Studia Scientiarum Mathematicarum Hungarica

AUXILIO
CONSILII INSTITUTI MATHEMATICI
ACADEMIAE SCIENTIARUM HUNGARICAE

REDIGIT
L. FEJES TÓTH

ADIUVANTIBUS
Á. CSÁSZÁR, I. CSISZÁR, A. HAJNAL, E. MAKAI,
P. RÉVÉSZ, O. STEINFELD, T. E. SCHMIDT,
J. SZABADOS, P. TURÁN, I. VINCZE

TOMUS XII.
FASC. 1–2.
1977

AKADÉMIAI KIADÓ, BUDAPEST

Studia Scientiarum Mathematicarum Hungarica

A Magyar Tudományos Akadémia matematikai folyóirata

Szerkesztőség: 1053 Budapest V., Reáltanoda u. 13—15.

Technikai szerkesztő: Deák E.

Kiadja az Akadémiai Kiadó, 1054 Budapest V., Alkotmány u. 21.

A *Studia Scientiarum Mathematicarum Hungarica* angol, német, francia vagy orosz nyelven közöl eredeti értekezéseket a matematika tárgyköréből. Félévenként jelenik meg, évi egy kötetben.

Előfizetési ára belföldre 120,— Ft, külföldre 165,— Ft. Megrendelhető a belföld számára az Akadémiai Kiadónál, a külföld számára pedig a Kultúra Könyv és Hírlap Külkereskedelmi Vállalatnál (1011 Budapest I., Fő u. 32.).

Cserekapcsolatok felvétele ügyében kérjük a MTA Matematikai Kutató Intézete Könyvtárához (1053 Budapest V., Reáltanoda u. 13—15) fordulni.

Közlésre szánt dolgozatokat kérjük két példányban a szerkesztőség címére küldeni.

Studia Scientiarum Mathematicarum Hungarica is a journal of the Hungarian Academy of Sciences publishing original papers on mathematics, in English, German, French or Russian.

It is published semiannually, making up one volume per year.

Editorial Office: 1053 Budapest V., Reáltanoda u. 13—15, Hungary.
Technical Editor: E. Deák

Orders may be placed with *Kultura* Trading Co. for Books and Newspapers, Budapest 62, P. O. B. 149 or with its representatives abroad.

For establishing exchange relations please write to the Library of the Mathematical Institute (1053 Budapest V., Reáltanoda u. 13—15.)

Papers intended for publication should be sent to the Editor in 2 copies.

Studia Scientiarum Mathematicarum Hungarica

Auxilio
Consilii Instituti Mathematici
Academiae Scientiarum Hungaricae

Redigit

L. Fejes Tóth

Adiuvantibus

Á. Császár, I. Csiszár, A. Hajnal, P. Révész, O. Steinfeld, T. E. Schmidt, J. Szabados, D. Szász, I. Vincze

Tomus XII

Akadémiai Kiadó, Budapest
1977

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Tomus XII

INDEX

80-293 — Szegedi Nyomda — F. v.: Dobó József igazgató

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 1—17

ÜBER ERZEUGENDENMENGEN ENDLICHER PARTITIONENVERBÄNDE

von

H. STRIETZ

Den Ausgangspunkt der Überlegungen zu dieser Arbeit bildete die Frage nach Möglichkeiten der Beschreibung der Elemente in dem Verband $\Pi(M)$ aller Partitionen einer endlichen Menge M — dieser Verband wird auch Partitionenverband genannt. Die folgenden Untersuchungen liefern einen Beitrag zu der Frage, wie ausgehend von möglichst wenigen Elementen von $\Pi(M)$ mit den Verbandsoperationen Infimums- und Supremumsbildung jedes andere Element von $\Pi(M)$ beschrieben werden kann, d. h. wenige Elemente $e_1, e_2, \ldots, e_k$ in $\Pi(M)$ zu finden, so daß zu jedem $x \in \Pi(M)$ ein Verbandsterm t mit $x = t(e_1, e_2, \ldots, e_k)$ existiert.

Hauptergebnis dieser Arbeit ist, daß eine kardinale Summe von Ketten genau dann einen Homomorphismus auf eine Erzeugendenmenge eines Partitionenverbandes $\Pi(M)$ mit $|M| \geqq 10$ besitzt, wenn die Summe aus wenigstens 3 Ketten besteht und im Fall, daß es genau drei Ketten sind, wenigstens eine Kette mehr als ein Element hat. Für endliche Partitionenverbände werden in diesem Zusammenhang explizit minimale Erzeugendenmengen angegeben.

Für das anschließende Problem der Klassifizierung minimaler Erzeugendenmengen wird ein Defektbegriff für Erzeugendenmengen eingeführt, mit dem u. a. gezeigt werden kann, daß die Anzahl nichtisomorpher minimaler Erzeugendenmengen mit zunehmender Elementenanzahl von M über alle Grenzen wächst.

1. Grundlagen und Hauptergebnis

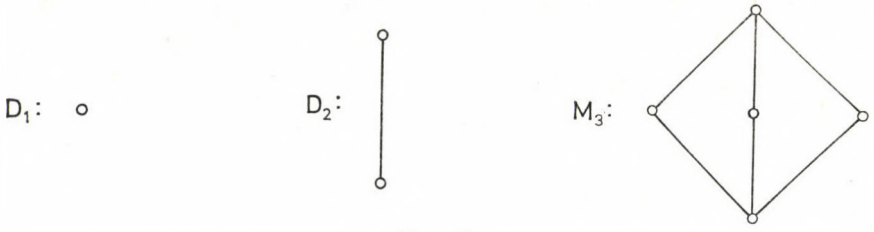
In dieser Arbeit sei stets $M := \{1, 2, \ldots, n\}$ eine endliche Menge. Eine Partition (von M) ist eine Zerlegung von M in nichtleere, paarweise disjunkte Teilmengen — „Blöcke" genannt —, so daß deren Vereinigung wieder M ergibt. In der Menge aller Partitionen einer Menge M ist durch das Enthaltensein der Blöcke einer Partition x in denen einer Partition y auf natürliche Weise eine Ordnungsrelation gegeben. Die halbgeordnete Menge aller Partitionen von M ist ein Verband und wird als Partitionenverband $\Pi(M)$ bezeichnet. Sind $x, y \in \Pi(M)$, so bezeichnet $x+y$ bzw. xy das Supremum bzw. das Infimum der Elemente x und y. Das größte Element von $\Pi(M)$, die Partition, die allein aus dem Block M besteht, wird mit **1** bezeichnet, die kleinste Partition, die aus genau n einelementigen Blöcken besteht, mit **0**. Eine Partition $x \in \Pi(M)$ mit k Blöcken K_i $(i \in I := \{1, \ldots, k\})$, soll allgemein wie folgt angegeben werden: $x = (K_i)_{i \in I}$.

$\Pi(M)$, $|M| \in \mathbf{N}$, ist ein vollständiger, einfacher, halbmodularer und atomistischer Verband (vgl. BIRKHOFF [1], S. 15 und 95, bzw. ORE [2], S. 573 ff).

Ist n eine natürliche Zahl, so bezeichne $\mathbf{n}$ die n-elementige Kette (total geordnete Menge). Weiter sei $\mathbf{n}_1+\mathbf{n}_2+\ldots+\mathbf{n}_k$ die kardinale (disjunkte) Summe der Ketten $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_k$ mit den Kardinalitäten $n_1, n_2, \ldots, n_k$.

Der Begriff einer Erzeugendenmenge, sowie der einer minimalen Erzeugendenmenge eines Verbandes, wird wie üblich verwendet. $\langle P\rangle = \Pi(M)$ bedeutet, daß die Menge P von Partitionen den zugehörigen Partitionenverband $\Pi(M)$ erzeugt.

Interessant für die folgenden Untersuchungen sind nur Verbände $\Pi(M)$ mit $|M| \geqq 4$, da $\Pi(\{1\}) \cong D_1$, $\Pi(\{1,2\}) \cong D_2$ und $\Pi(\{1,2,3\}) \cong M_3$ ist, wobei D_1 den einelementigen Verband, D_2 den zweielementigen Verband und M_3 den fünfelementigen modularen, nichtdistributiven Verband bezeichnet (vgl. Figur 1).



Figur 1.

Der in dieser Arbeit bewiesene Hauptsatz lautet nun:

SATZ 1.1 *Sei P kardinale Summe von Ketten und M eine endliche Menge mit $|M| \geqq 10$. So gilt: $\Pi(M)$ wird genau dann von einem homomorphen Bild von P erzeugt, wenn P eine isomorphe Kopie der halbgeordneten Mengen $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder $\mathbf{1}+\mathbf{1}+\mathbf{2}$ enthält.*

Eine Beweisrichtung erfolgt durch das Aufzeigen der Existenz von Erzeugendenmengen welche als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ bzw. zu $\mathbf{1}+\mathbf{1}+\mathbf{2}$ sind, die andere Beweisrichtung liefert der folgende Satz, der zusätzlich die Minimalität dieser Erzeugendenmengen zeigt.

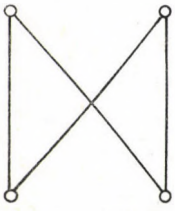
SATZ 1.2 (siehe WILLE [4], S. 455). *Für eine endliche halbgeordnete Menge P sind die folgenden Bedingungen äquivalent:*

(i) *Es gibt (bis auf Isomorphie) nur endlich viele einfache Verbände, die von einem homomorphen Bild von P erzeugt werden.*

(ii) *Jeder von einem homomorphen Bild von P erzeugte einfache Verband ist isomorph zu D_1, D_2 oder M_3.*

(iii) *P enthält keine der folgendenh albgeordneten Mengen: $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$, $\mathbf{1}+\mathbf{1}+\mathbf{2}$ und $\mathbf{1}+\mathbf{K}_2$.*

Dabei wird die halbgeordnete Menge $\mathbf{1}+\mathbf{K}_2$ durch folgendes Diagramm beschrieben:



Figur 2.

KOROLLAR 1.2.1. (a. a. O.) *Die einzigen einfachen Verbände, die durch ein homomorphes Bild zweier endlichen Ketten erzeugt werden, sind D_1 und D_2.*

KOROLLAR 1.2.2. (a. a. O.) *Die einzigen von 3 Elementen erzeugten einfachen Verbände sind D_1, D_2 und M_3.*

2. Minimale Erzeugendenmengen von $\Pi(M)$

In diesem Abschnitt werden minimale Erzeugendenmengen von $\Pi(M)$ angegeben, die als halbgeordnete Mengen isomorph zu $\mathbf{1+1+1+1}$ bzw. zu $\mathbf{1+1+2}$ sind. In den Beweisen zeigt sich, daß nach einer gewissen Anzahl von Operationen ausgezeichnete Atome von $\Pi(M)$ durch eine Rekursion erzeugt werden können. Atome von $\Pi(M)$ sind die Partitionen, die aus genau einem zweielementigen und $n-2$ einelementigen Blöcken bestehen, falls $|M|=n$ gilt. Diese ausgezeichneten Atome gestatten es, alle übrigen Atome, und damit auch alle anderen Elemente von $\Pi(M)$ zu erzeugen, da Partitionenverbände atomistisch sind.

Atome von $\Pi(M)$ werden wie folgt bezeichnet:

$$t_{\alpha,\beta} := (\alpha, \beta) \quad \alpha \neq \beta \quad \text{und} \quad \alpha, \beta \in M.$$

Einelementige Blöcke werden im weiteren aus Vereinfachungsgründen nicht geschrieben, es sei denn, Mißverständnisse wären möglich. Speziell wird in den Beweisen angegeben, wie die ausgezeichneten Atome $t_{1,i}$ und $t_{2,j}$ mit $2 \leqq i \leqq n$ und $3 \leqq j \leqq n$ erzeugt werden. Ein beliebiges anderes Atom $t_{\alpha,\beta} \in \Pi(M)$ erhält man dann durch:

$$(*) \qquad t_{\alpha,\beta} = (t_{1,\alpha} + t_{1,\beta})(t_{2,\alpha} + t_{2,\beta}), \quad \alpha, \beta \neq 1, 2.$$

SATZ 2.1. *Zu jedem endlichen Partitionenverband $\Pi(M)$, $|M| \geqq 10$, existieren Erzeugendenmengen $P \subseteq \Pi(M)$, die als halbgeordnete Mengen isomorph zu* $\mathbf{1+1+2}$ *sind.*

BEWEIS. Der Beweis von Satz 2.1 erfolgt in zwei Schritten: Zunächst werden Erzeugendenmengen für $\Pi(M)$ mit $|M| = n \equiv 1 \pmod 3$, $n \geqq 16$, angegeben, danach folgt auf einfache Weise der allgemeine Fall.

Sei $M = \{1, 2, \ldots, n\}$, $16 < n < \infty$ und $n \equiv 1 \pmod 3$. Dann wird $\Pi(M)$ von den folgenden Partitionen erzeugt:

$e_1 := (1, 2), (3), \ldots, (n),$

$e_2 := (1, 2, 11), (3, 4, 12), (5, 6, 8), (7, 9, 10), \ldots, (k+3, k+4, k+5), \ldots, (n),$

$e_3 := (1, 4, 6), (3, 5, 9), (2, 7, 12), \ldots, (k-2, k, k+5), \ldots, (n-5, n-3,) (n-2, n),$

$e_4 := (2, 3), (4, 8), (1, 5, 10), (6, 7, 13), \ldots, (k-1, k+1, k+6), \ldots$

$\ldots, (n-4, n-2), (n-1),$

dabei ist $k \equiv 1 \pmod 3$ und k durchläuft die Werte $10, 13, \ldots, n-6$. Diese Notation ist demnach so zu verstehen, daß nacheinander jeweils in e_2, e_3 und e_4 die möglichen Werte für k (abhängig von n) einzusetzen sind. Zunächst ist festzustellen:

(a) e_1, e_2, e_3, e_4 sind Partitionen von M, denn es ist $k-1, k+5 \equiv 0 \pmod 3$, $k, k+3$, $k+6 \equiv 1 \pmod 3$ und $k-2$, $k+1$, $k+4 \equiv 2 \pmod 3$. Die ersten explizit gegebenen Blöcke überprüft man direkt.

(b) $(\{e_1, e_2, e_3, e_4\}, \leqq) \cong \mathbf{1+1+2}$, denn es ist $e_1 < e_2$.

(c) $e_i e_j = \mathbf{0}$ für $i \neq j$ und $i, j \in \{2, 3, 4\}$, da die Elemente der durch die Variable k beschriebenen Blöcke in drei verschiedenen Blöcken der anderen Partitionen liegen, die ersten Blöcke überprüft man wiederum direkt.

In den folgenden Schritten (1) bis (45), die, um eine bessere Lesbarkeit zu gewährleisten, ausführlich angegeben werden, sollen die Atome $t_{1,i}$ und $t_{2,j}$ mit $2 \leqq i \leqq 13$ sowie $3 \leqq j \leqq 13$ erzeugt werden. Daran anschließend ergibt sich die Möglichkeit für Rekursionsformeln für $t_{1,i}$ und $t_{2,j}$ mit $14 \leqq i, j \leqq n$.

(1) $t_{6,7} = (e_1+e_3)e_4 = (6, 7)$

(2) $t_{3,5} = (e_1+e_4)e_3 = (3, 5)$

(3) $q_1 = (t_{6,7}+e_2)e_3 = (5, 9),\ (8, 10)$

(4) $t_{4,8} = (t_{3,5}+e_2)e_4 = (4, 8)$

(5) $q_2 = (t_{6,7}+e_2)e_4 = (6, 7),\ (5,10)$

(6) $q_3 = (q_1+q_2)e_2 = (5, 8),\ (9, 10)$

(7) $t_{5,10} = (q_1+q_3)q_2 = (5, 10)$

(8) $t_{6,8} = (t_{4,8}+e_3)e_2 = (6, 8)$

(9) $t_{4,6} = (t_{4,8}+t_{6,8})e_3 = (4, 6)$

(10) $q_4 = (q_3+e_4)e_3 = (1, 4),\ (5, 9),\ (8, 10)$

(11) $q_5 = (t_{6,7}+t_{6,8}+e_3)e_2 = (1, 2),\ (4, 12),\ (6, 8)\ (7, 10)$

(12) $q_6 = (t_{6,7}+t_{6,8}+e_3)e_4 = (4, 8),\ (1, 10),\ (6, 7)$

(13) $q_7 = q_2+q_5 = (1, 2),\ (4, 12),\ (5, 6, 7, 8, 10)$

(14) $t_{8,10} = q_1q_7 = (8, 10)$

(15) $t_{5,8} = (t_{5,10}+t_{8,10})q_3 = (5, 8)$

(16) $q_8 = (t_{8,10}+q_6)q_4 = (1, 4),\ (8, 10)$

(17) $q_9 = (t_{4,8}+t_{5,8}+t_{3,5})e_2 = (3, 4),\ (5, 8)$

(18) $q_{10} = (e_1+q_8+q_9)e_4 = (2, 3),\ (5, 10)$

(19) $q_{11} = (q_8+q_9)(e_1+q_{10}) = (1, 3),\ (5, 10)$

(20) $t_{1,10} = (t_{3,5}+q_{11})q_6 = (1, 10)$

(21) $t_{2,10} = (e_1+t_{1,10})(t_{3,5}+q_{10}) = (2, 10)$

(22) $t_{2,7} = (t_{2,10}+e_2)e_3 = (2, 7)$

(23) $t_{7,10} = (t_{2,7}+t_{2,10})e_2 = (7, 10)$

(24) $t_{1,7} = (t_{1,10}+t_{7,10})(e_1+t_{2,7}) = (1, 7)$

(25) $t_{1,4} = (t_{1,7}+t_{6,7}+t_{4,6})q_8 = (1, 4)$

(26) $t_{2,4} = (q_9+q_{10})(e_1+t_{1,4}) = (2, 4)$

(27) $t_{2,3} = (t_{24}+q_9)q_{10} = (2, 3)$

(28) $t_{1,3} = (t_{2,3}+e_1)q_{11} = (1, 3)$

(29) $t_{1,5} = (t_{1,3}+t_{3,5})e_4 = (1, 5)$

(30) $t_{2,5} = (e_1+t_{1,5})(t_{2,3}+t_{3,5}) = (2, 5)$

(31) $t_{1,6} = (t_{1,4}+t_{4,6})(t_{1,7}+t_{6,7}) = (1, 6)$

(32) $t_{2,6} = (t_{1,6}+e_1)(t_{2,4}+t_{4,6}) = (2, 6)$

(33) $t_{1,8} = (t_{1,4}+t_{4,8})(t_{1,6}+t_{6,8}) = (1, 8)$

(34) $t_{2,8} = (t_{1,8}+e_1)(t_{2,6}+t_{6,8}) = (2, 8)$

(35) $t_{1,9} = (t_{1,5}+q_1)(q_3+t_{1,10}) = (1, 9)$

(36) $t_{2,9} = (t_{1,9}+e_1)(q_3+t_{2,10}) = (2, 9)$

(37) $t_{2,11} = (t_{2,9}+e_4)e_2 = (2, 11)$

(38) $t_{1,11} = (t_{2,11}+e_1)(t_{1,9}+e_4) = (1, 11)$

(39) $q_{12} = (t_{2,3}+e_3)e_2 = (3, 12),\ (7, 9)$

(40) $t_{2,12} = (t_{2,3}+q_{12})(t_{2,4}+q_5) = (2, 12)$

(41) $t_{1,12} = (t_{2,12}+e_1)(t_{1,3}+q_{12}) = (1, 12)$

(42) $t_{6,13} = (t_{1,11}+e_3)e_4 = (6, 13)$

(43) $t_{11,13} = (t_{6,11}+t_{6,13})e_3 = (11, 13)$

(da: $t_{6,11} = (t_{1,6}+t_{1,11})(t_{2,6}+t_{2,11})$)

(44) $t_{1,13} = (t_{1,6}+t_{6,13})(t_{1,11}+t_{11,13}) = (1, 13)$

(45) $t_{2,13} = (t_{2,6}+t_{6,13})(t_{2,11}+t_{11,13}) = (2,13).$

Mit den folgenden Rekursionsformeln lassen sich nun alle weiteren $t_{1,i}$ und $t_{2,j}$ erzeugen:

(46) $t_{1,k+4} = [(t_{k+2,k+3}+e_4)e_2+t_{1,k+3}][(t_{k+2,k+3}+e_2)e_4+t_{1,k+2}]$

(47) $t_{2,k+4} = [(t_{k+2,k+3}+e_4)e_2+t_{2,k+3}][(t_{k+2,k+3}+e_2)e_4+t_{2,k+2}]$

(48) $t_{1,k+5} = [(t_{k-2,k+3}+e_2)e_3+t_{1,k-2}][(t_{k-2,k+3}+e_3)e_2+t_{1,k+3}]$

(49) $t_{2,k+5} = [(t_{k-2,k+3}+e_2)e_3+t_{2,k-2}][(t_{k-2,k+3}+e_3)e_2+t_{2,k+3}]$

(50) $t_{1,k+6} = [(t_{k-1,k+4}+e_3)e_4+t_{1,k-1}][(t_{k-1,k+4}+e_4)e_3+t_{1,k+4}]$

(51) $t_{2,k+6} = [(t_{k-1,k+4}+e_3)e_4+t_{2,k-1}][(t_{k-1,k+4}+e_4)e_3+t_{2,k+4}].$

Dabei durchläuft k nacheinander die Werte $10, 13, 16, \ldots, n-6$. Die in (46) bis (51) verwendeten Atome $t_{k+2,k+3}$, $t_{k-2,k+3}$ und $t_{k-1,k+4}$ sind (zum notwendigen Zeitpunkt) nach (*) aus bereits erzeugten Atomen konstruierbar. Mit (*) lassen sich weiterhin alle übrigen Atome von $\Pi(M)$, und damit alle anderen Elemente des Partitionenverbandes von M erzeugen, da die Atome $t_{1,i}$ und $t_{2,j}$ für $2 \leqq i \leqq n$

und $3 \leqq j \leqq n$ erzeugt werden können. Voraussetzung ist bisher allerdings noch $|M| \equiv 1 \pmod 3$ und $|M| \geqq 16$.

In den Erzeugungsschritten (1) bis (41) überzeugt man sich leicht davon, daß in diesen Partitionen Elemente von M, die größer als 12 sind, nur in einelementigen Blöcken auftreten. (c) gewährleistet dabei stets den Schnitt der übrigen Blöcke zu einelementigen Blöcken. Das Prinzip der Rekursionsformeln (46) bis (51) ist nun wie folgt (man vergleiche dazu (44) und (45), dort wird schon in zu (50) und (51) analoger Form vorgegangen):

Zum Beispiel: Zur Konstruktion des Atomes $t_{1,k+4}$ werden in e_2 und e_4 die das Element $k+4$ enthaltenden Blöcke gesucht. Nun sind die Erzeugenden gerade so gewählt, daß in diesen Blöcken zusammen mit $k+4$ noch andere kleinere (wenn man dabei in M die natürliche Ordnung wählt) Elemente α bzw. β von M liegen. Hier ist dies in e_2 das Element $\alpha = k+3$ und in e_4 das Element $\beta = k+2$. Verbindet man e_2 mit dem Atom $t_{k+2,k+3}$ und schneidet dann mit e_4, so hat diese Partition einen Block $(k+2, k+4)$ und eventuell einen weiteren mehrelementigen Block, der aber nicht die Elemente 1 oder 2 von M enthält. Entsprechend verfährt man bei der Verbindung von e_4 mit $t_{k+2,k+3}$ und dem Schnitt mit e_2. Verbindet man nun die entstandenen Partitionen mit $t_{1,k+2}$ bzw. $t_{1,k+3}$ und bildet das Infimum dieser beiden, so ergibt sich die gewünschte Partition.

An dieser Stelle ist festzustellen, die Erzeugung der Atome nach den Formeln (46) bis (51) ist nicht von der Kardinalität von M abhängig (M muß nur mindestens 10 Elemente enthalten).

Aufgrund dessen ergibt sich der allgemeine Fall: Ist $N = \{1, 2, \ldots, m\}$, $N \subseteq M$, $10 \leqq m \leqq n$, so bilden die angegebenen Erzeugenden e_1, e_2, e_3, e_4 eingeschränkt auf N eine Erzeugendenmenge $\{e_1', e_2', e_3', e_4'\}$ von $\Pi(N)$, d. h. es ist: $\langle e_1', e_2', e_3', e_4' \rangle = \Pi(N)$.

Für $10 \leqq m \leqq 16$ sind dies gerade die Schritte (1) bis (36) (für $m = 10$) bzw. bis (51) (für $m = 16$, wobei für k der Wert 10 einzusetzen ist). Man überzeugt sich leicht davon, bei einer Einschränkung von M auf N kommt es nie vor, daß ein Element von N zugleich in zwei einelementigen Blöcken von e_2, e_3 und e_4 liegt, d. h. die Formeln (46) bis (51) nicht mehr anwendbar würden.

Damit ist Satz 2.1 bewiesen.

Da es sich bei Partitionenverbänden um einfache Verbände handelt, ist mit Satz 1.2 festzustellen:

KOROLLAR 2.2. *Die Erzeugendenmenge* $\{e_1, e_2, e_3, e_4\}$ *(wie im Beweis zu Satz 2.1 angegeben) ist minimale Erzeugendenmenge von* $\Pi(M)$.

Mit den Erzeugenden e_1, e_2, e_3, e_4 aus Satz 2.1 läßt sich nun auch leicht der folgende Satz im Fall $|M| \geqq 10$ beweisen.

SATZ 2.3. *Zu jedem endlichen Partitionenverband* $\Pi(M)$ *mit* $|M| \geqq 4$, *existieren Erzeugendenmengen* $P \subseteq \Pi(M)$, *die als halbgeordnete Mengen isomorph zu* $\mathbf{1+1+1+1}$ *sind.*

BEWEIS. Sei $M = \{1, 2, \ldots, n\}$, dann bilden für $|M| \geqq 10$ die Partitionen e_1', e_2, e_3, e_4 (e_2, e_3, e_4 wie im Beweis zu Satz 2.1), mit $e_1' := (1, 2, 3), (4), \ldots, (n)$ die gewünschte Erzeugendenmenge von $\Pi(M)$, denn offenbar ist $(\{e_1', e_2, e_3, e_4\}, \leqq) \cong \cong \mathbf{1+1+1+1}$ und es gilt $e_1' e_2 = e_1$, so daß der Erzeugungsprozeß wie in Satz 2.1 verlaufen kann.

Für $|M|=4, 5, 6, 7, 8$ und 9 werden am Schluß von Abschnitt 3 Erzeugendenmengen für $\Pi(M)$ angegeben.

KOROLLAR 2.4. *Die Erzeugendenmenge* $\{e_1', e_2, e_3, e_4\}$ *(wie im Beweis zu Satz 2.3 angegeben) ist minimale Erzeugendenmenge von* $\Pi(M)$

In Satz 2.1 ist $|M|\geqq 10$ Vorbedingung. Es gilt

LEMMA 2.5. *Ist M eine vierelementige Menge, so existieren keine Erzeugendenmengen P von* $\Pi(M)$, *welche als halbgeordnete Mengen isomorph zu* $\mathbf{1+1+2}$ *sind.*

Eine offene Frage ist, ob $\Pi(M)$ mit $|M|=5, 6, 7, 8$ oder 9 eine Erzeugendenmenge besitzt, die als halbgeordnete Menge isomorph zu $\mathbf{1+1+2}$ ist.

Ein wirksames Hilfsmittel bei der Entscheidung, ob eine vorliegende halbgeordnete Menge eine Erzeugendenmenge sein könnte oder nicht, ist das

D_2-LEMMA 2.6 (siehe WILLE [4], S. 456). *Sei L ein einfacher Verband, der von einer endlichen Menge E erzeugt wird, und sei* $E=E_0\cup E_1$. *Dann folgt: Ist L nicht isomorph zu* D_2, *so gilt* $\sup E_0\geqq\inf E_1$.

Soll nun eine halbgeordnete Menge P von Partitionen von M mit $|M|\geqq 3$ eine Erzeugendenmenge sein, so ist für jede Partition $P=P_0\cup P_1$ die Bedingung $\sup P_0\geqq\inf P_1$ notwendige Voraussetzung.

Im BEWEIS ZU LEMMA 2.5 werden alle halbgeordneten Mengen P von Partitionen von $M=\{1, 2, 3, 4\}$ betrachtet, die isomorph zu $\mathbf{1+1+2}$ sind sowie die notwendige Bedingung des D_2-Lemmas erfüllen:

$e_1^1=(1,2)$	$e_1^2=(1,2)$	$e_1^3=(1,2)$
$e_2^1=(1,2),(3,4)$	$e_2^2=(1,2),(3,4)$	$e_2^3=(1,2,3)$
$e_3^1=(1,3),(2,4)$	$e_3^2=(1,3),(2,4)$	$e_3^3=(1,4),(2,3)$
$e_4^1=(1,4),(2,3)$	$e_4^2=(2,3)$	$e_4^3=(2,4),(1,3).$

Dies sind bis auf Isomorphie alle möglichen halbgeordneten Mengen von Partitionen von M, welche isomorph zu $\mathbf{1+1+2}$ sind, und zugleich Lemma 2.6 genügen. Eine einfache Überprüfung zeigt, daß es keine Erzeugendenmengen von $\Pi(\{1, 2, 3, 4\})$ sind.

Im Hinblick auf die genannten (ungelösten) Sonderfälle sind folgende Sätze interessant:

SATZ 2.7. *Zu jedem endlichen Partitionenverband* $\Pi(M)$ *mit* $|M|\geqq 4$ *existieren Erzeugendenmengen P, die als halbgeordnete Mengen isomorph zu* $\mathbf{1+2+2}$ *sind.*

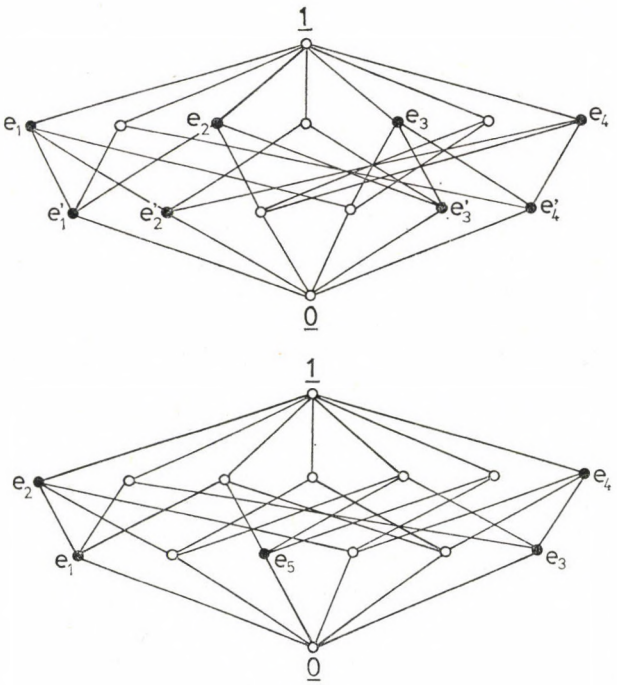
SATZ 2.8. *Zu jedem endlichen Partitionenverband* $\Pi(M)$ *mit* $|M|\geqq 5$ *existieren Erzeugendenmengen P, die als halbgeordnete Mengen isomorph zu* $\mathbf{1+1+3}$ *sind.*

BEWEIS VON SATZ 2.7 *und* 2.8. Ist $|M|\geqq 10$, so gewinnt man aus den in Satz 2.1 angegebenen Erzeugenden durch Hinzufügen geeigneter Partitionen auf einfache Weise entsprechende Erzeugendenmengen. Für die übrigen Fälle werden am Ende von Abschnitt 3 in einer Liste Erzeugendenmengen aufgeführt, die als halbgeordnete Mengen isomorph zu $\mathbf{1+2+2}$ bzw. zu $\mathbf{1+1+3}$ sind.

BEWEIS VON SATZ 1.1. Sei P kardinale Summe von Ketten und M eine endliche Menge mit $|M| \geqq 10$.

Werde $\Pi(M)$ von einem homomorphen Bild von P erzeugt. Da es eine abzählbar unendliche Anzahl endlicher Partitionenverbände gibt, die zudem einfach sind, ist Satz 1.2 anwendbar. Danach muß P eine isomorphe Kopie von $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$, $\mathbf{1}+\mathbf{1}+\mathbf{2}$ oder von $\mathbf{1}+\mathbf{K}_2$ enthalten. Da P kardinale Summe von Ketten ist, ist die Alternative $\mathbf{1}+\mathbf{K}_2$ nicht möglich. P enthält also wenigstens drei Ketten, und falls es genau drei Ketten sind, besteht mindestens eine von ihnen aus wenigstens zwei Elementen. P enthält demnach eine isomorphe Kople von $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder von $\mathbf{1}+\mathbf{1}+\mathbf{2}$.

Enthalte P eine isomorphe Kopie von $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder von $\mathbf{1}+\mathbf{1}+\mathbf{2}$. So gibt es einen Homomorphismus auf die in Satz 2.1 bzw. Satz 2.3 angegebenen Erzeugen-

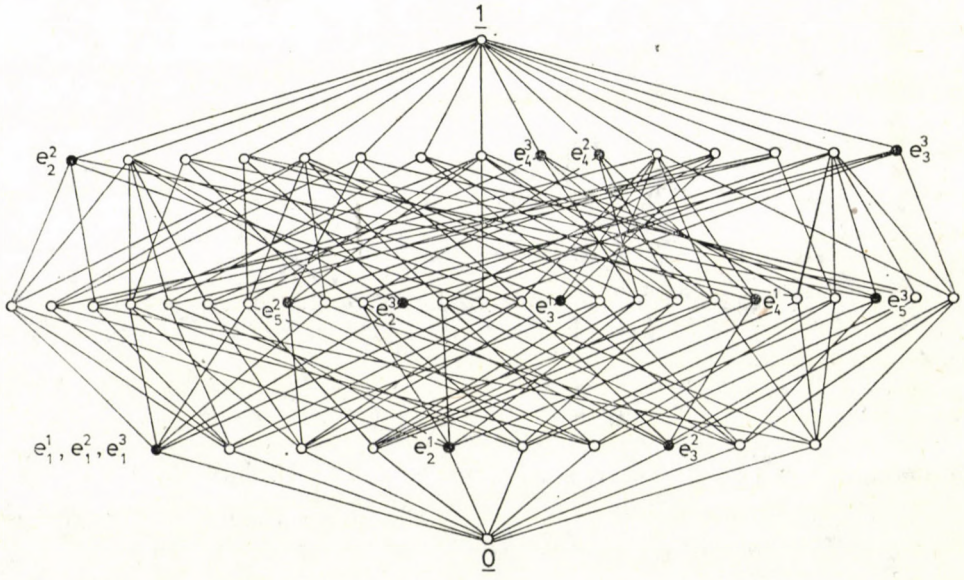


Figur 3.

denmengen von $\Pi(M)$ mit $|M| \geqq 10$. Das ist die zweite Beweisrichtung der Behauptung von Satz 1.1.

In Figur 3 und Figur 4 sind erzeugende Elemente in Partitionenverbände $\Pi(M)$ mit $|M|=3$ bzw. 4 eingezeichnet, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$ sind.

Figur 3: Erzeugendenmengen des Partitionenverbandes einer vierelementigen Menge. a) $\{e_1, e_2, e_3, e_4\}$ und $\{e_1', e_2', e_3', e_4'\}$ als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$, mit $\Delta(e_1, e_2, e_3, e_4)=2$ und $\Delta(e_1', e_2', e_3', e_4')=-2$. b) $\{e_1, e_2, e_3, e_4, e_5\}$ als halbgeordnete Menge isomorph zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$, mit $\Delta(e_1, e_2, e_3, e_4, e_5)=-\frac{1}{2}$. (Zur Definition des Defektes Δ einer Erzeugendenmenge siehe man Definition 3.3)



Figur 4.

Figur 4: Erzeugendenmengen des Partitionenverbandes einer fünfelementigen Menge. $(\{e_1^1, e_2^1, e_3^1, e_4^1\}, \leqq) \cong \mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ und $\Delta(e_1^1, \ldots, e_4^1)=-2$, $(\{e_1^2, e_2^2, e_3^2, e_4^2, e_5^2\}, \leqq) \cong \mathbf{1}+\mathbf{2}+\mathbf{2}$ und $\Delta(e_1^2, \ldots, e_5^2)=0$, sowie $(\{e_1^3, e_2^3, e_3^3, e_4^3, e_5^3\}, \leqq) \cong \mathbf{1}+\mathbf{1}+\mathbf{3}$ und $\Delta(e_1^3, \ldots, e_5^3)=1$.

3. Nichtisomorphe Erzeugendenmengen von $\Pi(M)$

Zu Beginn dieses Abschnittes werden halbgeordnete Mengen P von Partitionen von M, isomorph zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$, angegeben, die $\Pi(M)$ erzeugen, die aber zu den entsprechenden Erzeugendenmengen in Abschnitt 2 nicht isomorph sind. Zwei Erzeugendenmengen P und P' von $\Pi(M)$ heißen dabei isomorph, wenn P aus P' durch eine Permutation der Elemente von M entsteht. Die Konstruktionsmethode bestimmter ausgezeichneter Atome, mittels derer sich jedes weitere Element von $\Pi(M)$ erzeugen läßt, ist die gleiche wie in Satz 2.1.

Sei $M=\{1, 2, \ldots, n\}$ mit $n\geqq 12$ und $n\equiv 0 \pmod 4$, und seien ferner folgende Partitionen von M gewählt (vgl. STRIETZ [3], S. 257):

$$e_1 := (1,2,3,4), (5,6,7,8), \ldots, (k-3, k-2, k-1, k), \ldots, (n-3, n-2, n-1, n),$$

$$e_2 := (1,3), (2,4), (5,7), (6,8), \ldots, (k-3, k-1), (k-2, k), \ldots$$

$$\ldots, (n-3, n-1), (n-2, n),$$

$$e_3 := (1,2,5,6), (3,4,9,10), \ldots, (k-5, k-4, k+1, k+2), \ldots$$

$$\ldots, (n-5, n-4, n-1, n),$$

$$e_4 := (1,5), (2,6), (3,9), (2,10), \ldots, (k-5, k+1), (k-4, k+2), \ldots$$

$$\ldots, (n-5, n-1), (n-4, n),$$

$$e_5 := (1,4), (2,5), (3,7), (6,9), \ldots, (k-4, k-1), (k-2, k+1), \ldots$$

$$\ldots, (n-4), (n-2), (n-1), (n);$$

dabei ist $k\equiv 0 \pmod 4$ und k durchläuft die Werte $12, 16, \ldots, n-4$ (für $n=12$ existiert kein Wert für k). Diese Notation ist in gleicher Weise wie im Beweis von Satz 2.1 zu verstehen. Leicht ist festzustellen:

(a) $e_1, \ldots, e_5$ sind Partitionen von M,

(b) $\{e_1, \ldots, e_5\}$ als Teilmenge von $\Pi(M)$ genügt dem D_2-Lemma,

(c) $e_i e_j = \mathbf{0}$ für $i\neq j$ und $i, j\in\{2, 4, 5\}$.

Weiterhin sei $N\subseteqq M$, $N=\{1, 2, \ldots, m\}$ mit $n-4\leqq m\leqq n$. Dann gelten:

LEMMA 3.1. *Die Menge $\{e_1, e_2, e_3, e_4, e_5\}$ eingeschränkt auf N ist Erzeugendenmenge von $\Pi(N)$ und ist als halbgeordnete Menge isomorph zu* $\mathbf{1+2+2}$.

LEMMA 3.2. *Die Menge $\{e_1 e_3, e_2, e_4, e_5\}$ eingeschränkt auf N ist Erzeugendenmenge von $\Pi(N)$ und ist als halbgeordnete Menge isomorph zu* $\mathbf{1+1+1+1}$.

Zu bemerken ist, daß das „Konstruktionsprinzip" der Elemente dieser Erzeugendenmengen erheblich von dem der in Satz 2.1 verwendeten Erzeugenden abweicht. Will man hier eine konkrete Erzeugendenmenge für $\Pi(N)$ angeben, so ist, wenn $|N|=m$, die nächst größere Zahl $n\equiv 0 \pmod 4$ zu suchen (es sei denn $m\equiv 0 \pmod 4$), die Erzeugendenmenge gemäß obiger Vorschrift zu bilden, und schließlich auf N (unter Beibehaltung der Blöcke) einzuschränken. Eine weitere Einschränkung auf $N'\subset N$ mit $|N'|<n-4$ ergibt keine Erzeugendenmenge für $\Pi(N')$ mehr. Eine Erzeugendenmenge nach Satz 2.1 hingegen läßt sich beliebig einschränken, solange nur $|N'|\geqq 10$ gilt.

BEWEIS VON LEMMA 3.1. *und* 3.2. Es wird die gleiche Technik und Terminologie wie im Beweis zu Satz 2.1 verwendet. Zunächst ist:

$$e_1 e_3 = (1,2), (3,4), (5,6), (7,8), (9,10), \ldots, (k-1, k), (k+1, k+2), \ldots$$

$$\ldots, (n-1, n)$$

und es gilt:

(c′) $e_ie_j=\mathbf{0}$ für $i\neq j$ und $i, j\in\{2, 4, 5, 6\}$, wobei $e_6:=e_1e_3$. Aus der Feststellung $e_2<e_1$ und $e_4<e_3$, sowie (c′) folgen $(\{e_1, e_2, e_3, e_4, e_5\}, \leqq)\cong\mathbf{1}+\mathbf{2}+\mathbf{2}$ und $(\{e_2, e_4, e_5, e_6\}, \leqq)\cong\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$

(1) $e_6 = e_1e_3,$

(2) $t_{1,4} = e_1e_5 = (1, 4),$

(3) $t_{2,5} = e_3e_5 = (2, 5),$

(4) $q_1 = (t_{1,4}+e_2)e_6 = (1, 2),\ (3, 4),$

(5) $q_2 = (t_{2,5}+e_4)e_6 = (1, 2),\ (5, 6),$

(6) $t_{1,2} = q_1q_2 = (1, 2),$

(7) $t_{2,4} = (t_{1,4}+t_{1,2})e_2 = (2, 4),$

(8) $t_{1,5} = (t_{1,2}+t_{2,5})e_4 = (1, 5),$

(9) $t_{4,5} = (t_{1,4}+t_{1,5})(t_{2,4}+t_{2,5}) = (4, 5),$

(10) $t_{4,6} = (q_2+t_{4,5})(t_{2,4}+e_4) = (4, 6),$

(11) $t_{5,6} = (t_{4,5}+t_{4,6})q_2 = (5, 6),$

(12) $t_{1,6} = (t_{1,4}+t_{4,6})(t_{1,5}+t_{5,6}) = (1, 6),$

(13) $t_{2,6} = (t_{2,4}+t_{4,6})(t_{2,5}+t_{5,6}) = (2, 6),$

(14) $t_{3,7} = (t_{1,5}+e_2)e_5 = (3, 7),$

(15) $t_{4,7} = (q_1+t_{3,7})(t_{2,5}+e_2) = (4, 7),$

(16) $t_{1,3} = (t_{1,4}+t_{4,7}+t_{3,7})e_2 = (1, 3),$

(17) $t_{1,7} = (t_{1,3}+t_{3,7})(t_{1,4}+t_{4,7}) = (1, 7),$

(18) $t_{2,7} = (t_{1,2}+t_{1,7})(t_{2,4}+t_{4,7}) = (2, 7),$

(19) $t_{2,3} = (t_{2,7}+t_{3,7})(t_{1,2}+t_{1,3}) = (2, 3).$

Hier läßt sich festhalten: Ist $N'=\{1, 2, \ldots, 7\}$, so ist die angegebene Erzeugendenmenge eingeschränkt auf N' eine Erzeugendenmenge von $\Pi(N')$, denn Elemente von M, die größer als 7 sind, werden in den Schritten (1) bis (19) nicht benötigt. Das gleiche gilt für $N'\cup\{8\}$, $N'\cup\{8, 9\}$ und $N'\cup\{8, 9, 10\}$, denn

(20) $t_{1,8} = (t_{1,6}+e_2)(t_{1,7}+e_6) = (1, 8),$

(21) $t_{2,8} = (t_{2,6}+e_2)(t_{1,8}+t_{1,2}) = (2, 8),$

(22) $t_{1,9} = (t_{1,3}+e_4)(t_{1,6}+e_5) = (1, 9),$

(23) $t_{2,9} = (t_{1,2}+t_{1,9})(t_{2,3}+e_4) = (2, 9),$

(24) $t_{1,10} = (t_{1,4}+e_4)(t_{1,9}+e_6) = (1, 10),$

(25) $t_{2,10} = (t_{1,2}+t_{1,10})(t_{2,4}+e_4) = (2, 10).$

Es folgen nun Rekursionsformeln zur Erzeugung der Atome $t_{1,i}$ und $t_{2,j}$ für $11 \leqq i, j \leqq n$:

(26) $t_{1,k-1} = (t_{1,k-3}+e_2)(t_{1,k-4}+e_5)$,

(27) $t_{2,k-1} = (t_{2,k-3}+e_2)(t_{1,2}+t_{1,k-1})$,

(28) $t_{1,k} = (t_{1,k-2}+e_2)(t_{1,k-1}+e_6)$,

(29) $t_{2,k} = (t_{2,k-2}+e_2)(t_{1,2}+t_{1,k})$,

(30) $t_{1,k+1} = (t_{1,k-5}+e_4)(t_{1,k-2}+e_5)$,

(31) $t_{2,k+1} = (t_{2,k-5}+e_4)(t_{1,2}+t_{1,k+1})$,

(32) $t_{1,k+2} = (t_{1,k-4}+e_4)(t_{1,k+1}+e_6)$,

(33) $t_{2,k+2} = (t_{2,k-4}+e_4)(t_{1,2}+t_{1,k+2})$;

k durchläuft dabei die anfangs genannten Werte. Schließlich ist

(34) $t_{1,n-1} = (t_{1,n-3}+e_2)(t_{1,n-5}+e_4)$,

(35) $t_{2,n-1} = (t_{2,n-3}+e_2)(t_{1,2}+t_{1,n-1})$.

(36) $t_{1,n} = (t_{1,n-2}+e_2)(t_{1,n-4}+e_4)$,

(37) $t_{2,n} = (t_{2,n-2}+e_2)(t_{1,2}+t_{1,n})$.

Die Behauptung für die auf N eingeschränkte Erzeugendenmenge folgt mit gleicher Argumentation wie im Beweis zu Satz 2.1. Zum Beweis von Lemma 3.2 ist nur noch folgendes zu bemerken: Da $e_6 = e_1 e_3$ gilt, ist $e_1 = e_2 + e_6$ und $e_3 = e_4 + e_6$, das heißt, die Erzeugungsschritte (2) bis (37) verlaufen völlig analog.

Aus den bisherigen Überlegungen in Abschnitt 2 und 3 stellt sich die Frage nach einer Klassifizierung minimaler Erzeugendenmengen, ähnlich wie dies in der Arbeit von I. M. GELFAND und V. A. PONOMAREV in "Problems of linear algebra and classification of quadruples of subspaces in a finite dimensional vector space" durchgeführt wurde. Sie zeigten durch Einführung des Begriffes „Defekt eines Quadrupels von Untervektorräumen", daß nur Defekte mit den Werten $-2, -1, 0, 1$ und 2 auftreten können, was eine Klassifizierung der Erzeugendenquadrupel von Untervektorraumverbänden ermöglicht. In analoger Weise soll hier der Defekt einer Erzeugendenmenge von $\Pi(M)$ eingeführt werden. Zunächst einige Vorbemerkungen:

Unter dem *Rang* einer Partition p der n-elementigen Menge M versteht man die um eins verminderte Anzahl der Elemente einer maximalen Kette in $[\mathbf{0}, p] \subseteqq \subseteqq \Pi(M)$; dabei ist $[\mathbf{0}, p]$ das Intervall in $\Pi(M)$ mit $\mathbf{0}$ als kleinstem und p als größtem Element. Abkürzend wird der Rang von p mit $\mathrm{rg}(p)$ bezeichnet. Wie leicht zu beweisen ist, gilt:

$$(**) \qquad \mathrm{rg}(p) = n - c(p)$$

wobei $n = |M|$ und $c(p)$ die Anzahl der Blöcke von p ist.

DEFINITION 3.3. Der *Defekt* einer Elementefolge $(e_1, e_2, \ldots, e_k)$ von $\Pi(M)$ mit $|M| = n$ ist folgende rationale Zahl:

$$\Delta(e_1, e_2, \ldots, e_k) := \sum_{i=1}^{k} \mathrm{rg}(e_i) - \frac{k}{2}\,\mathrm{rg}(\mathbf{1})_{\Pi(M)} = \sum_{i=1}^{k} \mathrm{rg}(e_i) - \frac{k}{2}(n-1).$$

Ist $|\{e_1, \ldots, e_k\}|=k$, so ist $\Delta(e_1, \ldots, e_k)$ schon durch die Menge $\{e_1, \ldots, e_k\}$ festgelegt, weshalb dann auch im weiteren Verlauf vom Defekt der Menge $\{e_1, \ldots, e_k\}$ gesprochen werden soll.

Setzt man in die obige Beziehung die Formel $(**)$ ein, so ergibt sich der folgende gut zu verwendente Ausdruck:

$$(***) \qquad \Delta(e_1, e_2, \ldots, e_k) = \frac{k}{2}(n+1) - \sum_{i=1}^{k} c(e_i).$$

Sind die Defekte zweier Erzeugendenmengen von $\Pi(M)$ verschieden, so sind sie nicht isomorph, d. h. es existiert kein Automorphismus von $\Pi(M)$, der die eine Erzeugendenmenge in die andere überführt. Das bedeutet weiterhin, hat man Kenntnis von auftretenden Defekten minimaler Erzeugendenmengen eines Partitionenverbandes $\Pi(M)$, so erhält man eine untere Abschätzung der Anzahl nichtisomorpher minimaler Erzeugendenmengen von $\Pi(M)$. Zugleich ist aber der Defekt einer Erzeugendenmenge von $\Pi(M)$ auch ein „Maß" dafür, wie die Elemente der Erzeugendenmenge $\{e_1, e_2, \ldots, e_k\}$ „in der Mitte" von $\Pi(M)$ liegen, d.h. Defekt. Null bedeutet z. B., die Elemente der Erzeugendenmenge liegen „gemittelt in der Mitte von $\Pi(M)$". (Zu diesem Punkt vergleiche man Figur 3 und Figur 4.)

Im weiteren werden die Defekte von Erzeugendenmengen von $\Pi(M)$, die als halbgeordnete Mengen isomorph zu $\mathbf{1+1+2}$ oder zu $\mathbf{1+1+1+1}$ sind, untersucht. Dabei treten nur ganzzahlige Defekte auf.

Für die Erzeugendenmenge $\{e_1, e_2, e_3, e_4\}$ von $\Pi(M)$ mit $|M| \geqq 10$ aus Satz 2.1 wird zunächst der Defekt berechnet, dabei ist:

$$\mathrm{rg}(e_1)=1, \quad \text{für alle} \quad n \geqq 10$$

$$\mathrm{rg}(e_2) = \begin{cases} \frac{2}{3}n - \frac{2}{3}, & \text{für} \quad n \equiv 1 \pmod 3 \\ \frac{2}{3}n, & \text{für} \quad n \equiv 0 \pmod 3 \\ \frac{2}{3}n - \frac{1}{3}, & \text{für} \quad n \equiv 2 \pmod 3 \end{cases}$$

$$\mathrm{rg}(e_3) = \begin{cases} \frac{2}{3}n - \frac{2}{3}, & \text{für} \quad n \equiv 1 \pmod 3 \\ \frac{2}{3}n - 1, & \text{für} \quad n \equiv 0 \pmod 3 \\ \frac{2}{3}n - \frac{4}{3}, & \text{für} \quad n \equiv 2 \pmod 3 \end{cases}$$

$$\mathrm{rg}(e_4) = \begin{cases} \frac{2}{3}n - \frac{5}{3}, & \text{für} \quad n \equiv 1 \pmod 3 \\ \frac{2}{3}n - 2, & \text{für} \quad n \equiv 0 \pmod 3 \\ \frac{2}{3}n - \frac{4}{3}, & \text{für} \quad n \equiv 2 \pmod 3, \end{cases}$$

und damit ist

$$\Delta(e_1, e_2, e_3, e_4) = \sum_{i=1}^{4} \operatorname{rg}(e_i) - 2(n-1) = 0$$

für alle $n \geqq 10$.

Bisher wurden noch keine Erzeugendenmengen, welche als halbgeordnete Mengen isomorph zu $\mathbf{1+1+2}$ sind, gefunden, deren Defekt ungleich Null ist.

Ein wesentlich anderes Ergebnis erhält man aus der Untersuchung nichtisomorpher Erzeugendenmengen von $\Pi(M)$, die als halbgeordnete Mengen isomorph zu $\mathbf{1+1+1+1}$ sind.

SATZ 3.4. *$\Pi(M)$ sei der Partitionenverband der Menge M, und M habe $n \geqq 10$ Elemente. Dann existieren minimale Erzeugendenmengen P_i von $\Pi(M)$, die als halbgeordnete Mengen isomorph zu $\mathbf{1+1+1+1}$ sind, für die gilt:*

$$\Delta(P_i) = i \text{ für } i \in I = \{1, 2, \ldots, n-4\}.$$

BEWEIS. Für $\Pi(M)$ mit $|M| = n \geqq 10$, werden Erzeugendenmengen P_i angegeben, die als halbgeordnete Mengen isomorph zu $\mathbf{1+1+1+1}$ sind. Dabei wird $i \in I = \{1, 2, \ldots, n-4\}$ sein, und die Erzeugendenmengen werden so indiziert, daß gilt: $\Delta(P_i) = i$.

Sei $M = \{1, 2, \ldots, n\}$ mit $n \geqq 16$ und $n \equiv 1 \pmod 4$. Als Erzeugendenmengen von $\Pi(M)$ werden $P_i := \{e_1^i, e_2, e_3, e_4\}$ gewählt, wobei e_2, e_3, e_4 die im Beweis zu Satz 2.1 angegebenen Partitionen von M sind, und e_1^i wie folgt gewählt wird: $(e_1 = (1, 2), (3), \ldots, (n))$

$$e_1^1 := (1, 2, 3), \quad \text{also} \quad \operatorname{rg}(e_1^1) = 2 \quad \text{und} \quad e_1^1 e_2 = e_1,$$

$$e_1^2 := (1, 2, 3), (4, 6), \quad \text{also} \quad \operatorname{rg}(e_1^2) = 3 \quad \text{und} \quad e_1^2 e_2 = e_1,$$

$$e_1^3 := (1, 2, 3), (4, 6), (5, 7), \quad \text{also} \quad \operatorname{rg}(e_1^3) = 4 \quad \text{und} \quad e_1^3 e_2 = e_1,$$

$$\vdots$$

$$e_1^8 := (1, 2, 3, 8, 10), (4, 6, 9), (5, 7, 11, 12), \ldots, (k+3), (k+4), (k+5), \ldots, (n)$$

$$(\text{und es ist } \operatorname{rg}(e_1^8) = 9, \text{ sowie } e_1^8 e_2 = e_1),$$

$$\vdots$$

$$e_1^{n-4} := (1, 2, 3, 8, 10, \ldots, k+3, \ldots, n), (4, 6, 9, \ldots, k+4, \ldots), (5, 7, 11, 12, \ldots$$

$$\ldots, k+5, \ldots);$$

k durchläuft dabei die Werte $10, 13, \ldots, n-6$.

Es gilt also stets: (a) $\operatorname{rg}(e_1^i) = i+1$, (b) $e_1^i e_2 = e_1$ (in e_1^{i+1} hat genau ein Block genau ein Element von M mehr als in e_1^i, d. h., $\operatorname{rg}(e_1^{i+1}) = \operatorname{rg}(e_1^i) + 1$, außerdem sorgt man stets dafür, daß (b) gilt).

Wegen (b) ist $P_i := \{e_1^i, e_2, e_3, e_4\}$ Erzeugendenmenge von $\Pi(M)$ Diese P_i sind paarweise nicht isomorph. Weiterhin ergibt sich nun:

$$\Delta(P_i) = \Delta(e_1^i, e_2, e_3, e_4) = \sum_{j=1}^{4} \operatorname{rg}(e_j) - 2(n-1) =$$

$$= \sum_{j=2}^{4} \operatorname{rg}(e_j) + (i+1) - 2(n-1) = 2n - 3 + (i+1) - 2(n-1) = i.$$

Die Summe $\sum_{j=2}^{4} \mathrm{rg}\,(e_j)$ entnimmt man aus der Aufstellung zur Bestimmung des Defektes der Erzeugendenmenge in Satz 2.1.

Ist $N \subseteqq M$, $N=\{1, 2, \ldots, m\}$, $m \geqq 10$, so ist leicht zu prüfen, daß mit $e'_j = e_j|_N$ die gleiche Aussage für $\{e'_1, e'_2, e'_3, e'_4\}$ gilt.

KOROLLAR 3.5. *Die Menge der Defekte minimaler Erzeugendenmengen endlicher Partitionenverbände ist (nach oben) nicht beschränkt.*

Zum Abschluß dieses dritten Abschnittes wird die angekündigte Liste von Erzeugendenmengen von $\Pi(M)$ für $|M|=4, 5, 6, 7, 8$ und 9 angegeben:

$|M|=4$ (vgl. Figur 3):

$e_1^1 = (1, 2, 3)$ $\quad e_1^2 = (1, 2)$ $\quad \{e_1^2, \ldots, e_5^2\}$ ist minimale

$e_2^1 = (1, 2, 4)$ $\quad e_2^2 = (1, 2, 3)$ Erzeugendenmenge.

$e_3^1 = (1, 3, 4)$ $\quad e_3^2 = (3, 4)$

$e_4^1 = (2, 3, 4)$ $\quad e_4^2 = (2, 3, 4)$ $\quad \Delta(e_1^2, \ldots, e_5^2) = -1/2$

$\Delta(e_1^1, \ldots, e_4^1) = 2$ $\quad e_5^2 = (1, 4)$

$|M|=5$ (vgl. Figur 4):

$e_1^1 = (1, 2), (3, 4, 5)$, $\quad e_2^1 = (1, 3, 5), (2, 4)$, $\quad e_3^1 = (1, 5), (2, 3)$

$e_4^1 = (1, 4, (2, 5)$ mit $\Delta(e_1^1, \ldots, e_4^1) = 2$

$e_1^2 = (1, 2)$	$e_1^3 = (1, 2)$
$e_2^2 = (1, 2, 3), (4, 5)$	$e_2^3 = (1, 2), (3, 4)$
$e_3^2 = (3, 4)$	$e_3^3 = (1, 2), (3, 4, 5)$
$e_4^2 = (2, 3, 4), (1, 5)$	$e_4^3 = (2, 3), (1, 4, 5)$
$e_5^2 = (1, 4), (2, 5)$	$e_5^3 = (2, 4), (3, 5)$
$\Delta(e_1^2, \ldots, e_5^2) = 0$	$\Delta(e_1^3, \ldots, e_5^3) = 1$

$|M| = 6$

$e_1^1 = (1, 2), (3, 4), (5, 6)$, $\quad e_2^1 = (1, 6), (2, 3), (4, 5)$, $\quad e_3^1 = (1, 5), (2, 4)$,

$e_4^1 = (1, 4), (3, 6)$, $\quad$ mit $\quad \Delta(e_1^1, \ldots, e_4^1) = 0$

$e_1^2 = (1, 2)$	$e_1^3 = (1, 2)$
$e_2^2 = (1, 2, 3), (4, 5, 6)$	$e_2^3 = (1, 2), (3, 4), (5, 6)$
$e_3^2 = (3, 4)$	$e_3^3 = (1, 2, 3, 4), (5, 6)$
$e_4^2 = (2, 3, 4), (1, 5, 6)$	$e_4^3 = (1, 5), (2, 4), (3, 6)$
$e_5^2 = (1, 4)\ (2, 5), (3, 6)$	$e_5^3 = (1, 3), (2, 6), (4, 5)$
$\Delta(e_1^2, \ldots, e_5^2) = \frac{1}{2}$	$\Delta(e_1^3, \ldots, e_5^3) = \frac{3}{2}$

$|M|=7$ (für Erzeugendenmengen, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$ sind, siehe Lemma 3.1)

$e_1 = (1, 2)$

$e_2 = (1, 2), (3, 4), (5, 6, 7)$

$e_3 = (1, 2, 3, 4), (5, 6, 7)$

$e_4 = (1, 5), (2, 4, 7), (3, 6)$

$e_5 = (1, 3), (2, 6), (4, 5) \quad \Delta(e_1, \ldots, e_5) = 2$

$|M|=8$ (für Erzeugendenmengen, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$ sind, siehe Lemma 3.1)

$e_1 = (1, 2)$

$e_2 = (1, 2), (3, 4), (5, 6, 7)$

$e_3 = (1, 2, 3, 4), (5, 6, 7)$

$e_4 = (1, 5), (2, 3, 6), (4, 8)$

$e_5 = (1, 4), (2, 7), (3, 5), (6, 8) \quad \Delta(e_1, \ldots, e_5) = \frac{1}{2}$

$|M|=9$ (für Erzeugendenmengen, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$ oder zu $\mathbf{1}+\mathbf{2}+\mathbf{2}$ sind, siehe Lemma 3.1)

$e_1 = (1, 2)$

$e_2 = (1, 2), (3, 4), (5, 6, 7), (8, 9)$

$e_3 = (1, 2, 3, 4), (5, 6, 7), (8, 9)$

$e_4 = (1, 5, 9), (2, 3, 6), (4, 8)$

$e_5 = (1, 4), (2, 7), (3, 5), (6, 8) \quad \Delta(e_1, \ldots, e_5) = 1$

4. Schlußbemerkungen

Satz 1.1 liefert Aussagen darüber, wann eine kardinale Summe von Ketten homomorphe Bilder besitzt, die Erzeugendenmengen endlicher Partitionenverbände sind. Im Hinblick auf Satz 1.2 ist hier eine interessante Frage offen: „Existieren zu jedem endlichen Partitionenverband $\Pi(M)$ mit $|M| \geqq n_0$ Erzeugendenmengen P, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{K}_2$ sind?" Eine positive Beantwortung dieser Frage ließe dann folgenden Satz zu:

Sei P eine halbgeordnete Menge, und M eine endliche Menge mit $|M| \geqq n_0$. Dann gilt:

$\Pi(M)$ wird genau dann von einem homomorphen Bild von P erzeugt, wenn P eine isomorphe Kopie von $\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$, $\mathbf{1}+\mathbf{1}+\mathbf{2}$ oder $\mathbf{1}+\mathbf{K}_2$ enthält.

Im Zusammenhang mit der Frage der Klassifizierung minimaler Erzeugendenmengen durch den Begriff des Defektes wurde gezeigt, die Menge der Defekte

minimaler Erzeugendenmengen, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+$ $+\mathbf{1}+\mathbf{1}$ sind, ist nach oben nicht beschränkt. Offen ist hier das Problem einer unteren Schranke dieser Menge. Bisher liegen dazu nur Beispiele für $\Delta(P)=-2$ als kleinstem Defekt vor:

Ist $M=\{1, 2, 3, 4\}$, so bilden folgende Partitionen eine Erzeugendenmenge (vgl. Figur 3) $e_1=(1, 2)$, $e_2=(1, 3)$, $e_3=(3, 4)$, $e_4=(2, 4)$. Sofort überprüft man, daß gilt: $\langle e_1, \ldots, e_4\rangle=\Pi(M)$, $\Delta(e_1, \ldots, e_4)=-2$ und $(\{e_1, \ldots, e_4\}, \leqq)\cong\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}$.

Für $M=\{1, 2, 3, 4, 5\}$ bilden die folgenden Partitionen eine Erzeugendenmenge mit den gleichen Eigenschaften wie die vorher angegebene Erzeugendenmenge für $\Pi(\{1, 2, 3, 4\})$ (vgl. Figur 4): $e_1=(1, 2)$, $e_2=(2, 3)$, $e_3=(1, 4)$, $(3, 5)$ und $e_4=(1, 5)$, $(3, 4)$.

Weiterhin wurde schon darauf hingewiesen, daß bisher nur Erzeugendenmengen endlicher Partitionenverbände, die als halbgeordnete Mengen isomorph zu $\mathbf{1}+\mathbf{1}+\mathbf{2}$ sind, mit Defekt Null bekannt sind. Sollte sich bestätigen, daß für endliche Partitionenverbände zu $\mathbf{1}+\mathbf{1}+\mathbf{2}$ isomorphe Erzeugendenmengen nur vom Defekt Null möglich sind, könnte man hoffen, eine vollständige Klassifikation dieser Erzeugendenmengen zu bekommen.

LITERATURVERZEICHNIS

[1] BIRKHOFF, G.: *Lattice Theory*, Amer. Math. Soc. Colloquium Publications, vol. 25, third edition (1967).

[2] ORE, O.: Theory of Equivalence Relations, *Duke Math. Journal*, **9** (1942), 573—627.

[3] STRIETZ, H.: Finite Partition Lattices are Four-Generated, *Proceedings of the Lattice Theory Conference* (Ulm, 1976), 257—259.

[4] WILLE, R.: A Note on Simple Lattices, *Colloquia Mathematica Societatis János Bolyai* (Budapest, 1976), János Bolyai Mathematical Society and North-Holland Publishing Company, Amsterdam—Oxford—New York, 455—462.

Technische Hochschule, Fachbereich Algebra, 61—Darmstadt, Kant Platz 1, Bundesrepublik Deutschland

(Eingegangen am 27. April 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 19—27

ON SUPER-UNIVERSAL GRAPHS

by
J. PACH

1. Introduction

Let G^n be a non-directed graph with the vertex set $V(G^n)=\{P_0, P_1, \ldots, P_{n-1}\}$. Let us denote the number of edges of G^n by $e(G^n)$. The set $S(P_i)$ of the points adjacent to P_i is called the star of P_i. It is quite obvious to investigate graphs where the position of the stars is regular in a certain sense. D. J. KLEITMAN and J. SPENCER [1] have initiated the following definition.

DEFINITION. Let S be an n element set. If we have subsets $A_1, A_2, \ldots, A_k \subseteq S$, then they will be called k-independent if all the 2^k intersections

$$\bigcap_{i=1}^{k} B_i \quad (B_i = A_i \quad \text{or} \quad \bar{A}_i) \tag{1}$$

are non-empty. A collection of subsets is k-independent if every k of its members are k-independent.

Applying the above definition to the stars of a graph we get the following definition, first proposed by S. H. HECHLER [2].

DEFINITION. The graph G^n is *k-super-universal* if it has the property that for any k element subset A of $V(G_n)$ and any subset B of A there exists a point (not in A) joined to all points of B but to no point of $A \setminus B$.

In other words G^n is k-super-universal if for any k element subset A of $V(G^n)$

$$S(P_i) \setminus A \quad (P_i \in A)$$

are k-independent sets.

The question considered in this paper was raised by P. Erdős: At least how many edges must a k-super-universal graph with n vertices have? (An other question would be that at most how many edges can a k-super-universal graph have, but this problem is equivalent to the above one, because the complement of a k-super-universal graph has the same property.)

The case $k \geqq 3$ is essentially settled as follows: It is not difficult to find a k-super-universal graph having n vertices and no more than $c_k n \log_2 n$ edges. Let n points be divided into two parts, the first one having $2c_k \log_2 n$ points. Let the points of the second part be independent, and let us draw the other edges at random with probability 1/2. We can choose a k element subset A of the vertex set and a subset B of A in $\binom{n}{2} 2^k$ different ways. In any case the probability that there does not exist a point in the first part, which is joined to all points of B but to no point

of $A \setminus B$ is at most $(1-2^{-k})^{2c_k \log_2 n - k}$. Thus the probability that the obtained graph is not k-super-universal is not greater than

$$(2) \qquad \binom{n}{k} 2^k (1-2^{-k})^{2c_k \log_2 n - k} \leqq \frac{n^k}{k!} 2^k e^{-2^{-k}(2c_k \log_2 n - k)} \leqq \frac{e^{k2^{-k}} 2^k}{k!} n^{k - c_k 2^{1-k} \log_2 e}.$$

This quantity is arbitrary small (for a fixed k) if n is sufficiently large and $c_k \geqq 2k^{k-1}$. Since the probability that the obtained graph has no more than $c_k n \log_2 n$ edges is greater than 1/2, there exists a k-super-universal graph with n vertices and at most $k\, 2^{k-1} n \log_2 n$ edges $(n > n_0(k))$.

We denote the minimal number of edges of a k-super-universal graph of n vertices by $f(n, k)$. We obtained

$$(3) \qquad f(n, k) < k 2^{k-1} n \log_2 n \quad (n > n_0(k)).$$

On the other hand N. Sauer and M. Simonovits have proved (oral communication) that for $k \geqq 3$, this result is the best possible apart from the exact value of the constant. More precisely

$$(4) \qquad f(n, k) \geqq \frac{1}{2} n \log_2 (n-1) > \frac{1}{4} n \log_2 n \quad (k \geqq 3).$$

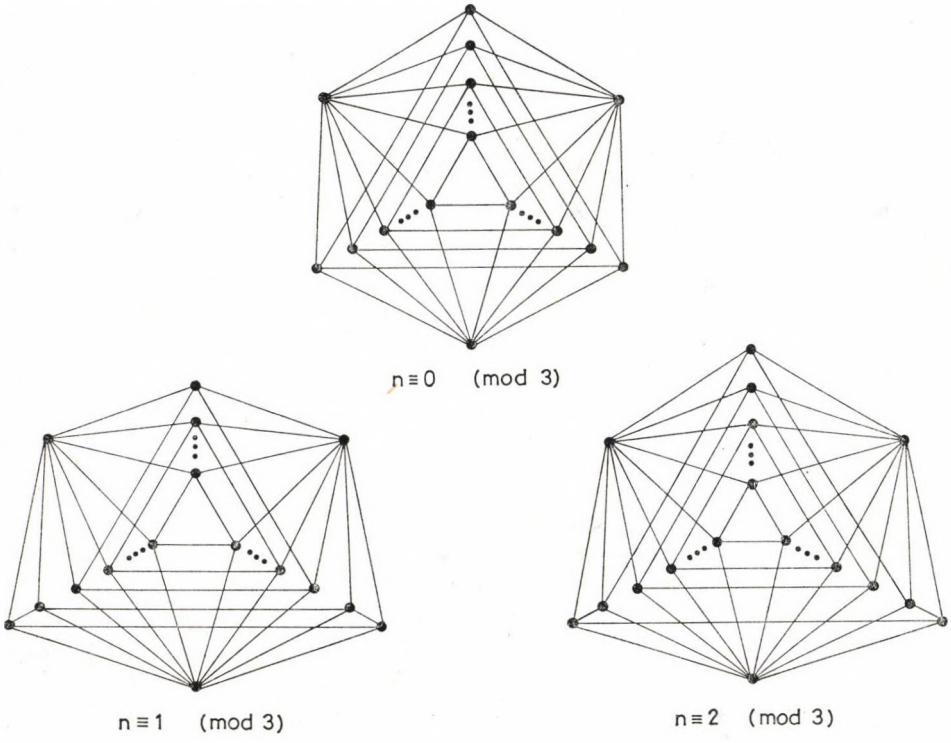


Fig. 1.

The proof of (4) is surprisingly simple: Let the graph G^n be k-super-universal. Let us consider a point P_0 of minimum degree in this graph. The neighbours of P_0 form the smallest star $S(P_0)$ of G^n. Evidently, $S(P_0)$ intersects the remaining $n-1$ stars of the graph, and all intersections are different. (At this point we have used that $k \geqq 3$.) Thus

$$|2^{S(P_0)}| \geqq n-1,$$

so the size of $S(P_0)$ (i.e. the minimum degree) is at least $\log_2(n-1)$ and (4) holds. In the second section of this paper we shall use this idea and the result of [1] to give a better lower bound for the function $f(n, k)$ in case $k \geqq 3$.

However, the case $k=2$ is essentially different from the above ones. In fact, for $n \geqq 12$, Erdős showed a 2-super-universal graph having n vertices and $3n-9$ edges if $n \equiv 0 \pmod 3$. By a slight modification of this construction we obtain a 2-super-universal graph of n vertices and $3n-9$ edges in case $n \equiv 1$ or $2 \pmod 3$. (See figure 1.) We thus have

$$f(n, 2) \leqq 3n-9 \quad (\text{if } n \geqq 12). \tag{5}$$

In the third section we shall substantially solve the problem in case $k=2$ proving the main result of this paper: $f(n, 2)=3n+O(1)$. In the same section we construct some 2-super-universal graphs of another type with $3n-9$ edges.

In the last section we shall discuss the adequate problem for hypergraphs.

2. A lower bound in case $k \geqq 3$

KLEITMAN AND SPENCER in their above-mentioned paper [1] raised the following problem: How large can a k-independent collection of subsets of an n element set be? Let us denote this maximal size by $g(n, k)$. They also proved that for a fixed k and $n > n_0(k)$

$$2^{d_1 n 2^{-k} k^{-1}} \leqq g(n, k) \leqq 2^{d_2 n 2^{-k}} \tag{6}$$

where d_1 and d_2 are absolute constants. Let us mention that there is a strong connection between this problem and our original one. If we consider the above probabilistic construction of a k-super-universal graph, we can find that those parts of the stars which belong to the first group of points, form a k-independent collection of subsets of the points of this group. Since there are $2c_k \log_2 n$ points in the first group and we have n stars, by $c_k = k2^k$ we get

$$n \leqq g(k2^{k+1} \log_2 n, k).$$

Clearly, this is equivalent to the left hand side of (6), with $d_1 = 1/2$.

Returning to our original problem, to give a lower bound for $f(n, k)$, we shall need the upper bound of (6). To make this paper self-contained we prove this statement in another brief way, in the form we need it:

$$g(n, k) \leqq 2^{n 2^{2-k}} + k. \tag{7}$$

The proof can be found in the Appendix.

If we have a k-super-universal graph G^n, where the smallest star (denoted by $S(P_1)$) contains r points, by (7) we get

$$n-1 \leqq g(r, k-1) \leqq 2^{r2^{3-k}}+k-1, \tag{8}$$

because there are $n-1$ intersection-sets of the smallest star with the remaining stars, and they form a $(k-1)$ independent collection of the subsets of $S(P_1)$. By (8) we obtain $r>2^{k-4}\log_2 n$ (if $n>n_0(k)$) and so $e(G^n)>2^{k-5}n\log_2 n$ holds. Thus we have the following.

THEOREM 1. *If* $k\geqq 3$ *and* $n\geqq n_0(k)$,

$$f(n, k) > 2^{k-5}n\log_2 n, \tag{9}$$

where $f(n, k)$, *defined above, denotes the minimum number of edges of a k-super-universal graph.* (Compare with (3) and (4).)

3. The case $k=2$

LEMMA 1. *The degree of any vertex of a 2-super-universal graph is at least* 4.

PROOF. Let P_0 be a fixed point. There exists a point P_1 adjacent to P_0. Applying the condition to the pair (P_0, P_1) we get a point P_2 joined to both P_0 and P_1, and a point P_3 joined to P_0 but not to P_1. In case there is an edge between P_2 and P_3, the condition applied to (P_0, P_2) guarantees a point P_4 adjacent to P_0 but not to P_2. (Consequently $P_4\neq P_1$, $P_4\neq P_3$.) If P_2 and P_3 are not connected by an edge, then P_0 and P_3 have a common neighbour (P_4). As P_1 and P_2 are not adjacent to P_3, $P_4\neq P_1$ and $P_4\neq P_2$. Thus P_0 is adjacent to at least 4 vertices (P_1, P_2, P_3, P_4) in both cases.

LEMMA 2. *Let* G^n *be a 2-super-universal graph and* $P_1\in V(G^n)$ *a vertex of degree* $n-a$. *Then* G^n *has at least* $3n-3a+\left[\frac{a}{2}\right]$ *edges.*

PROOF. Let $T(P_1)$ denote the set of vertices not adjacent to P_1. Thus $T(P_1)=$ $=V(G^n)\setminus S(P_1)\setminus P_1$, where $S(P_1)$, defined above, is the star of P_1. Any point P_2 of $S(P_1)$ is joined to at least one point of $T(P_1)$. (We have applied the condition to the pair (P_1, P_2).) Further, if P_3 is a point of $T(P_1)$, applying the condition to (P_1, P_3), we obtain a neighbour of P_3 in $T(P_1)$. Therefore and by Lemma 1 we get the following lower bound for the sum of the degrees of the points

$$2e(G^n) \geqq 1(n-a)+(n-a)4+\{(n-a)+(a-1)\},$$

which completes the proof.

By the help of the previous two simple lemmas we are able to prove the following statement, promised at the end of the first section.

THEOREM 2. *Let G^n be a 2-super-universal graph having n vertices. Then G^n has at least $3n-30$ edges. In other words*

$$f(n, 2) \geqq 3n-30. \tag{10}$$

(Compare with (5).)

PROOF. If each vertex of G^n has degree greater than 5, the graph has at least $3n$ edges. If each degree is at least 5, the number of edges is at least $3n-13$. We prove this assertion as follows: Let P_0 be a vertex of degree 5 (its neighbours are denoted by $P_1, \ldots, P_5$). Then each of the remaining $n-6$ vertices (i.e. each point of $T(P_0)$) is joined to at least one of P_i's ($i=1, \ldots, 5$), because, by the condition of the 2-super-universality, P_0 and these points have common neighbours. Therefore we get the following lower bound for the sum of the degrees of the vertices:

$$5+\{5+(n-6)\}+(n-6)5 = 6n-26.$$

The number of edges in G^n is at least the half of this value, $3n-13$. (We shall see that if every degree is at least 5, then the number of edges, in fact, is at least $3, 5n+O(1)$.)

Now, without loss of generality, we can assume that G^n has a vertex P_0 of degree 4. Let us denote the vertices adjacent to P_0 by $P_1, \ldots, P_4$. In the same way as above, we can say that each of the remaining $n-5$ points $(T(P_0))$ is joined to at least one of $P_1, \ldots, P_4$. Let us denote the subset of vertices adjacent only to P_i (among $P_1, \ldots, P_4$) by $I(P_i)$, and the set of points of $T(P_0)$ not adjacent to P_i by $I(\bar{P}_i)$. (Here $i=1, \ldots, 4$). By this definition we have (for example)

$$I(\bar{P}_1) \supseteq I(P_2) \cup I(P_3) \cup I(P_4). \tag{11}$$

Evidently, we can assume that at least one $I(P_i)$ contains a vertex of degree 4. Otherwise each of the $n-5$ points of $T(P_0)$ is either joined to at least 2 vertices of $P_1, \ldots, P_4$, or has degree at least 5. Hence for the number of edges we obtain

$$e(G^n) \geqq \frac{4+\left\{4+2(n-5)-\sum_{i=1}^{4}|I(P_i)|\right\}+\left\{4(n-5)+\sum_{i=1}^{4}|I(P_i)|\right\}}{2} = 3n-11. \tag{12}$$

Now let us suppose that $I(P_1)$ has a point of degree 4, and denote by $I'(P_1)$ the set of points of $I(P_1)$ of degree 4. We call a subset J of $I'(P_1)$ strongly independent, if it has the following two properties:

a) No two members of J are connected by an edge (i.e. the vertices of J are independent in G^n.)

b) No two members of J are connected by a path of length 2, avoiding P_1. (In other words no two members of J have a common neighbour, different from P_1.)

Let J be a maximal strongly independent set in $I'(P_1)$, and $I(J)$ denote the set of vertices different from P_1 and joined to some vertex of J. If $|J|=k$ then trivially $|I(J)|=3k$. (See figure 2.)

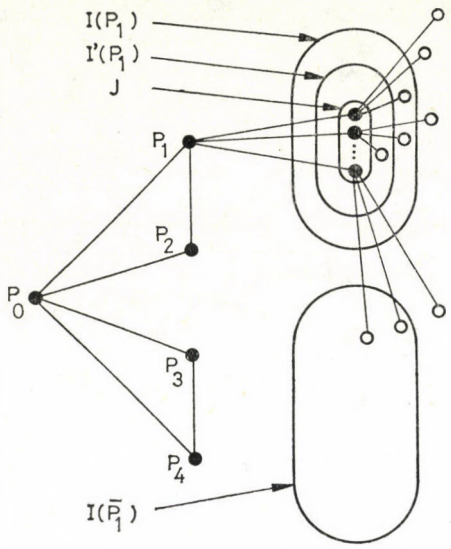


Fig. 2.

Now we are going to estimate the sum of the degrees of the points of $I(J)$. Each point of $I(\bar{P}_1)$ is adjacent to at least k points of $I(J)$. This is obvious, since a fixed point of $I(\bar{P}_1)$ and any point of J have a common neighbour, which is different from P_1, and these neighbours are different for distinct elements of J (by condition b of the strong independence). Further, by the maximality of J, each vertex belonging to $I'(P_1) \setminus J \setminus I(J)$ is joined to at least one vertex of $I(J)$. Finally each vertex of $I(J)$ is joined to one in J and to at least one of $P_1, \ldots, P_4$. Let us denote the number of elements of $I'(P_1) \cap I(J)$ by k_1. Thus we obtain the following lower bound for the sum of the degrees of the vertices belonging to $I(J)$:

(13)
$$A = k\,|I(\bar{P}_1)| + (|I'(P_1)| - k - k_1) + 3k + 3k.$$

We know that the number of vertices of $I(P_1) \setminus I'(P_1)$ not contained in $I(J)$ is at least

(14)
$$B = |I(P_1)| - |I'(P_1)| - 3k + k_1$$

and these vertices are of degree at least 5.

We have already shown (see (12)) that the sum of the degrees of the points $P_1, \ldots, P_4$ is at least

(15)
$$C = 4 + 2(n-5) - \sum_{i=1}^{4} |I(P_i)|.$$

Estimating by 4 the degree of the remaining points (not belonging to $\{P_0, \ldots, P_4\}$, $I(J)$ or $I(P_1) \setminus I'(P_1)$) we obtain the following lower bound for the number of edges of G^n.

$$e(G^n) \geqq \frac{4 + A + B + C + 4(n-5-3k)}{2}.$$

By (11), (13), (14), (15) we conclude

(16)
$$e(G^n) \geqq 3n - 16 + \frac{1}{2}(k-1)(|I(\bar{P}_1)| - 10).$$

Since $k \geqq 1$, in case $|I(\bar{P}_1)| \geqq 10$ the graph has at least $3n-21$ edges. On the other hand, if $|I(\bar{P}_1)| \leqq 9$ then P_1 is of degree at least $n-12$ and Lemma 2 implies $e(G^n) \geqq$ $\geqq 3n-30$. This completes the proof of Theorem 2.

We remark that we are unable to determide the smallest value c for which

$$f(n, 2) \geqq 3n - c.$$

Even if $f(n, 2)=3n-9$ (as Erdős conjectured) the graph shown by figure 1 is not the only extremal one. For $n\equiv 1$ or 3 (mod 4) we can construct a 2-super-universal graph with n vertices and $3n-9$ edges, which is essentially different from the graph of figure 1. (See figure 3 and figure 4. On figure 3 P_0 is joined to any point except $P_1, \ldots, P_6$ and P_i is adjacent to all points of $I(P_i)$.)

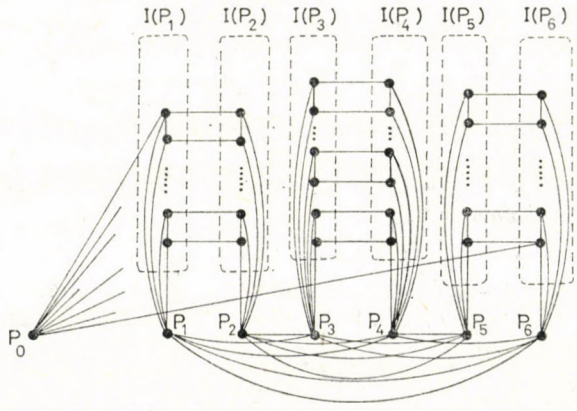


Fig. 3.

We mention that a planar graph with n vertices can have at most $3n-6$ edges. Theorem 2 states that a 2-super-universal graph must have at least $3n-30$ edges. Therefore and because both planarity and 2-super-universality are quite strong conditions, it is not surprising that a 2-super-universal graph (of sufficiantly large number of vertices) cannot be planar. This statement was proven by G. B. PURDY in [3].

4. Edge-universal hypergraphs

We shall formulate a problem for hypergraphs, which is very similar to the problems above for graphs.

By a hypergraph we mean a pair $(V, \mathcal{H})$, where V is a finite set and $\mathcal{H}$ is a family of different subsets of V. The elements of V are called points or vertices; the elements of $\mathcal{H}$ are the hyperedges or, in short, edges. $\varnothing$ will denote the empty set.

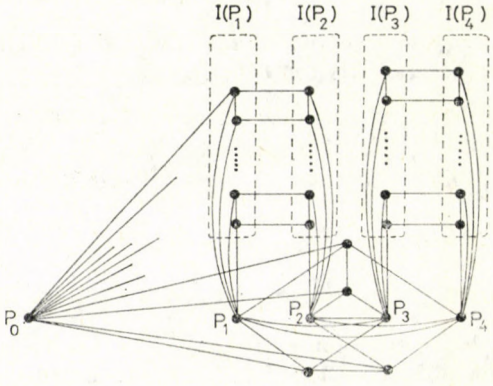


Fig. 4.

DEFINITION. A hypergraph $(V, \mathcal{H})$ is *k-edge-universal* if it has the following property: for any k element subset A of V and any subset B of A there exists an edge $H\in\mathcal{H}$, such that $B\subset H$ but $(A\setminus B)\cap H=\varnothing$.

(Compare this definition to the definition of the k-super-universal graph. At this point we note that the k-super-universal graphs are often called *k-point-universal,* which fact makes our term reasonable.)

Now the following question can be asked: At least how many edges must a k-edge-universal hypergraph of n vertices have? Let us denote this minimum number by $h(n, k)$. We can trace back this problem to another one, mentioned earlier.

THEOREM 3. *For $h(n, k)$, i.e. the minimal number of edges of a k-edge-universal graph*

$$h(n, k) = \min \{x | g(x, k) \geqq n\} \tag{17}$$

where $g(x, k)$, defined in section 2., is the maximal size of a k-independent collection of subsets of an x element set.

Let $(V, \mathscr{H})$ be a hypergraph of n vertices and m edges. The incidence matrix of $(V, \mathscr{H})$ is a matrix $A=(a_{ij})$ with m rows and n columns that represent the edges $H_i \in \mathscr{H}$ $(1 \leqq i \leqq m)$ and vertices $P_j \in V$ $(1 \leqq j \leqq n)$ respectively, such that

$$a_{ij} = \begin{cases} 0 & \text{if} \quad P_j \notin H_i, \\ 1 & \text{if} \quad P_j \in H_i. \end{cases} \tag{18}$$

The dual hypergraph of $(V, \mathscr{H})$, denoted by $(V^*, \mathscr{H}^*)$, is a hypergraph of m vertices and n edges, with the incidence matrix A^*, where A^* is the adjoint matrix of A. A hypergraph is called k-independent if its edges form a k-independent collection.

Now Theorem 3 is an immediate consequence of the following observation.

LEMMA 3. *A hypergraph $(V, \mathscr{H})$ is k-edge-universal if and only if its dual hypergraph $(V^*, \mathscr{H}^*)$ is k-independent.*

Using (6) with $d_1 = 1/2$ and (7), Theorem 3 gives the following lower and upper bounds for $h(n, k)$.

THEOREM 4. *For the minimum number of edges of a k-edge-universal hypergraph*

$$2^{k-3} \log_2 n \leqq h(n, k) \leqq k 2^{k+1} \log_2 n \tag{19}$$

holds.

Appendix

PROOF OF (7). If $k=2$ then (7) holds trivially. Let us suppose that our statement is true for $2, 3, \ldots, k-1$. Let $\{A, B, C, \ldots\}$ be a k-independent collection of $g(n, k)$ subsets of an n element set. We can assume that there is a member in this collection (say A), which has at most $\left[\frac{n}{2}\right]$ elements. (If we replace a subset in our collection with its complement, we get a k-independent collection again.) Since this fixed subset A intersects the remaining $g(n, k)-1$ members of the collec-

tion in different sets, and the intersection sets form a $(k-1)$-independent collection, we can state

$$g\left(\left[\frac{n}{2}\right], k-1\right) \geqq g(|A|, k-1) \geqq g(n, k)-1. \tag{20}$$

Thus

$$g(n, k) \leqq \left(2^{\frac{n}{2}2^{2-(k-1)}}+k-1\right)+1 = 2^{n2^{k-2}}+k$$

which completes the proof of (7).

Acknowledgement. The author wishes to thank M. Simonovits for his helpful suggestions and comments.

REFERENCES

[1] KLEITMAN, D. J. and SPENCER, J.: Families of k-independent sets, *Discrete Mathematics* **6** (1973), 255—262.

[2] HECHLER, S. H.: Large super-universal metric spaces, *Israel Journal of Mathematics* **14**, 2 (1973), 115—148.

[3] PURDY, G. B.: Planarity of 2-super-universal graphs, *Colloquia Mathematica Societatis János Bolyai 10. Infinite and finite sets*, Keszthely, Hungary (1973), Volume III. 1149—1157.

[4] ERDŐS, P., HECHLER, S. H. and KAINEN, P.: On finite super-universal graphs, *Discrete Mathematics* **24** (1978), 235—249.

Mathematical Institute of the Hungarian Academy of Sciences,
H—1053 Budapest, Reáltanoda u. 13—15, Hungary

(Received April 29, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 29—36

A GENERAL DECOMPOSITION THEOREM FOR ARTINIAN RINGS

by

A. WIDIGER

1. Introduction

We call a ring A artinian if it is right artinian (i.e. the minimum condition on right ideals of A is satisfied). A left and right artinian ring is called a two-sided artinian ring. A ring A is a strong artinian ring (short: SA-ring) if the underlying additive group $(A, +)$ of A satisfies the minimum condition on subgroups. Up to finite rings the structure of SA-rings was described in [7]. Call an ideal I of a ring A an SA-ideal if I is an SA-ring. One easily verifies that an artinian ring A contains a uniquely determined maximal SA-ideal $S(A)$ (cf. [2]). Recently DINH VAN HUYNH [2] has shown that a two-sided artinian ring A has a ring direct decomposition

$$A = B \boxplus S(A),$$

where B is an artinian ring with identity. From this result easily follow results of A. KERTÉSZ and the present author [7] and of the author [10].

In this note we prove the following general decomposition theorem on artinian rings: *Let A be an artinian ring. Then A has a group direct decomposition*

$$A = B \oplus C,$$

where B is an ideal and C a left ideal of A such that C is an SA-ring and $S(A) \subseteq C$; B is an artinian ring with right identity and has some other properties (cf. Theorem 1). From this result one easily obtains the above mentioned theorem of Dinh Van Huynh and other results.

Especially we obtain conditions for an artinian ring to possess a (right, left) identity.

2. Preliminaries

Let A be a ring. We shall denote the additive group of A by $(A, +)$, the Jacobson radical of A by $J(A)$, and the ring of all $n \times n$ matrices over A by A_n, respectively. The symbol $\boxplus$ stands for a ring theoretic direct sum, the symbols $\oplus$ and $\Sigma^{\oplus}$ for group theoretic direct sums.

A primary ring A is a ring with identity such that $A/J(A)$ is a full matrix ring over a division ring.

For a prime p, $Z(p^\infty)$ denotes the quasicyclic p-group and $r(A)$ denotes the right annihilator of the ring A. We remark that an SA-ring A is a torsion ring and

$$A = A(p_1) \boxplus \ldots \boxplus A(p_l)$$

where the $A(p_i)$-s, the so called p_i-components of A, are p_i-rings for distinct primes p_i. $(A(p_i), +)$ is a direct sum of a finite group and finitely many copies of $Z(p_i^\infty)$ lying in the annihilator of $A(p_i)$ (and thus of A).

PROPOSITION 1. *([6], p. 205) Let A be a semisimple artinian ring and M a (right) A-module. Then*

$$M = M_0 \oplus M_1$$

where M_0 is a trivial A-module and M_1 a completely reducible unitary A-module.

If A is a semisimple artinian ring and M a unitary simple A-module, then clearly M is isomorphic to a minimal right ideal of A. Therefore we have the

COROLLARY. *If A is a simple artinian ring which is not a zero ring, and M a unitary simple A-module, then M is finite iff A is finite (i.e. a full matrix ring over a finite field).*

We call an A-module M strong artinian if the underlying group of M satisfies the minimum condition on subgroups. The following proposition will be used later on:

PROPOSITION 2. *Let A be an artinian ring such that $A/J(A)$ is finite and let M be an artinian A-module. Then M is strong artinian.*

PROOF. We prove the proposition by induction on the nilpotence degree m of $J(A)$.

If $m=1$, then A is semisimple artinian, and by Proposition 1

$$M = M_0 \oplus MA$$

where M_0 is a trivial A-module and MA a unitary completely reducible A-module. Since M is an artinian A-module MA is artinian, i.e. MA is a finite direct sum of simple A-modules. Since A is finite, from the corollary to Proposition 1 it follows that MA is finite. M_0 is a trivial artinian A-module, hence strong artinian. Therefore M is strong artinian.

Now assume that the Proposition holds for every artinian ring B, $B/J(B)$ finite, with nilpotence degree of $J(B)<m$ and let m be the nilpotence degree of $J(A)$. Then $MJ(A)^{m-1}$ is a submodule of M and therefore also artinian. Defining

$$x\bar{a} = xa, \quad x \in MJ(A)^{m-1}, \quad a \in \bar{a} \in A/J(A)$$

$MJ(A)^{m-1}$ becomes an $A/J(A)$-module and the $A/J(A)$-submodules and A-submodules coincide.

From the first part of the proof it follows that $MJ(A)^{m-1}$ is strong artinian.

Now it suffices to prove that $M/MJ(A)^{m-1}$ is strong artinian. Similary as above the definition

$$\bar{y}\bar{a} = \bar{y}a, \ \bar{y} \in M/MJ(A)^{m-1}, \quad a \in \bar{a} \in A/J(A)^{m-1}$$

makes $M/MJ(A)^{m-1}$ to an $A/J(A)^{m-1}$-module and the A-submodules and the $A/J(A)^{m-1}$-submodules coincide. The nilpotence degree of $J(A/J(A)^{m-1})$ is $<m$ and by induction hypothesis $M/MJ(A)^{m-1}$ is strong artinian.

REMARK. Proposition 2 also holds (with the same proof) if $A=J(A)$, i.e. $A/J(A)=(0)$. Then $MA=(0)$ if $m=1$.

COROLLARY. *([1], Lemma II and [2], Hilfssatz 3.1) If A is an artinian ring such that $A/J(A)$ is finite, then A is strong artinian.*

3. Main theorem

THEOREM 1. *Let A be an artinian ring. There are left ideals B and C of A such that*

$$A = B \oplus C$$

where

(1) $BC=(0)$, i.e. B is a two-sided ideal of A,
(2) C is strong artinian,
(3) $S(A) \subseteq C$,
(4) B is an artinian ring with right identity e_R and $Ce_R=(0)$,
(5) $B/J(B)$ is a direct sum of full matrix rings over infinite division rings,
(6) B does not contain a proper SA-right ideal.

PROOF. Let

$$\bar{A} = A/J(A) = \bar{e}_1 \bar{A} \bar{e}_1 \boxplus \ldots \boxplus \bar{e}_n \bar{A} \bar{e}_n$$

with simple rings $\bar{e}_i \bar{A} \bar{e}_i$ $(i=1, \ldots, n)$. By Proposition 5 of [5] (p. 54) there exists a set of orthogonal idempotents $\{e_1, \ldots, e_n\}$ with $e_i \in \bar{e}_i$ $(i=1, \ldots, n)$. Consider the two-sided Peirce decomposition of A relative to the orthogonal idempotents $\{e_1, \ldots, e_n\}$ (cf. [5], p. 56):

$$A = \sum_{i,j=1}^{n} \oplus e_i A e_j \oplus (1-e)Ae \oplus eA(1-e) \oplus (1-e)A(1-e)$$

where

$$(1-e)Ae = \{ae-eae\colon\ a \in A\},$$

$$eA(1-e) = \{ea-eae\colon\ a \in A\},$$

$$(1-e)A(1-e) = \{a-ea-ae+eae\colon\ a \in A\}.$$

Define

$$(1-e)Ae_i = \{ae_i - eae_i\colon\ a \in A\}, \quad e_i A(1-e) = \{e_i a - e_i ae\colon\ a \in A\},$$

$(i=1, \ldots, n)$. Clearly then

$$(1-e)Ae = \sum_{i=1}^{n} \oplus (1-e)Ae_i, \quad eA(1-e) = \sum_{i=1}^{n} \oplus e_i A(1-e).$$

$e_i A e_i$ $(i=1, \ldots, n)$ is a primary artinian ring with radical $e_i J(A) e_i$ (cf. [5]).
Now assume that $e_i A e_i / e_i J(A) e_i \cong \bar{e}_i \bar{A} \bar{e}_i$ is infinite for $i=1, \ldots, t$ and finite for

$i=t+1, \ldots, n$. Let

$$B = \sum_{\substack{i=1 \\ j=1}}^{\substack{t \\ n}}{}^{\oplus} e_i A e_j \oplus \sum_{i=1}^{t}{}^{\oplus} (1-e) A e_i,$$

$$C = \sum_{\substack{i=1 \\ j=t+1}}^{\substack{n \\ n}}{}^{\oplus} e_i A e_j \oplus \sum_{i=t+1}^{n}{}^{\oplus} (1-e) A e_i \oplus eA(1-e) \oplus (1-e)A(1-e).$$

Then clearly $A=B \oplus C$.

Now F. Szász has proved [8] that $A(1-e)=\{a-ae\colon a \in A\}$ is an artinian ring. Since $A(1-e) \subseteqq J(A)$ is nilpotent, it follows from Proposition 2 that $A(1-e)$ is strong artinian.

Since $e_i A e_i$ is artinian from the corollary of Proposition 2 it follows that $e_i A e_i$ is strong artinian for $i=t+1, \ldots, n$. $e_i A e_j$, $i \neq j$, $j>t$ is a unitary $e_j A e_j$-right module. One easily verifies that if

$$U_1 \supset U_2 \supset \ldots$$

is a strictly descending chain of submodules of the $e_j A e_j$-right module $e_i A e_j$, then

$$U_1 A \supset U_2 A \supset \ldots$$

is a strictly descending chain of right ideals of A. Therefore $e_i A e_j$ is an artinian $e_j A e_j$-module and by Proposition 2 $e_i A e_j$ is strong artinian.

The same argument shows that $(1-e) A e_i$, $i=t+1, \ldots, n$ is strong artinian.

We have proved that C is strong artinian.

A simple calculation shows that B and C are left ideals of A. Now we shall prove $BC=(0)$. Write

$$eA(1-e) = \sum_{i=1}^{n}{}^{\oplus} e_i A(1-e).$$

If we prove $e_i A e_j=(0)$ whenever $i \leqq t$, $j>t$ and $e_i A(1-e)=(0)$ whenever $i \leqq t$, we are done.

Assume $J(A)^m=(0)$. We prove $e_i J(A)^{m-k} e_j=(0)$ by induction on k. If $k=0$, this is true.

Our induction hypothesis is $e_i J(A)^{m-k} e_j=(0)$ and we shall prove $e_i J(A)^{m-k-1} e_j=(0.)$

$e_i J(A)^{m-k-1} e_j$ is a unitary left $e_i A e_i$-module and is strong artinian (contained in C). In a natural way (see above) $e_i J(A)^{m-k-1} e_j$ becomes an $e_i A e_i / e_i J(A) e_i$-left module since

$$e_i J(A) e_i e_i J(A)^{m-k-1} e_j \subseteqq e_i J(A)^{m-k} e_j = (0).$$

But $e_i A e_i / e_i J(A) e_i$ is a matrix ring over an infinite division ring, therefore (since $e_i J(A)^{m-k-1} e_j$ is completely reducible) if $e_i J(A)^{m-k-1} e_j \neq (0)$, it is not strong artinian, a contradiction.

In case $k=m-1$ we have $e_i A e_j=(0)$ as asserted.

The same argument shows $e_i A(1-e)=(0)$, $i=1, \ldots, t$.

Thus (1) and (2) of the theorem are proved. We remark that the structure of C is

$$C = \sum_{i,j=t+1}^{n} \oplus e_i A e_j \oplus \sum_{i=t+1}^{n} \oplus (1-e) A e_i \oplus \sum_{i=t+1}^{n} \oplus e_i A (1-e) \oplus (1-e) A (1-e).$$

Now B is (since A is artinian) an artinian right A-module and from $BC=(0)$ we have that B is an artinian ring. Moreover $e_1+\ldots+e_t=e_R$ is a right identity of B with $Ce_R=(0)$. The validity of (5) follows from the denumeration of the e_i's. To prove (6) let R be a right ideal of B which is strong artinian, i.e. finite. Then $RJ(B)^{m-k-1}$ is a finite right ideal. If $RJ(B)^{m-k}=(0)$ then $RJ(B)^{m-k-1}$ is a unitary right $B/J(B)$-module, therefore it is completely reducible. From (5) it follows that every completely reducible $B/J(B)$-module $\neq(0)$ is infinite. Hence $RJ(B)^{m-k-1}=(0)$. We obtain $RB=R=(0)$.

From (6) we have $B \cap S(A)=(0)$. Then if $x \in S(A)$

$$x = b+c, \quad b \in B,\ c \in C,$$

$$xe_R = be_R + ce_R = b \in B \cap S(A) = (0),$$

i.e. $b=0$, $S(A) \subseteq C$. The proof is now complete.

4. Further results

If A is an artinian ring, then A contains a minimal two-sided ideal M with respect to the property that A/M is strong artinian. This ideal $M=M(A)$ is *uniquely determined,* for if M_1, M_2 are ideals of A with A/M_1, A/M_2 strong artinian, then by the first isomorphism theorem

$$M_1/(M_1 \cap M_2) \cong (M_1+M_2)/M_2 \subseteq A/M_2$$

is strong artinian and therefore $A/(M_1 \cap M_2)$ is strong artinian. From the main theorem we have $M(A)=B$, i.e.:

THEOREM 2. *$M(A)$ is an artinian ring with right identity satisfying* (5) *and* (6) *and a complement left ideal C satisfying* (2) *and* (3).
The decomposition

$$A = M(A) \oplus C$$

is a ring theoretic one iff C is a right ideal of A, i.e. iff

$$\left[\sum_{i=t+1}^{n} \oplus e_i A (1-e) \oplus (1-e) A (1-e)\right]\left[\sum_{j=1}^{t} \oplus (1-e) A e_j\right] = (0)$$

and

$$\sum_{\substack{i=t+1 \\ j=1}}^{\substack{n \\ t}} \oplus e_i A e_j = (0).$$

Then clearly $C=S(A)$; i.e. $S(A)$ is a ring direct summand of A; artinian rings with this property Dinh Van Huynh has called Z-rings. In [2] DINH VAN HUYNH has given a necessary and sufficient condition for an artinian ring to be a Z-ring.

PROPOSITION 3. *Let A be an artinian ring such that $M(A)$ has a left identity. Then*

$$A = M(A) \boxplus S(A).$$

PROOF. If e_R is the identity of $M(A)$ then $CM(A)=Ce_RM(A)=(0)$.

THEOREM 3. *([2], Satz 2.1) Let A be a left and right artinian ring. Then*

$$A = M(A) \boxplus S(A)$$

where $M(A)$ has an identity and $M(A)$ does not contain proper left or right SA-ideals.

PROOF. Follows by left-right symmetry from our main theorem.

THEOREM 4. *([7], Satz 3) Let A be an artinian ring such that $J(A)$ is itself an artinian ring. Then*

$$(*) \qquad A = S_{n_1}^{(1)} \boxplus \ldots \boxplus S_{n_k}^{(k)} \boxplus Q$$

where $S^{(i)}$ are infinite division rings and Q is strong artinian.

PROOF. By Proposition 2 $J(A)$ is strong artinian and therefore $J(A) \subseteqq S(A) \subseteqq C$. Hence $J(M(A))=(0)$, $M(A)$ has an identity and the Theorem follows from Proposition 3 and the Wedderburn—Artin theorem.

One easily verifies that an ideal of a ring as in $(*)$ is itself such a ring, i.e. is artinian. Thus the rings of Theorem 4 are precisely the rings each ideal of which is artinian, the so called hereditarily artinian rings.

THEOREM 5. *Let A be an artinian ring and I an ideal of A which is a left and right artinian ring. Then*

$$A = I^* \boxplus D$$

where I^* is two-sided artinian with (5) *and* (6), *$I^* \subseteqq I$, *I/I^* strong artinian.*

PROOF. Let $A=B \oplus C$ as in the main theorem. Theorem 3 leads to

$$I = M(I) \boxplus S(I).$$

Let e' be the identity of $M(I)$, $i' \in M(I)$, $a \in A$. Then

$$i'a = i'' + s, \quad i'' \in M(I), \quad s \in S(I)$$

since I is an ideal of A. It follows

$$e'i'a = i'a = e'i'' + e's = i'',$$

i.e. $s=0$. This shows that $I^*=M(I)$ is a right ideal of A. A similar argument on the left proves I^* to be an ideal of A with identity and therefore

$$A = I^* \boxplus D$$

and I^* has the required properties.

It is easy to prove that $S(I)$ is an ideal of A and therefore $S(I) \subseteqq S(A) \subseteqq C$.

Now if A is two-sided artinian, then

$$A = B \boxplus C$$

and if $I=I^* \boxplus S(I)$ is a two-sided artinian ideal of A, we have

$$A = I^* \boxplus B^* \boxplus C,$$

$S(I) \subseteq C$, thus $I^* \boxplus C$ is a ring theoretic direct summand of A containing I and $(I^* \boxplus C)/I$ strong artinian. This is theorem 2.2 of DINH VAN HUYNHS paper [2].

Let A be an artinian ring. Then A contains a maximal ideal N with respect to the property that N is a left and right artinian ring. By Theorem 5

$$A = N^* \boxplus D,$$

N^* two-sided artinian, $S(A) \subseteq D$.

From the maximality of N it follows that the maximal two-sided artinian ideal of D is $S(A)$, i.e. $D/S(A)$ does not contain two-sided artinian ideals $\neq (0)$.

Now N^* splits into finitely many direct indecomposable two-sided artinian rings which do not contain a two-sided artinian ideal $\neq(0)$ by Theorem 5 (and since $S(A) \subseteq D$). This is theorem 2.3 of DINH VAN HUYNH [2].

Now

$$D = B' \oplus C'$$

with B', C' as B, C in the main theorem. We prove that B' does not contain a two-sided artinian ideal $\neq(0)$:

Assume I is such an ideal. By Theorem 3

$$I = M(I) \boxplus S(I).$$

As in the proof of theorem 6 we obtain that $M(I)$, $S(I)$ are ideals of B'. But B' does not contain any strong artinian right ideal $\neq(0)$, therefore $S(I)=(0)$ and I has an identity, hence $I^2=I$. Then

$$DI = (B'+C')I = B'I+C'I \subseteq I+C'I = I+C'I^2 \subseteq I+B'I \subseteq I,$$

$$ID = I(B'+C') = IB' \subseteq I,$$

i.e. I is a two-sided artinian ideal of D, a contradiction. This shows that the study of artinian rings is essentially reduced to the study of left and right artinian rings with identity not containing a two-sided artinian ideal $\neq(0)$ and satisfying (5) and (6) and to the study of artinian rings with right identity satisfying (5) and (6) and not containing two-sided artinian ideals $\neq(0)$.

At least we consider the existence of a (one- or two-sided) identity in an artinian ring.

THEOREM 6. *An artinian ring A has a right identity iff $A/M(A)$ has a right identity.*

PROOF. Clear by the main theorem since $A/M(A) \cong C$.

COROLLARY. *If $J(C)=(0)$, i.e. $A/M(A)$ has zero radical, i.e. $J(A) \subseteq M(A)$, i.e. A does not contain a nilpotent artinian left ideal, then A has a right identity.*

The last statement of the Corollary is a result of F. SZÁSZ [8].

COROLLARY. ([3], [4], [9].) *If A is a torsion-free artinian ring, then A has a right identity.*

PROOF. $C=(0)$ under the assumption

THEOREM 7. *For an artinian ring A are equivalent*
(1) *$M(A)$ has a left identity and $A/M(A)$ has a left identity,*
(2) *$M(A)$ has an identity and $A/M(A)$ has a left identity,*
(3) *$r(M(A))=(0)$ and $S(A)$ has a left identity,*
(4) *$r(M(A))=(0)$ and $A/M(A)$ has a left identity.*
Then A has a left identity.

PROOF. The equivalence of (1) and (2) is obvious. From the structure of $M(A)=B$ we have at once

$$r(M(A)) = \sum_{\substack{t=t+1\\ j=1}}^{\substack{t\\ n}}{}^{\oplus} e_i A e_j \oplus \sum_{i=1}^{t}{}^{\oplus} (1-e) A e_i .$$

Then $r(M(A))=(0)$ iff $M(A)$ has an identity. Hence (2)$\Leftrightarrow$(4).

THEOREM 8. *For an artinian ring A are equivalent*
(1) *A has an identity.*
(2) *$A/M(A)$ has an identity and $r(A)=(0)$.*

PROOF. (1)$\Rightarrow$(2) is trivial. (2)$\Rightarrow$(1): Let $A=M(A)\oplus C$, e the identity of $C\cong A/M(A)$. Let $e'=e_{t+1}+\ldots+e_n$. Since $\bar{e}_{t+1}+\ldots+\bar{e}_n$ is the identity of $(A/M(A))/J(A/M(A))$, we have $e-e'\in J(C)$. Now $e-e'=(e-e')^2$, therefore $e=e'$. Thus

$$C = \sum_{i,\,j=t+1}^{n}{}^{\oplus} e_i A e_j .$$

From $r(A)=(0)$ we then have $\sum_{i=1}^{t}(1-e)Ae_i=(0)$. Then $e_1+\ldots+e_n$ is the identity of A.

COROLLARY. *Let A be an artinian ring not containing a nilpotent artinian left ideal $\neq(0)$. Then A has an identity iff $r(A)=(0)$.*

REFERENCES

[1] BARTOLOZZI, F.: Anelli artiniani con radicale a quadrato nullo, *Rendiconti Circolo Mat. Palermo* **19** (1970), 113—122.
[2] DINH VAN HUYNH: Über artinsche Ringe, *to appear*.
[3] HERSTEIN, I. N.: On torsion free Artin rings, *Ann. Univ. Sci. Budapest* **7** (1964), 97—98.
[4] HOPKINS, CH.: Rings with minimal condition for left ideals, *Ann. of Math.* **40** (1939), 712—730.
[5] JACOBSON, N.: Structure of rings, Providence 1956, 1964.
[6] KERTÉSZ, A.: *Vorlesungen über artinsche Ringe*, Budapest—Leipzig 1968.
[7] KERTÉSZ, A. und WIDIGER, A.: Artinsche Ringe mit artinschem Radikal, *J. Reine Angew. Math.* **242** (1970), 8—15.
[8] SZÁSZ, F.: Hinreichende Bedingung für die Existenz eines Rechtseinselementes in einem Ring, *Publ. Math. Debrecen* **14** (1967), 151—152.
[9] SZÁSZ, F.: Über artinsche Ringe, *Bull. Acad. Polon. Sci. Ser. Math. Astr. Phys.* **11** (1963), 351—354.
[10] WIDIGER, A.: Zur Zerlegung artinscher Ringe, *Publ. Math. Debrecen* **21** (1974), 193—196.

Martin-Luther-Universität, Sektion Mathematik,
DDR-402 Halle, Universitätsplatz 6

(Received May 4, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 37—39

RADICAL IDEALS OF PRINCIPAL CLASS

by

D. D. ANDERSON (Columbia)

E. DAVIS [1, Remark, p. 203] has shown that a prime P of principal class can be generated by an R-sequence of length ht (P).[1] Actually, the proof given by Davis also applies to radical ideals. In this note we offer a simple inductive proof of this result. We denote the Jacobson radical and the zero-divisors of a commutative ring R by $J(R)$ and $Z(R)$, respectively.

THEOREM 1. *Let R be a Noetherian ring and $I=(a_1, \ldots, a_n)$ a radical ideal of height n contained in $J(R)$. Then $a_1, \ldots, a_n$ is an R-sequence and $I_i=(a_1, \ldots, a_i)$ is a radical ideal of height i $(0 \leqq i \leqq n, I_0=(0))$. Moreover, each minimal prime ideal of I_{i-1} is contained in a minimal prime ideal of I_i.*

PROOF. Assume that $n=1$, so that $I=(a_1)$ is a height one radical ideal. Now $I=P_1 \cap \ldots \cap P_s$ where each P_i is a minimal prime ideal of I and ht $(P_i)=1$. Let $Q_i \subsetneqq P_i$ be a minimal prime ideal. Then $Q_1 \cap \ldots \cap Q_s \subset P_i \cap \ldots \cap P_s=(a_1)$, so that $Q_1 \cap \ldots \cap Q_s=C(a_1)$ for some ideal C. Since $C(a_1) \subset Q_i$ and $a_1 \notin Q_i$, $C \subset Q_i$ and hence $Q_1 \cap \ldots \cap Q_s=(Q_1 \cap \ldots \cap Q_s)(a_1)$. But $a_1 \in J(R)$, so by Nakayama's Lemma $Q_1 \cap \ldots \cap Q_s=0$. Thus $\{Q_1, \ldots, Q_s\}$ is the set of minimal primes of R and $Q_i \subsetneqq P_i$. Since R is reduced, $Z(R)=Q_i \cup \ldots \cup Q_s$ and hence $a_1 \notin Z(R)$.

Assume that $n>1$ and pass to $\bar{R}=R/(a_1)$. In $\bar{R}$, $\bar{I}=(a_1, \ldots, a_n)/(a_1)$ can be generated by $n-1$ elements and each prime ideal minimal over $\bar{I}$ has the form $P/(a_1)$ where P is minimal over I. Hence ht $(P)=n$. But by the Principal Ideal Theorem ht $\big(P/(a_1)\big)=n-1$ and hence ht $\big(I/(a)\big)=n-1$. By induction, $\bar{a}_2, \ldots, \bar{a}_n$ is an R-sequence, $\bar{R}$ is reduced (i.e., (a_1) is a radical ideal) and $\bar{I}_i=(a_1, \ldots, a_i)/(a_1)$ is a radical ideal of height $i-1$. Moreover, each minimal prime of $\bar{I}_i$ is contained in a minimal prime of $\bar{I}_{i+1}$. We show that ht $(I_i)=i$. Let Q be a minimal prime ideal of I_i $(i \geqq 1)$. Then $Q \subset P$ is a minimal prime ideal of I. Since $I_P=P_P, R_P$ is an n-dimensional regular local ring. In R_P, $I_{iP}=(a_1, \ldots, a_i)_P$ is a prime ideal of height i. But $Q \subset P$ and Q is minimal over I_i so $Q_P=I_{iP}$ and hence ht $(Q)=$ $=$ht $(Q_P)=i$. Thus rank $I_i=i$. In particular (a_1) is a height one radical ideal. By the case $n=1$, R is reduced, each minimal prime of R is contained in a minimal prime ideal of I_1 and $a_1 \notin Z(R)$. Hence $a_1, \ldots, a_n$ is an R-sequence.

---

[1] Following I. KAPLANSKY [2], let R be a commutative ring with identity 1. If A is an R-module (left module), then we denote the set of zero-divisors of R on A by $Z(A)$, i.e. $Z(A)=\{r \in R \mid ra=0, \exists a \neq 0, a \in A\}$. Then, a sequence $x_1, x_2, \ldots, x_n$ of elements of R is called an R-sequence (of length n) (i) $(x_1, x_2, \ldots, x)_n \neq R$, and (ii) $x_1 Z(R), \ldots, x_i Z(R/(x_1, \ldots, x_{i-1}))$ $(i=1, 2, \ldots, n)$.

COROLLARY 1.1. *Let* $I=(a_1, \ldots, a_n)$ *be a radical ideal of height n in a Noetherian ring. Then* $G(I)=n$ *and I can be generated by an R-sequence.*[2]

PROOF. By Theorem 135 in [2], there exists a maximal ideal $M \supset I$ with $G(I)=$ $=G(I_M)$. By Theorem 1, $G(I)=G(I_M)=n$. Thus by [1, Appendix] or Theorem 125 in [2], I can be generated by an R-sequence.

COROLLARY 1.2. *Let* $P=(a_1, \ldots, a_n)$ *be a prime ideal of height n in a Noetherian ring R. If* $P \subset J(R)$, *then* $a_1, \ldots, a_n$ *is an R-sequence. In any case,* $G(P)=n$ *and P can be generated by an R-sequence.*

We first proved Corollary 1.2 for R a local ring in answer to a query in [3]: If R is a local ring and $P=(a_1, \ldots, a_n)$ is a prime ideal of height n, must $a_1, \ldots, a_n$ be an R-sequence? Later E. Davis communicated to I. Kaplansky a proof of the query similar to ours and remarked that he had proved the result mentioned in the introduction.

By starting the induction from the other end (as pointed out the to author by J. Ohm) we have the following non-Noetherian result from which part of Theorem 1 may be derived. The proof is similar to Theorem 1 except we pass to $R/(a_n)$ instead of $R/(a_1)$. The implication ($\Rightarrow$) uses the fact that if x is contained in a height one prime of a domain, then $\bigcap_{n=1}^{n} (x)^n=(0)$ (Corollary 1.6 in [4]).

THEOREM 2. *Let R be a ring and* $I=(a_1, \ldots, a_n)$ *a radical ideal of R. Then* $(0) \subsetneqq (a_1) \subsetneqq \ldots \subsetneqq (a_1, \ldots, a_n)$ *is a chain of radical ideals with each minimal prime of* $(a_1, \ldots, a_i)$ *having height i and each minimal prime of* $(a_1, \ldots, a_{i-1})$ *being contained in a minimal prime of* $(a_1, \ldots, a_i)$ $(1 \leqq i \leqq n)$ *if and only if for* $i=1, \ldots, n$

(1) a_i *is not contained in any minimal prime ideal of* $(a_1, \ldots, a_{i-1})$ *that is contained in a minimal prime ideal of* $(a_1, \ldots, a_n)$, *and*

$$(2) \quad \bigcap_{s=1}^{\infty} [(a_1, \ldots, a_{i-1})+(a_i)^s] = (a_1, \ldots, a_{i-1}).$$

COROLLARY 2.1 (J. Ohm). *Let R be a ring and* $P=(a_1, \ldots, a_n)$ *a prime ideal of R. Then* $(0) \subsetneqq (a_1) \subsetneqq \ldots \subsetneqq (a_1, \ldots, a_n)=P$ *is a saturated chain of prime ideals if and only if for* $i=1, \ldots, n$

(1) a_i *is not contained in any minimal prime ideal of* $(a_1, \ldots, a_{i-1})$ *that is contained in P, and*

$$(2) \quad \bigcap_{s=1}^{\infty} [(a_1, \ldots, a_{i-1})+(a_i)^s] = (a_1, \ldots, a_{i-1}).$$

[2] Let $I \neq R$ be an ideal of a Noetherian commutative ring R. Then any two R-sequences of maximal length contained in I have the same length. This is $\aleph_0$, and it is called the *grade* of I, and is denoted by $G(I)$. Thus $G(I)$ may differ from the Brown—McCoy radical $\tilde{G}(I)$ of the ideal I. The term "grade" and notation $G(I)$ are by no means standard. N. Bourbaki and other French mathematicians use the word "profondeur", which may be translated as "depth". On the other hand, $G(I)=n$ is the *least* integer n for which $\mathrm{Ext}_R^n(R/I, R) \neq 0$ (cf. CARTAN—EILENBERG [5]).

REFERENCES

[1] DAVIS, E. D.: Ideals of the principal class, R-sequences and a certain monoidal transformation, *Pacific J. Math.* **20** (1967), 197—205.
[2] KAPLANSKY, I.: *Commutative Rings*, Revised edition, University of Chicago Press (1974).
[3] KAPLANSKY, I.: *Topics in Commutative Ring Theory*, mimeographed notes (1974).
[4] OHM, J.: Some counterexamples related to integral closure in $D[[X]]$, *Trans. Amer. Math. Soc.* **122** (1966), 321—333.
[5] CARTAN, H.—EILENBERG, S.: *Homological Algebra*, Princeton University Press (Princeton, 1956).

University of Missouri-Columbia, Department of Mathematics,
Columbia, Missouri 65201, USA

(Received June 30, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 41—46

ON THE SCHRÖDINGER EQUATION OF THE THREE BODY PROBLEM I.

by

E. MAKAI

1.

Let x_i, y_i, z_i $(i=1, 2, 3)$ be real variables, $\Delta_i\psi=\frac{\partial^2\psi}{\partial x_i^2}+\frac{\partial^2\psi}{\partial y_i^2}+\frac{\partial^2\psi}{\partial z_i^2}$, E constant, μ_1, μ_2, μ_3 nonnegative constants, $\sum\mu_k>0$, U a function of the nine variables $x_1, y_1, \ldots, z_3$ and let ψ satisfy the differential equation

$$-\sum_{k=1}^{3}\mu_k\Delta_k\psi+U\psi-E\psi=0. \tag{1.1}$$

In the following we shall suppose that U is a function of the three variables

$$\begin{aligned} r_1&=[(x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2]^{1/2}\\ r_2&=[(x_3-x_1)^2+(y_3-y_1)^2+(z_3-z_1)^2]^{1/2}\\ r_3&=[(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{1/2} \end{aligned} \tag{1.2}$$

only* and U is reasonably smooth having singularities of a "mild" type. It is known that in this case (1.1) has solutions depending only on r_1, r_2 and r_3 and these solutions $\psi=\psi(r_1, r_2, r_3)$ will be the object of this paper. Introducing the notations

$$\Delta^1\psi=\frac{\partial^2\psi}{\partial r_2^2}+\frac{2}{r_2}\frac{\partial\psi}{\partial r_2}+\frac{-r_1^2+r_2^2+r_3^2}{r_2r_3}\cdot\frac{\partial^2\psi}{\partial r_2\partial r_3}+\frac{\partial^2\psi}{\partial r_3^2}+\frac{2}{r_3}\frac{\partial\psi}{\partial r_3}\quad\text{(cycl.)} \tag{1.3}$$

it is known (cf. [2], esp. p. 233) that any function $\psi=\psi(r_1, r_2, r_3)$ satisfying (1.1) satisfies the differential equation

$$-\sum\mu_k\Delta^k\psi+U\psi-E\psi=0, \tag{1.4}$$

too. Eigenfunctions of (1.4) are solutions for which the Lebesgue integral

$$\iiint_{D_r}\psi^2 r_1r_2r_3\,dr_1\,dr_2\,dr_3$$

* Formulae of type (1.2) will be abbreviated throughout this paper in the following way: $r_1=[(x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2]^{1/2}$ (cycl.). By this it will be meant that the formula remains valid after a cyclic interchange of the indices 1, 2, 3.

is finite. Here the domain D_r of integration is defined by the triangle inequalities

$$r_2+r_3-r_1 \geqq 0, \quad r_3+r_1-r_2 \geqq 0, \quad r_1+r_2-r_3 \geqq 0. \tag{1.5}$$

In this paper we are dealing with solutions of (1.4), which are not necessarily eigenfunctions, though their domain of definition is still D_r. Consider a plane P: $\sum\alpha_i r_i=\beta$ in the real space r_1, r_2, r_3, the intersection of which with the closed domain D_r is not empty. Let D_0 be a plane domain in $P\cap D_r$. The theorem of Cauchy—Kowalewski states roughly, that if we prescribe values and normal derivatives of a solution of (1.4) in D_0, then there exists a domain D_1, the three dimensional measure of which in the space r_1, r_2, r_3 is positive, $D_0\subset D_1\subset D_r$, further a unique solution of (1.4) such that it satisfies the prescribed initial conditions and is analytic in D_1 provided certain conditions are met. These conditions are partly related to the smoothness of the initial data on D_0, partly to the structure of the coefficients of equations (1.4) on and near to D_0. If D_0 contains no point of the boundary of D_r and e.g. $U=\sum e_i/r_i$, e_i constants, then the theorem of Cauchy—Kowalewski is applicable, yet this is not within the scope of this paper.

The situation is quite different if we specialize the plane P so, that it should coincide with one of the boundary planes of the trihedral domain D_r, say, $\alpha_1=-1$, $\alpha_2=\alpha_3=1$, $\beta=0$, cf. (1.5). (This means that the points (x_i, y_i, z_i), $i=1, 2, 3$ lie on a straight line in the (x, y, z)-space.) The conditions of the theorem of Cauchy and Kowalewski are now not met, as we shall see it later. One cannot prescribe simultaneously on D_0 values and normal derivatives, so that they should guarantee the existence of an analytic solution of (1.4) satisfying the initial data in some three dimensional domain.

Instead, let us consider a sufficiently small closed part D_0 of the boundary of the trihedral domain D_r not containing any point of any edge of D_r and let us suppose that in the points of D_0 the function U is analytic*. *Then, prescribing on* D_0 *analytic initial values, these initial data alone determine uniquely an analytic solution of* (1.4) *in some closed domain* D_1. Here $D_0\subset D_1\subset D_r$ and the three dimensional measure of D_1 in the space r_1, r_2, r_3 is positive. This is the essential content of statement (i) is section 2.

Again, *if a solution of* (1.4) *is analytic in the interior of* D_1 *then* by statement (ii) of section 2 *this solution is either analytic on* D_0 *or exhibits a logarithmic singularity there*. Here, as in the previous paragraph, D_0 is part of the boundary of D_1 and of the boundary of D_r. This is related to T. KATO's result [4], that if U is sufficiently smooth, then the eigenfunctions of (1.4) are everywhere continuous in D_r and on the boundary of D_r, as well.

In the important special case $U=\sum e_i/r_i$ (e_1, e_2, e_3 constants), U is analytic everywhere in the trihedral domain D_r, save on its three edges and Kato's theorem holds.

* Analyticity means in this context not analyticity with respect to the variables r_1, r_2, r_3 but analyticity in the closely related variables s_1, s_2, s_3 defined by (2.1) for which $s_1+s_2=2r_3$ (cycl.) holds.

2.

Let us introduce the new variables

$$s_1 = r_2+r_3-r_1, \quad s_2 = r_3+r_1-r_2, \quad s_3 = r_1+r_2-r_3 \tag{2.1}$$

already used by C. L. PEKERIS [5] for the numerical approximation of an eigenfunction in a special case. By using these variables the domain D_r will be transformed into

$$D_s: \quad s_1 \geqq 0, \quad s_2 \geqq 0, \quad s_3 \geqq 0. \tag{2.2}$$

The quantities $\Delta^k\psi$ in the Schrödinger equation (1.4) become in the variables s_i more intricate, namely if $\psi(r_1, r_2, r_3)=\varphi(s_1, s_2, s_3)$ then we have

$$2r_1r_2r_3\Delta^1\psi = \sum_{i,j=1}^{3} b_{ij}^{(1)} \frac{\partial^2\varphi}{\partial s_i \partial s_j} + \sum_{i=1}^{3} b_i^{(1)} \frac{\partial\varphi}{\partial s_i} := \tag{2.3}$$

$$:= 2r_1r_2r_3\Delta^{(1)}\varphi \quad \text{(cycl.)}$$

where $b_{ij}^{1)}=b_{ji}^{(1)}$ and

$$b_{11}^{(1)} = s_1(s_2+s_3)(s_1+s_2+s_3), \quad b_{22}^{(1)} = -b_{23}^{(1)} = b_{33}^{(1)} = s_2s_3(s_2+s_3), \tag{2.4}$$

$$b_{12}^{(1)} = b_{13}^{(1)} = 0, \quad b_1^{(1)} = (s_2+s_3)(2s_1+s_2+s_3), \quad b_2^{(1)} = -b_3^{(1)} = s_3^2-s_2^2 \quad \text{(cycl.)}$$

yet one has the advantage that the equation (1.4) transformed by (2.3) into the form

$$-\sum \mu_k \Delta^{(k)}\varphi + (V-E)\varphi = 0, \quad U(r_1, r_2, r_3) = V(s_1, s_2, s_3), \tag{2.5}$$

becomes a particular case of a type of partial differential equations, which was investigated in recent years by several authors, namely by M. S. BAOUENDI and C. GOULAOUIC [1] and by V. H. FROIM [3]. As a simple application of their theorems we get the following two statements.

Suppose that the function $V(s_1, s_2, s_3)$ is an analytic function of the arguments s_1, s_2, s_3 in a complex neighbourhood N of $s_1=0$, $s_2=s_2^0(>0)$, $s_3=s_3^0(>0)$, i.e. in a complex neighbourhood of a boundary point of the real domain D_s not lying on any axis of coordinates. This occurs in the special case $U(r_1, r_2, r_3)=\sum e_i/r_i$ (e_i constants) equivalent to

$$V(s_1, s_2, s_3) = 2\left(\frac{e_1}{s_2+s_3}+\frac{e_2}{s_3+s_1}+\frac{e_3}{s_1+s_2}\right)$$

since if N is sufficiently small, each denominator is positive.

Further, let us define by the aid of the positive numbers R_1, R_2, R_3, where $R_2<s_2^0$, $R_3<s_3^0$ the complex domain

$$G: \quad |s_1| < R_1, \quad |s_2-s_2^0| < R_2, \quad |s_3-s_3^0| < R_3$$

and the domain

$$G_0: \quad 0 < |s_1| < R_1, \quad |\arg s_1| < \pi, \quad |s_2-s_2^0| < R_2, \quad |s_3-s_3^0| < R_3,$$

where we suppose that $G \subset N$. Then we have

(i) *For any function* $u(s_2, s_3)$ *analytic in a neighbourhood* N' *of* s_2^0, s_3^0 *there exists a uniquely determined solution* $\varphi(s_1, s_2, s_3)$ *of* (2.5) *analytic in a complex three dimensional neighbourhood* G *of* $0, s_2^0, s_3^0$ *with suitably chosen constants* R_1, R_2, R_3 *and satisfying the only boundary condition*

$$\varphi(0, s_2, s_3) = u(s_2, s_3).$$

(ii) *If a non-trivial solution* φ *of* (2.5) *is analytic in* G_0, *then it is of the form*

$$\varphi_0(s_1, s_2, s_3) + \varphi_1(s_1, s_2, s_3) \log s_1 \tag{2.6}$$

where φ_0 *and* φ_1 *are analytic in* G, *hence on the intersection of* G *and of the surface* $s_1 = 0$, *too*. On the latter point set $|\varphi_0(0, s_2, s_3)| + |\varphi_1(0, s_2, s_3)| \not\equiv 0$ holds. Further, $\varphi_1(0, s_2, s_3) \equiv 0$ implies $\varphi_1(s_1, s_2, s_3) \equiv 0$ and the functions $\varphi_0(0, s_2, s_3)$ and $\varphi_1(0, s_2, s_3)$ determine uniquely the solution of (2.6).

3.

Consider the partial differential equation

$$w_{11} + a_{12} w_{12} + a_{13} w_{13} + \frac{1}{z_1}(a_{22} w_{22} + a_{23} w_{23} + a_{33} w_{33}) + \tag{3.1}$$

$$+ \frac{1}{z_1}(a_1 w_1 + a_2 w_2 + a_3 w_3) + \frac{1}{z_1^2} a w = 0,$$

where $w = w(z_1, z_2, z_3)$, $w_i = \partial w / \partial z_i$, $w_{ij} = \partial^2 w / \partial z_i \partial z_j$, the quantities a_{ij}, a_i and a are analytic functions of z_1, z_2, z_3 in a complex neighbourhood N_0 of $z_1 = z_2 = z_3 = 0$, finally

$$a_1^0 = a_1(0, z_2, z_3) = 1, \quad a^0 = a(0, z_2, z_3) = 0. \tag{3.2}$$

Under these assumptions a special case of a theorem of BAOUENDI and GOULAOUIC [1] states, that there exists a neighbourhood N_0' of $z_1 = z_2 = z_3 = 0$ in which a unique analytic solution of (3.1) exists completely determined by the *only* initial condition

$$w(0, z_2, z_3) = \Phi(z_2, z_3), \tag{3.3}$$

where $\Phi(z_2, z_3)$ is an analytic function of z_2 and z_3, if $(0, z_2, z_3) \in N_0'$.

Further the theorem of FROIM [3] states that under the same assumptions there exist suitably small positive constants R_1, R_2, R_3, such that any analytic solution of (3.1) in

$$0 < |z_1| < R_1, \quad |\arg z_1| < \pi, \quad |z_2| < R_2, \quad |z_3| < R_3$$

is of the form

$$\Phi_0(z_1, z_2, z_3) + \Phi_1(z_1, z_2, z_3) \log z_1$$

where Φ_0 and Φ_1 are analytic in $|z_j| < R_j$ $(j = 1, 2, 3)$.

4.

For proving the statements of section 2 it is obviously enough to show that equation (2.5) becomes a particular case of (3.1) after the introduction of the new variables

$$z_1 = s_1, \quad z_2 = s_2 - s_2^0, \quad z_3 = s_3 - s_3^0 \tag{4.1}$$

and division by the coefficient of $\partial^2\varphi/\partial s_1^2$ $(=\partial^2 w/\partial z_1^2)$ taking into account that (2.5) is symmetric in the indices 1, 2, 3.

Using the notation

$$g(s_1, s_2, s_3) = \frac{1}{4}(s_1+s_2)(s_2+s_3)(s_3+s_1) = 2r_1r_2r_3 \tag{4.2}$$

equation (2.5) is equivalent to

$$\sum_{ij} c_{ij}\frac{\partial^2\varphi}{\partial s_i \partial s_j} + \sum_i c_i \frac{\partial\varphi}{\partial s_i} + c\varphi = 0, \tag{4.3}$$

where

$$c_{ij} = -\sum_k \mu_k b_{ij}^{(k)}, \quad c_i = -\sum_k \mu_k b_i^{(k)}, \quad c = g(s_1, s_2, s_3)(V-E), \tag{4.4}$$

and we are going to verify that (4.3) divided by c_{11} is indeed of the form (3.1) after a change of the variables s_j into z_j.

By (4.4) and (2.4) the coefficients c_{1j} are divisible by s_1, they are of the form $s_1 P_j(s_1, s_2, s_3)$, where $P_j(s_1, s_2, s_3)$ is a polynomial. The polynomial

$$P_1(s_1, s_2, s_3) = -\mu_1(s_2+s_3)(s_1+s_2+s_3) - \mu_2(s_1+s_3)s_3 - \mu_3(s_1+s_2)s_2$$

does not vanish in a neighbourhood of $s_1=0$, $s_2=s_2^0$, $s_3=s_3^0$, since both s_2^0 and s_3^0 are positive, hence $[P_1(s_1, s_2, s_3)]^{-1}=[P_1(z_1, z_2+s_2^0, z_3+s_3^0)]^{-1}$ is an analytic function of the variables z_j in a neighbourhood of the origin. Thus c_{12}/c_{11} and c_{13}/c_{11} are analytic functions of the z_j-s near the origin.

If $i, j>1$, then c_{ij} is a polynomial in the s_j-s, $c_{ij}/c_{11}=s_1^{-1}[c_{ij}/P_1(s_1, s_2, s_3)]$ and the quantity in the brackets is analytic. Similarly $c_i/c_{11}=s_1^{-1}[c_i/P_1(s_1, s_2, s_3)]$ and the quantity in the brackets is again analytic. In particular

$$a_1 = s_1\frac{c_1}{c_{11}} = \frac{-\mu_1(s_2+s_3)(2s_1+s_2+s_3) - \mu_2(s_3^2-s_1^2) - \mu_3(s_2^2-s_1^2)}{-\mu_1(s_2+s_3)(s_1+s_2+s_3) - \mu_2(s_1+s_3)s_3 - \mu_3(s_1+s_2)s_2}$$

thus

$$\lim_{z_1\to 0} a_1 = \lim_{s_1\to 0} a_1 = 1,$$

cf. (3.2). Finally

$$a = s_1^2\frac{c}{c_{11}} = s_1\frac{g(s_1, s_2, s_3)(V-E)}{P_1(s_1, s_2, s_3)}$$

is an analytic function of the z_j-s near the origin, since both s_1 and the quotient on the right hand side are analytic functions of the s_j-s near $0, s_2^0, s_3^0$. Moreover,

a satisfies condition (3.2), since

$$a^0 = \lim_{z_1 \to 0} a = \lim_{s_1 \to 0} a = 0.$$

Thus the theorems quoted in section 3 are applicable to equation (2.4).

REFERENCES

[1] BAOUENDI, M. S., and GOULAOUIC, C.: Cauchy problems with characteristic initial hypersurface, *Comm. Pure Appl. Math.* **26** (1973), 455—475.
[2] BETHE, H. A., and SALPETER, E. E.: Quantum Mechanics of one- and two-electron systems, *Encyclopedia of Physics*, vol. XXXV. Spinger, Berlin—Göttingen—Heidelberg, 1957.
[3] FROIM, V. H.: Linear scalar partial differential equations with regular singularities on a hyperplane, *Diff. Uravn.* **9** (1973), 533—541.
[4] KATO, T.: On the eigenfunctions of many-particle systems in Quantum Mechanics, *Comm. Pure Appl. Math.* **10** (1957), 151—177.
[5] PEKERIS, C. L.: Ground state of two-electron atoms, *Phys. Rev.* (2) **112** (1958), 1649—1658.

