' + e_2 + \gamma_{s'} + e_1 - \beta_{s'} \leq e'_3 - s' + e'_2 + \gamma_{s'} + e'_1 - \beta_{s'} \leq h$$

(letzteres da n' eine s' -Stelle hat). Also hätte n eine s' -Stelle, wo $s' < s$ ist, im Widerspruch zur Wahl von s .

Hilfssatz 4. *Es sei*

$$f = f(x_1, \dots, x_n) = \left(\alpha - \sum_{i=1}^n \alpha_i x_i \right) \prod_{i=1}^n x_i$$

eine Funktion reeller Variabler $x_1, \dots, x_n$, wo α und $\alpha_1, \dots, \alpha_n$ positive reelle Zahlen sind. Dann hat f in dem durch $x_i > 0$ ($i = 1, \dots, n$) definierten Gebiet G ihr Maximum bei

$$x_i = \frac{\alpha}{(n+1)\alpha_i} \quad (i = 1, \dots, n)$$

und hat dort den Wert

$$f\left(\frac{\alpha}{(n+1)\alpha_1}, \dots, \frac{\alpha}{(n+1)\alpha_n}\right) = \left(\frac{\alpha}{n+1}\right)^{n+1} \prod_{i=1}^n \frac{1}{\alpha_i}.$$

Beweis. Aufgrund der Substitution

$$x_i = \frac{\alpha}{(n+1)\alpha_i} y_i \quad (i = 1, \dots, n)$$

genügt es, die Funktion

$$g = g(y_1, \dots, y_n) = \left(n+1 - \sum_{i=1}^n y_i\right) \prod_{i=1}^n y_i$$

zu betrachten, wobei wir uns auf das Innere des durch

$$y_i = 0 \quad (i = 1, \dots, n) \quad \text{und} \quad \sum_{i=1}^n y_i = n + 1$$

eingeschlossenen Gebietes beschränken können. Es folgt mit

$$\frac{\partial g}{\partial y_j} = - \prod_{i=1}^n y_i + \left(n + 1 - \sum_{i=1}^n y_i \right) \prod_{i=1, i \neq j}^n y_i = 0 \quad \text{für } j = 1, \dots, n$$

$$\text{wegen } y_i \neq 0 \text{ und } \sum_{i=1}^n y_i \neq n + 1, \text{ daß}$$

$$y_j = y_k \quad \text{für alle } j \text{ und } k \text{ gilt,}$$

also $y_j = 1$ für alle j .

Die Maximaleigenschaft an dieser Stelle folgt formal leicht wegen

$$\frac{\partial^2 g}{\partial y_j^2} = -2 \quad \text{und} \quad \frac{\partial^2 g}{\partial y_j \partial y_k} = -1 \quad \text{für } j \neq k;$$

es gilt also für die Funktionaldeterminanten an dieser Stelle

$$D_\nu = \frac{\partial(g_{y_1}, \dots, g_{y_\nu})}{\partial(y_1, \dots, y_\nu)} = (-1)^\nu \begin{vmatrix} 2 & 1 & \dots & 1 \\ \vdots & & & \\ 1 & 1 & \dots & 2 \end{vmatrix}_{(\nu, \nu)} = (-1)^\nu (\nu + 1)$$

für $\nu = 1, \dots, n$. Es folgt die Behauptung.

Bemerkung. Für $n = 2$ diene dieser Hilfssatz implizit zur Auffindung geeigneter β und γ in § 2; in diesem § 3 benötigen wir ihn lediglich für $n = 1$; in § 4 wird er jedoch explizit für $n = 2$ gebraucht.

Satz 1¹⁾. *Es gibt ein h_0 so, daß für $h > h_0$ und jedem $\mathfrak{A} = \{a_1, a_2, a_3\}$ mit $a_1 = 1 < a_2 < a_3$ und $n_h(\mathfrak{A}) \geq \frac{h^3}{24}$ jede Zahl n aus $[1, n_h(\mathfrak{A})]$ eine s -Stelle mit einem $s \leq 3$ hat.*

Beweis. Nach Folgerung 4 existiert eine kleinste Zahl $s (\leq 24)$, so daß jede Zahl n aus $[1, n_h(\mathfrak{A})]$ eine höchstens s -Stelle hat. Annahme: $s \geq 4$.

Es folgt $sa_3 = \gamma_s a_2 - \beta_s$ mit $5 \leq \gamma_s \leq h$ und $0 \leq \beta_s \leq a_2 - 1$. Außerdem kann noch $s + \beta_s \geq \gamma_s + 1$ vorausgesetzt werden (andernfalls hätte kein n eine s -Stelle). Weiter gibt es nach Konstruktion von β_s und γ_s eine Zahl l mit $0 \leq l \leq s - 1$, so daß $\beta_s = s\beta_1 - la_2$ und $\gamma_s = s\gamma_1 - l$ ist.

Es sei nun

$$n_h(\mathfrak{A}) = \varepsilon_3 a_3 + \varepsilon_2 a_2 + \varepsilon_1 a_1 \geq \frac{h^3}{24}$$

die Euklidische Darstellung von $n_h(\mathfrak{A})$.

Nach Hilfssatz 3 existieren dann Zahlen e'_3, e'_2 und e'_1 mit

$$\begin{aligned} \varepsilon_3 - 1 &\leq e'_3 \leq \varepsilon_3, \\ \gamma_1 - 2 &\leq e'_2 \leq \gamma_1 - 1, \\ \beta_s &\leq e'_1 \leq a_2 - 1, \end{aligned}$$

¹⁾ Zusatz bei der Korrektur. Satz 1 und Folgerung 5 der vorliegenden Arbeit finden sich auch bei W. Klotz, Dissertation, Techn. Univ. Braunschweig (1968).

Herrn Dr. Härtter verdanke ich den Hinweis auf J. Riddell, Master's Thesis, University of Alberta (1960), in der sich weitere numerische Werte für $n_s(k)$ befinden.

so daß $n' = e'_3 a_3 + e'_2 a_2 + e'_1 a_1 \leq n_h(\mathfrak{A})$ eine s -Stelle hat. (Die Voraussetzung $e'_3 \geq s$ in Hilfssatz 3 folgt z. B. leicht so: Für $h > h_0$ folgt $a_3 \leq h a_2 + 1 \leq h(h+1) + 1 < 2h^2$ und $(\varepsilon_3 + 1)a_3 > n_h(\mathfrak{A}) \geq \frac{h^3}{21}$, also $e'_3 \geq \varepsilon_3 - 1 > \frac{h}{42} - 2 > 21 \geq s$.)

Es folgt für die Koeffizientensumme der Darstellung

$$n' = (e'_3 - s)a_3 + (e'_2 + \gamma_s)a_2 + (e'_1 - \beta_s)a_1$$

die Abschätzung

$$e'_3 - s + e'_2 + \gamma_s + e'_1 - \beta_s \leq h$$

oder schwächer

$$\varepsilon_3 - 1 - s + \gamma_1 - 2 + \gamma_s \leq h,$$

also wegen $\gamma_s = s\gamma_1 - l$ die Abschätzung

$$\varepsilon_3 \leq h + 3 + s + l - (s+1)\gamma_1.$$

Da weiter

$$\varepsilon_2 a_2 + \varepsilon_1 a_1 \leq a_3 - 1 \leq h a_2 \leq h(h+1)$$

und

$$a_3 = \gamma_1 a_2 - \beta_1 \leq \gamma_1 a_2$$

ist, folgt

$$n_h(\mathfrak{A}) \leq (h + 3 + s + l - (s+1)\gamma_1)\gamma_1 a_2 + O(h^2).$$

Nach Folgerung 4 folgt $l \leq s - 1 \leq 20$, also

$$n_h(\mathfrak{A}) \leq (h + 44 - (s+1)\gamma_1)\gamma_1 a_2 + O(h^2)$$

und nach Hilfssatz 4

$$(1) \quad n_h(\mathfrak{A}) \leq \frac{(h+44)^2}{4(s+1)} a_2 + O(h^2).$$

Nun gilt für $h > h_0$ nach Voraussetzung

$$n_h(\mathfrak{A}) \geq \frac{h^3}{21}$$

und daher

$$(2) \quad a_3 \geq \frac{h^2}{21}.$$

Nach Hilfssatz 2 folgt daher

$$n_h(\{a_1, a_2\}) = (h+3)a_2 - a_2^2 - 2 \geq a_3 - 1 \geq \frac{h^2}{21} - 1,$$

also

$$a_2^2 - (h+3)a_2 + \frac{h^2}{21} + 1 \leq 0$$

oder

$$a_2 \leq \frac{h}{2} \left(1 + \sqrt{\frac{17}{21}} \right) + O(1),$$

also

$$(3) \quad a_2 \leq \frac{19}{20} h + O(1).$$

Die Ungleichungen (1) und (3) geben für $s \geq 4$

$$n_h(\mathfrak{A}) \leq \frac{19}{400} h^3 + O(h^2) = \frac{399}{400} \frac{h^3}{21} + O(h^2),$$

also für genügend große h einen Widerspruch zur Voraussetzung.

Bemerkung. Auch wenn man die Voraussetzungen von Satz 1 zu $n_h(\mathfrak{A}) \geq \frac{4}{81} h^3$ verschärft und überdies statt (2) die schärfere Ungleichung

$$\varepsilon_3 a_3 \geq \frac{4}{81} h^3 - O(h^2)$$

benutzt, so erhält man doch erst für $s \geq 4$ einen Widerspruch zur Voraussetzung.

§ 4. Abschätzungen von $n_h(3)$ nach oben

Zur asymptotischen Abschätzung von $n_h(3)$ nach oben genügt es nach Satz 1 aus § 3, sich auf solche Zahlenmengen $\mathfrak{A} = \{a_1, a_2, a_3\}$ mit $a_1 = 1 < a_2 < a_3$ und $n_h(\mathfrak{A}) \geq \frac{h^3}{21}$ zu beschränken, für die jedes n aus $[1, n_h(\mathfrak{A})]$ eine 0-, 1-, 2- oder 3-Stelle hat. Für solche $\mathfrak{A}$ bezeichnen wir mit

$$n_h(\mathfrak{A}) = \varepsilon_3 a_3 + \varepsilon_2 a_2 + \varepsilon_1 a_1$$

die Euklidische Darstellung von $n_h(\mathfrak{A})$ und betrachten die vier Zahlen

$$\begin{aligned} n_0 &= (\varepsilon_3 - 1) a_3 + (\gamma_1 - 2) a_2 + (a_2 - 1) a_1, \\ n_1 &= (\varepsilon_3 - 1) a_3 + (\gamma_1 - 2) a_2 + (\beta_1 - 1) a_1, \\ n_2 &= (\varepsilon_3 - 1) a_3 + (\gamma_1 - 2) a_2 + (\beta_2 - 1) a_1, \\ n_3 &= (\varepsilon_3 - 1) a_3 + (\gamma_1 - 2) a_2 + (\beta_3 - 1) a_1. \end{aligned}$$

Die Fälle $\beta_i = 0$ für $1 \leq i \leq 3$, also die Fälle $\beta_1 = 0$, $\frac{a_2}{3}$, $\frac{a_2}{2}$ oder $\frac{2}{3} a_2$ werden gesondert betrachtet und zunächst ausgeschlossen. Es sind dann offenbar die obigen Ausdrücke für $n_0, \dots, n_3$ bereits die Euklidischen Darstellungen von $n_0, \dots, n_3$, und jede dieser vier Zahlen liegt in $[1, n_h(\mathfrak{A})]$. Daher hat jede dieser vier Zahlen eine s -Stelle mit $0 \leq s \leq 3$, sofern nur h genügend groß ist. Dieser Sachverhalt wird im folgenden zu Abschätzungen von ε_3 benutzt. Für die Summen mit a_2 und a_1 gilt, daß sie $\leq a_3 - 1$ sind, also durch $O(h^2)$ abgeschätzt werden können. Wegen $a_3 = \gamma_1 a_2 - \beta_1 a_1$ kann a_3 durch $a_3 \leq \gamma_1 a_2$ abgeschätzt werden. Um zu gegebenen β_1 und γ_1 Aussagen über $\beta_2, \beta_3, \gamma_2$ und γ_3 zu erhalten, unterscheiden wir zunächst vier Hauptfälle.

I. $0 < \beta_1 < \frac{a_2}{3}$

Es folgt $0 < 2\beta_1 < \frac{2}{3} a_2$, also $\beta_2 = 2\beta_1$ und $\gamma_2 = 2\gamma_1$

$0 < 3\beta_1 < a_2$, also $\beta_2 = 3\beta_1$ und $\gamma_3 = 3\gamma_1$

mit $\beta_1 < \beta_2 < \beta_3$.

II. $\frac{a_2}{3} < \beta_1 < \frac{a_2}{2}$

Es folgt $\frac{2}{3} a_2 < 2\beta_1 < a_2$, also $\beta_2 = 2\beta_1$ und $\gamma_2 = 2\gamma_1$

$a_2 < 3\beta_1 < \frac{3}{2} a_2$, also $\beta_3 = 3\beta_1 - a_2$ und $\gamma_3 = 3\gamma_1 - 1$

mit $\beta_3 < \beta_1 < \beta_2$.

$$\text{III. } \frac{a_2}{2} < \beta_1 < \frac{2}{3} a_2$$

Es folgt $a_2 < 2\beta_1 < \frac{4}{3} a_2$, also $\beta_2 = 2\beta_1 - a_2$ und $\gamma_2 = 2\gamma_1 - 1$

$$\frac{3}{2} a_2 < 3\beta_1 < 2a_2, \text{ also } \beta_3 = 3\beta_1 - a_2 \text{ und } \gamma_3 = 3\gamma_1 - 1$$

mit $\beta_2 < \beta_1 < \beta_3$.

$$\text{IV. } \frac{3}{2} a_2 < \beta_1 < a_2$$

Es folgt $\frac{4}{3} a_2 < 2\beta_1 < 2a_2$, also $\beta_2 = 2\beta_1 - a_2$ und $\gamma_2 = 2\gamma_1 - 1$

$$2a_2 < 3\beta_1 < 3a_2, \text{ also } \beta_3 = 3\beta_1 - 2a_2 \text{ und } \gamma_3 = 3\gamma_1 - 2$$

mit $\beta_3 < \beta_2 < \beta_1$.

Die weitere Schlußweise werde nur am kompliziertesten, nämlich dem Fall IV erläutert.

Da n_0 eine 0-, 1-, 2- oder 3-Stelle hat, unterscheiden wir weiter

IV A. n_0 hat eine 0-Stelle.

Dann folgt $\varepsilon_3 - 1 + \gamma_1 - 2 + a_2 - 1 \leq h$, also

$$\varepsilon_3 \leq h + 4 - \gamma_1 - a_2.$$

Nach den vorangehenden Betrachtungen wird daher

$$n_h(\mathfrak{A}) \leq (h + 4 - \gamma_1 - a_2)\gamma_1 a_2 + O(h^2),$$

also nach Hilfssatz 4 aus § 3

$$n_h(\mathfrak{A}) \leq \frac{h^3}{27} + O(h^2).$$

IV B. n_0 hat eine 1-Stelle.

Dann folgt $\varepsilon_3 - 2 + 2\gamma_1 - 2 + a_2 - 1 - \beta_1 \leq h$, also

$$(1) \quad \varepsilon_3 \leq h + 5 - 2\gamma_1 - a_2 + \beta_1.$$

Nun beachten wir, daß n_1 eine 0-, 2- oder 3-Stelle hat und unterscheiden daher weiter

IV B 1. n_1 hat eine 0-Stelle.

Dann folgt $\varepsilon_3 - 1 + \gamma_1 - 2 + \beta_1 - 1 \leq h$, also

$$(2) \quad \varepsilon_3 \leq h + 4 - \gamma_1 - \beta_1.$$

Die Addition von (1) und (2) liefert

$$\varepsilon_3 \leq h + \frac{9}{2} - \frac{3}{2} \gamma_1 - \frac{a_2}{2},$$

also

$$n_h(\mathfrak{A}) \leq \left(h + \frac{9}{2} - \frac{3}{2} \gamma_1 - \frac{a_2}{2}\right)\gamma_1 a_2 + O(h^2),$$

also

$$n_h(\mathfrak{A}) \leq \frac{4}{81} h^3 + O(h^2).$$

IV B 2. n_1 hat eine 2-Stelle.

Dann folgt $\varepsilon_3 - 3 + 3\gamma_1 - 3 + a_2 - 1 - \beta_1 \leq h$, also

$$(3) \quad \varepsilon_3 \leq h + 7 - 3\gamma_1 - a_2 + \beta_1.$$

Nun beachten wir, daß n_2 wegen $\beta_3 < \beta_2 < \beta_1$ eine 0- oder 3-Stelle hat und unterscheiden abermals weiter

IV B 2a. n_2 hat eine 0-Stelle.

Dann folgt $\varepsilon_3 - 1 + \gamma_1 - 2 + 2\beta_1 - a_2 - 1 \leq h$, also

$$(4) \quad \varepsilon_3 \leq h + 4 - \gamma_1 + a_2 - 2\beta_1.$$

Die Addition des Doppelten von (3) zu (4) liefert

$$\varepsilon_3 \leq h + 6 - \frac{7}{3}\gamma_1 - \frac{a_2}{3},$$

also

$$n_h(\mathfrak{A}) \leq \frac{h^3}{24} + O(h^2).$$

IV B 2b. n_2 hat eine 3-Stelle.

Dann folgt $\varepsilon_3 - 4 + 4\gamma_1 - 4 + a_2 - 1 - \beta_1 \leq h$, also

$$(5) \quad \varepsilon_3 \leq h + 9 - 4\gamma_1 - a_2 + \beta_1.$$

Nun beachten wir, daß n_3 wegen $\beta_3 < \beta_2 < \beta_1$ eine 0-Stelle hat, also

$$\varepsilon_3 - 1 + \gamma_1 - 2 + 3\beta_1 - 2a_2 - 1 \leq h$$

oder

$$(6) \quad \varepsilon_3 \leq h + 4 - \gamma_1 + 2a_1 - 3\beta_1.$$

Die Addition des Dreifachen von (5) zu (6) liefert

$$\varepsilon_3 \leq h + \frac{31}{4} - \frac{13}{4}\gamma_1 - \frac{a_2}{4},$$

also

$$n_h(\mathfrak{A}) \leq \frac{16}{13} \frac{h^3}{27} + O(h^2).$$

IV B 3. n_1 hat eine 3-Stelle.

Dann folgt

$$(7) \quad \varepsilon_3 \leq h + 9 - 4\gamma_1 - 2a_2 + 2\beta_1.$$

Dann beachten wir wieder, daß n_3 eine 0-Stelle hat, also (6) gilt. Eine geeignete Kombination von (7) und (6) gibt

$$\varepsilon_3 \leq h + 7 - \frac{14}{5}\gamma_1 - \frac{2}{5}a_2,$$

also

$$n_h(\mathfrak{A}) \leq \frac{25}{28} \frac{h^3}{27} + O(h^2).$$

IV C. n_0 hat eine 2-Stelle.

Es folgt

$$(8) \quad \varepsilon_3 \leq h + 7 - 3\gamma_1 - 2a_2 + 2\beta_1.$$

Da auch (6) gilt, folgt aus (8) und (6)

$$\varepsilon_3 \leq h + \frac{29}{5} - \frac{11}{5}\gamma_1 - \frac{2}{5}a_2,$$

also

$$n_h(\mathfrak{A}) \leq \frac{25}{22} \frac{h^3}{27} + O(h^2).$$

IV D. n_0 hat eine 3-Stelle.

Es folgt

$$(9) \quad \varepsilon_3 \leq h + 9 - 4\gamma_1 - 3a_2 + 3\beta_1.$$

Zusammen mit (6) folgt

$$\varepsilon_3 \leq h + \frac{13}{2} - \frac{5}{2}\gamma_1 - \frac{a_2}{2},$$

also

$$n_h(\mathfrak{A}) \leq \frac{4}{5} \frac{h^3}{27} + O(h^2).$$

Zu den Fällen I, II und III merken wir hier nur an, daß auch bei ihnen danach unterschieden wird, ob n_0 eine 0-, 1-, 2- oder 3-Stelle hat. Im Fall I zieht man daneben noch $\beta_1 < \frac{a_2}{3}$ heran oder beachtet, daß n_3 eine 0-, 1- oder 2-Stelle hat; es gilt stets

$$n_h(\mathfrak{A}) \leq \frac{16}{13} \frac{h^3}{27} + O(h^2) < \frac{h^3}{21} + O(h^2).$$

Im Fall II zieht man noch $\beta_1 < \frac{a_2}{2}$ heran und beachtet, daß n_2 eine 0-, 1- oder 3-Stelle hat; es gilt stets

$$n_h(\mathfrak{A}) \leq \frac{h^3}{21} + O(h^2).$$

Im Fall III zieht man noch $\beta_1 < \frac{2}{3}a_2$ heran und beachtet, daß n_1 eine 0- oder 2-Stelle hat; hat n_0 eine 1-Stelle und n_1 eine 0-Stelle, so folgen

$$(10) \quad \varepsilon_3 \leq h + 5 - 2\gamma_1 - a_2 + \beta_1,$$

$$(11) \quad \varepsilon_3 \leq h + 4 - \gamma_1 - \beta_1,$$

also (vgl. IV B 1)

$$n_h(\mathfrak{A}) \leq \frac{4}{81} h^3 + O(h^2).$$

Ansonsten folgt im Fall III

$$n_h(\mathfrak{A}) \leq \frac{8}{7} \frac{h^3}{27} + O(h^2) < \frac{h^3}{21} + O(h^2).$$

Schließlich sind noch die Fälle $\beta_1 = 0$, $\frac{a_2}{3}$, $\frac{a_2}{2}$ oder $\frac{2}{3}a_2$ zu betrachten. Ist $\beta_1 = \frac{2}{3}a_2$, so folgt $\beta_2 = \frac{a_2}{3}$ und $\beta_3 = 0$, also jedenfalls $\beta_1 \geq 1$ und $\beta_2 \geq 1$. Wir schließen dann weiter wie in Fall IV, nur daß wir statt n_3 nun sogleich $\beta_1 = \frac{2}{3}a_2$ heranziehen. Entsprechend werden die Fälle $\beta_1 = 0$, $\frac{a_2}{3}$ oder $\frac{a_2}{2}$ in die Betrachtungen der Fälle I, II oder III eingeordnet. Zusammen mit Folgerung 1 aus § 2 ergibt sich daher unmittelbar

$$\text{Folgerung 5. } n_h(3) = \frac{4}{81} h^3 + O(h^2).$$

§ 5. Die Extremalbasen zu $n_h(3)$ für $h > h_0$

Es sei im folgenden c eine beliebige reelle Konstante mit $\frac{1}{21} < c \leq \frac{4}{81}$; ferner sei h_0 eine hinreichend große Zahl und $h > h_0$; schließlich werde für alle auftretenden $\mathfrak{A} = \{a_1, a_2, a_3\}$ im folgenden stets $n_h(\mathfrak{A}) \geq ch^3$ vorausgesetzt. Die Existenz solcher $\mathfrak{A}$ ergibt sich aus Folgerung 1 aus § 2. Definiert man weiter ε_j als Koeffizienten der Euklidischen Darstellung von $n_h(\mathfrak{A})$, also

$$n_h(\mathfrak{A}) = \varepsilon_3 a_3 + \varepsilon_2 a_2 + \varepsilon_1 a_1,$$

ferner β_1 und γ_1 durch $a_3 = \gamma_1 a_2 - \beta_1$ mit $0 \leq \beta_1 \leq a_2 - 1$, so folgt nach den Fallunterscheidungen von § 4

$$(1) \quad \beta_1 \geq \frac{a_2}{2}$$

$$(2) \quad \varepsilon_3 \leq h + 4 - \gamma_1 - \beta_1$$

$$(3) \quad \varepsilon_3 \leq h + 5 - 2\gamma_1 - a_2 + \beta_1$$

(da ja andernfalls stets $n_h(\mathfrak{A}) \leq \frac{h^3}{24} + O(h^2)$ folgte). Weiter gibt es offenbar zwei positive, reelle Konstanten c_1 und c_2 mit

$$(4) \quad c_1 h < \min \{\varepsilon_3, a_2, \gamma_1, \beta_1\} < \max \{\varepsilon_3, a_2, \gamma_1, \beta_1\} < c_2 h.$$

Zunächst beweist man jedoch analog zu Hilfssatz 4 aus § 3

Hilfssatz 5. *Es sei*

$$f = f(x, y) = \left(h + \frac{11}{2} - \frac{3}{2}x - \frac{y}{2} \right) \left(xy - \frac{x+y}{2} \right)$$

eine Funktion reeller Variabler x und y . Dann nimmt f in dem durch $x \geq 1$, $y \geq 1$ definierten Gebiet ihr Maximum an der Stelle

$$x = \frac{2}{9}h + \frac{3}{2} + O\left(\frac{1}{h}\right)$$

$$y = \frac{2}{3}h + \frac{7}{2} + O\left(\frac{1}{h}\right)$$

an und hat dort den Wert

$$f = \frac{4}{81}h^3 + \frac{2}{3}h^2 + O(h).$$

Ist nun $\mathfrak{A}$ eine Zahlenmenge vom zuvor beschriebenen Typ, so zeigen wir

Hilfssatz 6. *Aus $a_2 + \gamma_1 - 2\beta_1 \neq 1$ folgt*

$$(9) \quad n_h(\mathfrak{A}) \leq \frac{4}{81}h^3 + \left(\frac{2}{3} - \frac{c_1^2}{2} \right) h^2 + O(h).$$

Beweis. a) Ist $a_2 + \gamma_1 - 2\beta_1 = 1 - n$ mit $n \geq 1$, so beachten wir, daß nach (2)

$$\varepsilon_3 \leq h + 4 - \gamma_1 - \beta_1 = h + \frac{9-n}{2} - \frac{3}{2}\gamma_1 - \frac{a_2}{2}$$

folgt; ferner gilt

$$a_3 = \gamma_1 a_2 - \frac{a_2 + \gamma_1 - 1 + n}{2}$$

sowie stets

$$\varepsilon_2 a_2 + \varepsilon_1 a_1 < a_3 < \gamma_1 a_2.$$

Es folgt

$$n_h(\mathfrak{A}) \leq \left(h + \frac{9}{2} - \frac{3}{2} \gamma_1 - \frac{a_2}{2} - \frac{n}{2} \right) \left(\gamma_1 a_2 - \frac{\gamma_1 + a_2}{2} - \frac{n-1}{2} \right) + \gamma_1 a_2,$$

unter Berücksichtigung von (4) also

$$n_h(\mathfrak{A}) \leq f(\gamma_1, a_2) - \frac{n}{2} \gamma_1 a_2 + O(h).$$

Nach Hilfssatz 5 und (4) folgt (9).

b) Ist $a_2 + \gamma_1 - 2\beta_1 = 1 + m$ mit $m \geq 1$, so beachten wir, daß nach (3) noch

$$\varepsilon_3 \leq h + 5 - 2\gamma_1 - a_2 + \beta_1 = h + \frac{9}{2} - \frac{3}{2} \gamma_1 - \frac{a_2}{2} - \frac{m}{2}$$

folgt und ähnlich zu a) daher

$$\begin{aligned} n_h(\mathfrak{A}) &\leq \left(h + \frac{9}{2} - \frac{3}{2} \gamma_1 - \frac{a_2}{2} - \frac{m}{2} \right) \left(\gamma_1 a_2 - \frac{\gamma_1 + a_2}{2} + \frac{m+1}{2} \right) + \gamma_1 a_2 \\ &\leq f(\gamma_1, a_2) - \frac{m}{2} \gamma_1 a_2 + O(mh). \end{aligned}$$

Nach Hilfssatz 5 und (4) folgt ebenfalls (9).

Ein Vergleich mit Folgerung 1 aus § 2 gibt daher unmittelbar

Folgerung 6. Für Extremalbasen zu $n_h(3)$ gilt (sofern nur $h > h_0$ ist)

$$(10) \quad a_2 = 2\beta_1 - \gamma_1 + 1.$$

Bemerkung. Diese Folgerung motiviert übrigens auch (2) in Hilfssatz 1 aus § 2.

Für den Rest dieses Paragraphen werde daher für alle auftretenden $\mathfrak{A}$ noch überdies (10) vorausgesetzt. Für derartige $\mathfrak{A}$ zeigen wir, daß die Zahl

$$(11) \quad z = (h + 4 - \gamma_1 - \beta_1) a_3 + (\gamma_1 - 2) a_2 + (\beta_1 - 1) a_1$$

größer als $n_h(\mathfrak{A})$ ist. Dazu genügt es nach Satz 1 aus § 3 zu zeigen, daß z keine 0-, 1-, 2- oder 3-Stelle hat. Wir bemerken zunächst, daß (11) bereits die Euklidische Darstellung von z ist. Für die Koeffizientensumme gilt

$$h + 4 - \gamma_1 - \beta_1 + \gamma_1 - 2 + \beta_1 - 1 = h + 1,$$

also besitzt z keine 0-Stelle. Da der Koeffizient von a_1 kleiner als β_1 ist, besitzt z keine 1-Stelle. Hätte z eine 2-Stelle, so auch wegen (2) die Zahl

$$n_1 = (\varepsilon_3 - 1) a_3 + (\gamma_1 - 2) a_2 + (\beta_1 - 1) a_1;$$

im Fall $\frac{a_2}{2} \leq \beta_1 < \frac{2}{3} a_2$ würde unmittelbar

$$n_h(\mathfrak{A}) \leq \frac{h^3}{27} + O(h^2)$$

folgen; im Fall $\beta_1 \geq \frac{2}{3} a_2$ würde stets

$$n_h(\mathfrak{A}) \leq \frac{h^3}{21} + O(h^2)$$

folgen (vgl. § 4, Fall IV B 2a und IV B 2b). Hätte z eine 3-Stelle, so würde $\beta_1 \geq \frac{2}{3} a_2$ folgen (denn für $\frac{a_2}{2} \leq \beta_1 < \frac{2}{3} a_2$ gilt $\beta_3 \geq \beta_1$); ferner hätte auch n_1 eine 3-Stelle; es würde

$$n_h(\mathfrak{A}) \leq \frac{25}{28} \frac{h^3}{27} + O(h^2)$$

folgen (vgl. § 4, Fall IV B 3).

Für alle hier betrachteten $\mathfrak{A}$, insbesondere also auch für die Extremalbasen zu $n_h(3)$ für $h > h_0$, gilt daher

$$(12) \quad n_h(\mathfrak{A}) \leq (h + 4 - \beta_1 - \gamma_1) a_3 + (\gamma_1 - 2) a_2 + (\beta_1 - 2) a_1$$

mit

$$(13) \quad a_1 = 1, a_2 = 2\beta_1 - \gamma_1 + 1 \text{ und } a_3 = \gamma_1 a_2 - \beta_1.$$

Daher folgt, wenn man die rechte Seite von (12) mit $F = F(\beta_1, \gamma_1)$ bezeichnet,

$$F = (h + 4 - \beta_1 - \gamma_1) (\gamma_1(2\beta_1 - \gamma_1 + 1) - \beta_1) + (\gamma_1 - 2) (2\beta_1 - \gamma_1 + 1) + \beta_1 - 2,$$

worin Hilfssatz 5 mit $\gamma_1 = x$ und $2\beta_1 = x + y - 1$ die Substitutionen

$$(14) \quad \begin{aligned} \varepsilon &= \gamma_1 - \frac{2}{9} h - \frac{3}{2} \\ \delta &= \beta_1 - \frac{4}{9} h - 2 \end{aligned}$$

nahelegt. Die reellen Zahlen ε und δ sind dabei so zu wählen, daß β_1 und γ_1 ganzzahlig ausfallen. Der weitere Gedankengang besteht nun darin, die optimalen ε und δ zu finden und für die zugehörigen β_1 und γ_1 die Eigenschaften (1) aus § 2, Hilfssatz 1 nachzuprüfen. Für $h > h_0$ gilt für Extremalbasen zu $n_h(3)$ nach (14) und (4) offenbar $\varepsilon = O(h)$ und $\delta = O(h)$. Mittels der Substitutionen $x_1 = \frac{3\varepsilon}{2h}$ und $x_2 = \frac{\delta}{h} - \frac{\varepsilon}{2h}$ macht man sich noch leicht $\varepsilon = o(h)$ und $\delta = o(h)$ klar (x und y reell). Denn es wird

$$\begin{aligned} F &= \frac{4}{81} h^3 ((1 - 3x_1 - 3x_2) (1 + 3x_1) (1 + 3x_2) + o(1)) \\ &= \frac{4}{81} h^3 (1 - 9x_2^2 (\xi^2 \eta + \xi \eta + 1) + o(1)), \end{aligned}$$

wo ξ durch $x_1 = \xi x_2$ und η durch $1 + 3x_2 = \eta$ bestimmt sind (ξ und η reell). Wegen

$$c_1 h < a_2 = 2\beta_1 - \gamma_1 + 1 = \frac{2}{3} h + \frac{7}{2} + 2hx_2 \leq h + 1$$

folgt für $h > h_0$ daher $\frac{c_1}{2} - \frac{1}{3} - o(1) < x_2 \leq \frac{1}{6}$, also $0 < \frac{3}{2} c_1 - o(1) < \eta \leq \frac{3}{2}$ und

damit $\xi^2 \eta + \xi \eta + 1 \geq \frac{5}{8}$, so daß für Extremalbasen zu $n_h(3)$ nach § 2 noch $x_2 = o(1)$

folgt und wegen $x_1 = O(1)$ und

$$F = \frac{4}{81} h^3 ((1 - 3x_1) (1 + 3x_1) + o(1))$$

auch $x_1 = o(1)$, also $\varepsilon = o(h)$ und $\delta = o(h)$.

Da nun

$$F = \frac{4}{81}h^3 + \frac{2}{3}h^2 + \frac{95}{36}h - \frac{h}{9}E - \frac{D}{8}$$

gilt mit

$$E = E(\varepsilon, \delta) = 7\varepsilon^2 + 2\varepsilon\delta + 4\delta^2$$

und

$$D = D(\varepsilon, \delta) = -8\varepsilon^3 + 8\varepsilon^2\delta + 16\delta^2\varepsilon + 28\varepsilon^2 + 8\varepsilon\delta + 16\delta^2 - 14\varepsilon + 18\delta + 1,$$

genügt es für die Extremalbasen zu $n_h(3)$ für $h > h_0$, lediglich die in oder wenigstens auf der Ellipse

$$(15) \quad E(\varepsilon, \delta) = c_3 \quad (c_3 \text{ eine geeignete reelle Zahl})$$

gelegenen Gitterpunkte der Form

$$(16) \quad \begin{aligned} \varepsilon &= \left\{ \frac{2}{9}h + \frac{3}{2} \right\} - \frac{2}{9}h - \frac{3}{2} + u \\ \delta &= \left\{ \frac{4}{9}h + 2 \right\} - \frac{4}{9}h - 2 + v \end{aligned}$$

zu betrachten, wo u und v über alle ganzen Zahlen laufen. Wählt man in (15) $c_3 = \frac{7}{4}$, so liegt für jedes $h \equiv 0 \pmod{9}$ genau ein Gitterpunkt im Innern von (15), nämlich wenn $u = v = 0$ ist. Für $h \equiv 0 \pmod{9}$ liegen keine Gitterpunkte der Form (16) im Innern von (15), jedoch zwei auf deren Rande, nämlich wenn $u = v = 0$ und wenn $u = -1, v = 0$ ist; für den letzteren Gitterpunkt ist jedoch D größer als für den mit $u = v = 0$. Daher erhält man für die optimalen ε und δ folgende Tabelle.

Für $h \pmod{9}$	ist	(ε, δ)	optimal mit	$12E(\varepsilon, \delta)$	und	$\frac{81}{8}D(\varepsilon, \delta)$
0		$\left(\frac{1}{2}, 0\right)$		21		0
1		$\left(\frac{5}{18}, -\frac{4}{9}\right)$		13		-62
2		$\left(\frac{1}{18}, \frac{1}{9}\right)$		1		26
3		$\left(-\frac{1}{6}, -\frac{1}{3}\right)$		9		0
4		$\left(-\frac{7}{18}, \frac{2}{9}\right)$		13		154
5		$\left(\frac{7}{18}, -\frac{2}{9}\right)$		13		-46
6		$\left(\frac{1}{6}, \frac{1}{3}\right)$		9		81
7		$\left(-\frac{1}{18}, -\frac{1}{9}\right)$		1		1
8		$\left(-\frac{5}{18}, \frac{4}{9}\right)$		13		170

Es folgt für die optimalen β_1 und γ_1 :

$$(17) \quad \begin{aligned} \beta_1 &= \left\{ \frac{4}{9}h + 2 \right\} = \left[\frac{5}{9}h \right] - \left[\frac{h}{9} \right] + 2, \\ \gamma_1 &= \left\{ \frac{2}{9}h + \frac{3}{2} \right\} = \left[\frac{2}{9}h \right] + 2 \end{aligned}$$

(vgl. Wahl 2 aus § 2). Nach § 2 sind für $h \geq 18$ mit (17) die Voraussetzungen von Hilfssatz 1 aus § 2 erfüllt; zusammen mit (12) folgt daher

Satz 2. Ist $h > h_0$, so erhält man die Extremalbasen zu $n_h(3)$ wie folgt:

Man wähle

$$\begin{aligned} \beta &= \left[\frac{5}{9}h \right] - \left[\frac{h}{9} \right] + 2, \\ \gamma &= \left[\frac{2}{9}h \right] + 2 \end{aligned}$$

und setze $a_1 = 1$, $a_2 = 2\beta - \gamma + 1$ und $a_3 = \gamma a_2 - \beta$.

Es gilt dann

$$n_h(3) = n_h(\mathfrak{A}) = (h + 4 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (\beta - 2)a_1.$$

Aus dem Beweis von Hilfssatz 1 aus § 2 folgt noch der

Zusatz 1. Für $h > h_0$ haben die Extremalbasen zu $n_h(3)$ in $[1, n_h(\mathfrak{A})]$ stets 0- oder 1-Stellen.

Weiter entnimmt man dem vorangehenden noch

Zusatz 2. Für $h > h_0$ gilt

$$(18) \quad \frac{4}{81}h^3 + \frac{2}{3}h^2 + \frac{66}{27}h \leq n_h(3) \leq \frac{4}{81}h^3 + \frac{2}{3}h^2 + \frac{71}{27}h - \frac{1}{81},$$

und in (18) gilt links das Gleichheitszeichen für $h \equiv 0 \pmod{9}$ und rechts das Gleichheitszeichen für $h \equiv 7 \pmod{9}$.

§ 6. Numerische Resultate für $h \leq 34$ ²⁾

Bei der nachfolgenden Tabelle ist zu beachten, daß in den mit ⁺) bezeichneten Fällen nicht alle $\mathfrak{A} = \{a_1, a_2, a_3\}$ mit $n_h(\mathfrak{A}) \geq a_3$ betrachtet wurden, sondern nur solche $\mathfrak{A}$, die überdies der einschränkenden Bedingung $h - 10 \leq a_2 \leq h + 1$ genügen. Mit dieser Einschränkung gibt die Tabelle für jedes $h \leq 34$ alle optimalen Mengen $\mathfrak{A}$ mit zugehörigen $n_h(\mathfrak{A})$.

²⁾ Die nachfolgenden numerischen Resultate wurden von Herrn E. Andersen an Rechenanlagen an Norwegens Technischer Hochschule in Trondheim erhalten. Für seinen beträchtlichen Einsatz und sein unermüdliches Eingehen auf die Wünsche des Verfassers sowie für seine Erlaubnis, die nachfolgende Tabelle hier zu veröffentlichen, sei ihm an dieser Stelle herzlich gedankt.

Tabelle (E. Andersen)

h	optimale	$a_1,$	$a_2,$	a_3	$n_h(\mathfrak{A})$
1		1,	2,	3	3
2		1,	3,	4	8
3		1,	4,	5	15
4		1,	5,	8	26
5		1,	6,	7	35
6		1,	7,	12	52
7		1,	8,	13	69
8		1,	9,	14	89
9		1,	9,	20	112
10		1,	10,	26	146
11	1, 9, 30 und	1,	10,	26	172
12		1,	11,	37	212
13		1,	13,	34	259
14		1,	12,	52	302
15		1,	12,	52	354
16		1,	15,	54	418
17		1,	14,	61	476
18		1,	15,	80	548
19		1,	18,	65	633
20		1,	17,	91	714
21 ⁺)		1,	17,	91	805
22 ⁺)		1,	20,	92	902
23 ⁺)		1,	18,	114	1012
24 ⁺)		1,	20,	127	1127
25		1,	20,	127	1254
26 ⁺)		1,	22,	140	1382
27 ⁺)		1,	21,	154	1524
28 ⁺)		1,	21,	154	1678
29 ⁺)		1,	23,	169	1841
30 ⁺)		1,	23,	169	2010
31 ⁺)		1,	25,	184	2188
32 ⁺)		1,	24,	200	2382
33 ⁺)		1,	26,	217	2584
34 ⁺)		1,	26,	217	2801

Die Angaben für $h = 1, 2, 3$ und 4 befinden sich bereits bei Stöhr [5].

Bemerkung. Man erhält die in der Tabelle angegebenen optimalen $\mathfrak{A}$ mit zugehörigen $n_h(\mathfrak{A})$ für $h = 2, 11, 15$ und 17 auch, indem man in § 2 die Wahl 1 heranzieht, wo man im Fall $h = 11$ die Menge $\{1, 9, 30\}$ erhält.

Für $h = 12, 14, 18, 20, 21, 23, 24, \dots, 34$ erhält man die entsprechenden Angaben auch aus § 5, Satz 2 oder § 2, Wahl 2, wo man für $h = 12$ und $h = 14$ die Voraussetzungen von Hilfssatz 1 aus § 2 nachprüft.

In den verbleibenden Fällen der Tabelle, also für $h = 3, 4, \dots, 10, 13, 16, 19, 22$ sowie für $h = 11$ im Fall der Menge $\{1, 10, 26\}$ hat man es mit anderen als den in § 5 (speziell Folgerung 6) betrachteten Mengen zu tun.

§ 7. Vorbemerkungen zu h -Mengen

Wir setzen wieder

$$(1) \quad a_3 = \gamma a_2 - \beta \text{ mit } 0 \leq \beta \leq a_2 - 1.$$

Bemerkung. Für h -Mengen $\mathfrak{A} = \{a_1, a_2, a_3\}$ gilt bekanntlich (Salié [4])

$$(2) \quad h + 3 - a_2 - \gamma = \left\langle \frac{1 - \beta}{a_2} \right\rangle.$$

Denn setzt man $\mathfrak{B} = \{a_1, a_2\}$, so folgt nach Definition der h -Mengen

$$n_h(\mathfrak{B}) \geq a_3 - 1 > n_{h-1}(\mathfrak{B}),$$

also nach (1) und § 3, Hilfssatz 2

$$(h + 3)a_2 - a_2^2 - 2 \geq \gamma a_2 - \beta - 1 > (h + 2)a_2 - a_2^2 - 2$$

und daher

$$h + 3 - a_2 - \gamma \geq \left\langle \frac{1 - \beta}{a_2} \right\rangle > h + 2 - a_2 - \gamma.$$

Der Fall $\beta = 0$ ist im Hinblick auf die Bestimmung von $l_h(3)$ von geringem Interesse. Denn nach (2) folgt $a_2 = h + 2 - \gamma$, also nach [1], Satz 1 und 2 (vgl. auch Salié [4], Satz 2) sowie § 3, Hilfssatz 4

$$n_h(\mathfrak{A}) = 2\gamma a_2 - 2 = 2(h + 2 - \gamma)\gamma - 2 \leq 2\left(\frac{h + 2}{2}\right)^2 - 2 = O(h^2).$$

Beim Studium von $l_h(3)$ kann daher o. B. d. A. $\beta \geq 1$ vorausgesetzt werden und

$$(3) \quad a_2 = h + 3 - \gamma.$$

§ 8. Abschätzungen von $l_h(3)$ für gerade h

Ist h gerade, so wählen wir in § 2, Hilfssatz 1

$$(1) \quad \gamma = \left\{ \frac{3 - \sqrt{3}}{6} (h + 2) + \frac{4}{3} \right\}$$

und erhalten aus § 2, (2) sowie § 7, (3)

$$a_2 = 2\beta - \gamma + 1 = h + 3 - \gamma,$$

also $\beta = \frac{h}{2} + 1$.

Es sind dann für alle geraden $h \geq 2$ die Voraussetzungen aus § 2, Hilfssatz 1 erfüllt, und wir erhalten

Folgerung 7. Für gerade $h \geq 2$ gilt

$$l_h(3) > \frac{\sqrt{3}}{36} h^3 + \frac{5 + 2\sqrt{3}}{12} h^2 + \frac{120 + 19\sqrt{3}}{72} h - \frac{133 - 18\sqrt{3}}{216}.$$

Bemerkung. Dieses Resultat folgt auch, wenn man bei Salié [4], Hilfssatz 2 oder Satz 4 in der dortigen Schreibweise

$$r = a_2 - \beta = \beta - \gamma + 1 = \frac{h}{2} + 2 - \left\{ \frac{3 - \sqrt{3}}{6} (h + 2) + \frac{4}{3} \right\}$$

und

$$f = \gamma - 1 = \left\{ \frac{3 - \sqrt{3}}{6} (h + 2) + \frac{1}{3} \right\}$$

setzt.

Wegen $l_h(3) \geq ch^3$ mit einem reellen $c > \frac{1}{21}$ können die Überlegungen der §§ 3 bis 5 teils direkt teils analog angewandt werden, wobei man gemäß (3) aus § 7 noch zusätzlich a_2 eliminieren kann, so daß nur γ als geeignet zu wählende Größe zurückbleibt. Wie in § 5 erhält man für Extremalbasen zu $l_h(3)$ für $h > h_0$ die Bedingungen

$$(2) \quad \beta \geq \frac{a_2}{2},$$

$$(3) \quad \varepsilon_3 \leq h + 4 - \beta - \gamma,$$

$$(4) \quad \varepsilon_3 \leq h + 5 - 2\gamma - a_2 + \beta = \beta - \gamma + 2$$

(da ja andernfalls $n_h(\mathfrak{A}) \leq \frac{h^3}{21} + O(h^2)$ gilt). Analog zu § 5, Folgerung 6 erhält man weiter für Extremalbasen zu $l_h(3)$ für $h > h_0$ die Bedingung $a_2 = 2\beta - \gamma + 1$ oder $2\beta = h + 2$, und für die hier interessierenden $\mathfrak{A}$ gilt in § 2, (3) das Gleichheitszeichen.

Setzt man mit diesen Beziehungen

$$\begin{aligned} F(\gamma) &= (h + 4 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (\beta - 2)a_1 \\ &= \left(\frac{h}{2} + 4 - \gamma\right) \left((h + 3)\gamma - \gamma^2 - \frac{h}{2} - 3\right) + 2, \end{aligned}$$

und

$$\varepsilon = \gamma - \frac{3 - \sqrt{3}}{6} h - \frac{7 - \sqrt{3}}{3},$$

wo die reelle Zahl ε so zu wählen ist, daß γ ganz ausfällt, so gilt für Extremalbasen zu $l_h(3)$ für gerade $h > h_0$ offenbar $\varepsilon = O(h)$. Mittels der Substitution $\varepsilon = hx$ (x reell) folgt wegen

$$F(\gamma) = \frac{\sqrt{3}}{36} h^3 \left(1 - 12\sqrt{3} x^2 \left(\frac{\sqrt{3}}{2} - x\right) + o(1)\right)$$

und $\gamma = \frac{3 - \sqrt{3}}{6} h + xh + O(1) \leq h$, also $x \leq \frac{3 + \sqrt{3}}{6} + O\left(\frac{1}{h}\right)$, also $\frac{\sqrt{3}}{2} - x > \frac{1}{15}$ für $h > h_0$, daß $x = o(1)$ oder $\varepsilon = o(h)$ ist.

Die Gleichung

$$(5) \quad F(\gamma) = \frac{\sqrt{3}}{36} h^3 + \frac{5 + 2\sqrt{3}}{12} h^2 + \frac{30 + 7\sqrt{3}}{18} h - \frac{11 - 9\sqrt{3}}{27} - \frac{\sqrt{3}}{2} h \varepsilon^2 - D,$$

wo $D = D(\varepsilon) = -\varepsilon^3 + \sqrt{3} \varepsilon^2 + \frac{\varepsilon}{3}$ gesetzt ist, lehrt daher, daß für Extremalbasen zu $l_h(3)$ für gerade $h > h_0$ der Ausdruck $\frac{\sqrt{3}}{2} h \varepsilon^2$ minimal zu werden hat.

Ähnlich zu § 5, Satz 2 folgt daher

Satz 3. Für gerade $h > h_0$ erhält man die Extremalbasen zu $l_h(3)$ wie folgt:

Man wähle

$$\beta = \frac{h}{2} + 1,$$

$$\gamma = \left\{ \frac{3 - \sqrt{3}}{6} h + \frac{7 - \sqrt{3}}{3} \right\}$$

und setze $a_1 = 1$, $a_2 = 2\beta - \gamma + 1$ und $a_3 = \gamma a_2 - \beta$. Es gilt dann

$$l_h(3) = n_h(\mathfrak{A}) = (h + 4 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (\beta - 2)a_1.$$

Zusatz 3. Für gerade $h > h_0$ haben die Extremalbasen zu $l_h(3)$ in $[1, n_h(\mathfrak{A})]$ stets 0- oder 1-Stellen.

Zusatz 4. Für gerade $h > h_0$ gilt

$$\begin{aligned} \frac{\sqrt{3}}{36} h^3 + \frac{5 + 2\sqrt{3}}{12} h^2 + \left(\frac{30 + 7\sqrt{3}}{18} - \frac{\sqrt{3}}{8} \right) h - \left(\frac{11 - 9\sqrt{3}}{27} + \frac{1 + 6\sqrt{3}}{24} \right) \\ < l_h(3) < \frac{\sqrt{3}}{36} h^3 + \frac{5 + 2\sqrt{3}}{12} h^2 + \frac{30 + 7\sqrt{3}}{18} h - \frac{11 - 9\sqrt{3}}{27} + \frac{1}{h}. \end{aligned}$$

§ 9. Abschätzungen von $l_h(3)$ für ungerade h

Es werde weiterhin o. B. d. A.

$$(1) \quad a_2 = h + 3 - \gamma$$

sowie $\beta \geq 1$ vorausgesetzt. Wir bemerken, daß für ungerade h die bisher benutzte Beziehung $a_2 = 2\beta - \gamma + 1$ wegen (1) nicht mit der Ganzzahligkeit von β verträglich ist; in (3) und (4) aus § 8 kann nun nicht gleichzeitig das Gleichheitszeichen stehen.

Dann zeigt man analog dem Beweis von Hilfssatz 1 aus § 2

Hilfssatz 7. Für ungerade $h \geq 3$ sei $2\beta = h + 1$ und γ beliebig mit

$$(2) \quad 2 \leq \gamma \leq \beta + 1$$

gegeben. Setzt man dann

$$(3) \quad a_1 = 1, \quad a_2 = 2\beta - \gamma + 2 \quad \text{und} \quad a_3 = \gamma a_2 - \beta,$$

so folgt

$$(4) \quad n_h(\mathfrak{A}) \geq (h + 3 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (a_2 - 2)a_1.$$

Beim Beweis hat man — einfacher als bei Hilfssatz 1 aus § 2 — im wesentlichen nur die beiden Fälle

$$1) \quad e_3 + e_2 = h + 2 - \beta$$

$$2) \quad e_3 + e_2 = h + 1 - \beta \quad \text{und} \quad e_1 = a_2 - 1$$

auf einen Widerspruch zu führen.

Ist h ungerade, so wählen wir in Hilfssatz 7

$$(5) \quad \gamma = \left\{ \frac{3 - \sqrt{3}}{6} (h + 2) + \frac{7}{6} \right\}.$$

Es sind dann für alle ungeraden $h \geq 3$ die Voraussetzungen aus Hilfssatz 7 erfüllt, und wir erhalten

Folgerung 8. Für ungerade $h \geq 3$ gilt

$$l_h(3) > \frac{\sqrt{3}}{36} h^3 + \frac{2 + \sqrt{3}}{6} h^2 + \frac{24 + 7\sqrt{3}}{18} h - \frac{34 - 9\sqrt{3}}{27}.$$

Für Extremalbasen zu $l_h(3)$ für ungerade $h > h_0$ folgt wie in § 5, Hilfssatz 6, daß die Fälle

$$a_2 + \gamma - 2\beta = -n \text{ mit } n \geq 1$$

oder

$$a_2 + \gamma - 2\beta = 2 + m \text{ mit } m \geq 1$$

nicht in Betracht kommen. Bezeichnet man durch $a_2 + \gamma - 2\beta = 0$, also durch $2\beta = h + 3$ gekennzeichnete h -Mengen mit $\mathfrak{A}'$, so folgt analog der Herleitung von (12) aus § 5 (vgl. auch Salié [4], Satz 4)

$$\begin{aligned} n_h(\mathfrak{A}') &\leq (h + 4 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (\beta - 2)a_1 \\ &= \left(\frac{h+7}{2} - \gamma\right) \left(\gamma(h+3-\gamma) - \frac{h+3}{2}\right) - 1 + 2\gamma - h - 4. \end{aligned}$$

Andererseits folgt für h -Mengen $\mathfrak{A}$ aus Hilfssatz 7

$$n_h(\mathfrak{A}) \geq \left(\frac{h+7}{2} - \gamma\right) \left(\gamma(h+3-\gamma) - \frac{h+3}{2}\right) - 1,$$

also $n_h(\mathfrak{A}) - n_h(\mathfrak{A}') \geq h + 4 - 2\gamma \geq 1$ für alle γ , die (2) erfüllen, und für die hier interessierenden $\mathfrak{A}$ zeigt man ähnlich den Betrachtungen von § 5, daß in (4) von Hilfssatz 7 das Gleichheitszeichen steht. Weiter zeigt man mittels der Substitution

$$\gamma = -\frac{3-\sqrt{3}}{6}(h+2) + \frac{7}{6} + \varepsilon,$$

wo die reelle Zahl ε so zu wählen ist, daß γ ganz ausfällt, daß die Funktion

$$\begin{aligned} F(\gamma) &= \left(\frac{h+7}{2} - \gamma\right) \left(\gamma(h+3-\gamma) - \frac{h+3}{2}\right) - 1 \\ &= \frac{\sqrt{3}}{36} h^3 + \frac{2 + \sqrt{3}}{6} h^2 + \frac{96 + 37\sqrt{3}}{72} h - \frac{64 - 63\sqrt{3}}{108} - \frac{\sqrt{3}}{2} h\varepsilon^2 - D, \end{aligned}$$

wo $D = D(\varepsilon) = -\varepsilon^3 + \sqrt{3} \varepsilon^2 + \frac{13}{12} \varepsilon$ gesetzt ist, unter der genannten Einschränkung für ε ihr Maximum an der Stelle

$$\gamma = \left\{ \frac{3-\sqrt{3}}{6}(h+2) + \frac{7}{6} \right\}$$

hat, und derartige γ erfüllen für alle ungeraden $h > h_0$ die Voraussetzung (2) aus Hilfssatz 7. Es folgt

Satz 4. Für ungerade $h > h_0$ erhält man die Extremalbasen zu $l_h(3)$ wie folgt:
Man wähle

$$\begin{aligned} \beta &= \frac{h+1}{2}, \\ \gamma &= \left\{ \frac{3-\sqrt{3}}{6} h + \frac{13-2\sqrt{3}}{6} \right\} \end{aligned}$$

und setze $a_1 = 1$, $a_2 = 2\beta - \gamma + 2$ und $a_3 = \gamma a_2 - \beta$. Es gilt dann

$$l_h(3) = n_h(\mathfrak{A}) = (h + 3 - \beta - \gamma)a_3 + (\gamma - 2)a_2 + (a_2 - 2)a_1.$$

Zusatz 5. Für ungerade $h > h_0$ haben die Extremalbasen zu $l_h(3)$ in $[1, n_h(\mathfrak{A})]$ stets 0- oder 1-Stellen.

Zusatz 6. Für ungerade $h > h_0$ gilt

$$\begin{aligned} \frac{\sqrt{3}}{36} h^3 + \frac{2 + \sqrt{3}}{6} h^2 + \left(\frac{96 + 37\sqrt{3}}{72} - \frac{\sqrt{3}}{8} \right) h - \left(\frac{64 - 63\sqrt{3}}{108} + \frac{5 + 3\sqrt{3}}{12} \right) \\ < l_h(3) < \frac{\sqrt{3}}{36} h^3 + \frac{2 + \sqrt{3}}{6} h^2 + \frac{96 + 37\sqrt{3}}{72} h - \frac{64 - 63\sqrt{3}}{108} + \frac{1}{h}. \end{aligned}$$

§ 10. Die Extremalbasen zu $g_h(3)$ für $h > h_0$

Der Vollständigkeit halber geben wir das Analogon der Sätze 2, 3 und 4 für reguläre Mengen an. Die hier benötigten Ergebnisse aus [1] werden vorausgesetzt. Aufgrund von [1], Folgerung 1 gilt für die Extremalbasen zu $g_h(3)$ in unserer Bezeichnung

$$a_3 = \gamma a_2 - 1.$$

Weiter folgt nach [1], Satz 1 und 3 für solche Zahlenmengen $\mathfrak{A}$

$$\begin{aligned} n_h(\mathfrak{A}) &= (h + 4 - \gamma - a_2)a_3 + (\gamma - 2)a_2 + (a_2 - 2)a_1 \\ &= (h + 5 - \gamma - a_2)\gamma a_2 + \gamma - h - 6 \\ &= \left(\frac{h + 5}{3} \right)^3 - 2 \frac{h + 5}{3} - 1 - \frac{h + 5}{3} E - D, \end{aligned}$$

wo $E = E(\varepsilon, \delta) = \varepsilon^2 + \varepsilon\delta + \delta^2$ und $D = D(\varepsilon, \delta) = \varepsilon^2\delta + \varepsilon\delta^2 - \delta$ gesetzt sind und wo die reellen Zahlen ε und δ so eingeschränkt sind, daß in

$$\gamma = \frac{h + 5}{3} + \delta,$$

$$a_2 = \frac{h + 5}{3} + \varepsilon$$

γ und a_2 ganz ausfallen. Dann betrachten wir die Ellipse

$$(1) \quad \varepsilon^2 + \varepsilon\delta + \delta^2 = \frac{1}{3}$$

und die in oder auf (1) gelegenen Gitterpunkte der Form

$$(2) \quad \begin{aligned} \varepsilon &= \left\{ \frac{h + 5}{3} \right\} - \frac{h + 5}{3} + u, \\ \delta &= \left\{ \frac{h + 5}{3} \right\} - \frac{h + 5}{3} + v, \end{aligned}$$

wo u und v ganzzahlig sind. Für $h \equiv 1 \pmod{3}$ liegt genau ein Gitterpunkt im Innern von (1). Für $h \not\equiv 1 \pmod{3}$ liegen keine Gitterpunkte der Form (2) im Innern von (1), jedoch genau drei auf dem Rande von (1). Es genügt daher, folgende Fälle zu betrachten:

Für $h \pmod{3}$	liefert	(ε, δ)	für	$3E(\varepsilon, \delta)$	und	$27D(\varepsilon, \delta)$
1		$(0, 0)$		0		0
0		$\left(\frac{1}{3}, \frac{1}{3}\right)$		1		—7
0		$\left(-\frac{2}{3}, \frac{1}{3}\right)$		1		—7
0		$\left(\frac{1}{3}, -\frac{2}{3}\right)$		1		20
2		$\left(-\frac{1}{3}, \frac{2}{3}\right)$		1		—20
2		$\left(\frac{2}{3}, -\frac{1}{3}\right)$		1		7
2		$\left(-\frac{1}{3}, -\frac{1}{3}\right)$		1		7

Wir erhalten daher

Satz 5. Für $h > h_0$ erhält man Extremalbasen zu $g_h(3)$ wie folgt:

Man wähle

$$\gamma = \left\lfloor \frac{2}{3}h \right\rfloor - \left\lfloor \frac{h}{3} \right\rfloor + 2,$$

$$a_2 = \left\lfloor \frac{h}{3} \right\rfloor + 2$$

und setze $a_1 = 1$ und $a_3 = \gamma a_2 - 1$. Es gilt dann

$$g_h(3) = n_h(\mathfrak{A}) = (h + 4 - \gamma - a_2)a_3 + (\gamma - 2)a_2 + (a_2 - 2)a_1.$$

Zusatz 7. Im Fall $h \equiv 0 \pmod{3}$ kann unter sonst gleichen Bedingungen auch

$$a_2 = \frac{h}{3} + 1$$

gewählt werden (vgl. [3], Fußnote 7), und dies sind dann alle Extremalbasen zu $g_h(3)$.

Zusatz 8. Für $h > h_0$ gilt

$$(3) \quad \left(\frac{h+5}{3}\right)^3 - 2\frac{h+5}{3} - 1 - \frac{h+5}{9} + \frac{7}{27} \leq g_h(3) \leq \left(\frac{h+5}{3}\right)^3 - 2\frac{h+5}{3} - 1,$$

und in (3) gilt links das Gleichheitszeichen für $h \equiv 0 \pmod{3}$, rechts für $h \equiv 1 \pmod{3}$.

Literatur

- [1] G. Hofmeister, Über eine Menge von Abschnittsbasen. Journ. f. die reine u. angew. Math. **213** (1964), 43—57.
- [2] G. R. Hofmeister, Zu einem Problem von Frobenius. Det Kgl Norske Vidensk. Selsk. Skr. 1966 Nr. 5, 37 S.
- [3] H. Rohrbach, Ein Beitrag zur additiven Zahlentheorie. Math. Zeitschr. **42** (1937), 1—30.
- [4] H. Salé, Reichweite von Mengen aus drei natürlichen Zahlen. Math. Annalen **165** (1966), 196—203.
- [5] A. Stöhr, Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe. I. Journ. f. die reine u. angew. Math. **194** (1955), 40—65.

Mathematisches Institut der Universität, 65 Mainz

Eingegangen 22. Oktober 1967

Injective hulls in the category of distributive lattices

By *B. Banaschewski* and *G. Bruns* at Hamilton/Ont.

To H. Rohrbach on his 65th Birthday

Introduction

This paper deals primarily with the relationship between the notions of injectivity and essential extension in the category D of all distributive lattices and their lattice homomorphisms, the main point being that in D the situation regarding these notions is fully analogous to what holds in the category of all left modules over a ring [4] or, for that matter, in the category of all partially ordered sets and order preserving mappings [2] (Propositions 2 and 3).

An interesting aspect of the results presented here seems to us the manner in which *Boolean* lattices make their appearance. Thus, it is proved that every distributive lattice has an essential extension which is Boolean (Proposition 1), and the injective hulls in D are shown to be the MacNeille completions of such Boolean essential extensions (Proposition 3).

The paper concludes with a discussion of injective hulls and essential extension of some particular types of distributive lattices and a result, already announced without proof in [2], which shows that essential extensions behave very badly in the category of *all* lattices and lattice homomorphisms¹).

1. Preliminaries

In the following, D will be the category of all distributive lattices and their lattice homomorphisms. General categorical notions will be used as in [6] and lattice theoretic concepts as in [3]. The complete atomic Boolean lattices will be called *power set lattices*.

One readily sees that the monomorphisms of D are precisely the one-to-one homomorphisms, and that any monomorphism $f: L \rightarrow M$ in D is actually an *embedding*, i. e., f decomposes into an isomorphism from L to the (set theoretically, not categorically defined) sublattice $\text{Im } (f)$ of M followed by the natural homomorphism $\text{Im } (f) \rightarrow M$ given by the set theoretic natural injection. A monomorphism $f: L \rightarrow M$ is called *essential* if any $g: M \rightarrow N$ for which $g \cdot f$ is a monomorphism is itself a monomorphism.

¹) The main results of this paper were presented at the meeting on universal algebra held at Oberwolfach, Germany, in July 1966. Financial support from the National Research Council of Canada on this occasion is gratefully acknowledged.

A lattice E is, as usual, called an *extension* of a lattice L iff the elements of L are elements of E , and the meet and join operations of L are the restrictions of the corresponding operations of E . E is called an *essential extension* of L iff the natural embedding $L \rightarrow E$ is essential. More explicitly, this means that any homomorphism $f: E \rightarrow M$ whose restriction to L is one-to-one is itself one-to-one. For two extensions E and F of a lattice L a homomorphism $f: E \rightarrow F$ over L is a homomorphism mapping L identically.

The following result concerning the existence of certain kinds of essential extensions will be of importance later on.

Proposition 1. *Every distributive lattice has an essential extension which is Boolean.*

Proof. If the distributive lattice L has only one element it is actually a (degenerate) Boolean lattice, and nothing has to be proved. Assume, therefore, that L has more than one element. Now, let B be any Boolean lattice containing L as a sublattice such that the zero and unit of B are the infimum and supremum, respectively, of L , and L generates B as a Boolean lattice. Such B do exist, for if E is any complete Boolean lattice containing L as a sublattice then the interval J in E from $\bigwedge L$ to $\bigvee L$ is again a Boolean lattice with L as sublattice, and the Boolean sublattice generated by L in J is of the required type. Note that for any such extension B of L , if L itself has a zero or a unit this will be the zero 0 or the unit e , respectively, of B , and in general $0 \neq e$ since L has more than one element. We now show that B is an essential extension of L .

Let $f: B \rightarrow L'$ be a lattice homomorphism, L' a distributive lattice, such that $f|L$ is one-to-one. Then, the sublattice $f(B)$ of L' , image under f , of B , is Boolean with $f(0)$ as its zero and $f(e)$ as its unit, and $\neg f(x) = f(\neg x)$. Now, each element of B is of the form $b = \bigvee x_i \wedge (\neg y_i)$ with finitely many x_i, y_i from $L^* = L \cup \{0, e\}$, and if $f(b) = 0$ then $f(x_i) \leq f(y_i)$ and thus $f(x_i) = f(x_i \wedge y_i)$ for each i . From this one obtains $x_i = x_i \wedge y_i$, and thus $x_i \leq y_i$, if $x_i, y_i \in L$, and the same clearly holds if $x_i = 0$ or $y_i = e$. On the other hand, if $x_i \in L$ and $y_i = 0$, for some i , then $f(x_i) = f(x_i \wedge x)$ and hence $x_i \leq x$ for all $x \in L$ which shows that also $x_i = 0$, and therefore again $x_i \leq y_i$. Analogously, if $x_i = e$ and $y_i \in L$, it follows that also $y_i = e$ and hence $x_i \leq y_i$. Finally, the case $x_i = e$ and $y_i = 0$ is excluded because it would imply $f(x) = f(0)$ for all $x \in L$ violating that $f|L$ is one-to-one. This shows that $x_i \leq y_i$ in all cases, and thus one obtains $x_i \wedge (\neg y_i) = 0$ for all i , hence $b = 0$. It follows that f is one-to-one.