Mathematical Institute of the Hungarian Academy of Sciences

(Received July 15, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 47—60

ON DISCRIMINANT FORM AND INDEX FORM EQUATIONS

by

K. GYŐRY and Z. Z. PAPP

Dedicated to Professor P. Erdős on his 65th birthday

1. Introduction

Let K be an algebraic number field of degree $k \geqq 3$ with discriminant D_K, and let $1, \alpha_1, \ldots, \alpha_m \in K$ be $m+1 \geqq 3$ linearly independent algebraic integers over $\mathbf{Q}$ [1] with heights $\leqq H$. Let d be a fixed non-zero rational integer and $\varkappa > 9(k-1)(k-2)/2$, $\varepsilon > 0$ be given numbers. As a considerable generalization of some earlier results (for references see Sections 2 and 3) the first named author proved in [9] that all solutions of the *discriminant form equation*

$$\mathrm{Discr}_{K/\mathbf{Q}}(\alpha_1 x_1 + \ldots + \alpha_m x_m) = d \tag{1}$$

in rational integers $x_1, \ldots, x_m$ satisfy

$$|\mathbf{x}| < H^{m-1} \exp\{c_1 |d|^{\varkappa}\} \tag{2}$$

and

$$|\mathbf{x}| < H^{m-1} \exp\{c_2 [|D_K|^{3(k-1)(k-2)} (|D_K|^{3(k-1)(k-2)/2} + \log |d|)]^{1+\varepsilon}\}, \tag{3}$$

where $|\mathbf{x}| = \max_{1 \leqq i \leqq m} (|x_i|)$ and $c_1 = c_1(k, \varkappa)$, $c_2 = c_2(k, \varepsilon)$ are effectively computable positive numbers. In [9] these estimates were deduced from an effective theorem on algebraic numbers with given discriminant ([9], Théorème 3; see also [10], Theorem 3), the proof of which was based on an effective estimate of STARK [37] for linear forms in the logarithms of algebraic numbers.

If in particular $m = k-1$ and $1, \alpha_1, \ldots, \alpha_{k-1}$ form a basis for an order O of K with discriminant D_O, then

$$\mathrm{Discr}_{K/\mathbf{Q}}(\alpha_1 x_1 + \ldots + \alpha_{k-1} x_{k-1}) = [F(x_1, \ldots, x_{k-1})]^2 D_O \tag{4}$$

where $F(x_1, \ldots, x_{k-1})$ is a form of degree $k(k-1)/2$ with rational integer coefficients (cf. HENSEL [15]). F is a product of linear forms with algebraic coefficients. It is called the *index form* of the basis $1, \alpha_1, \ldots, \alpha_{k-1}$ of O. An important consequence of the above estimates is that all solutions in rational integers $x_1, \ldots, x_{k-1}$ of the *index form equation*

$$F(x_1, \ldots, x_{k-1}) = a, \tag{5}$$

where $a \neq 0$ denotes a rational integer satisfy for example

$$\max_{1 \leqq i \leqq k-1} (|x_i|) < H^{k-2} \exp\{c_1 |a^2 D_O|^{\varkappa}\} \tag{6}$$

---

[1] As usual, $\mathbf{Q}$ and $\mathbf{Z}$ denote the field of rational numbers and the ring of rational integers.

with the above c_1 ([9], Corollaire to Théorème 7). The special case $a=\pm 1$ is of particular interest, because (6) provides then an effective algorithm to determine all the integers α in K for which $O=\mathbf{Z}[\alpha]$. (See also Corollaire 3.3 of [9]).

In the present paper we give p-adic generalizations of the above mentioned results on discriminant form and index form equations and improve the estimates (2), (3) and (6). As an application of these we obtain explicit lower bounds for the maximum norm of the prime ideal divisors of a discriminant form.

Our proofs depend on certain theorems concerning algebraic integers with given relative discriminant over a fixed algebraic number field (see GYŐRY [11], [12]), which have been obtained by using recent inequalities of BAKER [3], VAN DER POORTEN [28] and VAN DER POORTEN and LOXTON [30] on linear forms in the logarithms of algebraic numbers.

In our paper [13] some consequences of the main result (Theorem 1) of this paper are deduced from our results concerning generalized norm form equations [13], but the estimates obtained here are sharper in terms of certain parameters.

2. Effective estimates for the integer solutions of discriminant form equations

Before we state our theorems on discriminant form equations, we establish our notation, recall some standard definitions and refer briefly to the earlier results obtained on these equations in various special cases.

Let $K\supset L$ be algebraic number fields with $[K: L]=k\geqq 3$ and let $\mathbf{Z}_K$ and $\mathbf{Z}_L$ denote the rings of integers of K and L. There are k isomorphisms of K over L into the complex numbers; denote the images of an element α of K under these isomorphisms by $\alpha^{(1)}, \ldots, \alpha^{(k)}$. Let $M(\mathbf{x})=\alpha_1 x_1+\ldots+\alpha_m x_m$ be a linear form with coefficients in K such that $K=L(\alpha_1, \ldots, \alpha_m)$ [2]. Put $M^{(j)}(\mathbf{x})=\alpha_1^{(j)}x_1+\ldots+\alpha_m^{(j)}x_m$ for $j=1, \ldots, k$. The discriminant

$$\text{Discr}\,(M(\mathbf{x})) = \prod_{1\leqq i<j\leqq k}(M^{(j)}(\mathbf{x})-M^{(i)}(\mathbf{x}))^2 \tag{7}$$

is a form of degree $k(k-1)$ with coefficients in L. A form obtained in this way is called a *discriminant form over* L and will be written as $D_{K/L}(\alpha_1 x_1+\ldots+\alpha_m x_m)$. This notion dates back to Kronecker (see e.g. [19] and [17]). It is easy to see that any discriminant form over L is a product of norm forms with coefficients in L. Moreover, if K is a doubly transitive extension of L, then any discriminant form $D_{K/L}(\alpha_1 x_1+\ldots+\alpha_m x_m)$ is a norm form over L.

Let now $\delta\neq 0$ be a fixed element in L and consider the *discriminant form equation*

$$D_{K/L}(\alpha_1 x_1+\ldots+\alpha_m x_m)=\delta \tag{8}$$

in integers $x_1, \ldots, x_m$ of L. We may assume without loss of generality that $\alpha_1, \ldots, \alpha_m$ and δ are algebraic integers.

If $1, \alpha_1, \ldots, \alpha_m$ are linearly dependent over L and (8) has a solution $(x_1^0, \ldots, x_m^0)\in\mathbf{Z}_L^m$, then $(x_1^0+u_1, \ldots, x_m^0+u_m)$ are also solutions of (8) for any

[2] If $K\supsetneqq L(\alpha_1, \ldots, \alpha_m)$, then in (7) $\text{Discr}\,(M(\mathbf{x}))\equiv 0$.

$(u_1, \ldots, u_m) \in \mathbf{Z}_L^m$ satisfying $u_1\alpha_1 + \ldots + u_m\alpha_m \in \mathbf{Z}_L$, that is in this case (8) has infinitely many solutions in integers of L.

In what follows we shall suppose that $1, \alpha_1, \ldots, \alpha_m$ are linearly independent over L.

In the important particular case when $L=\mathbf{Q}$, $m=k-1$ and $1, \alpha_1, \ldots, \alpha_{k-1}$ is a basis for an order of K, the related equations (8) and (23) (i.e. (1) and (5)) were studied by many authors. The purpose of their investigations was, among others, to obtain results about algebraic integers with given discriminant or given index in various algebraic number fields K. We refer to the earlier results connected with the solvability of the equations in question in Section 3.

If $L=\mathbf{Q}$ and $m=2$, (8) leads to the Thue equation and a famous theorem of THUE [39] implies the finiteness of the number of solutions of (8)) (in case $k=3$ cf. DELONE [5] and NAGELL [21]). Further a celebrated theorem of BAKER [2] provides an algorithm for determining all the solutions. Moreover, in case of a large class of discriminant form equations in two unknowns the number of solutions or even the solutions themselves can be explicitly given by means of the well-known results of Delone, Nagell and Faddeev on cubic diophantine equations (see, e.g., NAGELL [20] and DELONE and FADDEEV [6]).

In the case $L=\mathbf{Q}$, $m \leqq k-1=3$ the finiteness of the number of solutions of (8) was proved by NAGELL [23], [24]. For a class of biquadratic number fields K his proof is effective [23].

As we mentioned in the introduction, in [9] it was proved in full generality that in case $L=\mathbf{Q}$ (8) has only finitely many solutions in rational integers $x_1, \ldots, x_m$ for any k and m with $2 \leqq m \leqq k-1$ and all solutions satisfy (2) and (3). It is probable that if $L=\mathbf{Q}$, the finiteness of the number of solutions of (8) can be deduced from well-known theorems of SCHMIDT [32], [33] on norm form equations,[3] but only in an ineffective form, without giving effective upper bounds for the sizes of the solutions.

Let us now return to the general case when in (8) $K \supset L$ are arbitrary algebraic number fields with $[K: L]=k \geqq 3$ and $[L: \mathbf{Q}]=n \geqq 1$. Denote by D_K and D_L respectively the absolute values of their discriminants. Let h_L be the class number of L. As before, $1, \alpha_1, \ldots, \alpha_m$ will denote linearly independent algebraic integers over L such that $K=L(\alpha_1, \ldots, \alpha_m)$ and $\max_{1 \leqq i \leqq m} (\overline{|\alpha_i|}) \leqq H$ [4]. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ be $s \geqq 0$ distinct prime ideals of L with norms $N(\mathfrak{p}_j)=p_j^{f_j}$, $j=1, \ldots, s$, where p_j denote rational primes not exceeding P (for $s=0$ let $P=2$). Write $\mathfrak{p}_j^{h_L}=(\pi_j)$ with some $\pi_j \in \mathbf{Z}_L$ for $j=1, \ldots, s$. Further let δ, β be non-zero integers in L with $|N_{L/\mathbf{Q}}(\delta)| \leqq d$ $(d \geqq 3)$, $|N_{L/\mathbf{Q}}(\beta)| \leqq b$.

The main result of this paper is as follows.

THEOREM 1. *Let $L, K, \alpha_1, \ldots, \alpha_m, \pi_1, \ldots, \pi_s, \delta$ and β be defined as above. Then for every solution of the equation*

$$D_{K/L}(\alpha_1 x_1 + \ldots + \alpha_m x_m) = \delta \pi_1^{z_1} \ldots \pi_s^{z_s} \tag{9}$$

[3] For $m \leqq k-1 \leqq 3$ this has recently been made by B. Knight (private communication of Professor W. M. Schmidt). See also W. M. SCHMIDT, *Proc. Internat. Congress Math.*, Vancouver, 1974, Vol. I, pp. 177—185.

[4] $\overline{|\alpha|}$ denotes the maximum of the absolute values of the conjugates of an algebraic number α.

in integers $x_1, \dots, x_m \in \mathbf{Z}_L$, $z_1, \dots, z_s \in \mathbf{Z}$ *with* $(x_1, \dots, x_m) \mid (\beta)$, $z_j \geqq 0$ *for* $j=1, \dots, s$ *there exists a unit* ε *in* L *such that*

$$\max\left\{|\overline{x_1\varepsilon}|, \dots, |\overline{x_m\varepsilon}|, (\mathfrak{p}_1^{f_1 z_1} \dots \mathfrak{p}_s^{f_s z_s})^{h_L/nk(k-1)}\right\} < b^{1/n} H^{m(n+1)-1} T_i \tag{10}$$

for $i=1, 2$, *where*

$$T_1 = \exp\left\{c_3\left[c_4(s+1)^{30}\left((D_L d^{1/k} P^{(s+1)n})^{\frac{3}{2}} (5^{skn} \log(D_L\, dP)\right)^{n+1}\right)^l\right]^{sl+4}\right\}$$

and

$$T_2 = \exp\left\{c_5 D_L^{1/2} P^{nl}\left[c_6(s+1)^{30}(\log P)^2 \left(D_K^{3/2k} (\log D_K)^n\right)^l\right]^{sl+4} \log d\,(\log\log d)^2\right\}$$

with $l=k(k-1)(k-2)$ *and effectively computable positive constants* c_3, c_4, c_5, c_6 *depending only on* n *and* k.

Theorem 1 is a $\mathfrak{p}$-adic generalization of Théorème 7 of [9] (which is quoted in the introduction). In case $m=2$ it can be deduced from the effective theorems of SPRINDŽUK and KOTOV ([35], [18], [36]) concerning the generalized Thue—Mahler equation (with estimates different from (10)).

It is clear from Theorem 1 that, for any given non-zero $\delta \in \mathbf{Z}_L$, one can effectively determine the set of all algebraic numbers $x_1, \dots, x_m$ of L satisfying (8), the denominators of which are divisible solely by powers of $\mathfrak{p}_1, \dots, \mathfrak{p}_s$.

If in Theorem 1 we take into consideration the sizes $|\overline{\delta}|, |\overline{\pi_1}|, \dots, |\overline{\pi_s}|$ as well, we can easily get an upper bound for $\max(|\overline{x_1}|, \dots, |\overline{x_m}|)$. However, the above formulation of Theorem 1 will be more useful in the course of its applications.

Of particular interest is the quite good dependence on H in each of our results. In our theorems D_K can be estimated from above by $D_L^k (2H)^{mkn(kn-1)}$.

An important consequence of Theorem 1 is the following.

COROLLARY 1. *Let* $L, K, \alpha_1, \dots, \alpha_m$ *be as in Theorem* 1. *Let* $x_1, \dots, x_m$ *be relatively prime integers in* L. *Suppose that* $D_{K/L}(\alpha_1 x_1 + \dots + \alpha_m x_m)$ *has* s *distinct prime ideal divisors in* L *and denote by* $N(\mathfrak{p})$ *the maximum norm of these prime ideals* $\mathfrak{p}$. *Then we have*

$$\log N(\mathfrak{p}) + s\log(s+1) + s\log\log N(\mathfrak{p}) > c_7 \log\log N \tag{11}$$

and

$$N(\mathfrak{p}) > c_8 \log\log N \tag{12}$$

provided that $N \geqq N_0$, *where* $N = \max_{1 \leqq i \leqq m} (|N_{L/\mathbf{Q}}(x_i)|)$ *and* c_7, c_8, N_0 *are effectively computable positive constants depending only on* H, n, k *and* D_K.

Consider now the important special case of Theorem 1 when $L=\mathbf{Q}$. As above, let K be an algebraic number field of degree $k \geqq 3$ with discriminant D_K, and let $1, \alpha_1, \dots, \alpha_m$ be linearly independent algebraic integers over $\mathbf{Q}$ such that $K=\mathbf{Q}(\alpha_1, \dots, \alpha_m)$ and $\max_{1 \leqq j \leqq m} (|\overline{\alpha_j}|) \leqq H$. Let b, d denote non-zero rational integers with $d \geqq 3$ and $p_1, \dots, p_s$, $s \geqq 0$ distinct rational primes not exceeding P (with $P=2$ if $s=0$). We signify by $|p_1, \dots, |p_s$ the usual valuations of $\mathbf{Q}$ defined by $p_1, \dots, p_s$, normalized so that $|p_j|_{p_j} = 1/p_j$.

COROLLARY 2. *Under the above assumptions all solutions of the equation*

$$D_{K/\mathbf{Q}}(\alpha_1 x_1+\ldots+\alpha_m x_m) = d p_1^{z_1} \ldots p_s^{z_s} \tag{13}$$

in rational integers $x_1, \ldots, x_m$, $z_1, \ldots, z_s$ *with* $(x_1, \ldots, x_m)|b$, $z_j \geqq 0$ *for* $j=1, \ldots, s$ *satisfy*

$$\max\left(|x_1|, \ldots, |x_m|, \quad (p_1^{z_1} \ldots p_s^{z_s})^{1/k(k-1)}\right) < |b| H^{2m-1} T_i \tag{14}$$

for $i=3, 4$, *where*

$$T_3 = \exp\left\{c_9\left[c_{10}(s+1)^{30}\left((|d|^{1/k} P^{s+1})^{3/2}(5^{3k}\log(|d|P))^2\right)^l\right]^{sl+4}\right\}$$

and

$$T_4 = \exp\left\{c_{11} P^l\left[c_{12}(s+1)^{30}(\log P)^2(|D_K|^{3/2k}\log|D_K|)^l\right]^{sl+4}\log|d|(\log\log|d|)^2\right\}$$

with $l=k(k-1)(k-2)$ *and effectively computable positive constants* $c_9, c_{10}, c_{11}, c_{12}$ *depending only on* k.

Corollary 2 implies the following

COROLLARY 3. *Let* K, $\alpha_1, \ldots, \alpha_m$, $p_1, \ldots, p_s$ *be defined as in Corollary 2. For any m-tuple of relatively prime rational integers* $x_1, \ldots, x_m$ *we have*

$$|D_{K/\mathbf{Q}}(\alpha_1 x_1+\ldots+\alpha_m x_m)| \prod_{j=1}^{s} |D_{K/\mathbf{Q}}(\alpha_1 x_1+\ldots+\alpha_m x_m)|_{p_j} > |\mathbf{x}|^{c_{13}(\log\log|\mathbf{x}|)^{-2}} \tag{15}$$

provided that $|\mathbf{x}| = \max_{1\leqq i\leqq m}(|x_i|) \geqq X_0$, *where* X_0 *and* c_{13} *are effectively computable positive numbers depending on* H, n, k, D_K, P *and* s, *but not on* $x_1, \ldots, x_m$.

Corollary 1 enables us to get some information about the arithmetical properties of those rational integers which can be represented by a given discriminant form.

COROLLARY 4. *Let* K, $\alpha_1, \ldots, \alpha_m$ *be as in Corollary 2 and let* D *be a rational integer with* s *distinct prime factors whose maximum is* P. *If there exist relatively prime rational integers* $x_1, \ldots, x_m$ *for which* $D_{K/\mathbf{Q}}(\alpha_1 x_1+\ldots+\alpha_m x_m)=D$, *then*

$$\log P + s\log(s+1) + s\log\log P > c_{14}\log\log|D|$$

provided that $|D| \geqq D_0$, *where* c_{14} *and* D_0 *are effectively computable positive constants which depend only on* H, k *and* D_K.

For simplicity Corollaries 3 and 4 are stated only for $L=\mathbf{Q}$, but in consequence of Theorem 1 they are also true in the general case when L is an arbitrary algebraic number field. Further, in view of $|D_K| \leqq (2H)^{k(k-1)^2}$ in Corollaries 3 and 4 we can derive the same estimates with constants independent of D_K.

Although our Theorem 1 remains valid for $s=0$, in this special case we can give slightly better upper bounds for the sizes of the solutions, in which all constants are explicitly computed.

THEOREM 2. *Let $L, K, \alpha_1, \ldots, \alpha_m$ and δ be as in Theorem* 1. *Then all solutions of the equation* (8) *in algebraic integers of L satisfy*

$$\max_{1\leqq i\leqq m}(|\overline{x_i}|) < |\overline{\delta}|^{\frac{1}{k(k-1)}}(3mH)^{mn-1}\exp\left\{(5nk^3)^{30nk^3}\left((dD_L^k)^{3/2}(\log(dD_L))^{nk}\right)^{\frac{3l}{k}}\right\} \tag{16}$$

and

$$\max_{1\leqq i\leqq m}(|\overline{x_i}|) < |\overline{\delta}|^{\frac{1}{k(k-1)}}(3mH)^{mn-1}\exp\left\{(5nl)^{30(nl+2)}\left(D_K(\log D_K)^{nk}\right)^{\frac{3l}{k}}(D_k^{3l/2k}+\log d)\right\} \tag{17}$$

where $l=k(k-1)(k-2)$.

Theorem 2 is a generalization of Théorème 7 of [9]. It implies that (8) has only finitely many solutions in integers $x_1, \ldots, x_m$ of L and all these solutions can be effectively determined.

In the special case $L=\mathbf{Q}$ (16) and (17) are sharper than (2) and (3).

If in particular K/L is normal, both (10) and (17) can be further sharpened. For example, by (28′) we get

$$\max_{1\leqq i\leqq m}(|\overline{x_i}|) < |\overline{\delta}|^{\frac{1}{k(k-1)}}(3mH)^{mn-1}\exp\left\{(5nk)^{30(nk+2)}D_K(\log D_K)^{3nk-1}(D_k^{1/2}+\log d)\right\} \tag{18}$$

in place of (17).

In (16), (17) and (18) one may take $d=|\overline{\delta}|^n$ and so for example (17) may be written in the form

$$\max_{1\leqq i\leqq m}(|\overline{x_i}|) < c_{15}H^{mn-1}|\overline{\delta}|^{c_{16}}$$

where the numbers c_{15}, c_{16} depend only on n, k and D_K and can be explicitly given.

Finally we remark that Theorem 2 is proved in our paper [13] as well as a special case of a result concerning generalized norm form equations (see [13], Theorem 4). However, the estimates of Theorem 2 are better than those obtained in [13]. In (16) and (17) and in the other estimates of the present paper the dependence on H is especially good.

An easy consequence of Theorem 2 is the following

COROLLARY 5. *Suppose $L, K, \alpha_1, \ldots, \alpha_m$ are as in Theorem* 2. *There exist effectively computable positive numbers $c_{17}=c_{17}(n, k, H, D_K)$, $c_{18}=c_{18}(n, k, D_K)$ such that*

$$|\overline{D_{K/L}(\alpha_1 x_1+\ldots+\alpha_m x_m)}| \geqq c_{17}\left\{\max_{1\leqq i\leqq m}(|\overline{x_i}|)\right\}^{c_{18}} \tag{19}$$

holds for any $(x_1, \ldots, x_m)\in\mathbf{Z}_L^m$.

If $L=\mathbf{Q}$, (19) is sharper than (15) in the special case $s=0$.

As we mentioned before, the earlier investigations on discriminant form and index form equations yielded results concerning algebraic integers with given discriminant or given index. We should note that in [9] and in the present paper such an argument would be circular.

3. Effective estimates for the integer solutions of index form equations

Let L and K be given algebraic number fields with the same properties as in Section 2, that is let $[L\colon \mathbf{Q}]=n\geqq 1$, $[K\colon L]=k\geqq 3$ and denote by D_L and D_K respectively the absolute values of their discriminants. Again, let h_L be the class number of L. Consider an order O of the field extension K/L (i.e. a subring of $\mathbf{Z}_K$ containing $\mathbf{Z}_L$ that has the full dimension k as a $\mathbf{Z}_L$-module) and suppose that O has a relative integral basis $1, \alpha_1, \ldots, \alpha_{k-1}$ over L. This holds for example in the following important special cases: (i) $L=\mathbf{Q}$, (ii) $O=\mathbf{Z}_K$ and $h_L=1$, (iii) $O=\mathbf{Z}_L[\alpha]$ for some $\alpha\in\mathbf{Z}_K$. Further results and many numerical examples and references can be found e.g. in BERWICK [4] and NARKIEWICZ [25].

Denote by $D_{K/L}(O)$ the principal ideal generated by $D_{K/L}(1, \alpha_1, \ldots, \alpha_{k-1})$ in L. Write $M(\mathbf{x})=x_1\alpha_1+\ldots+x_{k-1}\alpha_{k-1}$. We can easily see in the same way as in the case $L=\mathbf{Q}$ that the discriminant form $\mathrm{Discr}\,(M(\mathbf{x}))$ defined by (7) may be written in the form

$$D_{K/L}(\alpha_1 x_1+\ldots+\alpha_{k-1}x_{k-1}) = [F(x_1, \ldots, x_{k-1})]^2 D_{K/L}(1, \alpha_1, \ldots, \alpha_{k-1}) \tag{20}$$

where $F(x_1, \ldots, x_{k-1})\in\mathbf{Z}_L[x_1, \ldots, x_{k-1}]$ is a decomposable form of degree $k(k-1)/2$. It is called the *index form* of the basis, $1, \alpha_1, \ldots, \alpha_{k-1}$ of O over L.

There is an extensive literature of index forms and their applications; we refer the reader to the works by HENSEL [15], [16], [17], DELONE [5], NAGELL [21], [22], [23], DELONE and FADDEEV [6], NARKIEWICZ [25], PAYAN [26], GRAS [7], [8] and ARCHINARD [1] and thence to the literature there mentioned.

As is well-known, HENSEL's famous theorems on common index divisors (see, e.g., [15], [16], [17]) as well as their generalizations obtained by HASSE [14] and PLEASANTS [27] give necessary conditions for the solvability of index form equations (23) (see also (5)). In the important particular case $L=\mathbf{Q}$, $\delta=\pm 1$ (i.e. when $a=\pm 1$ in (5)) the solvability of these equations is settled for a great number of various special number fields K and orders O, and, in certain cases, all the solutions are explicitly given (see e.g. [5], [6], [23], [25], [26], [7], [8], [1]).

By virtue of (20) all the results enunciated in Section 2 can be stated for index forms in place of discriminant forms. As an immediate consequence of Theorem 1 we get the following p-adic generalization of the Corollaire to Théorème 7 of [9].

THEOREM 3. *Let $L, K, \pi_1, \ldots, \pi_s, \beta$ and δ be as in Theorem 1 and let O be an order in $\mathbf{Z}_K$ over $\mathbf{Z}_L$ having a relative integral basis $1, \alpha_1, \ldots, \alpha_{k-1}$ with $\max\limits_{1\leqq j\leqq k-1}(|\overline{\alpha_j}|)\leqq H$ and with index form $F(x_1, \ldots, x_{k-1})$. Then for every solution of the equation*

$$F(x_1, \ldots, x_{k-1}) = \delta\pi_1^{z_1}\ldots\pi_s^{z_s} \tag{21}$$

in integers $x_1, \ldots, x_{k-1}\in\mathbf{Z}_L$, $z_1, \ldots, z_s\in\mathbf{Z}$ with $(x_1, \ldots, x_{k-1})\mid(\beta)$, $z_i\geqq 0$ for $i=1, \ldots, s$ there exists a unit ε in L such that

$$\max\left\{|\overline{x_1\varepsilon}|, \ldots, |\overline{x_{k-1}\varepsilon}|, (p_1^{f_1 z_1}\ldots p_s^{f_s z_s})^{2h_L/nk(k-1)}\right\} < \tag{22}$$

$$< b^{1/n} H^{(k-1)(n+1)-1} \exp\left\{4c_5 D_L^{1/2} P^{nl}\left[c_6(s+1)^{30}(\log P)^2\left(D_K^{3/2k}(\log D_K)^n\right)^l\right]^{sl+4}\cdot\right.$$

$$\left.\cdot\log\left(dN(D_{K/L}(O))\right)\left[\log\log\left(dN(D_{K/L}(O))\right)\right]^2\right\}$$

with the constants $l=k(k-1)(k-2)$, c_5 and c_6 occurring in Theorem 1.

Theorem 3 remains valid even for $s=0$. However, in this special case we can obtain a more precise result by using Theorem 2. Indeed, in view of $|D_{K/L}(1, \alpha_1, \ldots, \alpha_{k-1})| \leqq k^k H^{2(k-1)}$ an easy consequence of Theorem 2 is as follows.

THEOREM 4. *Let L, K, δ be defined as in Theorem 2 and let O be an order in $\mathbf{Z}_K$ over $\mathbf{Z}_L$ having a relative integral basis $1, \alpha_1, \ldots, \alpha_{k-1}$ with $\max_{1 \leqq j \leqq k-1} (|\overline{\alpha_j}|) \leqq H$ and with index from $F(x_1, \ldots, x_{k-1})$. Then all solutions of the equation*

$$F(x_1, \ldots, x_{k-1}) = \delta \tag{23}$$

in $x_1, \ldots, x_{k-1} \in \mathbf{Z}_L$ satisfy

$$\max_{1 \leqq i \leqq k-1} (|\overline{x_i}|) < |\overline{\delta}|^{\frac{2}{k(k-1)}} (3kH)^{n(k-1)} \exp\left\{(5nl)^{30(nl+2)} \left(D_K (\log D_K)^{nk}\right)^{3l/k} \cdot \right. \\ \left. \cdot \left(D_k^{3l/2k} + \log (d^2 N(D_{K/L}(O)))\right)\right\}, \tag{24}$$

where $l = k(k-1)(k-2)$.

Theorem 4 generalizes the Corollaire to Théorème 7 of [9], improves (6) in the special case $L=\mathbf{Q}$ and implies that (23) has only finitely many solutions in integers $x_1, \ldots, x_{k-1}$ of L and all these solutions can be effectively determined.

By applying the other estimates of Section 2 it is easy to obtain upper bounds in (22) and (24) which do not depend on D_K.

In (22) and (24) one may take $d = |\overline{\delta}|^n$ and $N(D_{K/L}(O))$ can be estimated from above by $(k^k H^{2(k-1)})^n$. Therefore we get for example

$$\max_{1 \leqq i \leqq k-1} (|\overline{x_i}|) < c_{19} (H |\overline{\delta}|)^{c_{20}}$$

for every solution $(x_1, \ldots, x_{k-1}) \in \mathbf{Z}_L^{k-1}$ of (23), where the numbers $c_{19}, c_{20} > 0$ can be explicitly given in terms of n, k and D_K.

Finally we remark that Theorems 3 and 4 can be derived directly from the results obtained in [11] and [12] on algebraic integers with given index, too.

4. Preliminary results

We keep the notations of Section 2. Let S denote the multiplicative semigroup generated by $\pi_1, \ldots, \pi_s$ and U_L, U_L being the group of units of L. Obviously $S \subset \mathbf{Z}_L$, and if $s=0$, then $S = U_L$. Again, let δ be a non-zero integer in L with $|N_{L/\mathbf{Q}}(\delta)| \leqq d$ $(d \geqq 3)$.

We say that the algebraic integers α, α^* are $\mathbf{Z}_L$-*equivalent* if $\alpha - \alpha^* \in \mathbf{Z}_L$.

As we mentioned in the introduction, the proofs of our theorems are based on the following theorems which were obtained in [11] and [12] as corollaries of more general results concerning polynomials with given discriminant.

LEMMA 1. *Let L, S, δ be defined as above and let α be an algebraic integer with degree $k \geqq 3$ and discriminant $D(\alpha) \in \delta S$ over L. Then α is $\mathbf{Z}_L$-equivalent to an integer of the form $\eta \alpha^*$, where $\eta \in S$ and α^* is an integer satisfying*

$$|\overline{\alpha^*}| < \exp\left\{c_{21}\left[c_{22}(s+1)^{30}\left((D_L d^{1/k} P^{(s+1)n})^{3/2} (5^{skn} \log (D_L\, dP))^{n+1}\right)^l\right]^{sl+4}\right\} \tag{25}$$

with $l=k(k-1)(k-2)$ *and effectively computable positive constants* c_{21}, c_{22} *depending only on* n *and* k.

PROOF. This is Theorem 2 of [12].

Very recently TRELINA [40] has proved the above assertion in the special case $L=\mathbf{Q}$, $\delta=1$. Her estimate is weaker than (25) with $n=D_L=d=1$.

LEMMA 2. *Let* L, S, δ *be as in Lemma* 1 *and let* K *be an algebraic number field of degree* $k\geqq 3$ *over* L *with discriminant* $D_K=D_{K/\mathbf{Q}}$. *If* α *is an integer in* K *with* $D_{K/L}(\alpha)\in\delta S$, *then it is* $\mathbf{Z}_L$*-equivalent to an integer of the form* $\eta\alpha^*$, *where* $\eta\in S$ *and* α^* *is an integer such that*

(26)

$$\overline{|\alpha^*|} < \exp\left\{c_{23}D_L^{1/2}P^{nl}\left[c_{24}(s+1)^{30}(\log P)^2\left(|D_K|^{3/2k}(\log|D_K|)^n\right)^l\right]^{sl+4}\log d(\log\log d)^2\right\}$$

with $l=k(k-1)(k-1)$ *and with effectively computable positive constants* c_{23}, c_{24} *which depend only on* n *and* k.

PROOF. See the remark after the enunciation of Theorem 2 of [12].

The exponents of $(s+1)$ occurring in (25) and (26) are not explicitly given in [12]. However, replacing Theorems A and B [5] in [12] by Theorem 3 of [30] (see also [31]) and Theorem 4 of [28] respectively, we may take $c^*=30$ as exponent of $(s+1)$ throughout the paper [12]. In this case we get (26), (10), (14) without $(\log\log d)^2$.

LEMMA 3. *Let* L, δ *be as in Lemma* 1. *If* α *is an algebraic integer with degree* $k\geqq 3$ *and discriminant* δ *over* L, *then there exists an* α^* $\mathbf{Z}_L$*-equivalent to* α *such that*

$$\overline{|\alpha^*|} < \overline{|\delta|}^{\frac{1}{k(k-1)}}\exp\left\{(5nk^3)^{30nk^3}\left((dD_L^k)^{3/2}(\log dD_L)^{nk}\right)^{3(k-1)(k-2)}\right\}. \tag{27}$$

PROOF. This is Theorem 3A of [11].

LEMMA 4. *Let* L, δ *be defined as in Lemma* 1 *and let* K *be an algebraic number field of degree* $k\geqq 3$ *over* L *with discriminant* $D_K=D_{K/\mathbf{Q}}$. *If* α *is an integer in* K *with* $D_{K/L}(\alpha)=\delta$, *then there exists an* α^* $\mathbf{Z}_L$*-equivalent to* α *for which*

$$\overline{|\alpha^*|} < \overline{|\delta|}^{\frac{1}{k(k-1)}}\exp\left\{(5nl)^{30(nl+2)}\left(|D_K|(\log|D_K|)^{nk}\right)^{3l/k}\left(|D_K|^{3l/2k}+\log d\right)\right\} \tag{28}$$

holds with $l=k(k-1)(k-2)$.

PROOF. This is Theorem 3B in [11].

If in particular the extension K/L is normal, by estimate (8') of [11] we get the following better estimate

$$\overline{|\alpha^*|} < \overline{|\delta|}^{\frac{1}{k(k-1)}}\exp\left\{(5nk)^{30(nk+2)}|D_K|(\log|D_K|)^{3nk-1}\left(|D_K|^{1/2}+\log d\right)\right\} \tag{28'}$$

in place of (28).

[5] These very deep theorems are due to BAKER [3] and VAN DER POORTEN [28] (see also [38]).

We remark that the proofs of Lemmas 3 and 4 depend on a recent inequality of VAN DER POORTEN and LOXTON ([29], [30], [31]) for linear forms in the logarithms of algebraic numbers. Very recently WALDSCHMIDT [41] and Waldschmidt and van der Poorten (private communication) have obtained sharper estimates in terms of certain parameters.

5. Proofs

PROOF OF THEOREM 1. Consider an arbitrary solution of (9) in $x_1, \dots, x_m \in \mathbf{Z}_L$, $z_1, \dots, z_s \in \mathbf{Z}$ with $(x_1, \dots, x_m) | (\beta)$, $z_r \geqq 0$ for $r = 1, \dots, s$. Put

$$\alpha = \alpha_1 x_1 + \dots + \alpha_m x_m. \tag{29}$$

Obviously, α is a primitive integral element of K/L and

$$D_{K/L}(\alpha) = \delta \pi_1^{z_1} \dots \pi_s^{z_s}. \tag{30}$$

By Lemmas 1 and 2 α may be written in the form

$$\alpha = -x_0 + \alpha^* \varepsilon_1 \pi_1^{u_1} \dots \pi_s^{u_s}, \tag{31}$$

where $x_0 \in \mathbf{Z}_L$, $\varepsilon_1 \in U_L$, $u_1, \dots, u_s$ are non-negative rational integers and α^* satisfies

$$\overline{|\alpha^*|} < T_1' \quad \text{and} \quad \overline{|\alpha^*|} < T_2', \tag{32}$$

T_1', T_2' being the expressions occurring on the right sides of (25) and (26) respectively. Denote by

$$\alpha'^{(i)} = x_0 + \alpha_1^{(i)} x_1 + \dots + \alpha_m^{(i)} x_m = \alpha^{*(i)} \varepsilon_1 \pi_1^{u_1} \dots \pi_s^{u_s}, \quad i = 1, \dots, k, \tag{33}$$

the conjugates of $\alpha' = \alpha + x_0$ over L. Since $1, \alpha_1, \dots, \alpha_m$ are linearly indepenent over L, there are $i_0, \dots, i_m$ such that

$$\sigma = \begin{vmatrix} 1 & \alpha_1^{(i_0)} & \dots & \alpha_m^{(i_0)} \\ \vdots & & & \\ 1 & \alpha_1^{(i_m)} & \dots & \alpha_m^{(i_m)} \end{vmatrix} \neq 0. \tag{34}$$

Replacing the elements of the $(j+1)$-th column of σ by $\alpha^{*(i_0)}, \dots, \alpha^{*(i_m)}$, we get a determinant, say σ_j. From (33) we obtain

$$x_j \varepsilon_1^{-1} = \pi_1^{u_1} \dots \pi_s^{u_s} \sigma_j / \sigma \tag{35}$$

for $j = 1, \dots, m$. Let K^* be the smallest normal extension of L containing K and put $[K^* : L] = t$. Writing $\pi^{u_1} \dots \pi^{u_s} = \tau$, $(x_1, \dots, x_m) | (\beta)$ implies

$$|N_{K^*/\mathbf{Q}}(\tau)| \cdot N((\sigma_1, \dots, \sigma_m)) = |N_{K^*/\mathbf{Q}}(\sigma)| N((x_1, \dots, x_m)),$$

whence

$$|N_{L/\mathbf{Q}}(\tau)| \leqq |N_{K^*/\mathbf{Q}}(\sigma)|^{1/t} |N_{L/\mathbf{Q}}(\beta)|.$$

As is known, there are $\tau' \in \mathbf{Z}_L$ and $\varepsilon_2 \in U_L$ such that $\tau = \tau' \varepsilon_2$ and

$$\overline{|\tau'|} \leqq |N_{L/\mathbf{Q}}(\tau)|^{1/n} \exp\{c_{25} D_L^{1/2} (\log D_L)^{n-1}\} \leqq \tag{36}$$

$$\leqq b^{1/n} \overline{|\sigma|} \exp\{c_{25} D_L^{1/2} (\log D_L)^{n-1}\}$$

with an effectively computable positive constant c_{26} depending only on n.

Since all conjugates of σ and σ_j over $\mathbf{Q}$ can be written as determinants of the same form as above, hence

$$\overline{|\sigma|} \leqq \sqrt{m+1}\,(\sqrt{(m+1)H^2})^m = (m+1)^{\frac{m+1}{2}} H^m \tag{37}$$

and similarly

$$\overline{|\sigma_j|} \leqq (m+1)^{\frac{m+1}{2}} H^{m-1} T_i' \tag{38}$$

for $j = 1, \ldots, m$ and $i = 1, 2$. Putting $(\varepsilon_1 \varepsilon_2)^{-1} = \varepsilon$, from (35) we get

$$x_j' = x_j \varepsilon = \tau' \sigma_j / \sigma$$

for $j = 1, \ldots, m$. This implies

$$x_j'^t = N_{K^*/L}(x_j') = N_{K^*/L}(\tau \sigma_j) / N_{K^*/L}(\sigma),$$

whence

$$\overline{|x_j'|^t} \leqq \overline{|N_{K^*/L}(\tau' \sigma_j)|}\, \overline{|N_{K^*/L}(\sigma)|}^{n-1} \leqq \overline{|\tau' \sigma_j|^t}\, \overline{|\sigma|}^{t(n-1)}$$

and by (36), (37) and (38) we obtain

$$\overline{|x_j'|} \leqq \overline{|\tau'|}\, \overline{|\sigma_j|}\, \overline{|\sigma|}^{n-1} \leqq b^{1/n} H^{m(n+1)-1} T_i, \quad j = 1, \ldots, m,$$

with

$$T_i = T_i'^{c_{26}}$$

for $i = 1, 2$, where $c_{26} > 1$ denotes a suitable effectively computable constant depending only on n and k.

From (30) and (31) it follows

$$D_{K/L}(\alpha^*)(\varepsilon_1 \tau)^{k(k-1)} = \delta \pi_1^{z_1} \ldots \pi_s^{z_s}$$

and this yields

$$(p_1^{f_1 z_1} \ldots p_s^{f_s z_s})^{h_L} = |N_{L/\mathbf{Q}}(\pi_1^{z_1} \ldots \pi_s^{z_s})| \leqq |N_{L/\mathbf{Q}}(\tau)|^{k(k-1)} \left|N_{L/\mathbf{Q}}\left(D_{K/L}(\alpha^*)\right)\right| \leqq$$

$$\leqq (\overline{|\sigma|}^n b)^{k(k-1)} (2T_i')^{k(k-1)n}.$$

Thus we get

$$(p_1^{f_1 z_1} \ldots p_s^{f_s z_s})^{h_L / nk(k-1)} \leqq b^{1/n} H^m T_i$$

for $i = 1, 2$. Since T_1 and T_2 have the required forms, Theorem 1 is proved.

PROOF OF COROLLARY 1. Consider the prime ideal decomposition

$$\left(D_{K/L}(\alpha_1 x_1 + \ldots + \alpha_m x_m)\right) = \mathfrak{p}^{u_1} \ldots \mathfrak{p}^{u_s}$$

in L. Put $u_j = h_L z_j + r_j$, $0 \leqq r_j < h_L$ and $\mathfrak{p}_j^{h_L} = (\pi_j)$ with $\pi_j \in \mathbf{Z}_L$ for $j = 1, \ldots, s$. Then we have

$$D_{K/L}(\alpha_1 x_1 + \ldots + \alpha_m x_m) = \delta \pi_1^{z_1} \ldots \pi_s^{z_s} \tag{39}$$

with a suitable integer δ in L for which $(\delta)=\mathfrak{p}_1^{r_1}\dots\mathfrak{p}_s^{r_s}$ and $|N_{L/\mathbf{Q}}(\delta)|\leqq\mathscr{P}^{sn(h_L-1)}$ hold, where $\mathscr{P}=\max_j N(\mathfrak{p}_j)$. Applying now the estimate (10) of Theorem 1 with T_2 to (39), we get (11). Since $s\leqq\pi_K(\mathscr{P})\leqq c_{27}\mathscr{P}/\log\mathscr{P}$ with an effectively computable positive constant c_{27}, hence (12) follows at once from (11).

PROOF OF THEOREM 2. Let $(x_1,\dots,x_m)\in\mathbf{Z}_L^m$ be an arbitrary solution of (8). Then

$$\alpha=\alpha_1x_1+\dots+\alpha_mx_m$$

is a primitive integral element of K/L and

$$D_{K/L}(\alpha)=\delta.$$

Denote by T_5 and T_6 respectively the expressions occurring on the right sides of (27) and (28). By Lemmas 3 and 4 we have

$$\alpha=-x_0+\alpha^*$$

with a suitable $x_0\in\mathbf{Z}_L$, where

$$\overline{|\alpha^*|}<T_5 \quad\text{and}\quad \overline{|\alpha^*|}<T_6.$$

Following the proof of Theorem 1, we get now in the same way as there that

$$x_j=\sigma_j/\sigma$$

for $j=1,\dots,m$, whence

$$\overline{|x_j|}\leqq\overline{|\sigma_j|}\,\overline{|\sigma|}^{n-1}\leqq(3mH)^{mn-1}T_i$$

for $j=1,\dots,m$ and $i=5,6$. This completes the proof of Theorem 2.

Added in proof. The results of the present paper have recently been generalized by the first author (see *Ann. Acad. Sci. Fenn.*, Ser. A. I, **4** (1979), 341—355, and „Explicit upper bounds for the solutions of some diophantine equations", ibid., to appear).

REFERENCES

[1] ARCHINARD, G.: Extensions cubiques cycliques de Q dont l'anneau des entiers est monogène, *Enseignement Math.* **20** (1974), 179—203.
[2] BAKER, A.: Contributions to the theory of diophantine equations, *Philos. Trans. Roy. Soc. London*, Ser. A. **263** (1968), 173—208.
[3] BAKER, A.: The theory of linear forms in logarithms, *Proc. Conf. Transcendence, Cambridge, 1976, Advances in transcendence theory* (Edited by A. Baker and D. W. Masser), Academic Press, London and New York, pp. 1—27, 1977.
[4] BERWICK, W. E. H.: *Integral bases*, Reprinted by Stechert—Hafner Service Agency, New York and London, 1964.
[5] DELONE, B. N. (DELAUNAY): Über die Darstellung der Zahlen durch die binären kubischen Formen von negativer Diskriminante, *Math. Z.* **31** (1930), 1—26.
[6] DELONE, B. N. and FADDEEV, D. K.: *The theory of irrationalities of the third degree*, Amer. Math. Soc., Providence, 1964 (Translated from the Russian (1940) ed.)
[7] GRAS, M.-N.: Sur les corps cubiques cycliques dont l'anneau des entiers est monogène, *Ann. Sc. Univ. Besançon*, fasc. 6 (1973).

[8] GRAS, M.-N.: Lien entre le groupe des unités et la monogenéité des corps cubiques cycliques. *Publ. Math. Univ. Besançon*, fasc. 1 (1975—76).
[9] GYŐRY, K.: Sur les polynômes à coefficients entiers et de discriminant donné, III., *Publ. Math. Debrecen* **23** (1976), 141—165.
[10] GYŐRY, K.: Polynomials with given discriminant, *Coll. Math. Soc. János Bolyai 13, Debrecen, 1974. Topics in number theory* (Edited by P. Turán), North-Holland Publ. Comp., Amsterdam—Oxford—New York, 1976, pp. 65—78.
[11] GYŐRY, K.: On polynomials with integer coefficients and given discriminant, IV., *Publ. Math. Debrecen* **25** (1978), 155—167.
[12] GYŐRY, K.: On polynomials with integer coefficients and given discriminant, V., p-adic generalizations, *Acta Math. Acad Sci. Hung.* **32** (1978), 175—190.
[13] GYŐRY, K. and PAPP, Z. Z.: Effective estimates for the integer solutions of norm form and discriminant form equations, *Publ. Math. Debrecen* **25** (1978), 311—325.
[14] HASSE, H.: *Zahlentheorie*, Zweite Auflage, Akademie Verlag, Berlin, 1963.
[15] HENSEL, K.: Untersuchung der Fundamentalgleichung einer Gattung für eine reelle Primzahl als Modul und Bestimmung der Theiler ihrer Discriminante, *J. Reine Angew. Math.* **113** (1894), 61—83.
[16] HENSEL, K.: Arithmetische Untersuchungen über die gemeinsamen ausserwesentlichen Discriminantentheiler einer Gattung, *J. Reine Angew. Math.* **113** (1894), 128—160.
[17] HENSEL, K.: *Theorie der algebraischen Zahlen*, Teubner Verlag, Leipzig und Berlin, 1908.
[18] KOTOV, S. V.: The Thue-Mahler equation in relative fields (Russian), *Acta Arith.* **27** (1975), 293—315.
[19] KRONECKER, L.: Grundzüge einer arithmetischen Theorie der algebraischen Grössen, *J. Reine Angew. Math.* **92** (1882), 1—122.
[20] NAGELL, T.: *L'analyse indéterminée de degré supérieur*, Mémor. Sci. Math. Vol. 39, Paris, 1929.
[21] NAGELL, T.: Zur Theorie der kubischen Irrationalitäten, *Acta Math.* **55** (1930), 33—65
[22] NAGELL, T.: Quelques résultats sur les diviseurs fixes de l'index des nombres entiers d'un corps algébrique, *Arkiv för Mat.* **6** (1965), 269—289.
[23] NAGELL, T.: Sur les discriminants des nombres algébriques, *Arkiv för Mat.* **7** (1967), 265—282.
[24] NAGELL, T.: Quelques propriétés des nombers algébriques du quatrième degré, *Arkiv för Mat.* **7** (1969), 517—525.
[25] NARKIEWICZ, W.: *Elementary and analytic theory of algebraic numbers*, Polish Scientific Publishers, Warszawa, 1974.
[26] PAYAN, J. J.: Sur les classes ambiges et les ordres monogènes d'une extension cyclique de degré premier impair sur Q ou sur un corps quadratique imaginaire, *Arkiv för Mat.* **11** (1973), 239—244.
[27] PLEASANTS, P. A. B.: The number of generators of the integers of a number field, *Mathematika* **21** (1974), 160—167.
[28] VAN DER POORTEN, A. J.: Linear forms in logarithms in the p-adic case, *Proc.Conf. Transcendence, Cambridge, 1976. Advances in transcendence theory* (Edited by A. Baker and D. W. Masser), Academic Press, London and New York, pp. 29—57, 1977.
[29] VAN DER POORTEN, A. J. and LOXTON, J. H.: Computing the effectively computable bound in Baker's inequality for linear forms in logarithms, *Bull. Austral. Math. Soc.* **15** (1976), 33—57.
[30] VAN DER POORTEN, A. J. and LOXTON, J. H.: Multiplicative relations in number fields, *Bull. Austral. Math. Soc.* **16** (1977), 83—98.
[31] VAN DER POORTEN, A. J. and LOXTON, J. H.: Computing the effectively computable bound in Baker's inequality for linear forms in logarithms, and, Multiplicative relations in number fields: Corrigendum and addendum, *ibid.* **17** (1977), 151—155.
[32] SCHMIDT, W. M.: Linearformen mit algebraischen Koeffizienten, II., *Math. Ann.* **191** (1971) 1—20.
[33] SCHMIDT, W. M.: Norm form equations, *Annals of Math.* **96** (1972), 526—551.
[34] SHOREY, T. N., VAN DER POORTEN, A. J., TIJDEMAN, R. and SCHINZEL, A.: Applications of the Gelfond—Baker method to diophantine equations, *Proc. Conf. Transcendence, Cambridge, 1976. Advances in transcendence theory* (Edited by A. Baker and D. W. Masser), Academic Press, London and New York, pp. 59—77, 1977.
[35] SPRINDŽUK, V. G. and KOTOV, S. V.: An effective analysis of the Thue—Mahler equation in relative fields, (Russian), *Dokl. Akad. Nauk. BSSR* **17** (1973), 393—395.

[36] SPRINDŽUK, V. G. and KOTOV, S. V.: On approximation to algebraic numbers by algebraic numbers of a given field (Russian), *Dokl. Akad. Nauk. BSSR* **20** (1976), 581—584.
[37] STARK, H. M.: Further advances in the theory of linear forms in logarithms *Proc. Conf. held in Washington, 1972. Diophantine approximation and its applications* (Edited by C. F. Osgood), pp. 255—294, Academic Press, New York and London, 1973.
[38] STEWART, C. L.: *Divisor properties of arithmetical sequences,* Ph. D. Thesis, University of Cambridge, Cambridge, 1976.
[39] THUE, A.: Über Annäherungswerte algebraischer Zahlen, *J. Reine Angew. Math.*, **135** (1909), 284—305.
[40] TRELINA, L. A.: On algebraic integers with discriminants containing fixed prime divisors (Russian), *Mat. Zametki* **21** (1977), 289—296.
[41] WALDSCHMIDT, M.: A lower bound for linear forms in logarithms, *to appear.*

Mathematical Institute, Kossuth Lajos University, Debrecen, Hungary

(Received July 18, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 61—63

QUASI-HERMITE—FEJÉR INTERPOLATION

by

T. M. MILLS

1. Introduction

In 1959, PAUL SZÁSZ [6] introduced the quasi-Hermite—Fejér interpolation (QHFI) polynomials. These are defined in the following way.

Let $x_1, x_2, \ldots, x_n$ be n distinct interior points of the interval $I=[-1, 1]$. Then, for any function f: $I\to(-\infty, \infty)$, the QHFI polynomial is the unique polynomial $Q_{2n+1}(f, x)$ of degree $2n+1$ (or less) such that

$$(1)\qquad \begin{cases} Q_{2n+1}(f, x_k)=f(x_k), & k=1, 2, \ldots, n, \\ Q_{2n+1'}(f, x_k)=0, & k=1, 2, \ldots, n, \\ Q_{2n+1}(f, +1)=f(+1), \\ Q_{2n+1}(f, -1)=f(-1). \end{cases}$$

The case in which $x_1, x_2, \ldots, x_n$ are the zeros of certain Jacobi polynomials has been discussed in several papers. From the results of SZÁSZ [6], SÁNTA [5], POVCUN and PRIVALOV [4], and MILLS [3], we can compose the following result.

THEOREM 1. *Let $Q_{2n+1}(f, x)$ be the QHFI polynomial, based on the zeros of the Jacobi polynomial $P_n^{(\alpha,\beta)}(x)$, for a function $f\in C(I)$. If either* (i) $0\leqq\alpha<1$ *and* $0\leqq\beta<1$ *or* (ii) $\alpha=\beta=-1/2$, *then $Q_{2n+1}(f, x)$ converges to $f(x)$ uniformly on I.*

The aim of this paper is to extend this theorem to the cases $\alpha=-\beta=1/2$, and $\alpha=-\beta=-1/2$.

Let $f\in C(I)$ and ω the modulus of continuity of f. Now the main result of this paper may be stated.

THEOREM 2. *Let $Q_{2n+1}(f, x)$ be the QHFI polynomial, based on the zeros of $P_n^{(\frac{1}{2}, -\frac{1}{2})}(x)$, for a function $f\in C(I)$. Then, for $n\geqq n_1$,*

$$\|Q_{2n+1}(f)-f\| \leqq c_1 n^{-1} \sum_{k=1}^{n} \omega(f, 1/k)$$

where $\|\cdot\|$ denotes the uniform norm on I, and c_1 (and later $c_2, c_3, \ldots$) denotes a constant which is independent of n and f.

From the properties of $\omega(f, \delta)$ (see e.g. G. G. LORENTZ [2], pp. 43—44) we find that

$$n^{-1} \sum_{k=1}^{n} \omega(f, 1/k) \leqq c_2 \omega(f, (\log n)/n).$$

Thus, $Q_{2n+1}(f)$ converges to f uniformly on I and Theorem 1 is extended.

It will be clear from the proof that a similar theorem can be proved when $\alpha=-\beta=-1/2$.

I understand that P. VÉRTESI [8] has recently extended Theorem 1 to the cases where α, β satisfy both $-0.5\leqq\alpha, \beta<1$ and $|\alpha-\beta|\leqq 1$.

2. The Proof of Theorem 2

In a later paper, PAUL SZÁSZ [7] introduced another interpolation process. Consider the unique polynomial $R_{2n}(f, x)$ of degree $2n$, or less, such that

$$\text{(2)}\qquad \begin{cases} R_{2n}(f, x_k)=f(x_k), & k=1, 2, \ldots, n, \\ R'_{2n}(f, x_k)=0, & k=1, 2, \ldots, n, \\ R_{2n}(f, +1)=f(+1). \end{cases}$$

Szász proved that, under the assumptions of Theorem 2, $R_{2n}(f, x)$ converges to f uniformly on I as $\to\infty$. In a more recent paper, KUMAR and MATHUR [1] have shown that, for $n\geqq 2$,

$$\text{(3)}\qquad \|R_{2n}(f)-f\| \leqq c_3 n^{-1} \sum_{k=1}^{n} \omega(f, 1/k).$$

Now it follows from (a) the fact that the conditions (1) *uniquely* define $Q_{2n+1}(f, x)$ and (b) the properties (2) of $R_{2n}(f)$ that

$$\text{(4)}\qquad Q_{2n+1}(R_{2n}(f), x) \equiv R_{2n}(f, x).$$

In [8], VÉRTESI proves that $\{Q_{2n+1}: n=1, 2, 3, \ldots\}$ is a uniformly bounded sequence of linear operators.

Now we can prove Theorem 2. From (3), (4), and the uniform boundedness of Q_{2n+1}, we have

$$\|Q_{2n+1}(f)-f\| \leqq \|Q_{2n+1}(f)-Q_{2n+1}(R_{2n}(f))\|+\|Q_{2n+1}(R_{2n}(f))-R_{2n}(f)\|+$$

$$+\|R_{2n}(f)-f\| \leqq (\|Q_{2n+1}\|+1)\|R_{2n}(f)-f\| \leqq c_4 n^{-1} \sum_{k=1}^{n} \omega(f, 1/k).$$

3. Lower bounds for the error

Since Q_{2n+1} is not a positive operator, it is difficult to find general lower bounds for $\|Q_{2n+1}(f)-f\|$. However, examination of the error for a particular function shows us that Theorem 2 is best possible.

Let $g(x)=(1-x^2)|x|$, $-1\leqq x\leqq+1$. Now, from Theorem 2 and the properties of $\omega(g, \delta)$ we can obtain

$$\|Q_{2n+1}(g)-g\| \leqq c_{13}(\log n)/n, \quad n\geqq n_1.$$

On the other hand,

$$\|Q_{2n+1}(g)-g\| \geqq |Q_{2n+1}(g,0)-g(0)| = \left|\sum_{k=1}^{n} f(x_k)q_k(0)\right| =$$

$$= 2(2n+1)^{-2}\sum_{k=1}^{n}(|x_k|^{-1}-2|x_k|+|x_k|\cdot x_k) \geqq 2(2n+1)^{-2}\sum_{k=1}^{n}(|x_k|^{-1})-6n(2n+1)^{-2} \geqq$$

$$\geqq 2(2n+1)^{-2}\sum_{k=1}^{m}(|x_k|^{-1})-3(2n+1)^{-1} \geqq$$

$$\geqq \frac{1}{\pi(2n+1)}\int_0^{t_m}\sec t\,dt-3/(2n+1) \geqq (6\pi)^{-1}(\log n)/n$$

for $n=2m$ sufficiently large, where $t_m=\cos x_m$. Therefore we have

$$c_{14}\frac{\log n}{n} \leqq \|Q_{2n+1}(g)-g\| \leqq c_{13}\frac{\log n}{n}$$

for n even and sufficiently large. It is in this sense that Theorem 2 is best possible.

REFERENCES

[1] KUMAR, V. and MATHUR, K. K.: On the rapidity of convergence of a quasi-Hermite—Fejér interpolation polynomial, *Studia Sci. Math. Hung.* **9** (1974), 313—319.

[2] LORENTZ, G. G.: *Approximation of Functions*, Holt, Rinehart, and Winston (New York), 1966.

[3] MILLS, T. M.: A convergent quasi-Hermite—Fejér interpolation process, *Bull. Austral. Math. Soc.* **12** (1975), 267—276.

[4] POVCUN, L. P. and PRIVALOV, A. A.: A certain interpolation process of E. Egerváry and P. Turán, (in Russian), *Izv. Vyss. Učebn. Zaved. Matematika* **8 (147)** (1974), 82—88.

[5] SÁNTA, J.: Convergence theorems of quasi-Hermite—Fejér interpolation, *Publ. Math. Debrecen* **22** (1975), 23—29.

[6] SZÁSZ, P.: On quasi-Hermite—Fejér interpolation, *Acta Math. Acad. Sci. Hungar.* **10** (1959), 413—439.

[7] SZÁSZ, P.: The extended Hermite—Fejér interpolation formula with application to the theory of generalized almost-step parabolas, *Publ. Math. Debrecen.* **11** (1964), 85—100.

[8] VÉRTESI, P.: Hermite—Fejér type interpolation III, *Acta Math. Acad. Sci. Hungar.* (to appear).

Department of Mathematics, Bendigo College of Advanced Education, Bendigo, Victoria, Australia

(Received August 29, 1977; revised September 29, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 65—75

ON THE OPERATIONAL SOLUTION OF A CONVOLUTION TYPE INTEGRAL EQUATION OF THE THIRD KIND

by

T. FÉNYES

Introduction

In papers [1], [2] we have solved the integral equation

$$(t+a)f(t)+\int_0^t f(\tau)g(t-\tau)\,d\tau = h(t), \quad t \geqq 0, \tag{1}$$

applying Mikusiński's operational method. We have shown that (1) is solvable in the operator field for every real a and for every locally integrable $g(t)$, $h(t)$, if the quantity $g(0)=\lim_{t\to+0} g(t)$ exists and if the function $\frac{g(t)-g(0)}{t}$ is also locally integrable. We proved that under the above conditions, every solution of the corresponding algebraic differential equation

$$Df-af-gf=-h \tag{2}$$

is a finite order distribution having a support bounded on the left (see FENYŐ [5], WLOKA [6]). We have proved in [1], [2] that the corresponding homogeneous equation has nontrivial locally integrable solutions, if and only if, $a\leqq 0$, $g(0)<0$. In the case of $a>0$ (1) has exactly one locally integrable solution (this follows immediately from the Volterra theory) which can be written in the following form:

$$f(t) = U(t)+U(t)*G(t) \tag{3}$$

where $G(t) = \sum_{i=1}^{\infty} \left\{\frac{g(t)-g(0)}{-t}\right\}^i \frac{1}{i!}$

and

$$U(t) = \frac{h+h*G_0}{t+a} - g(0)(t+a)^{-g(0)-1}\int_0^t \frac{h+h*G_0}{(\tau+a)^{-g(0)+1}}\,d\tau$$

where $G_0(t) = \sum_{i=1}^{\infty} \left\{\frac{g(t)-g(0)}{t}\right\}^i \frac{1}{i!}$

In the case of $a=0$ (1) has no integrable solution in general. We have shown that the local integrability of the function $\frac{h(t)}{t}$ guarentees the existence of a locally integrable solution of (1). The solution formula (3) also holds for $a=0$, $g(0)\geqq 0$ if $\frac{h(t)}{t}$ is integrable. Moreover (3) also holds for $a=0$, $g(0)<0$ if $\frac{h(t)}{t}$ is integrable,

and if in (3) $\int_0^t \ldots$ is replaced by $\int_\varepsilon^t \ldots$ where $\varepsilon>0$ is arbitrary. (So we obtain infinitely many locally integrable solutions.)

In the more interesting case of $a<0$, the inhomogeneous equation (1) has not been discussed in [1], [2]. It is obvious that (1) is not solvable in general and it can be easily seen that the local integrability of $\frac{h(t)}{t+a}$ is neither necessary, nor sufficient to the solvability of (1). Moreover the following question arises: What is the explicit form of a solution of (1)? Assuming the existence of an integrable solution of a (1) type integral equation, it is obvious that the solution formula (3) does not hold for $a<0$ in general, since $(t+a)^{g(0)}$ has no meaning for not an integer $g(0)$ and letting $g(0)$ to be an integer we see that the integral occurring in $U(t)$ can have no meaning for $g(0)\leqq 0$.

In the first chapter of this paper we shall discuss (1) in the case of $a<0$. For $g(0)=0$ a simple necessary and sufficient solvability condition will be given. This condition will be only sufficient for negative integer $g(0)$. In the case of a non integer $g(0)$, the solution of (1) can be reduced to an Abel integral equation of the first kind.

In the second chapter we shall reject the above conditions made on the kernel $g(t)$ and give a simple sufficient condition referring to $g(t)$, which guarantees that the solutions of (2) are finite order distributions.

The main tool of the operational discussion of (1) is GESZTELYI's rule of the algebraic integration of operators being finit order distributions having a bounded support on the left. It can be formulized in different equivalent forms (see [1], [2], [3], [4]). For our purposes, the following formulization of Gesztelyi's rule will be very convenient.

Let $x=s^k e^{as}\{f\}$, $k=0, 1, 2, \ldots$; $-\infty<a<\infty$ *and let* f *be an arbitrary locally integrable function, then* x *is algebraic integrable and an integral of it reads as*

$$y=-s^{k+1}e^{as}\left\{(t-a)^k\int\frac{f(t)\,dt}{(t-a)^{k+1}}\right\} \tag{4}$$

where the function $\int\frac{f(t)\,dt}{(t-a)^{k+1}}$ *must take the value* 0 *for* $t=0$ *when* $a\neq 0$.

REMARK. This formulization of Gesztelyi's rule is based on the fact that the function

$$(t-a)^k\int\frac{f(t)\,dt}{(t-a)^{k+1}}$$

is locally integrable on $0\leqq t<\infty$. This is trivial for $a<0$, proved in [1] for $a=0$, and can be easily extended to the case of $a>0$.

Operational notations and symbols as in [1], [2] will be used in this paper. Sometimes the convolution will be denoted by $*$.

1. §. The solution of (1)

Instead of (1) we shall write

$$(t-a)f(t)+\int_0^t f(\tau)g(t-\tau)\,d\tau = h(t), \quad a>0. \tag{1.1}$$

The corresponding algebraic differential equation is of the form

$$Df+af-gf=-h \tag{1.2}$$

where D denotes the algebraic derivative. With the denotations of the introduction and the conditions made on $g(t)$ we have shown in [1] that the general operational solution of the corresponding homogeneous equation reads as

$$f = Ce^{-as}s^{g(0)}\exp\left\{\frac{g(t)-g(0)}{-t}\right\} = Ce^{-as}s^{g(0)}[1+\{G\}] \tag{1.3}$$

where C is an abitrary number. We can find a particular solution of (1.2) by applying the method of variation of parameters. We have

$$f_p = \left[-\int he^{as}s^{-g(0)}[1+\{G_0(t)\}]\right]e^{-as}s^{g(0)}(1+\{G(t)\}) \tag{1.4}$$

where $\int$ denotes the symbol of the algebraic integration.

In the sequel we shall distinguish the cases $g(0)\leqq 0$ and $g(0)>0$.

I. Let $g(0)=-k+\varepsilon\leqq 0$. Here k is a nonnegative integer, $0\leqq\varepsilon<1$. The algebraic integral occurring in (1.4) can be written as

$$\int he^{as}s^{k-\varepsilon}(1+\{G_0\}) = \int s^k e^{as}\{l^\varepsilon h + l^\varepsilon h * G_0\}. \tag{1.5}$$

By introducing the denotation

$$\{H(t)\} = \{l^\varepsilon h + l^\varepsilon h * G_0\} \tag{1.6}$$

the application of Gesztelyi's rule gives

$$\int s^k e^{as}\{H(t)\} = -s^{k+1}e^{as}\left\{(t-a)^k\int\frac{H(t)\,dt}{(t-a)^{k+1}}\right\}. \tag{1.7}$$

Substituting this to (1.4) we have

$$f_p = (1+\{G\})s^{\varepsilon+1}\left\{(t-a)^k\int\frac{H(t)\,dt}{(t-a)^{k+1}}\right\}. \tag{1.8}$$

Obviously the particular solution so obtained is a *second* order distribution in general. Taking into account (1.3) we can seethat every solution of (1.2) is a second order distribution in general.

It is important to investigate under which conditions is (1.8) identic with a locally integrable function? From the elements of the operational calculus follows that if $(t-a)^k\int\frac{H(t)\,dt}{(t-a)^{k+1}}$ is not absolutely continuous, then (1.8) cannot represent

a function. By assuming the differentiability of the above function we have

$$f_p = (1+\{G\})s^{\varepsilon}\left\{\frac{H(t)}{t-a}+k(t-a)^{k-1}\int\frac{H(t)\,dt}{(t-a)^{k+1}}\right\} \tag{1.9}$$

and for $g(0)=-k$ i.e. for $\varepsilon=0$ we obtain that

$$f_p = (1+\{G\})\left\{\frac{H(t)}{t-a}+k(t-a)^{k-1}\int\frac{H(t)\,dt}{(t-a)^{k+1}}\right\} \tag{1.10}$$

holds. It is obvious from (1.10) that for $k=0$ (1.10) is a locally integrable solution of (1.1) if and only if the function

$$\frac{h+h*G_0}{t-a} \tag{1.11}$$

is integrable on $\langle 0, \infty)$. In the case of $k>0$, $\varepsilon=0$, the integrability of (1.1) is only a sufficient condition guaranteeing the existence of a solution of (1.1). In fact, by taking into account the above Remark, and assuming the integrability of (1.11) and introducing the denotation $K(t)=\dfrac{H(t)}{t-a}$ we have

$$f_p = (1+\{G\})\left\{K(t)+k(t-a)^{k-1}\int\frac{K(t)\,dt}{(t-a)^{k}}\right\} \tag{1.12}$$

So is (1.12) locally integrable if (1.11) is. Necessity does not hold, since we can see from (1.10) that the integrability of (1.10) does not imply the integrability of $\dfrac{H(t)}{t-a}$.

Now comes the case of $\varepsilon\neq 0$. If (1.9) is a function, then it can be written in the form of an Abel type integral equation of the first kind as follows

$$\int_0^t f_p(\tau)\frac{(t-\tau)^{\varepsilon-1}}{\Gamma(\varepsilon)}\,d\tau = \frac{H(t)}{t-a}+k(t-a)^{k-1}\int\frac{H(t)\,dt}{(t-a)^{k+1}}+ \tag{1.13}$$

$$+G(t)*\left[\frac{H(t)}{t-a}+k(t-a)^{k-1}\int\frac{H(t)\,dt}{(t-a)^{k+1}}\right].$$

(1.13) has a well-known solution formula, if the right hand side of (1.13) is absolutely continuous (see [7]). But (1.13) can have a solution even if the right hand side of (1.13) has no integrable derivative.

II. Let $g(0)=k+\varepsilon>0$. $k=0, 1, 2, \ldots,$; $0\leqq\varepsilon\leqq 1$. Instead of (1.5) we have the following algebraic integral

$$\int he^{as}s^{-k-\varepsilon}(1+\{G_0\}) = \int (l^{k+\varepsilon}h+l^{k+\varepsilon}h*G_0)e^{as}. \tag{1.14}$$

Applying again Gesztelyi's rule about algebraic integration we have

$$\int \ldots = -se^{as}\left\{\int\frac{H_k(t)\,dt}{t-a}\right\} \tag{1.15}$$

where the function $H_k(t)$ is defined by

$$H_k(t) = l^{k+\varepsilon}h + l^{k+\varepsilon}h * G_0 \tag{1.16}$$

which coincides with $H(t)$ for $k=0$ $(H_0(t)=H(t))$.

By substituting (1.15), (1.16) into (1.14) instead of (1.8) we have

$$f_p = (1+\{G\})s^{k+\varepsilon+1}\left\{\frac{H_k(t)\,dt}{t-a}\right\}. \tag{1.17}$$

Obviously the particular solution so obtained is a distribution of order $k+2$ in general. Taking into account (1.3) we can see that every solution of (1.2) is a distribution of order $k+2$ in general.

The operator (1.17) can be a function only if the function $\int \frac{H_k(t)\,dt}{t-a}$ is absolutely continuous. By assuming this we obtain

$$f_p = (1+\{G\})s^{k+\varepsilon}\left\{\frac{H_k(t)}{t-a}\right\} \tag{1.18}$$

since by the definition $\int \frac{H_k(t)\,dt}{t-a}$ vanishes for $t=0$.

(1.18) can be a function only if $\frac{H_k(t)}{t-a}$ has a locally integrable k-th derivative. We need the following

LEMMA 1. *Let the k-th derivative of* $\frac{H_k(t)}{t-a}$ *be locally integrable. Then it is*

$$s^k\left\{\frac{H_k(t)}{t-a}\right\} = \left\{\frac{H(t)}{t-a} - \frac{k\int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a)^{k+1}}\right\}. \tag{1.19}$$

PROOF. The proof goes by total induction. First we can see that the i-th derivatives of $\frac{H_k(t)}{t-a}$ vanish at $t=0$ for $0\leqq i\leqq k-1$. (1.19) holds trivially for $k=1$, since $lH(t)=H_1(t)$. Let us assume that it holds for k, we show that it is also holding for $k+1$. Since

$$H_{k+1}(t) = l^{k+1}H(t) = l^k[lH(t)]$$

we have

$$s^{k+1}\left\{\frac{H_{k+1}(t)}{t-a}\right\} = s\left\{\frac{\int_0^t H(\tau)\,d\tau}{t-a} - \frac{k\int_0^t (\tau-a)^{k-1}\int_0^\tau H(u)\,du\,d\tau}{(t-a)^{k+1}}\right\}. \tag{1.20}$$

A partial integration gives that

$$s^{k+1}\left\{\frac{H_{k+1}(t)}{t-a}\right\} = \tag{1.21}$$

$$= s\left\{\frac{\int_0^t H(\tau)\,d\tau}{t-a} - \frac{k}{(t-a)^{k+1}}\left(\frac{(t-a)^k}{k}\int_0^t H(\tau)\,d\tau - \frac{1}{k}\int_0^t H(\tau)(\tau-a)^k\,d\tau\right)\right\} =$$

$$= s\left\{\frac{\int_0^t H(\tau)(\tau-a)^k\,d\tau}{(t-a)^{k+1}}\right\} = \left\{\frac{H(t)}{t-a} - \frac{(k+1)\int_0^t H(\tau)(\tau-a)^k\,d\tau}{(t-a)^{k+2}}\right\},$$

and the Lemma is proved. By assuming that the function $\frac{H_k(t)}{t-a}$ has a k-th derivative by the above Lemma we obtain the operator

$$f_p = (1+\{G\})s^{\varepsilon}\left\{\frac{H(t)}{t-a} - \frac{k\int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a)^{k+1}}\right\} \tag{1.22}$$

giving for $\varepsilon=0$ the function

$$f_p = (1+\{G\})\left\{\frac{H(t)}{t-a} - \frac{k\int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a)^{k+1}}\right\}. \tag{1.23}$$

We remark that now the integrability of $\frac{H(t)}{t-a}$ is neither sufficient, nor necessary to the solvability of (1.1) as it can be seen by trivial examples.

Finally for $\varepsilon=0$ we obtain from (1.22), (1.6) the Abel type integral equation of the first kind

$$\int_0^t f_p(\tau)\frac{(t-\tau)^{\varepsilon-1}}{\Gamma(\varepsilon)}\,d\tau = \frac{H(t)}{t-a} - \frac{k\int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a^{k+1}} + \tag{1.24}$$

$$+G(t)*\left[\frac{H(t)}{t-a} - \frac{k\int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a)^{k+1}}\right]$$

for the determination of $f_p(t)$.

We shall make of use of the following

LEMMA 2. *If the particular solution f_p of the algebraic differential equation* (1.2) *is not a function, i.e., an operator which is indentificable with a function, then the integral equation* (1.1) *has no locally integrable solution.*

PROOF. The statement is trivial for $g(0)<0$ since by (1.3) the corresponding homogeneous equation has nontrivial locally integrable solutions.

Let $g(0)=0$. By (1.3), (1.8) we have that every solution of (1.2) reads as

$$f = Ce^{-as}[1+\{G\}]+(1+\{G\})s\left\{\int \frac{H(t)}{t-a}\,dt\right\}$$

Let us assume that f is a function for some $C=C_1$.

Since by assumption $\int \frac{H(t)\,dt}{t-a}$ is not absolutely continuous we would obtain

$$\frac{\{f\}}{1+\{G\}} = \{F(t)\} = C_1 e^{-as}+s\left\{\int \frac{H(t)\,dt}{t-a}\right\}$$

and

$$\frac{C_1}{s} = e^{as}\left\{\int_0^t F(\tau)\,d\tau - \int \frac{H(t)\,dt}{t-a}\right\}$$

which may hold only if

$$\int_0^t F(\tau)\,d\tau - \int \frac{H(\tau)\,dt}{t-a} = 0 \quad \text{for} \quad 0 \leqq t \leqq a.$$

But this is impossible since $\int_0^t F(\tau)\,d\tau$ is absolutely continuous $\int \frac{H(t)\,dt}{t-a}$ is not in the neighbourhood of the point $t=a$.

Let now $g(0)>0$. By (1.3), (1.17) we have that every solution of (1.2) reads as

$$f = Ce^{-as}s^{g(0)}[1+\{G\}]+(1+\{G\})s^{g(0)+1}\left\{\int \frac{H_k(t)}{t-a}\,dt\right\}.$$

Let us assume that f is a function for some $C=C_1$. We would obtain that

$$\frac{fs^{-g(0)}}{1+\{G\}} = \{F(t)\} = C_1 e^{-as}+s\left\{\int \frac{H_k(t)\,dt}{t-a}\right\}$$

holds. So this case is reduced to the preceding one. Contradiction.

Taking into account (1.3) again we have proved the following

THEOREM 1. *Let us consider the integral equation*

$$(t-a)f(t)+\int_0^t f(\tau)g(t-\tau)\,d\tau = h(t), \quad a>0,\ t\geqq 0, \tag{1.25}$$

under the conditions that $g(0)=\lim_{t\to 0} g(t)$ *exists and the functions* $\frac{g(t)-g(0)}{t}$, $h(t)$ *are locally integrable. Moreover let*

$$G(t) = \sum_{j=1}^{\infty}\left\{\frac{g(t)-g(0)}{-t}\right\}^j, \quad G_0(t) = \sum_{j=1}^{\infty}\left\{\frac{g(t)-g(0)}{t}\right\}^j.$$

Let k be a nonnegative integer, $0 \leqq \varepsilon < 1$,

$$H_k(t) = \{l^{k+\varepsilon} h + l^{k+\varepsilon} h * G_0\} \quad (H_0(t) = H(t)).$$

I. *First let $g(0) = -k+\varepsilon \leqq 0$, then* (1.25) *has the following general operational solution*

$$f = Ce^{-as} s^{g(0)}(1+\{G\}) + f_p \tag{1.26}$$

where

$$f_p = (1+\{G\}) s^{1+\varepsilon} \left\{ (t-a)^k \int \frac{H(t)\,dt}{(t-a)^{k+1}} \right\}, \tag{1.27}$$

(1.27), (1.26) *are second order distributions in general.*

For $g(0)=0$ f_p is a function, if and only if, the function $\dfrac{H(t)}{t-a}$ is locally integrable.

For $g(0) = -k < 0$ the integrability of $\dfrac{H(t)}{t-a}$ implies the local integrability of f_p.

If f_p is a function for $\varepsilon = 0$, then it is of the form

$$f_p = (1+\{G\}) \left\{ \frac{H(t)}{t-a} + k(t-a)^{k-1} \int \frac{H(t)\,dt}{(t-a)^{k+1}} \right\}. \tag{1.28}$$

In the case of $\varepsilon \neq 0$ is (1.27) *a function, if and onl if, the function*

$$M(t) = \frac{H(t)}{t-a} + k(t-a)^{k-1} \int \frac{H(t)\,dt}{(t-a)^{k+1}}$$

is locally integrable and the Abel-type integral equation

$$\left\{ \int_0^t f_p(\tau) \frac{(t-\tau)^{\varepsilon-1}}{\Gamma(\varepsilon)} d\tau \right\} = (1+\{G\})\{M(t)\} \tag{1.29}$$

has a solution.

If f_p is a function, then for $g(0)=0$ it is the only locally integrable solution of (1.25), *and in the case of $g(0)<0$* (1.25) *has infinitely many locally integrable solutions.*

If f_p is not a function, then (1.25) *has no locally integrable solution.*

II. *Let $g(0) = k+\varepsilon > 0$.*

The general operational solution of (1.25) *is* (1.26) *where*

$$f_p = (1+\{G\}) s^{g(0)+1} \left\{ \int \frac{H_k(t)\,dt}{t-a} \right\}. \tag{1.30}$$

In this case (1.36), (1.30) *are distributions of order $k+2$ in general. If f_p is a function for $\varepsilon = 0$, it is of the form*

$$f_p = (1+\{G\}) \left\{ \frac{H(t)}{t-a} - \frac{k \int_0^t H(\tau)(\tau-a)^{k-1}\,d\tau}{(t-a)^{k+1}} \right\}. \tag{1.31}$$

In the case of $\varepsilon \neq 0$ *is* (1.30) *a function, if and only if, the function*

$$M(t) = \frac{H(t)}{t-a} - \frac{k \int_0^t H(\tau)(\tau-a)^{k-1}\, d\tau}{(t-a)^{k+1}}$$

is locally integrable and the Abel-type integral equation

$$\left\{ \int_0^t f_p(\tau) \frac{(t-\tau)^{\varepsilon-1}}{\Gamma(\varepsilon)}\, d\tau \right\} = (1+\{G\})\{M(t)\} \tag{1.32}$$

has a solution. If $\{x(t)\}$ *is a solution of* (1.25) *then*

$$\{x(t)\} = f_p$$

holds.

The occuring indefinite integrals must vanish at $t=0$.

COROLLARY. It follows from [1], [2] and [3], (1.28), (1.31) that for integer $g(0)$ the integral equation

$$(t-a)f(t) + \int_0^t f(\tau)g(t-\tau)\, d\tau = h(t)$$

has a particular solution formula which holds for $a>0$ and $a<0$ provided that it is solvable and this particular solution can be written as

$$f_p = (1+\{G\}) \left\{ \frac{H(t)}{t-a} - g(0)(t-a)^{-g(0)-1} \int \frac{H(t)\, dt}{(t-a)^{1-g(0)}} \right\} \tag{1.33}$$

where the indefinite integral must vanish at $t=0$.

2. §. A generalization of the kernel function $g(t)$

If the kernel function $g(t)$ does not satisfy the conditions introduced above, then the solutions of (1), if they exist, are not finite order distributions in general. This can be shown by trivial examples. At the application of the method of variation of parameters, we obtain algebraic integrals of operators being not finite order distributions in general. So Gesztelyi's rule is not applicable in general and since the problem of the algebraic integration is not quite solved, in the operational discussion of (1) we must restrict ourselves to interesting, but special cases. (For example see SCHATTE [8].)

The following question arises. Under what condition referring to the kernel $g(t)$ are the solutions of (1) distributions of finite order? We give now a sufficient condition guaranteeing this.

Let us consider the algebraic differential equation of the form

$$Df - af - bf = -h, \quad -\infty < a < \infty, \tag{2.1}$$

where it is not assumed a priori that the operator b is a function. Let us assume that the corresponding homogeneous equation has a solution of the form

$$f_h = e^{as}s^{\gamma}(1+\{x(t)\}) \tag{2.2}$$

where γ is a real number and $x(t)$ is a locally integrable function. The method of variation of parameters gives that a particular solution of the inhomogeneous equation (2.1) can be written as

$$f_p = -e^{as}s^{\gamma}(1+\{x(t)\})\int hs^{-\gamma}e^{-as}\frac{1}{1+\{x(t)\}} \tag{2.3}$$

where $\int$ denotes the algebraic integration.

Taking into account Gesztelyi's rule about algebraic integration, we can see that the solutions of (2.1) are finite order distributions for every function $h(t)$, if (2.2) holds. The only one we must to show that (2.1) is the operational form of an integral equation, i.e. that the operator b is a function:

$$b = \{g(t)\}.$$

Let us write the homogeneous differential equation

$$Df - af - bf = 0. \tag{2.4}$$

We have

$$Df_h = ae^{as}s^{\gamma}(1+\{x(t)\}) + e^{as}\gamma s^{\gamma-1}(1+\{x(t)\}) + e^{as}s^{\gamma}\{-tx(t)\}. \tag{2.5}$$

Substituting this into (2.4) we obtain

$$ae^{as}s^{\gamma}(1+x) + e^{as}\gamma s^{\gamma-1}(1+x) + e^{as}s^{\gamma}\{-tx(t)\} -$$

$$-ae^{as}s^{\gamma}(1+x) - be^{as}s^{\gamma}(1+x) = 0.$$

After simple calculation we have

$$b = \{g(t)\} = \frac{\gamma}{s} - \frac{\{tx(t)\}}{1+\{x(t)\}} = \{\gamma\} - \{tx(t)\}\sum_{i=0}^{\infty}(-1)^i\{x(t)\}^i \tag{2.6}$$

for every number γ, and for every function $\{x(t)\}$. It is obvious that $b=\{g(t)\}$ is not continuous in the neighbourhood of the origin in general, so the conditions made on $g(t)$ in the introduction are not satisfied in general.

On the other hand, if these conditions are satisfied, the function $\{x(t)\}$ in (2.6) reads as

$$\{x(t)\} = \exp\left\{\frac{g(t)-g(0)}{-t}\right\} - 1 = \sum_{i=1}^{\infty}\left\{\frac{g(t)-g(0)}{-t}\right\}^i\frac{1}{i!} = \{G(t)\} \tag{2.7}$$

and $\gamma=g(0)$. So it holds the following

THEOREM 2. *Let us consider the algebraic differential equation*

$$Df - af - gf = -h \tag{2.8}$$

where a is any real number, h is any locally integrable function. If there exists a real number γ and a locally integrable function $x(t)$ such that

$$g = \{g(t)\} = \frac{\gamma}{s} - \{tx(t)\} \sum_{i=0}^{\infty} (-1)^i \{x(t)\}^i \tag{2.9}$$

holds in the operational sense, then every operational solution of (2.8) *is a finite order distribution having a bounded support on the left. These solutions can be written by the corresponding formulas given in* [1], [2] *and in the first chapter of this paper, if we take the following substitutions in them:*

$$g(0) \leftrightarrow \gamma,$$

$$\{G(t)\} \leftrightarrow \{x(t)\},$$

$$\{G_0(t)\} \leftrightarrow \sum_{i=1}^{\infty} (-1)^i \{x(t)\}^i.$$

REFERENCES

[1] FÉNYES, T.: „Anwendung der Mikusińskischen Operatorenrechnung zur Lösung von Integralgleichungen dritter Art von Faltungstypus", *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **9** (1965), 365—399.

[2] FÉNYES, T.: A note on the solution of integral equations of convolution type of the third kind by application of the operational calculus of Mikusiński. *Studia Sci. Math. Hung.* **2** (1967), 81—89.

[3] FÉNYES, T.: On the operational solution of certain non-linear singular integral equations, *Studia Sci. Math. Hung.* **4** (1969), 69—91.

[4] GESZTELYI, E.: Anwendung der Operatorenrechnung auf lineare Differentialgleichungen mit Polynom-Koeffizienten, *Publ. Math. Debrecen* **10** (1963), 215—243.

[5] PENYŐ, I.: Über den Zusammenhang zwischen den Mikusińskischen Operatoren und den Distributionen, *Math. Nachr.* **19** (1958), 161—164.

[6] WLOKA, J.: Distributionen und Operatoren, *Math. Ann.* **140** (1960), 227—244.

[7] BUTZER, P. L.: Die Anwendung der Operatorenkalküls von Jan Mikusiński auf lineare Integralgleichungen vom Faltungstypus, *Archive for Rational Mechanics and Analysis* **2** (1958), 114—128.

[8] SCHATTE, P.: Funktionentheoretische Untersuchungen im Mikusńskischen Operatorenkörper, *Math. Nachr.* **35** (1967), 19—56.

Mathematical Institute of the Hungarian Academy of Sciences,
H-1053 Budapest, Reáltanoda u. 13—15, Hungary

(Received September 20, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 77—80

ON A PROBLEM OF L. FEJES TÓTH

by

G. O. H. KATONA

It is easy to see by induction that n lines in the Euclidean plane determine at most $\binom{n-1}{2}$ bounded domains. L. Fejes Tóth conjectured that an analogous statement holds for n convex sets. We prove in this note a slightly stronger theorem. Instead of convexity we need only that the pairwise intersections of the domain are connected.

Let E denote the Euclidean plane. If $A \subset E$, then $\delta(A)$ denotes the set of boundary points of A. The family of subsets A which are homeomorphic to the closed unit disc is denoted by $\mathscr{D}$. The elements of $\mathscr{D}$ are called *domains*. The family of closed connected sets is denoted by $\mathscr{C}$. A homeomorphic map of an interval is called an *arc*. If an arc is non-empty and is not a single point we call it *non-trivial*. If it is not ambiguous we name the arcs by their endpoints a, b: (a, b) if the arc is open and $[a, b]$ if it is closed.

THEOREM. *If* $A_1, \ldots, A_n \in \mathscr{D}$ *and* $A_i \cap A_j \in \mathscr{D}$ $(1 \leqq i, j \leqq n)$ *then*

$$E - \bigcup_{i=1}^{n} A_i$$

has at most $\binom{n-1}{2}$ *bounded connected componets.*

In the proofs we use the following trivial statements without proofs:

(i) Let $A \in \mathscr{D}$ and let $B, C \subset A$, $B \cap C = \varnothing$ be sets in the plane. If $x_1, y_1 \in \delta(B) \cap \delta(A)$. $x_2, y_2 \in \delta(C) \cap \delta(A)$ and x_1, x_2, y_1, y_2 lie on $\delta(A)$ in this order then either B or C is disconnected (Fig. 1).

(ii) Let $A \in \mathscr{D}$. If $B \subset A$ and $T \subset A$ are connected sets satisfying $B \cap T \neq \varnothing$, $B \cap (A - T) \neq \varnothing$ then $B \cap \delta(T) \cap \delta(A - T) \neq \varnothing$ (Fig. 2).

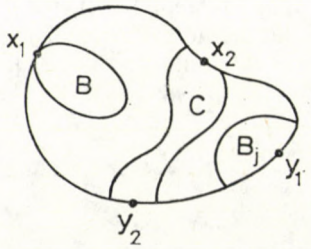


Fig. 1.

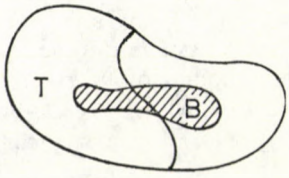


Fig. 2.

Let $A\in\mathcal{D}$, $B_1, \ldots, B_m\in\mathcal{C}$ be sets in the plane. The family of those connected components T of

$$A-\bigcup_{i=1}^{n} B_i$$

for which $T\cap\delta(A)$ contains at least two connected components is denoted by $\mathcal{T}_m=\mathcal{T}_m(A; B_1, \ldots, B_m)$. The set of the connected components of $T\cap\delta(A)$ $(T\in\mathcal{T}_m)$ is $\mathcal{V}_m=\mathcal{V}_m(A; B_1, \ldots, B_m)$.

LEMMA. *Let $A\in\mathcal{D}$, $B_1, \ldots, B_m\in\mathcal{C}$ be sets in the plane satisfying $B_i\subset A$ $(1\leqq i\leqq m)$ and $B_i\cap B_j=\varnothing$ $(1\leqq i<j\leqq m)$. Then $|\mathcal{V}_m|$ and $|\mathcal{T}_m|$ are finite and*

$$|\mathcal{V}_m|-|\mathcal{T}_m|\leqq m-1.$$

PROOF. 1. We first prove that there is an index j such that between the extremal points of $\delta(A)\cap B_j$ there is no element of B_i $(i\neq j)$ on $\delta(A)$ (Fig. 3, 4).

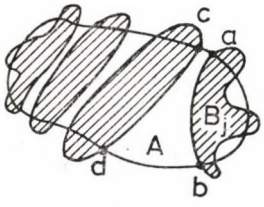


Fig. 3.

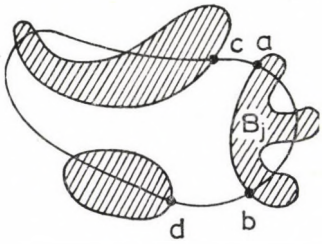


Fig. 4.

Choose k and two points a, b of $\delta(A)\cap B_k$ (and one of the arcs determined by them) in such a way that the number z of B_i's having non-empty intersection with (a, b) is >0 but minimal. If the condition that $z>0$ can not be satisfied, we are ready. On the arc $[a, b]$ take the maximum point c of $\delta(A)\cap B_k$ so that (a, c) does not contain any point of B_i $(i\neq k)$. On the other hand let d be the minimum point $(>c)$ of $\delta(A)\cap B_k$ on $[c, b]$. It is easy to see that (c, d) still contains a point from some B_i $(i\neq k)$. Thus, (c, d) satisfies the conditions required for (a, b), but it does not contain any point of B_k.

Choose now an index $l\neq k$ such that B_l has a point on the arc (c, d). Let e and f be the minimum and maximum points of B_l on the arc (c, d). If (e, f) contains a point of some B_i, $i\neq l$, then (e, f) contains points form at most $z-1$ different B_i, which contradicts the minimality of z. Consequently, (e, f) contains only points of B_l. If there is a point g of B_l outside of $[c, d]$, then we obtain a contradiction by (i), since $c, d\in B_k$, $e, g\in B_l$ and B_k and B_l are disjoint connected sets. It follows that there is no point of B_l on $\delta(A)$ outside of $[e, f]$, and there is no point of B_i $(i\neq l)$ in $[e, f]$. Our statement is proved.

2. Choose j according to section 1 of this proof. Let and b be the minimum and maximum points of $\delta(A)\cap B_j$ (Fig. 3, 4).

We shall use the following notations:

$$\mathcal{T}_{m-1}=\mathcal{T}_{m-1}(A; B_1, \ldots, B_{j-1}, B_{j+1}, \ldots, B_m),$$
$$\mathcal{V}_{m-1}=\mathcal{V}_{m-1}(A; B_1, \ldots, B_{j-1}, B_{j+1}, \ldots, B_m).$$

We claim that

$$|\mathscr{V}_m| - |\mathscr{V}_{m-1}| \leqq 2. \tag{1}$$

Denote by c the maximum of the points $\leqq a$ of $\delta(A) \cap (\bigcup_{i \neq j} B_i)$. Similarly, let d denote the minimum of the points $\geqq b$ in $\delta(A) \cap (\bigcup_{i \neq j} B_i)$. Since the B_i's are disjoint, we have $c < a$ and $b < d$. Let us show that there is no element of $\mathscr{V}_m$ in the arc (a, b). In the opposite case there is an arc $\alpha \subset T \cap \delta(A)$ in (a, b) for some $T \in \mathscr{T}_m$. There must be an $\alpha \neq \beta \subset T \cap \delta(A)$ by the definition of τ_m. If β is outside of (a, b), then we obtain the contradiction by (i) becaused $\alpha, \beta \subset T \cap \delta(A)$, $a, b \in \delta(A) \cap B_j$ and T and B_j are disjoint and connected. If B is also in (a, b) then there is a point e between α and β satisfying $e \in B_i$ for some i. It follows from the definition of j that $i = j$. We obtain the contradiction by (i) using the points a, e and one point of each of α and β.

It is easy to see that the elements of $\mathscr{V}_m$ and $\mathscr{V}_{m-1}$ lying outside of (c, d) coincide.

Thus $\mathscr{V}_m - \mathscr{V}_{m-1}$ contains at most 2 elements: the arcs (c, a) and (b, d). (1) is proved.

Equality holds in (1) only if both (c, a) and (b, d) belong to some members T_1 and T_2 of $\mathscr{V}_m$, but (c, d) does not belong to any member of $\mathscr{V}_{m-1}$ (Fig. 3).

We prove now that

$$|\mathscr{T}_m| - |\mathscr{T}_{m-1}| \geqq -1. \tag{2}$$

Let T be the connected component of $A - \bigcup_{i \neq j} B_i$ containing (c, d). If there is a point of B_j outside of T, we may apply (ii) with A, B_j and T: $B_j \cap \delta(T) \cap \cap \delta(A - T) \neq \varnothing$. However $\delta(T) \cap \delta(A - T) \subset \bigcup_{i \neq j} B_i$ since the B_i's are closed, thus the fact that $B_j \cap \delta(T) \cap \delta(A - T) \neq \varnothing$ contradicts the disjointness of B_i's. We have shown that $B_j \subset T$. As a consequence we obtain that B_j does not change the elements of $\mathscr{T}_{m-1} - \{T\}$: $\mathscr{T}_{m-1} - \{T\} \subset \mathscr{T}_m$. We have proved (2).

Equality holds in (2) only if $T \in \mathscr{T}_{m-1}$ but no subset of $T \in \mathscr{T}_m$. It means that $(c, a), (b, d) \notin \mathscr{V}_m$, and we have the stronger inequality $|\mathscr{V}_m| - |\mathscr{V}_{m-1}| \leqq 0$. (It can be proved that this situation does not occur.) In this case we have

$$(|\mathscr{V}_m| - |\mathscr{V}_{m-1}|) - (|\mathscr{T}_m| - |\mathscr{T}_{m-1}|) \leqq 1. \tag{3}$$

If both in (1) and (2) there is strict inequality (Fig. 4), then (3) is true, again. We still have to discuss the case when we have equality in (1) (Fig. 3). Then $|\mathscr{T}_m| - |\mathscr{T}_{m-1}| \geqq 1$ can be proved. Indeed, in this case $T \notin \mathscr{T}_{m-1}$, showing that $\mathscr{T}_{m-1} \subset \mathscr{T}_m$. On the other hand $T_1 \in \mathscr{T}_m - \mathscr{T}_{m-1}$. (3) is proved.

3. To prove the lemma we use induction over m. For $m = 1$ it is easy to see that $|\mathscr{V}_m| = |\mathscr{T}_m| = 0$. Suppose that $m > 1$ and the lemma is true for $m - 1$:

$$|\mathscr{V}_{m-1}| - |\mathscr{T}_{m-1}| \leqq m - 2. \tag{4}$$

The lemma follows from

$$|\mathscr{V}_m| - |\mathscr{T}_m| = (|\mathscr{V}_m| - |\mathscr{V}_{m-1}|) - (|\mathscr{T}_m| - |\mathscr{T}_{m-1}|) + (|\mathscr{V}_{m-1}| - |\mathscr{T}_{m-1}|)$$

and from (3) and (4).

PROOF OF THE THEOREM. We use induction over n. For $n=1$ the statement trivially holds. Let us suppose $n+1>1$ and the theorem is true for n, which means that the number of the bounded connected components of $E-\bigcap_{i=1}^{n} A_i$ is at most $\binom{n-2}{2}$. We have to prove that subtracting A_{n+1}, the number of the bounded connected components does not increase by more than $n-1$. Then the statement for $n+1$ follows from the identity $\binom{n-1}{2}+n-1=\binom{n}{2}$.

A_{n+1} does influence only those connected components of $E-\bigcup_{i=1}^{n} A_i$, which are non-disjoint to A_{n+1}. Denote these components by $C_1, \ldots, C_u$. The connected components of C_i-A_{n+1} are denoted by $D_{i1}, \ldots, D_{iw_i}$. Thus we have to prove that

$$\sum_{i=1}^{u} w_i - u \leqq n-1. \tag{5}$$

Construct a graph G_i whose vertices are on the one hand $D_{i1}, \ldots, D_{iw_i}$ and on the other hand the connected components $E_{i1}, \ldots, E_{it_i}$ of $C_i \cap A_{n+1}$. Two vertices are connected in G_i iff they have a common non-trivial arc on $\delta(A)$. If they have more such common arcs, then the are connected with more edges. It is easy to see that G_i is a connected graph, since C_i is connected. Denote the number of edges of G_i by v_i. The number of edges of a connected graph is greater than or equal to the number of vertices -1. That is,

$$w_i + t_i - 1 \leqq v_i \quad (1 \leqq i \leqq u)$$

whence

$$\sum_{i=1}^{u} w_i - u \leqq \sum_{i=1}^{u} v_i - \sum_{i=1}^{u} t_i. \tag{6}$$

Denote the connected components of $\bigcup_{i=1}^{n} (A_{n+1} \cap A_i)$ by $B_1, \ldots, B_m$. Obviously

$$m \leqq n \tag{7}$$

and by definition

$$\sum_{i=1}^{u} t_i = |\mathscr{T}_m|, \quad \sum_{i=1}^{u} v_i = |\mathscr{V}_m|. \tag{8}$$

Now (5) follows from (6), (8), the Lemma and (7). The proof is completed.

Recently M. Geréb, E. Győry and Gy. Szász found some other proofs and generalizations. They will be published in a forthcoming paper.

Mathematical Institute of the Hungarian Academy of Sciences,
H-1053 Budapest, Reáltanoda u. 13—15, Hungary

(Received October 15, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 81—88

CONCAVITY OF THE ZEROS OF BESSEL FUNCTIONS

by

Á. ELBERT

Let $\nu \geqq 0$ and let $j_{\nu k}$ denote the k^{th} positive zero of the Bessel function $J_\nu(x)$ of first kind. There are many results concerning the behaviour of $j_{\nu k}$ as a function of order ν, see e.g. in the books [1], [2] or in papers [3]—[6]. Here we shall present only those results which will be needed for our purpose.

First of all there is a very useful formula due to WATSON ([1], p. 508) which reads

$$\frac{dj_{\nu k}}{d\nu} = 2j_{\nu k}\int_0^\infty K_0(2j_{\nu k}\operatorname{sh} t)e^{-2\nu t}\,dt, \tag{1}$$

where $K_0(x)$ denotes the modified Bessel function of order zero. This formula is valid for all real ν (see the extended domain of definition of $j_{\nu k}$ below).

By (1) the function $j_{\nu k}$ is increasing as ν increases since $K_0(x)>0$ on $(0, \infty)$. About the asymptotic behaviour of $j_{\nu k}$ for large values of ν TRICOMI [4] proved that

$$j_{\nu k} \sim \nu + c_{1k}\nu^{\frac{1}{3}} + c_{2k}\nu^{-\frac{1}{3}} + \ldots \quad \text{for} \quad \nu \gg 1,\ k = 1, 2, \ldots, \tag{2}$$

where the c_{ik}'s are constants. In particular

$$\begin{aligned} c_{11} &= 1.855757\ldots, \\ c_{12} &= 3.2447\ldots, \\ c_{13} &= 4.3817\ldots. \end{aligned} \tag{3}$$

It is clear from the definition of $j_{\nu k}$ that the sequence $c_{11}, c_{12}, c_{13}, \ldots$ is nondecreasing.

The ratio $j_{\nu k}/\nu$ was studied independently by several authors (see [3], [5], [6]) and it turned out that this quotient is strictly decreasing. Combining this with (2) we have

$$0 < \frac{\nu}{j_{\nu k}} < 1 \quad \text{for} \quad \nu > 0,\ k = 1, 2, \ldots. \tag{4}$$

Now we shall extend the domain of definition of the functions $j_{\nu k}$ from $0 \leqq \nu < \infty$ to $-k < \nu < \infty$ for $k = 1, 2, \ldots$.

Let us consider the power series of the Bessel function $J_\nu(x)$

$$J_\nu(x) = \sum_{i=0}^{\infty}(-1)^i \frac{\left(\frac{x}{2}\right)^{2i+\nu}}{i!\Gamma(i+\nu+1)}.$$

Since $J_\nu(x)$ is analytical in both variables x and ν except, possibly, at $x=0$, therefore the zeros of $J_\nu(x)$ can be continued analytically as long as they remain positive. Consider first the neighbourhood $\nu=-l$ and $x=0$ on the plane (ν, x), where $l=1, 2, \ldots$. The function $f(\nu, x)$ defined by

$$(5)\qquad f(\nu,x)=(\nu+l)\Gamma(\nu+1)\left(\frac{x}{2}\right)^{-\nu}J_\nu(x)=(\nu+l)\left[1-\frac{\left(\frac{x}{2}\right)^2}{1!(\nu+1)}+\ldots\right.$$

$$\left.+(-1)^{l-1}\frac{\left(\frac{x}{2}\right)^{2l-2}}{(l-1)!(\nu+1)\ldots(\nu+l-1)}\right]+$$

$$+(-1)^l\frac{\left(\frac{x}{2}\right)^{2l}}{l!((\nu+1)\ldots(\nu+l-1)}\left[1-\frac{\left(\frac{x}{2}\right)^2}{(l+1)(\nu+l+1)}+\ldots\right]$$

is analytical in this neighbourhood provided it is sufficiently small. Since $f(-l, 0)=0$, $f_\nu'(-l, 0)=1\neq 0$, the equation $f(\nu, x)=0$ has a unique analytical solution of the form

$$\nu+l=\sum_{m=1}^{\infty}a_m x^m$$

for sufficiently small x's (see e.g. [7] on p. 192). It is easy to verify that $a_1=\ldots$ $\ldots=a_{2l-1}=0$, $a_{2l}=2^{-l}/[l!(l-1)!]$ and the a_m's are real. This function has an inverse function $x=j(\nu)$ (see [7], p. 190)

$$j(\nu)=\sum_{m=1}^{\infty}b_m[(\nu+l)^{1/(2l)}]^m,$$

where $b_1=2\cdot|[l!(l-1)!]^{1/(2l)}|>0$, b_m's are real.

The function $w=z^{1/(2l)}$ is a $2l$-valued function, and we choose here that branch which maps the positive real z's into positive real w's. Then the function $j(\nu)$ will be positive for sufficiently small positive values of $\nu+l$. If we chose another branch of $w=z^{1/(2l)}$, the function $j(\nu)$ given above would not be real and positive for small positive values of $\nu+l$. Hence by (5) the function $j(\nu)$ represents the only real and positive zero of $J_\nu(x)$ for such ν's, therefore $j(\nu)$ should coincide with one of the functions $\{j_{\nu k}\}_{k=1}^{\infty}$ in some right neighbourhood of $\nu=-l$. Due to the uniqueness of $j(\nu)$ in the above mentioned sense there is no other $j_{\nu k}$ which might vanish at $\nu=-l$.

A simple consideration shows that there are no non-integer negative values of ν at which any function from $\{j_{\nu k}\}_{k=1}^{\infty}$ vanishes. Since for a fixed k these functions are increasing, the function $j_{\nu 1}$ is the one which vanishes at $\nu=-1$ and which exists only for $-1<\nu<\infty$. Then the function $j_{\nu 2}$ should vanish at $\nu=-2$, and so on. Thus we have that on the interval $-k-1<\nu<-k$ $(k=0, 1, \ldots)$ only the functions $j_{\nu,k+1}, j_{\nu,k+2}, \ldots$ are defined and the value $j_{\nu,k+1}$ is the *first* positive real zero of the function $J_\nu(x)$. It may be of interest to remark here that by HURWITZ' theorem (see [1], p. 483) the function $J_\nu(x)$ for $-k-1<\nu<-k$

($k=1, 2, \ldots$) has exactly $2k$ complex valued zeros and all of the real zeros are $\pm j_{\nu,k+1}$, $\pm j_{\nu,k+2}$, $\ldots$, and if $\nu > -1$ then $J_\nu(x)$ has only real zeros.

We start with a lemma and then we shall prove the main result of our paper.

LEMMA. *Let* $j=j_{\nu k}$ *for* $k=1, 2, \ldots$. *Then the inequalities*

$$(\nu+k)\frac{dj_{\nu k}}{d\nu} \leqq j_{\nu k} \quad \text{for} \quad -k < \nu \leqq 0,$$

$$\left(\nu+\frac{1}{2}\right)\frac{dj_{\nu k}}{d\nu} < j_{\nu k} \quad \text{for} \quad \nu \geqq 0$$

are valid.

REMARK. The inequalities here imply that the ratio $j_{\nu k}\Big/\left(\nu+\frac{1}{2}\right)$ is strictly decreasing for $-k \leqq \nu < \infty$ which sharpens the result cited above concerning the decrease of the quotient $j_{\nu k}/\nu$. It would be of interest to show that also the ratio $j_{\nu k}/(\nu+k)$ is decreasing for $-k < \nu < \infty$.

PROOF. First we consider the case $-k < \nu \leqq 0$. Let $\varepsilon > 0$ and $k_\varepsilon = k/(1+2\varepsilon)$. Then we adapt the idea of Makai used in [5]. By Lommel's transformation of the Bessel differential equation (see [1]) we have that the function $w = w_\nu(z) = \sqrt{z}\, J_\nu(j_{\nu k} z^{\varepsilon/(\nu+k)})$ is a solution of the equation

$$w'' + \left[\left(\frac{\varepsilon j_{\nu k}}{\nu+k} z^{\frac{\varepsilon}{\nu+k}-1}\right)^2 - \left(\frac{\varepsilon\nu}{\nu+k}\right)^2 \frac{1}{z^2} + \frac{1}{4z^2}\right] w = 0.$$

Let us denote here the coefficient of w by $P_{\nu k}(z)$. Our aim is to show that the function $\varkappa(\nu) = j_{\nu k}/(\nu+k)$ is strictly decreasing for $-k_\varepsilon \leqq \nu \leqq 0$. By the continuity of $\varkappa(\nu)$ it is sufficient to show this on any open interval $(-r-1, -r) \cap [-k_\varepsilon, 0]$ where $r = 0, 1, \ldots, k-1$.

If we suppose the contrary, i.e. the function $\varkappa(\nu)$ is not strictly decreasing for $-k_\varepsilon \leqq \nu \leqq 0$, then by the continuity of $\varkappa(\nu)$ there would be an interval, say $(-r-1, -r) \cap (-k_\varepsilon, 0)$, on which there should exist two numbers ν', ν'' with $\nu' < \nu''$ and $\varkappa(\nu') \leqq \varkappa(\nu'')$. Then it is easy to see that the function $\nu^2/(\nu+k)^2$ is increasing for $-k < \nu < 0$ and

$$P_{\nu' k}(z) < P_{\nu'' k}(z) \quad \text{for} \quad 0 < z < 1.$$

The functions $w_{\nu'}(z)$, $w_{\nu''}(z)$ have the same number of zeros on $[0, 1]$. These are $z=0$, $z=1$ and the others correspond to $j_{\nu',r+1}, j_{\nu',r+2}, \ldots, j_{\nu',k-1}$ and $j_{\nu'',r+1}, j_{\nu'',r+2}, \ldots, j_{\nu'',k-1}$, resp. The existence of the zero at $z=0$ follows from $w_\nu(z)$ being of the order $z^{(1+2\varepsilon)(\nu+k_\varepsilon)/[2(\nu+k)]}$ in the right neighbourhood of $z=0$ and from $\nu', \nu'' > -k_\varepsilon$. Making use of the Sturmian comparison theorem we would have that the zeros of $w_{\nu''}(z)$ except $z=0$ should precede the corresponding ones of $w_{\nu'}(z)$ which is a contradiction because the $(k-r)$-th zeros are equal to 1 for both functions. This contradicition shows that $\varkappa(\nu)$ is strictly decreasing for $-k_\varepsilon < \nu \leqq 0$, hence $\varkappa'(\nu) \leqq 0$ there. Letting $\varepsilon \to +0$ we get the first inequality of Lemma.

Next we deal with the case $\nu>0$. Since $\operatorname{sh} t>t$ for $t>0$ and $K_0(x)$ is a strictly decreasing function we have from (1) that

$$j' < 2j\int_0^\infty K_0(2jt)e^{-2\nu t}\,dt = \int_0^\infty K_0(u)e^{-\frac{\nu}{j}u}\,du = \frac{\arccos\frac{\nu}{j}}{\sqrt{1-\frac{\nu^2}{j^2}}}, \tag{6}$$

where $j=j_{\nu k}$, $j'=dj_{\nu k}/d\nu$ and the value of the infinite integral is taken from [1] p. 388. Let $\alpha=\alpha(\nu)$ be defined by

$$\sin\alpha = \frac{\nu}{j}. \tag{7}$$

By (4) the function $\alpha(\nu)$ can be defined for all $\nu\geqq 0$. Since the ratio j/ν is strictly decreasing, the function $\alpha(\nu)$ is strictly increasing. So we have by (2) and by the definition

$$\alpha(0)=0, \quad \lim_{\nu\to\infty}\alpha(\nu)=\frac{\pi}{2},$$

and by (6)

$$j' < \frac{\frac{\pi}{2}-\alpha}{\cos\alpha}. \tag{8}$$

The function $\left(\frac{\pi}{2}-\alpha\right)\Big/\cos\alpha$ takes on its maximum value $\pi/2$ at $\alpha=0$ hence by (8) $j'<\pi/2$.

In the case $0\leqq\nu\leqq\frac{1}{2}$ we have

$$\left(\nu+\frac{1}{2}\right)\frac{j'}{j} < \frac{\frac{\pi}{2}}{j_{\nu k}} \leqq \frac{\frac{\pi}{2}}{j_{\nu 1}} \leqq \frac{\frac{\pi}{2}}{j_{01}} < 1$$

since $j_{01}=2.4048\ldots$. Thus the Lemma is true for these ν's.

Now we suppose that $\nu>\frac{1}{2}$. By (7) we have $j=\nu/\sin\alpha$ and $j'=(\sin\alpha-\nu\cos\alpha\cdot\alpha')/\sin^2\alpha$, hence the relation (8) implies

$$\alpha'\frac{\cos^2\alpha}{\sin\alpha\cos\alpha-\left(\frac{\pi}{2}-\alpha\right)\sin^2\alpha} > \frac{1}{\nu}. \tag{9}$$

The quotient of the left hand side is positive and therefore $\alpha'>0$. Moreover

$$\frac{\cos^2\alpha}{\sin\alpha\cos\alpha-\left(\frac{\pi}{2}-\alpha\right)\sin^2\alpha} < \frac{3}{\sin\alpha\cos\alpha} \quad \left(0<\alpha<\frac{\pi}{2}\right).$$

Indeed, this inequality is equivalent to the following statement: if $\Phi(\alpha)=3\cos\alpha-$ $-\cos^3\alpha-3\left(\frac{\pi}{2}-\alpha\right)\sin\alpha$ and $0<\alpha<\pi/2$, then $\Phi(\alpha)>0$. This, in turn, follows from $\Phi(\pi/2)=0$ and from

$$\Phi'(\alpha)=3\left(\frac{\pi}{2}-\alpha\right)\cos\alpha\left(\frac{\sin(\pi-2\alpha)}{\pi-2\alpha}-1\right)<0 \quad \text{if} \quad 0<\alpha<\frac{\pi}{2}.$$

Thus from (9) it follows that

$$\frac{\alpha'}{\sin\alpha\cos\alpha}>\frac{1}{3\nu}.$$

Integrating this from ν to ν_1 $(\nu<\nu_1<\infty)$ we have

$$\frac{\operatorname{tg}\alpha(\nu_1)}{\nu_1^{1/3}}>\frac{\operatorname{tg}\alpha(\nu)}{\nu^{1/3}},$$

thus the function $\operatorname{tg}\alpha(\nu)\cdot\nu^{-1/3}$ is increasing. Since $\sin\alpha=\nu/j_{\nu k}$ by (2) we have

$$\lim_{\nu\to\infty}\operatorname{tg}\alpha(\nu)\cdot\nu^{-1/3}=\frac{1}{\sqrt{2c_{1k}}},$$

and thus

$$\operatorname{tg}\alpha(\nu)<\frac{\nu^{1/3}}{\sqrt{2c_{1k}}}. \tag{10}$$

For proving the second part of the Lemma, it is sufficient by (7) and (8) to prove the validity of the inequaity

$$\left(\frac{\pi}{2}-\alpha\right)\operatorname{tg}\alpha<\frac{\nu}{\nu+\frac{1}{2}} \quad \text{for} \quad \nu>\frac{1}{2}. \tag{11}$$

Let $\bar{\alpha}=\bar{\alpha}_k(\nu)$ be defined by

$$\operatorname{tg}\bar{\alpha}=\frac{\nu^{1/3}}{\sqrt{2c_{1k}}}, \quad 0<\bar{\alpha}<\frac{\pi}{2}, \tag{12}$$

then by (10) we have that $\bar{\alpha}(\nu)>\alpha(\nu)$. Since $\left(\frac{\pi}{2}-\alpha\right)\operatorname{tg}\alpha$ is a strictly increasing function of α, by (11) it is sufficient to prove that

$$\left(\frac{\pi}{2}-\bar{\alpha}\right)\operatorname{tg}\bar{\alpha}<\frac{\nu}{\nu+\frac{1}{2}},$$

i.e.

$$\frac{\pi}{2}-\bar{\alpha}<\sqrt{2c_{1k}}\,\frac{\nu^{2/3}}{\nu+\frac{1}{2}}. \tag{13}$$

By the definition (12) of $\bar{\alpha}(\nu)$ we have

$$\frac{\pi}{2}-\bar{\alpha}=\frac{\pi}{2}-\operatorname{arctg}\frac{\nu^{1/3}}{\sqrt{2c_{1k}}}=\int_{\frac{\nu^{1/3}}{\sqrt{2c_{1k}}}}^{\infty}\frac{dx}{1+x^2}$$

and the inequality (13) is equivalent to

$$\sqrt{2c_{1k}}\,\frac{\nu^{2/3}}{\nu+\frac{1}{2}}-\int_{\frac{\nu^{1/3}}{\sqrt{2c_{1k}}}}^{\infty}\frac{dx}{1+x^2}>0 \quad \text{for} \quad \nu\geqq\frac{1}{2},\quad k=1,2,\ldots. \tag{14}$$

Let $\Psi=\Psi_k(\nu)$ denote the function on the left side. Then

$$\frac{d\Psi}{d\nu}=\frac{\psi(\nu)}{3\sqrt{2c_{1k}}\left(\nu+\frac{1}{2}\right)^2\left(\nu^{2/3}+\frac{\nu^{4/3}}{2c_{1k}}\right)},$$

where

$$\psi=\psi_k(\nu)=\left(-2c_{1k}+2\nu^{-\frac{1}{3}}+2c_{1k}\nu^{-1}+\frac{1}{4}\nu^{-\frac{4}{3}}\right)\nu^{\frac{4}{3}}.$$

It is easy to see that there exists only one positive value $\bar{\nu}_k$ for which $\psi_k(\bar{\nu}_k)=0$ and moreover

$$\psi_k(\nu)\gtrless 0 \quad \text{for} \quad 0<\nu<\bar{\nu}_k,\ \nu>\bar{\nu}_k.$$

Consequently the function $\Psi_k(\nu)$ takes on its minima on the interval $1/2<\nu<\infty$ at the endpoints of this interval. A simple consideration shows that $\lim_{\nu\to\infty}\Psi_k(\nu)=0$. The case $\nu=\frac{1}{2}$ requires a little computation. First let $k=1$. Then by (3) and (12) we have $\bar{\alpha}_1\left(\frac{1}{2}\right)=0.3908\ldots$ and $\Psi_1\left(\frac{1}{2}\right)=0.8228\ldots>0$. If $k\geqq 2$ then by (3) and (14) we have

$$\Psi_k\left(\frac{1}{2}\right)>\sqrt{2c_{1k}}\,\frac{1}{2^{2/3}}-\frac{\pi}{2}\geqq\sqrt{2c_{12}}\,\frac{1}{2^{2/3}}-\frac{\pi}{2}=1.6048-\frac{\pi}{2}>0.$$

Finally we get that $\Psi_k(\nu)>0$ for $\frac{1}{2}<\nu<\infty$ which completes the proof of the Lemma.

Having proved the Lemma we can deal with the concavity of the zeros.

THEOREM. *The zeros $j_{\nu k}$ considered as functions of ν are concave on the interval $-k<\nu<\infty$ for $k=1,2,\ldots.$*

PROOF. Putting $u=j_{\nu k}t$ in Watson's formula (1) and differentiating it with respect to ν we have

$$j'' = 2\int_0^\infty K_0'\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)e^{-\frac{2\nu}{j}u}\frac{d}{d\nu}\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)du -$$

$$-2\int_0^\infty K_0\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)e^{-\frac{2\nu}{j}u}\frac{d}{d\nu}\left(\frac{2\nu}{j}u\right)du.$$

An integration by parts gives for the first term on the right hand side that

$$\int_0^\infty K_0'\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)2\,\mathrm{ch}\,\frac{u}{j}\,\frac{e^{-\frac{2\nu}{j}u}\frac{d}{d\nu}\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)}{2\,\mathrm{ch}\,\frac{u}{j}}du =$$

$$= \left[K_0\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)\frac{e^{-\frac{2\nu}{j}u}\frac{d}{d\nu}\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)}{2\,\mathrm{ch}\,\frac{u}{j}}\right]_0^\infty -$$

$$-\int_0^\infty K_0\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)\frac{d}{du}\left[\frac{e^{-\frac{2\nu}{j}u}\frac{d}{d\nu}\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)}{2\,\mathrm{ch}\,\frac{u}{j}}\right]du.$$

The first term on the right hand side is 0 since

$$\frac{d}{d\nu}\left(2j\,\mathrm{sh}\,\frac{u}{j}\right) = 2j'\,\mathrm{sh}\,\frac{u}{j} - 2u\frac{j'}{j}\,\mathrm{ch}\,\frac{u}{j} \tag{15}$$

and for $K_0(x)$ we have

$$K_0(x) = \begin{cases} O\left(\log\frac{1}{x}\right) & \text{for} \quad x>0,\ x\sim 0 \\ o(e^{-x}) & \text{for} \quad x\gg 1. \end{cases}$$

A little manipulation yields that

$$j'' = -2\int_0^\infty K_0\left(2j\,\mathrm{sh}\,\frac{u}{j}\right)e^{-\frac{2\nu}{j}u}\left\{-2\nu\frac{j'}{j}\,\mathrm{th}\,\frac{u}{j} - \frac{j'}{j}\,\mathrm{th}^2\frac{u}{j} + 2\frac{u}{j}\right\}du. \tag{16}$$

Let the function in braces be denoted by I, then

$$I = 2t - 2\nu\frac{j'}{j}\,\mathrm{th}\,t - \frac{j'}{j}\,\mathrm{th}^2 t,$$

where $t=u/j$. If we show that $I>0$ then by (16) we prove the Theorem. Since $j'>0$ and $\operatorname{th} t<1$ therfore

$$I > 2t-2v\frac{j'}{j}\operatorname{th} t-\frac{j'}{j}\operatorname{th} t = 2t-(2v+1)\frac{j'}{j}\operatorname{th} t.$$

This implies that if $v\leqq -1/2$ then $I>2t>0$. If $-1/2<v<\infty$ then by the Lemma $\left(v+\frac{1}{2}\right)j'<j$ and making use of the inequality $\operatorname{th} t<t$ for $t>0$ we have

$$I > 2t\left[1-\left(v+\frac{1}{2}\right)\frac{j'}{j}\right] > 0.$$

Thus the proof of the Theorem is complete.

REFERENCES

[1] WATSON, G. N.: *A treatise on the theory of Bessel functions*, Second ed., Cambridge University Press, 1944.
[2] ERDÉLYI, A. et al., *Higher transcendental functions*, Vol. 2, McGraw-Hill, New York, 1954.
[3] LEWIS, J. T. and MULDOON, M. E.: Monotonicity and convexity properties of zeros of Bessel functions, *SIAM J. Math. Anal.* **8** (1977), 171—178.
[4] TRICOMI, F.: Sulle funzioni di Bessel di ordine e argomento pressoche uguali, *Atti Accad. Sci. Torino Cl. Sci. Fis. Mat. Nat.* **83** (1949), 3—20.
[5] MAKAI, E.: On zeros of Bessel functions (to appear).
[6] MCCANN, R. C.: Inequalities for the zeros of Bessel functions, *SIAM J. Math. Anal.* **8** (1977) 166—170
[7] BIEBERBACH, *Funktionentheorie*, I

Mathematical Institute of the Hungarian Academy of Sciences,
H-1053 Budapest, Reáltanoda u. 13—15.

(Received October 26, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 89—93

KERNEL OF A HOMOTOPY

by

P. DAS

1. Introduction

A quasigroup G is a system of three compositions called product ('$\cdot$'), right division ('$/$') and left division ('$\backslash$') such that $a\cdot b=c\Leftrightarrow c/b=a\Leftrightarrow c\backslash a=b$ for every $a, b, c\in G$. A topological quasigroup (G, τ) is a quasigroup G endowed with a topology τ w.r. to which the operations of G are continuous. A topological quasigroup having an identity element is said to be a topological loop.

A homotopy of a quasigroup G_1 into a quasigroup G_2 is a triple $h=(\alpha_1, \alpha_2, \alpha_3)$ where $\alpha_1, \alpha_2, \alpha_3$: $G_1\to G_2$ satisfy

$$(x)\alpha_1\cdot(y)\alpha_2=(x\cdot y)\alpha_3 \quad \text{for every} \quad x, y\in G_1. \tag{1}$$

Replacing x, y by y/x, x and x, $y\backslash x$ in (1) we get that

$$(y)\alpha_3/(x)\alpha_2=(y/x)\alpha_1, \tag{2}$$

$$(y)\alpha_3\backslash(x)\alpha_1=(y\backslash x)\alpha_2 \tag{3}$$

for every $x, y\in G_1$.

If (G_1, τ_1) and (G_2, τ_2) be topological quasigroups, then h is a homotopy if, in addition, $\alpha_1, \alpha_2, \alpha_3$: $(G_1, \tau_1)\to(G_2, \tau_2)$ are continuous.

In the present paper the notion of kernel of a homotopy has been introduced and some interesting generalisations of the results concerning the kernel of a homomorphism and those concerning the normal equivalence relation defined by a homomorphism have been obtained.

2. Kernel of a homotopy

(2.1) Let $h=(\alpha_1, \alpha_2, \alpha_3)$ be a homotopy of a quasigroup G_1 into a loop G_2. Let $K_i=\{x\in G_1;\ (x)\alpha_i=1\}$ for $i=1, 2, 3$. Then the triple (K_1, K_2, K_3) is said to be *the kernel of the homotopy h*. We note that when h is a homomorphism, $\alpha_1=\alpha_2=\alpha_3$ and then $K_1=K_2=K_3=K$ (say) is the kernel of the homomorphism h.

In what follows we shall study some relations between K_1, K_2 and K_3.

(2.2) *It holds that*

$$K_1\cdot K_2\subset K_3. \tag{4}$$

For, by (1), $x\in K_1, y\in K_2\Rightarrow(x)\alpha_1=1, (y)\alpha_2=1\Rightarrow(x\cdot y)\alpha_3=x\alpha_1\cdot y\alpha_2=1\cdot 1=1\Rightarrow$ $\Rightarrow x\cdot y\in K_3$.

Similarly using (2) and (3), it can be shown that

$$K_3/K_2 \subset K_1, \tag{5}$$

$$K_3\backslash K_1 \subset K_2. \tag{6}$$

Combining (4), (5), (6), we get that

$$K_1 \cdot K_2 = K_3; \quad K_3/K_2 = K_1; \quad K_3\backslash K_1 = K_2. \tag{7}$$

(2.3) *From* (2) *and* (3) *it follows that for every* $a\in G_1$, $a\backslash a\in K_2$ *if, and only if* $\alpha_1=\alpha_3$ *and* $a/a\in K_1$ *if, and only if* $\alpha_2=\alpha_3$.

(2.4) *If* $\alpha_1=\alpha_2$, *then* $K_1\cdot a=a\cdot K_1$ *for every* $a\in G_1$.

PROOF. Let $x\in K_1$ and $a\in G_1$ and let $y=(x\cdot a)\backslash a$. Then $(y)\alpha_1=(y)\alpha_2=$ $=(x\cdot a)\alpha_3\backslash(a)\alpha_1$ [by (3)]$=1$ ($\because x\in K_1$ and $\alpha_1=\alpha_2$). $\therefore y\in K_1$ and $x\cdot a=a\cdot y\in a\cdot K_1$. $\therefore K_1\cdot a\subset a\cdot K_1$. Similarly $a\cdot K_1\subset K_1\cdot a$. Combining we get that $K_1\cdot a=a\cdot K_1$.

(2.5) *If* $\alpha_1=\alpha_3$ *and if* α_2 *is a homomorphism, then*

$$K_1(a\cdot b) = (K_1\cdot a)\cdot b \tag{8}$$

for every $a, b\in G_1$.

PROOF. Let $x\in K_1$ and let $y=[(x\cdot a)\cdot b]/(a\cdot b)$. Then $(y)\alpha_1=[(x\cdot a)\cdot b]\alpha_3/(a\cdot b)\alpha_2$ [by (2)]$=1$. For, $[(x\cdot a)\cdot b]\alpha_3=(x\cdot a)\alpha_1\cdot(b)\alpha_2$ [by (1)]$=[(x)\alpha_1\cdot(a)\alpha_2]\cdot(b)\alpha_2$ [by (1) since $\alpha_1=\alpha_3$]$=(a\cdot b)\alpha_2$ [$\because x\in K_1$ and α_2 is a homomorphism].

$$\therefore y\in K_1 \text{ and } (x\cdot a)\cdot b=y\cdot(a\cdot b)\in K_1(a\cdot b), \quad \therefore (K_1\cdot a)\cdot b\subset K_1\cdot(a\cdot b). \tag{9}$$

Again let $x\in K_1$ and let $y\in G_1$ be such that $x\cdot(a\cdot b)=(y\cdot a)\cdot b$.

Then $[(y)\alpha_1\cdot(a)\alpha_2]\cdot(b)\alpha_2=(y\cdot a)\alpha_3\cdot(b)\alpha_2$ [by (1)]$=[(y\cdot a)\cdot b]\alpha_3$ [by (1) since $\alpha_1=\alpha_3$]$=[x\cdot(a\cdot b)]\alpha_3=(x)\alpha_1\cdot(a\cdot b)\alpha_2$ [by (1)]$=[1\cdot(a)\alpha_2]\cdot(b)\alpha_2$ [$\because x\in K_1$, and α_2 is a homomorphism]

$$\therefore (y)\alpha_1 = 1\cdot \therefore y\in K_1,$$

$$\therefore x\cdot(a\cdot b)\in(K_1\cdot a)\cdot b, \quad \therefore K_1\cdot(a\cdot b)\subset(K_1\cdot a)\cdot b. \tag{10}$$

From (9) and (10), we get (8).

Similarly it can be shown that if $\alpha_2=\alpha_3$ and if α_1 is a homomorphism, then

$$b\cdot(a\cdot K_2) = (b\cdot a)\cdot K_2 \tag{11}$$

for every $a, b\in G_1$.

(2.6) *Let* h *be a homomorphism with the kernel* K. *Then* $\alpha_1=\alpha_2=\alpha_3$ *and* $K_1=K_2=$ $=K_3=K$.

$\therefore$ It follows from (2.2) that $x, y\in K\Rightarrow x\cdot y$, y/x $y\backslash x\in K$. Also it follows from (2.3), (2.4) and (2.5) that for every $a, b\in G_1$, a/a, $a\backslash a\in K$; $K\cdot a=a\cdot K$; $K\cdot(a\cdot b)=$ $=(K\cdot a)\cdot b$; $b\cdot(a\cdot K)=(b\cdot a)\cdot K$. $\therefore K$ is a strictly normal subquasigroup of G_1. (A *strictly normal subquasigroup* of a quasigroup G_1 is a subquasigroup H of G_1 such that a/a, $a\backslash a\in H$; $H\cdot a=a\cdot H$; $(H\cdot a)\cdot b=H\cdot(a\cdot b)$ and $b\cdot(a\cdot H)=(b\cdot a)\cdot H$ for every $a, b\in G_1$ (DAS [3]).)

(2.7) *Let* $h=(\alpha_1, \alpha_2, \alpha_3)$ *be a homotopy of a topological quasigroup* (G_1, τ_1) *into a topological quasigroup* (G_2, τ_2) *and let* (K_1, K_2, K_3) *be the kernel of* h. *Then* $(x, y)\to x\cdot y$ $(x\in K_1,\ y\in K_2)$ *is a continuous mapping of* $K_1\times K_2$ *onto* K_3 [*considering* K_i *as subspaces of the topological space* (G_1, τ_1)].

PROOF. Let W be a neighbourhood of $x\cdot y$ in K_3. Then there exists a neighbourhood W' of $x\cdot y$ in G_1 such that $W=W'\cap K_3$. Since (G_1, τ_1) is a topological quasigroup, there exist neighbourhoods U', V' of x, y in G_1 such that $U'\cdot V'\subset W'$. Then $U=U'\cap K_1$, $V=V'\cap K_2$ are neighbourhoods of x, y in K_1, K_2 respectively and $U\cdot V\subset W$. $\therefore (x\cdot y)\to x\cdot y$ is a continuous mapping of $K_1\times K_2$ onto K_3.

Similarly it can be shown that $(x, y)\to y/x$ $(x\in K_2,\ y\in K_3)$ and $(x, y)\to y\backslash x$ $(x\in K_1,\ y\in K_3)$ are continuous mappings of $K_2\times K_3$ onto K_1 and of $K_1\times K_3$ onto K_2 respectively.

In particular, if h be a homomorphism with kernel K, it follows from (2.6) and (2.7) that K is a strictly normal subquasigroup of the topological quasigroup (G_1, τ_1).

3. Equivalence relations defined by a homotopy

(3.1) *Let* $h=(\alpha_1, \alpha_2, \alpha_3)$ *be a homotopy of a quasigroup* G_1 *into a quasigroup* G_2. *Let* $P_i=\{(a, b)\in G_1\times G_1;\ (a)\alpha_i=(b)\alpha_i\}$ *for* $i=1, 2, 3$. *Then* P_1, P_2, P_3 *are equivalence relations in* G_1 *which satisfy the following properties:*

(i) $(c\cdot a)P_3(c\cdot b)\Rightarrow aP_2b$. For, $(c\cdot a)P_3(c\cdot b)\Rightarrow(c\cdot a)\alpha_3=(c\cdot b)\alpha_3\Rightarrow(c)\alpha_1\cdot(a)\alpha_2=$

$=(c)\alpha_1\cdot(b)\alpha_2$ [by (1)]$\Rightarrow(a)\alpha_2=(b)\alpha_2\Rightarrow aP_2b$.

(ii) *Similarly* $(a\cdot c)P_3(b\cdot c)\Rightarrow aP_1b$.

(iii) $aP_1b,\ cP_2d\Rightarrow(a\cdot c)P_3(b\cdot d)$.

For, $aP_1b,\ cP_2d\Rightarrow(a)\alpha_1=(b)\alpha_1,\ (c)\alpha_2=(d)\alpha_2\Rightarrow(a\cdot c)\alpha_3=(a)\alpha_1\cdot(c)\alpha_2=$ $=(b)\alpha_1\cdot(d)\alpha_2=(b\cdot d)\alpha_3$ [by (1)]$\Rightarrow(a\cdot c)P_3(b\cdot d)$.

Using (2) *and* (3) *one can verify that*

(iv) $(c/a)P_1(c/b)\Rightarrow aP_2b$;

(v) $(c\backslash a)P_2(c\backslash b)\Rightarrow aP_1b$;

(vi) $(a/c)P_1(b/c)\Rightarrow aP_3b$;

(vii) $(a\backslash c)P_2(b\backslash c)\Rightarrow aP_3b$;

(viii) $aP_3b,\ cP_2d\Rightarrow(a/c)P_1(b/d)$;

(ix) $aP_3b,\ cP_1d\Rightarrow(a\backslash c)P_2(b\backslash d)$.

(3.2) *Let* G_1/P_i *be the quotient set of* G_1 *w.r. to* P_i *and let* (a, P_i) *be the element in* G_1/P_i *to which* a *belongs. Let* p_i *be the canonical mapping of* G *onto* G_1/P_i *and let*

ψ_i: $G_1/P_i \to G_2$ *be defined by the relation* $p_i\psi_i=\alpha_i$. *Then from* (1), (2) *and* (3), *we get that*

$$(x, P_1)\psi_1 \cdot (y, P_2)\psi_2 = (x \cdot y, P_3)\psi_3;$$

$$(y, P_3)\psi_3/(x, P_2)\psi_2 = (y/x), P_1\psi_1;$$

$$(y, P_3)\psi_3\backslash(x, P_1)\psi_1 = (y\backslash x, P_2)\psi_2.$$

(3.3) *For every* $x, y \in G_1$

$$(x, P_1) \cdot (y, P_2) = (x \cdot y, P_3) \quad \text{i.e.} \quad (x)p_1 \cdot (y)p_2 = (x \cdot y)p_3. \tag{12}$$

PROOF. $x' \in (x, P_1)$, $y' \in (y, P_2) \Rightarrow x'p_1x$, $y'p_2y \Rightarrow (x' \cdot y')p_3(x \cdot y)$ [by (3.1) (iii)] $\Rightarrow$ $\Rightarrow x' \cdot y' \in (x \cdot y, P_3)$.

$$\therefore (x, P_1) \cdot (y, P_2) \subset (x \cdot y, P_3). \tag{13}$$

Again $z \in (x \cdot y, P_3) \Rightarrow (x \cdot y')p_3(x \cdot y)$ where $y' = z\backslash x \Rightarrow y'p_2y$ [by (3.1) (i)] $\Rightarrow z =$ $= x \cdot y' \in (x, P_1) \cdot (y, P_2)$.

$$\therefore (x \cdot y, P_3) \subset (x, P_1) \cdot (y, P_2). \tag{14}$$

Combining (14) and (15), we get (13).

Similarly one can show that for every $x, y \in G_1$

$$(y, P_3)/(x, P_2) = (y/x, P_1) \quad \text{i.e.} \quad (y)p_3/(x)p_2 = (y/x)p_1 \tag{15}$$

and

$$(y, P_3)\backslash(x, P_1) = (y\backslash x, P_2) \quad \text{i.e.} \quad (y)p_3\backslash(x)p_1 = (y\backslash x)p_2. \tag{16}$$

(3.4) *Let* h *be a homomorphism. Then* $\alpha_1=\alpha_2=\alpha_3$; $P_1=P_2=P_3=P$ *(say);* $p_1=$ $=p_2=p_3=p$ *(say); and* $\psi_1=\psi_2=\psi_3=\psi$ *(say). It follows from* (3.1) *that* P *is a normal equivalence relation* (BRUCK [1]). *From* (3.2) *and* (3.3) *it follows that* G_1/P *is a quasigroup;* p *is a homomorphism of* G_1 *onto* G_1/P *and if* h *be surjective then* ψ *is an isomorphism of* G_1/P *onto* G_2.

(3.5) *Let* $h=(\alpha_1, \alpha_2, \alpha_3)$ *be a homotopy of a topological quasigroup* (G_1, τ_1) *into a topological quasigroup* (G_2, τ_2). *With the notations of* (3.1) *and* (3.2), τ_{1i}, *defined by the relation*

$$p_i\tau_{1i} = \tau_1 p_i \tag{17}$$

is a topology on G_1/P_i *for* $i=1, 2, 3$. *Also* p_i *is a continuous and open mapping of* (G_1, τ_1) *onto* $(G_1/P_i, \tau_{1i})$ *and if* α_i *be surjective and open, then* ψ_i *is a homeomorphism of* $(G_1/P_i, \tau_{1i})$ *onto* (G_2, τ_2) *for* $i=1, 2, 3$.

Also for every $x, y \in G_1$,

$$(x, P_1)\tau_{11} \cdot (y, P_2)\tau_{12} \leqq (x \cdot y, P_3)\tau_{13}.$$

For, $(x, P_1)\tau_{11} \cdot (y, P_2)\tau_{12} = (x)\tau_1 p_1 \cdot (y)\tau_1 p_2 = [(x)\tau_1 \cdot (y)\tau_1]p_3 \leqq$ [by (12)] $[(x \cdot y)\tau_1]p_3 = (x \cdot y, P_3)\tau_{13}$.

Similarly one can verify that for every $x, y \in G_1$,

$$(y, P_3)\tau_{13}/(x, P_2)\tau_{12} \leqq (y/x, P_1)\tau_{11}$$

and

$$(y, P_3)\tau_{13}\backslash(x, P_1)\tau_{11} \leqq (y\backslash x, P_2)\tau_{12}.$$

In particular, if h be a homomorphism, then it follows from (3.4) and above that $(G_1/P, \tau_1')$ where $\tau_{11}=\tau_{12}=\tau_{13}=\tau_1'$ (say) is a topological quasigroup, p is an open homomorphism of (G_1, τ_1) onto $(G_1/P, \tau_1')$ and if h is surjective and open, then ψ is an isomorphism of $(G_1/P, \tau_1')$ onto (G_2, τ_2).

(3.6) *Let P_1, P_2, P_3 be three equivalence relations in a topological quasigroup (G_1, τ_1) satisfying the conditions* (i)—(ix) *of* (3.1).

$'\cdot'$: $G_1/P_1\times G_1/P_2\to G_1/P_3$; $'/'$: $G_1/P_2\times G_1/P_3\to G_1/P_1$ and $'\backslash'$: $G_1/P_1\times G_1/P_3\to$ $\to G_1/P_2$ are defined by the relations (12), (15) *and* (16) *respectively where p_i are the canonical mappings of G_1 onto G_1/P_i. Topologies τ_{1i} are defined in G_1/P_i by the relation* (17). *Then $'\cdot'$, $'/'$, $'\backslash'$ are continuous.*

Also $(a)p_i=(b)p_i$ if, and only if aP_ib hold. Thus if $P_1=P_2=P_3=P$ (say), then G_1/P is a topological quasigroup and $p_1=p_2=p_3=p$ (say) is an open homomorphism of G_1 onto G_1/P such that normal equivalence relation defined by p is P.

Acknowledgement. I offer my grateful thanks to DR. A. C. CHOUDHURY for his kind help and guidance in the preparation of this paper.

REFERENCES

[1] BRUCK, R. H.: *A survey of binary systems*, Springer Verlag, 1958.
[2] DAS, P.: A note on homotopy and isotopy of topological groupoids, *Progress of Mathematics* **1** (1967), 21—28.
[3] DAS, P.: Topological Quasigroups *(Thesis submitted to the University of Calcutta)*, 1968.

Department of Mathematics, Visva-Bharati University, Santiniketan 731235, West Bengal, India

(Received May 12, 1972, revised Sept. 12, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 95—99

МНОГОТОЧЕЧНАЯ КРАЕВАЯ ЗАДАЧА ДЛЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ СВЕРХНЕЙТРАЛЬНОГО ТИПА

М. М. КОНСТАНТИНОВ и Д. Д. БАЙНОВ

1. Постановка задачи

Рассмотрим краевую задачу

$$\dot{x}(t) = f\big(t, x(t), \dot{x}(t), x(\tau_0), \dot{x}(\tau_0)\big), \quad t \in [t_1, t_N] = I \tag{1}$$

$$\sum_{i=1}^{N} A_i x(t_i) = \Gamma\big(x(t_1), \ldots, x(t_N)\big); \quad t_1, \ldots, t_N \in I,$$

где $x=(x^1, \ldots, x^n)$, $f=(f^1, \ldots, f^n)$, $\Gamma=(\Gamma^1, \ldots, \Gamma^n)$, а $A_i=(a_i^{jk})$-постоянные $(n\times n)$-матрицы. Преобразованный аргумент τ_0 определим следующим образом

$$\tau_k = \tau_k\big(t, x(t), \dot{x}(t), x(\tau_{k+1}), \dot{x}(\tau_{k+1})\big);$$

$$k = 0, \ldots, m-1; \quad \tau_m = \tau_m\big(t, x(t), \dot{x}(t)\big).$$

Предположим, что функции $f(t, \xi)$, $\tau_s(t, \xi)$, $s=0, \ldots, m$, где $\xi=(x, y, u, v)$, определены в области

$$Q = I\times D = I\times D_1\times D_2\times D_1\times D_2, \quad D_i = \{x: \|x\| \leqq d_i\}$$

($\|\cdot\|$ — некоторая норма в n-мерном вещественном пространстве R^n), а функция $\Gamma(\Xi)$, $\Xi=(\Xi_1, \ldots, \Xi_n)$ в области D_1^N.

Положим $\sup\{\|f(t, \xi)\|: \xi\in D\}=F(t)$. Далее будем предполагать, что выполнены условия (A):

A1. Матрица $A=A_1+\ldots+A_N$ — неособая.

A2. Выполнены неравенства $a\alpha+\Phi+\gamma\leqq d_1$, $\omega\leqq d_2$, где

$$a = \|A^{-1}\|, \quad \alpha = \min\left\{\sum_{i=1}^{N} a_i \left|\int_{t_s}^{t_i} F(t)\,dt\right|: s\in\overline{1, N}\right\} =$$

$$= \sum_{i=1}^{N} a_i \left|\int_{t_q}^{t_i} F(t)\,dt\right|, \quad a_i = \|A_i\|,$$

$$\Phi = \max\left\{\int_{J_1} F(t)\,dt, \int_{J_2} F(t)\,dt\right\} = \int_{J} F(t)\,dt,$$

$$J_1 = [t_1, t_q], \quad J_2 = [t_q, t_N],$$

$$\gamma = \sup\{\|A^{-1}\Gamma(\Xi)\|: \Xi\in D_1^N\},$$

$$\omega = \sup\{F(t): t\in I\}.$$

A3. В области Q функции $f(t, \xi)$ и $\tau_s(t, \xi)$, $s=0, \ldots, m$, удовлетворяют условиям Липшица

$$\|f(t, \xi)-f(\bar{t}, \bar{\xi})\| \leqq E|t-\bar{t}|+L_1\|x-\bar{x}\|+M_1\|y-\bar{y}\|+$$

$$+L_2\|u-\bar{u}\|+M_2\|v-\bar{v}\|; \quad \bar{\xi}=(\bar{x}, \bar{y}, \bar{u}, \bar{v});$$

$$|\tau_s(t, \xi)-\tau_s(\bar{t}, \bar{\xi})| \leqq \varepsilon|t-\bar{t}|+\lambda_1\|x-\bar{x}\|+\mu_1\|y-\bar{y}\|+\lambda_2\|u-\bar{u}\|+\mu_2\|v-\bar{v}\|.$$

A4. В области D_1^N функция $\Gamma(\Xi)$ удовлетворяет условиям Липшица

$$\|\Gamma(\Xi)-\Gamma(\bar{\Xi})\| \leqq \sum_{i=1}^{N} \gamma_i\|\Xi_i-\bar{\Xi}_i\|, \quad \bar{\Xi}=(\bar{\Xi}_1, \ldots, \bar{\Xi}_N),$$

причем $\gamma_0=a(\gamma_1+\ldots+\gamma_N)<1$.

A5. Кообласть функций $\tau_s(t, \xi)$, $s=0, \ldots, m$ принадлежит интервалу I.

2. Существование решений краевой задачи

Теорема 1. *Пусть выполнены условия* (A). *Пусть кроме того*

$$M_1^* = 1-M_1 > 0, \quad \lambda_2^* = 1-\lambda_2\omega > 0, \quad c_1 > 0,$$

$$c_1^2 \geqq 4c_0c_2, \quad \mu_2c_1 \leqq 2\lambda_2^*c_2,$$

причем $\mu_2c_1<2\lambda_2^*c_2$ *при* $c_1^2=4c_0c_2$, *где*

$$c_0 = \lambda_2^*E^*+L_2^*\varepsilon^*, \quad c_1 = \mu_2E^*+M_1^*\lambda_2^*-M_2\varepsilon^*-\mu_1L_2^*,$$

$$c_2 = \mu_1M_2+\mu_2M_1^*, \quad E^* = E+L_1\omega, \quad \varepsilon^* = \varepsilon+\lambda_1\omega, \quad L_2^* = L_2\omega.$$

Тогда краевая задача (I) *имеет хотя бы одно непрерывно дифференцируемое решение* x. *и* $\|x(t)\| \leqq d_1$ *при* $t \in I$.

Доказательство. Рассмотрим пространство B непрерывных n-мерных функций $z: I \to R^n$ с метрикой, порожденной нормой

$$\|z\|_B = \sup\{\|z(t)\|: t \in I\}.$$

Определим Ω как множество функций $z \in B$, удовлетворяющих условиям

$$\|z(t)\| \leqq F(t), \tag{2}$$

$$\|z(t)-z(\bar{t})\| \leqq \beta|t-\bar{t}|, \tag{3}$$

где

$$\beta = \frac{1}{2c_2}(c_1-\sqrt{c_1^2-4c_0c_2}).$$

Очевидно множество Ω выпукло, замкнуто и компактно.

Пусть оператор Π действует в Ω по формуле

$$\Pi z(t) = f(t, x(t), z(t), x(\tau_0), z(\tau_0)),$$

где

$$x(t) = b + \int_{t_q}^{t} z(s)\, ds$$

и

$$b = U(b, z) = -A^{-1} \sum_{i=1}^{N} A_i \int_{t_q}^{t_i} z(t)\, dt +$$

$$+ A^{-1} \Gamma \left(b + \int_{t_q}^{t_1} z(t)\, dt, \ldots, b + \int_{t_q}^{t_N} z(t)\, dt \right).$$

Отметим, что в силу условий A2 и A4 уравнение $b = U(b, z)$ имеет единственное решение $b^* = b^*(z)$, $\|b^*\| \leqq a\alpha + \gamma$.

Непосредственно проверяется, что операторное уравнение $z = \Pi z$ эквивалентно краевой задаче (1).

Покажем, что оператор Π преобразует множество Ω в себя. Действительно, функция Πz удовлетворяет условию (2) при каждом $z \in \Omega$, а $x(t)$ остается в области D_1 при $t \in I$.

Оценим число $\Delta = \|\Pi z(t) - \Pi z(\bar{t})\|$ при фиксированных $t, \bar{t} \in I$ и $z \in \Omega$. Имеем

$$\Delta \leqq E|t - \bar{t}| + L_1 \|x(t) - x(\bar{t})\| + M_1 \|z(t) - z(\bar{t})\| + \tag{4}$$

$$+ L_2 \|x(\tau_0) - x(\bar{\tau}_0)\| + M_2 \|z(\tau_0) - z(\bar{\tau}_0)\|,$$

где через $\bar{\tau}_s$, $s = 0, \ldots, m$, обозначена функция τ_s, в которой t следует заменить на $\bar{t}$.

Так как $\|x(t) - x(\bar{t})\| \leqq \omega |t - \bar{t}|$, то из (4) получаем

$$\Delta \leqq (E^* + M_1\beta)|t - \bar{t}| + (L_2^* + M_2\beta)|\tau_0 - \bar{\tau}_0|.$$

Аналогичным образом

$$|\tau_k - \bar{\tau}_k| \leqq (\varepsilon^* + \mu_1\beta)|t - \bar{t}| + (\lambda_2\omega + \mu_2\beta)|\tau_{k+1} - \bar{\tau}_{k+1}|;$$

$$k = 0, \ldots, m-1; \quad |\tau_m - \bar{\tau}_m| \leqq (\varepsilon^* + \mu_1\beta)|t - \bar{t}|.$$

Следовательно

$$|\tau_0 - \bar{\tau}_0| \leqq \frac{\varepsilon^* + \mu_1\beta}{\lambda_2^* - \mu_2\beta} |t - \bar{t}|, \tag{5}$$

так как в силу условий теоремы $\lambda_2^* > \mu_2\beta$.

Подставляя (5) в (4) приходим к оценке

$$\Delta \leqq \left(E^* + M_1\beta + (L_2^* + M_2\beta) \frac{\varepsilon^* + \mu_1\beta}{\lambda_2^* - \mu_2\beta} \right) |t - \bar{t}| = \beta |t - \bar{t}|.$$

Итак условие (3) имеет место для функции Πz, $z \in \Omega$.

Покажем наконец, что оператор Π непрерывен на Ω. Пусть $z, \tilde{z} \in \Omega$. Тогда

$$
\begin{aligned}
(6) \qquad \|\Pi z - \Pi \tilde{z}\| &\leqq L_1\|x-\tilde{x}\| + M_1\|z-\tilde{z}\| + \\
&+ L_2\|x(\tau_0)-\tilde{x}(\tau_0)\| + L_2\|\tilde{x}(\tau_0)-\tilde{x}(\tilde{\tau}_0)\| + \\
&+ M_2\|z(\tau_0)-\tilde{z}(\tau_0)\| + M_2\|\tilde{z}(\tau_0)-\tilde{z}(\tilde{\tau}_0)\| \leqq \\
&\leqq (L_1+L_2)\|x-\tilde{x}\| + (M_1+M_2)\|z-\tilde{z}\| + \\
&+ (L_2^* + M_2\beta)|\tau_0-\tilde{\tau}_0|,
\end{aligned}
$$

где через $\tilde{\tau}_s$, $s=0, \ldots, m$, обозначена функция τ_s, в которой x и z следует заменить на $\tilde{x}$ и $\tilde{z}$.

Аналогичным образом имеем

$$
\begin{aligned}
|\tau_k-\tilde{\tau}_k| &\leqq (\lambda_1+\lambda_2)\|x-\tilde{x}\| + (\mu_1+\mu_2)\|z-\tilde{z}\| + \\
&+ (\lambda_2\omega+\mu_2\beta)|\tau_{k+1}-\tilde{\tau}_{k+1}|; \quad k=0, \ldots, m-1; \\
|\tau_m-\tilde{\tau}_m| &\leqq \lambda_1\|x-\tilde{x}\| + \mu_1\|z-\tilde{z}\|,
\end{aligned}
$$

откуда

$$
(7) \qquad |\tau_0-\tilde{\tau}_0| \leqq \frac{(\lambda_1+\lambda_2)\|x-\tilde{x}\|_B + (\mu_1+\mu_2)\|z-\tilde{z}\|_B}{\lambda_2^* - \mu_2\beta}.
$$

С другой стороны

$$
\|x-\tilde{x}\|_B \leqq \|b-\tilde{b}\| + m_I\|z-\tilde{z}\|_B, \quad m_I = \operatorname{mes} I,
$$

$$
\|b-\tilde{b}\| \leqq \frac{a\sigma}{1-\gamma_0}\|z-\tilde{z}\|_B,
$$

$$
\sigma = \sum_{i=1}^{N}(a_i+\gamma_i)|t_i-t_q|.
$$

Следовательно

$$
(8) \qquad \|x-\tilde{x}\|_B \leqq \left(m_I + \frac{a\sigma}{1-\gamma_0}\right)\|z-\tilde{z}\|_B = T\|z-\tilde{z}\|_B.
$$

Доставляя (7) и (8) в (6) находим

$$
(9) \qquad \|\Pi z - \Pi\tilde{z}\|_B \leqq (M+TL)\|z-\tilde{z}\|_B,
$$

где

$$
(10) \qquad M = M_1 + M_2 + \frac{L_2^* + M_2\beta}{\lambda_2^* - \mu_2\beta}(\mu_1+\mu_2),
$$

$$
L = L_1 + L_2 + \frac{L_2^* + M_2\beta}{\lambda_2^* - \mu_2\beta}(\lambda_1+\lambda_2).
$$

Итак мы показали, что оператор Π, непрерывный на Ω, преобразует это выпуклое, замкнутое и компактное множество в себя. Отсюда в силу принципа Шаудера о неподвижной точке получаем утверждение теоремы.

3. Единственность решения краевой задачи

Теорема 2. *Пусть выполнены условия теоремы 1. Пусть кроме того* $M+TL<1$.

Тогда краевая задача (1) *имеет единственное непрерывно дифференцируемое решение* x, *причем* $\|x\|_B \leqq d_1$, $\dot{x} \in \Omega$ *и* $\lim\limits_{p\to\infty} x_p(t)=x(t)$, $x_p=\Pi x_{p-1}$, *равномерно по* $t\in I$ *при* $\|x_0\|_B \leqq d_1$ *и* $\dot{x}_0 \in \Omega$.

Утверждение теоремы непосредственно следует из (9) и (10) в силу принципа Банаха о неподвижной точке.

4. Некоторые замечания

1. Запаздывание вида $\tau_0(\cdot)$ (т. н. «итерированное» запаздывание) введено впервые в [1], [2] при исследовании некоторых вопросов фундаментальной теории начальных задач для уравнения сверхнейтрального типа (т. е. дифференциальное уравнение нейтрального типа с запаздыванием, зависящим от производной решения [3]).

2. Теорема 1 имеет глобальный характер, потому что там никакие специальные ограничения на интервале I не накладываются. Кроме того, по-видимому, эту теорему нельзя заметно улучшить.

3. Теорему 2 можно сформулировать и следующим образом: пусть выполнены условия теоремы 1. Пусть кроме того $M<1$. Тогда если число $\operatorname{mes} I$ достаточно мало, то краевая задача (1) имеет единственное решение. Следовательно эта теорема докального существования и единственности решения. Условия теоремы 2 можно несколько улучшить подходящим выбором индекса q или применением другой метрики в пространстве n-мернхы функций (см. напр. [4]).

ЛИТЕРАТУРА

[1] Константинов М. М., Байнов Д. Д.: Теоремы существования и единственности решения некоторых дифференциальных уравнений сверхнейтрального типа, *Publ. Inst. Math. (Beograd)*, **14** (28) (1972),

[2] Константинов М. М., Байнов Д. Д.: Существование и единственность решения некоторых экстремально — дифференсиальных систем сверхнейтрального типа, *Arch. Math. (Brno)*, **4** VIII, (1972),

[3] Скрипник В. П.: Об уравнениях с преобразованным аргументом нейтрального типа УМЖ, т. 22, № 5, 1970.

[4] Bielicki A.: Une remarque sur la méthode de Banach—Cacciopoli—Tikhonov, *Bull. Acad. Polon. Sci.*, IV, 5, 1956.

Высший машинно-электротехнический институт имени В. И. Ленина — София
Пловдивский университет имени П. Хилендарского

(Поступила 10 августа 1974)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 101—116

ON VALID ASSERTIONS — IN PROBABILITY LOGIC

by

M. FERENCZI

§ 0.

The well-known Löwenheim theorem which was the starting point of the Löwenheim—Skolem theorem group, deals with reducing the problem whether a formula of first-order logic holds on every model to the other problem whether this formula holds on certain (namely, countable) models.

Hitherto research in probability logic, a generalization of first-order logic, has essentially created the concepts necessary for the discussion of an analogous problem. The resulting theorems are similar to the Löwenheim theorem. Using such concepts (or slight modifications thereof), the present paper is aimed at answering a question of the said type. To this extent, the present work is related to the research in probability logic presented in (2) and (4). Parallelly to the proofs of existence theorems, we shall, however, pay special attention to the search of practical procedures testing universal satisfaction. In spite of probability theoretic aspects, our point of view will be consistently that of model theory, using the terminology of probability systems (probability models), which, in a sense, are generalizations of models.

Naturally, the analogy with the classical Löwenheim reduction question group is only partial. In probability logic from the point of view of the universal satisfiability of the formulas — similarly to the case of Boolean models (cf. (10)) — the problem of reducing the class of models arises not only through the reduction of the power of the universe, but parallel with this, through the restrictions imposed upon the structures, the latter essentially characterizing the range of truth values. Already this provides a justification for first restricting our investigation to propositional language. Beyond this we shall show the exceptional relevance of the case of propositional language to our problem by proving that, somewhat surprisingly, the problem of universal satisfiability in the predicate language can be reduced, in some sense, to the case of propositional language.

Through the propositional calculus, we shall see, beyond their theoretical interest, the immediate use of the results. Namely, we shall state a reduction theorem involving a practical method for deciding universal satisfiability.

§ 1 gives the necessary concepts and definitions; these involve the finitary logic language, the concept of probability systems corresponding to ordinary models defined for this language; the concept of valuation on such systems; the concept of probability assertions corresponding to propositions of first order logic, along with the concept of their universal satisfaction.

In § 2 the case of probability logic based on classical propositional language is discussed in the following order:

In § 2.1 we prove the first mentioned reduction theorem. This is followed by hints to some further possibility of reduction. Finally, we shall construct an example which, in a certain sense indicates the bounds of further reduction. In § 2.2 we shall give a sufficient condition for the universal satisfaction of propositions of probability logic. The corresponding theorem is based again on the reduction of the involved circle of models, which is the leading idea of the paper.

This theorem provides a useful tool for decidings universal satisfiability in concrete cases.

Thus we obtain a procedure based on entirely new theoretical considerations, which, for a wide range of propositions is easier to handle than the previous methods (4), (7).

§ 3 deals with the case of probability logic based on classical predicate language. It is shown here that the problem of universal satisfiability of a probability assertion in the predicate language can be reduced to the universal satisfiability — in a certain sense — of a probability assertion of the propositional language. Finally, the problem of universal satisfiability is investigated under a different aspect, within the framework of the proposition and the models of the original predicate language, when we raise the question how the original lass of probability models can be restricted from the point of view of universal satisfaction of probability assertions. For probability systems a comprehensive answer to this question is given by a necessary and sufficient condition of universal satisfaction. In fact, this answer includes both the problem of reducing the power of the universe and that of characterizing the Boolean algebras, which represent the set of truth values. We shall use the infinitary language (containing denumerable conjunctions and disjunctions) with a reference to the results of (4), involving the infinitary language.

§ 1.

Let us consider a first order calculus $\mathfrak{S}$, which contains denumerable individual variables: $x_1, x_2, \ldots, x_n, \ldots$; the logical constans $\wedge$, $\vee$, $\neg$, $\forall$, $\exists$ and $=$, finitely many, (namely l) relation symbols $P_1, P_2, \ldots, P_l$ with respective number of arguments $i(1), i(2), \ldots, i(l)$, $(i(j) \geqq 0, j=1, 2, \ldots, l)$.

Enlarging our language by the set C of individual constants, let us denote the resulting language by $\mathfrak{S}(C)$. Let $\mathscr{S}$ and $\mathscr{S}(C)$ denote the set of formulas without free variable belonging to $\mathfrak{S}$ and $\mathfrak{S}(C)$ respectively; and assuming a standard system of deductions denote by $\mathscr{S}/\equiv$ and $\mathscr{S}(C)/\equiv$ the Lindenbaum—Tarski algebras of these respective sets of sentences.

By a *probability system* belonging to the language $\mathfrak{S}$ we mean an $(l+4)$-tuple $V=\{B, E, P_1^0, \ldots, P_l^0, \mathscr{A}, m\}$ — where B is an arbitrary non-empty set, $\mathscr{A}$ a Boolean σ-algebra, m a positive probability measure on $\mathscr{A}$ (recall that a probability measure m is positive when $m(a)=0$ iff $a=0$), E is a function of two variables and the P_j^0-s are functions mapping the $i(j)$'th Cartesian power of B into $\mathscr{A}$. The function E takes the unit element I of the Boolean algebra if $b_1=b_2$ $(b_1, b_2 \in B)$ and the element 0, otherwise. More precisely we have thus defined — just for the sake of simplicity — a probability system with the so called strict identity. It is perhaps worthwile to call the reader's attention to the relationship of the present definition to Kripke models.

If the language contains only propositional variables then it suffices to require in the definition, as for as $\mathscr{A}$ is concerned that it be simply a Boolean algebra, and m be finitely additive.

The valuation of the formulas of $\mathscr{S}(B)$ may be defined by a usual recursive method:

$$f(b_1 = b_2) = E(b_1, b_2), \quad f(P_j(b_1, b_2, \ldots, b_{i(j)})) = P_j^0(b_1, b_2, \ldots, b_{i(j)})$$
$$(b_{i(j)} \in B \quad j = 1, 2, \ldots, l),$$
$$f(\neg\varphi) = I - f(\varphi),$$
$$f\left(\bigwedge_{m=1}^{M} \varphi_m\right) = \bigwedge_{m=1}^{M} f(\varphi_m), \quad f\left(\bigvee_{m=1}^{M} \varphi_m\right) = \bigvee_{m=1}^{M} f(\varphi_m),$$
$$f(\forall x \psi(x)) = \bigwedge_{b \in B} f(\psi(b)), \quad f(\exists x \psi(x)) = \bigvee_{b \in B} f(\psi(b)).$$

(The last two definitions are meaningful because the strict positiveness of m implies the completeness of $\mathscr{A}$ ((3) 15. §)). The valuation function f thus indirectly assigns to the formulas of the language a real value, too. This assignment will be denoted simply by $m(\varphi)$ instead of $m(f(\varphi))$ whenever this does not lead to confusion, and by $m_f(\varphi)$ otherwise.

By a *probability assertion* of a language $\mathfrak{S}$ we mean an $(N+1)$-tuple $\langle \Phi, \psi_1, \psi_2, \ldots, \psi_N \rangle$ where $\psi_1, \psi_2, \ldots, \psi_N \in \mathscr{S}$ and Φ is an algebraic formula of N free variables (i.e. a quantifier free formula of the algebraic language: containing denumerable individual constants 0, $+1$, -1 and the binary function symbols $+$ and $\cdot$).

We say that a probability system V *is a model of* $\vartheta = \langle \Phi, \psi_1, \psi_2, \ldots, \psi_N \rangle$ (or that ϑ holds on V) (see (4). 6 §), if the formulas $\psi_1, \psi_2, \ldots, \psi_N$ can be valuated on V and the N-tuple $(m(\psi_1), m(\psi_2), \ldots, m(\psi_N))$ satisfies Φ on the model $\langle \mathrm{Re}, \leqq, +, \cdot, 0, +1, -1 \rangle$ (where Re denotes the set of real numbers). ϑ *is valid if it holds on every* V. We know that every algebraic formula is equivalent to disjunctions of conjunctions of certain polynomial inequalities $P \geqq 0$ or $P > 0$ (disjunctive normal form). An algebraic formula is closed, if it is equivalent to the disjunction of conjunctions of formulas of the type $P \geqq 0$. Closed algebraic formulas and especially the formulas of the type $P \geqq 0$ proved to be particularly important in the study of probability logic.

To the origin of the notions and definitions used in this paragraph we refer to the above mentioned works of Gaifman, Scott and Krauss.

§ 2.

1.

In this paragraph we turn to discussion of our problem group in case of probability logic based on classical propositional language.

Thus our language contains only the logical constants $\wedge$, $\vee$, $\neg$ and the propositional symbols $X_1, X_2, \ldots, X_n$ $(1 = n)$, the latter also being called propositional variables.

In the case of propositional language it is sufficient to require that the Boolean algebra of the probability system be simply a Boolean algebra (not necessarily a Boolean σ-algebra) and that m be simply a finitely additive probability measure on this Boolean algebra. Moreover, since in this case the universe of the probability system plays no role, and the language does not contain the symbol $=$, the probability systems will be denoted simply as $(n+2)$-tuples $\langle X_1^0, X_2^0, \ldots, X_n^0, \mathscr{A}, m\rangle$. Clearly, in the present case the valuation is defined only for the formulas of the propositional language. *A probability system is said to have k atoms* if its Boolean algebra has k atoms.

In this part and more importantly in the next one we shall make use of several wellknown notions concerning polynomials of several variables. It is well-known that any polynomial of the variables $\lambda_1, \lambda_2, \ldots, \lambda_N$ is a sum of finitely many terms of the form $c_{k_1 k_2 \ldots k_N} \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_N^{k_N}$, where $k_1, k_2, \ldots, k_N$ are non-negative integers, and the coefficients $c_{k_1 k_2 \ldots k_N}$ are arbitrary non-negative reals. A polynomial is in the reduced form if to every N-tuple of real numbers $k_1, k_2, \ldots, k_N$ there is at most one term of the polynomial in which every λ_i is at the k_i' th power. The degree of the term $c_{k_1 k_2 \ldots k_N} \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_N^{k_N}$ is the number $k_1+k_2+\ldots+k_N$. The degree of the polynomial itself is the maximum of the degrees of its term. A polynomial is homogeneous if all its terms have the same degree.

Let $\Phi(\lambda_1, \lambda_2, \ldots, \lambda_N)$ be an arbitrary algebraic formula, let $\varphi_1(X_1, X_2, \ldots, X_n)$, $\varphi_2(X_1, X_2, \ldots, X_n), \ldots, \varphi_N(X_1, X_2, \ldots, X_n)$ be formulas of the propositional calculus, where $X_1, X_2, \ldots, X_n$ are propositional variables, whereas $\lambda_1, \lambda_2, \ldots, \lambda_N$ are variables of the algebraic language mentioned in § 1.

THEOREM 2.1. *The probability assertion $\vartheta = \langle \Phi, \varphi_1, \varphi_2, \ldots, \varphi_N\rangle$ holds on every probability system iff it holds on every system with 2^n atoms.*

PROOF. The necessity part is obvious.

We prove the sufficiency. It is known that a probability assertion $\langle \Psi, \psi_1, \psi_2, \ldots, \psi_N\rangle$ for which $\vdash \neg(\psi_i \wedge \psi_j)$ $(i \neq j)$,

$$\vdash \bigvee_{i=1}^{N} \psi_i \quad \text{and} \quad \{i: \vdash \neg\psi_i\} = \{L+1, L+2, \ldots, N\} \qquad \oplus$$

is true on every probability system, if for arbitrary non-negative real numbers p_i, $i=1, 2, \ldots, L$ such that $p_1+p_2+\ldots+p_L=1$ the formula $\Psi(p_1, p_2, \ldots, p_L, 0, \ldots, 0)$ in true on the model $\langle \mathrm{Re}, \leqq, +, \cdot, 0, +1, -1\rangle$ (cf. (4), (6)).

Suppose that ϑ holds on every system having 2^n atoms. Consider the 2^n elementary conjunctions of the variables $X_1, X_2, \ldots, X_n$. Denotes them by $\Gamma_p(X_1, X_2, \ldots, X_n)$ $(p=1, 2, \ldots, 2^n)$. We have $\vdash \neg(\Gamma_i \wedge \Gamma_j)$ $(i \neq j)$ and $\vdash \bigvee_{i=1}^{2^n} \Gamma_i$.

One can uniquely represent the formulas $\varphi_k(X_1, X_2, \ldots, X_n)$ $(k=1, 2, \ldots, N)$ as sums of certain Γ_p'-s, as it results from the existence of the disjunctive normal form.

Consider now the normal forms of Φ in the algebraic language. Suppose that all the polynomials it contains have N variables.

Choosing any of them, P say, by additivity of the measure m we see that there is a polynomial P_{2^n}' of 2^n variables such that in every probability system

$V=\langle X_1^0, X_2^0, \ldots, X_1^0, \mathscr{A}, m\rangle$ we have

$$P\big(m(\varphi_1(X_1^0, \ldots, X_n^0)), \ldots, m(\varphi_N(X_1^0, \ldots, X_n^0)) =$$

$$= P'_{2^n}(m(\Gamma_1(X_1^0, \ldots, X_n^0)), \ldots, m(\Gamma_{2^n}(X_1^0, \ldots, X_n^0))).$$

Denote by Φ' the formula of 2^n variables obtained upon replacing all the polynomials P in the normal form of Φ by P'_{2^n}.

One sees that ϑ and $\vartheta'=\langle\Phi', \Gamma_1(X_1, \ldots, X_n), \ldots, \Gamma_{2^n}(X_1, \ldots, X_n)\rangle$ simultaneously hold or do not hold on every individual probability system.

To the probability assertion ϑ' one can already apply theorem $\oplus$.

Let the numbers $p_1, p_2, \ldots, p_{2^n} \geqq 0$ be arbitrary with $\sum_{i=1}^{2^n} p_i = 1$. We shall show that there exist elements $A_1, A_2, \ldots, A_n$ in the Boolean algebra $\mathscr{B}(2^n)$ of 2^n atoms and a probability measure μ on $\mathscr{B}(2^n)$ so that

$$\mu(\Gamma_i(A_1, A_2, \ldots, A_n)) = p_i, \quad i = 1, 2, \ldots, 2^n.$$

By a suitable choice of the elements A_m we first assure that $\Gamma_i(A_1, \ldots, A_n)$ be the i'th atom a_i of the algebra ($i=1, 2, \ldots, 2^n$). Then the choice of a μ with the desired properties is straightforward.

We choose the elements A_m as follows: representing the element A_m as a sum of atoms, let a_i be included into this sum iff X_m is not negated in Γ_i. Hence a_i is included into the representation of Γ_i, and therefore no Γ_i can be the element 0, even though their pairwise products are all 0. However, there is only one system of elements of this kind over an algebra of 2^n atoms, namely, the system of the atoms. (This means that we have specified the elements A_m ($m=1, 2, 3, \ldots, n$).)

By assumption $\langle A_1, A_2, \ldots, A_n, B(2^n), \mu\rangle$ satisfies ϑ. Thus $\langle a_1, a_2, \ldots, a_{2^n}, B(2^n), \mu\rangle$ satisfies ϑ', and hence ϑ' meets the conditions of theorem $\oplus$. Accordingly, ϑ' holds on every probability system, implying that ϑ, too, holds on every probability system. Thus our assertion is proved. □

REMARKS. a) It is easy to see how this proof may be carried out in an equally simple way without using theorem $\oplus$.

b) Obviously, to a given probability assertion ϑ more than one probability assertion ϑ' can be found so that ϑ' satisfies the conditions of theorem $\oplus$ and is equivalent to ϑ in the said sense. (This is true even in the case of probability logic based on the language of predicate calculus.) However, in the case of transformations different from the elementary conjunctions used in the above train of thoughts, we can no more be sure that in a probability system having as many atoms as the number of arguments of the new probability assertion, the probabilities of the formulas involved in the assertion really run over all the possible values specified by theorem $\oplus$. We shall immediately return to this problem.

In the knowledge of the preceding theorem we may raise the question, whether the problem of satisfaction of probability assertions on systems having 2^n atoms can be reduced to examining systems with less atoms. The following answers this question in a certain respect:

THEOREM 2.2. *The substitution values of a polynomial P involved in a probability assertion over arbitrary probability systems with 2^n atoms are determined already by the values of the polynomial taken over systems with k atoms, where $k=\operatorname{grad} P$ and $k<2^n$.*

(The theorem will follow from the proof of the theorem treated in the next section.)

It is easy to set up a formula relating the values taken over systems with k resp. 2^n atoms, however the investigation of such a formula doesn't result in any simplification compared to the original problem involving systems with 2^n atoms.

On the other hand, recalling that by Theorem 2.1, if a probability assertion holds on every system having 2^n atoms, then it holds on every system having an algebra of a greater (or, of course, smaller) cardinality, one may ask whether in case of $k=\operatorname{grad} P<2^n$ a similar statement follows from the probability assertion being satisfied on every probability system with k atoms.

We give a negative answer by presenting an example of a probability assertion which holds on every system of 2^n-1 atoms (and thus on every system with less atoms, too) while it does not hold on some systems with 2^n atoms.

Let the elementary conjunctions of the variables $X_1, X_2, \ldots, X_n$ $(n\geqq 2)$ be: $\Gamma_1(X_1, \ldots, X_n)$, $\Gamma_2(X_1, \ldots, X_n)$, $\ldots$, $\Gamma_\alpha(X_1, \ldots, X_n)$. Consider the probability assertion:

$$\langle(\alpha-1)^{k-1}(\lambda_1^k+\ldots+\lambda_\alpha^k)-1\geqq 0,\quad \Gamma_1(X_1, \ldots, X_n), \ldots, \Gamma_\alpha(X_1, \ldots, X_n)\rangle,$$

where $\alpha=2^n$, and $\alpha>k$. As it results from the first theorem of this part, there exists a probability system with α atoms and substitutions $X_1, X_2, \ldots, X_n$ on this system so that

$$p(\Gamma_1)=p(\Gamma_2)=\ldots=p(\Gamma_\alpha)=\frac{1}{\alpha}.$$

Then the value of the polynomial involved in the above expression is:

$$\left(\frac{\alpha-1}{\alpha}\right)^{k-1}-1,$$

a negative number.

However even for arbitrary probability systems of $\alpha-1$ atoms we have $\sum_{i=1}^{2^n} p(\Gamma_i)=1$ for every substitution $X_1, X_2, \ldots, X_n$ and probability measure p. Further, there exist at most $\alpha-1$ pairwise disjoint elements different from zero and hence for some j, $p(\Gamma_j)=0$. If e.g. $j=\alpha$ then the value of the expression will be

$$(\alpha-1)^k\left[\frac{1}{\alpha-1}\left(p^k(\Gamma_1)+\ldots+p^k(\Gamma_{\alpha-1})\right)-\left(\frac{1}{\alpha-1}\left(p(\Gamma_1)+\ldots+p(\Gamma_{\alpha-1})\right)\right)^k\right]$$

which, according to a well-known identity, is never negative.

2.

We shall use the notation of the preceding part. For the sake of simplicity we shall require that the algebraic formula involved in the probability assertions has the form $P \geqq 0$, say, with P being a polynomial of degree k. In a certain sense, this will not restrict generality.

Next we give a sufficient condition for the satisfaction of a probability assertion on every probability system, for the case of $k = \text{grad } P < 2^n$ where P is the polynomial involved in the assertion; and "n" is the number of propositional variables. (This is an important case also, because for probability assertions arising in the practice 2^n is a large number with respect to "k"; — and making investigation on Boolean algebras of greater cardinality requires more and more computational work.) One of the advantages of the present method is the fact, that it is not necessary to transform the given probability assertion into an equivalent assertion (unlike e.g. applying theorem $\oplus$). Such a transformation requires a lot of calculation, since it generally increases exponentially the number of variables in the polynomial. Moreover, our method is of model theoretic character.

DEFINITION. By a "*saturated segment*" associated with the variables $y_1, y_2, \ldots, y_m$ of a reduced polynomial of several variables, we mean the sum of those terms of the polynomial in which these and only these variables appear with a positive exponent.

DEFINITION. By a "*veritable realization*" of the propositional variables $X_1, X_2, \ldots, X_n$ on an algebra of k atoms $a_1, a_2, \ldots, a_k$ we mean an assignment of the variables to the elements of the algebra with the property that expressing the elements substituted into $X_1, X_2, \ldots, X_m$ as a sum of atoms, for every pair a_i, a_j with $i \neq j$ there exist an "m" such that one and only one of a_i, a_j appears in X_m.

Let $\vartheta = \langle \Phi(\lambda_1, \ldots, \lambda_N), \varphi_1(X_1, \ldots, X_n), \ldots, \varphi_N(X_1, \ldots, X_n) \rangle$ be a probability assertion where Φ has the form $P \geqq 0$ and P is a polynomial of degree "k" with N variables ($k < 2^n$).

Before stating the next theorem let us remark that checking the assertion on the examples following the proof before reading the latter might help the reader in a better understanding of the theorem and the involved genuinely practical procedure.

THEOREM 2.3. *Let the elements $a_1, a_2, \ldots, a_k$ be the atoms of a fixed algebra, and consider an arbitrary but fixed "veritable realization" of the variables $X_1, X_2, \ldots, X_n$ on this algebra. Let "p" be the parameter probability defined on the algebra. Convert the polynomial obtained from P by substitution on the algebra into a reduced homogeneous polynomial R_k of degree "k" of the variables $p(a_1), p(a_2), \ldots, p(a_k)$. Clearly, this is possible. Let us consider a saturated segment of this polynomial, a polynomial of "l" variables, say ($l = 1, 2, \ldots, k$). Let us regard this as a polynomial substituted on the probability system with "l" atoms, rather than on a probability system with "k" atoms. If the latter polynomial considered as a probability assertion after having substituted the probabilities, holds (— is non-negative) for arbitrary systems with "l" atoms; moreover, the same is true for every saturated segment of R_k and this for every "veritable realization" of the variables $X_1, X_2, \ldots, X_n$ then the probability assertion Φ holds on every probability system.*

REMARKS. a) It is evident that if the conditions hold for every substitution (i.e. not only for the veritable ones) on a system with k atoms, then the assertion is a fortiori true, though we have made unnecessary substitutions, too.

b) As it will become clear from the proof of the theorem, substitutions arising from earlier ones by interchanging the roles of two atoms (a_i and a_j, say) in every substitution, (regarding the latter as sums of atoms) are superfluous, as well. Thus, e.g. for every substitution in which a_i is involved and a_j is not, the other involves a_j and does not involve a_i; or if both a_i and a_j are involved then the other substitution also involves both, and so on.

However omitting such substitutions practically amounts to taking into consideration the symmetry conditions — which for a given probability assertion would emerge anyhow.

PROOF. Notice first that the substitution values of the polynomial P on a system of k atoms ($k=\operatorname{grad} P$) do uniquely determine a homogeneous polynomial of the variables $p_1=p(a_1), p_2=p(a_2), \ldots, p_k=p(a_k)$, i.e. the polynomial is uniquely determined not only for the values $p_1+\ldots+p_k=1$, but for arbitrary $p_1 \geqq 0, \ldots, p_k \geqq 0$. This can be seen e.g. by writing P into a homogeneous form P^1 depending on $p_1, p_2, \ldots, p_k$, using the property

$$P^1(tp_1, tp_2, \ldots, tp_k) = t^k P^1(p_1, p_2, \ldots, p_k)$$

of homogeneous expressions. This means that P has a unique transformation into a homogeneous form, and thus P^1 is uniquely determined.

Let us start now from an arbitrary but fixed probability system

$$V = \langle X_1^0, X_2^0, \ldots, X_n^0, \mathscr{A}, m \rangle.$$

We shall prove that under the conditions of the theorem the probability assertion ϑ holds on this system.

Applying the reasoning already used, with the same notation to the polynomial P appearing in ϑ we have

$$P\big(m(\varphi_1(X_1^0, \ldots, X_n^0)), \ldots, m(\varphi_N(X_1^0, \ldots, X_n^0))\big) =$$
$$= P'_{2^n}\big(m(\Gamma_1(X_1^0, \ldots, X_n^0)), \ldots, m(\Gamma_{2^n}(X_1^0, \ldots, X_n^0)\big)$$

where P'_{2^n} is a polynomial of 2^n variables.

Let us pick arbitrary elementary conjunction formulas of our language and let us consider the sum of those terms of

$$P^*_{2^n}\big(m(\Gamma_1(X_1^0, \ldots, X_n^0)), \ldots, m(\Gamma_{2^n}(X_1^0, \ldots, X_n^0))\big)$$

in which at least one of the variables

$$m\big(\Gamma_{i_1}(X_1^0, \ldots, X_n^0)\big), \ldots, m\big(\Gamma_{i_k}(X_1^0, \ldots, X_n^0)\big)$$

appears and no other variable occurs. (These are variables inasmuch the underlying probability system is arbitrary.) Let $R_k\big(m(\Gamma_{i_1}), \ldots, m(\Gamma_{i_k})\big)$ be the resulting segment, which itself is a homogeneous polynomial.

On the other hand we claim that picking arbitrary conjunction formulas $\Gamma_{i_1}, \Gamma_{i_2}, \ldots, \Gamma_{i_k}$ of the language, by a suitable choice of $X_1, X_2, \ldots, X_n$ we can achieve

on the algebra of k atoms that $\Gamma_{i_1}, \Gamma_{i_2}, \ldots, \Gamma_{i_k}$ respectively be equal to $a_1, a_2, \ldots, a_k$ while the other conjunction terms become 0.

The corresponding substitution is the following: If $1 \leqq m \leqq n$ we substitute the elements into X_m in their form of sums of atoms, then a_j $(1 \leqq j \leqq k)$ should appear in the element substituted into X_m iff X_m appears in Γ_{i_j} in an unnegated form. This way we achieve that the atomic representation of Γ_{i_j} contains a_j and thus it is not 0, whereas the Γ_{i_j}'s are pairwise disjoint. However there is only one system of elements with this property over an algebra having k atoms; namely the system of the atoms. Hence the remaining elements Γ_i; are necessarily 0. This realizes the desired transformation.

Observe also that having selected k different Γ_i's the given realization is easily proved to be veritable. Conversely, every veritable realization is equivalent to a choice of k Γ_i's and to their transformation into atoms — as it appears from reversing the previous deduction. (The nonveritable realizations amount to selecting less than k Γ_i's depending on the various X_m's. In this case the Γ_i's do not necessarily transform only into atoms, rather, they transform into a system of pairwise disjoint elements of the k-atom Boolean algebra which has as many elements as there are atoms appearing in a different way in the substitution. The above implies that the substitutions — mentioned in Remark b) are different transformations connecting the same systems, i.e. the system Γ_i of k elements and the probability system having k atoms, so that their separate discussion is not needed.)

Continuing with the main idea of the proof, let us repeat the transformation applied to the polynomial P, instead of the system V, for the system of k atoms with the parameter probability measure p, assigning to the X_m's $(m=1, 2, \ldots, n)$ the values corresponding to the already defined veritable realization (thus exploiting that the conditions of the theorem are valid for an arbitrary realization on a system with k atoms). Then the polynomial $P^*_{2^n}$ is transformed into the polynomial $R_k(p(a_1), \ldots, p(a_k))$ where p is the same as before.

(At this point we observe that if we substitute into X_m the values of the algebra of k atoms resulting from the above defined veritable realization in the original form

$$P(m(\varphi_1(X_1, \ldots, X_n)), \ldots, m(\varphi_N(X_1, \ldots, X_n)))$$

rather than in the new form $P^*_{2^n}$, we still arrive at the function, $R_k(p(a_1), \ldots, p(a_k))$, since

$$P(m(\varphi_1(X_1, \ldots, X_n)), \ldots, m(\varphi_N(X_1, \ldots, X_n))) =$$

$$= P^*_{2^n}(m(\Gamma_1(X_1, \ldots, X_n)), \ldots, m(\Gamma_{2^n}(X_1, \ldots, X_n)))$$

holds for every probability system. Here we also make use of the said uniqueness of the transformation of a polynomial into a homogeneous one.)

Continuing with the original idea of the proof, we see that regarding the assertions $(S_l \geqq 0)$ defined by the saturated segments of $R_k(p(a_1), \ldots, p(a_k))$, (e.g. $S_l(p(a_{j_1}), \ldots, p(a_{j_l}))$ which depends on $p(a_{j_1}), \ldots, p(a_{j_l})$) in their form after substitution on an arbitrary probability system with "l" atoms, (i.e. $p(a_{j_1})+p(a_{j_2})+\ldots$ $\ldots+p(a_{j_l})=1$) the conditions of the theorem guarantee that the assertions hold on the probability system with "l" atoms. Using homogeneity, we also have:

$$S_l(tp(a_{j_1}), \ldots, tp(a_{j_l})) = t^k S_l(p(a_{j_1}), \ldots, p(a_{j_l}))$$

thus the non-negativity of $S_l(p(a_{j_1}), \ldots, p(a_{j_l}))$ follows for arbitrary $p(a_{j_1}) \geqq 0, \ldots$ $\ldots, p(a_{j_l}) \geqq 0$ (not only for the case $p(a_{j_1}) + \ldots + p(a_{j_l}) = 1$).

Returning to the original system V and going backward we conclude that substituting on the original arbitrary, but fixed probability system V the segment

$$S_l(m(\Gamma_{j_1}(X_1^0, \ldots, X_n^0)), \ldots, m(\Gamma_{j_l}(X_1^0, \ldots, X_n^0)))$$

will again be non-negative for the choice

$$p(a_{j_1}) = m(\Gamma_{j_1}(X_1^0, \ldots, X_n^0)), \ldots, p(a_{j_l}) = m(\Gamma_{j_l}(X_1^0, \ldots, X_n^0)).$$

Still regarding the polynomial P in its substituted form over the system V the polynomial $P_{2^n}^*$ can be decomposed into disjoint saturated segment, each of which, regarded as being substituted over V can be one obtained in the above way, whence its non-negativity follows by our deduction, i.e. starting from some formulas $\Gamma_{i_1}, \ldots$ $\ldots, \Gamma_{i_k}$, carrying out the proper substitution on the algebra of k atoms, we use the conditions of the theorem for some appropriate segment R_k of the resulting polynomial, with the parameter probability measure.

Thus P'_{2^n} is a sum of non-negative terms. □

As a first illustration of how our theorem can be applied let us examine the satisfaction of the probability assertion:

$$\langle 2\lambda_1 \cdot \lambda_2 + (\lambda_3 + \lambda_4 + \lambda_5)^2 - \lambda_1^2 - \lambda_2^2 - (\lambda_6 + \lambda_7 + \lambda_8)(\lambda_9 + \lambda_{10} + \lambda_{11}) + (\lambda_{12} + \lambda_{13} + \lambda_{14})^2 \geqq 0,$$

$$X \wedge Y \wedge Z,\ X \vee Y \vee Z,\ X, Y, Z,\ X \wedge Y,\ X \wedge Z,\ Y \wedge Z,\ X \vee Y,\ X \vee Z,\ Y \vee Z,\ \neg X \wedge \neg Y,$$

$$\neg X \wedge \neg Z,\ \neg Y \wedge \neg Z \rangle.$$

As the polynomial is of degree 2, the investigation can be reduced to the algebra of 2 atoms with corresponding probabilities p_1 and p_2 ($p_1 + p_2 = 1$).

On this algebra, taking symmetry into account, one has the following substitutions for the variables X, Y, Z:

$$0, 0, 0 \quad 0, 0, I \quad 0, 0, a \quad 0, I, I \quad 0, a, a \quad 0, I, a \quad 0, a, b$$

$$I, I, I \quad I, I, a \quad I, a, a \quad I, a, b$$

$$a, a, a \quad a, a, b$$

Among these I, I, I; $0, 0, 0$; $I, 0, 0$ and $I, I, 0$ are not veritable realizations (i.e. in every term of these substituions either both or none of a and b are present). Let us examine e.g. the substitutions $X = 0$, $Y = 0$, $Z = a$. Substituting into the variables of the polynomial the probabilities previously obtained we get the polynomial $p_1^2 - p_1^2 + (1 + 2p_2)^2$. The reduced homogeneous form of this is: $p_1^2 + 9p_2^2 + 6p_1p_2$ (this can be obtained using the relation $p_1 + p_2 = 1$). The saturated segments: p_1^2, $9p_2^2$, $6p_1 \cdot p_2$ are necessarily non-negative.

Similarly to the other cases, the homogeneous forms are the following:

$$0, a, b : p_1^2 + p_2^2 + p_1p_2; \quad 0, a, I : p_2^2; \qquad I, a, a : p_2^2$$

$$0, a, a : 9p_2^2; \qquad I, I, a \quad \text{and} \quad I, a, b : 0$$

$$a, a, a : 9p_2^2; \qquad a, a, b : p_2^2$$

Since in all cases the coefficients of the terms are always non-negative, according to our theorem the probability assertion holds on every probability system.

As a next example let us examine the inequality:

$$V_{l-1}^{(n)}(V_1^{(l)}-l+1)+\frac{1}{l-1}\binom{n-1}{l-1}\cdot 2\left[V_2^{(n)}-\binom{V_1^{(n)}}{2}\right]\leqq$$

$$\leqq l\cdot V_l^{(n)}\leqq V_{l-1}^{(n)}(V_1^{(n)}-l+1)+\binom{n-1}{l-1}\cdot 2\cdot\left[V_2^{(n)}-\binom{V_1^{(n)}}{2}\right] \quad (l=2,3,\ldots,n).$$

Here

$$V_l^{(n)}=\sum_{1\leqq i_1\leqq\ldots\leqq i_l\leqq n} m(X_{i_1}^0\wedge X_{i_2}^0\wedge\ldots X_{i_l}^0)$$

and $X_1^0,\ldots,X_n^0$ are arbitrary realizations of the corresponding propositional variables; m is an arbitrary probability measure and the summation is extended to every combination without repetition of the numbers $1,2,\ldots,n$ $(i_1,i_2,\ldots,i_n)$.

Our inequality is nothing but a probability assertion after substitution of appropriate probabilities.

Introducing the notation:

$$W_j^{(n)}=jV_j^{(n)}-V_{j-1}^{(n)}(V_1^{(n)}-j+1) \quad (j=2,3,\ldots,n)$$

the inequality can be rewritten into the form:

$$\frac{1}{l-1}\binom{n-1}{l-1}W_2^{(n)}\leqq W_l^{(n)}\leqq\binom{n-1}{l-1}W_2^{(n)}$$

(It is known that $W_j^{(n)}$ is non-negative.)

However,

$$(0)\qquad \frac{n-(j-1)}{j}W_j^{(n)}\leqq W_{j+1}^{(n)}\leqq\frac{n-(j-1)}{j-1}W_j^{(n)} \quad (j=2,3,\ldots,n).$$

Since here the polynomial of the probabilities is of second degree, it suffices to limit ourselves to algebras with two atoms (let them be again a and b put $p_1=p(a)$, $p_2=p(b)$).

If we consider an arbitrary but fixed, not necessarily veritable realization of the variables $X_1,X_2,\ldots,X_n$ on this algebra, where n_1, n_2, n_3 and n_4 of the X's is equal to I, a, b and 0 respectively $(n_1+n_2+n_3+n_4=n)$, then:

$$W_j^{(n)}=j\left(p_1\binom{n_1+n_2}{j}+p_2\binom{n_1+n_3}{j}\right)-$$

$$-\left(p_1\binom{n_1+n_2}{j-1}+p_2\binom{n_1+n_3}{j-1}\right)(p_1(n_1+n_2)+p_2(n_1+n_3)-j+1).$$

However, writing (0) into a reduced homogeneous form using $p_1+p_2=1$ the coefficient of p_1^2 and p_2^2 will be 0; — while between the coefficients of $p_1\cdot p_2$ the desired inequality holds. Thus (0) holds for every probability system, indeed.

Finally, we have to notice the fact, that the original inequality can be obtained as the product of the inequalities of (0) for $j=2, 3, \ldots, l-1$ — taking into consideration that the involved quantities are non-negative.

Using the theorem 2.3 the best-know probabilistic inequalities (more precisely, those which also define a probability assertion) can be proved. (A collection of such inequalities can be found e.g. in [6].) It is easily seen that all the inequalities which can be verified by the method of [7] and [8], can be proved by applying our theorem as well, since they are by far special cases of our theorem.

§ 3.

According to our program in this paragraph we shall examine the case of probability logic based on the language of first-order predicate calculus.

Next we show that in case of the predicate language the examination of universal satisfaction of a probability assertion is equivalent to the examination of universal satisfaction in a certain sense — of a probability assertion of the propositional language.

Let $\psi_1, \psi_2, \ldots, \psi_N$ be arbitrary closed formulas of the predicate language $\mathfrak{S}$, and let Φ be again a quantifier-free algebraic formula with N free variables.

THEOREM 3.1. *Let $\vartheta=\langle\Phi, \psi_1, \psi_2, \ldots, \psi_N\rangle$ be an arbitrary probability assertion and let $\Theta=\langle\Phi, X_1, X_2, \ldots, X_N\rangle$ be the probability assertion assigned to ϑ by letting $X_1, X_2, \ldots, X_N$ be propositional variables. Then there exist a finitary, atomic probability system and a suitable fixed valuation of the variables $X_1, X_2, \ldots, X_N$ on this system with the property that ϑ is satisfied on every probability system iff Θ is satisfied under the said valuation for arbitrary probabilities on the above finitary atomic probability system.*

PROOF. Consider the elementary conjunctions $\Gamma_1(X_1, \ldots, X_N), \Gamma_2(X_1, \ldots, X_N), \ldots$ $\ldots, \Gamma_{2^N}(X_1, \ldots, X_N)$ of the variables $X_1, X_2, \ldots, X_N$ and the formulas $\Gamma_1(\psi_1, \ldots, \psi_N), \ldots$ $\ldots, \Gamma_{2^N}(\psi_1, \ldots, \psi_N)$. Then analogously to the proof of the theorem 2.1 one can construct an algebraic formula Φ' of 2^N variables, such that

$$\vartheta' = \langle\Phi', \Gamma_1(\psi_1, \ldots, \psi_N), \ldots, \Gamma_{2^N}(\psi_1, \ldots, \psi_N)\rangle \quad \text{and} \quad \vartheta$$

do or do not hold simultaneously on every probability model. Now, to the second one of these probability assertions one can already apply theorem $\oplus$. Suppose furthermore, that among the formulas $\Gamma_1, \Gamma_2, \ldots, \Gamma^{2^n}$ those which are identically false or precisely the $(L+1)$'th, $(L+2)$'th... and the (2^N)'th.

On the other hand, using the idea of the proof of the theorem 2.3 we can assure that for a suitable valuation of the variables $X_1, X_2, \ldots, X_N$ over the algebra of L atoms $\Gamma_1, \Gamma_2, \ldots, \Gamma_L$ transform into the elements $a_1, a_2, \ldots, a_L$ of the algebra of L atoms whereas $\Gamma_{L+1}, \Gamma_{L+2}, \ldots, \Gamma_{2^n}$ transform into zero. Hence, with a suitable definition of the probabilities one can achieve that, with the above valuation, one assigns arbitrary probabilities $p_1, p_2, \ldots, p_L \geqq 0$ to $\Gamma_1, \Gamma_2, \ldots, \Gamma_L$, respectively, (satisfying the sole condition $p_1+p_2+\ldots+p_L=1$), while assigning probability zero to the remaining Γ_j's.

Consequently, if $\langle \Phi', \Gamma_1(X_1, \ldots, X_N), \ldots, \Gamma_{2^n}(X_1, \ldots, X_N)\rangle$ holds under the said fixed valuation of the variables $X_1, X_2, \ldots, X_N$ over the algebra of L atoms for arbitrary probabilities, then, in virtue of theorem $\oplus$, ϑ' holds on every probability system.

However, the probability assertions $\langle \Phi', \Gamma_1(X_1, \ldots, X_N), \ldots, \Gamma_{2^n}(X_1, \ldots, X_N)\rangle$ $\langle \Phi, X_1, \ldots, X_N\rangle$ are equivalent, i.e. have the same probability models. Thus the given valuation of the variables $X_1, X_2, \ldots, X_N$ on the algebra of L atoms, and, after having chosen arbitrary probabilities, the valuation of the probability assertion can be carried out with Θ as well.

Hence if Θ holds for the said valuation of the variables $X_1, X_2, \ldots, X_N$ over the algebra of L atoms for every possible choice of the probabilities on this algebra, then Θ too, holds on every probability model. □

The rest of this section deals with a completely different approach to universal satisfaction of a probability assertion.

Let T be the set of natural numbers. Let $\mathscr{S}^\infty(T)/\equiv$ denote the Lindenbaum—Tarski σ-algebra of the language containing the denumerable conjunction $\wedge$ and disjunction $\vee$. Let α' be an arbitrary probability measure on $\mathscr{S}^\infty(T)/\equiv$ and let $\mathscr{S}_\alpha^\infty(T)/\equiv$ be the quotient algebra modulo the σ-ideal $\{\varphi/\equiv : \alpha'(\varphi)=0\}$ of this algebra (where $\varphi/\equiv$ is the class containing φ). Let α denote the measure on the quotient algebra, obtained from α'. Let us define on T the functions $E', P_1', P_2', \ldots, P_l'$ in such a manner that the value $P_j'(t_1, t_2, \ldots, t_{i(j)})$ be the image in $\mathscr{S}_\alpha^\infty(T)/\equiv$ of the elements $P_j(t_1, t_2, \ldots, t_{i(j)})$ under the canonic homomorphism.

Denote by H the consisting of the systems having both a finite universe and a finite Boolean algebra, as well as of the systems

$$\langle T, E', P_1', P_2', \ldots, P_l', \mathscr{S}_\alpha^\infty(T)/\equiv, \alpha\rangle.$$

The first observation is that the systems:

$$\langle T, E', P_1', P_2', \ldots, P_l', \mathscr{S}_\alpha^\infty(T)/\equiv, \alpha\rangle$$

can in fact be regarded as probability systems; to this end it suffices to verify that α is a strictly positive probability on $\mathscr{S}_\alpha^\infty(T)/\equiv$ (cf. (3) 15. §).

Suppose now that the fixed probability assertion ϑ holds on every system of H. Let V be an arbitrary but fixed probability system, upon which ϑ is defined. Now we show that our hypothesis guarantees that ϑ holds on V, too. This is done separating the cases of V's universe being finite, denumerable or having a larger cardinality.

If the universe is finite, then the Boolean subalgebra of the original algebra generated by the values assigned to the atomic formulas is also finite. If we modify the original probability system so that we limit ourselves to this finite subalgebra and to the measure defined on it, then by our assumption ϑ holds on this system, and hence it holds on the original system as well.

Let us suppose that B is denumerable. Then the reasoning of the authors of [4] can be essentially repeated for the case of finite language and the corresponding probability systems, as follows; Limiting ourselves for example to the set of natural numbers, let V be an arbitrary probability system:

$$\langle T, E, P_1^0, P_2^0, \ldots, P_l^0, \mathscr{A}, m\rangle.$$

The valuation function f generates a σ-homomorphism of $\mathscr{S}^{\infty}(T)/\equiv$ into $\mathscr{A}$ (cf. [4], lemma 4.1); thus at the same time, it generates also probability α' on $\mathscr{S}^{\infty}(T)/\equiv$. Because of the strict positivity of the original measure, the kernel of this homomorphism is the σ-ideal $\{\varphi/\equiv : \varphi \in \mathscr{S}^{\infty}(T)/\equiv,\ \alpha'(\varphi)=0\}$ (cf. [3] § 15). The quotient algebra of $\mathscr{S}^{\infty}(T)/\equiv$ modulo the kernel of the homomorphism (as a σ-ideal) is isomorphic to the range of the homomorphism, which is σ-algebra. Consider the system

$$\langle T, E', P_1', \ldots, P_l', \mathscr{S}_{\alpha}^{\infty}(T)/\equiv, \alpha\rangle$$

so defined. The measure associated with the quantifierfree formulas of $\mathscr{S}^{\infty}(T)$ has not changed with respect to the original system, since α' (resp. α) were obtained by means of the corresponding σ-homomorphism. Since our systems are probability systems, α and m satisfy on $\mathscr{S}^{\infty}(T)$ the so called Gaifman condition: $\alpha'(\exists x\varphi) = \sup_R \alpha'\left(\bigvee_{b\in R} \varphi(b)\right)$ where R is taken over the set of finitary subsets of B and hence the probabilities of the quantifierfree formulas determine the probabilities on $\mathscr{S}^{\infty}(T)$.

Let now the cardinality of the universe be greater than denumerable. We shall show that in examining the satisfaction of ϑ on such a system, it is sufficient to consider a system with countable universe. It is sufficient to show that if a probability assertion holds on a probability system with infinite universe, then it holds on a probability system with countable universe as well. So, let V be a probability system with a fixed infinite universe, upon which the probability assertion $\langle \Phi, \psi_1, \psi_2, \ldots, \psi_N\rangle$ holds; and let h be the σ-homomorphism belonging to V. Consider now the arbitrary but fixed prenex forms $\psi_1^*, \psi_2^*, \ldots, \psi_N^*$ belonging to $\psi_1, \psi_2, \ldots, \psi_N$. Because of $\vdash \psi_k \leftrightarrow \psi_k^*$ $(k=1, 2, \ldots, N)$ takes the same values on ψ_k^* as on the original formulas. Let the quantifier blocks of $\psi_1^*, \psi_2^*, \ldots, \psi_N^*$ have $t_1, t_2, \ldots, t_N$ elements, respectively. Because of the definition of valuation,

$$h(\psi_k^*) = \nabla_1^{(k)} x_1 \nabla_2^{(k)} x_2 \ldots \nabla_{t_k}^{(k)} x_{t_k} \varphi \quad (k = 1, 2, \ldots, N)$$

where the ∇-s can be the operations $\vee$ and $\wedge$, taken over the A universe of V; φ is the kernel of the prenex formula.

We claim that there exist a countable subset E_0 containing E_0 (i.e. for E_0, too), there exist countable sets $F_1^{(k)} \subseteq \ldots \subseteq F_{t_k}^{(k)}$ also containing E_0' with the property that performing the operations $\nabla_1^{(k)}, \nabla_2^{(k)}, \ldots, \nabla_{t_k}^{(k)}$ on the respective sets $F_1^{(k)}, F_2^{(k)}, \ldots, F_{t_k}^{(k)}$ rather than on A we obtain the original values corresponding to the Boolean algebra.

$\nabla_1^{(k)}$ is one of the operations $\vee$, $\wedge$ taken over some infinite subset of A (i.e. of $\mathscr{A}$). Since the measure defined on $\mathscr{A}$ is strictly positive, $\mathscr{A}$ satisfies the countable chain condition (cf. [3], § 15). Hence $\mathscr{A}$ (and thus, also A) has a countable subset $F^{(k,0)}$ upon which we can restrict ourselves to the valuation of the operation $\nabla_1^{(k)}$, (clearly every subset containing $F^{(k,0)}$ is suitable).

Let $E_0 = \bigcup_{k=1}^{n} F^{(k,0)}$. Consequently for every countable subset E_0' containing E_0, the operations $\nabla_1^{(k)}$ $(k=1, 2, \ldots, N)$ give the original value. Let $F_1^{(k)} = E_0'$ $(k=1, 2, \ldots, N)$. For every fixed element of $F_1^{(k)}$ (i.e. of E_0') substituted into x_1 in φ in case of the infinite operation $\nabla_2^{(n)}$ we can again limit ourselves to a countable subset of $\mathscr{A}$ or A, and consequently to every subset containing A. To every element of E_0' one can assign such a countable subset, and since $F_1^{(k)}$ is countable; the union of these sets is also countable. Adding E_0 to this union; the resulting set can be

chosen as $F_2^{(k)}$. Then fixing an arbitrary element of $F_1^{(k)}$ and $F_2^{(k)}$, respectively the operation $\nabla_3^{(k)}$ can be applied; and again there exists a countable subset of A upon which it is sufficient to take $\nabla_3^{(k)}$. Since $F_1^{(k)} \times F_2^{(k)}$ is countable; the union of countable sets assigned to pairs of elements in the above manner is countable as well.

Adding $F_2^{(k)}$ to this union we can define $F_3^{(k)}$. We can proceed until t_k. The series of sets $F_1^{(k)} \subseteq F_2^{(k)} \subseteq \ldots \subseteq F_{t_k}^{(k)}$ $(k=1, 2, \ldots, N)$ and E_0 obtained in this way obviously meet the requirements.

Let us define now the following series of sets. Let E_1 be the union of the above sets $F_{t_1}^{(1)}, F_{t_2}^{(2)}, \ldots, F_{t_N}^{(N)}$. Consequently, E_1 is countable. Since E_1 is countable and contains E_0, the above procedure can be repeated for the choice $F_1^{(k)} = E_1$ $(k=1, 2, \ldots, N)$. Let E_2 be the union of the newly obtained $F_{t_1}^{(1)}, \ldots, F_{t_N}^{(N)}$ sets, etc. We claim that $B = \bigcup_{i=0}^{\infty} E_i$ will be a countable set with the property, that choosing B as the universe of V and limiting ourselves only to B at the valuation h assigns to $\psi_1, \psi_2, \ldots, \psi_N$ the same value of $\mathscr{A}$ as in the case of the universe A. Because of $E_0 \subseteq B$ the operation $\nabla_1^{(k)}$ gives the same result on both the sets A and B. Let b be an arbitrary but fixed element of B. There exists a set E which contains b; but then $F_1^{(k)}$ corresponding to E_{i+1}, contains it, too. But $F_2^{(k)}$ is constructed in such a manner that for every element of $F_1^{(k)}$ (including b), $\nabla_k^{(2)}$ should give on $F_2^{(k)}$ (and on every countable subset containing $F_2^{(k)}$) the same value as for A. Since this is true for every element of B in applying $\nabla_2^{(k)}$ we can limit ourselves to the set B. Similarly, fixing two arbitrary elements of B, there exists a set E_j which contains these elements, and thus $F_1^{(k)}$ and $F_2^{(k)}$, constructed in the $(j+1)$-th step contain then as well. $F_3^{(k)}$ has the property that for every element of $F_1^{(k)} \times F_2^{(k)}$, $\nabla_3^{(k)}$ gives the same value on $F_3^{(k)}$, or on a set containing $F_3^{(k)}$, as on A. This reasoning is valid for arbitrary pair of elements in B; etc. Thus we obtained the desired assertion.

Summing up the above, we have proved the following theorem:

THEOREM 3.2. *A probability assertion ϑ holds on every probability system iff it holds on every system having a finite universe and finite Boolean algebra as well as on the systems*

$$\langle T, E', P_1', P_2', \ldots, P_l', \mathscr{S}_\alpha^\infty(T)/\equiv, \alpha \rangle.$$

REMARKS. a) The conclusion of the theory summarizing the train of thought above can be already proved by paragraph (4). 6, nevertheless, we have provided a direct proof.

b) The Gaifman condition — or its generalization — means that in every probability system one can give a direct relation expressing the measure of formulas containing quantifiers, by the measure of quantifier free formulas.

Applying this fact to the probability systems of the present theorem, a further necessary and sufficient condition can be given for the universal satisfaction of a probability assertion. This condition not only reduces the problem of satisfaction of a probability assertion to its satisfaction on certain systems, but, it every reduces the problem to the satisfaction of quantifier free formulas of these systems.

c) If we choose a language without identity, similarly to the case of ordinary first-order logic, it is true that if a probability assertion holds on every probability system with universe A then it holds on every probability system the universe of

which has a smaller power. Thus in case of probability systems without identity, the necessary and sufficient condition a probability assertion to hold on every probability system is that it should hold on every probability system with denumerable universe.

REFERENCES

[1] GAIFMAN, H.: Concerning measures on Boolean algebras, *Pacific J. of Mathematics* **14** (1964), 61—73.
[2] GAIFMAN, H.: Concerning measures in first order calculi, *Israel J. of Mathematics* **2** (1964), 1—18.
[3] HALMOS, P. R.: *Lectures on Boolean algebras,* Van Nostrand, Princeton, 1963.
[4] SCOTT, D. and KRAUSS, P.: *Assigning probabilities to logical formulas. Aspects of inductive logic,* North Holland, Amsterdam, 1966.
[5] CHANG, C. C., and KEISLER, H. J.: *Model Theory,* North Holland, Amsterdam, 1973.
[6] FRÉCHET, M.: *Les probabilités associées à un système d'événements compatible et dépendants.* I., II., Hermann, Paris, 1940, 1943.
[7] GALAMBOS, J.—RÉNYI, A. On quadratic inequalities in probability theory, *Studia Sci. Math. Hung.* **3** (1968), 351—358.
[8] PARZEN, E.: *Modern probability theory and its applications,* Wiley, New York, 1960.
[9] HENKIN, L. and TARSKI, A.: Cylindric algebras, *Lattice theory, Proceedings of symposia in pure mathematics,* Vol. II, pp. 83—113. American Mathematical Society, Providence, R. I., 1961.
[10] RASIOWA, H. and SIKORSKI, R.: Algebraic Treatment of the Notion of Satisfiability, *Fundamenta Mathematicae* **40** (1953), 62—95.

University of Technology, Department of Mathematics, Budapest XI, Stoczek u. 2. H. III, Hungary 1111

(Received October 4, 1975: revised March 1, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 117—124

ON RS_u-INTEGRAL

by

A. G. DAS and B. K. LAHIRI

1. Introduction and definitions

The following definitions are known [2].

Let a', a, b, b' be fixed real numbers such that $a'<a<b<b'$ and let Δ denote the subdivision of the form

$$x_{-1} < a = x_0 < x_1 < \ldots < x_k = b < x_{k+1},$$

where $a' \leqq x_{-1} < a$, $b < x_{k+1} \leqq b'$. The norm of Δ, $\|\Delta\|$, is the number $\max (x_i - x_{i-1})$. The functions f, g, φ, u are defined at least in $[a, b]$, u always being strictly increasing.

DEFINITION 1. If $x \in [a, b)$ and $\lim\limits_{h \to 0+} \dfrac{g(x+h)-g(x)}{u(x+h)-u(x)}$ exists, we denote it by $g_u^+(x)$. A corresponding definition holds for $g_u^-(x)$, where $x \in (a, b]$. When $g_u^-(x) = g_u^+(x)$, we say g is *u-differentiable* at x and denote the common value by $g_u'(x)$.

CONDITION A. Suppose that $g_u^-(b)$ and $g_u^+(a)$ exist. The functions g, u are defined in $[a', a]$ and $[b, b']$ such that u is strictly increasing on $[a', b']$ and

$$\frac{g(x)-g(y)}{u(x)-u(y)} = g_u^+(a) \quad \text{for all} \quad x, y \in [a', a]$$

and

$$\frac{g(x)-g(y)}{u(x)-u(y)} = g_u^-(b) \quad \text{for all} \quad x, y \in [b, b'].$$

DEFINITION 2. For $x, y \in [a', b']$, $x \neq y$

$$g_u(x, y) = \frac{g(x)-g(y)}{u(x)-u(y)}$$

is called the *u-incrementary ratio* of g.

DEFINITION 3. A function g is *u-convex* on $[a, b]$ if for $a \leqq \alpha \leqq \xi \leqq \beta \leqq b$

$$g(\xi) \leqq \frac{u(\xi)-u(\alpha)}{u(\beta)-u(\alpha)}\, g(\beta) + \frac{u(\beta)-u(\xi)}{u(\beta)-u(\alpha)}\, g(\alpha).$$

DEFINITION 4. Let $\varepsilon > 0$ be arbitrarily small. Then

$$\int_a^b f(x) \frac{d^2 g(x)}{du(x)}$$

is the number I, if it exists uniquely, and there is a number $\delta(\varepsilon)$ such that for $x_{i-1} \leqq \leqq \xi_i \leqq x_{i+1}$, $i=1, 2, \ldots, k-1$, $\xi_0=a$, $\xi_k=b$

$$\left|I-\sum_{i=0}^{k} f(\xi_i)[g_u(x_{i+1}, x_i)-g_u(x_i, x_{i-1})]\right| < \varepsilon$$

whenever $\|\Delta\|<\delta(\varepsilon)$. When the integral exists, it is said $(f, g)\in RS_u\ [a, b]$.

In this paper along with some new results we obtain certain modifications of the results of [2]. The following definition is needed for the purpose.

DEFINITION 5. We consider a Δ subdivision and make the following definitions:

$$M_i = \sup_{x_{i-1}\leqq x\leqq x_{i+1}} f(x), \quad i = 1, 2, \ldots, k-1;$$

$$m_i = \inf_{x_{i-1}\leqq x\leqq x_{i+1}} f(x), \quad i = 1, 2, \ldots, k-1;$$

$$M_0 = \sup_{x_0\leqq x\leqq x_1} f(x), \qquad m_0 = \inf_{x_0\leqq x\leqq x_1} f(x);$$

$$M_k = \sup_{x_{k-1}\leqq x\leqq x_k} f(x), \quad m_k = \inf_{x_{k-1}\leqq x\leqq x_k} f(x);$$

$$S_\Delta = \sum_{i=0}^{k} M_i[g_u(x_{i+1}, x_i)-g_u(x_i, x_{i-1})],$$

$$s_\Delta = \sum_{i=0}^{k} m_i[g_u(x_{i+1}, x_i)-g_u(x_i, x_{i-1})].$$

Then

$$S_\Delta - s_\Delta = \sum_{i=0}^{k} O_i[g_u(x_{i+1}, x_i)-g_u(x_i, x_{i-1})]$$

where $O_i=M_i-m_i$, $i=0, 1, 2, \ldots, k$, is called the oscillatory sum corresponding to the subdivision Δ and is denoted by ω_Δ.

We shall often, for the sake of simplicity, use the notation $d(g; x_{i-1}, x_i, x_{i+1})$ for the expression $[g_u(x_{i+1}, x_i)-g_u(x_i, x_{i-1})]$. If no confusion arises we write $d(x_{i-1}, x_i, x_{i+1})$ for $d(g; x_{i-1}, x_i, x_{i+1})$.

2. Existence theorem and other results

RUSSELL [2] proved the following existence theorem: if f is continuous on $[a', b']$, g is u-convex on $[a, b]$ and g, u satisfy condition A, then $(f, g)\in RS_u\ [a, b]$.

We note that the theorem may be proved under weaker conditions on f. Our theorem runs as follows:

THEOREM 1. *If f is continuous and g is u-convex on $[a, b]$ and g, u satisfy condition* A, *then* $(f, g)\in RS_u\ [a, b]$.

PROOF. For the proof we only construct M_i, m_i $(i=0, 1, \ldots, k)$, S_Δ, s_Δ as in Definition 5 and make the definitions

$$U = \inf_{\Delta}^{*} S_\Delta \quad \text{and} \quad L = \sup_{\Delta} s_\Delta .$$

We do not go into the details, because the detailed proof may be carried on now step by step as made by RUSSELL [2].

Throughout we shall assume that f is bounded and g is u-convex on $[a, b]$ and g, u satisfy condition A.

The following theorem is vital to prove Theorems 3 and 5.

THEOREM 2. *A necessary and sufficient condition that* $(f, g)\in RS_u\ [a, b]$ *is that the oscillatory sum tends to zero as the norm of the subdivision tends to zero.*

PROOF. We only prove the necessary part because the sufficient part is routine.

If possible, suppose that the oscillatory sum does not tend to zero as the norm of the subdivision tends to zero. Then there exists $d>0$ and a sequence of subdivisions Δ_n with $\|\Delta_n\|\to 0$ as $n\to\infty$ such that $S_{\Delta_n}-s_{\Delta_n}>d$, $n=1, 2, \ldots$. Let Δ_n be given by

$$\Delta_n\colon\ x_{n,-1} < a = x_{n,0} < x_{n,1} < \ldots < x_{n,k_n} = b < x_{n,k_n+1},$$

where $a'\leqq x_{n,-1}<a$, $b<x_{n,k_n+1}\leqq b'$; and let

$$M_{n,i} = \sup_{x_{n,i-1}\leqq x\leqq x_{n,i+1}} f(x), \quad i = 1, 2, \ldots, k_n-1;$$

$$m_{n,i} = \inf_{x_{n,i-1}\leqq x\leqq x_{n,i+1}} f(x), \quad i = 1, 2, \ldots, k_n-1;$$

$$M_{n,0} = \sup_{x_{n,0}\leqq x\leqq x_{n,1}} f(x), \qquad m_{n,0} = \inf_{x_{n,0}\leqq x\leqq x_{n,1}} f(x),$$

$$M_{n,k_n} = \sup_{x_{n,k_n-1}\leqq x\leqq x_{n,k_n}} f(x), \quad m_{n,k_n} = \inf_{x_{n,k_n-1}\leqq x\leqq x_{n,k_n}} f(x),$$

$$S_{\Delta_n} = \sum_{i=0}^{k_n} M_{n,i}\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1}),$$

$$s_{\Delta_n} = \sum_{i=0}^{k_n} m_{n,i}\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1}).$$

Let $\varepsilon>0$ be arbitrary. Since $g'_u(a)$ and $g'_u(b)$ exist (by condition A), there exists $\delta>0$ such that

$$d(x_{n,-1}, a, \xi) < \varepsilon/3(M-m+1)$$

and

$$d(\eta, b, x_{n,k_n+1}) < \varepsilon/3(M-m+1)$$

for all $\xi\in(a, a+\delta)$ and $\eta\in(b-\delta, b)$, where M, m are the upper and lower bounds of f in $[a, b]$. Since $\|\Delta_n\|\to 0$ we may assume $x_{n,1}$ belonging to $(a, a+\delta)$, x_{n,k_n-1} to $(b-\delta, b)$, $\xi_{n,0}=a$, $\xi_{n,k_n}=b$, so that

$$f(\xi_{n,0})\, d(x_{n,-1}, x_{n,0}, x_{n,1})$$

and

$$f(\xi_{n,k_n})\, d(x_{n,k_n-1}, x_{n,k_n}, x_{n,k_n+1})$$

differ from $m_{n,0}d(x_{n,-1}, x_{n,0}, x_{n,1})$ and

$$m_{n,k_n}d(x_{n,k_n-1}, x_{n,k_n}, x_{n,k_n+1})$$

respectively by less than $\varepsilon/3$. Also we may choose $\xi_{n,i}$ in $[x_{n,i-1}, x_{n,i+1}]$, $i=1, 2, \ldots$ $\ldots, k_n-1$, such that

$$|f(\xi_{n,i})-m_{n,i}| < \varepsilon/3[g'_u(b)-g'_u(a)+1],$$

so that

$$\sum_{i=0}^{k_n-1} f(\xi_{n,i})\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1})$$

differs from

$$\sum_{i=0}^{k_n-1} m_{n,i}\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1})$$

by less than $\varepsilon/3$.

It therefore follows that the approximating sum for the RS_u-integral,

$$\sum_{i=0}^{k_n} f(\xi_{n,i})\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1})$$

differs from s_{Δ_n} by less than ε. Likewise, it is possible to choose $\zeta_{n,i}$ in $[x_{n,i-1}, x_{n,i+1}]$, $i=1, 2, \ldots, k_n-1$,

$$\zeta_{n,0} = a, \quad \zeta_{n,k_n} = b,$$

such that

$$\sum_{i=0}^{k_n} f(\zeta_{n,i})\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1})$$

differs from S_{Δ_n} by a small number. Consequently as $\|\Delta_n\| \to 0$ the limit of

$$\sum_{i=0}^{k_n} f(\xi_{n,i})\, d(x_{n,i-1}, x_{n,i}, x_{n,i+1})$$

does not exist. This proves the theorem.

THEOREM 3. *Let $g'_u(c)$ exist where $a<c<b$. If $(f, g)\in RS_u$ $[a, b]$, then $(f, g)\in RS_u$ $[a, c]$ and RS_u $[c, b]$. Conversely if $(f, g)\in RS_u$ $[a, c]$ and RS_u $[c, b]$, then $(f, g)\in RS_u$ $[a, b]$. In either case*

$$\int_a^b f(x)\frac{d^2g(x)}{du(x)} = \int_a^c f(x)\frac{d^2g(x)}{du(x)} + \int_c^b f(x)\frac{d^2g(x)}{du(x)}.$$

PROOF. Let $\varepsilon>0$ be arbitrary. We consider subdivisions Δ_1 and Δ_2 of $[a, c]$ and $[c, b]$ respectively as

$$x_{-1} < a = x_0 < x_1 < \ldots < x_{p-1} < x_p = c < x_{p+1}$$

and

$$x_{p-1} < c = x_p < x_{p+1} < \ldots < x_k = b < x_{k+1},$$

where

$$a' \leqq x_{-1} < a, \quad c < x_{p+1} < c+\delta_1, \quad c-\delta_1 < x_{p-1} < c, \quad b < x_{k+1} \leqq b'$$

and $\delta_1(\varepsilon)=\delta_1>0$ is such that

$$d(\beta, c, \alpha) < \varepsilon/4(M-m+1)$$

for $c<\alpha<c+\delta_1$, $c-\delta_1<\beta<c$; M and m being the upper and lower bounds of $f(x)$ in $[a, b]$. Let $\Delta=\Delta_1 \cup \Delta_2$ so that

$$\Delta: x_{-1} < a = x_0 < x_1 < \ldots < x_{p-1} < c = x_p < x_{p+1} < \ldots < x_k = b < x_{k+1}.$$

The oscillatory sum corresponding to Δ is given by

$$\omega_\Delta = \sum_{i=0}^{k} O_i\, d(x_{i-1}, x_i, x_{i+1}) =$$

$$= \sum_{i=0}^{p-1} O_i\, d(x_{i-1}, x_i, x_{i+1}) + O'_c\, d(x_{p-1}, c, x_{p+1}) +$$

$$+ O''_c\, d(x_{p-1}, c, x_{p+1}) + \sum_{i=p+1}^{k} O_i\, d(x_{i-1}, x_i, x_{i+1}) +$$

$$+ (O_p - O'_c - O''_c)\, d(x_{p-1}, c, x_{p+1})$$

where O'_c and O''_c are the oscillations of $f(x)$ in $[x_{p-1}, c]$ and $[c, x_{p+1}]$ respectively.

Let ω'_{Δ_1} and ω''_{Δ_2} denote the oscillatory sums over $[a, c]$ and $[c, b]$ corresponding to the respective subdivisions Δ_1 and Δ_2. Then

$$\omega_\Delta = \omega'_{\Delta_1} + \omega''_{\Delta_2} + (O_p - O'_c - O''_c)\, d(x_{p-1}, c, x_{p+1}). \tag{1}$$

We first suppose that $(f, g) \in RS_u\ [a, b]$. Then there exists a $\delta_2(\varepsilon)>0$ such that

$$\omega_\Delta < \varepsilon/4 \quad \text{whenever} \quad \|\Delta\| < \delta_2(\varepsilon).$$

Since each of O'_c, O''_c, O_p is $\leqq M-m$, it follows that

$$(O'_c + O''_c - O_p)\, d(x_{p-1}, c, x_{p+1}) < 3\varepsilon/4. \tag{2}$$

Hence by relation (1), we obtain

$$\omega'_{\Delta_1} + \omega''_{\Delta_2} < \varepsilon/4 + 3\varepsilon/4 = \varepsilon$$

whenever norm of each subdivision Δ_1, Δ_2 is $<\delta=\min(\delta_1, \delta_2)$. Consequently

$$\omega'_{\Delta_1} < \varepsilon, \quad \omega''_{\Delta_2} < \varepsilon \quad \text{whenever} \quad \|\Delta_1\| < \delta(\varepsilon), \quad \|\Delta_2\| < \delta(\varepsilon).$$

Hence by Theorem 2, it follows that

$$(f, g) \in RS_u[a, c] \quad \text{and} \quad RS_u[c, b].$$

Conversely, if $(f, g) \in RS_u\ [a, c]$ and $RS_u\ [c, b]$ then by (1), (2) and Theorem 2, it easily follows that $(f, g) \in RS_u\ [a, b]$.

We now establish the equality. Let $\varepsilon>0$ be arbitrary. There exists a $\delta(\varepsilon)>0$ such that for any subdivision

$$x_{-1} < a = x_0 < x_1 < \ldots < x_{p-1} < x_p = c < x_{p+1} < \ldots < x_k = b < x_{k+1}$$

where $a' \leqq x_{-1} < a$, $b < x_{k+1} \leqq b'$, with c as a point of subdivision and of norm $< \delta(\varepsilon)$, we have

$$\left| \sum_{i=0}^{p} f(\xi_i)\, d(x_{i-1}, x_i, x_{i+1}) - \int_a^c f(x) \frac{d^2 g(x)}{du(x)} \right| < \varepsilon/4,$$

$$\left| \sum_{i=p}^{k} f(\xi_i)\, d(x_{i-1}, x_i, x_{i+1}) - \int_c^b f(x) \frac{d^2 g(x)}{du(x)} \right| < \varepsilon/4,$$

$$\left| \sum_{i=0}^{k} f(\xi_i)\, d(x_{i-1}, x_i, x_{i+1}) - \int_a^b f(x) \frac{d^2 g(x)}{du(x)} \right| < \varepsilon/4,$$

where

$$x_{i-1} \leqq \xi_i \leqq x_{i+1}, \quad i = 1, 2, \ldots, p-1, p+1, \ldots, k-1;$$

$$\xi_0 = a, \quad \xi_k = b, \quad \xi_p = c.$$

Since

$$\sum_{i=0}^{p} + \sum_{i=p}^{k} = \sum_{i=0}^{k} + f(\xi_p)\, d(x_{p-1}, c, x_{p+1}),$$

we deduce that

$$\left| \int_a^b f(x) \frac{d^2 g(x)}{du(x)} - \int_a^c f(x) \frac{d^2 g(x)}{du(x)} - \int_c^b f(x) \frac{d^2 g(x)}{du(x)} \right| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves the theorem.

NOTE. RUSSELL [2] obtained this theorem only in one part. His theorem runs as follows.

THEOREM. *If $(f, g) \in RS_u[a, c]$ and $RS_u[c, b]$, where $a < c < b$, and if f is continuous and g is u-differentiable at c, with g having bounded u-incrementary ratios in some neighbourhood of c, then $(f, g) \in RS_u[a, b]$, and*

$$\int_a^b f(x) \frac{d^2 g(x)}{du(x)} = \int_a^c f(x) \frac{d^2 g(x)}{du(x)} + \int_c^b f(x) \frac{d^2 g(x)}{du(x)}.$$

Our assumptions in Theorem 3 are that (i) f is bounded in $[a, b]$, (ii) g is u-convex in $[a, b]$, (iii) g, u satisfy condition A, (iv) $g'_u(c)$ exists.

It is not difficult to show that (ii), (iii) and Lemma 1.2 of [2] together imply that g has bounded u-incrementary ratios in $[a, b]$. But we do not need the continuity of f at c even to prove that part of Theorem 3 which Russell obtained.

Let $(f, g) \in RS_u[a, b]$ and $g'_u(x)$ exist in (a, b). By Theorem 3, $(f, g) \in RS_u[a, x]$ for $x \in (a, b]$. Let

$$\varphi(x) = \int_a^x f(t) \frac{d^2 g(t)}{du(t)}, \quad a < x \leqq b \quad \text{with} \quad \varphi(a) = 0.$$

THEOREM 4. *Let f be continuous in $[a, b]$ and let $g_u(x)$ exist and be u-convex in $[a, b]$. Then $\varphi(x)$ is u-differentiable in $[a, b]$ except perhaps an enumerable set, and*

$$\varphi'_u(x) = f(x) g''_u(x) \quad \text{on} \quad [a, b] - E$$

where E is an enumerable set and $g''_u(x)$ represents the twice u-derivative of $g(x)$.

To prove the theorem we require the following lemmas of which the proof of Lemma 1 can be constructed similarly as in [1].

LEMMA 1. *If g is u-convex on $[a, b]$, then both $g_u^+(x)$, $g_u^-(x)$ exist everywhere in (a, b). Further $g_u'(x)$ exists on $[a, b]$ except atmost an enumerable set.*

LEMMA 2. *Let $a \leqq c < d \leqq b$ and $g_u'(c)$, $g_u'(d)$ exist, then*

$$\int_c^d \frac{d^2 g(x)}{du(x)} = g_u'(d) - g_u'(c).$$

PROOF. We consider a subdivision

$$\Delta' : y_{-1} < c = y_0 < y_1 < \ldots < y_k = d < y_{k+1}.$$

Since g is u-convex on $[a, b]$ and g, u satisfy condition A, $\int_a^b \frac{d^2 g(x)}{du(x)}$ exists by Theorem 1 and hence by Theorem 3, $\int_c^d \frac{d^2 g(x)}{du(x)}$ exists. Let $\varepsilon > 0$ be arbitrary. There exists a $\delta(\varepsilon) > 0$ such that

$$\left| \int_c^d \frac{d^2 g(x)}{du(x)} - \sum_{i=0}^{k} d(y_{i-1}, y_i, y_{i+1}) \right| < \varepsilon/2$$

whenever $\|\Delta'\| < \delta(\varepsilon)$

i.e.
$$\left| \int_c^d \frac{d^2 g(x)}{du(x)} - [g_u(y_{k+1}, d) - g_u(c, y_{-1})] \right| < \varepsilon/2$$

whenever $\|\Delta'\| < \delta(\varepsilon)$. We obtain, in the limit,

$$\int_c^d \frac{d^2 g(x)}{du(x)} = g_u'(d) - g_u'(c).$$

We now prove the theorem.

By Lemma 1, $g_u'(x)$ exists on $[a, b] - E$, where E is an enumerable set. For x, $x+h$ $(h \neq 0)$ in $[a, b]$

$$\varphi(x+h) - \varphi(x) = \int_x^{x+h} f(t) \frac{d^2 g(t)}{du(t)} =$$

$$= f(x+\theta h) \int_x^{x+h} \frac{d^2 g(t)}{du(t)}, \quad 0 \leqq \theta \leqq 1$$

$$= f(x+\theta h)[g_u'(x+h) - g_u'(x)], \quad 0 \leqq \theta \leqq 1,$$

by Lemma 2, and so

$$\frac{\varphi(x+h) - \varphi(x)}{u(x+h) - u(x)} = f(x+\theta h) \frac{g_u'(x+h) - g_u'(x)}{u(x+h) - u(x)}, \quad 0 \leqq \theta \leqq 1.$$

Hence as $h \to 0$, we obtain $\varphi_u'(x) = f(x) g_u''(x)$ for all x in $[a, b] - E$. This proves the theorem.

THEOREM 5. *Let $g(x)$ be u-convex on $[a, b]$ and g, u satisfy condition* A. *Let $\{f_n(x)\}$ be a sequence of functions which converges uniformly to $f(x)$ on $[a, b]$. If for all n, $(f_n, g) \in RS_u[a, b]$, then $(f, g) \in RS_u[a, b]$ and*

$$\lim_{n \to \infty} \int_a^b f_n(x) \frac{d^2 g(x)}{du(x)} = \int_a^b f(x) \frac{d^2 g(x)}{du(x)}.$$

We omit the proof of the theorem because the construction of the proof with the help of Theorem 2 is not difficult.

REFERENCES

[1] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G.: *Inequalities*, Cambridge, 1964, p. 91.
[2] RUSSELL, A. M.: Functions of bounded second variation and Stieltjes-type integrals, *J. London Math. Soc.* (2) **2** (1970), p. 193.

Department of Mathematics, University of Kalyani, West Bengal, India

(Received March 18, 1976; revised August 1, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 125—144

НА ПУТИ К ГИПОТЕЗЕ ШАНУЭЛЛА. АЛГЕБРАИЧЕСКИЕ КРИВЫЕ ВБЛИЗИ ТОЧКИ

ОБШАЯ ТЕОРИЯ ЦВЕТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
Г. В. ЧУДНОВСКИЙ

Как видно из заглавия, нас интересует какая-то определенная точка и кривые, проходящие возле нее. Нетрудно догадаться, что существование какого-то особенного семейства, проходящего возле точки, связано с особенной арифметической природой точки. Так оно и есть — рассматривается точка (в пространстве), координаты которой — базис трансцендентности для семейства чисел, известних благодаря гипотезе Шануэлла [2]:

$$\alpha_1, \ldots, \alpha_n, \quad e^{\alpha_1}, \ldots, e^{\alpha_n}.$$

Такое внимание к алгебраическим кривым и их семействам связано с одним явлением, которое всегда казалось автору неожиданным. Речь идет о известном более 20 лет существовании плотной последовательности алгебраических кривых, проходящих вблизи этой «ключевой» точки трансцендентных чисел. Этим явлением, конечно, всегда пользовались при исследовании. Так, в работах Ленга [10], Гельфонда [5], Броунвейля [3] и автора [6] семейство алгебраических кривых играло заметную роль. Затем, в исследованиях автора [7], [8], по гипотезе Шануэлла, различные семейства кривых и гиперповерхностей вблизи базиса трансцендентности вообще стали основным объектом исследования.

В настоящей работе будет детально рассмотрена двумерная ситуация и интересующие нас множества кривых без всякой связи с их присхождениями. Эта работа является вступлением в длинную (по ограниченную) последовательность работ автора, посвященных оценкам степеней трансцендентности полей, возникающих из гипотезы Шануэлла. Получаемые в конце концов оценки будут близки к требуемым и значительно лучше оценок [7].

§ 0. Предварительный материал

Все обозначения, определения и терминология стандартны [10], [14]. Остановимся только на одном определении. Для полинома $P(x_1, \ldots, x_k) \in C[x_1, \ldots, x_k]$ через $H(P)$ обозначается максимум модулей коэффициентов $P(x_1, \ldots, x_k)$, а через $d(P)$ (степень $P(x_1, \ldots, x_k)$) — $d_{x_k}(P)+\ldots+d_{x_k}(P)$. Полагая $t(P)=\max\{\ln(HP), d(P)\}$ назовем $t(P)$ — типом $P(x_1, \ldots, x_k)$.

Сразу же укажем, что все рассмотрения в работе необходимо проводить для случая одновременно растущих высот и степеней полиномов. Поэтому воспользуемся старой идеей [5], [10] и просто будем исследовать все с точки зрения роста типа полиномов. Это значительно важнее, чем предполагать

ограниченность степени, как иногда делают. Более того, рассмотрения типа важно и для применений к алгебраической независимости [5]. Заметим, что осуществить перенос наших соображений на случай одновременно применяющихся высот и степеней очень просто. Поэтому будем обращать внимание только на $t(P)$ и больше к этому вопросу не будем возвращаться.

Чтобы стало ясно, с какими именно последовательностями полиномов придется иметь дело, укажем на канонический пример.

Пример 0.1. Пусть $\theta=(\theta_1, \ldots, \theta_n)\in C^n$, где $\theta_1, \ldots, \theta_n$ алгебраически независимы над Q. Тогда для всякого натурального H существует полином $P_{H,\vec{\theta}}=P_H$, $P_H(x_1, \ldots, x_n)\in Z[x_1, \ldots, x_n]$ такой, что

1) $P_H(x_1, \ldots, x_n) \not\equiv 0$;

2) $t(P_H) \leqq H$;

3) $|P_H(\theta_1, \ldots, \theta_n)| < \exp(-c_0 H^{n+1})$,

где $c_0=c_0(n)>0$ — постоянная, зависящая только от n.

Семейство гиперповерхностей $\{P_{H,\vec{\theta}}(x_1, \ldots, x_n)=0\colon H\geqq 1\}$, удовлетворяющее условиям 1)—3), назовем семейством Дирихле вблизи точки $\vec{\theta}$.

Такое название объясняется тем, что существование полиномов, удовлетворяющих 1)—3), обеспечивается «принципом ящиков» Дирихле. Соответствующее доказательство см. [19], [18]. Впрочем, можно, дать много интересных (не эквивалентных) доказательств 0.1. Через теорему Миньковского 0.1 выводится в [19].

Теперь уже ясно, что нас интересует. Речь идет о последовательностях полиномов $P_H(x_1, \ldots, x_n)\in Z[x_1, \ldots, x_n]$, удовлетворяющих условиям 1), 2) из 0.1, а вместо 3) — более сильному. Например,

3′) $|P_H(\theta_1, \ldots, \theta_n)| < \exp(-H^{n+1}\varphi(H))$,

где $\varphi(H)$ монотонно возрастающая неотрицательная функция; $\lim\limits_{H\to\infty} \varphi(H)=\infty$.

Тут и возникает основной вопрос, ключевой для исследований по гипотезе Шануэлла.

Вопрос 0.2. Для каких точек $\vec{\theta}=(\theta_1, \ldots, \theta_n)\in C^n$ существует семейство полиномов $\{P_H(x_1, \ldots, x_n)\in Z[x_1, \ldots, x_n]\colon H\geqq 1\}$, удовлетворяющих 1), 2) и 3′)?

Поскольку условия 1)—3′) накладывают серьезные арифметико-аналитические ограничения на $\vec{\theta}$, то Ленг [10] предположил, естественно, что $\theta_1, \ldots, \theta_n$, удовлетворяющие условиям 0.2, должны быть алгебраически зависимы над Q.

У него было веское основание сделать это предположение. Речь идет о признаке трансцендентности одного числа (т. е. $n=1$), доказанном им [10], [11], а ранее не в полной общности установленном Гельфондом [5].

Признак 0.3. Пусть $\theta\in C$. Допустим также, что F — монотонно возрастающая неотрицательная функция натурального аргумента H, причем $\lim\limits_{H\to\infty} F(H)=\infty$ и существует постоянная a_0, для которой $F(H+1)\leqq a_0 F(H)$ для

всех $H \geqq a_1$. Если при всяком $H \geqq a_2$ существует такой полином $P_H(x) \in Z[x]$, $P_H(x) \not\equiv 0$, $t(P_H) \leqq F(H)$, что

$$|P_H(\theta)| < \exp\left(-C_1 F(H)^2\right),$$

для некоторой постоянной $C_1 = C_1(a_0, a_1, a_2) > 0$, то число θ — алгебраическое.

В частности, при $F(H) \equiv H$: $H \geqq 1$ получаем, что при $n=1$ всякое $\theta \in C$, удовлетворяющее условиям вопроса 0.2, должно быть алгебраическим.

Поэтому появление предположения Ленга выглядело естественным. Им никто специально не занимался (считая простым), пока Э. Бомбьери (см. [11]) не заметил, что оно просто неверно. Более того, неверен никакой его вариант с любой функцией $\varphi(H)$ в 3′).

Дело тут в следующем. При исследовании сингулярных линейных форм Дж. Касселс [1] доказал такое утверждение:

Лемма 0.4. *Для всякой монотонной неотрицательной функции* $\varphi(t): t \in R^+$ *существуют два алгебраически независимых числа* θ_1, θ_2 *такие, что неравенство*

$$|x\theta_1 + y\theta_2 + z| < \exp\left(-\varphi(H)\right)$$

разрешимо в целых рациональных числах x, y, z; $X = \max(|x|, |y|, |z|) \not\equiv 0$, $X \leqq H$, *для всех* $H \geqq c_2$. *Здесь* $c_2 = c_2(\varphi) \geqq 0$ — *некоторая постоянная.*

Следовательно, при $n \geqq 2$ уже нельзя рассчитывать на точный ответ на вопрос 0.2. Тем не менее, точки $\vec{\theta}$, удовлетворяющие 0.2, имеют специфическую арифметическую природу. Речь идет о т. н. полях «конечного типа транцендентности», такие внесенными в заголовок. Приведем

Определение 0.5. Пусть $K = Q(\vartheta_1, \ldots, \vartheta_d, \omega)$ подполе C степени трансцендентности d над Q, а $\vartheta_1, \ldots, \vartheta_d$ его базис трансцендентности. Поле K имеет тип трансцендентности $\leqq \tau (< \infty)$, если существует такая постоянная $C(K) > 0$, что для всякого полинсма $P(x_1, \ldots, x_d) \in Z[x_1, \ldots, x_d]$, $P(x_1, \ldots, x_d) \not\equiv 0$ имеем

$$|P(\vartheta_1, \ldots, \vartheta_d)| > \exp\left(-C(K) t(P)^{\tau}\right). \tag{0.1}$$

Замечание 0.6. Поле K имеет тип трансцендентности $\leqq (\tau, \tau')$, если (0.1) заменяется на

$$|P(\vartheta_1, \ldots, \vartheta_d)| > \exp\left(-C(K) \cdot t(P)^{\tau} (\ln t(P))^{\tau'}\right). \tag{0.2}$$

Эти определения принадлежат [10], [14], [3].

Как показывает пример 0.1, поле $K \subset C$ степени трансцендентности d не может иметь типа трансценцентности $\leqq d+1$. Условие конечности типа трансцендентности является арифметическим ограничением, не имеющим всегда места. Однако для почти всех чисел $\theta \in C^n$ поле $K_{\vec{\theta}} = Q(\theta_1, \ldots, \theta_n)$ имеет конечный тип трансцендентности; при $n=1$ этот тип даже равен двум. Пот этому крайне ценно (хотя бы для приложений), что координатам θ_i числа $\vec{\theta} = (\theta_1, \ldots, \theta_n)$, удовлетворяющего 0.2, отвечают поля $Q(\theta_i)$ типа трансцендентности > 2.

Точнее, существует результат, принадлежащий Д. Броунвейлю [3]. Приведем его в той форме, в какой его используем в работе:

Теорема 0.7. *Пусть* $\theta_1, \theta_2 \in C$ *и* θ_1, θ_2 *алгебраически независимы над* Q. *Если для всякого* $H \geqq 1$ *существует полином* $P_H(x, y) \in Z[x, y]$, $P_H(x, y) \not\equiv 0$, $t(P_H) \leqq H$ *такой, что*

$$|P_H(\theta_1, \theta_2)| < \exp(-c_3 H^{\mu}) \tag{0.3}$$

для какой-то постоянной $c_3 = c_3(\theta_1, \theta_2) > 0$, *то поле* $Q(\theta_1)$ *(и поле* $Q(\theta_2)$*) имеет тип трансцендентности* $\geqq \mu/2$. *Если же* (0.3) *заменено на*

$$|P_H(\theta_1, \theta_2)| < \exp\left(-c_3 H^{\mu} \varphi(H)\right) \tag{0.4}$$

для неограниченной неотрицательной функции $\varphi(t)$: $t \in R^+$, *то* $Q(\theta_1)$ *(и* $Q(\theta_2)$*) имеет тип трансцендентности* $> \mu/2$.

В последнем случае это означает, что для любого $c_4 > 0$ неравенства

$$|P(\theta_1)| < \exp\left(-c_4 t(P)^{\mu/2}\right)$$

разрешимы в бесконечном числе различных полиномов $P(x) \in Z[x]$, $P(x) \not\equiv 0$.

Как будет ясно из дальнейшего, это очень слабое утверждение даже по сравнению с применяемыми результатами. Точный характер оценки будет обсуждаться в этой и в последующих работах автора на ту же тему.

Для усиления этого результата в последующих работах на основе методов, излагаемых ниже, будет привлекаться разнообразная алгебраическая, аналитическая и комбинаторная техника. Понятно, что максимум, на что можно рассчитывать, это на то, что тип $Q(\theta_1)$ не менее μ. Но и эта надежда слишком радужная, как показывает пример семейства Дирихле, отвечающий точке (θ_1, θ_2) «общего положения» из C^2. Тогда $\mu = 3$, хотя $Q(\theta_1)$ и $Q(\theta_2)$ имеют «почти всегда» тип $\tau = 2$. Во всяком случае, тип $Q(\theta_1)$ будет ближе к μ, чем к $\mu/2$.

В конце работы подробно объясняется, зачем нужны все оценки такого типа.

§ 1. Плотное семейство кривых. Раскраска

В этой части работы анализируется о общих позиций семейство кривых, подходящих близко к точке $\vec{\theta} \in C^2$, удовлетворяющих условиям вопроса 0.2. Кривые $P_H(x, y) = 0$, отвечающие различным $H \geqq 1$ «раскрашиваются» в два цвета: красный и синий. Раскраска носит условный характер и введена для иллюстративности. Она будет полезна также при непосредственном исследовании семейств кривых, относящихся к гипотезе Шануэлла.

Из всех вспомогательных фактов нам нужен пока лишь один:

Лемма 1.1. *Пусть* $P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)$, $P(x_1, \ldots, x_n) \in C[x_1, \ldots, x_n]$. *Если* $P(x_1, \ldots, x_n) = P_1(x_1, \ldots, x_n) \times \ldots \times P_m(x_1, \ldots, x_n)$, *то для высоты* $H(P)$ *полинома* $P(x_1, \ldots, x_n)$ *имеем оценку снизу*

$$H(P) \geqq e^{-d(P)} H(P_1) \ldots H(P_m).$$

Доказательство этой леммы; историю и ее уточнения можно найти в [5], [14]. Приведем следствие из 1.1, которое и будет использоваться.

Следствие 1.2. *Пусть* $P_1(x_1, \ldots, x_n)$, $P_2(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$. *Если полином* $(P_1(x_1, \ldots, x_n))^s$; $s \geqq 1$, *делит* $P_2(x_1, \ldots, x_n)$, *то для типа* $t(P_1)$ *полинома* $P_1(x_1, \ldots, x_n)$ *получаем оценку*

$$t(P_1)s \leqq 2t(P_2).$$

Теперь можно перейти к непосредственному исследованию семейств полиномов. Вначале докажем одну лемму в духе [3], [6], но примечательную тем, что исключения переменных не происходит. Метод доказательства этой леммы кочует из одной статьи в другую с незначительными изменениями.

Лемма 1.3. *Пусть* $P(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$ *и* $\vec{\theta} = (\theta_1, \ldots, \theta_n) \in C^n$. *Если*

$$|P(\theta_1, \ldots, \theta_n)| \leqq \varepsilon < 1 \tag{1.1}$$

То либо существует делитель $R(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$ *полинома* $P(x_1, \ldots, x_n)$, *являющийся степенью неприводимого над* Q *полинома, для которого*

$$|R(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3} \tag{1.2}$$

либо существуют два взанимо простых делителя $P_1(x_1, \ldots, x_n)$, $P_2(x_1, \ldots, x_n) \in$ $\in Z[x_1, \ldots, x_n]$ *полинома* $P(x_1, \ldots, x_n)$, *для которых*

$$|P_1(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3}, \quad |P_2(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3}. \tag{1.3}$$

Доказательство. Обозначим через $R_1^{h_1}, \ldots, R_k^{h_k}$ все различные делители $P(x_1, \ldots, x_n)$, являющиеся степенями неприводимых над Q полиномов из $Z[x_1, \ldots, x_n]$. Тогда, применяя лемму Гаусса, получаем

$$P(x_1, \ldots, x_n) = A(R_1(x_1, \ldots, x_n))^{h_1} \ldots (R_k(x_1, \ldots, x_n))^{h_k}, \tag{1.4}$$

где $A \in Z$, $A \neq 0$. Для устранения тривиальных случаев предположим, что $\theta_1, \ldots, \theta_n$ алгебраически незавивсимы или, точнее, что

$$P(\theta_1, \ldots, \theta_n) \neq 0. \tag{1.5}$$

Если бы (1.5) не имело места, то получалось бы сразу же $R_i(\theta_1, \ldots, \theta_n) = 0$ для некоторого i, $1 \leqq i \leqq k$ и (1.2) имеет место при $R(x_1, \ldots, x_n) = R_i(x_1, \ldots, x_n)$. Поэтому можно считать, что (1.5) имеет место.

Далее нумерацию полиномов $R_i : 1 \leqq i \leqq k$ выбираем таким образом, чтобы

$$|R_1(\theta_1, \ldots, \theta_n)|^{h_1} \leqq \ldots \leqq |R_k(\theta_1, \ldots, \theta_n)|^{h_k}. \tag{1.6}$$

Далее, т. к. $\varepsilon < 1$, то из (1.1), (1.4) следует, что для некоторого l, $1 \leqq l \leqq k$ имеем

$$|R_1(\theta_1, \ldots, \theta_n)|^{h_1} \leqq \ldots \leqq |R_l(\theta_1, \ldots, \theta_n)|^{h_l} < |A|^{-1} \leqq \tag{1.7}$$

$$\leqq |R_{l+1}(\theta_1, \ldots, \theta_n)|^{h_{l+1}} \leqq \ldots \leqq |R_k(\theta_1, \ldots, \theta_n)|^{h_k}.$$

Понятно, что существует такое i, $1 \leqq i \leqq l$, что имеем, с одной стороны

$$|A| \cdot |R_1(\theta_1, \ldots, \theta_n)|^{h_1} \ldots |R_i(\theta_1, \ldots, \theta_n)|^{h_i} < \tag{1.8}$$

$$< |R_{i+1}(\theta_1, \ldots, \theta_n)|^{h_{i+1}} \ldots |R_k(\theta_1, \ldots, \theta_n)|^{h_k},$$

а с другой

$$|A| \cdot |R_1(\theta_1, \ldots, \theta_n)|^{h_1} \ldots R_{i-1}(\theta_1, \ldots, \theta_n)|^{h_{i-1}} \geqq \tag{1.9}$$

$$\geqq |R_i(\theta_1, \ldots, \theta_n)|^{h_i} \ldots |R_k(\theta_1, \ldots, \theta_n)|^{h_k}.$$

Воспользуемся сначала (1.8). Обозначаем $P_1'(x_1, \ldots, x_n) = A \cdot (R_1(x_1, \ldots, x_n))^{h_1} \ldots (R_i(x_1, \ldots, x_n))^{h_i}$ и $P_2'(x_1, \ldots, x_n) = (R_{i+1}(x_1, \ldots, x_n))^{h_{i+1}} \ldots (R_k(x_1, \ldots, x_n))^{h_k}$. Тогда $P_1'(x_1, \ldots, x_n)$, $P_2'(x_1, \ldots, x_n)$ — взаимно простые делители $P(x_1, \ldots, x_n)$. Допуская, что (1.3) для них не имеет места, получаем, согласно (1.8), что

$$|R_{i+1}(\theta_1, \ldots, \theta_n)|^{h_{i+1}} \ldots |R_k(\theta_1, \ldots, \theta_n)|^{h_k} > \varepsilon^{1/3}. \tag{1.10}$$

Воспользуемся теперь (1.9). Полагаем теперь

$$P_1''(x_1, \ldots, x_n) = A \cdot (R_1(x_1, \ldots, x_n))^{h_1} \ldots (R_{i-1}(x_1, \ldots, x_n))^{h_{i-1}}$$

и $P_2''(x_1, \ldots, x_n) = (R_i(x_1, \ldots, x_n))^{h_i} \ldots (R_k(x_1, \ldots, x_k))^{h_k}$. Снова $P_1''(x_1, \ldots, x_n)$, $P_2''(x_1, \ldots, x_n)$ — взаимно простые делители $P(x_1, \ldots, x_n)$. Поэтому, если (1.3) не имеет для них места, то по (1.9):

$$|A| \cdot |R_1(\theta_1, \ldots, \theta_n)^{h_1} \ldots R_{i-1}(\theta_1, \ldots, \theta_n)|^{h_{i-1}} > \varepsilon^{1/3}. \tag{1.11}$$

Перемножая (1.10) и (1.11), получаем:

$$|A| \cdot |R_1(\theta_1, \ldots, \theta_n)|^{h_1} \ldots |R_{i-1}(\theta_1, \ldots, \theta_n)|^{h_{i-1}} \times \tag{1.12}$$

$$\times |R_{i+1}(\theta_1, \ldots, \theta_n)|^{h_{i+1}} \ldots |R_k(\theta_1, \ldots, \theta_n)|^{h_k} > \varepsilon^{2/3}.$$

Это означает, согласно (1.4), что

$$|P(x_1, \ldots, x_n)| > \varepsilon^{2/3} |R_i(\theta_1, \ldots, \theta_n)|^{h_i}. \tag{1.13}$$

Последнее неравенство перепишем, используя (1.1):

$$\varepsilon^{1/3} > |R_i(\theta_1, \ldots, \theta_n)|^{h_i}. \tag{1.14}$$

Согласно (1.4), неравенство (1.14) доказывает (1.2) при $R(x_1, \ldots, x_n) = (R_i(x_1, \ldots, x_n))^{h_i}$. Лемма 1.3 доказана.

Лемму 1.3 можно переписать в таком виде:

Предложение 1.4. *Допустим, что существует полином* $P(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$ *такой, что для заданных* $(\theta_1, \ldots, \theta_n) \in C^n$ *имеет место неравенство*

$$|P(\theta_1, \ldots, \theta_n)| \leqq \varepsilon < 1,$$

и $t(P) \leqq T$. *Тогда либо существует полином* $R(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$, *являющийся степенью неприводимого над* Q *полинома* $S(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$ *с*

$$|R(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3}$$

вида $R(x_1, \ldots, x_n) = (R(x_1, \ldots, x_n))^s$, *причем* $t(S)s \leqq 2T$; $t(R) \leqq 2T$; *либо существуют два взаимно простых полинома* $P_1, (x_1, \ldots, x_n)$ *и* $P_2(x_1, \ldots, x_n)$ *из* $Z[x_1, \ldots, x_n]$, *для которых*

$$|P_1(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3}, \quad |P_2(\theta_1, \ldots, \theta_n)| \leqq \varepsilon^{1/3}$$

и $t(P_1), t(P_2) \leqq 2T$.

В таком виде предложение 1.4 следует из 1.3 и 1.2.

Изложенных предварительных сведений достаточно для начала исследования.

Рассматриваем для простоты самый интересный — плоский случай. Зафиксируем точку $\vec{\theta}=(\theta_1, \theta_2)\in C^2$ такую, что θ_1 и θ_2 алгебраически независимые над Q числа. Допустим, что существует множество алгебраических кривых, заданных уравнениями $P_H(x, y)=0$, проходящих вблизи $\vec{\theta}$ и удовлетворяющих следующим условиям:

(Ц. 1). Для всякого натурального числа H, $H\geqq c_5\geqq 1$ (c_5-какая-либо постоянная) существует полином $P_H(x, y)\in Z[x, y]$ такой, что

1) $P_H(x, y)\not\equiv 0$;

2) имеет место оценка типа $P_H(x, y)$:

$$t(P_H)\leqq c_6 H, \tag{1.15}$$

где c_6-некоторая постоянная;

3) для какой-то неограниченной неотрицательной монотонно возрастающей функции $\varphi(t): t\in R^+$ имеем оценки

$$|P_H(\theta_1, \theta_2)|<\exp\left(-H^{\mu}\varphi(H)\right) \tag{1.16}$$

для всех $H>c_7\geqq 1$.

Разумеется, в условиях (Ц. 1) основное внимание надо обратить на неравенства (1.16). Входящая в показатель степени в правой части (1.16) постоянная μ привлекает особое внимание. Она отвечает утверждениям типа теоремы 0.7; но в настоящей работе ее вид несущественен. Впрочем, для устранения тривиального случая — семейства Дирихле 0.1 — можно предположить, что $\mu\geqq 3$. Каким же образом можно преобразовать семейство кривых $\{P_H(x, y)=0: H\in N\}$ к более удобному виду. Идеальным явился бы тот случай, когда все кривые были бы неприводимыми и продолжали бы удовлетворять условиям (Ц. 1). Удовлетворить условиям 1)—2) в (Ц. 1) очень просто; с условием 3) дело значительно сложнее. Поскольку $P_H(x, y)$ может состоять из значительного количества неприводимых компонент: $P_H=R_1\ldots R_k$, то значения R_i в точке $\vec{\theta}$ могут быть значительно больше, чем значения P_H в точке $\vec{\theta}$. В соответствии с этим и вводим деление чисел H (т. е. соответствующих полиномов P_H) на «красные» и «синие».

К «синим» условно отнесем пока те H, для которые существует $P_H(x, y)\in Z[x, y]$, удовлетворяющий условиям 1)—3) из (Ц. 1), с, быть может, другими постоянными.

К «красным» естественно отнести все остальные H. Естественно возникающая задача — преобразовать полиномы, отвечающие «красным», в полиномы более интересные, чем просто элементы $Z[x, y]$.

Оказывается, и это составляет основной результат работы, что *для каждого «красного» H можно выбрать такой полином $P_H(x, y)\in Z[x, y]$ удовлетворяющий условиям* 1)—3) (Ц. 1), *что $P_H(x, y)$ взаимно прост с $P_{H-1}(x, y)$ и с $P_{H+1}(x, y)$.*

Соответствующее точное утверждение будет доказано ниже.

Для доказательства всех необходимых утверждений для построения таких «красных» полиномов, нужно простое вспомогательное утверждение. Мы же приведем одно общее утверждение, которое будет использоваться и в последующих работах.

Лемма 1.5. *Допустим, что* $\{P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)$*-некоторое семейство полиномов из* $C[x_1, \ldots, x_n]$*, не имеющих неконстантного общего делителя. Обозначим через* $R(x_1, \ldots, x_n)$, $S(x_1, \ldots, x_n)$ *два произвольные полинома из* $C[x_1, \ldots, x_n]$ *максимум степеней которых не более* d*. Тогда существуют такие целые рациональные числа* $A_1, \ldots, A_m$*, не все равные нулю, для которых*

$$|A_i| \leqq \left((m-1)d+1\right)^{m-1}: 1 \leqq i \leqq m \tag{1.17}$$

и полином

$$A_1 P_1(x_1, \ldots, x_n) + \ldots + A_m P_m(x_1, \ldots, x_n),$$

взаимно прост как с $R(x_1, \ldots, x_n)$*, так и с* $S(x_1, \ldots, x_n)$:

Доказательство. Под взаимной простотой двух полиномов из $C[x_1, \ldots, x_n]$ понимаем, разумеется, отсутствие у них общего неконстантного делителя.

Обозначим через $h_1(x_1, \ldots, x_n), \ldots, h_s(x_1, \ldots, x_n)$ и $q_1(x_1, \ldots, x_n), \ldots, q_t(x_1, \ldots, x_n)$-все различные негриводимые над C делители, соответственно, $R(x_1, \ldots, x_n)$ и $S(x_1, \ldots, x_n)$. Тогда, по определению степени, $s \leqq d(R)$ и $t \leqq d(S)$, т. е.

$$s \leqq d, \quad t \leqq d. \tag{1.18}$$

Для всякого целого рационального a, $a \neq 0$, обозначим через $P^a(x_1, \ldots, x_n)$ полином

$$P^a(x_1, \ldots, x_n) = P_1(x_1, \ldots, x_n) + aP_2(x_1, \ldots, x_n) + \ldots + a^{m-1} P_m(x_1, \ldots, x_n). \tag{1.19}$$

Покажем, что для какого-то a, $a \neq 0$, $|a| \leqq (m-1)d+1$, полином $P^a(x_1, \ldots, x_n)$ взаимно прост как с $R(x_1, \ldots, x_n)$ так и с $S(x_1, \ldots, x_n)$. Допустим, что ни для какого a, $a \neq 0$, $|a| \leqq (m-1)d+1$ это не имеет места. Тогда, по определению h_i и q_j для всякого a, $a \neq 0$, $|a| \leqq (m-1)d+1$ полином $P^a(x_1, \ldots, x_n)$ делится в $C[x_1, \ldots, x_n]$ на какой-то из полиномов $h_i(x_1, \ldots, x_n)$ или $q_j(x_1, \ldots, x_n)$: $1 \leqq i \leqq s$, $1 \leqq j \leqq t$.

Определяем отображение F множества $\{-(m-1)d-1, \ldots, -1, 1, \ldots, (m-1)d+1\}$ в $\{1, \ldots, s, s+1, \ldots, s+t\}$ так: при $|a| \leqq (m-1)d+1$, $a \neq 0$ полагаем

$$F(a) = \begin{cases} i, & \text{если } P^a \text{ делится на } h_i; \\ s+j, & \text{если } P^a \text{ делится на } q_j. \end{cases} \tag{1.20}$$

Поскольку выполнено (1.18), то $s+t \leqq 2d$ в то время как мощность множества $\{-(m-1)d-1, \ldots, -1, 1, \ldots, (m-1)d+1\}$ равна $2(m-1)d+1$. Следовательно, какая-то точка из $\{1, \ldots, s+t\}$ имеет более, чем $m-1$ преобразов. Это означает, что для m различных чисел $a_i: i=1, \ldots, m$, $a_i \neq 0$ и $|a_i| \leqq \leqq (m-1)d+1: i=1, \ldots, m$ имеем $F(a_1) = \ldots = F(a_m)$.

Согласно (1.20) это значит, что все полиномы $P^{a_i}(x_1, \ldots, x_n): i=1, \ldots, m$ делятся на один и тот же неприводимый над C полином $r(x_1, \ldots, x_n)$ (равный какому-то из $h_{i_0}(x_1, \ldots, x_n)$ или $q_{j_0}(x_1, \ldots, x_n)$).

Учитывая определение (1.19), запишем полученную информацию в виде серии равенств:

$$P_1(x_1, \ldots, x_n)+\ldots+a_1^{m-1}P_m(x_1, \ldots, x_n) = r_1(x_1, \ldots, x_n)\cdot r(x_1, \ldots, x_n), \tag{1.21}$$

$$P_1(x_1, \ldots, x_n)+\ldots+a_m^{m-1}P_m(x_1, \ldots, x_n) = r_m(x_1, \ldots, x_n)\cdot r(x_1, \ldots, x_n),$$

где $r(x_1, \ldots, x_n), \ldots, r_m(x_1, \ldots, x_n)\in C[x_1, \ldots, x_n]$.

Рассмотрим систему (1.21) как систему уравнений в $C[x_1, \ldots, x_n)$ относительно неизвестных $P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)$.

Определитель этой системы — это определитель Вандермонда:

$$\begin{vmatrix} 1 & a_1 & a_1^2 \ldots a_1^{m-1} \\ 1 & a_2 & a_2^2 \ldots a_2^{m-1} \\ \ldots\ldots\ldots\ldots\ldots \\ 1 & a_m & a_m^2 \ldots a_m^{m-1} \end{vmatrix} \neq 0,$$

т. к. все a_i — различны. Поэтому, решая (1.21) по правилу Крамера, получаем непосредственно серию равенств

$$\begin{cases} P_1(x_1, \ldots, x_n) = S_1(x_1, \ldots, x_n)\cdot r(x_1, \ldots, x_n), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ P_m(x_1, \ldots, x_n) = S_m(x_1, \ldots, x_n)\cdot r(x_1, \ldots, x_n), \end{cases} \tag{1.22}$$

где $S_i(x_1, \ldots, x_m)\in C[x_1, \ldots, x_m]: i=1, \ldots, m$. Но это означает, что семейство полиномов $\{P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)\}$ имеет общий делитель $r(x_1, \ldots, x_n)$. Это противоречит предположению леммы.

Следовательно, для какого-то целого рационального a, $a\neq 0$, $|a|\leqq$ $\leqq(m-1)d+1$ полином $P^a(x_1, \ldots, x_n)$ из (1.19) взаимно прост как $R(x_1, \ldots, x_n)$, так и с $S(x_1, \ldots, x_n)$. Поскольку

$$\max\{1, |a|, \ldots, |a|^{m-1}\} \leqq ((m-1)d+1)^{m-1},$$

то лемма доказана при $A_i=a^{i-1}: i=1, \ldots, m$.

Замечание 1.6. На самом деле оценка $|A_i|\leqq((m-1)d+1)^{m-1}$ является чрезмерно завышенной, она удовлетворительна только в используемом ниже случае $m=2$. На самом деле оценка $\max(|A_i|: 1\leqq i\leqq m)$ может быть выбрана в виде $O(d^{1/m})$; соответствующий результат нетрудно получить.

Приведем следствие из 1.5 в интересующем нас случае $m=2$:

Следствие 1.7. *Пусть $P_1(x_1, \ldots, x_n)$, $P_2(x_1, \ldots, x_n)$ — два взаимно простых полинома из $C[x_1, \ldots, x_n]$. Если $R(x_1, \ldots, x_n)$, $S(x_1, \ldots, x_n)$ — произвольные полиномы, то существуют целые рациональные числа a, b такие, что $|a|+$ $+|b|\neq 0$, $\max(|a|, |b|)\leqq \max(d(R), d(S))$ и полином*

$$aP_1(x_1, \ldots, x_n)+bP_2(x_1, \ldots, x_n)$$

взаимно прост как с $R(x_1, \ldots, x_n)$, так и с $S(x_1, \ldots, x_n)$.

Теперь произведем предварительную раскраску чисел H, $H \geqq c_5$.

Определение 1.8. Натуральное число $H, H \geqq c_5$ называем синим ($H \in B$), если у полинома $P_H(x, y)$ из (Ц. 1) существует делитель $P'_H(x, y)$, являющийся степенью неприводимого полинома над Q, и такой, что вместо (1.16) выполнено неравенство

$$|P'_H(x, y)| < \exp\left(-1/3H\varphi(H^{\mu})\right). \tag{1.23}$$

В противном случае H называем красным ($H \in R$). Получаем разбиение $N \setminus \{1, \ldots, [c_5]\} = R \cup B$ на два непересекающихся подмножества.

В соответствии с этим определением в следующем параграфе будет доказана теорема о представлении плотной последовательности (Ц. 1) в виде цветной. Как уже отмечалось, необходимость в такого рода утверждениях будет объяснена в § 3.

§ 2. Теорема о существовании цветных последовательностей

В настоящем параграфе докажем основную теорему работы.

Теорема 2.1. *Допустим, что существует семейство полиномов* $\{P_H(x, y): H \geqq c_5\}$, *удовлетворяющее условиям* 1)—3) (Ц. 1). *Тогда существует семейство полиномов* $\{R_H(x, y): H \geqq c_8\}$ *из* $Z[x, y]$, *удовлетворяющих следующим условиям:*

(Ц. 2). *Для всякого натурального* $H \geqq c_8$ $R_H(x, y)$ — *полином из* $Z[x, y]$, *причем выполнены условия:*

1) $R_H(x, y) \not\equiv 0$;

2) *имеется оценка типа* $R_H(x, y)$:

$$t(R_H) \leqq 4c_6 H = c_9 H; \tag{2.1}$$

3) *для неограниченной неотрицательной функции* $\varphi(t): t \in R^+$ *из* (Ц. 1) *имеем оценки*

$$|R_H(\theta_1, \theta_2)| < \exp\left(-1/4H^{\mu}\varphi(H)\right) \tag{2.2}$$

для всех $H \geqq c_{10}$;

4) *для любого* $H \geqq c_8 + 1$, *либо* $R_H(x, y)$ *является степенью неприводимого над* Q *полинома, либо* $R_H(x, y)$ *является взаимно простим как с* $R_{H-1}(x, y)$, *так и с* $R_{H+1}(x, y)$. *В первом случае* H *называется «синим», а во втором «красным».*

В частности, для семейства полиномов $\{R_H(x, y): H \geqq c_8\}$ из $Z[x, y]$, удовлетворяющего 1)—3) (Ц. 2) выполнены также условия 1)—3) (Ц. 1), если заменить c_5 на c_8, c_6 на c_9, c_7 на c_{10} иф ункцию $\varphi(t)$ на $1/4\varphi(t)$.

Доказательство теоремы 2.1. Допустим, что существует семейство полиномов $\{P_H(x, y): H \geqq c_5\}$, удовлетворяющее условиям (Ц. 1) 1)—3).

Учтем определение 1.8. Покажем, что «синие» и «красные» натуральные числа в смысле определения 1.8 являются, соответственно, «синими» и «красными» в смысле 4) (Ц. 2).

Начнем с определения постоянных c_8, c_9, c_{10}. Далее, т. к. $\varphi(t): t \in R^+$ неограниченная неотрицательная монотонно возрастающая функция, то существует такое $H_0 \geqq c_5$, что при $H \geqq H_0$ имсем $\varphi(H) \geqq 48(c_6+1)$ и $H \cdot c_6 \geqq$ $\geqq 48(1+c_6)$. Тогда полагаем $c_8 = c_{10} = H_0$.

Теперь уже можно определять полиномы $R_H(x, y) \in Z[x, y]$ при $H \geqq c_8$. Определение $R_H(x, y)$, наиболее просто в случае «синего» H в смнсле определения 1.8. Действительно, допустим, что H — «синее», т. е. $H \in B$. Согласно 1.8 это означает, что существует полином $S(x, y) \in Z[x, y]$, являющийся делителем $P_H(x, y)$ и степенью неприводимого над Q полинома, такой, что выполняется оценка

$$|S(\theta_1, \theta_2)| < \exp\left(-1/3H^{\mu}\varphi(H)\right). \tag{2.3}$$

В этом случае полагаем $S(x, y) = R_H(x, y)$, т. е. определяем $R_H(x, y)$ равенством

$$R_H(x, y) \equiv S(x, y), \quad \text{если} \quad H \in B. \tag{2.4}$$

Понятно, что условие 1) выполнено; 4) также выполнено, поскольку $S(x, y)$ является степенью неприводимого над Q полинома; 3) выполняется согласно (2.3). Осталось проверить 2). Но согласно (Ц. 1) 2) имеем оценку $t(P_H) \leqq c_6 H$, а $S(x, y)$ является делителем $P_H(x, y)$. Значит, слествие 1.2 дает оценку $t(R_H) = t(S) \leqq 2c_6 H < 4c_6 H = c_9 H$. Поэтому для полинома $R_H(x, y)$, определенного равенством (2.4) выполняются все условия 1)—4) (Ц. 2). Поэтому случай «синего» H в смысле определения 1.8, $H \in B$, рассмотрен.

Обратимся к «красному» H, $H \in R$ в смысле 1.8. Это означает, что $H \notin B$, т. е. ни для какого делителя $P_H(x, y)$, являющегося степенью неприводимого полинома из $Z[x, y]$, не имеет место оценка (2.3). Применяем лемму 1.3 в случае $n=2$, $P(x, y) \equiv P_H(x, y)$ и $\varepsilon = \exp\left(-H^{\mu}\varphi(H)\right)$. При этом $\varepsilon < 1$, т. к. $H \geqq 1$ и $\varphi(H) > 0$. Применяя лемму 1.3 и учитывая тот факт, что $H \notin B$ заключаем, что *существуют два взаимно простых делителя* $P_H(x, y)$ — *полиномы* $P_1^H(x, y)$ *и* $P_2^H(x, y)$ *из* $Z[x, y]$ *такие, что*

$$|P_1^H(\theta_1, \theta_2)| \leqq \exp\left(-1/3H^{\mu}\varphi(H)\right); \quad |P_2^H(\theta_1, \theta_2)| \leqq \exp\left(-1/3H^{\mu}\varphi(H)\right). \tag{2.5}$$

Оставшуюся часть построения полиномов $\{R_H(x, y): H \geqq c_8\}$ следует провести индукцией по натуральным $H \geqq c_8$.

Во-первых, для всех «синих» H, $H \in B$, определяем, как и в (2.4), $R_H(x, y) \in$ $\in Z[x, y]$ так, что все условия (Ц. 2] выполнены.

Допустим, что для всех «красных» H', т. е. $H' \notin B$, и таких, что $c_8 \leqq H' < H$ построены полиномы $R_{H'}(x, y) \in Z[x, y]$, удовлетворяющие условиям 1)—4) (Ц. 2].

Покажем, что при $H \geqq c_8$ существует и полином $R_H(x, y) \in Z[x, y]$ также удовлетворяющий условиям 1)—4) (Ц. 2).

Допустим сначала, что $H \in B$. Тогда полином $R_H(x, y)$, удовлетворяющий (Ц. 2), действительно определен выше. Поэтому рассматриваем тот случай, когда H — «красный», $H \in R$, $H \notin B$ и $H \geqq c_8$. Воспользуемся в дальнейшем

доказательстве следствием 1.7. Построение полинома $R_H(x, y)$ должно вестись таким образом, чтобы учитывались уже существующие «синие» полиномы. Поэтому различаем два случая.

Первый случай. Число $H+1$ — «красное», т. е. полином $R_{H+1}(x, y)$ пока не построен.

Рассмотрим два взаимно простых полинома $P_1^H(x, y)$, $P_2^H(x, y) \in Z[x, y]$, являющиеся делителями $P_H(x, y)$ и удовлетворяющие неравенствам (2.5). Сразу же заметим, что согласно лемме 1.1 или следствию 1.2 мы имеем следующую оценку типов: $P_1^H(x, y)$ и $P_2^H(x, y)$:

$$\max\left(t(P_1^H), t(P_2^H)\right) \leqq t_1(P_1^H)+t(P_2^H) \leqq 2t(P_H) \leqq 2c_6 H. \tag{2.6}$$

Применяя следствие 1.7, получаем два (не нулевых вдвоем) целых рациональных числа a, b; таки, что полином

$$aP_1^H(x, y)+bP_2^H(x, y) \tag{2.7}$$

взаимно прост с $R_{H-1}(x, y)$, причем

$$\max(|a|, |b|) \leqq d+1, \tag{2.8}$$

где $d=d(R_{H-1})$. Отметим, что в случае $H=[c_8+1]$, вместо $R_{H-1}(x, y)$ можно взять единичный (константный) полином. Теперь уже ясно, что в качестве полинома $R_H(x, y)$ можно выбрать полином 2.7):

$$R_H(x, y) \equiv aP_2^H(x, y)+bP_2^H(x, y). \tag{2.9}$$

Поскольку $|a|+|b| \neq 0$, а $P_1^H(x, y)$ и $P_2^H(x, y)$ — взаимно просты, то $R_H(x, y) \neq 0$, т. е. условие 1) (Ц. 2) выполнено. Оценка типа $R_H(x, y)$ очень проста:

$$\begin{aligned} t(R_H) &\leqq t(P_1^H)+t(P_2^H)+2\ln\left(\max(|a|, |b|)\right) \leqq \\ &\leqq t(P_1^H)+t(P_2^H)+2\ln\left(d(R_{H-1})+1\right) \leqq \\ &\leqq t(P_1^H)+t(P_2^H)+2\ln\left(4c_6(H-1)+1\right). \end{aligned}$$

Поскольку $Hc_6 \geqq 48$ при $H \geqq H_0$ (см. (1.15) и 2) (Ц. 1)), то $\ln\left(4c_6(H-1)+1\right) \leqq$ $\leqq c_6 H$. Отсюда

$$t(P_1^H)+t(P_2^H)+2c_6 H \geqq t(R_H). \tag{2.10}$$

Если учесть (2.6), то (2.10) дает требуемую в 2) (Ц. 2) оценку:

$$t(R_H) \leqq 4c_6 H = c_9 H. \tag{2.11}$$

Условие 3) также имеет место. В самом деле,

$$\begin{aligned} |R_H(\theta_1, \theta_2)| &\leqq \max(|a|, |b|) \cdot \max\left(|P_1^H(\theta_1, \theta_2)|, |P_2^H(\theta_1, \theta_2)|\right) \leqq \\ &\leqq \left(d(R_{H-1})+1\right) \cdot \max\left(|P_1^H(\theta_1, \theta_2)|, |P_2^H(\theta_1, \theta_2)|\right). \end{aligned}$$

Воспользуемся неравенствами (2.5), получаем:

$$|R_H(\theta_1, \theta_2)| \leqq \exp\left(\ln\left(d(R_{H-1})+1\right)-1/3H^n\varphi(H)\right). \tag{2.12}$$

Но при $H \geqq c_{10}$, $\varphi(H) \geqq 48\,(1+c_6)$ по определению c_{10}, а т. к. и $\mu \geqq 3$, то при $H \geqq c_{10}$,

$$1/2H^{\mu}\varphi(H) \geqq 48/12H^{\mu}c_6 > 4c_6H > \ln(4c_6H) \geqq \ln\big(d(R_{H-1})+1\big).$$

Отсюда и из (2.12) вытекает сразу же

$$|R_H(\theta_1, \theta_2)| \leqq \exp\big(-1/4H^{\mu}\varphi(H)\big). \tag{2.13}$$

Значит, 3) (Ц. 2) также имеет место. Наконец, условие 4) (Ц. 2) выполняется, т. к. $R_H(x, y)$ и $R_{H-1}(x, y)$ — взаимно просты, а $H+1$ — не «синее». Так как $H+1$ — красное, то и $R_{H+1}(x, y)$ будет взаимно просто с $R_H(x, y)$ по построению. Этим первый случай рассмотрен.

Второй случай. Допустим, что $H+1$ — «синее», т. е. полином $R_{H+1}(x, y) \in Z[x, y]$ удовлетворяющий 1)—4) (Ц. 2) уже существует. Доказательство в этом случае ничем не отличается от предыдущего. Снова применяем следствие 1.7 к паре взаимно простых полиномов $P_1^H(x, y)$, $P_2^H(x, y) \in Z[x, y]$, удовлетворяющих условиям (2.5) и (2.6).