A Boolean extension B of a distributive lattice L of the type considered in the above proof will be called a Boolean lattice *generated by* L . It will be shown later that any two such extensions of L are isomorphic over L .

2. Injective distributive lattices

The aim of this section is to characterize the injective objects in the category D . We begin by establishing the existence of non-trivial injectives in D .

Lemma 1. *The two-element chains are injective in D .*

Proof. Let C be such a chain, 0 its zero and e its unit, and consider any lattice homomorphism $f: L \rightarrow C$ where L is a sublattice of a distributive lattice E . If f is constant then clearly it has an extension to E . Otherwise, $A = f^{-1}(0)$ is an ideal of L and $F = f^{-1}(e)$ a filter in L such that $A \cap F = \emptyset$. Now, let B be the ideal of E generated by A , and G the filter in E generated by F ; evidently, $B = \{x \mid x \in E, x \leq a \text{ for some } a \in A\}$ and $G = \{x \mid x \in E, x \geq a \text{ for some } a \in F\}$, and therefore $c \in B \cap G$ implies that $c \leq a$ for some $a \in A$ as well as $b \leq c$ for some $b \in F$ from which one derives that $b \in A \cap F$, a

contradiction. It follows that, also, $B \cap G = \emptyset$, and by a well-known theorem [7] there then exists a prime ideal P of E such that $P \supseteq B$ and $P \cap G = \emptyset$. Now, define $g: E \rightarrow C$ by

$$g(x) = \begin{cases} 0 & \text{if } x \in P \\ e & \text{if } x \notin P. \end{cases}$$

This is a lattice homomorphism [3], and since $P \cap L = A$ one has that g extends f .

Even one non-trivial injective, together with the fact that D has products, provides a large supply of injectives since products of families of injectives are injective; in fact, the power set lattices are exactly the products of families of two-element chains, and therefore one has:

Corollary 1. *The power set lattices are injective in D .*

Moreover, since every distributive lattice has a representation by sets, i. e., is isomorphic to a sublattice of a power set lattice, one obtains:

Corollary 2. *Every object in D can be embedded in an injective one.*

Next one has:

Corollary 3. *Every complete Boolean lattice is injective in D .*

Proof. If B is such a lattice and E a power set lattice containing B as a sublattice then by [5] there exists a Boolean homomorphism $f: E \rightarrow B$ mapping B identically. Now, E is injective in D by Corollary 1, and f clearly belongs to D ; this makes B a retract in D of an injective in D , and hence B is itself injective.

Finally, we obtain the following uniqueness result in relation to Lemma 1:

Corollary 4. *Any two Boolean lattices generated by a distributive lattice L are isomorphic over L .*

Proof. Let A and B be the Boolean lattices in question and let C and D be MacNeille completions of A and B respectively. By Corollary 3 there exist lattice homomorphisms $f: C \rightarrow D$ and $g: D \rightarrow C$, each mapping L identically. Then f and g preserve zero and unit and, hence, are Boolean homomorphisms. Now consider the set S of all those elements $x \in A$ which satisfy $f(x) \in B$ and $g(f(x)) = x$. Since S is a Boolean sublattice of A containing L it is identical with A . It follows that f maps A into B and $g \circ f$ maps A identically. By symmetry g maps B into A and $f \circ g$ maps B identically. This proves that $f|A$ is an isomorphism of A onto B .

We now turn to the characterization of injectivity in D .

Proposition 2. *The following are equivalent for $L \in D$:*

- (1) *L is complete and Boolean.*
- (2) *L is injective.*
- (3) *L is a retract of every extension.*
- (4) *L has no proper essential extensions.*

Proof. (1) $\Rightarrow$ (2) is given by the above Corollary 3, and (2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (4) are simple consequences of the definitions. The only step requiring more detailed consideration is (4) $\Rightarrow$ (1). For this, let B be a Boolean lattice generated by L and C a MacNeille completion of B . Then $L < C$ iff L is not itself complete and Boolean. We show that C is an essential extension of L . Let $f: C \rightarrow M$ be any lattice homomorphism into an $M \in D$ whose restriction to L is one-to-one. Then, $f|B$ is one-to-one since B is an essential extension of L by Lemma 1. Moreover, the sublattice N of M given by the image of C

under f is Boolean, and f determines a Boolean homomorphism $C \rightarrow N$. Now, C is an essential extension of B in the category of Boolean lattices and Boolean homomorphisms [2] and hence f is one-to-one on all of C .

Remark. The equivalence of (1) and (2) was also announced in [1].

3. Injective hulls of distributive lattices

We now go into a closer study of those particular injective extensions for any $L \in D$ which were used in the proof of Proposition 1. These turn out to have in several respects, exactly the same properties as the injective hulls in the category of all left modules over a given ring.

Proposition 3. *The following are equivalent for an extension E of a distributive lattice L :*

- (1) *E is a MacNeille completion of a Boolean lattice generated by L .*
- (2) *E is an injective essential extension of L .*
- (3) *E is an injective extension of L not containing any properly smaller such extension of L .*
- (4) *E is an essential extension of L not contained in any properly larger such extension of L .*

Proof. (1) $\Rightarrow$ (2). That E is an essential extension of L is contained in the proof of Proposition 1, and its injectivity is given by Proposition 2.

(2) $\Rightarrow$ (3). If E' is an injective extension of L contained in an essential injective extension E of L then there exists a lattice homomorphism $f: E \rightarrow E'$ which maps E' identically and which, since $E' \cong L$, maps E one-to-one. This implies that $E' = E$.

(3) $\Rightarrow$ (4). An essential extension E' of L which contains an extension E of L is also an essential extension of E . Hence if E is injective then $E' = E$ by Proposition 2. It thus remains to be shown that E is an essential extension of L if it contains no properly smaller injective extensions of L . For this, let C be a MacNeille completion of a Boolean lattice B generated by L . Then, by the injectivity of E there exists a lattice homomorphism $f: C \rightarrow E$ which maps L identically, and by the essentialness of C over L f must map C one-to-one. Hence $f(C)$ is isomorphic to C , thus injective, and therefore $E = f(C)$. It follows that E is an essential extension of L .

(4) $\Rightarrow$ (1). Consider a MacNeille completion C of a Boolean lattice generated by E . By the transitivity of essential extensions, C is then an essential extension of L and hence $C = E$.

In view of Proposition 3, it is appropriate to call the MacNeille completions of the Boolean lattices generated by a distributive lattice L the *injective hulls* of L .

For these one now has:

Corollary 1. *The injective hulls of a distributive lattice L are unique up to a unique isomorphism over L .*

Proof. Let B be a Boolean lattice generated by L and E a MacNeille completion of B . Then, consider any lattice automorphism $f: E \rightarrow E$ mapping L identically. Clearly, f is a Boolean homomorphism, and therefore it maps B identically since B consists of the elements of the type $\bigvee (x_i \wedge (\neg y_i))$ with finitely many x_i, y_i from $L \cup \{0, e\}$. Now, since every element $c \in E$ is the join of all $b \in B$ with $b \leq c$ and since $b \leq c$ holds iff $b = f(b) \leq f(c)$ it follows that $f(c) = c$ for every $c \in E$, i. e. f is the identity map.

Now consider two injective hulls, E and F , of L . The proof that (3) $\Rightarrow$ (4) then shows there exists an isomorphism $f: E \rightarrow L$ over L , and for any other such isomorphism $g: E \rightarrow L$, $g^{-1} \circ f$ is the identity mapping on E , as was just proved; therefore $f = g$.

It might be observed that the counterpart of this corollary usually does not hold in categories of modules: For instance, for a cyclic group of prime order p , the injective hulls in the category of abelian groups are the groups of type p^∞ , and their endomorphism rings are isomorphic to the ring of p -adic integers in such a way that every p -adic integer congruent one modulo p determines an automorphism which maps the original cyclic group identically.

A further property of injective hulls in D is given in:

Corollary 2. *An injective hull E of a distributive lattice L can be embedded over L into any injective extension of L , and any essential extension of L can be embedded over L into E .*

Proof. If $F \geq L$ is injective then the natural injection $L \rightarrow F$ extends to a homomorphism $E \rightarrow F$ which must be one-to-one since E is an essential extension of L . Similarly, if F is any essential extension of L then the natural injection $L \rightarrow E$ extends to a homomorphism $F \rightarrow E$ which is one-to-one since F is an essential extension of E .

4. Particular classes of distributive lattices

In the following we discuss a few special types of distributive lattices in terms of essential extensions and injective hulls in the category D .

Let L be a distributive lattice, let B be a Boolean lattice generated by L and let a and b be elements of L such that $a < b$. Then it is easy to see that the element $b \wedge (-a)$ is an atom of B iff the interval $[a, b]$ is a jump in L , i. e., if there are no elements $x \in L$ such that $a < x < b$. Two jumps $[a, b], [c, d]$ determine the same atom iff $b \wedge (-a) = d \wedge (-c)$ which can be expressed in terms of L by the equations

$$(1) \quad a \vee d = b \vee c \text{ and } a \wedge d = b \wedge c.$$

Furthermore, for every atom p in B there exists a jump $[a, b]$ in L such that $p = b \wedge (-a)$. To prove this let us note first that there exist elements $u, v \in L$ with $p \not\leq u$ and $p \leq v$. This follows from the fact that $p \leq u$ for all $u \in L$ would imply $p = 0$ and that $p \not\leq v$, i. e., $v \wedge p = 0$ for all $v \in L$ would imply $v \leq -p$, that is $-p = e$ and, again, $p = 0$. Since p is an atom its general form $p = \bigvee (x_i \wedge (-y_i))$ reduces to $p = x \wedge (-y)$ with $x, y \in L \cap \{0, e\}$. Clearly $x \neq 0$ and $y \neq e$. It follows

$$p = (v \wedge x) \wedge ((-y) \wedge (-u)) = (v \wedge x) \wedge (-(u \vee y))$$

where $v \wedge x$ and $u \vee y$ belong to L . Put $a = u \vee y$, $b = (u \vee y) \vee (v \wedge x)$. Then $a < b$ and $b \wedge (-a) = p$; hence $[a, b]$ is a jump which determines the atom p .

It follows from these considerations that the atoms in B are in a one-one correspondence with equivalence classes of jumps in L defined by the equivalence relation (1). It is now easy to prove the following:

Proposition 4. *The injective hull of a distributive lattice L is a power set lattice iff every proper interval in L contains a jump.*

Proof. Let E be an injective hull of L . Then E is a MacNeille completion of a Boolean lattice B generated by L . Since every element of E is a join of elements of B

the atoms of E coincide with the atoms of B . It follows that E is atomic and hence a power set lattice iff B is atomic.

Assume now first that B is atomic and consider a proper interval $[a, b]$ in L . Then there exists an atom $p \in B$ with $p \not\leq a$ and $p \leq b$ and by the above considerations there exists a jump $[u, v]$ in L with $p = v \wedge (-u)$. The interval $[a \vee (b \wedge u), a \vee (b \wedge v)]$ is then contained in $[a, b]$. But a simple calculation shows $(a \vee (v \wedge b)) \wedge (- (a \vee (u \wedge b))) = p$, which proves that the interval is a jump.

Assume next that every proper interval in L contains a jump and consider an element $b \neq 0$ in B . Then $x \wedge (-y) \leq b$ for suitable $x, y \in L \cup \{0, e\}$ with $y < x$. By assumption there exists a jump $[u, v]$ in $[y, x]$. It follows that $p = v \wedge (-u)$ is an atom contained in b , which proves the proposition.

Any two different jumps in a chain are not equivalent under the equivalence relation (1). Since there are well ordered sets with an arbitrary number of jumps and since a power set lattice is completely determined by the number of its atoms, one has the following:

Corollary. *The power set lattices are exactly the injective hulls of well ordered sets.*

The distributive lattices which are, in a sense, most unlike Boolean lattices are the chains; nonetheless, essential extension bridges the gap between the two, and in particular the above corollary exhibits a significant class of Boolean lattices as injective hulls of chains. The wider question whether every Boolean lattice, or more generally every distributive lattice, is an essential extension of a chain has a negative answer. As was pointed out to us by R. Sikorski, every chain in a free Boolean lattice is at most countable. This then implies that a free Boolean lattice generated by more than $2^{\aleph_0}$ elements is not an essential extension of a chain. We do not know, however, whether every complete Boolean lattice is an essential extension of a chain. Another class of essential extensions of chains is given by the following:

Proposition 5. *Every at most countable distributive lattice L is an essential extension of a chain.*

Proof. Assume first that L is finite. Let C be a maximal chain in L and let B be a Boolean lattice generated by L . We show that B is also generated by C . Clearly, the zero 0 and unit e of L belong to C and are, therefore, the meet and join respectively of C in B . For any $c \in C$, $c \neq e$, let c' be the successor of c in C . Clearly

$$e = \bigvee \{c' \vee (-c) \mid c \in C, c \neq e\}.$$

It follows for any $a \in B$: $a = a \wedge e = \bigvee \{a \wedge c' \wedge (-c) \mid c \in C, c \neq e\}$. But since $c' \wedge (-c)$ is an atom of B , the meet $a \wedge c' \wedge (-c)$ is either zero or $c' \wedge (-c)$. Consequently, a is the join of all $c' \wedge (-c)$ for which $a \wedge c' \wedge (-c)$ is different from zero and, hence, is contained in the Boolean sublattice of B generated by C . This shows that C generates B . It now follows directly from Proposition 1 and the uniqueness statement given by Corollary 4 of Proposition 2 that B is an essential extension of C . This implies that L is an essential extension of C , for if $f: L \rightarrow M$ is a homomorphism in D which is one-to-one on C , then for any extension $g: B \rightarrow N$ of f where $N \cong M$ is an injective hull of M , $g \upharpoonright C$, therefore g itself, and hence also $f = g \upharpoonright L$, are all one-to-one.

Now let L be countable. Then $L = \bigcup L_k$ where $L_1, L_2, L_3, \dots$ is an increasing sequence of finite sublattices of L , and a sequence $C_1, C_2, C_3, \dots$ of chains in L can be defined by taking C_1 to be any maximal chain of L_1 and, for each $k > 1$, C_k a maximal chain of L_k containing C_{k-1} . $C = \bigcup C_k$ is then a chain in L and any homomorphism

$f: L \rightarrow M$ in D which is one-to-one on C is one-to-one on C_k , hence one-to-one on L_k by the first part of this proof, and therefore one-to-one on all of L .

To provide an example which is not covered by the preceding discussions we now prove:

Proposition 6. *The complete, atomless Boolean lattice R of all regular open subsets of the closed real unit interval I is an injective hull and hence an essential extension of a chain isomorphic to I .*

Proof. For the chain $C \subseteq R$ consisting of $\emptyset$, the intervals $[0, \alpha[$ for $0 < \alpha < 1$, and I one readily sees that I and C are isomorphic lattices. On the other hand, every finite union of intervals open in I can be obtained by means of the Boolean operations in R from members of C since the complement in R of $[0, \alpha[$ is $]\alpha, 1]$ and finite meets are intersections, and this means that the Boolean lattice generated by C is just the sublattice R_0 of R consisting of the finite unions of intervals open in I . Finally, every $U \in R$ is the union, and hence also the supremum in R , of the $V \subseteq U$ in R_0 , which shows that R is a MacNeille completion of R_0 .

Restriction, of the above argument, to intervals with rational endpoints only proves that R is also an essential extension of the closed *rational* unit interval, and since the presence or absence of endpoints makes no difference here, this can also be formulated as follows:

Corollary. *The injective hulls of any countable chain which is dense in itself are isomorphic to the Boolean lattice of all regular open subsets of the closed real unit interval.*

We now consider another class of distributive lattices, the free ones with finite basis.

Proposition 7. *The injective hulls of a free distributive lattice with n free generators are power set lattices with $2^n - 2$ atoms.*

Proof. Let B be a Boolean lattice with n free generators $x_1, \dots, x_n$ and L the sublattice of B generated by $X = \{x_1, \dots, x_n\}$. Then L is freely generated by X as a distributive lattice, since any mapping $f_0: X \rightarrow M$, $M \in D$, may first be extended to a Boolean homomorphism $g: B \rightarrow B'$, B' a Boolean lattice generated by M , and the restriction $g|L$ then provides the homomorphism $f: L \rightarrow M$ extending f_0 .

Now, let $a = x_1 \wedge \dots \wedge x_n$ and $b = (-x_1) \wedge \dots \wedge (-x_n)$; then a and $-b$ are also the infimum and supremum, respectively, of L in B . To see that the interval $[a, -b]$ of B , considered as a Boolean lattice, is generated by L and hence an injective hull of L one observes in addition that the mapping defined by $\varphi(x) = (a \vee x) \wedge (-b)$, $x \in B$, which maps $[a, -b]$ one-to-one and thus B onto $[a, -b]$, is a Boolean lattice homomorphism and therefore carries the generating set X of B to a generating set $\varphi(X) = X$ of $[a, -b]$. Moreover, the kernel of the Boolean epimorphism $\varphi: B \rightarrow [a, -b]$ is the ideal $[0, a \vee b]$ which has four elements since both a and b are atoms in B . This shows that $[a, -b]$ has $2^{2^n} \cdot 2^{-2}$ elements and therefore $2^n - 2$ atoms.

A combination of the result just obtained with the arguments used in the proof of Proposition 5 leads to:

Corollary. *The maximal chains in a free distributive lattice with n free generators have $2^n - 1$ elements.*

Proof. Since an injective hull E of such a lattice L is also an injective hull of any maximal chain C of L , C must have as many jumps as E has atoms.

5. Essential extensions in the category of all lattices

In order to contrast our results concerning distributive lattices with a very different situation we now consider the category L of all lattices and lattice homomorphisms.

Proposition 8. *Every lattice with at least two elements has arbitrarily large essential extensions.*

Proof. Let L be such a lattice and consider first a minimal extension L^* of L such that L^* has a least element 0 and a greatest element e . Then, L^* is an essential extension of L :

Nothing has to be shown if $L = L^*$. Suppose, then, that $0 \notin L$, and let $f: L^* \rightarrow M$ be a lattice homomorphism such that $f|L$ is one-to-one. If $f(0) = f(a)$ for some non-zero $a \in L^*$ then $a = e$ implies that $f(x) = f(0)$ for all $x \in L$, contradicting the fact that $f|L$ is one-to-one, and $a \neq e$ implies $a \in L$, and then

$$f(a \vee x) = f(a) \vee f(x) = f(0) \vee f(x) = f(0) = f(a)$$

shows that $x \geq a$ for all $x \in L$, contradicting the fact that $0 \notin L$. The case $e \notin L$ is analogous.

Now, let A be an arbitrary set disjoint from L^* , and define a lattice extension E of L as follows: E consists of the elements of L^* together with those of A , and its partial order is defined by setting $0 < a < e$ for all $a \in A$ and each $a \in A$ incomparable with every element distinct from a , 0 , and e . One readily sees that E is a lattice containing L^* as a sub-lattice; in particular, $x \wedge a = 0$ and $x \vee a = e$ for any $a \in A$ and $x \neq 0, a, e$ in E . E is an essential extension of L^* : If $f: E \rightarrow M$ is a lattice homomorphism, $f|L^*$ one-to-one, and for some $a \in A$ $f(a) = f(x)$, $x \neq a$ in E , then $f(0) = f(x \wedge a) = f(x) \wedge f(a) = f(a)$ and $f(e) = f(x \vee a) = f(x) \vee f(a) = f(a)$, hence $f(0) = f(e)$, a contradiction.

Since E is an essential extension of L^* and L^* an essential extension of L one has that E is an essential extension of L , and clearly, E can be made arbitrarily large since A was arbitrary.

References

- [1] R. Balbes, Projective and injective distributive lattices. Notices AMS 13 (1966), 740.
- [2] B. Banaschewski and G. Bruns, Categorical characterization of the MacNeille completion. Arch. der Math. 18 (1967), 369—377.
- [3] G. Birkhoff, Lattice Theory, Amer. Math. Soc. Coll. Publ. XXV (1948).
- [4] B. Eckmann und A. Schopf, Über injektive Moduln. Arch. der Math. 4 (1953), 75—78.
- [5] P. Halmos, Lectures on Boolean Algebras. Princeton 1963.
- [6] B. Mitchell, Category Theory. New York, 1966.
- [7] M. H. Stone, Topological representations of distributive lattices and Brouwerian logics. Casopis pro pestovani matematiky a fysiky, 67 (1937), 1—25.

Department of Mathematics, McMaster University, Hamilton, Ontario, Canada

Eingegangen 13. November 1967

Die Bildräume abgeschlossener Operatoren

Von *Gottfried Köthe* in Frankfurt

Hans Rohrbach zum fünfundsechzigsten Geburtstag

Kato bewies in [5] den folgenden für die Störungstheorie wichtigen Satz: E und F seien Banachräume, A ein abgeschlossener linearer Operator mit Definitionsbereich $D[A]$ in E und mit einem Bildraum $R[A]$ endlicher Codimension in F . Dann ist $R[A]$ abgeschlossen in F .

Goldberg gab in [3], S. 100, einen einfacheren Beweis, der gleichzeitig ein weitergehendes Resultat ergab.

Wir geben hier einen Beweis des Satzes von Kato unter wesentlich allgemeineren Voraussetzungen über E und F (Satz 1) und leiten unter Benutzung von Ergebnissen von Amemiya und Komura [1] ein schärferes Resultat ab, das auch für Banachräume neu zu sein scheint (Satz 2).

Für den Beweis von Satz 1 stützen wir uns auf die folgenden Tatsachen:

(1) Jeder lineare Teilraum endlicher Codimension eines tonnelierten Raumes ist tonneliert (Dieudonné [2]).

(2) Ist E tonneliert und F ein Ptákraum, so ist jede abgeschlossene lineare Abbildung von E in F stetig (Graphensatz von A. und W. Robertson [6], vgl. auch Horváth [4], S. 301).

Satz 1. *E sei ein Ptákraum, F sei tonneliert, A sei eine abgeschlossene lineare Abbildung von $D[A] < E$ in F . Hat der Bildraum $R[A]$ endliche Codimension in F , so ist A offen und $R[A]$ abgeschlossen in F .*

Bemerkung. Wir setzen nicht voraus, daß der Definitionsbereich $D[A]$ dicht in E ist.

Beweis. Ist $N[A]$ der Kern von A , so induziert A eine eindeutige Abbildung $\hat{A}$ von $D[\hat{A}] = D[A]/N[A] < E/N[A]$ in F . Bekanntlich ist $\hat{A}$ ebenfalls abgeschlossen und es ist $R[\hat{A}] = R[A]$. Da $E/N[A]$ als Quotientenraum eines Ptákraumes wieder ein Ptákraum ist und da $R[A]$ nach (1) tonneliert ist, ist die auf $R[A]$ definierte lineare Abbildung $\hat{A}^{-1}$ abgeschlossen und nach [2] stetig. Dann ist aber $\hat{A}$ und damit auch A offen als Abbildung auf $R[A]$, also auch als Abbildung in F .

Schließlich ist $R[A]$ abgeschlossen in F : Sei $\hat{A}\hat{x}_\alpha$, $\hat{x}_\alpha \in D[\hat{A}]$, ein gegen $y \in F$ konvergierendes Cauchysystem in $R[A]$. Da $\hat{A}^{-1}$ stetig ist, ist $\hat{x}_\alpha$ ein Cauchysystem in $E/N[A]$, das wegen der Vollständigkeit des Ptákraumes $E/N[A]$ einen Limes $\hat{x}_0$ besitzt. Aus der Abgeschlossenheit von $\hat{A}^{-1}$ folgt $y \in D[A^{-1}] = R[A]$, also die Abgeschlossenheit von $R[A]$. —

Amemiya und Komura bewiesen in [1] die beiden Sätze:

(3) Ein tonnelierter metrisierbarer Raum ist nicht die Vereinigung von abzählbar vielen absolutkonvexen abgeschlossenen und nirgends dichten Teilmengen.

(4) Ein linearer Teilraum abzählbarer Codimension eines tonnelierten metrisierbaren Raumes ist tonneliert.

H hat abzählbare Codimension in E , wenn es abzählbar viele, modulo H linear unabhängige Elemente x_i , $i = 1, 2, \dots$, in E gibt, so daß jedes Element von E die Summe eines Elementes aus H und einer Linearkombination endlich vieler x_i ist.

Satz 2. *E sei ein Ptákraum, F sei tonneliert und metrisierbar, A sei eine abgeschlossene lineare Abbildung von $D[A] \subset E$ in F . Hat der Bildraum $R[A]$ höchstens abzählbare Codimension in F , so ist A offen und $R[A]$ abgeschlossen in F und von endlicher Codimension.*

Beweis. Der Fall, daß $R[A]$ endliche Codimension besitzt, ist in Satz 1 behandelt worden. Es bleibt zu zeigen, daß $R[A]$ niemals abzählbare Codimension besitzen kann.

Nehmen wir an, daß $R[A]$ abzählbare Codimension hat. Dann ist $R[A]$ nach (4) tonneliert und der Beweis von Satz 1 ergibt genau so, daß $R[A]$ abgeschlossen ist. Dann ist aber $F/R[A]$ von abzählbarer Dimension, tonneliert und metrisierbar, da sich diese beiden Eigenschaften auf Quotientenräume übertragen.

Ein tonnelierter metrisierbarer Raum ist aber nach (3) niemals von abzählbarer Dimension, denn er wäre sonst die Vereinigung der abzählbar vielen absolutkonvexen, abgeschlossenen und nirgendsdichten Mengen M_n , M_n die Menge aller $\sum_1^n \alpha_k x_k$, $\sum_1^n |\alpha_k| \leq n$, wobei die $x_1, x_2, \dots$, eine abzählbare Basis des Raumes bilden.

Literaturverzeichnis

- [1] I. Amemiya und Y. Komura, Über nicht-vollständige Montelräume, erscheint in Math. Annalen.
- [2] J. Dieudonné, Sur les propriétés de permanence de certains espaces vectoriels topologiques, Ann. Soc. Polon. Math. **25** (1952), 50—55.
- [3] S. Goldberg, Unbounded Linear Operators, New York 1966.
- [4] J. Horváth, Topological Vector Spaces and Distributions. I, Massachusetts 1966.
- [5] T. Kato, Perturbation Theory for Nullity, Deficiency and other Quantities of Linear Operators, J. d'Analyse Math. **6** (1958), 273—322.
- [6] A. and W. Robertson, On the Closed Graph Theorem, Proc. Glasgow Math. Assoc. **3** (1956), 9—12.

6 Frankfurt/Main, Parkstraße 14

Eingegangen 13. November 1967

Differenzen von Mengen nichtnegativer ganzer Zahlen

Von *Erich Härtter* in Mainz

Herrn Prof. Dr. H. Rohrbach zum 65. Geburtstag gewidmet

1. Neben der Addition von Mengen wurden in der additiven Zahlentheorie auch andere Kompositionen von Mengen wie z. B. Produkte von Zahlenmengen (Niven [8]), erweiterte Basen (Rohrbach [11]) untersucht (siehe auch Stöhr [12], Härtter [4]). Wir wollen hier Differenzen von Mengen ganzer bzw. nichtnegativer ganzer Zahlen betrachten.

Rédei-Rényi [10] nennen eine Menge $\mathfrak{A}$ ganzer Zahlen (Abschnitts-) Differenzbasis für n , wenn jedes $z \in [1, n]^1$ als $z = a_i - a_j$ ($a_i, a_j \in \mathfrak{A}$) darstellbar ist. A. Brauer [1], Erdős-Gál [2], Leech [7], Haselgrove-Leech [6] und Wichmann [14] untersuchen eingeschränkte (Abschnitts-) Differenzbasen; dabei heißt eine Menge $\mathfrak{A}$ eingeschränkte (Abschnitts-) Differenzbasis für n , wenn $\mathfrak{A} \subseteq [0, n]$ und jedes $z \in [1, n]$ als $z = a_i - a_j$ ($a_i, a_j \in \mathfrak{A}$) darstellbar ist. (Vgl. auch Ostmann [9], S. 177.) Diese Betrachtungen bezogen sich alle auf das Abschnittsproblem.

Stöhr [12] nennt eine Menge $\mathfrak{A}$ nichtnegativer ganzer Zahlen (universelle) Differenzbasis für $\mathfrak{N} = \{1, 2, 3, \dots\}$, wenn jedes $z \in \mathfrak{N}$ als $z = a_i - a_j$ ($a_i, a_j \in \mathfrak{A}$) darstellbar ist²⁾.

2. Wir treffen für die Differenzbildung zweier Mengen $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{N}$ folgende

Definition 1. Unter der Differenz $\mathfrak{A} - \mathfrak{B}$ von $\mathfrak{A}$ und $\mathfrak{B}$ verstehen wir die Menge $\mathfrak{D} = \{a - b \mid a \in \mathfrak{A}; b \in \mathfrak{B}; b \leq a\} \subseteq \mathfrak{N}$.

Es drängt sich die Frage auf, ob für die Subtraktion zweier Mengen ein Analogon zum Mannschen Satz gilt. Daß dies nicht der Fall ist, zeigt folgendes Beispiel: Sei die Menge $\mathfrak{A} = \{0\} \cup \{x \equiv 1 \pmod{m}\}$ ⁴⁾ ($m > 2$), und die Menge $\mathfrak{B}$ sei gleich $\mathfrak{A}$. Dann ist, wie man sofort sieht, $\mathfrak{D} = \mathfrak{A} - \mathfrak{B} = \{x \equiv 0 \pmod{m}\} \cup \{x \equiv 1 \pmod{m}\}$, und die zugehörigen finiten (Schnirelmannschen) Dichten⁵⁾ sind $D(\mathfrak{A}) = \frac{1}{m} = D(\mathfrak{B})$ sowie $D(\mathfrak{D}) = \frac{1}{m-1}$. Wegen $m > 2$ wird $2m - 2 > m$, also

$$D(\mathfrak{A}) + D(\mathfrak{B}) = \frac{2}{m} > \frac{1}{m-1} = D(\mathfrak{D}).$$

¹⁾ Die Menge der ganzen Zahlen aus dem Intervall $a \leq \xi \leq b$ bezeichnen wir mit $[a, b]$; entsprechend bedeutet (a, b) die Menge der ganzen Zahlen aus dem Intervall $a < \xi < b$.

²⁾ Offenbar ist 0 immer darstellbar. — Entsprechend wird man (universelle) Differenzbasen für eine Menge $\mathfrak{M} \subseteq \mathfrak{N}$ definieren.

³⁾ Die Menge der natürlichen Zahlen wird mit $\mathfrak{N}$ bezeichnet, die der nichtnegativen ganzen Zahlen mit $\mathfrak{Z}$.

⁴⁾ Für die Zahlenmenge, die aus den nichtnegativen Vertretern der Restklasse $x \equiv a \pmod{m}$ besteht, schreiben wir kurz $\{x \equiv a \pmod{m}\}$.

⁵⁾ Ist $A(x) = \sum_{\substack{0 \leq a \leq x \\ a \in \mathfrak{A}}} 1$ die Anzahlfunktion der Menge $\mathfrak{A}$, dann ist $D(\mathfrak{A}) = \lim_{x \in \mathfrak{N}} \frac{A(x)}{x}$ die finite Dichte von $\mathfrak{A}$.

Da $\mathfrak{A} \subseteq \mathfrak{D} = \mathfrak{A} - \mathfrak{B}$ ist für $0 \in \mathfrak{B}$, gilt dann natürlich trivialerweise $D(\mathfrak{A}) \leq D(\mathfrak{D})$. Die folgende Aussage über die Subtraktion zweier Mengen benutzt einen modifizierten Dichtebegriff (vgl. Härtter [5]):

Satz 1. Seien $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{J}$; $0 \in \mathfrak{B}$; sei ferner $\lim_{x=1,2,\dots} \frac{A(2x) - A(x)}{x} = \tilde{D}(\mathfrak{A}) = \tilde{\alpha}$,
 $\lim_{x=1,2,\dots} \frac{B(x)}{x} = D(\mathfrak{B}) = \beta$ und $\tilde{\alpha} + \beta > 1$. Dann gilt $\mathfrak{A} - \mathfrak{B} \supseteq \mathfrak{N}$.

Zum Beweis schicken wir zunächst einen Hilfssatz voraus:

Hilfssatz 1. Seien $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{J}$; $0 \in \mathfrak{B}$; $\mathfrak{A} - \mathfrak{B} \not\supseteq \mathfrak{N}$, d. h. es gibt $0 < x \in \overline{\mathfrak{A} - \mathfrak{B}}$. Dann gilt für alle $0 \leq y < x$

$$x - y \geq A(2x - y) - A(x) + B(x - y).$$

Beweis. Aus $0 \leq y \leq x - 1$ folgt $x - y \geq 1$. Es sei $\mathfrak{B} \cap [1, x - y] = \{b_1, \dots, b_r\}$, also $B(x - y) = r$ ($r \geq 0$). Nun betrachten wir die Zahlen $x, x + b_1, \dots, x + b_r$; diese sind $\notin \mathfrak{A}$, da sonst

$$(1) \quad (x + b_e) - b_e = x \in \mathfrak{A} - \mathfrak{B}$$

wäre, aber nach Voraussetzung ist $x \notin \mathfrak{A} - \mathfrak{B}$. Wegen $1 \leq b_e \leq x - y$ haben wir $x < x + b_e \leq 2x - y$, woraus

$$\bar{A}(2x - y) - \bar{A}(x - 1) \geq r + 1$$

und weiter

$$(2x - y) - A(2x - y) - (x - 1) + A(x - 1) \geq r + 1$$

folgt. Daraus wird

$$(2) \quad \begin{aligned} x - y &\geq A(2x - y) - A(x - 1) + r \\ &\geq A(2x - y) - A(x - 1) + B(x - y). \end{aligned}$$

Wegen $x \notin \mathfrak{A} - \mathfrak{B}$ ist auch $x \notin \mathfrak{A}$, d. h. $A(x - 1) = A(x)$, was in (2) eingesetzt die Behauptung liefert.

Beweis von Satz 1. Wäre $\mathfrak{A} - \mathfrak{B} \not\supseteq \mathfrak{N}$, dann gäbe es $0 < x \in \overline{\mathfrak{A} - \mathfrak{B}}$. Wir benutzen Hilfssatz 1 mit $y = 0$ und erhalten dann

$$x \geq A(2x) - A(x) + B(x) \geq \tilde{\alpha}x + \beta x = (\tilde{\alpha} + \beta)x > x,$$

einen Widerspruch.

Eine Aussage über die Dichte der Differenzmenge $\mathfrak{D}$ zweier Mengen $\mathfrak{A}$ und $\mathfrak{B}$ gibt der

Satz 2. Seien $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{J}$; $0 \in \mathfrak{A} \cap \mathfrak{B}$; $\mathfrak{D} = \mathfrak{A} - \mathfrak{B}$; wir setzen $D(\mathfrak{A}) = \alpha$, $D(\mathfrak{B}) = \beta$ und $D(\mathfrak{D}) = \delta$. Ist $\mathfrak{A} = \{a_0 < a_1 < a_2 < \dots < a_k < \dots\}$, so gibt es zu jedem $x \in \mathfrak{N}$ ein N , so daß $a_N \leq x < a_{N+1}$ wird. Dann möge $a_N \geq Rx$ gelten mit einer Konstanten R für alle $x \in \mathfrak{N}$.

Für die Dichte der Differenzmenge gilt dann

$$\delta \geq \alpha\{1 - \beta\} + \beta R.$$

Beweis. Wir setzen $d_k = a_{k+1} - a_k - 1$ ($k = 0, 1, 2, \dots$). Dann enthält die Differenzmenge $\mathfrak{D} = \mathfrak{A} - \mathfrak{B}$ im Intervall (a_k, a_{k+1}) mindestens $B(d_k)$ Differenzen $a_{k+1} - b_j$ ($b_j \in \mathfrak{B}$), die nicht schon in $\mathfrak{A}$ liegen. Also wird

$$\begin{aligned} D(x) &\geq A(x) + \sum_{k=0}^{N-1} B(d_k) \geq A(x) + \beta \sum_{k=0}^{N-1} d_k = A(x) + \beta \sum_{k=0}^{N-1} (a_{k+1} - a_k - 1) \\ &= A(x) + \beta a_N - \beta A(x) \geq A(x) \{1 - \beta\} + \beta R x; \end{aligned}$$

Division durch x und Übergang zum fin für $x = 1, 2, \dots$ liefert die Behauptung.

3. Wie Stöhr [12], S. 127, bemerkt hat, gibt es Differenzbasen $\mathfrak{A}$, für die die Anzahlfunktion $A(x)$ beliebig langsam gegen ∞ geht. Wenn man von einer Differenzbasis nun verlangt, daß ihre Anzahlfunktion in irgendeinem Sinn klein bleiben soll, so wird man geeignete Zusatzforderungen an die Differenzbasis stellen. Wir wollen definieren:

Definition 2. Seien $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{Z}$; unter der eingeschränkten Differenz $\mathfrak{A} \ominus \mathfrak{B}$ von $\mathfrak{A}$ und $\mathfrak{B}$ verstehen wir die Menge $\mathfrak{D} = \left\{ a - b \mid a \in \mathfrak{A}; b \in \mathfrak{B}; b \leq \frac{a}{2} \right\} \subseteq \mathfrak{Z}$.

Für $0 \in \mathfrak{B}$ haben wir $\mathfrak{A} \subseteq \mathfrak{D}$. — Mit Definition 2 ist gleichwertig

$$\mathfrak{D} = \mathfrak{A} \ominus \mathfrak{B} = \{ a - b \mid a \in \mathfrak{A}; b \in \mathfrak{B}; b \leq a - b \}.$$

Wir wollen bemerken, daß Satz 1 auch für die eingeschränkten Differenzen $\mathfrak{A} \ominus \mathfrak{B}$ gilt, denn Hilfssatz 1 gilt auch für $\mathfrak{A} \ominus \mathfrak{B}$, da in (1) $b_e \leq x - y \leq x$ ist. —

Das folgende einfache Beispiel zeigt, daß man aus $D(\mathfrak{A}) + D(\mathfrak{B}) = \alpha + \beta > 1$ nicht $\mathfrak{A} \ominus \mathfrak{B} \geq \mathfrak{N}$ folgern kann: Sei $\mathfrak{A} = [0, t] \cup \{2t + 3, \infty\}$ mit $t > 0$ und $\mathfrak{B} = \mathfrak{Z}$. Dann wird $D(\mathfrak{A}) = \alpha = \frac{t}{2t+2}$ und $D(\mathfrak{B}) = \beta = 1$. Obwohl $\alpha + \beta > 1$ ist, ist

$$t + 1 \notin \mathfrak{A} \ominus \mathfrak{B}.$$

Definition 3. Eine Zahlenmenge $\mathfrak{A} \subseteq \mathfrak{Z}$ heißt eingeschränkte Differenzbasis für eine Menge $\mathfrak{M} \subseteq \mathfrak{Z}$, wenn $\mathfrak{A} \ominus \mathfrak{A} \supseteq \mathfrak{M}$ ist.

Das bedeutet, daß jedes $z \in \mathfrak{M}$ darstellbar ist als $z = a_i - a_j$ ($a_i, a_j \in \mathfrak{A}$) mit $a_j \leq \frac{a_i}{2}$, was sofort $a_j \leq z$ und $a_i \leq 2z$ nach sich zieht.

Offenbar enthält jede eingeschränkte Differenzbasis für $\mathfrak{Z}$ die Elemente 0 und 1, weil für 0 nur die Darstellung $0 = 0 - 0$ und für 1 nur die Darstellungen $1 = 1 - 0$ bzw. $1 = 2 - 1$ möglich sind. Wir geben einige Beispiele für eingeschränkte Differenzbasen für $\mathfrak{Z}$:

- 1.) $\mathfrak{A}_1 = \mathfrak{Z}$;
- 2.) $\mathfrak{A}_2 = \{0\} \cup \{x \equiv 1 \pmod{2}\}$;
- 3.) $\mathfrak{A}_3 = \{1\} \cup \{x \equiv 0 \pmod{2}\}$;
- 4.) $\mathfrak{A}_4 = \{0, 1, 2, 5, 6, 7, 8, 17, 18, \dots, 32, 65, \dots\} = \{0, 1\} \cup \bigcup_{n=0}^{\infty} [2^{2n} + 1, 2^{2n+1}]$.

In den ersten drei Beispielen ist die Basiseigenschaft trivial, im letzten Beispiel sieht man sie so ein: Ist $z \notin \mathfrak{A}_4$, so ist $z \in [2^{2n+1}, 2^{2n+2}]$ mit einem geeigneten n . Dann läßt sich aber z darstellen als $z = a - 2^{2n+1}$ mit $a \in [2^{2n+2} + 1, 2^{2n+3}]$, und es ist $2^{2n+1} < z$.

Als Folgerung aus den Beispielen 2 oder 3 ergibt sich

Satz 3. Die Klasse der eingeschränkten Differenzbasen für $\mathfrak{Z}$ hat die Mächtigkeit des Kontinuums.

Beweis. Jedes $\mathfrak{A}$ mit $\mathfrak{A}_2 \subseteq \mathfrak{A} \subseteq \mathfrak{Z}$ (oder $\mathfrak{A}_3 \subseteq \mathfrak{A} \subseteq \mathfrak{Z}$) ist eingeschränkte Differenzbasis für $\mathfrak{Z}$. Die Klasse dieser Mengen $\mathfrak{A} \subseteq \mathfrak{Z}$ ist kontinuummächtig.

Eine weitere einfache Aussage über eingeschränkte Differenzbasen macht

Satz 4. Jede eingeschränkte Differenzbasis $\mathfrak{A}$ für $\mathfrak{M} = [0, n]$ enthält für alle natürlichen Zahlen $x \leq n$ aus jedem Intervall $[x, 2x]$ mindestens ein Element.

Beweis. Angenommen, für ein $x \leq n$ wäre $\mathfrak{A} \cap [x, 2x] = \emptyset$. Dann könnte aber x nicht dargestellt werden als $x = a_i - a_j$ mit $a_j \leq x$; $a_i, a_j \in \mathfrak{A}$.

Die oben genannten Beispiele von eingeschränkten Differenzbasen hatten positive Dichte. In dem folgenden Satz wollen wir eine eingeschränkte Differenzbasis mit der Dichte 0 konstruieren und genauer zeigen

Satz 5. Es gibt eingeschränkte Differenzbasen $\mathfrak{A}$ für $\mathfrak{Z}$ mit $\frac{A(x)}{\sqrt{x}} < 4$ für alle $x \in \mathfrak{Z}$.

Beweis. Sei $\mathfrak{A}_1$ die Menge aller nichtnegativen ganzen Zahlen, die in der 4-adischen Zifferndarstellung⁶⁾ nur die Ziffern 1 oder 3 haben, also

$$\begin{aligned}\mathfrak{A}_1 &= \left\{ \sum_{v=0}^n \alpha_v 4^v \right\} \text{ mit } \alpha_v = 1 \text{ oder } 3 \\ &= \{1, 3, 5, 7, 13, 15, \dots\};\end{aligned}$$

sei ferner $\mathfrak{A}_2$ die Menge aller nichtnegativen ganzen Zahlen, die in der 4-adischen Zifferndarstellung nur die Ziffern 0 oder 1 haben, also

$$\begin{aligned}\mathfrak{A}_2 &= \left\{ \sum_{v=0}^n \beta_v 4^v \right\} \text{ mit } \beta_v = 0 \text{ oder } 1 \\ &= \{0, 1, 4, 5, \dots\}.\end{aligned}$$

Dann bilden wir $\mathfrak{A} = \mathfrak{A}_1 \cup \mathfrak{A}_2$ und zeigen, daß $\mathfrak{A}$ eingeschränkte Differenzbasis für $\mathfrak{Z}$ ist.

I. Ist $z \in \mathfrak{Z}$, dann stellen wir $z > 0$ im 4-adischen Ziffernsystem dar als

$$z = \sum_{v=0}^n c_v 4^v,$$

wobei $0 \leq c_v \leq 3$ und $c_n \neq 0$ ist und n von z abhängt. Nun schreiben wir die Ziffern als Differenzen und zwar $c_v = 0 = 1 - 1$, $c_v = 1 = 1 - 0$, $c_v = 2 = 3 - 1$, $c_v = 3 = 3 - 0$, so daß also allgemein $c_v = \alpha_v - \beta_v$ wird mit $\alpha_v = 1$ oder 3 und $\beta_v = 0$ oder 1. Somit folgt

$$z = \sum_{v=0}^n (\alpha_v - \beta_v) 4^v = \sum_{v=0}^n \alpha_v 4^v + \sum_{v=0}^n \beta_v 4^v,$$

und dabei ist die erste Summe $\in \mathfrak{A}_1$ und die zweite Summe $\in \mathfrak{A}_2$, also z Differenz $a_i - a_j$ zweier Elemente $\in \mathfrak{A}$.

$z = 0$ hat die triviale Darstellung $z = 0 - 0$.

II. Wir zeigen nun, daß in den so konstruierten Darstellungen $z = a_i - a_j$ gilt $a_j \leq z$. Für $c_n = 3$ haben wir

$$z \geq 3 \cdot 4^n \text{ und } a_j \leq 0 \cdot 4^n + 1 \cdot 4^{n-1} + \dots + 1 = \frac{1}{3} (4^n - 1) < \frac{1}{3} 4^n;$$

für $c_n = 2$ haben wir $z \geq 2 \cdot 4^n$ und $a_j \leq 1 \cdot 4^n + 1 \cdot 4^{n-1} + \dots + 1 = 4^n + \frac{1}{3} (4^n - 1) < \frac{4}{3} 4^n$;

für $c_n = 1$ haben wir $z \geq 4^n$ und $a_j \leq 0 \cdot 4^n + 1 \cdot 4^{n-1} + \dots + 1 = \frac{1}{3} 4^n$. Also in jedem Fall ist $a_j \leq z$.

⁶⁾ Zifferndarstellungen zur Konstruktion von Basen wurden schon von Stöhr [13] benutzt.

III. Es sei $4^{s-1} \leq x < 4^s$. Da $\mathfrak{A}_1 \cap \mathfrak{A}_2 = \left\{ \sum_{v=0}^n 1 \cdot 4^v \right\}$ ist, folgt

$$A(x) \leq A(4^s - 1) = A_1(4^s - 1) + A_2(4^s - 1) - s = 2^s + 2^s - s = 2^{s+1} - s$$

und weiter

$$\frac{A(x)}{\sqrt{x}} \leq \frac{A(4^s - 1)}{\sqrt{4^{s-1}}} = \frac{2^{s+1} - s}{2^{s-1}} = 4 - \frac{s}{2^{s-1}} < 4.$$

4. Minimalbasen kann man nun ganz analog wie bei der Addition von Mengen definieren (vgl. Stöhr [12]). Wir wollen hier folgende Definition betrachten:

Definition 4. Eine eingeschränkte Differenzbasis $\mathfrak{A}$ für $\mathfrak{M}$ heißt Minimalbasis, wenn keine echte Teilmenge $\mathfrak{A}^*$ von $\mathfrak{A}$ eingeschränkte Differenzbasis für $\mathfrak{M}$ ist.

Über die Mächtigkeit der Klasse der Minimalbasen für $\mathfrak{Z}$ gilt folgender

Satz 6. Die Klasse K der Minimalbasen für $\mathfrak{Z}$ ist von der Mächtigkeit des Kontinuums.

Beweis. Wir werden eine Klasse $K' \subseteq K$ von Minimalbasen für $\mathfrak{Z}$ konstruieren und zeigen, daß auch die Klasse K' schon von der Mächtigkeit des Kontinuums ist. Die Konstruktion geht induktiv vor⁷⁾: Wir setzen die Menge $\mathfrak{A}_1 = \{0, 1\}$, und für $\mathfrak{A}_2$ lassen wir die beiden Möglichkeiten $\mathfrak{A}_2 = \{0, 1, 2\}$ und $\mathfrak{A}_2 = \{0, 1, 3\}$ zu.

Sei nun $\mathfrak{A}_n = \{a_0 = 0, a_1 = 1, a_2, \dots, a_k\}$ ($a_k < a_{k+1}$) als Minimalbasis für $[0, n]$ ($n \geq 2$) schon konstruiert. Wenn nun auch $n+1$ darstellbar ist als $n+1 = a_i - a_j$ mit $a_i, a_j \in \mathfrak{A}_n$ und $a_j \leq n+1$, dann setzen wir $\mathfrak{A}_{n+1} = \mathfrak{A}_n$. Andernfalls bestimmen wir den Index r so, daß $a_r < n+1 < a_{r+1}$ ist, und für $a_k < n+1$ nehmen wir $r = k$; dann setzen wir $\mathfrak{A}_{n+1}^* = \mathfrak{A}_n \cup \{a_{k+1}\}$, wobei entweder

$$(I) \quad a_{k+1} = n+1 + a_r$$

oder

$$(II) \quad a_{k+1} = n+1 + a_{r-1}$$

ist. Dabei soll Konstruktionsmöglichkeit (II) nur erlaubt sein, wenn $n+1 > 10$ ($n'+1$) ist, wobei beim Induktionsschritt $n'+1$ zum letztenmal beide Konstruktionsmöglichkeiten bestanden.

Dann ist $n+1$ eindeutig darstellbar als $n+1 = a_{k+1} - a_r$ bzw. $n+1 = a_{k+1} - a_{r-1}$, d. h. a_{k+1} wird zur Darstellung von $n+1$ benötigt, und es ist $a_r < n+1$ nach Konstruktion.

Nun bilden wir — indem wir etwaige überflüssige Elemente streichen — aus $\mathfrak{A}_{n+1}^*$ eine darin enthaltene Minimalbasis $\mathfrak{A}_{n+1}$ für $[0, n+1]$. Im Fall (I) wird $\mathfrak{A}_{n+1} = \mathfrak{A}_{n+1}^*$, denn weil für $r < r$

$$a_{k+1} - a_r = n+1 + a_r - a_r > n+1 + a_r - a_r = n+1$$

ist, kann a_{k+1} die Darstellung von Zahlen $< n+1$ nicht mehr beeinflussen. Im Fall (II) bleibt sicher $a_{k+1} \in \mathfrak{A}_{n+1}$ und $\{a_0, a_1, \dots, a_{k'+1}\} \subset \mathfrak{A}_{n+1}$, wobei $a_{k'+1}$ das Element ist, das beim Induktionsschritt $n'+1$ durch $\mathfrak{A}_{n'+1}^* = \mathfrak{A}_{n'} \cup \{a_{k'+1}\}$ eingeführt wurde. Daß $a_{k+1} \in \mathfrak{A}_{n+1}$ bleibt, ist trivial, denn sonst wäre $n+1$ nicht darstellbar. Daß $\{a_0, a_1, \dots, a_{k'+1}\} \subset \mathfrak{A}_{n+1}$ bleibt, sieht man folgendermaßen: Nach Satz 4 enthält jede so konstruierte eingeschränkte Differenzbasis $\mathfrak{A}_{n+1}^*$ für $[0, n+1]$ im Intervall

⁷⁾ Diese verläuft ähnlich wie die induktive Konstruktion in [3].

$[2(n' + 1), 10(n' + 1)]$ mindestens 2 Elemente. Also ist $a_r > a_{r-1} > n' + 1$. Es ist

$$a_{k'+1} \leq 2(n' + 1);$$

da für $a_k \leq 2(n' + 1)$

$$\begin{aligned} a_{k+1} - a_k &\geq n + 1 + a_{r-1} - a_k > n + 1 + n' + 1 - 2(n' + 1) \\ &= n + 1 - (n' + 1) > n + 1 \end{aligned}$$

wird, kann a_{k+1} die Darstellung von Zahlen $\leq n' + 1$ nicht mehr beeinflussen.

Sei nun $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \dots \subseteq \mathfrak{A}_n \subseteq \dots$ eine Folge von auf diese Weise konstruierten und auseinander hervorgehenden Mengen. Dann bilden wir $\mathfrak{A} = \bigcup_{n=1}^{\infty} \mathfrak{A}_n$; aufgrund der Konstruktion ist $\mathfrak{A}$ Differenzbasis und Minimalbasis für $\mathfrak{Z}$.

Da bei der induktiven Konstruktion unendlich oft die Möglichkeit zwischen zwei Arten der Konstruktion besteht, gibt es also kontinuierlich viele Folgen von solchen auseinander hervorgehenden Mengen und damit auch kontinuierlich viele Minimalbasen.

Literaturverzeichnis

- [1] *A. Brauer*, A problem in additive number theory and its application in electrical engineering. Journ. of the Elisha Mitchell Scientific Soc. **61** (1945), 55—66.
- [2] *P. Erdős* and *I. S. Gál*, On the representation of $1, 2, \dots, N$ by differences. Nederl. Akad. Wetensch. Proc. **51** (1948), 1155—1158; Indagationes Math. **10** (1948), 379—382.
- [3] *E. Härtter*, Ein Beitrag zur Theorie der Minimalbasen. Journ. reine u. angew. Math. **196** (1956), 170—204.
- [4] *E. Härtter*, Eine Verallgemeinerung des Basisbegriffs in der additiven Zahlentheorie. I, II. Journ. reine u. angew. Math. **205** (1960/61), 197—220, **209** (1962), 211—220.
- [5] *E. Härtter*, Über den Dichtebegriff in der additiven Zahlentheorie. Math. Ann. **162** (1965), 27—53.
- [6] *C. B. Haselgrove* and *J. Leech*, Note on restricted difference bases. Journ. London Math. Soc. **32** (1957), 228—231.
- [7] *J. Leech*, On the representation of $1, 2, \dots, n$ by differences. Journ. London Math. Soc. **31** (1956), 160—169.
- [8] *I. Niven*, The asymptotic density of sequences. Bull. Amer. Math. Soc. **57** (1951), 420—434.
- [9] *H. H. Ostmann*, Additive Zahlentheorie. I. Ergeb. Math., Berlin-Göttingen-Heidelberg 1956.
- [10] *L. Rédei* und *A. Rényi*, Über die Darstellung der Zahlen $1, 2, \dots, N$ mittels Differenzen. Mat. Sbornik **66** (1949), 385—389.
- [11] *H. Rohrbach*, Ein Beitrag zur additiven Zahlentheorie. Math. Zeitschr. **42** (1937), 1—30.
- [12] *A. Stöhr*, Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe. I, II. Journ. reine u. angew. Math. **194** (1955), 40—65; 111—140.
- [13] *A. Stöhr*, Eine Basis h -ter Ordnung für die Menge aller natürlichen Zahlen. Math. Zeitschr. **42** (1937), 739—743.
- [14] *B. Wichmann*, A note on restricted difference bases. Journ. London Math. Soc. **38** (1963), 465—466.

Mathematisches Institut der Universität, 65 Mainz

Eingegangen 15. November 1967

Über arithmetische Progressionen der Länge p

Von *Harald Scheid* in Mainz

Herrn Professor Dr. H. Rohrbach zum 65. Geburtstag

1. Mit $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{M}$ werden Teilmengen der Menge $\mathfrak{N} = \{0, 1, 2, 3, \dots\}$ der nicht-negativen ganzen Zahlen bezeichnet. Die Anzahlfunktion von $\mathfrak{M}$ heie $M(x)$, die natrliche Dichte $\delta(\mathfrak{M})$ (sofern diese existiert) und die obere asymptotische Dichte $\bar{\delta}(\mathfrak{M})$. Man nennt $\delta_\alpha(\mathfrak{M}) = \lim_{x \rightarrow \infty} \frac{M(x)}{x^\alpha}$ fr $0 < \alpha \leq 1$ die x^α -Dichte von $\mathfrak{M}$, sofern dieser Grenzwert existiert. Eine Folge $\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{M}_3, \dots$ von Mengen *konvergiert gegen die Menge* $\mathfrak{M}$, wenn fr alle $k \in \mathfrak{N}$ ein $N(k) \in \mathfrak{N}$ derart existiert, da fr alle $n \geq N(k)$ die Abschnitte $\mathfrak{M}_n \cap [0, k]$ und $\mathfrak{M} \cap [0, k]$ gleich sind (vgl. [2] und [3]). „Pr (h)“ steht zur Abkrzung fr „arithmetische Progression der Lnge h “, wobei h eine natrliche Zahl ≥ 3 ist und die Lnge einer arithmetischen Progression die Anzahl ihrer Glieder bedeutet. Die Menge aller Teilmengen von $\mathfrak{N}$, welche keine Pr (h) enthalten, werde mit Γ_h bezeichnet. Ferner sei $\bar{\Gamma}_h$ die Menge aller $\mathfrak{M} \subseteq \mathfrak{N}$ mit der Eigenschaft, da fr jedes $x \notin \mathfrak{M}$ die Menge $\mathfrak{M} \cup \{x\}$ eine Pr (h) enthlt.

Es wird vermutet (vgl. [1]), da fr jedes h alle Mengen aus Γ_h die natrliche Dichte Null besitzen. Dies ist fr $h = 3$ von Roth [4] bewiesen worden. Wre die Vermutung fr beliebig groe h (z. B. fr beliebig groe Primzahlen) bewiesen, so wre damit gezeigt, da jede Menge $\mathfrak{M} \subseteq \mathfrak{N}$ mit $\bar{\delta}(\mathfrak{M}) > 0$ beliebig lange arithmetische Progressionen enthlt.

In der vorliegenden Arbeit wird eine spezielle Menge $\mathfrak{A} \in \Gamma_p$ (p ungerade Primzahl) mit $\delta(\mathfrak{A}) = 0$ und $\delta_\alpha(\mathfrak{A}) = \infty$ fr alle α mit $0 < \alpha < \frac{\log(p-1)}{\log p}$ konstruiert, welche die Eigenschaft hat, da jede $\mathfrak{A}$ echt umfassende Menge eine Pr (p) besitzt, also nicht in Γ_p liegt. Fat man Γ_p bezglich der mengentheoretischen Inklusion als teilweise geordnete Menge auf, so ist also $\mathfrak{A}$ ein maximales Element von Γ_p . Ferner wird gezeigt, da jedes maximale Element $\mathfrak{C}$ von Γ_p Grenzwert einer Folge von Elementen $\mathfrak{C}_i$ von $\bar{\Gamma}_p$ ist, fr welche $\delta(\mathfrak{C}_i) = 0$ gilt.

2. Fr eine ungerade Primzahl p und $n = 1, 2, 3, \dots$ sei

$$\mathfrak{B}_n = \bigcup_{i=1}^n \bigcup_{r=1}^\infty [rp^i - p^{i-1}, rp^i),$$

wobei

$$[a, b) = \{x \in \mathfrak{N} : a \leq x < b\}$$

ist. Die Mengen $\mathfrak{B}_n$ bilden eine konvergente Folge; ihr Grenzwert heie $\mathfrak{B}$. Ist $\mathfrak{A}_n$ die Komplementrmenge von $\mathfrak{B}_n$ in $\mathfrak{N}$, so existiert auch $\mathfrak{A} = \lim_{n \rightarrow \infty} \mathfrak{A}_n$, und $\mathfrak{A}$ ist die Kom-

plementärmenge von $\mathfrak{B}$ in $\mathfrak{N}$. Die Menge $\mathfrak{A}$ besteht aus allen $a \in \mathfrak{N}$, in deren p -adischer Darstellung

$$(1) \quad a = a_i p^i + a_{i-1} p^{i-1} + \cdots + a_1 p + a_0$$

die Ziffer $p - 1$ nicht vorkommt. Denn $\mathfrak{B}_n$ besteht aus genau denjenigen natürlichen Zahlen a , in deren p -adischer Darstellung (1) eine der Ziffern $a_0, a_1, \dots, a_{n-1}$ den Wert $p - 1$ hat.

Satz 1. Die Menge $\mathfrak{A}$ enthält keine $Pr(p)$.

Beweis. Es genügt der Nachweis, daß $\mathfrak{A}_n$ keine $Pr(p)$ zu einem Modul $k < p^n$ enthält; für $n \rightarrow \infty$ ergibt sich daraus die Behauptung des Satzes. Man gehe induktiv vor. Die Menge $\mathfrak{A}_1$ enthält keine $Pr(p)$ zu einem Modul $k < p$, denn die Kongruenz

$$a + xk \equiv -1 \pmod{p}$$

ist für jedes $a \in \mathfrak{N}$ (eindeutig $\pmod{p}$) lösbar, es gibt also ein $x_0 \in \mathfrak{N}$, $x_0 < p$ mit $a + x_0 k \in \mathfrak{B}_1$. Jetzt folgt der Induktionsschluß. Wegen $\mathfrak{A}_n < \mathfrak{A}_{n-1}$ und der Induktionsvoraussetzung enthält $\mathfrak{A}_n$ keine $Pr(p)$ zu einem Modul $k < p^{n-1}$. Sei also $p^{n-1} \leq k < p^n$ und

$$k = cp^{n-1} + d$$

mit $1 \leq c < p$ und $0 \leq d < p^{n-1}$. Ist $d > 0$, dann gilt mit

$$(2) \quad a + xk = a + xcp^{n-1} + xd \in \mathfrak{A}_n$$

auch

$$a + xd \in \mathfrak{A}_{n-1},$$

denn $\mathfrak{A}_n < \mathfrak{A}_{n-1}$ und $\mathfrak{A}_{n-1}$ besteht aus lauter vollständigen Restklassen $\pmod{p^{n-1}}$. Daher kann nach Induktionsvoraussetzung x in (2) höchstens $p - 1$ aufeinanderfolgende Zahlen aus $\mathfrak{N}$ durchlaufen. Ist $d = 0$ und setzt man

$$a = -a' + ep^{n-1}$$

mit $0 < a' \leq p^{n-1}$, dann ist also

$$a + xk = -a' + (e + xc)p^{n-1}.$$

Die Kongruenz

$$e + xc \equiv 0 \pmod{p}$$

ist (eindeutig $\pmod{p}$) lösbar, es gibt also ein $x_0 \in \mathfrak{N}$ $x_0 < p$, und ein $s \in \mathfrak{N}$, $s \neq 0$, so daß $a + x_0 k = -a' + sp^n$ im Intervall $[sp^n - p^{n-1}, sp^n)$ liegt. Es ist dann also $a + x_0 k \in \mathfrak{B}_n$, womit Satz 1 bewiesen ist.

Satz 2. Ist $u \in \mathfrak{B}$, dann enthält $\mathfrak{A} \cup \{u\}$ eine $Pr(p)$.

Beweis. Sei $n > \frac{\log u}{\log p}$, also $u < p^n$, und sei

$$u = a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \cdots + a_1 p + a_0$$

mit

$$0 \leq a_i < p \quad (i = 0, 1, 2, \dots, n-1)$$

die p -adische Darstellung von u . Man setze

$$\varepsilon_i = \begin{cases} 1 & \text{für } a_i = p-1 \\ 0 & \text{für } a_i < p-1 \end{cases} \quad (i = 0, 1, 2, \dots, n-1)$$

und definiere

$$k = \varepsilon_{n-1}p^{n-1} + \varepsilon_{n-2}p^{n-2} + \cdots + \varepsilon_1p + \varepsilon_0.$$

Wegen $u \notin \mathfrak{A}$ ist $k > 0$. Für $r = 0, 1, 2, \dots, p-1$ ist

$$(3) \quad u - rk = \varrho_{n-1}p^{n-1} + \varrho_{n-2}p^{n-2} + \cdots + \varrho_1p + \varrho_0$$

mit

$$\varrho_i = \begin{cases} a_i & \text{für } a_i < p-1 \\ p-1-r & \text{für } a_i = p-1 \end{cases} \quad (i = 0, 1, 2, \dots, n-1).$$

Die p Zahlen in (3) sind alle nichtnegativ, bilden also eine $\text{Pr}(p)$ in $\mathfrak{A}$. Da alle p -adischen Ziffern ϱ_i ($i = 0, 1, 2, \dots, n-1$) von $u - rk$ ($r = 1, 2, \dots, p-1$) von $p-1$ verschieden sind, ist also

$$u - rk \in \mathfrak{A} \quad (r = 1, 2, \dots, p-1).$$

Damit ist Satz 2 bewiesen.

Satz 3. Es gilt $\delta(\mathfrak{A}) = 0$ und $\delta_\alpha(\mathfrak{A}) = \infty$ für alle α mit $0 < \alpha < \frac{\log(p-1)}{\log p}$.

Beweis. Zunächst gilt

$$A(p^n) = (p-1)^n,$$

denn $A(p) = p-1$ und $A(p^n) = (p-1)A(p^{n-1})$. Für

$$m = a_n p^n + a_{n-1} p^{n-1} + \cdots + a_1 p + a_0$$

mit $0 \leq a_i < p$ und $a_n \geq 1$ gilt also

$$\frac{A(m)}{m} \leq \frac{A(p^{n+1})}{p^n} = (p-1) \left(\frac{p-1}{p} \right)^n$$

und

$$\frac{A(m)}{m^\alpha} \geq \frac{A(p^n)}{p^{\alpha(n+1)}} = \frac{1}{p^\alpha} \left(\frac{p-1}{p^\alpha} \right)^n.$$

Daraus folgen unmittelbar die Behauptungen von Satz 3.

Satz 4. Jedes maximale Element $\mathfrak{C}$ von Γ_p ist Grenzwert einer Folge von Elementen $\mathfrak{C}_i$ von $\bar{\Gamma}_p$, deren natürliche Dichten Null sind.