Получаем при $d = \max\big(d(R_{H-1}), d(R_{H+1})\big)$ два целых рациональных числа a, b не равных вместе нулю, таких, что полином

$$aP_1^H(x, y) + bP_2^H(x, y) \tag{2.14}$$

взаимно прост как с $R_{H-1}(x, y)$ так и с $R_{H+1}(x, y)$, причем

$$\max(|a|, |b|) \leqq d+1 = \max\big(d(R_{H-1}), d(R_{H+1})\big)+1. \tag{2.15}$$

Полином $R_H(x, y)$ полагаем равным полиному из (2.14):

$$R_H(x, y) \equiv aP_1^H(x, y) + bP_2^H(x, y). \tag{2.16}$$

Тогда $R_H(x, y) \not\equiv 0$, т. к. $P_1^H(x, y)$ и $P_2^H(x, y)$ — взаимно просты. Оценка тила $R_H(x, y)$ вида

$$t(R_H) \leqq t(P_1^H) + t(P_2^H) + 2\max\big(\ln(d(R_{H-1})+1,\ \ln\big(d(R_{H+1})+1\big)\big) \leqq t(P_1^H) + t(P_2^H) +$$

$$+ 2\max\big(\ln(4c_6(H-1)+1), \ln(2c_6(H+1)+1)\big).$$

Однако, как мы уже знаем, $\ln\big(4c_6(H-1)+1\big) \leqq c_6H$, а $\ln\big(2c_6(H+1)+1\big) \leqq c_6H$ при любом $H > 1$, т. к. $\ln(3n+1) \leqq n$ при всех $n \geqq 2$. Поэтому

$$t(R_H) \leqq t(P_1^H) + t(P_2^H) + 2c_6H. \tag{2.17}$$

Используя 2.6), получаем из (2.17)

$$t(R_H) \leqq 4c_6H, \tag{2.18}$$

т. е. условие 2) (Ц. 2) доказано. Аналогично устанавливаем 3) (Ц. 2):

$$|R_H(\theta_1, \theta_2)| \leqq \max(|a|, |b|) \cdot \max\big(|P_1^H(\theta_1, \theta_2)|,\ |P_2^H(\theta_1, \theta_2)|\big).$$

Применяя (2.5) и (2.15) выводим отсюда

$$|R_H(\theta_1, \theta_2)| \leqq \exp\big(\max(\ln(d(R_{H-1})+1), \ln(d(R_{H+1})+1)) - 1/3H^{\mu}\varphi(H)\big). \tag{2.19}$$

Поскольку при $H \geqq c_{10}$, $\varphi(H) \geqq 48c_6$, то при $H \geqq c_{10}$

$$1/12H^{\mu}\varphi(H) \geqq 4c_6 H^{\mu} \geqq 4c\ H >$$

$$> \ln\big(4c_6(H-1)+1\big)+\ln\big(2c_6(H+1)+1\big) \geqq \ln\big(d(R_{H-1})+1\big)+\ln\big(d(R_{H+1})+1\big).$$

Поэтому (2.19) означает, что при $H \geqq c_{10}$

$$|R_H(\theta_1, \theta_2)| < \exp\big(-1/4H^{\mu}\varphi(H)\big) \tag{2.20}$$

и 3) (Ц. 2) имеет место. Наконец, $R_H(x, y)$ по построению взаимно прост как с $R_{H-1}(x,y)$, так и с $R_{H+1}(x, y)$. Все условия (Ц. 2) выполнены.

Таким образом, для всех «красных» $H \geqq c_8$ построены полиномы $R_H(x, y) \in Z[x, y]$, удовлетворяющие условиям 1)—4) (Ц. 2.) Теорема 2.1 доказана.

Теорему 2.1 для некоторых технических целей удобно представить в несколько другом виде, в котором разделение полиномов на «синие» и «красные», проявляющееся в пункте 4) (Ц. 2), уже исчезает. Формулировка следующей теоремы кое в чем может напомнит работы Дэвенпорта и Шмидта [20], [21], т. к. по духу рассматриваемые здесь и в [20], [21] задачи сходны.

Интуитивно идею формулируемого ниже утверждения можно пояснить следующим образом. Допустим, что существует последовательность полиномов $\{R_H(x, y): H \geqq c_8\}$, удовлетворяющая условиям (Ц. 2). Постараемся выделить среди всех натуральных чисел $H \geqq c_8 \geqq 1$ такую бесконечную последовательность $X_n: n=1, 2, \ldots$, чтобы выполнялись следующие условия:

а) полиномы $R_{X_n}(x, y)$ и $R_{X_{n+1}}(x, y)$ взаимно просты для любых $n=1, 2, 3, \ldots$;

б) для всякого $n=1, 2, \ldots$, если $X_n+1 < X_{n+1}$, то $R_{X_n}(x, y)$ — является степенью неприводимого полинома из $Z[x, y]$, причем при $X_n < H < X_{n+1}$ всякий $R_H(x, y)$ является степенью того же самого полинома.

Реализацией этого соображения и служит устанавливаемая ниже

Теорема 2.2. *Допустим, что существует семейство полиномов* $\{P_H(x, y): H \geqq c_5\}$ *из* $Z[x, y]$, *удовлетворяющих условиям* 1)—3) (Ц. 1). *Тогда существует семейство полиномов* $\{R_H(x, y): H \geqq c_8\}$ *из* $Z[x, y]$, *удовлетворяющих следующим условиям:*

(Ц. 3). *Для всякого натурального* $H \geqq H_0 = [c_8+1]$ *выполняются все условия* 1)—4) *из* (Ц. 2). *Кроме того существует возрастающая бесконечная последовательность* $X_n: n < \infty$ *натуральных чисел такая, что*

$$H_0 = X_0 < X_1 < X_2 < \ldots < X_n < \ldots : n < \infty \tag{2.21}$$

и выполняются следующие условия:

i) *для любого* $n=0, 1, 2$ *полиномы* $R_{X_n}(x, y)$ *и* $R_{X_{n+1}}(x, y)$ — *взаимно просты;*

ii) *если* $X_n+1<X_{n+1}$, *то полином* $R_{X_n}(x, y)$ *является степенью неприводимого над* Q *полинома* $S_{X_n}(x, y)\in Z[x, y]$;

iii) *если* $X_n<H<X_{n+1}$, *то полином* $R_H(x, y)$ *также является степенью полинома* $S_{X_n}(x, y)$ *из* ii);

iv) *для всякого* $n=0, 1, 2, \ldots$ *имеют место неравенства, связывающие тип* $t(R_{X_n})$ *полинома* $R_{X_n}(x, y)$ *с* $|R_{X_n}(\theta_1, \theta_2)$:

$$t(R_{X_n}) \leqq c_9 X_n; \tag{2.22}$$

$$|R_{X_n}(\theta_1, \theta_2)| < \exp\left(-1/4X_n^{\mu}\varphi(X_n)\right); \tag{2.23}$$

v) *для всякого* $n=0, 1, 2, \ldots$ *существует такое натуральное* $s_n\geqq 1$, *что выполняются неравенства, связывающие* X_n *с* X_{n+1}:

$$s_n \cdot t(R_{X_n}) \leqq c_{11} X_{n+1}; \tag{2.24}$$

$$|R_{X_n}(\theta_1, \theta_2)| < \exp\left(-1/4(X_{n+1}-1)^{\mu}\varphi(X_{n+1}-1)/s_n\right), \tag{2.25}$$

где c_{11} — *некоторая постоянная,* $c_{11}=18c_9$.

Замечание 2.3. Интуитивно s_n можно представить себе как ту степень, в которую нужно возвести $R_{X_n}(x, y)$, чтобы получить $R_{X_{n+1}-1}(x, y)$ при $X_n\leqq \leqq X_{n+1}-1$ (см. iii)).

Доказательство теоремы 2.2. Допустим, что существует семейство $\{P_H(x, y): H\geqq c_5\}$, удовлетворяющее условиям 1)—3) (Ц. 1). Тогда согласно теореме 2.2 существует семейство полиномов $\{R_H(x, y): H\geqq c_8\}$, удовлетворяющее условиям 1)—4) (Ц. 2). Сохраняем постоянные c_8, c_9, c_{10}, введенные в 2.1 и объясненный в начале доказательства теоремы 2.1 их смысл.

Последовательность $X_n: n=0, 1, 2, \ldots$ натуральных чисел, для которой выполняются условия i)—v) (Ц. 3), строим по индукции. Во-первых, полагаем

$$X_0 = H_0 = [c_8+1]. \tag{2.26}$$

Тогда условия i)—iii), v) тривиальны, а iv) является переформулировкой условий 2) и 3) из (Ц. 2).

Теперь предположим, что построена конечная последовательность

$$H_0 = X_0 < X_1 < \ldots < X_m \tag{2.27}$$

натуральных чисел, $m\geqq 0$, такая, что для всех $n\leqq m$ выполняются все условия i)—v), неравенства (2.22)—(2.25). Построим теперь $X_{m+1}>X_m$, чтобы продлить последовательность (2.27). Покажем сначала, что существует такое натуральное $Y>X_m$, что выполнено условие

(2.28) полином $R_Y(x, y)$ взаимно прост с $R_{X_m}(x, y)$.

Действительно, предположим, что никакое $Y>X_m$ не удовлетворяет (2.28). Тогда применим условие 4) (Ц. 2). Получаем при $Y=X_m+1$, что $R_{X_{m+1}}(x, y)$ — является степенью неприводимого над Q полинома $T_{X_{m+1}}(x, y)\in Z[x, y]$ и что $R_{X_m}(x, y)$ также является степенью неприводимого над Q полинома

$T_{X_m}(x, y) \in Z[x, y]$. При этом, т. к. $R_{X_m}(x, y)$ и $R_{X_m+1}(x, y)$ не взаимно просты, то $T_{X_m}(x, y) \equiv T_{X_m+1}(x, y)$. Далее применяем 4) (Ц. 2) при $Y = X_m + 2$ и т. д., получаем, что все полиномы

$$R_{X_m}(x, y),\ R_{X_m+1}(x, y),\ \ldots,\ R_H(x, y)$$

при $H > X_m$ являются степенями одного и того же неприводимого полинома $T_{X_m}(x, y)$. Снова, используя отрицание (2.28) при $Y = H + 1$ и 4) (Ц. 2), заключаем, что и $R_{H+1}(x, y)$ является степенью $T_{X_m}(x, y)$. Оконвательно все полиномы

$$R_{X_m}(x, y),\ R_{X_m+1}(x, y),\ \ldots,\ R_H(x, y),\ \ldots$$

при всех $H > X_m$ являются степенями полинома $T_{X_m}(x, y) \equiv T_0(x, y) \in Z[x, y]$. Тогда имеем для $H > X_m$

$$R_H(x, y) \equiv (T_0(x, y))^{m_H} \tag{2.29}$$

при $m_m \geqq 1$. Используя следствие 1.2 и 2) (Ц. 2), выводим из (2.29), что при $H > X_m$,

$$m_H t(T_0) \leqq 2c_9 H. \tag{2.30}$$

Из (2.29) и 3) (Ц. 2) вытекает также, что при всех $H > X_m$

$$|T_0(\theta_1, \theta_2)| < \exp\left(-1/4H^{\mu}\varphi(H)/m_H\right). \tag{2.31}$$

Комбинируя (2.30) и (2.31), приходим при $H > X_m$ к неравенству:

$$|T_0(\theta_1, \theta_2)| < \exp\left(-1/4H^{\mu}\varphi(H)t(T_0)/2c_9 H\right). \tag{2.32}$$

Однако $\mu > 3$, а $\lim\limits_{H \to \infty} \varphi(H) = \infty$. Следовательно, переходя в (2.32) к пределу при $H \to \infty$, получаем

$$T_0(\theta_1, \theta_2) = 0,$$

а т. к. $T_0(x, y) \not\equiv 0$ по (2.29) и 1) (Ц. 2), то θ_1, θ_2 — алгебраически независимы над Q вопреки предположению (Ц. 1).

Значит, сделанное предположение о ложности (2.28) для всех $Y > X_m$ не имеет места, и какое-то $Y > X_m$, удовлетворяющее (2.28), существует. Обозначим через X_{m+1} наименьшее $Y = X_{m+1} > X_m$, удовлетворяющее (2.28).

Такое $X_{m+1} < \infty$, $X_m < X_{m+1}$ существует. Покажем, что опо является искомым.

Во-первых, по определению $X_{m+1} = Y$ удовлетворяет (2.28). Этим доказано свойство i).

Покажем ii) и iii). Пусть $X_{m+1} > X_m + 1$. Поскольку X_{m+1} — наименгшее $Y > X_m$, удовлетворяющее (2.28), то $Y = X_m + 1$ уже не удовлетворяет (2.28). Следовательно, по свойству 4) (Ц. 2) полином $R_{X_m}(x, y)$ является степенью неприводимого над Q полинома $S_{X_m}(x, y) \in Z[x, y]$. Поэтому ii) установлено. Допустим, однако, что iii) не имеет места. Тогда обозначим через H — такое наименьшее натуральное H, $X_m < H < X_{m+1}$, что полином $R_H(x, y)$ не является степенью $S_{X_m}(x, y)$. В этом случае полином $R_{H-1}(x, y)$ является степенью $S_{X_m}(x, y)$. Поскольку $Y = H < X_{m+1}$ не удовлетворяет (2.28), то полином $R_H(x, y)$ не является взаимно простым с $R_{X_m}(x, y)$ т. е. с $S_{X_m}(x, y)$. Поэтому

$R_H(x, y)$ имеет общий делитель и с $R_{H-1}(x, y)$. Согласно 4) (Ц. 2) это означат что $R_H(x, y)$ является степенью неприводимого полинома $S'(x, y)$. Но $R_H(x, y)$ имеет общий делитель со степенью $S_{X_m}(x, y)$, а это значит, что $S'(x, y) \equiv \equiv S_{X_m}(x, y)$. Свойство iii) также доказано.

Свойство iv), как уже отмечалось, является нереформулировкой свойств 2) и 3) из (Ц. 2), а неравенства (2.22) и (2.23), эквивалентны, соответственно, (2.1) и (2.2).

Осталось доказать v). Во первых, отметим, что при $X_{m+1}=X_m+1$ неравенства (2.24) и (2.25) следуют при $s_m=1$ из (2.22) и (2.23) для любого $c_{11} \geqq c_9$. Поэтому рассматриваем случай $X_{m+1}>X_m+1$; см. ii) и iii). Мы имеем согласно ii), что

$$R_{X_m}(x, y) = (S_{X_m}(x, y))^q \tag{2.33}$$

для некоторого натурального $q \geqq 1$. Далее, применяя iii) при $H=X_{m+1}-1$, мы получаем

$$R_H(x, y) = (S_{X_m}(x, y))^p \tag{2.34}$$

для $H=X_{m+1}-1>X_m$, где p — натуральное число, $p \geqq 1$. Теперь обозначаем через $s_m \geqq 1$ — наименьшее натуральное число, большее p/q. Иначе говоря,

$$p/q \leqq s_m, \quad s_m - p/q < 1. \tag{2.35}$$

Покажем, что при таком выборе s_m выполняются (2.24)—(2.25) при $n=m$. В самом деле, из (2.33) и леммы 1.1 непосредственно получаем

$$1/3qt(S_{X_m}) \leqq t(R_{X_m}) \leqq 3qt(S_{X_m}). \tag{2.36}$$

Аналогично, из (2.34) и 1.1—1.2 следует

$$1/3pt(S_{X_m}) \leqq t(R_H) \tag{2.37}$$

при $H=X_{m+1}-1$.

Из (2.36) следует

$$s_m t(R_{X_m}) \leqq 3s_m qt(S_{X_m}). \tag{2.38}$$

Привлекая (2.35), получаем из (2.38)

$$s_m t(R_{X_m}) \leqq 3(p/q+1)qt(S_{X_m})$$

при $p \geqq q$, или

$$s_m t(R_{X_m}) \leqq 3qt(S_{X_m})$$

при $p<q$.

Следовательно, в любом случае, имеем

$$s_m t(R_{X_m}) \leqq \max\{6p, 3q\}\, t(S_{X_m}). \tag{2.39}$$

Учитывая (2.36) и iv), получаем

$$3qt(S_{X_m}) \leqq 9t(R_{X_m}) \leqq 9c_9 X_m < 9c_9 X_{m+1}; \tag{2.40}$$

а учитывая (2.37) и 2) (Ц. 2), получаем

$$6pt(S_{X_m}) \leqq 18t(R_H) \leqq 18c_9(X_{m+1}-1). \tag{2.41}$$

Поэтому, полагая $c_{11}=18c_9 \geqq c_9, c_6$ (т. к. $c_9=4c_6>c_6$), полчуаем, объединяя (2.39)—(2.41), что

$$s_m t(R_{X_m}) \leqq c_{11} X_m. \tag{2.42}$$

Теперь (2.24) эквивалентно (2.24). Покажем (2.25).

Учитывая (2.33) и (2.34), получаем

$$|R_{X_m}(\theta_1, \theta_2)| = |R_H(x, y)|^{q/p} \tag{2.43}$$

при $H=X_{m+1}-1$. Используя 3) (Ц. 2), получаем из (2.43):

$$|R_{X_m}(\theta_1, \theta_2)| < \exp\left(-q/4p \cdot (X_{m+1}-1)^{\mu} \varphi(X_{m+1}-1)\right). \tag{2.44}$$

Но по (2.35),

$$s_m \geqq p/q, \quad \text{т.е.} \quad q/p \geqq 1/s_m \quad \text{или же} \quad -q/4p \leqq -1/4s_m.$$

Значит, (2.44) дает

$$|R_{X_m}(\theta_1, \theta_2)| < \exp\left(-\frac{1}{4s_m}(X_{m+1}-1)^{\mu} \varphi(X_{m+1}-1)\right). \tag{2.45}$$

Поэтому v) доказано. Следовательно, X_{m+1}, построенное выше, удовлетворяет условиям i)—v) и последовательность чисел (2.27), удовлетворяющих (Ц. 3), неограниченно продолжается. Этим доказана теорема 2.2.

Именно в таком преобразованном виде, будут использоваться последовательности полиномов, которые мы назвали цветными.

§ 3. Некоторые объяснения

В этом параграфе поясним, зачем нужна теорема 2.1 или 2.2, описывающая свойства плотных последовательностей, удовлетворяющих условиям (Ц). В первую очередь плотные последовательности, удовлетворяющие условию (Ц. 1) возникают при исследованиях по гипотезе Шануэлла в стиле [3], [5], [6], [7].

В самом деле, например, для множества Шануэлла $\{\beta_j, e^{\alpha_i \beta_j}\}$, где $\alpha_1, \ldots, \alpha_N$ и $\beta_1, \ldots, \beta_M$ — два набора линейно независимых чисел, возникает последовательности типа (Ц. 1), только для полиномов от n переменных, где $n=\deg Q(\beta_j, e^{\alpha_i \beta_j})$. В качестве параметра μ здесь появляется известный (см. [6], [7]) инвариант, связывающий M и N:

$$\mu = \varkappa_2 = \frac{M(N+1)}{M+N}.$$

Именно последовательность $\mu=\varkappa_2$, удовлетворяющая (Ц. 1) вместе с теоремой 0,7 была применена для построения первых примеров чисел вида α^{β}, среди которых есть тир алгебраически независимых. По поводу результатов этого типа см Д. Броунвейль [3] и [14].

Затем эти результаты были усилены в работе автора [6], где вместо условия (Ц. 1) использовались значительно более тонкие условия [12]. Эти условия также приводят к последовательностям цветных полиномов и будут

изучаться в общем случае позднее. Необходимо только отметить, что появление этих тонких условий связано с дополнительной оценкой меры линейной независимости чисел α_i и β_j, в то время как результаты [3], опирающиеся на 0.7, этого не предполагают. Поэтому устранение предположения о мере линейной независимости α_i и β_j связано с дополнительными исследованиями.

Однако теорема 0.7 оказалась полезной при изучении алгебраической независимости $4-x$ чисел. Опираясь на метод автора [6], М. Вальдшмидт [15] дал пример множества чисел вида α^β, содержащего 4 алгебраически независимых числа. Хотя этот пример и хуже результатов [7], [8], но представляет интерес относительная простота рассуждений.

Поэтому естественно ожидать, что применение теоремы типа 0.7 в сочетании с более сильными алгебраическими соображениями приведет к установлению, в частности, новых результатов об алгебраической независимости чисел вида α^β. Именно так и будет, как мы увидим в последующих работах.

Существует и еще одно соображение в пользу изучения цветных последовательностей. Кроме гипотезы Шануэлла существуют и другие задачи, относящиеся к алгебраической независимости значений мероморфных функций. В частности, есть проблема, аналогичная гипотезе Шануэлла, для случая значений эллиптических функций и абелевых интегралов. Однако, в отличие от экспоненты, в этом случае уже нельзя рассчитывать на выполнимость условий, более сильных чем (Ц. 1). Только последовательности типа (Ц. 1) существуют после аналитической части соответствующих доказательств.

Поэтому исследования последовательностей, удовлетворяющих условиям (Ц. 1)—(Ц. 3), а зашем и последовательностей, удовлетворяющих более сложным условиям, может дать дополнительную информацию для изучения эллиптических функций. Информация эта, как и в случае экспоненты, будет состоять в том, что какое-то поле, порожденное заданными постоянными, имеет нетривиальную нижнюю границу для типа τ своей трансцендентности. Если к тому жн из других соображений известна небольшая оценка сверху этого типа τ, то мы приходит к новой информации о степени трансцендентности рассматриваемых полей.

Итак, мы видим, что исследование последовательностей, названных условно «цветными», важно для многих задач теории трансцендентных чисел.

ЛИТЕРАТУРА

[1] Касселс, Дж. В. С.: *Введение в теорию диофантовых приближений*, ИЛ (Москва, 1961).

[2] Ax, I.: On Schanuel's conjecture, *Ann of. Math.* **93** (1971), 252—258.

[3] Brownawell, W. D.: Gelfond's method for algebraic independence, *Trans. Amer. Math. Soc.*, **210** (1975).

[4] Lang, S.: *Diophantine geometry*, Addison-Wesley, 1962.

[5] Гельфонд, А. О.: *Трансцендентные и алгебраические числа*, Гостехиздат, (Москва, 1952).

[6] Чудновский, Г. В.: Алгебраическая независимость нескольких значений показательной функции, *Матем. заметки*, **15** (4), (1974).

[7] Чудновский, Г. В.: *Некоторые аналитические методы в теории трансцендентных чисел*, Препринт ИМ—74—8 (Киев, 1974).

[8] Чудновский, Г. В.: *Аналитические методы в диофактовых приближениях*, Препринт ИМ—74—9 (Киев, 1974).

[9] Cijsouw, P.: *Transcendent measures*, Academic Service, 1972.

[10] LANG, S.: *Introduction to transcendental numbers*, Addison-Wesley, 1966.
[11] LANG, S.: Transcendental numbers and diophantine approximations, *Bull. Amer. Math. Soc.* **77** (1971).
[12] TIJDEMAN, R.: An auxiliary result in the theory of transcendental numbers, *J. Number Theory* **5** (1973).
[13] TIJDEMAN, R.: On the Gelfond—Baker method and its applications, *Bull. Amer. Math. Soc.* (to appear).
[14] WALDSCHMIDT, M.: Nombres transcendants. *Lecture Notes in Math.*, v. 402, Springer-Verlag, 1974.
[15] WALDSCHMIDT, M.: *Independance algebrique par la methode de G. V. Čudnovskij*, NG 8 Sem. Delange—Pisot—Poitou, 1974—1975.
[16] BROWNAWELL, W. D., WALDSCHMIDT, M.: The algebraic independence of certain numbers to algebraic powers (to appear).
[17] BROWNAWELL, D.: Sequences of diophantine approximations, *J. Number Theory*, **6** (1974), 11—21.
[18] Фельдман, Н. И., Шидловский, А. Б.: Развитие и современное состояние теории трансцендентных чисел, *УМП*, **22** (3) (1967).
[19] SCHMIDT, W.: Approximation to algebraic numbers, *Monat. Math.* (1973).
[20] DAVENPORT, H. and SCHMIDT, W.: Approximation to real numbers by quadratic irrationals, *Acta Arith.* **13** (1967), 169—176.
[21] DAVENPORT, H. and SCHMIDT, W.: Approximation to real numbers by algebraic integers, *Acta Arith.* **15** (1969), 393—416.

Ул. Тарасовская 10, кв. 17, 252 033, Киев — 33, СССР

(Поступило 4 апреля 1976)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 145—157

НА ПУТИ К ГИПОТЕЗЕ ШАНУЭЛЛА. АЛГЕБРАИЧЕСКИЕ КРИВЫЕ ВБЛИЗИ ТОЧКИ

II. ПОЛЯ КОНЕЧНОГО ТИПА ТРАНСЦЕНДЕНТНОСТИ И ЦВЕТНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ. РЕЗУЛЬТАНТЫ

Г. В. ЧУДНОВСКИЙ

Эта работа является непосредственным продолжением предыдущей, на которую будем кратко ссылаться как на [1]. Определения из [1] также будут использоваться, часто без дополнительных объяснений.

В ходе настоящего исследования общая схема «цветных» последовательностей из [1] будет применяться для изучения полей, возникающих при добавлении к Q координат θ_i рассматриваемой точки $\vec{\theta} \in C^n$. Как и выше [1] рассматриваем плоский случай ($n=2$) и вместо отдельного учета степени и высоты рассматривается сразу тип полинома. Однако методы, применяемые в работе общие и допускают простой перенос на пространственный случай ($n>3$). Целям такого переноса, который будет осуществлен в других работах этого цикла, служат также леммы общего плана, приводившиеся в [1], а также леммы этой работы.

Существенной особенностью работы является использование исключения неизвестной с помощью результантов по Сильвестру. Этот метод намеренно выбран как основной в работе, т. к. он хорошо известен и давно уже применялся в исследованиях на эту тему. Поэтому демонстрация новых мощных следствий, полученных даже на основе старого и неточного метода, показывает возможности, получаемые на основе точного изучения свойств цветных последовательностей.

В последующих работах мы разовьем и несравненно улучшим конкретные оценки, выводимые ниже. Однако эта работа является ключевой для понимания сути и схемы всех рассуждений на ту же тему. Поэтому мы сознательно идем на которое ухудшение оценок.

§ 0. Некоторое обсуждение

В § 3 [1] уже обсуждались причины, вынуждающие заниматься свойствами «цветных» последовательностей. Там же указывалось, что существование цветной последовательности вблизи точки $\vec{\theta}$ влечет нетривиальные оценки снизу типа τ трансцендентности полей, порожденных координатами θ_i. Одна (и единственная до настоящего времени) оценка такого типа τ через показатель μ у цветной последовательности приводилась в теореме 0.7 [1]. В этой работе дадим принципиальное улучшение этого результата. Вначале напомним основное

Определение 0.1. Пусть $K=Q(\vartheta_1, \ldots, \vartheta_d, \omega)$ — подполе C степени трансцендентности d над Q, а $\vartheta_1, \ldots, \vartheta_d$ — его базис трансцендентности. Говорят, что поле K имеет тип трансцендентности $\leqq\tau$ $(<\infty)$, если существует такая постоянная $C(K)>0$, что для всякого полинома $P(x_1, \ldots, x_d)\in Z[x_1, \ldots, x_d]$, $P(x_1, \ldots, x_d)\not\equiv 0$ имеем

$$|P(\vartheta_1, \ldots, \vartheta_d)| > \exp\left(-C(K)t(P)^\tau\right). \tag{0.1}$$

Если в правой части (0.1) вместо $t(P)^\tau$ стоит $t^\tau(P)(\ln t(P))^{\tau'}$ то говорят, что K имеет тип $\leqq(\tau, \tau')$.

Именно в рамках этого определения ведется все дальнейшее исследование. Поскольку мы рассматриваем плоский случай, то исследования бугут относиться к случаю $d=1$, т. е. полей степени трансцендентности 1. Для таких полей при $\tau<\infty$, $\tau\geqq 2$ и «почти все» такие поля (но не все) имеют $\tau=2$.

Поля конечного типа трансцендентности возникают в связи с исследованием цветных последовательностей. Поэтому напомним фундаментальное определение цветной последовательности, данное в [1]. На это определение мы будем часто ссылаться в дальнейшем.

Прежде всего раз и навсегда зафиксируем точку $\vec{\theta}\in C^2$, $\vec{\theta}=(\theta_1, \theta_2)$, такую, что θ_1 и θ_2 — алгебраически независимы над Q. Фиксируем также постоянную $\mu>0$.

Определение 0.2. Будем говорить, что (вбилзи $\vec{\theta}$) выполняется условие (Ц) с фиксированным μ, если существуют такие постоянные $c_1, c_2, c_3>0$, что выполнено исловие

(Ц. 1). Существует семейство $\{P_H(x, y): H\geqq c_1\}$ полиномов из $Z[x, y]$, занумерованных всеми натуральными числами $\geqq c_1$, такое, что

1) $P_H(x, y)\not\equiv 0$ для всех натуральных $H\geqq c_1$;

2) для всех $H\geqq c_1$ имеем

$$t(P_H) = \max\{\ln H(P_H), d(P_H)\} \leqq c_2 H; \tag{0.2}$$

3) для фиксированной неограниченной нестрицательной монотонно возрастающей функции $\varphi(t): t\in R^+$ и любого натурального $H\geqq c_3$ имеем оценку $P_H(x, y)$ в $\vec{\theta}$:

$$|P_H(\theta_1, \theta_2)| < \exp\left(-H^\mu\varphi(H)\right). \tag{0.3}$$

Сделаем сразу же важное замечание. Поскольку семейство Дирихле для $\vec{\theta}$ (см. пример 0.1 [1]) показывает выполнение [Ц] для любой точки $\vec{\theta}$ при всяком $\mu<3$, то естественно, для отбрасывания тривиальных возможностей, сразу же положить $\mu\geqq 3$. Это мы и сделаем.

Предположение 0.3. *Будет предполагаться, что в определении* (Ц) *из* (0.2) *всегда* $\mu\geqq 3$.

Впрочем, для настоящей работы естественнее бы выглядела гипотеза $\mu\geqq 4$.

Первым результатом, с которым можно сравнивать все наши, является уже упоминавшаяся теореме Д. Броунвейля [3] — теорема 0.7 [1]. Сформулируем ее:

Теорема 0.4 [3]. *Допустим, что для* $\vec{\theta}=(\theta_1, \theta_2)$ *выполнено условие* (Ц) *с заданным* μ. *Тогда ни поле* $Q(\theta_1)$, *ни поле* $Q(\theta_2)$ *не имеют типа трансцендентности* $\leqq \mu/2$.

Кратко обсудим, насколько возможно улучшение этого результата. Как видно из (Ц), наилучшее на что можно расяитывать — это оценка снизу вида $\tau>\mu$ (если бы у $P_H(x, y)$ отсутствовала каждая из переменных бесконечно часто). Но случай $\mu<3$, имеющий (согласно 0.1 [1]) место для всех $\vec{\theta}\in C^2$ и невозможность $\tau>2$ для всех трансцендентных θ, убеждают в нереальности этого предположения. Видно, что и утверждение типа $\tau>\mu-1+\varepsilon$ для любого $\varepsilon>0$ также нельзя доказать. С другой стороны, в этой работе уже доказано утверждение вида $\tau\geqq\mu-2$, что значительно лучше 0.4 во всех нетривиальных случаях (при $\mu\geqq 4$).

В конце работы кратко говорится о возможных обобщениях и применениях доказанных результатов.

§ 1. Вспомогательные утверждения

В этом параграфе приведем несколько вспомогательных утверждений из [3], [5], [1], необходимых для исследований.

Во-первых, эта лемма об оценке типа делителя полинома — лемма 1.1 [1]. Ее варианты и доказательства приведены в [14].

Лемма 1.1. *Пусть* $P(x_1, \ldots, x_n)$, $P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)$ — *полиномы из* $C[x_1, \ldots, x_n]$. *Обозначим через* $d(P)$ — *степень* $P(x_1, \ldots, x_n)$ $d(P)=d_{x_1}(P)+\ldots+d_{x_n}(P)$. *Если*

$$P(x_1, \ldots, x_n) = P_1(x_1, \ldots, x_n)\ldots P_m(x_1, \ldots, x_n), \tag{1.1}$$

то для высот $H(P), H(P_1), \ldots, H(P_m)$ *имеем*

$$H(P)\cdot e^{d(P)} \geqq H(P_1)\ldots H(P_m). \tag{1.2}$$

Из леммы 1.1, в частности, вытекает, что при соблюдении условия (1.1) для типов $t(P), t(P_1), \ldots, t(P_m)$ получаем вместо (1.2) оценку

$$t(P_1)+\ldots+t(P_m) \leqq 3t(P). \tag{1.3}$$

Поэтому лемма 1.1 будет использоваться скорее не в форме оценки (1.2), а в форме оценки (1.3).

Поскольку мы используем результанты, необходимы результаты о них. Для нас оказывается неважен вид или описание результанта двух полиномов. Необходимо на деле лишь использовать следующую лемму.

Лемма 1.2. *Допустим, что* $P(x, y)$, $S(x, y)\in Z[x, y]$ — *два ненулевых полинома, не имеющих неконстантных общих делителей*[1]. *Тогда существует*

[1] Такие полиномы, мы, как и в [1], называем просто взаимно простыми.

полином $R(x)\in Z[x]$ — *результант* $R(P, S)_y$ *полиномов* $P(x, y)$ *и* $S(x, y)$ *по* y — *такой, что выполняются следующие условия:*

1) $R(x) \not\equiv 0$;

2) *тип* $t(R)$ $R(x)$ *оценивается через степени* $d(P)$ *и* $d(S)$ *и типы* $t(P)$ *и* $t(S)$ y $P(x, y)$ *и* $S(x, y)$:

$$t(R) \leqq c_4(d(P)t(S)+d(S)t(P)). \tag{1.4}$$

В частности,

$$t(R) \leqq 2c_4 t(P)t(S). \tag{1.5}$$

В (1.4)—(1.5) $c_4>0$ — *некоторая абсолютная постоянная;*

3) *для любых комплексных чисел* ϑ_1, ϑ_2 *имеем оценку:*

$$\max\{|P(\vartheta_1, \vartheta_2)|,\ |S(\vartheta_1, \vartheta_2)|\} \geqq \exp\left(-c_5(d(P)t(S)+d(S)t(P))\right)\cdot|R(\vartheta_1)|, \tag{1.6}$$

где $c_5>0$ — *некоторая абсолютная постоянная.*

В частности, имеем

$$\max\{|P(\vartheta_1, \vartheta_2)|,\ |S(\vartheta_1, \vartheta_2)|\} \geqq \exp\left(-2c_5 t(P)t(S)\right)\cdot|R(\vartheta_1)|. \tag{1.7}$$

Построение полинома $R(x)$ через коэффициенты у $P(x, y)$ и $S(x, y)$ как полиномов по y есть во многих местах [3]. Результат 1) общеизвестен. Оценки (1.4)—(1.5) и (1.6)—(1.7) есть в статье [3], там же изложен простой метод доказательства. Наконец, в работе автора [8] содержится лемма 1.2 и даже значительно более общие утверждения. Правда, в [8] нет конструкции $R(x)$, но это и не нужно для целей исследования.

Наконец, исследуя цветные посцедовательности, мы опираемся на результаты работы [1]. Из всех результатов доказанных [1], нам совершенно необходима теорема 2.2 [1], являющаяся основой всех построений.

Теорема 1.3 (теорема 2.2 [1]). *Допустим, что выполняется условие* (Ц) *для заданного* $\vec{\theta}\in C^2$ *и при* $\mu(\geqq 3)$. *Тогда существуют такие постоянные* $c_6, c_7, c_8>0$, *эффективно определяемые по* c_1, c_2, c_3, μ *и* $\varphi(t)$, *что выполняется следующее условие:*

(Ц. 4). *Существует такая бесконечная возрастающая последовательность* $X_n: n=0, 1, 2, \ldots$ *натуральных чисел:*

$$[c_6+1] = X_0 < X_1 < \ldots < X_n < \ldots <: n < \infty \tag{1.8}$$

и последовательность соответствующих полиномов $R_{X_n}(x, y)\equiv R_n(x, y)$:

$$R_0(x, y),\ R_1(x, y),\ \ldots,\ R_n(x, y),\ \ldots: n < \infty \tag{1.9}$$

из $Z[x, y]$ *такие, что выполняются все следующие условия:*

1) *для любого* $n=0, 1, 2, \ldots R_n(x, y)\not\equiv 0$;

2) *для любого* $n=0, 1, 2, \ldots$ *полиномы* $R_n(x, y)$ *и* $R_{n+1}(x, y)$ — *взаимно просты*;

3) *если* $X_n+1<X_{n+1}$: $n=0, 1, 2, \ldots$, *то полином* $R_n(x, y)$ *является степенью неприводимого над* Q *полинома* $S_n(x, y)$ *из* $Z[x, y]$;

4) *для всякого* $n=0, 1, 2, \ldots$ *имеют место неравенства*

$$t(R_n) \leqq c_7 X_n; \tag{1.10}$$

$$|R_n(\theta_1, \theta_2)| < \exp\left(-1/4 X_n^{\mu} \varphi(X_n)\right). \tag{1.11}$$

5) *Для любого* $n=0, 1, 2, \ldots$ *существует такое натуральное* $s_n \geqq 1$, *что выполняются неравенства, связывающие* X_n *и* X_{n+1};

$$s_n t(R_n) \leqq c_8 X_{n+1}; \tag{1.12}$$

$$|R_n(\theta_1, \theta_2)| < \exp\left(-1/4 (X_{n+1}-1)^{\mu} \varphi(X_{n+1}-1)/s_n\right). \tag{1.13}$$

Приведенная формулировка теоремы 2.2 [1] незначительно отличается от оригинальной. Здесь только из всех условий (Ц. 2), выполненных для) $\{R_H(x, y)\}$ оставлено одно — 1) (Ц. 2) — это условие 1). Далее вместо $R_{X_n}(x, y)$ пишем просто $R_n(x, y)$. Условие i) (Ц. 3) — это 2) (Ц. 4) — ii) (Ц. 3) — 3) (Ц. 4); iv) и v) (Ц. 3) — это 4) и 5) (Ц. 4), а iii) (Ц. 3) просто опущено.

Теперь можно просто отказаться от определения 0.2 цветной последовательности и вместо этого использовать последовательность полиномов $\{R_n(x, y): n=0, 1, 2, \ldots\}$, существование которых постулировано в теореме 1.3.

Теперь уже можно приступить к изложению основного результата работы.

В заключение этого предварительного параграфа приведем еще два утверждения (уже о полиномах от одной переменной), используемые только как вспомогательные утверждения в § 2. Одна лемма — довольно известное утверждение.

Лемма 1.4. *Пусть* $P(x) \in Z[x]$, *а* θ — *произвольное комплексное число, причем*

$$|P(\theta)| \leqq \varepsilon < H(P)^{-4d(P)}, \tag{1.14}$$

где $H(P)$, $d(P)$ — *высота и степень* $P(x)$. *Тогда существует делитель* $S(x)$ *полинома* $P(x)$, *являющийся степенью неприводимого над* Q *полинома* $P_0(x) \in Z[x]$,

$$S(x) = \left(P_0(x)\right)^s, \tag{1.15}$$

такой, что

$$|S(\theta)| \leqq \varepsilon H(P)^{4d(P)} < 1. \tag{1.16}$$

Для полинома $P_0(x) \in Z[x]$ получаем оценки

$$t(P_0) \leqq 3t(P)/s; \tag{1.17}$$

$$|P_0(\theta)| \leqq \varepsilon^{1/s} H(P)^{4d(P)/s} < 1. \tag{1.18}$$

Понятно, что (1.17)—(1.18) следуют непосредственно из (1.15)—(1.16) и леммы 1.1—(1.3). Доказательства леммы 1.4 приводятся в [5], [10], [14], [9].

Вторая лемма — это фактически результанты дла полиномов от одной переменной. Эта лемма получается, если предположить просто, что $P(x, y)$ и $S(x, y)$ в лемме 1.2 не зависят от x. Тем не менее, для полноты сформулируем отдельное утверждение:

Лемма 1.5. *Допустим, что* $P(x)$, $S(x) \in Z[x]$ *ненулевые полиномы без общих неконстантных делителей. Тогда для всякого комплексного числа* θ *имеем*

$$\max\{|P(\theta)|, |S(\theta)|\} \geqq \exp\left(-c_9(d(P)t(S)+d(S)t(P))\right) \geqq \exp\left(-2c_9 t(P)t(S)\right), \tag{1.19}$$

где $c_9 > 0$ *абсолютная постоянная.*

Как отмечалось, 1.5 получается из 1.2, если $P(x, y)$, $S(x, y)$ не зависят от x. В этом случае (1.19) следует из (1.6)—(1.7), т. е. $R(x) \equiv R_0 \in Z$ и $|R_0| \geqq 1$. Непосредственное доказательство леммы 1.5 проводится в [5], [9].

§ 2. Типы трансцендентности полей, связанных с цветными последовательностями

Вспомогательные утверждения, установленные в § 1, дают все необходимое для доказательства основного результата работы — теоремы 2.1, значительно обобщающей теорему 0.4.

Теорема 2.1. *Допустим, что выполняется условие* (Ц) *для точки* $\vec{\theta} = (\theta_1, \theta_2) \in C^2$ *при заданном* μ. *Тогда ни поле* $Q(\theta_1)$, *ни поле* $Q(\theta_2)$ *не имеют типа трансцендентности* $\leqq \mu - 2$.

Замечание 2.1′. Эта теорема точнее 0.4 во всех нетривиальных случаях — $\mu \geqq 4$. При $\mu < 4$ как 0.4, так и 2.1 ничего не дают — поля типа трансцендентности < 2 — алгебраические.

Доказательство теоремы 2.1. Предположим, что выполняется условие (Ц) из 0.2 для точки $\vec{\theta} = (\theta_1, \theta_2)$ и заданного μ. Учитывая 0.3 и сделанное замечание 2.2, можно считать, что $\mu \geqq 4$. Впрочем, это ограничение не нужно.

Применим теорему 1.3. Получаем бесконечную возрастающую последовательность натуральных чисел

$$X_0 < X_1 < X_2 < \ldots < X_n < \ldots : n < \infty \tag{2.1}$$

и последовательность полиномов $R_0(x, y), \ldots, R_n(x, y), \ldots : n < \infty$ удовлетворяющих условиям (Ц. 4).

Идея всех дальнейших рассуждений состоит в рассмотрении результантов полиномов $R_n(x, y)$ и $R_{n+1}(x, y)$ по y. Поскольку $R_n(x, y)$ и $R_{n+1}(x, y)$ взаимно просты по 2) (Ц. 4), то результант $R_{n,n+1}(x) \not\equiv 0$ существует, причем в силу леммы 1.2 имеют место хорошие оценки $|R_{n,n+1}(\theta_i)|$. Однако, эти оценки будут тем лучше, чем больше X_{n+1} по сравнению с X_n. Поэтому приходится рассматривать одновременно все результанты $R_{n,n+1}(x)$.

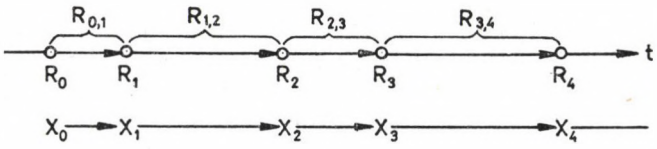


Рис. 1

На рис. 1 изображена симболическая таблица; на оси t отложены типы $t(R_n)$ полиномов $R_n(x, y)$, отвечающих точкам $X_n \in N$. Стрелка от R_n к R_{n+1}, от X_n к $X_{n+1}-1$ означает, что согласно (Ц. 3)—(Ц. 4) (см. условия 3), 5) (Ц. 4) и замечание 2.3 [1]) между X_n и $X_{n+1}-1$ все полиномы являются степенями того же неприводимого полинома, степенью которого является $R_n(x, y)$ при $X_n+1<X_{n+1}$ по 3) (Ц. 4). Показатель $s_n \geqq 1$ в 5) (Ц. 4) — это степень, в которую нужно возвести $R_n(x, y) \equiv R_{X_n}(x, y)$, чтобы получить полином $R_{X_{n+1}}(x, y)$ из (Ц. 3). Таким образом, рис. 1 как бы возвращает нас к чисто «цветной» последовательности из (Ц. 3).

Реализуем описанную выше идею. Согласно 1), 2) (Ц. 4) $R_n(x, y)$ и $R_{n+1}(x, y)$ — ненулевые взаимно простые полиномы. Поэтому согласно лемме 1.2 для любого $n=0, 1, 2, \ldots$ существует полином $R_{n,n+1}(x)$ — результант $R_n(x, y)$ и $R_{n+1}(x, y)$ по y — удовлетворяющий условиям леммы 1.2. Это означает, что

$$R_{n,n+1}(x) \in Z[x], \quad R_{n,n+1}(x) \not\equiv 0,$$

причем имеют место оценки

$$t(R_{n,n+1}) \leqq 2c_4 t(R_n) t(R_{n+1}); \tag{2.2}$$

$$|R_{n,n+1}(\theta_1)| \leqq \exp\left(2c_5 t(R_n) t(R_{n+1})\right) \cdot \max\{|R_n(\theta_1, \theta_2)|, |R_{n+1}(\theta_1, \theta_2)|\}. \tag{2.3}$$

С помощью (2.2.—(2[3] и (Ц. 4) докажем теперь, что поле $Q(\theta_1)$ не имеет типа трансцендентности $\leqq \mu-2$. Так как задача симметрична, то поменяв θ_1 и θ_2 местами, получим, что тип трансцендентности $Q(\theta_2)$ больше $\mu-2$.

$$|R_{n,n+1}(\theta_1)| \leqq \exp\left(2c_5 t(R_n) t(R_{n+1}) - 1/4 X_n^{\mu} \varphi(X_n)\right). \tag{2.4}$$

Используя неравенства (1.10)—(1.11) из (Ц. 4), преобразуем (2.2)—(2.3). Получаем

$$t(R_{n,n+1}) \leqq c_{10} X_n X_{n+1}, \tag{2.5}$$

где $c_{10}=2c_4 \cdot c_7^2>0$. Кроме того, поскольку $\varphi(t)$ — монотонно возрастающая последовательность, то из (1.11) и (2.3) следует

$$|R_{n,n+1}(\theta_1)| \leqq \exp\left(c_{11} X_n X_{n+1} - 1/4 X_n^{\mu} \varphi(X_n)\right), \tag{2.6}$$

где $c_{11}=2c_5 c_7>0$.

Вернемся к доказательству того, что поле $Q(\theta_1)$ имеет тип трансцендентности $>\mu-2$. Допуцтим, напротив, что поле $Q(\theta_1)$ имеет тип трансцендентности $\leqq \mu-2$. Согласно определению 0.1 это означает, что существует такая постоянная $C_\theta>0$, что для всех полиномов $P(x) \in Z[x]$, $P(x) \not\equiv 0$ справедлива оценка

$$|P(\theta_1)| > \exp\left(-C_\theta t(P)^{\mu-2}\right). \tag{2.7}$$

Мы покажем, что предположение (2.7) приводит к противоречию.

Для этого выведем из предположения (2.7) — о том, что $Q(\theta_1)$ имеет $\tau \leqq \mu-2$ — следующее ограничение на рост отношения $\overline{\lim\limits_{n\to\infty}} \ln(X_{n+1})/\ln X_n$. В лемме 2.2 показывается, что при $\tau \leqq \mu-2$ этот предел не более $\mu-3$. После доказательства 2.2 покажем, что на деле этот предел больше $\mu-3$. Нужно заметить, что на деле существование оценки сверху для $\overline{\lim\limits_{n\to\infty}} \ln(X_{n+1})/\ln X_n$

можно вывести даже из оценки $\tau<\mu$ для любого $\mu>4$. Это соображение будет реализовано в дальнейших работах автора.

Лемма 2.2. *Существует такое* $n_0 \geqq 0$, *что при всех* $n \geqq n_0$

$$X_{n+1} \leqq X_n^{\mu-3}. \tag{2.8}$$

Доказательство леммы 2.2. Допустим, напротив, что существует бесконечная последовательность $n_k<\infty$: $n_0<n_1<\ldots<n_k<\ldots: k<\infty$ натуральних чисел, такая, что

$$X_{n_k+1} > X_{n_k}^{\mu-3}: \; k<\infty. \tag{2.9}$$

Обратимся к свойству 5) (Ц. 4) вместо (2.6). Для натурального $s_n \geqq 1$ имеем:

$$s_n t(R_n) \leqq c_8 X_{n+1}; \tag{2.10}$$

$$|R_n(\theta_1, \theta_2)| < \exp\left(-1/4(X_{n+1}-1)^{\mu}\varphi(X_{n+1}-1)/s_n\right). \tag{2.11}$$

Следовательно, из (2.3) получаем

$$|R_{n,n+1}(\theta_1)| \leqq \exp\left(c_{12}X_{n+1}^2/s_n\right)\times \tag{2.12}$$

$$\times\max\left\{\exp\left(-1/4(X_{n+1}-1)^{\mu}\varphi(X_{n+1}-1)/s_n\right), \exp\left(-1/4X_{n+1}^{\mu}\varphi(X_{n+1})\right)\right\}.$$

Поскольку $\varphi(t): t\in R^+$ — монотонно возрастающая функция и $s_n \geqq 1$, то из (2.12) следует

$$|R_{n,n+1}(\theta_1)| \leqq \exp\left(c_{12}X_{n+1}^2/s_n - 1/4(X_{n+1}-1)^{\mu}\varphi(X_{n+1}-1)/s_n\right). \tag{2.13}$$

Поскольку $\mu \geqq 4$, а $\varphi(t): t\in R^+$ неограниченно возрастающая функция при $t\to\infty$, то существует такое $n_1<\infty$ и такое $c_{14}>0$, что (2.13) даёт при $n\geqq n_1$,

$$|R_{n,n+1}(\theta_1) \leqq \exp\left(-c_{14}X_{n+1}^{\mu}\varphi(X_{n+1}-1)/s_n\right). \tag{2.14}$$

В (2.14) постоянные $n_1<\infty$ и $c_{14}>0$ находятся из условия:

$$\varphi(X_{n_1+1}-1) \geqq 1+c_{12}; \quad X_{n_1+1} > 2\mu; \quad c_{14} = 1/8\cdot 2^{\mu}.$$

Отметим, что ни здесфь, ни в других местах работы параметры не выбираются наилучшим образом; их выбор подчинён только конкретным требованиям.

Применим (2.14) к последовательности X_{n_k} из (2.9). Выберем $k_0<\infty$ так, чтобы $n_{n_0} \geqq n_1$. Тогда (2.14) записывается в виде

$$|R_{n_k,n_k+1}(\theta_1)| \leqq \exp\left(-c_{14}X_{n_k+1}^{\mu}\varphi(X_{n_k+1}-1)s_{n_k}^{-1}\right) \tag{2.15}$$

при $k\geqq k_0$. Оценим теперь тип $t(R_{n,n+1})$ непосредственно из (2.2) и (1.10):

$$t(R_{n,n+1}) \leqq 2c_4 t(R_n)t(R_{n+1}) \leqq 2c_4 c_7 t(R_n)\cdot X_{n+1}. \tag{2.16}$$

Учтем, что по построению $R_{n,n+1}(x)\not\equiv 0$ для любого n. Поэтому к полиному $R_{n,n+1}(x)$ можно применить неравенство (2.7). Получаем согласно (2.16):

$$|R_{n,n+1}(\theta_1)| \geqq \exp\left(-c_9(c_{15}t(R_n)X_{n+1})^{\mu-2}\right) \tag{2.17}$$

при $c_{15}=2c_4c_7$. Сравним (2.15) с (2.17) при $n=n_k$ для любого $k\geqq k_0$. Получим

$$c_{16}t(R_{n_k})^{\mu-2}X_{n_k+1}^{\mu-2}\geqq c_{14}X_{n_k+1}^{\mu}\varphi(X_{n_k+1}-1)\cdot s_{n_k}^{-1}. \tag{2.18}$$

Неравенство (2.18) перепишем в виде

$$c_{17}s_{n_k}t(R_{n_k})^{\mu-2}\geqq X_{n_k+1}^{2}\varphi(X_{n_k+1}-1) \tag{2.19}$$

при $k\geqq k_0$. Согласно (2.10),

$$s_{n_k}t(R_{n_k})\leqq c_8X_{n_k+1}.$$

Поэтому (2.19) принимает вид

$$c_8c_{17}t(R_{n_k})^{\mu-3}\geqq X_{n_k+1}\varphi(X_{n_k+1}-1) \tag{2.20}$$

при $k\geqq k_0$. Учтем неравенство (1.10) для $t(R_{n_k})$:

$$t(R_{n_k})\leqq c_7X_{n_k}.$$

Тогда из (2.20) следует

$$c_{18}X_{n_k}^{\mu-3}\geqq X_{n_k+1}\varphi(X_{n_k+1}-1) \tag{2.21}$$

при $c_{18}=c_8c_{17}c_7$. Выбираем $k_1\geqq k_0$ таким образом, чтобы $\varphi(X_{n_{k_1}+1}-1)>c_{18}$. Тогда при $k\geqq k_1$ имеем $\varphi(X_{n_k+1}-1)>c_{18}$. Следовательно, (2.21) при $k\geqq k_1$ $(\geqq k_0)$ дает

$$X_{n_k}^{\mu-3}>X_{n_k+1}. \tag{2.22}$$

Последнее неравенство противоречит (2.9). Следовательно, (2.9) невозможно и (2.8) имеет место. Лемма 2.2 доказана.

Покажем, что для всех полиномов $R_{n,n+1}(x)$ оценки (2.2)—(2.3) вместе с леммой 1.4 приводят к выделению у $R_{n,n+1}(x)$ нетривиального неприводимого делителя, с хорошей оценкой сверху в точке θ_1.

Для этого при любом $n\geqq n_1$ рассмотрим неравенство (2.13), являющееся являющееся прямым следствием (2.3) и (2.10)—(2.11). Перепишем это неравенство для $n\geqq n_1$:

$$|R_{n,n+1}(\theta_1)|\leqq\exp\left(-c_{14}X_{n+1}^{\mu}\varphi(X_{n+1}-1)\cdot s_n^{-1}\right). \tag{2.23}$$

Учтем, что согласно (2.10) имеем

$$c_8^{-1}t(R_n)\leqq s_n^{-1}X_{n+1}. \tag{2.24}$$

Применяя (2.23) и 2.24), получаем

$$|R_{n,n+1}(\theta_1)|\leqq\exp\left(-c_{19}X_{n+1}^{\mu-1}t(R_n)\varphi(X_{n+1}-1)\right) \tag{2.25}$$

для всех $n\geqq n_1$ при $c_{19}=c_{14}c_8^{-1}$. Теперь параллельно с (2.25) рассмотрим оценку $t(R_{n,n+1})$. Именно согласно (2.2) и (1.10)—(1.11) получаем:

$$t(R_{n,n+1})\leqq 2c_4t(R_n)t(R_{n+1})\leqq 2c_4c_7X_{n+1}t(R_n).$$

Следовательно, при $c_{20}=2c_4c_7$ получаем

$$t(R_{n,n+1})\leqq c_{20}t(R_n)X_{n+1}. \tag{2.26}$$

Найдем такое $n_2 \geqq n_1$, чтобы при $n \geqq n_2$:

$$4(t(R_{n,n+1}))^2 < c_{19}/2X_{n+1}^{\mu-1} t(R_n)\varphi(X_{n+1}-1). \tag{2.27}$$

Для этого в силу (2.26) нужно, чтобы

$$4c_{20}^2 t(R_n)^2 X_{n+1}^2 < c_{19}/2X_{n+1}^{\mu-1} t(R_n)\varphi(X_{n+1}-1).$$

Последнее неравенство эквивалентно следующему

$$4c_{20}^2 t(R_n) < c_{19}/2X_{n+1}^{\mu-3}\varphi(X_{n+1}-1). \tag{2.28}$$

Для доказательства (2.28) согласно (1.10) достаточно, чтобы

$$4c_{20}^2 c_7 < c_{19}/2\varphi(X_{n+1}-1). \tag{2.29}$$

В самом деле, если (2.29) выполнено, то $\mu \geqq 4$, т. е. $X_{n+1}^{\mu-4} \geqq 1$ и

$$4c_{20}^2 c_7 < c_{19}/2X_{n+1}^{\mu-4}\varphi(X_{n+1}-1).$$

Так как $X_n < X_{n+1}$, то

$$4c_{20}^2(c_7 X_n) < c_{19}/2 \cdot X_{n+1}^{\mu-3}\varphi(X_{n+1}-1).$$

Тогда (1.10) даёт (2.28).

Теперь выбираем n_2 так, чтобы $n_2 > n_1$ и

$$4c_{20}^2 c_7 \cdot (c_{19}/2)^{-1} < \varphi(X_{n_2+1}-1). \tag{2.30}$$

Поскольку $\varphi(t): t \in R^+$ — монотонная функция, то из (2.30) вытекает (2.29) для всех $n \geqq n_2$. Поэтому выполнено (2.28) и (2.27) для всех $n \geqq n_2$.

Неравенство (2.27) гарангирует возможность применения леммы 1.4 к $R_{n,n+1}(x)$. В обозначеинях леммы 1.4 имеем $P(x) = R_{n,n+1}(x)$ и согласно (2.25),

$$\varepsilon = \exp\left(-c_{19} X_{n+1}^{\mu-1} t(R_n)\varphi(X_{n+1}-1)\right). \tag{2.31}$$

Условие (2.27) гарантирует выполнение при $n \geqq n_2$ усло-условия (1.14):

$$|R_{n,n+1}(\theta_1)| \leqq \varepsilon < \exp\left(-4(t(R_{n,n+1}))^2\right).$$

Теперь согласно лемме 1.4 для всякого $n \geqq n_2$ существует такой неприводимый полином над Q, $S_n(x) \in Z[x]$, что $(S_n(x))^{f_n}$ является делителем $R_{n,n+1}(x)$ и выполнено (1.17)—(1.18) при $P_0(x) = S_n(x)$. Следовательно, при $n \geqq n_2$:

$$t(S_n) \leqq 3t(R_{n,n+1})/f_n \tag{2.32}$$

$$|S_n(\theta)| \leqq \varepsilon_n^{1/f_n} \cdot \exp\left(4(t(R_{n,n+1}))^2/f_n\right). \tag{2.33}$$

Учитывая (2.26) м (2.31), перепишем неравенства (2.32)—(2.33) в виде

$$t(S_n) \leqq c_{21} t(R_n) \cdot X_{n+1}/f_n; \tag{2.34}$$

$$|S_n(\theta_1)| \leqq \exp\left(-c_{19} X_{n+1}^{\mu-1} t(R_n)\varphi(X_{n+1}-1) \cdot f_n^{-1} + 4(t(R_{n,n+1}))^2 \cdot f_n^{-1}\right) \tag{2.35}$$

при $c_{21} = 3c_{20}$. Согласно (2.27) при $n \geqq n_2$ неравенство (2.35) записывается в виде

$$|S_n(\theta_1)| \leqq \exp\left(-c_{22} X_{n+1}^{\mu-1} t(R_n) \cdot \varphi(X_{n+1}-1) \cdot f_n^{-1}\right) \tag{2.36}$$

при $c_{22}=c_{19}/2$ и при $n\geqq n_2$. Учитывая (2.34), получаем

$$-t(R_n)\cdot X_{n+1}f_n^{-1}\leqq -c_{21}^{-1}t(S_n).$$

Поэтому из (2.36) следует при $n\geqq n_2$:

$$|S_n(\theta_1)|\leqq \exp\left(-c_{23}X_{n-1}^{\mu-2}t(S_n)\varphi(X_{n+1}-1)\right) \tag{2.37}$$

с

$$c_{23}=c_{22}\cdot c_{21}^{-1}.$$

Для заданного $n\geqq n_2$ через $m(n)$, $n<m(n)<\infty$ обозначим такое наименьшее m, что $S_n\neq\pm S_m$. Тогда

$$S_n=\pm S_m:\ n\leqq m<m(n),\quad S_n\neq\pm S_{m(n)}. \tag{2.38}$$

Такое $m(n)<\infty$ существует для любого n. Если бы его не существовало, то по (2.37) мы бы имели для всех m, $n\geqq m<\infty$:

$$|S_n(\theta_1)|\leqq \exp\left(-c_{23}X_{m+1}^{\mu-2}t(S_n)\varphi(X_{m+1}-1)\right).$$

При $m\to\infty$ это дает $S_n(\theta_1)=0$, что противоречит трансцендентности θ_1. Следовательно, $m(n)<\infty$, удовлетворяющее (2.38), существует. Теперь, согласно (2.37) и (2.38), получаем

$$|S_n(\theta_1)|\leqq \exp\left(-c_{23}X_{m(n)}^{\mu-2}t(S_n)\varphi(X_{m(n)}-1)\right) \tag{2.39}$$

и

$$|S_{m(n)}(\theta_1)|\leqq \exp\left(-c_{23}X_{m(n)+1}^{\mu-2}t(S_{m(n)})\cdot\varphi(X_{m(n)}-1)\right). \tag{2.40}$$

Согласно (2.38) полиномы $S_n(x)$ и $S_{m(n)}(z)$, являющиеся неприводимыми над Q, — взаимно просты и не имеют общих неконстантных делителей. Следовательно, можно применить лемму 1.5. Согласно лемме 1.5 получаем

$$\max\{|S_n(\theta_1)|,\ |S_{m(n)}(\theta_1)|\}\geqq \exp\left(-2c_9t(S_n)t(S_{m(n)})\right). \tag{2.41}$$

Будем сравнивать (2.39)—(2.40) с (2.41). Если $t(S_{m(n)})\leqq t(S_n)$, то (2.39)—(2.40) даёт

$$\max\{|S_n(\theta_1)|,\ |S_{m(n)}(\theta_1)|\}\leqq \exp\left(-c_{23}X_{m(n)}^{\mu-2}t(S_{m(n)})\varphi(X_{m(n)}-1)\right) \tag{2.42}$$

при $t(S_{m(n)})\leqq t(S_n)$. Тогда из (2.41) получаем

$$c_{23}X_{m(n)}^{\mu-2}\varphi(X_{m(n)}-1)\leqq 2c_9t(S_n). \tag{2.43}$$

Однако, по определению, $m(n)>n$. Значит, из (2.43) следует

$$c_{24}X_{n+1}^{\mu-2}\varphi(X_{n+1}-1)<t(S_n) \tag{2.44}$$

при $c_{24}=c_{23}(2c_9)^{-1}$. Но по (2.34) имеем для любого $n\geqq n_2$,

$$t(S_n)\leqq c_{21}t(R_n)X_{n+1}\leqq c_{21}c_7X_nX_{n+1}\leqq c_{21}c_7X_{n+1}^2. \tag{2.45}$$

Но неравенства (2.44) и (2.45) несовместимы при $n\geqq n_3$, где $n_3\geqq n_2$ такое, что

$$\varphi(X_{n_3+1}-1)>c_{21}c_7(c_{24})^{-1}.$$

Следовательно, случай $t(S_{m(n)})\leqq t(S_n)$ невозможен для всех $n\geqq n_3$.

Поэтому рассмотрим случай $n \geqq n_3$ и $t(S_n) \leqq t(S_{m(n)})$. В этом случае из (2.39)—(2.40) вместо (2.42) получаем

$$\max\{|S_n(\theta_1)|, |S_{m(n)}(\theta_1)|\} \leqq \exp\left(-c_{23} X_{m(n)}^{\mu-2} t(S_n) \varphi(X_{m(n)}-1)\right) \tag{2.46}$$

при всех $n \geqq n_3$. Сравнивая (2.41) с (2.46), получаем

$$c_{23} X_{m(n)}^{\mu-2} \varphi(X_{m(n)}-1) \leqq 2c_9 t(S_{m(n)}). \tag{2.47}$$

Теперь (2.34) вместе с (2.47) даёт

$$c_{23} X_{m(n)}^{\mu-2} \varphi(X_{m(n)}-1) \leqq 2c_9 c_{21} c_7 X_{m(n)} X_{m(n)+1}. \tag{2.48}$$

при $m(n) > n \geqq n_3 \geqq n_1$. При $c_{25} = 2c_9 c_{21} c_7 c_{23}^{-1}$ получаем из (2.48)

$$X_{m(n)}^{\mu-3} \varphi(X_{m(n)}-1) \leqq c_{25} X_{m(n)+1}. \tag{2.49}$$

Неравенство (2.49) выполнено для любого $n \geqq n_3$.

Теперь можно сравнить (2.49) с леммой 2.2, являющейся следствием оценки (2.7). В этом месте мы пользуемся предположением о том, что поле $Q(\theta_1)$ имеет тип трансцендентности $\leqq \mu-2$ *Нужно подчеркнуть, что все выкладки от* (2.1) *до* (2.6) *и* (2.10)—(2.14), (2.23)—(2.49) *никак не зависят от этого предположения, а являются следствием условий* (Ц. 4)[2].

Найдем такое $n_4 \geqq n_3$, чтобы

$$\varphi(X_{n_4}-1) > c_{25}. \tag{2.50}$$

Поскольку $\varphi(t)$ — монотонно возрастающая и $m(n) > n$, то при всех $n \geqq n_4$ из (2.50) следует

$$\varphi(X_{m(n)}-1) > c_{25} \tag{2.51}$$

при $n \geqq n_4$. Тогда из (2.49) и (2.51) выводим при $n \geqq n_4 \geqq n_3$:

$$X_{m(n)}^{\mu-3} < X_{m(n)+1}. \tag{2.52}$$

Это противоречит неравенству (2.8) для любого $n \geqq n_4, n_0$ т. к. $m(n) > n$ Таким образом, лемма 2.2, а вместе с ней и предположение (2.7) неверно.

Поэтому поле $Q(\theta_1)$ не может иметь тип трансцендентности $\leqq \mu-2$. Аналогично (в силу симметрии) и для $Q(\theta_2)$. Теорема 2.1 доказана полностью.

§ 3. Дальнейшие работы этого цикла

Результат, доказанный в настоящей работе, наилучший с точки зрения результантов. Монжно даже выдвинуть предположение о существовании чисел $\theta_1, \theta_2 \in C$ последовательности $X_i : i < \infty$ и полиномов $P_i(x, y) \in Z[x, y]$, для которых

$$t(P_i) \leqq X_i;\ |P_i(\theta_1, \theta_2)| < \exp(-X_{i+1}^{\mu-1} X_i); \tag{3.1}$$

$$\lim_{i \to \infty} \ln X_{i+1} / \ln X_i = \chi < \infty. \tag{3.2}$$

[2] Такое замечание необходимо сделать, т. к. величина $\mu-2$ в теореме 2.1 может быть улучшена на основе (2.23)—(2.49) — см. дальше.

Условия (3.1)—(3.2) гарантируют, что система полиномов

$$\{P_H(x, y) = (P_i(x, y))^{[H/X_i]}: X_i < H \leqq X_{i+1}; 1 \leqq i < \infty\}$$

цветная. Для этого семейства результанты дают оценку типа τ у $Q(\theta_1)$: $\tau \geqq \mu - 2$. В случае $\chi = \mu - 3$ так бы и было. Однако, случай $\chi = \mu - 3$ *невозможен*. Оценка $\tau \geqq \mu - 2$ *может быть улучшена*.

Доказательству подобного утверждения будет посвящена следующай работа. В ней результанты уже не будут основным инструментом. Техника особенностей на алгебраических поверхностях выдвигается на первый план.

ЛИТЕРАТУРА

[1] Касселс, Дж. В. С.: *Введение в теорию диофантовых приближений*, ИЛ (Москва, 1961).
[2] Ax, I.: On Schanuel's conjecture, *Ann. of Math.* **93** (1971), 252—258.
[3] Brownawell, W. D.: Gelfond's method for algebraic independence, *Trans. Amer. Math. Soc.* **210** (1975).
[4] Lang, S.: *Diophantine geometry*, Addison-Wesley, 1962.
[5] Гельфонд, А. О.: *Трансцендентные и алгебраические числа*, Гостехиздат (Москва, 1952).
[6] Чудновский, Г. В.: Алгебраическая независимость нескольких значений показательной функции, *Матем. заметки* **15** (4) (1974).
[7] Чудновский, Г. В.: *Некоторые аналитические методы в теории трансцендентных чисел*, Препринт ИМ—74—8 (Киев, 1974).
[8] Чудновский, Г. В.: *Аналитические методы в диофактовых приближениях*, Препринт ИМ—74—9 (Киев, 1974).
[9] Cijsouw, P.: *Transcendence measures*, Academic Service, 1972.
[10] Lang, S.: *Introduction to transcendental numbers*, Addison—Wesley, 1966.
[11] Lang, S.: Transcendental numbers and diophantine approximations, *Bull. Amer. Math. Soc.* **77** (1971).
[12] Tijdeman, R.: An auxiliary result in the theory of transcendental numbers, *J. Number Theory* **5** (1973).
[13] Tijdeman, R.: On the Gelfond—Baker method and its applications, *Bull. Amer. Math. Soc.* (to appear).
[14] Waldschmidt, M: Nombres transcendants, *Lecture Notes in Math.* v. 402, Springer-Verlag, 1974.
[15] Waldschmidt, M.: *Independance algebrique par la methode de G. V. Čudnovskij*, NG 8 Sem. Delange—Pisot—Poitou, 1974—1975.
[16] Brownawell, W. D., Waldschmidt, M.: The algebraic independence of certain numbers to algebraic powers (to appear).
[17] Brownawell, D.: Sequences of Diophantine Approximations, *J. Number Theory* **6** (1974), 11—21.
[18] Фельдман, Н. И., Шидловский, А. Б.: Развитие и современное состояние теории трансцендентных чисел, *УМП*, **22** (3) (1967).
[19] Schmidt, W.: Approximation to algebraic numbers, *Monat. Math.* 1973.
[20] Davenport, H. and Schmidt, W.: Approximation to real numbers by quadratic irrationals, *Acta Arith.* **13** (1967), 169—176.
[21] Davenport, H. and Schmidt, W.: Approximation to real numbers by algebraic integers, *Acta Arith.* **15** (1969), 393—416.

Ул. Тарасовская 10, кв. 17, 252 033, Киев — 33, СССР

(Поступило 4 апрекя 1976)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 159—160

SOME THEOREMS ON FIXED POINTS

by

B. FISHER

In a paper by KANNAN, see [2], he proved the following theorem:

THEOREM 1. *If T is a mapping of the complete metric space X into itself, satisfying the inequality*

$$d(Tx, Ty) \leqq c\{d(x, Tx)+d(y, Ty)\}$$

for all x, y in X, where $0 \leqq c < \frac{1}{2}$, then T has a unique fixed point.

The following theorem was proved in [1]:

THEOREM 2. *If T is a mapping of the complete metric space X into itself, satisfying the inequality*

$$d(Tx, Ty) \leqq c\{d(x, Ty)+d(y, Tx)\}$$

for all x, y in X, where $0 \leqq c < \frac{1}{2}$, then T has a unique fixed point.

We now prove the following theorem:

THEOREM 3. *If T is a mapping of the complete metric space X into itself, satisfying the inequality*

$$\{d(Tx, Ty)\}^2 \leqq c\{d(x, Tx)d(x, Ty)+d(y, Ty)d(y, Tx)\}$$

for all x, y in X, where $0 \leqq c < \frac{1}{2}$, then T has a unique fixed point.

PROOF. Let x be an arbitrary point in X. Then

$$\{d(T^n x, T^{n+1}x)\}^2 \leqq c\{d(T^{n-1}x, T^n x)d(T^{n-1}x, T^{n+1}x)+0\} \leqq$$

$$\leqq cd(T^{n-1}x, T^n x)\{d(T^{n-1}x, T^n x)+d(T^n x, T^{n+1}x)\}.$$

This inequality implies that

$$d(T^n x, T^{n+1}x) \leqq \frac{1}{2}\{c+(c^2+4c)^{1/2}\}\, d(T^{n-1}x, T^n x)$$

for $n=1, 2, \ldots$ and since $c < \frac{1}{2}$, we have

$$\frac{1}{2}\{c+(c^2+4c)^{1/2}\} < 1.$$

It follows easily that the sequence $\{T^n x\}$ is a Cauchy sequence in the complete metric space X and so has a limit z in X.

We now have

$$\{d(T^n x, Tz)\}^2 \leqq c\{d(T^{n-1}x, T^n x)\, d(T^{n-1}x, Tz) + d(z, Tz)\, d(z, T^n x)\}$$

and on letting n tend to infinity, we see that

$$\{d(z, Tz)\}^2 \leqq 0.$$

It follows that $Tz = z$ so that z is a fixed point of T.

Now suppose that z' is a second fixed point of T. Then

$$\{d(z, z')\}^2 = \{d(Tz, Tz')\}^2 \leqq 0$$

and so $z = z'$. The fixed point is therefore unique. This completes the proof of the theorem.

We finally prove a similar theorem for compact metric spaces:

THEOREM 4. *If T is a continuous mapping of the compact metric space X into itself, satisfying the inequality*

$$\{d(Tx, Ty)\}^2 < \frac{1}{2}\{d(x, Tx)\, d(x, Ty) + d(y, Ty)\, d(y, Tx)\}$$

for all distinct x, y in X, then T has a unique fixed point.

PROOF. Since d and T are continuous functions and X is compact there exists a point z in X such that

$$d(z, Tz) = \inf\{d(x, Tx) : x \in X\}.$$

We will now suppose that $Tz \neq z$. Then

$$\{d(Tz, T^2 z)\}^2 < \frac{1}{2}\{d(z, Tz)\, d(z, T^2 z) + 0\} \leqq$$

$$\leqq \frac{1}{2} d(z, Tz)\{d(z, Tz) + d(Tz, T^2 z)\}.$$

This inequality implies that

$$d(Tz, T^2 z) < d(z, Tz),$$

contradicting the definition of z. It follows that we must have $Tz = z$ so that z is a fixed point of T.

Now suppose that T has a second distinct fixed point z'. Then

$$\{d(z, z')\}^2 = \{d(Tz, Tz')\}^2 < 0,$$

giving a contradiction. The fixed point is therefore unique. This completes the proof of the theorem.

REFERENCES

[1] FISHER, B.: A fixed point theorem, *Math. Mag.* **48** (1975), 223—225.
[2] KANNAN, R.: Some results on fixed points, *Bull. Calcutta Math. Soc.* **60** (1968), 71—76.

Department of Mathematics, University of Leicester

(Received October 23, 1976)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 161—167

A NOTE ON KOLMOGOROV'S LAW OF ITERATED LOGARITHM

by

P. MAJOR

Summary: Let the sequence of independent random variables $X_1, X_2, \dots$ satisfy the conditions of Kolmogorov's law of iterated logarithm. Then the partial sums $S_n = \sum_{i=1}^{n} X_i$, $n=1, 2, \dots$ can be approximated by an appropriate Wiener process. This implies a Strassen type law of iterated logarithm.

Introduction

Kolmogorov's law of iterated logarithm (see e.g. [1]) states the following result:

Let $X_1, X_2, \dots$ be independent random variables $EX_i=0$, $EX_i^2=\sigma_i^2$, $i=1, 2, \dots$, $S_n = \sum_{i=1}^{n} X_i$, $B_n = \sum_{i=1}^{n} \sigma_i^2$, $n=1, 2, \dots$. Let $B_n \to \infty$. Assume the existence of a numerical sequence M_n, $n=1, 2, \dots$, such that

$$M_n = o\left(\sqrt{\frac{B_n}{\log\log B_n}}\right)$$

and

$$P(|X_n| \leqq M_n) = 1.$$

Then the relation

$$\limsup \frac{S_n}{\sqrt{2B_n \log\log B_n}} = 1 \quad \textit{with pr. 1}$$

holds true.

Let us define the process $S(t)$, $t \geqq 0$ in the following way: $S(B_n)=S_n$, $(B_0=0,\ S(0)=0)$ and $S(t) = S_n \frac{B_{n+1}-t}{B_{n+1}-B_n} + S_{n+1} \frac{t-B_n}{B_{n+1}-B_n}$ if $B_n < t < B_{n+1}$. We prove the following

THEOREM. *Let the sequence of independent random variables $X_1, X_2, \dots$ satisfy the conditions of Kolmogorov's law of iterated logarithm. If the probability space where the X_i-s are given is sufficiently rich, one can construct a standard Wiener process $W(t)$ such that*

$$\lim_{t\to\infty} \frac{|W(t)-S(t)|}{\sqrt{t \log\log t}} = 0 \quad \textit{with probability } 1.$$

Since $B_{n+1}-B_n \leqq M_n^2$, and $M_n^2 = o(B_n)$ Strassen's law of iterated logarithm for Wiener process (see [2]) yields the following

COROLLARY. *Define*

$$S_n(t) = \frac{S(B_n t)}{\sqrt{2B_n \log\log B_n}}, \quad n = 1, 2, \ldots, 0 \leqq t \leqq 1.$$

This sequence of functions is relatively compact in the Banach space $C[0, 1]$, *and its limit points agree with the set* K, $K=\{f(t), 0\leqq t\leqq 1; f(0)=0, f(t)$ *is absolute continuous,* $\int_0^1 \dot{f}^2(t)dt \leqq 1\}$ *with probability* 1.

This corollary contains Kolmogorov's law of iterated logarithm as a special case.

PROOF OF THE THEOREM. We need an estimate of $P(S_n - S_m > x)$. Though this estimate is very similar to those needed in the proof of Kolmogorov's law of iterated logarithm, for the sake of completeness we prove it. Our estimation is based on an idea of FELLER (see [3]).

LEMMA. *Let* ε, δ, L *be arbitrary positive numbers. Under the conditions of the Theorem we have for every large* n

$$c \exp\left[-(1+\delta)\frac{x^2}{2(B_n - B_m)}\right] \leqq P(S_n - S_m > x) \leqq \exp\left[-(1-\delta)\frac{x^2}{2(B_n - B_m)}\right]$$

if $n > m$, $B_n - B_m > \varepsilon B_n$, $0 \leqq x \leqq L\sqrt{B_n \log\log B_n}$. *(c is a universal constant.)*

PROOF. First we estimate the moment generating function of $S_n - S_m$ from below and from above. Let $0 \leqq t \leqq t_0 = K\sqrt{\frac{\log\log B_n}{B_n}}$ where $K = 2L/\varepsilon$.

We have for $j \leqq n$ $tM_j < \delta/3$ if n is sufficiently large. Thus

$$E \exp tX_j = 1 + \sum_{k=2}^{\infty} \frac{t^k}{k!} EX_j^k \leqq 1 + \sigma_j^2 \frac{t^2}{2}\left(1 + \frac{t}{3} M_j + \frac{t^2}{12} M_j^2 + \ldots\right) \leqq$$

$$\leqq 1 + \frac{t^2}{2}\sigma_j^2\left(1 + \frac{t}{2} M_j\right) \leqq \exp\left[\left(1 + \frac{\delta}{2}\right)\sigma_j^2 \frac{t^2}{2}\right],$$

and

$$E \exp tX_j \geqq 1 + \frac{t^2}{2}\sigma_j^2\left(1 - \frac{t}{3} M_j - \frac{t^2}{12} M_j^2 - \ldots\right) \geqq \exp\left[\left(1 - \frac{\delta}{2}\right)\sigma_j^2 \frac{t^2}{2}\right].$$

These estimations imply that

$$\exp\left[\left(1 - \frac{\delta}{2}\right)\frac{t^2}{2}(B_n - B_m)\right] \leqq E \exp\left[t(S_n - S_m)\right] \leqq \exp\left[\left(1 + \frac{\delta}{2}\right)\frac{t^2}{2}(B_n - B_m)\right].$$

Define the probability distributions

$$F_j^t(dx) = \frac{\exp(tx) F_j(dx)}{\int \exp(tx) F_j(dx)}, \quad j = 1, 2, \ldots$$

and

$$G_{n,m}^t(dx) = (F_m^t * \ldots * F_n^t)(dx),$$

where $F_j(x)$ is the distribution function of X_j, and $*$ means convolution.

For any Borel set H the equation

$$P(S_n - S_m \in H) = E[\exp t(S_n - S_m)] \int_H \exp(-tx)\, G^t_{n,m}(dx)$$

holds. (These formulae follow from the basic properties of conjugated distributions, see e.g. [4] Chapter XVI. 6.)

Choose

$$t = \frac{x}{(B_n - B_m)\left(1+\frac{\delta}{2}\right)}.$$

Since

$$t \leqq \frac{L\sqrt{B_n \log\log B_n}}{\varepsilon B_n} \leqq t_0$$

the estimation

$$P(S_n - S_m > x) \leqq E\exp[t(S_n - S_m)]\exp(-tx) \leqq \exp\left[-\frac{(1-\delta)x^2}{2(B_n - B_m)}\right]$$

holds true.

Set

$$E^t_j = \int x F^t_j(dx), \quad (D^t_j)^2 = \int x^2 F^t_j(dx) - (E^t_j)^2$$

and

$$E^t_{m,n} = \sum_{j=m}^{n} E^t_j, \quad (D^t_{m,n})^2 = \sum_{j=m}^{n} (D^t_j)^2.$$

In order to get an estimate from below first we show that

(1) $$\left(1-\frac{\delta}{2}\right)(B_n - B_m) < (D^t_{m,n})^2 < \left(1+\frac{\delta}{2}\right)(B_n - B_m)$$

and

(2) $$\left(1-\frac{\delta}{2}\right) t(B_n - B_m) < E^t_{m,n} < \left(1+\frac{\delta}{2}\right) t(B_n - B_m)$$

if $t < t_0$.

(3) $$(D^t_j)^2 \leqq \int_{-M_j}^{M_j} \exp(tx)\, x^2 F_j(dx) \leqq \left(1+\frac{\delta}{2}\right)\int x^2 F_j(dx) = \left(1+\frac{\delta}{2}\right)\sigma_j^2$$

if $j \leqq n$ since $tM_n \leqq \delta/4$.

Similarly, exploiting that $tM_n \to 0$, as $n \to \infty$, we obtain that

(4) $$\int x^2 F^t_j(dx) \geqq \left(1-\frac{\delta}{4}\right)\sigma_j^2$$

and

$$\left|\int_0^{M_j} x F^t_j(dx) - \int_0^{M_j} x F_j(dx)\right| \leqq \frac{\delta}{4}\int_0^{M_j} x F_j(dx),$$

$$\left|\int_{-M_j}^{0} x F^t_j(dx) - \int_{-M_j}^{0} x F_j(dx)\right| \leqq \frac{\delta}{4}\int_{-M_j}^{0} |x| F_j(dx).$$

The last two inequalities imply

$$(E_j^t)^2 \leqq \left[\frac{\delta}{4}\int |x| F_j(dx)\right]^2 \leqq \frac{\delta^2}{16}\sigma_j^2. \tag{5}$$

(3), (4) and (5) give that

$$\left(1-\frac{\delta}{2}\right)\sigma_j^2 < (D_j^t)^2 < \left(1+\frac{\delta}{2}\right)\sigma_j^2 \quad \text{if} \quad j \leqq n.$$

Summing up this inequality from m to n we obtain (1).

Since

$$\frac{d}{dt} E_{m,n}^t = (D_{m,n}^t)^2$$

relation (1) implies (2).

Let us choose $\bar{t}$ as the solution of the equation $E_{m,n}^t = x + 2\sqrt{B_n - B_m}$. ($E_{m,n}^t$ is monotonically increasing in t therefore the equation has a unique solution.)

$$\frac{\left(1-\frac{\delta}{2}\right)(x+2\sqrt{B_n-B_m})}{B_n-B_m} < \bar{t} < \frac{\left(1+\frac{\delta}{2}\right)(x+2\sqrt{B_n-B_m})}{B_n-B_m}$$

because of (2).

Relation (1) and the Chebyshev inequality yield that

$$\int_x^{x+4\sqrt{B_n-B_m}} \exp(-\bar{t}x)\, G_{n,m}^{\bar{t}}(dx) \geqq$$

$$\geqq \exp\left[-\bar{t}\left(x+4\sqrt{B_n-B_m}\right)\right]\left[G_{n,m}^{\bar{t}}\left(x+4\sqrt{B_n-B_m}\right) - G_{n,m}^{\bar{t}}(x)\right] \geqq$$

$$\geqq \frac{2}{3}\exp\left(-\bar{t}x - 4\bar{t}\sqrt{B_n-B_m}\right) \geqq \frac{1}{100}\exp(-\bar{t}x).$$

Thus we can make the following estimation:

$$P(S_n - S_m > x) \geqq P\left(|S_n - S_m - x - 2\sqrt{B_n-B_m}| < 2\sqrt{B_n-B_m}\right) =$$

$$= E\exp[E(S_n - S_m)] \int_x^{x+4\sqrt{B_n-B_m}} \exp(-\bar{t}x)\, G_{n,m}^{\bar{t}}(dx) \geqq$$

$$\geqq \frac{1}{100}\exp\left[\left(1-\frac{\delta}{2}\right)\frac{\bar{t}^2}{2}(B_n - B_m) - \bar{t}x\right] \geqq c\exp\left[-\frac{(1-\delta)x^2}{2(B_n-B_m)}\right].$$

This estimation completes the Proof of the lemma.

Similar inequality holds in the case $0 \geqq x \geqq -L\sqrt{B_n \log\log B_n}$.

Set $G_{n,m}(x) = P(S_n - S_m > x)$, and let α be a uniformly distributed random variable in [0, 1] independent of the S_i-s.

Define

$$\eta = \tilde{G}_{n,m}(S_n - S_m)$$

and

$$T_n - T_m = \Phi^{-1}\left(\frac{\eta}{\sqrt{B_n - B_m}}\right),$$

where $\tilde{G}_{n,m}(x) = G_{n,m}(x) + \alpha\big(G_{n,m}(x+0) - G_{n,m}(x)\big)$, and $\Phi(x)$ is the standard normal distribution function. η is uniformly distributed in [0, 1] and $T_n - T_m$ has a normal distribution functions with expectation 0 and variance $B_n - B_m$.

Our lemma has the following

COROLLARY. *Let ε, η and L be arbitrary positive numbers. Let $B_n - B_m > \varepsilon B_n$. We have for every sufficiently large n*

$$|(S_n - S_m) - (T_n - T_m)| < \eta \sqrt{B_n \log\log B_n}$$

on the set $|S_n - S_m| < L\sqrt{B_n \log\log B_n}$.

PROOF. Since the Lemma holds for every $\delta > 0$ we have

$$\Phi\left(\frac{x - \eta\sqrt{B_n \log\log B_n}}{\sqrt{B_n - B_m}}\right) \leqq G_{n,m}(x) \leqq \Phi\left(\frac{x + \eta\sqrt{B_n \log\log B_n}}{\sqrt{B_n - B_m}}\right),$$

if $|x| < L\sqrt{B_n \log\log B_n}$. This relation implies the corollary.

PROOF OF THE THEOREM. Given any $\varepsilon > 0$ we show that for $n > n(\varepsilon)$ there exists a Wiener process such that

$$P\left(\sup_{t \geqq B_n} \frac{|S(t) - S(B_n) - W(t - B_n)|}{\sqrt{t \log\log t}} > \varepsilon\right) < \varepsilon. \tag{6}$$

Define a sequence of integers $n_0, n_1, \ldots$ in such a way that $n_0 = n$ and $\left(1 + \frac{\varepsilon}{20}\right) B_{n_k} < B_{n_{k+1}} < \left(1 + \frac{\varepsilon}{10}\right) B_{n_k}$. Let us construct a sequence of random variables $T_{n_k} - T_{n_{k-1}}$, $k = 1, 2, \ldots$ as it was done in the corollary. We may assume that the random variables $T_{n_k} - T_{n_{k-1}}$, $k = 1, 2, \ldots$ are independent. If the probability space is rich enough, a Wiener process $W(t)$, $t \geqq 0$ can be constructed in such a way that $W(B_{n_k} - B_n) = T_{n_k} - T_n$ for any $k \geqq 0$. We claim that this Wiener process satisfies relation (6).

First we show that

$$P\left(\sup_k \frac{|S(B_{n_k}) - S(B_n) - W(B_{n_k} - B_n)|}{\sqrt{B_{n_k} \log\log B_{n_k}}} > \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{2}. \tag{7}$$

Indeed,

$$|S(B_{n_k}) - S(B_n) - W(B_{n_k} - B_n)| = \left|\sum_{j=1}^{k} (S_{n_j} - S_{n_{j-1}}) - (T_{n_j} - T_{n_{j-1}})\right|.$$

But the last sum is less than

$$\eta \sum_{j=1}^{k} \sqrt{B_{n_j} \log\log B_{n_j}} < \eta \cdot \frac{20}{\varepsilon} \sqrt{B_{n_k} \log\log B_{n_k}}$$

if $|S_{n_j} - S_{n_{j-1}}| < L\sqrt{B_{n_j} \log\log B_{n_j}}$ for every $j=1, 2, \ldots$.

But choosing L sufficiently large the Lemma implies that the exceptional set has very little probability. Thus choosing $\eta=\varepsilon^2/40$ we obtain relation (7).

To finish the proof of relation (6) it is enough to show that

$$\sum_{k=0}^{\infty} P\left(\sup_{n_k<t<n_{k+1}} \frac{|S(t)-S(B_{n_k})|}{\sqrt{B_{n_k} \log\log B_{n_k}}} > \frac{\varepsilon}{4} \right) < \frac{\varepsilon}{4},$$

and

$$\sum_{k=0}^{\infty} P\left(\sup_{n_k<t<n_{k+1}} \frac{|W(t-B_n)-W(B_{n_k}-B_n)|}{\sqrt{B_{n_k} \log\log B_{n_k}}} > \frac{\varepsilon}{4} \right) < \frac{\varepsilon}{4}.$$

Using a well-known estimation about the supremum of independent random variables one gets that the first sum is less than

$$2 \sum_{k=0}^{\infty} P\left(|S(B_{n_{k+1}}) - S(B_{n_k})| > \frac{\varepsilon}{5} \sqrt{B_{n_k} \log\log B_{n_k}}\right) \leqq$$

$$\leqq 2 \sum_{k=0}^{\infty} P\left(|S_{n_{k+1}} - S_{n_k}| > \frac{\varepsilon}{5} \cdot \frac{10}{\varepsilon} (\log k + c(n))\right),$$

where $C(n) \to \infty$ as $n \to \infty$. If n is sufficiently large then this sum is less than $\varepsilon/4$ because of the Lemma. The second relation may be proved similarly.

One can choose a monotone sequence r_k, $k=1, 2, \ldots$ in such a way that relation (6) is satisfied with $\varepsilon=\frac{1}{k^2}$ for $n \geqq r_k$, and the relations

$$P\left(|S_{r_k}| > \sqrt{B_{r_k} \log\log\log B_{r_k}}\right) < \frac{1}{k^2}, \tag{8}$$

$$1 - \Phi\left(\sqrt{\log\log\log B_{r_k}}\right) < \frac{1}{k^2} \tag{8'}$$

also hold. (Here again we apply the Lemma.) One can construct a sequence of Wiener processes

$$W^k(t), \quad 0 \leqq t \leqq B_{r_{k+1}} - B_{r_k}, \quad k = 1, 2, \ldots$$

which satisfy

$$P\left(\sup_{B_{r_k}<t<B_{r_{k+1}}} \frac{|S(t)-S(B_{r_k})| - W(t-B_{r_k})}{\sqrt{t \log\log t}} > \frac{1}{k^2} \right) < \frac{1}{k^2}. \tag{6'}$$

We may assume that the processes $W^k(t)$, $k=1, 2, \ldots$ are independent. Let $W^0(t)$, $0<t<B_{r_1}$ be an arbitrary Wiener process, independent of the $W^k(t)$-s, $k=1, 2, \ldots$. Define $W(t)$, $t\geqq 0$ as

$$W(t)=\sum_{j=0}^{k-1} W^j(B_{r_{j+1}}-B_{r_j})+W^k(t-B_{r_k}) \quad \text{if} \quad B_{r_k}\leqq t<B_{r_{k+1}}.$$

We claim that this $W(t)$ satisfies the Theorem. Let us first observe that

$$\frac{S(B_{r_k})}{\sqrt{B_{r_k}\log\log B_{r_k}}}\to 0, \quad \frac{W(B_{r_k})}{\sqrt{B_{r_k}\log\log B_{r_k}}}\to 0 \quad \text{with probability 1}$$

because of (8), (8′) and the Borel—Cantelli lemma. Thus

$$\frac{S(B_{r_k})-W(B_{r_k})}{\sqrt{B_{r_k}\log\log B_{r_k}}}\to 0 \quad \text{with probability 1.}$$

On the other hand using again the Borel—Cantelli lemma and relation (6′) one gets that

$$\lim_{k\to\infty}\sup_{B_{r_k}\leqq t<B_{r_{k+1}}}\frac{|S(t)-S(B_{r_k})-W(t)-W(B_{r_k})|}{\sqrt{t\log\log t}}=0 \quad \text{with probability 1.}$$

These last two relations imply the theorem.

REFERENCES

[1] PETROV, V. V.: *On sums of independent random variables* (in Russian), Moscow, Nauka, 1972.
[2] STRASSEN, V.: An invariance principle for the law of iterated logarithm, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **3** (1964), 211—226.
[3] FELLER, W.: Limit theorems for probabilities of large deviations, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **14** (1969), 1—20.
[4] FELLER, W.: *An Introduction to Probability Theory and its Applications, Volume II*, John Wiley and Sons, Inc. New York. London. Sydney 1966.

Mathematical Institute of Hungarian Academy of Sciences

(Received April 13, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 169—184

HOLOÏDES FACTORIELS

par

J.-E. PIN

Abstract

Let H be a commutative monoid and suppose that the relation "divide" is an order on H. Then we say that H is an holoid and write $\leqq$ for the relation "divide": $a \leqq b \leftrightarrow \exists x \in H\ ax = b$.

Dubreil, Fuchs, Mitsch and Bosbach studied certain holoids in which every element has a unique factorization (possibly reduced) into irreducible, prime or maximal elements. We give a specific meaning to the words "reduction" and "reduced". Then we study a new family of holoids, called factorial — a concept which generalizes the previous holoids with "unique factorization" —. The most meaningful difference is that we don't suppose any chain condition. However we have again the "good" properties of these holoids: existence of l.c.m., existence of a minimum solution to the equation $ax = b$ in case $a \leqq b$ and prove that result: "If H is factorial, it is factorial too with respect of l.c.m. as a law of composition".

Introduction

Un monoïde commutatif H dans lequel la relation « divise » est une relation d'ordre est appelé un holoïde. (Cf. BOSBACH [1] DUBREIL [6].) C'est l'ordre « naturel » (FUCHS [7] et MITSCH [10]). On notera dans ce cas $\leqq$ la relation « divise »:

$$a \leqq b \Leftrightarrow \exists x \in H \quad ax = b$$

BOSBACH, DUBREIL, FUCHS et MITSCH ont étudié certains demi-groupes (non nécessairement commutatifs) dans lesquels tout élément possède une décomposition unique — éventuellement « réduite » — en produit de facteurs irréductibles premiers ou maximaux [cf. 1, 2, 3, 6, 7, 9, 10]. Nous donnons une signification précise aux mots « réduction » et « irréductible » puis nous étudions un nouveau type d'holoïdes — dits factoriels — notion qui généralise les holoïdes à « décomposition unique » déjà connus. Ces holoïdes factoriels ne vérifient a priori *aucune condition de chaine* ni la règle de simplification. Nous retrouvons néanmoins en partie les « bonnes » propriétés de ces holoïdes: existence du ppcm, existence d'une solution minimum à l'équation $ax = b$ dans le cas $a \leqq b$ (notion proche mais plus faible que celle de résiduel (Dubreil) ou de quotient (Fuchs)), caractérisation des diviseurs d'un élément et enfin le résultat suivant: Si H est factoriel, H est factoriel pour la loi $\vee$ (ppcm).

NOTATIONS. Soit H un holoïde de neutre e. On notera $\leqq$ la relation divise et $a < b$ si $a \leqq b$ et $a \neq b$.

— Si $(x_i)_{i \in I}$ est une famille finie, on note $\prod_{i \in I} x_i$ le produit des x_i. En particulier $\prod_{i \in \emptyset} x_i = e$.

— S'il existe un plus petit majorant m (au sens de la relation $\leqq$) de la famille $(x_i)_{i\in I}$, m est appelé plus petit commun multiple (en abrégé ppcm) de $(x_i)_{i\in I}$.

— S'il existe un plus grand minorant d, d est appelé pgcd de la famille $(x_i)_{i\in I}$.

A quelques nuances près on a repris la terminologie de BOSBACH [1, 2].

§ 1. Le concept de réduction

Soit H un holoïde de neutre e. I désigne un ensemble fini.

DÉFINITION 1. On dit que x est *irréductible* si pour tout I fini

$$x = \prod_{i\in I} x_i \Rightarrow \exists i \in I, \quad x_i = x.$$

DÉFINITION 2. On dit que x est *premier* si pour tout I fini

$$x \leqq \prod_{i\in I} x_i \Rightarrow \exists i \in I, \quad x \leqq x_i.$$

EXEMPLE 1. e n'est pas irréductible car $e = \prod_{i\in\varnothing} x_i$.

EXEMPLE 2. Dans le $\vee$-demi-treillis de la figure 1, p, q et r sont irréductibles; p et q sont premiers mais r ne l'est pas.

REMARQUE. On déduit immédiatement de la définition que tout élément premier est irréductible. Mais la réciproque est en général inexacte (cf. exemple 2). Examinons quelques propriétés élémentaires des irréductibles (cf. BOSBACH [1]).

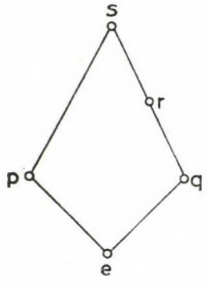


Figure 1

(1) PROPOSITION. *Si x est irréductible et si $a<x$ alors $ax=x$.*

□ En effet si $a<x$ il existe b tel que $x=ab$. Comme $x\neq a$, $x=b$. D'où $ax=x$. □

(2) PROPOSITION. *Les conditions suivantes sont équivalentes:*

i) *x est irréductible.*

ii) *Pour tout I fini, pour toute famille $(x_i)_{i\in I}$*

$$(\forall\, i\in I \;\; x_i < x) \Rightarrow \prod_{i\in I} x_i < x.$$

□ PREUVE. C'est une conséquence immédiate de (1). □

DÉFINITION 3. On appelle *décomposition de x* une famille $D=(x_i)_{i\in I}$ d'éléments irréductibles dont le produit est x.

On utilisera, suivant le contexte, l'une des notations suivantes pour désigner une décomposition D de x:

$$x = \prod_{i\in I} x_i, \quad x = \prod_{D}, \quad x = \prod_{y\in\mathscr{I}} y^{n_y(x)}.$$

Seule la dernière notation demande des explications. L'ensemble indexant $\mathscr{I}$ est l'ensemble des irréductibles de H; $n_y(x)$ est le nombre d'éléments de D égaux à y.

On utilise parfois une notation de ce genre pour écrire la décomposition d'un entier en facteurs premiers.

Soient D et D' deux décompositions de x. D est dite équivalente à D' et on note $D\sim D'$ si D et D' ont les mêmes facteurs. Formellement: $D=(x_i)_{i\in I}\sim D'=(x'_i)_{i\in I'}$ si et seulement si existe une bijection σ de I vers I' telle que pour tout $i\in I$, $x_i=x'_{\sigma(i)}$.

Nous arrivons aux deux définitions les plus importantes.

DÉFINITION 4. Soient $D=(x_i)_{i\in I}$ et $D'=(x'_i)_{i\in I'}$ deux décompositions de x. On dit que D est *plus réduite* que D' et on note DRD' s'il existe une injection σ de I vers I' telle que pour tout $i\in I$: $x_i\leqq x'_{\sigma(i)}$. σ est appelée injection de réduction.

Par abus de notation on notera parfois $\sigma(x_i)$ au lieu de $x'_{\sigma(i)}$. On voit facilement que R est une relation de préordre sur l'ensemble des décompositions de x.

DÉFINITION 5. On dit que H est *factoriel*, lorsque pour tout élément x de H, l'ensemble des décompositions de x a un élément *minimum* qu'on appelle décomposition réduite de x.

N. B. Cette notion généralise les holoïdes « halbprimkanonisch » de Bosbach et les « primfaktorzerlegungen » de Fuchs.

Le résultat suivant éclaire la définition 5.

(3) PROPOSITION. $(DRD'$ *et* $D'RD)\Leftrightarrow D\sim D'$.

En effet, si $D\sim D'$ il est clair que DRD' et $D'RD$. Réciproquement supposons que $D=(x_i)_{i\in I}RD'=(x'_i)_{i\in I'}$ et $D'RD$. Il existe alors des injections de réduction $\sigma\colon I\to I'$ et $\sigma'\colon I'\to I$. Comme I et I' sont finis, σ et σ' sont bijectives. On a pour tout $i\in I$ $x_i\leqq x'_{\sigma(i)}\leqq x_{\sigma'\circ\sigma(i)}$. Posons $\tau=\sigma'\circ\sigma$ et supposons qu'il existe $i_0\in I$ tel que $x_{i_0}<x_{\tau(i_0)}$. Puisque τ est bijective, il existe $n\geqq 1$ tel que $\tau^n(i_0)=i_0$ et donc $x_{i_0}<x_{\tau(i_0)}\leqq\ldots$ $\ldots\leqq x_{\tau^n(i_0)}=x_{i_0}$. Contradiction. Donc $x_i=x'_{\sigma(i)}=x_{\tau(i)}$ pour tout $i\in I$ et $D\sim D'$.

REMARQUES. Soit $D=\prod_{y\in\mathscr{I}} y^{n_y(x)}$ une décomposition réduite de x. Soient $y_0<y$ deux irréductibles. Alors $n_{y_0}(x)=0$ ou $n_y(x)=0$. Autrement dit deux irréductibles distincts et comparables ne peuvent figurer simultanément dans une décomposition réduite.

Il résulte de [3] que deux décompositions réduites de x ont les mêmes facteurs à l'ordre près. Dans un holoïde factoriel on parlera donc de *la* décomposition réduite d'un élément (qui n'est définie en fait qu'à l'ordre près des facteurs).

Voici un critère permettant de comparer deux décompositions mises sous forme exponentielle.

(4) THÉORÈME. *Pour que* $D=\prod_{y\in\mathscr{I}} y^{n_y(a)}$ *soit plus réduite que* $D'=\prod_{y\in\mathscr{I}} y^{n_y(b)}$ *il faut et il suffit que, pour toute partie* H *de* $\mathscr{I}$, *on ait*

$$\sum_{y\in H} n_y(a)\leqq\sum_{\substack{y'\geqq y\\ y\in H}} n_{y'}(b).$$

□ Cela résulte du lemme des mariages. Soit A l'application

$$I\to\mathscr{P}(I'),\quad i\to A(i)=\{j\,|\,x_i\leqq x'_j\}.$$

DRD' si et seulement si il existe une injection $\sigma\colon I\to I'$ telle que pour tout $i\in I$, $\sigma(i)\in A(i)$.

D'après le lemme des mariages il faut et il suffit que, pour toute partie K de I

$$\mathrm{Card}\left(\bigcup_{i\in K} A(i)\right) \geqq \mathrm{Card}\, K.$$

Posons $\overline{K}=\{i\in I,\ \exists j\in K,\ x_i=x_j\}$. Il est clair que $\mathrm{Card}\,\overline{K}\geqq\mathrm{Card}\,K$ et que $\mathrm{Card}\left(\bigcup_{i\in K} A(i)\right)=\mathrm{Card}\left(\bigcup_{i\in \overline{K}} A(i)\right)$. Donc $DRD'\Leftrightarrow$ pour toute partie K de I $\mathrm{Card}\left(\bigcup_{i\in K} A(i)\right)\geqq\mathrm{Card}\,\overline{K}$ ce qui n'est rien d'autre qu'une formulation différente du théorème. □

Ce théorème permettrait de démontrer par le calcul certains des énoncés des § 2, 3, 4 et 5. Donnons tout de suite un résultat simple, mais utile:

(5) Proposition. *Soit $x=\prod_{y\in\mathscr{I}} y^{n_y(x)}$ une décomposition réduite de x. Alors toute décomposition $a=\prod_{y\in\mathscr{I}} y^{n_y(a)}$ — avec, pour tout $y\in\mathscr{I}$, $n_y(a)\leqq n_y(x)$ — est une décomposition réduite de a.*

□ La démonstration est immédiate. □

Nous allons maintenant donner une caractérisation des décompositions réduites. Pour cela nous aurons besoin d'une proposition.

(6) Proposition. *Si $D=(x_i)_{i\in I}$ est la décomposition réduite de x, si $D'=(x'_i)_{i\in I'}$ est une décomposition quelconque de x, il existe une injection de réduction σ de D dans D' telle que $x'_{\sigma(i)}=x'_{\sigma(j)}\Rightarrow x_i=x_j$.*

□ On procède par récurrence sur $\mathrm{Card}\,D'=n$. C'est évident pour $n=0$ ou 1. Supposons le résultat acquis jusqu'à $n-1$. Soit σ une injection de réduction de D dans D' et supposons que $x'_{\sigma(i_1)}=x'_{\sigma(i_2)}$ avec $x_{i_1}\neq x_{i_2}$. Puisque D est réduite x_{i_1} et x_{i_2} sont incomparables (cf. la remarque suivant (3)) et donc $x_{i_1}<x'_{\sigma(i_1)}$, $x_{i_2}<x'_{\sigma(i_2)}=x'_{\sigma(i_1)}$ d'où $x_{i_1}x_{i_2}<x'_{\sigma(i_1)}$ d'après (2). On a donc:

$$x=\prod_{i\in I} x_i = x_{i_1}x_{i_2}\prod_{i\in I-\{i,i_2\}} x_i \leqq x'_{\sigma(i_1)}\prod_{i\in I-\{i_1,i_2\}} x'_{\sigma(i)} = \prod_{i\in I-\{i_2\}} x'_{\sigma(i)} \leqq x$$

et $D'-\{x'_{\sigma(i_2)}\}$ est encore une décomposition de x. Or $\mathrm{Card}\,(D'-\{x'_{\sigma(i_2)}\})=n-1$ et l'hypothèse de récurrence permet de conclure facilement. □

Avant d'énoncer le théorème précisons une terminologie: Si σ est une injection de réduction de $D=(x_i)_{i\in I}$ dans $D'=(x'_i)_{i\in I'}$ on appelle image dans $\mathscr{I}$ de σ le sous-ensemble de $\mathscr{I}$: $\{x'_{\sigma(i)}\ i\in I\}$.

(7) Théorème. (Caractérisation des décompositions réduites.) *Pour que $x=\prod_y y^{n_y(x)}=D$ soit la décomposition réduite de x, il faut et il suffit que pour toute décomposition D' de x il existe, pour chaque $y\in\mathscr{I}$, des injections de réduction σ_y de $y^{n_y(x)}$ dans D', d'images dans $\mathscr{I}$ deux à deux disjointes.*

□ La condition est suffisante: on peut construire une injection de réduction de D dans D' en « recollant » les σ_y.

La condition est nécessaire: d'après (6) il existe une injection de réduction σ de D dans D' telle que

$$(\alpha)\qquad x'_{\sigma(i)}=x'_{\sigma(j)}\Rightarrow x_i=x_j.$$

Pour chaque $y\in\mathscr{J}$, σ induit une injection de réduction σ_y de $y^{n_y(x)}$ dans D'. Les images des σ_y dans $\mathscr{J}$ sont deux à deux disjointes d'après (α). □

Ce résultat nous sera très utile au paragraphe 4 pour la démonstration du théorème fondamental (27).

Avant de terminer ce paragraphe, donnons un exemple d'holoïde factoriel. Il s'agit du $\vee$-demi-treillis de la figure 2 ci-dessous. Les éléments irréductibles sont a, b les x_n et les y_n. Il n'y a que 2 éléments premiers: a et y_1. En effet y_2, par exemple, n'est pas premier car $y_2\leqq ay_1=x$ mais $y_2\nleqq a$, $y_2\nleqq y_1$. La décomposition réduite de x est $x=ay_1$. Les éléments minimaux sont a et y_1.

Cet holoïde ne vérifie ni la condition de chaîne ascendante ni la condition de chaîne descendante: c'est là une différence essentielle avec les holoïdes étudiés par Bosbach ou avec les demi-groupes à décomposition unique en facteurs premiers de Fuchs et Dubreil—Jacotin.

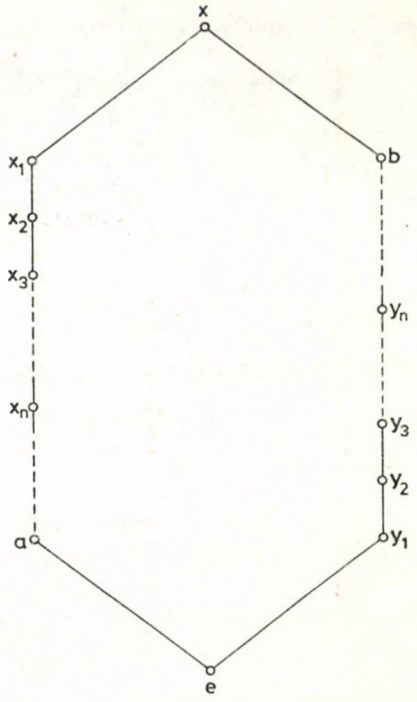


Figure 2

§ 2. Diviseurs d'un élément dans un holoide factoriel

En voici une première caractérisation:

(8) THÉORÈME. (Première caractérisation des diviseurs d'un élément.) *Soit un holoïde factoriel. Soit $x\leqq z$ et $z=\prod_{y\in\mathscr{J}} y^{n_y(z)}$ la décomposition réduite de z. Alors $x=x_1x_2$ où x_1 est absorbé par z et où x_2 admet une décomposition réduite de la forme $x_2=\prod_{y\in\mathscr{J}} y^{n_y(x_2)}$ avec, pour tout $y\in\mathscr{J}$, $n_y(x_2)\leqq n_y(z)$.*

Soit $x=\prod_{y\in\mathscr{J}} y^{n_y(x)}=D_1$ une décomposition de x. Posons

$$\begin{cases} x_1=\prod_{y\in\mathscr{J}} y^{n_y(x_1)}, \\ x_2=\prod_{y\in\mathscr{J}} y^{n_y(x_2)}, \end{cases} \quad \text{où} \quad \begin{cases} n_y(x_1)=\big(n_y(x)-n_y(z)\big)^+, \\ n_y(x_2)=\inf\big(n_y(x),\, n_y(z)\big). \end{cases}$$

En vertu de (5) x_2 vérifie bien les conditions de l'énoncé. De plus $x_1x_2=x$. Reste à montrer que $x_1z=z$. Soit a tel que $ax=z$ et soit $\prod_{y\in\mathscr{J}} y^{n_y(a)}=D_2$ une décomposition de a. On a:

$$\prod_{y\in\mathscr{J}} y^{n_y(z)}=\prod_{y\in\mathscr{J}} y^{n_y(a)+n_y(x)}=\prod_{D_1\dot{\cup}D_2}$$

où $D_1\dot{\cup}D_2$ désigne l'union disjointe des familles D_1 et D_2.

Mais puisque $\prod_{y\in\mathscr{J}} y^{n_y(z)}$ est réduite, il existe une injection de réduction σ de $\prod_{y\in\mathscr{J}} y^{n_y(z)}$ dans $\prod_{D_1\cup D_2} = \prod_{y\in\mathscr{J}} y^{n_y(x)+n_y(a)}$. Posons

$$z_1 = \prod_{\sigma^{-1}(D_2)}, \quad z_2 = \prod_{\sigma^{-1}(D_1)\cap\{y\mid\sigma(y)=y\}}, \quad z_3 = \prod_{\sigma^{-1}(D_1)\cap\{y\mid\sigma(y)>y\}}.$$

On a

$$z \leqq zx_1 = z \prod_{y\in\mathscr{J}} y^{[n_y(x)-n_y(z)]^+} \leqq z \prod_{y\in\mathscr{J}} y^{(n_y(x)-n_y(z_2))}.$$

En effet $n_y(z_2)\leqq n_y(z)$ et donc

$$(n_y(x)-n_y(z))^+ \leqq (n_y(x)-n_y(z_2))^+ = n_y(x)-n_y(z_2)$$

car $n_y(x)\geqq n_y(z_2)$. Comme $z=z_1z_2z_3$ on obtient

$$z \leqq z_1z_2z_3 \prod y^{[n_y(x)-n_y(z_2)]}.$$

Or d'après (1) z_3 est absorbé par $\prod_{y\in\mathscr{J}} y^{n_y(x)-n_y(z_2)}$. Donc

$$z \leqq zx_1 \leqq z_1z_2 \prod_{y\in\mathscr{J}} y^{[n_y(x)-n_y(z_2)]} \leqq z_1x \leqq ax = z.$$

Donc $zx_1=z$. □

Voici quelques conséquences de ce théorème.

(9) COROLLAIRE 1. *Soit y_0 un irréductible et soit $z=\prod_{y\in\mathscr{J}} y^{n_y(z)}$ la décomposition réduite de z. On suppose que $y_0^m\leqq z$. Alors $m\leqq n_{y_0}(z)$ ou bien z absorbe y_0.*

□ Reprenons la démonstration de (8) avec $x=y_0^m$. Si $m>n_{y_0}(z)$ alors $x_1=y_0^{m-n_{y_0}(z)}$ est absorbé par z; donc y_0 et par conséquent y_0^m sont absorbés par z. □

(10) COROLLAIRE 2. *Soit $a=\prod_{y\in\mathscr{J}} y^{n_y(a)}$ une décomposition de a (pas nécessairement réduite). Pour montrer que $a\leqq x$ il suffit de montrer que $y^{n_y(a)}\leqq x$ pour tout $y\in\mathscr{J}$.*

□ C'est une conséquence immédiate du corollaire 1. □

Citons encore un corollaire qui situe bien la différence entre éléments irréductible et premier.

(11) COROLLAIRE 3. *Soit y_0 un irréductible. Si $y_0\leqq ab$, alors $y_0\leqq a$, $y_0\leqq b$ ou y_0 est absorbé par ab.*

□ Soient

$$ab = \prod_{y\in\mathscr{J}} y^{n_y(ab)}, \quad a = \prod_{y\in\mathscr{J}} y^{n_y(a)}, \quad b = \prod_{y\in\mathscr{J}} y^{n_y(b)}$$

les décompositions réduites de ab, a et b respectivement. D'après (9) y_0 est absorbé par ab, ou $n_{y_0}(ab)\geqq 1$. Plaçons-nous dans ce dernier cas:

$\prod_{y\in\mathscr{J}} y^{n_y(ab)}$ est plus réduite que $\prod_{y\in\mathscr{J}} y^{n_y(a)+n_y(b)}$. D'après (4), appliqué à $H=\{y_0\}$,

$$1 \leqq n_{y_0}(ab) \leqq \sum_{y\equiv y_0} (n_y(a)+n_y(b)).$$

Donc $\sum_{y\equiv y_0} n_y(a)\geqq 1$ ou $\sum_{y\equiv y_0} n_y(b)\geqq 1$ et $y_0\leqq a$ ou $y_0\leqq b$. □

Enfin on a le

(12) COROLLAIRE 4. *Soient y_1 et y_2 deux irréductibles. Si $y_1^{n_1}=y_2^{n_2}$ $(n_1, n_2\in\mathbf{N}^*)$ alors $y_1=y_2$.*

□ En effet soit $z=y_1^{n_1}=y_2^{n_2}$ et D la décomposition réduite de z. Supposons $y_1\neq y_2$. De (10) on déduit alors $y_1^{n_1}y_2^{n_2}\leqq z$, d'où $y_1z=z$ et $y_2z=z$ et par conséquent ni y_1 ni y_2 ne figurent dans D. Soit σ l'injection de réduction de D dans $y_1^{n_1}$. Pour tout $x\in D$ on a $x\leqq\sigma(x)=y_1$ donc $x<y_1$ (puisque y_1 ne figure pas dans D). On en déduit d'après (2) $z=\prod_D <y_1\leqq y_1^{n_1}=z$. Contradiction. Donc $y_1=y_2$.

§ 3. Associés minima

Nous introduisons maintenant une notion proche — *mais distincte* comme on va le voir — de la notion de « quotient » (FUCHS [7]) ou de « résiduel » (DUBREIL [6]). La terminologie est due à BOSBACH [2].

DÉFINITION 6. Soit $x\leqq z$ et soit A l'ensemble des a tels que $ax=z$ (appelés associés de x dans z). On appelle *associé minimum de x dans z* un élément minimum de A relativement à $\leqq$. Si cet élément existe on le note $z:x$.

Si a est le « quotient » (au sens de Fuchs) de z par x alors on a l'éqiuvalence $a\leqq t\Leftrightarrow z\leqq xt$. On en déduit facilement que a est l'associé minimum de x dans z mais *la réciproque n'est pas vraie* ainsi que le montre l'exemple 2: r est l'associé minimum de q dans r, $r\leqq qp$ mais $q\not\leqq p$.

On sait (DUBREIL [6] partie 2 chap. 5, FUCHS [7] chap. 12), que dans un holoïde à décomposition en facteurs premiers unique (mit eindeutigen Primfaktorzerlegungen) il existe des résiduels (Quotienten). Voici un résultat analogue pour les holoïdes factoriels.

(13) THÉORÈME. *Soit H un holoïde factoriel. Soient $x\leqq z$ et $x=\prod_{y\in\mathscr{P}} y^{n_y(x)}$ $z=\prod_{y\in\mathscr{P}} y^{n_y(z)}$ les décompositions réduites de x et z. Alors $z:x$ existe et sa décomposition réduite est $\prod_{y\in\mathscr{P}} y^{n_y(z:x)}$ avec $n_{y_0}(z:x)=0$ s'il existe $y>y_0$ tel que $n_y(x)>0$, $n_{y_0}(z:x)=(n_{y_0}(z)-n_{y_0}(x))^+$ sinon.*

Posons $x'=\prod_{y\in\mathscr{P}} y^{n_y(x')}$ avec $n_{y_0}(x')=0$ s'il existe $y>y_0$ tel que $n_u(x)>0$, $n_{y_0}(x')=(n_{y_0}(z)-n_{y_0}(x))^+$ sinon.

On va montrer $z\leqq xx'$ puis $xx'\leqq z$ et enfin $x'=z:x$.

a) $z\leqq xx'$.

D'après (10) il suffit de montrer que, pour tout $y_0\in\mathscr{P}$ $y_0^{n_{y_0}(z)}\leqq xx'$.

— S'il existe $y>y_0$ tel que $n_y(x)>0$, y_0 est absorbé par x d'après (1) donc par xx' et c'est démontré.

— Sinon $n_{y_0}(x')=(n_{y_0}(z)-n_{y_0}(x))^+$ et $n_{y_0}(z)\leqq n_{y_0}(x)+n_{y_0}(x')$ donc $y_0^{n_{y_0}(z)}\leqq xx'$.

b) $xx'\leqq z$.

D'après (10) il suffit de prouver que, pour tout $y_0\in\mathscr{P}$ $y_0^{n_{y_0}(x)+n_{y_0}(x')}\leqq z$.

— S'il existe $y>y_0$ tel que $n_y(x)>0$ alors $n_{y_0}(x')=0$ et $n_{y_0}(x)=0$ d'après la remarque suivant (3). Donc $0=n_{y_0}(x)+n_{y_0}(x')\leqq n_{y_0}(z)$.

— Sinon $n_{y_0}(x') = (n_{y_0}(z) - n_{y_0}(x))^+$ d'où $n_{y_0}(x) + n_{y_0}(x') = \max(n_{y_0}(x), n_{y_0}(z))$. Si $n_{y_0}(x) \leqq n_{y_0}(z)$ c'est terminé et si $n_{y_0}(x) > n_{y_0}(z)$, (11) appliqué à l'inégalité $y_0^{n_{y_0}(x)} \leqq z$ montre que y_0 est absorbé par z et donc: $y_0^{n_{y_0}(x)+n_{y_0}(x')} \leqq z$. D'où $xx' = z$.

c) $x' = z:x$.

Supposons que $ax = z$ et soit $a = \prod_{y \in \mathscr{I}} y^{n_y(a)}$ la décomposition réduite de a. Il s'agit de montrer que $x' \leqq a$. Là encore il suffira de prouver que $y_0^{n_{y_0}(x')} \leqq a$ pour tout $y_0 \in \mathscr{I}$. Le seul cas à étudier est celui où $n_{y_0}(x') > 0$: on a donc $\sum_{y > y_0} n_y(x) = 0$. Appliquons (4) avec $H = \{y_0\}$ à $\prod_{y \in \mathscr{I}} y^{n_y(z)} \, R \prod_{y \in \mathscr{I}} y^{n_y(x)+n_y(a)}$. Il vient:

$$n_{y_0}(z) \leqq \sum_{y \geqq y_0} (n_y(a) + n_y(x)) = n_{y_0}(x) + \sum_{y \geqq y_0} n_y(a).$$

D'où

$$n_{y_0}(x') = (n_{y_0}(z) - n_{y_0}(x))^+ \leqq \sum_{y \geqq y_0} n_y(a).$$

D'après (4) $y_0^{n_{y_0}(x')} \, R \prod_{y \in \mathscr{I}} y^{n_y(a)}$ et donc $y_0^{n_{y_0}(x')} \leqq a$.

Enfin la décomposition de x' est réduite, d'après (5). □

REMARQUES. Les résultats suivants sont *faux* en général

— $(ax:x) = a$ prendre $a = x = x^2 \neq e$, $x^2:x = x:x = e$.

— Si $(x:a) = b$, $(x:b) = a$. Dans l'exemple 2 $s:r = p$ mais $s:p = q$.

— $z(y:x) = zy:x$. Dans l'exemple 2 $r(p:p) = r$, $rp:p = s:p = q$ (cependant on a toujours $z(y:x) \geqq zy:x$).

— $(a:x_1x_2) = (a:x_1):x_2 = (a:x_2):x_1$ (prendre $a = a^2 = x_1 = x_2$).

Le premier membre existe, mais ni le second, ni le troisième n'ont de sens.

— Si $x \leqq a \leqq b$ alors $(a:x) \leqq (b:x)$. Considérons en effet le ∨-demi-treillis représenté par la figure 3.

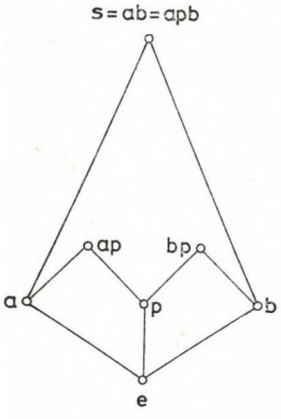


Figure 3

On a $a \leqq ap \leqq s$, $ap:a = p$, $s:a = b$, mais p et b sont incomparables.

En revanche on a le résultat suivant:

(14) PROPOSITION. *Si* $a \leqq b \leqq x$ *alors* $(x:b) \leqq (x:a)$.

□ Soient

$$x = \prod_{y \in \mathscr{I}} y^{n_y(x)}, \quad a = \prod_{y \in \mathscr{I}} y^{n_y(a)}, \quad b = \prod_{y \in \mathscr{I}} y^{n_y(b)}$$

les décompositions réduites de x, a et b respectivement. D'après (10) il suffit de prouver que pour tout $y_0 \in \mathscr{I}$, $y_0^{n_{y_0}(x:b)} \leqq x:a$. D'après (13) le seul cas où $n_{y_0}(x:b) \neq 0$ est celui où

$$\sum_{y > y_0} n_y(b) = 0, \quad n_{y_0}(x) > n_{y_0}(b).$$

Dans ce cas b n'absorbe pas y_0 car sinon on aurait

$$b(x:b) = x = b \prod_{\substack{y \in \mathscr{I} \\ y \neq y_0}} y^{n_y(x:b)} \quad \text{d'où} \quad n_{y_0}(x:b) = 0.$$

Par conséquent y_0 n'est pas absorbé par b — ni par a — et $\sum_{y>y_0} n_y(a)=0$. (9), appliqué à l'inéqualité $y_0^{n_{y_0}(a)} \leqq b$ montre que $n_{y_0}(a) \leqq n_{y_0}(b)$. Il vient alors:

$$0 < n_{y_0}(x)-n_{y_0}(b) \leqq n_{y_0}(x)-n_{y_0}(a) \leqq (n_{y_0}(x)-n_{y_0}(a))^+.$$

Comme $\sum_{y>y_0} n_y(a)=0$, $n_{y_0}(x:a)=(n_{y_0}(x)-n_{y_0}(a))^+$. On a donc montré que $n_{y_0}(x:b) \leqq$ $\leqq n_{y_0}(x:a)$ ce qui achève la démonstration. □

Nous allons maintenant caractériser les associés minima des différents diviseurs de z, qu'on appelle également *diviseurs minima* de z.

(15) PROPOSITION. *Soit* $z=\prod_{y\in\mathscr{S}} y^{n_y(z)}$ *la décomposition réduite de* z. *Les diviseurs minima de* z *sont les éléments* a *dont la décomposition réduite est de la forme* $a=\prod_{y\in\mathscr{S}} y^{n_y(a)}$ *avec* $n_y(a) \leqq n_y(z)$ *pour tout* $y\in\mathscr{S}$.

□ D'après (13) les diviseurs minima sont tous de cette forme.

Réciproquement soit a un élément de la forme indiquée ci-dessus. Posons:

$$x = \prod_{y\in\mathscr{S}} y^{n_y(x)} \quad \text{avec} \quad n_y(x) = n_y(z)-n_y(a).$$

Il est clair que $x \leqq z$. On va montrer que $z:x=a$.

Soit $y_0\in\mathscr{S}$ — S'il existe $y>y_0$ tel que $n_y(z)>0$ alors $n_{y_0}(z)=0$ d'après la remarque suivant (3) et $n_{y_0}(a)=0$ d'après l'hypothèse. Donc $n_{y_0}(z:x)=0=n_{y_0}(a)$.

— Sinon

$$n_{y_0}(z:x) = [n_{y_0}(z)-[n_{y_0}(z)-n_{y_0}(a)]]^+ = n_{y_0}(a).$$

On conclut à l'aide de (10). □

De (8) et (15) on déduit la seconde caractérisation des diviseurs d'un élément.

(16) THÉORÈME. *Soit* $z\in H$. *Tout diviseur de* z *est produit d'un diviseur minimum de* z *et d'un élément absorbé par* z. *Réciproquement tout élément qui se factorise de cette manière est un diviseur de* z.

§ 4. P.P.C.M.

On notera $\bigvee_{i=1}^{n} x_i$ le plus petit commun multiple (p.p.c.m.) d'une famille d'éléments $(x_i)_{i=1}^n$ — s'il existe. — On sait (cf. DUBREIL [6] et FUCHS [7]) que dans un holoïde à décomposition en facteurs premiers unique le p.p.c.m. existe. Comme on va le voir, ce résultat est conservé dans les holoïdes factoriels.

(17) THÉORÈME. *Soit* $(x_i)_{i=1}^m$ *une famille finie d'éléments d'un holoïde factoriel* H. *Soit* $x_i=\prod_{y\in\mathscr{S}} y^{n_y(x_i)}$ *une décomposition — pas nécessairement réduite — de* x_i. *Alors le p.p.c.m. des* $(x_i)_{i=1}^m$ *existe et on a:*

$$\bigvee_{i=1}^{n} x_i = \prod_{y\in\mathscr{S}} y^{\max_{i=1}^m n_y(x_i)}$$

(décomposition non réduite en général).

□ Posons $z=\prod_{y\in\mathscr{I}} y^{\max_{i=1}^{m} n_y(x_i)}$. Il est clair que $x_i\leqq z$ pour $1\leqq i\leqq m$.

Réciproquement soit $a\geqq x_i$ pour $1\leqq i\leqq m$ et soit $a=\prod_{y\in\mathscr{I}} y^{n_y(a)}$ la décomposition réduite de a. Soit $y_0\in\mathscr{I}$. Puisque $x_i\leqq a$, on a d'après (9).

— Soit $n_{y_0}(x_i)\leqq n_{y_0}(a)$ pour $i=1, \ldots, m$ et donc $\max_{i=1}^{m} n_{y_0}(x_i)\leqq n_{y_0}(a)$.

— Soit a absorbe y_0 et donc également $y_0^{\max_{i=1}^{m} n_{y_0}(x_i)}$. Dans les deux cas $y_0^{\max_{i=1}^{m} n_{y_0}(x_i)}\leqq a$. Donc $z\leqq a$ d'après (10). □

REMARQUE. En revanche deux éléments n'ont pas toujours de p.g.c.d. dans un holoïde factoriel. Considérons en effet le $\vee$-demi-treillis représenté par la figure 4 (x et y n'ont pas de p.g.c.d.).

Outre les propriétés classiques (commutativité, associativité, idempotence), $\vee$ possède une propriété de distributivité.

(18) PROPOSITION. $z(\bigvee_{i=1}^{n} x_i)=\bigvee_{i=1}^{n}(zx_i)$.

Soient $z=\prod_{y\in\mathscr{I}} y^{n_y(z)}$ et $x_i=\prod_{y\in\mathscr{I}} y^{n_y(x_i)}$ des décompositions de z et x_i. On a:

$$z\left(\bigvee_{i=1}^{n} x_i\right)=\prod_{y\in\mathscr{I}} y^{(n_y(z)+\max_{i=1}^{m} n_y(x_i))}=\prod_{y\in\mathscr{I}} y^{\max_{i=1}^{m}(n_y(z)+n_y(x_i))}=\bigvee_{i=1}^{n}(zx_i).$$

Voici une autre propriété du p.p.c.m.

(19) PROPOSITION. Si $m=\bigvee_{i=1}^{n} x_i$, $m^q=\bigvee_{i=1}^{n} x_i^q$.

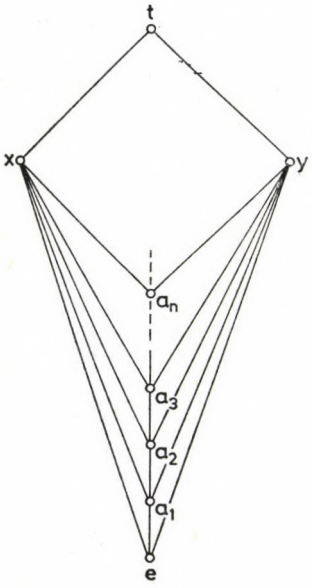


Figure 4

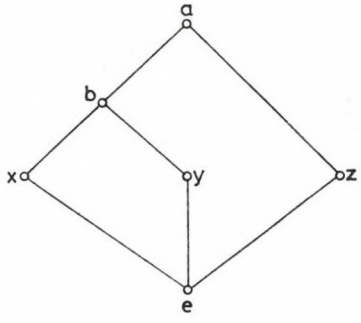


Figure 5

Preuve: c'est évident à l'aide de (17).

Nous allons maintenant examiner les relations entre p.p.c.m. et diviseurs minima: ce sera l'objet des propositions qui suivent.

(20) PROPOSITION. *Soit* $(x_i)_{i=1}^n$ *une famille d'éléments de* H, m *leur p.p.c.m. Soient* $(m:x_i)=\prod_{y\in\mathscr{J}} y^{n_y(m:x_i)}$ *les décompositions réduites des* $m:x_i$. *Alors pour tout* $y_0\in\mathscr{J}$, $\prod_{i=1}^m n_{y_0}(m:x_i)=0$. *De plus cette condition caractérise le p.p.c.m. parmi les multiples communs aux* x_i.

On a: $\prod_{y\in\mathscr{J}} y^{n_y(m)}\ R\ \prod_{y\in\mathscr{J}} y^{\max_{i=1}^n n_y(x_i)}$. En appliquant (4) à $H=\{y_0\}$ on obtient $n_{y_0}(m)\leqq\sum_{y\geqq y_0}\max_{i=1}^n n_y(x_i)$.

— S'il existe $y>y_0$ tel que $\max_{i=1}^n n_y(x_i)>0$, il existe i_0 tel que $n_y(x_{i_0})>0$ et donc $n_{y_0}(m:x_{i_0})=0$.

— Sinon $n_{y_0}(m)\leqq\max_{i=1}^n n_{y_0}(x_i)$. Il existe i_0 tel que $n_{y_0}(m)=n_{y_0}(x_{i_0})$ et donc $n_{y_0}(m:x_{i_0})=0$.

Supposons maintenant $x_i\leqq m'$ pour tout i et $\prod_{i=1}^n n_{y_0}(m':x_i)=0$ pour tout $y_0\in\mathscr{J}$. Soit $m'=\prod_{y\in\mathscr{J}} y^{n_y(m')}$ la décomposition réduite de m'. Soit i_0 tel que $n_{y_0}(m':x_{i_0})=0$.

Premier cas. Il existe $y>y_0$ avec $n_y(x_{i_0})>0$. Alors y_0 est absorbé par x_{i_0} et donc par m. Donc $y_0^{n_{y_0}(m')}\leqq m$.

Deuxième cas. $\sum_{y>y_0} n_y(x_{i_0})=0$.

$$0=n_{y_0}(m':\ x_{i_0})=(n_{y_0}(m')-n_{y_0}(x_{i_0}))^+$$

donc

$$n_{y_0}(m')\leqq n_{y_0}(x_{i_0})\leqq\max_{i=1}^n n_{y_0}(x_i)\quad\text{et}\quad y_0^{n_{y_0}(m')}\leqq m.$$

On conclut à l'aide de (10) que $m'\leqq m$ et donc $m'=m$ par définition du p.p.c.m.

(21) PROPOSITION. *Soient* x_1 *et* x_2 *des éléments de* H, m *leur p.p.c.m. Alors* $(m:x_1)x_1=m$, $(m:x_2)x_2=m$ *et* $(m:x_1)\vee(m:x_2)=(m:x_1)(m:x_2)$.

Les deux premières égalités resultent uniquement de la définition 6. La troisième égalité résulte de la proposition (20). En effet, on a pour tout $y_0\in\mathscr{J}$, $n_{y_0}(m:x_1)=0$ ou $n_{y_0}(m:x_2)=0$. D'où

$$n_{y_0}(m:\ x_1)+n_{y_0}(m:\ x_2)=\max\left(n_{y_0}(m:\ x_1),\ n_{y_0}(m:\ x_2)\right)$$

et donc $(m:x_1)(m:x_2)=(m:x_1)\vee(m:x_2)$.

REMARQUE. Si $x\leqq a$ et $x\leqq b$ on n'a pas en général $(a:x)\vee(b:x)=(a\vee b):x$. En effet considérons le $\vee$-demi-treillis représenté par la figure 5:

$$a:\ x=z\quad\text{et}\quad b:\ x=y,\quad z\vee y=a\quad\text{mais}\quad(a\vee b):\ x=a:\ x=z.$$

Nous introduisons maintenant la notion d'adjoint. C'est l'analogue du « a^{Δ} » de FUCHS [7 page 256]. On trouvera plus loin des propriétés voisines du « a^{Δ} ».

DÉFINITION 7. Soit $x \leqq z$. On appelle *adjoint de x dans z* l'élément $\bar{x} = z:(z:x)$. Remarquons tout d'abord ceci: $(x:z) = z$ donc $\bar{x} = z:(z:x) \leqq x$.

L'objet de la proposition suivante est le calcul de $\bar{x}$, $x:\bar{x}$ et $z:\bar{x}$. Soient

$$x = \prod_{y \in \mathcal{P}} y^{n_y(x)}, \quad z = \prod_{y \in \mathcal{P}} y^{n_y(z)}, \quad \bar{x} = \prod_{y \in \mathcal{P}} y^{n_y(\bar{x})},$$

$$(x:\bar{x}) = \prod_{y \in \mathcal{P}} y^{n_y(x:\bar{x})} \quad (z:\bar{x}) = \prod_{y \in \mathcal{P}} y^{n_y(z:\bar{x})}$$

les décompositions réduites de x, z, $\bar{x}$, $x:\bar{x}$, et $z:\bar{x}$ respectivement

(22) PROPOSITION.

a) $\begin{cases} n_{y_0}(\bar{x}) = n_{y_0}(z) & \text{s'il existe} \quad y > y_0 \quad \text{tel que} \quad n_y(x) > 0, \\ n_{y_0}(\bar{x}) = \operatorname{Min}(n_{y_0}(x), n_{y_0}(z)) & \text{sinon,} \end{cases}$

b) $n_{y_0}(x:\bar{x}) = (n_{y_0}(x) - n_{y_0}(z))^+$,

c) $n_{y_0}(z:\bar{x}) = n_{y_0}(z:x)$.

□ a) Tout d'abord supposons $n_{y_0}(z)=0$. Alors d'après (13) $n_{y_0}(z:x)=0$ et $n_{y_0}(\bar{x})=0$: a) est vérifié.

Supposons $n_{y_0}(z)>0$. Alors $n_y(z)=0$ pour $y>y_0$ et donc $n_y(z:x)=0$ pour $y>y_0$. On distingue alors deux cas.

— Il existe $y>y_0$ tel que $n_y(x)>0$; alors $n_{y_0}(z:x)=0$ d'après (13) et $n_{y_0}(\bar{x}) = (n_{y_0}(z)-0)^+ = n_{y_0}(z)$ toujours d'après (13).

— $n_y(x)=0$ pour $y>y_0$ d'où $n_{y_0}(z:x) = (n_{y_0}(z) - n_{y_0}(x))^+$.

$$n_{y_0}(\bar{x}) = (n_{y_0}(z) - (n_{y_0}(z) - n_{y_0}(x))^+)^+ = \begin{cases} n_{y_0}(x) & \text{si} \quad n_{y_0}(z) \geqq n_{y_0}(x), \\ n_{y_0}(z) & \text{si} \quad n_{y_0}(x) \geqq n_{y_0}(z), \end{cases}$$

$$n_{y_0}(\bar{x}) = \operatorname{Min}(n_{y_0}(x), n_{y_0}(z)).$$

b) Si $n_{y_0}(x)=0$ la formule est évidente.

Si $n_{y_0}(x)>0$, alors $n_y(x)>0$ pour $y>y_0$ donc $n_{y_0}(\bar{x}) = \operatorname{Min}(n_{y_0}(x), n_{y_0}(z))$ et $n_{y_0}(\bar{x})=0$ pour $y>y_0$ d'après a). Par conséquent

$$n_{y_0}(x:\bar{x}) = n_{y_0}(x) - \operatorname{Min}(n_{y_0}(x), n_{y_0}(z)) = (n_{y_0}(x) - n_{y_0}(z))^+,$$

c) Si $n_{y_0}(z)=0$ la formule est évidente.

Si $n_{y_0}(z)>0$, $n_y(\bar{x})=0$ pour $y>y_0$ (même raisonnement qu'au b)).

On en déduit

$$n_{y_0}(z:x) = (n_{y_0}(z) - n_{y_0}(\bar{x}))^+ = \begin{cases} 0 & \text{s'il existe} \quad y > y_0 \quad \text{tel que} \quad n_y(x) > 0, \\ (n_{y_0}(z) - n_{y_0}(x))^+ & \text{sinon} \end{cases}$$

$= n_{y_0}(z:x)$ □

REMARQUE. Si x est un diviseur minimum de z, on déduit de b) que $\bar{x}=x$.

(23) COROLLAIRE. *Si $\bar{x}$ est l'adjoint de x dans z*

a) $z:\bar{x} = z:x$,

b) *$x:\bar{x}$ est absorbé par z.*

Le a) résulte du c) de (22).

Le b) résulte du b) de (22) et de (9): si $n_{y_0}(x:\bar{x})>0$, $n_{y_0}(x)>n_{y_0}(z)$. D'après (9) appliqué à $y_0^{n_{y_0}(x)} \leqq z$, y_0 est absorbé par z, donc $y_0^{n_{y_0}(x:\bar{x})}$ est absorbé par z.

REMARQUE. On retrouve ainsi le théorème (16).
Voici quelques autres propriétés de $\bar{x}$.

(24) PROPOSITION. *Si* $\bigvee_{i=1}^{n} x_i = m$ *et si* $\bar{x}_i$ *est l'adjoint de* x_i *dans* m, $\bigvee_{i=1}^{n} \bar{x}_i = m$.

Posons $m' = \bigvee_{i=1}^{n} \bar{x}_i$; puisque $\bar{x}_i \leqq x_i \leqq m$, $m' \leqq m$. D'autre part $(m:\bar{x}_i) = (m:x_i)$ d'après (23). D'où $m = m'$ d'après (20).

(25) THÉORÈME. *Soient a et b des diviseurs de z.*

i) $\bar{a} \leqq a$,

ii) $\bar{\bar{a}} = \bar{a}$,

iii) $a \leqq b \Rightarrow \bar{a} \leqq \bar{b}$

iv) $\bar{a} \vee \bar{b} = \overline{a \vee b}$.

i) A déjà été démontré (après la définition 7).
ii) D'après (23) on a $\bar{\bar{a}} = z:(z:\bar{a}) = z:(z:a) = \bar{a}$.
iii) D'après (14) on a $a \leqq b \leqq z \Rightarrow (z:b) \leqq (z:a) \leqq z \Rightarrow z:(z:a) \leqq z:(z:b)$, soit encore $\bar{a} \leqq \bar{b}$.

(iv) On utilise (22) pour calculer les décompositions réduites de $\overline{a \vee b}$ et $\bar{a} \vee \bar{b}$. Soient

$$a = \prod_{y \in \mathscr{S}} y^{n_y(a)}, \quad b = \prod_{y \in \mathscr{S}} y^{n_y(b)}, \quad z = \prod_{y \in \mathscr{S}} y^{n_y(z)}$$

les décompositions réduites de a, b, et z respectivement. D'après (22) et (17), on a: $\bar{a} \vee \bar{b} = \prod_{y \in \mathscr{S}} y^{n_y(\bar{a} \vee \bar{b})}$ avec

$$n_{y_0}(\bar{a} \vee \bar{b}) = \begin{cases} n_{y_0}(z) & \text{s'il existe } y > y_0 \text{ tel que } n_y(a) > 0 \text{ ou } n_y(b) > 0 \\ \max\left(\min\left(n_{y_0}(a), n_{y_0}(z)\right), \min\left(n_{y_0}(b), n_{y_0}(z)\right)\right) = \\ = \min\left(n_{y_0}(z), \max\left(n_{y_0}(a), n_{y_0}(b)\right)\right) & \text{sinon.} \end{cases}$$

Cette décomposition est *à priori* non réduite, mais puisque pour tout $y \in \mathscr{S}$ $n_y(\bar{a} \vee \bar{b}) \leqq \leqq n_y(z)$, la décomposition est réduite d'aprés (5). Donc $\bar{a} \vee \bar{b}$ est un diviseur minimum de z. Soit $a \vee b = \prod_{y \in \mathscr{S}} y^{n_y(a \vee b)}$ la décomposition réduite de $a \vee b$. D'après (22) on a:

$$n_{y_0}(\bar{a} \vee \bar{b}) = n_{y_0}(z) \quad \text{s'il existe } y > y_0 \text{ tel que } n_y(a \vee b) > 0,$$
$$= \min\left(n_{y_0}(z), n_{y_0}(a \vee b)\right) \quad \text{sinon.}$$

— Or, s'il existe $y > y_0$ tel que $n_y(a \vee b) > 0$ $\sum_{y > y_0} n_y(a \vee b) > 0$, or d'après (4) appliqué à $H = \{y \in \mathscr{S} \mid y > y_0\}$,

$$D = \prod_{y \in \mathscr{S}} y^{n_y(a \vee b)} \quad \text{et} \quad D' = \prod_{y \in \mathscr{S}} y^{\max(n_y(a), n_y(b))}.$$

$$0 < \sum_{y > y_0} n_y(a \vee b) \leqq \sum_{y > y_0} \max\left(n_y(a), n_y(b)\right)$$

donc il existe $y>y_0$ tel que $n_y(a)>0$ ou $n_y(b)>0$. On en déduit $n_{y_0}(\bar{a}\vee\bar{b})=n_{y_0}(z)=$ $=n_{y_0}(\overline{a\vee b})$.

— Si $\sum_{y>y_0} n_y(a\vee b)=0$, $n_{y_0}(\bar{a}\vee\bar{b})=\min\left(n_{y_0}(z), n_{y_0}(a\vee b)\right)$. Or d'après (4) appliqué à $H=\{y_0\}$ à D et à D' $n_{y_0}(a\vee b)\leqq\max\left(n_{y_0}(a), n_{y_0}(b)\right)$ donc $n_{y_0}(\overline{a\vee b})\leqq n_{y_0}(\bar{a}\vee\bar{b})$. On en déduit finalement $\overline{a\vee b}\leqq\bar{a}\vee\bar{b}$.

Réciproquement, on a $\bar{a}\leqq a\leqq a\vee b$, $\bar{b}\leqq b\leqq a\vee b$ donc $\bar{a}\vee\bar{b}\leqq a\vee b$ d'où d'après iii) $\overline{\bar{a}\vee\bar{b}}\leqq\bar{a}\vee\bar{b}$ mais comme $\bar{a}\vee\bar{b}$ est un diviseur minimum de z, on a $\overline{\bar{a}\vee\bar{b}}=\bar{a}\vee\bar{b}$.

Conséquence. Soit Z l'ensemble des diviseurs de z. L'opérateur de Z dans Z. défini par $x\mapsto\bar{x}$ est un opérateur de fermeture pour la relation $\leqq$, compatible avec la loi $\vee$ (cf. Fuchs (7) page 257 pour des propriétés analogues).

H muni de la loi $\vee$ est un holoïde. L'ordre est en effet le même que dans $(H, \cdot)$ puisque: $a\leqq b\Leftrightarrow a\vee b=b$. On peut donc définir des éléments irréductibles pour la loi $\vee$, qu'on appellera éléments $\vee$-irréductibles. On introduit de façon analogue les notions de $\vee$-décompositions, d'holoïde $\vee$-factoriel, etc.

Voici une caractérisation des éléments $\vee$-irréductibles.

(26) Proposition. *x est $\vee$-irréductible si et seulement si x est une puissance (non nulle) d'irréductible.*

Soit $y\in\mathscr{I}$ et $n\in\mathbf{N}^*$. Supposons que $y^n=\bigvee_{i\in I} a_i$. Soient $\bar{a}_i$ les adjoints de a_i dans $y^n\cdot y^n=\bigvee_{i\in I}\bar{a}_i$ d'après (24). Puisque $\bar{a}_i$ est un diviseur minimum de y^n, $\bar{a}_i=y^{n_i}$ avec $n_i\leqq n$ d'après (15). Comme il est clair que $\bigvee_{i\in I} y^{n_i}=y^{\max_{i\in I} n_i}=y^n$ ou bien il existe $i_0\in I$ tel que $\max_{i\in I} n_i=n_{i_0}$ ou bien $I=\emptyset$. Le second cas est exclus car $n>0$. Donc: $y^{n_{i_0}}=$ $=\bar{a}_{i_0}=y^n$. Mais comme $\bar{a}_{i_0}\leqq a_{i_0}\leqq y^n$ $a_{i_0}=y^n$ donc y^n est $\vee$-irréductible.

Réciproquement soit x *$\vee$-irréductible.* Si $x=\prod_{y\in\mathscr{I}} y^{n_y(x)}$ est une décomposition de x, il vient $x=\bigvee_{y\in\mathscr{I}} y^{n_y(x)}$ d'après (17). Donc $x=y_0^{n_{y_0}(x)}$ pour un $y_0\in\mathscr{I}$.

Enonçons maintenant le théorème le plus important.

(27) Théorème. *Si H est factoriel, H est $\vee$-factoriel.*

□ Soit $D=\prod_{y\in\mathscr{I}} y^{n_y(x)}$ la décomposition réduite de x. Alors $x=\bigvee_{y\in\mathscr{I}} y^{n_y(x)}=D^v$ est une $\vee$-décomposition de x (d'après (17)).

Considérons une autre $\vee$-décomposition de x; soit $x=\bigvee_{y\in\mathscr{I}} y^{n'_y(x)}=D'^v$. On a alors pour la même raison $x=\prod_{y\in\mathscr{I}} y^{n'_y(x)}=D'$. Puisque D est réduite (7) s'applique: pour tout $y_0\in\mathscr{I}$ il existe une injection de réduction σ_{y_0} de $y_0^{n_{y_0}(x)}\to D'$ et les images dans $\mathscr{I}$ des σ_{y_0} sont deux à deux disjointes.

— Si l'image dans $\mathscr{I}$ de σ_{y_0} contient un $y>y_0$ on pose

$$\sigma^v(y_0^{n_{y_0}(x)})=y_0^{n'_y(x)}.$$

— Sinon on pose

$$\sigma^v(y_0^{n_{y_0}(x)})=y_0^{n'_{y_0}(x)}.$$

σ^v est une injection de D^v dans D'^v. En effet, si

$$\sigma^v(y_1^{n_{y_1}(x)}) = x_1^{n_1} = \sigma^v(y_2^{n_{y_2}(x)}) = x_2^{n_2}$$

alors $x_1 = x_2$ d'après (12). Mais x_1 et x_2 sont dans l'image dans $\mathscr{F}$ de σ_{y_1} et σ_{y_2} respectivement donc $y_1 = y_2$. Enfin il est clair que $x \leqq \sigma^v(x)$ donc $x \vee \sigma^v(x) = \sigma^v(x)$ et x est inférieur ou égal à $\sigma^v(x)$ pour l'ordre associé à la loi $\vee$.

Donc σ^v est une injection de réduction de D^v dans D'^v et D^v est la $\vee$-décomposition réduite de x.

REMARQUE. La réciproque du théorème fondamental est fausse. Considérons en effet l'holoïde $H = \{e, p, q, z\}$ dont la table est

	p	q	z
p	z	z	z
q	z	z	z
z	z	z	z

$\vee$	p	q	z
p	p	z	z
q	z	q	z
z	z	z	z

H est $\vee$-factoriel: p et q sont $\vee$-irréductible, $z = p \vee q$.
H n'est pas factoriel: $pq = p^2 = q^2 = z$.

(28) COROLLAIRE. *L'équation en x $a \vee x = m$ (pour $a \leqq m$) admet une solution minimum.*

□ C'est la traduction du théorème (13). □

§ 5. P.G.C.D.

Nous ne mentionnerons dans ce paragraphe qu'une seule propriété.

(29) PROPOSITION. *Soit H un holoïde factoriel, x et z des éléments de H. Si le p.g.c.d de x et z existe (on le note $x \wedge z$), on a $(x \wedge z)(x \vee z) = xz$.*

□ Soient $x = \prod_{y \in \mathscr{F}} y^{n_y(x)}$ et $z = \prod_{y \in \mathscr{F}} y^{n_y(z)}$ les décompositions réduites de x et z respectivement.

Posons $t = \prod_{y \in \mathscr{F}} y^{n_y(t)}$ où $n_y(t) = \inf(n_y(x), n_y(z))$. On a $t \leqq x$ et $t \leqq z$ donc $t \leqq x \wedge z$. Or

$$\prod_{y \in \mathscr{F}} y^{\inf(n_y(x), n_y(z))} \prod_{y \in \mathscr{F}} y^{\max(n_y(x), n_y(z))} = xz.$$

Donc $xz = t(x \vee y) \leqq (x \wedge y)(x \vee y)$.

Réciproquement soit d tel que $d \leqq x$ et $d \leqq z$. Soit $d = \prod_{y \in \mathscr{F}} y^{n_y(x)}$ la décomposition réduite de d. Soit $y_0 \in \mathscr{F}$ (9) appliqué aux inéquations $y_0^{n_{y_0}(d)} \leqq x$ et $y_0^{n_{y_0}(d)} \leqq z$ conduit à la discussion suivante:

1er cas. $n_{y_0}(d) \leqq n_{y_0}(x)$ et $n_{y_0}(d) \leqq n_{y_0}(z)$.

Alors $n_{y_0}(d) \leqq n_{y_0}(t)$. D'où

$$n_{y_0}(d) + \max(n_{y_0}(x), n_{y_0}(z)) \leqq n_{y_0}(x) + n_{y_0}(z).$$

2ème cas. y_0 est absorbé suit par x, soit par z donc par xz. Alors

$$y_0^{n_{y_0}(d)+\max(n_{y_0}(x),n_{y_0}(z))} \leqq xz.$$

On conclut d'après (10) que $(x \vee z)(x \wedge z) \leqq xz$.

REMARQUE. En général $p(x \wedge z) \neq px \wedge pz$ (même si les deux membres sont définis).

Remerciements. Je tiens à remercier mon ami Jacques Van de Wiele qui a largement contribué à la génèse de cet article et Messieurs les Professeurs K. Keimel, B. Bosbach, G. Lallement et H. Mitsch pour leurs précieux conseils.

BIBLIOGRAPHIE

[1] BOSBACH, B.: Charakterisierungen von Halbgruppen mit eindeutigen Halbprimfaktorzerlegungen, *Math. Annalen* **139** (1960), 184—196.
[2] BOSBACH, B.: Charakterisierungen von Halbgruppen mit eindeutigen Halbprimfaktorzerlegungen unter Berücksichtigung der Verbände und Ringe, *Math. Annalen* **141** (1960), 193—209.
[3] BOSBACH, B.: Arithmetische Halbgruppen, *Math. Annalen* **144** (1961), 239—252.
[4] BOSBACH, B.: Transzendente Ringerweiterungen mit eindeutigen Faktorzerlegungen, *Math. Annalen* **178** (1968), 299—301.
[5] DILWORTH, R. P.: In Lattice structure theory (1961) Proc. of Symp. in Pure Math. Vol. II p. 3—17.
[6] DUBREIL—JACOTIN—LESIEUR—CROISOT: *Leçons sur la théorie des treillis et des structures algébriques ordonnées.*
[7] FUCHS, L.: *Teilweise geordnete algebraische Strukturen,* Studia Mathematica — Mathematische Lehrbücher, Band XIX, Vandenhoeck & Ruprecht, Göttingen, 1966.
[8] GRÄTZER, G.: *Lattice theory,* Freeman and Co. (1971).
[9] MITSCH, H.: Rechsteilweise geordnete Halbgruppen, *Beitäge zur Algebra und Geometrie* **2** (1974), 61—72.
[10] MITSCH, H.: Rechtsteilweise geordnete Halbgruppen mit Teilbarkeitsordnungen, *Beiträge zur Algebra und Geometrie* **3** (1974), 23—35.

Institut de Programmation, Université Paris VI et CNRS, Tour 55—65, 4 Place Jussieu, 75 230 Paris, Cedex 05, France

(Reçu le 11 avril 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 185—192

ON THE DEGREE OF APPROXIMATION BY MATRIX MEANS

by

HUZOOR H. KHAN and A. WAFI*

1.

Let $f(x)$ be a 2π-periodic function integrable in the sense of Lebesgue over $(-\pi, \pi)$. Let its Fourier series be given by

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx). \tag{1.1}$$

Let $(\lambda_{n,k})$ $(n=0, 1, 2, \ldots, k=0, 1, \ldots, n;\ \lambda_{n,0}=1)$ be triangular matrix of real or complex numbers. Let

$$\sigma_n(f, x) = \sum_{k=0}^{n} \lambda_{n,k} u_k = \sum_{k=0}^{n} \Delta\lambda_{n,k} S_k \tag{1.2}$$

where S_k denotes the n^{th} partial sum of (1.1). Let

$$\Delta\lambda_{n,k} = \lambda_{n,k} - \lambda_{n,k+1}$$

and

$$\Delta^2\lambda_{n,k} = \Delta\lambda_{n,k} - \Delta\lambda_{n,k+1}.$$

A series Σu_n with partial sum s_n is said to be summable to a finite limit s if the sequence $\{\sigma_n(x)\}$ tends to s as n tends to infinity. The necessary and sufficient conditions for matrix means (denoted by (Λ)) to be regular are that

(i) there is a constant M such that

$$\lim_{n\to\infty} \sum_{k=0}^{n} |\Delta\lambda_{n,k}| < M \quad \text{for every} \quad k,$$

(ii) $\lim_{n\to\infty} \Delta\lambda_{n,k} = 0$ for every k,

(iii) $\lim_{n\to\infty} \sum_{k=0}^{n} \Delta\lambda_{n,k} = 1.$

In particular, if

$$\Delta\lambda_{n,k} = \begin{cases} \dfrac{p_{n-k}}{P_n} & (k \leqq n) \quad \text{(is non-negative and non-decreasing)}, \\ 0 & (k > n), \end{cases}$$

* Sponsored by the CSIR, New Delhi, India under the SRF grant No. 7/112(537)/76 — EMR — I

where $P_n=p_0+p_1+p_2+\ldots+p_n\to\infty$ as $n\to\infty$ and $\{p_n\}$ is non-negative non-increasing generating sequence, then σ_n defined by (1.2) is the same as Nörlund means generated by the sequence of coefficients $\{p_n\}$ which is usually written as (N, p_n) means.

Similarly, if

$$\Delta\lambda_{n,k}=\begin{cases}\dfrac{\binom{n-k+\alpha-1}{\alpha-1}}{\binom{n+\alpha}{\alpha}}, & \alpha>0, \quad \text{for} \quad k\leqq n\\ 0, & \text{for} \quad \kappa>n\end{cases}$$

then σ_n mean is the same as (C, α) mean, the familiar Cesàro means of order $\alpha>0$.

2.

The following theorem on the degree of approximation of a function $f\in\mathrm{Lip}\,\alpha$, by the (C, δ) means of its Fourier series, is due to G. ALEXITS [1].

THEOREM 2.1. *If a periodic function $f\in\mathrm{Lip}\,\alpha$ for $0<\alpha\leqq 1$, then the degree of approximation of the (C, δ)-means of its Fourier series for $0<\alpha<\delta\leqq 1$ is given by*

$$\max_{0\leqq x\leqq 2\pi}|f(x)-\sigma_n^{(\delta)}(x)|=O(1/n^{\alpha}),$$

and for $0<\alpha\leqq\delta\leqq 1$, is given by

$$\max_{0\leqq x\leqq 2\pi}|f(x)-\sigma_n^{(\delta)}(x)|=O\left(\frac{\log n}{n^{\alpha}}\right),$$

where $\sigma_n^{(\delta)}$ are the (C, δ)-means of the partial sum of (1.1).

Later on HOLLAND, SAHNEY and TZIMBALARIO [2] extended theorem (2.1) on the degree of approximation to a function by Nörlund means of its Fourier series belonging to $C^*[0, 2\pi]$, the class of all continuous functions on $[0, 2\pi]$, periodic and of period 2π. Their theorem states as follows:

THEOREM 2.2. *If $w(t)$ is the modulus of continuity of $f\in C^*[0, 2\pi]$, then the degree of a approximation of f by the Nörlund means of the Fourier series for f is given by*

$$E_n=\max_{0\leqq t\leqq 2\pi}|f(t)-T_n(t)|=O\left\{\frac{1}{P_n}\sum_{k=1}^{n}\frac{P_k w(1/k)}{k}\right\}$$

where T_n are the (N, p_n)-means of the Fourier series for f.

HOLLAND, SAHNEY and TZIMBALARIO [2] have shown that the theorem (2.2) reduces to Theorem (2.1) if we deal with Cesàro means of order δ and consider a function $f\in\mathrm{Lip}\,\alpha$, $0<\alpha\leqq 1$.

3.

In this paper we have given the answer of the open problem (i) imposed by HOLLAND etc. [2] by using a more general operator (matrix means) of which (N, p_n) is a special case for the Fourier series.

Our Theorem may be stated as follows:

THEOREM 3.1. *If $\{\Delta\lambda_{n,k}\}_{k=0}^{n}$ is non-negative and non-decreasing sequence with respect to k and if $w(t)$ is the modulus of continuity of $f\in C^*[0, 2\pi]$, then the degree of approximation of f by matrix means of the Fourier series for f is given by*

$$\max_{0\leqq t\leqq 2\pi}|f(x)-\sigma_n(f,x)| = O\left\{\sum_{k=1}^{n}\frac{\Delta\lambda_{n,n-k}w(1/k)}{k}\right\}$$

where $\sigma_n(f, x)$ are the matrix means of the Fourier series of (1.1).

In order to prove the theorem (3.1) we need the following Lemma [3].

LEMMA 3.1. *If $\{\Delta\lambda_{n,k}\}_{k=0}^{n}$ is non-negative and non-decreasing sequence with respect to k, then for $0\leqq a<b\leqq\infty$, $0\leqq t\leqq\pi$ for every n*

$$\left|\sum_{k=a}^{b}\Delta\lambda_{n,n-k}e^{i(n-k)t}\right| < Bt^{-1}\Delta\lambda_{n,n-\tau},$$

where τ is the integral part of $1/t$.

PROOF OF THEOREM 3.1. Let us write

$$S_k(x) = \frac{1}{2}a_0+\sum_{v=1}^{n}(a_v\cos vx+b_v\sin vx),$$

then

$$S_k(x)-f(x) = \frac{1}{2\pi}\int_0^{\pi}[f(x+t)+f(x-t)-2f(x)]\frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}}\,dt,$$

and

$$\sigma_n(f,x)-f(x) = \sum_{k=0}^{n}\Delta\lambda_{n,k}\{S_k(x)-f(x)\} =$$

$$= \frac{1}{2\pi}\sum_{k=0}^{n}\Delta\lambda_{n,k}\int_0^{\pi}[f(x+t)+f(x-t)-2f(x)]\frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}}\,dt =$$

$$= \int_0^{\pi}J_n(t)[f(x+t)+f(x-t)-2f(x)]\,dt,$$

where

$$J_n(t) = \frac{1}{2\pi}\sum_{k=0}^{n}\Delta\lambda_{n,k}\frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}}.$$

Now

$$|\sigma_n(f, x)-f(x)| \leqq \int_0^{\pi} |f(x+t)+f(x-t)-2f(x)|\,|J_n(t)|\,dt =$$

$$= \Big[\int_0^{\pi/n} + \int_{\pi/n}^{\delta} + \int_{\delta}^{\pi}\Big] [|f(x+t)+f(x-t)-2f(x)|] \cdot |J_n(t)|\,dt.$$

It is clear that $\Phi(t) \leqq \omega(t)$, when

$$\Phi(t) = [f(x+t)+f(x-t)-2f(x)]$$

and therefore

$$|\sigma_n(f, x)-f(x)| \leqq \Big[\int_0^{\pi/n} + \int_{\pi/n}^{\delta} + \int_{\delta}^{\pi}\Big] w(t)\,|J_n(t)|\,dt = I_1+I_2+I_3 \quad \text{(say)}.$$

For evaluating I_1, we notice that

$$I_1 = \int_0^{\pi/n} w(t) \left| \frac{1}{2\pi} \sum_{k=0}^{n} \Delta\lambda_{n,k} \frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O\big(w(1/n)\big) \int_0^{\pi/n} \left| \sum_{k=0}^{n} \Delta\lambda_{n,k} \frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O\big(w(1/n)\big) \int_0^{\pi/n} \left| \sum_{k=0}^{n} \Delta\lambda_{n,k}(2k+1) \right| dt$$

uniformly in $0 < t \leqq \pi/n$. Now applying Abel's lemma, we get

$$= O\big(w(1/n)\big) \int_0^{\pi/n} \left\{ \sum_{k=0}^{n-1} \left(\sum_0^k |\Delta\lambda_{n,k}| \right) |2k+1-2k-2| + \right.$$

$$\left. + (2n+1) \sum_0^n |\Delta\lambda_{n,k}| \right\} dt =$$

$$= O\big(w(1/n)\big) \int_0^{\pi/n} \{2Mn+(2n+1)M\}\,dt =$$

$$= O\big(w(1/n)\big).$$

Next,

$$I_2 = \int_{\pi/n}^{\delta} w(t) \left| \frac{1}{2\pi} \sum_{k=0}^{n} \Delta\lambda_{n,k} \frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= \int_{\pi/n}^{\delta} w(t) \frac{1}{2\pi} \left| \sum_{k=0}^{n} \Delta\lambda_{n,k} \frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O(1) \int_{\pi/n}^{\delta} w(t) \left| \sum_{k=0}^{n} \Delta\lambda_{n,n-k} \frac{\sin\left(\frac{2n-2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O(1) \int_{\pi/n}^{\delta} w(t) \left| \sum_{k=0}^{n} \Delta\lambda_{n,n-k} \frac{\sin\left(n-k+\frac{1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O(1) \int_{\pi/n}^{\delta} w(t) \frac{\Delta\lambda_{n,n-\tau}}{t} dt$$

(by lemma 3.1), where τ is the integral part of $1/t$.

By regularity condition of the matrix means, we have

$$\sum_{k=0}^{n} \frac{\Delta\lambda_{n,n-k}}{k} w(1/k) \geqq w(1/n) \sum_{k=0}^{n} \Delta\lambda_{n,n-\tau} = O\big(w(1/n)\big).$$

Hence

$$I_2 = O(1) \int_{n/\pi}^{1/\delta} \frac{w(1/t)}{1/t} \Delta\lambda_{n,n-t}(dt/t^2),$$

$$= O\left(\int_{n/\pi}^{1/\delta} \frac{w(1/t)}{t} \Delta\lambda_{n,n-t}\, dt \right),$$

$$= O\left(\sum_{k=1}^{n} w(1/k) \frac{\Delta\lambda_{n,n-k}}{k} \right).$$

Finally, for the evaluation of I_3, we have

$$I_3 = \int_{\delta}^{\pi} w(t) \left| \frac{1}{2\pi} \sum_{k=0}^{n} \Delta\lambda_{n,k} \frac{\sin\left(\frac{2k+1}{2}\right)t}{\sin\frac{t}{2}} \right| dt =$$

$$= O\left(\int_{\delta}^{\pi} w(t) \left\{ \sum_{k=0}^{n-1} |\Delta^2\lambda_{n,k}| \left| \frac{\sin^2\left(\frac{2k+1}{2}\right)t}{\sin^2\frac{t}{2}} \right| + \frac{1}{\pi} |\Delta^2\lambda_{n,n}| \left| \frac{\sin^2\left(\frac{2n+1}{2}\right)t}{\sin^2\frac{t}{2}} \right| \right\} dt \right) =$$

$$= O\left(\int_{\delta}^{\pi} w(t) |\Omega(t)| \, dt \right).$$

$$\max_{0 \leqq \delta \leqq t \leqq \pi} |\Omega(t)| < \frac{1}{2\pi \sin^2\frac{\delta}{2}} \left[\sum_{k=0}^{n-1} |\lambda_{n,k}| + o(1) \right] = o(1) \quad \text{as} \quad n \to \infty.$$

Hence

$$I_3 = o(1) \quad \text{as} \quad n \to \infty$$

which is dominated by the bound for I_2. Adding the bounds for I_1, I_2, I_3, we have

$$|f(x) - \sigma_n(f, x)| = O\left\{ \sum_{k=1}^{n} \frac{\Delta\lambda_{n,n-k} w(1/k)}{k} \right\}$$

which is the best order of approximation when $w(1/k)$ is decreasing.

4.

Using similar techniques, the following theorem for conjugate Fourier series is quite obvious:

THEOREM 4.1. *If $\{\Delta\lambda_{n,k}\}_{k=0}^{n}$ is non-negative and non-decreasing sequence with respect to k and if $w(t)$ is modulus of continuity of $\tilde{f} \in C^*[0, 2\pi]$, then the degree of approximation of $\tilde{f}$ by the matrix means of the conjugate Fourier series of $\tilde{f}$ is given by*

$$\max_{0 \leqq t \leqq 2\pi} |\tilde{f}(x) - \tilde{\sigma}_n(\tilde{f}, x)| = O\left\{ \sum_{k=1}^{n} \frac{\Delta\lambda_{n,n-k} w(1/k)}{k} \right\},$$

where $\tilde{\sigma}_n(\tilde{f}, x)$ are the matrix means of the conjugate Fourier series of $\tilde{f}$ defined as

$$\tilde{f}(x) \sim \sum_{n=1}^{\infty} (b_n \cos nx - a_n \sin nx) \equiv \sum_{n=1}^{\infty} B_n(x).$$

5.

Let $\{p_n\}$ and $\{q_n\}$ be non-negative, non-increasing generating sequence for (N, p_n, q_n) method such that

$$P_n = p_0+p_1+\ldots+p_n \to \infty \quad \text{as} \quad n \to \infty,$$

$$Q_n = q_0+q_1+\ldots+q_n$$

and

$$R_n = p_0 q_n+p_1 q_{n-1}+\ldots+p_n q_0 \to \infty \quad \text{as} \quad n \to \infty.$$

A given series $\sum_{n=0}^{\infty} g_n$ with the sequence of partial sums $\{s_n\}$ is said to be summable (N, p_n, q_n) to g, provided that

$$t_n^{p\cdot q}(f, x) = \frac{1}{R_n}\sum_{k=0}^{n} R_{n-k} g_k,$$

$$= \frac{1}{R_n}\sum_{k=0}^{n} r_{n-k} S_k \to g \quad \text{as} \quad n \to \infty,$$

and $t_n^{p\cdot q}(f, x)$ is called the generalized Nörlund operator.

Theorem (2.2) can further be extended for generalized Nörlund operator as follows:

THEOREM 5.1. *If $w(t)$ is the modulus of continuity of $f \in C^*[0, 2\pi]$, then the degree of approximation of f by the generalized Nörlund means of the Fourier series of f is given by*

$$\max_{0\leq t\leq 2\pi} |f(x)-t_n^{p\cdot q}(f, x)| = O\left\{\frac{1}{R_n}\sum_{k=1}^{n}\frac{R_k w(1/k)}{k}\right\}$$

where $t_n^{p\cdot q}(f, x)$ are the (N, p_n, q_n) means of Fourier series of f.

REMARKS. (1) If the matrix means is replaced by Cesàro means, Nörlund means and generalized Nörlund means in particular, then our theorem (3.1) deduces [where $\left\{\frac{p_{n-k}}{P_n}\right\}_{k=0}^{n}$ and $\left\{\frac{r_{n-k}}{R_n}\right\}_{k=0}^{n}$ are always non-negative and non-decreasing]:

(i) Theorem (2.1) with an improved order

(ii) Theorem (2.2), when $\Delta\lambda_{n,k}=\frac{p_{n-k}}{P_n}$

and

(iii) Theorem (5.1) when $\Delta\lambda_{n,k}=\frac{r_{n-k}}{R_n}$, respectively.

(2) When $q_n=1$, then Theorem (5.1) reduces directly to Theorem (2.2).

Acknowledgement

The authors take the opportunity of expressing their warmest thanks to the Referee, Dr. I. Joó, for his suggestion to improve the statement of the theorem.

REFERENCES

[1] ALEXITS, G.: Über die Annäherung einer stetigen Funktions durch die Cesaròschen Mittel ihrer Fourierreihe, Math. Annalen **100** (1928), 264—277.

[2] HOLLAND, A. S. B., SAHNEY, B. N. and TZIMBALARIO, J.: On the degree of approximation of a class of functions by means of Fourier series, *Acta Sci. Math.* **38** (1976), 69—72.

[3] KISHOR, N.: On the absolute matrix summability of a Fourier series, *Indian Journal of Mathematics* **13** (1971), 99—110.

Department of Mathematics, Aligarh Muslim University, Aligarh — 202001, India

(Received April 28, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 193—197

RINGS WITH CONSTRAINTS ON NILPOTENT ELEMENTS AND COMMUTATORS

by

M. S. PUTCHA and A. YAQUB

Abstract. Suppose R is an associative ring with the property that, for every x in R, there exists a positive integer $n=n(x)$ and a polynomial $f(\lambda)=f_x(\lambda)$ with integer coefficients such that $x^n=x^{n+1}f(x)$. Suppose, further, that (i) the nilpotent elements of R commute with each other, and (ii) every additive commutator $xy-yx$ is in the center of R. Then R is commutative. Moreover, examples are given which show that this theorem need not be true if either (i) or (ii) is deleted. The proof utilizes the structure theory of rings and yields the following structure of R:

$R\cong$ a subdirect sum of local commutative rings and nil commutative rings.

A well-known theorem of HERSTEIN [1] asserts that if R is an associative ring with the property that, for every x in R, there exists a positive integer $n=n(x)$ and a polynomial $f(\lambda)=f_x(\lambda)$ with integer coefficients such that $x^n=x^{n+1}f(x)$, and if all the nilpotent elements of R are in the center of R, then R is commutative. Example 1 below shows that the above constraint on the nilpotent elements of R (i.e., that they are central) cannot be weakened by replacing it with the hypothesis that these nilpotent elements *commute with each other*. However, by adding another constraint on the additive commutators, $xy-yx$, of R, commutativity follows. In fact, we prove the following

THEOREM 1. *Suppose R is an associative ring with the property that, for every x in R, there exists a positive integer $n=n(x)$ and a polynomial $f(\lambda)=f_x(\lambda)$ with integer coefficients such that $x^n=x^{n+1}f(x)$, (both n and $f(\lambda)$ depend on x). Suppose, further, that*

(i) *The nilpotent elements of R commute with each other; and*

(ii) *For all x, y in R, the additive commutator $xy-yx$ is in the center of R.*

Then R is commutative.

We also give examples which show that Theorem 1 need not be true if we delete either (i) or (ii). Our proof utilizes the structure theory of rings and yields the structure of the ground ring R given in Theorem 2.

In preparation for the proof of Theorem 1, we first establish the following

LEMMA 1. *Suppose R is an associative ring which satisfies all the hypotheses of Theorem* 1. *Then any homomorphic image of R inherits all of these hypotheses.*

PROOF. Suppose $\sigma\colon R\to\bar{R}$ is a homomorphism of R onto $\bar{R}$. Clearly, $\bar{R}$ inherits from R all the hypotheses in Theorem 1 with the possible exception of hypothesis (i). To prove that $\bar{R}$ satisfies (i) also, suppose that $\bar{x}$ and $\bar{y}$ are any nilpotent elements of $\bar{R}$, and suppose

$$(\bar{x})^r = 0, \quad (\bar{y})^s = 0. \tag{1}$$

Let x and y be pre-image elements in R of $\bar{x}$ and $\bar{y}$, respectively, under the homomorphism σ. By hypothesis, there exists a positive integer n and a polynomial $f(\lambda)$ with integer coefficients such that $x^n = x^{n+1}f(x)$. Since $x^{n-1}(x-x^2f(x))=0$, $(x-x^2f(x))^n = x^{n-1}(1-xf(x))^{n-1}(x-x^2f(x))=0$. Therefore, $x-x^2f(x)$ is nilpotent. Similarly, there exists a polynomial $g(\lambda)$ with integer coefficients such that $y-y^2g(y)$ is nilpotent. Hence, applying hypothesis (i) to the ground ring R, we conclude that

(2) $$x-x^2f(x) \quad \text{commutes with} \quad y-y^2g(y), \quad (x, y \in R).$$

Repeating this argument for $x^2f(x)$ (instead of x), *keeping y fixed*, we conclude that there exists a polynomial $f_2(\lambda)$ with integer coefficients such that (see (2))

(3) $$x^2f(x)-x^4f_2(x) \quad \text{commutes with} \quad y-y^2g(y).$$

Combining (2) and (3), we see that

$$x-x^4f_2(x) \quad \text{commutes with} \quad y-y^2g(y).$$

Continuing this process, we see that, for every positive integer l, there exists a polynomial $f_l(\lambda)$ with integer coefficients such that

(4) $$x-x^{2^l}f_l(x) \quad \text{commutes with} \quad y-y^2g(y).$$

Choose l_0 such that $2^{l_0} > r$. Then, by (4),

(5) $$x-x^{2^{l_0}}f_{l_0}(x) \quad \text{commutes with} \quad y-y^2g(y), \quad (2^{l_0} > r).$$

We now fix x in (5), and repeat the above argument (which led to (5)) to y to conclude the following: For every positive integer $\varkappa$, there exists a polynomial $g_\varkappa(\lambda)$ with integer coefficients such that

(6) $$x-x^{2^{l_0}}f_{l_0}(x) \quad \text{commutes with} \quad y-y^{2^\varkappa}g_\varkappa(y).$$

Now, choose $\varkappa_0$ such that $2^{\varkappa_0} > s$. Then, by (6), $x-x^{2^{l_0}}f_{l_0}(x)$ commutes with $y-y^{2^{\varkappa_0}}g_{\varkappa_0}(y)$, $(2^{l_0} > r,\ 2^{\varkappa_0} > s)$. This reflects in $\bar{R}$ as follows:

(7) $$\bar{x}-(\bar{x})^{2^{l_0}}f_{l_0}(\bar{x}) \quad \text{commutes with} \quad \bar{y}-(\bar{y})^{2^{\varkappa_0}}g_{\varkappa_0}(\bar{y}), \quad (2^{l_0} > r,\ 2^{\varkappa_0} > s).$$

Combining (7) and (1), we conclude that $\bar{x}$ commutes with $\bar{y}$, and the lemma is proved.

LEMMA 2. *Suppose R is an associative ring with the property that, for all x, y in R, $xy-yx$ is in the center of R. Then all the idempotents of R are in the center of R.*

PROOF. Suppose that e_1, e_2 are any idempotent elements in R. By hypothesis,

$$(e_1e_2)e_1-e_1(e_1e_2) \quad \text{commutes with} \quad e_1,$$

and hence

$$e_1(e_1e_2e_1-e_1e_2) = (e_1e_2e_1-e_1e_2)e_1 = 0.$$

Therefore, $e_1e_2=e_1e_2e_1$. Similarly, by considering the additive commutator $e_1(e_2e_1)-(e_2e_1)e_1$, we conclude that $e_2e_1=e_1e_2e_1$. Hence,

(8) $$e_1e_2 = e_2e_1 \quad \text{for all idempotents} \quad e_1, e_2 \text{ in } R.$$

Now, suppose that $x\in R$ and e is any idempotent element in R. It is readily verified that $e+ex-exe$ is idempotent, and hence, by (8),

$$e(e+ex-exe) = (e+ex-exe)e.$$

Therefore, $ex=exe$. A similar argument shows that $xe=exe$, and hence $ex=xe$. This proves the lemma.

LEMMA 3. *Suppose R is an associative subdirectly irreducible ring which satifies all the hypotheses of Theorem* 1. *Then either R is a nil commutative ring, or R is a local ring in the sense that every element in R is either nilpotent or has an inverse in R.*

PROOF. Suppose that $x\in R$. Then, by hypothesis, there exists a positive integer n and a polynomial $f(\lambda)$ with integer coefficients such that

$$x^n = x^{n+1}f(x).$$

An easy induction shows that

$$x^n = x^{n+r}\{f(x)\}^r, \quad \text{for all positive integers } r. \tag{9}$$

In particular,

$$x^n = x^{2n}\{f(x)\}^n. \tag{10}$$

Now, let $e=x^n\{f(x)\}^n$. Then $e^2=e$.

By Lemma 2, all of the idempotent elements of R are in the center of R. Since R is subdirectly irreducible, an easy argument [2, Lemma 9] shows that the only possible idempotents in R are 0 and 1, and hence

$$e = 0 \quad \text{or} \quad e = 1. \tag{11}$$

If R does *not* have an identity, then $e=0$ and hence $x^n\{f(x)\}^n=0$. Therefore, by (10), $x^n=0$. In this case, R is a nil commutative ring. On the other hand, if R has an identity 1, then, by (11),

$$x^n\{f(x)\}^n = 0 \quad \text{or} \quad x^n\{f(x)\}^n = 1.$$

Hence, recalling (10), x is nilpotent or $x^{-1}\in R$. This proves the lemma.

LEMMA 4. *Suppose R is an associative subdirectly irreducible ring with identity* 1 *which satisfies all the hypotheses of Theorem* 1. *Then the set N of nilpotent elements of R forms an ideal, and R/N is a field of prime characteristic.*

PROOF. By Lemma 3, R is a local ring. It follows at once that N is an ideal in R and R/N is a division ring. In fact, by Herstein's Theorem [1], R/N is a field. Moreover, R/N *cannot* be of characteristic zero in view of the "$x^n=x^{n+1}f(x)$" hypothesis. This proves the lemma.

We are now in a position to prove Theorem 1.

PROOF OF THEOREM 1. It is well-known that the ground ring R is isomorphic to a subdirect sum of subdirectly irreducible rings R_i ($i\in\Gamma$) each of which inherits all the hypotheses of Theorem 1 (Lemma 1). Hence, by Lemma 3, either R_i is a nil *commutative* ring, or R_i is a local ring. Hence we may *assume that the ground*

ring R is a subdirectly irreducible local ring with identity 1. Let N be the set of nilpotent elements of R. By Lemma 4, the characteristic of R/N is a prime p. Hence $p \cdot 1 \in N$. Thus $(p \cdot 1)^m = 0$ for some positive integer m. Therefore

$$\text{characteristic of } R = p^m. \tag{12}$$

Let $\bar{0} \neq \bar{b} \in R/N$. Then, in view of the "$x^n = x^{n+1} f(x)$" hypothesis, the subring, $\langle \bar{b} \rangle$, generated by $\bar{b}$ is a finite field. Hence,

$$\langle \bar{b} \rangle = GF(p^k). \tag{13}$$

Therefore,

$$(\bar{b})^{p^k} = \bar{b}, \quad \text{and} \quad (\bar{b})^{p^{mk}} = \bar{b}. \tag{14}$$

Now, let $a \in N$ and $b \in R$. By hypothesis (ii), $ab - ba$ commutes with b, and hence by an easy induction (which we omit)

$$b^n a - ab^n = nb^{n-1}(ba - ab),$$

for all positive integers n. In particular,

$$b^{p^{mk}} a - ab^{p^{mk}} = p^{mk} b^{p^{mk}-1}(ba - ab) = 0,$$

by (12). Hence,

$$ab^{p^{mk}} = b^{p^{mk}} a, \quad (a \in N,\ b \in R). \tag{15}$$

But, by (14),

$$b^{p^{mk}} - b \in N,$$

and hence, by hypothesis (i),

$$a(b^{p^{mk}} - b) = (b^{p^{mk}} - b)a, \quad (a \in N,\ b \in R). \tag{16}$$

Combining (15) and (16), we obtain

$$ab = ba, \quad \text{for all} \quad a \in N \quad \text{and all} \quad b \in R.$$

Hence, by HERSTEIN's Theorem [1], R is commutative, and Theorem 1 is proved.

In the process of proving Theorem 1, we also proved the following

THEOREM 2. *Suppose R is an associative ring which satisfies all the hypotheses of Theorem* 1. *Then R has the following structure:*

$R \cong$ a subdirect sum of nil commutative rings and local commutative rings.

We conclude with the following two examples which show that the above two theorems need not be true if we delete either hypothesis (i) or (ii).

EXAMPLE 1. Suppose

$$R = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \middle| 0, 1 \in GF(2) \right\}.$$

It is easily verified that R satisfies all the hypotheses in Theorem 1 (and Theorem 2) *except* hypothesis (ii). Also, R is *not* commutative. Thus, hypothesis (ii) *cannot* be deleted from the above theorems.

EXAMPLE 2. Let

$$R = \left\{ \begin{pmatrix} a & b & c \\ 0 & a & d \\ 0 & 0 & a \end{pmatrix} \middle| \, a, b, c, d \in GF(p) \right\}.$$

It can be checked that R satisfies all the hypotheses in Theorem 1 (and Theorem 2) *except* hypothesis (i). However, R is *not* commutative. Thus, hypothesis (i) *cannot* be deleted from the above theorems, even if we were to further assume that the ground ring R has an identity.

REFERENCES

[1] HERSTEIN, I. N.: A note on rings with central nilpotent elements, *Proc. A.M.S.* **5** (1954), 620.
[2] HERSTEIN, I. N.: The structure of a certain class of rings, *Amer. J. Math.* **75** (1953), 864—871.

AMS 1970 *subject classification.* Primary 16A70, 16A48; Secondary 16A38.

Department of Mathematics, North Carolina State University, Raleigh
and University of California, Santa Barbara

(Received May 12, 1977; revised August 30, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 199—206

ON SOLUTIONS OF LINEAR SECOND ORDER DIFFERENTIAL EQUATIONS

by

Á. ELBERT

Consider the second order differential equation

$$y''+p(x)y=0 \quad (' = d/dx), \tag{1}$$

where $p(x)$ is a positive monotonic function with continuous first and piecewise continuous second order derivatives on (a, b), $-\infty < a < b \leqq \infty$. Throughout this paper we assume, that

$$D(x) = D_p(x) = 2p^{5/2}(x)[p^{-1/2}(x)]'' = \frac{3}{2}p'^2(x) - p(x)p''(x) \gtrless 0, \tag{2}$$

i.e. $D(x)$ does not change its sign on (a, b).

A simple example for $p(x)$ which fulfils (2) is any positive monotonic concave function. To explain our aim, we consider a *solution* of (1) with such a concave function $p(x)$. Suppose the solutions has consecutive zeros at $u, v \in [a, b]$, then by a result of E. MAKAI [5]

$$\int_u^v \sqrt{p(x)}\,dx \leqq \pi.$$

On the other hand if $u, w \in [a, b]$ are consecutive zeros of the *derivative* of *another solution* of (1) with the same $p(x)$ then from [2] we know that

$$\int_u^w \sqrt{p(x)}\,dx \geqq \pi.$$

Comparing these two inequalities we conclude that $v \leqq w$. In what follows we generalize this fact showing that $v \leqq w$ is a consequence of the inequality $D_p(x) \geqq 0$. We shall introduce two functions $\varphi_1(x)$, $\varphi_2(x)$ (see e.g. in [2]) which can be considered as polar angles in a polar coordinate system. These functions are increasing if $D_p(x)$ does not change its sign in (a, b) and if $D_p(x)p'(x) > 0$ then the difference $\varphi_1(x) - \varphi_2(x)$ will be estimated depending on the sign of $D_p(x)$. As an application we solve a problem posed by W. LEIGHTON in [3] which seems to be still unsolved (see [4], pp. 462).

The interested reader may find similar results in J. VOSMANSKÝ's paper [6] (see esp. Theorem 5.1. on pp. 61—62) where the class of functions differs somewhat from ours.

Suppose that the differential equation (1) has solutions $y_1(x)$, $y_2(x)$ with the initial conditions

$$y_1(a) = 0, \quad y_1'(a) = 1,$$

$$y_2(a) = 1, \quad y_2'(a) = 0.$$

Denote by $a_0=a, a_1, \ldots, a_{n_1}$ all the consecutive zeros of $y_1(x)$ in $[a, b]$ and similarly by $a_0'=a, a_1', \ldots, a_{n_2}'$ those of $y_2'(x)$. Let $n_1, n_2 \geqq 1$. Since $p(x)>0$ in (a, b) therefore by (1) $y_1'(x)$ takes on the value zero exactly once, say at $x=\bar{a}_j$, in (a_{j-1}, a_j), $j=1, 2, \ldots, n_1$. Similarly $y_2(x)$ vanishes at $\bar{a}_j'$ where $a_{j-1}'<\bar{a}_j'<a_j'$, $j=1, 2, \ldots, n_2$.

It is possible that $y_1'(x)$ may have a zero on $(a_{n_1}, b]$ which will be denoted by $\bar{a}_{n_1+1}$, and let $c_1=\bar{a}_{n_1+1}$ if this value exists at all, otherwise let $c_1=a_{n_1}$. Concerning the solution $y_2(x)$ it is possible that there is a value $\bar{a}_{n_2+1}^1 \in (a_{n_2}', b]$ satisfying $y_2(\bar{a}_{n_2+1}')=0$. In this case let $c_2=\bar{a}_{n_2+1}'$, otherwise $c_2=a_{n_2}'$.

We introduce the continuous functions $\varphi_i(x)$ and $\varrho_i(x)$ by

$$\operatorname{tg} \varphi_i(x) = \frac{\sqrt{p(x)}\, y_i(x)}{y_i'(x)}, \tag{3}$$

$$\varrho_i^2(x) = y_i'^2(x) + p(x) y_i^2(x), \quad x \geqq a, \quad i = 1, 2.$$

Since with a given $\varphi_i(x)$ the functions $\varphi_i(x)+k\pi$ $(k=\pm 1, \pm 2, \ldots)$ also fulfil (3) for the sake of uniqueness we choose those which satisfy

$$\varphi_1(a) = 0, \quad \varphi_2(a) = \frac{\pi}{2}. \tag{4}$$

The geometric meaning of the functions just introduced is clear: $\varrho_i(x)$ and $\varphi_i(x)$ are the polar coordinates of the point, the Cartesian coordinates of which are $X=y_i'(x)$, $Y=\sqrt{p(x)}\, y_i(x)$.

The solution $y_1(x)$ is concave on (a, a_1) hence $y_1'(x)$ is strictly decreasing there thus the curve $(y_1'(x), \sqrt{p(x)}\, y_1(x))$ on the plane XY crosses the Y-axis only once at $x=\bar{a}_1$ and at $x=a_1$ it crosses the X-axis. These are crosses in the proper sense, since at $x=\bar{a}_1$ the function $y_1'(x)$ changes its sign and at $x=a_1$ the function $y_1(x)$ does it. On (a_1, a_2) the curve passes below the X-axis while $y_1'(x)$ is strictly increasing. Proceeding on this way we conclude that the curve $(y_1'(x), \sqrt{p(x)}\, y_1(x))$ crosses alternatively the X- and Y-axes when $\varphi_1(x)$ assumes the values $0, \pi/2, \pi, \ldots$. A similar statement holds for the curve $(y_2'(x), \sqrt{p(x)}\, y_2(x))$, too. Thus

$$\varphi_1(a_j) = j\pi, \quad \varphi_1(\bar{a}_{j+1}) = \left(j+\frac{1}{2}\right)\pi,$$

$$\varphi_2(a_j') = \left(j+\frac{1}{2}\right)\pi, \quad \varphi_2(\bar{a}_j') = j\pi, \quad j = 0, 1, \ldots. \tag{5}$$

Differentiating the first formula of (3) and making use of (1) we have

$$\varphi_i' = \sqrt{p} + \frac{p'}{4p} \sin 2\varphi_i \quad (i = 1, 2), \tag{6}$$

since by (3)

$$\sin\varphi_i = \frac{\sqrt{p}y_i}{\varrho_i}, \quad \cos\varphi_i = \frac{y_i'}{\varrho_i} \quad (i = 1, 2). \tag{7}$$

THEOREM 1. *If $D_p(x)$ does not change its sign on (a, b) then $\varphi_i'(x)>0$ on (a, c_i) for $i=1, 2$.*

PROOF. We remark that if $p(x)=$const.>0 then $p'(x)\equiv 0$, $D_p(x)\equiv 0$ and $\varphi_i'(x)=\sqrt{p}>0$ on (a, b) for $i=1, 2$. Hence the condition $D_p(x)>0$ or $D_p(x)<0$ throughout (a, b) is not necessary for the validity of the inequality $\varphi_i'(x)>0$.

Instead of $\varphi_i'(x)$ we shall consider the function $V_i(x)$ introduced by

$$V_i(x) = \frac{\varphi_i'}{p}\varrho_i^2 = \sqrt{p}y_i^2 + \frac{y_i'^2}{\sqrt{p}} + \frac{p'}{2p^{3/2}} y_i y_i' \quad (i = 1, 2). \tag{8}$$

This function has a derivative of rather simple form

$$V_i'(x) = -\frac{1}{2}\frac{D_p}{p^{5/2}} y_i y_i'. \tag{9}$$

Thus the function $V_i(x)$ takes on its local extrema on the closed interval $[a, c_i]$ where either $y_i(x)=0$ or $y_i'(x)=0$, and we see that the local minima among them are positive with the possible exceptions at $x=a$ and $x=b$. By (9) the limit $\lim_{x\to a+0} V_i(x)=V_i(a)$ always exists, though it may not be finite. Since for $x\in(a, b)$ $0<p(x)<\infty$ and $p'(x)$ is continuous, therefore $V_i(x)>0$ if $y_i(x)y_i'(x)=0$. Furthermore, if $p(x)$ has these properties at $x=a$, too, then, of course, $V_i(a)>0$. Otherwise a detailed discussion is needed.

Suppose first that $D_p(x)>0$. Then by (9) $V_1(a)$ is local maximum of $V_1(x)$ and $V_2(a)$ is a local minimum of $V_2(x)$. We shall show that $V_2(a)\geqq 0$. By (2) the function $p'p^{-3/2}$ is strictly decreasing and $\lim_{x\to a+0} p'p^{-3/2}$ exists. If this limit is finite then we conclude from (8) that $V_2(a)\geqq 0$. If $\lim_{x\to a+0} p'p^{-3/2}=\infty$ then in a right neighbourhood of a $p'(x)>0$ and the function $p(x)$ is increasing. Thus $\lim_{x\to a+0} p(x)=p(a)$ exists and it is finite. By the definition of $y_2(x)$ and by (1) we have

$$y_2'(x) = -\int_a^x p(\xi)y_2(\xi)\,d\xi.$$

Thus $y_2(x)$ is decreasing and positive on $[a, \bar{a}_1')$ hence if x is sufficiently near to a, we have

$$0 > \frac{p'}{2p^{3/2}} y_2 y_2' = -\frac{p'}{2p^{3/2}} y_2 \int_a^x p(\xi)y_2(\xi)\,d\xi >$$

$$> -\int_a^x \frac{p'(\xi)}{2p^{1/2}(\xi)} y_2^2(\xi)\,d\xi = -\left[\sqrt{p(\xi)}y_2^2(\xi)\right]_a^x + \int_a^x \sqrt{p(\xi)}[y_2^2(\xi)]'\,d\xi >$$

$$> -\left[\sqrt{p(\xi)}y_2^2(\xi)\right]_a^x + \sqrt{p(x)}\int_a^x [y_2^2(\xi)]'\,d\xi = \sqrt{p(a)} - \sqrt{p(x)},$$

and therefore

$$\lim_{x\to a+0} \frac{p'}{p^{3/2}} y_2 y_2' = 0$$

and $V_2(a) \geqq 0$.

If $D_p(x) < 0$, then $V_1(a)$ is a local minimum for $V_1(x)$. Now the function $p' p^{-3/2}$ is strictly increasing. Hence $\lim_{x\to a+0} p' p^{-3/2}$ exists. If this is finite then by (8) $V_1(a) \geqq 0$. If $\lim_{x\to a+0} p' p^{-3/2} = -\infty$ then in a sufficiently small right neighbourhood of a $p' < 0$ and the function p is decreasing hence $\lim_{x\to a+0} \frac{1}{\sqrt{p}}$ exists and it is finite. By the initial conditions of $y_1(x)$ we have

$$y_1(x) = \int_a^x y_1'(\xi)\, d\xi > 0 \quad \text{for} \quad a < x < a_1$$

and $y_1'(x)$ is decreasing and positive on $(a, \bar{a}_1)$ therefore if x is sufficiently near to a we have

$$0 > \frac{p'}{2p^{3/2}} y_1 y_1' = \frac{p'}{2p^{3/2}} y_1' \int_a^x y_1'(\xi)\, d\xi >$$

$$> \int_a^x \frac{p'(\xi)}{2p^{3/2}(\xi)} [y_1'(\xi)]^2\, d\xi = \left[-\frac{y_1'^2(\xi)}{\sqrt{p(\xi)}} \right]_a^x + \int_a^x \frac{[y_1'^2(\xi)]'}{\sqrt{p(\xi)}}\, d\xi >$$

$$> \left[-\frac{y_1'^2(\xi)}{\sqrt{p(\xi)}} \right]_a^x + \frac{1}{\sqrt{p(x)}} \int_a^x [y_1'^2(\xi)]'\, d\xi = \frac{1}{\sqrt{p(a)}} - \frac{1}{\sqrt{p(x)}},$$

and thus

$$\lim_{x\to a+0} \frac{p'}{p^{3/2}} y_1 y_1' = 0,$$

and $V_1(a) \geqq 0$.

If any of the relations $c_1 = b$ or $c_2 = b$ holds then the proof runs in similar manner as in the case $x = a$ and the proof of Theorem 1 is complete.

Now we proceed to comparing the zeros of $y_1(x)$ with those of $y_2'(x)$.

THEOREM 2. *If $D_p(x) \geqq 0$, then for the zeros $a_1, a_2, \ldots, a_{n_1}$ of $y_1(x)$ and for the zeros $a_1', a_2', \ldots, a_{n_2}'$ of $y_2'(x)$ the inequalities*

$$a_j \leqq a_j', \quad j = 1, 2, \ldots, \min(n_1, n_2)$$

hold.

PROOF. Let $y(x)$ be a solution of (1). Then $Y(x) = y'(x) p^{-1/2}(x)$ satisfies the differential equation

$$Y'' + \left(p(x) - \frac{1}{2} \frac{D_p(x)}{p^2(x)} \right) Y = 0 \tag{10}$$

(see [1]). If we choose $y(x) = y_2(x)$ then $\tilde{Y}(x) = y_2'(x) p^{-1/2}(x)$ has the same zeros as $y_2'(x)$, i.e. $a, a_1', \ldots, a_{n_2}'$. This is clear for those a_j'-s which are in the open interval (a, b), because p is continuous there. By our assumptions the function p is monotonic hence the limit $\lim_{x\to a+0} p^{-1/2}$ exists and is either finite or infinite. In former case $\tilde{Y}(a) = 0$.

In the latter case $\lim_{x\to a+0} p=0$ and since p is positive and monotonic in (a, b), p is increasing in this interval. By (1) we have for $a<x<\bar{a}_1'$

$$0 > y_2' p^{-1/2} = p^{-1/2} \int_a^x y_2''(\xi)\, d\xi = -p^{-1/2} \int_a^x p(\xi) y_2(\xi)\, d\xi > -p^{1/2} \int_a^x y_2(\xi)\, d\xi,$$

thus in this case, too, we have

$$\lim_{x\to a+0} y' p^{-1/2} = \lim_{x\to a+0} \tilde{Y}(x) = 0.$$

A similar argument is needed if $a_{n_2}'=b$.

Suppose first that $D_p(x)>0$. For $y_1(a)=\tilde{Y}(a)=0$, the differential equation (10) is a Sturm minorant to (1) hence by the Sturmian comparison theorem we can compare the solution $y_1(x)$ of (1) and the solution $\tilde{Y}(x)$ of (10) and we obtain

$$a_j < a_j' \quad j = 1, 2, \ldots, n_2.$$

In the case $D_p(x)<0$ the proof runs in a similar manner which completes the proof of Theorem 2.

It may be of interest to determine those p-s for which $D_p\equiv 0$. Since $D_p=2p^{5/2}(p^{-1/2})''=0$ therefore $p^{-1/2}=Ax+B$ where A and B are real constants. If $A=0$ then $p=$const. if $A\neq 0$ then $p=(Ax+B)^{-2}$. In the former case it is clear that we can choose $b=\infty$ and $a_1, a_2, \ldots$ and $a_1', a_2', \ldots$ exists and $a_j=a_j'$ $(j=1, 2, \ldots)$. In the latter case it may happen that there are no a_j-s and no a_j'-s. This occurs if $A\geqq 2$ when the solutions are nonoscillatory. Therefore one should be careful with supposing the existence of any a_j or a_j' for a given $p(x)$ at all.

Let us consider the function $\varphi_1(x)-\varphi_2(x)$. By making use of the formulas (7) the Wronskian

$$W(y_1, y_2) = y_1 y_2' - y_1' y_2 = -1$$

will have the form

$$W = \varrho_1 \varrho_2 p^{-1/2} \sin(\varphi_1-\varphi_2) = -1.$$

From this it follows immediately that the function $\sin(\varphi_1-\varphi_2)$ does not change its sign on (a, b). By (4) $\varphi_1(a)-\varphi_2(a)=-\pi/2$ thus we conclude that

$$-\pi < \varphi_1(x)-\varphi_2(x) < 0. \tag{11}$$

What is really surprising the function $\cos(\varphi_1-\varphi_2)$, too, does not change its sign under appropriate conditions which enables us to get a sharper estimate for the difference $\varphi_1-\varphi_2$ as the next theorem shows.

THEOREM 3. *If the functions $p'(x)$ and $D_p(x)$ do not change their signs on (a, b) and if $p'(x)$. $D_p(x)>0$ then*

$$-\frac{\pi}{2} < \varphi_1(x)-\varphi_2(x) < 0 \quad \text{if} \quad D_p(x) > 0$$

and

$$-\pi < \varphi_1(x)-\varphi_2(x) < -\frac{\pi}{2} \quad \text{if} \quad D_p(x) < 0$$

for all $a<x<b$.

PROOF. Let the function $\psi(x)$ be defined by

$$\psi(x) = y_1' y_2' + p y_1 y_2$$

then by (1) we have

$$\psi' = p' y_1 y_2 \tag{12}$$

and by (7)

$$\cos(\varphi_1 - \varphi_2) = \frac{\psi}{\varrho_1 \varrho_2}. \tag{13}$$

To prove Theorem 3 it is sufficient to show that the function ψ is positive if $p'>0$ and $D_p>0$ or it is negative if $p'<0$ and $D_p<0$. By (12) the function ψ has local extrema where $y_1(x)=0$ or $y_2(x)=0$, i.e. at $a_0, a_1, \ldots$ and at $\bar{a}_1', \bar{a}_2', \ldots$, respectively, then the two sequences are interlacing:

$$a_0 < \bar{a}_1' < a_1 < \ldots < \bar{a}_j' < a_j < \ldots (\leqq) b. \tag{14}$$

In order to show this we put $x=\bar{a}_j'$ into (11) and by (5) we have

$$\varphi_1(\bar{a}_j') - \varphi_2(\bar{a}_j') < 0 = \varphi_1(a_j) - \varphi_2(\bar{a}_j'),$$

hence $\varphi_1(\bar{a}_j') < \varphi_1(a_j)$, thus by Theorem 1 $\bar{a}_j' < a_j$. Similarly, putting $x=\bar{a}_{j+1}'$ into (11) we obtain

$$\varphi_1(\bar{a}_{j+1}') - \varphi_2(\bar{a}_{j+1}') > -\pi = \varphi_1(a_j) - \varphi_2(\bar{a}_{j+1}'),$$

hence $\varphi_1(\bar{a}_{j+1}') > \varphi_1(a_j)$, and so $a_j < \bar{a}_{j+1}'$, as we stated.

By (7) the function ψ' in (12) can be written in the form

$$\psi' = \frac{p'}{p} \varrho_1 \varrho_2 \sin\varphi_1 \sin\varphi_2. \tag{15}$$

From this and from (5), (14) follows that if $p'>0$ then the function ψ has local minima at $a_0, a_1, \ldots$, and if $p'<0$ then ψ has local maxima also at $a_0, a_1, \ldots$. Considering ψ at these places we have by (7) and (5)

$$\psi(a_j) = y_1'(a_j) y_2'(a_j) = \varrho_1 \varrho_2 \cos\varphi_1(a_j) \cos\varphi_2(a_j) = (-1)^j \varrho_1 \varrho_2 \cos\varphi_2(a_j).$$

If $D_p>0$ on (a, b) then by Theorem 2 and by (14) $\bar{a}_j < a_j < a_j'$ thus by (5)

$$j\pi < \varphi_2(a_j) < \left(j+\frac{1}{2}\right)\pi,$$

and so $\psi(a_j)>0$ for $j=1, 2, \ldots$.

Similarly we get $\psi(x)<0$ for $a<x<b$ if $D_p<0$, which proves Theorem 3.

COROLLARY. *Under the conditions of Theorem* 3 *the sequence of the inequalities*

$$\ldots < a_{j-1}' < \bar{a}_j < \bar{a}_j' < a_j < a_j' < \ldots \quad (j = 1, 2, \ldots) \quad \text{if} \quad D_p(x) > 0, \tag{16}$$

$$\ldots < a_{j-1} < \bar{a}_j' < \bar{a}_j < a_j' < a_j < \ldots \quad (j = 1, 2, \ldots) \quad \text{if} \quad D_p(x) < 0 \tag{17}$$

holds. The sequence breaks down where the next value in question does not exist in $[a, b]$.

PROOF. With respect to Theorem 2 and (14) only the validity of the inequalities

$$a'_{j-1} < \bar{a}_j < \bar{a}'_j \quad (j = 1, 2, \ldots)$$

is still to be proved in the first case. Applying Theorem 3 for $x=\bar{a}_j$ and taking into consideration the values of $\varphi_2(a'_{j-1})$ and $\varphi_2(\bar{a}'_j)$ from (5) we get immediately the desired inequalities. In the second case only the relations

$$\bar{a}'_j < \bar{a}_j < a'_j \quad (j = 1, 2, \ldots)$$

are new and their proof is similar to the former case.

As an application of our results we solve the problem posed by W. LEIGHTON in [3] which reads:

"Consider the (modified) Bessel equation

$$y'' + p(x)y = 0,$$

where

$$p(x) = 1 + \frac{1-4n^2}{4x^2},$$

and suppose there is a non zero solution with zeros at $x=a$ and $x=b$ (that is, b is conjugate to a), $0<a<b$. Prove that if $y(x)$ is any solution such that $y(a)\neq 0$ then the integral

$$\int_b^a [y'^2(x) - p(x)y^2(x)]\,dx > 0$$

when $n^2>1/4$ and negative when $n^2<1/4$."

First we remark that this Bessel differential equation belongs to the class treated in this paper since $p(x)$ is monotonic,

$$D_p(x) = \frac{3}{2}\frac{4n^2-1}{x},$$

$$D_p(x)\,p'(x) = \frac{3(4n^2-1)^2}{4x^7} > 0 \quad (x > 0),$$

and $p(x)$ is nonnegative if $n^2 \leqq 1/4$. If, however, $n^2>1/4$ we should make a restriction on a to ensure $p(x)\geqq 0$:

$$a \geqq \frac{\sqrt{4n^2-1}}{2} \quad \text{if} \quad n^2 > \frac{1}{4}$$

instead of $a>0$.

Let $y_1(x)$ be a solution which has zeros at a and at b and $y'_1(a)=1$. With our notations let $a=a_0$ and $b=a_{n_1}$ $(n_1 \geqq 1)$. Let $y_2(x)$ be defined as above. Then the solution $y(x)$ in question can be written as

$$y(x) = \alpha y_1(x) + \beta y_2(x),$$

where α, β are suitable constants, $\beta\neq 0$, since $y(a)\neq 0$. Consider the integral

$$\Phi = \int_a^b [y'^2 - py^2]\,dx. \tag{18}$$

Since $py=-y''$, integrating by parts we obtain that

$$\Phi = [y(x)y'(x)]_a^b = \alpha\beta[y_1'(b)y_2(b)-y_1'(a)y_2(a)]+\beta^2 y_2(b)y_2'(b).$$

Taking the Wronskian $W(y_1, y_2)=y_1y_2'-y_1'y_2\equiv -1$ at $x=a$ and $x=b$, we get that

$$-y_1'(b)y_2(b) = -y_1'(a)y_2(a) = -1,$$

thus we have

$$\Phi = \beta^2 y_2(b)y_2'(b).$$

By (7) Φ can be expressed as

$$\Phi = \frac{\beta^2}{2}\varrho_2^2(b)p^{-1/2}(b)\sin 2\varphi_2(b).$$

Taking into consideration the relations (16), (17) we have for $b=a_{n_1}$

$$\bar{a}_{n_1}' < b < a_{n_1}' \quad \text{if} \quad |n| > 1/2,$$

or

$$a_{n_1}' < b < \bar{a}_{n_1+1}' \quad \text{if} \quad |n| < 1/2,$$

hence by (5)

$$n_1\pi < \varphi_2(b) < \left(n_1+\frac{1}{2}\right)\pi \quad \text{if} \quad |n| > 1/2$$

or

$$\left(n_1+\frac{1}{2}\right)\pi < \varphi_2(b) < (n_1+1)\pi \quad \text{if} \quad |n| < \frac{1}{2}$$

thus $\operatorname{sign}\Phi=\operatorname{sign}\sin 2\varphi_2(b)=\operatorname{sign}\left(|n|-\frac{1}{2}\right)$, which was to be proved. □

Finally we remark that we have, in fact, proved a more general statement than required in Leighton's problem, namely that if (i) a and b are two zeros of a solution to (1), (ii) $p(x)$ is a positive monotonic function on (a, b) (iii) $p'(x)D_p(x)>0$, on (a, b) then $\Phi D_p(x)>0$ on (a, b) where Φ is defined by (18).

REFERENCES

[1] Борувка, O.: О колеблющихся интегралах дифференциалных линейных уравнений 2-ого порядка, *Čech. Math. J.* **3** (78) (1953), 199—251.
[2] ELBERT, Á.: On the solutions of the differential equation $y''+q(x)y=0$, where $[q(x)]^v$ is concave, II. *Studia Sci. Math. Hung.* **4** (1969), 257—266.
[3] LEIGHTON, W.: Advanced problems, N^r 5794, *The American Math. Monthly*, **78** (1971), 411.
[4] *The American Math. Monthly*, **81** (1974), 462.
[5] MAKAI, E.: Über Eigenwertabschätzungen bei gewissen homogenen linearen Differentialgleichungen zweiter Ordnung, *Compositio Math.* **6** (1939), 368—374.
[6] VOSMANSKÝ, J.: Monotonic properties of zeros and extremants of the differential equation $y''+q(t)y=0$, *Arch. Math. (Brno)*, **6** (1970), 37—73.

Mathematical Research Institute of the Hungarian Academy of Sciences,
Budapest 5, Reáltanoda u. 13—15. H—1053

(Received June 30. 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 207—214

OPTIMUM CONFIDENCE BANDS FOR DENSITY FUNCTIONS

by

R.-D. REISS

Introduction

The different classes of confidence bands for density functions which are known by now do not contain an optimum procedure (see e.g. [1], Corollary, page 1072, and Applications (i), pages 1077—1079, and [5], Sections 4 and 5). Every confidence band in [5], Section 5, is defined pointwise by means of two order statistics which depend on a predetermined integer $m(n)$ for every sample size n. By enlarging this class of confidence bands we can find the optimum order of the integers $m(n)$ as n goes to infinity.

§ 1. Preliminaries

Denote by $\mathbf{R}$ (respectively, $\mathbf{N}$) the set of all real numbers (positive integers). Let P^n denote the independent product of n identical probability measures P. $\mathbf{R}^n$ denotes the Euclidean n-space. The i-th order statistic $Z_{i:n}:\mathbf{R}^n \to \mathbf{R}$, for the sample size n, is defined by $Z_{i:n}(x_1, \ldots, x_n) = z_{i:n}$ where $z_{1:n} \leqq \ldots \leqq z_{n:n}$ are the components of $(x_1, \ldots, x_n) \in \mathbf{R}^n$ arranged in the increasing order.

Given $m(n) \in \{1, \ldots, n-1\}$ and $m(n) < i \leqq n - m(n)$ define

$$p_{m(n),n}(y) := \frac{2m(n)}{n(Z_{i+m(n):n} - Z_{i-m(n):n})} \tag{1.1}$$

for $Z_{i:n} \leqq y < Z_{i+1:n}$. The density estimator $p_{m(n),n}$ was proposed by RÉVÉSZ ([6], Definition 3.2).

Let $0 < \gamma_1 < \gamma_2 < 1$. Let P be a probability measure with the distribution function F and the Lebesgue-density p. Let ξ_i be a solution of the equation $F(y) = \gamma_i$ for $i = 1, 2$. Assume that for some constants $0 < c_1 < c_2 < \infty$

$$c_1 \leqq p(x) \leqq c_2 \tag{1.2}$$

for every $x \in [\xi_1, \xi_2]$. Thus, the inverse function F^{-1} of F exists on (γ_1, γ_2).

Assume that F^{-1} has a third derivative on (γ_1, γ_2) such that

$$|(p \circ F^{-1})(F^{-1})^{(3)}| \leqq A \tag{1.3}$$

on (γ_1, γ_2) for some constant $A > 0$.

EXAMPLE 1.4. The distribution function F with $F(x) = A^{-1/2} \log x$ for $1 \leqq x \leqq \exp(A^{1/2})$ has the property

$$(p \circ F^{-1})(F^{-1})^{(3)} = A \quad \text{on} \quad (0, 1).$$

Let $\mathscr{P}(A, c_1, c_2, \gamma_1, \gamma_2)$ be the family of all probability measures P which fulfill (1.2) and (1.3) (given the constants A, c_1, c_2, γ_1, γ_2). Since our interest mainly concerns the dependence of the optimum sequence $m(n)$, $n\in\mathbf{N}$, from the number A we shall write $\mathscr{P}_A$ in place of $\mathscr{P}(A, c_1, c_2, \gamma_1, \gamma_2)$ for brevity. Hereafter we shall always assume that the number c_1 (respectively, c_2) is sufficiently small (large) so that the probability measure in Example 1.4 is an element of $\mathscr{P}_A$. We remark that our asymptotic results will not depend on the particular values of c_1 and c_2.

Given $\gamma_1<\beta_1<\beta_2<\gamma_2$ and $A>0$ define

$$a_{m(n),n}(t) := \frac{A}{6}\frac{m(n)^2}{n^2}+\left(\log\left(\frac{n}{m(n)}(\beta_2-\beta_1)\right)\right)^{1/2} m(n)^{-1/2}+$$

$$+\left(2t+\log\log\frac{n}{m(n)}-\log\pi\right)\Big/\left(4\left(m(n)\log\frac{n}{m(n)}\right)^{1/2}\right).$$

Keep in mind that $a_{m(n),n}(t)$ also depends on A and $\beta_2-\beta_1$.

§ 2. The results

The density function will be evaluated between the β_1- and β_2-quantile where $0<\beta_1<\beta_2<1$. In statistical problems the quantiles are unknown and, therefore, they are replaced by their consistent estimators $Z_{[n\beta_1]:n}$ and $Z_{[n\beta_2]:n}$. Put $C_n=[Z_{[n\beta_1]:n}, Z_{[n\beta_2]:n}]$.

THEOREM 2.1. *Let $m(n)\in\mathbf{N}$, $n\in\mathbf{N}$, be a sequence such that*

$$n^{1/2+\varepsilon} \leqq m(n) \leqq n^{1-\varepsilon} \quad \text{for some} \quad \varepsilon>0. \tag{2.2}$$

Let $0<\gamma_1<\beta_1<\beta_2<\gamma_2<1$ and $A>0$. Then

$$\begin{matrix}\overline{\lim}\\ \underline{\lim}\\ n\in\mathbf{N}\end{matrix}\Big\{\inf_{P\in\mathscr{P}_A} P^n\{p_{m(n),n}(y)(1-a_{m(n),n}(t)) \leqq p(y) \leqq$$

$$\leqq p_{m(n),n}(y)(1+a_{m(n),n}(t)),\quad y\in C_n\}\Big\}\begin{matrix}\geqq e^{-e^{-t}}\\ \leqq e^{-(1/2)e^{-t}}\end{matrix}$$

uniformly for every $t\in\mathbf{R}$.

The proof of Theorem 2.1 is postponed until Section 3, (3.6). If $n^{1/2+\varepsilon}\leqq m(n)\leqq n^{4/5-\varepsilon}$, $n\in\mathbf{N}$, for some $\varepsilon>0$ then one can easily derive from Lemma 3.1, (3.3), that uniformly for every $P\in\mathscr{P}_A$

$$\lim_{n\in\mathbf{N}} P^n\{p_{m(n),n}(y)(1-a_{m(n),n}(t)) \leqq p(y) \leqq p_{m(n),n}(y)(1+a_{m(n),n}(t)),\ y\in C_n\} = e^{-e^{-t}} \tag{2.3}$$

for every $t\in\mathbf{R}$. We remark that (2.3) is a correction to [5], Theorem 5.3, according to the correction to [1], Theorem 3.1, in [2].

We conjecture that

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{p_{m(n),n}(y)(1-a_{m(n),n}(t)) \leqq p(y) \leqq \tag{2.4}$$

$$\leqq p_{m(n),n}(y)(1+a_{m(n),n}(t)),\ y\in C_n\} = e^{-(1/2)e^{-t}}$$

for every $t\in\mathbf{R}$ if $n^{4/5}\leqq m(n)\leqq n^{1-\varepsilon}$, $n\in\mathbf{N}$, for some $\varepsilon>0$. We do not know whether (2.4) holds true.

The purpose of this paper is to characterize optimum asymptotic confidence bands of the form

$$[p_{m(n),n}(y)(1-a^n),\ p_{m(n),n}(y)(1+a_n)],\quad y\in C_n,$$

at the level $1-\alpha$, $\alpha\in(0,1)$: consequently, a_n, $n\in\mathbf{N}$, is a sequence with the property

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{p_{m(n),n}(y)(1-a_n) \leqq p(y) \leqq p_{m(n),n}(y)(1+a_n),\ y\in C_n\} \geqq 1-\alpha. \tag{2.5}$$

It follows from Theorem 2.1 that

$$a_n := a_{m(n),n}(-\log(-\log(1-\alpha))) \quad \text{fulfills (2.5).} \tag{2.6}$$

The performance of confidence bands is measured by the probability that

$$[p_{m(n),n}(y)(1-a_n),\ p_{m(n),n}(y)(1+a_n)]_{y\in C_n}$$

does not cover functions q which deviate from the true Lebesgue-density p by the amount of $2p(y)b_n$ at some point $y\in C_n$ where $b_n>0$, $n\in\mathbf{N}$, is a suitably chosen sequence. The sequences $m(n)$, $n\in\mathbf{N}$ which are optimum in this sense have also the property of minimizing (asymptotically) the values of $a_{m(n),n}$.

THEOREM 2.7. *Let* $0<\gamma_1<\beta_1<\beta_2<\gamma_2<1$, $A>0$ *and* $\alpha\in(0,1)$. *There exists a sequence* $a_n>0$, $n\in\mathbf{N}$, *such that*

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{p_{m(n),n}(y)(1-a_n) \leqq p(y) \leqq p_{m(n),n}(y)(1+a_n),\ y\in C_n\} \geqq 1-\alpha, \tag{2.8}$$

and for every $B>\left(\frac{125}{48}A\right)^{1/5}$

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{[p_{m(n),n}(y)(1-a_n),\ p_{m(n),n}(y)(1+a_n)] \subset \tag{2.9}$$

$$\subset \left[p(y)\left(1-B\left(\frac{\log n}{n}\right)^{2/5}\right),\ p(y)\left(1+B\left(\frac{\log n}{n}\right)^{2/5}\right)\right],\ y\in C_n\} > 0$$

iff

$$m(n) \sim \left(\frac{9}{20}\right)^{1/5} A^{-2/5}(\log n)^{1/5} n^{4/5}. \tag{2.10}$$

Theorem 2.7 will be proved in (3.7). We remark that

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{[p_{m(n),n}(y)(1-a_n),\ p_{m(n),n}(y)(1+a_n)] \subset$$

$$\subset \left[p(y)\left(1-B\left(\frac{\log n}{n}\right)^{2/5}\right),\ p(y)\left(1+B\left(\frac{\log n}{n}\right)^{2/5}\right)\right],\ y\in C_n\} = 1$$

for every $B>\left(\frac{125}{48}\right)A\Big)^{1/5}$ if $m(n)$, $n\in\mathbf{N}$, fulfills (2.10) and a_n is defined as in (2.6).

§ 3. Auxiliary results and proofs

Let

$$a'_{m(n),n}(t) := \left(\log\left(\frac{n}{m(n)}(\beta_2-\beta_1)\right)\right)^{1/2} m(n)^{-1/2} + \frac{2t+\log\log\frac{n}{m(n)}-\log\pi}{4\left(m(n)\log\frac{n}{m(n)}\right)^{1/2}}.$$

The proofs of Theorems 2.1 and 2.7 are based on

LEMMA 3.1. *Let* $0<\gamma_1<\beta_1<\beta_2<\gamma_2<1$ *and* $c_1, c_2, C>0$. *Assume that* $m(n)$, $n\in\mathbf{N}$, *fulfills condition* (2.2). *Uniformly for every probability measure* P *which fulfills* (1.2) *(for* c_1, c_2*) and has a Lebesgue-density* p *such that*

$$|p(x)-p(y)| \leqq C|x-y| \quad \text{for every} \quad x, y\in(\xi_1, \xi_2) \tag{3.2}$$

the following holds true: For every $t\in\mathbf{R}$

$$\lim_{n\in\mathbf{N}} P^n\left\{\max_{[n\beta_1]\leqq i\leqq[n\beta_2]}\left|\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-e_{i,n}\right| \leqq a'_{m(n),n}(t)\right\} = e^{-e^{-t}}, \tag{3.3}$$

and

$$\lim_{n\in\mathbf{N}} P^n\left\{\max_{[n\beta_1]\leqq i\leqq[n\beta_2]}\left(\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-e_{i,n}\right) \leqq a'_{m(n),n}(t)\right\} = e^{-(1/2)e^{-t}} \tag{3.4}$$

where $e_{i,n}$ *is defined by*

$$e_{i,n} := \frac{n}{2m(n)}\, p\left(F^{-1}\left(\frac{i}{n}\right)\right)\left\langle F^{-1}\left(\frac{i+m(n)}{n}\right)-F^{-1}\left(\frac{i-m(n)}{n}\right)\right\rangle.$$

PROOF. I. The proof of (3.3) is based on results of BICKEL and ROSENBLATT ([1], Theorem 3.1) and KIEFER ([3], Theorem 2).

Define the "natural" kernel type estimator f_n by

$$f_n^{(x_1,\ldots,x_n)}(y) = \frac{1}{2nb(n)}\sum_{i=1}^{n} 1_{[-1,1]}\left(\frac{y-x_i}{b(n)}\right)$$

for every sample $(x_1, \ldots, x_n)\in\mathbf{R}^n$ and $y\in\mathbf{R}$ where $b(n)>0$, $n\in\mathbf{N}$, is a sequence such that

$$n^{\varepsilon} \leqq b(n)^{-1} \leqq n^{1/2-\varepsilon}, \quad n\in\mathbf{N},$$

for some $\varepsilon>0$. Let Q denote the uniform distribution on $(0, 1)$. We have

$$\lim_{n\in\mathbf{N}} Q^n\left\{\sup_{\beta_1\leqq y\leqq\beta_2}|(2nb(n))^{1/2}(f_n(y)-1) \leqq\right. \tag{1}$$

$$\left.\leqq\left(2\log\frac{\beta_2-\beta_1}{b(n)}\right)^{1/2}+\frac{2t+\log\log b(n)^{-1}-\log\pi}{2(2\log b(n)^{-1})^{1/2}}\right\} = e^{-e^{-t}}$$

for every $t\in\mathbf{R}$. ((1) can be proved in the same way as [1], Theorem 3.1. (1) is slightly more general than [1], Theorem 3.1, as far as the sequence $b(n)$, $n\in\mathbf{N}$, and the interval $[\beta_1, \beta_2]$ are concerned; furthermore, Q does not fulfill the conditions (A2) and (A3) in [1]. Notice that Theorem 3.1 in [1] is corrected in [2].)

It is a consequence of the Bernstein inequality that for every $s \in \mathbf{N}$

$$Q^n\left\{(nb(n))^{1/2} \sup_{\frac{i}{n} \leqq y \leqq \frac{i+1}{n}} \left|f_n\left(\frac{i}{n}\right) - f_n(y)\right| \geqq \frac{\log n}{(nb(n))^{1/2}}\right\} \leqq$$

$$\leqq Q^n\left\{(nb(n))^{-1/2} \sum_{i=1}^{n} \left(1_{\left[\frac{i}{n}-b(n), \frac{i+1}{n}-b(n)\right] \cup \left[\frac{i}{n}+b(n), \frac{i+1}{n}+b(n)\right]}(x_i) - \frac{2}{n}\right) \geqq\right.$$

$$\left. \geqq \frac{\log n}{4(nb(n))^{1/2}}\right\} = O(n^{-s})$$

uniformly for every i with $nb(n) < 1 < n - nb(n)$.

This together with (1) implies

$$\lim_{n \in \mathbf{N}} Q^n\left\{\max_{[n\beta_1] \leqq i \leqq [n\beta_2]} \left|(2nb(n))^{1/2}\left(f_n\left(\frac{i}{n}\right) - 1\right)\right| \leqq\right. \tag{2}$$

$$\left. \leqq \left(2 \log \frac{\beta_2 - \beta_1}{b(n)}\right)^{1/2} + \frac{2t + \log\log b(n)^{-1} - \log \pi}{2(2 \log b(n)^{-1})^{1/2}}\right\} = e^{-e^{-t}}$$

for every $t \in \mathbf{R}$.

Let $\hat{f}_n$ be defined as f_n with $1_{[-1,1)}$ in place of $1_{[-1,1]}$. It is clear that (2) also holds for $\hat{f}_n$ in place of f_n. Notice that

$$\hat{f}_n(y) = \frac{1}{2b(n)} \langle F_n(y + b(n)) - F_n(y - b(n)) \rangle$$

where F_n denotes the empirical distribution function for the sample size n.

(2) will be applied with $b(n) = m(n)/n$ (with $\hat{f}_n$ in place of f_n).

It follows from [3], Theorem 2, (see also [5], (5.2)) that

$$\lim_{n \in \mathbf{N}} Q^n\left\{\frac{m(n)}{n^{1/4+\delta}} \max_{m(n) < i < n - m(n)} \left|\left(\frac{1}{p_{m(n),n}(Z_{i:n})} - 1\right) + \left(\hat{f}_n\left(\frac{i}{n}\right) - 1\right)\right| \geqq \varepsilon\right\} = 0$$

for every $\delta > 0$ and $\varepsilon > 0$.

This together with (2) implies

$$\lim_{n \in \mathbf{N}} Q^n\left\{\max_{[n\beta_1] \leqq i \leqq [n\beta_2]} \left|\frac{1}{p_{m(n),n}(Z_{i:n})} - 1\right| \leqq a'_{m(n),n}(t)\right\} = e^{-e^{-t}} \tag{3}$$

for every $t \in \mathbf{R}$.

Using the inverse probability integral transformation for order statistics we get

$$P^n * \left(\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}\right)_{i=[n\beta_1]}^{[n\beta_2]} = \tag{4}$$

$$= Q^n * \left(p(F^{-1}(Z_{i:n})) \frac{n(F^{-1}(Z_{i+m(n):n}) - F^{-1}(Z_{i-m(n):n}))}{2m(n)}\right)_{i=[n\beta_1]}^{[n\beta_2]}$$

where F^{-1} denotes the generalized inverse of F. ($P^n * g$ denotes the measure induced by P^n and the measurable function g.)

Define

$$B_n := \bigcap_{i=[n\beta_1]}^{[n\beta_2]} \left(\left\{ \left| Z_{i:n} - \frac{i}{n} \right| \leqq \frac{\log n}{n^{1/2}} \right\} \cap \left\{ \left| Z_{i+m(n):n} - Z_{i-m(n):n} - \frac{2m(n)}{n} \right| \leqq \frac{m(n)^{1/2}}{n} \log n \right\} \right).$$

It follows from [4], Lemmas (7.9) and (7.10) that

$$\lim_{n \in \mathbf{N}} Q^n(B_n) = 1. \tag{5}$$

The following considerations hold uniformly for all $P \in \mathscr{P}_A$. (1.2) and (3.2) imply that uniformly on B_n and uniformly for every $r \in \mathbf{N}$ with $[n\beta_1] - m(n) \leqq r \leqq [n\beta_2] + m(n)$

$$F^{-1}(Z_{r:n}) = F^{-1}\left(\frac{r}{n}\right) + \frac{1}{p \circ F^{-1}\left(\frac{r}{n}\right)} \left(Z_{r:n} - \frac{r}{n} \right) + O\left(\frac{(\log n)^2}{n} \right). \tag{6}$$

Moreover, as the reader will easily verify

$$p(F^{-1}(Z_{i:n})) = p\left(F^{-1}\left(\frac{i}{n} \right) \right) + O\left(\frac{\log n}{n^{1/2}} \right) \tag{7}$$

uniformly on B_n, and

$$\left| \frac{1}{p(F^{-1}(y))} - \frac{1}{p(F^{-1}(x))} \right| = O(|x-y|) \tag{8}$$

for every $x, y \in (\gamma_1, \gamma_2)$. By (7), (8) and (6), applied for $r = i + m(n)$ and $r = i - m(n)$,

$$\frac{n}{2m(n)} p(F^{-1}(Z_{i:n})) \langle F^{-1}(Z_{i+m(n):n}) - F^{-1}(Z_{i-m(n):n}) \rangle = \tag{9}$$

$$= \frac{1}{p_{m(n),n}(Z_{i:n})} + e_{i,n} + O\left(\frac{\log n}{n^{1/2}} + \frac{(\log n)^2}{m(n)} \right)$$

uniformly on B_n.

(3.3) is an immediate consequence of (3), (4), (5), and (9).

II. The proof of (3.4) is based on the following result:

$$\lim_{n \in \mathbf{N}} Q^n \left\{ \sup_{\beta_1 \leqq y \leqq \beta_2} ((2nb(n))^{1/2}(f_n(y) - 1)) \leqq \left(2 \log \frac{\beta_2 - \beta_1}{b(n)} \right)^{1/2} + \right. \tag{10}$$

$$\left. + \frac{2t + \log \log b(n)^{-1} - \log \pi}{2(2 \log b(n)^{-1})^{1/2}} \right\} = e^{-\frac{1}{2} e^{-t}}$$

for every $t \in \mathbf{R}$. (To prove (10) use [1], Theorem A1, (A.2), instead of [1], Corollary A1, which was applied to prove [1], Theorem 3.1) (3.4) follows from (10) in the same way as (3.3) from (1).

REMARK 3.5. The conditions (1.2) and (1.3) imply condition (3.2) with some constant C which only depends on c_1, c_2, A and $\gamma_2 - \gamma_1$. Thus, Lemma 3.1 is applicable to the probability measures $P \in \mathscr{P}_A$.

(3.6) PROOF OF THEOREM 2.1. For every $P\in\mathscr{P}_A$ let $e_{i,n}$ be defined as in Lemma 3.1. Uniformly for all i with $[n\beta_1]\leqq i\leqq[n\beta_2]$

$$|e_{i,n}-1|\leqq\frac{A}{6}\left(\frac{m(n)}{n}\right)^2+O\left(\left(\frac{m(n)}{n}\right)^3\right). \tag{1}$$

Let Q be a probability measure such that

$$e_{i,n}-1=\frac{A}{6}\left(\frac{m(n)}{n}\right)^2+O\left(\left(\frac{m(n)}{n}\right)^3\right) \tag{2}$$

uniformly for all i with $[n\beta_1]\leqq i\leqq[n\beta_2]$ (see Example 1.4). Lemma 3.1 together with (1) and (2) implies

$$Q^n\left\{\max\left(\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-e_{i,n}\right)\leqq a'_{m(n),n}(t)\right\}=$$

$$=Q^n\left\{\max\left(\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-1\right)\leqq a_{m(n),n}(t)\right\}+o(1)\geqq$$

$$\geqq\inf_{P\in\mathscr{P}_A}P^n\left\{\max\left|\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-1\right|\leqq a_{m(n),n}(t)\right\}+o(1)\geqq$$

$$\geqq\inf_{P\in\mathscr{P}_A}P^n\left\{\max\left|\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-e_{i,n}\right|\leqq a'_{m(n),n}(t)\right\}+o(1)$$

for every $t\in\mathbf{R}$ where the maximum ranges over all i with $[n\beta_1]\leqq i\leqq[n\beta_2]$.

Thus, applying Lemma 3.1 again we find

$$\left.\begin{matrix}\overline{\lim_{n\in\mathbf{N}}}\\ \underline{\lim_{n\in\mathbf{N}}}\end{matrix}\right\}\left\{\inf_{P\in\mathscr{P}_A}P^n\left\{\max_{[n\beta_1]\leqq i\leqq[n\beta_2]}\left|\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-1\right|=a_{m(n),n}(t)\right\}\right\}\begin{matrix}\geqq e^{-e^{-t}}\\ \leqq e^{-\frac{1}{2}e^{-t}}\end{matrix} \tag{3}$$

uniformly for every $t\in\mathbf{R}$.

By (1) and (3.3)

$$\lim_{n\in\mathbf{N}}P^n\{p_{m(n),n}(Z_{i:n})\geqq c_1/2,\ [n\beta_1]\leqq i\leqq[n\beta_2]\}=1, \tag{4}$$

and

$$\lim_{n\in\mathbf{N}}P^n\left\{|Z_{i+1:n}-Z_{i:n}|\leqq\frac{1}{n^{1-\varepsilon}},\ [n\beta_1]\leqq i\leqq[n\beta_2]\right\}=1 \tag{5}$$

for every $\varepsilon>0$ uniformly for all $P\in\mathscr{P}_A$.

(5) follows from (5), proof of Lemma 3.1, by means of the inverse probability integral transformation.

Thus,

$$\lim_{n\in\mathbf{N}}P^n\left\{\max_{Z_{i:n}\leqq y\leqq Z_{i+1:n}}\left|\frac{p(Z_{i:n})}{p_{m(n),n}(Z_{i:n})}-\frac{p(y)}{p_{m(n),n}(y)}\right|\leqq\right. \tag{7}$$

$$\left.\leqq\frac{1}{n^{1-\varepsilon}},\ [n\beta_1]\leqq i\leqq[n\beta_2]-1\right\}=1$$

for every $\varepsilon>0$ uniformly for all $P\in\mathscr{P}_A$.

(3) and (7) together imply the assertion.

(3.7) PROOF OF THEOREM 2.7. Theorem 2.1 implies that

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\left\{\left|\frac{p(y)}{p_{m(n),n}(y)}-1\right| \leqq a_n,\ y\in C_n\right\} \geqq 1-\alpha \tag{1}$$

iff

$$a_n \geqq a_{m(n),n}(t_0), \quad n\in\mathbf{N}$$

for some suitably chosen $t_0\in\mathbf{R}$. Put $b_n := B\left(\frac{\log n}{n}\right)^{2/5}$ where $B>\left(\frac{125}{48}A\right)^{1/}$. According to (1)

$$\lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\{[p_{m(n),n}(y)(1-a_n),\ p_{m(n),n}(y)(1+a_n)] \subset \tag{2}$$

$$\subset [p(y)(1-b_n),\ p(y)(1+b_n)],\ y\in C_n\} =$$

$$= \lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\left\{\frac{1+a_n}{1+b_n} \leqq \frac{p(y)}{p_{m(n),n}(y)} \leqq \frac{1-a_n}{1-b_n},\ y\in C_n\right\} =$$

$$= \lim_{n\in\mathbf{N}} \inf_{P\in\mathscr{P}_A} P^n\left\{\left|\frac{p(y)}{p_{m(n),n}(y)}-1\right| \leqq b_n-a_n+O\left(\left(\frac{\log n}{n}\right)^{4/5}\right),\ y\in C_n\right\} > 0$$

iff

$$b_n-a_n \geqq a_{m(n),n}(t_1) \quad \text{for some} \quad t_1\in\mathbf{R}.$$

Thus, (1) and (2) hold for some sequence a_n, $n\in\mathbf{N}$, iff

$$b_n \geqq a_{m(n),n}(t_0)+a_{m(n),n}(t_1) \tag{3}$$

for some $t_0, t_1\in\mathbf{R}$.

By elementary calculations it is seen that (3) holds for every $B>\left(\frac{125}{48}A\right)^{1/5}$ iff $m(n)$, $n\in\mathbf{N}$, fulfills condition (2.10).

REFERENCES

[1] BICKEL, P. J. and ROSENBLATT, M.: On some global measures of the deviation of density function estimates, *Ann. Statist.* **1** (1973), 1071—1095.

[2] BICKEL, P. J. and ROSENBLATT, M.: Corrections to "On some global measures of the deviations of density function estimates", *Ann. Statist.* **3** (1975), 1370.

[3] KIEFER, J.: Deviations between the sample quantile process and the sample d.f., in: *Proc. First Int. Conf. Nonpar. Inf.* (1969), 299—319, Cambridge Univ. Press, 1970.

[4] REISS, R.-D.: The asymptotic normality and asymptotic expansions for the joint distribution of several order statistics, in: *Coll. Math. Soc. János Bolyai* **11** (1974), 297—340, North Holland, 1975.

[5] REISS, R.-D.: Approximate distributions of the maximum deviation of histograms, *Metrika* **25** (1978), 9—26.

[6] RÉVÉSZ, P.: On empirical density functions, *Period. Math. Hungar.* **2** (1972), 85—110.

Institut für Mathematische Stochastik, Universität Freiburg, Hermann-Herder-Str. 10, 7800 Freiburg, West Germany

(Received September 28, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 215—227

CAUCHY PROBLEMS FOR SYSTEMS OF LINEAR SINGULAR PARTIAL DIFFERENTIAL EQUATIONS

by

A. SZÉP

1. Introduction

The immediate predecessors of the present paper are those of M. S. BAOUENDI and C. GOULAOUIC [1], [2]. Actually we shall deal with systems of the form

$$\sum_{k=0}^{N} A_k(x)(tD_t)^k u - \sum_{j=0}^{N} C_{N-j} t(D_t t)^j u = f. \tag{1.1}$$

Here $x=(x_1, x_2, \ldots, x_m)\in \mathbf{C}^m$, t a scalar, $D_t=\partial/\partial t$, the $A_k(x)$-s and C_{N-j}-s are $n\times n$ matrices, f and u are n-vectors. The components of f and of the unknown u depend on t and x, the entries of the $A_k(x)$-s only on x, whereas the entries of the C_{N-j}-s are linear differential operators with respect to the x_j-s, and, moreover, depend on t in a way to be specified later. The references to [1] and [2] provide a great variety of problems which can be reduced to the form (1.1). The precise results can be described only after introducing suitable function spaces. We shall deal separately with two different cases: analytic (A) and continuous (C). This distrinction will be made concerning the time variable t which is allowed to be complex in case (A).

Analyticity with respect to x is assumed throughout. The reason for dealing with the two cases separately is natural regarding the simple ordinary differential equation

$$ty'+\alpha y = g(t). \tag{1.2}$$

In case (A) ($g(t)$ analytic at $t=0$) equation (1.2) has a unique analytic solution at $t=0$ iff $\alpha\neq 0, 1, 2, \ldots$ while if $g(t)$ is merely continuous (case (C)) and $\operatorname{Re}\alpha<0$ then there exists a unique solution continuously differentiable for $t\neq 0$ such that $\lim_{t\to 0} ty'(t)$ exists.

We emphasize that in contrast to the title for (1.1) no initial conditions are prescribed. We shall immediately see how initial conditions can be "built in" a given equation.

2. Assumption and tools

We introduce some local concepts and still do not make assumptions on domains of definition beyond necessity. The real variable t varies in a neighbourhood of $t=0$, $x\in\mathbf{C}^m$. For a multiindex $\alpha=(\alpha_1, \alpha_2, \ldots, \alpha_m)$ with nonnegative integer entries we define

$$D_x^\alpha = \partial^{|\alpha|}/\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\ldots\partial_m^{\alpha_m}, \quad (\alpha)=\alpha_1+\alpha_2+\ldots+\alpha_m.$$

DEFINITION 2.1. A *differential monomial* is a formal expression $c(t, x)t^l D_t^p D_x^\beta$ where $c(t, x)$ is continuous in t at $t=0$ and satisfies $c(0, x) \not\equiv 0$.

The *weight* of the monomial is $p-l$. (Here p and l are non-negative integers.)

We shall start with the following Cauchy problem:

$$Bu = f \tag{2.1}$$

where $B=(b_{ik})_{i,k=1}$ denotes an $n\times n$ matrix whose entries are linear differential operators representable as sums of monomials.

We shall make some assumptions on the weight and order of monomials occurring in B.

(A1) The monomials of minimal weight do not contain differentiation with respect to the coordinates x_i.

Suppose that the highest order of differentiation with respect to t is N (>0).

If the minimal weight in (A1) is positive, that is there exists a monomial of minimal weight of the form $c(t, x)t^k D_t^N D_x^\beta$ $(0 \leqq k < N)$, then we may prescribe Cauchy-conditions

$$D_t^l u|_{t=0} = \varphi_l(x) \quad l = 0, 1, \ldots, N-k-1.$$

Assuming differentiability of the solutions this is equivalent to

$$u(t, x) = \sum_{l=0}^{N-k-1} \frac{t^l}{l!} \varphi_l(x) + t^{N-k} v(t, x). \tag{2.2}$$

Substitute (2.2) into (2.1). We can see that concerning $v(t, x)$ as unknown function, B contains no term of positive weight and assumption (A1) holds.

Assume that the coefficient functions in the operator matrix B are suitably differentiable with respect to t. Applying the identity $D_t t = tD_t + 1$ and using Taylor expansion for the coefficient functions, we arrive at the form (1.1). Note that the right hand side obtained is different from the original one in (1.1), it contains a simple linear expression of the functions $\varphi_l(x)$ and their derivatives that are supposed to exist. Concerning the operator matrices C_{N-j} we make the following assumption.

(A2) Let there be given non-negative integers $m_1, m_2, \ldots, m$. Suppose that

$$C_{N-j} = (c_{ik}^{N-j})_{i,k=1}^n \quad (j = 0, 1, \ldots, N-1)$$

and

$$\operatorname{ord} c_{N-j}^{ik} \leqq m_i - m_k + N - j.$$

We turn to introduce the function spaces necessary for stating and proving our theorems.

Let Ω be a bounded non-void subset of $\mathbf{C}^m$. We define its polycylindrical hull of radius s as the union

$$\Omega_s = \bigcup_{a \in \Omega} B(a, s)$$

where $s>0$ is fixed and $B(a, s)$ is defined by

$$B(a, s) = \{z:\ z \in \mathbf{C}^m,\ z = (z_1, z_2, \ldots, z_m),\ |z_i - a_i| < s\}.$$

Denote by $E_s(\Omega)$ the set of complex valued functions analytic in Ω_s and continuous on $\overline{\Omega_s}$. $E_s(\Omega)$ is a Banach-space with respect to the supremum norm. Denote by $\mathscr{H}^m$

the set of complex valued functions analytic on $\mathbf{C}^m$ and by $X_s = X_s(\Omega)$ the closure of $\mathscr{H}^m$ in $E_s(\Omega)$. For an $x \in X_s$, its norm will be denoted simply by $\|x\|_s$. Moreover, consider the space of power series

$$x = \sum_{l=0}^{\infty} x_l t^l$$

such that $x_l \in X_s$ $(l=0, 1, \ldots)$ and

$$\|x\|_{s,T} = \sum_{k=0}^{\infty} \|x_l\|_s T^l < +\infty. \tag{2.3}$$

The set of all such power series will be again a Banach space which we denote by $X_{s,T}$, with the norm (2.3). The elements of this space are analytic $\{(t) < T\} \times \Omega_s \subset \subset \mathbf{C}^{m+1}$ by Weierstrass' theorem.

We note that $X_{s,T}$ is a natural generalization of the space used in [4] to prove Lettenmeyer's theorem. A straighforward calculation shows that $X_{s,T}$ is also a Banach-algebra with respect to the given norm.

An elementary application of the Cauchy integral formula gives the following

LEMMA 2.1. (M. S. BAOUENDI, C. GOULAOUIC, [1]) *For any i, differentiation $\partial/\partial z_i$ is a bounded linear map from X_s into X_r $(s > r > 0)$ and for $u \in X_s$*

$$\left\| \frac{\partial u}{\partial z_i} \right\|_r \leqq \frac{1}{s-r} \|u\|_s. \tag{2.4}$$

Applying (2.4) we have

LEMMA 2.2. (C) *Let P be a linear differential operator of the form*

$$P = \sum_{|\alpha| \leqq j} c_\alpha(t, x) D_x^\alpha$$

then if $0 < s_1 < s_2 < s_0$, $0 < T$, $u \in C([-T, T], X_{s_2})$ and for all $|\alpha| \leqq j$, $c_\alpha \in C([-T, T], X_{s_0})$ then

$$\sup_{|t| \leqq T} \|Pu\|_{s_1} \leqq \frac{M_c}{(s_2 - s_1)^j} \sup_{(t) \leqq T} \|u\|_{s_2} \tag{2.5}$$

with a suitable constant M_c depending only on the coefficients c_α but not on s_1, s_2.

LEMMA 2.3. (A) *Let P (formally), s_0, s_1, s_2, T be the same as above, $u \in X_{s_2,T}$ and for all α, $|\alpha| \leqq j$, $c_\alpha \in X_{s_0,T}$. Then*

$$\|Pu\|_{s_1,T} \leqq \frac{M_A}{(s_2 - s_1)^j} \|u\|_{s_2,T} \tag{2.6}$$

with a suitable constant M_A depending only on the coefficients c_α but not on s_1, s_2.

PROOF. Expand u into a power series with respect to the variable t. Differentiating the coefficients, Lemma 2.1 can be applied. For estimating the products the Banach-algebra-property of the $X_{s,T}$ space can be used.

REMARK 2.1. Actually only inequalities (2.5) and (2.6) are needed for the proofs, hence the results immediately extend to pseudodifferential operators or, more generally, to systems of operator equations on scales of Banach spaces.

Consider the principal part in (1.1), i.e.

$$\mathscr{L}u = \sum_{k=0}^{N} A_k(x)(tD_t)^k u.$$

Note that here A_k denotes an $n\times n$ matrix with entries depending analytically on x and $A_N \equiv I$ (identity).

DEFINITION 2.2. The λ-matrix (or matrix pencil)

$$Q(\lambda) = \sum_{k=0}^{N} A_k(x)\lambda^k \tag{2.7}$$

will be called the characteristic polynomial (or indicial polynomial) of the operator $\mathscr{L}$ (or corresponding to equation (1.1)).

DEFINITION 2.3. Let W be an arbitrary Banach-space, $C([-T, T], W)$ the space of functions mapping $[-T, T]$ continuously into W. The operator tD_t is densely defined on $C([-T, T], W)$. Define

$$C^N = C^N([-T, T], W) = \bigcap_{k=0}^{N} \{w: (tD_t)^k w \in C(-[T, T], W)\}.$$

LEMMA 2.4. *Suppose that the matrices $A_k(x)$ are scalar multiples of the identity, $A_k(x)=a_k(x)I$, $k=0, 1, \ldots, N$, and for the roots $\lambda_1(x), \lambda_2(x), \ldots, \lambda_N(x)$ of (the polynomial $Q(\lambda)$ the inequality*

$$\operatorname{Re}\lambda_i(x) < -\Lambda < 0 \quad (i = 1, 2, \ldots, N)$$

holds for the values x under consideration. Then there exists a linear integral operator H_c so that if $f=f(t, x)$ is continuous in t for $|t|\leqq T$, then $\mathscr{L}H_c f=f$ and $u=H_c f\in C^N$ is unique. Moreover if $a_k\in X_s$ $(k=0, 1, \ldots, N-1)$ then the operator $(D_t t)^N H_c$ is a bounded linear map from $C([-T, T], X_s)$ into itself for any $0<T_1\leqq T$, with operator norm uniformly bounded for $0<T_1\leqq T$.

PROOF. Suppose $\mathscr{L}u=f$. By the assumption we have essentially n-times the same Euler-type singular ordinary differential equation containing the variable x as a parameter. For one copy of the equation (with u and f scalars) introduce the new (scalar) variables $y_1, y_2, \ldots, y_N$ as follows (cf. [5], p. 84):

$$y_j = (tD_t)^{j-1}u \quad (j = 1, 2, \ldots, N).$$

Then

$$ty_j' = y_{j+1} \quad (j = 1, 2, \ldots, N)$$

$$ty_N' = -a_0(x)y_1 - a_1(x)y_2 - \ldots - a_{N-1}(x)y_{N-1} + f.$$

It follows easily that $y \in C$ iff $u \in C^N$. Introduce the matrix

$$L = \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & 0 & 1 & & & \\ & & & \ddots & \ddots & & \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ -a_0(x) & -a_1(x) & \dots & & & & -a_N(x) \end{bmatrix}$$

and the vectors $\tilde{f}=(0, 0, \dots, 0, f)$, $y=(y_1, y_2, \dots, y_N)$. We have the following ordinary inhomogeneous linear system

$$ty' = Ly + \tilde{f}.$$

It is well known that the corresponding homogeneous equation has the fundamental solution $t^L = e^{L \log t}$ and variation-of-constant yields

$$y(t) = t^L c + t^L \int_0^t s^{-L-I} \tilde{f}(s, x)\, ds = \tag{2.8}$$

$$= t^L c + \int_0^1 \sigma^{-(L+I)} \tilde{f}(\sigma t, x)\, d\sigma.$$

By (2.7) and the definition of Λ

$$\|\sigma^{-(L+I)}\| \leqq K\sigma^{-\Lambda-1}$$

where $\|\cdot\|$ denotes some matrix-norm and K is a positive constant depending on the type of this norm and on the difference

$$-\Lambda - \sup_{i,x} \lambda_i(x) > 0,$$

where x varies in the domain under consideration. Thus the integral in (2.8) converges and depends continuously on t.

On the other hand all coordinates of y remain bounded iff $c=0$ in (2.8). Thus

$$u = Hf = P_1 \int_0^1 \sigma^{-L-I} \tilde{f}(\sigma t, x)\, d\sigma \tag{2.9}$$

is uniquely determined where P_1 denotes projection on the first coordinate. The last assertion follows from the last row of (2.8).

REMARK 2.2. (2.9) shows that u depends analytically on x whenever L and f do.

REMARK 2.3. Essentially the same statement was proved in [1] without deducing the direct representation (2.9).

LEMMA 2.5. (C) [1]. *For a constant* $\alpha>0$, *an arbitrary Banach-space* W, $v\in C([-\alpha,\alpha], W)$ $1\leqq k\leqq N$ *introduce the operator*

$$\mathscr{H}_C^k v = \int_0^1 \int_0^1 \dots \int_0^1 v(\sigma_1\sigma_2 \dots \sigma_k t)\, d\sigma_1\, d\sigma_2 \dots d\sigma_k. \tag{2.10}$$

Then $\mathscr{H}_C^k$ *is a linear bounded operator from* $C([-\alpha,\alpha], W)$ *into* $C^k([-\alpha,\alpha], W)$.

PROOF. Induction yields the statement.

In the sequel we give the analytic counterpart of the above results.

DEFINITION 2.4. For a positive integer N denote by $X_{s,T}^N$ the space of power series

$$x = \sum_{k=0}^{\infty} x_k t^k,$$

such that $x_k\in X_s$ and

$$\|x\|_{s,T}^N = \sum_{k=0}^{\infty} \|x_k\|_s (k+1)^N T^k < +\infty. \tag{2.11}$$

Again, $X_{s,T}^N$ is a Banach-space of analytic (in t and x) functions with the norm (2.11).

LEMMA 2.6. *Allow the* A_k*-s in the operator* $\mathscr{L}$ *to be arbitrary* $n\times n$ *matrices with entries belonging to* X_s. *Suppose that for any non-negative integer* p, $Q(p)=\sum_{l=0}^{N} A_l(x)p^l$ *is invertible, moreover* $\|Q(p)^{-1}\|<B_0$ *for* $x\in\bar{\Omega}_s$ *where* $\|\cdot\|$ *is some matrix norm and* B_0 *denotes a constant. For convenience for any Banach space* S *denote by* $\bar{S}=S\times S\times\dots\times S$ *(n-times) the product space endowed with the norm defined as the sum of norms. Then there exists a linear operator* H_A *such that if* $f\in\bar{X}_{s,T}$ *then* $\mathscr{L}H_A f=f$ *and* $u=H_A f\in\overline{X_{s,T}^N}$ *is unique. Moreover the operator* $(D_t t)^N H_A$ *is a bounded linear map from* $\bar{X}_{s,T}$ *into itself for any* $0<T_1\leqq T$ *with operator norm uniformly bounded for* $0<T_1\leqq T$.

PROOF. Let $f=\sum_{k=0}^{\infty} x_k t^k$, $x_k\in\bar{X}_s$. By the uniqueness theorem for power series the only possibility for $u=H_A f$ is

$$u = H_A f = \sum_{k=0}^{\infty} Q(k)^{-1} x_k t^k.$$

Since

$$Q(k) = k^N\left(I+\frac{1}{k}A_{N-1}(x)+\dots+\frac{1}{k^N}A_0(x)\right),$$

if k is large enough, $k\geqq k_0$ the matrix in the bracket is close to the identity, with respect to a given matrix norm, uniformly for $x\in\bar{\Omega}_s$. Thus

$$(Q(k))^{-1} = \frac{1}{k^N}\left(I+\frac{1}{k}A_{N-1}(x)+\dots+\frac{1}{k^N}A_0(x)\right)^{-1},$$

whence

$$(k+1)^N(Q(k))^{-1} = \left(\frac{k+1}{k}\right)^N \left(I+\frac{1}{k}A_{N-1}(x)+\ldots+\frac{1}{k^N}A_0(x)\right)^{-1}.$$

For $k \geqq k_0$ the last matrix is bounded in norm, while for $k=0, 1, \ldots, k_0-1$ we have to take the maximum of finitely many quantities to get

$$\sup_{\substack{k=0,1,\ldots \\ x\in\bar{\Omega}_s}} (k+1)^N \|Q(k)^{-1}\| < +\infty,$$

whence our statement follows.

DEFINITION 2.5. For $x \in X_{s,T}$ we introduce the operator

$$\mathscr{H}_A^l: \sum_{k=0}^{\infty} x_k t^k \to \sum_{k=0}^{\infty} \frac{1}{(k+1)^l} x_k t^k \quad (l=0, 1, 2, \ldots).$$

LEMMA 2.7. (A) *$\mathscr{H}_A^l$ is a bounded linear operator from $X_{s,T}$ into $X_{s,T}^l$ (where $X_{s,T}^0 = X_{s,T}$) $(D_t t)^l \mathscr{H}_A^l = I$ (identity) for $l=0, 1, \ldots$.*

3. Existence and uniqueness results

Using the results of the previous section we can prove the following theorems for the system

$$\sum_{k=0}^{N} A_k(x)(tD_t)^k u - \sum_{j=0}^{N} C_{N-j} t(D_t t)^j u = f. \tag{3.1}$$

THEOREM 3.1. *Suppose that $T>0$, $s_0>0$ and*

(i) *The matrices A_k are scalars,*

$$A_k = a_k I, \quad a_k \in X_{s_0}, \quad A_N \equiv I.$$

(ii) *The roots $\lambda_i(x)$ of the polynomial*

$$Q(\lambda) = \sum_{k=0}^{N} a_k(x)\lambda^k \tag{3.2}$$

satisfy

$$\operatorname{Re}\lambda_i(x) < -\Lambda < 0 \quad (i=1, 2, \ldots, N)$$

for $x \in \bar{\Omega}_{s_0}$.

(iii) *The matrices C_{N-j} contain linear differential operator entries with coefficients belonging to*

$$C([-T, T], X_{s_0}).$$

(iv) *For the order of these operators assumption* (A2) *holds.*

(v) *The right hand side $f=f(t, x)$ is an n-dimensional vector valued function with coordinates in $C([-T, T], X_{s_0})$.*

Then there exists a unique $u=u(t, x)=(u_1, u_2, \ldots, u_n)$ such that for a suitable $0<\varepsilon<T$, $0<s_1<s_0$,

$$u_k \in C^N([-\varepsilon, \varepsilon], X_{s_1}).$$

Moreover this solution can be obtained by a suitable successive approximation process.

THEOREM 3.2. *Suppose that* $T>0$, $s_0>0$ *and*

(i) *The matrices* A_k *are scalars,*

$$A_k = a_k I \quad (k = 0, 1, \ldots, N-1), \quad a_k \in X_{s_0}, \quad A_N \equiv I.$$

(ii) *For the polynomial* $Q(\lambda)$ *in* (3.2)

$$\sup_{\substack{p=0,1,\ldots \\ x\in\Omega_{s_0}}} \frac{1}{|Q(p)|} = B_1 < +\infty$$

holds.

(iii) *The matrices* C_{N-j} *contain linear differential operator entries with coefficients belonging to* $X_{s_0,T}$.

(iv) *For the order of these operators assumption* (A2) *holds.*

(v) *The right hand side* $f=f(t, x)$ *is an n dimensional vector valued function with coordinates in* $X_{s_0,T}$.

Then there exists a unique $u=u(t, x)=(u_1, u_2, \ldots, u_n)$ *such that for a suitable* $0<\varepsilon<T$, $0<s_1<s_0$,

$$u_k \in X^N_{s_1,\varepsilon} \quad (k = 1, 2, \ldots, n).$$

Moreover this solution can be obtained by a suitable successive approximation process.

THEOREM 3.3. (i′) *Let* A_k *be arbitrary matrices,* $A_N\equiv I$ *and the entries of all* A_k *belong to* X_{s_0} ($s_0>0$, $T>0$). *Suppose*

(ii′) *For the matrix pencil*

$$Q(\lambda) = \sum_{k=0}^{N} A_k(x)\lambda^k$$

$$\sup_{\substack{p=0,1,\ldots \\ x\in\Omega_{s_0}}} \|Q(p)^{-1} = B_1 < +\infty$$

holds.

(iii′)=(iii) *in Theorem* 3.2.

(iv′) *For the order of these operators assumption* (A2) *holds with* $m_1=m_2= \ldots = \ldots =m_n=0$.

(v′)=(v) *in Theorem* 3.2.

Then we have the some conclusion as in Theorem 3.2.

PROOF OF THEOREM 3.1.

(a) By Lemma 2.4, equation (3.1) is equivalent to

$$u = H_C\left[f+\sum_{j=0}^{N-1} C_{N-j}t(D_t t)^j u\right]$$

or

$$u_i = H_C(f_i)+H_C\left[\sum_{j=0}^{N-1}\sum_{k=1}^{n} c^{ik}_{N-j}t(D_t t)^j u_k\right] \quad (i = 1, 2, \ldots, n). \tag{3.3}$$

(b) Define

$$u_{i,0} = 0, \quad u_{i,1} = H_C(f_i) \quad (i = 1, 2, \ldots, n),$$

$$u_{i,p+1} = H_C(f_i)+H_C\left(\sum_{j=0}^{N-1}\sum_{k=1}^{n} c^{ik}_{N-j}t(D_t t)u_{k,p}\right) \tag{3.4}$$

and, for $p \geqq 1$

$$v_{i,p} = (D_t t^N) \quad (u_{i,p+1} - u_{i,p}).$$

(c) We shall prove by induction the inequality

$$\|v_{i,p}(t)\|_s \leqq K \frac{C^p e^{N_p}(t)^p p^{m_i}}{(s_0 - s)^{N_p + m_i}}, \tag{3.5}$$

where

$$K = \overline{K} \max_i \|f_i\|_{C([-T,T], X_{s_0})},$$

$\overline{K}$ is a finite bound existing according to the last assertion of Lemma 2.4, $C = NM\overline{K}n$, M denotes the maximum of the M_c-s of Lemma 2.2, evaluated for all linear differential operators occuring in the C_{N-j}-s.

Taking the difference of (3.4) for p and $p+1$ and using Lemma 2.5 we get

$$v_{i,p+1}(t) = (D_t t)^N H_C \left[\sum_{j=0}^{N-1} \sum_{k=1}^{n} c_{N-j}^{ik} t \mathscr{H}_c^{N-j}(v_{k,p})(t) \right].$$

Taking X_s-norms ($s < s_0$) and using again the last assertion of Lemma 2.4 we get

$$\|v_{i,p+1}\|_s \leqq \overline{K} \sum_{j=0}^{N-1} \sum_{k=1}^{n} \sup_{|\tau| < |t|} \|c_{N-j}^{ik} \mathscr{H}_C^{N-j}(v_{k,p})(\tau)\|_s. \tag{3.6}$$

(d) Consider a single term in the right hand side of (3.6) and suppose that (3.5) holds for p. By Lemma 2.2, for $|\tau| \leqq |t|$, $0 < \eta < s_0 - s$,

$$\|c_{N-j}^{ik} \tau \mathscr{H}_C^{N-j}(v_{k,p})(\tau)\|_s \leqq \frac{M|t|}{\eta^{N-j+m_i-m_k}} \|\mathscr{H}_C^{N-j}(v_{k,p})(\tau)\|_{s+\eta} \leqq$$

$$\leqq \frac{M|t|}{\eta^{N-j+m_i-m_k}} \mathscr{H}_C^{N-j}(\|v_{k,p}(\tau)\|_{s+\eta})).$$

Here we made use of the simple inequality

$$\|\mathscr{H}_C^k(y(\tau))\|_s \leqq \mathscr{H}_C^k(\|y(\tau)\|_s)$$

valid for $y \in C([-\varepsilon_0, \varepsilon_0], X_s)$, $\varepsilon_0 > 0$.

Using (3.5) for $v_{k,p}$ and integrating we get

$$\|c_{N-j}^{ik} \tau \mathscr{H}_{ik}^{N-j}(v_{k,p})(\tau)\|_s \leqq \frac{M|t|}{\eta^{N-j+m_i-m_k}} \cdot \frac{KC^p e^{N_p}(t)^p p^{m_k}(p+1)^{j-N}}{(s_0 - s - \eta)^{N_p + m_k}}. \tag{3.7}$$

(e) Choose $\eta = s_0 - s$. Then (3.7) can be estimated by

$$K \frac{M|t|^{p+1} e^{N_p} C^p}{(s_0 - s)^{(N+1)p + m_i - j}} \cdot \frac{(p+1)^{N-j+m_i-m_k} p^{m_k}(p+1)^{j-N}}{\left(\frac{p}{p+1}\right)^{N_p - m_k}} =$$

$$= K \frac{M|t|^p e^{N_p} C^p}{(s_0 - s)^{(N+1)p + m_i - j}} \left(\frac{p+1}{p}\right)^{N_p} (p+1)^{m_i}.$$

Here $\left(1 + \frac{1}{p}\right) < e$, and by $s_0 - s < 1$, summation in (3.6) yields only the factor Nn.

(f) Fix $s<s_0$, $s_0-s<1$ and choose $\varepsilon>0$ so that

$$\frac{Ce^N\varepsilon}{s_0-s}<1. \tag{3.8}$$

By (3.5) this implies that

$$\sum_{p=1}^{\infty} v_{i,p}(t) \quad (i=1,2,\ldots,n)$$

converges in $C([-\varepsilon,\varepsilon],X_s)$, whence u_p converges in $C^N([-\varepsilon,\varepsilon],X_s)$ to a limit u which is a solution of (3.1).

The proof of uniqueness is standard: if u_i $(i=1,2)$ were solutions with the same f then the difference $u=u_1-u_2$ satisfied the homogeneous equation. For $v=(D_t t)^N u$ we get from (3.4)

$$v_i(t)=(D_t t)^N H_C\left[\sum_{j=0}^{N-1}\sum_{k=1}^{n} c_{N-j}^{ik}\, t\mathscr{H}_C^{N-j}(v_k)(t)\right].$$

Repeating the same argument as in the proof of (3.5) we get

$$\|v_i(t)\|_s \leqq \bar{K}\max_i \|v_i\|_{C([-T,T],X_{s_0})}\frac{C^p e^{N_p}(t)\,p^{m_i}}{(s_0-s)^{N_p+m_i}}.$$

Suppose that inequality (3.8) holds and let $p\to\infty$. Thus $v\equiv 0$, i.e. $(D_t t)^N u\equiv 0$. Define $w=(D_t t)^{N-1}u$. By the definition of the space C^N, w is continuous. It is easily seen that the only continuous solution of $D_t t w$ is $w=0$. Repeating this argument $(N-1)$-times we get $u\equiv 0$. Q.e.d.

PROOF OF THEOREM 3.2. Note that the time variable is allowed to be complex, $|t|\leqq T$. Steps (a) and (b) are the same as above, with H_A instead of H_c and Lemma 2.6 as reference, with scalar A_k matrices.

Steps (c) and (d) should be modified as follows. (c′):

$$\|v_{i,p}\|_{s,|t|} \leqq K\frac{C^p e^{N_p}(t)\,p^{m_i}}{(s_0-s)^{N_p+m_i}}, \tag{3.9}$$

where

$$K=\bar{B}\max_{1\leqq i\leqq n}\|f\|_{s_0,T},$$

$C=NnM\bar{B}$. Here $\bar{B}$ denotes the finite bound existing by the last assertion of Lemma 2.6 while M denotes the maximum of all M_A-s of Lemma 2.3 evaluated for all linear operators occurring in the C_{N-j}-s.

From the modified recursion (3.4) we get

$$v_{i,p+1}(t)=(D_t t)^N H_A\left[\sum_{j=0}^{N-1}\sum_{k=1}^{n} c_{N-j}^{ik}\, t\mathscr{H}_A^{N-j}(v_{k,p})(t)\right].$$

Taking $X_{s,|t|}$ norms $(|\tau|\leqq|t|)$.

$$\|v_{i,p+1}\|_{s,|t|} \leqq \bar{B}\sum_{j=0}^{N-1}\sum_{k=1}^{n}\|c_{N-j}^{ik}\,\tau\mathscr{H}_A^{N-j}(v_{k,p})(\tau)\|_{s,|t|}. \tag{3.10}$$

Consider a single term in the right hand side of (3.10) and suppose that (3.9) holds for p. By Lemma 2.3, for $|t|\leqq T$, $0<\eta<s_0-s$,

$$\|c_{N-j}^{ik}\tau\mathscr{H}_A^{N-j}(v_{k,p})(\tau)\|_{s,|t|}\leqq\frac{M|t|}{\eta^{N-j+m_i+m_k}}\|\mathscr{H}_C^{N-j}(v_{k,p})(\tau)\|_{s+\eta,|t|}=$$

$$=\frac{M|t|}{\eta^{N-j+m_i+m_k}}\mathscr{H}_C^{N-j}(\|v_{k,p}(\tau)\|_{s+\eta,|t|}).$$

Here we made use of the equality

$$\|\mathscr{H}_A^l(y(\tau))\|_{s,|t|}=\mathscr{H}_C^l(\|y\|_{s,|t|}),\tag{3.11}$$

where the operator $\mathscr{H}_c^l$ should be applied to the continuous function $\psi(t')=\|y\|_{s,t'}$, $t'>0$, and $y\in X_{s,\varepsilon_0}$, $\varepsilon_0>0$, $l=1, 2, \ldots$. Using (3.9) and integrating we get

$$\|c_{N-j}^{ik}\tau\mathscr{H}_A^{N-j}(v_{k,p})(\tau)\|_{s,|t|}\leqq\frac{M|t|}{\eta^{N-j+m_j-m_k}}\frac{KC^pe^{N_p}|t|^p p^{m_k}(p+1)^{j-N}}{(s_0-s-\eta)^{N_p+m_k}}.$$

The remainder of the proof of Theorem 3.1 requires no essential changes. If ε satisfies (3.8) then

$$\sum_{k=1}^{\infty}v_{i,p}(t)$$

converges in $X_{s,\varepsilon}$, whence u_p converges in $X_{s,\varepsilon}^N$ to a limit u which is a solution of (3.1). Q.e.d.

PROOF OF THEOREM 3.3. In the sequel u, v denote vector valued functions, the norm of a vector is the sum of norms of its coordinates, H_A denotes a matrix-type operator (see Lemma 2.6). (3.1) is equivalent to

$$u=H_A\left[f+\sum_{j=0}^{N+1}C_{N-j}+(D_t t)^j u\right].$$

Defining u_p and v_p similarly as in the proof of Theorem 3.1, we claim

$$\|v_p\|_{s,T}\leqq K\frac{C^pe^{N_p}|t|^p}{(s_0-s)^{N_p}},$$

where $K=\overline{K}\|f\|_{s,T}$, $\overline{K}$ denotes the bound existing by the last assertion of Lemma 2.6, $C=NKM$ where M is a positive constant for which

$$\|C_{N-j}y\|_{s_1,T}\leqq\frac{M}{(s_2-s_1)^{N-j}}\|y\|_{s_2,T}\quad(j=0, 1, \ldots, N-1).\tag{3.12}$$

Such an M can be obtained by summing the constants M_A in (2.6) for fixed j (for each operator-entry of C_{N-j}) and taking maximum in j. $s_0-s<1$ should be assumed throughout. We have

$$v_{p+1}(t)=(D_t t)^N H_A\left[\sum_{j=0}^{N-1}C_{N-j}t\mathscr{H}_A^{N-j}(v_p)(t)\right]$$

whence

$$\|v_{p+1}\|_{s,|t|} \leqq \overline{K} \sum_{j=0}^{N-1} \|C_{N-j}\tau \mathscr{H}_A^{N-j}(v_p)(\tau)\|_{s,|t|}.$$

Using (3.11) and (3.12) we get

$$\|v_{p+1}\|_{s,|t|} \leqq M\overline{K} \sum_{j=0}^{N-1} \frac{|t|^{p+1}}{\eta^{N-j}} \frac{1}{(p+1)^{N-j}} \frac{KC^p e^{N_p}}{(s_0-s-\eta)^{N_p}}.$$

The remaining steps of the proof are the same as above. Q.e.d.

Finally, as a corollary of Theorems 3.1 and 3.2 we prove the following

THEOREM 3.4. *Consider equation* (3.1). *Suppose that* $T>0$, $s_0>0$ *and*

(i) *The matrices* A_k *are scalars,* $A_k=a_kI$, $(k=0, 1, \ldots, N-1)$, $A_N\equiv I$.

(ii) *The roots* $\lambda_i(x)$ *of the polynomial* $Q(x)$ ((3.2)) *satisfy*

$$\operatorname{Re}\lambda_i(x) < \Lambda \quad (i = 1, 2, \ldots, N)$$

for $x\in\overline{\Omega}_{s_0}$. *(No restriction on the sign of the constant* Λ.*)*

(iii) *The matrices* C_{N-j} *contain linear differential operator entries with coefficients belonging to* $X_{s_0,T}$.

(iv) *For the order of these operators assumption* (A2) *holds.*

(v) *The right hand side* $f=f(t, x)$ *is an n-dimensional vector valued function which is* $\varkappa$ *times differentiable, or, more precisely*

$$f = \sum_{l=0}^{\varkappa-1} f_l(x)t^l + t^{\varkappa}\tilde{f}(f, x) \tag{3.13}$$

where $f_l\in X_{s_0}$, $\tilde{f}(t, x)\in C([-T, T], X_{s_0})$.

(vi)=(ii) *of Theorem* 3.2.

(vii) $\Lambda<\varkappa$.

Then (3.1) *has a unique solution* $u\in C^N([-\varepsilon, \varepsilon], X_s)$ *for a suitable* $s<s_0$ *and* $0<\varepsilon<T$.

PROOF. For short denote the t-polynomial in (3.13) by f_1. Thus equation (3.1) may be considered in the form

$$\mathscr{P}_u = f_1 + t^{\varkappa}\tilde{f}.$$

By Theorem 3.2 there exists a $u_1\in X_{s,\varepsilon}$ such that $\mathscr{P}_{u_1}=f_1$. Note that after a change of variable $u=t^{\varkappa}v$ the weight of each monomial in the principal part of $\mathscr{P}$ attains $\varkappa$ and after cancelling the factor $t^{\varkappa}$ we get an equation

$$\tilde{\mathscr{P}}\cdot v = \tilde{f} \tag{3.14}$$

where $\tilde{\mathscr{P}}$ satisfies (A1) and (A2) and the corresponding characteristic roots are shifted into the negative half-plane. Hence Theorem 3.1 applies and (3.14) admits a solution $v\in C([-\varepsilon_2, \varepsilon_2], X_{s_2})$. Putting $\varepsilon=\min(\varepsilon_1, \varepsilon_2)$, $s=\min(s_1, s_2)$ and using the inclusion $X_{s,\varepsilon}\subset C([-\varepsilon, \varepsilon], X_s)$ we see that $u_1+t^{\varkappa}v$ is a solution of (3.1). Uniqueness follows directly from the uniqueness statement of Theorem 3.2. Q.e.d.

REMARK 3.1. If (vi) does not hold, i.e. there are nonnegative integer characteristic values, then Theorem 3.4 fails to be true. For the solvability of (3.1) the Taylor polynomial of f should satisfy certain semi-algebraic, (i.e. algebraic with respect to the variable t) conditions — these are the compatibility conditions in [1]. In this case a Fredholm-type alternative holds.

REMARK 3.2. If all conditions but (v) hold, i.e. the only failure is the non-smoothness of f, then a solution to (3.1) still exists in the distribution sense.

REMARK 3.3. If the order of differentiation with respect to the variables x exceeds the bounds of (A2), then similar results can be proved for functions belonging to suitable Gevrey-spaces (see [6]).

REFERENCES

[1] BAOUENDI, M. S., GOULAOUIC, C.: Cauchy problems with characteristic initial hypersurface, *Comm. Pure Appl. Math.* **26** (1973), 455—475.
[2] BAOUENDI, M. S., GOULAOUIC, C.: Singular nonlinear Cauchy problems, *J. Diff. Equ.* **22** (1976), 268—291.
[3] VOLEVIC', L. R.: On a problem of linear programming related to differential equations, *Uspekhi Mat. Nauk*. **18**, 3 (1963), 155—162.
[4] HARRIS, W. A. JR., SIBUYA, Y., WEINBERG, L.: Holomorphic solutions of linear differential systems at singular points, *Arch. Rat. Mech. and Anal.* **35** (1969), 245—248.
[5] HARTMAN, P.: *Ordinary differential equations*, Wiley ,New York, 1964.
[6] STEINBERG, S.: Existence and uniqueness of hyperbolic equations which are not necessarily strictly hyperbolic, *J. Diff. Equ.* **17** (1975), 119—153.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest V, Reáltanoda u. 13—15, 1053, Hungary

(Received October 26, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 229—231

APPRAISING THE CENTRALITY OF VERTICES IN TREES

by

P. J. SLATER

Abstract. Pollák and Ádám have asked if there exists a tree in which the center, centroid and end vertex centroid are pairwise disjoint. An affirmative answer is given, and it is shown that these sets can be made non-collinear and arbitrarily far apart in trees.

Let T be a tree with vertex set $V(T)$. The terminology and notation of [2] will be used whenever possible. In particular, the distance in T between vertices u and v, denoted $d(u, v)$, is the number of edges in the unique u to v path. The *eccentricity* of $u \in V(T)$, denoted $e(u)$, is defined as

$$e(u) = \max \{d(u, v): v \in V(T)\},$$

and the *distance* of $u \in V(T)$, denoted $d(u)$, is defined as

$$d(u) = \sum_{v \in V(T)} d(u, v).$$

Further, $bw(u)$ denotes the *branch weight* of u, and $bw(u)$ is defined to be the largest number of vertices in a component of $T-u$.

The sets of vertices at which these functions are minimized are respectively known as the *center, centroid* and *branch weight centroid* of T. An early study of these measures of centrality showed the following.

THEOREM 1 (JORDAN [3]). *Every tree has a center consisting of either one vertex or two adjacent vertices; every tree has a centroid consisting of either one vertex or two adjacent vertices.*

More recently the last two sets were shown to be identical.

THEOREM 2 (ZELINKA [5]). *For any tree T the branch-weight centroid of T is the centroid of T.*

In many places, such as DEO [2], it has been noted that when using trees one is "generally interested in computing the sum of the distances of all end vertices from a given vertex u". That is, if $P \subseteq V(T)$ is the set of vertices of degree one in T, then one is interested in $\sum_{v \in P} d(u, v)$. More general measures of centrality in which

This work was supported by the U.S. Department of Energy (DOE), Contract No. AT(29—1)—789. By acceptance of this article, the publisher and/or recipient acknowledges the U.S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

S is any subset of $V(T)$ are examined in [4]. Two of the functions examined are the following. The *S-distance* of u is defined as

$$d_S(u) = \sum_{v \in S} d(u, v),$$

and the subset of vertices with minimum S-distance is called the *S-centroid.* The *S-branch weight* of u, denoted $bw_S(u)$, is defined as max $|S \cap B|$ where B is the vertex set of a component of $T-u$, and the *S-branch weight centroid* is the set of vertices in T of minimum S-branch weight.