Beweis. Für ein festes $i \in \mathfrak{N}$ sei $y_i^{(1)}$ das kleinste $y \notin \mathfrak{C}$ mit $y \leq p^i$ und $(\mathfrak{C} \cup \{y_i^{(1)}\}) \cap [0, p^i] \in \Gamma_p$, $y_i^{(2)}$ das kleinste $y \notin \mathfrak{C}$ mit $y_i^{(1)} < y_i^{(2)} \leq p^i$ und

$$(\mathfrak{C} \cup \{y_i^{(1)}, y_i^{(2)}\}) \cap [0, p^i] \in \Gamma_p, \dots,$$

$y_i^{(r_i)}$ das kleinste $y \notin \mathfrak{C}$ mit $y_i^{(r_i-1)} < y_i^{(r_i)} \leq p^i$ und $(\mathfrak{C} \cup \{y_i^{(1)}, \dots, y_i^{(r_i)}\}) \cap [0, p^i] \in \Gamma_p$, wobei die Zahl r_i größtmöglich sein soll. Die Menge $\mathfrak{Y}_i = \{y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(r_i)}\}$ kann auch leer sein. Man setze $\mathfrak{S}_i^{(0)} = (\mathfrak{C} \cup \mathfrak{Y}_i) \cap [0, p^i]$, bilde die Schnirelmann-Summen $\mathfrak{S}_i^{(j)} = \mathfrak{S}_i^{(0)} + \{j p^i\}$ ($j = 0, 1, 2, \dots$) und setze $\mathfrak{C}_i = \bigcup_{j \in \mathfrak{N}} \mathfrak{S}_i^{(j)}$, wobei $\mathfrak{N}$ die oben definierte spezielle Menge aus Γ_p ist. Wegen $C_i(p^{i+n}) = (p-1)^n S_i^{(0)}(p^i) \leq (p-1)^n p^i$ erkennt man sofort $\delta(\mathfrak{C}_i) = 0$. Ist $\mathfrak{Y}_i = \emptyset$ für alle $i \geq i_0$, so ist

$$\mathfrak{S}_i^{(0)} = \mathfrak{C} \cap [0, p^i] = \mathfrak{C}_i \cap [0, p^i]$$

für alle $i \geq i_0$, also $\mathfrak{C}_i \xrightarrow{i \rightarrow \infty} \mathfrak{C}$. Gibt es aber eine unendliche Folge $i_1, i_2, \dots$ mit $\mathfrak{Y}_{i_s} \neq \emptyset$ ($s = 1, 2, \dots$), so enthält $\mathfrak{C} \cup \{y_{i_s}^{(1)}\}$ eine $\text{Pr}(p)$ durch $y_{i_s}^{(1)}$, denn $\mathfrak{C}$ soll maximal sein. Das letzte Glied dieser $\text{Pr}(p)$ sei $\leq p^N$. Ist $i_s > N$, dann folgt $y_{i_s}^{(1)} > y_{i_s}^{(1)}$, so daß man mit

$m_i = y_{i_t}^{(1)}$ eine monoton wachsende Folge $m_1 < m_2 < \dots$ erhält, wobei stets $m_i \leq p^i$ ist. Wegen

$$\mathfrak{C}_{i_t} \cap [0, m_i) = \mathfrak{S}_{i_t}^{(0)} \cap [0, m_i) = \mathfrak{C} \cap [0, m_i)$$

folgt aus $i \geq i_t$ die Beziehung

$$\mathfrak{C}_i \cap [0, m_i) = \mathfrak{C} \cap [0, m_i),$$

also wieder $\mathfrak{C}_i \xrightarrow{i \rightarrow \infty} \mathfrak{C}$. Es fehlt noch der Nachweis, daß $\mathfrak{C}_i \in \bar{\Gamma}_p$ gilt. Ist

$$x \notin \mathfrak{C}_i, x \in [jp^i, (j+1)p^i]$$

und $j \in \mathfrak{A}$, dann ist $\mathfrak{C}_i \cup \{x\} \notin \Gamma_p$, denn dann existiert eine $\text{Pr}(p)$ in $[jp^i, (j+1)p^i]$. Sei also $x \notin \mathfrak{C}_i$, $x \in (jp^i, (j+1)p^i]$ mit $j \notin \mathfrak{A}$ und zunächst $u = x - jp^i \in \mathfrak{S}_i^{(0)}$. Da in $\mathfrak{A} \cup \{j\}$ eine $\text{Pr}(p)$: $j_1, j_2, \dots, j_p$ durch j existiert, ist $u + j_1p^i, u + j_2p^i, \dots, u + j_pp^i$ eine $\text{Pr}(p)$ in $\mathfrak{C}_i \cup \{x\}$. Sei nun $u = x - jp^i \notin \mathfrak{S}_i^{(0)}$. In $\mathfrak{S}_i^{(0)} \cup \{u\}$ existiert eine $\text{Pr}(p)$: $u_1, u_2, \dots, u_p$ durch u (es sei $u_m = u$), in $\mathfrak{A} \cup \{j\}$ existiert eine $\text{Pr}(p)$: $j_1, j_2, \dots, j_p$ durch j (es sei $j_n = j$). Dann ist

$$u_\mu + j_\nu p^i \in \mathfrak{C}_i \cup \{x\} \quad (\mu, \nu = 1, 2, \dots, p).$$

Nimmt man o. B. d. A. nun $m \leq n$ an, dann liegt in $\mathfrak{C}_i \cup (x)$ die $\text{Pr}(p)$

$$u_{n-m+1} + j_1p^i, u_{n-m+2} + j_2p^i, \dots, u_{n-m+p} + j_pp^i.$$

Damit ist Satz 4 bewiesen.

3. Bemerkungen Nur in Satz 1 wurde benutzt, daß p eine Primzahl ist. Die Aussage „ $\delta(\mathfrak{A}) = 0$ “ aus Satz 2 läßt sich als die bekannte Tatsache interpretieren, daß fast alle (d. h. alle bis auf eine Ausnahmемenge der natürlichen Dichte Null) natürlichen Zahlen in ihrer p -adischen Darstellung die Ziffer $p-1$ enthalten. Dies gilt selbstverständlich auch für jede andere p -adische Ziffer, so daß man allgemeiner sagen kann, daß fast alle natürlichen Zahlen in ihrer p -adischen Darstellung alle p -adischen Ziffern enthalten.

Literatur

- [1] H.-H. Ostmann, Additive Zahlentheorie. II, Berlin 1956, S. 98f.
- [2] H.-H. Ostmann, Untersuchungen über den Summenbegriff in der additiven Zahlentheorie, Math. Ann. **120** (1948), 165–196.
- [3] H. Rohrbach und B. Volkmann, Zur Konvergenz von Mengenfolgen, Math. Ann. **124** (1951/52), 298–302.
- [4] K. Roth, Sur quelques ensembles d'entiers, C. R. Acad. Paris **234** (1952), 388–390.

Mathematisches Institut der Universität, 65 Mainz

Eingegangen 15. November 1967

Zur Berechnung der Mahlerschen Funktionen w_n

Von *Rainer Güting* in Kampala (*Uganda*)

Herrn Professor Dr. H. Rohrbach zum 65. Geburtstag gewidmet

1. Einleitung

Es sei $P(x) = a_p x^p + \dots + a_0$ ($a_p \neq 0$) ein Polynom mit beliebigen komplexen Koeffizienten vom Grad p , der Höhe $h(P) = \max |a_i|$, und der Größe $s(P) = \sum |a_i|$. Es sei $\mathfrak{P}(n, h)$ die Menge aller Polynome mit ganzrationalen Koeffizienten vom Grad $\leq n$ und der Höhe $\leq h$. Mahler [4] (siehe auch [6]) hat die folgenden Funktionen eingeführt. Ist γ eine beliebige reelle oder komplexe Zahl, so sei

$$\begin{aligned}w_n(h, \gamma) &= \min_{\substack{P \in \mathfrak{P}(n, h) \\ P(\gamma) \neq 0}} |P(\gamma)|, \\w_n(\gamma) &= \overline{\lim}_{h \rightarrow \infty} \frac{\log w_n(h, \gamma)}{\log h}, \\w(\gamma) &= \overline{\lim}_{n \rightarrow \infty} \frac{w_n(\gamma)}{n}.\end{aligned}$$

Gibt es einen Index n , für den $w_n = \infty$ ist, dann sei $\mu(\gamma)$ der kleinste solche Index; andernfalls sei $\mu(\gamma) = \infty$. Mahler nennt eine Zahl γ

eine A -Zahl, falls $w = 0, \mu = \infty$,
eine S -Zahl, falls $0 < w < \infty, \mu = \infty$,
eine T -Zahl, falls $w = \infty, \mu = \infty$,
eine U -Zahl, falls $w = \infty, \mu < \infty$ ist.

Ein Hauptproblem dieser Einteilung der Zahlen in vier Klassen besteht darin, für ein gegebenes reelles oder nicht reelles γ den Wert der Funktion w_n zu bestimmen. Für reelles γ und $n = 1$ wurde diese Aufgabe in einer früheren Arbeit des Verfassers [1] mit Hilfe der Kettenbruchentwicklung von γ gelöst: Ist $[a_0; a_1, a_2, \dots]$ die regelmäßige Kettenbruchentwicklung der Zahl γ und p_n/q_n der n -te Näherungsbruch von γ , so ist

$$w_1(\gamma) = 1 + \overline{\lim}_{n \rightarrow \infty} \frac{\log a_n}{\log q_{n-1}},$$

wenn der Grenzwert auf der rechten Seite endlich ist. Andernfalls ist $w_1(\gamma) = \infty$.

Die vorliegende Arbeit behandelt nun für reelles γ den Fall $n \geq 2$. Ist

$$\gamma^n = \sum_{i=1}^{\infty} 2^{-pi}$$

die dyadische Entwicklung von γ^n , so gilt stets die Abschätzung

$$w_n(\gamma) \geq \overline{\lim}_{i \rightarrow \infty} \frac{p_{i+1}}{p_i} - 1.$$

Existiert ferner der Grenzwert $\lim_{i \rightarrow \infty} \frac{p_{i+1}}{p_i}$ und ist er genügend groß, so wird auch eine obere Schranke für $w_n(\gamma)$ angegeben [s. Satz 4]. Ähnliche Ergebnisse folgen, falls die regelmäßige Kettenbruchentwicklung der Zahl γ^n bekannt ist (vgl. Sätze 7 und 8). Auch Aussagen über gewisse Klassen nicht-reeller Zahlen sind möglich. Als Beispiel dazu dient Satz 6. Ein zweites Problem betrifft die Beziehung zwischen $w_n(\gamma)$ und $w_m(\gamma)$ für $n < m$. Außer der trivialen Beziehung $w_n(\gamma) \leq w_m(\gamma)$ ist auch bekannt, daß es zu jedem natürlichen n Zahlen γ gibt, für die $w_n(\gamma) < \infty$, aber $w_{n+1}(\gamma) = \infty$ ist. Daher ist auch folgender Satz von Interesse: Zu beliebigen natürlichen Zahlen n und m ($n < m$) und zu genügend großem c gibt es stets Zahlen γ , für die $w_n(\gamma) = w_{n+1}(\gamma) = \dots = w_m(\gamma) = c$ ist (Satz 5). Aus dem vorangehenden Resultat ergibt sich, daß jede genügend große reelle Zahl c ($c > n^2 + 3n - 2$) unter den Werten der Funktion w_n vorkommt. Grundlegend ist ein allgemeiner Satz über Approximationen mit Hilfe von Polynomen des Grades $\leq p$, der für reelles und nicht reelles γ gilt. Sätze über Polynome mit mehrfachen Nullstellen (siehe [2]) werden angewandt, um ohne die einschränkende Bedingung auszukommen, daß die Approximationspolynome irreduzibel sind. Als weitere Anwendung wird noch eine hinreichende Bedingung dafür angegeben, daß eine Zahl T -Zahl ist. Schließlich werden noch für gewisse U_l -Zahlen von LeVeque [3] obere Schranken für w_n ($n < l$) gegeben.

2. Hilfsbetrachtungen

Hilfssatz 1. *Es seien P und R beliebige Polynome von den Graden p und r . Ferner sei α eine Nullstelle der Ordnung k von P . Dann gilt*

$$(1) \quad |R(\alpha)| \geq \frac{|\text{Res}(P, R)|^{\omega/k} \max(1, |\alpha|)^r}{2^{\omega pr/k} s(P)^{\omega r/k} s(R)^{-1 + \omega p/k}}$$

mit $\omega = 1$. Die Ungleichung (1) gilt auch mit $\omega = 1/2$, wenn P und R reelle Koeffizienten haben und α nicht reell ist.

Beweis. Siehe [2], Theorem 1'.

Aus Hilfssatz 1 folgt sogleich

Hilfssatz 2. *Unter den Voraussetzungen von Hilfssatz 1 gilt für Polynome P und R mit ganzzahligen Koeffizienten und ohne gemeinsame Nullstellen die Abschätzung*

$$(2) \quad |R(\alpha)| \geq c_1 h(P)^{-\omega r/k} h(R)^{1 - (\omega p/k)}$$

mit einer von den Höhen $h(P)$ und $h(R)$ unabhängigen Konstanten c_1 .

Beweis. Da P und R keine gemeinsamen Nullstellen haben, verschwindet die Resultante der beiden Polynome nicht. Sie ist ganzzahlig, da P und R ganzzahlige Koeffizienten haben, also mindestens vom Betrage 1.

Hilfssatz 3. Es sei γ eine beliebige komplexe Zahl. Ferner sei P ein Polynom mit ganzzahligen Koeffizienten. Ist dann α eine Nullstelle von P mit minimaler Entfernung von γ und k die Vielfachheit von α , so gilt

$$(3) \quad |\gamma - \alpha|^k \leq c_2(p, k) |\operatorname{Im} \alpha|^{-k/2} \max(1, |\alpha|)^{-p+3k} s(P)^{(p-3k)/2k} |P(\gamma)|,$$

wenn α nicht reell ist; und für beliebiges α gilt

$$(4) \quad |\gamma - \alpha|^k \leq c_3(p, k) \max(1, |\alpha|)^{-p+2k} s(P)^{(p-2k)/k} |P(\gamma)|$$

mit nur von p und k abhängigen Konstanten c_2 und c_3 .

Beweis. Siehe [2], Korollar zu Theorem 4'.

Hilfssatz 4. Die Voraussetzungen seien die gleichen wie in Hilfssatz 3, jedoch sei P ein Polynom mit genügend großer Höhe, das die Ungleichung

$$(5) \quad |P(\gamma)| \leq h(P)^{-a}$$

($a > p - 2$) erfüllt. Dann gilt mit einer nur von γ und p abhängigen Konstanten $c_4(\gamma, p)$ die Ungleichung

$$(6) \quad |\gamma - \alpha| \leq c_4(\gamma, p) h(P)^{[\omega p - \omega - 1 - a]/k}$$

mit $\omega = 1$, falls γ reell, und mit $\omega = 1/2$, falls γ nicht reell ist.

Beweis. Wir wenden Hilfssatz 3 an. Aus (4) und (5) erhalten wir zunächst die Ungleichung

$$|\gamma - \alpha|^k \leq c_3(p, k) \max(1, |\alpha|)^{-p+2k} s(P)^{p/k} s(P)^{-2} h(P)^{-a}.$$

Wegen $p \geq k$ und $h(P) \leq s(P) \leq (p+1)h(P)$ folgt daraus

$$\begin{aligned} |\gamma - \alpha|^k &\leq c_3(p, k) \max(1, |\alpha|)^k [(p+1)h(P)]^{p/k} h(P)^{-2-a} \\ &\leq c_3(p, k) (p+1)^{p/k} \max(1, |\alpha|)^k h(P)^{p-2-a}. \end{aligned}$$

Es ergibt sich also mit einer nur von p abhängigen Konstanten c_5 (man benutze nötigenfalls die Beziehung $1 \leq k \leq p$)

$$(7) \quad |\gamma - \alpha| \leq c_5(p) \max(1, |\alpha|) h(P)^{(p-2-a)/k}.$$

Wegen $a > p - 2$ können wir $h(P)$ als so groß voraussetzen, daß $c_5(p) h(P)^{(p-2-a)/p} < 1/2$ ist. Dann gilt auch $c_5(p) h(P)^{(p-2-a)/k} < 1/2$.

Dann folgt aus (7) die Abschätzung $|\alpha| \leq \max\left(\frac{1}{2} + |\gamma|, 2|\gamma|\right)$. Es gilt also mit einer nur von γ und p abhängigen Konstanten c_6 die Ungleichung

$$(8) \quad |\gamma - \alpha| \leq c_6 h(P)^{(p-2-a)/k}.$$

Ist γ reell, so ist damit die Behauptung bewiesen. Ist γ nicht reell, so ist für genügend großes $h(P)$ wegen (8) sicher $|\gamma - \alpha| \leq \frac{|\operatorname{Im} \gamma|}{2}$, woraus folgt, daß auch α nicht reell ist. Wir können also auch die Abschätzung (3) von Hilfssatz 3 anwenden. Die Behauptung folgt dann ohne weiteres unter Benutzung der Ungleichungen $|\operatorname{Im} \alpha| \geq \frac{|\operatorname{Im} \gamma|}{2}$, $|\alpha| \leq 2|\gamma|$, $1 \leq 2k \leq p$ und (5).

Hilfssatz 5. *Es seien Q und R Polynome mit ganzzahligen Koeffizienten und ohne gemeinsame Nullstellen, und es sei $Q(\alpha) = R(\beta) = 0$. Dann gilt*

$$\begin{aligned} |\alpha - \beta| &\geq \frac{|\operatorname{Res}(Q, R)|^\omega \max(1, |\alpha|) \max(1, |\beta|)}{\min(q, r) s(Q)^{\omega r} s(R)^{\omega q}} \\ &\geq \frac{1}{\min(q, r) [(q+1)h(Q)]^{\omega r} [(r+1)h(R)]^{\omega q}}. \end{aligned}$$

Beweis. Siehe [2], Theorem 6.

Hilfssatz 6. *Ist $PQ = R$, so gilt für die Höhen*

$$h(P)h(Q) \leq c_7(r)h(R),$$

wobei die Konstante $c_7(r)$ nur vom Grad r des Polynoms R abhängt.

Beweis. Die verlangte Ungleichung ist eine direkte Konsequenz von [5], (7).

3. Hauptteil

Wir betrachten eine beliebige komplexe Zahl γ . Ist γ reell, so setzen wir $\omega = 1$; ist aber γ nicht reell, so setzen wir $\omega = 1/2$. Es sei p eine beliebige natürliche Zahl. Ferner seien a und b positive Zahlen mit

$$(9) \quad b > a > 3\omega p - 1 - \omega.$$

Wir sagen, daß γ eine T_p -Folge hat, wenn es eine Folge $\{P_j\}$ von Polynomen vom Grade $\leq p$ und mit den Höhen h_j gibt, die die Ungleichungen

$$(10) \quad h_j^{-b} \leq |P_j(\gamma)| \leq h_j^{-a},$$

$$(11) \quad h_{j-1} < h_j \leq h_{j-1}^{b+1}$$

befriedigen. Wir zeigen folgenden

Satz 1. *Hat γ eine T_p -Folge, so gilt für jedes natürliche n mit $p \leq n < \frac{a+1}{\omega} - 2p + 1$ die Abschätzung*

$$(12) \quad a \leq w_n(\gamma) \leq \frac{(b+1)^2 \omega p}{a+1+\omega-\omega p-\omega n} - 1.$$

Beweis. Nach (10) gilt für unendlich viele j

$$|P_j(\gamma)| \leq h_j^{-a}.$$

Demnach ist für unendlich viele h

$$w_n(h, \gamma) \leq h^{-a},$$

und daraus folgt $w_n(\gamma) \geq a$ für jedes $n \geq p$. Um die rechte Seite von (12) zu beweisen, benutzen wir die Tatsache, daß für jede Zahl γ und für jedes n die Gleichung $\tilde{w}_n(\gamma) = w_n(\gamma)$ gilt, wobei $\tilde{w}_n$ wie w_n definiert ist, jedoch unter Beschränkung auf irreduzible Polynome ([7], Remark 2. Man beachte, daß der Beweis auch für algebraische Zahlen gültig ist). Es genügt also, $\tilde{w}_n(\gamma)$ abzuschätzen. Es sei nun P ein beliebiges irreduzibles Polynom

des Grades $\leq n$ und von genügend großer Höhe h . Es sei j festgelegt durch die Ungleichung

$$(13) \quad h_{j-1} \leq h < h_j.$$

Ist nun P ein Teiler von P_{j-1} , so gilt

$$(14) \quad |P(\gamma)| |Q(\gamma)| = |P_{j-1}(\gamma)| \geq h_{j-1}^{-b} \geq h^{-b}$$

für ein gewisses Polynom Q mit ganzrationalen Koeffizienten. Bezeichnen wir die Höhe von Q mit g , so gilt nach Hilfssatz 6 unter Benutzung von (13)

$$hg \leq c_7(p) h_{j-1} \leq c_7(p) h.$$

Also ist $g \leq c_7(p)$. Daraus folgt die Abschätzung

$$|Q(\gamma)| \leq gq \max(1, |\gamma|)^q \leq c_7(p) p \max(1, |\gamma|)^p,$$

und wegen (14) gilt dann mit einer von h unabhängigen Konstanten c_8

$$(15) \quad |P(\gamma)| \geq c_8 h^{-b}.$$

Ist aber P kein Teiler von P_{j-1} , so bezeichnen wir mit α_{j-1} eine Nullstelle von P_{j-1} mit minimalem Abstand von γ , und es sei k die Ordnung von α_{j-1} als Nullstelle von P_{j-1} . Wegen

$$(16) \quad a > 2\omega p + \omega n - 1 - \omega > p - 2$$

sind die Voraussetzungen von Hilfssatz 4 erfüllt. Demnach ist

$$(17) \quad |\gamma - \alpha_{j-1}| \leq c_4 h_{j-1}^{(\omega p - \omega - 1 - a)/k}.$$

Wir können also schreiben

$$\gamma = \alpha_{j-1} + t h_{j-1}^{(\omega p - \omega - 1 - a)/k} \text{ mit } |t| \leq c_4.$$

Da P irreduzibel und kein Teiler von P_{j-1} ist, so haben P und P_{j-1} keine gemeinsamen Nullstellen. Daher folgt aus Hilfssatz 2 die Abschätzung

$$(18) \quad |P(\alpha_{j-1})| \geq c_1 h_{j-1}^{-\omega n/k} h^{1-\omega p/k}$$

mit einer von h_{j-1} und h unabhängigen Konstanten c_1 . Es gilt somit

$$(19) \quad \begin{aligned} |P(\gamma)| &\geq |P(\alpha_{j-1})| - c_9 h h_{j-1}^{(\omega p - \omega - 1 - a)/k} \\ &\geq c_1 h_{j-1}^{-\omega n/k} h^{1-\omega p/k} - c_9 h h_{j-1}^{(\omega p - \omega - 1 - a)/k}, \\ |P(\gamma)| &\geq h_{j-1}^{-\omega n/k} h (c_1 h^{-\omega p/k} - c_9 h_{j-1}^{(\omega p + \omega n - \omega - 1 - a)/k}) \end{aligned}$$

mit von h_{j-1} und h unabhängigen Konstanten c_1 und c_9 . Ist jetzt

$$h^{\omega p} < (c_1/2c_9) h_{j-1}^{a+1+\omega-\omega p-\omega n},$$

so folgt aus (13) und (19)

$$(20) \quad |P(\gamma)| \geq \frac{c_1}{2} h_{j-1}^{-\omega n/k} h^{1-\omega p/k} \geq \frac{c_1}{2} h^{1-(\omega n + \omega p)/k} \geq \frac{c_1}{2} h^{1-\omega n-\omega p}.$$

Ist aber

$$(21) \quad h^{\omega p} \geq \frac{c_1}{2c_3} h_{j-1}^{a+1+\omega-\omega p-\omega n},$$

so betrachten wir das Polynom P_j . Ist dann P ein Teiler von P_j , so gilt für ein gewisses Polynom R mit ganzrationalen Koeffizienten die Abschätzung

$$|P(\gamma)| |R(\gamma)| = |P_j(\gamma)| \geq h_j^{-b}.$$

Nach Hilfssatz 6 gilt für die Höhen

$$h \cdot h(R) \leq c_7(p) h_j.$$

Es gilt also

$$|R(\gamma)| \leq p \frac{c_7(p) h_j}{h} \max(1, |\gamma|)^p.$$

Demnach finden wir

$$|P(\gamma)| \geq h_j^{-b-1} h c_{10},$$

wobei die Konstante c_{10} nur von p und γ abhängt. Aus den Ungleichungen (11) und (21) bekommt man dann

$$(22) \quad |P(\gamma)| \geq c_{11} h^{1-(b+1)\omega p/(a+1+\omega-\omega p-\omega n)}.$$

Ist aber P kein Teiler von P_j , so bezeichnen wir mit α_j wieder eine Nullstelle von P_j mit minimalem Abstand von γ und mit k' die Ordnung von α_j . Es gelten dann die (17) und (18) entsprechenden Ungleichungen (man ersetze $j-1$ durch j und k durch k'), und so erhält man statt (19) die Ungleichung

$$|P(\gamma)| \geq h_j^{-\omega n/k'} h (c_1 h^{-\omega p/k'} - c_9 h_j^{(\omega p + \omega n - \omega - 1 - a)/k'}).$$

Unter Berücksichtigung von (13) und (16) finden wir jetzt

$$|P(\gamma)| \geq h_j^{-\omega n/k'} h (c_1 h^{-\omega p/k'} - c_9 h^{(\omega p + \omega n - \omega - 1 - a)/k'}).$$

Ebenfalls wegen (16) ist $\omega p + \omega n - \omega - 1 - a < -\omega p$. Wir können also h als so groß voraussetzen, daß

$$|P(\gamma)| \geq \frac{c_1}{2} h_j^{-\omega n/k'} h^{1-\omega p/k'}$$

ist. Benutzen wir nun die Ungleichungen (11) und (21), so haben wir

$$(23) \quad \begin{aligned} |P(\gamma)| &\geq \frac{c_1}{2} h_{j-1}^{-\omega n(b+1)/k'} h^{1-\omega p/k'} \\ &\geq c_{12} h^{1-(\omega p/k')-\omega^2 p n(b+1)/k'(a+1+\omega-\omega p-\omega n)} \\ &\geq c_{12} h^{1-\omega p-\omega^2 p n(b+1)/(a+1+\omega-\omega p-\omega n)}. \end{aligned}$$

Die Konstante c_{12} ist dabei unabhängig von den Höhen h_{j-1} , h_j und h . Faßt man nun (15), (20), (22) und (23) zusammen, so sieht man: Für jedes irreduzible Polynom P von einem Grad $\leq n$ und von genügend großer Höhe h gilt mit einer von h unabhängigen Konstanten c_{13} :

$$(24) \quad |P(\gamma)| \geq c_{13} h^{-A}$$

mit $A = \max \left[b, \omega n + \omega p - 1, -1 + \frac{(b+1)^2 \omega p}{a+1+\omega-\omega p-\omega n}, \omega p - 1 + \frac{\omega^2 p n(b+1)}{a+1+\omega-\omega p-\omega n} \right].$

Nun ist jedoch

$$(25) \quad \omega n + \omega p - 1 < a < b$$

wegen (16). Ferner ist auch leicht zu sehen, daß b kleiner ist als das dritte Glied in der eckigen Klammer. Aber auch das vierte Glied in der eckigen Klammer läßt sich durch das dritte Glied abschätzen; denn es gilt wegen $p \geq \frac{1}{2} + \frac{1}{2\omega}$, (9) und (16) die Ungleichung $\omega n < a < b$ und daraus folgt

$$a + 1 + \omega - \omega p - \omega n + \omega n(b + 1) < (b + 1)(1 + \omega n) < (b + 1)^2.$$

Also ist

$$A = \frac{(b + 1)^2 \omega p}{a + 1 + \omega - \omega p - \omega n} - 1.$$

Somit gilt für genügend großes h stets

$$\tilde{w}_n(h, \gamma) \geq c_{13} h^{-4},$$

und daraus folgt wegen $\tilde{w}_n(\gamma) = w_n(\gamma)$ die rechte Ungleichung von (12), was zu beweisen war.

Satz 2. *Hat γ eine T_p -Folge von irreduziblen Polynomen, deren Koeffizienten keinen gemeinsamen Faktor haben, so besteht für jedes natürliche n mit $p \leq n < \frac{a+1}{\omega} - 2p + 1$ die Ungleichung*

$$a \leq w_n(\gamma) \leq \max \left[b, \omega p - 1 + \frac{\omega^2 p n (b + 1)}{a + 1 + \omega - \omega p - \omega n} \right].$$

Der Beweis verläuft ähnlich wie der Beweis des Satzes 1 (mit $k = k' = 1$), jedoch läßt sich die Abschätzung (22) vermeiden: Ist P ein Teiler des Polynoms P_j , so folgt, da jetzt beide Polynome irreduzibel und ihre Koeffizienten ganzzahlig sind, daß P_j ein rationales Vielfaches von P ist. Weil ferner nach Voraussetzung die Koeffizienten von P_j keinen gemeinsamen Faktor haben, so muß sogar P ein ganzes Vielfaches von P_j sein, was aber wegen der Bedingung $h < h_j$ unmöglich ist. Man findet also die Ungleichung (24) mit

$$A = \max \left[b, \omega n + \omega p - 1, \omega p - 1 + \frac{\omega^2 p n (b + 1)}{a + 1 + \omega - \omega p - \omega n} \right],$$

und unter Benutzung von (25) folgt daraus die Behauptung.

Korollar. *Es seien die Voraussetzungen von Satz 2 erfüllt. Ferner sei*

$$a \geq \omega^2 p^2 + 3\omega p - 1 - \omega.$$

Dann gilt für jedes n mit $p \leq n \leq (a + 1 + \omega - 2\omega p)/(\omega^2 p + \omega)$ die Abschätzung

$$a \leq w_n(\gamma) \leq b.$$

Beweis. Wir können wie folgt abschätzen:

$$\begin{aligned} \omega p - 1 + \frac{\omega^2 p n (b + 1)}{a + 1 + \omega - \omega p - \omega n} &\leq \omega p - 1 + \frac{\omega^2 p n (b + 1)}{\omega^2 p n + \omega p} \\ &= \frac{\omega^2 p n + \omega p - 1 + \omega n b}{\omega n + 1} < b, \end{aligned}$$

wobei die letzte Ungleichung aus $\omega^2 p n + \omega p - 1 \leq \omega^2 p n + 2\omega p + \omega n - 1 - \omega \leq a < b$ folgt. Also ist das Maximum in Satz 2 gleich b .

Satz 3. *Es sei p eine natürliche Zahl und γ eine reelle Zahl im Einheitsintervall. Ferner sei*

$$\gamma^p = \sum_{i=1}^{\infty} 2^{-p_i}$$

die dyadische Entwicklung von γ^p . Dann gilt für $n \geq p$ die Abschätzung

$$w_n(\gamma) \geq \overline{\lim}_{i \rightarrow \infty} \frac{p_{i+1}}{p_i} - 1.$$

Beweis. Wir setzen $\overline{\lim}_{i \rightarrow \infty} \frac{p_{i+1}}{p_i} = d$. Dann gibt es zu jedem $\varepsilon > 0$ unendlich viele i , für die $p_{i+1} > p_i(d - \varepsilon)$ ist. Für diese i betrachten wir dann die Polynome

$$(26) \quad P_i(x) = 2^{p_i} x^p - 2^{p_i} \sum_{j=1}^i 2^{-p_j}.$$

P_i hat die Höhe $h_i = 2^{p_i}$, und es ist

$$|P_i(\gamma)| = 2^{p_i} \sum_{j=i+1}^{\infty} 2^{-p_j} \leq 2^{1+p_i-p_{i+1}} \leq 2^{p_i(1-d+\frac{1}{p_i}+\varepsilon)} = h_i^{(1-d+\varepsilon+\frac{\log 2}{\log h_i})}.$$

Es gibt also zu jedem $\varepsilon > 0$ unendlich viele Polynome des Grades p derart, daß

$$|P(\gamma)| \leq h^{-(d-1-\varepsilon-\log 2/\log h)}, \quad h = h(P),$$

ist, und daraus folgt $w_n(\gamma) \geq w_p(\gamma) \geq d - 1$, wie behauptet.

Satz 4. *Die Voraussetzungen seien dieselben wie in Satz 3. Existiert ferner der Grenzwert $d = \lim_{i \rightarrow \infty} \frac{p_{i+1}}{p_i}$ und ist $d > 3p - 1$, so gilt für jedes natürliche n mit $p \leq n < d - 2p + 1$ die Beziehung*

$$w_n(\gamma) \leq \frac{d^2 p}{d + 1 - p - n} - 1.$$

Beweis. Es sei ε so gewählt, daß $0 < \varepsilon < \frac{1}{2}(d + 1 - 3p)$ ist. Aus dem Beweis des Satzes 3 ergibt sich, daß die Polynome P_i in (26) für alle genügend großen i die Ungleichung

$$|P_i(\gamma)| \leq h_i^{-(d-1-2\varepsilon)}$$

erfüllen. Weiter ist $p_{i+1} \leq p_i(d + \varepsilon)$ für alle genügend großen i . Also ist

$$|P_i(\gamma)| \geq 2^{p_i-p_{i+1}} \geq 2^{p_i(1-d-\varepsilon)} = h_i^{-(d-1+\varepsilon)}.$$

Schließlich gilt $h_{i+1} = 2^{p_{i+1}} \leq 2^{p_i(d+\varepsilon)} = h_i^{d+\varepsilon}$. Also hat die Zahl γ für jedes $\varepsilon > 0$ eine T_p -Folge mit $a = d - 1 - 2\varepsilon$ und $b = d - 1 + \varepsilon$. Das Ergebnis folgt nun aus Satz 1.

Wir können nun folgenden Existenzsatz beweisen:

Satz 5. *Es seien p und q ($p \leq q$) beliebige natürliche Zahlen. Ferner sei c eine genügend große reelle Zahl. Dann gibt es reelle Zahlen γ , für die*

$$w_p(\gamma) = w_{p+1}(\gamma) = \cdots = w_q(\gamma) = c$$

ist. Insbesondere nimmt die Funktion w_p bei gegebenem p alle Werte $c > p^2 + 3p - 2$ an.

Beweis. Es sei $c > pq + q + 2p - 2$. Man wähle zunächst p_1 beliebig, aber relativ prim zu p . Ist weiter p_i für $i \geq 1$ schon gefunden, so sei p_{i+1} irgendeine zu p teilerfremde natürliche Zahl im Intervall $[(c+1)p_i, (c+1)p_i + p]$. Dann hat die Zahl $\gamma = \left(\sum_{i=1}^{\infty} 2^{-p_i} \right)^{1/p}$ die verlangte Eigenschaft. Um das zu zeigen, wollen wir das Korollar zu Satz 2 anwenden. Wir benutzen dazu für $i = 1, 2, \dots$ die Polynome (26). Aus der Konstruktion der Zahlen p_i ergibt sich zunächst die Beziehung

$$d = \lim_{i \rightarrow \infty} \frac{p_{i+1}}{p_i} = c + 1.$$

Wählt man ε so, daß $0 < \varepsilon < \frac{1}{2}(c - pq - 2p - q + 2)$ ist, so sieht man wie im Beweis von Satz 4, daß die Polynome (26) für $i > i_0(\varepsilon)$ eine T_p -Folge für γ bilden, und zwar mit $a = c - 2\varepsilon$ und $b = c + \varepsilon$. Ferner folgt aus der Verallgemeinerung des Eisensteinschen Irreduzibilitätskriteriums von G. Dumas unter Verwendung der Voraussetzung $(p, p_i) = 1$ die Irreduzibilität der Polynome (26). Offensichtlich haben die Koeffizienten der P_i auch keine gemeinsamen Faktoren, und schließlich ist

$$a = c - 2\varepsilon > pq + q + 2p - 2 \geq p^2 + 3p - 2.$$

Das Korollar zu Satz 2 ergibt dann für jedes n mit $p \leq n \leq \frac{c + 2 - 2p}{p + 1}$ die Gleichung $w_n(\gamma) = c$, und da auf Grund der Wahl von q und von c die Ungleichung $p \leq n \leq q$ gilt, so ist damit die erste Behauptung bewiesen. Die zweite Behauptung folgt einfach daraus, daß im Fall $q = p$ die Bedingung für c sich zu $c > p^2 + 3p - 2$ spezialisiert.

Als Anwendung der Sätze 1 und 2 auf nicht reelle Zahlen läßt sich mit denselben Methoden leicht zeigen der

Satz 6. *Ist m eine natürliche Zahl und $\gamma = \left(- \sum_{j=1}^{\infty} 2^{-p_j} \right)^{1/2m}$, existiert der Grenzwert $d = \lim_{i \rightarrow \infty} \frac{p_{i+1}}{p_i}$ und ist $d > 3m - \frac{1}{2}$, so gilt für alle n mit $2m \leq n < 2d - 4m + 1$ die Abschätzung*

$$d - 1 \leq w_n(\gamma) \leq \frac{2md^2}{2d + 1 - 2m - n} - 1.$$

Ist überdies $d > m^2 + 3m - \frac{1}{2}$ und ist $(p, p_i) = 1$ für alle i , so gilt für alle n mit $2m \leq n \leq \frac{2d - 4m + 1}{m + 1}$ die Gleichung

$$w_n(\gamma) = d - 1.$$

Satz 7. Es sei p eine natürliche Zahl. Es sei $[0; a_1, a_2, \dots]$ die regelmäßige Kettenbruchentwicklung der Zahl γ^p ($\gamma < 1$) und q_i der i -te Näherungsnenner von γ^p ; dann ist für $n \geq p$

$$w_n(\gamma) \geq 1 + \overline{\lim}_{i \rightarrow \infty} \frac{\log a_{i+1}}{\log q_i}.$$

Beweis. Es sei $D = \overline{\lim}_{i \rightarrow \infty} \frac{\log a_{i+1}}{\log q_i}$. Dann gibt es zu jedem $\varepsilon > 0$ unendlich viele i , für die die Ungleichung $a_{i+1} > q_i^{D-\varepsilon}$ gilt. Nach der Theorie der Kettenbrüche gilt dann für unendlich viele Näherungsbrüche r_i/q_i die Abschätzung

$$\left| \gamma^p - \frac{r_i}{q_i} \right| < \frac{1}{a_{i+1} q_i^2} < \frac{1}{q_i^{D+2-\varepsilon}}.$$

Also ist $|q_i \gamma^p - r_i| < q_i^{-D-1+\varepsilon}$ für solche i . Wegen $\gamma < 1$ ist $q_i < r_i$. Es gibt also unendlich viele Polynome P des Grades p , für die

$$|P(\gamma)| < h(P)^{-D-1+\varepsilon}$$

ist, und daraus folgt $w_p(\gamma) \geq D + 1$.

Satz 8. Es seien die Voraussetzungen von Satz 7 erfüllt. Existiert ferner der Grenzwert $D = \lim_{i \rightarrow \infty} \frac{\log a_{i+1}}{\log q_i}$ und ist $D > 3p - 3$, so gilt für alle n mit $p \leq n < D - 2p + 3$ die Abschätzung

$$D + 1 \leq w_n(\gamma) \leq \frac{p(D+2)^2}{D - p - n + 3} - 1.$$

Sind ferner die Polynome $q_i x^p - r_i$ irreduzibel, ist $D \geq p^2 + 3p + 3$ und ist $Q = \left\lfloor \frac{D - 2p - 3}{p + 1} \right\rfloor$, so besteht die Gleichung

$$w_p(\gamma) = w_{p+1}(\gamma) = \dots = w_Q(\gamma) = D + 1.$$

Beweis. Wir betrachten wieder die Näherungsbrüche r_i/q_i von γ^p und bilden die Polynome $P_i(x) = q_i x^p - r_i$ mit den Höhen $h_i = h(P_i) = q_i$. Entsprechend der Definition von D gibt es zu jedem $\varepsilon > 0$ eine Zahl $I > 0$, so daß für alle $i > I$ die Ungleichung $a_{i+1} \leq q_i^{D+\varepsilon}$ befriedigt ist. Aus der Kettenbruchtheorie folgt dann für jedes $i > I$ die Beziehung

$$|P_i(\gamma)| \geq \frac{1}{q_i + q_{i+1}} \geq \frac{1}{(a_{i+1} + 2)q_i} \geq h_i^{-D-1-2\varepsilon}.$$

Wie im Beweis von Satz 7 ergibt sich ferner unter Benutzung der Existenz des Grenzwertes $\lim_{i \rightarrow \infty} \frac{\log a_{i+1}}{\log q_i}$ die Abschätzung

$$|P_i(\gamma)| < h_i^{-D-1+\varepsilon}$$

für alle genügend großen i . Ferner gilt für die Höhen

$$h_i = a_i h_{i-1} + h_{i-2} \leq h_{i-1}^{D+2+2\varepsilon}.$$

Es gibt also zu jedem $\varepsilon > 0$ eine T_p -Folge mit $a = D + 1 - \varepsilon$ und $b = D + 1 + 2\varepsilon$. Und schließlich ist $(p_i, q_i) = 1$ für alle $i \geq 1$. Aus Satz 1 und dem Korollar zu Satz 2 folgt jetzt die Behauptung. Man beachte, daß die Voraussetzung der Irreduzibilität der Polynome $q_i x^p - r_i$ im Fall $p = 1$ trivialerweise erfüllt ist.

4. Anwendungen auf T - und U -Zahlen

Wir wollen uns nun dem Fall zuwenden, daß γ für verschiedene Zahlen p T_p -Folgen hat. Dann gibt es zu jeder solchen Zahl p Zahlen a und b , wie sie in (9) bis (11) auftreten. Wir bezeichnen sie mit a_p und b_p . Wir sagen, daß γ eine T -Folge hat, wenn folgende Bedingungen erfüllt sind: 1. Es gebe unendlich viele p , für die die Zahl γ T_p -Folgen hat.

2. Für die Zahlen a_p , die dabei auftreten, gelte die Beziehung $\overline{\lim}_{p \rightarrow \infty} \frac{a_p}{p} = \infty$.

Satz 9. *Hat γ eine T -Folge, so ist γ eine T -Zahl.*

Beweis. Auf Grund von Satz 1 gilt für unendlich viele p die Abschätzung: $w_p \geq a_p$. Ferner gilt $w(\gamma) = \overline{\lim}_{p \rightarrow \infty} \frac{w_p}{p} \geq \overline{\lim}_{p \rightarrow \infty} \frac{a_p}{p} = \infty$. Also ist γ weder A - noch S -Zahl. Andererseits können wir auch zeigen, daß für jedes r die Funktion $w_r(\gamma) < \infty$ ist. Denn da die Zahl γ für unendlich viele p eine T_p -Folge hat, gibt es ein $p > r$, zu dem es eine T_p -Folge für γ gibt. Es seien a_p und b_p die entsprechenden in den Voraussetzungen von Satz 1 auftretenden Konstanten. Dann gilt, wiederum nach Satz 1 (mit $n = p$), daß w_p endlich ist. Wegen $w_r \leq w_p$ ist aber auch w_r endlich. Also ist γ auch keine U -Zahl.

In einer Arbeit von LeVeque ([3], Theorem 4) werden Zahlen γ betrachtet, zu denen es eine Folge $P_i(x)$ von Polynomen des Grades m_i und der Höhe h_i gibt mit den folgenden Eigenschaften:

1. $h_1 \leq h_2 \leq \dots$
2. $|P_i(\gamma)| = h_i^{-\sigma_i}$, $\lim_{i \rightarrow \infty} \sigma_i = \infty$.
3. Alle P_i sind irreduzibel.
4. Es gibt ein m , so daß $m_i \leq m$ für alle i ist.
5. Es gibt ein $\delta > 0$, so daß $h_{i+1} < h_i^{\sigma_i/\delta}$ für alle i ist.

Es folgt dann (s. [3]) für $l = \underline{\lim} m_i$, daß $w_l(\gamma) = \infty$ und $w_n(\gamma) < \infty$ für $n < l$ ist¹⁾. Wir wollen diese Aussage verschärfen, indem wir für $n < l$ obere Schranken für die Zahlen $w_n(\gamma)$ explizit angeben:

Satz 10. *Gibt es zu der Zahl γ eine Folge P_i mit den Eigenschaften 1. bis 5., so gilt für $n < l = \underline{\lim} m_i$ die Abschätzung*

$$w_n(\gamma) \leq \omega m - 1 + \omega^2 m n / \delta$$

mit $\omega = 1$, falls γ reell, und mit $\omega = 1/2$, falls γ nicht reell ist.

Wir zeigen zunächst den

Hilfssatz 7. *Für jede Folge, die die Bedingungen 1. bis 5. erfüllt, gilt die Ungleichung²⁾ $\delta \leq \omega m$.*

Beweis. Nehmen wir an, $\{P_i\}$ sei eine solche Folge. Wir bezeichnen mit α_i eine Nullstelle von P_i mit minimalem Abstand von γ .

¹⁾ Als Beispiel für solche Zahlen gibt LeVeque die Zahlen $\zeta^{1/l}$ an ($l = 1, 2, \dots$), wo ζ von der Form

$$\zeta = \sum_{k=1}^{\infty} 2^{-t_k}$$

ist. Hier sind die Zahlen t_k Primzahlen und so gewählt, daß die Folge $\log t_{k+1} / \log t_k$ über alle Schranken monoton wächst. Für die Zahlen $\zeta^{1/l}$ zeigt dann übrigens auch Satz 3, daß $w_l = \infty$ ist.

²⁾ Für die Zahlen $\zeta^{1/l}$ von LeVeque findet man die Bedingung 5 nur erfüllt, wenn $\delta < 1$ gewählt wird. Da $\zeta^{1/l}$ reell ist, ergibt sich also Übereinstimmung mit dem obigen Hilfssatz.

Auf Grund von Hilfssatz 4 gilt

$$|\gamma - \alpha_i| \leq c_4(\gamma, m) h_i^{\omega m - \omega - 1 - \sigma_i}.$$

Daraus folgt:

$$|\alpha_i - \alpha_{i+1}| \leq |\gamma - \alpha_i| + |\gamma - \alpha_{i+1}| \leq c_4(\gamma, m) (h_i^{\omega m - \omega - 1 - \sigma_i} + h_{i+1}^{\omega m - \omega - 1 - \sigma_{i+1}}).$$

Ferner, da σ_i mit i gegen unendlich strebt, so ist bei genügend großem i

$$(27) \quad |\alpha_i - \alpha_{i+1}| \leq 2c_4(\gamma, m) h_i^{\omega m - 1 - \omega - \sigma_i}.$$

Andererseits gilt aber nach Hilfssatz 5 mit einer nur von m abhängigen Konstanten $c_{14}(m)$

$$|\alpha_i - \alpha_{i+1}| \geq c_{14}(m) h_i^{-\omega m} h_{i+1}^{-\omega m}.$$

Benutzen wir nun Eigenschaft 5, so finden wir

$$|\alpha_i - \alpha_{i+1}| \geq c_{14}(m) h_i^{-\omega m - \omega m \sigma_i / \delta}.$$

Da die h_i auch gegen unendlich streben (es gibt nur endlich viele Polynome des Grades m und der Höhe $\leq h_0$), so liefert die letzte Ungleichung in Verbindung mit (27) die Beziehung

$$-\omega m - \omega m \sigma_i / \delta \leq \omega m - 1 - \omega - \sigma_i,$$

das heißt $\delta \leq \frac{\omega m \sigma_i}{\sigma_i - 2\omega m + 1 + \omega}$.

Da auch σ_i gegen unendlich strebt, kann diese Ungleichung nur bestehen, falls $\delta \leq \omega m$ ist.

Beweis des Satzes 10. Es sei $1 \leq n < l$, und $Q(x)$ sei ein beliebiges irreduzibles Polynom vom Grade $\leq n$ und der Höhe g , die als genügend groß vorausgesetzt wird. Dann läßt sich wegen Eigenschaft 1 der Folge $\{P_i\}$ eine Zahl i so bestimmen, daß

$$(28) \quad h_i \leq g < h_{i+1}$$

ist. Für jedes i sei ferner α_i eine Nullstelle von P_i mit minimalem Abstand von γ . Wir können annehmen, daß $Q(\alpha_i) \neq 0$ ist. Es sind nämlich nur endlich viele Polynome P_i vom Grade $n < l = \lim_{i \rightarrow \infty} m_i$. Ist also h genügend groß, so sind alle P_i mit $h_i > h$ vom Grade $\geq l$, und weil Q und P_i irreduzibel sind, aber von verschiedenen Graden, ist $Q(\alpha_i) \neq 0$. Nach Hilfssatz 2 ist

$$(29) \quad |Q(\alpha_i)| \geq c_1 h_i^{-\omega n} g^{1 - \omega m_i} \geq c_1 h_i^{-\omega n} g^{1 - \omega m},$$

wobei c_1 von g und h_i unabhängig ist. Da $\lim_{i \rightarrow \infty} \sigma_i = \infty$ ist, so folgt bei genügend großem g (was dann auch ein genügend großes i nach sich zieht), daß $\sigma_i > m - 2 \geq m_i - 2$ ist. Nach Hilfssatz 4 ist also ($k = 1$)

$$|\gamma - \alpha_i| \leq c_4(\gamma, m_i) h_i^{\omega m_i - \omega - 1 - \sigma_i}.$$

Schätzt man jetzt m_i durch m ab, so erhält man mit einer neuen Konstanten $c_{15}(\gamma, m)$ die Ungleichung

$$|\gamma - \alpha_i| \leq c_{15}(\gamma, m) h_i^{\omega m - \omega - 1 - \sigma_i},$$

die für jedes genügend große i gültig ist. Wir schließen jetzt genau wie vorher

$$\begin{aligned} |Q(\gamma)| &= |Q(\alpha_i + th_i^{\omega m - \omega - 1 - \sigma_i})|, \quad |t| \leq c_{15}(\gamma, m), \\ |Q(\gamma)| &\geq |Q(\alpha_i)| - c_{16}(\gamma, m) g h_i^{\omega m - \omega - 1 - \sigma_i}. \end{aligned}$$

Das heißt, unter Benutzung von (29) erhalten wir mit von γ und den Höhen h_i unabhängigen Konstanten c_1 und c_{16} die Abschätzung

$$(30) \quad |Q(\gamma)| \geq c_1 h_i^{-\omega n} g^{1-\omega m} - c_{16} g h_i^{\omega m - \omega - 1 - \sigma_i} = h_i^{-\omega n} g (c_1 g^{-\omega m} - c_{16} h_i^{\omega m + \omega n - \omega - 1 - \sigma_i}).$$

Gilt nun $h_i^{\omega m + \omega n - \omega - 1 - \sigma_i} < \frac{c_1}{2c_{16}} g^{-\omega m}$, so folgt

$$(31) \quad |Q(\gamma)| \geq \frac{c_1}{2} h_i^{-\omega n} g^{1-\omega m} \geq \frac{c_1}{2} g^{1-\omega m - \omega n}.$$

Gilt jedoch

$$(32) \quad h_i^{\omega m + \omega n - \omega - 1 - \sigma_i} \geq \frac{c_1}{2c_{16}} g^{-\omega m},$$

so benutzen wir die Tatsache, daß $Q(\alpha_{i+1}) \neq 0$ ist. Man erhält statt (30) die Abschätzung

$$|Q(\gamma)| \geq h_{i+1}^{-\omega n} g (c_1 g^{-\omega m} - c_{16} h_{i+1}^{\omega m + \omega n - \omega - 1 - \sigma_{i+1}}).$$

Wendet man jetzt (28) auf das zweite Glied in der Klammer an, so folgt

$$|Q(\gamma)| \geq h_{i+1}^{-\omega n} g (c_1 g^{-\omega m} - c_{16} g^{\omega m + \omega n - \omega - 1 - \sigma_{i+1}}).$$

Ist g genügend groß, so gilt wegen $\lim_{i \rightarrow \infty} \sigma_i = \infty$ die weitere Abschätzung

$$|Q(\gamma)| \geq \frac{1}{2} c_1 h_{i+1}^{-\omega n} g^{1-\omega m}.$$

Jetzt schließlich ziehen wir die Eigenschaft 5 heran. Es folgt dann unter Verwendung von (32)

$$|Q(\gamma)| \geq \frac{1}{2} c_1 h_i^{-\omega n \sigma_i / \delta} g^{1-\omega m} \geq \frac{1}{2} c_1 \left(\frac{c_1}{2c_{16}} \right)^{\frac{\omega n \sigma_i / \delta}{\sigma_i + 1 + \omega - \omega m - \omega n}} g^{1-\omega m - \frac{\omega^2 m n \sigma_i / \delta}{\sigma_i + 1 + \omega - \omega m - \omega n}}.$$

Es sei nun $\varepsilon > 0$ beliebig. Wir können dann i so groß wählen, daß

$$1 < \frac{\sigma_i}{\sigma_i + 1 + \omega - \omega m - \omega n} < 1 + \varepsilon$$

ist. Dann ist

$$(33) \quad |Q(\gamma)| \geq c_{17} g^{1-\omega m - [\omega^2 m n (1+\varepsilon)/\delta]}$$

mit einer von g unabhängigen Konstanten c_{17} (c_{17} hängt von $\gamma, \omega, m, n, \varepsilon$ und δ ab). Nimmt man nun (31) und (33) zusammen, so sieht man: Es gilt für jedes irreduzible Polynom Q von genügend großer Höhe die Abschätzung

$$|Q(\gamma)| \geq c_{18} g^{1-\omega m - \max(\omega n, \omega^2 m n (1+\varepsilon)/\delta)}$$

mit einer von g unabhängigen Konstanten c_{18} . Für genügend großes h ist daher

$$\tilde{w}_n(h, \gamma) \geq c_{18} h^{1-\omega m - \max(\omega n, \omega^2 m n(1+\varepsilon)/\delta)},$$

woraus nun $\tilde{w}_n(\gamma) \leq \omega m - 1 + \omega n \max(1, \omega m(1+\varepsilon)/\delta)$ folgt. Das gilt aber für jedes $\varepsilon > 0$. Außerdem gilt $w_n(\gamma) = \tilde{w}_n(\gamma)$. Also haben wir

$$w_n(\gamma) \leq \omega m - 1 + \omega n \max(1, \omega m/\delta).$$

Schließlich ergibt sich noch aus Hilfssatz 7, daß das Maximum auf der rechten Seite gleich $\omega m/\delta$ ist. Damit ist Satz 10 vollständig bewiesen.

Literatur

- [1] *R. Güting*, Über den Zusammenhang zwischen rationalen Approximationen und Kettenbruchentwicklungen, *Math. Zeitschr.* **90** (1965), 382—387.
- [2] *R. Güting*, Polynomials with multiple zeros, *Mathematika* **14** (1967), 181—196.
- [3] *W. J. LeVeque*, On Mahler's U -numbers, *J. London Math. Soc.* **28** (1953), 220—229.
- [4] *K. Mahler*, Zur Approximation der Exponentialfunktion und des Logarithmus. I, *J. reine angew. Math.* **166** (1932), 118—136.
- [5] *K. Mahler*, An application of Jensen's formula to polynomials, *Mathematika* **7** (1960), 98—100.
- [6] *T. Schneider*, Einführung in die transzendenten Zahlen, Berlin 1957.
- [7] *V. G. Sprindžuk*, Beweis der Mahlerschen Vermutung über das Maß der Menge der S -Zahlen, *Izv. Akad. Nauk SSSR, Ser. Mat.* **29** (1965), 379—436 (Russisch).

Makerere University College, P. O. Box 7062, Kampala, Uganda

Eingegangen 1. Februar 1968

Über die alternierenden Verfahren von H. A. Schwarz und C. Neumann

Von *Stefan Hildebrandt* in Mainz

Herrn Prof. Dr. H. Rohrbach zum 65. Geburtstag gewidmet

1. In letzter Zeit sind für das alternierende Verfahren von H. A. Schwarz und eine Reihe von ähnlichen iterativen Prozessen Konvergenzbeweise mit Hilfe von Variationsmethoden oder mit den verwandten Projektionsmethoden im Hilbertraum gegeben worden¹⁾. Mit diesen Iterationsverfahren kann man die Lösung von Randwertaufgaben bei elliptischen Gleichungen für ein Gebiet G konstruieren, falls sich die entsprechenden Randwertaufgaben für G überdeckende Teilgebiete $G_1, G_2, \dots$ lösen lassen. Bei der analogen Aufgabe, ein Randwertproblem für den Durchschnitt G von Gebieten $G_1, G_2, \dots$ zu lösen, für die die Lösung bekannt ist, hat erstmals Carl Neumann [12] ein Iterationsverfahren angegeben. Von A. Pfluger [16] stammt ein Konvergenzbeweis im Falle der Potentialgleichung auf Grund eines Variationsprinzips. Einen anderen Beweis — für ein Randwertproblem der Elastizitätstheorie — hat S. L. Sobolew [18] angedeutet²⁾.

In dieser Arbeit wollen wir mit den früher entwickelten Projektionsmethoden die Konvergenz einer Klasse von Verfahren begründen, die als Verallgemeinerung des Neumannschen Verfahrens für „positiv definite“ Differentialoperatoren aufgefaßt werden können. Diese Konvergenzbeweise haben gegenüber den auf das Maximumprinzip gegründeten den Vorteil, daß sie auch auf elliptische Gleichungen höherer Ordnung und auf Systeme anwendbar sind. Ferner sei bemerkt, daß sich das Neumannsche Verfahren mit dem Trefftzschen kombinieren läßt. Das ist ein Gegenstück zu der in [7] angegebenen Kombination des alternierenden Verfahrens mit dem Ritz-Prozeß (vgl. [7], Nr. 6 und 9). Ferner weisen wir auf unsere Resultate in Nr. 3 über die „minimale Öffnung“ von N Unterräumen hin, die mit der gleichmäßigen Konvergenz des Neumannschen Verfahrens zusammenhängen. Weiterhin studieren wir in Nr. 6, wie das Verfahren für semidefinite, stark elliptische Differentialoperatoren, z. B. für den polyharmonischen Operator Δ^p , modifiziert werden muß, um Konvergenz zu erzielen. Abschließend zeigen wir im Abschnitt 7, daß das alternierende Verfahren von Schwarz im Falle von N Gebieten $G_1, G_2, \dots, G_N$ im allgemeinen „gleichmäßig“ konvergiert; insbesondere haben wir eine Fehlerabschätzung.

Herrn Dipl.-Math. Bernd Schmidt danke ich für Diskussionen über die Öffnung von Unterräumen, die insbesondere zu Satz 5 führten.

¹⁾ Vgl. [7], [8] und die dort angegebenen Literaturhinweise.

²⁾ Wir zeigen später, daß das von Sobolew angegebene Verfahren mit dem Neumannschen übereinstimmt.

2. Das Neumannsche Verfahren für Orthogonalprojektoren

Bezeichne H einen reellen oder komplexen Hilbertraum mit den Elementen $x, y, z, \dots$, dem Skalarprodukt (x, y) und der Norm $\|x\| = (x, x)^{\frac{1}{2}}$. Weiter seien M_i , $i \in J$, endlich oder abzählbar unendlich viele Unterräume von H , $M_0 = \bigcap_{i \in J} M_i$, und $M = \bigvee_{i \in J} M_i$ sei der von den M_i aufgespannte Unterraum von H . Bezeichnen wir mit $U^\perp = H \ominus U$ das orthogonale Komplement eines Unterraumes U , so gilt (vgl. [13], S. 56)

$$(2.1) \quad M_0^\perp = \bigvee_{i \in J} M_i^\perp \text{ und } M^\perp = \bigcap_{i \in J} M_i^\perp.$$

Die orthogonalen Projektoren von H auf die Unterräume M_i , M_0 , M bzw. $M_i^\perp$, $M_0^\perp$, $M^\perp$ nennen wir Q_i , Q_0 , Q bzw. P_i , P_0 , P , also

$$(2.2) \quad I = P_i + Q_i = P_0 + Q_0 = P + Q, \quad i \in J,$$

und

$$(2.3) \quad P \leq P_i \leq P_0, \quad Q_0 \leq Q_i \leq Q, \quad i \in J.$$

Wir definieren mit F. E. Browder [2]: Eine Folge $\{i_n\}$ von Indizes $i_n \in J$ heißt *zulässige Indexfolge*, wenn es zu jedem Index $i \in J$ und zu jeder natürlichen Zahl n zwei natürliche Zahlen $r = r(n, i)$ und $k = k(i)$ gibt, so daß $i_{n+r} = i$ und $r(n, i) \leq k(i)$ für alle $n = 1, 2, \dots$ ist³⁾. Dann sind die folgenden Resultate bekannt:

Satz 1. Mit R_n bezeichnen wir die Kontraktionen $R_n = Q_{i_n} Q_{i_{n-1}} \cdots Q_{i_1}$.

(i) Für jede zulässige Indexfolge $\{i_n\}$ konvergieren die R_n schwach gegen Q_0 : $R_n \rightarrow Q_0$ für $n \rightarrow \infty$ (Browder [2]).

(ii) Wenn $J = \{1, 2, \dots, N\}$ eine endliche Indexmenge ist, so gilt $R_n \rightarrow Q_0$ für jede Indexfolge $\{i_n\}$, $i_n \in J$, in der jeder Index aus J unendlich oft auftritt (Amemiya-Ando [4]).

(iii) Wenn $J = \{1, 2, \dots, N\}$ und die zulässigen Indexfolgen $1, 2, \dots, N, 1, 2, \dots, N, \dots$ bzw. $1, 2, \dots, N, \dots, 2, 1, 1, 2, \dots, N, \dots, 2, 1, \dots$ gewählt werden, so haben wir sogar starke Konvergenz: $R_n \rightarrow Q_0$ (von Neumann [13], Hildebrandt [7], Halperin [6]).

(iv) Wenn $J = \{1, 2, \dots, \infty\}$ und $\{i_n\}$ die zulässige Indexfolge $1, 2, 1, 3, 1, 2, 1, 4, 1, 2, 1, 3, 1, 2, 1, 5, \dots$ bezeichnet, so gilt ebenfalls $R_n \rightarrow Q_0$ (Stummel [19]).

(v) Falls $J = \{1, 2, \dots, N\}$, so ist die folgende Bedingung für die gleichmäßige Konvergenz $\|R_n - Q_0\| \rightarrow 0$ für eine zulässige Indexfolge notwendig und hinreichend:

$$(2.4) \quad \text{Jedes Element } z \in M_0^\perp \text{ läßt sich in der Form } z = \sum_{i=1}^N z_i \text{ mit } z_i \in M_i^\perp \text{ (} i = 1, \dots, N \text{) darstellen, d. h. } M_0^\perp = M_1^\perp + M_2^\perp + \cdots + M_N^\perp.$$

(Hildebrandt und Schmidt [8], vgl. ferner [7], [15], [17] und [19].)

(Die gleichmäßige Konvergenz $\|R_n - Q_0\| \rightarrow 0$ bezeichnen wir auch durch $R_n \Rightarrow Q_0$).

Beachten wir nun die Relationen (2.2), so folgt

$$(2.5) \quad I - R_n \rightarrow P_0, \quad I - R_n \rightarrow P_0 \text{ bzw. } I - R_n \Rightarrow P_0,$$

je nachdem, welche der Voraussetzungen von (i) bis (v) wir wählen. Wegen $R_n = Q_{i_n} Q_{i_{n-1}} \cdots Q_{i_1} = (I - P_{i_n}) R_{n-1}$ erhalten wir dann

$$(2.6) \quad I - R_1 = P_{i_1}, \quad I - R_n = (I - R_{n-1}) + P_{i_n} \{I - (I - R_{n-1})\} \text{ für } n \geq 2.$$

³⁾ Die in [7] betrachteten regulären Klasseneinteilungen von Indizes enthalten die zulässigen Indexfolgen als Spezialfall (vgl. [17]).

Somit ergibt sich der

Satz 2. *Definiert man für einen beliebigen Vektor $g \in H$ die Folge $\{x_n\}$ durch*

$$(2.7) \quad x_0 = 0, \quad x_n = x_{n-1} + P_{i_n}\{g - x_{n-1}\} \text{ für } n \geq 1,$$

oder, was dazu äquivalent ist,

$$(2.7') \quad x_n = \sum_{k=1}^n z_k, \quad z_1 = P_{i_1}g, \quad z_k = P_{i_k}\left(g - \sum_{j=1}^{k-1} z_j\right), \text{ für } j \geq 2$$

und setzen wir $u = P_0g$, so gilt

$$x_n \rightharpoonup u, \quad x_n \rightarrow u \text{ oder } \|x_n - u\| \leq \|Q_0 - R_n\| \|g\| \rightarrow 0,$$

je nachdem, welche der Voraussetzungen (i) bis (v) von Satz 1 wir machen.

Den durch (2.7) bzw. (2.7') definierten Algorithmus wollen wir *Neumannsches Verfahren* nennen.

Als eine erste Anwendung des Satzes 2 erwähnen wir das prediction problem für einen q -variaten stationären stochastischen Prozeß. Mit Hilfe des alternierenden Verfahrens können wir die erzeugende Funktion des stochastischen Prozesses konstruieren. Für $q = 2$ haben dies bereits Wiener und Masani ausgeführt (vgl. [21], Teil II, S.104—111).

3. Gleichmäßige Konvergenz. Öffnung von Unterräumen

In [8] wurde folgendes Resultat bewiesen:

$$\textbf{Satz 3.} \quad \|Q_N Q_{N-1} \cdots Q_1 - Q_0\| < 1 \Leftrightarrow M_0^\perp = M_1^\perp + M_2^\perp + \cdots + M_N^\perp.$$

Bei den Anwendungen, die wir im Auge haben, ist an Stelle von (2.4) die duale Voraussetzung

$$(3.1) \quad M = M_1 + M_2 + \cdots + M_N$$

erfüllt, die nach Satz 3 zu

$$(3.1') \quad \|P_N P_{N-1} \cdots P_1 - P\| < 1$$

äquivalent ist. Wir fragen nun, ob die Bedingungen (2.4) und (3.1) äquivalent sind. Dies ist in der Tat für $N = 2$ der Fall, sogar im Banachraum, was seit langem wohlbekannt ist. (Bezüglich der Literatur und Anwendungen auf Stabilitätsaussagen für Semi-Fredholmoperatoren verweisen wir auf Kato [9]). Wir geben zunächst einen neuen kurzen Beweis dieses Resultates, der als Nebenresultat einige interessante Formeln für den orthogonalen Projektor P_0 auf den „Span“ $M_0^\perp = M_1^\perp \vee M_2^\perp$ von zwei Unterräumen $M_1^\perp$ und $M_2^\perp$ liefert, falls $M_1^\perp + M_2^\perp$ abgeschlossen und somit gleich $M_0^\perp$ ist. Danach zeigen wir anhand eines Beispiels, daß die Bedingungen (2.4) und (3.1) für $N \geq 3$ nicht mehr äquivalent sind, und schließlich geben wir einige Bedingungen an, unter denen wenigstens (2.4) aus (3.1) folgt.