THEOREM 3 ([4]). *For any tree T and any $S \subseteq V(T)$ the S-branch weight centroid of T is the S-centroid.*

THEOREM 4 ([4]). *For any subset S of the vertices of a tree T the S-centroid of T consists of the vertices of a path in T.*

In ÁDÁM [1] five functions measuring the centrality of a vertex are considered. The sets of vertices determined to be most central by these functions are the center in two cases, the centroid in one case, and the P-centroid in two cases (where P is the collection of end vertices). [1] concludes with a question suggested by G. POLLÁK.

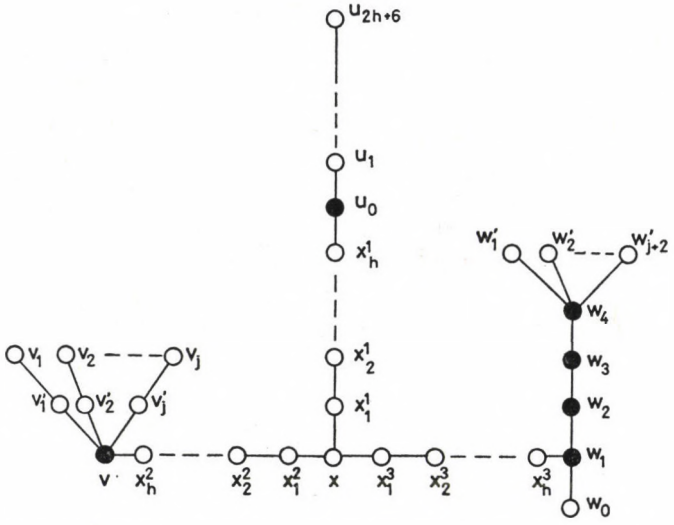


Figure 1. A tree with disjoint center, centroid and end vertex centroid.

PROBLEM. Does there exist a tree in which the center, centroid and P-centroid are pairwise disjoint?

THEOREM 5. *For any $k \geqq 1$ there exists a tree T_k for which the center, centroid and P-centroid are pairwise separated by distances greater than k, and no path in T contains a vertex of each of these sets.*

PROOF. Consider tree T_k in Figure 1. Selecting $h \geqq k/2$ and $j=5h+15$, it suffices to show that the center is $\{u_0\}$, the centroid is $\{v\}$, and the P-centroid is $\{w_1, w_2, w_3, w_4\}$.

Since $e(u_0)=2h+6$, $e(u_i)=2h+6+i$, and $y \neq u_i$ $(0 \leqq i \leqq 2h+6)$ implies that $d(y, u_{2h+6})=e(y) \geqq 2h+7$, one has that $\{u_0\}$ is the center.

Making use of Theorem 2 to determine the centroid, one has $bw(v)=2j$, $bw(v_i)=4j$ and $bw(v_i')=4j-1$ for $1 \leqq i \leqq j$, and for any other vertex y one has $bw(y) \geqq 2j+1$. Hence $\{v\}$ is the centroid.

Similarly one can make use of Theorem 3 to determine the P-centroid. One has $bw_P(w_i)=j+2$ for $1 \leqq i \leqq 4$; for $0 \leqq i \leqq j+2$ one has $bw_P(w_i')=2j+3$; and for any other $y \in V(T)$ one has $bw_P(y) \geqq j+3$. Hence $\{w_1, w_2, w_3, w_4\}$ is the P-centroid.

REFERENCES

[1] ÁDÁM, A.: The centrality of vertices in trees, *Studia Sci. Math. Hung.* **9** (1974), 285—303.
[2] DEO, N.: *Graph Theory with Applications to Engineering and Computer Science*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1974.
[3] JORDAN, C.: Sur les assemblages des lignes, *J. Reine Angew. Math.* **70** (1869), 185—190.
[4] SLATER, P.: Centers to centroids in graphs, *Journal of Graph Theory* **2** (1978), 209—222.
[5] ZELINKA, B.: Medians and peripherians of trees, *Archivum Mathematicum (Brno)* (1968), 87—95.

Sandia Laboratories, Albuquerque, New Mexico 87185

(Received January 5, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 233—244

A STRONG LAW OF LARGE NUMBERS AND SOME APPLICATIONS

by

L. GYŐRFI, Z. GYŐRFI and I. VAJDA

In this paper a simple generalization of the strong law of large numbers is presented, and its interesting applications have been found in various fields, namely:

— density estimation based on non-stationary and dependent samples,

— for Neyman—Pearson decision the estimation of error probability of second-kind,

— estimation of asymptotic error probability of convergent sequences of decision functions,

— estimation of optimal risk for risk minimizing procedures.

1. The Law of Large Numbers

Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space and H a separable Hilbert space with the norm $\|\cdot\|$ and the inner product $(\cdot, \cdot)$. Let $\mathscr{H}$ denote the σ-algebra of Borel subsets of H. Let ξ be a random element taking values in $(H, \mathscr{H})$, for which $\mathbf{E}\|\xi\| < +\infty$ and let $\mathscr{F} \subseteq \mathscr{A}$ be a σ-algebra. Then the definition of the conditional expectation of ξ with respect to $\mathscr{F}$ is the following: For all $f \in H$ $(f, \mathbf{E}[\xi/\mathscr{F}]) = \mathbf{E}[(f, \xi)/\mathscr{F}]$, and the expectation of ξ is $\mathbf{E}[\xi/\{\emptyset, \Omega\}]$ ([1], [2], [3]).

Let $\{\xi_n\}$, $n=1, 2, \ldots$ be a sequence of random elements taking values in $(H, \mathscr{H},)$ such that $\mathbf{E}\|\xi_n\| < +\infty$, $n=1, 2, \ldots$. Moreover, let $\{\mathscr{F}_n\}$, $n=1, 2, \ldots$ be monotonically increasing sequence of σ-algebras such that ξ_n is measurable with respect to $\mathscr{F}_n$, $n=1, 2, \ldots$.

THEOREM. *Assume that there exists a random element ξ for which*

$$\lim_{n\to\infty} \|\mathbf{E}[\xi_{n+1}/\mathscr{F}_n] - \xi\| = 0 \quad a.s. \tag{1'}$$

or

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}[\xi_{i+1}/\mathscr{F}_i] - \xi \right\| = 0 \quad a.s. \tag{1''}$$

Finally, let $\{X_n\}$, $n=1, 2, \ldots$ be a sequence of random elements defined by

$$X_{n+1} = X_n(1-\gamma_n) + \gamma_n \xi_{n+1}, \quad n = 1, 2, \ldots \tag{2}$$

where X_1 stands for an arbitrary random element for which $\mathbf{E}\|X_1\| < +\infty$, and $\{\gamma_n\}$, $n=1, 2, \ldots$ is a sequence of positive numbers for which

$$\prod_{n=1}^{\infty} (1-\gamma_n) = 0 \tag{3}$$

and

$$\sum_{n=1}^{\infty} \gamma_n^2 \mathbf{E}\|\mathbf{E}[\xi_{n+1}/\mathscr{F}_n]-\xi_{n+1}\|^2 < +\infty. \tag{4}$$

Then (1′), (2), (3) *and* (4) *imply that*

$$\lim_{n\to\infty} \|X_n-\xi\| = 0 \quad a.s. \tag{5}$$

and, for $\gamma_n=\frac{1}{n+1}$, (1″), (2), (3) *and* (4) *imply* (5).

2. Applications

2.1. The Kolmogorov Theorem and the Law of Large Numbers of Komlós and Révész

In this section let H be the $\mathbf{R}^N$ space. Let $\{\xi_n\}$, $n=1, 2, \ldots$ be a sequence of independent random vectors with mean $\{m_n\}$, $n=1, 2, \ldots$ and variance $d_n^2=\mathbf{E}\|\xi_n-m_n\|^2$, $n=1, 2, \ldots$, so that

$$\lim_{n\to\infty} \left\| \frac{1}{n}\sum_{i=1}^{n} m_i - m \right\| = 0 \tag{6}$$

and

$$\sum_{n=1}^{\infty} \left(\frac{d_n}{n+1}\right)^2 < +\infty$$

then, by the Theorem, it is easy to verify that if

$$X_n = \frac{1}{n}\sum_{n=1}^{n} \xi_i \tag{7}$$

then

$$\lim_{n\to\infty} \|X_n-m\| = 0 \quad \text{a.s.} \tag{8}$$

(Since if $\xi_1=X_1$ and $\gamma_n=\frac{1}{n+1}$, then $X_{n+1}=X_n(1-\gamma_n)+\gamma_n\xi_{n+1}$, $n=1, 2, \ldots$.)

This proposition does not guarantee the convergence (8) in case of $d_n=\sqrt{1+n}$.

Now let $m_n=m$, $n=1, 2, \ldots$. RÉVÉSZ and KOMLÓS [4] showed that if we consider, instead of (7), the following weighted average

$$X_n = \frac{\sum_{i=1}^{n} \xi_i/d_i^2}{\sum_{i=1}^{n} 1/d_i^2} \tag{9}$$

then the sufficient condition for (8) is

$$\sum_{i=1}^{\infty} 1/d_i^2 = +\infty. \tag{10}$$

Now we show that (8) can be proved by the Theorem also in case of $\lim_{n\to\infty} \|m_n - m\| = 0$. In order to verify this we have only to consider the choice

$$\gamma_n = \frac{1/d_{n+1}^2}{\sum_{i=1}^{n+1} 1/d_i^2}, \quad n = 1, 2, \ldots.$$

(9) can be written in the form

$$X_{n+1} = X_n(1-\gamma_n)+\gamma_n\xi_{n+1}, \quad n = 1, 2, \ldots$$

and (10) implies that

$$\prod_{n=1}^{\infty}(1-\gamma_n) = 0$$

and by the Abel—Dini theorem

$$\sum_{n=1}^{\infty}\gamma_n^2 d_n^2 < +\infty.$$

2.2. Density Estimation by Non Stationary and Dependent Samples

Let $\{\eta_n\}$, $n=1, 2, \ldots$ be a sequence of random vectors taking values in $(\mathbf{R}^N, \mathscr{B}^N)$. Suppose that the conditional distributions

$$\mathbf{Q}_n(A) = \mathbf{P}(\eta_n \in A/\eta_1, \ldots, \eta_{n-1}), \quad n = 1, 2, \ldots, A \in \mathscr{B}^N$$

are regular and dominated by the Lebesgue measure λ, $n=1, 2, \ldots$. f_n denotes the corresponding density

$$f_n = \frac{d\mathbf{Q}_n}{d\lambda}, \quad n = 1, 2, \ldots.$$

Now let $H=L_2(\mathbf{R}^N, \mathscr{B}^N, \lambda)$ and suppose that $f_n \in L_2$, $n=1, 2, \ldots$ a.s. Choose the function $h \in L_2$ such that

$$\int h(x)\lambda(dx) = 1$$

and, if $\{c_n\}$, $n=1, 2, \ldots$ is a sequence tending to 0 then let

$$K_n(x, y) \triangleq \frac{1}{c_n^N} h\left(\frac{x-y}{c_n}\right), \quad n = 1, 2, \ldots, x, y \in \mathbf{R}^N.$$

Finally, define $\{X_n\}$, $n=1, 2, \ldots$ as a sequence of random elements taking values in L_2:

(11) $$X_{n+1} = X_n(1-\gamma_n)+\gamma_n K_n(\cdot, \eta_{n+1}), \quad n = 1, 2, \ldots$$

where $X_1 = x \in L_2$ is arbitrary and $\{\gamma_n\}$, $n=1, 2, \ldots$ is a sequence of positive numbers and, for $\{\gamma_n\}$, $n=1, 2, \ldots$, satisfying (3).

The properties of the algorithm (11) has been often scrutinized when $\{\eta_n\}$, $n=1, 2, \ldots$ is a sequence of independent and identically distributed random vectors, and under some further conditions

$$\lim_{n\to\infty} \|X_n - f\| = 0, \quad f_n = f, \quad n = 1, 2, \ldots \quad \text{a.s.} \tag{12}$$

is proved ([5], [6], [7]).

Now, we consider the non-stationary, dependent case:

COROLLARY 1. *If for a* $f \in L_2$

$$\lim_{n\to\infty} \|f_n - f\| = 0 \quad a.s. \tag{13}$$

then (12) *holds for the sequence* $\{X_n\}$, $n=1, 2, \ldots$ *defined by* (11) *supposing that*

$$\sum_{n=1}^{\infty} \frac{\gamma_n^2}{c_n^N} < +\infty, \quad \lim_{n\to\infty} c_n = 0. \tag{14}$$

COROLLARY 2. *Assume a* $f \in L_2$, *for which*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^{n} f_i - f \right\| = 0 \quad a.s. \tag{15}$$

Choose $\gamma_n = \dfrac{1}{n+1}$, $n=1, 2, \ldots$ *and the sequence* $\{c_n\}$, $n=1, 2, \ldots$ *satisfies* (14), *furthermore*

$$\lim_{n\to\infty} \frac{c_n}{c_{n+1}} = 1 \tag{16}$$

and

$$\sup_n \frac{1}{n} \sum_{i=1}^{n} i \left| \frac{c_i^N}{c_{i+1}^N} - 1 \right| < +\infty \tag{17}$$

and there exists an $\varepsilon > 0$ *and* $b > 0$ *such that for all* $|a-1| < \varepsilon$

$$\int |a^N h(ax) - h(x)| \lambda(dx) \leqq b|a^N - 1| \tag{18}$$

then, for the sequence $\{X_n\}$, $n=1, 2, \ldots$, (12) *holds.*

Considering the following three lemmas and the Theorem, the statements of the Corollary 1 and 2 are obvious:

LEMMA 1. (13) *and* $\lim_{n\to\infty} c_n = 0$ *implies that*

$$\lim_{n\to\infty} \|\mathbf{E}[K_n(\cdot, \eta_{n+1})/\eta_1, \ldots, \eta_n] - f\| = 0 \quad \text{a.s.}$$

LEMMA 2. (14) *implies that*

$$\sum_{n=1}^{\infty} \gamma_n^2 \mathbf{E} \|K_n(\cdot, \eta_{n+1}) - \mathbf{E}[K_n(\cdot, \eta_{n+1})/\eta_1, \ldots, \eta_n]\|^2 < +\infty.$$

LEMMA 3. *Under the conditions of the Corollary* 2

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[K_i(\cdot,\eta_{i+1})/\eta_1,\ldots,\eta_i]-f\right\|=0 \quad \text{a.s.}$$

(The proof of these lemmas can be found in the Appendix.)

REMARK 1. For the sequence $c_n=(1/n)^\alpha$, $n=1, 2, \ldots$ the assumptions of Corollary 2 hold if $0<\alpha<1/N$.

REMARK 2. For example, (18) is met for

$$h(x)=\begin{cases}1 & \text{if} \quad x\in[-1/2, 1/2],\\ 0 & \text{otherwise}\end{cases}$$

and for

$$h(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad (N=1).$$

REMARK 3. For the sample $\eta_1, \eta_2, \ldots, \eta_n$ satisfiyng the conditions of Corollary 2 let us consider the case when $\eta_1, \eta_2, \ldots, \eta_n$ is strictly stationary, ergodic, Markov sequence of random variable such that the transition density — denoted by g — is square integrable and its L_2 norm has finite expectation. In this case $f_n(x, \eta_1, \eta_2, \ldots, \eta_{n-1})=g(x, \eta_{n-1})$. The average $\frac{1}{n}\sum_{k=1}^{n} g(\cdot, \eta_{k-1})$ tends to the common density of η_n in L_2 norm since the following version of ergodic theorem can be verified by a simple generalization of [12]. Theorem 3: let $\varepsilon_1, \varepsilon_2, \ldots$ be a strictly stationary and ergodic sequence of random element taking values in a separable Hilbert space and having expectation, then

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i-\mathbf{E}\varepsilon_1\right\|=0.$$

2.3. Estimation Error Probability of Second-Kind for Neyman—Pearson Decision

Let η and ζ be random elements taking values in the measurable space $(Y, \mathcal{Y})$. Let denote $\mathbf{Q}_\eta$ and $\mathbf{Q}_\zeta$ their distribution respectively.

The Neyman—Pearson problem is the following: Considering a given number $0\leqq\alpha\leqq 1$, determine the set $A\in\mathcal{Y}$ for which

$$\mathbf{Q}_\zeta(A)=\inf_{B\in\mathcal{Y}^\alpha}\mathbf{Q}_\zeta(B)$$

where

$$\mathcal{Y}^\alpha=\{B|B\in\mathcal{Y},\ 1-\mathbf{Q}_\eta(B)\leqq\alpha\}$$

According to the Neyman—Pearson lemma the solution is as follows: let μ be a measure on $(Y, \mathcal{Y})$ such that $\mathbf{Q}_\eta, \mathbf{Q}_\zeta\ll\mu$ and $f_\eta=\frac{d\mathbf{Q}_\eta}{d\mu}$ and $f_\zeta=\frac{d\mathbf{Q}_\zeta}{d\mu}$ denote the Radon—Nikodym derivatives and

$$A=A(c)\overset{\triangle}{=}\left\{y/y\in Y, \frac{f_\eta(y)}{f_\zeta(y)}>c\right\}$$

where c is a real for which $A(c)\in\mathscr{Y}^{\alpha}$ and for all $B\in\mathscr{Y}^{\alpha}$ $\mathbf{Q}_{\eta}(B)\cong\mathbf{Q}_{\eta}(A(c))$. Consider the case when there exists a c^{α} which is the solution of $1-\mathbf{Q}_{\eta}(A(c))=\alpha$. Assume a sequence of random elements $\{\eta_n\}$, $n=1, 2, \ldots$, having the same distribution as η, and a sequence $\{C_n\}$, $n=1, 2, \ldots$ of real random variables so that C_n is measurable with respect to $\{\eta_1, \ldots, \eta_n\}$ and

$$\lim_{n\to\infty} C_n = c^{\alpha} \quad \text{a.s.} \tag{19}$$

(For instance the sequence $\{C_n\}$, $n=1, 2, \ldots$ can be generated by the following stochastic approximation algorithm:

$$C_{n+1} = C_n - \lambda_n[\chi_{\{\eta_{n+1}\notin A(C_n)\}} - \alpha], \tag{20}$$

$n=1, 2, \ldots$ $C_1=c$; $\chi_{\{\cdot\}}$ stands for the indicator of the event $\{\cdot\}$, and the sequence $\{\lambda_n\}$, $n=1, 2, \ldots$ fulfils $\sum_{n=1}^{\infty}\lambda_n=+\infty$, $\sum_{n=1}^{\infty}\lambda_n^2<+\infty$.) moreover let a sequence $\{\zeta_n\}$, $n=1, 2, \ldots$ of random elements be given having the same distribution as ζ. Let us suppose that the pairs $\{(\eta_n, \zeta_n)\}$, $n=1, 2, \ldots$ are independent.

The question is how to estimate $\mathbf{Q}_{\zeta}(A(c^{\alpha}))$, the error of the second-kind when the sequence $\{C_n\}$, $n=1, 2, \ldots$ and the sequence $\{\zeta_n\}$, $n=1, 2, \ldots$ are only given.

COROLLARY 3. *If* $\mathbf{Q}_{\zeta}(A(c))$ *is continuous in* c^{α} *then, for the sequence* $\{X_n\}$, $n=1, 2, \ldots$, *defined by* $X_1=x_1$,

$$X_{n+1} = X_n(1-\gamma_n)+\gamma_n\chi_{\{\zeta_{n+1}\in A(C_n)\}}, \quad n=1, 2, \ldots$$

we have

$$\lim_{n\to\infty} X_n = \mathbf{Q}_{\zeta}(A(c^{\alpha})) \quad a.s.$$

provided $\{\gamma_n\}$, $n=1, 2, \ldots$,

$$\sum_{n=1}^{n}\gamma_n = +\infty, \quad \sum_{n=1}^{n}\gamma_n^2 < +\infty. \tag{21}$$

This statement is a trivial consequence, observing that

$$\mathbf{E}[\chi_{\{\zeta_{n+1}\in A(C_n)\}}|\zeta_1, \ldots, \zeta_n, \eta_1, \ldots, \eta_n] = \mathbf{Q}_{\zeta}(A(C_n)).$$

2.4. Estimation of the Asymptotic Error Probability of Convergent Sequence of Decision Functions

Consider the following Bayesian decision problem: let η be the observation, a random element taking values in the measurable space $(Y, \mathscr{Y})$ and ϱ, a random variable taking values in the set $\{1, 2, \ldots, M\}$. An arbitrary D: $Y\to\{1, 2, \ldots, M\}$ measurable function is called decision function and

$$\mathbf{P}(D(\eta)\neq\varrho)$$

stands for its error probability. The minimization of this error probability in D is the Bayes decision which uses the conditional distributions of ϱ given η. However for practical problems (i.e. pattern classification) these distributions are unknown and only a labelled sample $\{(\eta_n, \varrho_n)\}$, $n=1, 2, \ldots$ is given, where $\{(\eta_n, \varrho_n)\}$, $n=1, 2, \ldots$

are independent and have the same distribution as (η, ϱ). Solving this problem, a sequence of decision functions $\{D_n\}$, $n=1, 2, \ldots$ is generated so that, if $y \in Y$, then the random variables $D_n(y)$ is measurable with respect to $\{(\eta_1, \varrho_1), \ldots, (\eta_n, \varrho_n)\}$.

Suppose now that, for a $D: Y \to \{1, 2, \ldots, M\}$ decision function, the sequence $\{D_n\}$, $n=1, 2, \ldots$ converges to D in the sense that, if $P_n = \mathbf{P}(D_n(\eta) \neq \varrho | (\eta_1, \varrho_1), \ldots, (\eta_n, \varrho_n))$ and $P_D = \mathbf{P}(D(\eta) \neq \varrho)$, then

$$\lim_{n \to \infty} P_n = P_D \quad \text{a.s.} \tag{22}$$

COROLLARY 4. *If the sequence of real random variables* $\{X_n\}$, $n=1, 2, \ldots$ *is given by*

$$X_{n+1} = X_n(1-\gamma_n) + \gamma_n \chi_{\{D_n(\eta_{n+1}) \neq \varrho_{n+1}\}}, \quad n = 1, 2, \ldots$$

then (22) *implies that*

$$\lim_{n \to \infty} X_n = P_D \quad \text{a.s.}$$

supposing that (21) *holds for the sequence* $\{\gamma_n\}$, $n=1, 2, \ldots$.

(Obviously, because

$$\mathbf{E}[\chi_{\{D_n(\eta_{n+1}) \neq \varrho_{n+1}\}} | (\eta_1, \varrho_1), \ldots, (\eta_n, \varrho_n)] = P_n).$$

Such kind of statement is given by WOLVERTON with respect to the nearest neighbor decision rule [8], and by L. GYŐRFI [9] generally with assumption differing from (22).

2.5. *Estimation of Optimal Risk for Risk Minimizing procedures*

Let $(Y, \mathcal{Y})$ be a measurable space and L be a real function defined on the product space $(Y \times \mathbf{R}^N, \mathcal{Y} \times \mathcal{B}^N)$. Supposing that η is a random element taking values in $(Y, \mathcal{Y})$, we define the real function K as

$$K = \mathbf{E}L(\eta, \cdot)$$

The cost-minimization task is to find the extremum of K. Suppose that the gradient of the function $L(y, \cdot)$ exists for all $y \in Y$, and it is denoted by f:

$$f(y, \cdot) = \operatorname{grad} L(y, \cdot), \quad y \in Y.$$

Finally, let

$$R = \mathbf{E}f(\eta, \cdot).$$

Suppose that the gradient of K exists and it equals to R. If we are given a sequence of independent and identically distributed random elements $\{\eta_n\}$, $n=1, 2, \ldots$, having the same distribution as η, then the usual stochastic approximation can be used:

$$\theta_{n+1} = \theta_n - \lambda_n f(\eta_{n+1}, \theta_n).$$

Assume that the conditions of the convergence of $\{\theta_n\}$, $n=1, 2, \ldots$ are met ([11]) then

$$\lim_{n \to \infty} \|\theta_n - \theta\| = 0 \quad \text{a.s.} \tag{23}$$

It would often be necessary to know the optimum risk value $K(\theta)$. Now we show how to estimate $K(\theta)$ if the observation of the $\{\eta_n\}$, $n=1, 2, \ldots$ and $\{\theta_n\}$, $n=1, 2, \ldots$ are only possible. Ultimately, suppose that (23) holds for the sequence $\{\theta_n\}$, $n=1, 2, \ldots$ and θ_n is measurable with respect to $\{\eta_1, \ldots, \eta_n\}$, $n=1, 2, \ldots$.

COROLLARY 5. *If the function K is continuous at the point θ then, for the sequence $\{X_n\}$, $n=1, 2, \ldots$ defined by*

$$X_{n+1} = X_n(1-\gamma_n)+\gamma_n L(\eta_{n+1}, \theta_n), \quad n = 1, 2, \ldots,$$

$X_1=x\in\mathbf{R}^1$, it holds that

$$\lim_{n\to\infty} X_n = K(\theta) \quad \text{a.s.}$$

supposing

$$\sum_{n=1}^{\infty} \gamma_n = +\infty$$

and

$$\sum_{n=1}^{\infty} \gamma_n^2 \mathbf{E}[L(\eta_{n+1}, \theta_n)-K(\theta_n)]^2 < +\infty.$$

(Observe that $\mathbf{E}[L(\eta_{n+1}, \theta_n)|\eta_1, \ldots, \eta_n]=K(\theta_n)$.)

3. Appendix

First of all we need the following generalization of the Toeplitz theorem proved by FRITZ [10].

LEMMA 4. *Consider an array $\{\mathbf{C}_{n,i}\}$, $i, n=1, 2, \ldots$ of bounded linear operators on H. Let $\mathbf{C}_{n,i}=0$ if $i>n$. Assume that the subsequent conditions are valid:*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{C}_{n,i} = \mathbf{C} \tag{24}$$

pointwise in H, where $\mathbf{C}$ is a bounded linear operator on H,

$$\lim_{n\to\infty} \|\mathbf{C}_{n,i}\| = 0, \quad i = 1, 2, \ldots \tag{25}$$

moreover

$$\sup_n \sum_{i=1}^{n} \|\mathbf{C}_{n,i}\| < +\infty \tag{26}$$

then, if for a sequence $\{f_n\}$, $n=1, 2, \ldots$ and an element f in H

$$\lim_{n\to\infty} \|f_n-f\| = 0,$$

then

$$\lim_{n\to\infty} \left\| \sum_{i=1}^{n} \mathbf{C}_{n,i} f_i - \mathbf{C}f \right\| = 0. \tag{27}$$

If furthermore

$$\sup_n \sum_{i=1}^{n} i \, \|\mathbf{C}_{n,i}-\mathbf{C}_{n,i+1}\| < +\infty \tag{28}$$

then

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^{n} f_i - f \right\| = 0$$

implies (27).

PROOF OF THEOREM. Introducing the notation

$$Y_{n+1} = Y_n(1-\gamma_n)+\gamma_n \mathbf{E}[\xi_{n+1}/\mathscr{F}_n], \quad n = 1, 2, \ldots, \tag{29}$$

$Y_1=\mathbf{E}X_1$, it is easy to check by induction that

$$Y_{n+1} = \sum_{i=1}^{n} \mathbf{C}_{n,i}\mathbf{E}[\xi_{i+1}/\mathscr{F}_i], \quad n = 1, 2, \ldots$$

where

$$\mathbf{C}_{n,i} = \begin{cases} \gamma_i \prod_{j=i+1}^{n} (1-\gamma_j)\mathbf{I} & \text{if} \quad i < n, \\ \gamma_i \mathbf{I} & \text{if} \quad i = n, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that conditions (24), (25) and (26) of the Toeplitz theorem (Lemma 4) hold ($\mathbf{C}=\mathbf{I}$) so (1′) implies

$$\lim_{n\to\infty} \|Y_n-\xi\| = 0 \quad \text{a.s.} \tag{30}$$

If $\gamma_n = \dfrac{1}{n+1}$ then for $\{\mathbf{C}_{n,i}\}$, n, $i=1, 2, \ldots$ (28) holds too, so (1″) really implies (30).

Let $U_n=X_n-Y_n$, $n=1, 2, \ldots$. Then

$$U_{n+1} = \sum_{i=1}^{n} \mathbf{C}_{n,i}(\xi_{i+1}-\mathbf{E}[\xi_{i+1}/\mathscr{F}_i]), \quad n = 1, 2, \ldots.$$

Because of (30) we have only to show, that

$$\lim_{n\to\infty} \|U_n\| = 0 \quad \text{a.s.}$$

Obviously

$$\mathbf{E}\|U_n\|^2 = \mathbf{E}\left\| \sum_{i=1}^{n-1} \mathbf{C}_{n-1,i}(\xi_{i+1}-\mathbf{E}[\xi_{i+1}|\mathscr{F}_i]) \right\|^2 = \tag{31}$$

$$= \mathbf{E} \sum_{i=1}^{n-1} \|\mathbf{C}_{n-1,i}(\xi_{i+1}-\mathbf{E}[\xi_{i+1}|\mathscr{F}_i])\|^2 =$$

$$= \sum_{i=1}^{n} \|\mathbf{C}_{n-1,i}\|^2 \mathbf{E}\|\xi_{i+1}-\mathbf{E}[\xi_{i+1}/\mathscr{F}_i]\|^2.$$

Applying the definition of $\{\mathbf{C}_{n,i}\}$, $n, i=1, 2, \ldots$ and (3), (4), Lebesgue's dominated convergence theorem implies that

$$\lim_{n\to\infty} \mathbf{E}\|U_n\|^2 = 0. \tag{32}$$

Consider the sequence of random variables $\{V_n\}$, $n=1, 2, \ldots$ defined as

$$V_n = \|U_n\|^2 - \sum_{i=1}^{n-1} \gamma_i^2 \mathbf{E}[\|\xi_{i+1} - \mathbf{E}[\xi_{i+1}|\mathscr{F}_i]\|^2|\mathscr{F}_i],$$

$n=1, 2, \ldots$. It is easy to check that the sequence $\{(V_n, \mathscr{F}_n)\}$, $n=1, 2, \ldots$. Is a supermartingale and because of (4) and (32) there exists a number M such that

$$\mathbf{E}|V_n| \leqq M < +\infty, \quad n = 1, 2, \ldots.$$

This implies that there is a real random variable V for which

$$\lim_{n\to\infty} V_n = V.$$

However, (4) implies that

$$\sum_{i=1}^{\infty} \gamma_i^2 \mathbf{E}\left[\|\xi_{i+1} - \mathbf{E}[\xi_{i+1}|\mathscr{F}_i]\|^2|\mathscr{F}_i\right] < +\infty \quad \text{a.s.}$$

therefore

$$\lim_{n\to\infty} \|U_n\|^2 = V + \sum_{n=1}^{\infty} \gamma_n^2 \mathbf{E}\left[\|\xi_{n+1} - \mathbf{E}[\xi_{n+1}/\mathscr{F}_n]\|^2/\mathscr{F}_n\right].$$

PROOF OF LEMMA 1. Let $\{\mathbf{L}_n\}$, $n=1, 2, \ldots$ be a sequence of the linear operators in L_2 so that for any $g \in L_2$

$$\mathbf{L}_n g = \int K_n(\cdot, y) g(y) \lambda(dy), \quad n = 1, 2, \ldots. \tag{33}$$

Then, as

$$\mathbf{E}[K_n(\cdot, \eta_{n+1})/\eta_1, \ldots, \eta_n] = \mathbf{L}_n f_n, \quad n = 1, 2, \ldots. \tag{34}$$

We have to show that for (14) and $\lim_{n\to\infty} c_n = 0$

$$\lim_{n\to\infty} \|\mathbf{L}_n f_n - f\| = 0 \quad \text{a.s.}$$

This follows, however, from the fact that, for the sequence $\{\mathbf{L}_n\}$, $n=1, 2, \ldots$,

$$\lim_{n\to\infty} \|\mathbf{L}_n f - f\| = 0 \quad \text{a.s.}$$

(see e.g. [7]) and that

$$\|\mathbf{L}_n f_n - \mathbf{L}_n f\| \leqq \|\mathbf{L}_n\| \|f_n - f\| \leqq \int |h(x)| \lambda(dx) \|f_n - f\|.$$

PROOF OF LEMMA 2. Because of condition (14) it is enough to show that

$$\mathbf{E}\left[\|K_n(\cdot, \eta_{n+1}) - \mathbf{E}[K_n(\cdot, \eta_{n+1})/\eta_1, \ldots, \eta_n]\|^2/\eta_1, \ldots, \eta_n\right] \leqq \frac{\|h\|^2}{c_n^N}.$$

Obviously

$$\mathbf{E}[\|K_n(\cdot, \eta_{n+1}) - \mathbf{E}[K_n(\cdot, \eta_{n+1})/\eta_1, \ldots \eta_n]\|^2/\eta_1, \ldots, \eta_n] \leqq$$

$$\leqq \mathbf{E}[\|K_n(\cdot, \eta_{n+1})\|^2/\eta_1, \ldots, \eta_n] =$$

$$= \frac{1}{c_n^N} \iint h^2\left(\frac{x-y}{c_n}\right) f_n(y)\lambda(dx)\lambda(dy) = \frac{\|h\|^2}{c_n^N}.$$

PROOF OF LEMMA 3. Let $\{\mathbf{C}_{n,i}\}$, $n, i = 1, 2, \ldots$ be an array of operators in L_2 defined as

$$\mathbf{C}_{n,i} = \begin{cases} \frac{1}{n}\mathbf{L}_i, & \text{if} \quad i \leqq n, \\ 0, & \text{otherwise,} \end{cases}$$

where the sequence $\{\mathbf{L}_n\}$, $n = 1, 2, \ldots$, is defined by (33), then we have to prove that

$$\lim_{n\to\infty} \left\| \sum_{i=1}^{n} \mathbf{C}_{n,i} f_i - f \right\| = 0 \quad \text{a.s.} \tag{35}$$

For this purpose we use Lemma 4. In the proof of Lemma 1 we have shown that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{C}_{n,i} = \mathbf{I}$$

pointwise, and

$$\|\mathbf{C}_{n,i}\| \leqq \frac{1}{n} \int |h(x)|\,\lambda(dx).$$

First of all, for any $g \in L_2$,

$$\|\mathbf{C}_{n,i}g - \mathbf{C}_{n,i+1}g\| = \frac{1}{n^2} \int \left\{ \int \left[\frac{1}{c_{i+1}^N} h\left(\frac{x-y}{c_{i+1}}\right) - \frac{1}{c_i^N} h\left(\frac{x-y}{c_{i-1}}\right) \right] g(y)\lambda(dy) \right\}^2 \lambda(dx) =$$

$$= \frac{1}{n^2} \int \left\{ \int \left[\frac{c_i^N}{c_{i+1}^N} h\left(\frac{c_i}{c_{i+1}} z\right) - h(z) \right] g(x - c_i z)\lambda(dz) \right\}^2 \lambda(dx) =$$

$$= \frac{1}{n^2} \iint \left\{ \left[\frac{c_i^N}{c_{i+1}^N} h\left(\frac{c_i}{c_{i+1}} u\right) - h(u) \right] \left[\frac{c_i^N}{c_{i+1}^N} h\left(\frac{c_i^N}{c_{i+1}^N} v\right) - h(v) \right] \cdot \right.$$

$$\left. \cdot \int g(x - c_i u)\, g(x - c_i v)\lambda(dx) \right\} \lambda(du)\lambda(dv) \leqq$$

$$\leqq \frac{1}{n^2} \|g\|^2 \left\{ \int \left| \frac{c_i^N}{c_{i+1}^N} h\left(\frac{c_i}{c_{i+1}} z\right) - h(z) \right| \lambda(dz) \right\}^2.$$

Thus from condition (18) we obtain

$$\|\mathbf{C}_{n,i+1} - \mathbf{C}_{n,i}\| \leqq \frac{b}{n} \left| \frac{c_i^N}{c_{i+1}^N} - 1 \right|.$$

Therefore (18) implies that

$$\sup_n \sum_{i=1}^{n} i \|\mathbf{C}_{n,i} - \mathbf{C}_{n,i+1}\| < +\infty.$$

We have verified the conditions of Lemma 4, consequently (35) is met.

REFERENCES

[1] MOURIER, A.: Éléments aleatoires dans un espace de Banach, *Ann. Inst. H. Poincaré* **13** (1952)
[2] DRIML, M., HANS, O.: Conditional Expectation for Generalized Random Variables, *Trans of the Second Prague Conference on Information Theory, Statistical Decision Functions and Random Processes* (1959).
[3] GRENANDER, U.: *Probabilities on Algebraic Structures*, Almqvist and Wiksell, Stockholm.
[4] KOMLÓS, J., RÉVÉSZ, P.: On the Weighted Averages of Independent Random Variables, *Publications of the Math. Inst. of the Hung. Acad. Sci.* **9** (1964).
[5] WOLVERTON, C. T., WAGNER, T. J.: Asymptotically Optimal Discriminant Functions for Pattern Classification, *IEEE Trans. on Information Theory* **15** (1969), 258—265.
[6] REJTŐ, L., RÉVÉSZ, P.: Density Estimation and Pattern Recognition, *Problems of Control and Information Theory* **2** (1973), 67—80.
[7] GYÖRFI, L.: On some Estimation Problems in Pattern Recognition, *Problems of Control and Information Theory* **3** (1974), 11—17.
[8] WOLVERTON, C. T.: Strong Consistency of the Asymptotic Error Probability of the Nearest Neighbor Rule, *IEEE Trans. on Information Theory* **19** (1973), 119—120.
[9] GYÖRFI, L.: On the Estimation of Asymptotic Error Probability, *IEEE Trans. on Information Theory* **20** (1974), 277—278.
[10] FRITZ, J.: Learning from an Ergodic Training Sequence, *Limit Theorems of Probability Theory, ed. Révész, Nort-Holland* (1974).
[11] VENTER, J.: On Dvoretzky Stochastic Approximation Theorem, *Ann. Math. Stat.* **37** (1966), 1534—1544.
[12] BECK, A.: On the Strong Law of Large Numbers, *Ergodic Theory, ed. F. B. Wright, Academic Press* (1963).

Technical University of Budapest, Budapest XI, Stoczek u. 2, 1111, Hungary

(Received February 16, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 245—249

ON REPLACING COMPOSITE HYPOTHESES BY SIMPLE ONES

by

N. KUSOLITSCH

Summary: HUBER and STRASSEN had shown in 1973 (cf. [1]) that minimax-tests between two composite hypotheses can be reduced to tests between simple hypotheses of a fixed representative pair which is independent of the level of significance, if the two composite hypotheses are described in terms of 2-alternating capacities. In this paper we extend their result to the case of more than two composite hypotheses $\theta_1, \ldots, \theta_n$ showing that they can be replaced by simple hypotheses $P_1^\alpha, \ldots, P_n^\alpha$. In addition we give an example showing that, in general, these simple hypotheses cannot be chosen independently of the level of significance, even if all the composite hypotheses $\theta_1, \ldots, \theta_n$ are described by 2-alternating capacities.

Introduction: Let M be a Polish space and $\mathfrak{A}$ the system of its Borel sets. Further let $\mathfrak{p}$ be the set of all probability distributions on $(M, \mathfrak{A})$ and $\theta_1, \ldots, \theta_n$ n disjoint subsets of $\mathfrak{p}$. The generalized Neyman—Pearson problem for composite hypotheses may be formulated as follows: Given the n hypotheses $\theta_1, \ldots, \theta_n$ and a vector of significance $\alpha=(\alpha_1, \ldots, \alpha_{n-1})$ $(\alpha_i \in [0, 1] i=1(1)n-1)$ one has to find a test $\varphi_\alpha \in \Phi := \{\varphi\ (M, \mathfrak{A}) \to ([0, 1], \mathfrak{B} \cap [0, 1])\}$ satisfying both

$$\sup_{P \in \theta_i} \int \varphi_\alpha \, dP \leqq \alpha_i \quad \forall\, i = 1(1)n-1$$

and

$$\inf_{\theta_n} \int \varphi_\alpha \, dP = \sup_{\varphi \in \Phi_\alpha} \inf_{\theta_n} \int \varphi \, dP$$

with

$$\Phi_\alpha := \left\{\varphi \in \Phi\colon \sup_{\theta_i} \int \varphi \, dP \leqq \alpha_i \forall\, i = 1(1)n-1\right\}$$

LEHMANN [2] had given a solution to this problem for the case of simple hypotheses.

The case of composite hypotheses, $n=2$, had been successfully attacked by HUBER and STRASSEN [1]. They showed, under certain assumptions, that the composite hypotheses θ_1 and θ_2 can be replaced by properly chosen representative simple hypotheses P_1^α and P_2^α. This means that in finding optimal decision functions one can reduce the case of two composite hypotheses to that of two simple ones. In this paper we will extend this result to the case $n \geqq 2$. More precisely, we will show for a class of composite hypotheses that they can be represented by single measures $P_1^\alpha, \ldots, P_n^\alpha$, $P_i^\alpha \in \theta_i$, such that

$$\sup_{\Phi_\alpha} \inf_{\theta_n} \int \varphi \, d P = \sup_{\varphi\colon \int \varphi \, dP_i^\alpha \leqq \alpha_i,\ \forall i=1(1)n-1} \int \varphi \, dP_n^\alpha$$

holds true (see Theorem 4).

Huber and Strassen also proved that the pair P_1^α, P_2^α can be chosen independently of the "level-vector" α. Unfortunately this result can not be extended to the general $n \geqq 2$ case, as is shown by an example.

The replacement of the composite hypotheses by an n-tupel of distributions: We assume all θ_i to be convex and dominated by a σ-finite measure μ. We will consider the topology on $\mathfrak{p}$, for which all functionals $L_f(P) := \int f dP$ with bounded and continuous f: $M \to \mathbf{R}$ are continuous. This topology will be called weak topology, and we assume all θ_i to be compact in this sense.

We will use the minimax-theorem of the game theory and a theorem on the Lagrange-multipliers. They are recalled as Theorems 1 and 2:

THEOREM 1 [4]. *Under the assumptions:*
a) *U, V convex subsets of linear topological spaces,*
b) *U compact,*
c) *f: $U \times V \to \mathbf{R}$ concave-convex,*
d) *$f(\cdot, v)$: $U \to \mathbf{R}$ upper semicontinuous (USC) for every $v \in V$,*
the following equality holds true:

$$\sup_U \inf_V f(u, v) = \inf_V \sup_U f(u, v).$$

Checking the assumptions we get the following important

COROLLARY 1.
$$\sup_{\Phi} \left[\inf_{\theta_n} \int \varphi \, dP + \sum_{i=1}^{n-1} k_i \left(\alpha_i - \sup_{\theta_i} \int \varphi \, dP\right)\right] =$$

$$\inf_{\substack{\times \\ i=1}}^{n} \theta_i \sup_{\Phi} \left[\int \varphi \, dP_n + \sum_{i=1}^{n-1} k_i \left(\alpha_i - \int \varphi \, dP_i\right)\right] \quad \forall \alpha_i \in (0, 1),\ k_i \geqq 0,\ i = 1(1)n-1.$$

PROOF. Φ is a weak* compact subset of the set $L_\infty(\mu)$ of μ-ae (= almost everywhere) bounded functions.

Furthermore $f(\varphi;\ P_1, \ldots, P_n) := \int \varphi \, dP_n + \sum_{i=1}^{n-1} k_i(\alpha_i - \int \varphi \, dP_i)$ is linear and therefore concave-convex in φ and $(P_1, \ldots, P_n)$. Because the set $c(\mu)$ of bounded, countable additive set functions is isometrical isomorphic to the set $L_1(\mu)$ of μ-integrable functions, the equality $\lim_{n \to \infty} \int \varphi_n dP = \int \varphi \, dP \; \forall P \in \mathfrak{p} \cap c(\mu)$ holds true for every sequence of tests (φ_n), which is weakly convergent to a test φ. Therefore $f(\varphi;\ P_1, \ldots, P_n)$ is an USC function with respect to the weak* topology on $L_\infty(\mu)$ for every fixed n-tupel $(P_1, \ldots, P_n)$. So all assumptions of Theorem 1 are satisfied.

THEOREM 2 [3]. *Let X be a linear vector space, Ω a convex subset of X, f a real-valued concave functional on Ω and $\{g_i;\ i = 1(1)m\}$ a set of real-valued convex mappings on Ω. Assume that Ω contains a point x_1 for which $g_i(x_1) < 0 \ \forall i = 1(1)m$, and assume that $\mu_0 := \sup \{f(x)\colon x \in \Omega,\ g_i(x) \leqq 0 \ \forall i = 1(1)m\}$ is finite. Then there are nonnegative numbers k_i^0 $i = 1(1)m$ such that* $\mu_0 = \sup_{x \in \Omega} \left[f(x) - \sum_{i=1}^{m} k_i^0 g_i(x)\right]$.

Furthermore if the supremum μ_0 is achieved by $x_0 \in \Omega$, $g_i(x_0) \leqq 0 \ \forall i = 1(1)m$, then $\sum_{i=1}^{m} k_i^0 g_i(x_0) = 0$ *and* $f(x_0) - \sum_{i=1}^{m} k_i^0 g_i(x_0) = \mu_0$.

COROLLARY 2. *For every vector* $\alpha:=(\alpha_1, \ldots, \alpha_{n-1})\in(0, 1]^{n-1}$ *there exist nonnegative numbers* k_i^α $i=1(1)n-1$, *such that*

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP=\sup_{\Phi}\Big[\inf_{\theta_n}\int\varphi\,dP+\sum_{i=1}^{n-1}k_i^\alpha\Big(\alpha_i-\sup_{\theta_i}\int\varphi\,dP\Big)\Big].$$

PROOF. The corollary follows immediately from Theorem 2 if we set

$$X=L_\infty(\mu),\ x_1=\varphi_1\equiv 0,\ \Omega=\Phi,\ f(\cdot)=\inf_{\theta_n}\int\cdot\,dP\quad\text{and}\quad g_i(\cdot)=$$

$$=\sup\int\cdot\,dP-\alpha_i,\ i=1(1)n-1.$$

THEOREM 3. *For every vector* $\alpha\in(0, 1]^{n-1}$ *and all nonnegative numbers* k_i, $i=1(1)n-1$, *there exists an n-tupel* $(P_1^0, \ldots, P_n^0)$ *from* $\overset{n}{\underset{i=1}{\times}}\theta_i$ *such that*

$$\sup_{\Phi}\Big[\int\varphi\,dP_n^0+\sum_{i=1}^{n-1}k_i\Big(\alpha_i-\int\varphi\,dP_i^0\Big)\Big]=\inf_{\overset{n}{\underset{i=1}{\times}}\theta_i}\sup_{\Phi}\Big[\int\varphi\,dP_n+\sum_{i=1}^{n-1}k_i\Big(\alpha_i-\int\varphi\,dP_i\Big)\Big].$$

PROOF. Let Φ_c be the set of continuous decision functions φ. The function $f(\varphi;.,\ldots,.)$ on $\overset{n}{\underset{i=1}{\times}}\theta_i$ is continuous with respect to the weak topology. Therefore $\sup_{\Phi_c}f(\varphi;.,\ldots,.)$ is lower semicontinuous (LSC). Taking into account that the continuous functions are dense in $L_1(\nu)$ (ν a regular measure) one can show that $\sup_{\Phi_c}f(\varphi;\ P_1,\ldots,P_n)=\sup_{\Phi_n}f(\varphi;\ P_1,\ldots,P_n)$ for every fixed n-tupel $(P_1,\ldots,P_n)\in\overset{n}{\underset{i=1}{\times}}\theta_i$. Because of that $\sup_{\Phi}f(\varphi;.,\ldots,.)$ is LSC, too. This implies the existence of an n-tupel $(P_1^0,\ldots,P_n^0)\in\overset{n}{\underset{i=1}{\times}}\theta_i$ such that $\sup_{\Phi}f(\varphi;\ P_1^0,\ldots,P_n^0)=\inf_{\overset{n}{\underset{i=1}{\times}}\theta_i}\sup_{\Phi}f(\varphi;P_1,\ldots,P_n)$, since all θ_i are weakly compact.

By means of these theorems we can formulate our main result.

THEOREM. *For every vector* $\alpha\in(0, 1]^{n-1}$, *there is an n-tupel* $(P_1^\alpha,\ldots,P_n^\alpha)\in\overset{n}{\underset{i=1}{\times}}\theta_i$ *for which*

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP=\min_{\overset{n}{\underset{i=1}{\times}}\theta_i}\ \sup_{\varphi:\int\varphi\,dP_i\leqq\alpha_i\ \forall i=1(1)n-1}\int\varphi\,dP_n=\sup_{\varphi:\int\varphi\,dP_i^\alpha\leqq\alpha_i\ \forall i=1(1)n-1}\int\varphi\,dP_n^\alpha.$$

PROOF. Because of Corollary 2 there are $n-1$ nonnegative numbers k_i^α for which

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP=\sup_{\Phi}\Big[\inf_{\theta_n}\int\varphi\,dP_n+\sum_{i=1}^{n-1}k_i^\alpha\Big(\alpha_i-\sup_{\theta_i}\int\varphi\,dP_i\Big)\Big].$$

It follows from Corollary 1 that this expression is equal to $\inf_{\overset{n}{\underset{i=1}{\times}}\theta_i}\sup_{\Phi}\Big[\int\varphi\,dP_n+$

$+\sum_{i=1}^{n-1}k_i^\alpha\Big(\alpha_i-\int\varphi\,dP_i\Big)\Big]$, and Theorem 3 implies the existence of an n-tuple $(P_1^\alpha,\ldots,P_n^\alpha)$

such that the infimum above is equal to $\sup_{\Phi}\left[\int\varphi\,dP_n^{\alpha}+\sum_{i=1}^{n-1}k_i^{\alpha}\left(\alpha_i-\int\varphi\,dP_i^{\alpha}\right)\right]$. So we get

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP=\sup_{\Phi}\left[\int\varphi\,dP_n^{\alpha}+\sum_{i=1}^{n-1}k_i^{\alpha}\left(\alpha_i-\int\varphi\,dP_i^{\alpha}\right)\right].$$

This implies

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP\geqq\sup_{\varphi:\,\int\varphi\,dP_i^{\alpha}\leqq\alpha_i\ \forall i=1(1)n-1}\int\varphi\,dP_n^{\alpha}.$$

On the other hand, we get

$$\sup_{\Phi_\alpha}\inf_{\theta_n}\int\varphi\,dP\leqq\sup_{\Phi_\alpha}\int\varphi\,dP_n\leqq\sup_{\varphi:\,\int\varphi\,dP_i\leqq\alpha_i\ \forall i=1(1)n-1}\int\varphi\,dP_n\quad\forall(P_1,\ldots,P_n)\in\bigtimes_{i=1}^{n}\theta_i,$$

which completes the proof.

The above used technic of Lagrange-multipliers is adequate to the dual-programing technic used by KRAFFT and WITTING [5] to solve the case for $n=2$.

HUBER and STRASSEN had shown [1] that the representation $(P_1^{\alpha}, P_2^{\alpha})$ may be chosen independent of α in the case $n=2$, provided the hypotheses are described by 2-alternating capacities. In what follows we will present an example, showing that, in general, this condition does not ensure the existence of level-independent representation.

At first we recall a definition.

DEFINITION. Let M be as before and let 2^M denote its power set. A function $v\colon 2^M\to[0,1]$ is called 2-alternating capacity if

(i) $v(\emptyset)=0$, $v(M)=1$.
(ii) $A\subset B\Rightarrow v(A)\leqq v(B)$.
(iii) If $A_i\subset M$, $i=0,1,\ldots$ and $A_n\uparrow A_0$ then $v(A_n)\uparrow v(A_0)$.
(iv) If $F_n, F_0\in\mathfrak{F}\ \forall n\in\mathbf{N}$, $F_n\downarrow F_0$ then $v(F_n)\downarrow v(F_0)$ where $\mathfrak{F}$ is the system of closed subsets of M.
(v) $v(A\cup B)+v(A\cap B)\leqq v(A)+v(B)$.

Let v_i $i=1,2$ be two 2-alternating capacities and let the hypotheses θ_1, θ_2 be of the form $\theta_i=\{P\in\mathfrak{p};\ P(A)\leqq v(A)\ \forall A\in\mathfrak{A}\}$. Then the main result of HUBER—STRASSEN [1] is that there exists a pair of "uniformly least favourable" distributions $(P_0, Q_0)\in\theta_1\times\theta_2$, such that

$$\sup_{\varphi:\,\sup_{\theta_1}\int\varphi\,dP\leqq\alpha}\inf_{\theta_2}\int\varphi\,dQ=\sup_{\varphi:\,\int\varphi\,dP_0\leqq\alpha}\int\varphi\,dQ_0\quad\forall\alpha\in[0,1].$$

The following counterexample shows that this result cannot be extended even to the case $n=3$.

EXAMPLE. Let $M=\{1,2\}$, $\mathfrak{A}=2^M$, $\theta_1=\{P_1:=(1,0)\}$, $\theta_2=\{P_2=(0,1)\}$, $\theta_3=\{Q=(q,1-q)\colon q\in[1/3,2/3]\}$.

These hypotheses can be described by the following 2-alternating capacities.

A	$v_1(A)$	$v_2(A)$	$v_3(A)$
M	1	1	1
$\emptyset$	0	0	0
$\{1\}$	1	0	2/3
$\{2\}$	0	1	2/3

It is easy to see that these three functions are really 2 alternating capacities.

We will show by direct computation that no triple (P_1, P_2, Q) can represent θ_1, θ_2, θ_3 for the "level-vectors" $\alpha'=(1/6, 1)$ and $\alpha''=(1, 1/6)$ at the same time. First we consider the case $\alpha'=(1/6, 1)$:

$$\sup_{\varphi:\int\varphi\,dP_1\leqq\frac{1}{6}}\inf_{\theta_3}\int\varphi\,dQ=\sup_{\varphi:\varphi(1)=\frac{1}{6}}\inf_{q\in\left[\frac{1}{3},\frac{2}{3}\right]}[\varphi(1)q+\varphi(2)(1-q)]=$$

$$=\sup_{\varphi:\varphi(1)=\frac{1}{6}}\left(\frac{1}{6}\cdot\frac{2}{3}+1\cdot\frac{1}{3}\right)=\frac{8}{18}=\sup_{\varphi:\varphi(1)=\frac{1}{6}}\int\varphi Q'\quad\text{with}\quad Q'=\left(\frac{2}{3},\frac{1}{3}\right).$$

But $\sup_{\varphi:\int\varphi\,dP_1\leqq\frac{1}{6}}\int\varphi\,dQ''$ with $Q''=\left(\frac{1}{3},\frac{2}{3}\right)$ is equal to $\frac{13}{18}$.

Analogous we get for $\alpha''=(1, 1/6)$

$$\sup_{\varphi:\int\varphi\,dP_2\leqq\frac{1}{6}}\inf_{\theta_3}\int\varphi\,dQ=\sup_{\varphi:\varphi(2)=\frac{1}{6}}\int\varphi\,dQ''=\frac{8}{18}<\sup_{\varphi(2)=\frac{1}{6}}\int\varphi\,dQ'=\frac{13}{18}.$$

REFERENCES

[1] HUBER, P. J.—STRASSEN, V.: Minimax Tests and the Neyman—Pearson Lemma for Capacities, *Ann. Stat.* **1** (1973), (251—263).

[2] LEHMANN, E. L.: *Testing Statistical Hypotheses*, J. Wiley & Sons, New York, 1959.

[3] LUENBERGER, D. G.: *Optimization by Vector Space Methods*, J. Wiley & Sons, New York, 1969.

[4] SION, M.: On General Minimax Theorems, *Pac. Journ. Math.* **8** (1958), (171—175.)

[5] WITTING, H.—KRAFFT, O.: Optimale Tests und ungünstige Verteilungen, *Z. Wahrscheinlichkeitstheorie* **7** (1967), (289—302).

Institut für Statistik, Technische Universität Wien, A—1040 Wien 4, Argentinierstr. 8/7

(Received March 3, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 251—256

A CATEGORY-THEORETICAL CHARACTERIZATION OF SURJECTIVE HOMOMORPHISMS OF PARTIAL ALGEBRAS

by

ANA PÁSZTOR

INTRODUCTION. A central problem in the investigation of partial algebras is the importation of the Birkhoff-variety concept and the theory of the Galois connections related to it from total algebras to partial algebras (cf. SLOMINSKI [10], KERKHOFF [7], HÖFT [5], BURMEISTER [3], POYTHRESS [8], JOHN [6], ANDRÉKA—NÉMETI [1, 2]). This importation was done with tools of partial algebra, but with tools of category theory too. The second way of importation resulted in a generalized identity concept which is the function of three operators: H — homomorphic image, S — subalgebra and P — product operators (cf. ANDRÉKA—NÉMETI [1, 2]). Choosing for H the weak homomorphic image operator and for S the strong subalgebra operator, Andréka—Németi got a generalized version of the so called "strong identities" (cf. HÖFT [5], BURMEISTER [3], JOHN [6]) which were the first step in the investigation of partial algebras (see SCHMIDT [9]).

In the last ten years the category theorists tried to find an adequate category theoretical generalization of the concept of weak homomorphic image or with other words of that of surjective homomorphism, since they discovered that in (total) varieties it is different from that of epimorphism (cf. in variety of semigroups). The class of strong epimorphisms (H_s) is in total varieties exactly the class of surjective homomorphisms, but even in similarity classes of partial algebras they doesn't coincide. In ANDRÉKA—NÉMETI [1, 2] the authors give the definition of relative epimorphism, which is a good generalization of the surjective epimorphism, but which is, since defined by help of special objects, somewhat too "local". I tried in this paper to give a "global" definition of relative epimorphism, which is easy to work with. The factorization pair of relative epimorphism, the relative monomorphism, gives the category theoretical generalization of relative subalgebra concept (cf. ANDRÉKA—NÉMETI [1, 2]).

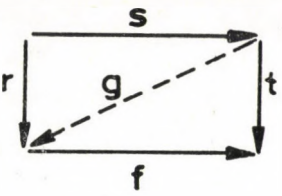
Symbols and notations

$\mathscr{C}$	— a category
Mor $\mathscr{C}$	— the class of morphisms of $\mathscr{C}$
$\lvert\mathscr{C}\rvert$	— the class of objects of $\mathscr{C}$
1_a	— the identity morphism of a
$fg\colon a\to c$	— the composition of $f\colon a\to b$ and $g\colon b\to c$
Is	— the class of all isomorphisms of $\mathscr{C}$
Epi	— the class of all epimorphisms of $\mathscr{C}$
Mono	— the class of all monomorphisms of $\mathscr{C}$
MN	$= \{fg \mid f\in M,\ g\in N\}$, where $M, N\subseteq \mathrm{Mor}\,\mathscr{C}$
$M\mathscr{A}$	$= \{a\in\lvert\mathscr{C}\rvert \mid \exists\, a \xrightarrow{f} b,\ f\in M,\ b\in\mathscr{A}\}$

$\underset{\leftarrow}{M}\mathscr{A}$ $= \{a \in |\mathscr{C}| \mid \exists\, a \xleftarrow{f} b,\ f \in M,\ b \in \mathscr{A}\}$ and $M \subseteq \operatorname{Mor}\mathscr{C}$, $\mathscr{A} \subseteq |\mathscr{C}|$

$\underset{i\in I}{P}\langle f_i\rangle$ — the product of f_i-s $(i \in I)$

$\Lambda(\mathscr{E})$ — for $\mathscr{E} \subseteq \operatorname{Mor}\mathscr{C}$ — is, by STRECKER [11], the class of such $f \in \operatorname{Mor}\mathscr{C}$, which satisfy the condition of $\mathscr{E}$-lower diagonalizability, i.e. for any commutative diagram



with $s \in \mathscr{E}$, there exists g, which makes both triangles commute.

$(\mathscr{H}, S)$-category — a category which is uniquely $(\mathscr{H}, S)$-factorizable (see HERRLICH—STRECKER [4], definition 17.15) and $\mathscr{H}\mathscr{H} \subseteq \mathscr{H}$, $SS \subseteq S$

P^t — the class of all t-type partial algebras

$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \ldots$ — partial algebras, i.e. $\mathfrak{A} = \langle A, f_i^{\mathfrak{A}}\rangle_{i\in I}$ with $t(i) \in \omega$ $(i \in I)$

$f(\mathfrak{A})$ $= \langle f(A), f(f_i^{\mathfrak{A}})\rangle_{i\in I}$ for $f\colon \mathfrak{A} \to \mathfrak{B} \in \operatorname{Mor} P^t$

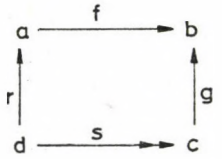
pr_A — the projection map from a product to the factor A

S_s — the class of strong monomorphisms (in the sense of HERRLICH—STRECKER [4] p. 265.) i.e. $S_s = \Lambda(\mathrm{Epi})$

t — an arbitrary similarity type (for algebras), i.e. $t\colon I \to \omega$, where I is the set of operation symbols

ω — the set of natural numbers

DEFINITION 1. $f\colon a \to b$ is a *relative epimorphism* iff for any $g\colon c \to b$, (f, g) has a weak source-pair (r, s) with $s \in \mathrm{Epi}$, i.e. there is an $r\colon d \to a$ and an $s\colon d \to c$ with $rf = sg$ and $s \in \mathrm{Epi}$.



NOTATION. H_r is the class of relative epimorphisms in $\mathscr{C}$.

THEOREM 2.

(1) $H_r \subseteq \mathrm{Epi}$.

(2) *H_r is closed under composition, i.e. $H_r H_r \subseteq H_r$.*

(3) *H_r is closed under left cancellation, i.e. if for any $f, g \in \operatorname{Mor}\mathscr{C}$ $fg \in H_r$, then $g \in H_r$.*

(4) *H_r is closed under right cancellation w.r.t. monomorphisms, i.e. if for any $f, g \in \operatorname{Mor}\mathscr{C}$ $fg \in H_r$ and $g \in \mathrm{Mono}$, then $f \in H_r$.*

(5) *H_r is closed under formation of pullbacks.*

(6) *If $\mathscr{C}$ has pullbacks, then for any $S \subseteq \operatorname{Mor}\mathscr{C}$ closed under formation of pullbacks and for any $\mathscr{A} \subseteq |\mathscr{C}|$ $\underset{\rightarrow\leftarrow}{S}\underset{\leftarrow\rightarrow}{H_r}\mathscr{A} \subseteq H_r S\mathscr{A}$.*

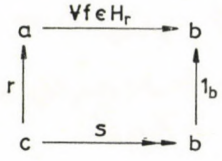
(7) *If $\mathscr{C}$ has enough H_r-projective objects, i.e. if*

$$(\forall a \in |\mathscr{C}|)(\exists b \in P_j(H_r))(\exists f\colon b \to a) \quad f \in H_r,$$

then H_r is closed under formation of products, i.e. $PH_r \subseteq H_r$.

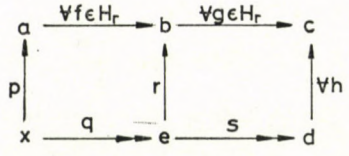
PROOF.

(1) Let $f \in H_r$ be arbitrary. Let (r, s) be a weak source-pair of $(f, 1_b)$ with $s \in \mathrm{Epi}$.

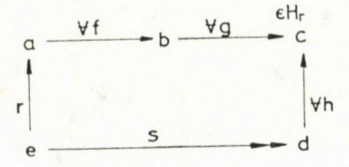


Then, since Epi is closed under left cancellation and $rf = s1_b = s \in \mathrm{Epi}$, $f \in \mathrm{Epi}$.

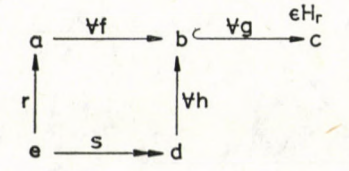
(2) Let $f\colon a \to b$, $g\colon b \to c \in H_r$ and $h\colon d \to c$ be arbitrary. (r, s) is a weak source-pair of (g, h) with $s \in \mathrm{Epi}$ and (p, q) of (f, r) with $q \in \mathrm{Epi}$. Since Epi is closed under composition, $qs \in \mathrm{Epi}$. Now, $p(fg) = qrg = (qs)h$. Hence, $fg \in H_r$.



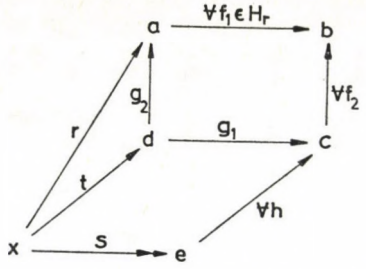
(3) Let $f\colon a \to b$, $g\colon b \to c \in \mathrm{Mor}\,\mathscr{C}$ with $fg \in H_r$ and $h\colon d \to c$ be arbitrary. (r, s) is a weak source-pair of (fg, h) with $s \in \mathrm{Epi}$. Then (rf, s) is a weak source-pair for (g, h). Thus, $g \in H_r$.



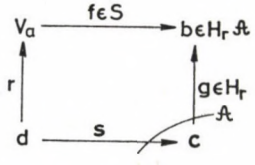
(4) Let $f\colon a \to b$, $g\colon b \to c$ be arbitrary with $fg \in H_r$ and $g \in \mathrm{Mono}$. For any $h\colon d \to b$ let (r, s) be a weak source-pair of (fg, hg) with $s \in \mathrm{Epi}$. $g \in \mathrm{Mono}$ implies that (r, s) is a weak source-pair of (f, h) $(rfg = shg \Rightarrow rf = sh)$. Hence, $f \in H_r$.



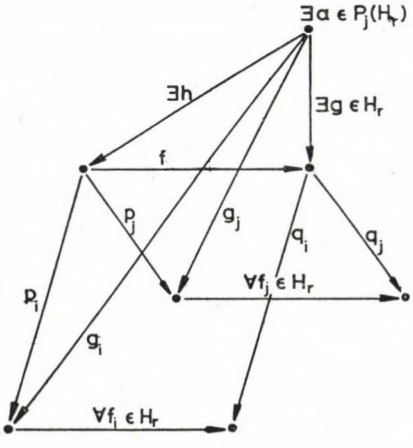
(5) For any $f_1\colon a\to b\in H_r$ and $f_2\colon c\to b\in \mathrm{Mor}\,\mathscr{C}$ let (g_1, g_2) denote their pullback. Consider an arbitrary $h\colon e\to c\in\mathrm{Mor}\,\mathscr{C}$. Denote with (r, s) a weak source-pair of (f_1, hf_2), where $s\in\mathrm{Epi}$. Now (t, s) is a weak source-pair of (g_1, h), where t is the unique morphism in the definition of pullback. Thus, $g_1\in H_r$.

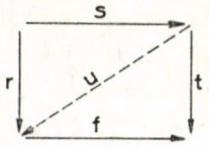


(6) Let $a\in \underset{\leftarrow}{S}\underset{\rightarrow}{H_r}\mathscr{A}$. Then there are b, f, g, c as in the diagram, where $f\in S$, $g\in H_r$ and $c\in\mathscr{A}$. If (r, s) denotes their pullback, then $s\in S$ and by (5) $r\in H_r$; hence $a\in H_r S\mathscr{A}$.



(7) Let $\langle f_i\rangle_{i\in I}\in {}^I H_r$ be arbitrary. Let $f=\underset{i\in I}{P}\langle f_i\rangle$. For codo f there is a H_r-projective object a and $g\colon a\to \mathrm{codo}\, f\in H_r$. Since $(\forall i\in I) f_i\in H_r$, $(\forall i\in I)(\exists g_i) g_i\cdot f_i=g\cdot q_i$, where the q_i-s are projections. But $\mathrm{do}\, f=\underset{i\in I}{P}\langle \mathrm{do}\, f_i\rangle$ implies that there is an h with $(\forall i\in I) h\cdot p_i=g_i$, where the p_i-s are projections. Now $(\forall i\in I) h\cdot f\cdot q_i=h\cdot p_i\cdot f_i=$ $=g_i\cdot f_i=g\cdot q_i$. Since $\langle q_i\rangle_{i\in I}$ is a mono-source, $h\cdot f=g\in H_r$ holds, which implies by (3) $f\in H_r$.



DEFINITION 3. $f \in \operatorname{Mor} \mathscr{C}$ is called a *relative monomorphism* iff for any commutative diagram 

with $s \in H_r$, there is a unique u which makes both triangles commute.

NOTATION. S_r is the class of all relative monomorphisms of $\mathscr{C}$.

THEOREM 4. *In any category* $\mathscr{C}$,
(1) $S_r \supseteq S_s$, *i.e. the strong monomorphisms are also relative monomorphisms.*
(2) $S_r S_r \subseteq S_r$, *i.e.* S_r *is closed under composition.*
(3) S_r *is closed under right cancellation, i.e. for any* f, g *with* $fg \in S_r$, $f \in S_r$.
(4) S_r *is closed under formation of pullbacks.*
(5) S_r *is closed under formation of intersections.*
(6) S_r *is closed under formation of products, i.e.* $PS_r \subseteq S_r$.

PROOF. See Strecker [11] with $S_r = \Lambda(H_r)$.

THEOREM 5. (Representation theorem for P^t.) $f\colon \mathfrak{A} \to \mathfrak{B} \in \operatorname{Mor} P^t$ *is a relative epimorphism in* P^t *iff it is onto.*

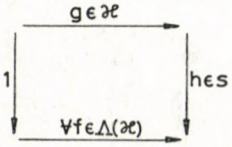
PROOF. 1) Suppose $f\colon \mathfrak{A} \to \mathfrak{B} \in \operatorname{Mor} P^t$ is a relative epimorphism in P^t. We want to prove $f(A) = B$. If $B = \emptyset$ then we are ready. Let $b \in B$ be arbitrary and consider $g\colon \{b\} \to B$ where $g = \mathrm{id}_{\{b\}}$. Since $f \in H_r$ there are $r\colon \mathfrak{C} \to \mathfrak{A}$ and $s\colon \mathfrak{C} \to \{b\}$ with $rf = sg$ and $s \in \operatorname{Epi}$. Now for any $c \in C$, $f(r(c)) = g(s(c)) = b$, hence $b \in f(A)$.

2) Let $f\colon \mathfrak{A} \to \mathfrak{B}$ be onto. We want to prove that $f \in H_r$. Therefore consider an arbitrary $g\colon \mathfrak{C} \to \mathfrak{B}$. Let $D = \{(a, c) \in A \times C \mid f(a) = g(c)\}$. Since f is onto $\mathrm{pr}_c \restriction D\colon D \to C$ is epi. Now $(\mathrm{pr}_A \restriction D, \mathrm{pr}_c \restriction D)$ gives a source-pair for (f, g) which proves $f \in H_r$.

NOTE. The above characterization of H_r in P^t works both if empty universes are allowed in P^t as well as in the case when empty universes are not allowed in P^t. Specifically, it works for the category "Cat" of all small categories.

LEMMA. *Let* $\mathscr{H} \subseteq \operatorname{Epi}$ *and* $\operatorname{Is} \subseteq S$. *If* $\mathscr{C}$ *is an* $(\mathscr{H}, S)$*-category, then* $S = \Lambda(\mathscr{H})$.

PROOF. 1) By the proof of HERRLICH—STRECKER [4] theorem 33.3, if $\mathscr{C}$ is a $(\mathscr{H}, S)$-category then $S \subseteq \Lambda(\mathscr{H})$.
2) Now let $f \in \Lambda(\mathscr{H})$ be arbitrary and let $f = gh$ with $g \in \mathscr{H}$ and $h \in S$.



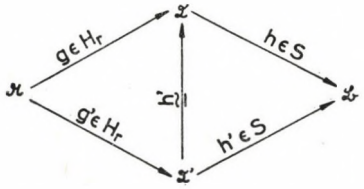
By definition of $\Lambda(\mathscr{H})$, there is a k with $gk = 1$ and $kf = h$. Hence g is a section and thus an isomorphism. This implies $gh = f \in S$.

THEOREM 6. (Representation theorem for P^t.) *$f\colon \mathfrak{A}\to\mathfrak{B}\in \mathrm{Mor}\, P^t$ is a relative monomorphism iff f is injective and $f(\mathfrak{A})$ is a relative subalgebra of $\mathfrak{B}$, i.e. iff f is injective and*

$$(*)\qquad (\forall i\in I)(\forall a\in {}^{t(i)}A)\big(f_i^{\mathfrak{B}}(f(a))=f(b)\to f_i^{\mathfrak{A}}(a)=b\big).$$

PROOF. Let us denote with S the class of those morphisms, which are injective and satisfy $(*)$. We'll show that P^t is a (H_r, S)-category which implies by the lemma that $S=A(H_r)=S_r$. By theorem 2, $H_rH_r\subseteq H_r$ and it is easy to see that also $SS\subseteq S$ holds. Now we show that any $f\colon \mathfrak{A}\to\mathfrak{B}\in\mathrm{Mor}\, P^t$ has up to isomorphism a unique factorization $f=gh$ with $g\in H_r$ and $h\in S$.

Denote by $\mathfrak{C}=\langle f(A), f_i^{\mathfrak{B}}\cap {}^{(t(i)+1)}f(A)\rangle_{i\in I}$ and let $g\colon \mathfrak{A}\to\mathfrak{C}$ be defined by $g=f$ and $h\colon \mathfrak{C}\to\mathfrak{B}$ by $h=\mathrm{id}_{f(A)}$. Since $(\forall i\in I)\ \big(f(f_i^{\mathfrak{A}})\subseteq f_i^{\mathfrak{C}}\subseteq f_i^{\mathfrak{B}}\big)$, g and h are homomorphisms. By theorem 5 $g\in H_r$ and $h\in S$ since it is injective and satisfies $(*)$. Now let $f=g'h'$ with $g'\in H_r$ and $h'\in S$.



Since $f(A)=h'\big(g'(A)\big)=h'(C')$ and $h'\in S$, $h'(\mathfrak{C}')=\langle f(A), f_i^{\mathfrak{B}}\cap {}^{(t(i)+1)}f(A)\rangle_{i\in I}=\mathfrak{C}$. Hence h' is an isomorphism between $\mathfrak{C}'$ and $\mathfrak{C}$. Now, $(\forall a\in A)\ g(a)=f(a)=h'\big(g'(a)\big)$, hence $g=g'h'$. This, by $g'\in\mathrm{Epi}$, implies $h'h=h'$.

REFERENCES

[1] ANDRÉKA, H.—NÉMETI, I.: Generalisation of variety and quasivariety-concept to partial algebras through category theory, *Preprint* 1976, *Math. Inst. Hung. Acad. Sci.* (Part of its content can be found in [2].)

[2] ANDRÉKA, H.—NÉMETI, I.: A general axiomatisability theorem using cone-injective subcategories, *Proc. Coll. Univ. Alg. 77 Esztergom.*

[3] BURMEISTER, P.: Primitive Klassen partieller Algebren, *Habilitationsschrift, Universität Bonn,* 1971.

[4] HERRLICH, H.—STRECKER, G. E.: *Category Theory,* Allyn and Bacon Inc., Boston. 1973.

[5a] HÖFT, H.: *Equations in partial algebras,* Dissertation, University of Houston, 1970.

[5b] HÖFT, H.: Weak and strong equations in partial algebras, *Algebra Universalis* **3** (1973), 203—215.

[6] JOHN, R.: *Gültigkeitsbegriffe für Gleichungen in partiellen Algebren,* T. H. Darmstadt, Dissertation, 1976.

[7] KERKHOFF, R.: Gleichungsdefinierbare Klassen partieller Algebren, *Math. Annalen* **185** (1970), 112—133.

[8] POYTHRESS, V. S.: Partial morphisms on partial algebras, *Algebra Universalis* **3** (1973), 182—202.

[9] SCHMIDT, J.: *Universal Algebra,* Univ. of Georgia, Athens, Ga. 1967/68.

[10] SLOMINSKI, J.: Peano-algebras and quasi algebras, *Diss. Mathematical, Rozprawy Mat.* **57** (1968).

[11] STRECKER, G. E.: Epireflective operators *vs* perfect morphism and closed classes of epimorphisms, *Bull. Austr. Math. Soc.* **7** (1972), 359—366.

Institut für Informatik, D—7000 Stuttgart 1, Azenbergstr. 12

(Received March 31, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 257—265

ON THE SCHRÖDINGER EQUATION OF THE THREE-BODY PROBLEM, II

by

E. MAKAI

1. Let $P_i(x_i, y_i, z_i)$, $i=1, 2, 3$ be three points in the 3-space,

$$r_1 = [(x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2]^{1/2}, \quad r_2 = [(x_3-x_1)^2+(y_3-y_1)^2+(z_3-z_1)^2]^{1/2},$$

$$r_3 = [(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{1/2}, \quad \Delta_i = \partial^2/\partial x_i^2+\partial^2/\partial y_i^2+\partial^2/\partial z_i^2$$

and consider those solutions of the Schrödinger equation

$$\left(-\sum_{i=1}^{3} \mu_i \Delta_i + U - E\right)\psi = 0 \tag{1.1}$$

[μ_i, E constants, $U=U(r_1, r_2, r_3)$] which depend only on the quantities r_1, r_2, r_3. In the special case

$$\mu_1 = \mu_2 = 1/2, \quad \mu_3 = 0, \quad U = -2r_1^{-1}-2r_2^{-1}+\tfrac{1}{2}r_3^{-1} \tag{1.2}$$

T. KINOSHITA [4] observed that introducing the quantities

$$s = r_1+r_2, \quad p = \frac{r_3}{r_1+r_2}, \quad q = \frac{r_1-r_2}{r_3} \tag{1.3}$$

equation (1.1) admits formal solutions of the type

$$e^{-ks} \sum_{l,m,n=0}^{\infty} c_{lmn} s^l p^m q^n, \tag{1.4}$$

or what amounts to the same, formal solutions of the type

$$\sum_{l,m,n=0}^{\infty} C_{lmn} s^l p^m q^n. \tag{1.5}$$

He conjectured that the first eigenfunction of equation (1.1) in the special case (1.2) is of the form

$$e^{-ks} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{m} c_{lmn} s^l p^m q^n = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{m} C_{lmn} s^l p^m q^n. \tag{1.6}$$

The question arose whether there exist formal solutions to (1.1) of the form (1.6) and subsequently it was shown by G. MUNSCHY and PH. PLUVINAGE [6] that in the special case (1.2) there do exist formal solutions of type (1.6).

In the present paper by an admittedly longer analysis we derive Theorem 2, a consequence of which, namely Theorem 1 generalizes the result of Munschy and Pluvinage to the case where

(i) μ_1, μ_2, μ_3, E are any constants with the restriction $\mu_1+\mu_2\neq 0$ and

(ii) the function $L^0=w\cdot(U-E)$ with $w=(1-p^2q^2)p^2s^2$ is of the form

$$L^0=\sum_{k=0}^{\infty} p^k L_k^0(s,q). \tag{1.7}$$

Here $L_0^0(s,q)=0$ and for $k>0$ $L_k^0(s,q)$ is a polynomial in q of degree $k-1$:

$$L_k^0(s,q)=\sum_{\nu=0}^{k-1}\gamma_{k\nu}(s)\,q^\nu \tag{1.7'}$$

where the functions $\gamma_{k\nu}(s)$ are infinitely many times differentiable in some interval $s_1<s<s_2$.

For the sake of convenience we shall assume that

$$L_k^0(s,q)=0 \quad \text{if} \quad k>4. \tag{1.7''}$$

It is easily seen that the function

$$U-E=\sum e_i r_i^{-1}-E \quad (e_i=\text{const.}) \tag{1.8}$$

satisfies conditions (1.7), (1.7′) and (1.7″). Indeed by (1.3)

$$wr_1^{-1}=2p^2s(1-pq), \quad wr_2^{-1}=2p^2s(1+pq), \quad wr_3^{-1}=ps(1-p^2q^2).$$

THEOREM 1. *Under assumptions* (1.7) *and* (1.7′) *and if* $\mu_1+\mu_2\neq 0$ *equation* (1.1) *has formal solutions of the type*

$$\psi=\sum_{m=0}^{\infty}\sum_{n=0}^{m}\alpha_{mn}(s)\,p^m q^n, \tag{1.9}$$

where the $\alpha_{mn}(s)$*'s are infinitely many times differentiable in some interval* $s_1<s<s_2$.

THEOREM 2. *If* $\mu_1+\mu_2\neq 0$, (1.7) *and* (1.7′) *hold, then the most general functions* $g_m(s,q)$ *in the formal solution*

$$\psi=\sum_{m=0}^{m} g_m(s,q)\,p^m \tag{1.10}$$

of equation (1.1) *are of the form*

$$g_{mn}(s,q)=\sum_{n=0}^{m}\left[\alpha_{mn}(s)\,q^n+\beta_{mn}(s)q^n\log\frac{1+q}{1-q}\right]. \tag{1.11}$$

Alternatively, if $P_n(x)$ is the Legendre polynomial of degree n and

$$Q_n(x)=\frac{1}{2}P_n(x)\log\frac{1+x}{1-x}-\frac{2n-1}{1\cdot n}P_{n-1}(x)-\frac{2n-5}{3(n-1)}P_{n-3}(x)-\ldots \tag{1.12}$$

$$(-1<x<1)$$

is the Legendre function of the second kind [2; vol. I. p. 141] then *the coefficients of the most general formal solution* (1.10) *can be expressed in the form*

$$g_{mn}(s, q) = \sum_{n=0}^{m} [a_{mn}(s) P_n(q) + b_{mn}(s) Q_n(q)]. \tag{1.11'}$$

The functions $a_{mn}(s)$ *and* $b_{mn}(s)$ *(and thus* $\alpha_{mn}(s)$ *and* $\beta_{mn}(s)$*, too) satisfy recurrence formulas: for a fixed m and n,* $n<m$, $a_{mn}(s)$ *and* $b_{mn}(s)$ *are uniquely defined by the choice of the functions*

$$a_{00}(s), a_{11}(s), \ldots, a_{m-1,m-1}(s); \quad b_{00}(s), b_{11}(s), \ldots, b_{m-1,m-1}(s). \tag{1.13}$$

The functions $a_{mm}(s)$ *and* $b_{mm}(s)$, $m=0, 1, 2, \ldots$ *can be arbitrarily chosen, provided only they are infinitely many times differentiable in* $s_1<s<s_2$. *Finally all functions* $a_{mn}(s)$ *and* $b_{mn}(s)$ *[*$\alpha_{mn}(s)$ *and* $\beta_{mn}(s)$*] are of class* $C^\infty(s_1, s_2)$.

COROLLARY. If $b_{mm}(s)=0$ $(m=0, 1, \ldots)$ then all $b_{mn}(s)$'s vanish. In this special case (1.10) is of the form (1.9) generalizing the result of Munschy and Pluvinage. If, however, $b_{00}(s)=b_{11}(s)=\ldots=b_{m-1,m-1}(s)=0$, but $b_{mm}(s)\neq 0$, then the functions $g_m(s, q)$, $g_{m+1}(s, q)$, ... exhibit logarithmic singularities at $q=\pm 1$.

Note that by (1.3) $q=\pm 1$ corresponds to a degenerated shape of the triangle $P_1P_2P_3$ the points P_1, P_2 and P_3 lying on a straight line. If (1.8) holds, then by a general theorem of V. H. FROIM [1], [5] solutions of (1.1) analytic in a complex neighbourhood of $q=\pm 1$, i.e. in a domain $|q\mp 1|<\varepsilon_1$, $|p-p_0|<\varepsilon_2$, $|s-s_0|<\varepsilon_3$ are either regular at $q=\pm 1$ or they have a logarithmic singularity there. On the other hand, T. KATO [7] has shown that if (1.8) holds, then the eigenfunctions of (1.1) are continuous everywhere, thus in a neighbourhood of $q=\pm 1$, too.

If (1.7) and (1.7') hold, then by (1.9) formal solutions of (1.1) regular everywhere in the domain $r_1+r_2\geqq r_3$, $r_2+r_3\geqq r_1$, $r_3+r_1\geqq r_2$ can be written in the form

$$\psi = \sum_{n\leqq m} \frac{\alpha_{mn}(r_1+r_2)}{(r_1+r_2)^m} r_3^{m-n}(r_1-r_2)^n \tag{1.13}$$

so they are formal power series of the variables r_3 and r_1-r_2. In view of the above said, equation (1.1) has formal power series solutions of the form

$$\psi = \sum_{n\leqq m} \frac{\alpha^*_{mn}(r_2+r_3)}{(r_2+r_3)^m} r_1^{m-n}(r_2-r_3)^n \quad \text{and} \quad \psi = \sum_{n\leqq m} \frac{\alpha^{**}_{mn}}{(r_3+r_1)^m} r_2^{m-n}(r_3-r_1)^n \tag{1.14}$$

too, at least in the case (1.8).

2. We begin with two definitions.

DEFINITION 1. A function $u(s, q)$ is of class Π_n or Π_n^*, if it is of the form

$$\sum_{\nu=0}^{n} \alpha_\nu(s) q^\nu, \quad \text{or} \quad \sum_{\nu=0}^{n} \left[\alpha_\nu(s) q^\nu + \beta_\nu(s) q^\nu \log\frac{1+q}{1-q}\right],$$

respectively. Here the functions $\alpha_\nu(s)$, $\beta_\nu(s)$ are of class $C^\infty(s_1, s_2)$ and any of them may vanish identically.

LEMMA 1. *The class Π_n and Π_n^* coincides with the class of functions of the form*

$$\sum_{\nu=0}^{n} a_\nu(s) P_\nu(q) \quad \text{and} \quad \sum_{\nu=0}^{n} [a_\nu(s) P_\nu(q) + b_\nu(s) Q_\nu(q)],$$

respectively, where $a_\nu(s) \in C^\infty(s_1, s_2)$ $b_\nu(s) \in C^\infty(s_1, s_2)$.

This statement follows from q^ν being expressible as a linear consequence of the Legendre polynomials $P_0(q), P_1(q), \ldots, P_\nu(q)$ and from (1.12).

DEFINITION 2. A linear operator M is of the type $T(n, n')$ if it maps any function of class Π_n into a function of class $\Pi_{n'}$ and any function of class Π_n^* into a function of class $\Pi_{n'}^*$. If M maps any function of class Π_n^* into the identically 0 function, we shall write $M \in T(n, -1)$.

LEMMA 2. *If* $M \in T(n, n')$ *and* $N \in T(n', n'')$ *then* $NM \in T(n, n'')$.

LEMMA 3. *If* $M \in T(n, n')$ *and* $N \in T(n, n')$ *then* $M+N \in T(n, n')$.

Before giving examples of various operators of the above defined type we quote some known formulas [7, with a somewhat different notation]. Let R_n denote either $P_n(q)$ or $Q_n(q)$. Then one has

$$(1-q^2) R_n'' - 2qR_n' + n(n+1) R_n = 0 \quad (n = 0, 1, \ldots), \tag{2.1}$$

$$(q^2-1) R_n' = nqR_n - nR_{n-1} \quad (n = 1, 2, \ldots), \tag{2.2}$$

$$(2n+1) qR_n = (n+1) R_{n+1} + nR_{n-1} \quad (n = 1, 2, \ldots), \tag{2.3}$$

$$(q^2-1) P_0'(q) = 0, \quad (q^2-1) Q_0'(q) = -1, \tag{2.2'}$$

$$qP_0(q) = P_1(q), \quad qQ_0(q) = Q_1(q) + 1. \tag{2.3'}$$

3. In this section $\varkappa$ denotes any non-negative integer; the following examples will be needed later in section 5.

EXAMPLE 1. The operator L_k^0 defined by (1.7'), i.e. multiplication by L_k^0 is an operator of type $T(\varkappa, \varkappa+k-1)$, $k=0, 1, \ldots$. By Lemmas 2 and 3 it is sufficient to observe, that $q^\nu \in T(\varkappa, \varkappa+\nu)$. This, however, follows directly from Definition 1.

EXAMPLE 2. The operator

$$M_{1\varkappa} = (1-q^2) \frac{\partial^2}{\partial q^2} - 2q \frac{\partial}{\partial q} + \varkappa(\varkappa+1)$$

is of type $T(\varkappa, \varkappa-1)$. Indeed, using (2.1) and Lemma 1, one has

$$M_{1\varkappa} \sum_{\nu=0}^{\varkappa} [a_\nu(s) P_\nu(q) + b_\nu(s) Q_\nu(q)] = \sum_{\nu=0}^{\varkappa-1} [\varkappa(\varkappa+1) - \nu(\nu+1)][a_\nu(s) P_\nu(q) + b_\nu(s) Q_\nu(q)].$$

EXAMPLE 3. The operator

$$M_{2\varkappa} = (q^2-1) \frac{\partial}{\partial q} - \varkappa q$$

is of type $T(\varkappa, \varkappa)$. Indeed by (2.2) and (2.2′)

$$M_{2\varkappa} R_\nu = (\nu-\varkappa) q R_\nu - \nu R_{\nu-1} \quad (\nu = 1, 2, \ldots, \varkappa),$$

$$M_{2\varkappa} P_0(q) = -\varkappa q, \quad M_{2\varkappa} Q_0(q) = -1-\varkappa q Q_0(q),$$

thus, if $\delta(\varkappa, \nu)=1-\operatorname{sgn}(\varkappa-\nu)^2$ denotes Kronecker's delta one has

$$M_{2\varkappa} P_\nu(q) \in \Pi_{\nu+1-\delta(\varkappa,\nu)}, \; M_{2\varkappa} Q_\nu(q) \in \Pi^*_{\nu+1-\delta(\varkappa,\nu)} \quad (\nu = 0, 1, \ldots, \varkappa).$$

Finally by Lemma 1 one gets the statement.

EXAMPLE 4. The operator

$$M_{3\varkappa} = (1-q^2) q \frac{\partial^2}{\partial q^2} - (q^2+1) \frac{\partial}{\partial q} + \varkappa^2 q$$

is of type $T(\varkappa, \varkappa)$. Indeed, by (2.1)

$$M_{3\varkappa} R_\nu = (q^2-1) R'_\nu + [\varkappa^2 - \nu(\nu+1)] q R_\nu$$

and then by (2.2), (2.3), (2.2′), (2.3′) and Lemma 1 one has

$$M_{3\varkappa} P_\nu(q) \in \Pi_{\nu+1-\delta(\varkappa,\nu)}, \; M_{3\varkappa} Q_\nu(q) \in \Pi^*_{\nu+1-\delta(\varkappa,\nu)} \quad (\nu = 0, 1, \ldots, \varkappa).$$

EXAMPLE 5. The operator

$$M_{4\varkappa} = (1-q^2) q^2 \frac{\partial^2}{\partial q^2} - 2q \frac{\partial}{\partial q} + \varkappa^2 (q^2-1) - \varkappa (q^2+3)$$

is of type $T(\varkappa, \varkappa+1)$. By (2.1) and (2.2) one has for $\nu=1, 2, \ldots, \varkappa$

$$M_{4\varkappa} R_\nu = 2q(q^2-1) R'_\nu + [(\varkappa^2-\varkappa-\nu^2-\nu) q^2 - \varkappa^2 - 3\varkappa] R_\nu =$$

$$= [(\varkappa^2-\varkappa-\nu^2+\nu) q^2 - \varkappa^2 - 3\varkappa] R_\nu - 2\nu q R_{\nu-1},$$

hence by (2.3)

$$M_{4\varkappa} P_\nu(q) \in \Pi_{\nu+2-2\delta(\nu,\varkappa)}, \; M_{4\varkappa} Q_\nu(q) \in \Pi^*_{\nu+2-2\delta(\nu,\varkappa)} \quad (0 < \nu < \varkappa).$$

The two statements in the last row hold in the case $\nu=0$, too, thus by Lemma 1 $M_{4\varkappa} \in T(\varkappa, \varkappa+1)$ follows.

4. Let the functions $a^k_{ij} = a^k_{ij}(s, q)$ be of class C^∞ in the rectangle R: $s_1 < s < s_2$, $q_1 < q < q_2$, where the closed interval $[q_1, q_2]$ does not contain any of the points 1 and -1. Consider the formal differential equation

$$Lu = \sum_{k=0}^{\infty} p^k L_k u(s, p, q) = 0 \tag{4.1}$$

where $L_k = L_k(D_s, pD_p, D_q)$ stands for the formal polynomial of the operators $D_s = \partial/\partial s$, $pD_p = p\partial/\partial p$, $D_q = \partial/\partial q$

$$L_k = a^k_{11} D_s^2 + a^k_{12} D_s pD_p + a^k_{13} D_s D_q + a^k_{22} (pD_p)^2 + a^k_{23} pD_p D_q + a^k_{33} D_q^2 +$$
$$+ a^k_{10} D_s + a^k_{20} pD_p + a^k_{30} D_q + a^k_{00}.$$

Here $D_s^2=\partial^2/\partial s^2$, $D_sD_q=\partial^2/\partial s\partial q$, etc. We seek a formal solution of (4.1) in the form of the power series

$$u=\sum_{\varkappa=0}^{\infty}p^{\varkappa}u_{\varkappa}(s,q). \tag{4.2}$$

Denoting by $L_{k\varkappa}$ the operator

$$L_k(D_s,\varkappa,D_q)=a_{11}^kD_s^2+a_{12}^kD_s\varkappa+a_{22}^k\varkappa^2+a_{13}^kD_sD_q+a_{23}^k\varkappa D_q+a_{33}^kD_q^2+$$
$$+a_{10}^kD_s+a_{20}^k\varkappa+a_{30}^kD_q+a_{00}^k$$

one has

$$L_kp^{\varkappa}u_{\varkappa}(s,q)=p^{\varkappa}L_{k\varkappa}u_{\varkappa}(s,q) \tag{4.3}$$

and if a rearrangement of the double sum is permitted one gets from (4.1) that

$$Lu=\sum_k\sum_{\varkappa}p^kL_kp^{\varkappa}u_{\varkappa}=\sum_k\sum_{\varkappa}p^{k+\varkappa}L_{k\varkappa}u_{\varkappa}=$$

$$=\sum_{m=0}^{\infty}p^m\sum_{k=0}^{m}L_{k,m-k}u_{m-k}=0.$$

A formal power series solution of (4.1) in the rectangle R and of the form (4.2) is an infinite sequence $\{u(s,q)\}_{\varkappa=-\infty}^{\infty}$ of functions, for which

(i) $u_{\varkappa}=0$ if $\varkappa<0$,

(ii) $u_{\varkappa}(s,q)\in C^{\infty}(R)$ and

(iii)

$$L_{0m}u_m+L_{1m-1}u_{m-1}+\ldots+L_{m0}u_0=0 \quad (m=0,1,\ldots) \tag{4.4}$$

holds.

THEOREM 3. *Suppose that*
(i) $L_0=c[(pD_p)^2+pD_p+(1-q^2)D_q^2-2qD_q]$, $c\neq0$;
(ii) *the operators* $L_{k\varkappa}$ *are of type* $T(\varkappa,\varkappa+k-1)$.
Then each formal solution (4.2) *of* (4.1) *in the interval* (s_1,s_2) *is of the form*

$$u=\sum_{m=0}^{\infty}p^m\sum_{n=0}^{m}[a_{mn}(s)P_n(q)+b_{mn}(s)Q_n(q)].$$

Here $a_{mn}(s),b_{mn}(s)\in C^{\infty}(s_1,s_2)$.

PROOF. It is known that if

$$f(q)=\sum_{n=0}^{m-1}[a_nP_n(q)+b_nQ_n(q)]$$

then the most general solution of the inhomogeneous ordinary differential equation

$$(1-q^2)\frac{d^2y}{dq^2}-2q\frac{dy}{dq}+m(m+1)y+f(q)=0 \tag{4.5}$$

is

$$y=\sum_{n=0}^{m-1}\frac{a_nP_n(q)+b_nQ_n(q)}{n(n+1)-m(m+1)}+AP_m(q)+BQ_m(q), \tag{4.6}$$

where A and B are arbitrary constants. By assumption (i) of the theorem and by (4.2), (4.3) one has

$$L_{0m} = c[(1-q^2)D_q^2 - 2qDq + m(m+1)] \quad (c \neq 0).$$

On the other hand, from the recurrence formula (4.4) in the case $m=0$

$$[(1-q^2)D_q^2 - 2qD_q]u_0 = 0$$

hence by (4.6) $u_0 = a_{00}(s)P_0(q) + b_{00}(s)Q(q)$ and

$$u_m = \sum_{n=0}^{m} [a_{mn}(s)P_n(q) + b_{mn}(s)Q_n(q)] \tag{4.7}$$

holds in the case $m=0$.

Next we use an induction in m. We suppose that the functions $u_\varkappa$ ($\varkappa = 0, 1, \ldots, m-1$) are either of class $\Pi_\varkappa$ or of class $\Pi_\varkappa^*$. Then by assumption (ii) each term of the sum

$$f(s, q) = \sum_{k=1}^{m} L_{k,m-k} u_{m-k}$$

is either of class Π_{m-1} or of class Π_{m-1}^*, thus by (4.4) and (4.5) formula (4.7) holds for m.

REMARK 1. The construction shows that if $u_0, u_1, \ldots, u_{m-1}$ are already known, then the functions $a_{mm}(s)$ and $b_{mm}(s)$ can be arbitrarily chosen provided they are of class $C^\infty(s_1, s_2)$ and the functions $a_{mn}(s)$, $b_{mn}(s)$ $(n<m)$ are uniquely determined by $u_0, u_1, \ldots, u_{m-1}$, i.e. by the choice of $a_{00}(s), \ldots, a_{m-1,m-1}(s), b_{00}(s), \ldots$ $\ldots, b_{m-1,m-1}(s)$.

REMARK 2. By the foregoing construction, if

$$b_{mm}(s) = 0 \quad (m = 0, 1, 2, \ldots), \tag{4.8}$$

then $b_{mn}(s)=0$ for all m and n and each of the functions (4.7) is a polynomial in q the degree of which does not exceed m. If, however, (4.8) does not hold, then at least one of the functions (4.7) has a logarithmic singularity at $q=1$ and $q=-1$.

5. Theorems 1 and 2 will be proved, if one shows that when U satisfies (1.7) and (1.7′), then equation (1.1) fulfils both assumptions of Theorem 3. Though this is true, it is easier to verify that equation (1.1) when multiplied by $w=(1-p^2q^2)p^2s^2$ is of the form specified in Theorem 3. Starting from the well known formula

$$\Delta_1 v(r_1, r_2, r_3) = \frac{\partial^2 v}{\partial r_2^2} + \frac{-r_1^2+r_2^2+r_3^2}{r_2 r_3}\frac{\partial^2 v}{\partial r_2 \partial r_3} + \frac{\partial^2 v}{\partial r_3^2} + \frac{2}{r_2}\frac{\partial v}{\partial r_2} + \frac{2}{r_3}\frac{\partial v}{\partial r_3}$$

and two similar formulas for $\Delta_2 v$ and $\Delta_3 v$ a straightforward but tedious calculation shows that

$$w\Delta_i u(s, p, q) = \sum_{k=0}^{4} p^k L_k^i u(s, p, q) \tag{5.1}$$

where the operators $L_k^i = L_k^i(D_s, pD_p, D_q)$ are formal polynomials of the operators D_s, pD_p, D_q with coefficients depending on s and q:

$$L_0^1 = L_0^2 = (pD_p)^2 + pD_p + (1-q^2)D_q^2 - 2qD_q,$$

$$L_1^1 = -L_1^2 = 2q(pD_p)^2 + 2(1-q^2)qD_q^2 - 2spqD_sD_p +$$
$$+ 2s(q^2-1)D_sD_q - 2(q^2+1)D_q,$$

$$L_2^1 = L_2^2 = s^2D_s^2 + (q^2-1)(pD_p)^2 + (1-q^2)q^2D_q^2 - 2sq^2D_spD_p +$$
$$+ 2sq(q^2-1)D_sD_q + 4sD_s - (q^2+3)pD_p - 2qD_q,$$

$$L_3^1 = -L_3^2 = 2q[-(pD_p)^2 + sD_s \cdot pD_p + 2sD_s - 2pD_p],$$

$$L_4^1 = L_4^2 = q^2[-s^2D_s^2 - (pD_p)^2 + 2sD_spD_p - pD_p],$$

$$L_0^3 = L_1^3 = L_3^3 = 0,$$

$$L_2^3 = 4[s^2D_s^2 - 2sD_spD_p + (pD_p)^2 + 2sD_s - pD_p + (1-q^2)D_q^2 - 2qD_q],$$

$$L_4^3 = 4[-s^2D_s^2 + 2sD_spD_p - (pD_p)^2 - pD_p].$$

Hence equation (1.1) multiplied by $-w$ is of the form (4.1) with

$$L_k = \sum_{i=0}^{3} \mu_i L_k^i, \quad \mu_0 = -1,$$

cf. (1.7) and (1.7″). Note that L_0 meets assumption (i) of Theorem 3 with $c = \mu_1 + \mu_2$.

For calculating the recurrence formula (4.4) we introduee the quantities $L_{k\varkappa}^i$ defined by

$$L_{k\varkappa}^i u(s, q) = p^{-\varkappa} L_k^i p^{\varkappa} u(s, q) = p^{-\varkappa} L_k^i(D_s, pD_p, D_q) u(s, q) = L_k^i(D_s, \varkappa, D_q) u(s, q)$$

and so we have

$$L_{k\varkappa} = \sum_{i=0}^{3} \mu_i L_{k\varkappa}^i.$$

Again a straightforward calculation shows that

$$L_{k\varkappa}^i \in T(\varkappa, \varkappa+k-1) \quad (\varkappa = 0, 1, 2, \ldots), \tag{5.2}$$

thus by Lemma 3 $L_{k\varkappa} \in T(\varkappa, \varkappa+k-1)$ and assumption (ii) of Theorem 3 holds. We indicate only a part of the results of these calculations:

$$L_{0\varkappa}^1 = L_{0\varkappa}^2 = M_{1\varkappa},$$

$$L_{1\varkappa}^1 = -L_{1\varkappa}^2 = 2sD_sM_{2\varkappa} + 2M_{3\varkappa},$$

$$L_{2\varkappa}^1 = L_{2\varkappa}^2 = s^2D_s^2 + 4sD_s + 2sD_sqM_{2\varkappa} + M_{4\varkappa},$$

$$L_{2\varkappa}^3 = 4(s^2D_s^2 + 2s(1-\varkappa)D_s + M_{1\varkappa} - 2\varkappa],$$

$$L_{4\varkappa}^3 = 4[-s^2D_s^2 + 2\varkappa sD_s - \varkappa(\varkappa+1)],$$

with the notations of section 3. By sections 2 and 3 these formulas yield (5.2). Those operators $L_{k\varkappa}^i$ which are not given here explicitely, fulfil trivially (5.2).

REFERENCES

[1] FROIM, V. H.: Linear scalar partial differential equations with regular singularities on a hyperplane. Diff. Uravn. **9** (1973), 533—541.
[2] HEINE, E.: *Handbuch der Kugelfunctionen,* Reprint of the second edition, Würzburg, Physica 1961.
[3] KATO, T.: On the eigenfunctions of many-particle systems in Quantum Mechanics, *Comm. Pure Appl. Math.* **10** (1957), 151—177.
[4] KINOSHITA, T.: Ground state of the helium atom, *Phys. Rev.* **105** (1957), 1490—1502.
[5] MAKAI, E.: On the Schrödinger equation of the three-body problem, I, *Studia Sci. Math. Hung.*
[6] MUNSCHY, G.—PLUVINAGE, PH.: Approximation numérique des premiers états S de l'hélium, *J. de Phys.* **23** (1962), 184—192.
[7] WHITTAKER, G. T., WATSON, G. N.: *A course of modern analysis,* 4th ed. Cambridge 1935.

Mathematical Institute of the Hungarian Academy of Sciences
Budapest V, Reáltanoda u. 13—15, 1053.

(Received April 6, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977) 267—280

LÖSUNG GEWÖHNLICHER ANFANGSWERTAUFGABEN SINGULÄREN TYPS UND SINGULÄRER NICHTLINEARER GLEICHUNGSSYSTEME

von

S. FRIVALDSZKY

Summary. The two problems pointed out in the headline are apparently far from each other A special (non-polynomial) Hermite type interpolation is developed (Chapter 1) and forms the bases of a predictor-corrector method founded on a non-polynomial fitting. The latter method is suitable to the numerical evaluation of the initial problem of systems of ordinary differential equations with solution curve changing abruptly towards the end of the solution interval (Chapter 2). To such a problem reduces — by means of the continuation method (otherwise called Davidenko method) — also the numerical evaluation of those nonlinear systems of equations, for which the Jacobi matrix has a singular point near the wanted root.

1. Eine nichtpolynomiale Interpolation Hermiteschen Typs

1.1. In Äquidistanten Grundpunkten

$$x_k = x_0 + kh, \quad k = 0, 1, 2, \ldots, \quad h > 0$$

seien die Werte

$$Y_{n-j} = Y(x_{n-j}),\ Y'_{n-j} = Y'(x_{n-j}), \quad j = 0, 1, \ldots, m, \quad n \geqq m \geqq 1$$

einer Funktion $Y(x)$ und ihrer Ableitung $Y'(x)$ bekannt.

Nehmen wir an, daß für gewisse Werte $s > x_n$ und p das Produkt $Y(x)(s-x)^p$ im Intervall $[x_0, s]$ genügend vielmal differenzierbar ist, oder genauer ausgedrückt, daß

$$Y(x) = \frac{u(x)}{(s-x)^p} \quad (x < s) \tag{1.1}$$

ist mit

$$u(x) \in C_{r+1}[x_0, s], \quad u(s) \neq 0, \quad r \geqq 1,$$

$$p \neq 0, -1, \ldots, -r.$$

1.2. Wir beginnen mit der Abschätzung der Parameter der Formel unter (1.1) bzw. des Parameterpaares (c_n, p) in einem Grundpunkt x_n, wobei

$$c_n = \frac{h}{s - x_n}. \tag{1.2}$$

Zwischen den zu den verschiedenen Grundpunkten x_{n-j} gehörenden Parametern c_{n-j} bestehen im Sinne des oben gesagten die Beziehungen

$$c_{n-j} = \frac{c_n}{1 + c_n j}, \quad c_n = \frac{c_{n-j}}{1 - c_{n-j} j}, \quad j = \pm 1, \pm 2, \ldots. \tag{1.3}$$

Als Derivierte der Funktion unter (1.1) erhalten wir

$$Y'(x)=\frac{u'(x)(s-x)+u(x)p}{(s-x)^{p+1}}=\frac{v(x)}{(s-x)^{p+1}} \tag{1.4}$$

mit $v(x)\in C_r[x_0, s]$ und $v(s)\neq 0$.

Werden die linke bzw. rechte Seite der Gleichung unter (1.4) durch die entsprechende Seite der Gleichung unter (1.1) dividiert und dann die Stützpunkte x_{n-j} $(j=0, 1, \ldots, m)$ eingesetzt, so ergibt sich — unter Berücksichtigung der Definition unter (1.2) und auch der Folgen von (1.3) —

$$\frac{hY'_{n-j}}{Y_{n-j}}=\frac{hu'_{n-j}}{u_{n-j}}+\frac{pc_n}{1+c_nj}, \quad j=0, 1, \ldots, m. \tag{1.5}$$

Sind $(s-x_n)$ und h genügend klein, so haben wir $u(x)\neq 0$ im Intervall $[x_{n-m}, x_n]$, weshalb die Definition

$$\alpha(x)=\begin{cases}\log u(x) & (u(x)>0),\\ \log[-u(x)] & (u(x)<0)\end{cases}$$

sinnvoll ist; für diese Funktion gelten dann

$$\alpha'(x)=\frac{u'(x)}{u(x)}, \quad \alpha(x)\in C_{r+1}[x_{n-m}, x_n].$$

Jetzt werden die in den Grundpunkten x_{n-j} $(j=0, 1, \ldots, m)$ angenommenen Werte dieser Ableitung in die entsprechenden Gleichungen unter (1.5) eingesetzt und dann die folgende Linearkombination derselben gebildet:

(1.6)

$$\beta_n^{(m)}=\sum_{j=0}^{m}(-1)^j\binom{m}{j}\frac{hY'_{n-j}}{Y_{n-j}}=h\sum_{j=0}^{m}(-1)^j\binom{m}{j}\alpha'_{n-j}+p\sum_{j=0}^{m}(-1)^j\binom{m}{j}\frac{c_n}{1+c_nj}.$$

Die Größenordnung der auf der rechten Seite von (1.6) stehenden ersten Summe kann bei einer Funktion

$$\alpha(x)\in C_{r+1}[x_{n-m}, x_n] \quad \text{nur} \quad O(h^{\min(r,m)+1})$$

sein.

Um dies zu beweisen (was mit ganz einfachen Kunstgriffen gelingt) und die obige Summe genau abzuschätzen, benutzen wir das folgende Lemma:

(a)
$$\sum_{j=0}^{m}(-1)^j\binom{m}{j}\frac{1}{z+j}=\frac{m!}{\prod_{i=0}^{m}(z+i)} \quad (z>0),$$

(1.7) (b)
$$\sum_{j=1}^{m}(-1)^j\binom{m}{j}j^k\frac{1}{z+j}=\frac{(-1)^k z^k m!}{\prod_{i=0}^{m}(z+i)} \quad (1\leqq k\leqq m, z>0),$$

(c)
$$\sum_{j=1}^{m}(-1)^j\binom{m}{j}j^{m+1}\frac{1}{z+j}=(-1)^{m+1}m!\left(\frac{z^{m+1}}{\prod_{i=0}^{m}(z+i)}-1\right) \quad (z>0);$$

es kann durch die Partialbruchzerlegungen der rechten Seiten dieser Gleichungen leicht eingesehen werden.

Die erste Summe auf der rechten Seite unter (1.6) läßt im Falle $r>m$, auf Grund von (1.7) (c) eine gute Abschätzung zu, die durch den Grenzübergang $z=0$ erreicht wird. Diese Abschätzung ergibt annäherungsweise $h^{m+1}\alpha_n^{(m+1)}$. Auch im Falle $r\leqq m$ könnten wir für den obigen Ausdruck eine (allerdings praktisch unbrauchbare) obere Schranke angeben.

Deswegen werden wir jetzt die erste Summe der rechten Seite der Gleichung unter (1.6) vernachlässigen, die zweite aber mittels der Beziehung unter (1.7), (a) umformen, die wir an der Stelle $z=1/c_n>0$ anwenden. So erhalten wir die Annäherung

$$\beta_n^{(m)} = p\frac{m!}{\prod_{i=0}^{m}\left(\frac{1}{c_n}+i\right)} = p\frac{m!\,c_n^{m+1}}{\prod_{i=0}^{m}(1+c_n i)}. \tag{1.8}$$

Auch im Grundpunkt x_{n-1}, d. h. für die Linearkombination $\beta_{n-1}^{(m)}$, kann eine ähnliche Formel aufgeschrieben werden. Davon und von der Beziehung unter (1.8) ausgehend ergibt sich leicht das Parameterpaar (c_n, p), wenn auch noch die erste Beziehung unter (1.3) benutzt wird. Wir haben also

$$c_n = \frac{1}{m+1}\left[\frac{\beta_n^{(m)}}{\beta_{n-1}^{(m)}}-1\right], \quad p = \frac{\beta_n^{(m)}}{m!}\frac{\prod_{i=1}^{m}(1+c_n i)}{c_n^{m+1}}. \tag{1.9}$$

1.3. Nachdem das Parameterpaar (c_n, p) berechnet ist, kann die Interpolation bezüglich der Funktion $Y(x)$ auf die polynomiale Interpolation bezüglich $u(x)$ zurückgeführt werden. Das Verfahren besteht darin, daß man für die Funktion $u(x)$ eine geeignet gewählte polynomiale Interpolationsformel oder eine Linearkombination solcher Interpolationsformeln aufsetzt, wobei die Koeffizienten auch von den Parametern (c_n, p) abhängen können. Das aber kann so erzielt werden: aus der Beziehung

$$u(x) = Y(x)(s-x)^p \quad (x<s), \tag{1.10}$$

die wir durch Umformung der Beziehung unter (1.1) erhalten, ergeben sich durch Differenzierung Ausdrücke für einige der ersten Ableitungen der Funktion $u(x)$. Werden diese wieder in die interpolierende Beziehung für $u(x)$ eingesetzt, so entsteht eine Interpolationsformel für die Funktion $Y(x)$. Bei diesem Verfahren müssen auch die Beziehungen unter (1.3) benutzt werden, damit in der endlich erhaltenen Interpolationsformel nur ein, zu einem Punkt x_n gehörender Parameter c_n vorkomme.

1.4. Gehen wir z. B. von der Interpolationsformel

$$u_{n+1} = \sum_{i=0}^{L} a_i u_{n-i} + h\sum_{i=-1}^{L} d_i u'_{n-i} + t_n \tag{1.11}$$

aus, wobei die Koeffizienten a_i $(i=0, 1, \ldots, L)$ und d_i $(i=-1, 0, 1, \ldots, L)$ den $r+1$ linearen Gleichungen

$$\sum_{i=0}^{L} a_i = 1, \quad -\sum_{i=0}^{L} i a_i + \sum_{i=-1}^{L} d_i = 1, \quad \text{usw.} \tag{1.12}$$

$(r+1 \leqq 2L+3)$ genügen, mit deren Hilfe der Wert des Fehlergliedes t_n beliebig klein gemacht werden kann, und $d_{-1} \geqq 0$ ist (s. [1]). Das Fehlerglied t_n kann häufig in der Form

$$t_n = Ch^{r+1}u^{(r+1)}(\eta) \quad \text{mit} \quad x_{n-L} < \eta < x_{n+1} \tag{1.13}$$

angeben werden.

Durch Derivierung der Gleichung unter (1.10) erhalten wir

$$u'(x) = Y'(x)(s-x)^p - pY(x)(s-x)^{p-1}. \tag{1.14}$$

Aus den Beziehungen unter (1.10) und (1.14) ergeben sich die Funktionswerte unter (1.11) in allen Grundpunkten x_k, und es können sodann — mit Hilfe der ersten Beziehung unter (1.3) — sämtliche Parameter c_k durch den Parameter c_n ausgedrückt werden. Die so erhaltene Gleichung dividieren wir noch durch $h^p c_n^{-p}$ und ordnen sie dann in geeigneter Weise. Dadurch ergibt sich die Beziehung

$$Y_{n+1} = \sum_{i=0}^{L} A_i(c_n, p) Y_{n-i} + h \sum_{i=-1}^{L} D_i(c_n, p) Y'_{n-i} + T_n, \tag{1.15}$$

wobei

$$A_i(c_n, p) = \frac{a_i(1+c_n i) - d_i p c_n}{(1-c_n)+d_{-1}pc_n} \frac{(1+c_n i)^{p-1}}{(1-c_n)^{p-1}}, \tag{1.16}$$

$$D_i(c_n, p) = \frac{d_i(1+c_n i)}{(1-c_n)+d_{-1}pc_n} \frac{(1+c_n i)^{p-1}}{(1-c_n)^{p-1}},$$

$$T_n = t_n \frac{c_n^p}{h^p} \frac{1}{(1-c_n)^{p-1}} \frac{1}{(1-c_n)+d_{-1}pc_n}$$

gelten.

1.5. Weiter kann auch der Interpolationsausdruck für die Ableitungen höherer Ordnung der Funktion $Y(x)$ aufgeschrieben werden, indem man sich auf die Ableitungen erster Ordnung und evtl. auf die Funktionswerte stützt. Möchte man sich aber in der Interpolationsformel auf die Werte der Funktion $Y(x)$ nicht stützen, so besteht noch die Möglichkeit, die Ableitung von der aus (1.4) folgenden Gleichung

$$v(x) = Y'(x)(s-x)^{p+1}$$

— anstelle von (1.10) — ausgehend zu führen.

So erhalten wir z. B. die Interpolationsausdrücke

$$Y''_n = \frac{1}{2h}[Y'_{n+1}(1-c_n)^{p+1} - Y'_{n-1}(1+c_n)^{p+1}] + \frac{(p+1)c_n}{h} Y'_n + \frac{h^2}{6}\left(\frac{c_n}{h}\right)^{p+1} v'''(\eta_1) \tag{1.17}$$

mit

$$x_{n-1} < \eta_1 < x_{n+1}$$

und

$$Y'''_n = \frac{1}{h^2}\{Y'_{n+1}(1+c_n)^{p+1}[1+(p+1)c_n] + Y'_{n-1}(1+c_n)^{p+1} \cdot \tag{1.18}$$

$$\cdot [1-(p+1)c_n] - Y'_n[2-(p+1)(p+2)c_n^2]\} +$$

$$+\left(\frac{c_n}{h}\right)^{p+1}\left[-\frac{h^2}{12}v^{(IV)}(\eta_2) + \frac{(p+1)c_n}{3} h v'''(\eta_1)\right]$$

mit

$$x_{n-1} < \eta_1, \eta_2 < x_{n+1}$$

sowie den — sich auch auf die Funktionswerte von $Y(x)$ stützenden — Ausdruck

$$(1.19) \qquad Y_n'' = \frac{1}{h^2}\left\{Y_{n+1}(1-c_n)^p\left[2+\frac{pc_n}{2(1-c_n)}\right]+Y_{n-1}(1+c_n)^p\cdot\right.$$

$$\left.\cdot\left[2-\frac{pc_n}{2(1+c_n)}\right]-Y_n[4+p(p-1)c_n^2]\right\}-\frac{1}{2h}\cdot$$

$$\cdot[Y'_{n+1}(1-c_n)^p-Y'_{n-1}(1+c_n)^p-4pc_nY'_n]+\frac{h^4}{360}\left(\frac{c_n}{h}\right)^p u^{(\mathrm{VI})}(\eta_3)$$

mit

$$x_{n-1} < \eta_3 < x_{n+1},$$

von denen sich der erste an die Interpolationsformel

$$v'_n = \frac{v_{n+1}-v_{n-1}}{2h}-\frac{h^2}{6}v'''(\eta_1),$$

der zweite an die Linearkombination dieser und der Beziehung

$$v''_n = \frac{v_{n+1}-2v_n+v_{n-1}}{h^2}-\frac{h^2}{12}v^{(\mathrm{IV})}(\eta_2)$$

und endlich der dritte an die Beziehung

$$u''_n = \frac{2(u_{n+1}-2u_n+u_{n-1})}{h^2}-\frac{1}{2h}(u'_{n+1}-u'_{n-1})+\frac{h^4}{360}u^{(\mathrm{VI})}(\eta_3)$$

anlehnt, unter den Voraussetzungen $r\geqq 3$, $r\geqq 4$ bzw. $r\geqq 5$.

1.6. Schließlich drücken wir noch den Deriviertenwert

$$Y'_{n-\varepsilon} = Y'(x_n-\varepsilon h)$$

wobei $0<\varepsilon<1$ ist, mit den Werten Y'_n und Y'_{n-1} aus. Von der Beziehung

$$v_{n-\varepsilon} = (1-\varepsilon)v_n+\varepsilon v_{n-1}-\frac{\varepsilon(1-\varepsilon)}{2}h^2v''(\eta_4)$$

mit

$$x_{n-1} < \eta_4 < x_n$$

leitet man — ähnlich wie unter 1.5 — das Ergebnis

$$(1.20) \qquad Y'_{n-\varepsilon} = \frac{1}{(1+\varepsilon c_n)^{p+1}}[(1-\varepsilon)Y'_n+\varepsilon Y'_{n-1}\cdot(1+c_n)^{p+1}]-$$

$$-\left(\frac{c_n}{1+\varepsilon c_n}\right)^{p+1}\frac{\varepsilon(1-\varepsilon)h^{1-p}}{2}v''(\eta_4)$$

ab. Die Bedeutung des Index ist dabei dem obigen analog.

Für unsere bisherigen Ergebnisse werden wir im folgenden Verwendung finden.

2. Lösung der gewöhnlichen Anfangswertaufgabe von singulärem Typ

2.1. Nehmen wir an, daß sich einige Funktionskomponenten der Lösung der Anfangswertaufgabe

$$\mathbf{y}' = F(\mathbf{y}, x), \quad \mathbf{y}(x_0) = \mathbf{y}_0 \tag{2.1}$$

gegen das Ende des zu untersuchenden Intervalls stark ändern, was bei Anwendung der üblichen, im allgemeinen auf polynomialen Ansatz gegründeten Lösungsmethoden zu großen Formelfehlern führen kann. In solchen Fällen können wir auch Verfahren benutzen, die auf nichtpolynomialem Ansatz beruhen, oder wir können im Verlauf der Lösung auf solche Verfahren übergehen. Es existieren tatsächlich Methoden dieser Art (s. [2], [3], [4]); allerdings zieht ihre Anwendung praktische und evtl. auch theoretische Schwierigkeiten nach sich (s. [5]).

Wir verfolgen einen neuen Weg, indem wir die sich singulär verhaltenden Funktionskomponenten der Aufgabe unter (2.1) mit einer Summe der Form

$$y(x) = Y(x) + \sum_{i=0}^{l} B_i(\bar{x}-x)^i, \quad \bar{x} < s \tag{2.2}$$

annähern, wo die Funktion $Y(x)$ dieselbe ist wie die unter (1.1) und der festgesetzte Punkt $\bar{x}$ in der Nähe der Singularitätsabszisse s der vorliegenden Funktion liegt.

2.2. Bei jenen Ansatzfunktionen der Art unter (2.2), wo $p>0$ gilt, genügt es, den zum Wert $l=-1$ gehörenden Ausdruck zu benutzen, während man im Falle $p<-1$ an den Verfahren festhalten kann, die auf polynomialer Ansatzfunktion beruhen. Im Falle $-1<p<0$ lohnt sich es mit einem genaueren, zu den Werten $l=0, 1, 2, \ldots$ gehörenden Ausdruck zu versuchen.

Bei $l=-1$ läßt sich der Ansatz unter (2.2) einfach anwenden, wenn einige Funktions- und Ableitungswerte (in äquidistanten Grundpunkten) der sich singulär verhaltenden Funktion $y(x)$, aus denen die unter (1.6) definierten Ausdrücke $\beta_k^{(m)}$ berechnet werden können, schon bekannt sind. Auf diese Weise erhalten wir aus der Beziehung unter (1.9) eine Abschätzung der Parameter (c_n, p) im Punkt x_n (d. i. der Punkt größter Abszisse von den obigen Grundpunkten). Hiernach können wir die gesuchte Funktion mit einer, auf dem Ansatz unter (2.2) beruhenden Prediktor-Korrektor-Methode berechnen (s. (1.15) und (1.16)), die wir gewöhnlich nach dem bis zum Grundpunkt x_n angewendeten polynomialen Gegenstück wählen, und die auch eine obere Schranke für den Formelfehler liefert:

$$T_n^{\text{korr}} = C^{\text{korr}} \frac{Y_{n+1}^{\text{korr}} - Y_{n+1}^{\text{pr}}}{C^{\text{pr}}\left(1 + \frac{d_{-1}^{\text{korr}} c_n p}{1-c_n}\right) - C^{\text{korr}}},$$

wenn die Beziehung unter (1.13) besteht.

2.3. Es sei x_N der im Verlauf der vorigen Aufgabe vorgeschriebene letzte Berechnungsgrundpunkt, u. zw. soll er kleiner sein als die Singularitätsabszissen sämtlicher Funktionen. Zunächst führen wir die Berechnung bis zum Grundpunkt x_{N-1} nach der polynomialen Lösungsmethode (der Vektor $\mathbf{y}'_N$ läßt sich wohl nur sehr ungenau berechnen), und dann errechnen wir für jede sich singulär verhaltende Funktionskomponente die Parameter (c_{N-1}, p). Die Neuberechnung der letzten paar Vektoren

$\{\mathbf{y}_n\}$, $\{\mathbf{y}'_n\}$ erfolgt mit der auf dem polynomialen Ansatz beruhenden bzw. der dieser entsprechenden Prediktor-Korrektor-Methode unter (1.15) und (1.16), bei jeder Funktion mit den ihr entsprechenden Parametern (c_n, p), die mit Hilfe von (1.3) aus den Parametern (c_{N-1}, p) gewonnen werden. Das Verfahren wird iteriert, und der letzte Prediktor-Schritt, die Berechnung des Vektors $\mathbf{y}_N$ erfolgt erst, wenn in den vorangehenden Grundpunkten schon hinreichend genaue Funktions- und Ableitungswerte ermittelt worden sind.

2.4. Anstelle des obigen Verfahrens kann der Ansatz unter (2.2), bei $l=0$, mit verhältnismäßig wenig Mehrarbeit auf sich singulär verhaltende Funktionen angewendet werden. Dabei gewinnt man für jede Funktion und zumeist in jedem Schritt den Parameter B_0 durch Auflösung einer nichtlinearen Gleichung, die in einem Punkt x_n aus der Forderung entsteht, daß die Parameter c_n und c_{n-1} der Bedingung (1.3) genügen. Diese Bedingung kann mittels der Beziehung (1.9) derart ausgedrückt werden:

$$(2.3) \qquad (m+2)\frac{\beta_n^{(m)}}{\beta_{n-1}^{(m)}} - \frac{m\beta_{n-1}^{(m)} + \beta_n^{(m)}}{\beta_{n-2}^{(m)}} - 1 = 0,$$

worin durch Anwendung des Ausdrucks unter (2.2) (mit $l=0$) und der Ableitung derselben von der Funktion $Y(x)$ zur Funktion $y(x)$ übergegangen worden ist.

Eine ähnliche Substitution ist auch im Falle vorzunehmen, daß man bei $l \geqq 0$ die auf nichtpolynomialem Ansatz beruhende Prediktor-Korrektor-Formel unter (1.15) und (1.16) anwendet.

Wir lösen die Gleichung unter (2.3) im Grundpunkt x_{N-1}.

2.5. Von einem genaueren Ansatz dürfen nur dann auch genauere Funktionswerte $\{y_n\}$ erwartet werden, wenn zu Beginn der Lösung der Aufgabe unter (2.1) der Fehler, der sich im Verlauf des polynomialen Ansatzes angehäuft hat, genügend klein ist. Widrigenfalls kann sich der geerbte Fehler schon im Grundpunkt x_{N-1} in bedeutendem Maße anhäufen, und wenn wir auf noch so genaue Ansatzfunktionen übergehen.

Wenn der obenerwähnte, während des polynomialen Ansatzes angehäufte Fehler klein ist, so können wir es nach dem bisher beschriebenen Verfahren auch mit einer genaueren Ansatzfunktion versuchen, die sich aus der Beziehung unter (2.2), bei $l=1$ ergibt.

Zuerst bestimmen wir, mit Hilfe der momentanen Parameter (c_{N-1}, p, B_0) und des dazugehörenden Ansatzes unter (2.2), bei $l=0$, die Näherungswerte der Ableitungen zweiter Ordnung $y''_{N-2}, y''_{N-3}, \ldots$. Das geschieht mit Hilfe des Ausdrucks unter (1.17) oder des genaueren Ausdrucks unter (1.19), wobei nicht nur der Ansatz unter (2.2) (mit $l=0$) und die Ableitungen derselben benutzt werden, sondern auch die Beziehung unter (1.3). Dann setzen wir für den Parameter B_1 im Grundpunkt x_{N-2} eine Gleichung, ähnlich derjenigen unter (2.3), auf, die auf den an die Beziehungen unter (1.6) und (1.9) erinnernden Eigenschaften der Ableitungen höherer Ordnung der Funktion $Y(x)$ beruht. Sind nämlich die Schrittlänge h genügend klein und $1 \leqq k \leqq r-1$, so besteht die folgende Näherungsgleichung:

$$(2.4) \qquad \beta_n^{(k,m)} = \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{hY_{n-j}^{(k+1)}}{Y_{n-j}^{(k)}} = (p+k)\frac{m!\, c_n^{m+1}}{\prod_{i=0}^{m}(1+c_n i)}.$$

Die Behauptungen unter (1.9) und (2.3) können auch auf die Ausdrücke für die $\beta_n^{(k,m)}$ bezogen formuliert werden. Die letztere Behauptung kann auf eine Gleichung für den Parameter B_k führen, wenn man nämlich die dem Ansatz unter (2.2) entsprechende Substitution bei $l=k$ vornimmt und die nötigen Ableitungswerte $\{y_n^{(k+1)}\}$, $\{y_n^{(k)}\}$ schon kennt.

Wir kommen jetzt auf die Bestimmung des Parameters B_1 zurück. Da die Näherungswerte der zweiten Ableitungen $\{y_n''\}$ bekannt sind, ergibt das oben beschriebene Verfahren, im Grundpunkt x_{N-2} angewendet, einen Näherungswert des Parameters B_1. Die hiezugehörenden genaueren Parameter (c_{N-1}, p, B_0) erhalten wir dann in der üblichen Weise — aus der Gleichung unter (2.3) und dem Ausdruck unter (1.9) — bloß muß jetzt die Substitution unter (2.2) bei $l=1$ durchgeführt werden. Die so ermittelten Parameter ermöglichen eine genauere Bestimmung der zweiten Ableitungen $\{y_n''\}$, aus denen wieder ein genauerer Parameterwert B_1 errechnet werden kann, usw.

2.6. Endlich behandeln wir, als Spezialfall, das Singularitätsproblem bei der Anfangswertaufgabe der Differentialgleichungen höherer Ordnung.

Nehmen wir die Anfangswertaufgabe

$$y^{(M)} = F(x, y, y', \ldots, y^{(M-1)}),$$
$$y^{(k)}(x_0) = y_{0,k}, \quad k = 0, 1, \ldots, M-1,$$

die auf bekannte Weise zum Ausdruck unter (2.1) umgeformt werden kann. Es steht auch fest, daß die Funktion $y(x)$ und ihre Ableitungen sich nur zugleich singulär verhalten können, in dem Sinne, daß sie sich mit solchen Funktionen unter (2.2) gut annähern lassen, wo der Wert des Parameters p nicht eingeschränkt ist. Im Singularitätsfall werden die Ansatzfunktion (2.2) bei $l=M-1$ und ihre Ableitungen — die ebenfalls von der Gestalt (2.2) sind — eine Annäherung der Funktion $y(x)$ und ihrer Ableitungen abgeben, und demzufolge können die Koeffizienten B_k $(k=M-1, M-2, \ldots, 1, 0)$ aus den im Verlauf der Rechnung erhaltenen Ableitungen höherer Ordnung unmittelbar bestimmt werden. Die Prediktor-Korrektor-Methode wird in einem Punkt x_n so angewendet, daß die nötigen Werte in der Reihenfolge

$$y_{n+1}^{(M-1)\,\mathrm{pred}}, y_{n+1}^{(M-2)\,\mathrm{korr}}, y_{n+1}^{(M-3)\,\mathrm{korr}}, \ldots, y_{n+1}^{\prime\,\mathrm{korr}},$$
$$y_{n+1}^{\mathrm{korr}}, y_{n+1}^{(M)}, y_{n+1}^{(M-1)\,\mathrm{korr}}$$

berechnet werden, was ein sehr rasches Verfahren ergibt.

3. Lösung eines nichtlinearen Gleichungssystems mit der Fortsetzungs- (Davidenko-) Methode

3.1. Wir suchen eine der Lösungen eines nichtlinearen Gleichungssystems

$$F(\mathbf{y}) = \mathbf{0}, \tag{3.1}$$

ausgehend von einem gegebenen Punkt $\mathbf{y}_0$, der evtl. auch eine gute Annäherung der gesuchten Wurzel sein kann. Die beiden Punkte werden durch eine Vektorfunktion $\mathbf{y}(x)$ verbunden; eine solche ergibt sich etwa aus der Beziehung

$$G_x(\mathbf{y}) = F(\mathbf{y}) + (x-1)\,F(\mathbf{y}_0) = \mathbf{0}. \tag{3.2}$$

Im Falle gewisser, dem Gleichungssystem unter (3.1) aufgelegten Bedingungen (s. [6]) bestimmt das Gleichungssystem unter (3.2) eindeutig die obige Vektorfunktion, welche der Anfangswertaufgabe

$$\mathbf{y}'(x) = -F'(\mathbf{y}(x))^{-1} F(\mathbf{y}_0), \tag{3.3}$$

$$\mathbf{y}(0) = \mathbf{y}_0$$

entspricht. Hier kommt die Inverse der Jacobischen Matrix (Ableitungsmatrix) F' vor.

3.2. Ist die Jacobische Matrix in der Nähe der gesuchten Wurzel $\mathbf{y}(1)$ singulär, so konvergieren die numerischen Verfahren überhaupt nicht oder nur sehr langsam. Diese Schwierigkeit tritt bei der Aufgabe unter (3.3) in der Art auf, daß einige Komponenten der Vektorfunktion $\mathbf{y}(x)$ in der Nähe der Abszisse $x=1$ eine starke Änderung zeigen. Das gegebene Problem kann so auf die im Kapitel 2. dargelegte Aufgabe zurückgeführt werden, das Lösungsverfahren weicht aber von dem dort beschriebenen folgendermaßen ab:

3.3. a) Die Berechnung eines jeden Ableitungsvektors ist mit der Lösung eines linearen Gleichungssystems verbunden, dessen Koeffizientenmatrix (Jacobische Matrix) vorerst bestimmt werden muß. Wir nehmen an, daß die Elemente der Matrix analytisch, in einfacher Form dargestellt werden können. Bietet sich keine andere Möglichkeit, so lösen wir das lineare Gleichungssystem auf direktem Weg, durch Darstellung der Koeffizientenmatrix als Produkt einer unteren und einer oberen Dreiecksmatrix (d. i. die sog. $L \cdot U$-Gestalt). Diese Zerlegung ist eigentlich der die meiste Arbeit verlangende Teil der ursprünglichen Aufgabe unter (3.3) (s. [7]).

In gewissen Fällen können auch die Iterationsmethoden zur Lösung herangezogen werden, wodurch die Berechnungen oft eine bedeutende Beschleunigung erfahren. (S. [8].)

b) Die Lösungsgrundpunkte wählen wir so:

$$x_n = nh, \quad n = 0, 1, 2, \ldots N$$

mit $Nh=1$.

c) Ist eine Annäherung $\hat{\mathbf{y}}_n$ des Vektors $\mathbf{y}_n$ bekannt, so kann dieselbe beliebig genau gemacht werden. Wir dürfen daher annehmen, daß wir bei der Lösung der Aufgabe unter (3.3) genaue Funktions- und Ableitungswerte erhalten. Der Vektor $\mathbf{y}_n$ ist nämlich bei $x=x_n$ eine Wurzel der Gleichung (3.2), die vom Anfangswert $\hat{\mathbf{y}}_n$ ausgehend mit der Newtonschen Methode rasch gelöst werden kann, wenn nur der Grundpunkt x_n nicht zu nahe zu 1 ausfällt, d. h. die Schrittlänge h im Verhältnis zur Genauigkeit des obigen Anfangswertes nicht zu klein ist. Die Lösung der Gleichung unter (3.2) erhält man am schnellsten mittels der Variante

$$\hat{\mathbf{y}}_n^{(m+1)} = \hat{\mathbf{y}}_n^{(m)} - F'(\hat{\mathbf{y}}_n)^{-1}\left[F(\hat{\mathbf{y}}_n^{(m)}) + [x_n - 1]F(\mathbf{y}_0)\right], \quad m = 1, 2, \ldots, M-1, \tag{3.4}$$

$$\hat{\mathbf{y}}_n^{(1)} = \hat{\mathbf{y}}_n, \; \mathbf{y}_n = \hat{\mathbf{y}}_n^{(M)}$$

des Newtonschen Iterationsverfahrens, da wir bei der Bestimmung des Vektors $\hat{\mathbf{y}}_n$ (im Korrektorschritt) die Jacobische Matrix $F'(\hat{\mathbf{y}}_n)$ schon als Produkt einer unteren und einer oberen Dreiecksmatrix dargestellt und diese Zerlegung gespeichert haben. Das Lösen des bei der Iteration (3.4) auftretenden linearen Gleichungs-

systems reduziert sich daher auf das Lösen von zwei linearen Gleichungssystemen mit Dreiecksmatrizen.

Jetzt erhalten wir auch den genauen Ableitungsvektor $\mathbf{y}'_n$ nicht aus der Lösung des linearen Gleichungssystems unter (3.3), sondern mittels der Iteration

$$\hat{\mathbf{y}}_n'^{(m+1)} = \hat{\mathbf{y}}_n'^{(m)} - F'(\hat{\mathbf{y}}_n)^{-1}[F'(\mathbf{y}_n)\hat{\mathbf{y}}_n'^{(m)} + F(\mathbf{y}_0)], \quad m = 1, 2, \ldots, P-1, \tag{3.5}$$

$$\hat{\mathbf{y}}_n'^{(1)} = \hat{\mathbf{y}}'_n, \quad \mathbf{y}'_n = \hat{\mathbf{y}}_n'^{(P)},$$

wobei wir in Kenntnis der Matrizenzerlegung der Gestalt $L \cdot U$ nur lineare Gleichungssysteme mit Dreiecksmatrizen angeben und die Matrix $F'(\mathbf{y}_n)$ nur ein einziges Mal herzustellen brauchen. Die Iteration unter (3.4) als auch diejenige unter (3.5) konvergiert schnell, wenn nur der Vektor $\hat{\mathbf{y}}_n$ eine hinreichend genaue Annäherung darstellt.

d) Da die gesuchte Wurzel, der Vektor $\mathbf{y}(1)=\mathbf{y}_N$, verhältnismäßig genau zu bestimmen ist, genügt es oft nicht, bei den Ansätzen der Form (2.2) die Werte $l=-1$ oder $l=0$ zu wählen.

3.4. Wir führen die Berechnungen zur Lösung der Aufgabe unter (3.3) zuerst bis zum Grundpunkt x_{N-1}; wo es nötig ist, wenden wir den Ansatz unter (2.2) an, vorläufig bei $l=-1$ oder $l=0$. Erst dann bestimmen wir die obigen Vektoren genauer, in einigen letzten Grundpunkten $\{x_n\}$, nach 3.3. c).