Satz 4. *$M_1 + M_2$ ist dann und nur dann abgeschlossen, wenn $M_1^\perp + M_2^\perp$ abgeschlossen ist.*

Beweis. Sei $M = M_1 + M_2$ und somit $\|P_2 P_1 - P\| < 1$. Setzen wir nun $P'_i = P_i - P$, ($i = 0, 1, 2$), so gilt

$$(3.2) \quad \|P'_2 P'_1\| < 1,$$

und wegen $I - (Q_2 Q_1)^n \rightarrow P_0$ mit $n \rightarrow \infty$ und $I = P + Q$ folgt

$$(3.3) \quad Q - [(Q - P'_2)(Q - P'_1)]^n \rightarrow P'_0 \text{ für } n \rightarrow \infty.$$

Aus den Relationen $P'_i = QP'_i = P'_i Q$, ($i = 1, 2$), ergibt sich dann leicht mittels Induktion

$$\begin{aligned} & [(Q - P'_2)(Q - P'_1)]^n \\ &= \{Q - P'_2 + (P'_1 P'_2) - P'_2(P'_1 P'_2) + (P'_1 P'_2)^2 - \dots - P'_2(P'_1 P'_2)^{n-1}\} (Q - P'_1). \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} (3.4) \quad & Q - [(Q - P'_2)(Q - P'_1)]^n \\ &= P'_1 + P'_2 \left[\sum_{v=0}^{n-1} (P'_1 P'_2)^v \right] (Q - P'_1) - P'_1 P'_2 \left[\sum_{v=0}^{n-2} (P'_1 P'_2)^v \right] (Q - P'_1). \end{aligned}$$

Aus (3.2) bis (3.4) folgt dann

$$P'_0 = P'_2(I - P'_1 P'_2)^{-1} \cdot (Q - P'_1) - P'_1 P'_2(I - P'_1 P'_2)^{-1} (Q - P'_1) + P'_1.$$

Mittels der Operatoren

$$S_1 = -P'_2(I - P'_1 P'_2)^{-1} \cdot (Q - P'_1) + I, \quad S_2 = (I - P'_1 P'_2)^{-1} (Q - P'_1)$$

bekommen wir dann wegen $P_0 = P'_0 + P$ die Darstellung

$$(3.5) \quad P_0 = P + P'_1 S_1 + P'_2 S_2,$$

aus der sich unmittelbar $M_0^\perp = M_1^\perp + M_2^\perp$ ergibt. Damit ist die eine Hälfte des Satzes bewiesen. Die Umkehrung erfolgt durch Dualisierung.

Bemerkung. Aus (3.2) bis (3.4) lassen sich noch die folgenden Formeln ableiten:

Wenn $P'_i = P_i - P$ und $Q'_i = I - P'_i$ gesetzt und $M = M_1 + M_2$ angenommen werden, so gilt

$$\begin{aligned} (3.6) \quad & P_0 = I - Q_2(I - P'_1 P'_2)^{-1} \cdot Q_1 = I - Q_1(I - P'_2 P'_1)^{-1} \cdot Q_2 \\ &= P + \{I - Q'_1(I - P'_2 P'_1)^{-1}\} + \{I - Q'_2(I - P'_1 P'_2)^{-1}\}. \end{aligned}$$

Satz 5. Für $N > 2$ sind die Bedingungen

$$„M_1 + M_2 + \dots + M_N \text{ ist abgeschlossen}“$$

und

$$„M_1^\perp + M_2^\perp + \dots + M_N^\perp \text{ ist abgeschlossen}“$$

nicht äquivalent, falls $\dim H = \infty$ ist.

Beweis. Wir wählen drei Unterräume M_1 , M_2 und M_3 von H mit den folgenden Eigenschaften:

(i) $\dim M_1 < \infty$.

(ii) $M_1 \subset M_2$, d. h. $Q_1 Q_2 = Q_2 Q_1 = Q_1$.

(iii) $M_2 + M_3$ ist nicht abgeschlossen und $M_2 \cap M_3 = \{0\}$ ⁴⁾. Dabei bezeichne Q_i den Orthogonalprojektor von H auf M_i und P_i den von H auf $M_i^\perp$.

⁴⁾ Bekanntlich existieren in jedem unendlichdimensionalen Hilbertraum solche Unterräume.

Nun gilt

$$(3.7) \quad \|Q_1 Q_2 Q_3\| = \|Q_1 Q_3\| < 1.$$

Anderenfalls gäbe es eine Folge x_n mit $\|x_n\| = 1$ und $\|Q_1 Q_3 x_n\| \rightarrow 1$. Ohne Beschränkung der Allgemeinheit könnten wir $x_n \rightarrow x_0$ annehmen. Wegen der Vollstetigkeit von $Q_1 Q_3$ und der schwachen Unterhalbstetigkeit der Norm folgte

$$1 = \|Q_1 Q_3 x_0\| \leq \|x_0\| \leq \lim_{n \rightarrow \infty} \|x_n\| = 1$$

und daher $x_0 = Q_1 x_0 = Q_3 x_0$, d. h. $x_0 \in M_1 \cap M_3 = \{0\}$, oder $x_0 = 0$, was ein Widerspruch zu $\|x_0\| = 1$ ist.

Aus (3.7) und Satz 3 folgt, daß $M_1^\perp + M_2^\perp + M_3^\perp$ abgeschlossen ist, während wegen (ii) und (iii) $M_1 + M_2 + M_3 = M_2 + M_3$ nicht abgeschlossen ist. Damit ist Satz 5 bewiesen.

Satz 6. $M_1^\perp + M_2^\perp + \dots + M_N^\perp$ ist abgeschlossen, wenn die folgenden zwei Bedingungen erfüllt sind:

(a) $M_1 + M_2 + \dots + M_N$ ist abgeschlossen.

(b) Aus $x_1 + x_2 + \dots + x_N = 0$ für $x_i \in M_i$, ($i = 1, 2, \dots, N$) folgt $x_i \in M_0 = \bigcap_{j=1}^N M_j$, ($i = 1, 2, \dots, N$).

*Beweis*⁵⁾. Wir bezeichnen mit Q'_i den orthogonalen Projektor von H auf $M'_i = M_i \ominus M_0$, d. h. $Q'_i = Q_i - Q_0$. Wäre nun $M_1^\perp + M_2^\perp + \dots + M_N^\perp$ nicht abgeschlossen, so folgte aus Satz 3

$$(3.8) \quad \|Q'_N Q'_{N-1} \dots Q'_1\| = 1.$$

Setzen wir $P'_i = I - Q'_i$, so gilt für die Funktion

$$(3.9) \quad A(x) := \|P'_2 x\|^2 + \|P'_3 Q'_2 x\|^2 + \dots + \|P'_N Q'_{N-1} \dots Q'_3 Q'_2 x\|^2$$

die Relation

$$A(x) + \|Q'_N Q'_{N-1} \dots Q'_2 x\|^2 = \|x\|^2$$

und daher

$$(3.10) \quad \|Q'_N Q'_{N-1} \dots Q'_1\|^2 = \sup_{x \in M'_1} \frac{\|Q'_N Q'_{N-1} \dots Q'_1 x\|^2}{\|x\|^2} = 1 - \inf_{x \in M'_1} \frac{A(x)}{\|x\|^2}.$$

Dann folgt aus (3.8) und (3.10)

$$(3.11) \quad \inf_{x \in M'_1, \|x\|=1} A(x) = 0.$$

Wir führen nun die *minimale Öffnung* der Unterräume $M'_1, M'_2, \dots, M'_N$ ein durch

$$(3.12) \quad \varepsilon(M'_1, M'_2, \dots, M'_N) := \inf_{\substack{x_i \in M'_i \\ (i=1,2,\dots,N)}} \frac{\left\| \sum_{i=1}^N x_i \right\|}{\left(\sum_{i=1}^N \|x_i\|^2 \right)^{\frac{1}{2}}}.$$

⁵⁾ Vgl. [7], Beweis von Lemma 0.1, [16] und [4], Lemma 1.

Dann erhalten wir für beliebiges $x \in M'_1$ die Abschätzung

$$\begin{aligned} \varepsilon(M'_1, \dots, M'_n) &\leq \frac{\|x + 0 + 0 + \dots + (-Q'_N Q'_{N-1} \dots Q'_1 x)\|}{(\|x\|^2 + \|Q'_N Q'_{N-1} \dots Q'_1 x\|^2)^{\frac{1}{2}}} \\ &\leq \frac{\|x - Q'_2 x + Q'_2 x - Q'_3 Q'_2 x + Q'_3 Q'_2 x - \dots - Q'_N Q'_{N-1} \dots Q'_1 x\|}{\|x\|} \\ &\leq \frac{\|P'_2 x\| + \|P'_3 Q'_2 x\| + \dots + \|P'_N Q'_{N-1} \dots Q'_2 x\|}{\|x\|} \\ &\leq \sqrt{N-1} \frac{\sqrt{A(x)}}{\|x\|} \end{aligned}$$

oder

$$(3.13) \quad 0 \leq \varepsilon(M'_1, M'_2, \dots, M'_N) \leq \sqrt{N-1} \left\{ \inf_{x \in M'_1, \|x\|=1} A(x) \right\}^{\frac{1}{2}}.$$

Aus (3.11) und (3.13) erhalten wir schließlich

$$(3.14) \quad \varepsilon(M'_1, M'_2, \dots, M'_N) = 0.$$

Wir bilden jetzt den Produktraum $\mathfrak{A} = M'_1 \times M'_2 \times \dots \times M'_N$ mit den Elementen $X = \{x_1, x_2, \dots, x_N\}$, $x_i \in M'_i$, ($i = 1, 2, \dots, N$), der mit dem Skalarprodukt $(X, Y) = \sum_{i=1}^N (x_i, y_i)$ und der Norm $\|X\| = (X, X)^{\frac{1}{2}} = \left(\sum_{i=1}^N \|x_i\|^2 \right)^{\frac{1}{2}}$ zu einem Hilbertraum wird. Dann definieren wir durch

$$(3.15) \quad TX = \sum_{i=1}^N x_i, \quad X = \{x_1, x_2, \dots, x_N\} \in \mathfrak{A}$$

eine Abbildung von $\mathfrak{A}$ auf den linearen Raum $U = M'_1 + M'_2 + \dots + M'_N$. Wegen Voraussetzung (b) ist U die direkte Summe der Unterräume $M'_1, M'_2, \dots, M'_N$ und somit $TX = 0 \Leftrightarrow X = 0$. Daher ist T eine eindeutige und wegen

$$\|TX\| \leq \sum_{i=1}^N \|x_i\| \leq \sqrt{N} \left(\sum_{i=1}^N \|x_i\|^2 \right)^{\frac{1}{2}} = \sqrt{N} \|X\|$$

auch stetige Abbildung von $\mathfrak{A}$ auf U . Nach einem bekannten Satz von Banach ist also U genau dann abgeschlossen, wenn T^{-1} stetig ist, was genau dann der Fall ist, wenn

$$0 < \inf_{X \in \mathfrak{A}} \frac{\|TX\|}{\|X\|} = \varepsilon(M'_1, M'_2, \dots, M'_N)$$

gilt. Also ergibt sich schließlich aus (3.14), daß $M'_1 + \dots + M'_N$ und somit auch $M_1 + M_2 + \dots + M_N$ nicht abgeschlossen wäre, womit wir einen Widerspruch zu (a) gefunden hätten.

Bemerkung. Da für $N = 2$ bereits aus $M'_1 + M'_2 = \{0\}$ folgt, daß $M'_1 + M'_2$ eine direkte Summe ist, ohne daß wir (b) voraussetzen müssen, so ergibt sich aus Satz 6 ein weiterer Beweis für Satz 4.

Satz 7. Wenn einer der Räume $M'_i = M_i \ominus M_0$ endlichdimensional ist, so ist $M_1^1 + M_2^1 + \dots + M_N^1$ abgeschlossen.

Beweis. Wir können o. B. d. A. annehmen, daß $\dim M_1 < \infty$ ist. Dann zerlegen wir die lineare Mannigfaltigkeit $U = M_2^1 + \dots + M_N^1$ in der Form

$$U = U_1 + U_2, \quad U_1 \subset M_1^1, \quad U_2 \subset M_1.$$

Dann ist $M_1^1 + \dots + M_N^1 = M_1^1 + U_2$, wobei M_1^1 abgeschlossen und U_2 endlich-dimensional und orthogonal zu M_1^1 ist. Hieraus ergibt sich leicht die Abgeschlossenheit von $M_1^1 + U_2$ und damit von $M_1^1 + \dots + M_N^1$.

Mit den Sätzen 4, 6, 7 haben wir Bedingungen für die Räume $M_1, M_2, \dots, M_N$ gefunden, die (vgl. Satz 3) die gleichmäßige Konvergenz der Operatoren R_n gegen Q_0 sichern und damit eine Fehlerabschätzung für das in Satz 2 definierte Neumannsche Verfahren liefern, während die Bedingung von Satz 3 an die M_i^1 gestellt wird.

4. Das Neumannsche Verfahren für das Dirichletproblem bei positiv definiten linearen Differentialgleichungen

Wir betrachten im d -dimensionalen Raum E_d der Punkte $t = (t^1, t^2, \dots, t^d)$ eine endlich oder abzählbar unendliche Familie $\{G_i\}_{i \in J}$ von beschränkten Gebieten $G_i \subset E_d$ und die Mengen

$$(4.1) \quad G_0 = \bigcap_{i \in J} G_i, \quad G = \bigcup_{i \in J} G_i,$$

die wir ebenfalls als beschränkte Gebiete des E_d annehmen. Es ist also

$$(4.2) \quad G_0 \subseteq G_i \subseteq G.$$

Dann führen wir in bekannter Weise die Hilberträume $H_s(\Omega)$ und $\dot{H}_s(\Omega)$ als Vervollständigung der Klassen $C^s(\bar{\Omega})$ und $C_c^s(\Omega)$ in der Norm

$$(4.3) \quad \|\varphi\|_{s,\Omega} = \left\{ \int_{\Omega} \sum_{k=1}^r \sum_{|\alpha| \leq s} |D_{\alpha} \varphi^k|^2 dt \right\}^{\frac{1}{2}}, \quad (s = 0, 1, 2, \dots),$$

ein. Dabei bezeichnet $C^s(\bar{\Omega})$ die Klasse der auf $\bar{\Omega}$ s -mal stetig differenzierbaren reellwertigen Vektoren $\varphi(t) = (\varphi^1(t), \varphi^2(t), \dots, \varphi^r(t))$, und $C_c^s(\Omega)$ besteht aus den Vektoren $\in C^s(\bar{\Omega})$, die einen kompakten, in der beschränkten, offenen Menge Ω gelegenen Träger haben.

Wir wollen nun die Ergebnisse von Nr. 2 und Nr. 3 auf den Raum $H = H_p(G)$ und seine Unterräume

$$(4.4) \quad M_i = \dot{H}_p(G_i)$$

anwenden⁶⁾.

Aus (4.1) und (4.2) folgt dann

$$(4.5) \quad M_0 = \dot{H}_p(G_0) = \bigcap_{i \in J} M_i \quad \text{und} \quad M = \dot{H}_p(G) = \bigvee_{i \in J} M_i.$$

(Der Beweis von $\dot{H}_p(G) = \bigvee_{i \in J} \dot{H}_p(G_i)$ folgt wie in [7] mit einer geeigneten Zerlegung der Eins auf G .)

Wir betrachten nun auf $\bar{G}$ einen (stark elliptischen) Differentialoperator L der Ordnung $2p$, dessen Koeffizienten $a_{jk}^{\alpha\beta}(t)$ „hinreichend oft“ stetig differenzierbar auf $\bar{G}$ sind:

$$(4.6) \quad Lx = \sum_{|\alpha|, |\beta| \leq p} (-1)^{|\beta|} D_{\beta} [a^{\alpha\beta} D_{\alpha} x].$$

⁶⁾ Daß wir $\dot{H}_p(G_i)$ in natürlicher Weise als Unterraum von $H_p(G)$ deuten können, ist evident.

Dabei bezeichne α einen Multi-Index $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$, $\alpha_j \geq 0$ ($j = 1, 2, \dots, d$), $|\alpha| = \sum_{j=1}^d \alpha_j$, $D_\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \dots D_d^{\alpha_d}$, $D_1 = \frac{\partial}{\partial t^1}$, etc. ...; $x = (x^1(t), x^2(t), \dots, x^r(t))$, und $a^{\alpha\beta}$ sind $r \times r$ -Matrizen $(a_{jk}^{\alpha\beta})$. Wir setzen nun voraus, daß die L zugeordnete Bilinearform

$$(4.7) \quad B(x, y) = \int_G \sum_{i,k=1}^r \sum_{|\alpha|, |\beta| \leq p} a_{jk}^{\alpha\beta} D_\alpha x^j D_\beta y^k dt$$

auf $H_p(G)$ symmetrisch und positiv definit ist. Da $B(x, y)$ offensichtlich auch stetig auf $H_p(G)$ ist (Anwendung der Schwarzschen Ungleichung), haben wir

$$(4.8) \quad B(x, y) = B(y, x) \text{ für } x, y \in H_p(G)$$

und

$$(4.9) \quad c \cdot \|x\|_{p,G}^2 \leq B(x, x) \leq c' \cdot \|x\|_{p,G}^2 \text{ für } x \in H_p(G) \text{ mit } 0 < c \leq c'.$$

Damit definiert $(x, y) := B(x, y)$ ein Skalarprodukt auf $H_p(G)$, dessen zugeordnete Norm $\|x\| = \sqrt{B(x, x)}$ zur Norm $\|x\|_{p,G}$ äquivalent ist. Somit bildet $H = H_p(G)$ auch bezüglich (x, y) einen Hilbertraum, und die Mannigfaltigkeiten M_0, M_i, M bleiben in der neuen Norm $\|x\|$ abgeschlossen. Wir denken uns daher im folgenden H und seine Unterräume mit dem neuen Skalarprodukt (x, y) und der zugeordneten Norm $\|x\|$ versehen.

Wir wollen nun untersuchen, was die Anwendung eines Projektors $P_i = I - Q_i$ auf ein Element $f \in H$ bedeutet, wenn Q_i wie bisher die Orthogonalprojektion von H auf M_i bezeichnet. Wir zerlegen

$$f = P_i f + Q_i f = u_i + v_i, \quad u_i \in M_i^\perp, \quad v_i \in M_i.$$

Dann ist $u_i = f - v_i$ eindeutig charakterisiert als Lösung der Gleichung

$$B(f - v_i, v) = 0 \text{ für alle } v \in M_i = \dot{H}_p(G_i); \quad v_i \in M_i.$$

Somit ist

$$u_i = \begin{cases} f & \text{in } G - G_i \\ u & \text{in } G_i, \end{cases}$$

wobei u die Lösung der Randwertaufgabe

$$Lu = 0 \quad \text{in } G_i$$

$$D_\alpha u = D_\alpha f \text{ auf } \partial G_i \text{ für } |\alpha| \leq p - 1 \text{ im verallgemeinerten Sinne}$$

bezeichnet, wenn wir den Regularitätssatz für die Lösungen elliptischer Differentialgleichungen („Weylsches Lemma“, vgl. z. B. [14]) beachten. Die Randbedingung ist als $u - f \in \dot{H}_p(G_i)$ zu verstehen und kann auch im klassischen Sinne gelesen werden, falls f und ∂G_i „hinreichend glatt“ sind.

Wenden wir jetzt Satz 2 auf unser Beispiel an, so ergibt sich folgendes Resultat:

Satz 8. Definieren wir für einen beliebigen Vektor $g \in H = H_p(G)$ und eine zulässige Folge von Indizes $i_n \in J$ die Folge $\{x_n\}$ durch

$$(4.10) \quad x_0 = 0, \quad x_n = x_{n-1} + z_n,$$

wobei

$$(4.11) \quad z_n = \begin{cases} g - x_{n-1} & \text{in } G - G_{i_n} \\ \zeta_n & \text{in } G_{i_n} \end{cases}$$

und ξ_n die Lösung der Randwertaufgabe

$$(4.12) \quad \begin{aligned} L\xi_n &= 0 && \text{in } G_{i_n} \\ D_\alpha \xi_n &= D_\alpha[g - x_{n-1}] && \text{auf } \partial G_{i_n} \text{ i. v. S. für } |\alpha| \leq p-1 \end{aligned}$$

ist, so gilt in H : $x_n \rightarrow u$, $x_n \rightarrow u$ oder $\|x_n - u\| \leq \|R_n - Q_0\| \cdot \|g\| \rightarrow 0$ für $n \rightarrow \infty$, je nachdem, welche der Bedingungen (i) bis (v) von Satz 1 erfüllt ist. Das Grenzelement $u = P_0 g$ ist die Lösung der Randwertaufgabe

$$(4.13) \quad \begin{aligned} Lu &= 0 && \text{in } G_0 = \bigcap_{i \in J} G_i \\ u &= g && \text{auf } \partial G_0 \text{ i. v. S.} \end{aligned}$$

Auf Grund des Rellichschen Satzes haben wir in jedem Fall

$$(4.14) \quad \lim_{n \rightarrow \infty} \|x_n - u\|_{0, G} = 0.$$

Ferner existieren nach bekannten Ergebnissen aus der Theorie der elliptischen Differentialgleichungen a priori Schranken der Form

$$\|x\|_{2p+s, G'} \leq \text{const} \cdot \{\|x\|_{0, G_0} + \|f\|_{s, G_0}\}$$

für die schwachen Lösungen x von

$$Lx = f \text{ in } G_0$$

für jedes Teilgebiet $G' \subset \bar{G}' \subset G_0$ (wofür wir $G' \subset\subset G_0$ schreiben). Damit folgt

$$(4.15) \quad \|x_n - u\|_{2p+s, G'} \leq \text{const} \cdot \|x_n - u\|_{0, G_0} \rightarrow 0 \text{ mit } n \rightarrow \infty$$

für alle $s \leq s_0$, wobei s_0 von der Differenzierbarkeit der Koeffizienten $a_{jk}^{\alpha\beta}$ abhängt und $= \infty$ gewählt werden kann, wenn $a_{jk}^{\alpha\beta} \in C^\infty(G)$ ist. Aus der Sobolewschen Ungleichung

$$\|v\|_{k-\left[\frac{d}{2}\right]-1, \Omega'} \leq \text{const} \|v\|_{k, \Omega}, \quad \Omega' \subset\subset \Omega,$$

folgt dann⁷⁾

$$(4.16) \quad \lim_{n \rightarrow \infty} \sup_{G'} |D^\alpha x_n(t) - D^\alpha u(t)| = 0 \text{ für } |\alpha| \leq 2p + s_0 - \left[\frac{d}{2}\right] - 1$$

und $G' \subset\subset G_0$, wobei $|x| = \left(\sum_{k=1}^r |x^k|^2\right)^{\frac{1}{2}}$ für $x = (x^1, x^2, \dots, x^r)$ gesetzt ist.

Mit anderen Worten:

Wir erhalten nicht nur Konvergenz im Mittel, sondern auch punktweise gleichmäßige Konvergenz von $x_n(t)$ und allen Ableitungen bis zu einer gewissen Ordnung gegen die Lösung u der Randwertaufgabe (4.13) und deren entsprechende Ableitungen in jedem echten Teilgebiet von G_0 . Falls die Voraussetzung (2.4) aus Satz 1 erfüllt ist, erhalten wir sogar eine Fehlerabschätzung.

Wir betrachten jetzt noch den Spezialfall von zwei Gebieten G_1 und G_2 , $G_0 = G_1 \cap G_2$, $G = G_1 \cup G_2$ und wählen

$$i_n = \begin{cases} 1 & \text{für } n \text{ ungerade} \\ 2 & \text{für } n \text{ gerade.} \end{cases}$$

⁷⁾ $\|v\|_{s, \Omega} := \sup_{x \in \Omega, |\alpha| \leq s} |D_\alpha v(x)|.$

Satz 9. Dann können wir die durch (2. 7) bzw. (4. 10) bis (4. 12) definierte Folge $\{x_n\}$ auch auf folgende Weise erzeugen:

$$(4. 17) \quad x_n = \sum_{\nu=1}^n (-1)^{\nu-1} w_\nu \\ w_1 = P_1 g, w_2 = P_2(-g + w_1), w_n = P_n w_{n-1} \text{ für } n > 2.$$

Beweis. Es ist $x_1 = \{I - (I - P_1)\}g = P_1 g = w_1$

$$\begin{aligned} x_2 &= \{I - (I - P_2)(I - P_1)\}g \\ &= \{I - (I - P_1)\}g - \{-P_2 + P_2 P_1\}g \\ &= w_1 - w_2. \end{aligned}$$

Angenommen, wir hätten bereits

$$x_{2k} = \{I - [(I - P_2)(I - P_1)]^k\}g = w_1 - w_2 + w_3 - \cdots + w_{2k-1} - w_{2k}$$

bewiesen. Dann folgt aus $P_1 w_{2\nu-1} = w_{2\nu-1}$ und $P_1 w_{2\nu} = w_{2\nu+1}$, $\nu \geq 1$,

$$\begin{aligned} x_{2k+1} &= \{I - (I - P_1)[(I - P_2)(I - P_1)]^k\}g \\ &= x_{2k} - P_1\{w_1 - w_2 + w_3 - \cdots - w_{2k}\} + P_1 g \\ &= x_{2k} - w_1 + w_3 - w_3 + \cdots + w_{2k-1} - w_{2k-1} + w_{2k+1} + w_1 \\ &= x_{2k} + w_{2k+1}, \end{aligned}$$

und aus $P_2 w_{2\nu} = w_{2\nu}$ und $P_2 w_{2\nu+1} = w_{2\nu+2}$ für $\nu \geq 1$ ergibt sich

$$\begin{aligned} x_{2k+2} &= \{I - (I - P_2)(I - P_1)[(I - P_2)(I - P_1)]^k\}g \\ &= x_{2k+1} - P_2\{w_1 - w_2 + w_3 - \cdots - w_{2k} + w_{2k+1}\} + P_2 g \\ &= x_{2k+1} + (P_2 g - P_2 w_1) + w_2 - w_4 + w_4 - \cdots - w_{2k} + w_{2k} - w_{2k+2} \\ &= x_{2k+1} - w_{2k+2}, \end{aligned} \quad \text{q. e. d.}$$

Bei Anwendung auf das Randwertproblem liefert der durch (4. 17) definierte Algorithmus gerade das von Sobolew⁸⁾ angegebene Verfahren, das also mit dem Neumannschen Verfahren zusammenfällt.

Als eine Anwendung von Satz 4 skizzieren wir noch das folgende Beispiel: Sei auf einer kompakten Riemannschen Mannigfaltigkeit $\mathfrak{M}$ ein „Ringgebiet“ R als Durchschnitt $R = G_1 \cap G_2$ von zwei Gebieten $G_1, G_2 \subset \mathfrak{M}$ dargestellt. Auf $\mathfrak{M}$ denken wir uns einen Differentialoperator L der Ordnung $2p$ mit positiv definiter symmetrischer Bilinearform gegeben. Für die geeignet zu definierenden Räume $H = H_p(\mathfrak{M})$ und $M_i = \dot{H}_p(G_i)$, ($i = 1, 2$), folgt mittels einer Zerlegung der Eins auf $\mathfrak{M}$: $H = M_1 + M_2$. (Vgl. [7].)

Dann ist nach Satz 4 auch $(M_1 \cap M_2)^\perp = M_1^\perp + M_2^\perp$. Andererseits ist

$$M_1 \cap M_2 = \dot{H}_p(R),$$

womit gezeigt ist:

Satz 10. $\dot{H}_p(R)^\perp = \dot{H}_p(G_1)^\perp + \dot{H}_p(G_2)^\perp$, d. h. jede Lösung von $Lu = 0$ in R läßt sich in R als Summe $u = u_1 + u_2$ einer Lösung u_1 von $Lu_1 = 0$ in G_1 und einer Lösung u_2 von $Lu_2 = 0$ in G_2 darstellen.

(Vgl. auch Pfluger [16], S. 135—137.)

⁸⁾ Vgl. [18], S. 246, Formeln (19) bis (23).

5. Kombination des Neumannschen mit dem Trefftz'schen Verfahren

Wir beweisen zunächst eine einfache Verallgemeinerung von Satz 4, (i), die wir in etwas anderer Form bereits in einer früheren Arbeit angegeben haben (vgl. [7], Satz 6. 1). In der Beweismethode folgen wir Stummel [19].

Satz 11. Sei $\{Q_i\}_{i \in J}$ eine Menge von endlich oder abzählbar unendlich vielen orthogonalen Projektoren im Hilbertraum H , ferner seien $\{i_n\}$ eine zulässige Indexfolge ($i_n \in J$) und $\{A_k\}$ eine Folge von Orthogonalprojektoren mit den folgenden Eigenschaften:

(a) Für $n < m$ und $i_n = i_m = i \in J$ gilt $A_n \geq A_m \geq Q_i$

(b) $A_n \rightarrow Q_i$ für $i_n = i$ und $n \rightarrow \infty$.

Dann konvergieren die Operatoren $R_k = A_k A_{k-1} \cdots A_1$ für $k \rightarrow \infty$ schwach gegen den Orthogonalprojektor Q_0 von H auf $M_0 = \bigcap_{k=1}^{\infty} U_k = \bigcap_{i \in J} M_i$, wobei $U_k = A_k H$ und $M_i = Q_i H$ gesetzt ist.

Beweis. Setzen wir $D_k = R_k - Q_0$, $P_0 = I - Q_0$, $D_k = R_k P_0 = A_k R_{k-1} P_0 = A_k D_{k-1}$ und $B_k = I - A_k$, so folgt für $r = r(n, i)$

$$(5.1) \quad \|B_{n+r} D_n x\| \leq \sum_{v=n+1}^{n+r} \|D_{v-1} x - D_v x\| \leq \sum_{v=n+1}^{n+k(i)} \|D_{v-1} x - D_v x\|.$$

Wir geben uns nun irgendeinen Index $i \in J$ vor und betrachten diejenigen natürlichen Zahlen l , für die $i_l = i$ ist. Fixieren wir nun ein solches l und betrachten die natürlichen Zahlen $n > l$. Zu n und i wählen wir $r(n, i) \leq k(i)$, so daß $i_l = i_{n+r} = i$ und $l < n + r$, also auch $A_l \geq A_{n+r}$ oder $B_l \leq B_{n+r}$ ist. Wegen (5.1) folgt dann

$$(5.2) \quad \|B_l D_n x\|^2 \leq k(i) \sum_{v=n+1}^{n+k(i)} \|D_{v-1} x - D_v x\|^2 = k(i) \{ \|D_n x\|^2 - \|D_{n+k(i)} x\|^2 \}.$$

Da $\|D_{n+1} x\| \leq \|D_n x\|$ ist, existiert $\lim_{n \rightarrow \infty} \|D_n x\|$, und wir erhalten aus (5.2)

$$(5.3) \quad \lim_{n \rightarrow \infty} \|B_l D_n x\| = 0 \text{ für alle } l \text{ und für } x \in H.$$

Die Relation $M_0 = \bigcap_{k=1}^{\infty} U_k = \bigcap_{i \in J} M_i$ ist offensichtlich richtig. Aus $M_0^\perp = \bigcap_{l=1}^{\infty} U_l^\perp$ folgt, daß die Elemente $u = \sum_{l=1}^m u_l$, $u_l \in U_l^\perp$, in $M_0^\perp$ dicht liegen. Somit gilt für $x \in H$ wegen $B_l u_l = u_l$

$$\begin{aligned} |(u, D_n x)|^2 &= \left| \left(\sum_{l=1}^m u_l, D_n x \right) \right|^2 = \left| \sum_{l=1}^m (u_l, B_l D_n x) \right|^2 \\ &\leq \left(\sum_{l=1}^m \|u_l\| \cdot \|B_l D_n x\| \right)^2 \leq \sum_{l=1}^m \|u_l\|^2 \sum_{l=1}^m \|B_l D_n x\|^2. \end{aligned}$$

Wegen (5.3) bekommen wir hieraus $\lim_{n \rightarrow \infty} (u, D_n x) = 0$ für alle $u \in M_0^\perp$. Da $D_n x \in M_0^\perp$ und $\|D_n x\| \leq \|x\|$ ist, so ergibt sich schließlich $D_n x \rightarrow 0$ oder $R_n x \rightarrow Q_0 x$ für alle $x \in H$.

Wir wollen nun Satz 11 auf das im ersten Teil von Nr. 4 betrachtete Beispiel anwenden.

Der Operator $B_k = I - A_k$ ist der orthogonale Projektor von H auf $U_k^\perp \subset M_j^\perp$, wenn $i_k = j$ ist. Die Projektoren B_k mit $i_k = j$ konvergieren monoton wachsend gegen P_j . Um das Verfahren zu realisieren, könnten wir eine orthonormale Basis $\{\psi_\nu^{(j)}\}_{\nu=1,2,\dots}$

von $M_j^\perp$, d. h. ein H -orthonormales System von Vektoren $\psi \in H$, die in G_j Lösungen von $L\psi = 0$ sind, wählen und

$$(5.4) \quad U_{k_q}^\perp = \bigvee_{v=1}^q \psi_v^{(j)}$$

setzen, wenn $k_1 < k_2 < \dots < k_q < \dots$ die Teilfolge der natürlichen Zahlen mit $i_{k_q} = j$ bezeichnet. Dann definieren wir für einen beliebigen Vektor $g \in H$ die Folge $\{x_n\}$ durch

$$(5.5) \quad x_0 = 0, \quad x_n = x_{n-1} + B_n\{g - x_{n-1}\} = x_{n-1} + \sum_{v=1}^q B(g - x_{n-1}, \psi_v^{(j)}) \cdot \psi_v^{(j)},$$

wenn $n = k_q$ und $i_n = j$ ist und $B(x, y)$ die dem Differentialoperator zugeordnete Bilinearform bezeichnet. Bekanntlich liefert $B_n\{g - x_{n-1}\}$ gerade eine Trefftzsche Näherungslösung der Aufgabe (4.11), (4.12), vgl. [11] und [20]. Für diese Folge bekommen wir dann wegen Satz 11

$$x_n \rightharpoonup u = P_0 g \text{ in } H \text{ für } n \rightarrow \infty,$$

wobei u die Lösung von (4.13) ist. Die an Satz 8 anschließenden Bemerkungen über punktweise Konvergenz im Inneren von G_0 gelten wortwörtlich auch für unsere Folge (5.5).

Dieses Resultat dürfte besonders für numerische Zwecke wichtig sein.

6. Semidefinite stark elliptische Operatoren

In vielen wichtigen Fällen, z. B. bei der Potentialgleichung $\Delta u = 0$ und bei der biharmonischen Gleichung $\Delta^2 u = 0$, ist der Differentialoperator L zwar stark elliptisch, aber die assoziierte Bilinearform ist nur semidefinit. Im folgenden wollen wir untersuchen, wie das Neumannsche Verfahren zu modifizieren ist, um auch in diesen Fällen Konvergenz zu erhalten. Dazu betrachten wir die folgende Situation, die bei unseren Anwendungen erfüllt ist:

Sei $\mathfrak{H}$ ein reeller linearer Raum⁹⁾, auf dem zwei symmetrische Bilinearformen $(x, y)_0$ und $(x, y)_1$ mit den Eigenschaften

$$(6.1) \quad \|x\|_0^2 := (x, x)_0 > 0, \quad \|x\|_1^2 := (x, x)_1 \geq 0 \text{ für } x \in \mathfrak{H}, x \neq 0$$

definiert sind. Dann bildet

$$(6.2) \quad (x, y) := (x, y)_0 + (x, y)_1$$

ein Skalarprodukt auf $\mathfrak{H}$ mit der zugeordneten Norm

$$(6.3) \quad \|x\| = (x, x)^{\frac{1}{2}} = \{\|x\|_0^2 + \|x\|_1^2\}^{\frac{1}{2}}.$$

Wir setzen jetzt folgendes voraus:

- (i) $\mathfrak{H}$ ist bezüglich der Norm $\|x\|$ vollständig und bildet daher einen Hilbertraum mit dem Skalarprodukt (x, y) .
- (ii) $\|x\|_0^2$ ist eine vollstetige quadratische Form auf dem Hilbertraum $\mathfrak{H}$ mit der Norm $\|x\|$.
- (iii) Wir betrachten eine Familie $\{\mathfrak{U}_i\}_{i \in J}$ von endlich oder abzählbar unendlich vielen Unterräumen $\mathfrak{U}_i$ von $\mathfrak{H}$ und bilden $\mathfrak{U}_0 = \bigcap_{i \in J} \mathfrak{U}_i$, $\mathfrak{U} = \bigvee_{i \in J} \mathfrak{U}_i$, also $\mathfrak{U}_0 \subseteq \mathfrak{U}_i \subseteq \mathfrak{U}$.

⁹⁾ Hier wie auch in Nr. 4 und 5 gelten analoge Ergebnisse auch im komplexen Fall.

Dann gelte

$$(6.4) \quad k \cdot \|u\|_0^2 \leq \|u\|_1^2 \text{ für alle } u \in \mathfrak{U}, k > 0.$$

Weiterhin denken wir uns auf $\mathfrak{H}$ eine beschränkte, symmetrische, nichtnegative und elliptische Bilinearform $B(x, y)$ definiert, d. h. es gilt auf $\mathfrak{H}$:

- (a) $B(x, y) = B(y, x)$ (Symmetrie)
- (b) $|B(x, y)| \leq K \cdot \|x\| \cdot \|y\|$ (Beschränktheit)
- (c) $0 \leq B(x) := B(x, x)$ (Semidefinitheit)
- (d) $B(x) \geq c \cdot \|x\|^2 - c_0 \cdot \|x\|_0^2$ (Elliptizität).

Offensichtlich folgen (b), (c), (d) aus der schärferen Voraussetzung

$$(b') \quad c_1 \cdot \|x\|_1^2 \leq B(x) \leq c_2 \cdot \|x\|^2; \quad c_1, c_2 > 0.$$

Um die folgenden Betrachtungen zu vereinfachen, denken wir uns sogar (a), (b') vorausgesetzt, obwohl alles auch funktionierte, wenn wir nur (a) bis (d) annahmen und dafür (iii) geeignet modifizierten.

Der Nullraum $\mathfrak{N}$ von $B(x, y)$ läßt sich folgendermaßen charakterisieren:

$$(6.5) \quad \mathfrak{N} := \{x: B(x) = 0\} = \{x: B(x, y) = 0 \text{ für alle } y \in \mathfrak{H}\}.$$

Aus dem Spektralsatz folgt, daß sich $\mathfrak{H}$ eindeutig in der Form

$$(6.6) \quad \mathfrak{H} = \mathfrak{N} \oplus \mathfrak{D}$$

zerlegen läßt, wobei die Unterräume $\mathfrak{N}$ und $\mathfrak{D}$ nicht nur bezüglich (x, y) , sondern auch bezüglich $B(x, y)$ orthogonal sind:

$$(6.7) \quad (x, y) = 0, \quad B(x, y) = 0 \text{ für } x \in \mathfrak{N}, \quad y \in \mathfrak{D}.$$

Ferner gilt

$$(6.8) \quad B(x) > 0 \text{ für } x \in \mathfrak{D}, \quad x \neq 0.$$

Hieraus folgt wegen (ii), (c) und (d) in bekannter Weise

$$(6.9) \quad B(x) \geq \gamma \cdot \|x\|^2 \text{ für } x \in \mathfrak{D}, \quad \gamma > 0,$$

und

$$(6.10) \quad \dim \mathfrak{N} = \varrho < \infty.$$

Bezeichne $\psi_1, \psi_2, \dots, \psi_\varrho$ eine $(\cdot, \cdot)$ -normierte Basis von $\mathfrak{N}$. Dann definieren wir den Unter-
raum

$$(6.11) \quad M := \{x = u - \Pi u, u \in \mathfrak{U}\},$$

wobei

$$(6.12) \quad \Pi u = \sum_{j=1}^{\varrho} (u, \psi_j) \cdot \psi_j$$

die Orthogonalprojektion von u auf $\mathfrak{N}$ ist.

Offensichtlich ist M eine lineare Mannigfaltigkeit, so daß wir noch die Abgeschlossenheit von M zu beweisen haben¹⁰⁾. Aus (6.3) und (6.4) folgt zunächst

$$(6.13) \quad \|u\|_0 \leq q \cdot \|u\| \text{ für } u \in \mathfrak{U} \text{ mit } 0 < q = (1 + k)^{-\frac{1}{2}} < 1.$$

¹⁰⁾ Im allgemeinen braucht nämlich die Orthogonalprojektion eines Unterraumes auf einen anderen Unter-
raum nicht abgeschlossen zu sein.

Andererseits haben wir analog zu (6. 5)

$$(6. 14) \quad \mathfrak{N}_1 := \{x : \|x\|_1 = 0\} = \{x : (x, y)_1 = 0 \text{ für alle } y \in \mathfrak{H}\},$$

und wegen (b') ist $\mathfrak{N} \subseteq \mathfrak{N}_1$, somit gilt

$$(6. 15) \quad (u, x)_1 = 0 \text{ für } x \in \mathfrak{N}.$$

Aus (6. 2), (6. 13) und (6. 15) erhalten wir dann

$$\frac{|(u, x)|}{\|u\| \cdot \|x\|} = \frac{|(u, x)_0|}{\|u\| \cdot \|x\|_0} \leq \frac{\|u\|_0 \cdot \|x\|_0}{\|u\| \cdot \|x\|_0} \leq q < 1 \text{ für } u \in \mathfrak{U}, x \in \mathfrak{N},$$

also

$$(6. 16) \quad |(u, x)| \leq q \cdot \|u\| \cdot \|x\| \text{ für alle } u \in \mathfrak{U}, x \in \mathfrak{N}.$$

Setzen wir nun $x = \Pi u$ für ein beliebiges $u \in \mathfrak{U}$, so folgt aus (6. 16)

$$\|\Pi u\|^2 = (u, \Pi u) \leq q \cdot \|u\| \cdot \|\Pi u\|$$

oder

$$(6. 17) \quad \|\Pi u\| \leq q \cdot \|u\| \text{ für } u \in \mathfrak{U}$$

und daher

$$(6. 18) \quad (1 - q) \cdot \|u\| \leq \|u - \Pi u\| \leq (1 + q) \cdot \|u\| \text{ für } u \in \mathfrak{U}, 0 < q < 1.$$

Also ist M das normisomorphe Bild des Unterraumes $\mathfrak{U}$ von $\mathfrak{H}$ und somit abgeschlossen.

Definieren wir analog

$$(6. 19) \quad M_i = \{x = u - \Pi u, u \in \mathfrak{U}_i\}, \quad M_0 = \{x = u - \Pi u, u \in \mathfrak{U}_0\},$$

so sind wegen $\mathfrak{U}_0 \subseteq \mathfrak{U}_i \subseteq \mathfrak{U}$ auch M_0 und M_i Unterräume mit

$$(6. 20) \quad M_0 \subseteq M_i \subseteq M \subseteq \mathfrak{D}, \quad M_0 = \bigcap_{i \in J} M_i, \quad M = \bigvee_{i \in J} M_i.$$

Wegen (b) und (6. 9) sind $\|x\|$ und $\sqrt{B(x)}$ auf $\mathfrak{D}$ äquivalente Normen. Wir verstehen unter H den Hilbertraum $\mathfrak{D}$ mit dem Skalarprodukt $B(x, y)$. Dann können wir auf M_0, M_i, M, H unsere Resultate von Nr. 2 und 3 anwenden. Bezeichnen wir die identische Abbildung auf H mit Δ , die B -orthogonale Projektion von H auf M_0, M_i, M mit Q_0, Q_i, Q und mit $P_0 = \Delta - Q_0, P_i = \Delta - Q_i, P = \Delta - Q$ die B -orthogonale Projektion von H auf die B -Orthogonalkomplemente $H \ominus M_0 \supseteq H \ominus M_i \supseteq H \ominus M$, so folgt für eine zulässige Indexfolge $\{i_n\}$ mit $i_n \in J$ und $R_n = Q_{i_n} Q_{i_{n-1}} \cdots Q_{i_1}$

$$R_n \xrightarrow{H} Q_0, \quad R_n \xrightarrow{H} Q_0 \text{ oder } R_n \xrightarrow{H} Q_0.$$

Wegen der Äquivalenz von $\sqrt{B(x)}$ und $\|x\|$ erhalten wir daher

$$(6. 21) \quad \Delta - (\Delta - P_{i_n}) (\Delta - P_{i_{n-1}}) \cdots (\Delta - P_{i_1}) \xrightarrow{\mathfrak{D}}, \xrightarrow{\mathfrak{D}} \text{ oder } \xrightarrow{\mathfrak{D}} P_0.$$

Wir definieren nun auf $\mathfrak{H}$ die Operatoren $T_i, i \in J$, indem wir für ein beliebiges $y \in \mathfrak{H}$

$$(6. 22) \quad T_i y = y - u_i$$

setzen, wobei $u_i \in \mathfrak{U}_i$ als Lösung der Aufgabe

$$(6. 23) \quad B(y - u_i, \varphi) = 0 \text{ für alle } \varphi \in \mathfrak{U}_i$$

gewählt wird. Da $B(\varphi) \geq K \cdot \|\varphi\|^2, K > 0$, für alle $\varphi \in \mathfrak{U}$ ist wegen (b') und (6. 4), so gibt es nach dem Satz von Fréchet-Riesz genau eine Lösung u_i von (6. 23). Daher ist

T_i wohldefiniert. Analog definieren wir T_0 und T , indem wir oben u_i durch u_0 bzw. u und $\mathfrak{U}_i$ durch $\mathfrak{U}_0$ bzw. $\mathfrak{U}$ ersetzen.

Wir bezeichnen jetzt mit $\Delta = I - \Pi$ die Orthogonalprojektion von $\mathfrak{S}$ auf $\mathfrak{D}$. Da die Restriktion von Δ auf $\mathfrak{D}$ gerade die Identität ist, erhalten wir keinen Widerspruch zu unserer früheren Definition von Δ .

Andererseits ist für ein beliebiges Element $y \in \mathfrak{D}$ die Projektion $P_i y = y - m_i$ mit $m_i \in M_i$ eindeutig charakterisiert durch die Gleichung

$$(6.24) \quad B(y - m_i, \varphi) = 0 \text{ für alle } \varphi \in M_i.$$

Da $B(\varphi, \psi) = B(\varphi + x, \psi) = B(\varphi, \psi + x)$ für $x \in \mathfrak{N}$ ist, folgt aus (6.24)

$$(6.25) \quad B(y - m_i, \varphi) = 0 \text{ für alle } \varphi \in \mathfrak{U}_i,$$

da M_i und $\mathfrak{U}_i$ eineindeutig aufeinander bezogen sind. Vergleichen wir (6.23) und (6.25), so folgt aus der eindeutigen Lösbarkeit dieser Gleichungen $m_i = u_i - \Pi u_i = \Delta u_i$ und daher wegen $\Delta y = y$ für $y \in \mathfrak{D}$:

$$P_i y = y - m_i = \Delta y - \Delta u_i = \Delta(y - u_i) = \Delta T_i y \text{ für } y \in \mathfrak{D}.$$

Wenn wir noch

$$(6.26) \quad P_j x = P x = P_0 x = 0 \text{ für } x \in \mathfrak{N}$$

setzen, so folgt $P_i = \Delta T_i \Delta$, und entsprechendes gilt für P_0 und P , also:

$$(6.27) \quad P_0 = \Delta T_0 \Delta, \quad P_i = \Delta T_i \Delta, \quad P = \Delta T \Delta.$$

Somit ergibt sich aus (6.21)

$$(6.28) \quad \Delta - (\Delta - \Delta T_{i_n} \Delta) (\Delta - \Delta T_{i_{n-1}} \Delta) \cdots (\Delta - \Delta T_{i_1} \Delta) \xrightarrow{\mathfrak{D}}, \xrightarrow{\mathfrak{D}} \text{ oder } \xrightarrow{\mathfrak{D}} \Delta T_0 \Delta,$$

oder

$$(6.29) \quad \{I - (I - \Delta T_{i_n}) (I - \Delta T_{i_{n-1}}) \cdots (I - \Delta T_{i_1})\} \Delta \xrightarrow{\mathfrak{S}}, \xrightarrow{\mathfrak{S}} \text{ oder } \xrightarrow{\mathfrak{S}} \Delta T_0 \Delta.$$

Konstruieren wir also für ein beliebiges Element $g \in \mathfrak{S}$ die Folge $\{x_n\}$ durch die Vorschrift

$$(6.30) \quad \hat{g} = \Delta g, \quad x_1 = \Delta T_{i_1} \hat{g}, \quad x_n = x_{n-1} + \Delta T_{i_n} \{\hat{g} - x_{n-1}\},$$

so folgt $x_n \xrightarrow{\mathfrak{S}}, \xrightarrow{\mathfrak{S}} \text{ oder } \xrightarrow{\mathfrak{S}} \hat{w} = \Delta T_0 \Delta g = \Delta T_0 \hat{g} = P_0 \hat{g}$. Es ist leicht zu sehen, daß sich die gesuchte Lösung $w = T_0 g = g - u_0$, $u_0 \in \mathfrak{U}_0$ der „Randwertaufgabe“

$$(6.31) \quad B(g - u_0, \varphi) = 0 \text{ für alle } \varphi \in \mathfrak{U}_0$$

von dem Grenzelement $\hat{w}$ nur um ein Element $\hat{n} \in \mathfrak{N}$ unterscheidet:

$$(6.32) \quad w = \hat{w} + \hat{n}, \quad \hat{w} = \lim_{n \rightarrow \infty} x_n, \quad \hat{n} \in \mathfrak{N}.$$

Fassen wir unsere Ergebnisse zusammen:

Satz 12. *Sei $\mathfrak{S}$ ein Hilbertraum und $\{\mathfrak{U}_i\}_{i \in J}$ eine höchstens abzählbar unendliche Familie von Teilräumen von $\mathfrak{S}$, die (i) bis (iii) genügen. Auf $\mathfrak{S}$ sei eine Bilinearform $B(x, y)$ mit den Eigenschaften (a), (b') definiert. Ferner sei eine zulässige Folge von Indizes $i_n \in J$ gegeben.*

Ist dann $\hat{g}$ für ein beliebiges Element $g \in \mathfrak{S}$ die Orthogonalprojektion Δg von g auf das Komplement $\mathfrak{D} = \mathfrak{S} \ominus \mathfrak{N}$ des Nullraumes $\mathfrak{N}$ von $B(x, y)$ in $\mathfrak{S}$ und definieren wir eine

Folge $\{x_n\}$ durch

$$x_0 = 0, \quad x_n = x_{n-1} + \Delta T_{i_n} \{\hat{g} - x_{n-1}\},$$

wobei $T_{i_n} \{\hat{g} - x_{n-1}\}$ die eindeutig bestimmte Lösung $\hat{g} - x_{n-1} - u_{i_n}$, $u_{i_n} \in \mathfrak{U}_{i_n}$, der „abstrakten Randwertaufgabe“

$$(6.33) \quad B(\hat{g} - x_{n-1} - u_{i_n}, \varphi) = 0 \text{ für alle } \varphi \in \mathfrak{U}_{i_n}$$

bezeichnet, so gilt

$$x_n \xrightarrow[\mathfrak{S}]{} \hat{w} \text{ oder } \xrightarrow[\mathfrak{S}]{} \hat{w}$$

wobei sich $\hat{w}$ nur um ein Element $\hat{n} \in \mathfrak{N}$ von der Lösung $w = g - u_0$, $u_0 \in \mathfrak{U}_0$, der „Randwertaufgabe“

$$(6.31) \quad B(g - u_0, \varphi) = 0 \text{ für alle } \varphi \in \mathfrak{U}_0$$

unterscheidet:

$$(6.32) \quad w = \hat{w} + \hat{n}, \quad \hat{n} \in \mathfrak{N}.$$

Da $\dim \mathfrak{N} = \varrho < \infty$ und somit $\hat{n} = \sum_{\nu=1}^{\varrho} \hat{n}^{\nu} \cdot \psi_{\nu}$ ist, führt die Bestimmung von $\hat{n}$ auf ein Problem mit endlich vielen Parametern $\hat{n}^{\nu}$, die sich im allgemeinen leicht aus den „Randbedingungen“ ermitteln lassen, falls man eine Basis $\psi_1, \psi_2, \dots, \psi_{\varrho}$ von $\mathfrak{N}$ kennt.

Greifen wir auf das erste Beispiel von Nr. 4 zurück, erlauben jetzt aber, daß die dem Differentialoperator L zugeordnete Bilinearform anstelle der Definitheit auf $H_p(G)$ den schwächeren Bedingungen (a), (b') genügt. Dabei haben wir $\mathfrak{S}$ mit $H_p(G)$, $\mathfrak{U}_i$ mit $\dot{H}_p(G_i)$, $\mathfrak{U}_0$ mit $\dot{H}_p(G_0)$ und $\mathfrak{U}$ mit $\dot{H}_p(G)$ zu identifizieren. Dann sind (6.31) und (6.33) in der Tat Randwertaufgaben für den Differentialoperator L .

Wir identifizieren $(x, y)_0$ mit $\int_G x \cdot y dt$ und $(x, y)_1$ mit $\int_G \sum_{|\alpha|=p} D_{\alpha} x \cdot D_{\alpha} y dt$. Dann ist (i) leicht zu sehen, (ii) folgt aus dem Satz von Rellich, und (iii) ist eine wohlbekannte Ungleichung.

Die Bedingungen (a), (b') sind für viele Probleme, z. B. für $\Delta u = 0$, $\Delta^2 u = 0, \dots$, $\Delta^p u = 0$ unter Dirichletschen Randbedingungen und für die Gleichungen der linearen Elastizitätstheorie erfüllt¹¹⁾.

7. Das alternierende Verfahren von H. A. Schwarz

Wie schon eingangs bemerkt, ist die Anwendung von Variations- bzw. Projektionsmethoden auf das Schwarzsche alternierende Verfahren in einer Reihe von Arbeiten eingehend untersucht worden. In [7] ergab sich als wesentlich neues Resultat, daß es im Falle von endlich vielen Gebieten $G_1, G_2, \dots, G_N$ möglich ist, zu einer Fehlerabschätzung für das alternierende Verfahren zu gelangen. Allerdings war diese Abschätzung an die Möglichkeit geknüpft, auf $G = \bigcup_{i=1}^N G_i$ eine Zerlegung der Eins: $\sum_{i=1}^N \eta_i(t) = 1$ auf G , $\eta_i(t) \in C^{\infty}(\bar{G})$, $\eta_i = 0$ in $G - G_i$, konstruieren zu können, für die

$$|D_{\alpha} \eta_i(t)| \leq \text{const}, \quad i = 1, \dots, N, \quad |\alpha| \leq p, \quad t \in \bar{G}$$

¹¹⁾ Δ bezeichnet hier den d -dimensionalen Laplace-Operator.

gilt. Dies bedeutet, grob gesprochen, daß zwei Gebiete G_i und G_j mit nichtleerem Durchschnitt in den Teilen ihres Randes, der zu G gehört, glatt oder „gürtelförmig“ verschmolzen sein müssen. Daraus ergab sich eine wesentliche Einschränkung für die Möglichkeit, Fehlerschranken aufstellen zu können, da sich nichtkonvexe Gebiete G mit „Ecken“ im Rand im allgemeinen nicht in der gewünschten Weise aus konvexen Elementargebieten G_i aufbauen lassen. Beispielsweise können wir ein L - oder T -förmiges Gebiet nicht durch gürtelförmige Verschmelzung von Rechtecken herstellen.

In diesem Abschnitt wollen wir nun zeigen, daß sich die entscheidende Voraussetzung (2.4) von Satz 1, (v) der vorliegenden Arbeit beim alternierenden Verfahren schon unter ganz geringen Annahmen über die G_i nachweisen läßt, ohne daß es nötig wäre, die obige Zerlegung der Eins konstruieren zu können. Auf diese Weise erhalten wir in allen für die Anwendung interessanten Fällen Fehlerabschätzungen für das Schwarzsche Verfahren und damit gleichmäßige Konvergenz.

Wir benutzen jetzt die Bezeichnungen von Abschnitt 4. Wie dort seien $G_1, G_2, \dots, G_N$ und $G = \bigcup_{i=1}^N G_i$ Gebiete im E_d , L ein stark elliptischer Differentialoperator auf G von der Form (4.6) und $B(x, y)$ die zu L gehörige Bilinearform, die auf $H_p(G)$ stetig und symmetrisch sowie auf $\dot{H}_p(G)$ positiv definit ist:

$$(7.1) \quad B(x, y) = B(y, x) \text{ und } |B(x, y)| \leq c' \cdot \|x\|_{p,G} \cdot \|y\|_{p,G} \text{ für alle } x, y \in H_p(G),$$

$$(7.2) \quad B(x) = B(x, x) \geq c \cdot \|x\|_{p,G}^2 \text{ für alle } x \in \dot{H}_p(G), c > 0.$$

Dann schreiben wir die erste Randwertaufgabe

$$(7.3) \quad \begin{aligned} Lx &= g && \text{in } G \\ D_\alpha x &= D_\alpha f, \quad |\alpha| \leq p-1, && \text{auf } \partial G \end{aligned}$$

in der (unter geeigneten Voraussetzungen an L , f , g und G äquivalenten) Form:

Bestimme zu gegebenen $f \in H_p(G)$, $g \in H_0(G)$ ein $u \in \dot{H}_p(G)$, so daß $x = f + u$ der Relation

$$(7.4) \quad B(x, \varphi) = l(\varphi) \text{ für alle } \varphi \in \dot{H}_p(G) \text{ genügt,}$$

wobei $l(\varphi)$ das auf $H_p(G)$ stetige lineare Funktional $(g, \varphi)_{0,G}$ bezeichnet.

Bekanntlich hat die Aufgabe (7.4) unter den Voraussetzungen (7.1) und (7.2) genau eine Lösung $x = f + u$. Das alternierende Verfahren ist nun folgendermaßen definiert:

Wir betrachten eine zulässige Folge $\{i_n\}$ von Indizes $i_n \in J = \{1, 2, \dots, N\}$ und setzen $x_0 = f$. Darauf definieren wir x_n mittels x_{n-1} durch

$$(7.5) \quad x_n = \begin{cases} x_{n-1} & \text{in } G - G_{i_n} \\ z_n & \text{in } G_{i_n} \end{cases} \in H_p(G),$$

wobei $z_n \in H_p(G_{i_n})$ die Lösung der Randwertaufgabe

$$(7.6) \quad \begin{aligned} Lz_n &= g && \text{in } G_{i_n} \\ D_\alpha z_n &= D_\alpha x_{n-1}, \quad |\alpha| \leq p-1, && \text{auf } \partial G_{i_n} \end{aligned}$$

bezeichnet. Mit anderen Worten:

Wir bestimmen $x_n = x_{n-1} + u_n$ mit $u_n \in \dot{H}_p(G_{i_n})$ als Lösung von

$$(7.7) \quad B(x_n, \varphi) = l(\varphi) \quad \text{für alle } \varphi \in \dot{H}_p(G_{i_n}).$$

Setzen wir

$$(7.8) \quad y_n = x_n - x,$$

so ergeben sich die Relationen

$$(7.9) \quad \begin{aligned} B(y_n, \varphi) &= 0 && \text{für alle } \varphi \in \dot{H}_p(G_{i_n}) \\ y_n &= y_{n-1} + u_n && u_n \in \dot{H}_p(G_{i_n}). \end{aligned}$$

Wir nehmen nun $H = \dot{H}_p(G)$ als zugrundeliegenden Hilbertraum mit dem Skalarprodukt $(x, y) := B(x, y)$, dessen zugehörige Norm $\|x\| = \sqrt{B(x)}$ wegen (7.1) und (7.2) zu $\|x\|_{p,G}$ äquivalent ist. Ferner setzen wir $M_i^\perp = \dot{H}_p(G_i)$, $M_i = \dot{H}_p(G_i)^\perp$. Dann gilt, wie in [7] gezeigt wurde,

$$(7.10) \quad M_0^\perp = \bigvee_{i=1}^N M_i^\perp = \bigvee_{i=1}^N \dot{H}_p(G_i) = \dot{H}_p(G) = H$$

und daher

$$(7.11) \quad M_0 = \{0\} \quad \text{und} \quad Q_0 = 0.$$

Aus (7.9) folgt $y_n = Q_{i_n} y_{n-1}$ und daher mit $f - x = -u$ wegen Satz 1:

$$(7.12) \quad x_n - x = y_n = R_n(-u) \xrightarrow{H} Q_0(-u) = 0 \quad \text{für } n \rightarrow \infty.$$

Insbesondere folgt nach (v) von Satz 1: $\|R_n\| \rightarrow 0$ für $n \rightarrow \infty$, falls anstelle von (7.10) sogar gilt

$$(7.13) \quad \dot{H}_p(G) = \dot{H}_p(G_1) + \dot{H}_p(G_2) + \cdots + \dot{H}_p(G_N).$$

Aus (7.1), (7.2) und (7.4) folgt

$$c \|u\|_{p,G}^2 \leq B(u) \leq |l(u)| + |B(f, u)| \leq \{\|g\|_{0,G} + c' \cdot \|f\|_{p,G}\} \|u\|_{p,G}$$

und daher

$$\|u\|_{p,G} \leq K := c^{-1} \|g\|_{0,G} + \frac{c'}{c} \cdot \|f\|_{p,G}.$$

Damit bekommen wir die Fehlerabschätzung

$$(7.14) \quad \|x - x_n\|_{p,G} \leq K \cdot \|R_n\|, \quad K = K(f, g, c, c'), \quad \lim_{n \rightarrow \infty} \|R_n\| = 0.$$

Für die zulässige Indexfolge $1, 2, \dots, N, 1, 2, \dots, N, \dots$ folgt insbesondere

$$(7.15) \quad \|R_n\| \leq K \cdot q^{\left[\frac{n}{N}\right]}, \quad 0 < q < 1,$$

für eine geeignete Zahl q , die von c, c', n, p und der geometrischen Gestalt der Gebiete $G, G_1, \dots, G_N$ abhängt.

Wir skizzieren nun den Beweis von (7.13) und betrachten als erstes den Fall $N = 2: G = G_1 \cup G_2$. Sei $G'_1 = G - G_2$ und $S = \partial G_2 - \{\partial G \cap \partial G_2\}$. Ferner bezeichne $\eta \in C^\infty(\overline{G'_1})$ eine Funktion, die auf S gleich Eins ist und außerhalb eines δ -Streifens S_δ um S in G'_1 verschwindet, $\delta > 0$ hinreichend klein.

Wir nehmen nun eine Funktion $u \in \dot{H}_p(G)$ und bilden die Funktion

$$v(t) = \eta(t) \cdot u(t) \in H_p(G'_1),$$

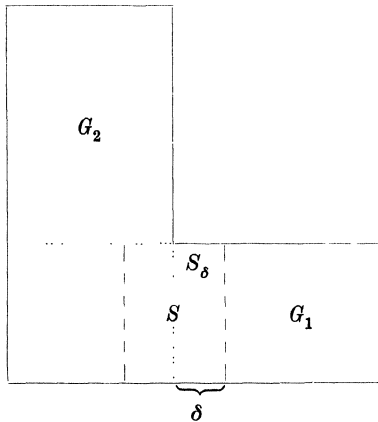
die außerhalb S_δ in G'_1 verschwindet. Wenn nun S hinreichend glatt und $\delta > 0$ genügend klein ist, so kann man $v(t)$ mittels Aufbiegen von S zu einer Hyperebene und mit einer bekannten Fortsetzungsformel von Lions zu einer Funktion $w(t)$ auf G_1 fortsetzen, die in $\dot{H}_p(G_1)$ liegt. Darauf bilden wir

$$u_1(t) = \begin{cases} u(t) & t \in G'_1 \\ w(t) & \text{für } t \in G_1 - G'_1 \\ 0 & t \in G - G_1. \end{cases}$$

Nach Konstruktion liegt $u_1(t)$ in $\dot{H}_p(G_1)$ und $u_2 := u - u_1$ liegt in $\dot{H}_p(G_2)$. Damit haben wir gezeigt, daß sich jedes $u \in \dot{H}_p(G)$ in der Form

$$u = u_1 + u_2 \text{ mit } u_i \in \dot{H}_p(G_i), \quad (i = 1, 2)$$

schreiben läßt, womit (7.13) für $N = 2$ bewiesen ist. Der allgemeine Fall läßt sich auf den eben behandelten durch Induktion zurückführen. Ferner wollen wir ohne Beweis bemerken, daß die Konstante q in (7.13) leicht abgeschätzt werden kann, sobald $\delta > 0$, $\eta(t)$, die Lionssche Fortsetzungsformel und die C^p -Norm für die Aufbiegetransformation von S bekannt sind.



Insbesondere können wir für $d = 2$ ein L -förmiges Gebiet G aus zwei Rechtecken G_1 und G_2 so zusammensetzen, daß die Voraussetzungen unserer Konstruktion erfüllt sind und somit

$$\dot{H}_p(G) = \dot{H}_p(G_1) + \dot{H}_p(G_2)$$

folgt. Damit erhalten wir also beispielsweise Fehlerabschätzungen für die Konvergenz des alternierenden Verfahrens für die polyharmonische Gleichung

$$\begin{aligned} \Delta^p x &= g & \text{in } G \\ \frac{\partial^v x}{\partial n^v} &= 0, \quad v = 0, 1, \dots, p-1, & \text{auf } \partial G, \end{aligned}$$

wobei n die Normale auf G bezeichnet, mit den beiden Rechtecken G_1 und G_2 als Iterationsgebieten.

Wir bemerken ferner, daß sich das alternierende Verfahren in naheliegender Weise auch auf andere Randwertprobleme anwenden läßt, sobald diese auf definite quadratische Variationsprobleme führen.

Eine weitere Anwendung ist die Konstruktion harmonischer Differentialformen mit vorgeschriebenen Perioden auf Riemannschen Mannigfaltigkeiten, worauf wir an anderer Stelle eingehen werden.

Literaturverzeichnis

- [1] *I. Anemiyi* and *T. Ando*, Convergence of random products of contractions in Hilbert space, *Acta Sci. Math.* (Szeged) **26** (1965), 239—244.
- [2] *F. E. Browder*, On some approximation methods for solutions of the Dirichletproblem for linear elliptic equations of arbitrary order, *J. Math. Mech.* **7** (1958), 69—80.
- [3] *K. O. Friedrichs*, On the differentiability of solutions of linear elliptic differential equations, *Commun. Pure and Appl. Math.* **6** (1953), 299—325.
- [4] *I. C. Gochberg* und *A. S. Markus*, Zwei Sätze über die Öffnung von Unterräumen eines Banachraumes, *Uspechi Mat. Nauk* **14** (89) (1959), 135—140. (In Russisch.)
- [5] *P. R. Halmos*, Introduction to Hilbert space and the theory of spectral multiplicity, New York 1957.
- [6] *I. Halperin*, The product of projection operators, *Acta Sci. Math.* (Szeged) **23** (1962), 96—99.
- [7] *S. Hildebrandt*, Einige konstruktive Methoden bei Randwertaufgaben für lineare partielle Differentialgleichungen und in der Theorie harmonischer Differentialformen. I, *J. f. reine u. angew. Math.* **213** (1963), 66—88.
- [8] *S. Hildebrandt* und *B. Schmidt*, Zur Konvergenz von Operatorprodukten im Hilbertraum, *Math. Z.* **105** (1968), 62—71.
- [9] *T. Kato*, Perturbation theory for linear operators, Berlin-Heidelberg-New York 1966.
- [10] *L. Lichtenstein*, Enzyklopädie Math. Wiss. II C 3. Neuere Entwicklung der Potentialtheorie, Konforme Abbildung (vgl. insbesondere Nr. 24, S. 257).
- [11] *K. Maurin*, Bemerkungen über die Methoden von Trefftz und Ritz, *Bull. Acad. Polonaise Sci., Cl. III*, **3**, Nr. 11 (1955), 573—577.
- [12] *C. Neumann*, Zur Theorie des logarithmischen und Newtonschen Potentials, Berichte über die Verhandlungen der Königl. Sächs. Gesellschaft der Wiss. Leipzig, Math.-Phys. Klasse, Band **22** (1870), Zweite Mitteilung, 264—321.
- [13] *J. v. Neumann*, Functional Operators, Vol. II, *Annals of Math. Studies*, Nr. 22 (1952), Princeton.
- [14] *L. Nirenberg*, Remarks on strongly elliptic partial differential equations, *Comm. Pure Appl. Math.* **8** (1955), 643—674.
- [15] *W. V. Petryshyn*, Referat Nr. 3840 in *Math. Reviews* **31**, Nr. 4 (1966) über die Arbeit von Stummel [19].
- [16] *A. Pfluger*, Theorie der Riemannschen Flächen, Berlin-Göttingen-Heidelberg 1957.
- [17] *B. Schmidt*, Zur Konvergenz unendlicher Produkte von Operatoren im Hilbertraum, Diplom-Arbeit, Mainz 1966.
- [18] *S. L. Sobolew*, L'algorithme de Schwarz dans la théorie de l'élasticité, *Doklady Akad. Nauk SSSR* **4** (13) (1936), 243—246.
- [19] *F. Stummel*, Zur Konvergenz des Balayage-Verfahrens in Hilbertschen Räumen, *Math. Z.* **86** (1964), 136—144.
- [20] *F. Stummel*, Näherungsmethoden zur Lösung elliptischer partieller Differentialgleichungen, Vortrag auf der DMV-Tagung 1963, Berichte des Hahn-Meitner-Instituts (1963), HMI-B 31.
- [21] *N. Wiener* und *P. Masani*, The prediction theory of multivariate stochastic processes. I, II, *Acta Math.* **98** (1957), 111—150; **99** (1958), 93—137.

Mathematisches Institut der Universität, 65 Mainz

Eingegangen 26. Februar 1968

Periodische Lösungen von nichtlinearen elliptischen Differentialgleichungen

Von *Ulrich Staude* in Mainz

Herrn Professor Dr. H. Rohrbach zum 65. Geburtstag gewidmet

In einer früheren Arbeit [4] habe ich periodische Lösungen in der Nachbarschaft von $u(x, t) \equiv 0$ einer nichtlinearen elliptischen Differentialgleichung mit der Methode der unendlich vielen Veränderlichen untersucht. Dabei wurden die Lösungen aus der Randwertaufgabe eines nichtlinearen Systems von abzählbar unendlich vielen gewöhnlichen Differentialgleichungen bestimmt. Schwierigkeiten ergaben sich dabei bei den Abschätzungen der Greenschen Funktionen, was eine Übertragung auf kompliziertere Gleichungen oder Systeme als kaum durchführbar erscheinen ließ.

Es lag daher nahe, die modernen Methoden der Behandlung von elliptischen Differentialgleichungen (siehe etwa S. Agmon [1], S. Hildebrandt [2]) zur Lösung des Problems heranzuziehen. In dieser Untersuchung ist dies nun für einen etwas allgemeineren linearen Teil als in [4] durchgeführt. Es war möglich, kompliziertere Nichtlinearitäten, bei denen auch die zweiten Ableitungen $\frac{\partial^2 u}{\partial x \partial t}$ und $\frac{\partial^2 u}{\partial t^2}$ auftreten, zu behandeln, was mit der Methode in [4] nicht möglich war. Genauso läßt sich der Fall untersuchen, daß bei den nichtlinearen Termen $u \frac{\partial^2 u}{\partial x^2}$ auftritt, dazu wäre es aber notwendig gewesen, diese Glieder zeitweise gesondert zu betrachten. Jedoch gelang es nicht, mit dieser Methode Ausdrücke der Gestalt $\frac{\partial^2 u}{\partial x^2} D^1 u$ oder $\frac{\partial^2 u}{\partial x^2} D^2 u$ zu behandeln, wobei $D^1 u$ eine erste Ableitung, $D^2 u$ eine zweite Ableitung von u bedeutet.

Die Forderung, daß die Koeffizienten des linearen Teils nicht von t abhängen und in x analytisch sein sollen sowie die Analytizität der Koeffizienten des nichtlinearen Teils, wurde nur gestellt, um die Darstellung zu vereinfachen. Dies geschah ebenso bei der Beschränkung darauf, daß die Nichtlinearitäten in u und seinen Ableitungen homogen von 2. Ordnung vorausgesetzt sind.

Die Übertragung auf stark elliptische Systeme, wobei dann x ein Vektor in einem n -dimensionalen Gebiet ist, soll einer späteren Arbeit vorbehalten bleiben.

1. Problemstellung

Wir suchen Lösungen $u(x, t)$ der partiellen Differentialgleichung

$$(1.1') \quad \frac{\partial}{\partial x} \left\{ a_{11}(x; \eta) \frac{\partial u}{\partial x} \right\} + a_{22}(x; \eta) \frac{\partial^2 u}{\partial t^2} + a_1(x; \eta) \frac{\partial u}{\partial x} + a(x; \eta) u \\ = \eta f_0(x, t; \eta) + \sum_{\substack{j, k, l, m, n=0 \\ j+k+l+m+n=2}}^2 a_{jklmn}(x, t; \eta) u^j \left(\frac{\partial u}{\partial x} \right)^k \left(\frac{\partial u}{\partial t} \right)^l \left(\frac{\partial^2 u}{\partial x \partial t} \right)^m \left(\frac{\partial^2 u}{\partial t^2} \right)^n,$$

die die Bedingungen

$$(1.2) \quad u(0, t) = u(1, t) = 0,$$

$$(1.3') \quad u(x, t) = u(x, t + T), \quad T = \frac{2\pi}{\lambda}, \quad \lambda \neq 0$$

erfüllen.

Dabei soll die Periode T entweder durch die in t mit derselben Periode T periodischen Funktionen auf der rechten Seite erzwungen werden oder aber sich frei einstellen, falls die rechte Seite nicht explizit von t abhängt. Im folgenden werden wir beide Fälle zusammen behandeln.

Zur Vereinfachung reduzieren wir die Periode durch die Transformation

$$(1.4) \quad \tau = \lambda t$$

auf die Periode 2π in τ . Mit $\frac{\partial}{\partial t} = \lambda \frac{\partial}{\partial \tau}$ und $\frac{\partial^2}{\partial t^2} = \lambda^2 \frac{\partial^2}{\partial \tau^2}$ erhalten wir, wenn wir dann wieder t statt τ schreiben:

$$(1.1) \quad L(x, t; \lambda, \eta) u := \frac{\partial}{\partial x} \left\{ a_{11}(x; \eta) \frac{\partial u}{\partial x} \right\} + \lambda^2 a_{22}(x; \eta) \frac{\partial^2 u}{\partial t^2} + a_1(x; \eta) \frac{\partial u}{\partial x} + a(x; \eta) u \\ = \eta f_0(x, t; \eta) + \sum_{\substack{j, k, l, m, n=0 \\ j+k+l+m+n=2}}^2 \lambda^{l+m+2n} a_{jklmn}(x, t; \eta) u^j \left(\frac{\partial u}{\partial x} \right)^k \left(\frac{\partial u}{\partial t} \right)^l \left(\frac{\partial^2 u}{\partial x \partial t} \right)^m \left(\frac{\partial^2 u}{\partial t^2} \right)^n \\ = \eta f_0(x, t; \eta) + f(x, t, u, D^1 u, D^2 u; \lambda, \eta) = g(x, t, u, D^1 u, D^2 u; \lambda, \eta).$$

Je nachdem wie wir es benötigen werden, werden wir die Argumente in den Funktionen angeben.

Wir bezeichnen mit

$$\tilde{G}: \left\{ \begin{array}{l} x \in [0, 1] \\ -\infty < t < +\infty \end{array} \right\}, \quad G: \left\{ \begin{array}{l} x \in [0, 1] \\ t \in [0, 2\pi] \end{array} \right\}.$$

η sei ein kleiner Parameter. Alle Funktionen seien bezüglich x, t, η für $(x, t) \in G$ und

$$(1.5) \quad |\eta| \leq K_0$$

analytisch. Funktionen, die explizit von t abhängen, sollen in t mit der Periode 2π periodisch sein.

Der Operator L soll für alle η , die (1.5) erfüllen, für alle $(x, t) \in \tilde{G}$ und $\lambda > 0$ elliptisch sein. Dafür können wir ohne Beschränkung der Allgemeinheit annehmen, daß

$$(1.6) \quad a_{11}(x; \eta) \geq a_{11} > 0, \quad a_{22}(x; \eta) \geq a_{22} > 0$$

für $x \in [0, 1]$ und $|\eta| \leq K_0$.

Nach (1. 2), (1. 3') und (1. 4) suchen wir somit Lösungen $u(x, t)$ der Differentialgleichung (1. 1), die die Bedingungen

$$(1. 2) \quad u(0, t) = u(1, t) = 0,$$

$$(1. 3) \quad u(x, t) = u(x, t + 2\pi)$$

erfüllen. Wir wollen solche Lösungen in der Nachbarschaft von $u(x, t) \equiv 0$ in Abhängigkeit von dem Parameter η bestimmen. Ist (1. 1) von t unabhängig, so wird sich λ , also damit die Periode T der ursprünglichen Gleichung (1. 1') nach (1. 3') in Abhängigkeit von η ergeben.

Für spätere Rechnungen führen wir noch folgende Bezeichnungen ein:

$$(1. 7) \quad L^0 := \frac{\partial}{\partial x} \left\{ a_{11}(x; \eta) \frac{\partial u}{\partial x} \right\} + \lambda^2 a_{22}(x; \eta) \frac{\partial^2 u}{\partial t^2},$$

$$(1. 8) \quad L^1 := L - L^0 = a_1(x; \eta) \frac{\partial u}{\partial x} + a(x; \eta) u,$$

$$(1. 9) \quad L_0 := L(\lambda^0; \eta = 0) = \frac{\partial}{\partial x} \left\{ a_{11}(x; 0) \frac{\partial u}{\partial x} \right\} + \lambda^{0^2} a_{22}(x; 0) \frac{\partial^2 u}{\partial t^2} \\ + a_1(x; 0) \frac{\partial u}{\partial x} + a(x; 0) u,$$

$$(1. 10) \quad L_1 = L_1(\lambda^0, 0; \varepsilon, \eta) := L(\lambda^0 + \varepsilon, \eta) - L_0(\lambda^0, 0).$$

Hieraus ergibt sich, was unter L_0^0 und L_0^1 zu verstehen ist. L^0 ist der elliptische Hauptteil von L . L_0 ist das konstante Glied der Entwicklung von L nach ε und η an der Stelle $\lambda^0, \eta = 0$; L_1 ist der Rest, in dem somit alle Glieder ε oder η als Faktor haben.

2. Die lineare Aufgabe

Die Funktionen $u(x, t)$, die die Bedingungen (1. 2) und (1. 3)

$$(2. 1) \quad u(0, t) = u(1, t) = 0$$

$$(2. 2) \quad u(x, t) = u(x, t + 2\pi)$$

erfüllen und nach beiden Veränderlichen in G beliebig oft differenzierbar sind, bilden den linearen Raum $C_{0, 2\pi}^\infty$. In diesem Raum führen wir die beiden Skalarprodukte

$$(2. 3) \quad (u, v)_0 = \int_0^{2\pi} \int_0^1 u(x, t) v(x, t) dx dt$$

und

$$(2. 4) \quad (u, v)_1 = (u, v)_0 + \left(\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x} \right)_0 + \left(\frac{\partial u}{\partial t}, \frac{\partial v}{\partial t} \right)_0$$

ein. Bezüglich der sich aus diesen Skalarprodukten ergebenden Normen

$$(2. 5) \quad \|u\|_0^2 = (u, u)_0$$

und

$$(2. 6) \quad \|u\|_1^2 = (u, u)_1$$

vervollständigen wir $C_{0, 2\pi}^\infty$ zu den Hilberträumen H_0 und H_1^0 .

Mit dem in (1. 1) definierten Operator L mit $\lambda > 0$, der auf einem in H_0 dichten Teilraum definiert ist, bilden wir die Bilinearform

$$(2. 7) \quad B(u, \chi) = (Lu, \chi)_0$$

für alle $\chi \in C_{0,2\pi}^\infty$.

Explizit ergibt sich

$$(2. 8) \quad B(u, \chi) = (Lu, \chi)_0 = - \left(a_{11}(x; \eta) \frac{\partial u}{\partial x}, \frac{\partial \chi}{\partial x} \right)_0 - \lambda^2 \left(a_{22}(x; \eta) \frac{\partial u}{\partial t}, \frac{\partial \chi}{\partial t} \right)_0 \\ + \left(a_1(x; \eta) \frac{\partial u}{\partial x}, \chi \right)_0 + (a(x; \eta) u, \chi)_0.$$

Somit ist $B(u, \chi)$ eine bezüglich der Norm $\| \cdot \|_1$ beschränkte Bilinearform:

$$(2. 9) \quad | B(u, \chi) | = | (Lu, \chi)_0 | \leq K \| u \|_1 \| \chi \|_1.$$

Dann läßt sich $B(u, \chi)$ zu einer beschränkten Bilinearform mit derselben Schranke K auf alle $u \in H_1^0$, $\chi \in H_1^0$ fortsetzen. Da sie beschränkt ist, läßt sie sich als ein Skalarprodukt auf H_1^0 darstellen:

$$B(u, \chi) = (v, \chi)_1.$$

Dabei gehört zu jedem $u \in H_1^0$ genau ein $v \in H_1^0$.

Durch die Relation

$$(2. 10) \quad (Lu, \chi)_0 = B(u, \chi) = (v, \chi)_1 = (Tu, \chi)_1$$

haben wir dann zu L einen auf H_1^0 definierten linearen, beschränkten Operator T erklärt, denn nach (2. 9) gilt

$$| (Tu, \chi)_1 | = | B(u, \chi) | \leq K \| u \|_1 \| \chi \|_1,$$

also

$$(2. 11) \quad \| Tu \|_1 \leq K \| u \|_1.$$

Trennen wir nun von dem Operator L den elliptischen Hauptteil ab und bezeichnen ihn wie in (1. 7) mit

$$L^0 := \frac{\partial}{\partial x} \left\{ a_{11}(x; \eta) \frac{\partial u}{\partial x} \right\} + \lambda^2 a_{22}(x; \eta) \frac{\partial^2 u}{\partial t^2},$$

so ergibt sich für alle $\lambda > 0$ für die $L^0(\lambda, \eta)$ nach obiger Methode entsprechenden Operatoren $T^0(\lambda, \eta)$, wenn man in (2. 8) nur die beiden ersten Glieder auf der rechten Seite berücksichtigt:

$$| (T^0 u, \chi)_1 | \leq K \| u \|_1 \| \chi \|_1,$$

somit

$$(2. 12) \quad \| T^0 u \|_1 \leq K \| u \|_1$$

und

$$(2. 13) \quad | (T^0 u, u)_1 | \geq 2\alpha \left(\left\| \frac{\partial u}{\partial x} \right\|_0^2 + \left\| \frac{\partial u}{\partial t} \right\|_0^2 \right)$$

mit $2\alpha \leq \min(a_{11}, \lambda^2 a_{22})$.

Wegen der Rand- und Periodizitätsbedingungen für die Funktionen $u(x, t)$ erhält man die Abschätzung

$$(2.14) \quad \|u\|_0^2 \leq \frac{1}{2} \left\| \frac{\partial u}{\partial x} \right\|_0^2,$$

denn

$$u^2(x, t) = \left(u(0, t) + \int_0^x \frac{\partial u}{\partial x} dx \right)^2 \leq \int_0^x dx \int_0^x \left(\frac{\partial u}{\partial x} \right)^2 dx \leq x \int_0^1 \left(\frac{\partial u}{\partial x} \right)^2 dx,$$

somit

$$\int_0^{2\pi} \int_0^1 u^2(x, t) dx dt \leq \frac{1}{2} \int_0^{2\pi} \int_0^1 \left(\frac{\partial u}{\partial x} \right)^2 dx dt.$$

Aus (2.13) und (2.14) ergibt sich dann wegen (2.4) und (2.6)

$$(2.15) \quad |(T^0 u, u)_1| \geq \alpha \|u\|_1^2.$$

Der dem Rest $L^1 = L - L^0$ entsprechende Operator T^1 ist dann nach dem Rellichschen Auswahlssatz ein auf H_1^0 vollstetiger Operator.

Somit gilt für den auf H_1^0 erklärten Operator $T = T^0 + T^1$ für alle reellen $\lambda > 0$ und $|\eta| \leq K_0$, daß T^0 ein beschränkter, definiter Operator ist und T^1 ein vollstetiger Operator ist. Damit sind die Voraussetzungen von S. Hildebrandt-E. Wienholtz [3], Theorem 3 (siehe auch S. Hildebrandt [2], S. 139) erfüllt, und wir haben

Satz 2.1 (Fredholmalternative). *Unter den obigen Voraussetzungen über T^0 und T^1 gilt:*

- a) *Die Gleichung $Tu = h$ ist genau dann lösbar, wenn $h \in H_1^0 \ominus N(T^*)$ ist. Die Lösung u ist modulo $N(T)$ eindeutig bestimmt.*
- b) $\dim N(T) = \dim N(T^*) < \infty$.

Aus dem Beweis dieses Satzes bei S. Hildebrandt-E. Wienholtz [3] folgt noch

$$(2.16) \quad |(Tu, Tu)_1| \geq k \|u\|_1^2$$

für alle $u \in H_1^0 \ominus N(T)$.

λ^0 heißt Eigenwert zu $\eta = \eta^0$ der zu (1.1), (1.2), (1.3) gehörenden homogenen Aufgabe, wenn es nichttriviale Lösungen der Aufgabe

$$(2.17) \quad L(\lambda, \eta)\varphi = 0$$

$$(2.18) \quad \varphi(0, t) = \varphi(1, t) = 0$$

$$(2.19) \quad \varphi(x, t) = \varphi(x, t + 2\pi)$$

gibt zu $\lambda = \lambda^0$, $\eta = \eta^0$. Wir interessieren uns für die Eigenwerte $\lambda^0 > 0$.

Auf Grund der Analytizität von L sind die Eigenlösungen von (2.17), (2.18), (2.19) mit denen von

$$T(\lambda^0, \eta^0)\varphi = 0$$

in H_1^0 identisch, denn

$$0 = (L(\lambda^0, \eta^0)\varphi, \chi)_0 = (T(\lambda^0, \eta^0)\varphi, \chi)_1$$

für alle $\chi \in H_1^0$. Dann folgt aus Satz 2.1, daß zu festem η^0 die reellen Eigenwerte $\lambda^0 > 0$ dieser Aufgabe (2.17), (2.18), (2.19) nur endliche Vielfachheit haben.

Es sei M die Vielfachheit, und $\varphi^1, \dots, \varphi^M$ eine Basis des linearen Raums der Eigenlösungen von η^0, λ^0 , dann hat nach Satz 2. 1

$$Tu = h$$

genau dann eine Lösung, wenn

$$(h, \varphi^m)_1 = 0, \quad m = 1, \dots, M$$

gilt. Das bedeutet, daß

$$(2. 20) \quad Lu = g$$

genau dann eine Lösung hat, wenn

$$(2. 21) \quad (g, \varphi^m)_0 = 0, \quad m = 1, \dots, M,$$

ist.

3. Hilfssätze und Abschätzungen

Neben den in (2. 5) und (2. 6) erklärten Normen $\|u\|_0$ und $\|u\|_1$ führen wir noch folgende Normen bzw. Halbnormen ein:

$$(3. 1 a) \quad \|\widetilde{u}\|_\nu^2 = \sum_{j=0}^\nu \left\| \frac{\partial^j u}{\partial x^{j-1} \partial t^j} \right\|_0^2,$$

$$(3. 1 b) \quad \|\widetilde{u}\|_\nu'^2 = \left\| \frac{\partial^2 u}{\partial x^2 \partial t^{\nu-2}} \right\|_0^2 + \left\| \frac{\partial^2 u}{\partial x \partial t^{\nu-1}} \right\|_0^2 + \left\| \frac{\partial^2 u}{\partial t^\nu} \right\|_0^2, \quad \nu \geq 3,$$

$$(3. 1 c) \quad \|u\|_\nu^2 = \sum_{j=0}^\nu \|\widetilde{u}\|_j^2,$$

$$(3. 1 d) \quad \|u\|_2^{*2} = \|u\|_1^2 + \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2,$$

$$(3. 1 e) \quad \|u\|_\nu'^2 = \|u\|_2^2 + \sum_{j=3}^\nu \|\widetilde{u}\|_j'^2, \quad \nu \geq 3.$$

Diese Normen und Halbnormen brauchen wir im folgenden bis zur Ordnung $\nu = 5$.

Zwischen den einzelnen Normen und Halbnormen bestehen folgende Beziehungen:

$$(3. 2 a) \quad \|\widetilde{u}\|_\nu^2 \leq \|u\|_\nu^2,$$

$$(3. 2 b) \quad \|u\|_2^{*2} \leq \|u\|_2^2,$$

$$(3. 2 c) \quad \|u\|_2^2 \leq \|u\|_3'^2,$$

$$(3. 2 d) \quad \|u\|_\nu'^2 \leq \|u\|_{\nu+1}^2.$$

Für spätere Rechnungen wollen wir noch einige Abschätzungen zusammenstellen in den speziellen Formen, die wir dann benötigen. Wir wollen also nicht die jeweils schärfsten Abschätzungen angeben. Um nicht zuviel Konstanten bei den Rechnungen verwenden zu müssen, werden wir für fast alle auftretenden Konstanten die Bezeichnung C wählen. Aus dem Zusammenhang wird dann hervorgehen, wovon die Konstante nicht abhängt, bezüglich welcher Variablen C konstant sein soll.

A. Sobolewskes Lemma. Dieses lautet speziell für ein Rechteckgebiet — unter Verwendung der Bezeichnung (3. 1 d)

$$(3. 3) \quad |u(x, t)|^2 \leq C \left(\|u\|_1^2 + \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2 \right) = C \|u\|_2^{*2}.$$

Es gehen hierbei also auf der rechten Seite nicht noch die beiden 2. Ableitungen $\frac{\partial^2 u}{\partial x^2}$ und $\frac{\partial^2 u}{\partial t^2}$ ein, sondern als einzige 2. Ableitung $\frac{\partial^2 u}{\partial x \partial t}$.

Der Vollständigkeit halber sei hier der Beweis für ein Rechteckgebiet aufgeführt:

$u(x, t)$ sei eine Funktion, die in einem Rechteck R der Kantenlänge $2a$ und $2b$ definiert sei und dort zweimal stetig differenzierbar sei. Ohne Beschränkung der Allgemeinheit sei $a \leq b$. Jedem Punkt (x_0, t_0) des Rechtecks R können wir dann als Eckpunkt eines kleinen Quadrates Q mit der Seitenlänge a ansehen, dessen Kanten parallel zu den Seiten des Rechtecks R laufen und das ganz in R liegt. Wir nehmen nun an, daß (x_0, t_0) der linke untere Eckpunkt eines solchen Quadrates Q ist. Wir verschieben das Achsenkreuz noch so, daß (x_0, t_0) in $(0, 0)$ übergeht, und zwei Quadratseiten mit den Koordinatenachsen übereinstimmen.

Wir betrachten eine Abschneidefunktion $\xi(r)$ für $0 \leq r \leq a$ mit den folgenden Eigenschaften:

$$\xi(r) : \begin{cases} 1 & \text{für } 0 \leq r \leq \frac{a}{2} \\ 0 \leq \xi(r) \leq 1 & \text{für } \frac{a}{2} \leq r \leq a \\ 0 & \text{für } a \leq r, \end{cases}$$

$\xi(r) \in C^1[0, a]$. Für die Ableitungen von $\xi(r)$ gilt

$$|\xi(r)|, \quad \left| \frac{d\xi(r)}{dr} \right| < C_1,$$

wobei C_1 nur von der Wahl von $\xi(r)$ abhängt. Mit $v(x, t) = u(x, t)\xi(x)\xi(t)$ gilt dann, da $v(a, a) = v(0, a) = v(a, 0) = 0$:

$$\begin{aligned} u(0, 0) &= v(0, 0) - v(a, 0) - v(0, a) + v(a, a) = \int_0^a \int_0^a \frac{\partial^2 v}{\partial x \partial t} dx dt \\ &= \int_0^a \int_0^a \frac{\partial^2}{\partial x \partial t} u(x, t) \xi(x) \xi(t) dx dt = \int_0^a \int_0^a \left\{ \frac{\partial^2 u(x, t)}{\partial x \partial t} \xi(x) \xi(t) \right. \\ &\quad \left. + \frac{\partial u(x, t)}{\partial x} \xi(x) \frac{d\xi(t)}{dt} + \frac{\partial u(x, t)}{\partial t} \frac{d\xi(x)}{dx} \xi(t) + u(x, t) \frac{d\xi(x)}{dx} \frac{d\xi(t)}{dt} \right\} dx dt, \end{aligned}$$

somit

$$\begin{aligned} |u(0, 0)|^2 &\leq C_1^4 \left\{ \int_0^a \int_0^a \left(\left| \frac{\partial^2 u}{\partial x \partial t} \right| + \left| \frac{\partial u}{\partial x} \right| + \left| \frac{\partial u}{\partial t} \right| + |u| \right) dx dt \right\}^2 \\ &\leq C \left\{ \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2 + \|u\|_1^2 \right\} = C \|u\|_2^{*2}. \end{aligned}$$

Analoge Rechnungen ergeben sich, wenn (x_0, t_0) ein anderer Eckpunkt eines ganz in R liegenden Quadrates ist.

Für den bezüglich der Norm $\|u\|_2^*$ vervollständigten Raum der in R definierten, zweimal stetig differenzierbaren Funktionen gilt dann dieselbe Ungleichung (siehe etwa S. Hildebrandt [2], S. 58), also für alle $u(x, t)$, für die die rechte Seite existiert.

B. Da $\frac{\partial^n u(0, t)}{\partial t^n} = \frac{\partial^n u(1, t)}{\partial t^n} = 0$ ist, und alle auftretenden Funktionen mit der Periode 2π periodisch sind in t , gilt

$$(3.4) \quad \left(\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial t^2} \right)_0 = \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2.$$

Beweis.

$$\begin{aligned} \left(\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial t^2} \right)_0 &= \int_0^{2\pi} dt \int_0^1 \frac{\partial^2 u}{\partial x^2} \frac{\partial^2 u}{\partial t^2} dx = \int_0^{2\pi} dt \left[\frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial t^2} \right]_0^1 - \int_0^1 \frac{\partial u}{\partial x} \frac{\partial^3 u}{\partial t^2 \partial x} dx \\ &= - \int_0^1 dx \left[\frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial x \partial t} \right]_0^{2\pi} - \int_0^{2\pi} \left(\frac{\partial^2 u}{\partial x \partial t} \right)^2 dt = \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2. \end{aligned}$$

C. Mit

$$(3.5) \quad 0 < a_{11} \leq a_{11}(x) \leq A_{11}, \quad 0 < a'_{11} \leq \left| \frac{da_{11}(x)}{dx} \right| \leq A'_{11}$$

gilt

$$\begin{aligned} (3.6) \quad \left\| \frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\} \right\|_0^2 &= \left\| a_{11}(x) \frac{\partial^2 u}{\partial x^2} \right\|_0^2 + \left\| \frac{da_{11}(x)}{dx} \frac{\partial u}{\partial x} \right\|_0^2 \\ &+ 2 \left(a_{11}(x) \frac{\partial^2 u}{\partial x^2}, \frac{da_{11}(x)}{dx} \frac{\partial u}{\partial x} \right)_0 \geq a_{11}^2 \left\| \frac{\partial^2 u}{\partial x^2} \right\|_0^2 + a_{11}'^2 \left\| \frac{\partial u}{\partial x} \right\|_0^2 - 2A_{11}A'_{11} \left\| \frac{\partial^2 u}{\partial x^2} \right\|_0 \left\| \frac{\partial u}{\partial x} \right\|_0. \end{aligned}$$

D. Durch mehrmalige partielle Integration analog dem Beweis von (3. 4) ergibt sich

$$\left(\frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\}, a_{22}(x) \frac{\partial^2 u}{\partial t^2} \right)_0 = \left(a_{11}(x) a_{22}(x) \frac{\partial^2 u}{\partial x \partial t}, \frac{\partial^2 u}{\partial x \partial t} \right)_0 + \left(a_{11}(x) \frac{da_{22}(x)}{dx} \frac{\partial^2 u}{\partial x \partial t}, \frac{\partial u}{\partial t} \right)_0.$$

Berücksichtigen wir (3. 5) und

$$0 < a_{22} \leq a_{22}(x), \quad \left| \frac{da_{22}(x)}{dx} \right| \leq A'_{22},$$

so erhalten wir

$$(3.7) \quad \left(\frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\}, a_{22}(x) \frac{\partial^2 u}{\partial t^2} \right)_0 \geq a_{11} a_{22} \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2 - A_{11} A'_{22} \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0 \left\| \frac{\partial u}{\partial t} \right\|_0.$$

E. Für L^0 , das in (1. 7) definiert ist, gilt mit endlichem, reellem $\lambda > 0$, $|\eta| < K_0$

$$(3.8) \quad \|L^0 u\|_0^2 \geq C_1 \|\widetilde{u}\|_2^2 - C_2 \|\widetilde{u}\|_2 \|u\|_1.$$

Beweis.

$$\begin{aligned} \|L^0 u\|_0^2 &= \left\| \frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\} + \lambda^2 a_{22}(x) \frac{\partial^2 u}{\partial t^2} \right\|_0^2 \\ &= \left\| \frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\} \right\|_0^2 + 2\lambda^2 \left(\frac{\partial}{\partial x} \left\{ a_{11}(x) \frac{\partial u}{\partial x} \right\}, a_{22}(x) \frac{\partial^2 u}{\partial t^2} \right)_0 + \lambda^4 \left\| a_{22}(x) \frac{\partial^2 u}{\partial t^2} \right\|_0^2. \end{aligned}$$

Dies ist nach (3. 6) und (3. 7)

$$\begin{aligned} &\geq a_{11}^2 \left\| \frac{\partial^2 u}{\partial x^2} \right\|_0^2 + 2\lambda^2 a_{11} a_{22} \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2 + \lambda^4 a_{22}^2 \left\| \frac{\partial^2 u}{\partial t^2} \right\|_0^2 \\ &\quad - 2A_{11}A'_{11} \left\| \frac{\partial^2 u}{\partial x^2} \right\|_0 \left\| \frac{\partial u}{\partial x} \right\|_0 + a_{11}'^2 \left\| \frac{\partial u}{\partial x} \right\|_0^2 - 2\lambda^2 A_{11}A'_{22} \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0 \left\| \frac{\partial u}{\partial t} \right\|_0. \end{aligned}$$

Somit nach der Definition (3. 1 a) von $\|\widetilde{u}\|_2^2$

$$\begin{aligned} \|L^0 u\|_0^2 &\geq C_1 \left(\left\| \frac{\partial^2 u}{\partial x^2} \right\|_0^2 + \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0^2 + \left\| \frac{\partial^2 u}{\partial t^2} \right\|_0^2 \right) - C_2 \|u\|_1 \left(\left\| \frac{\partial^2 u}{\partial x^2} \right\|_0 + \left\| \frac{\partial^2 u}{\partial x \partial t} \right\|_0 \right) \\ &\geq C_1 \|u\|_2^2 - C_2 \|u\|_1 \|\widetilde{u}\|_2. \end{aligned}$$

F. Die analogen Rechnungen wie in C. bis E. lassen sich für $v = \frac{\partial^n u}{\partial t^n}$ statt u durchführen. Da $\frac{\partial^n}{\partial t^n} L^0 u = L^0 \frac{\partial^n u}{\partial t^n} = L^0 v$ ist, erhalten wir

$$(3. 9) \quad \left\| \frac{\partial^n}{\partial t^n} L^0 u \right\|_0^2 \geq C_1 \|u\|_{n+2}'^2 - C_2 \left\| \frac{\partial^n u}{\partial t^n} \right\|_1 \|\widetilde{u}\|_{n+2}'^2,$$

wobei $\|\widetilde{u}\|_{n+2}'^2$ in (3. 1 b) definiert ist.

G. Abschätzung der nichtlinearen Terme und ihrer Ableitungen nach t . Neben den in (3. 1) eingeführten Normen und Halbnormen benötigen wir für diesen und den nächsten Abschnitt noch folgende Bezeichnungen:

$$\begin{aligned} D^0 u &= |u|, \\ \widetilde{D^1 u} &= \left| \frac{\partial u}{\partial x} \right| + \left| \frac{\partial u}{\partial t} \right|, & D^1 u &= D^0 u + \widetilde{D^1 u}, \\ \widetilde{D^2 u} &= \left| \frac{\partial^2 u}{\partial x \partial t} \right| + \left| \frac{\partial^2 u}{\partial t^2} \right|, & D^2 u &= D^1 u + \widetilde{D^2 u}, \\ \widetilde{D^3 u} &= \left| \frac{\partial^3 u}{\partial x \partial t^2} \right| + \left| \frac{\partial^3 u}{\partial t^3} \right|, & D^3 u &= D^2 u + \widetilde{D^3 u}, \\ \widetilde{D^4 u} &= \left| \frac{\partial^4 u}{\partial x \partial t^3} \right| + \left| \frac{\partial^4 u}{\partial t^4} \right|, & D^4 u &= D^3 u + \widetilde{D^4 u}. \end{aligned} \quad (3. 10)$$

Mit Hilfe dieser Bezeichnungen ergibt sich, da nach (1. 1)

$$f(u) = \sum_{\substack{j, k, l, m, n=0 \\ j+k+l+m+n=2}}^2 \lambda^{l+m+2n} a_{jklmn}(x, t; \eta) u^j \left(\frac{\partial u}{\partial x} \right)^k \left(\frac{\partial u}{\partial t} \right)^l \left(\frac{\partial^2 u}{\partial x \partial t} \right)^m \left(\frac{\partial^2 u}{\partial t^2} \right)^n$$

und die $a_{jklmn}(x, t; \eta)$ stetig sind

$$\begin{aligned} |f(u)| &\leq C(D^2 u)^2, \\ \left| \frac{\partial f(u)}{\partial t} \right| &\leq C D^2 u D^3 u, \\ \left| \frac{\partial^2 f(u)}{\partial t^2} \right| &\leq C((D^3 u)^2 + D^2 u D^4 u), \\ \left| \frac{\partial^3 f(u)}{\partial t^3} \right| &\leq C(D^3 u D^4 u + D^2 u D^5 u). \end{aligned} \quad (3. 11)$$

Wegen des Sobolewschen Lemmas (3. 3) und der in (3. 1) eingeführten Bezeichnungen für die Halbnormen und Normen gilt:

$$\begin{aligned} D^2 u &\leq C \|u\|_4', & D^3 u &\leq C \|u\|_5', \\ (3. 12) \quad \|D^2 u\|_0 &\leq C \|u\|_2, & \|D^3 u\|_0 &\leq C \|u\|_3', \\ \|D^4 u\|_0 &\leq C \|u\|_4', & \|D^5 u\|_0 &\leq C \|u\|_5'. \end{aligned}$$

Da

$$(3. 13) \quad \|fg\|_0 \leq \mathbf{Max}_{(x, t) \in G} |f| \|g\|_0$$

ist, folgt aus (3. 11) und (3. 12)

$$(3. 14) \quad \begin{aligned} \|f(u)\|_0 &\leq C \|u\|'_4 \|u\|_2, & \left\| \frac{\partial f(u)}{\partial t} \right\|_0 &\leq C \|u\|'_4 \|u\|'_3, \\ \left\| \frac{\partial^2 f(u)}{\partial t^2} \right\|_0 &\leq C (\|u\|'_5 \|u\|'_3 + \|u\|'^2_4), & \left\| \frac{\partial^3 f(u)}{\partial t^3} \right\|_0 &\leq C \|u\|'_5 \|u\|'_4, \end{aligned}$$

also wegen (3. 2c) und (3. 2d)

$$(3. 15) \quad \left\| \frac{\partial^v f(u)}{\partial t^v} \right\|_0 \leq C \|u\|'^2_5, \quad v = 0, 1, 2, 3.$$

H. Abschätzung der Differenzen der nichtlinearen Terme. Da

$$a_1 a_2 - b_1 b_2 = \frac{1}{2} \{ (a_1 - b_1) (a_2 + b_2) + (a_2 - b_2) (a_1 + b_1) \}$$

ist und hier nur Differenzen von Gliedern zweiter Ordnung auftreten, erhalten wir mit den Bezeichnungen (3. 10)

$$\begin{aligned} |f(u_1) - f(u_2)| &\leq C D^2(u_1 - u_2) (D^2 u_1 + D^2 u_2), \\ \left| \frac{\partial f(u_1)}{\partial t} - \frac{\partial f(u_2)}{\partial t} \right| &\leq C \{ D^2(u_1 - u_2) (D^3 u_1 + D^3 u_2) + D^3(u_1 - u_2) (D^2 u_1 + D^2 u_2) \}, \\ \left| \frac{\partial^2 f(u_1)}{\partial t^2} - \frac{\partial^2 f(u_2)}{\partial t^2} \right| &\leq C \{ D^2(u_1 - u_2) (D^4 u_1 + D^4 u_2) + D^3(u_1 - u_2) (D^3 u_1 + D^3 u_2) \\ &\quad + D^4(u_1 - u_2) (D^2 u_1 + D^2 u_2) \}, \\ \left| \frac{\partial^3 f(u_1)}{\partial t^3} - \frac{\partial^3 f(u_2)}{\partial t^3} \right| &\leq C \{ D^2(u_1 - u_2) (D^5 u_1 + D^5 u_2) + D^3(u_1 - u_2) (D^4 u_1 + D^4 u_2) \\ &\quad + D^4(u_1 - u_2) (D^3 u_1 + D^3 u_2) + D^5(u_1 - u_2) (D^2 u_1 + D^2 u_2) \}. \end{aligned}$$

Somit gilt wegen (3. 12), (3. 13), (3. 2c) und (3. 2d)

$$(3. 16) \quad \left\| \frac{\partial^v f(u_1)}{\partial t^v} - \frac{\partial^v f(u_2)}{\partial t^v} \right\|_0 \leq C \|u_1 - u_2\|'_5 (\|u_1\|'_5 + \|u_2\|'_5)$$

für $v = 0, 1, 2, 3$.

4. Die nichtlineare Aufgabe

Wir nehmen für das Folgende an, daß $\lambda = \lambda^0$ ein reeller, positiver Eigenwert der homogenen Aufgabe (2. 17), (2. 18), (2. 19) zu $\eta = 0$ ist. Dann hat nach Satz 2. 1 die homogene Aufgabe zu $\lambda^0, \eta = 0$ nur endlich viele Eigenfunktionen $\varphi^1, \dots, \varphi^M$, die wir als orthonormiert annehmen können:

$$(4. 1) \quad (\varphi^m, \varphi^n)_0 = \delta^{mn}, \quad m, n = 1, \dots, M.$$

Gemäß (1. 9), (1. 10) zerlegen wir

$$L(\lambda, \eta) = L_0(\lambda^0, 0) + L_1(\lambda^0, 0; \varepsilon, \eta)$$

und erhalten somit aus (1. 1) folgende Gleichung

$$(4. 2) \quad L_0 u = -L_1 u + g(u).$$

Die zu $L_0(\lambda^0, 0)$ adjungierte Aufgabe habe die orthonormierten Eigenfunktionen $\psi^m, m = 1, \dots, M$,

$$(4. 3) \quad (\psi^m, \psi^n)_0 = \delta^{mn}, \quad m, n = 1, \dots, M.$$

Da L_0 ein elliptischer Operator ist mit analytischen Koeffizienten und auch der Rand von $\tilde{G}$ analytisch ist, sind die Eigenfunktionen $\varphi^m, \psi^m \in C^\infty(\tilde{G})$, somit gilt

$$(4.4) \quad \left\| \frac{\partial^v \varphi^m}{\partial x^{v-i} \partial t^j} \right\|_0, \left\| \frac{\partial^v \psi^m}{\partial x^{v-i} \partial t^j} \right\|_0 \leq C.$$

Die Nullräume von $L_0(\lambda^0, 0)$ bzw. $L_0^*(\lambda^0, 0)$ bezeichnen wir mit $N(L_0)$ bzw. $N(L_0^*)$

$$N(L_0) = \bigvee_{m=1}^M \varphi^m, \quad N(L_0^*) = \bigvee_{m=1}^M \psi^m.$$

Wir definieren Projektionsoperatoren

$$(4.5) \quad \begin{aligned} P: H_0 &\rightarrow H_0 \ominus N(L_0^*) = N(L_0^*)^\perp, \\ Q: H_0 &\rightarrow N(L_0^*), \quad P = I - Q. \end{aligned}$$

Jedes Element $u \in H_0$ können wir zerlegen in

$$(4.6) \quad u = v + w, \quad v \in N(L_0)^\perp, \quad w \in N(L_0).$$

Wir suchen nun Lösungen der nichtlinearen Aufgabe (4.2) in der Umgebung von $u(x, t) \equiv 0$.

Da L_0 nicht invertierbar ist, gehen wir zur „modifizierten“ Aufgabe

$$(4.7a) \quad L_0 v = P\{-L_1 u + g(u)\},$$

$$(4.7b) \quad \sum_{m=1}^M (\varphi^m, w)_0 \psi^m = \sum_{m=1}^M \zeta_m \psi^m + Q\{-L_1 u + g(u)\}$$

über, wobei (4.7b) mit

$$w = \sum_{m=1}^M w_m \varphi^m, \quad w_m = (w, \varphi^m)_0$$

äquivalent ist zu

$$(4.7b') \quad w_m = \zeta_m + (-L_1 u + g(u), \varphi^m)_0, \quad m = 1, \dots, M.$$

Dabei sind die $\zeta_m, m = 1, \dots, M$, unbestimmte Parameter.

Diese modifizierte Aufgabe ist eindeutig lösbar in dem Sinne, daß für jedes u , für das die rechte Seite erklärt ist, und jedes M -Tupel $\{\zeta_m\}$, die wir auf der rechten Seite einsetzen, wir genau ein $v \in N(L_0)^\perp$ und ein $w \in N(L_0)$ erhalten, denn da $P\{-L_1 u + g(u)\} \in N(L_0^*)^\perp$ liegt, gibt es genau ein $v \in N(L_0)^\perp$, so daß $L_0 v = P\{-L_1 u + g(u)\}$ ist. Die Eindeutigkeit der Lösung von (4.7b) bzw. (4.7b') ist klar. Es braucht aber natürlich nicht $u = v + w$ zu sein.

Notwendig und hinreichend dafür, daß eine so gewonnene Lösung

$$u(x, t; \zeta_1, \dots, \zeta_M) = v + w$$

der modifizierten Aufgabe (4.7a), (4.7b) auch eine Lösung der ursprünglichen Aufgabe (4.2) ist, ist wegen der Gültigkeit der Fredholmalternative

$$-L_1 u + g(u) \in N(L_0^*)^\perp.$$

Das bedeutet dann nach (4.5), daß

$$P\{-L_1 u + g(u)\} = -L_1 u + g(u), \quad Q\{-L_1 u + g(u)\} = 0$$

sind. Daraus folgt wegen (4. 6) und (4. 7b) als notwendige und hinreichende Bedingungen dafür, daß eine Lösung der modifizierten Aufgabe (4. 7a), (4. 7b) auch Lösung der ursprünglichen Aufgabe (4. 2) ist:

$$(4. 8) \quad (u, \varphi^m)_0 = (w, \varphi^m)_0 = \zeta_m, \quad m = 1, \dots, M.$$

Diese Gleichungen sind die Verzweigungsgleichungen.

Wir wollen nun das Gleichungssystem (4. 7a), (4. 7b) durch sukzessive Approximation lösen.

Wir beginnen mit

$$(4. 9) \quad u^{(0)} = 0.$$

Bestimmen dann $v^{(n+1)} \in N(L_0)^\perp$ durch

$$(4. 10) \quad L_0 v^{(n+1)} = P\{-L_1 u^{(n)} + g(u^{(n)})\}$$

und $w^{(n+1)} \in N(L_0)$ durch

$$(4. 11) \quad w_m^{(n+1)} = \zeta_m + (-L_1 u^{(n)} + g(u^{(n)}), \varphi^m)_0,$$

$$(4. 12) \quad w^{(n+1)} = \sum_{m=1}^M w_m^{(n+1)} \varphi^m.$$

Die $M + 1$ Gleichungen (4. 10), (4. 11) sind bei jedem Schritt eindeutig lösbar, denn da $u^{(0)} = 0$ analytisch ist, liefert jeder Schritt wieder eine analytische Funktion, also sind die rechten Seiten bei jedem Schritt erklärt.

Um zu sichern, daß die sukzessive Approximation ein eindeutig bestimmtes Element liefert, wenden wir den Banachschen Fixpunktsatz an auf den durch (4. 10), (4. 11), (4. 12) definierten Operator

$$(4. 13) \quad A : (v^{(n)}, w^{(n)}) \rightarrow (v^{(n+1)}, w^{(n+1)})$$

oder

$$(4. 14) \quad U = Au$$

mit

$$(4. 15) \quad U = V + W, \quad u = v + w,$$

$$(4. 16) \quad v, V \in N(L_0)^\perp, \quad w, W \in N(L_0).$$

Dazu müssen wir zeigen, daß

1. jedes Element einer abgeschlossenen Teilmenge $\mathfrak{M}$ des vollständigen Raumes $\mathfrak{U}$ durch den Operator A wieder in ein Element von $\mathfrak{M}$ überführt wird:

$$(4. 17) \quad u = v + w \in \mathfrak{M}, \quad U = Au = V + W \in \mathfrak{M}$$

mit

$$(4. 18a) \quad L_0 V = -PL_1 u + Pg(u),$$

$$(4. 18b) \quad W_m = \zeta_m + (-L_1 u + g(u), \varphi^m)_0, \quad W = \sum_{m=1}^M W_m \varphi^m;$$

2. der Operator A in $\mathfrak{M}$ kontrahierend ist, d. h. es gibt ein q mit $0 \leq q < 1$ derart, daß für alle $u_1, u_2 \in \mathfrak{M}$ gilt

$$(4. 19) \quad \|U_1 - U_2\| = \|Au_1 - Au_2\| \leq \|V_1 - V_2\| + \|W_1 - W_2\| \leq q \|u_1 - u_2\|;$$

3. der Ausgangspunkt der sukzessiven Approximation in $\mathfrak{M}$ liegt:

$$(4.20) \quad u^{(0)} \in \mathfrak{M}.$$

Den Raum $\mathfrak{U}$ bestimmen wir folgendermaßen:

Wir nehmen die Funktionen $u(x, t)$, die folgende Bedingungen erfüllen:

a) $u(x, t)$ und alle Ableitungen bis zur zweiten Ordnung sowie die Ableitungen

$$\frac{\partial^3 u}{\partial x^2 \partial t}, \frac{\partial^3 u}{\partial x \partial t^2}, \frac{\partial^3 u}{\partial t^3}, \frac{\partial^4 u}{\partial x^2 \partial t^2}, \frac{\partial^4 u}{\partial x \partial t^3}, \frac{\partial^4 u}{\partial t^4}, \frac{\partial^5 u}{\partial x^2 \partial t^3}, \frac{\partial^5 u}{\partial x \partial t^4}, \frac{\partial^5 u}{\partial t^5}$$

$$\text{sind in } \tilde{G} = \left\{ \begin{array}{l} x \in [0, 1] \\ -\infty < t < +\infty \end{array} \right\} \text{ stetig,}$$

$$\text{b) } u(0, t) = u(1, t) = 0,$$

$$\text{c) } u(x, t) = u(x, t + 2\pi).$$

Diese lineare Menge schließen wir bezüglich der in (3.1e) eingeführten Norm $\|u\|'_5$ ab.

Nach (2.10) und $L_0 = L(\lambda^0, 0)$ gilt

$$(L_0 V, \chi)_0 = (T_0 V, \chi)_1,$$

also mit $\chi = T_0 V$ nach (2.16), da $V \in N(L_0)^\perp$,

$$k \|V\|_1^2 \leq |(T_0 V, T_0 V)_1| = |(L_0 V, T_0 V)_0|,$$

wegen (4.18a) ist dies

$$(4.21) \quad \begin{aligned} &\leq |(PL_1 u, T_0 V)_0| + |(Pg(u), T_0 V)_0| \\ &\leq \|L_1 u\|_0 \|T_0 V\|_1 + \|g(u)\|_0 \|T_0 V\|_1, \end{aligned}$$

da P ein Projektionsoperator ist und $\|u\|_0 \leq \|u\|_1$.

Für

$$(4.22) \quad |\varepsilon| + |\eta| \leq K \leq K_0$$

gilt wegen (1.10)

$$(4.23) \quad \|L_1 u\|_0 \leq CK \|u\|_2.$$

Hiermit und da T_0 nach (2.12) ein auf H_1^0 beschränkter Operator ist, folgt aus (4.21)

$$(4.24) \quad k \|V\|_1^2 \leq C(K \|u\|_2 + \|g(u)\|_0) \|V\|_1.$$

Somit wegen der Definition von $g(u)$ in (1.1) und wegen (3.15)

$$(4.25) \quad \|V\|_1 \leq C(K \|u\|_2 + \|g(u)\|_0) \leq C(K \|u\|'_5 + \|u\|'^2_5 + K \|f_0\|_0).$$

Setzen wir nun entsprechend (1.7) und (1.8) $L_0 = L_0^0 + L_0^1$, so erhalten wir aus (4.18a)

$$(4.26) \quad L_0^0 V = -L_0^1 V - PL_1 u + Pg(u).$$

Bilden wir nun bei dieser Gleichung das Skalarprodukt $(\cdot, L_0^0 V)_0$, so erhalten wir

$$\begin{aligned} \|L_0^0 V\|_0^2 &= (-L_0^1 V, L_0^0 V)_0 + (-PL_1 u, L_0^0 V)_0 + (Pg(u), L_0^0 V)_0 \\ &\leq \|L_0^1 V\|_0 \|L_0^0 V\|_0 + \|L_1 u\|_0 \|L_0^0 V\|_0 + \|g(u)\|_0 \|L_0^0 V\|_0 \\ &\leq C(\|V\|_1 \|V\|_2 + K \|u\|_2 \|V\|_2 + \|g(u)\|_0 \|V\|_2), \end{aligned}$$

denn aus (1. 8) und (1. 9) folgt

$$\|L_0^1 V\|_0 \leq C \|V\|_1.$$

Wegen (3. 8) erhalten wir dann

$$\|\widetilde{V}\|_2^2 \leq C(\|V\|_1 + K\|u\|_2 + \|g(u)\|_0)\|V\|_2$$

und nach (1. 1) und (3. 15)

$$(4. 27) \quad \|\widetilde{V}\|_2^2 \leq C(\|V\|_1 + K\|u\|_5' + K\|f_0\|_0 + \|u\|_5'^2)\|V\|_2.$$

Differenzieren wir (4. 26) nach t , so erhalten wir

$$(4. 28) \quad L_0^0 \frac{\partial V}{\partial t} = -L_0^1 \frac{\partial V}{\partial t} + \frac{\partial}{\partial t} [P\{-L_1 u + g(u)\}].$$

Nach (4. 5) ist $P = I - Q$, somit gilt

$$P\{-L_1 u + g(u)\} = -L_1 u + g(u) - \sum_{m=1}^M (-L_1 u + g(u), \psi^m)_0 \psi^m.$$

Dann ist

$$(4. 29) \quad \begin{aligned} \frac{\partial}{\partial t} P\{-L_1 u + g(u)\} &= -L_1 \frac{\partial u}{\partial t} + \frac{\partial}{\partial t} g(u) \\ &\quad - \sum_{m=1}^M (-L_1 u + g(u), \psi^m)_0 \frac{\partial \psi^m}{\partial t}. \end{aligned}$$

Wegen (4. 4), (4. 23), (1. 1) und (3. 15) ergibt sich

$$(4. 30) \quad \begin{aligned} |(-L_1 u + g(u), \psi^m)_0| &\leq \| -L_1 u + g(u) \|_0 \| \psi^m \|_0 \\ &\leq C(K\|u\|_5' + K\|f_0\|_0 + \|u\|_5'^2). \end{aligned}$$

Aus (4. 28), (4. 29), (4. 30), (1. 1) und (3. 15) folgt

$$\left\| L_0^0 \frac{\partial V}{\partial t} \right\|_0 \leq C \left(\|V\|_2 + K\|u\|_5' + K \left\| \frac{\partial f_0}{\partial t} \right\|_0 + \|u\|_5'^2 \right).$$

Dann liefert die zu (4. 26) bis (4. 27) analoge Rechnung bei Bildung des Skalarproduktes von (4. 28) mit $\left(\cdot, L_0^0 \frac{\partial V}{\partial t} \right)_0$, weil nach (1. 7), (1. 9) und (3. 1e)

$$\left\| L_0^0 \frac{\partial V}{\partial t} \right\|_0 \leq C \|V\|_3'$$

ist,

$$(4. 31) \quad \|\widetilde{V}\|_3'^2 \leq C \left(\|V\|_2 + K\|u\|_5' + \left\| \frac{\partial f_0}{\partial t} \right\|_0 + \|u\|_5'^2 \right) \|V\|_3'.$$

Die weiteren Differentiationen von (4. 26) nach t liefern analog

$$(4. 32) \quad \|\widetilde{V}\|_4'^2 \leq C \left(\|V\|_3' + K\|u\|_5' + \left\| \frac{\partial^2 f_0}{\partial t^2} \right\|_0 + \|u\|_5'^2 \right) \|V\|_4',$$

$$(4. 33) \quad \|\widetilde{V}\|_5'^2 \leq C \left(\|V\|_4' + K\|u\|_5' + K \left\| \frac{\partial^3 f_0}{\partial t^3} \right\|_0 + \|u\|_5'^2 \right) \|V\|_5'.$$

Wir setzen

$$(4. 34) \quad \|f_0\|_0^2 + \left\| \frac{\partial f_0}{\partial t} \right\|_0^2 + \left\| \frac{\partial^2 f_0}{\partial t^2} \right\|_0^2 + \left\| \frac{\partial^3 f_0}{\partial t^3} \right\|_0^2 = \|f_0\|_{3,t}^2.$$

Aus (4. 25) und (4. 34) folgt dann

$$(4. 35) \quad \|V\|_1^2 \leq C(K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2).$$

Nach (3. 1 a) und (3. 1 c) ist

$$\|V\|_2^2 = \|V\|_1^2 + \widetilde{\|V\|_2^2},$$

also haben wir mit (4. 27), (4. 34) und (4. 35)

$$\|V\|_2^2 \leq C(K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2).$$

Unter weiterer Berücksichtigung von (4. 31), (4. 32) und (4. 33) schätzen wir dann nacheinander $\|V\|_3'$, $\|V\|_4'$ und $\|V\|_5'$ ab und erhalten schließlich

$$(4. 36) \quad \|V\|_5' \leq C(K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2).$$

Als nächstes wollen wir W , den Anteil von U in $N(L_0)$ abschätzen. Aus (4. 18b) ergibt sich

$$\begin{aligned} |W_m| &\leq |\zeta_m| + (\|L_1 u\|_0 + \|g(u)\|_0) \|\psi^m\|_0 \\ &\leq |\zeta_m| + C(K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2) \end{aligned}$$

nach (4. 4), (4. 23), (1. 1) und (3. 15).

Somit wegen (4. 18b) und (4. 4) und wenn

$$(4. 37) \quad \sum_{m=1}^M |\zeta_m| \leq Z$$

$$(4. 38) \quad \|W\|_0, \|W\|_1, \|W\|_2, \|W\|_\nu' \leq C(Z + K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2),$$

mit $\nu = 3, 4, 5$.

Aus (4. 38) für $\nu = 5$, (4. 36), (4. 15) und (4. 16) erhalten wir dann

$$(4. 39) \quad \|U\|_5' \leq C(Z + K \|u\|'_5 + K \|f_0\|_{3,t} + \|u\|_5'^2).$$

Nun wollen wir gemäß (4. 19) abschätzen

$$(4. 40) \quad \|U_1 - U_2\| \leq \|V_1 - V_2\| + \|W_1 - W_2\|.$$

Diese Abschätzung geschieht wieder bezüglich der Norm $\|\cdot\|_5'$.

Aus (4. 18a) ergibt sich

$$L_0(V_1 - V_2) = -PL_1(u_1 - u_2) + P\{g(u_1) - g(u_2)\}.$$

Die Abschätzungen verlaufen dann genauso wie in (4. 21) bis (4. 36), nur haben wir statt

$$V: V_1 - V_2 \quad \text{jetzt} \quad u: u_1 - u_2$$

zu setzen und bei den nichtlinearen Termen statt

$$g(u) = \eta f_0 + f(u) \quad \text{jetzt} \quad f(u_1) - f(u_2),$$

da sich bei der Differenzbildung ηf_0 weghebt.

Dann erhalten wir analog zu (4. 36) wegen (3. 16)

$$(4. 41) \quad \|V_1 - V_2\|_5' \leq C(K + \|u_1\|'_5 + \|u_2\|'_5) \|u_1 - u_2\|'_5.$$

Für die Abschätzung von $\|W_1 - W_2\|_5'$ erhalten wir nach (4. 18b)

$$w_{m,1} - w_{m,2} = (-L_1 u_1 + L_1 u_2 + f(u_1) - f(u_2), \psi^m)_0,$$

da sich die ζ_m wegheben. Damit analog zu (4. 38) mit (3. 16)

$$(4. 42) \quad \|W_1 - W_2\|'_5 \leq C(K + \|u_1\|'_5 + \|u_2\|'_5) \|u_1 - u_2\|'_5.$$

Somit erhalten wir aus (4. 40), (4. 41) und (4. 42)

$$(4. 43) \quad \|U_1 - U_2\|'_5 \leq C(K + \|u_1\|'_5 + \|u_2\|'_5) \|u_1 - u_2\|'_5.$$

Die Ungleichungen (4. 39) und (4. 43) sind dann die beiden Ungleichungen, aus denen wir schließen können, daß die beiden ersten Bedingungen des Banachschen Fixpunktsatzes (4. 17) und (4. 19) erfüllt sind. Denn wählen wir

$$(4. 44) \quad Z \leq \frac{1}{64C^2}, \quad K \leq \text{Min}\left(\frac{1}{4C}, \frac{1}{16\|f_0\|_{3,t}}\right), \\ K \leq K_0$$

so sind für

$$(4. 45) \quad \mathfrak{M} := \left\{u \mid \|u\|'_5 \leq \frac{1}{8C}\right\}$$

die beiden Bedingungen

$$\|U\|'_5 \leq \frac{1}{8C}$$

und

$$C(K + \|u_1\|'_5 + \|u_2\|'_5) \leq \frac{1}{2} = q < 1$$

erfüllt.

Die dritte Bedingung (4. 20) $u^{(0)} \in \mathfrak{M}$ ist auch erfüllt, denn

$$\|u^{(0)}\|'_5 = 0 \leq \frac{1}{8C}.$$

Wir erhalten somit für jedes hinreichend kleine $M + 2$ -Tupel $(\varepsilon, \eta, \zeta_m)$, $m = 1, \dots, M$, von Parametern, das nach (4. 22), (4. 37) und (4. 44) die Bedingungen

$$(4. 46) \quad |\varepsilon| + |\eta| \leq K \leq K_0, \quad \sum_{m=1}^M |\zeta_m| \leq Z$$

erfüllt, eine Lösung der modifizierten Aufgabe (4. 7 a), (4. 7 b), die nach (4. 45)

$$(4. 47) \quad \|u(x, t; \varepsilon, \eta, \zeta_m)\|'_5 \leq \frac{1}{8C}$$

erfüllt.

Nun ist nicht jede Lösung der modifizierten Aufgabe eine Lösung der ursprünglichen Aufgabe, sondern nur wenn noch die Verzweigungsgleichungen (4. 8)

$$(\varphi^m, u(x, t; \varepsilon, \eta, \zeta_m))_0 = \zeta_m, \quad m = 1, \dots, M,$$

erfüllt sind. In allen M Gleichungen treten nur Glieder auf, die die Parameter von 1. oder höherer Ordnung enthalten. Es treten also keine freien Glieder auf. Somit ist

$$(\varepsilon, \eta, \zeta_m) = (0, 0, 0)$$

eine Lösung der Verzweigungsgleichungen. Wir suchen nun Lösungen dieser Verzweigungsgleichungen in der Umgebung des Punktes $(0, 0, 0)$, die die Bedingungen (4. 46) erfüllen.

Die in ζ_m linearen Glieder von u lauten

$$\sum_{m=1}^M \zeta_m \varphi^m.$$

Das bedeutet für die Verzweigungsgleichungen, da nach (4. 1)

$$(\varphi^m, \varphi^n)_0 = \delta^{mn}$$

ist, daß sich die Glieder, die nur ζ_m , und diese nur von erster Ordnung enthalten, auf beiden Seiten wegheben.

Jedes $M + 2$ -Tupel $(\varepsilon, \eta, \zeta_m)$, das die Verzweigungsgleichungen erfüllt und für das die sukzessive Approximation eine Lösung liefert, ergibt eine schwache Lösung der ursprünglichen Aufgabenstellung (1. 1).

Berücksichtigen wir (4. 47), so liefert die Anwendung des Sobolewschen Lemmas in der Gestalt (3. 3) wegen der Definition (3. 1 e) von $\|u\|'_5$, daß

$$u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial x \partial t}, \frac{\partial^2 u}{\partial t^2}, \frac{\partial^3 u}{\partial x \partial t^2}, \frac{\partial^3 u}{\partial t^3} \in C^0(G)$$

sind, also wegen der Periodizität in $C^0(\tilde{G})$. Dann sind alle Glieder auf der linken Seite der Gleichung (1. 1) bis auf das Glied $\frac{\partial}{\partial x} \left\{ a_{11}(x; \eta) \frac{\partial u}{\partial x} \right\}$ und alle Glieder auf der rechten Seite in $C^0(\tilde{G})$, dann muß aber auch dieses Glied in $C^0(\tilde{G})$ sein. Somit sind alle sich nach diesem Verfahren ergebenden schwachen Lösungen auch Lösungen im klassischen Sinne.

Literatur

- [1] *S. Agmon*, Lectures on elliptic boundary value problems, Princeton 1965.
- [2] *S. Hildebrandt*, Lineare elliptische Differentialgleichungen, Vorlesungsausarbeitung, Mainz 1966.
- [3] *S. Hildebrandt* und *E. Wienholtz*, Constructive proofs of representation theorems in separable Hilbert space, Comm. Pure Appl. Math. **17** (1964), 369—373.
- [4] *U. Stade*, Periodische Lösungen bei nichtlinearen partiellen Differentialgleichungen vom elliptischen Typ, Math. Ann. **165** (1966), 117—134.

Mathematisches Institut der Universität, 65 Mainz

Eingegangen 27. Februar 1968

Dominant dimension of semi-primary rings

By Bruno J. Müller¹⁾ at Hamilton, Ontario

Herrn Professor Dr. H. Rohrbach zum 65. Geburtstag gewidmet

The notion of dominant dimension, as introduced for finite-dimensional algebras over fields by Nakayama [15] and Tachikawa [17], generalizes immediately to arbitrary rings: right-dominant-dimension of R ($\text{r.dom.dim. } R$) $\geq n$ if there exists an exact sequence $0 \rightarrow R \rightarrow X_1 \rightarrow \cdots \rightarrow X_n$ of R -right-modules and $\cdot$ -homomorphisms where the X_i are projective and injective (left-dominant-dimension accordingly). Our main result is the equality of these two dimensions, if both are positive and if R is semi-primary (or two-sided perfect). The observation leading to this result is the fact that the endomorphism-ring of the injective hull of such a ring R is an artinian ring possessing a Morita-duality. A classification of semi-primary rings of positive dimension, analogous to the results in [12], [13] for algebras, is then obtained.