Es ist nämlich folgendes von Bedeutung:

Wird irgendeine, von der approximativen Gleichung unter (1.8) abgeleitete Beziehung angewendet, so soll jeder dort auftretende Funktions- und Ableitungswert mit einem Formelfehler gleicher Art belastet sein, d. h. sie sollen nach derselben Ansatzfunktion berechnet werden. Wenn wir also diese Werte sofort nach ihrer approximativen Bestimmung genauer machen würden, so müßten wir im weiteren mit allen Funktions- und Ableitungswerten dasselbe tun. Sind wir im Verlauf der Rechnung dennoch zur Änderung der Ansatzfunktion gezwungen, so müssen wir für einige Schritte die wiederholte Berechnung der Parameter einstellen.

3.5. Im ersten Schritt wenden wir bei $l=1$ die obige Ansatzfunktion an. Wir setzen für jeden Parameter B_1 die durch das nachstehende Verfahren festgesetzte Gleichung auf.

Wir bestimmen für irgendeinen Wert B_1 auf die übliche Weise die dazugehörenden Parameter (c_{N-1}, p, B_0), deren Gesamtheit, bei $l=1$, eine Ansatzfunktion nach (2.2) festsetzt. Durch die letztere werden die Parameter der Prediktor-Korrektor-Formel, deren allgemeine Gestalt unter (1.15) und (1.16) zu finden ist, bestimmt. Der Parameter B_1 ist so zu wählen, daß der im Grundpunkt x_{N-2} aufgesetzte obige Korrektor als Ergebnis den genauen Wert von y_{N-1} ergebe, d. h.

$$y_{N-1}^{\text{korr}} = y_{N-1} \tag{3.6}$$

sei.

Es lohnt sich im allgemeinen nicht, bei einem auf gegebene Schrittlänge h aufgebauten System von Funktions- und Ableitungswerten eine genauere Ansatzfunktion (2.2) anzuwenden. Man benötigt dazu ein Wertsystem, das sich auf Grundpunkte $x_N - nh^*$ $(n=1, 2, \ldots)$ mit einer bedeutend kleineren Schrittlänge h^* stützt. Wir werden im folgenden eben ein solches Wertsystem bestimmen.

3.6. Zu diesem Zweck berechnen wir zuerst die Werte der Vektoren

$$\mathbf{y}_{n-\varepsilon} = \mathbf{y}(x_n - \varepsilon h), \quad n = N-1, N-2, \ldots, N-L-1 \tag{3.7}$$

und der dazugehörigen Ableitungsvektoren $\mathbf{y}'_{n-\varepsilon}$, wo ε eine festgesetzte Zahl mit $0<\varepsilon\ll 1$ und L die auf die Grundpunkte der bisher benutzten Prediktor-Korrektor-Formel bezogene Zahl (s. unter (1.15)) bedeuten.

Die Näherungswerte der Vektoren unter (3.7) berechnen wir mit Hilfe der Vektoren $\mathbf{y}_n$, $\mathbf{y}'_n$ komponentenweise, u. zw. so, daß wir im Grundpunkt x_n und mit der Schrittlänge εh, die Eulersche Prediktor-Methode, nach dem polynomialen bzw. dem momentanen Ansatz unter (2.2), aufschreiben. Die Näherungswerte der dazugehörenden Ableitungen bilden wir nach der Formel unter (1.20).

Die so erhaltenen Näherungswerte können wir auf übliche Weise, nach 3.3. c), genau machen, allerdings mit der Abänderung, daß wir in den Formeln unter (3.4) und (3.5) an Stelle der im Punkt $\hat{\mathbf{y}}_{n-\varepsilon}$ berechneten Inverse der Jacobischen Matrix die Matrix $F'(\hat{\mathbf{y}}_n)^{-1}$ setzen, weil ja die $L\cdot U$-Zerlegung der letzteren schon bekannt und ε eine kleine Zahl sind.

Wir berechnen den Näherungswert des Vektors $\mathbf{y}_{N-\varepsilon}$ mit Hilfe der Vektoren unter (3.7) und der dazugehörenden Ableitungen, u. zw. so, daß wir für jede Komponente den, auf dem entsprechenden Ansatz beruhenden, bisher angewendeten Prediktor im Punkte $x_{N-1-\varepsilon}=x_{N-1}-\varepsilon h$ aufschreiben. Als Ergebnis erwarten wir einen so guten Näherungswert, aus dem die genauen Vektoren $\mathbf{y}_{N-\varepsilon}$, $\mathbf{y}'_{n-\varepsilon}$, mit wenig Rechnungsarbeit, auf die übliche Weise gewonnen werden können.

Die neuen Grundpunkte bezeichnen wir, im Einklang mit den bisher benutzten Bezeichnungen, mit

$$x_{N-n\varepsilon^*} = x_N - nh\varepsilon^*, \quad n = 1, 2, \ldots, m+5, \tag{3.8}$$

wobei $(m+5)\varepsilon^*=\varepsilon$ ist, und die Zahl m ist durch den Ausdruck unter (1.8) bestimmt. Ähnlich bezeichnen wir auch die Funktions- und Ableitungsvektoren in den obigen Grundpunkten. Diese Werte berechnen wir für jede Komponente mit der nichtpolynomialen Eulerschen Prediktor-Korrektor-Methode, vom Grundpunkt $x_{N-\varepsilon}$ ausgehend und mit Anwendung der entsprechenden Ansatzfunktion.

Ähnlich wie bei dem bisher Beschriebenen können wir auch im Punkt $x_{N-\bar{\varepsilon}}=$ $=x_N-\bar{\varepsilon}h$ $(0<\bar{\varepsilon}\ll\varepsilon^*)$ die Näherungswerte der Funktions- und Ableitungsvektoren bestimmen (und zwar mit Hilfe des nichtpolynomialen Analogons der Eulerschen Prediktor-Korrektor-Methode) und dann genaumachen auf Grund von 3.3. c).

3.7. Sind diese Berechnungen durchgeführt, so lassen sich in den Grundpunkten $x_{N-n\varepsilon^*}$ $(n=2, 3, \ldots, m+4)$ die zweite und die dritte Ableitung der gegebenen Funktionskomponente mittels des Ausdrucks unter (1.17) oder (1.19) bzw. jenes unter (1.18), auf Grund des momentan benutzten Ansatzes, berechnen. Nach 2.5 kann man mit Hilfe der Beziehung unter (2.4) zwei Gleichungen aufschreiben, so wie wir es bei dem Aufsetzen der Gleichung unter (2.3) getan haben. Die Lösung der ersten Gleichung — mit $k=2$ — ergibt den Wert des Parameters B_2, während die zweite Gleichung — mit $k=1$ — den Wert des Parameters B_1 liefert. In beiden Gleichungen ersetzen wir die Funktion $Y(x)$ und ihre Ableitungen derart durch die Funktion $y(x)$ und deren Ableitungen, daß wir den Ansatz unter (2.2) bei $l=2$ anwenden.

Den Parameter B_0 berechnen wir aber nicht aus der Gleichung (2.3), sondern mittels eines Verfahrens. Der Parameter B_0 ist demnach so zu wählen, daß, nach

Bestimmung der ihm entsprechenden Parameter $(c_{N-\varepsilon^*}, p)$, der in dem Grundpunkt $x_{N-\varepsilon^*}$ mit der Schrittlänge $(\varepsilon^*-\bar{\varepsilon})h$ aufgeschriebene, auf dem zu den fünf obigen Parametern gehörenden Ansatz unter (2.2) beruhende Eulersche Korrektor als Ergebnis den genauen Wert

$$y_{N-\bar{\varepsilon}}^{\text{korr}} = y_{N-\bar{\varepsilon}} \tag{3.9}$$

ergibt.

3.8. Im Besitz eines hinreichend genauen Ansatzes läßt sich der Funktionswert y_N — die entsprechende Komponente der Wurzel von der Gleichung unter (3.1) — durch einen im Grundpunkt $x_{N-\varepsilon^*}$ aufgeschriebenen Prediktor-Schritt gut bestimmen, wenn nur $c_{N-\varepsilon^*} \leqq 1$ ausfällt. Ist das Letztere nicht der Fall — was nur als Folge der ungenügenden Genauigkeit der Rechnung geschehen kann —, so besteht die Möglichkeit, $y_N = y(s)$ als guten Näherungswert zu nehmen. Wir könnten im Sinne des obigen die bisherigen Berechnungsverfahren durch Hinzunahme der entsprechenden Bedingung $c_{N-1} \leqq 1$ oder $c_{N-\varepsilon^*} \leqq 1$ auch ergänzen.

4. Beispiele

4.1. Die im 2. Kapitel empfohlene Methode soll nun anhand der folgenden Anfangswertaufgabe dargestellt werden:

$$\begin{aligned} &x'(3ye^{xy-1}-2x)+y'(3xe^{xy-1}-2) = a_{14}, \\ &y'(4y-3ze^{yz-1})+z'(2-3ye^{yz-1}) = a_{24}, \\ &z'(3ze^{xz-1}-2)+x'(3xe^{xz-1}-2z) = a_{34}, \end{aligned} \tag{4.1}$$

wo auf der rechten Seite der Wert der Vektorfunktion (f_1, f_2, f_3) unter (4.2) für $x=y=z=0.85$ steht. Annäherungsweise haben wir

$$a_{14} = a_{34} \cong 0.149\,473\,305, \quad a_{24} \cong 0.128\,026\,695.$$

Die Lösung wird in den Grundpunkten $t_n = 0.025n$ $(n=1, 2, \ldots, 39)$ gesucht.

a) Die Berechnung führen wir — nach der geeignet gewählten Startphase — mit der folgenden Prediktor-Korrektor-Methode aus:

$$L = 2, \quad \mathbf{a} = (1, 0, 0), \quad \mathbf{d} = (0{,}23/12, -4/3, 5/12)$$

und

$$\mathbf{a} = (1, 0, 0), \quad \mathbf{d} = (5/12, 2/3, -1/12, 0).$$

Die Iteration haben wir so lange weitergeführt, bis der Fehler 10^{-7} nicht mehr überstieg.

b) Bei der zweitmaligen Ausführung der Berechnung haben wir in jedem Grundpunkt nur einen Korrektor-Schritt getan und dann auf die letzten 6 Grundpunkte gestützt die Parameter (c_{39}, p, B_0), $(m=3)$, der Ansatzfunktion unter (2.2), $l=0$, bestimmt.

Im Sinne von 2.4 haben wir die Durchrechnung der Aufgabe (4.1) in den letzten 6 Grundpunkten nach dem nichtpolynomialen Analogon der Prediktor-Korrektor-

Methode unter a) wiederholt. Um das Resultat genauer zu machen, wiederholten wir den unter b) beschriebenen Ansatz bzw. die Durchrechnung der Aufgabe (4.1). Jetzt zeigten der Prediktor- und der Korrektorwert kaum mehr eine Abweichung voneinander. (Den letzteren haben wir jeweils nur einmal berechnet.)

Das Resultat ist in der Tabelle 1 zusammengestellt.

	Genauer Wert	Der Korrektor nach 4.1.a)	Der Korrektor nach 4.1.b)
x (0.975)	1.019 369 9	1.019 108 4	1.019 332 6
y (0.975)	0.978 507 7	0.978 756 8	0.978 535 7
z (0.975)	0.978 587 0	0.978 846 6	0.978 624 4

Tabelle 1

Der Aufwand an Rechenoperationen ist bei den Berechnungen unter 4.1.a) und 4.1.b) ungefähr gleich groß. Die Fehler der Werte in der letzten Spalte entstehen fast gänzlich aus der Anhäufung der Fehler bis zum Grundpunkt t_{33}.

4.2. Das im 3. Kapitel beschriebene Verfahren soll durch die folgende Aufgabe veranschaulicht werden. Wir suchen eine Lösung des nichtlinearen Gleichungssystems

$$f_1 = x^2+2y-3e^{xy-1} = 0$$

(4.2) $$f_2 = 2y^2+2z-3e^{yz-1}-1 = 0$$

$$f_3 = z^2+2x-3e^{xz-1} = 0$$

vom Punkt $x=y=z=0.85$ ausgehend. Bei Anwendung der Newton—Raphson-Methode divergiert das Verfahren. Mit der Fortsetzungs-Methode beginnend gelangen wir zur Anfangswertaufgabe unter (4.1), wenn wir die Parametrisierung unter (3.2) anwenden. Führen wir in den gegebenen Grundpunkten die Berechnungen nach 4.1.a) bis zum Grundpunkt t_{39} durch (wobei jeder Korrektor nur einmal ausgerechnet wird), so gelangen wir in die Nähe der Lösung $x=y=z=1$, wo die Jacobische Matrix des Gleichungssystems singulär wird. Wir führen zuerst das Genauermachen nach 3.3.c) in den letzten 6 Grundpunkten, und dann die Abschätzung der gesuchten Wurzel mit Hilfe der Ansatzfunktion unter (2.2), $l=0$ (s. die 1. Zeile der Tabelle 2) mit der Wahl $m=4$ durch. Wir werden in der Tabelle nur die wiederholte Annäherung des Endwertes x_{40} angeben, da der momentane Ansatz unter (2.2) für die Endwerte y_{40} und z_{40} jeweils bessere Abschätzungen als für den Wert x_{40} ergibt. Im Laufe der Berechnungen wird die Bedingung $c_{39} \leqq 1$ nicht gestellt, weil sonst die Ergebnisse viel zu günstig ausfallen würden. So erhalten wir nur dort einen Näherungswert für x_{40}, wo diese Bedingung erfüllt ist.

Wir berechnen in den neuen Grundpunkten die Werte des Funktions- und des Deriviertenvektors bei $\varepsilon=1/10$ und $\varepsilon^*=\varepsilon/10$ (anstelle von $\varepsilon^*=\varepsilon/9$). Auf diese gestützt ergibt der Ansatz (2.2), $l=1$ die 2. Zeile der Tabelle 2, wenn der Parameter B_1 von

der Bedingung ausgehend berechnet wird, die im Grundpunkt $x_{N-\varepsilon^*}$ aufgeschrieben ist und jener unter (3.6) entspricht. Wird an Stelle der letzten Bedingung diejenige unter (3.9) gesetzt ($\bar{\varepsilon}=10^{-4}$), so ergibt die vorige Berechnung die 3. Zeile der Tabelle 2.

Endlich bestimmen wir, mit Hilfe der erhaltenen Parameter ($c_{40-\varepsilon^*}, p, B_0, B_1$) die Derivierten zweiter und dritter Ordnung. Nach der vollständigen Durchführung des im 3. Kapitel beschriebenen Verfahrens erhalten wir die 4. Zeile der Tabelle 2.

Laufende Nummer	Abschätzung für den Wert von x_{40}
1.	1.002 469 981
2.	1.000 002 346
3.	1.000 000 824
4.	1.000 000 093

Tabelle 2

Zum Abschluß möchte ich Herrn Prof. A. Békéssy für seine wertvollen Anregungen (vor mehreren Jahren), mich mit solchen Problemen zu befassen, und Frau Katalin Demetrowitsch für das sorgfältige Durchlesen des Manuskripts dieser Arbeit meinen herzlichen Dank aussprechen.

LITERATUR

[1] RALSTON, A.: *A first course in numerical analysis*, McGraw-Hill, Inc. 1965.
[2] LAMBERT, J. D., SHAW, B.: On the Numerical Solution of $y'=f(x, y)$ by a Class of Formulae Based on Rationale Approximation, *Mathematics of Computation* **19** (1965), 456—461.
[3] LAMBERT, J. D., SHAW, B.: A Method for the Numerical Solution of $y'=f(x, y)$ Based on a Self Adjusting Non-Polynomial Interpolant, *Mathematics of Computation* **20** (1966), 11—20.
[4] LAMBERT, J. D., SHAW, B.: Generalisation of Multistep Methods for Ordinary Differential Equations, *Numerische Mathematik* **8** (1966), 250—263.
[5] FRIVALDSZKY, S.: Ein Verfahren zur Berechnung der Lösung mit singulärem Verhalten bei Differentialgleichungen erster Ordnung, I—II—III, *Studia Sci. Math. Hung.* **10** (1975), 1—18.
[6] ORTEGA, J. M., RHEINBOLDT, W. C.: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York and London, 1970.
[7] WESTLAKE, J. R.: *A handbook of matrix inversion and solution of linear equations*, John Wiley & Sons, New York, 1968.
[8] VARGA, R. S.: *Matrix Iterative Analysis*, Englewood-Cliffs Prentice Hall, 1962.

Universitätsrechenzentrum, Budapest, Dimitrov tér 8, H—1093

(Eingegangen am 5. Juni, 1978)

L. B. Kovács

COMBINATORIAL METHODS OF DISCRETE PROGRAMMING

Discrete programming deals with optimization problems in which all or some of the variables take integer values.

Applicability: This field is a rapidly developing one because of its wide direct applicability and because of its close links with many other mathematical subjects.

Concise, comprehensive: This is a concise yet comprehensive textbook of the theory and practice of combinatorial methods of discrete programming with emphasis being placed on efficiency.

Coverage: After the presentation of many models one chapter is devoted to each of the basic methods, viz. implicit enumerations, branch and bound algorithms, dynamic programming. The following chapters consist of refinements such as Benders decomposition and a number of algebraic methods. Heuristic methods, specially structured problems, and complex algorithms are provided in further chapters. Finally, there is an annotated bibliography on the recent developments of discrete programming.

Descriptions, illustrative examples: All of the methods are described in theoretical frameworks but they can also be found in the form of step-by-step algorithms. Their illustration by means of numerical examples is an important asset of the volume.

In English — Approx. 180 pages — 17×25 cm — Cloth
ISBN 963 05 2004 4

AKADÉMIAI KIADÓ, Budapest
Publishing House of the Hungarian Academy of Sciences

Printed in Hungary

A kiadásért felel az Akadémiai Kiadó igazgatója — Műszaki szerkesztő: Botyánszky Pál
A kézirat a nyomdába érkezett: 1979. I. 12. — Terjedelem: 24,75 (A/5) ív, 31 ábra

79-215—Szegedi Nyomda — F.v.: Dobó József igazgató

INDEX

Die *Studia Scientiarum Mathematicarum Hungarica* ist eine Halbjahrsschrift der Ungarischen Akademie der Wissenschaften. Sie veröffentlicht Originalbeiträge aus dem Bereich der Mathematik in deutscher, englischer, französischer oder russischer Sprache. Es erscheint jährlich ein Band.

Adresse der Redaktion: 1053 Budapest V., Reáltanoda u. 13—15, Ungarn.
Technischer Redaktor: E. Deák

Bestellbar bei Buch- und Zeitungs-Aussenhandelsunternehmen *Kultúra* (Budapest 62, P. O. B. 149), oder bei den Vertretungen im Ausland.

Austauschabmachungen können mit der Bibliothek des Mathematischen Instituts (1053 Budapest V., Reáltanoda u. 13—15) getroffen werden.

Die zur Veröffentlichung bestimmten Manuskripte sind in zwei Exemplaren an die Redaktion zu schicken.

Studia Scientiarum Mathematicarum Hungarica est une revue biannuelle de l'Académie Hongroise des Sciences publiant des essais originaux, en francais, anglais, allemand ou russe, du domaine des mathématiques.

Rédaction: 1053 Budapest V., Reáltanoda u. 13—15, Hongrie.
Rédacteur technique: E. Deák

On s'abonne chez *Kultúra*, Société pour le Commerce de Livres et Journaux (Budapest 62, P. O. B. 149) ou chez ses représentants à l'étranger.

Pour établir des relations d'échange on est prié de s'adresser à la Bibliothèque de l'Institut de Mathématique (1053 Budapest V., Reáltanoda u. 13—15).

On est prié d'envoyer les articles destinés à la publication en deux exemplaires à l'adresse de la Rédaction

---

Studia Scientiarum Mathematicarum Hungarica — выходит два раза в год в Издании Академии Наук Венгрии. Журнал публикует оригинальные исследования в области математики на русском, немецком, английском, и французском языках. Отдельные выпуски составляют ежегодно один том.

Адрес редакции: 1053 Budapest V., Reáltanoda u. 13—15, Венгрия.
Технический редактор: E. Deák.

Подписка на журнал принимается Внешнеторговым предприятием „Култура" (Budapest 62, P. O. B. 149) или его представительствами за границей.

По поводу отношения обмена просим обрашаться к Библиотеке Института Математики (1053 Budapest V., Reáltanoda u. 13—15).

Работы, преднасначенные для опубликования в журнале следует направлять по адресу редакции в двух экземплярах.

All the reviews of the Hungarian Academy of Sciences may be obtained among others from the following bookshops:

ALBANIA

Ndermarja Shtetnore e Botimeve
Tirana

AUSTRALIA

A. Keesing
Box 4886, GPO
Sidney

AUSTRIA

Globus Buchvertrieb
Salzgries 16
Wien I.

BELGIUM

Office International de Librairie
30, Avenue Marnix
Bruxelles 5
Du Monde Entier
5, Place St. Jean
Bruxelles

BULGARIA

Raznoiznos
1 Tzar Assen
Sofia

CANADA

Pannonia Books
2 Spadina Road
Toronto 4, Ont.

CHINA

Waiwen Shudian
Peking
P.O.B. Nr. 88.

CZECHOSLOVAKIA

Artia A. G.
Ve Smeckách 30
Praha II.
Postova Novinova Sluzba
Dovoz tisku
Vinohradska 46
Praha 2
Postova Novinova Sluzba
Dovoz tlace
Leningradska 14
Bratislava

DENMARK

Ejnar Munksgaard
Nörregade 6
Kopenhagen

FINLAND

Akateeminen Kirjakauppa
Keskuskatu 2
Helsinki

FRANCE

Office International de Documentation et Libraire
48, rue Gay Lussac
Paris 5

GERMAN DEMOCRATIC REPUBLIC

Deutscher Buchexport und Import
Leninstraße 16.
Leipzig C. I.
Zeitungvertriebsamt
Clara Zetkin Straße 62.
Berlin N. W.

GERMAN FEDERAL REPUBLIC

Kunst und Wissen
Eich Bieber
Postfach 46.
7 Stuttgart S.

GREAT BRITAIN

Collet's Subscription Dept.
44—45 Museum Street
London W. C. I.
Robert Maxwell and Co. Ltd.
Waynflete Bldg. The Plain
Oxford

HOLLAND

Swetz and Zeitlinger
Keizersgracht 471—487
Amsterdam C.
Martinus Nijhof
Lange Voorhout 9
The Hague

INDIA

Current Technical Literature
Co. Private Ltd.
Head Office:
India House OPP.
GPO Post Box 1374
Bombay I.

ITALY

Santo Vanasia
71 Via M. Macchi
Milano
Libreria Commissionaria Sanson
Via La Marmora 45
Firenze

JAPAN

Nauka Ltd.
2 Kanada-Zimbocho 2-ehome
Chiyoda-ku
Tokyo
Maruzen and Co. Ltd.
P.O. Box 605
Tokyo

Far Eastern Booksellers
Kanada P.O. Box 72
Tokyo

KOREA

Chulpanmul
Korejskoje Obschestvo po Exportu Importu Proizvedenij Pechati
Phenjan

NORWAY

Johan Grund Tanum
Karl Johansgatan 43
Oslo

POLAND

Export- und Import-Unternehmen RUCH
ul. Wilcza 46.
Warszawa

ROUMANIA

Cartimex
Str. Aristide Briand 14—18.
Bucuresti

SOVIET UNION

Mezhdunarodnaja Kniga
Moscow
G—200

SWEDEN

Almquist and Wiksell
Gamla Brogatan 26
Stockholm

USA

Stechert Hafner Inc.
31 East 10th Street
New York 3 N. Y.
Walter J. Johnson
111 Fifth Avenue
New York 3 N. Y.

VIETNAM

Xunhasaba
Service d'Export et d'Import des Livres et Périodiques
19. Tran Quoc Toan
Hanoi

YUGOSLAVIA

Forum
Vojvode Misiva broj 1.
Novi Sad
Jugoslovenska Kniga
Terazije 27.
Beograd

20. V. 1980.

Index: 26 748

HU ISSN 0081—6906

Scientiarum Mathematicarum Hungarica

AUXILIO
CONSILII INSTITUTI MATHEMATICI
ACADEMIAE SCIENTIARUM HUNGARICAE

REDIGIT
L. FEJES TÓTH

ADIUVANTIBUS
Á. CSÁSZÁR, I. CSISZÁR, A. HAJNAL,
P. RÉVÉSZ, O. STEINFELD, T. E. SCHMIDT,
J. SZABADOS, D. SZÁSZ, I. VINCZE

MUS XII.
SC. 3—4.
7

AKADÉMIAI KIADÓ, BUDAPEST

Studia Scientiarum Mathematicarum Hungarica

A Magyar Tudományos Akadémia matematikai folyóirata

Szerkesztőség: 1053 Budapest V., Reáltanoda u. 13—15.

Technikai szerkesztő: Deák E.

Kiadja az Akadémiai Kiadó, 1054 Budapest V., Alkotmány u. 21.

A *Studia Scientiarum Mathematicarum Hungarica* angol, német, francia vagy orosz nyelven közöl eredeti értekezéseket a matematika tárgyköréből. Félévenként jelenik meg, évi egy kötetben.

Előfizetési ára belföldre 120,— Ft, külföldre 165,— Ft. Megrendelhető a belföld számára az Akadémiai Kiadónál, a külföld számára pedig a Kultúra Könyv és Hírlap Külkereskedelmi Vállalatnál (1011 Budapest I., Fő u. 32.).

Cserekapcsolatok felvétele ügyében kérjük az MTA Matematikai Kutató Intézete Könyvtárához (1053 Budapest V., Reáltanoda u. 13—15.) fordulni.

Közlésre szánt dolgozatokat kérjük két példányban a szerkesztőség címére küldeni.

Studia Scientiarum Mathematicarum Hungarica is a journal of the Hungarian Academy of Sciences publishing original papers on mathematics, in English, German, French or Russian.

It is published semiannually, making up one volume per year.

Editorial Office: 1053 Budapest V., Reáltanoda u. 13—15, Hungary.
Technical Editor: E. Deák

Orders may be placed with *Kultura* Trading Co. for Books and Newspapers, Budapest 62, P.O.B. 149 or with its representatives abroad.

For establishing exchange relations pelase write to the Library of the Mathematical Institute (1053 Budapest V., Reáltanoda u. 13—15.)

Papers intended for publication should be sent to the Editor 2 copies.

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 281—293.

SCHEIBENPACKUNGEN MIT NACH UNTEN BESCHRÄNKTER NACHBARNZAHL

von

J. LINHART

Herrn Prof. L. Fejes Tóth zum 65. Geburtstag gewidmet.

Einleitung

Unter einer *Scheibe* verstehen wir eine offene, beschränkte und konvexe Teilmenge der euklidischen Ebene. Unter der *Newtonschen Zahl* N einer Scheibe S verstehen wir die maximale Anzahl von disjunkten, zu S kongruenten Scheiben, welche sich mit S in Berührung bringen lassen. Eine Packung von zu S kongruenten Scheiben heißt *Maximalpackung,* wenn jede Scheibe genau N andere berührt.

Von L. FEJES TÓTH stammt die Vermutung, daß es Maximalpackungen nur für $N \leqq 21$ gibt und 21 die genaue obere Schranke ist [2, p. 201]. P. GÁCS [3] gelang es zu zeigen, daß für N jedenfalls eine obere Schranke existiert. Seine Methode lieferte jedoch eine Schranke von der Größenordnung 10^8.

In der vorliegenden Arbeit wird nun bewiesen, daß tatsächlich $N \leqq 21$ gilt. Ob 21 wirklich bestmöglich ist, bleibt noch offen. Die verwendete graphentheoretische Methode liefert außerdem interessante Abschätzungen für die Nachbarnzahl in allgemeineren Packungen, welche das Resultat eines früheren Aufsatzes [4] verallgemeinern und verbessern.

I. Endliche (n, g)-Packungen

Für die Zwecke dieses Kapitels scheint es natürlicher zu sein, auch Packungen auf der Kugelfläche in die Betrachtungen mit einzubeziehen. Wir setzen dabei zusätzlich voraus, daß der Durchmesser einer Scheibe auf der Kugel $< \pi$ sei.

DEFINITION. Seien n und g natürliche Zahlen, und $g \geqq 3$. Unter einer (n, g)-*Packung* verstehen wir eine Menge paarweise disjunkter Scheiben (in der Ebene bzw. auf der Kugel), in der jede Scheibe mindestens n andere berührt, und höchstens g Scheiben einen Randpunkt gemeinsam haben.

SATZ 1. *Für endliche* (n, g)*-Packungen gilt:*

$$n < 2g \quad \textit{für} \quad g \geqq 3 \tag{1}$$

und

$$n < \frac{g^2-3g+18}{g-6} \quad \textit{für} \quad g \geqq 7. \tag{2}$$

BEMERKUNGEN. Die Ungleichung (1) wurde in [4] nur für den Fall bewiesen, daß jede Scheibe genau n andere berührt (sogenannte *n-Nachbarnpackungen*). Es wurde außerdem gezeigt, daß sie für $g \leqq 6$ genau ist. Die Ungleichung (2) ist für

$g \geqq 12$ besser als (1). Beispielsweise ergibt sich

(3) $$n < 2g-3 = 21 \quad \text{für} \quad g = 12$$

und

(4) $$n < g+4 \quad \text{für} \quad g \geqq 42.$$

Was die Genauigkeit dieser Ungleichungen betrifft, seien außer den in [4] genannten Beispielen noch die folgenden erwähnt:

1) Das sphärische Netz des Sternpolyeders $\left\{5, \frac{5}{2}\right\}$ (siehe [1]) ist eine (17, 10)-Packung, d. h. $n=2g-3$. Hier ist also (1) noch verhältnismäßig gut.

2) Das Sternpolyeder $\left\{3, \frac{5}{2}\right\}$ liefert eine (19, 12)-Packung (vgl. (3)).

3) Zeichnet man auf der Kugel einen Großkreis und verbindet die zugehörigen Pole durch g verschiedene Halbkreise, so erhält man für beliebiges $g \geqq 4$ eine $(g+2, g)$-Packung. Es wäre interessant, zu wissen, ob es für große g auch $(g+3, g)$-Packungen gibt.

Es sei bemerkt, daß die Beispiele 1) und 3) sogar n-Nachbarnpackungen sind.

Beweis des Satzes. Wir betrachten eine endliche (n, g)-Packung und ordnen ihr folgendermaßen einen ebenen paaren Graphen G zu: Aus jeder Scheibe wählen wir einen (inneren) Punkt aus. Die Menge dieser Punkte bezeichnen wir mit A, ihre Elemente heißen A-Ecken. Wir definieren weiters eine Menge B so: Wenn q Scheiben ($q \geqq 3$) einen Randpunkt y gemeinsam haben, dann sei $y \in B$ (so ein y nennen wir auch *Berührungsecke*). Wenn zwei Scheiben einander berühren, aber nicht in einer Berührungsecke, dann wählen wir einen gemeinsamen Randpunkt aus und nehmen ihn zur Menge B hinzu. Die Elemente von B heißen B-Ecken. Jede A-Ecke wird nun geradlinig mit allen B-Ecken, die auf dem Rand der zugehörigen Scheibe liegen, verbunden. Auf diese Weise erhalten wir einen schlichten, ebenen, paaren Graphen G mit folgenden Eigenschaften:

a) Jeder Scheibe entspricht umkehrbar eindeutig eine A-Ecke, welche in dieser Scheibe liegt.

b) Jede A-Ecke ist mit mindestens n anderen A-Ecken durch je zwei Kanten verbunden.

c) Der Grad der B-Ecken ist $\leqq g$.

Wir können o. B. d. A. annehmen, daß G zusammenhängend ist (andernfalls betrachten wir statt G die einzelnen Zusammenhangskomponenten).

Unter einem *Vierkreis* von G verstehen wir einen Kreis (d. h. einfach geschlossenen Kantenzug) von G, welcher aus vier Ecken und vier Kanten besteht. In der Ebene sind diesbezüglich die Begriffe „Inneres" und „Äußeres" sinnvoll. Auf der Kugel wollen wir unter dem Inneren eines Vierkreises dasjenige der beiden in Frage kommenden Gebiete verstehen, in dem die (kürzeste) Verbindungsstrecke der beiden B-Ecken des Vierkreises liegt. Eine Fläche von G, deren Rand ein Vierkreis ist, nennen wir *Viereck*.

Lemma 1. *Im Inneren eines Vierkreises von G liegen keine Ecken und Kanten von G (daher ist jeder Vierkreis Rand eines Vierecks).*

BEWEIS. Seien x_1, x_2 die beiden A-Ecken des Vierkreises K, S_1, S_2 die entsprechenden Scheiben, und y_1, y_2 die beiden B-Ecken. Wir zeigen zunächst, daß das ganze Innere von K zu $\overline{S_1} \cup \overline{S_2}$ gehört ($\overline{S}$ bezeichnet die abgeschlossene Hülle von S). Sei z ein Punkt im Inneren von K. Dann ist entweder (i) z in dem Dreieck $x_1y_1y_2$ oder (ii) in dem Dreieck $x_2y_1y_2$. (Auf der Kugel sind die Seiten dieser Dreiecke laut Voraussetzung alle $<\pi$.) Im Falle (i) betrachten wir die Gerade (bzw. den Großkreis) durch x_1 und z. Sie schneidet den Rand des Dreiecks auf der Seite $\overline{y_1y_2}$. Der Schnittpunkt heiße w, also $z \in \overline{x_1w}$. Da $y_1, y_2 \in \overline{S_1}$, ist wegen der Konvexität von S_1 auch $w \in \overline{S_1}$ und daher auch $z \in \overline{S_1}$. Im Falle (ii) ergibt sich analog $z \in \overline{S_2}$, also jedenfalls $z \in \overline{S_1} \cup \overline{S_2}$. Daher liegt keine A-Ecke und keine Berührungsecke innerhalb von K. Es kann aber auch keine B-Ecke vom Grad 2 im Inneren von K liegen, denn nach Definition der B-Ecken dürften sich dann S_1 und S_2 in sonst keiner B-Ecke berühren, also auch nicht in y_1 und y_2.

LEMMA 2. *Man kann G so modifizieren, daß alle Eckengrade $\geqq 3$ sind und dennoch die Eigenschaften a), b), c) erhalten bleiben, und auch der modifizierte Graph schlicht, eben, paar und zusammenhängend ist.*

BEWEIS. Wenn eine A-Ecke Grad 1 hat, kann die entsprechende Scheibe höchstens $g-1$ Nachbarn haben, und die Ungleichungen (1) und (2) sind trivialerweise erfüllt. Diesen Fall können wir daher ausschließen. B-Ecken vom Grad 1 sind per def. unmöglich. Eine A-Ecke x vom Grad 2 kann nach Lemma 1 nicht auf dem Rand zweier Vierecke liegen. Es ist daher stets möglich, in einer der beiden Flächen (bzw. in der einen Fläche), auf deren Rand x liegt, eine zusätzliche B-Ecke einzufügen und sie mit x und zwei anderen A-Ecken am Rand dieser Fläche zu verbinden. Wenn bei einer solchen Einfügung eine 2-gradige B-Ecke auf dem Rand eines Vierecks zu liegen kommt, lassen wir sie mit den beiden zugehörigen Kanten weg. Die von x jetzt ausgehenden drei Kanten werden davon nicht betroffen, denn sonst hätte die x entsprechende Scheibe x höchstens $1+(g-1)$ Nachbarn gehabt, was wir ausschließen können. (Beim ursprünglichen Graphen liegt keine zweigradige B-Ecke auf dem Rand eines Vierecks, wie sich aus ihrer Definition ergibt.)

Auf diese Weise kann man erreichen, daß alle A-Ecken einen Grad $\geqq 3$ haben und die genannten Eigenschaften von G erhalten bleiben. Es ist dabei nur zu überlegen, daß es nie vorkommen kann, daß eine zwei-gradige A-Ecke an zwei Vierecken liegt. Beim ursprünglichen Graphen kann dies, wie gesagt, wegen Lemma 1 nicht der Fall sein; also müßte es im Laufe der Modifikation eintreten. Die Kanten zweier solcher Vierecke können nicht bei der Modifikation entstanden sein, denn sonst wäre eine der beiden B-Ecken samt den drei Kanten neu hinzugekommen und daher die betrachtete A-Ecke vorher vom Grad 1 gewesen, was wir ausgeschlossen haben. (Auch während der Modifikation kann eine A-Ecke niemals Grad 1 erhalten, denn die Anzahl der mit ihr über je zwei Kanten verbundenen A-Ecken wird bei der Modifikation nicht kleiner, also niemals $\leqq g-1$.) Also gehören die beiden Vierecke zum ursprünglichen Graphen, im Widerspruch zu Lemma 1.

Wenn alle A-Ecken Grad $\geqq 3$ haben, verbinden wir jede verbleibende B-Ecke y vom Grad 2 mit einer noch nicht mit y verbundenen A-Ecke am Rand derselben Fläche. Das ist stets möglich, da y nicht am Rand von zwei Vierecken liegt. Wenn dabei eine andere zwei-gradige B-Ecke an den Rand eines Vierecks gerät, wird sie wieder sofort mit ihren beiden Kanten weggelassen. Wenn dadurch eine A-Ecke

Grad 2 bekommt, wird zunächst wieder, wie vorhin beschrieben, eine B-Ecke vom Grad 3 eingefügt, sodaß von dieser A-Ecke wieder drei Kanten ausgehen. Erst dann wird mit der Behandlung der zwei-gradigen B-Ecken fortgefahren. Bei dieser Vorgangsweise kann es wieder nicht vorkommen, daß eine zwei-gradige A-Ecke am Rand von zwei Vierecken liegt, denn sonst hätte sich die weggelassene B-Ecke innerhalb eines dieser beiden Vierecke befunden, d. h. sie wäre mit ihren beiden Kanten am Rand von zwei Vierecken gelegen. Das kann nie eintreten, da eine zwei-gradige B-Ecke schon weggelassen wird, wenn sie am Rand von *einem* Viereck liegt, und niemals zwei solche Vierecke auf einmal entstehen können.

Auf diese Weise kann man erreichen, daß alle Eckengrade $\geqq 3$ werden und der Graph trotzdem die gewünschten Eigenschaften behält.

Der gemäß Lemma 2 modifizierte Graph ist i. a. nicht mehr geradlinig. Es gibt jedoch einen dazu isomorphen geradlinigen Graphen mit genau entsprechenden Flächen [5], den wir der Einfachheit halber von nun an betrachten wollen.

Ein Tripel (z_1, z, z_2) von Ecken einer Fläche p heißt (orientierter) *Winkel* von p in der Ecke z, wenn $\overline{z_1 z}$ und $\overline{z_2 z}$ Kanten auf dem Rand von p sind und es ein $\varepsilon>0$ gibt, so daß in einem (geometrischen) Kreis $K_\varepsilon(z)$ mit Radius ε um z die Strecke $\overline{z_1 z} \cap K_\varepsilon(z)$ in die Strecke $\overline{z_2 z} \cap K_\varepsilon(z)$ um z in positivem Sinn so gedreht werden kann, daß die Drehung ganz innerhalb von p verläuft. Den bei der Drehung überstrichenen Bereich nennen wir den zugehörigen Winkelbereich. Wenn es in einer Ecke von p mehrere (etwa m) Winkel von p gibt, so sprechen wir von einer mehrfachen (m-fachen) Ecke, sonst von einer einfachen Ecke von p. (Für mehrfache Ecken scheint die Winkeldefinition von ORE [5] nicht verwendbar zu sein.)

LEMMA 3. *Sei p eine Fläche mit s Ecken, der Vielfachheit nach gezählt (wir sagen dann einfach s-Eck). Dann können diese Ecken derart in eine Folge $z_1, \ldots, z_s$ angeordnet werden, daß jede Ecke ihrer Vielfachheit entsprechend viele Indizes erhält, und je drei aufeinanderfolgende Ecken einen Winkel von p bilden (dabei heißen z. B. auch z_s, z_1, z_2 aufeinanderfolgend, d. h. die Indizes sind modulo s zu verstehen).*

BEWEIS. Es wird im wesentlichen behauptet, daß die Ecken von p im Uhrzeigersinn numeriert werden können. Dies folgt z. B. daraus, daß man das (einfach zusammenhängende) Gebiet p konform auf das Innere eines Kreises abbilden kann. Diese Abbildung läßt sich nämlich stetig auf den Rand fortsetzen, wobei die m-fachen Randpunkte von p zu m verschiedenen Randpunkten des Kreises gehören ([6], S. 364).

LEMMA 4. *Jede Fläche von G hat mindestens eine einfache A-Ecke.*

BEWEIS. Wir numerieren die Ecken gemäß Lemma 3 und nehmen an, daß es keine einfache A-Ecke gibt. Wir betrachten eine A-Ecke z_k, bei welcher die minimale Differenz von Indizes (modulo s) auftritt, also $z_k = z_{k+d}$, und wenn für eine A-Ecke $z_j = z_{j+t}$ gilt, dann folgt $|t| \geqq |d|$. Da Grad $(z_{k+1}) \neq 1$, ist $z_{k+2} \neq z_k$, also sicher $d \geqq 4$ (o. B. d. A. $d>0$). Nach Annahme ist z_{k+2} auch eine Mehrfachecke. Es ist daher möglich, einen Punkt u_1 des Winkelbereichs $(z_{k+1}, z_{k+2}, z_{k+3})$ innerhalb von p mit einem Punkt u_2 eines anderen Winkelbereichs von z_{k+2} zu verbinden und diesen Weg durch die beiden Strecken $\overline{u_1 z_{k+2}}$ und $\overline{z_{k+2} u_2}$ zu einer geschlossenen Jordankurve J durch z_{k+2} zu ergänzen (Abb. 1). Der geschlossene Kantenzug $(z_k, z_{k+1}, \ldots, z_{k+d})$ tritt bei z_{k+2} von einem der beiden durch J bestimmten Gebiete

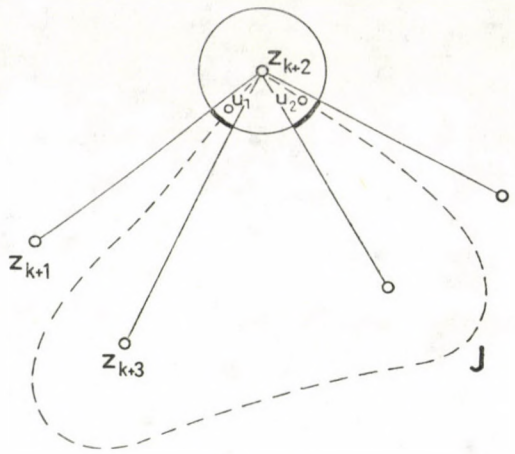


Abbildung 1.

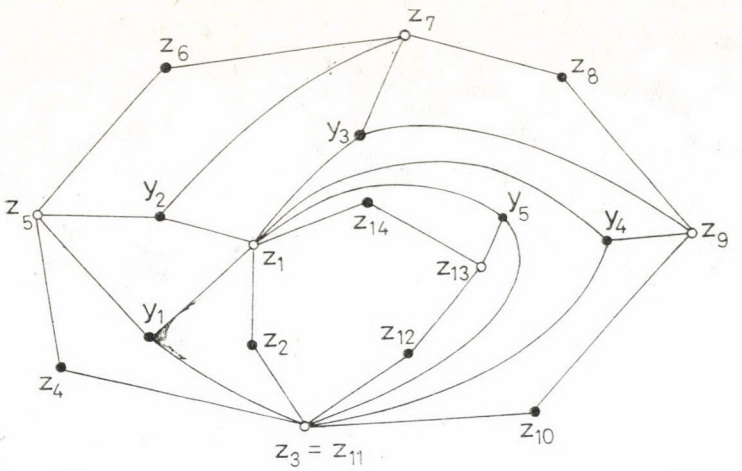
in das andere über. Da er J nur im Punkt z_{k+2} schneiden kann und wieder zu $z_k = z_{k+d}$ zurückgeht, muß er mindestens noch ein zweitesmal durch z_{k+2} laufen, d. h. $z_{k+2} = z_{k+l}$ mit $2 < l < d$, im Widerspruch zur Minimalität von d.

LEMMA 5. *Wir können G weiters so modifizieren, daß alle Flächen Vierecke sind, ohne die Eigenschaften a), b), c) und die von Lemma 2 zu ändern, und daß auch der modifizierte Graph schlicht, eben, paar und zusammenhängend ist.*

BEWEIS. Man kann jedes $2s$-Eck p mit $s \geqq 3$ in folgender Weise in Vierecke zerlegen: Sei z_1 gemäß Lemma 4 eine einfache A-Ecke von p, und von z_1 ausgehend seien die Ecken von p gemäß Lemma 3 im Uhrzeigersinn numeriert. Wir fügen in p eine B-Ecke y_1 ein und verbinden sie mit den A-Ecken z_1, z_3, z_5 durch sich nicht überschneidende Kanten derart, daß die zu einer Ecke z_i führende Kante durch den Winkelbereich (z_{i-1}, z_i, z_{i+1}) geht und durch keinen anderen zu z_i gehörigen Winkelbereich. Dadurch wird p in zwei Vierecke und ein $2(s-1)-$Eck p_1 zerlegt (Abb. 2). Im Inneren von p_1 fügen wir nun eine B-Ecke y_2 ein und verbinden sie in analoger Weise mit z_1, z_5, z_7 durch Kanten in p_1, usw. Schließlich wird y_{s-2} im Inneren eines 6-Ecks p_{s-3} gewählt und mit z_1, z_{2s-3}, z_{2s-1} verbunden, wodurch p in Vierecke zerlegt ist. Die Einfachheit von z_1 garantiert, daß dabei keine Mehrfachkanten auftreten. Die Grade der A-Ecken werden nicht kleiner, die Grade der alten B-Ecken bleiben unverändert, und die neuen B-Ecken y_i haben Grad $3 \leqq g$.

Wir betrachten von nun an den gemäß Lemma 2 und 5 abgeänderten Graphen G. Wir beweisen zunächst die Ungleichung (1). Bezeichnungen:

a... Anzahl der A-Ecken,
b... Anzahl der B-Ecken,
b_q... Anzahl der B-Ecken vom Grad q,
k... Anzahl der Kanten.



○ ... A-Ecken

● ... B-Ecken

Abbildung 2.

Jede A-Ecke ist mit mindestens n anderen A-Ecken über je zwei Kanten verbunden. Wenn zwei A-Ecken zu demselben Viereck gehören, dann sind sie auf diese Weise doppelt verbunden. Bei jeder B-Ecke vom Grad q treten $\frac{q(q-1)}{2}$ Verbindungen auf, davon mindestens q Doppelverbindungen. Da die Doppelverbindungen zu zwei B-Ecken zählen, erhalten wir für jede B-Ecke vom Grad q höchstens $\frac{q(q-1)}{2}-\frac{q}{2}=\frac{q(q-2)}{2}$ zu zählende Verbindungen. Die Gesamtzahl der Verbindungen muß wegen b) $\geqq\frac{na}{2}$ sein, also

$$na \leqq \sum_{q=3}^{g} q(q-2)b_q \leqq g \sum_{q=3}^{g}(q-2)b_q. \tag{5}$$

Aus dem Eulerschen Polyedersatz folgt $k<2(a+b)$. $k=\sum qb_q$, also $\sum(q-2)b_q<$ $<2a$, und daher folgt aus (5): $na<g\cdot 2a$, somit $n<2g$.

Zum Beweis der Ungleichung (2) verwenden wir außer G noch folgenden ebenen Graphen H: Die Eckenmenge von H sei B. Wenn zwei B-Ecken zum selben Viereck gehören, dann verbinden wir sie durch eine Kante innerhalb dieses Vierecks. Die Grade der B-Ecken in H stimmen mit denen in G überein. Auf diese Weise liegt jede A-Ecke x vom Grade s in einer Fläche von H mit der Seitenzahl s, die wir

mit $\bar{x}$ bezeichnen. $\bar{x}$ enthält sonst keine A-Ecke. Umgekehrt enthält jede Fläche von H genau eine A-Ecke. Da die A-Ecken genau den Scheiben entsprechen, ist also jeder Fläche von H genau eine Scheibe zugeordnet und umgekehrt. Auf H wenden wir nun das Konzept der „Eulerschen Beiträge" an [5, p. 54ff.]. Sei W_p die Menge der Winkel einer Fläche p in einem ebenen Graphen, $s(p)$ die Seitenzahl (=Winkelanzahl) und $q(\delta)$ der Grad der zu einem Winkel δ gehörigen Ecke. Dann heißt

$$C(p) := 1 - \frac{s(p)}{2} + \sum_{\delta \in W_p} \frac{1}{q(\delta)}$$

der Eulersche Beitrag von p. Aus dem Eulerschen Polyedersatz ergibt sich [5], daß

$$\sum_{p \in F} C(p) > 0, \tag{6}$$

wenn F die Menge aller Flächen des Graphen bedeutet. Es muß daher mindestens eine Fläche p_0 geben mit

$$C(p_0) > 0. \tag{7}$$

Wir betrachten nun eine A-Ecke x vom Grad s. Jede mit x in G verbundene B-Ecke δ vom Grad $q(\delta)$ liefert $q(\delta)-1$ A-Nachbarn von x. Mindestens s solcher Nachbarn gehören jedoch zu zwei B-Ecken, also gilt wegen b) $\sum_{\delta \in W_{\bar{x}}} (q(\delta)-1) - s \geqq n$. Da $\sum_{\delta \in W_{\bar{x}}} 1 = s$, folgt daraus:

$$\sum_{\delta \in W_{\bar{x}}} q(\delta) \geqq n + 2s. \tag{8}$$

Für $0 < t_1 \leqq t_2$ und $0 \leqq h < t_1$ ist $\frac{1}{t_1} + \frac{1}{t_2} \leqq \frac{1}{t_1 - h} + \frac{1}{t_2 + h}$. Durch wiederholte Anwendung dieses Prinzips erhält man aus (8) unter Beachtung von $3 \leqq q(\delta) \leqq g$:

$$\sum_{\delta \in W_{\bar{x}}} \frac{1}{q(\delta)} \leqq \underbrace{\frac{1}{3} + \ldots + \frac{1}{3}}_{(s-2)\text{-mal}} + \frac{1}{(n+2s)-(s-2)\cdot 3 - g} + \frac{1}{g} =$$

$$= \frac{s-2}{3} + \frac{1}{g} + \frac{1}{n-g-s+6}, \quad \text{falls} \quad n-g-s+6 \geqq 3 \quad \text{ist.}$$

In diesem Fall ergibt sich aus (7) die Ungleichung

$$1 - \frac{s}{2} + \frac{s-2}{3} + \frac{1}{g} + \frac{1}{n-g-s+6} > 0 \quad \text{für} \quad s = s(p_0).$$

Die linke Seite ist eine monoton fallende Funktion von s, wir können daher s durch 3 ersetzen und erhalten

$$\frac{1}{g} + \frac{1}{n-g+3} > \frac{1}{6}, \quad \text{d.h.} \quad n < \frac{g^2-3g+18}{g-6}.$$

Wir zeigen nun noch, daß der Fall $n-g-s+6<3$ nicht eintreten kann. Wir können $n>g+3$ annehmen, da sonst die Aussage des Satzes in trivialer Weise folgt. Also wäre $6<s$, und falls $n-s+3\geqq 3$, hätten wir

$$\sum \frac{1}{q(\delta)} \leqq \frac{s-1}{3}+\frac{1}{(n+2s)-(s-1)\cdot 3} = \frac{s-1}{3}+\frac{1}{n-s+3},$$

und aus (7) folgte $1-\frac{s}{2}+\frac{s-1}{3}+\frac{1}{n-s+3}>0$.

Wegen Monotonie müßte diese Ungleichung auch für $s=6$ gelten, und es wäre $-\frac{1}{3}+\frac{1}{n-3}>0$, also $n<6$, im Widerspruch zu $n>g+3$ und $g>6$. Schließlich wäre noch der Fall $n-s+3<3$, also $n<s$ zu behandeln. Da hätten wir $1-\frac{s}{2}+\frac{s}{3}>0$, somit $n<s<6$, also auch einen Widerspruch.

II. Unendliche (n, g)-Packungen in der Ebene

Sei M eine unendliche (n, g)-Packung in der Ebene, und G ein zugehöriger Graph (wie in I.). z sei ein beliebiger fester Punkt, und K_r sei ein offener Kreis mit Mittelpunkt z und Radius r, auf dessen Rand keine B-Ecke von G liegt. M_r' sei die Menge der Scheiben von M, die in K_r enthalten sind und $M_r'':=\{S\cap K_r \mid \emptyset\neq S\cap K_r\neq S\in M\}$; schließlich sei $M_r:=M_r'\cup M_r''$. Die Scheiben von M_r' haben in M_r genausoviele Nachbarn wie in M, also mindestens n. Weiters seien $a(r)$, $a'(r)$ bzw. $a''(r)$ die Anzahlen der Elemente von M_r, M_r' bzw. M_r''.

SATZ 2. *Wenn in einer unendlichen (n, g)-Packung in der Ebene $a(r)$ endlich ist (für jedes $r>0$, für das es definiert ist), und*

$$\lim_{r\to\infty} \frac{a''(r)}{a'(r)} = 0, \tag{9}$$

dann gilt

$$n \leqq 2g \qquad \text{für} \quad g\geqq 3 \tag{10}$$

und

$$n \leqq \frac{g^2-3g+18}{g-6} \quad \text{für} \quad g\geqq 7. \tag{11}$$

BEMERKUNG. Die Voraussetzungen des Satzes sind sinnvoll, da G nur abzählbar viele B-Ecken enthält und daher nur abzählbar viele $r>0$ ausgeschlossen werden.

BEWEIS. Sei $\varepsilon>0$ beliebig vorgegeben. Wir wählen r so groß, daß $\frac{a''(r)+1}{a'(r)}<\varepsilon$ ist. Wie in I. konstruieren wir zu M_r einen schlichten ebenen paaren Graphen G_r, für den a) und c) gelten. Wenn G_r nicht zusammenhängend ist, dann gibt es ein Gebiet f von G_r, dessen Rand Ecken aus mindestens zwei Zusammenhangskom-

ponenten Z_1 und Z_2 von G_r enthält. Wir fügen innerhalb von f eine neue B-Ecke ein und verbinden sie mit einer A-Ecke des Z_1-Randes von f und einer A-Ecke des Z_2-Randes von f. Dadurch wird die Anzahl der Zusammenhangskomponenten von G_r um eins erniedrigt. Da G_r endlich ist, wird er auf diese Weise nach endlich vielen Schritten zusammenhängend. Die obengenannten Eigenschaften von G_r werden dabei nicht geändert.

G_r kann A-Ecken vom Grad 1 oder mit weniger als $g+1$ zweiten Nachbarn enthalten und daher nicht unmittelbar gemäß Lemma 2 modifiziert werden. Solche A-Ecken liegen notwendigerweise am Rand des Außengebiets von G_r. Wir ändern daher G_r folgenderweise ab: Seien $x_1, \ldots, x_t$ die A-Ecken des Außengebiets von G_r, in dem in Lemma 3 definierten Umlaufsinn.

Damit wir auf G_r ohne Schwierigkeiten die Modifikationen von Lemma 2 anwenden können, ist es zweckmäßig, vorher die mehrfachen A-Ecken des Außengebiets p_0 von G_r folgenderweise zu eliminieren: Wenn x_i eine mehrfache A-Ecke von p_0 ist, fügen wir in p_0 eine neue B-Ecke ein und verbinden sie mit x_{i-1}, x_i und x_{i+1} in den entsprechenden Winkelbereichen. (Wenn dabei eine B-Ecke vom Grad 2 an den Rand eines Vierecks kommt, wird sie mit ihren Kanten weggelassen.) Anschließend numerieren wir die A-Ecken von p_0 neu gemäß Lemma 3. Auf diese Weise verringern wir schrittweise die Gesamtmultiplizität der mehrfachen A-Ecken von p_0, bis schließlich keine mehr vorhanden sind. Wenn dann zwei aufeinanderfolgende A-Ecken x_i, x_{i+1} über eine einfache zweigradige B-Ecke des Außengebiets verbunden sind, bezeichnen wir diese B-Ecke mit y_i. Andernfalls verbinden wir x_i und x_{i+1} durch zwei zusätzliche Kanten über eine neu eingefügte B-Ecke y_i außerhalb von G_r. Führt man dies sukzessive für $i=1, \ldots, t$ durch, so wird schließlich das Außengebiet von G_r von dem geschlossenen Kantenzug $(x_1, y_1, x_2, \ldots$ $\ldots, x_t, y_t, x_1)$ berandet. Damit die neuen zwei-gradigen B-Ecken y_i bei der Modifikation nach Lemma 2 nicht weggelassen werden, fügen wir in p_0 eine neue A-Ecke x_0 ein und verbinden sie innerhalb von p_0 der Reihe nach mit $y_1, \ldots, y_t$. Dadurch bekommen die y_i Grad 3. Lemma 1 gilt in einem modifizierten Sinn auch für den so erhaltenen Graphen: Wenn nach dem Einfügen von x_0 ein Vierkreis Ecken oder Kanten sowohl im Inneren als auch im Äußeren enthielte, so müßte x_0 eine Ecke dieses Vierkreises sein; die andere A-Ecke des Vierkreises wäre dann aber eine mehrfache Ecke von p_0 gewesen, was wir oben ausgeschlossen haben. Wir können daher diesen Graphen gemäß Lemma 2 und 5 modifizieren, denn die Tatsache, daß die A-Ecken mindestens $g+1$ zweite Nachbarn haben, wird dabei nur für die von $x_0, \ldots, x_t$ verschiedenen A-Ecken verwendet.

$a'(r)$ ist dann kleiner oder gleich der Anzahl der A-Ecken von G_r, welche mit mindestens n anderen A-Ecken aus G_r durch je zwei Kanten verbunden sind. Die Gesamtzahl der (zu zählenden) Verbindungen in G_r ist daher $\geqq \frac{n \cdot a'(r)}{2}$. Wir haben also in (5) a durch $a'(r)$ zu ersetzen und erhalten dann:

$$n \cdot a'(r) < g \cdot 2(a(r)+1) = 2g(a'(r)+a''(r)+1),$$

also $n<2g(1+\varepsilon)$. Da ε beliebig war, folgt (10). Wir konstruieren nun zu G_r einen ebenen Graphen H_r wie in I. Den Eulerschen Beitrag einer Fläche p von H_r bezeichnen wir mit $C_r(p)$. Da $s(p) \geqq 3$ und $q(\delta) \geqq 3$, ist $C_r(p) \leqq \frac{1}{2}$ für alle Flächen von H_r.

Sei nun F_r' bzw. F_r'' die Menge der Flächen von H_r, deren zugehörige Scheiben aus M_r' bzw. M_r'' sind. Die Fläche, die zu der hinzugefügten A-Ecke x_0 gehört, rechnen wir zu F_r''. Aus (6) erhalten wir dann:

$$\sum_{p\in F_r'} C_r(p)+\sum_{p\in F_r''} C_r(p)>0$$

und daher $a'(r)\cdot\max_{p\in F_r'} C_r(p)+(a''(r)+1)\cdot\frac{1}{2}>0$, also $\max_{p\in F_r'} C_r(p)>-\frac{\varepsilon}{2}$, d.h. es gibt eine Fläche $p_r\in F_r'$ mit

$$C_r(p_r)>-\frac{\varepsilon}{2}. \tag{12}$$

Da für die Flächen aus F_r' die Ungleichung (8) gilt, erhalten wir wie in I., indem wir (7) durch (12) ersetzen:

$$1-\frac{s}{2}+\frac{s-2}{3}+\frac{1}{g}+\frac{1}{n-g-s+6}>-\frac{\varepsilon}{2} \quad \text{für} \quad s=s(p_r).$$

(Der Fall $n-g-s+6<3$ läßt sich wie in I. ausschließen.) Wir können wieder $s=3$ setzen und erhalten

$$\frac{1}{g}+\frac{1}{n-g+3}>\frac{1}{6}-\frac{\varepsilon}{2}.$$

Die linke Seite ist nun nicht mehr von ε abhängig, also folgt $\frac{1}{g}+\frac{1}{n-g+3}\geqq\frac{1}{6}$ und daraus (11).

III. Maximalpackungen

Eine n-Nachbarnpackung kongruenter Scheiben heißt *maximal*, wenn n gleich der Newtonschen Zahl N dieser Scheiben ist.

SATZ 3. *Für jede maximale n-Nachbarnpackung in der Ebene gilt*

$$n\leqq 21.$$

BEMERKUNG. Wie schon in der Einleitung erwähnt wurde, ist 21 vermutlich die genaue Schranke, und zwar scheint das zum Archimedischen Mosaik (3, 12, 12) duale Mosaik [2, Abb. 135] eine maximale 21-Nachbarnpackung zu sein.

Zum Beweis des Satzes:

LEMMA 6. *Jede maximale n-Nachbarnpackung erfüllt die Voraussetzungen von Satz 2.*

BEWEIS. Die Endlichkeit von $a(r)$ ist trivial. Sei nun d der Durchmesser und f der Flächeninhalt der Scheiben von M. Diejenigen Scheiben von M, die zu M_r'' beitragen, liegen nicht ganz innerhalb von K_r; folglich sind sie in dem zu K_r konzentrischen Kreisring mit den Radien $r-d$ und $r+d$ enthalten, und es ist

$$a''(r)\leqq\frac{4\pi dr}{f}. \tag{13}$$

In jedem Quadrat mit Seitenlänge $3d$ muß mindestens eine Scheibe S aus M ganz enthalten sein, denn sonst könnte man in dem konzentrischen und homothetischen Quadrat mit Seitenlänge d eine zusätzliche Scheibe S_0 einfügen, die zu den anderen disjunkt ist, und man könnte dann S_0 so verschieben, daß sie eine Scheibe S_1 von M gerade berührte. Dadurch würde man die Nachbarnzahl von S_1 um 1 erhöhen, im Widerspruch zur Maximalität der Packung. Wenn nun das Quadrat innerhalb von K_r liegt, gehört S zu M'_r. Da in K_r mindestens $\left[\frac{r}{3d}\right]^2$ disjunkte solche Quadrate Platz haben, folgt:

$$a'(r) \geqq \left[\frac{r}{3d}\right]^2. \tag{14}$$

Aus (13) und (14) ergibt sich die Richtigkeit von (9), und auch die Unendlichkeit der Packung.

LEMMA 7. *Die Newtonsche Zahl einer Scheibe mit Minimalwinkel* $\alpha \leqq \frac{2\pi}{15} = 24°$ *ist* $\geqq 23$.

BEWEIS. Sei E die zu diesem Winkel α gehörige Ecke („Spitze") der Scheibe. D sei ein Randpunkt der Scheibe, für den $\overline{DE}$ gleich dem Durchmesser ist. Wir nehmen an, daß $\overline{DE}=1$ ist. Bei E legen wir 9 Nachbarn mit ihrer Spitze an, so daß sie in einem Winkelbereich von $9 \cdot 24° = 216°$ liegen, und die Symmetrale dieses Winkelbereichs möge gleich der Symmetralen von α in S sein. Bei D legen wir weitere 6 Nachbarn mit ihrer Spitze an, sodaß der entsprechende Winkelbereich $= 6 \cdot 24° = 144°$ ist und die Symmetrale parallel zu der vorhin genannten ist (Abb. 3). In den Bereichen, die in der Zeichnung die Nummern 13 und 20 tragen, können wir zwei weitere Nachbarn anlegen. Diese werden unter Umständen S nicht im Punkt D berühren. Wir legen nun durch den Mittelpunkt M der Strecke $\overline{DE}$ eine Gerade parallel zu der Trennlinie der Bereiche 8 und 9. Dadurch wird die Begrenzung des Bereichs 10 definiert. Dieser umfaßt ein Dreieck mit den Winkeln 24°, 60° und 96°, und die kleinste Seite ist $\geqq \frac{1}{2}\frac{\cos 18°}{\cos 6°}$. Die zweitlängste Seite ist daher $\geqq \frac{1}{2}\frac{\cos 18° \cdot \sin 60°}{\cos 6° \cdot \sin 24°} > 1$, also können wir hier sicher einen Nachbarn unterbringen. Analog wird der Bereich 23 konstruiert. Die Trennlinie der Bereiche 11 und 12 wird parallel zu der von 9 und 10 gelegt, und zwar so, daß ihr Schnittpunkt mit der Trennlinie 10/11 auf dem Rand von S zu liegen kommt. Es ist nun leicht zu sehen, daß der Bereich 12 nicht kleiner als der Bereich 10 ist, also können wir hier auch einen Nachbarn anlegen, und analog beim Bereich 21.

BEWEIS DES SATZES. Für Scheiben mit Minimalwinkel α kann man die Newtonsche Zahl folgendermaßen nach unten abschätzen [3]:

$$N \geqq \left[\frac{2\pi}{\alpha}\right] + \left[\frac{\pi}{\alpha}\right] - 1.$$

Für Maximalpackungen folgt daraus wegen $g \leqq \frac{2\pi}{\alpha}$: $n = N \geqq g + \left[\frac{g}{2}\right] - 1$. Wegen

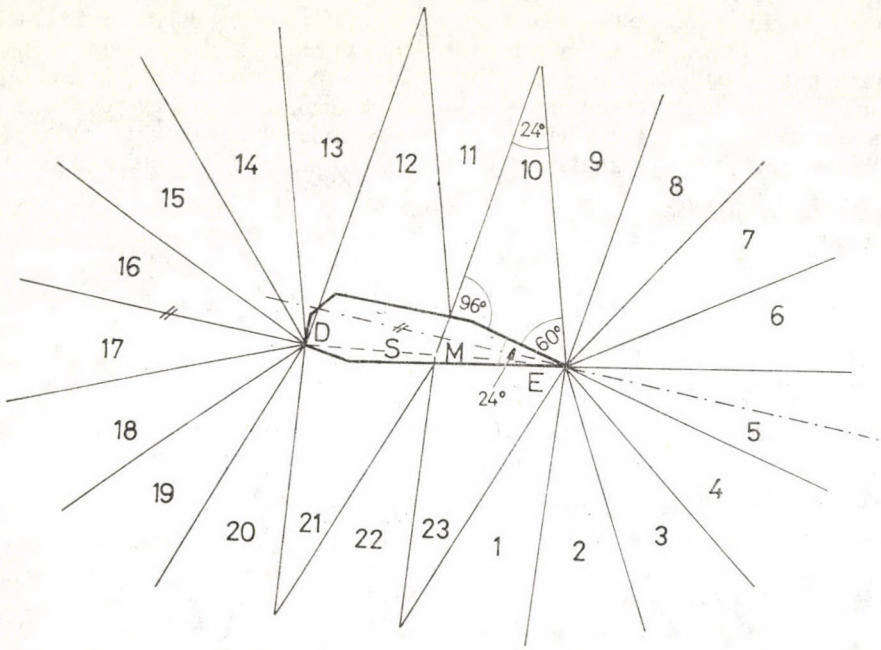


Abbildung 3.

Lemma 6 können wir Satz 2 anwenden. Das gibt, falls $g \geqq 7$ ist:

$$\left[\frac{3g}{2}\right] - 1 \leqq \frac{g^2 - 3g + 18}{g-6},$$

woraus $g \leqq 15$ folgt. Für $g \leqq 14$ ist $n \leqq 21$ nach (10) und (11), also bleibt nur der Fall $g = 15$ zu untersuchen. Hier muß $\alpha \leqq \frac{2\pi}{15}$ sein, also $N \geqq 23$ nach Lemma 7. Andererseits ergibt sich $N = n \leqq 22$ aus (11). Dieser Fall kann also gar nicht eintreten, womit der Satz bewiesen ist.

Abschließend möchte ich Herrn E. MAKAI jun. für zahlreiche wertvolle Verbesserungen meinen aufrichtigen Dank aussprechen.

LITERATURVERZEICHNIS

[1] FEJES TÓTH, L.: *Reguläre Figuren*, Akadémiai Kiadó, Budapest, 1965.
[2] FEJES TÓTH, L.: *Lagerungen in der Ebene, auf der Kugel und im Raum*, 2. Aufl., Springer-Verlag, Berlin—Heidelberg—New York, 1972.
[3] GÁCS, P.: Packing of convex sets in the plane with a great number of neighbours, *Acta Math. Acad. Sci. Hung.* **23** (1972), 383—388.
[4] LINHART, J.: Endliche n-Nachbarnpackungen in der Ebene und auf der Kugel, *Periodica Math. Hung.* **5** (1974), 303—306.
[5] ORE, O.: *The Four-Color Problem*, Academic Press, New York—London, 1967.
[6] BEHNKE, H., SOMMER, F.: *Theorie der analytischen Funktionen einer komplexen Veränderlichen*, Springer-Verlag, Berlin—Heidelberg—New York, 1965.

II. Lehrkanzel für Mathematik, Universität Salzburg,
Petersbrunnstraße 19, A—5020 Salzburg

(Eingegangen am 15. Februar 1975;
revidierte Fassung am 3. November 1979)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 295—300.

A NOTE ON THE PRODUCT OF DISTRIBUTIONS

by

B. FISHER

In the following we define a sequence $\{f_n\}$ of infinitely differentiable functions to be *regular* on the open interval (a, b) if the limit of the sequence

$$\{(f_n, \varphi)\} = \left\{\int_a^b f_n(x)\,\varphi(x)\,dx\right\}$$

exists for all test functions φ with support contained in (a, b).

The *product* of two distributions f and g on the open interval (a, b) was defined in [1] as the limit of the sequence $\{f_n g_n\}$, provided this sequence is regular on (a, b), where

$$f_n = f * \delta_n, \quad g_n = g * \delta_n, \quad \delta_n(x) = n\varrho(nx)$$

for $n=1, 2, \ldots$ and ϱ is a fixed infinitely differentiable function having the following properties:

(1) $\varrho(x) = 0$, for $|x| \geqq 1$,

(2) $\varrho(x) \geqq 0$,

(3) $\varrho(x) = \varrho(-x)$,

(4) $\int_{-1}^{1} \varrho(x)\,dx = 1$.

It is obvious that $\{\delta_n\}$ is a regular sequence converging to the Dirac delta-function $\delta(x)$.

It was then proved in [1] that this definition of the product was in agreement with the usual definition of the product when it exists, but with this definition of the product, further products of distributions are defined which are not defined by the usual definition. For example, it was proved that

$$x_+^r\,\delta^{(r)}(x) = \frac{1}{2}(-1)^r r!\,\delta(x),$$

$$x_-^r\,\delta^{(r)}(x) = \frac{1}{2} r!\,\delta(x)$$

for $r=0, 1, 2, \ldots$, where x_+^r and x_-^r are the summable functions defined by

$$x_+^r = \begin{cases} x^r & \text{for} \quad x > 0, \\ 0 & \text{for} \quad x \leqq 0, \end{cases}$$

$$x_-^r = \begin{cases} |x|^r & \text{for} \quad x < 0, \\ 0 & \text{for} \quad x \geqq 0. \end{cases}$$

It follows immediately from these equations that

$$|x|^{2r-1}\delta^{(2r-1)}(x) = (x_+^{2r-1} + x_-^{2r-1})\delta^{(2r-1)}(x) = 0 \tag{1}$$

for $r=1, 2, \ldots$ and

$$(\operatorname{sgn} x \cdot x^{2r})\delta^{(2r)}(x) = (x_+^{2r} - x_-^{2r})\delta^{(2r)}(x) = 0 \tag{2}$$

for $r=0, 1, 2, \ldots$.

We now prove some further results. First of all we have

PROPOSITION 1.

$$(\operatorname{sgn} x \cdot \ln |x|)\delta(x) = 0, \tag{3}$$

where $\operatorname{sgn} x \cdot \ln |x|$ *is the summable function defined by*

$$\operatorname{sgn} x \cdot \ln |x| = \begin{cases} \ln x & \text{for} \quad x > 0, \\ -\ln |x| & \text{for} \quad x < 0. \end{cases}$$

PROOF. Putting

$$(\operatorname{sgn} x \cdot \ln |x|)_n = (\operatorname{sgn} x \cdot \ln |x|)^* \delta_n(x) =$$

$$= \int_{-1/n}^{x} \ln (x-t)\delta_n(t)\, dt - \int_{x}^{1/n} \ln (t-x)\delta_n(t)\, dt,$$

we note that $(\operatorname{sgn} x \cdot \ln |x|)_n \delta_n(x)$ is an odd function and has its support contained in the interval $(-1/n, 1/n)$. It follows that

$$\int_{-\infty}^{\infty} (\operatorname{sgn} x \cdot \ln |x|)_n \delta_n(x)\, dx = \int_{-1/n}^{1/n} (\operatorname{sgn} x \cdot \ln |x|)_n \delta_n(x)\, dx = 0.$$

Further

$$\int_{-1/n}^{1/n} \left| x(\operatorname{sgn} x \cdot \ln |x|)_n \delta_n(x) \right| dx \leqq$$

$$\leqq \int_{-1/n}^{1/n} |x|\, \delta_n(x) \int_{-1/n}^{x} |\ln (x-t)|\, \delta_n(t)\, dt\, dx +$$

$$+ \int_{-1/n}^{1/n} |x|\, \delta_n(x) \int_{x}^{1/n} |\ln (t-x)|\, \delta_n(t)\, dt\, dx.$$

Putting $k=\sup \varrho(x)$ we have

$$\int_{-1/n}^{1/n} |x|\, \delta_n(x) \int_{-1/n}^{x} |\ln (x-t)|\, \delta_n(t)\, dt\, dx \leqq$$

$$\leqq nk^2 \int_{-1/n}^{1/n} |\ln (x-t)|\, dt\, dx =$$

$$= nk^2 \int_{-1/n}^{1/n} (x+1/n)[1-\ln (x+1/n)]\, dx \leqq$$

$$\leqq 2k^2 \int_{-1/n}^{1/n} [1-\ln (x+1/n)]\, dx$$

which tends to zero as n tends to infinity.

Similarly

$$\lim_{n\to\infty} \int_{-1/n}^{1/n} |x|\, \delta_n(x) \int_{x}^{1/n} |\ln (t-x)|\, \delta_n(t)\, dt\, dx = 0.$$

It follows that

$$\lim_{n\to\infty} \int_{-1/n}^{1/n} \left|x(\operatorname{sgn} x \cdot \ln |x|_n\, \delta_n(x)\right| dx = 0.$$

Now let φ be an arbitrary test function with compact support. We have

$$\varphi(x) = \varphi(0)+x\varphi'(\xi x)$$

where $0\leqq\xi\leqq 1$. Thus if $K=\sup |\varphi'(x)|$,

$$\left|((\operatorname{sgn} x \cdot \ln |x|)\, \delta(x), \varphi)\right| =$$

$$= \lim_{n\to\infty} \left| \int_{-1/n}^{1/n} (\operatorname{sgn} x \cdot \ln |x|)_n\, \delta_n(x)[\varphi(0)+x\varphi'(\xi x)]\, dx \right| \leqq$$

$$\leqq \lim_{n\to\infty} \left\{ \varphi(0) \int_{-1/n}^{1/n} (\operatorname{sgn} x \cdot \ln |x|)_n\, \delta_n(x)\, dx + \right.$$

$$\left. + \int_{-1/n}^{1/n} \left|x(\operatorname{sgn} x \cdot \ln |x|)_n\, \delta_n(x)\varphi'(\xi x)\right| dx \right\} \leqq$$

$$\leqq \lim_{n\to\infty} \left\{ 0+K \int_{-1/n}^{1/n} \left|x(\operatorname{sgn} x \cdot \ln |x|)_n\, \delta_n(x)\right| dx \right\} = 0,$$

from what we have proved above. Thus

$$\lim_{n\to\infty} ((\operatorname{sgn} x \cdot \ln |x|)_n\, \delta_n(x), \varphi) = 0$$

for arbitrary test function φ and equation (3) follows. □

PROPOSITION 2.

$$(|x^{2r-1}|\ln|x|)\,\delta^{(2r-1)}(x)=0 \tag{4}$$

for $r=1,2,\ldots$ *and*

$$(\operatorname{sgn}x\cdot x^{2r}\ln|x|)\,\delta^{(2r)}(x)=0 \tag{5}$$

for $r=0,1,2,\ldots.$

PROOF. Assume that equation (5) holds for some r. Then since it is easily proved that

$$(|x^{2r+1}|\ln|x|)\,\delta^{(2r)}(x)=\lim_{n\to\infty}(|x^{2r+1}|\ln|x|)_n\,\delta_n^{(2r)}(x)=0$$

we have on differentiating that

$$\begin{aligned}\lim_{n\to\infty}\{&(2r+1)(\operatorname{sgn}x\cdot x^{2r}\ln|x|)_n\delta_n^{(2r)}(x)+(\operatorname{sgn}x\cdot x^{2r})_n\,\delta_n^{(2r)}(x)+\\&+(|x^{2r+1}|\ln|x|)_n\delta_n^{(2r+1)})(x)\}=\\&=(2r+1)(\operatorname{sgn}x\cdot x^{2r}\ln|x|)\,\delta^{(2r)}(x)+(\operatorname{sgn}x\cdot x^{2r})\,\delta^{(2r)}(x)+\\&+\lim_{n\to\infty}(|x^{2r+1}|\ln|x|)_n\,\delta_n^{(2r+1)}(x)=\\&=0+0+\lim_{n\to\infty}(|x^{2r+1}|\ln|x|_n\,\delta_n^{(2r+1)}(x)=0\end{aligned}$$

by our assumption and using equation (2). Thus equation (4) is true for $r+1$.

Similarly, differentiating the easily proved result

$$(\operatorname{sgn}x\cdot x^{2r+2}\ln|x|)\,\delta^{(2r+1)}(x)=\lim_{n\to\infty}(\operatorname{sgn}x\cdot x^{2r+2}\ln|x|)_n\,\delta_n^{(2r+1)}(x)=0$$

we get the result

$$(\operatorname{sgn}x\cdot x^{2r+2}\ln|x|)\,\delta^{(2r+2)}(x)=0$$

by what we have just proved and using equation (1).

Thus equation (5) is true for $r+1$. Since equation (5) holds when $r=0$, being equation (3), the proofs of equations (4) and (5) follow by induction. □

PROPOSITION 3.

$$H(x)|x^{-1}|=x_+^{-1}, \tag{6}$$

where $H(x)$ *denotes Heaviside's function and*

$$x_+^{2r-1}(\operatorname{sgn}x\cdot x^{-2r})=x_+^{-1}-\varphi(2r-1)\,\delta(x), \tag{7}$$

$$x_+^{2r}|x^{-2r-1}|=x_+^{-1}-\varphi(2r)\,\delta(x) \tag{8}$$

for $r=1,2,\ldots,$ *where* $\varphi(r)=\sum_{i=1}^{r}1/i.$

PROOF. Using the result

$$H(x)(\operatorname{sgn}x\cdot\ln|x|)=\lim_{n\to\infty}H_n(x)(\operatorname{sgn}x\cdot\ln|x|)_n=\ln x_+$$

we have on differentiating that

$$\lim_{n\to\infty}\{\delta_n(x)(\operatorname{sgn} x\cdot\ln|x|)_n+H_n(x)(|x^{-1}|)_n=x_+^{-1}$$

and using equation (3) we have

$$\lim_{n\to\infty}H_n(x)(|x^{-1}|)_n=x_+^{-1},$$

completing the proof of equation (6).

We will now assume that equation (8) holds for some r. Then using the result

$$x_+^{2r+1}|x^{-2r-1}|=\lim_{n\to\infty}(x_+^{2r+1})_n(|x^{-2r-1}|)_n=H(x)$$

we have on differentiating that

$$\lim_{n\to\infty}\{(2r+1)(x_+^{2r})_n(|x^{-2r-1}|)_n-(2r+1)(x_+^{2r+1})_n(\operatorname{sgn} x\cdot x^{-2r-2})_n\}=\delta(x).$$

By our assumption we now have

$$\lim_{n\to\infty}(x_+^{2r+1})_n(\operatorname{sgn} x\cdot x^{-2r-2})_n=x_+^{-1}-\varphi(2r)\delta(x)-\frac{1}{2r+1}\delta(x)=$$

$$=x_+^{-1}-\varphi(2r+1)\delta(x)$$

and so equation (7) holds for $r+1$.

Now using the result

$$x_+^{2r+2}(\operatorname{sgn} x\cdot x^{-2r-2})=\lim_{n\to\infty}(x_+^{2r+2})_n(\operatorname{sgn} x\cdot x^{-2r-2})_n=H(x)$$

we have on differentiating that

$$\lim_{n\to\infty}\{(2r+2)(x_+^{2r+1})_n(\operatorname{sgn} x\cdot x^{-2r-2})_n-(2r+2)(x_+^{2r+2})_n(|x^{-2r-3}|)_n\}=\delta(x).$$

By what we have just proved, we now have

$$\lim_{n\to\infty}(x_+^{2r+2})_n(|x^{-2r-3}|)_n=x_+^{-1}-\varphi(2r+1)\delta(x)-\frac{1}{2r+2}\delta(x)=x_+^{-1}-\varphi(2r+2)\delta(x)$$

and so equation (8) holds for $r+1$. Since equation (8) holds for the case $r=0$, being equation (6), equations (7) and (8) follow by induction. □

PROPOSITION 4.

$$x_-^{2r-1}(\operatorname{sgn} x\cdot x^{-2r})=-x_-^{-1}+\varphi(2r-1)\delta(x), \tag{9}$$

$$x_-^{2r}|x^{-2r-1}|=x_-^{-1}-\varphi(2r)\delta(x) \tag{10}$$

for $r=1,2,\ldots.$

PROOF. Replace x by $-x$ in equations (7) and (8) and equations (9) and (10) follow immediately. □

PROPOSITION 5.

$$x^{2r-1}(\operatorname{sgn} x \cdot x^{-2r}) = |x^{-1}| - 2\varphi(2r-1)\delta(x),$$
$$|x^{2r-1}|(\operatorname{sgn} x \cdot x^{-2r}) = x^{-1},$$
$$x^{2r}|x^{-2r-1}| = |x^{-1}| - 2\varphi(2r-1)\delta(x),$$
$$(\operatorname{sgn} x \cdot x^{2r})|x^{-2r-1}| = x^{-1}$$

for $r=1, 2, \ldots$.

PROOF. Using equations (7), (8), (9) and (10) we have

$$x^{2r-1}(\operatorname{sgn} x \cdot x^{-2r}) = (x_+^{2r-1} - x_-^{2r-1})(\operatorname{sgn} x \cdot x^{-2r}) =$$
$$= |x|^{-1} - 2\varphi(2r-1)\delta(x),$$
$$|x^{2r-1}|(\operatorname{sgn} x \cdot x^{-2r}) = (x_+^{2r-1} + x_-^{2r-1})(\operatorname{sgn} x \cdot x^{-2r}) = x^{-1},$$
$$x^{2r}|x^{-2r-1}| = (x_+^{2r} + x_-^{2r})|x^{-2r-1}| = |x^{-1}| - 2\varphi(2r-1)\delta(x),$$
$$(\operatorname{sgn} x \cdot x^{2r})|x^{-2r-1}| = (x_+^{2r} - x_-^{2r})|x^{-2r-1}| = x^{-1}.$$

REFERENCES

[1] FISHER, B.: The product of distributions, *Quart. J. Math. Oxford* (2) **22** (1971), 291—298.

Department of Mathematics, The University of Leicester, LE1 7RH, England

(Received February 18, 1975; revised September 7, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 301—321.

ОБ ОДНОМ МОДИФИЦИРОВАННОМ ДВУСТОРОННЕМ ИТЕРАЦИОННОМ МЕТОДЕ

J. HEGEDŰS

Работа примыкает к статье [1], где с помощью весов, примененных к парам верхних и нижних приближений решения линейного дифференциального уравнения в частных производных, установлена ускоренная сходимость полученных приближений к решению.

Мы приведем такую конструкцию (аналогичную конструкции в [1]) с весами, при которой на каждом шагу получаются снова двусторонние приближения решения, и в отличие от [1] дадим достаточные условия возможности выбора весов для соблюдения этого свойства. Результаты получены в случае начальной задачи (часть II работы) и общей изотонной (см. [2], [4]) краевой задачи (часть I) для нелинейного дифференциального уравнения n-го порядка.

Из приведенных в работе формул легко получить оценки для весов. Затронута и проблем квадратичной сходимости метода.

Несмотря на некоторое сходство двух задач (в I и II), мы привели все же два разных метода: в части I метод неравенств в интегральной форме, а в части II в дифференциальной форме. Дело в том, что и тот и другой метод имеет свои преимущества и недостатки по сравнению с другим. Второй метод можно применить и в части I, первый же трудно применим в части II. С другой стороны второй метод требует вычисления постоянных в соответствующих нижних и верхних оценках (для невязок) на каждом шагу для определения искомых величин: весов λ_p, μ_p (но он дает более точные результаты для определения λ_p, μ_p), а первый метод более удобен в силу того, что λ_p, μ_p определяются из явно выписанных линейных уравнений с одним неизвестным (этому, конечно, способствует и характер задачи в I).

Номера формул в части I и во Введении (кроме задач (1′), (1*)) идут без штриха, а в части II снабжены штрихами: (2′), (3′), ...

Введение

Введем некоторые обозначения и основные условия рассматриваемых задач.

Через $\mathscr{M}$ будем обозначать пространство $(n-1)$-раз непрерывно дифференцируемых на $[0, 1]$ функций, удовлетворяющих краевым соотв. начальным условиям рассматриваемых ниже задач, и вводим в $\mathscr{M}$ расстояние- ϱ и час-

тичное упорядочение $\overset{(B)}{\leqq}$ по следующему правилу:

$$\text{(A)} \qquad \varrho(u, v) = \sum_{i=0}^{n-1} \max_{x\in[0,1]} |u^{(i)}(x) - v^{(i)}(x)| \quad (u, v \in \mathscr{M}),$$

$$\text{(B)} \qquad u \overset{(B)}{\leqq} v \Leftrightarrow u^{(i)}(x) \leqq v^{(i)}(x) \quad (0 \leqq x \leqq 1;\ i = 0, \ldots, n-1;\ u, v \in \mathscr{M}).$$

Мы будем рассматривать две задачи ((1) и (1′)):

$$\text{(1)} \qquad \begin{aligned} &y^{(n)}(x) = f[y] \equiv f(x, y, \ldots, y^{(n-1)}) \quad (0 \leqq x \leqq 1,\ n \geqq 2,\ f \in C([0,1]\times R^n)), \\ &L_i y \equiv \sum_{k=0}^{n-1} [a_{ik} y^{(k)}(0) + b_{ik} y^{(k)}(1)] = 0 \quad (i = 0, \ldots, n-1) \end{aligned}$$

с такими постоянными a_{ik}, b_{ik}, чтобы однородная задача

$$y^{(n)} = 0, \quad L_i y = 0 \quad (i = 0, \ldots, n-1)$$

имела лишь тривиальное решение $y = 0$; и задачу

$$\text{(1′)} \qquad \begin{aligned} &y^{(n)}(x) = f[y] \equiv f(x, y, \ldots, y^{(n-1)}) \quad (0 \leqq x \leqq 1,\ n \geqq 1,\ f \in C([0,1]\times R^n)), \\ &y(0) = \ldots = y^{(n-1)}(0) = 0. \end{aligned}$$

Задачи (1), (1′) эквивалентны в $\mathscr{M}$ уравнениям

$$\text{(1*)} \qquad y = \mathrm{A}y \equiv \int_0^1 G(x, t) f(t, y(t), \ldots, y^{(n-1)}(t))\, dt$$

где

$$\text{[i]} \qquad \begin{cases} G(x, t) \text{ — функция Грина (в случае задачи (1)),} \\ G(x, t) = \begin{cases} 0 & x \leqq t \\ \dfrac{(x-t)^{n-1}}{(n-1)!} & x > t \end{cases} \text{ (в случае задачи (1′)).} \end{cases}$$

Предположим, что в обоих случаях $f(x, u_0, \ldots, u_{n-1})$ непрерывно дифференцируема по последним n аргументам, причем

$$\text{[ii]} \qquad 0 \leqq \frac{\partial f}{\partial u_i} \leqq N \qquad (i = 0, \ldots, n-1),$$

а в случае задачи (1) пусть еще и

$$\text{[iii]} \qquad 0 \leqq \frac{\partial^i G(x, t)}{\partial x^i} \qquad (i = 0, \ldots, n-1;\ 0 \leqq x, t \leqq 1).$$

Задача (1*) имеет ровно одно решение если например выполнено следующее условие (см. [3], [4], [5])

$$\text{[iv]} \qquad N \sum_{i=0}^{n-1} \max_{x\in[0,1]} \int_0^1 \frac{\partial^i G(x, t)}{\partial x^i}\, dt \leqq \theta < 1,$$

обеспечивающее, что $\mathbf{A}$ сжато отображает $\mathcal{M}$ в $\mathcal{M}$ относительно ϱ.

Из [i], [ii], [iii] вытекает, что задача (1*) является изотонной (см. [2], [4]), поэтому учитывая и [iv] можно сконструировать (см. [4], [5]) такие $z_1, w_1 \in \mathcal{M}$ с которыми последовательности

$$z_{p+1} = \mathbf{A}z_p, \quad w_{p+1} = \mathbf{A}w_p \qquad (p = 1, 2, \ldots) \tag{d}$$

обладают свойством $z_p \overset{(B)}{\leqq} y \overset{(B)}{\leqq} w_p$ $(p=1, 2, \ldots)$ и свойством

$$w_1 \overset{(B)}{\leqq} w_2 \overset{(B)}{\leqq} \ldots, \ldots \overset{(B)}{\leqq} z_2 \overset{(B)}{\leqq} z_1;\ w_p, z_p \xrightarrow{\varrho} y,\ p \to \infty \tag{C}$$

(см. (A), (B)), где y — решение задачи (1*).

Постановка задачи

Пусть $0 < \lambda_p, \mu_p < 1$ $(p=1, 2, \ldots)$ постоянные, $Z_1 = z_1$, $W_1 = w_1$ и

$$Z_{p+1} = \lambda_p \mathbf{A}Z_p + (1-\lambda_p)\mathbf{A}W_p, \quad W_{p+1} = (1-\mu_p)\mathbf{A}Z_p + \mu_p \mathbf{A}W_p \quad (p = 1, 2, \ldots). \tag{D}$$

Возникает следующий основной вопрос: можно ли выбрать z_1, w_1 и числа λ_p, μ_p на каждом шагу так, чтобы последовательности (D) обладали свойством (C) и свойством

$$Z_p \overset{(B)}{\leqq} z_p, \quad w_p \overset{(B)}{\leqq} W_p \quad (p = 1, 2, \ldots). \tag{C_1}$$

I.

Прежде, чем формулировать основной результат для задачи (1), введем одну вспомогательную функцию

$$\varphi(x, c) = c \int_0^1 G(x, t)\, dt,$$

где G—функция Грина задачи (1).

Известно, что функция Грина задачи (1) имеет вид (см. [6] стр. 12, 30, 31)

$$G(x, t) = \begin{cases} \sum_{i=0}^{n-1} B_i(t) x^i & (0 \leqq x \leqq t \leqq 1), \\ \sum_{i=1}^{n-1} A_i(t) x^i & (0 \leqq t \leqq x \leqq 1), \end{cases}$$

где $A_i(t)$, $B_i(t)$ полиномы степени не выше $n-1$. Из условий склеивания для $G(x, t)$ при $x=t$ (см. там же) можно вывести, что

$$A_i(t) - B_i(t) = \frac{(-1)^{n-1-i}}{(n-1)!} \binom{n-1}{i} t^{n-1-i} \qquad (i = 0, \ldots, n-1)$$

следовательно

$$\sum_{i=0}^{n-1} [A_i(t)-B_i(t)]x^i = \frac{(x-t)^{n-1}}{(n-1)!}.$$

Отсюда, и из структуры G, заменяя $A_i(t)$ на $[A_i(t)-B_i(t)]+B_i(t)$ при интегрировании $G(x, t)$ по t (при любом фиксированном x) получаем

$$\varphi(x, c) = c\left(\frac{x^n}{n!} + \sum_{i=0}^{n-1} x^i \int_0^1 B_i(t)\,dt\right) \equiv c\left(\frac{x^n}{n!} + \sum_{i=0}^{n-1} r_i x^i\right)$$

где все постоянные

$$r_i = \int_0^1 B_i(t)\,dt \geqq 0$$

в силу того, что $B_i(t) \geqq 0$ согласно [iii]. Покажем, что среди чисел r_i есть отличные от нуля. Действительно, в противном случае

$$B_0(t) \equiv \ldots \equiv B_{n-1}(t) \equiv 0$$

и

$$G(x, t) = \begin{cases} 0 & (0 \leqq x \leqq t \leqq 1), \\ \dfrac{(x-t)^{n-1}}{(n-1)!} & (0 \leqq t \leqq x \leqq 1), \end{cases}$$

откуда краевые условия (см [6] стр. 12, 30, 31)

$$L_i G(x, t) = 0 \qquad (i = 0, \ldots, n-1),$$

которые должны выполняться при любом $0 \leqq t \leqq 1$ приобретут вид

$$\begin{pmatrix} b_{0,0} & \cdots & b_{0,n-1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ b_{n-1,0} & \cdots & b_{n-1,n-1} \end{pmatrix} \begin{pmatrix} \dfrac{(1-t)^{n-1}}{(n-1)!} \\ \dfrac{(1-t)^{n-2}}{(n-2)!} \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}.$$

Здесь вектор-столбец левой части есть фундаментальная система (ф. с.) для уравнения

$$y^{(n)}(t) = 0 \qquad (0 \leqq t \leqq 1),$$

следовательно из линейной независимости ф. с. все $b_{i,j} = 0$; а в этом случае задача (1) сводится к начальной задаче (1′), которую будем рассматривать в части II.

Таким образом мы показали, что в случае, если краевые условия при $x = 1$ не вырождены (т. е. если в условиях $L_i y = 0$ не все $b_{i,j}$ равны нулю), то

$$\varphi(x, c) = c\left(\frac{x^n}{n!} + r_{i_1} x^{i_1} + \ldots + r_{i_s} x^{i_s}\right)$$

где

$$r_{i_1}, \ldots, r_{i_s} > 0$$

постоянные.

Теорема 1. *Если f удовлетворяет условию*

$$\frac{\partial f}{\partial u_{i_1}}+\ldots+\frac{\partial f}{\partial u_{i_s}}>0, \tag{2}$$

тогда наш основной вопрос решается положительно.

Доказательство. Отметим сразу, что вместо условия (2) можно было бы взять положительность той же суммы на каком-нибудь компакте, охватывающем решения y и его производных.

В качестве $z_1=Z_1$, $w_1=W_1$ возьмем решения уравнений

$$z_1=\mathbf{A}z_1+\varphi, \quad w_1=\mathbf{A}w_1-\varphi \quad (z_1, w_1\in\mathscr{M}). \tag{3}$$

Легко доказать по индукции, используя изотонность $\mathbf{A}$ и равенства $Z_1=z_1$, $W_1=w_1$ что первое из свойств (C): $W_1\overset{(B)}{\leqq}W_2\overset{(B)}{\leqq}\ldots$, $\ldots\overset{(B)}{\leqq}Z_2\overset{(B)}{\leqq}Z_1$ соблюдается при любых $0<\lambda_p, \mu_p<1$ поскольку $\mathbf{A}W_p\overset{(B)}{\leqq}W_p$, $Z_p\overset{(B)}{\leqq}\mathbf{A}Z_p$ (см. [3]) и

$$Z_{p+1}\overset{(B)}{\leqq}\lambda_p\mathbf{A}Z_p+(1-\lambda_p)\mathbf{A}Z_p=\mathbf{A}Z_p,$$

$$W_{p+1}\overset{(B)}{\geqq}(1-\mu_p)\mathbf{A}W_p+\mu_p\mathbf{A}W_p=\mathbf{A}W_p \qquad (p=1, 2, \ldots),$$

при этом очевидно, что второе из свойств (C): $Z_p, W_p\xrightarrow{\varrho}y$ $(p\to\infty)$ тоже соблюдается. Нам осталось обеспечить, то, чтобы все W_p $(p\geqq 2)$ остались минорантами, а все Z_p $(p\geqq 2)$ мажорантами решения. Докажем это по индукции по p.

Покажем для этого сначала, что к любым натуральным $m_1, k_1\geqq 2$ можно выбрать λ_1, μ_1 такими, чтобы соблюдались неравенства:

$$\lambda_1\mathbf{A}Z_1+(1-\lambda_1)\mathbf{A}W_1\overset{(B)}{\geqq}\mathbf{A}^{m_1}z_1=z_{1+m_1}, \tag{3}$$

$$(1-\mu_1)\mathbf{A}Z_1+\mu_1\mathbf{A}W_1\overset{(B)}{\leqq}\mathbf{A}^{k_1}w_1=w_{1+k_1} \tag{4}$$

т. е.

$$\lambda_1(\mathbf{A}Z_1-z_{1+m_1})\overset{(B)}{\geqq}(1-\lambda_1)(z_{1+m_1}-\mathbf{A}W_1), \tag{5}$$

$$\mu_1(w_{1+k_1}-\mathbf{A}W_1)\overset{(B)}{\leqq}(1-\mu_1)(\mathbf{A}Z_1-w_{1+k_1}). \tag{6}$$

В неравенствах (5), (6) выражения в скобках слева оцениваются снизу, а скобки в правых частях сверху в смысле (B) функциями $a\varphi(x, c)$, $b\varphi(x, c)$ соответственно, где $a, b>0$ постоянные, возможно зависящие от m_1, k_1. Чтобы это показать, заметим, что из свойства (C) последовательностей (d) и сжатости $\mathbf{A}$ следуют равенства:

$$\mathbf{A}Z_1-z_{1+m_1}=\sum_{j=1}^{m_1}(\mathbf{A}^jz_1-\mathbf{A}^{j+1}z_1)=\sum_{j=1}^{m_1}(\mathbf{A}^jz_1-\mathbf{A}^jz_2), \tag{7}$$

$$w_{1+k_1}-\mathbf{A}W_1=\sum_{l=1}^{k_1}(\mathbf{A}^{l+1}w_1-A^lw_1)=\sum_{l=1}^{k_1}(\mathbf{A}^lw_2-\mathbf{A}^lw_1), \tag{8}$$

и аналогично, с помощью $\mathbf{A}^j z_i \xrightarrow{\varrho} y$, $\mathbf{A}^l w_i \xrightarrow{\varrho} y$ $(i=1, 2;\ j, l \to \infty)$

$$z_{1+m_1} - \mathbf{A}W_1 = \sum_{j=m_1}^{\infty} (\mathbf{A}^j z_1 - \mathbf{A}^j z_2) + \sum_{l=1}^{\infty} (\mathbf{A}^l w_2 - \mathbf{A}^l w_1), \tag{9}$$

$$\mathbf{A}Z_1 - w_{1+k_1} = \sum_{j=1}^{\infty} (\mathbf{A}^j z_1 - \mathbf{A}^j z_2) + \sum_{l=k_1}^{\infty} (\mathbf{A}^l w_2 - \mathbf{A}^l w_1). \tag{10}$$

Оценим теперь отдельно каждый из членов рядов (7), ..., (10). Из самого выбора z_1, w_1 получаем, что

$$z_1 - z_2 = \varphi(x, c), \quad w_2 - w_1 = \varphi(x, c) \tag{11}$$

далее

(12)
$$\mathbf{A}z_1 - \mathbf{A}z_2 = \int_0^1 G(x, t)\{f[z_1(t)] - f[z_2(t)]\}\, dt = \int_0^1 G(x, t) \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| (z_1 - z_2)^{(i)}(t)\, dt,$$

(13)
$$\mathbf{A}w_2 - \mathbf{A}w_1 = \int_0^1 G(x, t)\{f[w_2(t)] - f[w_1(t)]\}\, dt = \int_0^1 G(x, t) \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| (w_2 - w_1)^{(i)}(t)\, dt,$$

и совершенно так же получаем:

(14)
$$\mathbf{A}^j z_1 - \mathbf{A}^j z_2 =$$
$$= \int_0^1 G(x, t)\{f[z_{1+j}(t)] - f[z_{2+j}(t)]\}\, dt = \int_0^1 G(x, t) \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| (z_{1+j} - z_{2+j})^{(i)}(t)\, dt,$$

$$\mathbf{A}^l w_2 - \mathbf{A}^l w_1 = \int_0^1 G(x, t) \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| (w_{2+l} - w_{1+l})^{(i)}(t)\, dt, \tag{15}$$

где $\dfrac{\partial f}{\partial u_i}\bigg|$ как и везде в дальнейшем, означает промежуточное по формуле Лагранжа значение этой производной.

Воспользуясь сжатостью ([iv]), из (11), ..., (15) легко доказать следующие оценки сверху (напр. по индукции):

$$\mathbf{A}^j z_1 - \mathbf{A}^j z_2 \overset{(B)}{\leqq} c\theta^j \varphi(x, 1), \tag{16}$$

$$\mathbf{A}^j w_2 - \mathbf{A}^j w_1 \overset{(B)}{\leqq} c\theta^j \varphi(x, 1). \tag{17}$$

Совершенно так же можно вывести и соответствующие оценки снизу левых частей (16), (17); надо лишь использовать условие (2) теоремы и вытекающее из него и [iv] неравенства:

$$c\theta \geqq \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| (z_1 - z_2)^{(i)} = \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg| \varphi^{(i)}(t, c) \geqq c \sum_{j=1}^{s} \frac{\partial f}{\partial u_{i_j}}\bigg| (i_j)!\, r_{i_j} \geqq ch > 0, \tag{18}$$

$$(19)\quad c\theta \geqq \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\Big|(w_2-w_1)^{(i)} = \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\Big|\varphi^{(i)}(t,c) \geqq c \sum_{j=1}^{s} \frac{\partial f}{\partial u_{i_j}}\Big|(i_j)!\, r_{i_j} \geqq ch > 0.$$

Существование положительного минимума ch последних сумм, вытекает из непрерывности частных производных f и того, что точки в которых берутся значения $\dfrac{\partial f}{\partial u_i}$ находятся в компактном множестве

$$\mathscr{D} = \{(x, u_0, \ldots, u_{n-1}) | 0 \leqq x \leqq 1,\ w_1 \leqq u_0 \leqq z_1, \ldots, w_1^{(n-1)} \leqq u_{n-1} \leqq z_1^{(n-1)}\}.$$

В итоге последних рассуждений получаем оценки снизу:

$$(20)\quad \mathbf{A}^j z_1 - \mathbf{A}^j z_2 \overset{(B)}{\geqq} ch^j \varphi(x,1),$$

$$(21)\quad \mathbf{A}^j w_2 - \mathbf{A}^j w_1 \overset{(B)}{\geqq} ch^j \varphi(x,1).$$

Подставив (16), (17), (20), (21) в соответствующие формулы (7), ..., (10) получаем наконец нужные для выбора λ_1, μ_1 неравенства

$$(22)\quad \mathbf{A}Z_1 - z_{1+m_1} \overset{(B)}{\geqq} c \sum_{j=1}^{m_1} h^j \varphi(x,1) = c\varphi(x,1) h \frac{1-h^{m_1}}{1-h} = \varphi(x,c) \frac{h}{1-h}(1-h^{m_1}),$$

$$(23)\quad w_{1+k_1} - \mathbf{A}W_1 \overset{(B)}{\geqq} c \sum_{l=1}^{k_1} h^j \varphi(x,1) = c\varphi(x,1) h \frac{1-h^{k_1}}{1-h} = \varphi(x,c) \frac{h}{1-h}(1-h^{k_1}),$$

$$(24)\quad z_{1+m_1} - \mathbf{A}W_1 \overset{(B)}{\leqq} c\varphi(x,1) \frac{\theta^{m_1}}{1-\theta} + c\varphi(x,1) \frac{\theta}{1-\theta} = \varphi(x,c) \frac{\theta}{1-\theta}(1+\theta^{m_1-1}),$$

$$(25)\quad \mathbf{A}Z_1 - w_{1+k_1} \overset{(B)}{\leqq} c\varphi(x,1) \frac{\theta}{1-\theta} + c\varphi(x,1) \frac{\theta^{k_1}}{1-\theta} = \varphi(x,c) \frac{\theta}{1-\theta}(1+\theta^{k_1-1}),$$

из которых по (5), (6) можно установить, что выбрав λ_1, μ_1 из уравнений

$$(26)\quad \frac{\lambda_1}{1-\lambda_1} = \frac{1+\theta^{m_1-1}}{1-h^{m_1}} \cdot \frac{\dfrac{\theta}{1-\theta}}{\dfrac{h}{1-h}}, \qquad \frac{\mu_1}{1-\mu_1} = \frac{1+\theta^{k_1-1}}{1-h^{k_1}} \cdot \frac{\dfrac{\theta}{1-\theta}}{\dfrac{h}{1-h}}$$

полученные Z_2, W_2 останутся мажорантой соотв. минорантой решения. В частности при $0<\theta \ll 1$, $m_1 \gg 1$, $k_1 \gg 1$; $h \approx \theta$ получаем, что λ_1, μ_1 близки к 0,5.

Покажем теперь, что λ_2, μ_2 можно выбрать такими, чтобы при наперед заданных $m_2 > m_1$, $k_2 > k_1$ выполнялись неравенства аналогичные (5), (6):

$$(27)\quad \lambda_2(\mathbf{A}Z_2 - z_{1+m_2}) \overset{(B)}{\geqq} (1-\lambda_2)(z_{1+m_2} - \mathbf{A}W_2),$$

$$(28)\quad \mu_2(w_{1+k_2} - \mathbf{A}W_2) \overset{(B)}{\geqq} (1-\mu_2)(\mathbf{A}Z_2 - w_{1+k_2}).$$

Выражения левых частей в скобках, учитывая $Z_2 \overset{(B)}{\geqq} z_{1+m_1}$, $W_2 \overset{(B)}{\leqq} w_{1+k_1}$ оцени-

ваются снизу по (7), ..., (21) так же, как и выше:

$$\mathbf{A}Z_2 - z_{1+m_2} \overset{(B)}{\geqq} \mathbf{A}z_{1+m_1} - z_{1+m_2} = z_{2+m_1} - z_{1+m_2} = \sum_{j=m_1+1}^{m_2-1} \mathbf{A}^j z_1 - \mathbf{A}^j z_2$$

откуда

$$\mathbf{A}Z_2 - z_{1+m_2} \overset{(B)}{\geqq} c \sum_{j=m_1+1}^{m_2-1} h^j \varphi(x,1) = \varphi(x,c)\, h^{m_1+1} \frac{1-h^{m_2-m_1-1}}{1-h} \tag{29}$$

и аналогично

$$w_{1+k_2} - \mathbf{A}W_2 \overset{(B)}{\geqq} w_{1+k_2} - \mathbf{A}w_{1+k_1} = w_{1+k_2} - w_{2+k_1} =$$

$$= \mathbf{A}^{k_2-1} w_2 - \mathbf{A}^{k_1+1} w_1 = \sum_{j=k_1+1}^{k_2-1} \mathbf{A}^j w_2 - \mathbf{A}^j w_1,$$

откуда

$$w_{1+k_2} - \mathbf{A}W_2 \overset{(B)}{\geqq} c \sum_{j=k_1+1}^{k_2-1} h^j \varphi(x,1) = \varphi(x,c)\, h^{k_1+1} \frac{1-h^{k_2-k_1-1}}{1-h}. \tag{30}$$

Оценки сверху скобок правых частей (27), (28) по (7), ..., (21) используя и $\mathbf{A}W_2 \overset{(B)}{\geqq} w_3$, $\mathbf{A}Z_2 \overset{(B)}{\geqq} z_3$ получаются следующим образом.

Аналогично (24), (25) выводим неравенства

$$z_{1+m_2} - \mathbf{A}W_2 \overset{(B)}{\geqq} z_{1+m_2} - w_3 = \sum_{j=m_2}^{\infty} \mathbf{A}^j z_1 - \mathbf{A}^j z_2 + \sum_{j=2}^{\infty} \mathbf{A}^j w_2 - \mathbf{A}^j w_1,$$

т. е.

$$z_{1+m_2} - \mathbf{A}W_2 \overset{(B)}{\geqq} c \sum_{j=m_2}^{\infty} \theta^j \varphi(x,1) + c \sum_{j=2}^{\infty} \theta^j \varphi(x,1) = \varphi(x,c) \frac{\theta^{m_2} + \theta^2}{1-\theta} \tag{31}$$

и соответственно

$$\mathbf{A}Z_2 - w_{1+k_2} \overset{(B)}{\geqq} z_3 - w_{1+k_2} = \sum_{j=2}^{\infty} \mathbf{A}^j z_1 - \mathbf{A}^j z_2 + \sum_{j=k_2}^{\infty} \mathbf{A}^j w_2 - \mathbf{A}^j w_1,$$

т. е.

$$\mathbf{A}Z_2 - w_{1+k_2} \overset{(B)}{\geqq} c \sum_{j=2}^{\infty} \theta^j \varphi(x,1) + c \sum_{j=k_2}^{\infty} \theta^j \varphi(x,1) = \varphi(x,c) \frac{\theta^2 + \theta^{k_2}}{1-\theta}. \tag{32}$$

В итоге, в качестве подходящих λ_2, μ_2 можно взять решения уравнений:

$$\frac{\lambda_2}{1-\lambda_2} = \frac{(\theta^{m_2}+\theta^2)(1-h)}{(h^{m_1+1}-h^{m_2})(1-\theta)}, \quad \frac{\mu_2}{1-\mu_2} = \frac{(\theta^{k_2}+\theta^2)(1-h)}{(h^{k_1+1}-h^{k_2})(1-\theta)}. \tag{33}$$

Отметим, что получающиеся отсюда λ_2, μ_2 вообще говоря не будут уже близки к 0,5.

Покажем наконец, что если $\lambda_1, \ldots, \lambda_{p-1}$; $\mu_1, \ldots, \mu_{p-1}$ выбраны подходящим образом, при соответствующих заданных $m_1, \ldots, m_{p-1}$; $k_1, \ldots, k_{p-1}$ то к любым $m_p > m_{p-1}$, $k_p > k_{p-1}$ можно подобрать такие λ_p, μ_p чтобы Z_{p+1}, W_{p+1} образовали мажорантно-минорантную пару решения, для чего мы потребуем, чтобы выполнялись неравенства

$$\lambda_p(\mathbf{A}Z_p - z_{1+m_p}) \overset{(B)}{\geqq} (1-\lambda_p)(z_{1+m_p} - \mathbf{A}W_p), \tag{34}$$

$$\mu_p(w_{1+k_p} - \mathbf{A}W_p) \overset{(B)}{\geqq} (1-\mu_p)(\mathbf{A}Z_p - w_{1+k_p}). \tag{35}$$

Из индуктивного предположения мы знаем, что

$$Z_p \overset{(B)}{\geqq} z_{1+m_{p-1}}, \quad W_p \overset{(B)}{\leqq} w_{1+k_{p-1}}$$

кроме того очевидно, что при любых $0<\lambda_p, \mu_p<1$: $Z_p \overset{(B)}{\geqq} z_p$, $w_p \overset{(B)}{\leqq} W_p$. Следовательно скобки левых частей (34), (35) можно оценить снизу так:

$$\mathbf{A}Z_p - z_{1+m_p} \overset{(B)}{\geqq} z_{2+m_{p-1}} - z_{1+m_p} \overset{(B)}{\geqq} \varphi(x, c) h^{m_{p-1}+1} \frac{1-h^{m_p-m_{p-1}-1}}{1-h}, \tag{36}$$

$$w_{1+k_p} - \mathbf{A}W_p \overset{(B)}{\geqq} w_{1+k_p} - w_{2+k_{p-1}} \overset{(B)}{\geqq} \varphi(x, c) h^{k_{p-1}+1} \frac{1-h^{k_p-k_{p-1}-1}}{1-h}. \tag{37}$$

Скобки правых частей (34), (35) будем оценивать сверху, используя индуктивные предположения $\mathbf{A}W_p \overset{(B)}{\geqq} w_{p+1}$, $\mathbf{A}Z_p \overset{(B)}{\leqq} z_{p+1}$ таким образом:

$$z_{1+m_p} - \mathbf{A}W_p \overset{(B)}{\leqq} z_{1+m_p} - w_{p+1} \overset{(B)}{\leqq} \varphi(x, c) \frac{\theta^{m_p}+\theta^p}{1-\theta}, \tag{38}$$

$$\mathbf{A}Z_p - w_{1+k_p} \overset{(B)}{\leqq} z_{p+1} - w_{1+k_p} \overset{(B)}{\leqq} \varphi(x, c) \frac{\theta^p+\theta^{k_p}}{1-\theta}. \tag{39}$$

Возвращаясь к (34), (35) устанавливаем, что λ_p, μ_p можно выбрать нужным образом, например как решения уравнений:

$$\frac{\lambda_p}{1-\lambda_p} = \frac{(\theta^{m_p}+\theta^p)(1-h)}{(h^{m_{p-1}+1}-h^{m_p})(1-\theta)}, \quad \frac{\mu_p}{1-\mu_p} = \frac{(\theta^{k_p}+\theta^p)(1-h)}{(h^{k_{p-1}+1}-h^{k_p})(1-\theta)} \tag{40}$$

и при таком выборе будем иметь $Z_{p+1} \overset{(B)}{\geqq} z_{1+m_p}$, $W_{p+1} \overset{(B)}{\leqq} w_{1+k_p}$ и значит (Z_{p+1}, W_{p+1}) мажорантно-минорантная пара решения, причем поскольку $0<\lambda_p, \mu_p<1$ при всех p, то эта пара дает лучшие приближения, чем исходная пара (z_{p+1}, w_{p+1}) в том смысле, что $Z_{p+1} \overset{(B)}{\geqq} z_{p+1}$, $w_{p+1} \overset{(B)}{\leqq} W_{p+1}$. Теорема доказана.

Замечание 1. Для простоты выкладок мы ввели z_1, w_1 как решения (точные решения) задач

$$z = \mathbf{A}z + \varphi(x, c), \quad w = \mathbf{A}w - \varphi(x, c), \tag{41}$$

решить которые вообще говоря не легче, чем исходную задачу $y=\mathbf{A}y$. Однако, приведенная выше конструкция мало изменится если в качестве z_1, w_1 возьмем такие приближения решений задач (41): $z_1, w_1 \in \mathscr{M}$ при которых

$$z_1 = \mathbf{A}z_1 + \tilde{\varphi}, \quad w_1 = \mathbf{A}w_1 - \tilde{\varphi} \qquad (a\varphi \overset{(B)}{\leqq} \tilde{\varphi} \overset{(B)}{\leqq} b\varphi), \tag{42}$$

где $0<a<b$ постоянные (см. Замечание 3). Во всех оценках снизу вместо $\varphi(x, c)$ будет фигурировать функция $a\varphi(x, c)$, а в оценках сверху $b\varphi(x, c)$, и тем самым исходя из этих z_1, w_1 тоже можно найти все λ_p, μ_p так, чтобы $\{Z_p\}, \{W_p\}$ обладали нужными свойствами. Изменятся лишь уравнения (40), где в правых частях появится множитель $\frac{b}{a}$.

Замечание 2. Если в теореме 1, z_1, w_1 взять как «неравноотстоящие» от y элементы $\mathscr{M}$, т. е.

$$z_1 = \mathbf{A}z_1 + \varphi(x, c), \quad w_1 = \mathbf{A}w_1 - \varphi(x, d) \quad c \neq d; \quad c, d > 0,$$

то в оценке левой части (35) появится $\varphi(x, d)$ вместо $\varphi(x, c)$ а правые части (34), (35) оцениваются выражениями вида $C_1\varphi(x, c) + C_2\varphi(x, d)$, т. е. в конечном счете и в этом случае в правых частях (40) появится некоторых новый множитель: некоторая постоянная C. Существование подходящих λ_p, μ_p следовательно и в этом случае доказано.

Замечание 3. Для построения z_1, w_1 удовлетворяющих (42), можно указать простую процедуру. Воспользуемся для этого техникой невязок (см. напр. [3], [4]), т. е. вместо интегральных форм неравенств выше перейдем к дифференциальным неравенствам.

Пусть z произвольный элемент из $\mathscr{M} \cap C^n$. Тогда невязка

$$\alpha(x) = z^{(n)}(x) - f[z(x)] \qquad (0 \leqq x \leqq 1)$$

непрерывна на [0, 1], следовательно

$$A_1 \leqq \alpha(x) \leqq A_2 \qquad (0 \leqq x \leqq 1),$$

где A_1, A_2 некоторые постоянные (допускаем и случай $A_1 < 0$). Покажем, что постоянную $c > 0$ можно выбрать настолько большой, чтобы имели

$$(43) \qquad 0 < a \leqq \big(z + \varphi(x, c)\big)^{(n)} - f[z + \varphi(x, c)] \equiv \tilde{\alpha}(x) \leqq b \qquad (0 \leqq x \leqq 1),$$

а это эквивалентно тому (см. [3], [4]), что первое уравнение из (42) выполняется для $z_1 = z + \varphi(x, c)$. Существование упомянутого только что $c > 0$ вытекает из условия сжатости [iv] и того, что применяя формулу Лагранжа для разности $f[z+\varphi] - f[z]$ в

$$\tilde{\alpha} = \big(z + \varphi(x, c)\big)^{(n)} - f[z + \varphi(x, c)] = \big(z + \varphi(x, c)\big)^{(n)} - \big(f[z + \varphi(x, c)] - f[z]\big) - f[z]$$

$$\tilde{\alpha} = \alpha + \varphi^{(n)} - \sum_{i=0}^{n-1} \frac{\partial f}{\partial u_i}\bigg|\varphi^{(i)}(x, c) \geqq \alpha(x) + c - c\theta =$$

$$= \alpha(x) + c(1-\theta) \geqq A_1 + c(1-\theta)$$

и

$$\tilde{\alpha}(x) \leqq \alpha(x) + c \qquad (0 \leqq x \leqq 1).$$

Конструкция w_1 для (42) совершенно такая же. Из двух полученных c нам осталось взять бóльшую.

Замечание 4. Пусть в конструкции теоремы $c > 0$ маленькое число $(\varphi(x, c) = c\varphi(x, 1))$. Расписав $\mathbf{A}z_1 - \mathbf{A}y$, $\mathbf{A}y - \mathbf{A}w_1$ подобно (12), (13) легко показать, что

$$h(z_1 - y) \overset{(B)}{\leqq} \mathbf{A}z_1 - \mathbf{A}y \overset{(B)}{\leqq} \theta(z_1 - y), \quad h(y - w_1) \overset{(B)}{\leqq} \mathbf{A}y - \mathbf{A}w_1 \overset{(B)}{\leqq} \theta(y - w_1),$$

а поскольку

$$z_1 - y = \mathbf{A}z_1 - \mathbf{A}y + \varphi(x, c), \quad y - w_1 = \mathbf{A}y - \mathbf{A}w_1 + \varphi(x, c)$$

то

$$O(c)=\frac{\varphi(x,\,c)}{1-h}\overset{(B)}{\leqq}z_1-y,\quad y-w_1\overset{(B)}{\leqq}\frac{\varphi(x,\,c)}{1-\theta}=O(c).$$

Если мы теперь захотим, чтобы Z_2-y, $y-W_2$ были величинами $O(c^2)$ то — в частном случае

$$f[y]\equiv\sum_{i=0}^{n-1}a_i(x)\,y^{(i)}+f_1(x)$$

в качестве λ_1 мы должны взять $\lambda_1=0{,}5$ (если искать λ_1 среди положительных $\lambda_1<1$), поскольку

$$Z_2-y=\lambda_1(z_1-\varphi)+(1-\lambda_1)(w_1+\varphi)-y=\varphi(1-2\lambda_1)+\lambda_1(z_1-y)+(1-\lambda_1)(w_1-y).$$

Аналогично получается, что μ_1 должно равняться 0,5.

Если здесь z_1, w_1 заменим на приближенные решения задач (3), то уже нелегко выяснить: можно ли $0<\lambda_1, \mu_1<1$ выбрать так, чтобы Z_2-y, $y-W_2$ были величинами порядка $O(c^2)$, тем более если захотим при этом еще и сохранить мажорантно-минорантное свойство Z_2, W_2; не говоря уже об общем случае нелинейной f.

II.

Рассмотрим теперь начальную задачу (1′). Согласно сказанному в самом начале, в этой части мы будем применять дифференциальные неравенства. В приведенной ниже теореме 1′ мы покажем лишь то, что $z_1, w_1, \lambda_p, \mu_p$ можно выбрать так, чтобы последовательности (D) обладали свойствами (C), (C_1). После доказательства теоремы укажем, как можно уточнить полученные результаты.

Замечание 1′. Применение метода части I для задачи (1′) вообще говоря невозможно, потому что вспомогательная функция $\varphi(x, 1)$ в этом случае равна $\frac{x^n}{n!}$. Если все же попытаться проделать оценки выражений

$$\mathbf{A}^j z_1-\mathbf{A}^j z_2,\quad \mathbf{A}^j w_2-\mathbf{A}^j w_1$$

подобно тому, как это в части I сделали, то в отличие от I здесь на каждом шагу появляется множитель x, который усложняет вид оценок (особенно неудобно с нижними оценками), может привести к тому, что производные высокого порядка ($j\gg 1$) станут большими. Сравнить нужные величины на аналогию I все же возможно, но только отдельно в двух интервалах: в $[0, \delta_p]$ и в $[\delta_p, 1]$ если δ_p достаточно мало; однако это приводит к необозримым формулам для выбора λ_p, μ_p.

Введем одно обозначение: пусть

$$\varrho_x(u,\,v)\equiv\sum_{i=0}^{n-1}|u^{(i)}(x)-v^{(i)}(x)|.$$

Лемма 1′. *Если выполнены следующие два условия:*

$$z(x)\in C^n[0,\,1]\cap\mathcal{M},\quad \alpha(x)\equiv z^{(n)}(x)-f[z(x)]\geqq 0\quad (0\leqq x\leqq 1),$$

тогда

$$z(x) \overset{(B)}{\geqq} y(x).$$

Аналогично, из того, что

$$w(x) \in C^n[0,1] \cap \mathscr{M}, \quad \beta(x) \equiv w^{(n)}(x) - f[w(x)] \leqq 0 \quad (0 \leqq x \leqq 1),$$

следует

$$w(x) \overset{(B)}{\geqq} y(x).$$

Более того, если при этом, с некоторым $p \geqq 1$ *натуральным*

$$mt^{p-1} \leqq \alpha(t) \leqq Mt^{p-1} \quad (0 \leqq t \leqq 1;\ m, M > 0 \text{ const.}),$$

то

$$m\frac{x^p}{p} \leqq \varrho_x(z, y) \leqq Mx^p \frac{\theta}{N(1-\theta)} \quad (0 \leqq x \leqq 1);$$

аналогично, из

$$Kt^{p-1} \leqq \beta(t) \leqq kt^{p-1} \quad (0 \leqq t \leqq 1;\ k, K < 0 \text{ const.})$$

следуют неравенства:

$$|k|\frac{x^p}{p} \leqq \varrho_x(y, w) \leqq |K| x^p \frac{\theta}{N(1-\theta)} \quad (0 \leqq x \leqq 1).$$

Доказательство. Из определения невязок α, β получаем

$$\int_0^1 G(x,t) z^{(n)}(t)\,dt = \int_0^1 G(x,t) f[z(t)]\,dt + \int_0^1 G(x,t)\alpha(t)\,dt, \tag{2'}$$

$$\int_0^1 G(x,t) w^{(n)}(t)\,dt = \int_0^1 G(x,t) f[w(t)]\,dt + \int_0^1 G(x,t)\beta(t)\,dt. \tag{3'}$$

Ограничимся пока (2′). В силу того, что $z \in \mathscr{M} \cap C^n$ левая часть (2′) есть $z(x)$, и значит

$$z(x) = \mathbf{A}z + \int_0^1 G(x,t)\alpha(t)\,dt \tag{4'}$$

откуда в силу неотрицательности G и ее производных, учитывая условие $\alpha \geqq 0$ получаем

$$z(x) - \mathbf{A}z = \int_0^1 G(x,t)\alpha(t)\,dt \overset{(B)}{\geqq} 0, \tag{5'}$$

откуда в силу того, что $\mathbf{A}^p z \searrow y$ при $p \to \infty$, получаем $z \overset{(B)}{\geqq} y$ (см. [3], [4]). Поскольку

$$y(x) - \mathbf{A}y = 0, \tag{6'}$$

то вычитав (6′) из (5′), по формуле Лагранжа получаем

$$(7') \quad z(x)-y(x)-\int_0^1 G(x,t)\sum_{i=0}^{n-1}\left.\frac{\partial f}{\partial u_i}\right|_{P(t)}\left(z^{(i)}(t)-y^{(i)}(t)\right)dt=\int_0^1 G(x,t)\alpha(t)\,dt,$$

откуда

$$(8') \quad z^{(j)}(x)-y^{(j)}(x)=\int_0^1 G^{(j)}(x,t)\sum_{i=0}^{n-1}\left.\frac{\partial f}{\partial u_i}\right|_{P(t)}\left(z^{(i)}(t)-y^{(i)}(t)\right)dt+\int_0^1 G^{(j)}(x,t)\alpha(t)\,dt$$

при всех $j=0, 1, \ldots, n-1$; $0\leqq x\leqq 1$.

Учитывая вид $G(x,t)$; $z-y\overset{(B)}{\geqq}0$, $\alpha\geqq 0$ отсюда вытекает, что $z^{(j)}(x)-y^{(j)}(x)$ не убывает по x. Используя это и вид $G(x,t)$ после просуммирования (8′) по всем j получаем:

$$(9') \quad \varrho_x(z,y)\leqq\left(1-N\sum_{i=1}^{n}\frac{x^i}{i!}\right)^{-1}\int_0^1\sum_{j=0}^{n-1}G^{(j)}(x,t)\alpha(t)\,dt\leqq$$

$$\leqq\max_{[0,x]}\alpha(t)\sum_{i=1}^{n}\frac{x^i}{i!}\left(1-N\sum_{i=1}^{n}\frac{x^i}{i!}\right)^{-1},$$

откуда

$$\varrho_x(z,y)\leqq\frac{1}{1-\theta}Mx^{p-1}x\sum_{i=1}^{n}\frac{x^{i-1}}{i!}\leqq\frac{\theta M}{N(1-\theta)}x^p\quad(0\leqq x\leqq 1).$$

С другой стороны из (8′) получаем

$$\sum_{j=0}^{n-1}z^{(j)}(x)-y^{(j)}(x)=\int_0^1\left(\sum_{j=0}^{n-1}G^{(j)}(x,t)\right)\sum_{i=0}^{n-1}\left.\frac{\partial f}{\partial u_i}\right|_{P(t)}\left(z^{(i)}(t)-y^{(i)}(t)\right)dt+$$

$$+\int_0^1\sum_{j=0}^{n-1}G^{(j)}(x,t)\alpha(t)\,dt,$$

следовательно

$$(10') \quad \varrho_x(z,y)\geqq\int_0^1\sum_{j=0}^{n-1}G^{(j)}(x,t)\alpha(t)\,dt=\int_0^x\sum_{i=0}^{n-1}\frac{(x-t)^i}{i!}\alpha(t)\,dt\geqq$$

$$\geqq\int_0^x mt^{p-1}\,dt=m\frac{x^p}{p}\qquad(0\leqq x\leqq 1).$$

Займемся наконец оценками для w. Из (3′) получаем

$$(11') \quad w(x)=\mathbf{A}w+\int_0^1 G(x,t)\beta(t)\,dt,$$

а значит используя $G^{(i)}(x,t)\geqq 0$, $\beta\leqq 0$ из (11′) вытекает

$$(12') \quad w(x)\overset{(B)}{\geqq}\mathbf{A}w$$

откуда, в силу того, что $\mathbf{A}^p w\nearrow y$ при $p\to\infty$, получаем $w\overset{(B)}{\leqq}y$ (см. [3], [4]).

Вычитав теперь из (6′) равенство (11′), после применения формулы Лагранжа получаем

$$(13') \quad y(x)-w(x)=\int_0^1 G(x,t)\sum_{i=0}^{n-1}\left.\frac{\partial f}{\partial u_i}\right|_{Q(t)}\left(y^{(i)}(t)-w^{(i)}(t)\right)dt-\int_0^1 G(x,t)\beta(t)\,dt,$$

откуда все получается как выше для $z-y$. Лемма доказана.

Наша цель показать (по индукции), что Z_1, W_1 можно подобрать так, чтобы при некоторых $0<\lambda_p, \mu_p<1$ все невязки

$$\alpha_p \equiv Z_p^{(n)}(x)-f[Z_p(x)], \quad \beta_p \equiv W_p^{(n)}(x)-f[W_p(x)] \qquad (0 \leqq x \leqq 1)$$

были неотрицательными соотв. неположительными. Поэтому очень важно иметь удобные формулы для выражения $\alpha_{p+1}, \beta_{p+1}$ через α_p, β_p. Выпишем сначала эти выражения для трех частных случаев f, затем перейдем к двум леммам и теореме 1′ касающихся уже случая общего f.

Замечание 2′. Если

$$f[y] \equiv \sum_{i=0}^{n-1} a_i(x)\,y^{(i)}+a(x) \equiv L_x y,$$

тогда

$$\alpha_{p+1}(x)=\int_0^1 L_x(G)\{\lambda_p\alpha_p(t)+(1-\lambda_p)\beta_p(t)\}\,dt$$

$$\beta_{p+1}(x)=\int_0^1 L_x(G)\{(1-\mu_p)\alpha_p(t)+\mu_p\beta_p(t)\}\,dt \qquad (0 \leqq x \leqq 1).$$

Замечание 3′. Если f выпуклая в том смысле, что при всех $0 \leqq \lambda \leqq 1$; $z, w \in C^n[0,1]$ имеет место неравенство

$$f[\lambda z+(1-\lambda)w] \leqq \lambda f[z]+(1-\lambda)f[w],$$

то

(14′)

$$\alpha_{p+1}(x) \geqq \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\left[\left.\frac{\partial f}{\partial u_i}\right|_{P(x)}\lambda_p\alpha_p(t)+\left.\frac{\partial f}{\partial u_i}\right|_{Q(x)}(1-\lambda_p)\beta_p(t)\right]dt \quad (0 \leqq x \leqq 1),$$

где $P(x), Q(x)$ точки $(n+1)$-мерного пространства $(x, \ldots)$ последние n координаты заключены между $(\mathbf{A}Z_p)^{(i)}(x)$ и $Z_p^{(i)}(x)$; $W_p^{(i)}(x)$ и $(\mathbf{A}W_p)^{(i)}(x)$ при каждом x; $i=0, \ldots, n-1$ соответственно.

Замечание 4′. В случае вогнутой f аналогично

$$(15') \quad \beta_{p+1}(x) \leqq \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\left[\left.\frac{\partial f}{\partial u_i}\right|_{P(x)}(1-\mu_p)\alpha_p(t)+\left.\frac{\partial f}{\partial u_i}\right|_{Q(x)}\mu_p\beta_p(t)\right]dt,$$

с точно теми же ограничениями на $P(x), Q(x)$ что и выше.

Перед формулировкой Теоремы 1′ докажем еще две леммы. Предположим при этом, что всюду ниже выполнено условие

$$\frac{\partial f}{\partial u_{n-1}} \geqq \theta_1 > 0 \tag{16′}$$

где конечно $\theta_1 < N < \theta$, ведь в силу [iv]

$$N\left(\frac{1}{n!}+\ldots+1\right) = \theta.$$

Лемма 2′. *Если существуют такие постоянные* $m_p, M_p > 0$, *с которыми*

$$m_p t^{p-1} \leqq \alpha_p(t) \leqq M_p t^{p-1} \qquad (0 \leqq t \leqq 1), \tag{17′}$$

тогда

$$\frac{\theta_1}{p} m_p x^p \leqq \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\alpha_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(x)} dt \leqq \theta M_p x^p \qquad (0 \leqq x \leqq 1), \tag{18′}$$

и аналогично, если при каких-нибудь постоянных $k_p, K_p < 0$

$$K_p t^{p-1} \leqq \beta_p(t) \leqq k_p t^{p-1} \qquad (0 \leqq t \leqq 1), \tag{19′}$$

то

$$\theta K_p x^p \leqq \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\beta_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(x)} dt \leqq \frac{\theta_1}{p} k_p x^p \qquad (0 \leqq x \leqq 1), \tag{20′}$$

где нам безразлично, в какой точке берутся частные производные f, *важно лишь то, что эти точки не зависят от* t; *от* x *могут зависеть.*

Доказательство. Поскольку на отрезке $[0, x]$ невязка $\alpha_p(t) \leqq M_p x^{p-1}$ то используя в (18′) явный вид G получаем:

$$\mathscr{I} \equiv \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\alpha_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(x)} dt \leqq N \int_0^x M_p x^{p-1}\left(\frac{(x-t)^{n-1}}{(n-1)!}+\ldots+1\right) dt.$$

После интегрирования, используя условие сжатости [iv] и то, что в случае задачи (1′) в [iv] левая часть достигает максимума при $x=1$, получим

$$\mathscr{I} \leqq M_p x^{p-1} N\left(\frac{x^n}{n!}+\ldots+x\right) \leqq M_p x^p N\left(\frac{1}{n!}+\ldots+1\right) = M_p x^p \cdot \theta.$$

Оценим теперь снизу $\mathscr{I}$, причем таким образом, что на отрезке $[0, x]$ выражение $\sum_{i=0}^{n-1} G^{(i)}(x,t)$ заменяем на единицу, а $\frac{\partial f}{\partial u_{n-1}}$ заменяем на θ_1; вместо остальных производных f пишем нуль а вместо $\alpha_p(t)$ напишем $m_p t^{p-1}$. Таким образом получаем

$$\mathscr{I} \geqq \theta_1 \int_0^x m_p t^{p-1} dt = \theta_1 \frac{m_p}{p} x^p \quad \text{ч.т.д.}$$

Неравенства (20′) доказываются совершенно так же.

Лемма 3′. *При условиях Леммы 2′ имеют место оценки:*

$$\theta_1(|k_p|+m_p)\frac{x^p}{p} \leqq f[Z_p(x)]-f[W_p(x)] \leqq \frac{\theta}{1-\theta}(M_p+|K_p|)x^p.$$

Доказательство. Применяя формулу Лагранжа имеем

$$f[Z_p]-f[W_p] = \sum_{i=0}^{n-1}\frac{\partial f}{\partial u_i}\Big|[Z_p^{(i)}-y^{(i)}+y^{(i)}-W_p^{(i)}],$$

откуда в силу

$$0 \leqq \frac{\partial f}{\partial u_i} \leqq N \qquad (i=0,\ldots,n-1)$$

и Леммы 1′ получаем доказываемую верхнюю оценку.

Что касается нижней оценки, то заметим, что из той же формулы Лагранжа

$$f[Z_p]-f[W_p] \geqq \frac{\partial f}{\partial u_{n-1}}\Big|[Z_p^{(n-1)}-y^{(n-1)}+y^{(n-1)}-W_p^{(n-1)}]$$

а поскольку из неравенств (8′) при $j=n-1$

$$Z_p^{(n-1)}(x)-y^{(n-1)}(x) \geqq \int_0^x G^{(n-1)}(x,t)\alpha_p(t)\,dt \geqq \int_0^x m_p t^{p-1}\,dt = m_p\frac{x^p}{p}$$

и аналогично

$$y^{(n-1)}(x)-W_p^{(n-1)}(x) \geqq |k_p|\frac{x^p}{p},$$

то получаем в силу $\dfrac{\partial f}{\partial u_{n-1}} \geqq \theta_1$ требуемую оценку:

$$f[Z_p(x)]-f[W_p(x)] \geqq \theta_1(|k_p|+m_p)\frac{x^p}{p}.$$

Теорема 1′. *В случае задачи* (1′), *при выполнении условия* (16′), *можно выбрать* z_1, w_1 *таким образом, что при некоторых* $0<\lambda_p, \mu_p<1$ *последовательности* (D) *будут обладать свойствами* (C) *и* (C_1). *Более того* z_1, w_1 *можно найти в явном виде, а в качестве* λ_p, μ_p *можно взять любое решение* $0<\lambda_p<1$, $0<\mu_p<1$ *систем неравенств* (36′), (37′) *соотв.* (38′), (39′).

Доказательство. Явная конструкция $z_1=Z_1$, $w_1=W_1$ дается в Замечании 3; для построенных там Z_1, W_1 выполняются неравенства

$$m_1 \leqq \alpha_1(x) \leqq M_1, \quad K_1 \leqq \beta_1(x) \leqq k_1 \tag{21′}$$

$$(0 \leqq x \leqq 1;\ K_1, k_1 < 0 < m_1, M_1 \text{ постоянные}).$$

Предположим, что уже доказано, то, что существуют такие $\lambda_1, \mu_1, \ldots$ $\ldots, \lambda_{p-1}, \mu_{p-1}$ с которыми при всех $j=1, 2, \ldots, p$

$$m_j x^{j-1} \leqq \alpha_j(x) \leqq M_j x^{j-1}, \quad K_j x^{j-1} \leqq \beta_j(x) \leqq k_j x^{j-1} \quad (0 \leqq x \leqq 1), \tag{22′}$$

где $K_j, k_j<0<m_j, M_j$ постоянные.

Покажем, что из последних из неравенств (22′):

(23′) $m_p x^{p-1} \leqq \alpha_p(x) \leqq M_p x^{p-1}, \quad K_p x^{p-1} \leqq \beta_p(x) \leqq k_p x^{p-1} \quad (0 \leqq x \leqq 1),$

$K_p, k_p < 0 < m_p, M_p$ постоянные,

вытекает, что существуют такие λ_p, μ_p при которых выполяются неравенства

(24′) $m_{p+1} x^p \leqq \alpha_{p+1}(x) \leqq M_{p+1} x^p, \quad K_{p+1} x^p \leqq \beta_{p+1}(x) \leqq k_{p+1} x^p \quad (0 \leqq x \leqq 1)$

с некоторыми постоянными $K_{p+1}, k_{p+1} < 0 < m_{p+1}, M_{p+1}$.