1. Lemmata

For any ring R we fix the following notation: H is the injective hull of the R -right-module R ; A is the endomorphism-ring of H_R ; Q is the endomorphism-ring of ${}_A H$, the complete right-quotient-ring of R (Lambek [9]). The direct sum of the injective hulls of the simple A -left-modules is denoted by U ; B is the endomorphism-ring of ${}_A U$; so U is an A — B -bimodule²⁾. For an A -left- (B -right-) module X , X^* means the U -dual $\text{Hom}(X, U)$, a B -right (A -left-) module. $X \rightarrow X^*$ is a contravariant equivalence between the categories of U -reflexive modules; and if U_B is an injective cogenerator and $\text{Endo}(U_B) = A$, then these categories are closed under sub-, factor-modules and contain the modules ${}_A A, B_B$; in this case we say that A is a ring with (Morita-) duality.

Rings R satisfying any of the following conditions have been called (right-) QF-3 ([3], [6], [18]):

- (1) There exists a minimal faithful right-module (a module which is a direct summand in every faithful module);
- (2) there exists an injective projective faithful right-ideal;
- (3) there exists an injective projective faithful right-module;
- (4) H is torsionless.

¹⁾ This work was supported in part by the National Research Council of Canada, Grant A-4033.

²⁾ The corresponding quantities obtained from the R -left-module R will be denoted by H', A', Q', U', B' .

The implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) hold for any ring; and if R is left-perfect all are equivalent. A left-perfect ring has $\text{r.dom.dim. } R > 0$ if and only if it is QF -3 and its right-socle is finite.

Lemma 1. *If $\text{r.dom.dim. } R \geq n$, then there exists an exact sequence*

$$0 \rightarrow R \rightarrow X_1 \rightarrow \cdots \rightarrow X_n$$

of finitely generated projective injective right modules.

Proof. Proposition 2.4 by Faith and Walker [5] shows that the injective hull of a finitely generated module is finitely generated if it is projective. It follows by induction that the first n terms of the minimal injective resolution of R_R , being projective if $\text{r.dom.dim. } R \geq n$, are finitely generated.

Lemma 2. *If $\text{r.dom.dim. } R > 0$ and $R = Q$, then $\text{r.dom.dim. } R \geq 2 + n$ if and only if there exists an exact sequence $P_{n+2} \rightarrow \cdots \rightarrow P_1 \rightarrow H \rightarrow 0$ of finitely generated projective A -left-modules P_i , and $\text{Ext}_A^k(H, H) = 0$ for all $0 \leq k \leq n$.*

Proof. $\text{r.dom.dim. } R \geq 2 + n$ yields an exact sequence $0 \rightarrow R \rightarrow H \rightarrow X_2 \rightarrow \cdots \rightarrow X_{2+n}$ of finitely generated projective injective R -right-modules, and the functor $\text{Hom}(-, H_R)$ produces ${}_A\text{Hom}_R(X_{2+n}, H) \rightarrow \cdots \rightarrow {}_A\text{Hom}_R(H, H) = A \rightarrow {}_A H \rightarrow 0$ where ${}_A\text{Hom}_R(X_i, H)$ is direct in $\oplus_j {}_A\text{Hom}_R(H, H) = \oplus_j {}_A A$ hence finitely generated projective, since $X_i < \oplus_j R < \oplus_j H$ implies that X_i is direct in $\oplus_j H$. Applying $\text{Hom}_A(-, H)$ to compute $\text{Ext}_A^k(H, H)$ for $1 \leq k \leq n$, and observing

$$\text{Hom}_A(\text{Hom}_R(X_i, H), H)_R \cong X_i \otimes_R \text{Hom}_A(H, H)_R = X_i \otimes_R R_R \cong X_{i,R},$$

we get $\text{Ext}_A^k(H, H) = 0$ for $1 \leq k \leq n$.

Conversely the conditions imply that the sequence

$$0 \rightarrow \text{Hom}_A(H, H)_R \rightarrow \text{Hom}_A(P_1, H)_R \rightarrow \cdots \rightarrow \text{Hom}_A(P_{2+n}, H)_R$$

is exact, and since the P_i are finitely generated projective and H_R is projective injective, $\text{Hom}_A(P_i, H)_R$ is projective injective and $\text{r.dom.dim. } R \geq 2 + n$.

Lemma 3. *If $\text{r.dom.dim. } R > 0$, then H_Q is the injective hull of Q_Q and*

$$\text{r.dom.dim. } Q > 0.$$

Proof. cf. Lambek [9]; the projectivity of H_R implies the projectivity of H_Q since every R -map $H \rightarrow R$ is a Q -map (dual-basis-lemma).

Lemma 4. *If $\text{r.dom.dim. } R > 0$, then $H \otimes_R H^* \cong U$ as A — B -bimodule and $\text{End}({}_R H^*) \cong B$ as ring. If furthermore ${}_A H$ is U -reflexive, then $\text{End}(H_B^*) \cong Q$ as ring.*

Proof. $H \otimes_R H^* \cong \text{Hom}_A(\text{Hom}_R(H, H), U) = \text{Hom}_A(A, U) \cong U$ since H_R is finitely generated projective.

$$\text{End}({}_R H^*) = \text{Hom}_R(H^*, \text{Hom}_A(H, U)) \cong \text{Hom}_A(H \otimes_R H^*, U) \cong \text{Hom}_A(U, U) = B;$$

and

$$\begin{aligned} \text{End}(H_B^*) &= \text{Hom}_B(H^*, \text{Hom}_A(H, U)) \cong \text{Hom}_A(H, \text{Hom}_B(H^*, U)) \\ &= \text{Hom}_A(H, H^{**}) \cong \text{Hom}_A(H, H) = Q. \end{aligned}$$

Lemma 5. *If X, Y are similar modules (each is a direct summand in a finite direct sum of copies of the other), then there exists a category-equivalence between $\text{End } X$ - and $\text{End } Y$ -modules that carries X into Y . The second commutators of X, Y are isomorphic.*

Proof. Since X, Y are both similar to $X \oplus Y$, we may confine ourselves to the situation $X \oplus X' = Y, Y \oplus Y' \cong \oplus_f X$. Let $e \in \text{End } Y$ be the projection of Y onto X with kernel X' , then $\text{End } X = e \text{End } Y e$ and $X = eY$. The $\text{End } Y$ -isomorphisms $\text{Hom}(Y \oplus Y', Y) \cong \oplus_f \text{Hom}(X, Y) = \oplus_f \text{Hom}(eY, Y) \cong \oplus_f \text{Hom}(Y, Y)e = \oplus_f \text{End } Ye$ and $\text{Hom}(Y \oplus Y', Y) \cong \text{End } Y \oplus \text{Hom}(Y', Y)$ show that $\text{End } Ye$ is an $\text{End } Y$ -progenerator, with endomorphism-ring $e \text{End } Y e \cong \text{End } X$. Therefore $\text{End } Ye \otimes_{\text{End } X} —$ is a category-equivalence, and its inverse

$$\text{Hom}_{\text{End } Y}(\text{End } Ye, \text{End } Y) \otimes_{\text{End } Y} — \cong e \text{End } Y \otimes_{\text{End } Y} —$$

maps Y onto $e \text{End } Y \otimes_{\text{End } Y} Y \cong eY = X$.

2. l.dom.dim. $R = \text{r.dom.dim. } R$

Theorem 6. *Let R be two-sided perfect, and $\text{r.dom.dim. } R > 0$. Then $\text{l.dom.dim. } R > 0$ if and only if A is a left-artinian ring with duality.*

Proof. Suppose first $\text{l.dom.dim. } R > 0$ so we have a left-right-symmetric situation. We observe that H_R finitely generated projective implies $A = \text{End}(H_R) \cong e R_n e$ where R_n is the $n \times n$ -matrix-ring over R and e is an idempotent in R_n ; hence A is two-sided perfect and $B = \text{End}({}_A U)$ is certainly semi-perfect. The number of isomorphism-types of indecomposable direct summands of ${}_R H^*$ coincides with the number of isomorphism-types of primitive idempotents of $\text{End}({}_R H^*) \cong B$; similarly for ${}_R H'$ and A' . Furthermore this number for A equals the number of simple A -left-modules, hence the number of (non-isomorphic) indecomposable direct summands of ${}_A U$ and the number of non-isomorphic primitive idempotents of $\text{End}({}_A U) = B$.

The module ${}_R H^*$ is faithful since ${}_A U$ is an injective cogenerator. Hence ${}_R R$ is contained in a direct product of copies of ${}_R H^*$ which is injective and therefore contains ${}_R H'$. The indecomposable direct summands of H' are injective hulls of simple R -left-ideals; therefore each of them is mapped monomorphically by some projection of the above product, and H^* contains a copy of each isomorphism-type of such injective hulls. Their (finite) sum in H^* is direct since they have non-isomorphic socles; hence it is injective and direct in H^* . We have obtained that the number of isomorphism-types of indecomposable direct summands of H' is not greater than that of H^* , which means that A' doesn't have more isomorphism-types of primitive idempotents than A . Symmetry implies equality, and therefore ${}_R H'$ and ${}_R H^*$ are similar.

We will show that U_B is a finitely generated cogenerator; this implies that A has duality (cf. [14]) and hence is left-artinian since it is perfect (Osofsky [16]). The isomorphism ${}_A U \cong {}_A H \otimes_R H^*$ and the projectivity of ${}_R H^*$ yield that the injective cogenerator ${}_A U$ is direct in a direct sum of copies of ${}_A H$, which is therefore itself a cogenerator; similarly H'_A is a cogenerator and so is H_B^* , since the similarity of ${}_R H'$ and ${}_R H^*$ leads to category-equivalence (Lemma 5). The exact functor $— \otimes_R H_B^*$ and $R_R < H_R$ give $H_B^* < H \otimes_R H_B^* \cong U_B$, so U_B is a cogenerator. $\text{Hom}_R(—, H^*)$ produces from $R_R < \oplus_f {}_R H^*$ a B -epimorphism $\oplus_f B \cong \oplus_f \text{End}({}_R H^*) \rightarrow \text{Hom}_R(R, H^*) \cong H_B^*$, hence H_B^* is finitely

generated. Then U_B is finitely generated because of $U_B \cong H \otimes_R H_B^*$ and H_R finitely generated projective.

Conversely we assume now that A is left-artinian with duality. As before we have ${}_R H^*$ injective, and it is also flat hence projective: ${}_A H$ is cyclic ($R_R < H_R$ and $\text{Hom}_R(-, H)$ yield an epimorphism ${}_A A \rightarrow {}_A H$) hence U -reflexive, and the exactness of

$$\text{Hom}_R(-, H) \cong \text{Hom}_R(-, H^{**}) = \text{Hom}_R(-, \text{Hom}_B(H^*, U)) \cong \text{Hom}_B(- \otimes_R H^*, U)$$

implies the exactness of $- \otimes_R H^*$ since U_B is a cogenerator. Since ${}_A H$ is artinian,

$\bigcap_{\alpha \in H^*} \text{Ker } \alpha = 0$ reduces to a finite intersection $\bigcap_{i=1}^n \text{Ker } \alpha_i = 0$, which implies ${}_R R < \bigoplus_f {}_R H^*$; hence $\text{l.dom.dim. } R > 0$.

Corollary 7. *In the situation of theorem 6, R is semi-primary.*

Proof. A left-artinian implies finite length for ${}_A H$, hence ascending and descending chain conditions for $\mathfrak{A}_l(H, R)$ (in the terminology of Faith [4]), hence the same for $\mathfrak{A}_r(H, R)$ which is equal to $\mathfrak{A}_r(R, R)$, the set of right-annulets of R , since $R_R < H_R < \Pi R_R$. The proposition on p. 185 of [4] yields the result.

Remark. Harada [6], Corollary 1 has shown that for a two-sided artinian ring R , $\text{r.dom.dim. } R > 0$ implies $\text{l.dom.dim. } R > 0$; therefore in this case A always has duality.

Lemma 8. *If $\text{r.dom.dim. } R > 0$ and A is left-artinian with duality, then the complete left-, right-quotient-rings Q' , Q of R are the same.*

Proof. $Q = \text{End}({}_A H)$ is semi-primary since ${}_A H$ is of finite length (Jacobson [7], Chap. 4), H_Q is the injective hull of Q_Q and projective (Lemma 3), and $\text{End}(H_Q) = A$, hence $\text{l.dom.dim. } Q > 0$ and ${}_Q H^*$ similar to the injective hull of ${}_Q Q$, by Theorem 6. We get $R < Q < \bigoplus_f H^*$ and ${}_R H'$ is direct in the injective module $\bigoplus_f {}_R H^*$:

$${}_R H' \oplus {}_R M \cong \bigoplus_f {}_R H^*.$$

Therefore ${}_R M$ is injective and an epimorphism $\bigoplus R \rightarrow M$ extends to $\bigoplus H' \rightarrow M$, consequently M is the sum of epimorphic images of H' . Bourbaki ([2], § 1) says that $H' \oplus M \cong \bigoplus_f H^*$ is actually a Q -decomposition (observe that the second commutator of $\bigoplus_f {}_R H^*$ is the same as that of ${}_R H^*$ which is Q , Lemma 4) and that we get an embedding of Q into the second commutator Q' of ${}_R H'$. Then ${}_Q H^*$ finitely generated projective yields $H^* < \bigoplus_f Q < \bigoplus_f Q' < \bigoplus_f H'$, and ${}_Q H^*$ is direct in $\bigoplus_f {}_Q H'$ since it is injective; altogether H^* is similar to H' as Q -, hence as R -module. Therefore their second commutators Q , Q' agree.

Lemma 9. *Consider the conditions:*

- (1) R is the endomorphism-ring of a left-generator-cogenerator of finite length;
- (2) $\text{r.dom.dim. } R \geq 2$;
- (3) $R = Q$.

Then (1) $\Rightarrow$ (2) $\Rightarrow$ (3) for any ring; and if R is semi-primary and $\text{r.dom.dim. } R, \text{l.dom.dim. } R > 0$, then they are equivalent and equivalent to their left-right-analoga.

Proof. If $R = \text{End}({}_C X)$ where ${}_C X$ is a generator-cogenerator of finite length, then C is left-artinian and possesses a left-injective cogenerator of finite length, hence it has duality (Morita [10], Th. 6. 3). (We see that (3) is left-right-symmetric:

$$R = \text{End}({}_C X) \cong \text{End}(X^*)$$

and X^* is a right-cogenerator-generator of finite length). An exact sequence $P_2 \rightarrow P_1 \rightarrow {}_C X \rightarrow 0$ with finitely generated projective P_i yields

$$0 \rightarrow {}_R R \rightarrow \text{Hom}_C(P_1, X) \rightarrow \text{Hom}_C(P_2, X),$$

and $\text{Hom}_C(P_i, X)_R$ is projective since ${}_C X$ is a generator, and injective (Morita [10], Th. 16. 2). Hence $\text{r.dom.dim. } R \geq 2$.

If $\text{r.dom.dim. } R \geq 2$, then H/R_R is torsionless, and for every $h \in H$, $h \notin R$ there exists $\alpha: H_R \rightarrow R_R < H_R$ (hence $\alpha \in A$) with $\alpha h \neq 0$, $\alpha R = 0$. Then $Q = \{h \in H \mid \alpha h = 0 \text{ for all } \alpha \in A, \alpha R = 0\}$ (Lambek [9]) = R .

If $R = Q = \text{End}({}_A H)$, and if the other assumptions are satisfied, then A is left-artinian with duality and $H_R, {}_R H^*$ are both finitely generated projective; but then ${}_A H, H_B^*$ are generators, ${}_A H$ is of finite length hence reflexive, and $\cong H^{**}$ hence a co-generator, by the duality.

Theorem 10. *If R is semi-primary and $\text{l.dom.dim. } R, \text{r.dom.dim. } R > 0$, then $\text{l.dom.dim. } R = \text{r.dom.dim. } R$.*

Proof. By the preceding lemma, either $\text{r.dom.dim. } R = \text{l.dom.dim. } R = 1$, or both dimensions are ≥ 2 , and $R = Q = Q'$. Then they are characterized by $\text{Ext}_A^k(H, H)$, $\text{Ext}_{A'}^k(H', H')$ (Lemma 2, observe A, A' artinian with duality). Actually ${}_R H'$ is similar to ${}_R H^*$, and this implies (Lemma 5) $\text{Ext}_{A'}^k(H', H') \cong \text{Ext}_B^k(H^*, H^*)$ which is $\cong \text{Ext}_A^k(H, H)$, by the duality.

3. Classification

If R is a semi-primary ring with $\text{dom.dim. } R \geq 2$, then $R = \text{End}({}_C X)$ where X is a generator-cogenerator and C is left-artinian with duality. From Morita [11] we know that this representation is essentially unique: if C is self-basic (which we may always assume) then C is determined up to isomorphism, and X up to semilinear isomorphism. X is the direct sum of the isomorphism-types of indecomposable projective or injective C -modules, and an arbitrary module Y of finite length. Hence the isomorphism-types of these rings R are classified by the pairs (C, Y) where C is an isomorphism-type of self-basic left-artinian rings with duality, and Y is a semilinear-isomorphism-type of C -left-modules of finite length.

To describe rings of dominant dimension one, we introduce as in [13], for any semi-primary ring S with $\text{dom.dim. } S \geq 2$, the class $C_0(S)$ of the proper subrings R of S which contain the identity 1 of S and appropriate injective projective faithful ideals eS, Sf . The same argument as in [13] shows that $C_0(S)$ consists, up to inner isomorphisms of S , precisely of the rings $S_0 < R \leq S$, where $S_0 = Z + eS + Sf + SfeS$ with “properly chosen” idempotents e, f ; Z is the subring generated by $1 \in S$.

Theorem 11. *$C_0(S)$ consists of the rings R satisfying $\text{r.dom.dim. } R = \text{l.dom.dim. } R = 1$ and having S as their right- and/or left-quotient-ring.*

Proof. If $R \in C_0(S)$ we see as in [13] (cf. Morita [10], Th. 17. 3), that R has the stated properties. Conversely if both dimensions are one, and if $S = Q$ is only the right-quotient-ring of R , then $A = \text{End}(H_R) = \text{End}(H_Q)$ which is left-artinian with duality, and Lemma 8 yields that the two quotient-rings are equal: $S = Q = Q'$. In $H < \oplus_f R < \oplus_f Q$ which is obtained from H_R finitely generated projective, H is actually a Q -submodule of $\oplus_f Q$: If x_k, δ_k are dual bases for H_R , then the embedding is explicitly

$H \ni h \rightarrow (\delta_k(h))_k \in \oplus R < \oplus Q$; and since for all $x \in H$ we have $x\delta_k \in A$, we get for $q \in Q$ $x\delta_k(hq) = (x\delta_k(h))q = x\delta_k(h)q$, hence $\delta_k(hq) = \delta_k(h)q$ since H_Q is faithful. Every isomorphism-type of indecomposable direct summands of H_Q , being the injective hull of a simple module, is mapped monomorphically onto some $X_i < R < Q$, by some projection $\oplus Q \rightarrow Q$. The sum of the X_i is direct hence Q -injective and of the form eQ with an idempotent e of Q . We have $eQ < R$, and eQ is obviously similar to H_Q hence an injective projective faithful Q -ideal.

Remark. A ring $R \in C_0(S)$ need not be semi-primary. Actually a ring R lies in $C_0(S)$ if and only if $\text{r.dom.dim. } R = \text{l.dom.dim. } R = 1$, and A is left-artinian with duality. This suggests to consider in general, among the rings R with $\text{r.dom.dim. } R, \text{l.dom.dim. } R > 0$, the larger class for which A is left-artinian with duality (instead of the semi-primary R). But this class is properly larger only if $\text{r.dom.dim. } R = \text{l.dom.dim. } R = 1$: If one dimension is ≥ 2 , then R is equal to its quotient-ring on that side, but then on both sides (Lemmata 9, 8) and the right-quotient-ring $\text{End } ({}_A H)$ is semi-primary since A is left-artinian.

4. Examples

$$(1) \quad R = \begin{bmatrix} K & K & K \\ & Z & K \\ & & K \end{bmatrix}$$

where Z is the ring of integers and K is the field of rationals, is a ring in $C_0(K_3)$ which is not semi-primary (cf. the preceding remark).

$$(2) \quad R = \begin{bmatrix} K & K \\ & Z \end{bmatrix}$$

with Z, K as before, is an example of a ring with $\text{r.dom.dim. } R > 0$, $A = K$ artinian with duality, but $\text{l.dom.dim. } R = 0$ (R not semi-primary) (cf. Theorem 6, Lemma 8).

(3) Let B be a semi-primary ring, U_B its minimal injective cogenerator and $A = \text{End } (U_B)$; then A and hence

$$R = \begin{bmatrix} B & 0 \\ U & A \end{bmatrix}$$

are semi-primary. R is right- QF -3; $\text{r.dom.dim. } R > 0$ holds if and only if the right-socle of B is finite; R is left- QF -3 and then $\text{l.dom.dim. } R > 0$ if and only if B has duality (with A). All these cases are possible (cf. Theorem 6).

(4) The algebra R over the field of rationals generated by 1 (identity), $e_k, x_k, k = 1, 2, \dots$ with $e_k^2 = e_k, e_k x_k = x_k e_{k-1} = x_k$, all other products zero (Osofsky [16]) has torsionless H , but no injective projective faithful right-module (cf. introductory remarks on QF -3 rings).

References

- [1] H. Bass, Finitistic dimension and a homological generalization of semi-primary rings, Trans. Amer. Math. Soc. **95** (1960), 466—488.
- [2] N. Bourbaki, *Eléments de Mathématique*, Algèbre Chap. 8, Paris 1958.
- [3] R. R. Colby and E. R. Rutter, Semiprimary QF -3 rings, Amer. Math. Soc. Notices **15** (1968), 164.
- [4] C. Faith, Rings with ascending chain condition on annihilators, Nagoya Math. J. **27** (1966), 179—191.

- [5] *C. Faith and E. Walker*, Direct sum representations of injective modules, *J. Algebra* **5** (1967), 203—221.
- [6] *M. Harada*, *QF-3* and semi-primary *PP*-rings I, II, *Osaka Math. J.* **2** (1965), 357—368; **3** (1966), 21—28.
- [7] *N. Jacobson*, *The Theory of Rings*, Amer. Math. Soc., New York 1943.
- [8] *J. P. Jans*, Projective injective modules, *Pac. J. Math.* **9** (1959), 1103—1108.
- [9] *J. Lambek*, On Utumi's ring of quotients, *Can. J. Math.* **15** (1963), 363—370.
- [10] *K. Morita*, Duality for modules and its applications to the theory of rings with minimum condition, *Science Reports Tokyo Kyoiku Daigaku* **6** (1958), 83—142.
- [11] *K. Morita*, Category-isomorphisms and endomorphism rings of modules, *Trans. Amer. Math. Soc.* **103** (1962), 451—469.
- [12] *B. J. Müller*, The classification of algebras by dominant dimension, *Can. J. Math.* **20** (1968), 398—409.
- [13] *B. J. Müller*, On algebras of dominant dimension one, *Nagoya J. Math.* **31** (1968), 173—183.
- [14] *B. J. Müller*, On Morita duality (to appear).
- [15] *T. Nakayama*, On algebras with complete homology, *Abh. Math. Sem. Univ. Hamburg* **22** (1958), 300—307.
- [16] *B. L. Osofsky*, A generalization of quasi-Frobenius rings, *J. Algebra* **4** (1966), 373—387.
- [17] *H. Tachikawa*, On dominant dimension of *QF-3* algebras, *Trans. Amer. Math. Soc.* **112** (1964), 249—266.
- [18] *L. Wu, H. Y. Mochizuki and J. P. Jans*, A characterization of *QF-3* rings, *Nagoya Math. J.* **27** (1966), 7—14.

Department of Mathematics, McMaster University, Hamilton, Ont., Canada

Eingegangen 12. Mai 1968

Structure theory for operators. I

By *R. G. Douglas*¹⁾ at Ann Arbor, Michigan

1. We begin in this paper the study of a structure theory for bounded operators on Hilbert space which is based on the notion of a canonical model. A canonical model for an operator is a “natural” representation of the operator in terms of simpler operators in a context in which more structure is present. We do not ask of such a representation that it make the answer to all questions transparent or even that all questions be answerable in terms of it. Indeed one aspect of such a study is the determination of the class of problems which can be successfully studied in its context. Thus it should not surprise us to find that we may need to study more than one such representation.

In the last several years a wealth of results have been obtained in operator theory that were explicitly or implicitly based on the notion of a canonical model. We are thinking particularly of the work of a number of Russian authors including M. S. Livšic, M. S. Brodskii and many others (cf. [4]) showing the existence of enough invariant subspaces for various non-self adjoint operators to establish integral representations for such operators, of the study of B. Sz.-Nagy and C. Foiaş of contractions ([14], [15], [17]—[24]), and the work of L. de Branges on Hilbert spaces of entire functions and of analytic functions (cf. [1]). (The preceding list is certainly not exhaustive and is included only to give some idea of the kind of operator theory in which we are interested.)

We have chosen to begin with an exposition of the Sz.-Nagy-Foiaş structure theory for contractions. This is perhaps the most elegant of the model theories to date. It is becoming clear, however, that while this theory provides a canonical model for every operator, this representation is not adequate to answer many questions in which we are interested. In particular, it seems to be the case that the structure of operators which are “almost unitary” is simpler or at least easier to discuss than that of an arbitrary operator. The Sz.-Nagy-Foiaş theory allows us, however, to study only such operators which in addition to being “almost unitary” are also contractions. (The technique of multiplying an operator by a constant less than one fails spectacularly in this instance because the resultant operator is no longer “almost unitary”.) We plan to consider more general models in a subsequent paper.

Our exposition of the structure theory of Sz.-Nagy and Foiaş is somewhat different from that which they gave. While the difference can be thought of as mainly one of emphasis and organization, we have been strongly influenced by several unpublished manuscripts of de Branges and Rovnyak. Also due to personal preference we prefer to

¹⁾ This research was performed while the author was a National Academy of Sciences Fellow supported by the Air Force Office of Scientific Research.

adopt a more geometrical point of view toward this theory than have either Sz.-Nagy and Foiaş or de Branges and Rovnyak. We feel that a certain economy of thought is gained in doing this and that certain aspects of the model are made more intuitive.

We do not complete our exposition of the Sz.-Nagy-Foiaş theory in this paper. This will be done in a second paper, an outline of which will be found at the end of this section. We now describe the contents of the present paper.

In § 2 we set forth certain notational conventions and describe various facts about bilateral and unilateral shifts which we shall need. The treatment of such things actually forms the first chapter in our study but we postpone the writing of such to a later date. A geometrical version of the Sz.-Nagy-Foiaş model is defined in § 3. The universality of this model for contractions is established in § 4, while in § 5 its relation to the strong unitary dilation is examined. We offer for completeness a full treatment of the latter notion based on our earlier results on canonical models. Thus we reverse the roles which the concepts of canonical model and strong unitary dilation played in the Sz.-Nagy-Foiaş derivation.

Finally, we introduce the characteristic operator function in § 6 and show how it can be used to define a canonical model. The explicit formula for the characteristic operator function in terms of the operator established in [22] is not mentioned. This along with the connection of the model with analytic functions will be considered in the second paper. A brief indication of what we plan to include in the second paper is: § 7 shift operators and vector-valued functions, § 8 the canonical model revisited, § 9 the spectrum, § 10 invariant subspaces and § 11 the de Branges-Rovnyak model and its connection with the Sz.-Nagy-Foiaş model.

We wish to express our appreciation to a number of friends who read and criticized an earlier draft of this paper.

2. Preliminaries and shift operators. We will consider only complex Hilbert spaces and operator will mean bounded linear operator. A subspace is a closed linear manifold and for $\mathfrak{M}$ a subspace of the Hilbert space $\mathfrak{H}$, we let $P_{\mathfrak{M}}$ denote the (orthogonal) projection of $\mathfrak{H}$ onto $\mathfrak{M}$. If $\mathfrak{M}$ is a linear submanifold of $\mathfrak{H}$, then we denote by $[\mathfrak{M}]^-$ the closure of $\mathfrak{M}$.

If T is an operator on $\mathfrak{H}$ and $\mathfrak{M}$ is a subspace of $\mathfrak{H}$, then $\mathfrak{M}$ is said to be an *invariant subspace* for T if $T\mathfrak{M} \subset \mathfrak{M}$. If we also have $T\mathfrak{M}^\perp \subset \mathfrak{M}^\perp$, then $\mathfrak{M}$ is said to be a *reducing subspace* for T . Lastly, a subspace $\mathfrak{M}$ is said to be a *full subspace* for T if the minimal reducing subspace for T containing $\mathfrak{M}$ is $\mathfrak{H}$; or equivalently, $\mathfrak{M}$ is a full subspace for T if and only if $\mathfrak{M}^\perp$ contains no nontrivial reducing subspace for T .

Let $\mathfrak{E}$ be a Hilbert space and $L_{\mathfrak{E}}$ denote the set of functions f from the integers $\mathbb{Z}$ to $\mathfrak{E}$ having the property that

$$\sum_{n=-\infty}^{\infty} \|f(n)\|^2 < \infty.$$

Then $L_{\mathfrak{E}}$ is a Hilbert space relative to pointwise addition and scalar multiplication, and the inner product

$$(f, g) = \sum_{n=-\infty}^{\infty} (f(n), g(n)).$$

The space $L_{\mathfrak{E}}$ can also be viewed as the direct sum of countably many copies of $\mathfrak{E}$. Our reasons for considering it as a space of functions will become apparent.

There is a natural operator U acting on $L_{\mathbb{G}}$ called the bilateral shift that is defined as follows: $(Uf)(n) = f(n-1)$ for f in $L_{\mathbb{G}}$ and n in $\mathbb{Z}$. It is easy to verify that U is a unitary operator and that its adjoint U^* merely "shifts" the values of a function in $L_{\mathbb{G}}$ in the other direction. There is a natural identification of $\mathbb{G}$ as a subspace of $L_{\mathbb{G}}$ so that an element x of $\mathbb{G}$ is mapped onto the function f , where $f(0) = x$ and $f(n) = 0$ for $n \neq 0$. We shall use $P_{\mathbb{G}}$ to denote the projection defined $P_{\mathbb{G}}f = f(0)$ for f in $L_{\mathbb{G}}$.

Let $H_{\mathbb{G}}$ denote the subspace of $L_{\mathbb{G}}$ consisting of those functions which vanish at the negative integers; more precisely, $H_{\mathbb{G}} = \{f \in L_{\mathbb{G}} \mid f(n) = 0 \text{ for } n < 0\}$. It is clear that $H_{\mathbb{G}}$ is an invariant subspace for U and that the operator $U_+ = U|_{H_{\mathbb{G}}}$ is an isometry. The operator U_+ is called the unilateral shift. We use $H_{\mathbb{G}}^{\perp}$ to denote $L_{\mathbb{G}} \ominus H_{\mathbb{G}}$.

If $\mathfrak{D}$ is a subspace of $\mathbb{G}$, then there is a natural way to identify $L_{\mathfrak{D}}$ and $H_{\mathfrak{D}}$ as subspaces of $L_{\mathbb{G}}$ and $H_{\mathbb{G}}$, respectively. Further, it is obvious that $L_{\mathfrak{D}}$ is a reducing subspace for U while $H_{\mathfrak{D}}$ is a reducing subspace for U_+ . Moreover, it turns out that every reducing subspace for U_+ is of the form $H_{\mathfrak{D}}$ for some subspace $\mathfrak{D}$ contained in $\mathbb{G}$. This is, of course, not the case for U since it is unitary.

A good reference concerning the basic facts about shift operators is [8].

3. Canonical models. We give in this section a geometrical definition of the Sz.-Nagy-Foiaş canonical model. This notion will be further refined in § 6 when we introduce the characteristic operator function. Along with giving the definition we now prove a technical lemma and then use it to establish the uniqueness of the canonical model.

Let $\mathfrak{H}$ be a Hilbert space and U_+ be the unilateral shift on $H_{\mathfrak{H}}$. Further, let $\mathfrak{K}$ be a Hilbert space and F be a unitary operator on $\mathfrak{K}$. We consider the operator $U_+^* \oplus F$ on $H_{\mathfrak{H}} \oplus \mathfrak{K}$ recalling that this is defined for $f \oplus g$ in $H_{\mathfrak{H}} \oplus \mathfrak{K}$ so that

$$(U_+^* \oplus F)(f \oplus g) = U_+^*f \oplus Fg.$$

We note that the action of powers of this operator eventually decreases the norm of the first component while preserving the norm of the second. Also from the result of von Neumann [26] on the decomposition of isometries, it follows that $(U_+^* \oplus F)^*$ is the most general isometry.

If $\hat{\mathfrak{H}}$ is a full invariant subspace for $U_+^* \oplus F$, then the operator $\hat{T} = (U_+^* \oplus F)|_{\hat{\mathfrak{H}}}$ is said to be a canonical model. (We use the circumflex because we will want to consider both an abstract operator T on $\mathfrak{H}$ and its canonical model $\hat{T}$ on $\hat{\mathfrak{H}}$ in later sections and this notation seems convenient.) The requirement that $\hat{\mathfrak{H}}$ be a full subspace is imposed to make the representation unique. It is obvious that the operator $\hat{T}$ is a contraction and we show in § 4 that every contraction is unitarily equivalent to a canonical model.

The preceding is a geometrical form of the Sz.-Nagy-Foiaş canonical model which was first suggested to us by an unpublished manuscript of de Branges and Rovnyak (cf. [2]).

We will establish the uniqueness of the model after proving the following lemma.

Lemma 1. *If $\hat{\mathfrak{H}}$ is an invariant subspace for $U_+^* \oplus F$ on $H_{\mathfrak{H}} \oplus \mathfrak{K}$, then $[P_{H_{\mathfrak{H}}} \hat{\mathfrak{H}}]^-$ and $[P_{\mathfrak{K}} \hat{\mathfrak{H}}]^-$ are invariant subspaces for U_+^* and F , respectively. More importantly, $\hat{\mathfrak{H}}$ is a full invariant subspace for $U_+^* \oplus F$ if and only if each of $[P_{H_{\mathfrak{H}}} \hat{\mathfrak{H}}]^-$ and $[P_{\mathfrak{K}} \hat{\mathfrak{H}}]^-$ is a full invariant subspace for the operators U_+^* and F , respectively.*

Proof. It is obvious that $[P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}]^-$ and $[P_{\mathfrak{F}} \hat{\mathfrak{H}}]^-$ are invariant subspaces and that if at least one is not a full subspace then neither is $\hat{\mathfrak{H}}$ since these statements do not depend upon the special nature of U_+^* and F but are valid for a direct sum of arbitrary operators.

Assume that $\mathfrak{G}$ is a reducing subspace for $U_+^* \oplus F$ containing $\hat{\mathfrak{H}}$. Then for $f \oplus g$ in $\mathfrak{G}$ we have

$$0 \oplus g = \lim_{n \rightarrow \infty} (U_+ \oplus F^*)^n (U_+^* \oplus F)^n (f \oplus g) \text{ is in } \mathfrak{G}$$

so that $[P_{\mathfrak{F}} \mathfrak{G}]^-$ is contained in $\mathfrak{G}$. This implies $[P_{H_{\mathfrak{L}}} \mathfrak{G}]^-$ is contained in $\mathfrak{G}$. But $[P_{H_{\mathfrak{L}}} \mathfrak{G}]^-$ reduces U_+^* and contains $[P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}]^-$, while $[P_{\mathfrak{F}} \mathfrak{G}]^-$ reduces F and contains $[P_{\mathfrak{F}} \hat{\mathfrak{H}}]^-$. Since $[P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}]^-$ and $[P_{\mathfrak{F}} \hat{\mathfrak{H}}]^-$ are full subspaces, it follows that $[P_{H_{\mathfrak{L}}} \mathfrak{G}]^- = H_{\mathfrak{L}}$ and $[P_{\mathfrak{F}} \mathfrak{G}]^- = \mathfrak{F}$ so that $\hat{\mathfrak{H}}$ is a full subspace for $U_+^* \oplus F$.

We now consider the uniqueness of canonical models. Suppose $\hat{T}_1 = (U_{1+}^* \oplus F_1)/\hat{\mathfrak{H}}_1$, where $\hat{\mathfrak{H}}_1$ is contained in $H_{\mathfrak{L}_1} \oplus \mathfrak{F}_1$, and $T_2 = (U_{2+}^* \oplus F_2)/\hat{\mathfrak{H}}_2$, where $\hat{\mathfrak{H}}_2$ is contained in $H_{\mathfrak{L}_2} \oplus \mathfrak{F}_2$, are canonical models so that T_1 is unitarily equivalent to T_2 . Let γ be an isomorphism from $\hat{\mathfrak{H}}_1$ onto $\hat{\mathfrak{H}}_2$ for which $\gamma T_1 = T_2 \gamma$. We will show that γ can be extended to an isomorphism Γ from $H_{\mathfrak{L}_1} \oplus \mathfrak{F}_1$ onto $H_{\mathfrak{L}_2} \oplus \mathfrak{F}_2$ so that $\Gamma \hat{\mathfrak{H}}_1 = \hat{\mathfrak{H}}_2$, $\Gamma H_{\mathfrak{L}_1} = H_{\mathfrak{L}_2}$, $\Gamma U_{1+}^* = U_{2+}^*$, $\Gamma \mathfrak{F}_1 = \mathfrak{F}_2$ and $\Gamma F_1 = F_2 \Gamma$. Canonical models related in this way are said to be isomorphic.

The extension Γ of γ is defined for $\{f_n\}_{n=0}^N$ in $\hat{\mathfrak{H}}_1$, so that

$$\Gamma \left(\sum_{n=0}^N (U_{1+} \oplus F_1^*)^n f_n \right) = \sum_{n=0}^N (U_{2+} \oplus F_2^*)^n \gamma f_n.$$

An easy computation

$$\begin{aligned} & \left\| \Gamma \left(\sum_{n=0}^N (U_{1+} \oplus F_1^*)^n f_n \right) \right\|^2 = \sum_{n,m=0}^N ((U_{2+} \oplus F_2^*)^n \gamma f_n, (U_{2+} \oplus F_2^*)^m \gamma f_m) \\ &= \sum_{\substack{n,m=0 \\ n \geq m}}^N (\gamma f_n, \gamma (U_{1+}^* \oplus F_1)^{n-m} f_m) + \sum_{\substack{n,m=0 \\ n < m}}^N (\gamma (U_{1+}^* \oplus F_1)^{m-n} f_n, \gamma f_m) \\ &= \left\| \sum_{n=0}^N (U_{1+} \oplus F_1^*)^n f_n \right\|^2 \end{aligned}$$

using the fact that both $U_{1+} \oplus F_1^*$ and $U_{2+} \oplus F_2^*$ are isometries, shows that Γ can be extended to an isomorphism between the minimal reducing subspaces for $U_{1+}^* \oplus F_1$, and $U_{2+}^* \oplus F_2$ containing $\hat{\mathfrak{H}}_1$ and $\hat{\mathfrak{H}}_2$, respectively. Thus, since both $\hat{\mathfrak{H}}_1$ and $\hat{\mathfrak{H}}_2$ are full subspaces, Γ is an isomorphism from $H_{\mathfrak{L}_1} \oplus \mathfrak{F}_1$ onto $H_{\mathfrak{L}_2} \oplus \mathfrak{F}_2$ which extends γ so that $\Gamma(U_{1+}^* \oplus F_1) = (U_{2+}^* \oplus F_2) \Gamma$.

If $f \oplus g$ is in $H_{\mathfrak{L}_1} \oplus \mathfrak{F}_1$ and $\Gamma(f \oplus g) = h \oplus k$ is in $H_{\mathfrak{L}_2} \oplus \mathfrak{F}_2$, then

$$\begin{aligned} \Gamma(0 \oplus g) &= \Gamma \left[\lim_{n \rightarrow \infty} (U_{1+} \oplus F_1^*)^n (U_{1+}^* \oplus F_1)^n (f \oplus g) \right] \\ &= \lim_{n \rightarrow \infty} (U_{2+} \oplus F_2^*)^n (U_{2+}^* \oplus F_2)^n (h \oplus k) = 0 \oplus k. \end{aligned}$$

Thus the remaining properties of Γ are easily established and we have proved:

Lemma 2. *Two canonical models are unitarily equivalent if and only if they are isomorphic.*

We state this as a lemma since it will be incorporated into the statement of Theorem 1 in the next section.

4. The representation theorem. Let $\mathfrak{H}$ be a Hilbert space and T be a contraction on $\mathfrak{H}$, that is, T is an operator on $\mathfrak{H}$ so that $\|T\| \leq 1$. We show in this section that T is unitarily equivalent to a canonical model. Toward this end we need to define several auxiliary operators and spaces.

Let D denote the positive square root of $I - T^*T$ and $\mathfrak{D}$ denote the closure of the range of D . Let $\mathfrak{L}$ be a Hilbert space isomorphic to $\mathfrak{D}$ and let λ be an isomorphism from $\mathfrak{D}$ onto $\mathfrak{L}$. Further, let U_+ be the unilateral shift on $H_{\mathfrak{L}}$. This will be the unilateral shift in the canonical model. Consider the mapping φ from $\mathfrak{H}$ to $H_{\mathfrak{L}}$ defined so that for f in $\mathfrak{H}$ and n in $\mathbb{Z}$ we have $(\varphi f)(n) = \lambda D T^n f$. We must show of course that φf is in $H_{\mathfrak{L}}$ but we postpone this till later.

We next obtain the unitary operator F in the canonical model we are constructing. From the operator inequality $T^{*n}(I - T^*T)T^n \geq 0$ which holds for each positive integer n , it follows that $\{T^{*n}T^n\}_{n=0}^{\infty}$ is a monotone decreasing sequence of positive operators. Thus there exists a positive contraction A on $\mathfrak{H}$ so that A^2 is the limit of the sequence $\{T^{*n}T^n\}_{n=0}^{\infty}$ in the strong operator topology. This operator is seen to satisfy the identity $T^{*n}A^2T^n = A^2$. Let $\mathfrak{A}$ denote the closure of the range of A , let $\mathfrak{F}^+$ be a Hilbert space isomorphic to $\mathfrak{A}$, and let μ be an isomorphism of $\mathfrak{A}$ onto $\mathfrak{F}^+$.

Consider the mapping F_+ defined on the image in $\mathfrak{F}^+$ of the range of A so that for f in $\mathfrak{H}$ we have $F_+(\mu A f) = \mu A T f$. The following computation shows that F_+ is not only well defined but is isometric: for f in $\mathfrak{H}$ we have

$$\|F_+(\mu A f)\|^2 = \|A T f\|^2 = (T^*A^2 T f, f) = \|A f\|^2 = \|\mu A f\|^2.$$

Thus F_+ can be uniquely extended to an isometry on $\mathfrak{F}^+$. Now every isometry has a unique minimal *extension* to a unitary operator. Let F on $\mathfrak{F}$ be such for F_+ on $\mathfrak{F}^+$, that is, $\mathfrak{F}^+$ is a subspace of $\mathfrak{F}$ and F is a unitary operator on $\mathfrak{F}$ so that $F_+ = F|_{\mathfrak{F}^+}$, and let μ now denote the isomorphism from $\mathfrak{A}$ into $\mathfrak{F}$.

Define a mapping ψ from $\mathfrak{H}$ to $\mathfrak{F}$ so that for f in $\mathfrak{H}$ we have $\psi f = \mu A f$. The following computation shows that not only are the mappings φ and ψ well defined but that the mapping $\xi = \varphi \oplus \psi$ from $\mathfrak{H}$ to $H_{\mathfrak{L}} \oplus \mathfrak{F}$ is an isomorphism. For f in $\mathfrak{H}$ we have

$$\begin{aligned} \|\xi f\|^2 &= \|\varphi f\|^2 + \|\psi f\|^2 \\ &= \lim_{N \rightarrow \infty} \left\{ \sum_{n=0}^N \|D T^n f\|^2 \right\} + \|A f\|^2 \\ &= \lim_{N \rightarrow \infty} \left\{ \sum_{n=0}^N (T^{*n}(I - T^*T)T^n f, f) \right\} + \|A f\|^2 \\ &= \|f\|^2 - \lim_{N \rightarrow \infty} (T^{*N+1}T^{N+1}f, f) + (A^2 f, f) = \|f\|^2. \end{aligned}$$

Thus ξ is an isomorphism from $\mathfrak{H}$ into $H_{\mathfrak{L}} \oplus \mathfrak{F}$ and another simple computation shows that $\xi T = (U_+^* \oplus F)\xi$. Hence, if we let $\hat{\mathfrak{H}}$ denote the range of ξ , then $\hat{\mathfrak{H}}$ is an invariant subspace for $U_+^* \oplus F$ and $\hat{T} = (U_+^* \oplus F)|_{\hat{\mathfrak{H}}}$ is unitarily equivalent to T . It remains only to show that $\hat{\mathfrak{H}}$ is a full subspace for $U_+^* \oplus F$ in order to conclude that T is unitarily equivalent to a canonical model.

Consider the linear manifold $P_{H_L}\hat{\mathfrak{H}}$ and let $H_{\mathfrak{G}}$ where $\mathfrak{G}$ is contained in $\mathfrak{L}$, be minimal reducing subspace for U_+^* containing $P_{H_{\mathfrak{L}}}\hat{\mathfrak{H}}$. Then $\mathfrak{G}$ contains the image under λ of the range of D . Since the latter is dense in $\mathfrak{L}$, $[P_{H_{\mathfrak{L}}}\hat{\mathfrak{H}}]^-$ is a full invariant subspace for U_+^* . Similarly, the linear manifold $P_{\mathfrak{H}}\hat{\mathfrak{H}}$ contains the image under μ of the range of A so that $[P_{\mathfrak{H}}\hat{\mathfrak{H}}]^-$ contains $\mathfrak{H}^+$. Since F is the minimal extension of F_+ we have that $[P_{\mathfrak{H}}\hat{\mathfrak{H}}]^-$ is a full invariant subspace for F . Thus we have shown that $\hat{T} = (U_+^* \oplus F)/\hat{\mathfrak{H}}$ is a canonical model. Combining this result with Lemma 2 we have:

Theorem 1. *Every contraction is unitarily equivalent to a unique canonical model.*

We now make several remarks after which we will conclude the section with a result concerning the connection between an operator and its canonical model.

The above theorem is due to Sz.-Nagy and Foiaş [22]. Our construction, however, is based in part on a construction of de Branges and Rovnyak [2]. This theorem and its proof also generalizes results due to Rota [10] and Foiaş [5].

Before continuing let us comment on our use of the auxiliary spaces $\mathfrak{L}$ and $\mathfrak{H}^+$. The reason for doing this is that from the natural identification of $\mathfrak{H}$ with $\hat{\mathfrak{H}}$ we obtain an identification of $\mathfrak{D}$ and $\mathfrak{A}$ with subspaces $\hat{\mathfrak{D}}$ and $\hat{\mathfrak{A}}$ of $H_{\mathfrak{L}} \oplus \mathfrak{H}$. These subspaces do not, however, coincide with $\mathfrak{L}$ and $\mathfrak{H}^+$ so we distinguished at the start between $\mathfrak{D}$ and $\mathfrak{L}$ and between $\mathfrak{A}$ and $\mathfrak{H}^+$ to avoid a possible source of confusion.

We conclude this section with the following corollary. It is this case which appears in [2].

Corollary. *Let $\hat{T} = (U_+^* \oplus F)/\hat{\mathfrak{H}}$, where $\hat{\mathfrak{H}}$ is contained in $H_{\mathfrak{L}} \oplus \mathfrak{H}$ be the canonical model for the contraction T on $\mathfrak{H}$. The subspace $\mathfrak{H} = \{0\}$ if and only if the sequence $\{T^n\}_{n=0}^{\infty}$ converges strongly to 0, while the subspace $\mathfrak{L} = \{0\}$ if and only if T is an isometry.*

Proof. The proof is immediate after we notice that $A = 0$ if and only if the sequence $\{T^n\}_{n=0}^{\infty}$ converges strongly to 0 and this follows from the identity $\|Af\| = \lim_{n \rightarrow \infty} \|T^n f\|$ which holds for f in $\mathfrak{H}$.

In particular, if T is a strict contraction ($\|T\| < 1$) on $\mathfrak{H}$, then the Corollary applies.

5. Strong unitary dilations. In this section we introduce the notion of a strong unitary dilation, establish its existence (using the existence of the canonical model) and uniqueness, and lastly, examine certain aspects of its relation to the operator itself. The latter analysis will be used in § 6.

Let W be a unitary operator on the Hilbert space $\mathfrak{R}$, $\mathfrak{H}$ be a subspace of $\mathfrak{R}$, and T be the compression of W to $\mathfrak{H}$, that is, $T = P_{\mathfrak{H}}W|_{\mathfrak{H}}$. Then W is said to be a *strong unitary dilation* of T if $T^{(n)} = P_{\mathfrak{H}}W^{(n)}|_{\mathfrak{H}}$ for all integers n , where we define $S^{(n)}$ for an operator S so that

$$S^{(n)} = \begin{cases} S^n & \text{if } n \geq 0 \\ S^{*-n} & \text{if } n < 0 \end{cases}.$$

The notion of strong unitary dilation is due to Sz.-Nagy [14] where he establishes the existence and uniqueness of such for all contractions. This notion specializes the concept of weak unitary dilation introduced by Halmos [7] for which only $T = P_{\mathfrak{H}}W|_{\mathfrak{H}}$ is assumed.

If W on $\mathfrak{K}$ is a strong unitary dilation of T on $\mathfrak{H}$ and $\mathfrak{K}_0$ is the minimal reducing subspace for W containing $\mathfrak{H}$, then $W/\mathfrak{K}_0$ is also a strong unitary dilation of T . If $\mathfrak{K} = \mathfrak{K}_0$, then W is said to be *minimal*. We first establish that if a contraction has a minimal strong unitary dilation, then it is unique. The argument is very similar to that given for Lemma 2 and will be only sketched.

If W_1 on $\mathfrak{K}_1$ and W_2 on $\mathfrak{K}_2$ are each minimal strong unitary dilations of T on $\mathfrak{H}$, then define a mapping Γ from $\mathfrak{K}_1$ to $\mathfrak{K}_2$ so that for each subset $\{f_n\}_{n=-N}^N$ of $\mathfrak{H}$ we set

$$\Gamma\left(\sum_{n=-N}^N W_1^n f_n\right) = \sum_{n=-N}^N W_2^n f_n.$$

An easy calculation shows that Γ is well defined and isometric, and hence can be extended to an isomorphism from $\mathfrak{K}_1$ onto $\mathfrak{K}_2$ for which $\Gamma W_1 = W_2 \Gamma$ and $\Gamma/\mathfrak{H}$ is the identity. Two such unitary dilations of T are said to be isomorphic. Consequently, we can speak of *the* minimal strong unitary dilation of a contraction after we have established that one exists. What we actually show is that each canonical model has a minimal strong unitary dilation.

Let $\hat{T} = (U_+^* \oplus F)/\hat{\mathfrak{H}}$ be a canonical model with $\hat{\mathfrak{H}}$ contained in $H_{\mathfrak{L}} \oplus \mathfrak{F}$. The minimal strong unitary dilation of $\hat{T}$ is $\hat{W} = U^* \oplus F$ on $\hat{\mathfrak{K}} = L_{\mathfrak{L}} \oplus \mathfrak{F}$. To see this let $f \oplus g$ be in $\hat{\mathfrak{H}}$; then for $n \geq 0$ we have

$$[P_{\hat{\mathfrak{H}}}(U^* \oplus F)^n - T^n](f \oplus g) = P_{\hat{\mathfrak{H}}}[(U^{*n} - U_+^{*n})f \oplus 0] = P_{\hat{\mathfrak{H}}}(P_{H_{\mathfrak{L}}}^\perp U^{*n}f \oplus 0) = 0$$

since $H_{\mathfrak{L}}^\perp$ is orthogonal to $\hat{\mathfrak{H}}$. Thus $\hat{T}^n = P_{\hat{\mathfrak{H}}} \hat{W}^n / \hat{\mathfrak{H}}$ for all positive n and taking adjoints we have $\hat{T}^{*n} = P_{\hat{\mathfrak{H}}} \hat{W}^{*n} / \hat{\mathfrak{H}}$ so that $\hat{W}$ is a strong unitary dilation of $\hat{T}$.

Now suppose $\hat{\mathfrak{K}}_0$ is a reducing subspace for $\hat{W}$ containing $\hat{\mathfrak{H}}$. If $f \oplus g$ is in $\hat{\mathfrak{H}}$, then $(U_+^* \oplus F)^n (f \oplus g)$ is in $\hat{\mathfrak{H}}$ so that

$$0 \oplus g = \lim_{n \rightarrow \infty} [(U \oplus F^*)^n (U_+^* \oplus F)^n (f \oplus g)] \text{ is in } \hat{\mathfrak{K}}_0.$$

Thus $[P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}]^-$ and $[P_{\mathfrak{F}} \hat{\mathfrak{H}}]^-$ are contained in $\hat{\mathfrak{K}}_0$ and because the latter is a full invariant subspace for F we have $\mathfrak{F}$ contained in $\hat{\mathfrak{K}}_0$. Further, since $\hat{\mathfrak{K}}_0 \cap L_{\mathfrak{L}}$ contains $P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}$ and reduces U , and $U_+ = U/H_{\mathfrak{L}}$, we have $\bigvee_{n=0}^{\infty} U_+^n [P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}]$ contained in $\hat{\mathfrak{K}}_0$. The latter is a subspace of $H_{\mathfrak{L}}$ which both reduces U_+^* and contains $P_{H_{\mathfrak{L}}} \hat{\mathfrak{H}}$; thus it must be $H_{\mathfrak{L}}$. Since $H_{\mathfrak{L}}$ is contained in $\hat{\mathfrak{K}}_0$, the subspace $L_{\mathfrak{L}}$ is also. Hence $\hat{\mathfrak{K}} = \hat{\mathfrak{K}}_0$ so that $\hat{W}$ is minimal and we have proved the following theorem.

Theorem 2. *Every contraction has a unique minimal strong unitary dilation. Moreover, the minimal strong unitary dilation of the canonical model $\hat{T} = (U_+^* \oplus F)/\hat{\mathfrak{H}}$ is $\hat{W} = U^* \oplus F$ on $L_{\mathfrak{L}} \oplus \mathfrak{F}$.*

We begin our analysis of the relation between $\hat{T}$ and $\hat{W}$ with the following lemma.

Lemma 3. *The subspace $\hat{\mathfrak{H}} \oplus \hat{H}_{\mathfrak{L}}^\perp$ is the minimal invariant subspace for $\hat{W}$ containing $\hat{\mathfrak{H}}$.*

Proof. Let $\mathfrak{G}$ denote the minimal invariant subspace for $\hat{W}$ which contains $\hat{\mathfrak{S}}$. If $f \oplus g$ is in $\hat{\mathfrak{S}}$ with f in $H_{\mathfrak{L}}$ and g in $\mathfrak{J}$, then

$$\hat{W}(f \oplus g) = U^*f \oplus Fg = (U^*_+f \oplus Fg) + (P_{H_{\mathfrak{L}}^\perp}U^*f \oplus 0)$$

so that $\hat{W}(f \oplus g)$ is in $\hat{\mathfrak{S}} \oplus H_{\mathfrak{L}}^\perp$. Thus $\hat{\mathfrak{S}} \oplus H_{\mathfrak{L}}^\perp$ is an invariant subspace for $\hat{W}$ and therefore contains $\mathfrak{G}$.

Further, since $\hat{T}(f \oplus g) = (U^*_+f \oplus Fg)$ is in $\hat{\mathfrak{S}}$, then

$$(P_{H_{\mathfrak{L}}^\perp}U^*f \oplus 0) = \hat{W}(f \oplus g) - \hat{T}(f \oplus g)$$

is in $\mathfrak{G}$ for $f \oplus g$ in $\hat{\mathfrak{S}}$. If we let $\mathfrak{J}$ denote the closure of the manifold $\{f(0) \mid f \oplus g \text{ in } \hat{\mathfrak{S}}\}$, then $\hat{\mathfrak{S}}$ is easily seen to be contained in $H_{\mathfrak{J}} \oplus \mathfrak{J}$. However, $\hat{\mathfrak{S}}$ is a full subspace for $U^*_+ \oplus F$ while $H_{\mathfrak{J}} \oplus \mathfrak{J}$ is a reducing subspace for $U^*_+ \oplus F$ so that $\mathfrak{J} = \mathfrak{L}$. The identity $f(0) = U P_{H_{\mathfrak{L}}^\perp}U^*f$ which holds for f in $H_{\mathfrak{L}}$ shows that $\mathfrak{J}$ contains $U^*\mathfrak{L}$ and since $\mathfrak{G}$ is an invariant subspace for U^* we can conclude that $\mathfrak{G}$ contains $H_{\mathfrak{L}}^\perp$. Thus $\mathfrak{G} = \hat{\mathfrak{S}} \oplus H_{\mathfrak{L}}^\perp$ and the proof is complete.

Before continuing we wish to compare the preceding lemma to a result of Sarason. If B is an operator defined on a Hilbert space $\mathfrak{H}$, $\mathfrak{S}$ is a subspace of $\mathfrak{H}$, and P is the projection onto $\mathfrak{S}$, then $\mathfrak{S}$ is said to be a semi-invariant subspace for B if $PB^nP = (PBP)^n$ for n a positive integer, that is, if B is a strong dilation of the operator PBP defined on $\mathfrak{S}$. In [11] Sarason showed that $\mathfrak{S}$ is a semi-invariant subspace for B if and only if there exist invariant subspaces $\mathfrak{M}$ and $\mathfrak{N}$ for B so that $\mathfrak{N} < \mathfrak{M}$ and $\mathfrak{S} = \mathfrak{M} \ominus \mathfrak{N}$. The preceding lemma establishes one half of this result for the special case of minimal strong unitary dilations. We point out that it would be possible to use Sarason's result along with the existence of a strong unitary dilation for a contraction and the structure theorem for isometries to reach this stage of our progress toward the Sz.-Nagy-Foiaş model.

We next obtain a further decomposition of the space $\hat{\mathfrak{H}}$. Observe that the adjoint T^* is also a contraction on $\mathfrak{H}$ and so has a canonical model and minimal strong unitary dilation based on it. If we let $\hat{W}_*$ on $\hat{\mathfrak{H}}_*$ denote this minimal strong unitary dilation of T^* , then $\hat{W}_*$ is also a minimal strong unitary dilation of T and so must be isomorphic to $\hat{W}$ on $\hat{\mathfrak{H}}$ by Theorem 2. Now we can write $\hat{W}_* = U^*_+ \oplus F_*$ acting on $\hat{\mathfrak{H}}_* = L_{\mathfrak{L}'} \oplus \mathfrak{J}_*$. If we identify $\hat{\mathfrak{H}}$ and $\hat{\mathfrak{H}}_*$ using the isomorphism of Theorem 2 we have $L_{\mathfrak{L}'}$ contained in $\hat{\mathfrak{H}}$ and the bilateral shift U_* on $L_{\mathfrak{L}'}$ is the restriction of $\hat{W}$ to $L_{\mathfrak{L}'}$.

Since the bilateral shift U is the restriction of $\hat{W}^*$ to $L_{\mathfrak{L}}$, U and U_* act on $\hat{\mathfrak{H}}$ in "opposite directions". The concept of characteristic operator function is made more intuitive, however, if the two bilateral shifts involved act in the "same direction". In particular, it can be thought of as an analytic operator valued function in this case. Therefore, we proceed to "turn $L_{\mathfrak{L}'}$ around". This involves two things; firstly, we consider the mapping $n \rightarrow -n$ on the domain and secondly, we shift the coordinates by one. The use of the notation $L_{\mathfrak{L}'}$ for the original function space and $L_{\mathfrak{L}_*}$ for the altered one is strictly speaking not quite accurate. We observe, however, that in view of the convention adopted in § 2 identifying $\mathfrak{L}'_*$ and $\mathfrak{L}_*$ as subspaces of $L_{\mathfrak{L}'}$ and $L_{\mathfrak{L}_*}$, respectively, $\mathfrak{L}'_*$ and $\mathfrak{L}_*$ are not the same.

Being more precise, we imbed $L_{\mathfrak{L}_*}$ in $\hat{\mathfrak{H}}$ so that the function f in $L_{\mathfrak{L}_*}$ is mapped onto the function $\tilde{f}$ in $L_{\mathfrak{L}'}$ defined $\tilde{f}(n) = f(-n-1)$ for each integer n . The bilateral

shift U_* on $L_{\mathfrak{L}_*}$ is equal to $\hat{W}^*|L_{\mathfrak{L}_*}$ and under this transformation $H_{\mathfrak{L}_*}$ is mapped onto $H_{\mathfrak{L}_*}^\perp$.

We add that the relation between $L_{\mathfrak{L}}$ and $L_{\mathfrak{L}_*}$ is complex and it is this relation we study in the next section with the aid of the characteristic operator function. The following lemma establishes the promised decomposition of $\hat{\mathfrak{H}}$.

Lemma 4. $\hat{\mathfrak{H}} = H_{\mathfrak{L}_*} \oplus \hat{\mathfrak{H}} \oplus H_{\mathfrak{L}}^\perp$.

Proof. This identity is verified by showing that $H_{\mathfrak{L}_*} \oplus \hat{\mathfrak{H}} = H_{\mathfrak{L}} \oplus \mathfrak{H}$. Observe first that a subspace of $H_{\mathfrak{L}} \oplus \mathfrak{H}$ is invariant for $\hat{W}^* = U \oplus F^*$ if and only if it is invariant for $U_+ \oplus F^*$. Therefore, the minimal invariant subspace for $\hat{W}^*$ containing $\hat{\mathfrak{H}}$ must reduce $U_+^* \oplus F$ and hence must equal $H_{\mathfrak{L}} \oplus \mathfrak{H}$ since $\hat{\mathfrak{H}}$ is full. On the other hand from Lemma 3 it follows that $H_{\mathfrak{L}_*} \oplus \hat{\mathfrak{H}} (= H_{\mathfrak{L}_*}^\perp \oplus \hat{\mathfrak{H}})$ is the minimal invariant subspace for $\hat{W}^* (= \hat{W}_*)$ containing $\hat{\mathfrak{H}}$. Thus the result follows.

Note that each of the subspaces $\mathfrak{L}$ and $\mathfrak{L}_*$ is a wandering subspace [8] of $\hat{W}$. The preceding decomposition of $\hat{\mathfrak{H}}$ appears in [9] and [19], and is the geometrical key to the Schaffer matrix construction of the minimal strong unitary dilation of a contraction [12], and especially to the modification of that due to Sz.-Nagy [16]. We again point out that an alternate approach to this theory could be based on the Sz.-Nagy construction in [16].

We conclude this section with a final lemma after we introduce the following terminology. A contraction T on the Hilbert space $\mathfrak{H}$ is said to be *completely nonunitary* if for no nontrivial reducing subspace $\mathfrak{M}$ for T is $T|_{\mathfrak{M}}$ a unitary operator. It is easy to show that the subspace $\{f|Af = A_*f, f = f\}$ of $\mathfrak{H}$ is the largest reducing subspace for T on which the restriction of T is a unitary operator and that T is completely nonunitary on its orthogonal complement. The existence of such a decomposition is due to Sz.-Nagy and Foiaş [18] and Langer [27] and reduces the study of contractions relative to many questions to the study of completely nonunitary contractions. We identify the completely nonunitary part of a canonical model in the following lemma.

Lemma 5. *If $\hat{T}$ is a canonical model on $\hat{\mathfrak{H}}$, then $L_{\mathfrak{L}}^\perp \cap L_{\mathfrak{L}_*}^\perp = \mathfrak{M}$ is the largest reducing subspace for $\hat{T}$ for which $\hat{T}|_{\mathfrak{M}}$ is a unitary operator. In particular, $\hat{T}$ is a completely nonunitary contraction if and only if $L_{\mathfrak{L}} + L_{\mathfrak{L}_*}$ is dense in $\hat{\mathfrak{H}}$.*

Proof. The subspace $\mathfrak{M}$ is a reducing subspace for $\hat{W}$ which is contained in $\hat{\mathfrak{H}}$ by Lemma 4. Thus $\hat{T}|_{\mathfrak{M}}$ is a unitary operator. Moreover, if $\mathfrak{N}$ is a reducing subspace for $\hat{T}$ for which $\hat{T}|_{\mathfrak{N}}$ is a unitary operator, then it follows that $\mathfrak{N} \subset \mathfrak{H}$ and similarly $\mathfrak{N} \subset \mathfrak{H}_*$. Therefore, $\mathfrak{N}$ is contained in $\mathfrak{M}$ and the lemma follows.

Essentially this result appears in [19].

6. Characteristic operator functions. We consider in this section the characteristic operator function of an operator and its relation to the canonical model. We introduce this function in terms of the model and leave the derivation of the explicit form for it in terms of the operator to a later paper. Also we continue to stress the geometry of the situation and we will not at this stage consider any analytical aspects.

We first define the abstract notion of a characteristic operator function and show how we can use it to define a canonical model. Secondly, we show how to obtain from a canonical model a characteristic operator function so that if we repeat the first step

we will obtain an isomorphic model. We state these results as Theorem 3 and then derive several corollaries.

Let $\mathfrak{E}$ and $\mathfrak{E}_*$ be Hilbert spaces and Θ be an operator from $L_{\mathfrak{E}_*}$ to $L_{\mathfrak{E}}$ satisfying: (1) $\|\Theta\| \leq 1$, (2) $\Theta U_* = U\Theta$, (3) $\Theta H_{\mathfrak{E}_*} \subseteq H_{\mathfrak{E}}$, and (4) for $0 \neq x$ in $\mathfrak{E}_*$ we have $\|P_{\mathfrak{E}}\Theta x\| < \|x\|$. Such a triple $(\mathfrak{E}, \mathfrak{E}_*, \Theta)$ satisfying (1), (2), (3), and (4) is said to be a characteristic operator function. Condition (4) corresponds to the “purity” hypothesis of Sz.-Nagy and Foiaş [22] and is imposed to obtain uniqueness.

We now proceed to obtain the canonical model defined by $(\mathfrak{E}, \mathfrak{E}_*, \Theta)$.

If Δ denotes the positive square root of $I - \Theta^*\Theta$, then Δ is a positive contraction on $L_{\mathfrak{E}_*}$ which commutes with U_*^* . Moreover, if we let $\bar{\Delta}$ denote the projection of $L_{\mathfrak{E}_*}$ onto the closure of the range of Δ , then $\bar{\Delta}$ commutes with U_*^* , $\bar{\Delta}L_{\mathfrak{E}_*}$ is a reducing subspace for U_*^* , and $V = U_*^*\bar{\Delta}L_{\mathfrak{E}_*}$ is a unitary operator. Consider the operator $U_+^* \oplus V$ on $H_{\mathfrak{E}} \oplus \bar{\Delta}L_{\mathfrak{E}_*}$ and the subspace

$$\hat{\mathfrak{S}} = (H_{\mathfrak{E}} \oplus \bar{\Delta}L_{\mathfrak{E}_*}) \ominus \{\Theta f \oplus \Delta f \mid f \text{ in } H_{\mathfrak{E}_*}\}.$$

Further, define the mapping ζ from $L_{\mathfrak{E}_*}$ to $L_{\mathfrak{E}} \oplus \bar{\Delta}L_{\mathfrak{E}_*}$ so that for f in $L_{\mathfrak{E}_*}$ we have $\zeta f = \Theta f \oplus \Delta f$. Then ζ is an isomorphism for which $\zeta U_*^* = (U \oplus V^*)\zeta$ so that $\zeta H_{\mathfrak{E}_*}$ is an invariant subspace for $U_+ \oplus V^*$. But $\hat{\mathfrak{S}} = (H_{\mathfrak{E}} \oplus \bar{\Delta}L_{\mathfrak{E}_*}) \ominus \zeta H_{\mathfrak{E}_*}$ so that $\hat{\mathfrak{S}}$ is an invariant subspace for $U_+^* \oplus V$ and it remains only to show that $\hat{\mathfrak{S}}$ is a full subspace for $U_+^* \oplus V$ to conclude that $U_+^* \oplus V/\hat{\mathfrak{S}}$ is a canonical model.

The subspace $\hat{\mathfrak{S}}$ is a full subspace for $(U_+^* \oplus V)$ if and only if $\{\Theta h \oplus \Delta h \mid h \text{ in } H_{\mathfrak{E}_*}\}$ contains no nontrivial reducing subspace for $U_+^* \oplus V$. If the vector $\Theta f \oplus \Delta f$ is in a reducing subspace for $U_+^* \oplus V$ contained in $\{\Theta h \oplus \Delta h \mid h \text{ in } H_{\mathfrak{E}_*}\}$, then so is

$$0 \oplus \Delta f = \lim_{n \rightarrow \infty} [(U_+ \oplus V^*)^n (U_+^* \oplus V)^n (\Theta f \oplus \Delta f)].$$

If $\Delta f \neq 0$, then there exists g in $H_{\mathfrak{E}_*}$ so that $\Theta g = 0$ and $\Delta g = g = \Delta f$. Therefore, the subspace $\{g/g \text{ in } H_{\mathfrak{E}_*} \text{ and } \Delta g = g\}$ contains a nontrivial reducing subspace for U_*^* since its image under ζ does for $0 \oplus V$. This is impossible since $H_{\mathfrak{E}_*}$ contains no nontrivial reducing subspace for U_*^* so that we must have $\Delta f = 0$.

If $\Delta f = 0$, then $\Theta^*\Theta f = f$ so that subspace $\{\Theta f \oplus 0 \mid f \text{ in } H_{\mathfrak{E}_*} \text{ and } \Theta^*\Theta f = f\}$ contains a nontrivial reducing subspace for $U_+^* \oplus 0$. Thus there exists x in $\mathfrak{E}$ and g in $H_{\mathfrak{E}_*}$ so that $\Theta g = x$ and $\Theta^*\Theta g = g$ which implies $\|P_{\mathfrak{E}}\Theta g\| = \|g\|$. If we set $g = y + U_*h$, where y is in $\mathfrak{E}_*$ and h is in $H_{\mathfrak{E}_*}$, then $P_{\mathfrak{E}}\Theta g = P_{\mathfrak{E}}\Theta y + P_{\mathfrak{E}}\Theta U_*h = P_{\mathfrak{E}}\Theta y + P_{\mathfrak{E}}U\Theta h = P_{\mathfrak{E}}\Theta y$, since $U\Theta h$ is in $U\Theta H_{\mathfrak{E}_*} \subset UH_{\mathfrak{E}}$ and the latter subspace is orthogonal to $\mathfrak{E}$. Moreover, the resulting inequality $\|y\| \geq \|P_{\mathfrak{E}}\Theta y\| = \|P_{\mathfrak{E}}\Theta g\| = \|g\|$ implies $g = y$. This contradicts (4) so that $\hat{\mathfrak{S}}$ is a full subspace for $U_+^* \oplus V$ and $\hat{T} = U_+^* \oplus V/\hat{\mathfrak{S}}$ is a canonical model.

We remark before continuing that the canonical model just constructed is a completely nonunitary contraction. To see this observe that it follows from Lemma 4 and the uniqueness of canonical models that for $\hat{T}$ we have $L_{\mathfrak{L}} = L_{\mathfrak{E}}$ and $L_{\mathfrak{L}_*} = \zeta L_{\mathfrak{E}_*}$ so that $[L_{\mathfrak{L}} + L_{\mathfrak{L}_*}]^- = [L_{\mathfrak{E}} + \zeta L_{\mathfrak{E}_*}]^- = L_{\mathfrak{E}} \oplus \bar{\Delta}L_{\mathfrak{E}_*}$ and $\hat{T}$ is completely nonunitary by Lemma 5.

We have thus shown how to define a canonical model in terms of a characteristic operator function. We now go the other way. Starting with a canonical model we define

a characteristic operator function for which the previous construction recovers an isomorphic canonical model. This will establish a one-to-one and onto correspondence between canonical models of completely nonunitary contractions and characteristic operator functions.

Let T be a completely nonunitary contraction on $\mathfrak{H}$, D be the positive square root of $I - T^*T$, and D_* be the positive square root of $I - TT^*$. Further, let $\mathfrak{D}$ and $\mathfrak{D}_*$ denote the closures of the ranges of D and D_* , respectively, and let λ and λ'_* denote the isomorphisms from $\mathfrak{D}$ and $\mathfrak{D}_*$ on to $\mathfrak{L}$ and $\mathfrak{L}'_*$ defined in § 4. (To recall the definition of $\mathfrak{L}'_*$ and its relation to $\mathfrak{L}_*$ see the discussion preceding Lemma 4 in § 5.) The mapping $\lambda_* = U_*\lambda'_*$ is an isomorphism of $\mathfrak{D}_*$ onto $\mathfrak{L}_*$.

Let A and A_* be the unique extensions of λ and λ_* to mappings from $L_{\mathfrak{D}}$ and $L_{\mathfrak{D}_*}$ into $\hat{\mathfrak{H}}$ so that each is an isomorphism for which $AU = \hat{W}^*A$ and $A_*U_* = \hat{W}^*A_*$. The ranges of A and A_* are $L_{\mathfrak{L}}$ and $L_{\mathfrak{L}_*}$. Let Θ denote the operator from $L_{\mathfrak{D}_*}$ to $L_{\mathfrak{D}}$ defined $\Theta = A^*A_*$. Then $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ is the desired characteristic operator function.

We first establish that it satisfies (1) — (4). It is easily seen that (1) $\|\Theta\| \leq 1$ and (2) $\Theta U_* = U\Theta$. Further, the relation (3) $\Theta H_{\mathfrak{D}_*} \subset H_{\mathfrak{D}}$ holds because

$$\Theta H_{\mathfrak{D}_*} = A^*H_{\mathfrak{L}_*} \subset A^*(H_{\mathfrak{L}} \oplus \mathfrak{F}) = H_{\mathfrak{D}},$$

where the inclusion follows from Lemma 4. Lastly, if x is in $\mathfrak{D}_*$ so that $\|P_{\mathfrak{D}}\Theta x\| = \|x\|$, then $\|P_{\mathfrak{L}}A_*x\| = \|x\| = \|A_*x\|$. Thus, A_*x is in $H_{\mathfrak{L}_*}$ and also in $\mathfrak{L}$ so that $H_{\mathfrak{G}}$, where $\mathfrak{G}$ is the subspace of $\mathfrak{L}$ spanned by A_*x , is orthogonal to $\hat{\mathfrak{F}}$ by Lemma 4. Then $H_{\mathfrak{G}}$ is a reducing subspace for $U_+^* \oplus F$ orthogonal to $\hat{\mathfrak{F}}$ and hence must be trivial since $\hat{\mathfrak{F}}$ is full. Thus $x = 0$ and we have established (4) if $0 \neq x$ is in $\mathfrak{D}_*$, then $\|P_{\mathfrak{D}}\Theta x\| < \|x\|$. Therefore, $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ is a characteristic operator function and we next show that the canonical model it defines is unitarily equivalent to T .

Consider the mapping ω from $A L_{\mathfrak{D}_*}$ to $\hat{\mathfrak{H}}$ defined so that for f in $L_{\mathfrak{D}_*}$ we set $\omega(Af) = (A_* - A\Theta)f = (I - P_{L_{\mathfrak{L}}})A_*f$. We show that ω is well-defined by proving that it is isometric as follows: for f in $L_{\mathfrak{D}_*}$ we have

$$\begin{aligned} \|\omega(Af)\|^2 &= \|(I - P_{L_{\mathfrak{L}}})A_*f\|^2 = \|A_*f\|^2 - \|P_{L_{\mathfrak{L}}}A_*f\|^2 \\ &= \|f\|^2 - (P_{L_{\mathfrak{L}}}A_*f, A_*f) = \|f\|^2 - \|\Theta f\|^2 = \|Af\|^2. \end{aligned}$$

Thus ω can be extended to an isomorphism from $\bar{A}L_{\mathfrak{D}_*}$ into $\mathfrak{F} = (I - P_{L_{\mathfrak{L}}})\hat{\mathfrak{H}}$. Let Ω be the isomorphism $A \oplus \omega$ from $L_{\mathfrak{D}} \oplus \bar{A}L_{\mathfrak{D}_*}$ into $\hat{\mathfrak{H}}$. Note that $\Omega(U_+^* \oplus V) = (U_+^* \oplus F)\Omega$ and it remains only to show that Ω is onto and to identify the “subspaces $\hat{\mathfrak{F}}$ ” in each model.

If f is in $L_{\mathfrak{L}_*}$, then A_*^*f is in $L_{\mathfrak{D}_*}$ and $\Omega(\Theta A_*^*f \oplus A A_*^*f) = f$ so that $L_{\mathfrak{L}_*}$ is in the range of Ω . Since T is completely nonunitary it follows from Lemma 5 that Ω is onto. Moreover, it is clear that Ω takes the subspace $\{\Theta f \oplus Af \mid f \text{ in } H_{\mathfrak{D}_*}\}$ of $L_{\mathfrak{D}} \oplus \bar{A}L_{\mathfrak{D}_*}$ onto $H_{\mathfrak{L}_*}$ and $H_{\mathfrak{F}}$ onto $H_{\mathfrak{L}}$ so that “ $\Omega\hat{\mathfrak{F}} = \hat{\mathfrak{F}}$ ”. Thus the canonical model defined by $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ is isomorphic to the one we started with which was unitarily equivalent to T . We can summarize what we have established in the following theorem.

Theorem 3. *If T is a completely nonunitary contraction on the Hilbert space $\mathfrak{H}$, $\mathfrak{D} = [\sqrt{I - T^*T} \mathfrak{H}]^-$, and $\mathfrak{D}_* = [\sqrt{I - TT^*} \mathfrak{H}]^-$, then there exists a characteristic*

operator function $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ so that T is unitarily equivalent to the canonical model defined by $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$. That is, if $\Delta = \sqrt{I - \Theta^* \Theta}$, $\bar{\Delta}$ is the projection onto $[\Delta L_{\mathfrak{D}_*}]^\perp$, $V = U_*^* / \bar{\Delta} L_{\mathfrak{D}_*}$, and $\hat{\mathfrak{H}} = (H_{\mathfrak{D}} \oplus \bar{\Delta} L_{\mathfrak{D}_*}) \ominus \{\Theta f \oplus \Delta f \mid f \text{ in } H_{\mathfrak{D}_*}\}$, then T is unitarily equivalent to $(U_+^* \oplus V) / \hat{\mathfrak{H}}$. Moreover, two completely nonunitary contractions are unitarily equivalent if and only if their characteristic operator functions are isomorphic.

Proof. We have proved everything except the last statement which will follow from our earlier uniqueness criterion in Theorem 1 after we have defined what we mean by isomorphic characteristic operator functions. Two such $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ and $(\mathfrak{E}, \mathfrak{E}_*, \Phi)$ are said to be isomorphic if there exist isomorphisms ξ and ξ_* between $\mathfrak{D}$ and $\mathfrak{E}$ and between $\mathfrak{D}_*$ and $\mathfrak{E}_*$, respectively, so that the isomorphisms Ξ and Ξ_* which ξ and ξ_* establish between $L_{\mathfrak{D}}$ and $L_{\mathfrak{E}}$ and between $L_{\mathfrak{D}_*}$ and $L_{\mathfrak{E}_*}$ are such that $\Xi\Theta = \Phi\Xi_*$.

It is easy to see that isomorphic characteristic operator functions yield unitarily equivalent models and only a little harder to show that unitarily equivalent models yield isomorphic characteristic operator functions. We leave the details to the reader.

It goes almost without saying that the preceding theorem is due to Sz.-Nagy and Foiaş [22].

Before continuing we want to say something about the relation which exists between the characteristic operator functions for T and T^* . If $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ is the characteristic operator function for T , then $(\mathfrak{D}_*, \mathfrak{D}, \Theta^\sim)$ is the characteristic operator function for T^* , where $\Theta^\sim$ is defined from $L_{\mathfrak{D}}$ to $L_{\mathfrak{D}_*}$ as follows: for f in $L_{\mathfrak{D}}$ we have

$$(\Theta^\sim f)(n) = (\Theta^* \tilde{f})(-n - 1)$$

where $\tilde{f}$ is defined as in the paragraph preceding Lemma 4. In other words, we are “turning around” both $L_{\mathfrak{D}}$ and $L_{\mathfrak{D}_*}$. We will have more to say about this in our next paper.

We next establish several corollaries to our theorem. Before stating the first we need to give a definition. A unitary operator is said to be *absolutely continuous* if it is unitarily equivalent to the restriction of a bilateral shift to a reducing subspace. The usual definition is that the spectral measure of the unitary operator is absolutely continuous with respect to Lebesgue measure on the unit circle. That this is the case for a part of a bilateral shift is obvious and it is a theorem that an absolutely continuous unitary operator can be extended to a bilateral shift.

Corollary 1. *The minimal strong unitary dilation of a completely nonunitary contraction is absolutely continuous.*

Proof. It follows from the Theorem that the minimal strong unitary dilation of a completely nonunitary contraction T is unitarily equivalent to $(U^* \oplus U_*^*) / L_{\mathfrak{D}} \oplus \bar{\Delta} L_{\mathfrak{D}_*}$.

This result is due to Sz.-Nagy and Foiaş [18], [19]. See [11] for a proof based on the notion of semi-invariant subspace.

Corollary 2. *Let T be a completely nonunitary contraction on $\mathfrak{H}$. If either of the sequences $\{T^n\}_{n=0}^\infty$ or $\{T^{*n}\}_{n=0}^\infty$ converges strongly to 0 or if one of the spaces $\mathfrak{D}$ and $\mathfrak{D}_*$ is infinite dimensional, then the minimal strong unitary dilation of T is a bilateral shift.*

Proof. If the sequence $\{T^n\}_{n=0}^\infty$ converges strongly to 0, then $\mathfrak{H} = \{0\}$ by the Corollary to Theorem 1 so that $\Delta = 0$. Thus the minimal strong unitary dilation of T is U^* on

$L_{\mathfrak{D}}$. Similarly, if the sequence $\{T^{**n}\}_{n=0}^{\infty}$ converges strongly to 0, then the minimal unitary dilation of T is U_* on $L_{\mathfrak{D}_*}$.

Assume now that $\mathfrak{D}$ is infinite dimensional and its dimension is greater than or equal to that of $\mathfrak{D}_*$. Then the multiplicity function [6] of $(U^* \oplus U_*)/L_{\mathfrak{D}} \oplus \overline{\Delta}L_{\mathfrak{D}_*}$ is easy to describe: it takes the value 0 at sets of Lebesgue measure 0 and the cardinality of $\mathfrak{D}$ at sets of positive measure. It is not too difficult to show that any unitary operator with a multiplicity function of this form is a bilateral shift. We will not carry out the details of this, however, for it would take us too far afield.

This result is also due to Sz.-Nagy and Foiaş and generalizes as Corollary 1 does also, a result of Schreiber [13] which states that the minimal strong unitary dilation of a strict contraction is a bilateral shift having multiplicity equal to the dimension of the space on which the contraction is defined.

We add that it is possible for the minimal strong unitary dilation of a contraction to be a bilateral shift with none of the hypotheses of Corollary 2 holding. Consider, for example, the canonical model defined by the characteristic operator function $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ with each of $\mathfrak{D}$ and $\mathfrak{D}_*$ one-dimensional and Θ equal to one half times the natural isomorphism of $L_{\mathfrak{D}_*}$ onto $L_{\mathfrak{D}}$. In this case $\overline{\Delta} = I$.

Corollary 3. *Let T be a completely nonunitary contraction on $\mathfrak{H}$ and $(\mathfrak{D}, \mathfrak{D}_*, \Theta)$ be its characteristic operator function. Then (1) Θ is an isomorphism (co-isomorphism) if and only if the sequence $\{T^n\}_{n=0}^{\infty}$ ($\{T^{**n}\}_{n=0}^{\infty}$) converges strongly to 0 and (2) Θ is an isomorphism from $L_{\mathfrak{D}_*}$ onto $L_{\mathfrak{D}}$ if and only if both of the sequences $\{T^n\}_{n=0}^{\infty}$ and $\{T^{**n}\}_{n=0}^{\infty}$ converge strongly to 0.*

Proof. From the Corollary to Theorem 1 it follows that $\mathfrak{H} = \{0\}$ which is equivalent to $\Delta = 0$ if and only if the sequence $\{T^n\}_{n=0}^{\infty}$ converges strongly to 0. From this and the fact that $\Delta^2 = I - \Theta^*\Theta = 0$ if and only if Θ is an isomorphism the result follows.

Corollary 4. *If T is a contraction on $\mathfrak{H}$ for which the sequence $\{T^n\}_{n=0}^{\infty}$ converges strongly to 0, then T is unitarily equivalent to a model of the form $U_+^* |H_{\mathfrak{D}} \ominus \Theta H_{\mathfrak{D}_*}$.*

Added in proof: Since this paper was submitted a detailed exposition of this theory has been given in the book of Sz.-Nagy and Foiaş entitled *Analyse Harmonique des Operateurs de l'Espace de Hilbert*, Budapest-Paris.

Bibliography

- [1] *L. de Branges*, Some Hilbert spaces of analytic functions II, *J. Math. Anal. Appl.* **11** (1965), 44—72.
- [2] *L. de Branges* and *J. Rovnyak*, The existence of invariant subspaces, *Bull. Amer. Math. Soc.* **70** (1964), 718—721.
- [3] *L. de Branges* and *J. Rovnyak*, Correction to "The existence of invariant subspaces", *Bull. Amer. Math. Soc.* **71** (1965), 396.
- [4] *M. S. Brodskii* and *M. S. Livšic*, Spectral analysis of nonself adjoint operators and intermediate systems, *Ups. Mat. Nauk* **13** (1958), 3—85; *Transl. Amer. Math. Soc.* (2) **13** (1960), 265—346.
- [5] *C. Foiaş*, Un remarque sur le modele universel de contraction de G. C. Rota, *Com. Acad. R. P. R.* **13** (1963), 349—352.
- [6] *P. R. Halmos*, Introduction to Hilbert space, New York, 1951.
- [7] *P. R. Halmos*, Normal dilations and extensions of operators, *Summa Brasil Math.* **2** (1950), 125—134.
- [8] *P. R. Halmos*, Shifts on Hilbert spaces, *J. Reine Ang. Math.* **208** (1961), 102—112.
- [9] *I. Halperin*, The unitary dilation of a contraction operator, *Duke Math. J.* **28** (1961), 563—572.

- [10] *G. C. Rota*, On models for linear operators, *Comm. Pure Appl. Math.* **13** (1960), 469—472.
- [11] *D. E. Sarason*, On spectral sets having connected complement, *Acta Sci. Math. (Szeged)*. **26** (1965), 289—299.
- [12] *J. J. Schaffer*, On unitary dilations of operators, *Proc. Amer. Math. Soc.* **6** (1955), 322.
- [13] *M. Schreiber*, Unitary dilations of operators, *Duke Math. J.* **24** (1956), 579—594.
- [14] *B. Sz.-Nagy*, Sur les contractions de l'espace de Hilbert, *Acta Sci. Math. (Szeged)* **15** (1953), 87—92.
- [15] *B. Sz.-Nagy*, Sur les contractions de l'espace de Hilbert II, *Acta Sci. Math. (Szeged)* **18** (1957), 1—14.
- [16] *B. Sz.-Nagy*, On Schaffer's construction of unitary dilations, *Annales Univ. Budapest, sect. math.*, **3—4** (1960—61), 343—346.
- [17] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. III, *Acta Math. Sci. (Szeged)* **19** (1958), 26—45.
- [18] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. IV, *Acta Sci. Math. (Szeged)* **21** (1960), 251—259.
- [19] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. V, Translations bilaterales, *Acta Sci. Math. (Szeged)* **23** (1962), 106—129.
- [20] *B. Sz.-Nagy* and *B. Foiaş*, Sur les contractions de l'espace de Hilbert. VI, Calcul fonctionnel, *Acta Sci. Math. (Szeged)* **23** (1962), 130—167.
- [21] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. VII, Triangulations canoniques, Fonction minimum, *Acta Sci. Math. (Szeged)* **25** (1964), 12—37.
- [22] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. VIII. Fonctions caracteristiques, Modeles fonctionnels, *Acta Sci. Math. (Szeged)* **25** (1964), 38—71.
- [23] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. IX. Factorisations de la fonction caracteristique, Sous-espace invariants, *Acta Sci. Math. (Szeged)* **25** (1964), 283—316.
- [24] *B. Sz.-Nagy* and *C. Foiaş*, Sur les contractions de l'espace de Hilbert. X. Contractions similaires a des transformations unitaires, *Acta Sci. Math. (Szeged)* **26** (1965), 79—91.
- [25] *B. Sz.-Nagy* and *C. Foiaş*, Corrections et complements aux „Contractions. IX“, *Acta Sci. Math. (Szeged)* **26** (1965), 193—196.
- [26] *J. von Neumann*, Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren, *Math. Ann.* **102** (1929), 49—131.
- [27] *H. Langer*, Ein Zerspaltungssatz für Operatoren im Hilbertraum, *Acta Math. Sci. (Szeged)* **12** (1961), 441—445.

Department of Mathematics, University of Michigan, Ann Arbor, Mich., U. S. A.

Eingegangen 14. Februar 1967

Means over countable semigroups and almost-periodicity in l^1

By *Hans Günzler* in Göttingen

The space $l^1 = l^1(N)$ of all sequences with absolutely convergent sum has among general Banach spaces many special properties; see for example Day [6], p. 31, (12), p. 33, corollary 2 and 3, Lindenstrauß [15]¹). Another one has been added recently by Amerio [2], [4] Ricci and Rizzonelli [19]: If f is a function defined on the reals R with values in l^1 , then f is norm almost periodic if and only if it is almost periodic in the weak topology. In the following we will generalize this to arbitrary semigroups S instead of R and somewhat more general locally convex range spaces instead of l^1 .

We want to proceed similarly as in the case $S = R$. First it is enough in the situation considered to show that weakly a. p. functions f have totally bounded range $f(S)$ (theorem 3 below). This in turn will follow with the approximation theorem for a. p. functions on semigroups, if the weak spectrum of f is at most countable. For $S = R$ the latter has been shown by Amerio [2] using the fact that the mean $M(g)$ of an arbitrary continuous a. p. function g is determined by a sequence M_n of operators independent of g ,

$M_n g := \frac{1}{2n} \int_{-n}^n g \, dt$. Here the analogy breaks down, even for groups no such sequences

M_n are known and probably do not exist in general²). Fortunately, by a contradiction argument it is enough to treat countable semigroups, and here the existence of such M_n can be demonstrated; for groups this would be a purely combinatorial matter, in the semigroup case one has to work with unitary representations (§ 1). Thus, for countable semigroups a theory of a. p. functions with values in a locally convex space E is possible if E is only sequentially complete, topological completeness is not necessary.

Simultaneously we treat the connection between weak and strong continuity for functions in E generalizing l^1 and which are connected with quasi-Montel spaces (§ 2).

The results of Jacobs [10] about equivalence of weak and strong almost-periodicity (see also DeLeeuw and Glicksberg [7]) are of course not applicable in our situation: l^1 is typically non-reflexive, S need not have a unit, the functions considered are in general not of the form $f(s) = sx_0$, s (bounded) operator in E . There exist even weakly a. p. f defined on the integers with values in the Hilbert space l^2 and with a. p. norm $\|f\|$, but which are not norm a. p.; see also the counterexamples after theorem 4 below.

¹) Compare however with Bessaga and Pelczyński [22].

²) In the meantime the non-discrete case has been treated by Davis [23]; see also the references cited there and Emerson and Greenleaf [24].

Our terminology is that of Day [6], Kelley [11], Köthe [13]. Some of the lemmas following are probably known, since we could not find them in the literature we have indicated their proofs for the interested reader.

§ 1. Means on countable semigroups

A semigroup S here means a set of elements $a, b, \dots, s, \dots$ for which a binary operation $S \times S \rightarrow S$ is defined which is associative. For any such S which is also a topological space, and for any uniform space U , $A(S, U)$ denotes the set of all almost periodic (a. p.) functions $f: S \rightarrow U$ in the sense of [8]: f is a. p. iff f is continuous and for every neighborhood V of the uniformity there exists a partition $P_1, \dots, P_k$ of S , $S = \bigcup_{m=1}^k P_m$, such that whenever for some $c', d', x, y \in S_u$, $c'xd'$ and $c'y d'$ are in the same P_m , then $(f(cxd), f(cyd)) \in V$ for all $c, d \in S_u$ for which cxd and $cyd \in S$. $S_u := S$ if S has a unit u ($us = su = s$ for all $s \in S$), otherwise S_u denotes the semigroup obtained from S by formally adjoining a unit. If $S = S_u$ in the discrete topology and $U = \mathbb{C} =$ complex numbers, the above definition coincides with the one introduced by Maak [17], for groups it is equivalent to the usual one.

If $U = E$ is a locally convex and topologically complete complex linear space, then one has on $A(S, E)$ a unique mean $M: A(S, E) \rightarrow E$ which is linear, invariant, normalized and continuous with respect to uniform convergence on S .

A sequence $A_1, A_2, \dots$ will be called *mean-determining* for S iff each A_m is a finite sequence consisting of elements from S such that for each E as above and each $f \in A(S, E)$ one has

$$(1) \quad M(f) = \lim_{m \rightarrow \infty} M(f, A_m),$$

where

$$(2) \quad M(f, A) := \frac{1}{|A|} \sum_{a \in A} f(a), \quad |A| := \text{number of elements in } A.$$

Multiplicities are allowed, i. e. A_m may contain an a say k times, this a is then to be counted k times in $|A|$ and the sum.

We will first prove

Theorem 1. *Any at most countable semigroup has a mean-determining sequence.*

The essential point here is that a countable number of A_m 's suffice, independent of E . A similar statement with nets $(A_i)_{i \in I}$ and arbitrary semigroups is quite trivial. For special S , for example free countable (semi)groups, one can exhibit such (A_m) explicitly, with even a stronger property and without multiplicities (see the remarks after lemma 3 below). For the proof we need some lemmas:

Lemma 1. *Let M be a set of k complex unitary $n \times n$ matrices $U_1, \dots, U_k$ which is irreducible, i. e. admits no invariant subspace $\neq (0)$, $\neq \mathbb{C}^n$; define $V := \frac{1}{k} \sum_{i=1}^k U_i$, denote with $\|A\|$ the operator-norm $\sup_{|x|=1} |Ax|$ of a linear operator $A: \mathbb{C}^n \rightarrow \mathbb{C}^n$, $|x|^2 := \sum_{j=1}^n |x_j|^2$ for $x \in \mathbb{C}^n$. If then $\|V^m\| = 1$ for all m natural, $n = 1$ and $U_1 = U_2 = \dots = U_k$.*

Proof. There are $x_m \in C^n$, $|V^m x_m| = 1 = |x_m|$, $m = 1, 2, \dots$; choose $x_{m_k} \rightarrow y_0$, $|y_0| = 1$. Since $\|V^m\| \leq 1$ for each m , $|V^{m_k} y_0| \rightarrow 1$ and therefore $|V^m y_0| = 1$, $m = 0, 1, \dots$. Now if $z_1, \dots, z_k \in C^n$, $z_1 \neq 0$, and $\left| \sum_{i=1}^k z_i \right| = \sum_{i=1}^k |z_i|$, $z_i = \mu_i z_1$ with real $\mu_i \geq 0$, $i = 1, \dots, k$. Hence $\left| \sum_{i=1}^k U_i(V^m y_0) \right| = k = \sum_{i=1}^k |U_i(V^m y_0)|$ gives $U_i(V^m y_0) = V^{m+1} y_0$ for all m, i , or $U_i^m y_0 = V^m y_0$ for $m = 0, 1, \dots$, $i = 1, \dots, k$, where $y_0 \neq 0$. Define $L := \{y \in C^n : U_i^m y = V^m y \text{ for } i = 1, \dots, k, m = 0, 1, \dots\}$, a linear subspace of C^n of dimension ≥ 1 . One easily sees that $U_i L = L$, $i = 1, \dots, k$, and thus $L = C^n$, $U_1 = \dots = U_k$. Again using the irreducibility of $(U_1, \dots, U_k)$, n has to be 1. —

Lemma 2. *If $M = \{U_1, U_2, \dots\}$ is a countable irreducible set of complex unitary $n \times n$ matrices, then there exists a natural n_0 such that $U_1, \dots, U_{n_0}$ is irreducible.*

This follows by a simple contradiction argument, using that C^n is locally compact.

With S_ω we denote the free semigroup with a countable number of generators $s_1, s_2, \dots$. S_ω consists of all words $s = s_{i_1} \dots s_{i_m}$ with $m, i_1, \dots, i_m$ natural ≥ 1 , and the obvious definition of multiplication, it is equipped with the discrete topology. S_ω has no unit, $s = s^*$ iff $m = m^*$, $i_j = i_j^*$, $j = 1, \dots, m$; m is the length of s .

Lemma 3. *S_ω has a mean-determining sequence.*

Proof. Let $A_{k,i} := \{s_1 s_2 \dots s_i s : s \text{ word of length } k^*, \text{ with generators } s_1, \dots, s_k \text{ only}\}$, and put

$$(3) \quad A_m := \bigcup_{k=1}^m \bigcup_{i=0}^{k-1} A_{k,i}.$$

The $A_{k,i}$ appearing in the A_m are mutually disjoint, so $A_m < S_\omega$, one has multiplicities 1 only, $A_m < A_{m+1}$ for all m (but $\bigcup_1^\infty A_m \neq S_\omega$). Because of the approximation theorem for $A(S, E)$ it is sufficient to show $\lim_{m \rightarrow \infty} M(\beta, A_m) = M(\beta) = 0$ for each coefficient $\beta = \beta_{ij}$ of a unitary irreducible finite-dimensional representation $U = (\beta_{ij})$ of S_ω which is $\neq 1$ (compare with [8]). Now $|A_m|$, the number of elements in A_m , is given by $\sum_{k=1}^m k^{k+1}$, so $|A_{m-1}|/|A_m| \rightarrow 0$, $\left| \bigcup_{i=0}^{m-1} A_{m,i} \right|/|A_m| \rightarrow 1$ as $m \rightarrow \infty$, it is therefore enough to show $M\left(\beta, \bigcup_{i=0}^{k-1} A_{k,i}\right) \rightarrow 0$ as $k \rightarrow \infty$.

With $U_k := U(s_k)$, $V_k := \frac{1}{k} \sum_{i=1}^k U_i$, assume first that $\|V_k^m\| = 1$ for all natural k and m . Then with lemmas 1, 2 U is a representation of degree 1 and $U_k = e^{i\varphi}$ with real φ independent of k , $\varphi \neq 0 \pmod{2\pi}$. So $M\left(\beta, \bigcup_i A_{k,i}\right) = k^{-k-1} \sum_{j=0}^{k-1} k^{k^j} e^{i\varphi(k^j+i)}$ which obviously tends to 0 as $k \rightarrow \infty$.

Otherwise there are natural k_0, m_0 with $\|V_{k_0}^{m_0}\| =: v < 1$ (even for $n = 2$, m_0 is generally > 1). If $k > k_0$, though the V_{k_0}, U_j do not commute,

$$\begin{aligned} \|V_k^{m_0}\| &= \left\| \left(\frac{k_0}{k} V_{k_0} + \frac{1}{k} \sum_{j=k_0+1}^k U_j \right)^{m_0} \right\| \quad *) \\ &\leq \left(\frac{k_0}{k} \right)^{m_0} \cdot v + \sum_{i=1}^{m_0} \left(\frac{k_0}{k} \right)^{m_0-i} \cdot 1 \cdot \binom{m_0}{i} \left(\frac{1}{k} \right)^i (k - k_0)^i = 1 - \left(\frac{k_0}{k} \right)^{m_0} (1 - v). \end{aligned}$$

*) Here we use: If A, B are bounded linear operators in a normed linear space, then

$$\|(A+B)^m - A^m\| \leq (\|A\| + \|B\|)^m - \|A\|^m$$

for $m = 1, 2, \dots$

For $k \geq m_0$ therefore

$$\|V_k^{k^k}\| \leq 1 \cdot \|V_k^{m_0 \cdot k^{k-1}}\| \leq \left(1 - \left(\frac{k_0}{k}\right)^{m_0} (1-v)\right)^{k^{k-1}} \leq \left(k^{k-1} \left(\frac{k_0}{k}\right)^{m_0} (1-v)\right)^{-1} \xrightarrow[k \rightarrow \infty]{} 0.$$

So

$$\begin{aligned} \left\| M\left(U, \bigcup_{i=0}^{k-1} A_{k,i}\right) \right\| &\leq \frac{1}{k} \sum_{i=0}^{k-1} \left\| \frac{1}{k^{k^k}} \sum_{s \in A_{k,i}} U(s) \right\| = \frac{1}{k} \sum_{i=0}^{k-1} \|U_1 \cdots U_i(V_k)^{k^k}\| \\ &\leq \|V_k^{k^k}\| \rightarrow 0 \text{ as } k \rightarrow \infty. \end{aligned}$$

But if for a sequence $B_k = (b_{i,j,k})$ of complex $n \times n$ matrices $\|B_k\| \rightarrow 0$ in the operator norm, then obviously $b_{i,j,k} \rightarrow 0$ for each $i, j = 1, \dots, n$. —

Remarks. 1. For the free semigroup S_g with a finite number g of generators, the sets

$$(3g) \quad A_m := \bigcup_{i=0}^{m-1} \{s_1^i \cdot s : \text{length of } s = m\}$$

form a mean-determining sequence with $\bigcup_{m=1}^{\infty} A_m = S_g$.

2. For the free groups G_g respectively G_{ω} the same sets (3g) resp. (3) form a mean-determining sequence. This implies $M(f) = M(f|S_g)$ resp. $= M(f|S_{\omega})$ for any $f \in A(G_g, E)$ resp. $A(G_{\omega}, E)$, E as above.

In this case however one can find “better” sequences: There exist finite subsets $B_1, B_2, \dots$ of G_g such that for any fixed $s \in G_g$,

$$(4) \quad |(sB_m) \triangle B_m|/|B_m| \rightarrow 0 \text{ as } m \rightarrow \infty;$$

similarly for G_{ω} . $A \triangle B := (A - A \cap B) \cup (B - B \cap A)$, the B_m also satisfy $B_m < B_{m+1}$, $\bigcup_{m=1}^{\infty} B_m = G$.

Here we use the following, a result of Struble [21] generalizing

Lemma 4. *Let S be a semigroup, μ a left-invariant measure on a σ -ring $\mathcal{S}$ of subsets from S ; it is assumed that $sA \in \mathcal{S}$ implies $A \in \mathcal{S}$. Let $B_1, B_2, \dots$ be a sequence with $B_m \in \mathcal{S}$, $0 < \mu(B_m) < \infty$ such that*

$$(5) \quad \mu(sB_m \triangle B_m)/\mu(B_m) \rightarrow 0 \text{ as } m \rightarrow \infty$$

for any $s \in S$. If furthermore E is locally convex and topologically complete such that any $f \in A(S, E)$ is over each B_m μ -Bochner-integrable, then for any $f \in A(S, E)$

$$\lim_{m \rightarrow \infty} \frac{1}{\mu(B_m)} \int_{B_m} f d\mu \text{ exists} = M(f).$$

This does not help much, since even in the case of locally compact groups it is not clear²⁾ when such B_m exist, so we omit the proof. Natural examples are open convex cones $S < R^n$.

3. Free semigroups S_g, S_{ω} probably are not “ ω -exhaustibel” in the sense of lemma 4. However, for free abelian semigroups S which are at most countable there exist again finite sets B_m with (4) and $B_m < B_{m+1}$, $\bigcup_m B_m = S$.

4. Lemma 3 holds for arbitrary countable semigroups $S = \{s_1, s_2, \dots\}$ with formally the same A_m as in (3). Only the A_m are no longer subsets of S , the elements have to be counted according to their importance, their multiplicity in A_m given by the number of representations of the form admitted in A_m .

In somewhat different words, one has

Lemma 5. *If S is a countable discrete semigroup, then there exists a homomorphism $h: S_\omega \rightarrow S$ which is onto.*

This is trivial, $h(s) := a_{m_1} \cdots a_{m_k}$ for $s = s_{m_1} \cdots s_{m_k} \in S_\omega$, $S = \{a_1, a_2, \dots\}$ does everything.

Proof of theorem 1. With the h of lemma 5, for $f \in A(S, E)$ one has $f \circ h \in A(S_\omega, E)$. $M^*(f) := M(f \circ h)$ is linear, invariant, normalized and continuous on $A(S, E)$, so it coincides with $M(f)$ (compare with Maak [16], p. 40). Lemma 3 gives

$$M(f) = \lim_{m \rightarrow \infty} M(f \circ h, A_m) = \lim_{m \rightarrow \infty} M(f, A_m^*),$$

where A_m^* is the finite sequence $(h(s))_{s \in A_m}$ (compare with remark 4 above). —

Corollary. *Let S be a semigroup which is at most countable⁴⁾, E a locally convex complex linear space which is sequentially complete. Then the space $A(S, E)$ of almost-periodic functions $f: S \rightarrow E$ has a mean $M: A(S, E) \rightarrow E$ which is linear, invariant, normalized and continuous, all theorems for a. p. functions with values in topologically complete spaces (see [8]) are valid also for these $A(S, E)$.*

This can be helpful for the study of weakly a. p. functions in non-reflexive spaces, where the usual theory as outlined in [8] not applies. For example, $L^1(X, \mathcal{S}, \mu, R)$ is weakly sequentially complete, $(X, \mathcal{S}, \mu)$ being any measure space (Day [6], p. 108); though these spaces are in general not weakly quasicomplete, for countable or finite S all theorems on vector-valued a. p. functions still hold.

In a similar way one can show for arbitrary semigroups S and arbitrary cardinal numbers β : If S is β -exhaustible by a directed set $(B_i)_{i \in I}$ with $\text{card } I \leq \beta$ as in (5) above, if furthermore E is β -topologically complete, i. e. each totally bounded Cauchy net $(x_i)_{i \in I}$ with $\text{card } I \leq \beta$ converges in E , then for $f \in A(S, E)$, $M(f)$ exists $\in E$.

§ 2. Continuity in l^1 -spaces

We will say that a system (E, π_i) is a *locally convex space with projections* if E is a locally convex real or complex linear space and if the π_i are linear transformations: $E \rightarrow E$, $i \in$ some index set I . In the applications below the π_i are continuous projections but in general we do not assume this. A locally convex space with projections (E, π_i) is said to satisfy $(w\omega s)$ if for each subset $A \subset E$ the following holds: If A is a subset of a weakly sequentially compact set of E and if all $\pi_i(A)$ are subsets of sequentially compact sets, $i \in I$, then A is totally bounded. E satisfies $(w\omega s)$ if $(E, 0)$ does, $0x := 0$ for $x \in E$.

If (E, π_i) satisfies $(w\omega s)$, then for each sequence $(x_n)_n$ from E ,

(6) $x_n \rightarrow 0$ weakly and $\pi_i(x_n) \rightarrow 0$ for each $i \in I$ imply⁵⁾ $x_n \rightarrow 0$ as $n \rightarrow \infty$

(use lemma 7 below). The converse is true if either all π_i are w -s-sequentially closed, i. e. $y_n \rightarrow y_0$ weakly and $\pi_i(y_n) \rightarrow z$ imply $z = \pi_i(y_0)$, (y_n) sequence from E , or if for each $x \in E$ at most a countable number of $\pi_i(x)$ differ from 0. The connection with quasi-Montel spaces is discussed in § 3, especially⁸⁾.

⁴⁾ More generally, the corollary to theorem 1 holds for separable topological⁹⁾ semigroups (see [9], p. 272 below).

⁵⁾ Or equivalently: (x_n) has a (strong) Cauchy-subnet.

A topological linear space satisfies von Neumann's *countability axiom* (A_0) means there exists a countable number $V_1, V_2, \dots$ of neighborhoods of 0 in E with $\bigcap_{n=1}^{\infty} V_n = \{0\}$ (see von Neumann [18]).

Theorem 2. *Let X be a topological space, (E, π_i) a locally convex space with projections which satisfies $(w\omega s)$, f a function $X \rightarrow E$. Assume further that either X satisfies the first axiom of countability (A_1) or X is locally countably compact and E satisfies (A_0) . Then f is continuous if f is weakly continuous and all $\pi_i \circ f$ are continuous.*

The assumptions made on E , X and f are essential: If an (E, π_i) does not satisfy (6), then there exists a compact X with (A_1) and a weakly continuous $f: X \rightarrow E$ with continuous $\pi_i \circ f$ which is not continuous. Furthermore there exist X (even locally convex linear spaces) and (linear) $f: X \rightarrow l^1(N)$ which are weakly continuous but not continuous, whereas l^1 has $(w\omega s)$. Usually the π_i are continuous, then the continuity of the $\pi_i \circ f$ is also necessary.

Examples. 1. No infinite-dimensional reflexive Banach space has (6), so for $l^p(I)$, $L^p(X, \mathcal{S}, \mu)$, $1 < p < \infty$, theorem 2 becomes false. Here I is an arbitrary infinite set, μ a σ -finite measure on the σ -ring $\mathcal{S}$ from X , the functions are real or complex-valued. For X open $\subset \mathbb{R}^n$ and $\mu =$ Lebesgue-measure also L^1 has not (6), the same holds for $l^\infty(I)$.

2. Every Montel-space or nuclear space has $(w\omega s)$, then theorem 2 is even without lemma 7 rather obvious, (A_0) is not necessary. The spaces $\mathcal{D}'(\Omega)$ respectively $\mathcal{H}(\Omega)$ of complex-valued distributions resp. holomorphic functions on the open set $\Omega \subset \mathbb{R}^n$ resp. C^∞ are Montelspaces (see Schwartz [20], p. 74 and Köthe [13], p. 378); $\mathcal{D}'(\Omega)$ also has (A_0) , $\mathcal{H}(\Omega)$ is even metrisable. See also⁶⁾, ⁸⁾.

3. For Banach spaces the classical examples are the real or complex $l^1(I)$, I arbitrary set (see Day [6], p. 33 and Köthe [13], p. 283/4). More generally, one has the

Corollary. *If X has (A_1) or is locally compact, $f: X \rightarrow l^1(I, N_i)$ is weakly continuous and the coordinate functions $f_i: X \rightarrow N_i$ are all norm-continuous (for example, all N_i have (6)), then f is norm-continuous.*

With $l^1(I, N_i)$ we denote the substitution space $P_{\nu(I)}N_i$, where I is an arbitrary index set and the N_i are real respectively complex normed spaces (Day [6], p. 31); if the N_i are complete, $E := l^1(I, N_i)$ is too. This E becomes a locally convex space with projections (E, π_i) using the natural projections $\pi_i: E \rightarrow N_i$ onto the coordinate spaces N_i . The corollary then follows with

Lemma 6. *$l^1(I, N_i)$ with its natural system of projections has $(w\omega s)$ or equivalently (6). So $l^1(I, N_i)$ alone satisfies (6) if and only if each N_i does.*

The proof of lemma 6 is similar to the one when all $N_i = C$ (see Köthe [13], p. 284 for countable I), in addition one only has to use the Hahn-Banach theorem for the N_i ; we omit the details.

In the same way one can also show: *If X has (A_1) , $f: X \rightarrow l^1(I, N_i)$ is weakly uniformly continuous and all $\pi_i \circ f$ are uniformly norm-continuous, then f is uniformly norm-continuous; for $N_i = C$, $i \in I$, the f_i are automatically uniformly continuous.*

⁶⁾ The perfect sequence-spaces of Köthe also satisfy $(w\omega s)$, they are therefore weakly sequentially complete and even Quasi-Montel-spaces (Köthe [13], p. 415–419).

Proof of theorem 2. We will show a little more than stated, namely that (6) alone implies the continuity of f . The (A_1) -case is trivial. Let now E satisfy (A_0) and X be locally countably compact, i. e. each $x \in X$ has a countably compact neighborhood U_x . Given $x_0 \in X$, let U_0 be such a countably compact neighborhood of x_0 ; then $f(U_0)$ is weakly countably compact. Because of (A_0) , the theorem of Šmulian-Dieudonné-Schwartz⁷⁾ (Köthe [13], p. 314) can be applied, so $f(U_0)$ is weakly sequentially compact. With lemma 7 below it is enough to show that $f(U_0)$ is totally bounded. Assume therefore that (x_n) is a sequence of points from U_0 with $f(x_n) - f(x_m) \notin V_0$ for some neighborhood V_0 of 0 in E , $n \neq m$. One can assume that $f(x_n) \rightarrow f(y_0)$ weakly with some $y_0 \in U_0$. Then $\pi_i \circ f(x_n) \rightarrow \pi_i \circ f(y_0)$ for each $i \in I$: Else for some i , some neighborhood W_0 of 0 in E and a subsequence n_k one has $\pi_i \circ f(x_{n_k}) - \pi_i \circ f(y_0) \notin W_0$, $k = 1, 2, \dots$. U_0 being countably compact there is a subnet $(y_j)_j$ of $(x_{n_k})_k$ with $y_j \rightarrow z_0 \in U_0$, so $\pi_i \circ f(y_j) \rightarrow \pi_i \circ f(z_0)$ and $f(y_j) \rightarrow f(z_0)$ weakly. Since $f(x_{n_k}) \rightarrow f(y_0)$ weakly, $f(y_0) = f(z_0)$ and therefore $\pi_i \circ f(y_j) \rightarrow \pi_i \circ f(y_0)$, a contradiction with the assumed existence of the W_0 . The sequence $(f(x_n) - f(y_0))_n$ satisfies therefore (6), which shows that a V_0 as above cannot exist. —

Lemma 7. *Let A be a totally bounded subset of a locally convex real or complex linear space E . Then for A the induced uniformity coincides with the induced weak uniformity.*

Proof (no completeness assumptions are needed). First it is enough to show that for totally bounded A weak and strong topologies coincide; for the equivalence of these topologies in $A - A$, which is again totally bounded, implies that the weak and strong uniformities are the same in A . So let A be totally bounded and $(x_i)_{i \in I}$ be a net in A which converges weakly to $x_0 \in A$; we have to show that $x_i \rightarrow x_0$ strongly. If not, there exists a strong neighborhood U_0 of 0 and a subnet (y_j) of (x_i) with $y_j \rightarrow x_0$ weakly, no $y_j \in x_0 + U_0$. Since $y_j \in A$ which is totally bounded, there exists a (strong) Cauchy-subnet (z_k) of (y_j) with $z_k \notin x_0 + U_0$, $z_k \rightarrow x_0$ weakly.

Now generally, if $u_k \rightarrow u_0$ weakly and $(u_k)_k$ is a strong Cauchy-net in a locally-convex space, then $u_k \rightarrow u_0$ strongly: Let V be a convex strong neighborhood of 0 which is closed and $< U$ given, and choose k_0 such that for $k, k' \geq k_0$ one has $u_k - u_{k'} \in V$. Since in locally convex spaces such V are also weakly closed (Köthe [13], p. 245/6), $u_k - u_0 \in V$, i. e. $u_k \rightarrow u_0$ strongly. —

§ 3. Almost-periodicity in l^1 -spaces

In the following we will have to use

Theorem 3. *Let S be a semigroup and a topological space, E a locally convex real or complex linear space, $f: S \rightarrow E$ weakly almost periodic. Then f is almost periodic if and only if $f(c_0 S d_0)$ is totally bounded for some $c_0, d_0 \in S_u$.*

Weakly a. p. means of course $\varphi \circ f$ is continuous and a. p. $\in A(S, C)$ for each continuous linear functional φ on E , i. e. a. p. in the weak topology of E . For $S =$ additive group of the reals R , $E =$ Banach space, theorem 3 has been shown independently by Kopec [14] and Amerio [2]. In the general case, the strong continuity and strong almost-periodicity of weakly a. p. f follows at once from lemma 7 resp. corollary I.d of [9].

Our main result is now contained in

⁷⁾ This theorem is also an easy consequence of lemma 8 and 11 below.

Theorem 4. Let S be a semigroup and a topological space, (E, π_i) a locally convex space with projections which satisfies $(w\omega s)$, $f: S \rightarrow E$. Assume further that E has (A_0) and is sequentially complete. Then f is a. p. if f is weakly a. p. and all $\pi_i \circ f$ are a. p.

The definitions of (E, π_i) , $(w\omega s)$ and (A_0) can be found at the beginning of § 2. With the $l^1(I, N_i)$ defined in § 2 and lemma 6, applied to $l^1(I, B_i)$, one gets

Corollary I. If the N_i are real or complex normed spaces, S is as above, I an arbitrary set and $f: S \rightarrow l^1(I, N_i)$, then f is norm-a. p. if and only if it is weakly a. p. and all its projections $f_i: S \rightarrow N_i$ are norm-a. p., $i \in I$. So if all N_i satisfy (6), f is a. p. if and only if it is weakly a. p.

Corollary II. Let S be a convex semigroup of the additive group of the reals in the usual topology, I any set, B_i real or complex Banach spaces, $f: S \rightarrow l^1(I, B_i)$ be a. p., $t_0 \in S$, $F(t) := \int_{t_0}^t f \, dt$ (Bochner-integral) be bounded on S . If for each $i \in I$ and a. p. $g: S \rightarrow B_i$ the function $t \rightarrow \int_{t_0}^t g \, dt$ is a. p. whenever it is bounded, F is a. p.

Corollary I is a generalization of a result of Ricci and Rizzonelli [19]: They considered the special case $S =$ additive group of the reals, I countable, all $N_i = C$, i. e. $E = l^1(N)$.

Corollary II slightly generalizes a result of Amerio [4], case $p = 1$; the weak almost-periodicity of F follows from [9] and simple properties of the Bochner-Lebesgue integral. Similar statements hold also for convex semigroups $< R^n$, they can be extended to arbitrary semigroups and l^p -substitution spaces, $p < \infty$ (see [9], corollary 4*. a, 4*. b; f need only be Stepanoff-a. p.).

Condition $(w\omega s)$ is of course essential, also when all $\pi_i = 0$: There exist even functions $f_0: R \rightarrow \bigcap_{1 \leq p \leq \infty} l^p(N)$ and $g_0: R \rightarrow L^\infty([0, 1])$ such that f_0, g_0 are weakly a. p. as functions with values in l^p resp. L^p for each p , $1 < p < \infty$, but for no p , $1 \leq p \leq \infty$, are they norm-a. p. in l^p resp. L^p (Lebesgue measure is used). These l^p and L^p are also weakly sequentially complete. Using constructions of Amerio [1], p. 325 and [3], p. 377—382, one can also find functions h defined on R with values in $l^p(N)$, $1 < p < \infty$, resp. $L^2([0, 1])$ resp. $l^\infty(N)$ resp. $C([0, 1])$ which are weakly a. p. but not norm-a. p., in the cases l^p , $1 < p < \infty$, and L^2 they have even a. p. $\|h\|$.

The weak almost-periodicity of f cannot be weakened much further: Again using Amerio [1], p. 325, one can construct $f: R \rightarrow l^1(N)$ which are not norm-a. p., though $\|f\|$ and all components f_n of f are a. p.

The sequential completeness of E is necessary insofar as there exist locally convex E with (A_0) whose (sequential) completion has not (A_0) ; only if E is metrisable the sequential completeness is superfluous, $\bar{E}$ has then (6) too.

The non-complete case is the only really interesting one: For weakly quasi-complete E , for example Montel-spaces where (A_0) is superfluous too, theorem 4 is an almost immediate consequence of theorem 3, lemma 8 and 11. This applies to $\mathcal{D}(\Omega)$, $\mathcal{D}'(\Omega)$, $\mathcal{H}(\Omega)$ (see § 2).

In this direction, the quasi-Montel spaces of Kera [12] are also subsumed^{6), 8)} if one has (A_0) and weak sequential completeness. More generally, if (E, π_i) is a real or complex space with projections, let us say that it satisfies $(cwws)$ if for each $A < E$ which is contained in a weakly sequentially compact and absolutely convex set, and for which all $\pi_i(A)$ are subsets of sequentially compact and absolutely convex sets, one can conclude

that A is totally bounded; $l^1(I, N_i)$ with its π_i have $(cws)^8$). Then theorem 4 holds true too with (ws) weakened to (cws) , if the sequential completeness of E is strengthened to weak sequential completeness. We omit the proof, since for nice π_i conditions (cws) and (ws) are directly seen to be equivalent⁸).

Proof of theorem 4 (for complex E it could be somewhat simplified). By theorem 3 it is enough to show that $f(S)$ is totally bounded. Let us assume that this is not the case; then there exists a neighborhood U_0 of 0 in E and a sequence $(s_n)_{n=1,2,\dots}, s_n \in S$, such that $f(s_n) - f(s_m) \notin U_0$ for $n \neq m$. Denote by S_0 the countable semigroup generated by the s_n in S , in the discrete topology.

$f_i := \pi_i \circ f|_{S_0}$ is a. p. $\in A(S_0, E)$ and has its range in $E_i :=$ closed linear hull of $f_i(S_0)$, where E_i in the induced topology is separable, sequentially complete and has (A_0) . Because of lemma 8, E_i in its weak topology also has (A_0) . If now $M_k(\dots)$ denotes the finite mean $M(\dots, A_k)$ on S_0 of theorem 1, for any $\varphi \in E'_i$ and $g \in A(S_0, R)$,

$$\varphi(M_k(g \cdot f_i)) = M_k(g \cdot \varphi \circ f_i) \rightarrow M(g \cdot \varphi \circ f_i) \text{ as } k \rightarrow \infty,$$

so $(M_k(g \cdot f_i))_k$ is a weak Cauchy sequence in E_i . Since $g \cdot f_i$ is a. p., $(g \cdot f_i)(S_0)$ is totally bounded; therefore (see Köthe [13], p. 243) the convex hull H_i of $(g \cdot f_i)(S_0)$ is also totally bounded, one has $a_{i,k} := M_k(g \cdot f_i) \in H_i$. Because of lemma 7, $(a_{i,k})_k$ is a strong Cauchy sequence in $H_i \subset E_i$. E_i being sequentially complete, $a_{i,k} \rightarrow$ some $a_i \in E_i$.

One can therefore apply the second statement in lemma 9, the weak spectrum of $f_i: S_0 \rightarrow E_i$ is countable. Lemma 10 shows that $f_i(S_0)$ is sequentially precompact and therefore relatively sequentially compact. Lemma 11 yields then the sequential compactness of $\overline{f_i(S_0)}$.

With an analogous reasoning one shows that $(a_k)_k$, $a_k := M_k(g \cdot f_0)$, is a weak Cauchy sequence in $E_0 :=$ closed linear hull of $f_0(S_0)$, $f_0 := f|_{S_0}$, $g \in A(S_0, R)$. Since $\pi_i(a_k) = a_{i,k} \rightarrow a_i$, condition (ws) via (6) implies that $(a_k)_k$ is a strong Cauchy sequence, $a_k \rightarrow a \in E_0$ as $k \rightarrow \infty$. So again with theorem 1, lemma 8 and lemma 9, the weak spectrum of f_0 is countable. With lemma 10, applied to E_0 in its weak topology, $f_0(S_0)$ is weakly sequentially precompact since the two s_w appearing here coincide.

Consider now the s_n generating S_0 : There is a subsequence $(t_m)_m$ of $(s_n)_n$ such that $(f_0(t_m))_m$ is a weak Cauchy sequence, furthermore $f_i(t_m) \in K_i$ sequentially compact for all m . So there exists a sequence $(c_m)_m$, $c_m = f_0(t_m) - f_0(t_{k_m})$, such that $c_m \rightarrow 0$ weakly, $c_m \notin U_0$, but for each $i \in I$, $\pi_i(c_m) \in K_i - K_i$, which is again sequentially compact in E .

⁸) If (E, π_i) is a locally convex real or complex space with projections, if E satisfies (A_0) , if further either each π_i is w - s -sequentially closed (see the beginning of § 2) or if for each fixed $x \in E$ at most countably many $\pi_i(x)$ are $\neq 0$, then the following implications hold for (E, π_i) :

$$(6)^8 \Leftrightarrow (ws) \Leftrightarrow (ws) \Rightarrow (c_0ws) \Leftrightarrow (c_0ws) \Leftrightarrow (cws) \Leftrightarrow (cws).$$

If in addition E is weakly sequentially complete, all seven conditions are equivalent. Because of lemma 7 and 11, the "totally bounded" in the conclusion is always equivalent with relative compactness and relative sequential compactness. If the π_i are w - w -sequentially continuous (for example $\pi_i = 0$; no assumptions on E), then (6) or (ws) are equivalent with: If A is weakly sequentially precompact and all $\pi_i(A)$ are totally bounded, then A is totally bounded.

Here (ws) is obtained from (cws) by deleting the "sequentially" wherever it appears, similarly for (cws) , (c_0ws) ; the c_0 -conditions are obtained from the c -conditions by deleting twice the "absolutely". (c_0ws) is Kera's condition for $\pi_i = 0$.

The proof may be obtained by combining lemma 7, 8, 11 with: If in a locally convex space $x_n \rightarrow 0$, then the absolutely convex hull of the $x_1, x_2, \dots$ is sequentially precompact.

Since the c_m together with 0 obviously constitute a weakly sequentially compact set, by condition ($w\omega s$) this set is totally bounded. With lemma 7 one deduces $c_m \rightarrow 0$ strongly, the contradiction sought.

Lemma 8. *If E is a real or complex locally convex separable space with (A_0) , then the weak topology for E also satisfies (A_0) .*

The proof is standard: If U_n are absolutely convex neighborhoods of 0 with $\bigcap_{n=1}^{\infty} U_n = \{0\}$, p_n the corresponding seminorms, if further $\{x_m : m = 1, 2, \dots\}$ is dense in E , and $\varphi_{m,n} \in E'$ are such that $\varphi_{m,n}(x_m) = p_n(x_m)$ and $|\varphi_{m,n}(x)| \leq p_n(x)$ on E , then $V_{m,n} := \left\{ |\varphi_{m,n}| < \frac{1}{2} \right\}$ have only 0 in their intersection. —

Lemma 9. *Let S be a topological semigroup⁹⁾, E a locally convex real or complex linear space with (A_0) , $f : S \rightarrow E$ a. p. $\in A(S, E)$. Then the following two statements hold:*

If E is complex, if all Fourier components $P_{\omega}f \in A(S, E)$, $\omega \in \Omega(S)$, then the spectrum $\sigma(f)$ of f is at most countable.

If E_1 is a real or complex locally convex linear space, $E = E_1$ in its weak topology, and if for each $\omega \in \Omega(S)$, $U = (\beta_{kl}) \in \omega$, $\beta = \beta_{kl} \in U$ there exist¹⁰⁾ $b_1, b_2 \in E_1$ such that for arbitrary $\varphi \in E'_1$, $M((\operatorname{Re} \beta) \cdot \varphi \circ f) = \varphi(b_1)$, similarly for $\operatorname{Im} \beta$, then the weak spectrum of f , considered as a function $S \rightarrow E_1$, is at most countable.

Here $\Omega(S)$ denotes the set of all equivalence classes ω of continuous unitary irreducible finite-dimensional representations U of S ; $P_{\omega}f := s_{\omega} \cdot \operatorname{trace} (M(f \cdot U^*) \cdot U)$, independent of $U \in \omega$, s_{ω} = number of rows in $U \in \omega$. $P_{\omega}f$ is for $f \in A(S, E)$ well defined and $\in A(S, E)$ if E is complex and topologically complete [8]. For arbitrary complex locally convex E we define $P_{\omega}f := P_{\omega}\underline{f}$, where $\underline{f}$ is the a. p. function $f : S \rightarrow \bar{E}$ = completion of E , which is again locally convex.

The spectrum of f is defined by $\sigma(f) := \{\omega \in \Omega(S) : P_{\omega}f \neq 0\}$. If $f : S \rightarrow E$ is weakly a. p., E real or complex, then the weak spectrum of f is $\sigma_w(f) := \bigcup_{\varphi \in E'} \sigma(\varphi \circ f)$. If $f \in A(S, E)$ with complex E , then $\sigma(f) = \sigma_w(f)$.

There exist a. p. functions defined on the integers with values in a locally convex complex separable linear space with (A_0) which have an uncountable spectrum, so $P_{\omega}f(S) \subset E$ and of course (A_0) are essential.

Proof (Generally $\bar{E}$ will not have (A_0) , so the theorems of Bochner and von Neumann [5] are not applicable even if S is a group). The first statement can be shown with the summation theorem of [8], § 4; since we do not use this in the proof of theorem 4, we omit the details. For complex E_1 the second statement follows with the first. In the real case this argument is not possible, $P_{\omega}f$ is not defined, but then Amerio's proof in [2], p. 292 applies with some modifications: Since E_1 (real or complex) in its weak topology satisfies by assumption (A_0) , there are countably many $\varphi_1, \varphi_2, \dots \in E'_1$ such that $\varphi_n(x) = 0$ for all n implies $x = 0$. If $\sigma_0 := \bigcup_{n=1}^{\infty} \sigma(\varphi_n \circ f)$, then there exists a summation method which sums the Fourier series of all $g \in A(S, C)$ with $\sigma(g) \subset \sigma_0$. (see Bochner

⁹⁾ It is enough if S is a semigroup and a topological space such that for each fixed $a \in S$, the maps $s \rightarrow a \cdot s$ and $s \rightarrow s \cdot a$ are continuous on S .

¹⁰⁾ For complex E_1 , the assumptions about the existence of b_1, b_2 are equivalent with $P_{\omega}f \in A(S, E)$.

and von Neumann [5]; the semigroup case is treated in [8]). Therefore

$$\sum_{\omega} \gamma_{\omega, m} P_{\omega}(\varphi_n \circ f) \Rightarrow \varphi_n \circ f$$

as $m \rightarrow \infty$, n arbitrary natural. Applying $P_{\omega'}$ with $\omega' \notin \sigma_0$ one gets $P_{\omega'}(\varphi_n \circ f) = 0$ for all n . If E_1 is real, $\varphi_n \circ f$ is real-valued and so $M((\operatorname{Re} \beta) \cdot \varphi \circ f) = 0 = M((\operatorname{Im} \beta) \cdot \varphi \circ f)$ for any $\beta \in U \in \omega'$, $\varphi = \varphi_1, \varphi_2, \dots$. Since these expressions are $= \varphi(b_{1,2})$, $\varphi_n(b_{1,2}) = 0$ for all n , $b_{1,2} = 0$ and therefore $P_{\omega'}(\varphi \circ f) = 0$ für all $\varphi \in E'_1$, $\omega' \notin \sigma_0$. If E_1 is complex, a similar reasoning can be applied to $\varphi = \operatorname{Re} \varphi_n$, $\operatorname{Im} \varphi_n$, one gets $\sigma_w(f) \subset \sigma_0 \cup \overline{\sigma_0}$. —

Corollary. *Let E be a complex locally convex sequentially complete space with (A_0) , S as in lemma 9 and either countable or a convex open subsemigroup of the additive group R^n in the usual topology. Then an a. p. $f \in A(S, E)$ has a spectrum which is at most countable.*

This follows from theorem 1 respectively the end of remark 2 in § 1. For $S = R^1$, $E =$ complex Banach space E_1 in its weak topology, the above corollary can be found in Amerio [2], p. 292 (use lemma 8, $\sigma = \sigma_w$ and $E' = E'_1$).

Lemma 10. *If S is a topological semigroup⁹⁾, E a real or complex locally convex linear space, $f: S \rightarrow E$ a. p. with at most countable weak spectrum $\sigma_w(f)$, then $f(S)$ is sequentially precompact¹¹⁾.*

Proof (The “weak part” is due to Amerio [2]). Let $(s_n)_n$ be a sequence from S , $\omega_1, \omega_2, \dots$ be the weak spectrum of f . Choose a diagonal subsequence $(t_m)_m$ from $(s_n)_n$ such that for any $U \in \omega_k$ and any natural k , $U(t_m)$ converges as $m \rightarrow \infty$. Then for any $\omega \in \Omega(S)$ and $\varphi \in E'$, $(P_{\omega}(\varphi \circ f)(t_m))_m$ converges, $P_{\omega}(\dots)$ being a finite linear combination of components of $U \in \omega$. The summation theorem for $A(S, C)$ from [8], § 4, shows that $(\varphi \circ f(t_m))_m$ converges, $f(S)$ is weakly sequentially precompact. But since f is a. p., $f(S)$ is totally bounded in E . Lemma 7 gives then the desired result. —

Similarly, for example $(f(st_m))_m$ is even a Cauchy sequence with respect to uniform convergence in $s \in S$. One can further show that the absolutely-convex hull of $f(S)$ is sequentially precompact, under some additional assumptions this is equivalent with a countable spectrum.

Lemma 11. *Let U be a uniform space whose uniformity satisfies (A_0) , M a subset of U which is relatively countably compact. Then the uniformity induced in the closure $\overline{M}$ of M satisfies the first axiom of countability and $\overline{M}$ is compact.*

From this one deduces at once: If U is a uniform space with (A_0) for its uniformity, M a subset of U , then the three statements “ M is compact”, “ M is countably compact”, “ M is sequentially compact” are equivalent.

No completeness assumptions for U are necessary, but the condition on M in lemma 11 cannot be weakened to “ M is totally bounded”, also lemma 11 becomes false if only the induced uniformity for M satisfies (A_0) . A uniformity has (A_0) means the existence of a countable number of members U_n of the uniformity with $\bigcap_{n=1}^{\infty} U_n = \text{diagonal}$, or equivalently the existence of a coarser metrisable uniformity for U . The set M is

¹¹⁾ If f is only weakly a. p., $f(S)$ is weakly sequentially precompact by lemma 7 of [9]; if f is a. p., then by this lemma the induced uniformity in the totally bounded $f(S)$ is even metrisable.

relatively countably compact iff each sequence from M has a subnet which converges to some point of U .

Proof of lemma 11. Let U_n be $\in \mathcal{U}$ = uniformity for U with $\bigcap_{n=1}^{\infty} U_n = \text{diagonal}$. Choose $V_n \in \mathcal{U}$ such that $V_n^5 < U_n$, $V_{n+1} < V_n = V_n^{-1}$ for $n = 1, 2, \dots$. We will show first: If $z_n \in \bar{M}$, $(z_n, z_m) \in V_m$ for all natural $n \geq m$, $m = 1, 2, \dots$, then there exists a $z_0 \in \bar{M}$ such that $z_n \rightarrow z_0$. To do this, choose $y_n \in M$ with $(z_n, y_n) \in V_n$. Then there exists a subnet $(y_{n(i)})_i$ of $(y_n)_n$ and a $z_0 \in \bar{M}$ with $y_{n(i)} \rightarrow z_0$, so $(z_{n(i)}, z_0) \in V_n^2$ for $i \geq i_n$. With $(z_n, z_m) \in V_m$ one deduces $(z_n, z_0) \in V_n^3$ for each $n = 1, 2, \dots$. If (z_n) would not converge to z_0 there would exist a $U_0 \in \mathcal{U}$ and a subsequence (z_{n_k}) with $(z_{n_k}, z_0) \notin U_0^2$ for $k = 1, 2, \dots$. Choose $x_k \in M$ with $(z_{n_k}, x_k) \in U_0 \cap V_k$; for a suitable subnet $(x_{k(j)})$ of (x_k) , $x_{k(j)} \rightarrow x_0 \in \bar{M}$. Using the interpolation $z_0, z_{n_{k(j)}}, x_{k(j)}, x_0$ one gets $(z_0, x_0) \in V_n^5 < U_n$ for $n = 1, 2, \dots$ and so $x_{k(j)} \rightarrow x_0 = z_0$. But then for suitable j , $(z_{n_{k(j)}}, z_0) \in U_0^2$, the desired contradiction.

Next we show that $\mathcal{U} \cap (\bar{M} \times \bar{M})$ has $\{V_m : m = 1, 2, \dots\}$ as a base: If not, there exists a $V_0 \in \mathcal{U}$ and $x'_m, y'_m \in \bar{M}$ with $(x'_m, y'_m) \in V_m, \notin V_0^4$. Choose now $V'_m = V_m^{-1} \in \mathcal{U}$ and $x_m, y_m \in M$ with $V_m'^2 < V_m$, (x'_m, x_m) and $(y'_m, y_m) \in V_0 \cap V_m$. There exists a subsequence (x_{m_k}) and a $u_0 \in \bar{M}$ with $(x_{m_l}, u_0) \in V'_k$ for $l \geq k$, $k = 1, 2, \dots$, so

$$(x_{m_k}, x_{m_l}) \in V_k'^2 < V_k$$

if $l \geq k$. Our result above shows then that $x_{m_k} \rightarrow x_0$. Similarly, (y_{m_k}) has a subsequence $(y_{m_{k_n}})_{n=1,2,\dots}$ which converges to some $y_0 \in \bar{M}$. With suitable interpolations one deduces $(y_0, x_0) \in \text{each } V_m^5$ and so $x_0 = y_0$. But then $(x'_{m_{k_n}}, y'_{m_{k_n}}) \in V_0^4$ for n sufficiently large, which is again a contradiction.

With these two results one easily shows that $\bar{M}$ is complete. Since M is totally bounded, $\bar{M}$ is too and so $\bar{M}$ is compact and also sequentially compact. —

References

- [1] L. Amerio, Quasi-periodicità degli integrali ad energie limitata dell'equazione delle onde, con termine noto quasi-periodico. II, Rend. accad. naz. Lincei, Classe fis.-mat. **28** (1960), 322—327.
- [2] L. Amerio, Funzioni debolmente quasi-periodiche, Rend. Sem. Mat. di Padova **30** (1960), 288—301.
- [3] L. Amerio, Sull'integrazione delle funzioni quasi-periodiche astratte, Annali di Mat. pura et appl. **53** (1961), 371—382.
- [4] L. Amerio, Sull'integrazione delle funzioni $l^p(X_n)$ -quasi-periodiche, con $1 \leq p < +\infty$. Ricerche di Mat. **12** (1963), 3—12.
- [5] S. Bochner and J. von Neumann, Almost-periodic functions in groups. II, Trans. Amer. Math. Soc. **37** (1935), 21—50.
- [6] M. M. Day, Normed linear spaces, Ergebnisse d. Math., Berlin 1958.
- [7] K. DeLeeuw and I. Glicksberg, Applications of almost-periodic compactifications, Acta Math. **105** (1961), 63—97.
- [8] H. Günzler, Fourierreihen vektorwertiger fastperiodischer Funktionen auf Halbgruppen, Math. Inst. U. Göttingen, Sem. W. Maak, Bericht Nr. 5, 1961.
- [9] H. Günzler, Integration of almost-periodic functions, Math. Zeitschr. **102** (1967), 253—287.
- [10] K. Jacobs, Fastperiodizitätseigenschaften allgemeiner Halbgruppen in Banachräumen, Math. Zeitschr. **67** (1957), 83—92.
- [11] J. L. Kelley, General Topology, New York 1959.

- [12] K. Kera, On Quasi-Montel Spaces, Proc. Jap. Ac. **40** (1964), 633—637.
- [13] G. Köthe, Topologische lineare Räume. I, Berlin 1960.
- [14] J. Kopel, On linear differential equations in Banach spaces, Zeszyty Nauk. Uniw. Mickiewicza. Mat.-Chem. **1** (1957), 3—7 (Polish).
- [15] J. Lindenstrauss, On extreme points in $\mathbb{P}$, Israel J. of Math. **4** (1966), 59—61.
- [16] W. Maak, Fastperiodische Funktionen, Berlin 1950.
- [17] W. Maak, Fastperiodische Funktionen auf Halbgruppen, Acta Math. **87** (1952), 33—58.
- [18] J. von Neumann, On complete topological linear spaces, Trans. Amer. Math. Soc. **37** (1935), 1—20.
- [19] M. L. Ricci e P. Rizzonelli, Sulle funzioni $\mathbb{P}$ -quasi-periodiche, Istituto Lombardo Rend. Sci. A **95** (1961), 941—946.
- [20] L. Schwartz, Théorie des Distributions. I, Paris 1957.
- [21] R. A. Struble, Almost-periodic functions on locally compact groups, Proc. Nat. Ac. Sci. USA **39** (1953), 122—126.
- [22] C. Bessaga and A. Pelczyński, On extreme points in separable conjugate spaces, Israel J. Math. **4** (1966), 262—264.
- [23] H. W. Davis, On the mean value of Haar-measurable functions, Duke Math. J. **34** (1967), 201—214.
- [24] W. R. Emerson and F. P. Greenleaf, Covering Properties and Følner Conditions for Locally Compact Groups, Math. Zeitschr. **102** (1967), 370—384.

Mathematisches Institut der Universität, 34 Göttingen, Bunsenstr. 3—5

Eingegangen 6. März 1967

Arithmetische Funktionen über Halbordnungen. II

Von *Harald Scheid* in Mainz

Einleitung

Die vorliegende Arbeit ist die Fortsetzung einer in dieser Zeitschrift erschienenen Arbeit [22] und schließt sich in den Bezeichnungen und der Terminologie an diese an. Sie hat das Ziel, mit den in [22] entwickelten Methoden die von E. Cohen [4], [5], [6], [9], [21]¹⁾ gelieferten Aussagen über die Klassen der gleichmäßigen und der unitären Funktionen (mod r) in einen umfassenden Zusammenhang zu stellen. Durch Spezialisierung der gewonnenen Ergebnisse werden dann einige Aussagen über unitäre Funktionen (mod r) auf zahlentheoretische Funktionen von mehreren Veränderlichen übertragen. Im letzten Paragraphen wird eine Möglichkeit aufgezeigt, eine Multiplikativität für Funktionen über einer Quotientenhalbordnung zu erklären.

§ 1. Definitionen und Vorbetrachtungen

Es sei $\mathfrak{E}$ eine Hierarchie, welche genau aus den Elementen einer Menge $\mathfrak{M} = \{a, b, c, \dots, x, y, \dots\}$ besteht. Die Ordnungsrelation von $\mathfrak{E}$ werde mit „ $<$ “, diejenige der Quotientenhierarchie $Q(\mathfrak{E})$ mit „ $<$ “ bezeichnet. Ferner sei $\mathfrak{G}$ eine lokal endliche Halbordnung, aber nicht notwendig eine Hierarchie, welche ebenfalls genau aus den Elementen der Menge $\mathfrak{M}$ besteht. Die Ordnungsrelation von $\mathfrak{G}$ werde mit „ $\ll$ “, diejenige der Quotientenhalbordnung $Q(\mathfrak{G})$ mit „ $\ll$ “ bezeichnet. Es soll stets vorausgesetzt werden, daß $\mathfrak{E}$ inf-vollständig ist, d. h. jede nichtleere Teilmenge (insbesondere also jede unendliche Teilmenge) soll in $\mathfrak{E}$ ein Infimum besitzen. Daher ist $\mathfrak{E}$ nach unten beschränkt; das Schrankenelement von $\mathfrak{E}$ heiße o .

Definition 1. Die Hierarchie $\mathfrak{E}$ heißt *Erweiterungshierarchie* der lokal endlichen Halbordnung $\mathfrak{G}$, wenn für $a, b \in \mathfrak{M}$ aus $a \ll b$ stets $a < b$ folgt.

Beispiel 1. $\mathfrak{M}$ sei die Menge der natürlichen Zahlen. Die Relation $a < b$ bedeute „ a ist ein Teiler von b “ und $a \ll b$ bedeute „ a ist ein unitärer²⁾ Teiler von b “.

Beispiel 2. $\mathfrak{M}$ sei die Menge aller endlichen Mengen nichtnegativer ganzer Zahlen, welche die Null enthalten; die Elemente von $\mathfrak{M}$ mögen $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \dots$ heißen. Die Relation $\mathfrak{A} < \mathfrak{B}$ bedeute „ $\mathfrak{A}$ ist Teilmenge von $\mathfrak{B}$ “ und $\mathfrak{A} \ll \mathfrak{B}$ bedeute „es gibt ein $\mathfrak{A}' \in \mathfrak{M}$ derart, daß für die Schnirelmann-Summe $\mathfrak{A} + \mathfrak{A}' = \mathfrak{B}$ gilt.“

¹⁾ Die Zahlen [1] bis [20] beziehen sich auf das Literaturverzeichnis in [22].

²⁾ Ein Teiler a von b heißt *unitärer Teiler* von b , wenn $a \mid b$ und $\left(a, \frac{b}{a}\right) = 1$ ist.

Im folgenden sei $\mathfrak{E}$ stets eine Erweiterungshierarchie der lokal endlichen Halbordnung $\mathfrak{G}$. Quotienten aus $Q(\mathfrak{E})$ werden mit a / b und solche aus $Q(\mathfrak{G})$ mit $a // b$ bezeichnet. Es sei auch $\mathfrak{G}$ nach unten beschränkt; das Schranken-Element von $\mathfrak{G}$ ist dann ebenfalls gleich o . Die Abbildung

$$C = c // d \in Q(\mathfrak{G}) \rightarrow \bar{C} = c / d \in Q(\mathfrak{E})$$

ist eine eindeutige Abbildung von $Q(\mathfrak{G})$ in $Q(\mathfrak{E})$. Der Bildbereich von $Q(\mathfrak{G})$ werde mit $\overline{Q(\mathfrak{G})}$ bezeichnet. $Q(\mathfrak{G})$ und $\overline{Q(\mathfrak{G})}$ sind ordnungsisomorph. Die Urbilder von Elementen $A, B, \dots \in \overline{Q(\mathfrak{G})}$ werden mit $\tilde{A}, \tilde{B}, \dots$ bezeichnet. Für $a, b \in \mathfrak{M}$ definiere man

$$\inf^*(a, b) = \sup_{\mathfrak{E}} \{x : x < a, x \ll b\}$$

und

$$\sup^*(a, b) = \inf_{\mathfrak{E}} \{x : a < x, b \ll x\}.$$

Die Existenz von $\inf^*(a, b)$ folgt daraus, daß die Menge $\{x : x < a, x \ll b\}$ nicht leer und in dem Quotienten $\inf_{\mathfrak{E}}(a, b) / o$ enthalten ist. Die Existenz von $\sup^*(a, b)$ ist aufgrund der geforderten inf-Vollständigkeit von $\mathfrak{E}$ gewährleistet, falls die Menge $\{x : a < x, b \ll x\}$ nicht leer ist. Man füge nun zu $\mathfrak{E}$ das uneigentliche Element $u = \inf_{\mathfrak{E}} \emptyset$ hinzu und treffe die Festsetzungen $a < u$ und $a / u = \emptyset$ für alle $a \in \mathfrak{E}^3$. Ist dann die Menge $\{x : a < x, b \ll x\}$ leer, so setze man $\sup^*(a, b) = u$. Für $a, b \in \mathfrak{M}$ gilt

$$(1) \quad o < \inf^*(a, b) < \inf_{\mathfrak{E}}(a, b) < \sup_{\mathfrak{E}}(a, b) < \sup^*(a, b) < u.$$

Für $N = m / n \in Q(\mathfrak{E})$ und $A = a // b \in Q(\mathfrak{G})$ sei nun

$$\inf^*(N, A) = \sup_{Q(\mathfrak{E})} \{X \in \overline{Q(\mathfrak{G})} : X < N, \tilde{X} \ll A\}.$$

Dann gilt

$$(2) \quad \inf^*(N, A) < \inf^*(m, a) / \sup^*(n, b) < \inf_{\mathfrak{E}}(m, a) / \sup_{\mathfrak{E}}(n, b) = \inf_{Q(\mathfrak{E})}(N, \bar{A}).$$

Im allgemeinen liegt $\inf^*(N, A)$ nicht in $\overline{Q(\mathfrak{G})}$. Ist dies jedoch der Fall, so sei zur Abkürzung

$$\widehat{\inf}(N, A) = \widetilde{\inf^*(N, A)}.$$

Aus der Definition von $\inf^*(N, A)$ und aus (2) folgt, daß

$$(3) \quad \text{genau dann } X \ll \widehat{\inf}(N, A), \text{ wenn } X \ll A \text{ und } \bar{X} < N.$$

Es sei ϑ_0 eine Abbildung von $\mathfrak{M}$ in sich, die sowohl bezüglich der Ordnung in $\mathfrak{E}$ als auch bezüglich der Ordnung in $\mathfrak{G}$ ein Ordnungsisomorphismus ist. Dann ist ϑ_0 ein Verbandsisomorphismus von $\mathfrak{E}$, und wir wollen fordern, daß $\vartheta_0(\mathfrak{E})$ ein Teilverband von $\mathfrak{E}$ ist. Durch $\vartheta(a / b) = \vartheta_0(a) / \vartheta_0(b)$ bzw. durch $\vartheta(a // b) = \vartheta_0(a) // \vartheta_0(b)$ wird in $Q(\mathfrak{E})$ bzw. in $Q(\mathfrak{G})$ ein Ordnungsisomorphismus auf eine Teilmenge der jeweiligen Quotientenhalbordnung erklärt. Obwohl es sich um zwei Abbildungen in verschiedenen Bereichen handelt, bezeichnen wir sie mit dem gleichen Symbol ϑ , da durch das Argument A von $\vartheta(A)$ bestimmt ist, welche der beiden Abbildungen gemeint ist. Insbesondere ist für $X \in Q(\mathfrak{G})$

$$(4) \quad \overline{\vartheta(X)} = \vartheta(\bar{X}).$$

³⁾ Ist z. B. $\mathfrak{M}$ die Menge der natürlichen Zahlen, so wählt man $u = 0$.

Es sei nun

$$\lambda^*(N, A) = \sup_{Q(\mathfrak{G})} \{X \in Q(\mathfrak{G}) : \vartheta(X) < \inf^*(N, \vartheta(A))\}$$

und

$$\kappa^*(N, A) = \vartheta \lambda^*(N, A).$$

Offenbar gilt die Beziehung

$$(5) \quad \kappa^*(N, A) < \inf^*(N, \vartheta(A)).$$

Für $A, B \in Q(\mathfrak{G})$ und $M, N \in Q(\mathfrak{G})$ folgt aus $A \ll B$ und $M < N$ die Relation

$$(6) \quad \kappa^*(M, A) < \kappa^*(N, B).$$

Ist $\lambda^*(N, A)$ ein Element von $\overline{Q(\mathfrak{G})}$, dann auch $\kappa^*(N, A)$. In diesem Fall sei zur Abkürzung

$$\tilde{\lambda}(N, A) = \widetilde{\lambda^*(N, A)} \text{ und } \tilde{\kappa}(N, A) = \widetilde{\kappa^*(N, A)}.$$

Es folgt dann aus (3), daß für $X \in Q(\mathfrak{G})$

$$(7) \quad \text{genau dann } X \ll \tilde{\lambda}(N, A), \text{ wenn } X \ll A \text{ und } \vartheta(\bar{X}) < N.$$

Für $\inf^*(N, \vartheta(A)) \in \overline{Q(\mathfrak{G})}$ und $\lambda^*(N, A) \in \overline{Q(\mathfrak{G})}$ folgt aus (5)

$$(8) \quad \bar{\kappa}(N, A) \ll \inf(N, \vartheta(A)).$$

Lemma 1. Wenn $\lambda^*(N, A) \in \overline{Q(\mathfrak{G})}$ ist für ein festes $A \in Q(\mathfrak{G})$ und für alle $N \in Q(\mathfrak{G})$, dann ist für dieses feste A die Abbildung $(N, A) \rightarrow \lambda^*(N, A)$ eine Abbildung von $Q(\mathfrak{G})$ auf $\bar{A} / \emptyset$ und daher $(N, A) \rightarrow \tilde{\lambda}(N, A)$ eine Abbildung von $Q(\mathfrak{G})$ auf $A // \emptyset$.

Beweis. Aus (3) und (8) folgt $\tilde{\lambda}(N, A) \ll A$. Aus $C \ll A$ folgt $\vartheta(\bar{C}) < \vartheta(\bar{A})$; daher ist $\inf^*(\vartheta(\bar{C}), \vartheta(A)) = \vartheta(\bar{C})$, also $\lambda^*(\vartheta(\bar{C}), A) = \bar{C}$ und $\tilde{\lambda}(\vartheta(\bar{C}), A) = C$.

Bemerkung. Die Voraussetzung $\lambda^*(N, A) \in \overline{Q(\mathfrak{G})}$ ist nicht überflüssig, denn $(N, A) \rightarrow \lambda^*(N, A)$ ist bei festem A i. a. keine Abbildung von $Q(\mathfrak{G})$ auf $\bar{A} / \emptyset$, weil aus $C \in \bar{A} / \emptyset$ nicht $C \in \overline{Q(\mathfrak{G})}$ folgt.

Lemma 2. Es seien $A, B \in Q(\mathfrak{G})$ und $A \ll B$. Dann ist

$$\kappa^*(\kappa^*(N, B), A) = \kappa^*(N, A).$$

Ist außerdem $\lambda^*(N, A) \in \overline{Q(\mathfrak{G})}$, dann gilt

$$\tilde{\kappa}(\kappa^*(N, B), A) = \tilde{\kappa}(N, A).$$

Beweis. Aus (2), (5) und (6) folgt

$$\kappa^*(N, A) < \kappa^*(N, B) < N$$

und hieraus nach (6)

$$\kappa^*(\kappa^*(N, A), A) < \kappa^*(\kappa^*(N, B), A) < \kappa^*(N, A).$$

Nun ist nach (2) und (5) $\inf^*(\kappa^*(N, A), A) = \kappa^*(N, A)$, also auch

$$\kappa^*(\kappa^*(N, A), A) = \kappa^*(N, A),$$

woraus der erste Teil der Behauptung folgt. Mit $\kappa^*(N, A)$ ist also auch $\kappa^*(\kappa^*(N, B), A)$ ein Element von $\overline{Q(\mathfrak{G})}$, woraus die zweite Behauptung folgt.

Definition 2. Eine Funktion $f(N, A)$ über $Q(\mathfrak{G}) \times Q(\mathfrak{G})$ heißt ϑ - $\ast$ -gleichmäßig in $N \pmod{A}$, wenn für alle $N \in Q(\mathfrak{G})$ die Gleichung

$$f(N, A) = f(\kappa^*(N, A), A)$$

erfüllt ist. Die Klasse der ϑ - $\ast$ -gleichmäßigen Funktionen heiße E_{ϑ}^* .

Im Fall $\mathfrak{G} = \mathfrak{E}$ ist eine ϑ - $\ast$ -gleichmäßige Funktion $\pmod{A}$ auch ϑ -gleichmäßig $\pmod{A}$, da dann $\kappa(N, A) = \kappa^*(N, A)$ ist (vgl. [22], § 4). In diesem Fall müssen die Sätze der folgenden Paragraphen mit den entsprechenden in [22] hergeleiteten Sätzen übereinstimmen. Da man im Fall $\mathfrak{G} = \mathfrak{E}$ auf die inf-Vollständigkeit von $\mathfrak{E}$ verzichten kann ($\sup^*(a, b) = \sup(a, b)$ existiert stets), sind die Sätze und Definitionen dieser Arbeit Verallgemeinerungen derjenigen aus [22].

§ 2. Verallgemeinerte Ramanujan-Summen

Das Produkt von Quotienten aus $Q(\mathfrak{G})$ und die Konvolution von Funktionen über $Q(\mathfrak{G})$ werden mit *einem* Stern und die entsprechenden Bildungen für Quotienten aus $Q(\mathfrak{G})$ und Funktionen über $Q(\mathfrak{G})$ mit *zwei* Sternen geschrieben. Die Möbiusfunktion über $Q(\mathfrak{G})$ heiße μ , die Möbiusfunktion über $Q(\mathfrak{G})$ heiße μ^* . Die Funktion, deren Wert stets gleich 1 ist, heiße in beiden Fällen ν ; das Einselement der Konvolution heiße in beiden Fällen ε .

Die Quotientenhalbordnung $Q(\mathfrak{G})$ sei nun normiert, und die Norm werde mit N bezeichnet. Durch

$$(9) \quad (c_L^*(N, X) \ast \ast \nu(Y))(A) = \begin{cases} N(\vartheta(A)), & \text{wenn } \overline{\vartheta(A)} < N \\ 0 & \text{sonst} \end{cases}$$

und

$$(10) \quad (\nu(X) \ast \ast c_R^*(N, Y))(A) = \begin{cases} N(\vartheta(A)), & \text{wenn } \overline{\vartheta(A)} < N \\ 0 & \text{sonst} \end{cases}$$

oder gleichwertig damit durch

$$(11) \quad c_L^*(N, A) = \sum_{\substack{X \ast \ast Y = A \\ \overline{\vartheta(X)} < N}} N(\vartheta(X)) \mu^*(Y)$$

und

$$(12) \quad c_R^*(N, A) = \sum_{\substack{X \ast \ast Y = A \\ \overline{\vartheta(Y)} < N}} \mu^*(X) N(\vartheta(Y))$$

werden verallgemeinerte Ramanujan-Summen c_L^* und c_R^* definiert. Diese Ramanujan-Summen genügen wieder Orthogonalitätsrelationen, welche zu denjenigen für c_L und c_R ([22], § 3) völlig analog sind und in diese für $\mathfrak{G} = \mathfrak{E}$ übergehen. Es gilt nämlich

Satz 1. Es sei $A \in Q(\mathfrak{G})$, B ein oberer und C ein unterer Faktor von A . Dann gelten die Beziehungen

$$(13) \quad (c_L^*(\overline{\vartheta(X)}, X) \ast \ast c_R^*(\overline{\vartheta(C)}, Y))(A) = \begin{cases} N(\vartheta(A)), & \text{wenn } B \ast \ast C = A \\ 0 & \text{sonst} \end{cases}$$

und

$$(14) \quad (c_L^*(\overline{\vartheta(B)}, X) \ast \ast c_R^*(\overline{\vartheta(Y)}, C))(A) = \begin{cases} N(\vartheta(A)), & \text{wenn } B \ast \ast C = A \\ 0 & \text{sonst.} \end{cases}$$

Beweis. Es seien $A = a // b$, $B = a // c$ und $C = d // b$ mit $c, d \in A$. In der linken Seite von (13) ersetze man c_L^* durch (11) und wende dann (10) an:

$$\begin{aligned} \sum_{x \in a // b} c_R^*(\overline{\vartheta(C)}, x // b) &= \sum_{\substack{y \in a // c \\ \vartheta(a // y) < \vartheta(a // x)}} N(\vartheta(a // y)) \mu^*(y // c) \\ &= \sum_{y \in a // c} N(\vartheta(a // y)) \mu^*(y // c) \sum_{x \in y // b} c_R^*(\overline{\vartheta(d // b)}, x // b) \\ &= N(\vartheta(A)) \sum_{\substack{y \in a // c \\ \vartheta(y // b) < \vartheta(d // b)}} \mu^*(y // c). \end{aligned}$$

Ist $c \not\leq d$, so ist diese Summe leer. Ist aber $c \leq d$, so hat sie den Wert $N(\vartheta(A))\varepsilon(d // c)$, woraus (13) folgt. Formel (14) wird ebenso bewiesen.

§ 3. Eine Beziehung zwischen den Klassen E_ϑ und E_ϑ^*

Jeder Funktion $f(N, A)$ über $Q(\mathfrak{E}) \times Q(\mathfrak{G})$ kann man wegen der Ordnungsisomorphie von $Q(\mathfrak{G})$ und $\overline{Q(\mathfrak{G})}$ eineindeutig eine Funktion $g(N, \overline{A})$ über $Q(\mathfrak{E}) \times \overline{Q(\mathfrak{G})}$ durch die Festsetzung

$$(15) \quad g(N, \overline{A}) = f(N, A)$$

zuordnen. Es gilt

Satz 2. Die Funktion $f(N, A)$ sei ϑ - $\times$ -gleichmäßig (mod A). Dann ist $g(N, \overline{A})$ ϑ -gleichmäßig (mod $\overline{A}$).

Beweis. Wegen $\kappa(N, \overline{A}) < N$ ist nach (6)

$$(16) \quad \kappa^*(\kappa(N, \overline{A}), A) < \kappa^*(N, A).$$

Nach (2) und (5) ist $\kappa^*(N, A) < \inf(\overline{\vartheta(A)}, N)$, also $\kappa^*(N, A) < \kappa(N, \overline{A})$, woraus $\kappa^*(\kappa^*(N, A), A) < \kappa^*(\kappa(N, \overline{A}), A)$ folgt. Wegen $\kappa^*(\kappa^*(N, A), \overline{A}) = \kappa^*(N, A)$ gilt also auch

$$(17) \quad \kappa^*(N, A) < \kappa^*(\kappa(N, \overline{A}), A).$$

Aus (16) und (17) folgt

$$\kappa^*(\kappa(N, \overline{A}), A) = \kappa^*(N, A).$$

Daher gilt

$$g(\kappa(N, \overline{A}), \overline{A}) = f(\kappa(N, \overline{A}), A) = f(\kappa^*(\kappa(N, \overline{A}), A), A) = f(\kappa^*(N, A), A) = f(N, A) = g(N, \overline{A}),$$

womit die Behauptung bewiesen ist.

Nach diesem Satz muß also die einer ϑ - $\times$ -gleichmäßigen Funktion $f(N, A)$ gemäß obiger Festsetzung (15) zugeordnete Funktion $g(N, \overline{A})$ eine arithmetische Darstellung und, falls $Q(\mathfrak{E})$ normiert ist, eine Ramanujan-Darstellung besitzen ([22], § 5). Die Norm von $Q(\mathfrak{G})$ kann man durch die Festsetzung $N(\overline{X}) = N(X)$ auf $\overline{Q(\mathfrak{G})}$ übertragen; diese Norm sei auf ganz $Q(\mathfrak{E})$ fortgesetzt und werde ebenfalls mit N bezeichnet. Da nun c_L^* und c_R^* wegen (7) ϑ - $\times$ -gleichmäßig (mod A) sind, kann man die diesen Funktionen gemäß (15) zugeordneten Funktionen $\overline{c}_L$ und $\overline{c}_R$ durch die Ramanujan-Summen c_L und c_R ausdrücken. Unter gewissen einschränkenden Voraussetzungen erhält man dabei besonders einfache Formeln.

Satz 3. Ist $\inf^*(Y, A) \in \overline{Q(\mathfrak{G})}$ für alle unteren Faktoren Y von A , dann gilt

$$(18) \quad \bar{c}_L(N, A) = \sum_{\substack{X \star Y = \bar{A} \\ \inf^*(Y, A) = I}} c_L(N, X).$$

Ist $\inf^*(X, A) \in \overline{Q(\mathfrak{G})}$ für alle oberen Faktoren X von A , dann gilt

$$(18') \quad \bar{c}_R(N, A) = \sum_{\substack{X \star Y = \bar{A} \\ \inf^*(X, A) = I}} c_R(N, Y).$$

Beweis. Mit

$$\beta(X, Y) = \begin{cases} 1, & \text{wenn } X, Y \in \overline{Q(\mathfrak{G})} \\ 0 & \text{sonst,} \end{cases}$$

$$\chi_L(X, \bar{A}) = \begin{cases} 1, & \text{wenn } X \text{ oberer Faktor von } \bar{A} \text{ ist} \\ 0 & \text{sonst,} \end{cases}$$

$$\bar{\mu}(Y) = \begin{cases} \mu^*(\bar{Y}), & \text{wenn } Y \in \overline{Q(\mathfrak{G})} \\ 0 & \text{sonst} \end{cases}$$

und

$$g_L(X, A) = N(\vartheta(X)) \bar{\mu} \left[\frac{\bar{A}}{X} \right] \beta \left(X, \left[\frac{\bar{A}}{X} \right] \right) \chi_L(X, \bar{A})$$

ist

$$\bar{c}_L(N, A) = \sum_{\substack{X \star Y = \bar{A} \\ \vartheta(X) < N}} N(\vartheta(X)) \bar{\mu}(Y) \beta(X, Y) = (g_L(X, A) \star \nu(Y)) (\lambda(N, \bar{A})).$$

Ist $\lambda(N, \bar{A})$ ein oberer Faktor von $\bar{A}$, dann ist nach Satz 3 aus [22]

$$\bar{c}_L(N, A) = (c_L(N, X) \star \alpha_R(Y, \bar{A})) (\bar{A}),$$

wobei nach Satz 4 aus [22]

$$\alpha_R \left(\left[\frac{\bar{A}}{B} \right], \bar{A} \right) = \frac{1}{N(\vartheta(\bar{A}))} \sum_{X \star Y = \left[\frac{\bar{A}}{B} \right]} g_L \left(\left\{ \frac{A}{Y} \right\}, \bar{A} \right) N(\vartheta(Y))$$

ist. Setzt man hier den obigen Ausdruck für g_L ein, dann erhält man

$$\alpha_R \left(\left[\frac{\bar{A}}{B} \right], \bar{A} \right) = \sum_{X \star Y = \left[\frac{\bar{A}}{B} \right]} \bar{\mu}(Y) \beta \left(\left\{ \frac{\bar{A}}{B} \right\}, Y \right).$$

Mit $A = a // b$, $\left[\frac{\bar{A}}{B} \right] = c / b$ ($c \in \bar{A}$) lautet die rechte Seite der letzten Gleichung

$$\sum_{x \in c / b} \bar{\mu}(x / b) \beta(a / x, x / b) = \sum_{\substack{x \in a // b \\ x \in c / b}} \mu^*(x // b).$$

Aufgrund der Voraussetzung des Satzes ist $\inf^*(c, a) \in a // b$. Daher ist die Summationsbedingung der letzten Summe gleichbedeutend mit „ $x \in \inf^*(c, a) // b$ “, woraus

$$\alpha_R \left(\left[\frac{\bar{A}}{B} \right], \bar{A} \right) = \varepsilon(\inf^*(c, a) // b)$$

folgt. Da nun genau dann $\inf^*(c, a) = b$ gilt, wenn $\inf^*(c / b, a // b) = b / b$ ist, hat man damit Formel (18) bewiesen. Formel (18') folgt ebenso.

§ 4. Die Darstellungen der Klasse E_ϑ^*

Unter der Voraussetzung, daß $\lambda^*(N, A) \in \overline{Q(\mathfrak{G})}$ für alle $N \in Q(\mathfrak{E})$ gilt, erhält man in Analogie zu den Sätzen 2 und 3 aus [22] die folgenden Sätze:

Satz 4. Sei $f(N, A) \in E_\vartheta^*$ und $\xi * \eta = \varepsilon$. Dann hat $f(N, A)$ eine Darstellung der Form

$$(19) \quad f(N, A) = (g_L^*(X, A) * \xi(Y))(\tilde{\lambda}(N, A))$$

mit

$$(20) \quad g_L^*(D, A) = (f(\overline{\vartheta(X)}), A) * \eta(Y)(D)$$

für alle $D \triangleleft A$, und eine Darstellung der Form

$$(21) \quad f(N, A) = (\xi(X) * g_R^*(Y, A))(\tilde{\lambda}(N, A))$$

mit

$$(22) \quad g_R^*(D, A) = (\eta(X) * f(\overline{\vartheta(Y)}), A)(D)$$

für alle $D \triangleleft A$.

Satz 5. Sei $f(N, A) \in E_\vartheta^*$. Ist $\lambda^*(N, A)$ ein oberer Faktor von $\bar{A}$, dann ist

$$(23) \quad f(N, A) = (c_L^*(N, X) * \alpha_R^*(Y, A))(A),$$

wobei für alle unteren Faktoren C von A

$$(24) \quad \alpha_R^*(C, A) = \frac{1}{N(\vartheta(A))} (f(\overline{\vartheta(X)}), A) * c_R^*(\overline{\vartheta(C)}, Y)(A)$$

gilt. Ist $\lambda^*(N, A)$ ein unterer Faktor von $\bar{A}$, dann ist

$$(25) \quad f(N, A) = (\alpha_L^*(X, A) * c_R^*(N, Y))(A),$$

wobei für alle oberen Faktoren B von A

$$(26) \quad \alpha_L^*(B, A) = \frac{1}{N(\vartheta(A))} (c_L^*(\overline{\vartheta(B)}), X) * f(\overline{\vartheta(Y)}, A)(A)$$

gilt.

§ 5. Teilklassen der Klasse E_ϑ^*

Entsprechend den Teilklassen der Klasse E_ϑ ([22], § 6) kann man Teilklassen von E_ϑ^* definieren.

Definition 3. Eine Funktion $f(N, A)$ über $Q(\mathfrak{E}) \times Q(\mathfrak{G})$ heißt *vollständig ϑ - $*$ -gleichmäßig* in $N \pmod{A}$, wenn eine Funktion $F(X)$ über $Q(\mathfrak{E})$ derart existiert, daß für alle $N \in Q(\mathfrak{E})$ die Gleichung

$$f(N, A) = F(\kappa^*(N, A))$$

erfüllt ist. Die Klasse dieser Funktionen heiße VE_ϑ^* .

Wegen $\kappa^*(\kappa^*(N, A), A) = \kappa^*(N, A)$ ist die Klasse VE_ϑ^* in E_ϑ^* enthalten. In den Darstellungen der Funktionen aus VE_ϑ^* kann man in den Ausdrücken für $g_L^*, g_R^*, \alpha_L^*, \alpha_R^*$ jeweils $f(\overline{\vartheta(X)}, A)$ durch $F(\overline{\vartheta(X)})$ ersetzen, denn für $X \triangleleft A$ ist $\kappa^*(\vartheta(X), A) = \overline{\vartheta(X)}$.

Nun sei γ_0 eine Abbildung von $\mathfrak{M}$ in sich, die sowohl in $\mathfrak{G}$ als auch in $\mathfrak{E}$ ein Ordnungsendomorphismus mit den in [22], § 6 geforderten Eigenschaften (a), (b), (c) ist. Man definiere $\gamma(a/b) = \gamma_0(a)/\gamma_0(b)$ für $a/b \in Q(\mathfrak{E})$ und $\gamma(a//b) = \gamma_0(a) // \gamma_0(b)$ für $a//b \in Q(\mathfrak{G})$.

Definition 4. Eine Funktion $f(N, A)$ über $Q(\mathfrak{E}) \times Q(\mathfrak{G})$ heißt ϑ - $\ast$ -*primitiv* in $N \pmod{A}$, wenn für alle $N \in Q(\mathfrak{E})$ die Gleichung

$$f(N, A) = f(\kappa^*(N, \gamma(A)), A)$$

erfüllt ist. Die Klasse dieser Funktionen heie P_ϑ^* .

Definition 5. Eine Funktion $f(N, A)$ über $Q(\mathfrak{E}) \times Q(\mathfrak{G})$ heißt *vollständig ϑ - $\ast$ -primitiv* in $N \pmod{A}$, wenn eine Funktion $G(X)$ über $Q(\mathfrak{E})$ derart existiert, da für alle $N \in Q(\mathfrak{E})$ die Gleichung

$$f(N, A) = G(\kappa^*(N, \gamma(A)))$$

erfüllt ist. Die Klasse dieser Funktionen heie VP_ϑ^* .

Unter der Voraussetzung, da $\lambda^*(N, \gamma(A)) \in \overline{Q(\mathfrak{G})}$ für alle $N \in Q(\mathfrak{E})$ gilt, erhält man in Analogie zu den Sätzen 6 und 7 aus [22] die folgenden Sätze, wobei der Kürze wegen nur jeweils eine der beiden möglichen Darstellungen angegeben wird:

Satz 6. Sei $f(N, A) \in P_\vartheta^*$ und $\xi \ast \ast \eta = \varepsilon$. Dann hat $f(N, A)$ eine Darstellung der Form

$$(27) \quad f(N, A) = (g_L^*(X, A) \ast \ast \xi(Y))(\tilde{\lambda}(N, \gamma(A)))$$

mit

$$(28) \quad g_L^*(D, A) = (f(\overline{\vartheta(X)}, A) \ast \ast \eta(Y))(D)$$

für alle $D \triangleleft \gamma(A)$.

Satz 7. Sei $f(N, A) \in P_\vartheta^*$. Ist $\lambda^*(N, \gamma(A))$ ein oberer Faktor von $\gamma(A)$, dann ist

$$(29) \quad f(N, A) = (c_L^*(N, X) \ast \ast \alpha_R^*(Y, A))(\gamma(A)),$$

wobei für alle unteren Faktoren C von $\gamma(A)$

$$(30) \quad \alpha_R^*(C, A) = \frac{1}{N(\vartheta(\gamma(A)))} (f(\overline{\vartheta(X)}, A) \ast \ast c_R^*(\overline{\vartheta(C)}, Y))(\gamma(A))$$

ist.

§ 6. Unitäre Funktionen über freien Halbgruppen

Ordnet man eine freie abelsche Halbgruppe $\mathfrak{F}$ mit abzählbar unendlich vielen Erzeugenden durch die Teilbarkeit, dann entsteht eine inf-vollständige Hierarchie $\mathfrak{F}^0$. Ordnet man dagegen $\mathfrak{F}$ durch die unitäre Teilbarkeit, so entsteht lediglich eine lokal endliche Halbordnung $\mathfrak{F}^{00}$. Sowohl $\mathfrak{F}^0$ als auch $\mathfrak{F}^{00}$ besitzen ein kleinstes Element e , nämlich das Einselement der freien Halbgruppe. Es ist $\inf^*(a, b) = (a, b)_*$ der größte Teiler von a , der unitärer Teiler von b ist. Im allgemeinen sind $(a, b)_*$ und $(b, a)_*$ verschieden. Es gilt

$$\inf^*(n/e, a//e) = \inf^*(n, a)/e = (n, a)_*/e.$$

Die Enden von $\mathfrak{F}^0$ bzw. $\mathfrak{F}^{00}$ sind die Quotienten der Form a/e bzw. $a//e$. Ein Ende von $\mathfrak{F}^0$ ist also stets ein Element aus $\overline{Q(\mathfrak{F}^{00})}$. Ist N ein Ende von $\mathfrak{F}^0$ und A ein Ende von

$\mathfrak{F}^{00}$, dann sind auch $\inf^*(N, \vartheta(A))$, $\lambda^*(N, A)$ und $\kappa^*(N, A)$ Enden von $\mathfrak{F}^0$ und daher Elemente von $\overline{Q(\mathfrak{F}^{00})}$. Man definiere $\kappa_0^*(n, a)$ durch

$$\kappa_0^*(n, a) / e = \kappa^*(n / e, a // e).$$

Dann ist $\kappa_0^*(n, a)$ der größte Teiler von $(n, \vartheta_0(a))_*$, der ein ϑ_0 -Bild ist.

Als Beispiel werde nun die freie Halbgruppe der n -Tupel $\mathfrak{a} = (a_1, a_2, \dots, a_n)$ natürlicher Zahlen betrachtet. Das Rechnen mit diesen n -Tupeln ist komponentenweise erklärt; ebenso kann man die Teilbarkeit bzw. die unitäre Teilbarkeit von n -Tupeln durch die entsprechenden Relationen für ihre n Komponenten definieren. Man erhält so eine inf-vollständige Hierarchie $\mathfrak{N}$ und eine lokal endliche Halbordnung $\mathfrak{N}^*$, wobei $\mathfrak{N}$ eine Erweiterungshierarchie von $\mathfrak{N}^*$ ist. Die Abbildung ϑ_0 und γ_0 und die Norm N_0 seien in $\mathfrak{N}$ und in $\mathfrak{N}^*$ durch

$$\vartheta_0(\mathfrak{a}) = \mathfrak{a}^{\mathfrak{f}}, \quad (\mathfrak{f} \text{ festes Element aus } \mathfrak{N}),$$

$$\gamma_0(\mathfrak{a}) = \text{quadratfreier Kern von } \mathfrak{a}$$

und

$$N_0(\mathfrak{a}) = \prod_{i=1}^n a_i, \quad (\mathfrak{a} = (a_1, a_2, \dots, a_n)),$$

expliziert. Dann ist ⁴⁾

$$\kappa_0^*(n, \mathfrak{a}) = (n, \mathfrak{a}^{\mathfrak{f}})_{*\mathfrak{f}}.$$

Die Möbiusfunktion über $Q(\mathfrak{N})$ wurde in [22], § 9 berechnet. Für die Möbiusfunktion μ^* über $Q(\mathfrak{N}^*)$ findet man

$$\mu^*(\mathfrak{a} // \mathfrak{b}) = (-1)^{\omega\left(\frac{\mathfrak{a}}{\mathfrak{b}}\right)},$$

wobei $\omega(\mathfrak{x})$ die Anzahl der verschiedenen Erzeugenden von $\mathfrak{x}$ angibt. Es ist also

$$\mu^*(\mathfrak{a} // \mathfrak{b}) = \mu_0^*\left(\frac{\mathfrak{a}}{\mathfrak{b}}\right) = (-1)^{\sum_{i=1}^n (\omega_0(a_i) - \omega_0(b_i))},$$

wobei $\omega_0(a)$ die Anzahl der verschiedenen Primfaktoren der Zahl a ist.

Definition 6. Eine Funktion $f_0(n, \mathfrak{a})$ über $\mathfrak{N} \times \mathfrak{N}^*$ heißt $\mathfrak{f}$ -unitär in $n \pmod{\mathfrak{a}}$, wenn für alle $n \in \mathfrak{N}$ die Gleichung

$$f_0(n, \mathfrak{a}) = f_0(q^{\mathfrak{f}}, \mathfrak{a})$$

erfüllt ist, wobei zur Abkürzung

$$q^{\mathfrak{f}} = (n, \mathfrak{a}^{\mathfrak{f}})_{*\mathfrak{f}}$$

gesetzt ist.

Ein Beispiel für eine $\mathfrak{f}$ -unitäre Funktion $\pmod{\mathfrak{a}}$ ist die unitäre Ramanujan-Summe

$$c_{\mathfrak{f}}^*(n, \mathfrak{a}) = \sum_{\substack{\mathfrak{x} * \mathfrak{y} = \mathfrak{a} \\ \mathfrak{x}^{\mathfrak{f}} | n}} N_0(\mathfrak{x}^{\mathfrak{f}}) \mu_0^*(\mathfrak{y}).$$

Dabei ist $\mathfrak{x} * \mathfrak{y} = \mathfrak{a}$ eine abkürzende Schreibweise für $\mathfrak{x}\mathfrak{y} = \mathfrak{a}$ mit $(\mathfrak{x}, \mathfrak{y}) = e$, und $\mathfrak{x}^{\mathfrak{f}} | n$ bedeutet, daß $\mathfrak{x}^{\mathfrak{f}}$ ein Teiler von n ist. Es gilt

$$c_{\mathfrak{f}}^*(n, \mathfrak{a}) = c_R^*(n / e, \mathfrak{a} // e) = c_L^*\left([n, \mathfrak{a}^{\mathfrak{f}}] / \frac{\mathfrak{a}^{\mathfrak{f}}}{(n, \mathfrak{a}^{\mathfrak{f}})}, \mathfrak{a} // e\right).$$

⁴⁾ $(a, b)_{*k}$ ist die größte k -te Potenz, welche $(a, b)_*$ teilt.

Funktionen $\xi, \eta, \dots$ über $Q(\mathfrak{N}^*)$ seien nun in der Form

$$\xi(\mathfrak{a} // \mathfrak{b}) = \xi_0 \left(\frac{\mathfrak{a}}{\mathfrak{b}} \right), \quad \eta(\mathfrak{a} // \mathfrak{b}) = \eta_0 \left(\frac{\mathfrak{a}}{\mathfrak{b}} \right), \dots$$

durch Funktionen $\xi_0, \eta_0, \dots$ über $\mathfrak{N}^*$ definiert. Ist ferner $f_0(\mathfrak{n}, \mathfrak{a})$ eine $\mathfrak{k}$ -unitäre Funktion $(\text{mod } \mathfrak{a})$, dann ist die durch

$$f(\mathfrak{n} / \mathfrak{e}, \mathfrak{a} // \mathfrak{e}) = f_0(\mathfrak{n}, \mathfrak{a})$$

über dem kartesischen Produkt der Menge der Enden von $\mathfrak{N}$ mit der Menge der Enden von $\mathfrak{N}^*$ erklärte Funktion ϑ - $\times$ -gleichmäßig in $\mathfrak{n} / \mathfrak{e} \pmod{\mathfrak{a} / \mathfrak{e}}$ (vgl. Definition 2). Aus den Sätzen 4 und 5 erhält man nun folgende Darstellungen für $\mathfrak{k}$ -unitäre Funktionen, wobei die Verknüpfung „ $\times$ “ das unitäre Dirichlet-Produkt bedeutet:

Satz 8. Sei $f_0(\mathfrak{n}, \mathfrak{a})$ eine $\mathfrak{k}$ -unitäre Funktion $(\text{mod } \mathfrak{a})$ und $\xi_0 \times \eta_0 = \varepsilon_0$. Dann hat $f_0(\mathfrak{n}, \mathfrak{a})$ eine Darstellung der Form

$$(31) \quad f_0(\mathfrak{n}, \mathfrak{a}) = (g_0(\mathfrak{x}, \mathfrak{a}) \times \xi_0(\mathfrak{y}))(\mathfrak{q}^*)$$

mit

$$(32) \quad g_0(\mathfrak{b}, \mathfrak{a}) = (f_0(\mathfrak{x}^\dagger, \mathfrak{a}) \times \eta_0(\mathfrak{y}))(\mathfrak{b})$$

für alle unitären Teiler $\mathfrak{b}$ von $\mathfrak{a}$.

Satz 9. Sei $f_0(\mathfrak{n}, \mathfrak{a})$ eine $\mathfrak{k}$ -unitäre Funktion $(\text{mod } \mathfrak{a})$. Dann hat $f_0(\mathfrak{n}, \mathfrak{a})$ eine Darstellung der Form

$$(33) \quad f_0(\mathfrak{n}, \mathfrak{a}) = (c_\dagger^*(\mathfrak{n}, \mathfrak{x}) \times \alpha_0^*(\mathfrak{y}, \mathfrak{a}))(\mathfrak{a}),$$

wobei für alle unitären Teiler $\mathfrak{b}$ von $\mathfrak{a}$

$$(34) \quad \alpha_0^*(\mathfrak{b}, \mathfrak{a}) = \frac{1}{N_0(\mathfrak{a}^\dagger)} (f(\mathfrak{x}^\dagger, \mathfrak{a}) \times c_\dagger^*(\mathfrak{b}^\dagger, \mathfrak{y}))(\mathfrak{a})$$

ist.

Jede $\mathfrak{k}$ -unitäre Funktion über $\mathfrak{N} \times \mathfrak{N}^*$ ist, als Funktion über $\mathfrak{N} \times \mathfrak{N}$ betrachtet, auch $\mathfrak{k}$ -gleichmäßig. Insbesondere folgt aus Satz 3 die Formel

$$(35) \quad c_\dagger^*(\mathfrak{n}, \mathfrak{a}) = \sum_{\substack{\mathfrak{x}\mathfrak{y}=\mathfrak{a} \\ (\mathfrak{y}, \mathfrak{a})=\mathfrak{e}}} c_\dagger(\mathfrak{n}, \mathfrak{x}),$$

wobei die zweite Summationsbedingung gleichbedeutend mit $\gamma_0(\mathfrak{x}) = \gamma_0(\mathfrak{a})$ ist. Formel (35) ist eine Verallgemeinerung eines Satzes von Cohen [21].

§ 7. Einige Identitäten für Funktionen über freien Halbgruppen

Mit zwei multiplikativen Funktionen g und h über $\mathfrak{F}$ und mit

$$q^* = kgV\{x \in \mathfrak{F}^0 : \vartheta_0(x) \mid (n, \vartheta_0(a))_*\}$$

setze man ⁵⁾

$$(36) \quad f(n, a) = \sum_{d \parallel q^*} h(d) g\left(\frac{a}{d}\right) \mu^*\left(\frac{a}{d}\right)$$

und

$$(37) \quad F(a) = f(\vartheta_0(a), a) = \sum_{x \star y = a} h(x) g(y) \mu^*(y).$$

⁵⁾ $a \parallel b$ bedeutet, daß a unitärer Teiler von b ist. Die Summationsbedingung $x \star y = a$ ist gleichbedeutend mit $xy = a$, $(x, y) = e$.

Ist $g(p^t) \neq h(p^t)$ für alle Erzeugendenpotenzen p^t , dann ist $F(a) \neq 0$ für alle $a \in \mathfrak{F}$. Es gilt dann in Analogie zu Satz 16 aus [22] folgender Satz:

Satz 10. Für die durch (36) und (37) definierten Funktionen gelten folgende Identitäten:

$$(38) \quad \sum_{\substack{x \parallel a \\ (x, q^*)_* = e}} \frac{g(x)}{F(x)} = \frac{h(a)F(q^*)}{h(q^*)F(a)};$$

$$(39) \quad f(n, a) = -\frac{F(a)g\left(\frac{a}{q^*}\right)\mu^*\left(\frac{a}{q^*}\right)}{F\left(\frac{a}{q^*}\right)};$$

$$(40) \quad F(a) \sum_{\substack{x \parallel a \\ (x, q^*)_* = e}} \frac{h(x)}{F(x)} \mu^*\left(\frac{a}{x}\right) = \mu^*(a)f(n, a).$$

Der Beweis dieses Satzes verläuft analog zu demjenigen des Satzes 16 aus [22]. Die Identitäten (38) bis (40) kann man der Reihe nach *verallgemeinerte unitäre Landau-Identität*, *verallgemeinerte unitäre Hölder-Identität* und *verallgemeinerte unitäre Brauer-Rademacher-Identität* nennen. Sie wurden für den Spezialfall, daß $\mathfrak{F}$ die Menge der natürlichen Zahlen und ϑ_0 die identische Abbildung ist, zuerst von P. J. Mc Carthy [17] bewiesen.

§ 8. Multiplikativität von Funktionen über einer Quotientenhalbordnung

Das Rechnen mit zahlentheoretischen Funktionen, vor allem die Bestimmung ihrer Funktionswerte, ist besonders einfach, wenn die betreffenden Funktionen multiplikativ sind. Es soll nun versucht werden, den Begriff der Multiplikativität auch für Funktionen über der Quotientenhalbordnung $Q(\mathfrak{F})$ einer lokal endlichen Halbordnung $\mathfrak{F}$ zu erklären. Zu diesem Zweck wird zunächst die Teilerfremdheit zweier Quotienten definiert und die Zerlegbarkeit eines Quotienten in ein Produkt von nicht weiter zerlegbaren paarweise teilerfremden Quotienten bewiesen. Eine Funktion f über $Q(\mathfrak{F})$ soll dann multiplikativ heißen, wenn für $a/b, b/c \in Q(\mathfrak{F})$ die Beziehung

$$(41) \quad f(a/b)f(b/c) = f(a/c)f(b/b)$$

gilt, sobald a/b und b/c teilerfremd sind.

Bemerkung. Man kann die Multiplikativität einer zahlentheoretischen Funktion f durch das Bestehen der Gleichung

$$f(a)f(b) = f(kgV(a, b))f(ggT(a, b))$$

für alle natürlichen Zahlen a, b definieren. Ist $\mathfrak{F}$ eine Hierarchie, so liegt es daher scheinbar nahe, die Multiplikativität einer Funktion f über $Q(\mathfrak{F})$ durch das Bestehen der Gleichung

$$f(A)f(B) = f(\sup(A, B))f(\inf(A, B))$$

für alle $A, B \in Q(\mathfrak{F})$ zu erklären. Aus dieser Beziehung würde aber schon ohne eine Voraussetzung über a/b und b/c die Gleichung (41) folgen. Man hätte also ein Analogon zur vollständigen Multiplikativität zahlentheoretischer Funktionen.

Es sei nun $\mathfrak{H}$ eine lokal endliche Halbordnung mit einem kleinsten Element o . Jedem Element $k \in \mathfrak{H}$ sei eine ordnungsisomorphe Abbildung τ_k der Menge $\{a \in \mathfrak{H} : k < a\}$ auf eine Teilmenge von $\mathfrak{H}$ mit folgenden Eigenschaften zugeordnet:

- (a) $\tau_a(a) = o$ für alle $a \in \mathfrak{H}$;
- (b) $\tau_o(a) = a$ für alle $a \in \mathfrak{H}$;
- (c) aus $c < b < a$ folgt $\tau_b(a) < \tau_c(a)$.

Die Menge dieser Abbildungen heie T . Für $a/b \in Q(\mathfrak{H})$ und $k < b < a$ definiere man

$$\tau_k(a/b) = \tau_k(a) / \tau_k(b).$$

Dies ist eine Abbildung von $Q(\mathfrak{H})$ in sich mit folgenden Eigenschaften:

- (d) für $k < c$ folgt aus $a/b < c/d$ die Relation $\tau_k(a/b) < \tau_k(c/d)$, d. h. die Abbildung τ_k ist ordnungstreu;
- (e) aus $a/b < c/d$ folgt $\tau_b(a/b) < \tau_d(c/d)$;
- (f) ist $b \in a/c$ und $k < c$, dann gilt

$$\tau_k(a/c) = \tau_k(a/b) * \tau_k(b/c).$$

Definition 7. Zwei Quotienten $A = a/a'$ und $B = b/b'$ heißen *teilerfremd* (bezüglich T), wenn die Quotienten $\tau_{a'}(a/a')$ und $\tau_{b'}(b/b')$ nur das Element o gemeinsam haben, wenn also $\inf(\tau_{a'}(a), \tau_{b'}(b))$ existiert und gleich o ist.

Definition 8. Ein Quotient P , der nicht als Produkt $P = P_1 * P_2$ von zwei teilerfremden Quotienten P_1 und P_2 , von denen keiner ein Einheitsquotient ist, geschrieben werden kann, heißt *Primquotient* (bezüglich T).

Definition 9. Zwei Quotienten $A = a/a'$ und $B = b/b'$ heißen *äquivalent* (bezüglich T), wenn $\tau_{a'}(a) = \tau_{b'}(b)$ ist.

Lemma 3. Jeder Quotient $A \in Q(\mathfrak{H})$ kann als Produkt von paarweise teilerfremden Primquotienten geschrieben werden.

Beweis. Ist A nicht Primquotient, dann gibt es eine Zerlegung $A = A_1 * A_2$ in zwei zueinander teilerfremde Quotienten A_1 und A_2 . Sind dies keine Primquotienten, so kann man sie weiter in zueinander teilerfremde Quotienten zerlegen. Da A eine endliche Menge ist, bricht dieser Proze einmal ab. Die Primquotienten der so entstehenden Zerlegung sind paarweise teilerfremd; denn für $B_1 < A_1$ und $B_2 < A_2$ sind mit A_1 und A_2 nach (e) auch B_1 und B_2 teilerfremd.

Auf die oben beschriebene Art wird man im allgemeinen zu wesentlich verschiedenen Zerlegungen eines Quotienten in paarweise teilerfremde Primquotienten kommen. Es seien

$$a/b = x_0 / x_1 * x_1 / x_2 * \cdots * x_{r-1} / x_r$$

mit $x_0 = a$ und $x_r = b$ und

$$a/b = y_0 / y_1 * y_1 / y_2 * \cdots * y_{s-1} / y_s$$

mit $y_0 = a$ und $y_s = b$ zwei solche Zerlegungen, wobei keine Einheitsquotienten auftreten sollen. Diese beiden Zerlegungen sollen *nicht wesentlich verschieden* (bezüglich T) heißen, wenn $r = s$ ist und die Quotienten $\tau_{x_i}(x_{i-1}/x_i)$ eine Permutation der Quotienten $\tau_{y_i}(y_{i-1}/y_i)$ sind. Nicht wesentlich verschiedene Zerlegungen haben also äquivalente Primquotienten. Sind für alle $A \in Q(\mathfrak{H})$ je zwei Zerlegungen in ein Produkt von paar-

weise teilerfremden Primquotienten nicht wesentlich verschieden, so sage man, in $Q(\mathfrak{F})$ gelte die eindeutige Primzerlegung (bezüglich T). Wann dies der Fall ist, hängt von $\mathfrak{F}$ und von T ab.

Es werde nun vorausgesetzt, daß in $Q(\mathfrak{F})$ die eindeutige Primzerlegung (bezüglich T) gilt.

Definition 10. Eine Funktion f über $Q(\mathfrak{F})$ heißt *multiplikativ* (bezüglich T), wenn

- (i) $f(I) \neq 0$ für alle Einheitsquotienten I ,
- (ii) $f(P) = f(\tau_q(P))f(q|q)$ für alle Primquotienten $P = p|q$,
- (iii) $f(a|b)f(b|c) = f(a|c)f(b|b)$ gilt, sobald $a|b$ und $b|c$ zueinander teilerfremd sind.

Definition 11. Eine Funktion f über $Q(\mathfrak{F})$ heißt *endenmultiplikativ* (bezüglich T), wenn (a) und (b) aus Definition 10 gelten und (iii) aus Definition 10 nur für $c = o$ gilt.

Satz 11. Es sei $A = P_1 * P_2 * \cdots * P_r$ die Zerlegung von A in zueinander teilerfremde Primquotienten, wobei $A = a|b = x_0|x_r$ und $P_i = x_{i-1}|x_i$ ist. Wenn f multiplikativ (bezüglich T) ist, dann gilt

$$(42) \quad f(A) = f(b|b) \prod_{i=1}^r f(\tau_{x_i}(P_i)).$$

Ist f nur endenmultiplikativ, dann gilt (42) nur für $b = o$. Ist f endenmultiplikativ und ist außerdem A teilerfremd zu $B = b|o$ und $f(B) \neq 0$, dann gilt ebenfalls Formel (42).

Beweis. Nach (iii) von Definition 10 gilt

$$\prod_{i=1}^r f(P_i) = f(A) \prod_{i=1}^{r-1} f(x_i|x_i).$$

Andererseits ist nach (ii)

$$\prod_{i=1}^r f(P_i) = \prod_{i=1}^r f(\tau_{x_i}(P_i)) \prod_{i=1}^r f(x_i|x_i),$$

woraus nach (i) die Behauptung folgt. Ebenso erkennt man, daß für eine endenmultiplikative Funktion f Formel (42) gilt, falls $A = a|o$, also A ein Ende ist. Zum Beweis des letzten Teiles des Satzes sei $B = Q_1 * Q_2 * \cdots * Q_s$ die Zerlegung von B in zueinander teilerfremde Primquotienten. Dann gelten die Beziehungen

$$f(A * B) = f(o|o) \prod_{i=1}^r f(\tau_{x_i}(P_i)) \prod_{j=1}^s f(\tau_{x_j}(Q_j))$$

und

$$f(B) = f(o|o) \prod_{j=1}^s f(\tau_{x_j}(Q_j)),$$

woraus

$$f(A * B) = f(B) \prod_{i=1}^r f(\tau_{x_i}(P_i))$$

folgt. Andererseits gilt nach Definition 11

$$f(A * B) = \frac{f(A)f(B)}{f(b|b)},$$

so daß wegen $f(B) \neq 0$ auch die letzte Behauptung des Satzes bewiesen ist.

Beispiel 1. In $\mathfrak{F}^0$ definiere man die Abbildungsmenge T durch

$$\tau_k(a) = \frac{a}{k} \text{ für alle Teiler } k \text{ von } a.$$

Es sei f_0 eine Funktion über $\mathfrak{F}^0$, welche im üblichen Sinn multiplikativ ist, d. h. es ist $f_0(e) = 1$ und $f_0(ab) = f_0(a)f_0(b)$ für $(a, b) = e$. Die durch $f(a/b) = f_0\left(\frac{a}{b}\right)$ über $Q(\mathfrak{F}^0)$ definierte Funktion f ist dann multiplikativ (bezüglich T), denn es gilt $f(I) = f_0(e) \neq 0$, $f(a/b) = f\left(\frac{a}{b}/e\right)$ und ferner, da für (bezüglich T) teilerfremde Quotienten a/b und b/c auch $\left(\frac{a}{b}, \frac{b}{c}\right) = e$ und somit $f_0\left(\frac{a}{b}\right)f_0\left(\frac{b}{c}\right) = f_0\left(\frac{a}{c}\right)f_0\left(\frac{b}{b}\right)$ gilt,

$$f(a/b)f(b/c) = f(a/c)f(b/b),$$

falls a/b zu b/c teilerfremd (bezüglich T) ist.

In $Q(\mathfrak{F}^0)$ gilt natürlich die eindeutige Primzerlegung. Dabei sind die $\tau_{x_i}(x_{i-1})$ der Zerlegung von a/b in paarweise teilerfremde Primquotienten die maximalen Erzeugendenpotenzen von $\frac{a}{b}$.

Beispiel 2. Im Gegensatz zu Beispiel 1 sei jetzt eine Funktion f über $Q(\mathfrak{F}^0)$ durch $f(a/b) = \sum_{[x,b]=a} f_0(x)$ definiert. Ferner sei $\sum_{x|a} f_0(x) \neq 0$ für alle $a \in \mathfrak{F}$. Dann ist für $a \in \mathfrak{F}$

$$f(a/a) = \sum_{[x,a]=a} f_0(x) = \sum_{x|a} f_0(x) \neq 0.$$

Weiter gilt für ein beliebiges $a \in \mathfrak{F}$ und eine Erzeugendenpotenz p^t

$$\begin{aligned} f(ap^t/a) &= \sum_{[x,a]=ap^t} f_0(x) = \sum_{y|a} f_0(y p^t) = f_0(p^t)f(a/a) \\ &= f(\tau_a(ap^t/a))f(a/a). \end{aligned}$$

Ist $A = a/b$, $B = b/e$ und A zu B teilerfremd, d. h. $\left(\frac{a}{b}, b\right) = e$, dann gilt

$$\begin{aligned} f(A)f(B) &= \sum_{[x,b]=a} f_0(x) \sum_{[y,e]=b} f_0(y) = f_0(b) \sum_{[x,b]=a} f_0(x) \\ &= f_0(b) \sum_{[x,b]=\frac{a}{b}b} f_0(x) = f_0(b) \sum_{y|\frac{a}{b}} f_0\left(\frac{a}{b}y\right) \\ &= f_0(b)f_0\left(\frac{a}{b}\right) \sum_{y|b} f_0(y) = f_0(a) \sum_{y|b} f_0(y) \\ &= \sum_{[x,e]=a} f_0(x) \sum_{[y,b]=b} f_0(y) = f(A * B)f(b/b). \end{aligned}$$

Damit ist gezeigt, daß f eine endenmultiplikative Funktion über $Q(\mathfrak{F}^0)$ ist.

Literatur

- [21] E. Cohen, Unitary functions (mod r). I, Duke Math. J. **28** (1961), 475—485.
 [22] H. Scheid, Arithmetische Funktionen über Halbordnungen. I, J. Reine Angew. Math. **231** (1968), 192—214.