Для этого, и для дальнейших рассуждений, вычислим

$$\alpha_{p+1}(x) = Z_{p+1}^{(n)}(x) - f[Z_{p+1}(x)]$$

подставив здесь — в силу определения последовательностей (D)

$$Z_{p+1} = \lambda_p \mathbf{A} Z_p + (1-\lambda_p)\mathbf{A} W_p$$

в аргумент f, а в силу равенства

$$(\mathbf{A}z)^{(n)} = f[z] \quad (z \in \mathscr{M})$$

первый член α_{p+1} заменяем на

$$Z_{p+1}^{(n)}(x) = \lambda_p f[Z_p(x)] + (1-\lambda_p) f[W_p(x)].$$

Таким образом

$$\alpha_{p+1} = \lambda_p f[Z_p] + (1-\lambda_p) f[W_p] - f[\lambda_p \mathbf{A} Z_p + (1-\lambda_p)\mathbf{A} W_p]$$

откуда разбив вычитаемое на две части

$$\alpha_{p+1} = \lambda_p \big(f[Z_p] - f[\lambda_p \mathbf{A} Z_p + (1-\lambda_p)\mathbf{A} W_p]\big) + (1-\lambda_p)\big(f[W_p] - f[\lambda_p \mathbf{A} Z_p + (1-\lambda_p)\mathbf{A} W_p]\big).$$

После применения формулы Лагранжа к двум разностям

$$\alpha_{p+1}(x) = \lambda_p \sum_{i=0}^{n-1} \left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1}, Z_p)(x)} \{Z_p^{(i)}(x) - [\lambda_p (\mathbf{A} Z_p)^{(i)}(x) + (1-\lambda_p)(\mathbf{A} W_p)^{(i)}(x)]\} +$$

(25′)

$$+ (1-\lambda_p) \sum_{i=0}^{n-1} \left.\frac{\partial f}{\partial u_i}\right|_{(W_p, Z_{p+1})(x)} \{W_p^{(i)}(x) - [\lambda_p (\mathbf{A} Z_p)^{(i)}(x) + (1-\lambda_p)(\mathbf{A} W_p)^{(i)}(x)]\},$$

где $\dfrac{\partial f}{\partial u_i}$ берутся в некоторых точках $(n+1)$-мерного пространства $(x, \ldots)$ с последними n координатами между $Z_{p+1}^{(j)}(x)$ и $Z_p^{(j)}(x)$, $W_p^{(j)}(x)$ и $Z_{p+1}^{(j)}(x)$ $(j=0, \ldots, n-1)$ соответственно. Порядок записи Z_{p+1}, Z_p и W_p, Z_{p+1} при любых $0 < \lambda_p, \mu_p < 1$ правилен, поскольку (используя индукцию)

$$Z_{p+1} \overset{(B)}{\leqq} \lambda_p \mathbf{A} Z_p + (1-\lambda_p)\mathbf{A} Z_p = \mathbf{A} Z_p \overset{(B)}{\leqq} Z_p$$

и так же

$$W_p \overset{(B)}{\leqq} \mathbf{A} W_p = (1-\mu_p)\mathbf{A} W_p + \mu_p \mathbf{A} W_p \overset{(B)}{\leqq} Z_{p+1}.$$

Разобъем теперь $Z_p^{(i)}, W_p^{(i)}$ в фигурных скобках (25′) на две части с коэффициентами $\lambda_p, (1-\lambda_p)$. Таким образом после прибавления и вычитания

некоторых членов, первая фигурная скобка примет вид:

$$\lambda_p[Z_p^{(i)}-(\mathbf{A}Z_p)^{(i)}]+(1-\lambda_p)[Z_p^{(i)}-(\mathbf{A}Z_p)^{(i)}+(\mathbf{A}Z_p)^{(i)}-(\mathbf{A}W_p)^{(i)}], \tag{26'}$$

а вторая фигурная скобка (25′) преобразуется к виду:

$$\lambda_p[(\mathbf{A}Z_p)^{(i)}-(\mathbf{A}W_p)^{(i)}+(\mathbf{A}W_p)^{(i)}-W_p^{(i)}]+(1-\lambda_p)[(\mathbf{A}W_p)^{(i)}-W_p^{(i)}]. \tag{27'}$$

Подставим эти выражения в (25′) и заметим, что

$$Z_p^{(i)}(x)-(\mathbf{A}Z_p)^{(i)}(x)=\int_0^1 G^{(i)}(x,t)\big(Z_p^{(n)}(t)-f[Z_p(t)]\big)\,dt=\int_0^1 G^{(i)}(x,t)\alpha_p(t)\,dt, \tag{28'}$$

$$W_p^{(i)}(x)-(\mathbf{A}W_p)^{(i)}(x)=\int_0^1 G^{(i)}(x,t)\beta_p(t)\,dt, \tag{29'}$$

$$(\mathbf{A}Z_p)^{(i)}(x)-(\mathbf{A}W_p)^{(i)}(x)=\int_0^1 G^{(i)}(x,t)\big(f[Z_p(t)]-f[W_p(t)]\big)\,dt, \tag{30'}$$

и с помощью последних, α_{p+1} преобразуется к виду:

$$\alpha_{p+1}(x)=\int_0^1\sum_{i=0}^{n-1}G^{(i)}(x,t)\left\{\lambda_p\alpha_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}+(1-\lambda_p)\beta_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,Z_{p+1})(x)}+\right. \tag{31'}$$

$$\left.+\lambda_p(1-\lambda_p)\left(\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}-\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,Z_{p+1})(x)}\right)\big(f[Z_p(t)]-f[W_p(t)]\big)\right\}dt.$$

Для дальнейшего нам удобнее будет, когда частные производные при α_p, β_p взяты в общих точках, поэтому перепишем (31′) в виду

$$\alpha_{p+1}(x)=\int_0^1\sum_{i=0}^{n-1}G^{(i)}(x,t)\left\{\lambda_p\alpha_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}+(1-\lambda_p)\beta_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}+\right.$$

$$+(1-\lambda_p)\left(\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,Z_{p+1})(x)}-\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}\right)\beta_p(t)+ \tag{32'}$$

$$\left.+\lambda_p(1-\lambda_p)\left(\left.\frac{\partial f}{\partial u_i}\right|_{(Z_{p+1},Z_p)(x)}-\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,Z_{p+1})(x)}\right)\big(f[Z_p(t)]-f[W_p(t)]\big)\right\}dt.$$

Аналогично получаем и формулу для невязки β_{p+1}:

$$\beta_{p+1}(x)=\int_0^1\sum_{i=0}^{n-1}G^{(i)}(x,t)\left\{\mu_p\beta_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,W_{p+1})(x)}+(1-\mu_p)\alpha_p(t)\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,W_{p+1})(x)}+\right.$$

$$+(1-\mu_p)\left(\left.\frac{\partial f}{\partial u_i}\right|_{(W_{p+1},Z_p)(x)}-\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,W_{p+1})(x)}\right)\alpha_p(t)+ \tag{33'}$$

$$\left.+\mu_p(1-\mu_p)\left(\left.\frac{\partial f}{\partial u_i}\right|_{(W_{p+1},Z_p)(x)}-\left.\frac{\partial f}{\partial u_i}\right|_{(W_p,W_{p+1})(x)}\right)\big(f[Z_p(t)]-f[W_p(t)]\big)\right\}dt.$$

Воспользуемся теперь индуктивными предположениями (23′) и тем, что частные производные f и их разности не превосходят числа N. Мы получаем:

$$\alpha_{p+1}(x) \leqq \lambda_p \theta M_p x^p - (1-\lambda_p)\frac{\theta_1}{p}|k_p|x^p + (1-\lambda_p)\theta|K_p|x^p +$$

$$+\lambda_p(1-\lambda_p)N^2 \int_0^1 \sum_{i=0}^{n-1} G^{(i)}(x,t)\varrho_t(Z_p, W_p)\,dt,$$

где последний член (без множителя $\lambda_p(1-\lambda_p)$) не превосходит (используя и Лемму 1′ и условие [iv])

$$N^2\varrho_x(Z_p, W_p)\int_0^x \sum_{j=0}^{n-1}\frac{(x-t)^j}{j!}\,dt \leqq N\left(N\sum_{i=1}^{n}\frac{x^i}{i!}\right)(M_p+|K_p|)x^p\frac{\theta}{N(1-\theta)} \leqq$$

$$\leqq (M_p+|K_p|)\frac{\theta^2}{1-\theta}x^{p+1}.$$

В итоге получаем:

(34′)

$$\alpha_{p+1}(x) \leqq x^p\left(\lambda_p M_p\theta - (1-\lambda_p)\frac{\theta_1}{p}|k_p| + (1-\lambda_p)\theta|K_p| + \lambda_p(1-\lambda_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x\right).$$

Аналогично оценивается (с помощью Леммы 2′) α_{p+1} снизу:

$$\alpha_{p+1}(x) \geqq \lambda_p\frac{\theta_1}{p}m_p x^p - (1-\lambda_p)\theta|K_p|x^p - (1-\lambda_p)\theta|K_p|x^p -$$

(35′)
$$-\lambda_p(1-\lambda_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x^{p+1} \geqq$$

$$\geqq x^p\left(\lambda_p\frac{\theta_1}{p}m_p - 2(1-\lambda_p)\theta|K_p| - (1-\lambda_p)\lambda_p(M_p+|K_p|)\frac{\theta^2}{1-\theta}x\right).$$

Из (34′), (35′) получаем два условия на λ_p при всех x $(0 \leqq x \leqq 1)$:

(36′) $$\lambda_p M_p\theta - (1-\lambda_p)\frac{\theta_1}{p}|k_p| + (1-\lambda_p)\theta|K_p| + \lambda_p(1-\lambda_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x \geqq 0,$$

(37′) $$\lambda_p\frac{\theta_1}{p}m_p - 2(1-\lambda_p)\theta|K_p| - \lambda_p(1-\lambda_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x > 0,$$

которые при λ_p близких к единице выполняются (как это и можно было ожидать).

Приведем наконец оценки для β_{p+1} на основе Лемм 1′, 2′, 3′ и условия [iv]. Из (33′), если на последнем шагу воспользуемся оценкой типа проделанной перед (34′), получаем

$$\beta_{p+1}(x) \geqq -\mu_p\theta|K_p|x^p + (1-\mu_p)\frac{\theta_1}{p}m_p x^p - (1-\mu_p)\theta M_p x^p -$$

$$-\mu_p(1-\mu_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x^{p+1},$$

откуда получаем условие на μ_p:

$$(38') \quad \mu_p\theta|K_p|-(1-\mu_p)\frac{\theta_1}{p}m_p+(1-\mu_p)\theta M_p+\mu_p(1-\mu_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x \geqq 0.$$

Оценим теперь сверху β_{p+1} (оценка последнего члена дается так же, как и выше, на основе формул перед (34′)):

$$\beta_{p+1}(x) \leqq -\frac{\theta_1}{p}|k_p|\mu_p x^p+(1-\mu_p)\theta M_p x^p+(1-\mu_p)\theta M_p x^p+$$

$$+\mu_p(1-\mu_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x^{p+1},$$

откуда получаем второе условие на μ_p:

$$(39') \qquad \mu_p\frac{\theta_1}{p}|k_p|-2(1-\mu_p)\theta M_p-\mu_p(1-\mu_p)(M_p+|K_p|)\frac{\theta^2}{1-\theta}x > 0.$$

Очевидно, что (38′), (39′) при μ_p достаточно близком к единице выполняются.

Выбрав теперь в (36′), (37′) какое-нибудь из подходящих λ_p, в качестве M_{p+1} можно будет взять значение левой части (36′) при $x=1$, а в качестве m_{p+1} значение левой части (37′) при $x=1$. Аналогично, подставив подходящее значение μ_p в (38′), (39′) при $x=1$ получаем — K_{p+1} соотв. — k_{p+1}.

Таким образом, из индуктивного предположения мы вывели, что существуют такие λ_p, μ_p с которыми при только что вычисленных постоянных M_{p+1}, m_{p+1}; K_{p+1}, k_{p+1}

$$m_{p+1}x^p \leqq \alpha_{p+1}(x) \leqq M_{p+1}x^p, \quad K_{p+1}x^p \leqq \beta_{p+1}(x) \leqq k_{p+1}x^p \quad (0 \leqq x \leqq 1).$$

Теорема доказана.

Замечание 5′. Из уравнений (36′), ..., (38′) можно определить наименьшее значение подходящих λ_p, μ_p, что приведет к дальнейшему улучшению сходимости процесса. Отметим также, что последние два члена в фигурных скобках (32′), (33′) имеют разный знак, и учет этого факта тоже улучшает условия на λ_p, μ_p.

Заметим наконец, что неравенства (36′), (38′) следуют из (37′) соотв. (39′); поэтому для определения λ_p, μ_p достаточно решить неравенства (37′), (39′) относительно квадратных трехчленов от λ_p соотв. μ_p.

Замечание 6′. Если f дважды непрерывно дифференцируема, то в формулах (32′), (33′) разности частных производных можно расписать по соответствующим формулам Лагранжа, и таким образом убедиться (на основе Леммы 1′), что последние два члена в (32′), (33′) в фигурных скобках имеют второй порядок малости если порядок α_p, β_p есть один.

Если при этом $\alpha_p \approx -\beta_p$, то выберая λ_p, μ_p близко к 0,5 можем получить, что α_{p+1}, β_{p+1} тоже будут иметь второй порядок малости. Конечно, при $\alpha_p=-\beta_p$, этот эффект обязательно соблюдается с λ_p, $\mu_p=0{,}5$.

В заключение автор выражает благодарность рецензенту А. Эльберту за ценное обсуждение.

ЛИТЕРАТУРА

[1] Ковач, Ю. И.: Модификации ускоренной сходимости к решению линейного дифференциального уравнения в частных производных с отклоняющимся аргументом, *Украинский Математический Журнал*, т. **26**, 5 (1974), 591—602.

[2] Вулих, Б. З.: *Введение в теорию полуупорядоченных пространств*, Физматгиз, Москва, 1961.

[3] Ковач, Ю. И., Hegedűs, J.: Об одном двустороннем итерационном методе решения краевой задачи с запаздыванием, *Acta Scientiarum Mathematicarum* **36** (1974), 69—89.

[4] Hegedüs, J.: On a two-sided iterative method, *Colloquia Mathematica Societatis János Bolyai* **15**, *Differential Equations*, Keszthely (Hungary), 1975, 277—290.

[5] Hegedüs, J.: On the problem of the choice of first approximants in a two-sided iteration method, *Acta Scientiarum Mathematicarum* **39** (1977), 273—289.

[6] Наймарк, М. А.: *Линейные дифференциальные операторы*, Гос. Изд.-во Технико-Теоретич. Литературы, Москва, 1954.

Больаи Институт, Сегедский Университет, Aradi vértanuk tere 1, H—6720 Szeged, Венгрия

(Поступила: 1 февраля 1976 г.)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 323—334.

ЧИСЛО КОНГРУЭНТНЫХ ШАРОВ, ЗАКРЫВАЮЩИХ ДАННЫЙ ШАР ТРЕХМЕРНОГО ПРОСТРАНСТВА, НЕ МЕНЬШЕ ЧЕМ 30

G. CSÓKA

Пусть $\mathbf{K}_0, \mathbf{K}_1, \ldots, \mathbf{K}_N$ — совокупность конгруэнтных шаров трехмерного евклидова пространства E^3, попарно не имеющих внутренних точек. Множество шаров $\mathbf{K}_1, \ldots, \mathbf{K}_N$ называется *облаком шара* $\mathbf{K}_0$, если любая полупрямая, исходящая из центра шара $\mathbf{K}_0$, имеет хотя бы одну общую точку с каким-либо из шаров $\mathbf{K}_1, \ldots, \mathbf{K}_N$. Говорят также, что шары $\mathbf{K}_1, \ldots, \mathbf{K}_N$ «закрывают» шар $\mathbf{K}_0$.

Вопрос А. Хорниха, после которого и стали рассматривать конструкцию облака, следующий: каково минимальное число N шаров облака?

Первый результат в связи с вопросом Хорниха получил Л. Фейеш Тот [1], доказав что число шаров облака $N \geqq 19$. Позже А. Хеппеш [2] показал, что $N \geqq 24$. Он использовал тот факт, что если шары образуют облако, то некоторые из них должны быть расположены довольно далеко от шара $\mathbf{K}_0$, так что их «тени» достаточно маленькие. Тенью шара $\mathbf{K}_i$, принадлежащего облаку шара $\mathbf{K}_0$ с центром в точке $\mathbf{O}$, называется проекция из точки $\mathbf{O}$ шара $\mathbf{K}_i$ на поверхность шара $\mathbf{K}_0$. Данзер [3] сконструировал облако из 42-х шаров, то есть получил оценку $N \leqq 42$.

В этой статье показано, что $N \geqq 30$. Чтобы получить названную оценку, исследуются проекции из точки $\mathbf{O}$ центров шаров облака на поверхность шара $\mathbf{K}_0$. Пусть $\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3$ — три таких шара облака, каждый из которых имеет хотя бы одну общую точку с некоторой прямой $\mathbf{OP}$, и пусть $\triangle = \mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3'$ — сферический треугольник, образованный центрами этих шаров (рис. 1). Оценив сверху площадь треугольника $\triangle$, мы и выведем нижнюю оценку числа N.

Рассмотрим кривую Вивиани, заданную следующим образом: в прямоугольной декартовой системе координат $(\mathbf{O}, x, y, z)$ возьмем круговой цилиндр радиуса 1, ось которого совпадает с осью $\mathbf{O}_z$, и на нем точку $\mathbf{O}_1(1, 0, \sqrt{3})$. Тогда пересечение поверхности этого цилиндра и поверхности шара центром в точке $\mathbf{O}_1$ и радиусом 2 образуют нужную нам кривую Вивиани. Заметим, что плоскости $z = \sqrt{3}$ и $y = 0$ являются плоскостями симметрии этой кривой. В дальнейшем мы будем рассматривать только ту половину этой кривой, которая лежит в полупространстве $z \geqq \sqrt{3}$. Её уравнения:

$$\text{(1)} \qquad \begin{aligned} x &= \frac{1}{2}\left(z^2 - 2\sqrt{3}z + 1\right), \\ y &= \pm\sqrt{1 - x^2}, \\ \sqrt{3} &\leqq z \leqq \sqrt{3} + 2. \end{aligned}$$

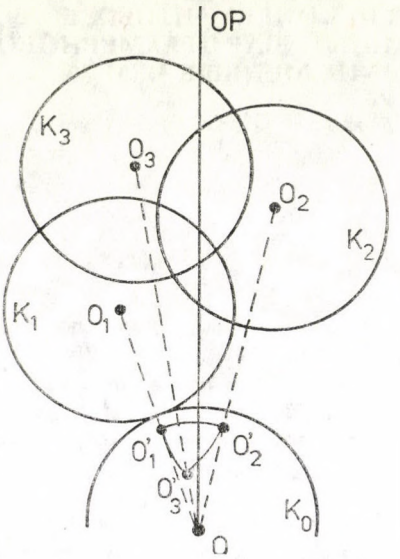


рис. 1

Очевидно, на кривой (1) лежат центры единичных шаров, касающихся единичного шара с центром в точке $\mathbf{O}_1$, оси $\mathbf{O}_z$ и не имеющие общих внутренних точек с единичным шаром $\mathbf{K}_0$, центром которого является точка $\mathbf{O}(0, 0, 0)$.

Рассмотрим точки

$$\mathbf{O}_1(1, 0, \sqrt{3}), \quad \mathbf{Q}(1, 0, \sqrt{3}+2), \quad \mathbf{O}_{20}(-1, 0, \sqrt{3}); \tag{2}$$

$$\mathbf{O}_{21}(0, -1, \sqrt{3}+\sqrt{2}), \quad \mathbf{O}_{31}(0, 1, \sqrt{3}+\sqrt{2}), \quad \mathbf{P}(0, 0, 1)$$

первые три из которых лежат на плоскости $\mathbf{O}_{xz}$, а остальные на плоскости $\mathbf{O}_{yz}$. Точки $\mathbf{O}_{20}$, $\mathbf{O}_{21}$, $\mathbf{O}_{31}$ и $\mathbf{Q}$ лежат на кривой (1).

Спроектируем из точки $\mathbf{O}$ на поверхности $x^2+y^2+z^2=1$ шара $\mathbf{K}_0$ кривую (1) и точки (2). Проекции точек (2) мы будем обозначать теми же буквами, что и сами точки, добавляя знак $'$ (рис. 2). Для дуг больших кругов на поверхности $\mathbf{K}_0$ получаем

$$\smile \mathbf{O}'_1\mathbf{O}'_{20} = 2 \cdot \sphericalangle(\mathbf{O}_1\mathbf{O}\mathbf{O}''_1) = 2 \arcsin\frac{1}{2} = 60^\circ$$

$$\smile \mathbf{O}'_{21}\mathbf{O}'_{31} = 2 \cdot \sphericalangle(\mathbf{O}_{31}\mathbf{O}\mathbf{O}''_{31}) = 2 \operatorname{arc\,tg}\frac{1}{\sqrt{3}+\sqrt{2}} = 35^\circ 16'$$

$$\smile \mathbf{P}\mathbf{Q}' = \sphericalangle(\mathbf{Q}\mathbf{O}\mathbf{Q}'') = \operatorname{arc\,tg}\frac{1}{\sqrt{3}+2} = 15^\circ 7'$$

где знак $''$ над буквой $\mathbf{M}$ обозначает ортогональную проекцию точки $\mathbf{M}$ на прямую $\mathbf{O}_z$. $\smile \mathbf{MN}$ обозначает как дугу большого круга (меньшую из дуг с концами $\mathbf{M}$ и $\mathbf{N}$) так и её угловую величину.

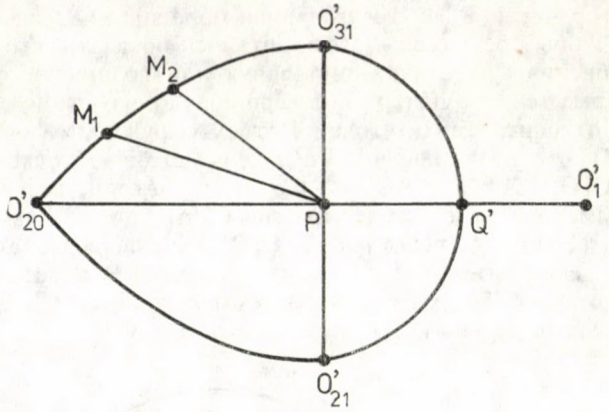


рис. 2

Так как проекция кривой (1) на плоскости $\mathbf{O}_{xz}$ есть дуга параболы, то функция $x=x(z)$ для проекции является графиком, строго возрастающая. Отсюда вытекает, что если точка $\mathbf{M}$ движется по проекции кривой (1) от $\mathbf{O}'_{20}$ к $\mathbf{Q}'$, то длина отрезка $\mathbf{PM}$ строго уменьшается (на рис. 2 $\mathbf{PM}_2<\mathbf{PM}_1$).

Теперь докажем три леммы, на основе которых будет получен наш основной результат.

Лемма А. *Пусть конгруэнтные шары* $\{\mathbf{K}_i\}$ $(i=1, \ldots, N)$ *образуют облако шара* $\mathbf{K}_0$ *и пусть* $\{\mathbf{O}'_i\}$ *— проекции их центров* $\mathbf{O}_i$ *из центра* $\mathbf{O}$ *шара* $\mathbf{K}_0$ *на его поверхность. Тогда существует такое разбиение поверхности шара* $\mathbf{K}_0$ *больших кругов на треугольники, что вершинами треугольников являются точки* $\mathbf{O}'_i$ *и для каждого треугольника* $\mathbf{O}'_{i_1}\mathbf{O}'_{i_2}\mathbf{O}'_{i_3}$ *существует полупрямая* $\mathbf{OP}(\mathbf{P}\in\triangle(\mathbf{O}'_{i_1}\mathbf{O}'_{i_2}\mathbf{O}'_{i_3})$, *которая имеет хотя бы по одной общей точке с каждым из трех шаров* $\mathbf{K}_{i_1}, \mathbf{K}_{i_2}, \mathbf{K}_{i_3}$ *облака.*

Доказательство. Считаем, что облако не имеет лишних шаров, то есть таких шаров, после выбрасывания которых из состава облака, последнее не перестает быть облаком. Кроме системы таких $\{\mathbf{O}'_i\}$, рассмотрим на поверхности $\mathbf{S}_0$ шара $\mathbf{K}_0$ тени $\mathbf{C}'_i$ $(i=1, 2, \ldots, N)$ шаров $\mathbf{K}_i$; через r_i обозначим сферический (угловой) радиус тени $\mathbf{C}'_i$. Произведем следующее разбиение I, поверхности $\mathbf{S}_0$: каждой точке $\mathbf{O}'_i$ соотнесем множество точек $\mathbf{X}$ поверхности, для каждой из которых величина $\smile\mathbf{O}'_i\mathbf{X}-r_i$ минимальна при $i=1, \ldots, N$ (так называемое «гиперболическое разбиение» или разбиение Фейеша Тота—Молнара [4]). Кривыми, производящими разбиение в этом случае являются дуги сферических гипербол и больших кругов.

Разбиение II. Поверхности $\mathbf{S}_0$ дугами больших кругов дуальное разбиению I, имеет своими вершинами множество точек $\{\mathbf{O}'_i\}$. Вершины всякой клетки разбиения II обладают тем свойством, что тени с центрами в этих вершинах покрывают вершину разбиения I, соответствующую данной клетке.

Если клетки разбиения II не являющиеся треугольниками, разобьем некоторым образом диагоналями этих клеток на треугольники, то получим

разбиение поверхности S_0 на треугольники обладающие указанными в лемме свойствами. Чтобы убедиться в этом, осталось показать, что в каждом треугольнике разбиения существует внутренняя или граничная точка, принадлежащая всем трем теням с центрами в вершинах треугольника. Пусть большой круг $OO_1'O_2'$ отделит общую точку P трех окружностей—теней C_1', C_2', C_3', от вершины O_3' (рис. 3) и пусть точка P' симметрична точке P относительно плоскости $OO_1'O_2'$. Тогда $P' \in C_1'$, $P' \in C_2'$ по симметрии; вследствии того, что плоскость симметрии отделила центр тени C_3' от точки P, имеем $P' \in C_3'$. Если точка P' попала внутрь треугольника $O_1'O_2'O_3'$, утверждение доказано. Если нет, то утверждение вытекает из того, что отрезок PP' пересекает треугольник $O_1'O_2'O_3'$ и состоит из общих точек всех трех окружностей C_1', C_2', C_3', поскольку множество таких точек выпуклое. Лемма доказана.

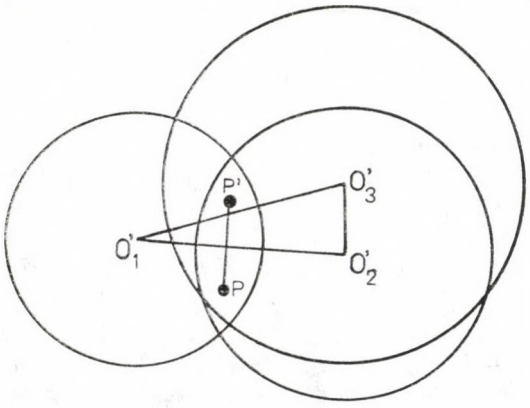


рис. 3

Лемма B. *Пусть единичные шары* K_0, K_1, K_2, K_3 *попарно не имеют общих внутренних точек; а полупрямая* l, *исходящая из центра* O *шара* K_0, *имеет хотя бы по одной общей точке с каждым из шаров* K_1, K_2, K_3. *Тогда площадь сферического треугольника* $O_1'O_2'O_3'$, *вершинами которого являются проекции из точки* O *на поверхность шара* K_0 *центров* O_1, O_2, O_3 *шаров* K_1, K_2, K_3, *будет максимальна в том и только том случае, когда каждый из шаров* K_1, K_2, K_3 *касается* l, *и когда расположение этих шаров таково, что ни один из них нельзя приблизить к точке* O, *перемещая параллельно прямой* l.

Доказательство. Если шар K_3 перемещаем параллельно прямой l, тогда O_{31}' заменяется точкой O_{32}' (рис. 4). Пусть $OO_1 \leqq OO_2 \leqq OO_3$. Так как тени шаров K_1, K_2, K_3 — сферические круги C_1, C_2 и C_3 имеют общую точку и наименьший из них C_3 ,легко видеть, что C_3 покрывает по крайней мере одну из двух общих граничных точек кругов C_1 и C_2. Обозначим эту точку через P_1. Если C_1 и C_2 касаются, тогда P_1 является точкой касания. Следовательно, существует полупрямая OP_1, только касающаяся шаров K_1 и K_2, и касающаяся или пересекающая шар K_3.

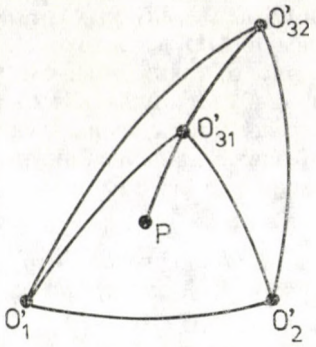


рис. 4

Покажем, что в этом последнем случае можно удалить точку $\mathbf{O}_3$ от $\mathbf{OP}_1$ так, что шары $\mathbf{K}_1, \mathbf{K}_2$ и $\mathbf{K}_3$ попрежнему не будут пересекаться внутренними точками, а площадь сферического треугольника $(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ увеличится. Для этого рассмотрим плоскость Σ, содержащую прямую $\mathbf{OO}_3$ и перпендикулярную плоскостьи $[\mathbf{OO}_1\mathbf{O}_2]$. Обозначим через $\mathbf{O}_1^*$ и $\mathbf{O}_2^*$ перпендикулярные проекции точек $\mathbf{O}_1$ и $\mathbf{O}_2$ на плоскости Σ. Точки $\mathbf{O}$, $\mathbf{O}_1^*$ и $\mathbf{O}_2^*$ будут точками общей прямой плоскости Σ и $[\mathbf{O}\,\mathbf{O}_1\mathbf{O}_2]$ (рис. 5). Очевидно, если передвигать центр $\mathbf{O}_3$ шара $\mathbf{K}_3$ на плоскости Σ по окружности с центром $\mathbf{O}_1^*$ и радиусом $\mathbf{O}_1^*\mathbf{O}_3$ или вне круга этой окружности, то шары $\mathbf{K}_3$ и $\mathbf{K}_1$ не будут иметь общих внутренних точек. То же самое имеет место и при движении точки $\mathbf{O}_3$ по окружности с центром $\mathbf{O}_2^*$ и радиусом $\mathbf{O}_2^*\mathbf{O}_3$ или вне круга этой окружности — шары $\mathbf{K}_3$ и $\mathbf{K}_2$ не пересекаются.

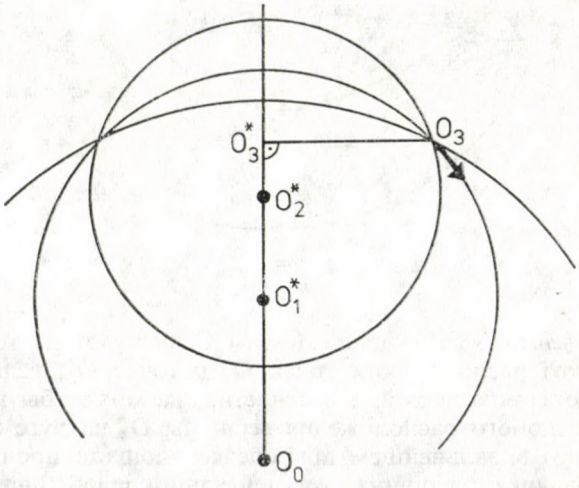


рис. 5

Поскольку точки $\mathbf{O}_1^*$ и $\mathbf{O}_2^*$ лежат внутри отрезка $\mathbf{OO}_3^*$ ($\mathbf{O}_3^*$ — ортогональная проекция точки $\mathbf{O}_3$ на прямую $\mathbf{OO}_2^*$), то точку $\mathbf{O}_3$ можно двигать по окружности радиусом $\mathbf{OO}_3$ (см. рис. 5) таким образом, что шар $\mathbf{K}_3$ не будет пересекаться с другими шарами и будет удаляться от прямой $\mathbf{OP}_1$, причем такое удаление может проводиться до положения касания $\mathbf{K}_3$ и прямой $\mathbf{OP}_1$. Поскольку при этом точка $\mathbf{O}_3'$ движется перпендикулярно дуге $\mathbf{O}_1'\mathbf{O}_2'$, то площадь треугольника $\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3'$ увеличивается. Тем самым первое утверждение леммы доказано.

Лемма C. *Пусть из начала координат* $\mathbf{O}(0, 0, 0)$ *на поверхность единичной сферы* $\mathbf{K}_0$ *с центром в точке* $\mathbf{O}$ *спроектированы кривая Вивиани* (1) *и точки* (2); *обозначим через* $\mathbf{G}'$ *проекцию точки* $\mathbf{G}$ *кривой* (1) *с координатами* $z=\sqrt{3}+0{,}05$, *а через* $(\mathbf{O}_{20}'\mathbf{G}')$, $(\mathbf{O}_{20}'\mathbf{Q}')$ *проекции дуг кривой* (1) *с концами соответственно* $\mathbf{O}_{20}'\mathbf{G}'$ *и* $\mathbf{O}_{20}'\mathbf{Q}'$. *Тогда площадь* $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ *всякого сферического треугольника с вершинами* $\mathbf{O}_1'$, $\mathbf{O}_2'$, $\mathbf{O}_3'$ *(рис. 6), где* $\mathbf{O}_2' \in (\mathbf{O}_{20}'\mathbf{G}')$; $\mathbf{O}_3' \in (\mathbf{O}_{20}'\mathbf{Q}')$ *удовлетворяет неравенству*

$$T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3') < 0{,}23. \tag{3}$$

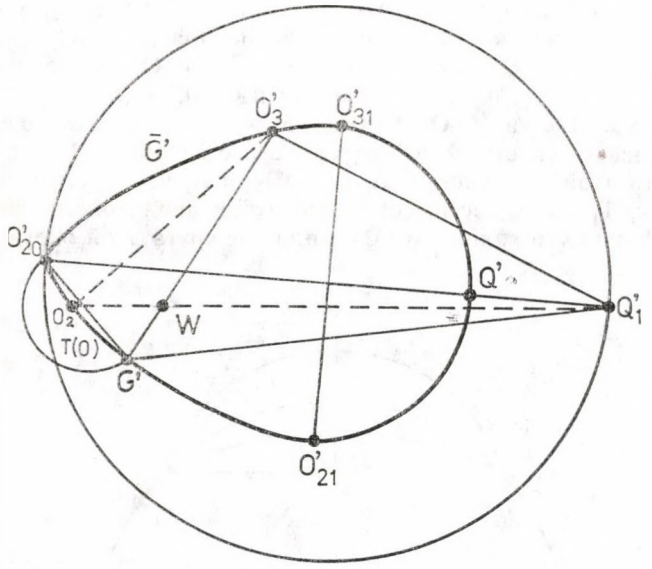


рис. 6

Доказательство. Утверждение Леммы C вытекает из того, что при достаточно близком расположении точки $\mathbf{O}_2'$ к точке $\mathbf{O}_{20}'$ величина $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ тоже будет достаточно малой, в частности, она может быть сделана меньше оценки (3) для любого расположения вершины $\mathbf{O}_3'$ на дуге $(\mathbf{O}_{20}'\mathbf{Q}')$. На основании оценки (3) в дальнейшем при оценке площади произвольного сферического треугольника с вершинами на проекции кривой Вивиани можно будет исключить из рассмотрения часть кривой, состоящей из дуги $(\mathbf{O}_{20}'\mathbf{G}')$ и ее

отражения $(\mathbf{O}'_{20}\overline{\mathbf{G}}')$ (см. рис. 6). Такое исключение нам потребовалось из-за неограниченности производной одной из рассматриваемых функций в точке $\mathbf{O}'_{20}$ проекции кривой (1). Переходим теперь к доказательству леммы.

Из уравнений (1) находим координаты точки $\mathbf{G}'$:

$$x = \frac{1}{2}\left[(\sqrt{3}+0{,}05)^2 - 2\sqrt{3}(\sqrt{3}+0{,}05)+1\right] \approx -0{,}9987,$$

$$y = \sqrt{1-0{,}9987^2} \approx 0{,}0498,$$

$$z = \sqrt{3}+0{,}05 \approx 1{,}7820$$

и оцениваем величину дуги большого круга $\smile\mathbf{G}'\mathbf{O}'_{20} < 1°35'$.

Пусть $\mathbf{O}'_2 \in (\mathbf{O}'_{20}\mathbf{G}')$ и $\mathbf{O}'_3 \in (\mathbf{O}'_{20}\mathbf{Q}')$. Обозначим через $\mathbf{W}$ точку пересечения дуг больших кругов $\smile\mathbf{G}'\mathbf{O}'_3$ и $\smile\mathbf{O}'_2\mathbf{O}'_1$. Для площадей получившихся сферических треугольников (рис. 6) имеем равенство:

$$T(\mathbf{O}'_2\mathbf{O}'_1\mathbf{O}'_3) = T(\mathbf{G}'\mathbf{O}'_1\mathbf{O}'_3) + T(\mathbf{O}'_2\mathbf{W}\mathbf{O}'_3) - T(\mathbf{G}'\mathbf{O}'_1\mathbf{W}). \tag{4}$$

Способом, который будет описан ниже, при доказательстве теоремы, получаем оценку $T(\mathbf{G}'\mathbf{O}'_1\mathbf{O}'_3) < 0{,}184$. Из нее и равенства (4) имеем

$$T(\mathbf{O}'_2\mathbf{O}'_1\mathbf{O}'_3) < 0{,}184 + T(\mathbf{O}'_2\mathbf{W}\mathbf{O}'_3).$$

Оценим теперь величину $T(\mathbf{O}'_2\mathbf{W}\mathbf{O}'_3)$. Из свойства кривой (1) о котором говорилось выше, перед Леммой A, следует, что $T(\mathbf{O}'_2\mathbf{W}\mathbf{O}'_3) < T(\mathbf{G}'\mathbf{O}'_{20}\mathbf{O}'_3) + T(0)$, где через $T(0)$ обозначена половина площади сферического круга радиуса $\frac{1}{2}(\smile\mathbf{G}'\mathbf{O}'_{20})$. Вычислив $T(0)$ находим, что $T(0) < 0{,}006$. Так как $T(\mathbf{G}'\mathbf{O}'_{20}\mathbf{O}'_3)$ меньше площади равнобедренного сферического треугольника δ со сторонами с 60°, 60° и 1°35′, поскольку $\smile\mathbf{O}'_{20}\mathbf{G}' < 1°35'$, $\smile\mathbf{O}'_{20}\mathbf{O}'_3 < 60°$ и $\smile\mathbf{G}'\mathbf{O}'_3 < 60°$, то вычислив площадь δ получаем неравенство $T(\mathbf{G}'\mathbf{O}'_{20}\mathbf{O}'_3) < 0{,}033$, из которого и следует неравенство (3). Переходим теперь к доказательству основного результата.

Теорема. *В пространстве E^3 любое облако состоит не менее, чем из 30 шаров.*

Доказательство. Пусть $\mathbf{K}_0$ — единичный шар с центром в точке $\mathbf{O}$, $\{\mathbf{K}_i\}$ $(i=1, \ldots, N)$ — облако шара $\mathbf{K}_0$, $\mathbf{O}_i$ — центр шара $\mathbf{K}_i$, $\mathbf{O}'_i$ — проекция точки $\mathbf{O}_i$ из точки $\mathbf{O}$ на поверхность $\mathbf{S}_0$ шара $\mathbf{K}_0$. Считаем, что облако $\{\mathbf{K}_i\}$ не имеет лишних шаров.

По Лемме A существует разбиение поверхности $\mathbf{S}_0$ на сферические треугольники, вершинами треугольников являются точки $\{\mathbf{O}'_i\}$ и всякому треугольнику соответствует полупрямая $\mathbf{OP}$ ($\mathbf{P}$ принадлежит к треугольнику), имеющая хотя бы по одной общей точке с шарами проекции центров которых являются вершинами треугольника. Будем искать верхнюю оценку площади треугольников такого разбиения.

Рассмотрим полупрямую $\mathbf{OP}$ соответствующую некоторому треугольнику $\triangle(\mathbf{O}'_1\mathbf{O}'_2\mathbf{O}'_3)$ разбиения; тем самым полупрямая $\mathbf{OP}$ имеет общие точки с каждым из шаров $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_3$ облака $\{\mathbf{K}_i\}$ (рис. 1). Согласно Лемме B в наших

поисках верхней оценки площади треугольника $\triangle(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ достаточно рассмотреть случай, когда шары $\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3$ касаются полупрямой **OP** и ни один из них нельзя приблизить к точке **O** движением параллельно прямой **OP**.

Предполагаем, что $\mathbf{OO}_1 \leqq \mathbf{OO}_2 \leqq \mathbf{OO}_3$, откуда следует, что шар $\mathbf{K}_1$ касается шара $\mathbf{K}_0$. Основная идея доказательства состоит в исследовании площади сферического треугольника, который получится после приближения по направлению прямой **OP** шара $\mathbf{K}_3$ к шару $\mathbf{K}_1$ до касания, допусти, что при таком приближении шары $\mathbf{K}_3$ и $\mathbf{K}_2$ могут оказаться пересекающимися. При таком приближении, как об этом говорилось при доказательстве Леммы B, площадь $\triangle(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ увеличилась. Точки $\mathbf{O}_2$ и $\mathbf{O}_3$ теперь лежат на кривой Вивиани, которая определяется пересечением цилиндра единичного радиуса с осью **OP** и шара радиуса 2 с центром в точке $\mathbf{O}_1$ (рис. 7). Прямоугольную систему координат выберем так, чтобы начало совпало с точкой **O**, а точка **P** имела координаты (0, 0, 1) то есть кривая Вивиани описывалась уравнениями (1). На рис. 7 указаны проекции точек (2) на поверхности шара $\mathbf{K}_0$. Теперь наша задача свелась к оценке сверху площади треугольника $\triangle(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$.

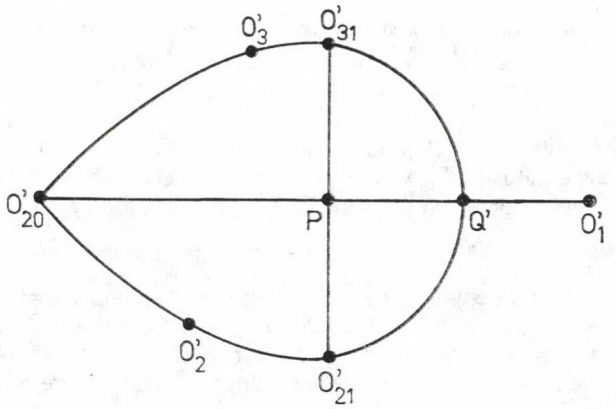


рис. 7

Оценку получим следующим способом: пусть точка $\mathbf{O}_2'(x_2, y_2, z_2)$ пробегает дугу $(\mathbf{O}_{20}'\mathbf{O}_{21}')$ проекции и при каждом фиксированном положении точки $\mathbf{O}_2'$ точка $\mathbf{O}_3'(x_3, y_3, z_3)$ пробегает дугу $(\mathbf{O}_{20}'\mathbf{O}_{31}'\mathbf{Q}')$. Тогда на основе (1) мы получим, что площадь $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ будет функцией двух переменных z_2 и z_3: $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3') = T(z_2, z_3)$. Областью определения функции $T(z_2, z_3)$ в плоскости (z_2, z_3) является квадрат с вершинами $(\sqrt{3}, \sqrt{3})$, $(\sqrt{3}, \sqrt{3}+2)$, $(\sqrt{3}+2, \sqrt{3})$, $(\sqrt{3}+2, \sqrt{3}+2)$.

Разделим этот квадрат прямыми параллельными его. Расстояния между любыми двумя из этих прямых возьмем равным данному числу d. Вдоль каждого из отрезков прямых попавшего в квадрат функция $T(z_2, z_3)$ будет функция только одной переменной: $T(c, z_3)$; $T(z_2, c)$, где $\sqrt{3} \leqq c \leqq \sqrt{3}+2$ (график одной из таких функций дан на рис. 8).

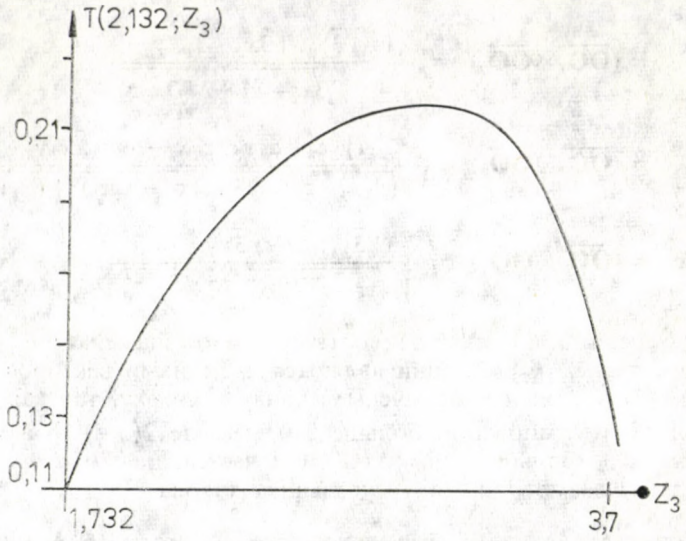


рис. 8

Вычислив значение функции $T(z_2, z_3)$ и значение ее частных производных в точках квадратной решетки со стороны d, которая образовалась от пересечения двух систем параллельных прямых, мы найдем и оценку для этой функции на всем квадрате, то есть получим оценку площади $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$. Результат вычисления дает верхнюю оценку 0,23. Следовательно, в разбиении поверхности шара $\mathbf{K}_0$ на треугольнике описанном в Лемме A имеется по меньшей мере l граней, где $l \geqq \frac{4\pi}{0{,}23}$. Обозначим через e число ребер, а через c число вершин разбиения. Так как все грани разбиения треугольники, то $3l=2e$. Из равенства $l+c=e+2$ (теорема Эйлера) получаем $c=\frac{1}{2}l+2 \geqq \frac{1}{2}\frac{4\pi}{0{,}23}+$ $+2 \geqq 29{,}3$. Учитывая, что c–целое число, получаем что $c \geqq 30$, то есть утверждение теоремы.

Переходим к подробному описанию расчетов для получения оценки $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$. Рассмотрим векторы $\overrightarrow{\mathbf{OO}}_1(1, 0, \sqrt{3})$, $\overrightarrow{\mathbf{OO}}_2(x_2, y_2, z_2)$ и $\overrightarrow{\mathbf{OO}}_3(x_3, y_3, z_3)$, где $\mathbf{O}_2$ и $\mathbf{O}_3$ точки кривой Вивиани. Рассмотрим также точки $\mathbf{O}_2' \in (\mathbf{O}_{21}'\mathbf{G}')$ и $\mathbf{O}_3' \in (\mathbf{Q}'\overline{\mathbf{G}}')$ (рис. 7). Мы не рассматриваем случаи, когда $\mathbf{O}_2' \in (\mathbf{O}_{21}'\mathbf{Q}')$ и $\mathbf{O}_3' \in (\mathbf{O}_{31}'\overline{\mathbf{G}}')$ так как точка $\mathbf{P}$ не попадет в треугольник $\triangle(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')$ если $\mathbf{O}_3' \in (\mathbf{O}_{31}'\mathbf{Q}')$ и когда $\mathbf{O}_3' \in (\overline{\mathbf{G}}'\mathbf{O}_{31}')$–из-за симметрии проекции кривой (1).

При данных условиях рассматриваемые векторы образуют правую тройку. Единичные векторы, идущие в направлениях их векторных произведений

следующие:

$$\bar{\mathbf{e}}_{12} = (\overrightarrow{\mathbf{OO}_1} \times \overrightarrow{\mathbf{OO}_2})^\circ = \frac{(-\sqrt{3}y_2,\ \sqrt{3}x_2 - z_2,\ y_2)}{\sqrt{4y_2^2 + 3x_2^2 + z_2^2 - 2\sqrt{3}x_2 z_2}},$$

$$\bar{\mathbf{e}}_{23} = (\overrightarrow{\mathbf{OO}_2} \times \overrightarrow{\mathbf{OO}_3})^\circ = \frac{(y_2 z_3 - z_2 y_3,\ z_2 x_3 - x_2 z_3,\ x_2 y_3 - y_2 x_3)}{\sqrt{(y_2 z_3 - z_2 y_3)^2 + (z_2 x_3 - x_2 z_3)^2 + (x_2 y_3 - y_2 x_3)^2}},$$

$$\bar{\mathbf{e}}_{31} = (\overrightarrow{\mathbf{OO}_3} \times \overrightarrow{\mathbf{OO}_1})^\circ = \frac{(\sqrt{3}y_3,\ z_3 - \sqrt{3}x_3,\ -y_3)}{\sqrt{4y_3^2 + z_3^2 + 3x_3^2 - 2\sqrt{3}x_3 z_3}}.$$

Обозначим через $U(z)$, $V(z)$, $R(z)$ соответственно знаменатели множителей в выражениях для $\bar{\mathbf{e}}_{12}$, $\bar{\mathbf{e}}_{23}$, $\bar{\mathbf{e}}_{31}$; они являются и длинами векторов векторного произведения. Эти длины ограничены и снизу и сверху, так как длины векторов, которые мы умножали, больше 2 и меньше $2(z+\sqrt{3})$, а углы между ними по Лемме C больше угла $\sphericalangle(\mathbf{G'OG'})$ и меньше 60°.

Площадь изучаемого нами треугольника такова:

$$T(z_2, z_3) = \arccos(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31}) + \arccos(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{23}) + \arccos(\bar{\mathbf{e}}_{23}\bar{\mathbf{e}}_{31}) - \pi.$$

Дифференциал функции $T(z_2, z_3)$ в том случае, когда z_2 и z_3 изменяются на величину d имеет вид:

$$\Delta_d T(z_2, z_3) = T'(c, z_3)d + T'(z_2, c)d,$$

где $T'(c, z_3)$ и $T'(z_2, c)$ частные производные.

При данных условиях ищем верхнюю оценку частной производной $\dfrac{\partial T}{\partial z_2}$ на всей области определения функции $T(z_2, z_3)$. Поскольку функция $T(z_2, z_3)$ симметрична относительно переменных z_2 и z_3, то полученная оценка будет оценкой и производной $\dfrac{\partial T}{\partial z_3}$. Обозначим $T' = \max \dfrac{\partial T}{\partial z_2} = \max \dfrac{\partial T}{\partial z_3}$, где $z_2 \in [\sqrt{3}+0{,}05,\ \sqrt{3}+\sqrt{2}]$ и $z_3 \in [\sqrt{3}+0{,}05,\ \sqrt{3}+2]$. Если T' существует, тогда разность между значением функции $T(z_2, z_3)$ в точке $z_2 = \sqrt{3}+0{,}05+kd$, $z_3 = \sqrt{3}+0{,}05+jd$ и значением ее в произвольной точке квадрата

$$[\sqrt{3}+0{,}05+kd,\ \sqrt{3}+0{,}05+(k+1)d] \times [\sqrt{3}+0{,}05+jd,\ \sqrt{3}+0{,}05+(j+1)d]$$

оценивается неравенством $\max(\Delta_d T(z_2, z_3)) \leqq T' 2d$. Переходим теперь к доказательству существования T'.

$$\frac{\partial T}{\partial z_3} = \frac{(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})'}{\sqrt{1-(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})^2}} - \frac{(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{23})'}{\sqrt{1-(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{23})^2}} - \frac{(\bar{\mathbf{e}}_{23}\bar{\mathbf{e}}_{31})'}{\sqrt{1-(\bar{\mathbf{e}}_{23}\bar{\mathbf{e}}_{31})^2}}.$$

Покажем, что каждое из трех слагаемых в отдельности ограничено. Поскольку выражения всех трех слагаемых аналогичны, достаточно провести доказательство для одного из них. Рассмотрим, например, первое слагаемое. После

подстановки значений векторов имеем

$$(5)\qquad -\frac{(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})'}{\sqrt{1-(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})^2}}=\frac{\left\{\frac{1}{U(z)R(z)}\left[4y_2y_3+(\sqrt{3}x_2-z_2)(\sqrt{3}x_3-z_3)\right]\right\}'}{\sqrt{1-(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})^2}}.$$

Величина $1-(\bar{\mathbf{e}}_{12}\bar{\mathbf{e}}_{31})$ ограничена снизу числом большим 0, потому что $\bar{\mathbf{e}}_{12}=(\overrightarrow{\mathbf{OO}}_1\times\overrightarrow{\mathbf{OO}}_2)^\circ$, $\bar{\mathbf{e}}_{31}=(\overrightarrow{\mathbf{OO}}_3\times\overrightarrow{\mathbf{OO}}_1)^\circ$ а угол между $\overrightarrow{\mathbf{OO}}_2$ и $\overrightarrow{\mathbf{OO}}_3$ не меньше угла $\sphericalangle(\mathbf{G}'\mathbf{O}\overline{\mathbf{G}}')$ (см. Лемму C).

Обратимся к производной, стоящей в числителе (5). В знаменателе этой производной стоит произведение $U(z)R(z)$, которое ограничено. Числитель производной имеет вид:

$$(4y_2y_3'+3x_2x_3'+z_2-\sqrt{3}x_2-\sqrt{3}z_2x_3')U(z)R(z)-$$

$$-(4y_2y_3-\sqrt{3}x_2z_3+3x_2x_3+z_2z_3-\sqrt{3}z_2x_3)U(z)R'(z).$$

Нужно доказать ограниченность сверху функций $x_3'(z)$, $y_3'(z)$ и $R'(z)$. Функция $x_3'(z_3)=z_3-\sqrt{3}$ ограничена и снизу и сверху, так как $z_3\in[\sqrt{3}+0{,}05,\sqrt{3}+2]$.

Функция $y_3'(z_3)=-\frac{x_3(z_3)}{\sqrt{1-x_3^2(z_3)}}x_3'(z_3)$ также ограничена, потому, что множители в числителе ограничены сверху, а знаменатель ограничен снизу на основе оценки $x_3(z_3)\leqq 0{,}9987$ (Лемма C). Наконец, ограниченность функции

$$R'(z_3)=\frac{1}{2\sqrt{R(z_3)}}(8y_3y_3'+2z_3+6x_3x_3'-2\sqrt{3}x_3-2\sqrt{3}z_3x_3')$$

вытекает из уже сказанного.

Подходящими оценками сверху выражения (5) находим конкретную величину T' и при выборе величины d достаточно малой сможем оценить сверху величину $\Delta_d T(z_2, z_3)$. На основе полученных оценок и была найдена оценка для площади треугольника $T(\mathbf{O}_1'\mathbf{O}_2'\mathbf{O}_3')<0{,}23$. Метод уже использовался ранее при доказательстве Леммы C. Теорема доказана.

Замечание. Легко видеть, что ход доказательства теоремы позволяет написать функцию оценки $N(R, r)$ числа шаров облака шара фиксированного радиуса R, состоящего из шаров радиуса r. В частности, мы нашли оценку числа $N(1, 1)$. Вероятно, полученную оценку можно улучшить оценивая сверху тень системы, состоящей более чем из трех шаров, хотя до сих пор это не удалось.

ЛИТЕРАТУРА

[1] FEJES TÓTH, L.: Verdeckung einer Kugel durch Kugeln, *Publ. Math. Debrecen* **6** (1959), 234—240.
[2] HEPPES, A.: On the number of spheres, which can hide a given sphere, *Canad. J. Math.* **19** (1967), 413—418.
[3] DANZER, L.: Drei Beispiele zu Lagerungsproblemen, *Arch. Math.* **11** (1960), 159—165.
[4] MOLNÁR, J.: Kreispackungen und Kreisüberdeckungen auf Flächen konstanter Krümmung, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 243—251.

Кафедра геометрии, Университет им. Л. Этвеша,
Múzeum krt. 6—8, H—1088 Budapest, Венгрия

(Поступила 31 мая 1976)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 335—336.

A GENERALIZATION OF FITTING-SUBGROUP

by

P. HERMANN

It is well-known by the theorem of Fitting, that the subgroup of a finite group generated by all nilpotent normal subgroups is nilpotent. The aim of this paper is to show the normal persistency of some properties, which are all weaker than nilpotence. However, we deal with the p-solvable case.

PROPOSITION 1. *Let* $\mathbf{t}=(p_1, p_2, \ldots)$ *be a fixed order on the set of the primes,* N *and* M *normal subgroups of the finite group* G. *If* N *and* M *have Sylow-towers of the type* **t**, *then* NM *also has a Sylow-tower of the type* **t**.

PROOF. Induction on the order of G. We can suppose, that $G=NM$. Let p be the greatest prime factor of $|G|$ according to **t**, then, by the assumption, N and M have characteristic p-Sylow subgroups R and S, respectively. Factorizing by RS NS/RS and MR/RS are normal in G/RS, and they have Sylow-towers of the type **t**. Since $|G: RS|<|G|$, $G/RS=NS/RS \cdot MR/RS$ also has a Sylow-tower of the type **t**, and the same holds for G, because RS is the unique p-Sylow subgroup of G.

One can prove in the same way

PROPOSITION 2. *Let* p *be a prime. If* N *and* M *are normal* p*-nilpotent subgroups of the finite group* G, *then* NM *is also* p*-nilpotent.*

Before the final generalization we recall some facts about the finite one step non-nilpotent groups.

THEOREM 1. *Let* G *be a finite non* p*-nilpotent group, all of whose proper subgroups are* p*-nilpotent; then*

1) G *is solvable, moreover* $|G|=p^a q^b$ $(p\neq q)$.

2) *Every proper subgroup of* G *is nilpotent.*

3) G *has a normal* p*-Sylow subgroup* P, *and for any* q*-Sylow subgroup* Q, $G=\langle Q^x \mid x\in G\rangle$ (see [1]).

Groups like in Theorem 1 are called (p, q)-groups. It is clear, that a finite group G is p-nilpotent if and only if for every $q\in\pi(G)\setminus\{p\}$ it does not contain a (p, q)-group.

THEOREM 2. *Let* G *be a finite group,* p *and* q *fixed different primes,* N *and* M *normal subgroups of* G. *If* N *and* M *do not contain* (p, q)*-groups and* G *is* p*-solvable, then* NM *does not contain a* (p, q)*-group.*

PROOF. Let G be a counterexample, for which $|G|+|G:N|+|G:M|$ is minimal. Then by assumption, $G=NM$, and there exists a (p, q)-group U in G. Let us denote the unique p-Sylow subgroup of U by U_p and a q-Sylow subgroup of U by U_q. Let $L \lhd G$, $L \neq 1$, then $G/L = NL/L \cdot ML/L$ and NL/L, ML/L do not contain (p, q)-groups because of [2], so G/L does not contain a (p, q)-group, consequently UL/L is nilpotent by Theorem 1. The nilpotence of UL/L involves $U_qL/L \lhd UL/L$, so $UL/L = \langle (U_qL)^x \mid x \in U \rangle / L = U_qL/L$, which implies $U_p \leqq L$. In fact, $p \mid |L|$ and $O_{p'}(G)=1$, hence by the p-solvability of G $O_p(G)>1$. Taking $L=O_p(G)$ we get $U_p \leqq O_p(G)$, and the choice $L=M$ and $L=N$, respectively, gives $U_p \leqq N \cap M$, consequently, $U_q \not\leqq N$ and $U_q \not\leqq M$. Suppose that $N<K<G$ for some $K \lhd G$. Using $K = K \cap NM = N \cdot (K \cap M)$ we get by the minimality of G, that K does not contain a (p, q)-group; but $|G:K|<|G:N|$ implies in this case that $G=KM$ does not contain a (p, q)-group, contrary to the assumption. Thus G/N and G/M are simple p'-groups, hence $O_p(G) \leqq N \cap M$. Denote by Q a q-Sylow subgroup of G containing U_q, then $O_p(G) \leqq N \cap M$ implies $Q \cap N \leqq C_N(O_p(G))$ and $Q \cap M \leqq$ $\leqq C_M(O_p(G))$. Finally, we get by $U_q \leqq Q = (Q \cap M) \cdot (Q \cap N) \leqq C_G(O_p(G)) \leqq C_G(U_p)$ a contradiction, which completes the proof.

*

Acknowledgement: I should like to express my gratitude to Professor K. Corradi for his valuable suggestions.

REFERENCES

[1] HUPPERT, B.: *Endliche Gruppen*, Springer, Berlin, 1967.
[2] CORRADI, K.: On the p-supersolvability of finite groups (to appear).

Department of Algebra and Number Theory, Roland Eötvös University, Múzeum krt. 6—8, H—1088 Budapest, Hungary

(Received June 20, 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 337—343.

DICHTESTE KUGELPACKUNGEN IM OKTAEDER

von

GEORG GOLSER

1. Einleitung

Ein öfter behandeltes Problem der Geometrie der Zahlen bzw. der diskreten Geometrie ist die Angabe von dichtesten Kugelpackungen in konvexen Körpern im $\mathbf{R}^d$. Die vorliegende Arbeit befaßt sich mit diesem Problem für Oktaeder im $\mathbf{R}^3$.

Es seien $B, K_1, \ldots, K_n$ kompakte, konvexe Teilmengen des $\mathbf{R}^d$ mit nichtleerem Inneren. Das System $\{K_i\}$ heißt *Packung* in B, wenn $\bigcup_{i=1} K_i \subset B$ ist und die K_i paarweise keine gemeinsamen inneren Punkte haben. Von nun an seien die K_i Einheitskugeln, (d. h. Kugeln vom Radius 1). $\{K_i\}$ heißt *dichteste Packung* von n Einheitskugeln in B, wenn $\{K_i\}$ eine Packung in B ist und λB für $0<\lambda<1$ keine Packung von n Einheitskugeln enthält.

Im folgenden sei stets $d=3$. Den Arbeiten [8], [9], [10] von SCHAER entnimmt man die dichtesten Packungen von 3, 4, 5, 6, 8 und 9 Einheitskugeln in einem Würfel. Schaer löst nämlich das Problem, im Einheitswürfel 3, ..., 9 Punkte so anzuordnen, daß der kleinste ihrer gegenseitigen Abstände möglichst groß ist.

In [5] gibt GOLDBERG Packungen von zehn bis siebenundzwanzig Einheitskugeln im Würfel an. (Die Frage, ob sie dichteste Packungen sind, bleibt offen.)

In der Arbeit [2] von BLACHMANN wird die größte Anzahl von Einheitskugeln berechnet, die eine Packung in einer Kugel vom Radius $r \leqq 1+\sqrt{2}$ bilden; für $r>1+\sqrt{2}$ findet man in [2] Abschätzungen für diese Zahl. Blachmann gibt auch höher dimensionale Resultate an.

Von HORVÁTH [6] stammen Abschätzungen für dichteste Packungen von Einheitskugeln im Zylinder mit Radius $\leqq 2$. Die Resultate werden auch auf höhere Dimensionen verallgemeinert.

Im Abschnitt 2 dieser Arbeit wird die Seitenlänge des regulären Oktaeders angegeben, in dem drei Einheitskugeln am dichtesten gepackt sind.

Im 3. Abschnitt werden die Seitenlängen der regulären Oktaeder angegeben, die dichteste Packungen von jeweils 4, 5 und 6 Einheitskugeln enthalten. Davon folgt, daß ein reguläres Oktaeder, das eine Packung von vier Einheitskugeln enthält, stets eine solche von sechs Einheitskugeln enthält.

Für die Anregung und Betreuung dieser Arbeit danke ich Herrn Prof. Dr. P. M. GRUBER sehr herzlich.

2. Dichteste Packung von drei Einheitskugeln im Oktaeder

SATZ 1. *Die Seitenlänge des regulären Oktaeders, das eine dichteste Packung von drei Einheitskugeln enthält, ist* $(\sqrt{2}+3\sqrt{6})/2$.

Ohne Beweis geben wir eine Verallgemeinerung an. Seien $\alpha, \beta \in \mathbf{R}$, $1 \leqq \alpha \leqq \beta$ und sei

$$\gamma := \sqrt{1+\alpha^2+\beta^2} + \begin{cases} (\sqrt{3}+\alpha)/2 & \text{für} \quad \alpha \leqq \sqrt{3} \\ 2\alpha/\sqrt{1+\alpha^2} & \text{für} \quad \alpha \geqq \sqrt{3}. \end{cases}$$

Dann enthält $\{x \mid |x_1|+\alpha|x_2|+\beta|x_3| \leqq \gamma\}$ eine dichteste Packung von drei Einheitskugeln.

Zum Beweis von Satz 1 benötigen wir zwei Hilfssätze. Es sei $e_1 := (1, 0, 0)$, $e_2 := (0, 1, 0)$, $e_3 := (0, 0, 1)$ und die gewöhnliche euklidische Norm werde mit $\| \ \|$ bezeichnet. $\langle , \rangle$ und $\wedge$ bezeichnen das innere und das äußere Produkt. Im Hilfssatz 1 betrachten wir einen Bereich A, dessen Rand aus $n (\in \{5, 8\})$ glatten Flächenstücken besteht. Unter einer *Kante* von A verstehen wir einen nichtleeren Durchschnitt zweier dieser Flächenstücke.

HILFSSATZ 1. *Es seien* B, C *die Bereiche* $B := \{x \mid |x_1|+|x_2|+|x_3| \leqq 1\}$, $C := \{x \mid |x_1|+|x_2|+|x_3| \leqq 1, \|e_3+x\| \geqq \sqrt{2}\}$ *und sei* $A \in \{B, C\}$.

Dann gibt es Punkte q_1, q_2, q_3 *auf Kanten von* A *mit*

$$\min_{1 \leqq i < j \leqq 3} \|q_i - q_j\| = \sup_{\{p_1, p_2, p_3\} \subset A} \min_{1 \leqq i < j \leqq 3} \|p_i - p_j\|.$$

BEWEIS. Der Rand von A besteht aus den Flächenstücken $F_i := A \cap \{x \mid f_i(x) = 1\}$ $1 \leqq i \leqq n$, $(n \in \mathbf{N})$, wobei die $f_i : \mathbf{R}^3 \to \mathbf{R}$ folgendermaßen definiert sind:

1. $A := B$. $n := 8$, $f_i(x) := \langle (\varepsilon_1, \varepsilon_2, \varepsilon_3), x \rangle$ mit $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{1, -1\}$.
2. $A := C$. $n := 5$, $f_i(x) := \langle (\varepsilon_1, \varepsilon_2, 1), x \rangle$ mit $1 \leqq i \leqq 4$, $\varepsilon_1, \varepsilon_2 \in \{1, -1\}$ und

$$f_5(x) := 2 - \frac{1}{\sqrt{2}} \|e_3 + x\|.$$

Wir definieren eine Funktion $e : \mathbf{R}^3 \to \mathbf{R}$ durch $e(x) := \sum_{1 \leqq i \leqq n, \ x \in F_i} 1$ und beweisen die folgende Eigenschaft von A, die unten benötigt wird:

Zu je drei Punkten $p_1, p_2, p_3 \in A$ gibt es Punkte p_1', p_2', p_3' mit

(1)
$$\min_{1 \leqq i \leqq 3} e(p_i') \geqq 1, \quad \min_{1 \leqq i < j \leqq 3} \|p_i' - p_j'\| \geqq \min_{1 \leqq i < j \leqq 3} \|p_i - p_j\|.$$

Wegen $p_1 \in A$ gibt es ein $p_1' \in \{x \mid x = p_1 + \lambda(p_1 - p_2) \wedge (p_1 - p_3), \ \lambda \geqq 0\} \cap \{x \mid e(x) \geqq 1\}$. Für p_1' gilt: $e(p_1') \geqq 1$, $\|p_1' - p_2\| \geqq \|p_1 - p_2\|$, $\|p_1' - p_3\| \geqq \|p_1 - p_3\|$. Vom Punktetripel p_1', p_2, p_3 ausgehend schließt man analog weiter und beweist so (1).

Nun gilt folgende Aussage:

(2) Wenn $\{p_1, p_2, p_3\} \subset A$, $e(p_1) = 1$, $e(p_2) \in \{1, 2\}$ ist, dann gibt es Punkte $p_1', p_2', p_3' \in A$ mit

$$\min_{1 \leqq i < j \leqq 3} \|p_i' - p_j'\| > \min_{1 \leqq i < j \leqq 3} \|p_i - p_j\|.$$

Es ist $A = \{x \mid f_i(x) \leqq 1, 1 \leqq i \leqq n\}$. Wegen $e(p_1) = 1$ gibt es ein $j \in \{1, \ldots, n\}$ sodaß $f_j(p_1) = 1$, $\max_{1 \leqq i \leqq n, \ i \neq j} f_i(p_1) < 1$. Die Gerade

$$\left\{x \mid x = p_1 + \lambda(p_1 - p_2) \wedge \left(\frac{\partial f_j}{\partial x_1}(p_1), \frac{\partial f_j}{\partial x_2}(p_1), \frac{\partial f_j}{\partial x_3}(p_1)\right), \lambda \in \mathbf{R}\right\}$$

enthält einen Punkt p_1' mit

$$\|p_1'-p_2\| > \|p_1-p_2\| \tag{3}$$

und

$$p_1' \in A, \quad \|p_1'-p_3\| > \|p_1-p_3\|. \tag{4}$$

Wir greifen aus $\{1, \ldots, n\}$ zwei Elemente k, e heraus, daß

$$\max_{1 \leqq i \leqq n,\ i \neq k,\ i \neq e} f_i(p_2) \leqq f_k(p_2) \leqq f_e(p_2).$$

Nach (3) gibt es einen Punkt

$$p_2' \in \left\{x \mid x = p_2 + \lambda \left(\frac{\partial f_k}{\partial x_1}(p_2), \frac{\partial f_k}{\partial x_2}(p_2), \frac{\partial f_k}{\partial x_3}(p_2)\right) \wedge \left(\frac{\partial f_e}{\partial x_1}(p_2), \frac{\partial f_e}{\partial x_2}(p_2), \frac{\partial f_e}{\partial x_3}(p_2)\right),\ \lambda \in \mathbf{R}\right\}$$

mit

$$p_2' \in A, \|p_1'-p_2'\| > \|p_1-p_2\|, \|p_2'-p_3\| > \|p_2-p_3\|. \tag{5}$$

Mit $p_3' := p_3$ und (4), (5) ist (2) bewiesen.

Da die Menge A kompakt ist, nimmt die Funktion

$$m\colon A^3 \to \mathbf{R},\ m(p_1, p_2, p_3) := \min_{1 \leqq i < j \leqq 3} \|p_i - p_j\|$$

einen größten Wert an. Nach (1) gibt es Punkte g_1, g_2, g_3 mit

$$1 \leqq e(g_1) \leqq e(g_2) \leqq e(g_3), \quad m(g_1, g_2, g_3) = \sup_{\{p_1, p_2, p_3\} \subset A} m(p_1, p_2, p_3).$$

Wegen (2) können wir uns auf folgende zwei Fälle beschränken:

1. $\min\limits_{1 \leqq i \leqq 3} e(g_i) \geqq 2$. Hier gilt Hilfssatz 1 mit $q_i := g_i$, $1 \leqq i \leqq 3$.

2. $e(g_1) = 1$, $e(g_2) = e(g_3) = 3$. Dann ist $g_1 \in A \subset \{x \mid \|x\| \leqq 1\}$,

$\{g_2, g_3\} \subset \{e_1, e_2, e_3, -e_1, -e_2, -e_3\}$, daher $\min\limits_{1 \leqq i < j \leqq 3} \|g_i - g_j\| \leqq \sqrt{2}$.

Mit $q_i := e_i$, $i \in \{1, 2, 3\}$ ist Hilfssatz 1 bewiesen.

HILFSSATZ 2. *Es sei σ die Seitenlänge des regulären Oktaeders, das eine dichteste Packung von n Einheitskugeln enthält. Ferner sei $\delta := \sup \min\limits_{1 \leqq i < j \leqq n} \|p_i - p_j\|$, wobei das Supremum über alle Anordnungen von n Punkten $p_i \in \{x \mid |x_1| + |x_2| + |x_3| \leqq 1\}$ genommen wird. Dann gilt: $\sigma = \sqrt{6} + 2\sqrt{2}/\delta$.*

BEWEIS. Der Bereich $\left\{x \mid |x_1| + |x_2| + |x_3| \leqq \frac{\sigma}{\sqrt{2}}\right\}$ enthält eine Packung von n Einheitskugeln. Die Mittelpunkte dieser Kugeln liegen in $\left\{x \mid |x_1| + |x_2| + |x_3| \leqq \frac{\sigma}{\sqrt{2}} - \sqrt{3}\right\}$ und haben paarweise Abstand $\geqq 2$. Daher ist

$$\sigma \geqq \sqrt{6} + \frac{2\sqrt{2}}{\delta}. \tag{6}$$

Der Bereich $\left\{x \,|\, |x_1|+|x_2|+|x_3| \leqq \frac{2}{\delta}\right\}$ enthält n Punkte p_i mit paarweisem Abstand $\geqq 2$. Die Einheitskugeln mit Mittelpunkten in p_i bilden eine Packung in $\left\{x \,|\, |x_1|+|x_2|+|x_3| \leqq \frac{2}{\delta}+\sqrt{3}\right\}$. Daher gilt:

$$\sigma \leqq \sqrt{6}+\frac{2\sqrt{2}}{\delta}. \tag{7}$$

Aus (6), (7) folgt die Behauptung.

BEWEIS VON SATZ 1. Wir bezeichnen die abgeschlossene Strecke zwischen zwei Punkten x, y mit $[x, y]$. Sei $B:=\{x \,|\, |x_1|+|x_2|+|x_3| \leqq 1\}$, $e_1:=(1, 0, 0)$, $e_2:=(0, 1, 0)$, $e_3:=(0, 0, 1)$ und sei

$$\delta := \sup_{\{p_1, p_2, p_3\} \subset B} \min_{1 \leqq i < j \leqq 3} \|p_i - p_j\| \tag{8}$$

definiert. Die Punkte $(\sqrt{3}-1, 0, \sqrt{3}-2)$, $(1-\sqrt{3}, 0, \sqrt{3}-2)$, e_3 liegen in B und haben paarweise Abstand $2(\sqrt{3}-1)$. Daher ist

$$\delta \geqq 2(\sqrt{3}-1). \tag{9}$$

Im folgenden zeigen wir, daß

$$\delta \leqq 2(\sqrt{3}-1) \tag{10}$$

ist. Nach Hilfssatz 1 genügt es, (10) für Punkte p_1, p_2, p_3 zu beweisen, die auf Kanten von B liegen. O. B. d. A. sei $p_1 \in [e_1, e_2]$. Bei p_2 beschränken wir uns auf folgende Fälle:

1. $p_2 \in [-e_1, -e_2]$, 2. $p_2 \in [-e_1, e_2]$, 3. $p_2 \in [-e_1, -e_3]$.

Im ersten Fall nehmen wir $p_3 \in [-e_1, e_2]$ an, die Fälle 2, 3 unterteilen wir folgendermaßen: 2.1. $p_3 \in [e_1, -e_2]$, 2.2. $p_3 \in [-e_2, e_3]$, 3.1. $p_3 \in [e_1, -e_2]$, 3.2. $p_3 \in [-e_1, e_3]$, 3.3. $p_3 \in [-e_2, e_3]$. Die Fälle 1, 2.1 sind trivial — hier liegen p_1, p_2, p_3 in einem Quadrat. Fall 2.2 wird durch $\|p_1-p_3\| \leqq \|e_2+p_1\|$, $\|p_2-p_3\| \leqq \|e_2+p_2\|$ auf Fall 1 zurückgeführt. Ferner werden die Fälle 3.1, 3.2 durch Fall 2.2 erledigt.

Wir untersuchen nun Fall 3.3: $p_1 \in [e_1, e_2]$, $p_2 \in [-e_1, -e_3]$, $p_3 \in [-e_2, e_3]$.

Für passende $\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$ ist $p_1=(1-\lambda_1, \lambda_1, 0)$, $p_2=(-\lambda_2, 0, \lambda_2-1)$, $p_3=(0, \lambda_3-1, \lambda_3)$. Die Ungleichungen

$$\lambda_1^2+\lambda_2^2 > \lambda_1\lambda_2+\lambda_1, \quad \lambda_2^2+\lambda_3^2 > \lambda_2\lambda_3+\lambda_2, \quad \lambda_1^2+\lambda_3^2 > \lambda_1\lambda_3+\lambda_3 \tag{11}$$

stehen zu $0 \leqq \lambda_i \leqq 1$, $i \in \{1, 2, 3\}$ im Widerspruch. Daher gilt (11) nicht,

$$\min_{1 \leqq i < j \leqq 3} \|p_i - p_j\| = \sqrt{2} \min^{\frac{1}{2}} \{\lambda_1^2+\lambda_2^2-\lambda_1\lambda_2-\lambda_1+1, \lambda_2^2+\lambda_3^2-\lambda_2\lambda_3-\lambda_2+1,$$
$$\lambda_1^2+\lambda_3^2-\lambda_1\lambda_3-\lambda_3+1\} = \sqrt{2} < 2(\sqrt{3}-1).$$

Damit ist (10) gezeigt. Nach (9) ist $\delta=2(\sqrt{3}-1)$ und wegen (8) und Hilfssatz 2 der Beweis geführt.

3. Dichteste Packungen von 4, 5 und 6 Einheitskugeln im Oktaeder

SATZ 2. *Die Seitenlängen der regulären Oktaeder, die dichteste Packungen von vier, fünf und sechs Einheitskugeln enthalten, sind alle gleich* $2+\sqrt{6}$.

Zum Beweis benötigen wir den folgenden

HILFSSATZ 3. *Es sei* $\alpha\in\left]0, \frac{1}{2}\right]$, $p\in P:=\{x\,|\,|x_1|+|x_2|\leqq\alpha,\ x_3=\alpha-1\}$ *und* $q\in Q:=$ $:=\{x\,|\,|x_1|+|x_2|\leqq 1,\ |x_1|+|x_2|-x_3\leqq 1,\ |x_1|\leqq 1-\alpha,\ |x_2|\leqq 1-\alpha,\ x_3\geqq\alpha-1,\ \|e_3+x\|\leqq\sqrt{2}\}$. *Dann gilt:* $\|p-q\|\leqq\sqrt{2}$.

BEWEIS. Seien P, Q die im Satz definierten Bereiche. Für $\alpha=0$ gilt der Hilfssatz trivialerweise. Sei also $\alpha\in\left]0, \frac{1}{2}\right]$. Der Abstand von einem $p\in P$ zu $q\in Q$ ist am größten, wenn p Ecke von P ist. Aus Symmetriegründen ist

$$\sup_{p\in P,\,q\in Q}\|p-q\| = \sup_{q\in Q}\|p_0-q\|,\quad p_0:=(\alpha,0,\alpha-1). \tag{12}$$

Wir behaupten:

(13) Ist q keine Ecke von Q, so gibt es ein $q'\in Q$ mit $\|p_0-q'\|>\|p_0-q\|$, wobei wir unter "Ecke von Q" einen der unten angeführten Punkte q_i, $i\in\{1,\ldots,24\}$ verstehen.

Wir beschränken uns darauf, daß $q\in K:=Q\cap\{x\,|\,\|e_3+x\|=\sqrt{2}\}$. K ist ein Kugelstück, dessen Rand aus acht Kreisbögen besteht. Liegt q im Inneren von K, dann ist $\|p_0-q'\|>\|p_0-q\|$ für jedes $q'\in K\cap\{x\,|\,\langle p_0-q, x\rangle=\langle p_0-q, q\rangle, x\neq q\}$; wegen $\alpha>0$ ist diese Menge $\neq\{\ \}$. Sei nun q auf dem Rand von K, z.B. sei $q\in K\cap\{x\,|\,x_1=\alpha-1\}$. Durch $t\to x=(\alpha-1, \sqrt{1+2\alpha-\alpha^2}\sin t, \sqrt{1+2\alpha-\alpha^2}\cos t-1)$ wird ein Intervall $[-t_0, t_0]$ auf $K\cap\{x\,|\,x_1=\alpha-1\}$ abgebildet, wobei x für $t=\pm t_0$ Ecke von Q ist. $\|p_0-x\|^2=2+2\alpha-2\alpha\sqrt{1+2\alpha-\alpha^2}\cos t$ ist bei $t=\pm t_0$ am größten. Analog zeigt man (13), wenn q auf einem anderen zum Rand von K gehörenden Kreisbogen liegt.

Aus (12), (13) folgt: $\sup_{p\in P,\,q\in Q}\|p-q\| = \max_{1\leqq i\leqq 24}\|p_0-q_i\|$, wobei q_i, $i\in\{1,\ldots,24\}$ folgende Punkte sind: $(\pm(1-\alpha), \pm\alpha, 0)$, $(\pm(1-\alpha), \pm\alpha, \beta)$ mit $\beta:=\sqrt{1+2\alpha-2\alpha^2}-1$, $(\pm(1-\alpha), 0, -\alpha)$, $(\pm\alpha, \pm(1-\alpha), 0)$, $(\pm\alpha, \pm(1-\alpha), \beta)$, $(0, \pm(1-\alpha), -\alpha)$, $(0, \pm\alpha, \alpha-1)$, $(\pm\alpha, 0, \alpha-1)$. Für $i\in\{1,\ldots,24\}$, $\alpha\in\left]0, \frac{1}{2}\right]$ ist $\|p_0-q_i\|\leqq\sqrt{2}$ und damit Hilfssatz 3 bewiesen.

BEWEIS VON SATZ 2. Zunächst zeigen wir, daß

$$\min_{1\leqq i<j\leqq 3}\|p_i-p_j\|\leqq\sqrt{2},\ \text{wenn}\ \{p_1,p_2,p_3\}\subset A:=\{x\,|\,|x_1|+|x_2|+|x_3|\leqq 1,\ \|e_3+x\|\geqq\sqrt{2}\}. \tag{14}$$

Nach Hilfssatz 1 genügt es, (14) für Punkte p_1, p_2, p_3 zu beweisen, die auf Kanten von A liegen (Definition einer „Kante von A" bei Hilfssatz 1). Eine Kante von A zwischen zwei Ecken x, y von A sei mit $[x, y]$ bezeichnet. Wir beschränken uns auf folgende zwei Fälle:

1. $p_1 \in [e_1, e_3]$, $p_2 \in [-e_1, e_2]$, $p_3 \in [-e_1, -e_2]$.
Wir unterscheiden nochmals:

1.1. $\{p_2, p_3\} \subset \left\{x \mid x_1 \leqq -\frac{1}{3}\right\}$. p_2, p_3 liegen in

$$\left\{x \mid -x_1 + |x_2| + x_3 \leqq 1,\ x_1 \leqq -\frac{1}{3},\ x_3 \geqq 0\right\}$$

mit Durchmesser $< \sqrt{2}$.

1.2. $p_2 \in \left\{x \mid x_1 \geqq -\frac{1}{3}\right\}$. Dann ist

$$p_2 \in R := \left\{x \mid -x_1 + x_2 + x_3 = 1,\ -\frac{1}{3} \leqq x_1 \leqq 0,\ 0 \leqq x_3 \leqq \frac{1}{3}\right\}.$$

R ist ein Rhombus und hat folgende Ecken:

$$e_2, \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(-\frac{1}{3}, \frac{2}{3}, 0\right), \left(0, \frac{2}{3}, \frac{1}{3}\right).$$

Aus

$$\|e_3 + p_2\| = \|e_3 + e_2\| = \left\|e_3 + \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)\right\| = \sqrt{2}, \quad \left\|e_3 + \left(-\frac{1}{3}, \frac{2}{3}, 0\right)\right\| < \sqrt{2}$$

folgt:

(15) p_2 liegt im Dreieck mit den Ecken $e_2, \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right)$.

Daher ist

$$\|p_1 - p_2\| \leqq \max\left\{\|x - y\| \,\middle|\, x \in \left\{e_2, \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right)\right\},\ y \in \{e_1, e_3\}\right\} = \sqrt{2}.$$

2. $p_1 \in [e_1, e_2]$, $p_2 \in [-e_1, e_2]$, $p_3 \in [-e_1, -e_2]$. Wir geben einige Dreiecke $D_{11}, \ldots, D_{32}$ durch ihre Ecken an:

D_{11}: $e_2, \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $\quad D_{12}$: $e_1, \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, 0, \frac{1}{3}\right)$,

D_{21}: $e_2, \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $\quad D_{22}$: $-e_1, \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(-\frac{2}{3}, 0, \frac{1}{3}\right)$,

D_{31}: $-e_2, \left(-\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}\right), \left(0, -\frac{2}{3}, \frac{1}{3}\right)$, $\quad D_{32}$: $-e_1, \left(-\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}\right), \left(-\frac{2}{3}, 0, \frac{1}{3}\right)$.

Wir unterscheiden bei p_i, $1 \leqq i \leqq 3$ die Fälle $|x_1| \leqq \frac{1}{3}$ bzw. $|x_1| \geqq \frac{1}{3}$ und zeigen, ähnlich wie wir (15) bewiesen, daß $p_1 \in D_{11} \cup D_{12}$, $p_2 \in D_{21} \cup D_{22}$, $p_3 \in D_{31} \cup D_{32}$. Sei $i \in \{1, 2, 3\}$ und $\triangle_i$ ein Dreieck $\in \{D_{i1}, D_{i2}\}$, das p_i enthält. Dann gilt: $\min\limits_{1 \leqq i < j \leqq 3} \|p_i - p_j\| \leqq m := \min\limits_{1 \leqq i < j \leqq 3} \max\{\|x - y\| \,|\, x = \text{Ecke von } \triangle_i,\ y = \text{Ecke von } \triangle_j\}$. Es gibt acht Fälle dafür, daß $\triangle_1 \in \{D_{11}, D_{12}\}$, $\triangle_2 \in \{D_{21}, D_{22}\}$, $\triangle_3 \in \{D_{31}, D_{32}\}$ ist, und in jedem dieser Fälle ist $m \leqq \sqrt{2}$. Daher ist (14) bewiesen.

Wir behaupten weiter:

(16) Für je vier Punkte $\in B := \{x \mid |x_1| + |x_2| + |x_3| \leqq 1\}$ gilt $\min\limits_{1 \leqq i < j \leqq 4} \|p_i - p_j\| \leqq \sqrt{2}$.

Unter den Koordinatenbeträgen von p_1, p_2, p_3, p_4 betrachten wir den größten. O. B. d. A. sei dies der Betrag der x_3-Koordinate ξ von p_4, und $\xi<0$. Wir unterscheiden wieder zwei Fälle:

I. $\{x \mid \|e_3+x\| \leqq \sqrt{2}\}$ enthält einen Punkt $p \in \{p_1, p_2, p_3\}$.

Für passendes $\alpha \geqq 0$ ist $\xi=\alpha-1$. Wenn $\alpha \geqq \frac{1}{2}$ ist, liegen die Punkte p_i, $1 \leqq i \leqq 4$ in einem Würfel der Kantenlänge 1. In [8] wird gezeigt, daß daraus folgt:

$$\min_{1 \leqq i<j \leqq 4} \|p_i-p_j\| \leqq \sqrt{2}.$$

Wenn aber $\alpha \in \left[0, \frac{1}{2}\right]$ ist, dann gilt (16) wegen

$$p \in \{x \mid |x_1|+|x_2| \leqq 1,\ |x_1|+|x_2|-x_3 \leqq 1,\ |x_1| \leqq 1-\alpha,\ |x_2| \leqq 1-\alpha,\ x_3 \geqq \alpha-1,\ \|e_3+x\| \geqq \sqrt{2}\},$$

$$p_4 \in \{x \mid |x_1|+|x_2| \leqq \alpha,\ x_3=\alpha-1\}$$

und wegen Hilfssatz 3.

II. $\{p_1, p_2, p_3\} \subset \{x \mid \|e_3+x\| \geqq \sqrt{2}\}$. Hier folgt (16) sofort aus (14). Aus (16) folgt:

$$\delta := \sup_{\{p_1, p_2, p_3, p_4\} \subset B} \min_{1 \leqq i<j \leqq 4} \|p_i-p_j\| \leqq \sqrt{2}.$$

Da vier Ecken von B paarweise Abstand $\geqq \sqrt{2}$ haben, ist $\delta=\sqrt{2}$. Nach Hilfssatz 2 beträgt die Seitenlänge des Oktaeders mit einer dichtesten Packung von vier Einheitskugeln $\sigma=2+\sqrt{6}$. Da dieses Oktaeder sogar eine Packung von sechs Einheitskugeln enthält, ist Satz 2 bewiesen.

LITERATUR

[1] BARANOVSKIĬ, E. P.: Packings, coverings, partitionings and certain other distributions in spaces of constant curvature, *Progress math.* **9** (1971), 209—253.

[2] BLACHMANN, N. M.: The closest packing of equal spheres in a larger sphere, *Amer. Math. Monthly* **70** (1963), 526—529.

[3] FEJES TÓTH, L.: *Lagerungen in der Ebene, auf der Kugel und im Raum*, Berlin, Springer-Verlag, 1953.

[4] FEJES TÓTH, L.: *Regular Figures*, Oxford, Pergamon Press, 1964.

[5] GOLDBERG, M.: On the densest packing of equal spheres in a cube, *Math. Mag.* **44** (1971), 199—208.

[6] HORVÁTH, J.: The densest packing of an n-dimensional cylinder by unit spheres (Russian), *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **15** (1972), 139—143.

[7] SCHAER, I.: The densest packing of 9 circles in a square, *Can. Math. Bull.* **8** (1965), 273—277.

[8] SCHAER, I.: On the densest packing of spheres in a cube, *Can. Math. Bull.* **9** (1966), 265—270.

[9] SCHAER, I.: The densest packing of five spheres in a cube, *Can. Math. Bull.* **9** (1966), 271—274.

[10] SCHAER, I.: The densest packing of six spheres in a cube, *Can. Math. Bull.* **9** (1966), 275—280.

Institut für Analysis, Technische Universität
A—1040 Wien, Gußhausstr. 27

(Eingegangen am 1. Dezember 1977)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 345—369.

ON THE REMAINDER TERM OF THE PRIME NUMBER FORMULA III

Sign changes of $\pi(x)-\operatorname{li} x$

by

J. PINTZ

1. Riemann [14] asserted in his famous paper in 1859

$$\pi(x) < \operatorname{li} x = \int_0^x \frac{dt}{\log t} \qquad (x > 2). \tag{1.1}$$

About 50 years later E. SCHMIDT [15] proved that dealing with the function

$$\Pi(x) = \sum_{p^m \leq x} \frac{1}{m} = \sum_{\nu \geq 1} \frac{1}{\nu} \pi\left(x^{\frac{1}{\nu}}\right) \tag{1.2}$$

the analogous difference $\Pi(x)-\operatorname{li} x$ changes sign infinitely often as $x \to \infty$. This naturally could not decide Riemann's problem, since

$$\Pi(x)-\pi(x) \sim \frac{\sqrt{x}}{\log x}. \tag{1.3}$$

On the other hand he observed that if (1.1) is true then the famous Riemann conjecture concerning the zeros of $\zeta(s)$ is also true. This remark made naturally even more interesting the question whether (1.1) is true or not.

The truth of (1.1) seemed to be supported by the calculations of D. N. LEHMER [9], who showed in 1914 that (1.1) is valid for all $x < 10^7$. But LITTLEWOOD [10] in the same year disproved Riemann's assertion (1.1) proving that $\pi(x)-\operatorname{li} x$ has infinitely many sign changes.

2. However, after Littlewood's work the curious problem arose that it was known that $\pi(x)-\operatorname{li} x$ changes sign infinitely many times but no explicit bound X could be given that (1.1) would be false for any $x < X$. Namely, Littlewood's proof was completely ineffective and so it could not give any explicit upper bound for the first sign change of $\pi(x)-\operatorname{li} x$. The problem of the effectivization turned out to be a very deep one. Only 14 years later in 1955 could SKEWES [16] give the first explicite, however, incredibly large upper bound

$$e_4(7{,}705) \tag{2.1}$$

for the first sign change, where $e_4(x)$ means the four times iterated exponential function. In 1966 this was improved by SHERMAN LEHMAN [8] to

$$1{,}65 \cdot 10^{1165}. \tag{2.2}$$

3. Littlewood's proof also did not give any information about the problem

how often $\pi(x)-\operatorname{li} x$ changes his sign. So the more deep and general problem was: what can one say from the oscillatorical behaviour of various forms of the remainder term of the prime number formula.

Let

$$\begin{aligned} \Delta_1(x) &\overset{\text{def}}{=} \pi(x)-\operatorname{li} x, \\ \Delta_2(x) &\overset{\text{def}}{=} \Pi(x)-\operatorname{li} x, \\ \Delta_3(x) &\overset{\text{def}}{=} \theta(x)-x = \sum_{p \leqq x} \log p - x, \\ \Delta_4(x) &\overset{\text{def}}{=} \psi(x)-x = \sum_{p^m \leqq x} \log p - x, \end{aligned} \tag{3.1}$$

and let $V_i(Y)$ denote the number of sign changes of $\Delta_i(x)$ in $[2, Y]$.

The first result in this direction is due to PÓLYA [12] who showed

$$\overline{\lim_{Y\to\infty}} \frac{V_4(Y)}{\log Y} > 0 \tag{3.2}$$

but he could not prove anything for the more difficult case $i=1$. For $V_1(Y)$ the first and very important though conditional result was proved by INGHAM [3] in 1936. Let θ denote the least upper bound of the real parts of the zeros of $\zeta(s)$. Ingham showed that if there is a zero on the line $\sigma=\theta$ then for $Y>Y_0$ $\Delta_1(x)$ has a sign change in every interval of the form

$$[Y, c_0 Y] \tag{3.3}$$

with a constant c_0. Ingham's result trivially implies that if the condition is satisfied then

$$\underline{\lim_{Y\to\infty}} \frac{V_1(Y)}{\log Y} > 0. \tag{3.4}$$

We must note, however, that Ingham's condition is very deep. It is satisfied naturally if, e.g., the Riemann hypothesis is true, but it implies $\theta<1$, i.e., the so called quasi Riemann hypothesis. A second disadvantage of this beautiful theorem is that the constant c_0 and thus also the lower bound for the left side of (3.4) cannot be computed effectively even if we suppose the Riemann hypothesis.

4. The first unconditional lower bound for $V_1(Y)$ was given by S. KNAPOWSKI [4], [5] in 1961 and 1962 using Turán's method. He proved the completely effective lower estimate

$$V_1(Y) > e^{-35} \log_4 Y \quad \text{for} \quad Y > e_5(35) \tag{4.1}$$

where $\log_4 Y$ denotes the four times iterated logarithm function, i.e., $\log_4 Y = \log\log\log\log Y$. He also showed the sharper but ineffective inequality

$$\underline{\lim_{Y\to\infty}} \frac{V_1(Y)}{\log_2 Y} > 0. \tag{4.2}$$

These results were improved in 1974—76 by S. KNAPOWSKI and P. TURÁN [6], [7]. In the first work [6] they proved that $\pi(x)-\mathrm{li}\, x$ changes sign in every interval of the form

$$[Y,\ Y\exp\{\log^{3/4} Y(\log_2 Y)^4\}] \tag{4.3}$$

if $Y>Y_0$ ineffective constant. This implies for $V_1(Y)$ the ineffective inequality

$$\varlimsup_{Y\to\infty} \frac{V_1(Y)}{\log^{1/4} Y(\log_2 Y)^{-4}} > 0. \tag{4.4}$$

In the second work [7] they showed the existence of effectively computable positive constants c_1 and c_2 such that for $Y>c_1$ the inequality

$$V_1(Y) > c_2 \log_3 Y \tag{4.5}$$

holds.

This result was recently improved by the author [11] to

$$V_1(Y) > C_1(\log_2 Y)^{C_2} \quad \text{for} \quad Y > C_3 \tag{4.6}$$

where C_1, C_2 and C_3 are positive effective constants.

5. In this paper we shall prove using Turán's method

THEOREM 1. *For* $Y>Y_i$ $(1\leqq i\leqq 4)$ *the interval*

$$[Y,\ Y\exp\{63\sqrt{\log Y}\log_2 Y\}] \tag{5.1}$$

contains a sign-change of $\Delta_i(x)$ $(1\leqq i\leqq 4)$, *where* Y_1 *and* Y_3 *are ineffective constants,* Y_2 *and* Y_4 *are effective ones.*

This theorem already implies the partially ineffective inequality (see **10**)

$$\varlimsup_{Y\to\infty} \frac{V_i(Y)}{\sqrt{\log Y}(\log_2 Y)^{-1}} > 0. \tag{5.2}$$

However, using a more explicit form of Theorem 1 it will be possible to deduce from it the inequality (5.2) in an effective form for all $i\leqq 4$, namely we state further

THEOREM 2. *There exist effectively computable constants* c_3 *and* c_4 *such that for* $Y>c_3$ *the inequality*

$$V_i(Y) > c_4 \frac{\sqrt{\log Y}}{\log_2 Y} \qquad (1\leqq i\leqq 4) \tag{5.3}$$

holds.

The reason why we stated the seemingly weaker (because for $i=1, 3$ ineffective) Theorem 1 as a separate theorem is that it contains a localization for the sign changes of $\Delta_i(x)$, whereas Theorem 2 gives only a lower bound for the total number of sign changes without any localization (or more precisely with a very weak but effective localization, which one can read out from the proof, namely the interval $[\exp(c\sqrt{\log Y}(\log_2 Y)^{-1});\ Y]$.

We mention that part IV of this series will be devoted to the proof of the partially ineffective improvement of (5.2), namely, we shall prove

$$V_i(Y) > 10^{-11} \frac{\log Y}{(\log_2 Y)^3} \quad \text{for} \quad Y > Y_i \tag{5.4}$$

where the constants Y_i are effective for $i=2$ and 4 and ineffective for $i=1$ and 3.

6. We shall give only the proof for the most interesting and deep case $i=1$. Our proof implicitly contains the case $i=2$, too, with ineffective Y_2. In **24** we shall mention the slight changes in the course of proof of Theorem 1 which make possible to get for Y_2 an explicitly calculable value. The cases $i=3$ and 4 are very similar to the cases $i=1$ and 2, resp., they are even easier, so we do not work them out.

Theorem 1 will be the immediate consequence of the following

LEMMA 1. *If for a* $Z > c_5$ *(effective constant) the function* $\zeta(s)$ *has a zero* $\varrho^* = \beta^* + i\gamma^*$ *with*

$$\beta^* \geqq \frac{1}{2} + \frac{10 \log \gamma^*}{4\sqrt{\log Z}\,(\log_2 Z)^{-1}},$$

$$0 < \gamma^* \leqq \exp\left(\frac{1}{10}\sqrt{\log Z}\,(\log_2 Z)^{-1}\right) \tag{6.1}$$

then the interval

$$I(Z) = \left(Z \exp\left(-31\sqrt{\log Z}\log_2 Z\right),\ Z \exp\left(31\sqrt{\log Z}\log_2 Z\right)\right) \tag{6.2}$$

contains a sign-change of $\Delta_1(x)$.

Namely, it is easy to see that if the Riemann conjecture is not true, than any zero $\varrho^* = \beta^* + i\gamma^*$ with $\beta^* > \frac{1}{2}$ satisfies (6.1) if $Z > Z_0(\varrho^*)$, and thus in this case Theorem 1 follows from Lemma 1 indeed.

On the other hand if the Riemann conjecture is true then Ingham's quoted theorem (see 3.3) shows the validity of Theorem 1 even in a stronger form. (For the sake of completeness we note that Ingham's theorem with essentially unchanged proof is true for the cases $i=2, 3, 4,$ too.)

With the proof of our lemma we shall use the following notations:

$$L \overset{\text{def}}{=} \log Z, \tag{6.3}$$

$$M \overset{\text{def}}{=} 100(\log_2 Z)^2 = 100 \log^2 L, \tag{6.4}$$

$$\lambda \overset{\text{def}}{=} \frac{\sqrt{\log Z}}{10 \log_2 Z} = \sqrt{\frac{L}{M}}. \tag{6.5}$$

Let k be a real number to be determined later, which will be restricted at present only by

$$M \leqq k \leqq M\left(1 + \frac{1}{L}\right). \tag{6.6}$$

Let further

$$\mu \overset{\text{def}}{=} k\lambda^2, \tag{6.7}$$

$$A \overset{\text{def}}{=} \exp(\mu - 3k\lambda), \tag{6.8}$$

$$B \overset{\text{def}}{=} \exp(\mu + 3k\lambda), \tag{6.9}$$

$$f(x) \overset{\text{def}}{=} \Pi(x) - \lg x \pm \sqrt{x} \overset{\text{def}}{=} \Pi(x) - \sum_{2 \leqq n \leqq x} \frac{1}{\log n} \pm \sqrt{x}, \tag{6.10}$$

$$H(s) \overset{\text{def}}{=} \frac{\zeta'}{\zeta}(s) + \zeta(s) - 1 \mp \frac{1}{2\left(s - \frac{1}{2}\right)^2}, \tag{6.11}$$

where both in (6.10) and (6.11) the upper or in both the lower signs are meant.

Let us assume that $f(x)$ has no sign change in the interval

$$I' \overset{\text{def}}{=} [A, B] \subset I(Z). \tag{6.12}$$

We shall show that this assumption leads to contradiction, and thus proves the existence of

$$x', x'' \in I' \subset I(Z) \tag{6.13}$$

for which

$$\Pi(x) - \lg x' > \sqrt{x'} \tag{6.14}$$

and

$$\Pi(x'') - \lg x'' < -\sqrt{x''} \tag{6.15}$$

and so the inequalities

$$\pi(x') - \operatorname{li} x' > \frac{1}{2}\sqrt{x'} \tag{6.16}$$

and

$$\pi(x'') - \operatorname{li} x'' < -\frac{1}{2}\sqrt{x''} \tag{6.17}$$

hold owing to the trivial estimate

$$\Pi(x) - \pi(x) = O\left(\frac{\sqrt{x}}{\log x}\right) \tag{6.18}$$

and

$$\lg x = \operatorname{li} x + O(1). \tag{6.19}$$

7. We shall distinguish the following two cases.

Case A. There exists a zero $\varrho_0=\beta_0+i\gamma_0$ with

$$\beta_0 \geqq \frac{1}{2}+\frac{\log\gamma_0}{4\lambda}, \quad 0<\gamma_0 \leqq \lambda^5. \tag{7.1}$$

Then let $\varrho_1'=\beta_1'+i\gamma_1'$ be the zero with the maximal real part β_1' among those satisfying (7.1). Let successively ϱ_{n+1}' be the zero with the maximal real part β_{n+1}' satisfying

$$\gamma_n' \leqq \gamma_{n+1}' \leqq \gamma_n'+2\lambda, \quad \beta_{n+1}' \geqq \beta_n'+\frac{1}{\lambda} \tag{7.2}$$

if such a zero exists. Thus we get after at most $\left[\frac{\lambda}{2}\right]$ steps the existence of a zero $\varrho_N'=\beta_N'+i\gamma_N' \overset{\text{def}}{=} \varrho_1=\beta_1+i\gamma_1$ with

$$\beta_1 \geqq \frac{1}{2}+\frac{1}{\lambda}, \quad 0<\gamma_1 \leqq 2\lambda^5 \tag{7.3}$$

for which the domains

$$|t| \leqq \lambda^5, \quad \sigma > \beta_1 \tag{7.4}$$

and

$$|t-\gamma_1| \leqq 2\lambda, \quad \sigma > \beta_1+\frac{1}{\lambda} \tag{7.5}$$

are zero-free.

Case B. There is no zero satisfying (7.1).

Then let $\varrho_1=\beta_1+i\gamma_1$ be any zero satisfying (6.1); i.e., with our new notations in this case we have

$$\beta_1 \geqq \frac{1}{2}+\frac{\log\gamma_1}{4\lambda}, \quad \lambda^5<\gamma_1 \leqq e^{\lambda}. \tag{7.6}$$

8. Then in both cases our starting formula is:

$$\int_1^{\infty} f(x)\frac{d}{dx}(x^{-s}\log x)\,dx = H(s) \tag{8.1}$$

which can be proved easily by partial summation for $\sigma>1$.

Further we shall use the formula ($A>0$, B arbitrary complex)

$$\begin{aligned}\frac{1}{2\pi i}\int_{(2)} e^{As^2+Bs}\,ds &= \exp\left(-\frac{B^2}{4A}\right)\cdot\frac{1}{2\pi i}\int_{(2)} e^{\left(\sqrt{A}s+\frac{B}{2\sqrt{A}}\right)^2}\,ds = \\ &= \exp\left(-\frac{B^2}{4A}\right)\cdot\frac{1}{\sqrt{A}}\cdot\frac{1}{2\pi i}\int_{(0)} e^{z^2}\,dz = \frac{1}{2\sqrt{\pi A}}\exp\left(-\frac{B^2}{4A}\right).\end{aligned} \tag{8.2}$$

Using (8.1) with $s+i\gamma_1$ instead of s, multiplying both sides by $e^{ks^2+\mu s}$ and integrating along $\sigma=2$ we get our basic identity as follows:

(8.3)

$$U=\frac{1}{2\pi i}\int_{(2)} H(s+i\gamma_1)e^{ks^2+\mu s}\,ds=$$

$$=\frac{1}{2\pi i}\int_{(2)}\int_1^\infty f(x)\frac{d}{dx}(x^{-s-i\gamma_1}\log x\cdot e^{ks^2+\mu s})\,dx\,ds=$$

$$=\int_1^\infty f(x)\frac{d}{dx}\left\{x^{-i\gamma_1}\log x\cdot\frac{1}{2\pi i}\int_{(2)} e^{ks^2+(\mu-\log x)s}\,ds\right\}dx=$$

$$=\frac{1}{2\sqrt{\pi k}}\int_1^\infty f(x)\frac{d}{dx}\left\{x^{-i\gamma_1}\log x\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)\right\}dx=$$

$$=\frac{1}{2\sqrt{\pi k}}\int_1^\infty \frac{f(x)}{x}x^{-i\gamma_1}\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)\left\{-i\gamma_1\log x+1+\log x\,\frac{\mu-\log x}{2k}\right\}dx.$$

9. The main idea of the proof is that supposing that $f(x)$ does not change its sign in I' one can deduce an upper bound for the absolute value of the right side of (8.3); on the other hand one can give a lower estimate for the absolute value of the left side of (8.3) choosing k suitably in the interval (6.6) and these two estimations will contradict.

In course of the proof of the lower bound for the left side of (8.3) essential role will be played by a powersum theorem of Vera T. Sós—P. Turán, which we state as Theorem A of the Appendix.

10. We split the integral U on the right side of (8.3) into the following three parts:

(10.1) $$U=U_1+U_2+U_3$$

where

(10.2) $$U_1=\int_1^A,\quad U_2=\int_A^B,\quad U_3=\int_B^\infty.$$

Considering our notations (6.7)—(6.12) and (8.3) we have

(10.3)

$$|U_2|\leqq\frac{1}{2\sqrt{\pi k}}\int_A^B\frac{|f(x)|\log x}{x}\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)\left(\gamma_1+\frac{1}{\log x}+\frac{|\mu-\log x|}{2k}\right)dx\leqq$$

$$\leqq\frac{1}{2\sqrt{\pi k}}\int_A^B\frac{|f(x)|\mu\left(1+\frac{3}{\lambda}\right)}{x}\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)\left(\gamma_1+1+\frac{3k\lambda}{2k}\right)dx\leqq$$

$$\leqq\frac{2\mu(\gamma_1+\lambda)}{2\sqrt{\pi k}}\int_A^B\frac{|f(x)|}{x}\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)dx=$$

$$=\frac{2\mu(\gamma_1+\lambda)}{2\sqrt{\pi k}}\left|\int_A^B\frac{f(x)}{x}\exp\left(-\frac{(\log x-\mu)^2}{4k}\right)dx\right|$$

since $f(x)$ does not change its sign in $[A, B]$.

On the other hand we can trivially estimate

$$|U_4| \overset{\text{def}}{=} \left| \int_B^\infty \frac{f(x)}{x} \exp\left(-\frac{(\log x-\mu)^2}{4k}\right) dx \right| \leqq$$

$$\leqq \int_B^\infty \exp\left(-\frac{(\log x-\mu)^2}{4k}\right) dx =$$

$$= \int_{3k\lambda}^\infty \exp\left(\mu+y-\frac{y^2}{4k}\right) dy \leqq \tag{10.4}$$

$$\leqq \int_{3k\lambda}^\infty \exp(\mu+y-2y-2k\lambda^2)\, dy \leqq \int_0^\infty e^{-\mu-y}\, dy = e^{-\mu}$$

and analogously

$$|U_5| \overset{\text{def}}{=} \left| \int_1^A \frac{f(x)}{x} \exp\left(-\frac{(\log x-\mu)^2}{4k}\right) dx \right| \leqq$$

$$\leqq \int_1^{e^{\mu-3k\lambda}} \exp\left(-\frac{9k^2\lambda^2}{4k}\right) dx = e^{\mu-3k\lambda-\frac{9}{4}\mu} \leqq e^{-\mu}. \tag{10.5}$$

Naturally we have mutatis mutandis

$$U_1 \leqq e^{-\mu} \quad \text{and} \quad U_3 \leqq e^{-\mu}. \tag{10.6}$$

Thus using (10.1)—(10.6) we can change the intervals in the left and right side of (10.3) from $[A, B]$ to $[1, \infty)$ and so with the notation

$$K \overset{\text{def}}{=} \frac{1}{2\sqrt{\pi k}} \int_1^\infty \frac{f(x)}{x} \exp\left(-\frac{(\log x-\mu)^2}{4k}\right) dx \tag{10.7}$$

we get

$$|U| = |U_2| + o(1) = 2\mu(\gamma_1+\lambda)|K| + o(1). \tag{10.8}$$

11. With the upper estimate of $|K|$ our starting formula will be

$$\int_1^\infty \frac{f(x)}{x^{s+1}}\, dx = \frac{1}{s}\left\{\int_2^s \left(\frac{\zeta'}{\zeta}(z)+\zeta(z)\right) dz + h\right\} \pm \frac{1}{s-\frac{1}{2}} =$$

$$\overset{\text{def}}{=} \varphi(s) \pm \frac{1}{s-\frac{1}{2}} \tag{11.1}$$

which is valid for $\sigma > 1$ with a constant h. This can be proved easily by partial integration.

Now multiplication by $e^{ks^2+\mu s}$ and integration along the line $\sigma = 2$ gives:

$$K = \pm\frac{1}{2\pi i} \int_{(2)} \frac{e^{ks^2+\mu s}}{s-\frac{1}{2}}\, ds + \frac{1}{2\pi i} \int_{(2)} \varphi(s) e^{ks^2+\mu s}\, ds =$$

$$\overset{\text{def}}{=} \pm K_1 + K_2. \tag{11.2}$$

Applying Cauchy's theorem for K_1 we get

$$K_1 = e^{\frac{k}{4}+\frac{\mu}{2}} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-kt^2+i\mu t}}{it-\frac{1}{2}}\,dt = e^{\frac{k}{4}+\frac{\mu}{2}} + O(1). \tag{11.3}$$

12. Now in Case A we transform the way of integration in K_2 on the broken line l defined for $t \geqq 0$ by

$$\begin{aligned}
&I_1\colon \sigma = \frac{5}{4} && \text{for} \quad t \geqq \lambda,\\
&I_2\colon \beta_1 + \frac{1}{\mu} \leqq \sigma \leqq \frac{5}{4} && \text{for} \quad t = \lambda,\\
&I_3\colon \sigma = \beta_1 + \frac{1}{\mu} && \text{for} \quad 10 \leqq t \leqq \lambda, \qquad (12.1)\\
&I_4\colon \frac{1}{4} \leqq \sigma \leqq \beta_1 + \frac{1}{\mu} && \text{for} \quad t = 10,\\
&I_5\colon \sigma = \frac{1}{4} && \text{for} \quad 0 \leqq t \leqq 10,
\end{aligned}$$

and for $t \leqq 0$ by reflection on the real axis, because by the choice of ϱ_1 (see (7.3)—(7.5)) $\varphi(s)$ is regular right of the broken line l and on l. Thus we have

$$K_2 = \frac{1}{2\pi i}\int_{(l)} \varphi(s) e^{ks^2+\mu s}\,ds. \tag{12.2}$$

We shall use the fact, that if $\zeta(s)$ has no zero in the domain

$$\sigma > \beta\left(\geqq \frac{1}{2}\right), \quad |t| \leqq T+1 \tag{12.3}$$

then for

$$\sigma \geqq \beta+\eta, \quad 2 \leqq |t| \leqq T \tag{12.4}$$

one has

$$\left|\frac{\zeta'}{\zeta}(z)\right| = O\left(\frac{\log|t|}{\eta}\right). \tag{12.5}$$

(This follows easily from Satz 4.1 of PRACHAR [13] (p. 225) in case of $k=1$.)

Further we use the classical estimate

$$|\zeta(z)| = O\left(\sqrt{|t|}\right) \quad \text{for} \quad \sigma \geqq \frac{1}{2}, \quad |t| \geqq 10. \tag{12.6}$$

From (12.5) and (12.6) we get by easy computation for the integrals J_i on the intervals I_i $(1 \leqq i \leqq 5)$ the estimates

$$|J_1| = O\left(\exp\left(k \cdot \frac{25}{16} - k\lambda^2 + \frac{5}{4}\mu\right)\right) \leqq e^{\frac{\mu}{3}},$$

$$|J_2| = O\left(\log\lambda \cdot \mu \exp\left(k \cdot \frac{25}{16} - k\lambda^2 + \frac{5}{4}\mu\right)\right) \leqq e^{\frac{\mu}{3}},$$

$$|J_3| = O\left(\log\lambda \cdot \mu \exp\left(-98k + \mu\left(\beta_1 + \frac{1}{\mu}\right)\right)\right) \leqq e^{-97k+\mu\beta_1}, \tag{12.7}$$

$$|J_4| = O\left(\exp\left(-98k + \mu\left(\beta_1 + \frac{1}{\mu}\right)\right)\right) \leqq e^{-97k+\mu\beta_1},$$

$$|J_5| = O\left(\exp\left(\frac{k}{16} + \frac{\mu}{4}\right)\right) \leqq e^{\frac{\mu}{3}}.$$

Thus we have considering (11.2)—(11.3) and (12.7) the upper bound

$$|K| = O\left(e^{-97M+\mu\beta_1} + e^{\frac{k}{4}+\frac{\mu}{2}}\right). \tag{12.8}$$

Further, using (7.3), (10.8) and (12.8) we get

$$|U| \leqq e^{-96M\mu\beta_1} + e^{\frac{k}{4}+\mu\beta_1-\frac{\mu}{\lambda}+\lambda} \leqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{96M}}. \tag{12.9}$$

13. Now we shall give a lower bound for the absolute value of the integral U on the left of (8.3) by suitable choice of k using Theorem A of the Appendix. Shifting the line of integration to $\sigma = -\frac{1}{2}$ we get

$$U = \sum_{\varrho} \exp\{k[(\varrho - i\gamma_1)^2 + \lambda^2(\varrho - i\gamma_1)]\} \mp$$

$$\mp \frac{1}{2}\frac{d}{ds}(e^{ks^2+\mu s})_{s=\frac{1}{2}-i\gamma_1} + \tag{13.1}$$

$$+ \frac{1}{2\pi i}\int_{(-\frac{1}{2})} H(s+i\gamma_1)\exp(ks^2+\mu s)\,ds.$$

Easy computation shows that the last integral is $O(1)$ and the second residue is absolutely

$$\leqq \frac{1}{2}\left(2k\left|\frac{1}{2} - i\gamma_1\right| + \mu\right)e^{\frac{k}{4}+\frac{\mu}{2}-k\gamma_1^2} \leqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{99M}}. \tag{13.2}$$

Further we shall use that the number of zeros of $\zeta(s)$ in $T \leqq t \leqq T+1$ is

$$< c \log T \quad \text{where} \quad c = 15 \quad \text{for} \quad T > T_0 \tag{13.3}$$

(see, e.g., W. J. ELLISON—M. MENDÈS FRANCE [1] p. 165).

The number of zeros with

$$10 \leqq |\gamma - \gamma_1| \leqq 2\lambda \tag{13.4}$$

is owing to (13.3) and (7.3)

$$\leqq 2 \cdot 2\lambda \cdot c \cdot 6 \log \lambda \leqq \mu. \tag{13.5}$$

Thus for the contribution of such zeros to the infinite powersum we get by (7.5) the upper bound

$$\mu \exp\left\{k\left(\beta_1 + \frac{1}{\mu}\right)^2 - 100k + \mu\left(\beta_1 + \frac{1}{\mu}\right)\right\} \leqq \frac{e^{k\beta_1^2 + \mu\beta_1}}{e^{99M}}. \tag{13.6}$$

Using again (13.3) we have for the contribution of zeros with $|\gamma - \gamma_1| > 2\lambda$ the upper estimate

$$2 \sum_{n=[2\lambda]}^{\infty} c \log(\gamma_1 + n) \exp(k - kn^2 + \mu) = O(1). \tag{13.7}$$

So the number of remaining zeros with

$$|\gamma - \gamma_1| < 10 \tag{13.8}$$

is again by (13.3) and (6.5)

$$1 \leqq n < 20 \cdot c \log(\gamma_1 + 10) < 300 \log(2\lambda^5 + 10) < 900 \log L. \tag{13.9}$$

Now we can apply Theorem A of the Appendix for the numbers

$$\alpha_j = (\varrho_j - i\gamma_1)^2 + \lambda^2(\varrho_j - i\gamma_1) \tag{13.10}$$

with

$$|\gamma_j - \gamma_1| < 10 \tag{13.11}$$

and with the choice

$$a = M, \quad d = \frac{M}{L}. \tag{13.12}$$

So we get the existence of a k satisfying (6.6) for which

$$|W_1| = \left| \sum_{|\gamma_j - \gamma_1| < 10} \exp\{k[(\varrho_j - i\gamma_1)^2 + \lambda^2(\varrho_j - i\gamma_1)]\} \right| \geqq$$

$$\geqq \frac{e^{k\beta_1^2 + k\lambda^2\beta_1}}{(30L)^{900 \log L}} \geqq \frac{e^{k\beta_1^2 + \mu\beta_1}}{e^{1000 \log^2 L}} = \frac{e^{k\beta_1^2 + \mu\beta_1}}{e^{10M}}. \tag{13.13}$$

Thus (13.2), (13.6), (13.7) and (13.13) together imply

$$|U| \geqq \frac{e^{k\beta_1^2 + \mu\beta_1}}{2e^{10M}} \tag{13.14}$$

which contradicts to (12.9) and thus proves the lemma in Case A.

14. In Case B we transform the way of integration in K_2 on the broken line l defined for $t \geqq 0$ by

$$
\begin{aligned}
&I_1\colon\ \sigma = \frac{5}{4} && \text{for} \quad t \geqq \lambda,\\
&I_2\colon\ \alpha + \frac{1}{\mu} \leqq \sigma \leqq \frac{5}{4} && \text{for} \quad t = \lambda,\\
&I_3\colon\ \sigma = \alpha + \frac{1}{\mu} && \text{for} \quad 10 \leqq t \leqq \lambda,\\
&I_4\colon\ \frac{1}{4} \leqq \sigma \leqq \alpha + \frac{1}{\mu} && \text{for} \quad t = 10,\\
&I_5\colon\ \sigma = \frac{1}{4} && \text{for} \quad 0 \leqq t \leqq 10,
\end{aligned}
\tag{14.1}
$$

where

$$
\alpha = \frac{1}{2} + \frac{\log \lambda}{2\lambda} \tag{14.2}
$$

and for $t \leqq 0$ by reflection on the real axis, because there is no zero with (7.1) and so $\varphi(s)$ is regular right of l and on l.

Considering that the only change compared to the way in (12.1) is that β_1 is replaced by α, and so we get for K_2 analogously to (12.7) the estimate

$$
|K_2| = O\left(e^{-97M+\mu\alpha} + e^{\frac{\mu}{3}}\right) \leqq e^{k\beta_1^2+\mu\alpha}. \tag{14.3}
$$

Taking in account (7.6) and (14.2) we have

$$
\beta_1 - \alpha \geqq \frac{\log \gamma_1}{4\lambda} - \frac{\log \lambda}{2\lambda} \geqq \frac{\log \gamma_1}{4\lambda} - \frac{\frac{1}{5}\log \gamma_1}{2\lambda} > \frac{\log \gamma_1}{8\lambda}. \tag{14.4}
$$

Thus estimating here $|U|$ we get from (10.8), (14.3) and (14.4) the inequality

$$
|U| \leqq e^{2\lambda} \frac{e^{k\beta_1^2+\beta_1}}{e^{\mu \cdot \frac{\log \gamma_1}{8\lambda}}} \leqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{\frac{1}{9}M\lambda \log \gamma_1}}. \tag{14.5}
$$

To get a lower bound for $|U|$ we have to consider that in (13.1) the integral is again $O(1)$; the residue is owing to (7.6) absolutely

$$
\leqq \frac{1}{2}\left(2k\left|\frac{1}{2} - i\gamma_1\right| + \mu\right) e^{\frac{k}{4}+\frac{\mu}{2}-k\gamma_1^2} \leqq e^{2\lambda} \cdot e^{\frac{k}{4}+\frac{\mu}{2}-k\lambda^{10}} \leqq e^{\mu-\lambda^8\mu} \leqq 1. \tag{14.6}
$$

Further the infinite powersum belonging to zeros with $|\gamma - \gamma_1| \geqq 2\lambda$ is, as given by (13.7), $O(1)$.

Thus here again only the behaviour of the finite power sum belonging to the zeros with

$$
|\gamma - \gamma_1| < 2\lambda \tag{14.7}
$$

is interesting.

The number of terms is here by (13.3)

$$1 \leqq n < 4\lambda \cdot 15 \cdot \log(\gamma_1+2\lambda) < 61\lambda \log \gamma_1 \tag{14.8}$$

and thus proceeding as in (13.10)—(13.13) we get for our finite powersum by appropriate choice of k satisfying (6.6) the lower bound:

$$|W_2| = \Big|\sum_{|\gamma_j-\gamma_1|<2\lambda} \exp\{k[(\varrho_j-i\gamma_1)^2+\lambda^2(\varrho_j-i\gamma_1)]\}\Big| \geqq$$

$$\geqq \frac{e^{k\beta_1^2+k\lambda^2\beta_1}}{(30L)^{61\lambda\log\gamma_1}} \geqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{62\lambda\log\gamma_1\log L}} \geqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{\lambda\log\gamma_1\log^2 L}} = \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{\frac{1}{100}\lambda M\log\gamma_1}}. \tag{14.9}$$

This implies essentially the same estimate for $|U|$, namely taking in account the upper estimate of the integral, the residue and the contribution of the zeros with $|\gamma-\gamma_1|>2\lambda$ we have

$$|U| \geqq \frac{e^{k\beta_1^2+\mu\beta_1}}{2e^{\frac{1}{100}\lambda M\log\gamma_1}} \tag{14.10}$$

in contradiction to (14.5). Thus the proof of Case B is also finished; Lemma 1 is proved.

15. For the proof of Theorem 2 let first

$$\lambda_0 \overset{\text{def}}{=} \frac{1}{10}\frac{\sqrt{\log\sqrt{Y}}}{\log_2\sqrt{Y}}\left(>\frac{1}{20}\frac{\log Y}{\log_2 Y}\right). \tag{15.1}$$

In the course of proof we shall distinguish the following two cases.

Case I. There is a zero $\varrho^*=\beta^*+i\gamma^*$ with

$$\beta^* \geqq \frac{1}{2}+\frac{\log\gamma^*}{4\lambda_0}, \quad 0<\gamma^* \leqq e^{\lambda_0}. \tag{15.2}$$

This case is essentially settled already by Lemma 1. Namely, the zero ϱ^* for which (15.2) holds, satisfies the condition of Lemma 1 for any $Z \geqq \sqrt{Y}$ (if $Y>c_6$) and thus for any Z with

$$\sqrt{Y} \leqq Z \leqq Y \tag{15.3}$$

there is at least one sign change of $\Delta_1(x)$ in the interval

$$I^*(Z) = \left(\exp\left(\log Z-31\sqrt{\log Y}\log_2 Y\right), \exp\left(\log Z+31\sqrt{\log Y}\log_2 Y\right)\right) \tag{15.4}$$

because of (15.3) $I^*(Z)$ contains the interval $I(Z)$ given by (6.2).

Applying this for

$$Z_\nu = \sqrt{Y}\exp\left(62\nu\sqrt{\log Y}\log_2 Y\right), \quad 1 \leqq \nu \leqq \left[\frac{\sqrt{\log Y}}{124\log_2 Y}\right]-1 \tag{15.5}$$

we get at least

$$\left[\frac{\sqrt{\log Y}}{124\log_2 Y}\right]-1>\frac{1}{125}\cdot\frac{\sqrt{\log Y}}{\log_2 Y} \tag{15.6}$$

disjoint intervals contained in

$$[\sqrt{Y},\ Y] \tag{15.7}$$

such that every interval contains at least one sign change of $\Delta_1(x)$. Thus in Case I the total number of sign changes in $[2, Y]$ is

$$V_1(Y)>\frac{1}{125}\cdot\frac{\sqrt{\log Y}}{\log_2 Y} \tag{15.8}$$

which proves Theorem 2 for the Case I.

16. As the zeros of $\zeta(s)$ lie symmetrically to the line $\sigma=\frac{1}{2}$ we can formulate the other case as

Case II. All zeros $\varrho=\frac{1}{2}+\delta+i\gamma$ of $\zeta(s)$ with

$$|\gamma|\leqq e^{\lambda_0} \tag{16.1}$$

satisfy

$$|\delta|<\frac{\log|\gamma|}{4\lambda_0}. \tag{16.2}$$

In this case we shall prove that there are at least

$$c_7\lambda_0 \tag{16.3}$$

(c_7 positive effectively computable constant) sign changes of $\Delta_1(x)$ in $[2, e^{\frac{\lambda_0}{2}}]$ and thus we get the inequality

$$V_1(Y)\geqq V_1\left(e^{\frac{\lambda_0}{2}}\right)\geqq c_7\lambda_0>\frac{c_7}{20}\frac{\sqrt{\log Y}}{\log_2 Y} \tag{16.4}$$

which will prove Theorem 2 also in Case II.

In this case we shall use ideas of Littlewood, Ingham and Skewes, too.

17. Now we shall show that under the condition (16.1)—(16.2) the investigation of $\Delta_1(x)$ can be reduced to the investigation of the easier manageable $\Delta_4(x)$ (for the notation see (3.1)).

Introducing the notations

$$\Delta(x)=\int_1^x \Delta_4(\vartheta)\,d\vartheta, \tag{17.1}$$

$$\Delta_1^*(r)=\frac{\Delta_1(r)}{\left(\frac{\sqrt{r}}{\log r}\right)}, \tag{17.2}$$

$$\Delta_4^*(r)=\frac{\Delta_4(r)}{\sqrt{r}} \tag{17.3}$$

we shall use two well-known lemmata. (All the constants as well as those implied by the O and o symbols will be absolute, effective constants.)

LEMMA 2.

$$\Delta(u) = -\sum_{|\gamma|\leqq u^2} \frac{u^{\varrho+1}}{\varrho(\varrho+1)} + O(u). \tag{17.4}$$

The proof follows easily from Theorem 28 (p. 73) of INGHAM [2].

LEMMA 3. *For* $r \to \infty$ *we have*

$$\Delta_1^*(r) - \Delta_4^*(r) + 1 + o(1) = \frac{\Delta(r)}{r^{\frac{3}{2}} \log r} + \frac{\log r}{\sqrt{r}} \int_2^r \Delta(u) \frac{\log u + 2}{u^2 \log^3 u}\, du. \tag{17.5}$$

For the proof see, e.g., INGHAM [2], formula (33) in Theorem 35 (p. 104).

18. Combining this with Lemma 2 we get

LEMMA 4. *Under the condition* (16.1)—(16.2) *we have for*

$$u \leqq e^{\frac{\lambda_0}{2}} \tag{18.1}$$

the inequality

$$|\Delta(u)| \leqq c_8 u^{\frac{3}{2}}. \tag{18.2}$$

For the proof we consider (17.4). This gives using (18.1) and (16.1)—(16.2) the estimate

$$\begin{aligned} |\Delta(u) + O(u)| &\leqq \sum_{|\gamma|\leqq u^2} \frac{u^{\frac{3}{2}+\delta}}{\gamma^2} \leqq \sum_{|\gamma|\leqq u^2} \frac{u^{\frac{3}{2}} e^{\frac{\lambda_0}{2}\delta}}{\gamma^2} \leqq \\ &\leqq u^{\frac{3}{2}} \sum_{|\gamma|\leqq u^2} \frac{e^{\frac{\lambda_0}{2} \cdot \frac{\log|\gamma|}{4\lambda_0}}}{\gamma^2} \leqq u^{\frac{3}{2}} \sum_{\gamma} \frac{|\gamma|^{\frac{1}{8}}}{\gamma^2} = O\left(u^{\frac{3}{2}}\right) \end{aligned} \tag{18.3}$$

which proves the lemma.

Now using Lemmata 3 and 4 we get

LEMMA 5. *Under the condition* (16.1)—(16.2) *we have for*

$$r \leqq e^{\frac{\lambda_0}{2}} \tag{18.4}$$

the relation

$$\Delta_1^*(r) = \Delta_4^*(r) - 1 + o(1). \tag{18.5}$$

(By the $o(1)$ symbol we mean that the corresponding quantity is absolutely less than ε if $r > r_0(\varepsilon)$ and r satisfies (18.4).)

Owing to Lemma 3 it is enough to prove that the right side of (17.5) is $o(1)$. This is trivially true for the first term by (18.2) but again using (18.2) we have also for the integral on the right side of (17.5) the upper bound

$$\int_2^r c_8 u^{\frac{3}{2}} \frac{\log u + 2}{u^2 \log^3 u}\, du \leqq c_9 \frac{\sqrt{r}}{\log^2 r} = o\left(\frac{\sqrt{r}}{\log r}\right) \tag{18.6}$$

and thus the lemma is proved.

19. Due to Lemma 5 to guarantee a sign change for $\Delta_1(r)$ in an interval

$$J \subset [c_{10}, e^{\frac{\lambda_0}{2}}] \tag{19.1}$$

it is sufficient to show that

$$\max_{r \in J} \Delta_4^*(r) > \frac{3}{2} \tag{19.2}$$

and

$$\min_{r \in J} \Delta_4^*(r) < -\frac{3}{2}. \tag{19.3}$$

But the shortened form of the Riemann—van Mangoldt exact prime number formula gives for $r \leqq e^{\frac{\lambda_0}{2}}$

$$\Delta_4^*(r) = -\sum_{|\gamma| \leqq e^{\lambda_0}} \frac{r^{\delta+i\gamma}}{\varrho} + o(1) \tag{19.4}$$

(see, e.g., INGHAM [2] Theorem 29, (p. 77)) and thus $\Delta_4^*(r)$ can be treated easier than $\Delta_1^*(r)$.

Thus introducing the notation

$$G(v) \overset{\text{def}}{=\!=} \sum_{|\gamma| \leqq e^{\lambda_0}} \frac{e^{(\delta+i\gamma)v}}{\varrho} \tag{19.5}$$

$\Delta_1(r)$ has certainly a sign change in an interval

$$[e^{a_1}, e^{a_2}] \subset [c_{11}, e^{\frac{\lambda_0}{2}}] \tag{19.6}$$

if we can show that

$$\max_{a_1 \leqq v \leqq a_2} G(v) > 2 \tag{19.7}$$

and

$$\min_{a_1 \leqq v \leqq a_2} G(v) < -2. \tag{19.8}$$

(We remark that since the zeros of $\zeta(s)$ lie symmetrically to the real axis, $G(v)$ is real.)

20. Now we shall use an idea of INGHAM [3] which makes use of the Fejér-kernel

$$\int_{-\infty}^{\infty} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 e^{iuy}\, dy = \begin{cases} 2\pi(1-|u|) & \text{for } |u| \leqq 1, \\ 0 & \text{for } |u| \geqq 1 \end{cases} \tag{20.1}$$

and which makes possible to reduce the number of terms in $G(v)$ and so the effective application of Theorem B of the Appendix.

Let $A > 20$ and $B > 8$ be sufficiently large effective constants, B an integer, to be determined later, further ω any real number, satisfying

$$\log c_{11} + 1 \leqq \omega \leqq \frac{\lambda_0}{2} - 1.$$

We note that the further constants $c_v's$ with $12 \leqq v \leqq 22$ will be absolute effective positive constants whose values do not depend on A, B either.

Using the notation (19.5) we define the integral

$$I_1(\omega) \stackrel{\text{def}}{=} \int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 G\left(\omega + \frac{y}{A}\right) dy =$$

(20.2)

$$= \sum_{|\gamma| \leqq e^{\lambda_0}} \frac{e^{(\delta+i\gamma)\omega}}{\varrho} \int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 e^{iy\frac{\gamma}{A}} \cdot e^{\frac{\delta y}{A}} dy.$$

Further we define the integrals

$$I_2(\omega) \stackrel{\text{def}}{=} \sum_{|\gamma| \leqq e^{\lambda_0}} \frac{e^{(\delta+i\gamma)\omega}}{\varrho} \int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 e^{iy\frac{\gamma}{A}} dy \tag{20.3}$$

and

$$I_3(\omega) \stackrel{\text{def}}{=} \sum_{|\gamma| \leqq e^{\lambda_0}} \frac{e^{(\delta+i\gamma)\omega}}{\varrho} \int_{-\infty}^{\infty} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 e^{iy\frac{\gamma}{A}} dy. \tag{20.4}$$

We shall prove that

$$|I_1(\omega) - I_3(\omega)| \leqq c_{12} \tag{20.5}$$

and so with the use of Fejér's kernel we can show that in the investigation of the average of $G(v)$ in the interval $\left[\omega - \frac{1}{4}, \omega + \frac{1}{4}\right]$ in (20.2) only the contribution of the low zeros, i.e., those with $|\gamma| < A$ is essential.

To prove (20.5) we get with easy computation by partial integration

$$\left| \int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 \left(e^{\frac{\delta y}{A}} - 1\right) e^{iy\frac{\gamma}{A}} dy \right| \leqq$$

$$\left| \left[\frac{A}{i\gamma} e^{iy\frac{\gamma}{A}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 \left(e^{\frac{\delta y}{A}} - 1\right) \right]_{-\frac{A}{4}}^{\frac{A}{4}} \right| + \left| \int_{-\frac{A}{4}}^{\frac{A}{4}} \frac{A}{i\gamma} e^{iy\frac{\gamma}{A}} \cdot \frac{d}{dy} \left\{ \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 \left(e^{\frac{\delta y}{A}} - 1\right) \right\} dy \right| \leqq \tag{20.6}$$

$$2 \cdot \frac{A}{|\gamma|} \cdot \frac{1}{\left(\frac{A}{8}\right)^2} \left(e^{\frac{|\delta|}{4}} - 1\right) + \frac{A}{|\gamma|} \left(\frac{|\delta|}{A} e^{\frac{|\delta|}{4}} \cdot 2\pi + c_{13}\left(e^{\frac{|\delta|}{4}} - 1\right) \right) \leqq c_{14} \frac{A|\delta|}{|\gamma|}.$$

Thus we get

$$|I_1(\omega)-I_2(\omega)| \leqq \sum_{|\gamma|\leqq e^{\lambda_0}} \left|\frac{e^{(\delta+i\gamma)\omega}}{\varrho}\right| \left| \int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin\frac{y}{2}}{\frac{y}{2}}\right)^2 e^{iy\frac{\gamma}{A}} \left(e^{\frac{\delta y}{A}}-1\right) dy \right| \leqq$$

(20.7)

$$\leqq \sum_{|\gamma|\leqq e^{\lambda_0}} \frac{e^{\delta\omega}}{|\gamma|} \cdot c_{14} \frac{A|\delta|}{|\gamma|} \leqq c_{14} A \sum_{|\gamma|\leqq e^{\lambda_0}} \frac{e^{\frac{\log|\gamma|}{4\lambda_0}\cdot\frac{\lambda_0}{2}}}{|\gamma|} \cdot \frac{\log|\gamma|}{4\lambda_0} \cdot \frac{1}{|\gamma|} \leqq$$

$$\leqq \frac{c_{14}A}{4\lambda_0} \sum_{\gamma} \frac{\log|\gamma| \cdot |\gamma|^{\frac{1}{8}}}{|\gamma|^2} = \frac{c_{15}A}{4\lambda_0} \leqq c_{15}.$$

Further again by partial integration we have

$$\left| \int_{\frac{A}{4}}^{\infty} \left(\frac{\sin\frac{y}{2}}{\frac{y}{2}}\right)^2 e^{iy\frac{\gamma}{A}} dy \right| \leqq$$

(20.8)

$$\left| \left[\frac{A}{i\gamma} e^{iy\frac{\gamma}{A}} \left(\frac{\sin\frac{y}{2}}{\frac{y}{2}}\right)^2 \right]_{\frac{A}{4}}^{\infty} \right| + \left| \frac{A}{i\gamma} \int_{\frac{A}{4}}^{\infty} e^{iy\frac{\gamma}{A}} \frac{d}{dy} \left\{ \left(\frac{\sin\frac{y}{2}}{\frac{y}{2}}\right)^2 \right\} dy \right| \leqq$$

$$\frac{A}{|\gamma|} \cdot \frac{1}{\left(\frac{A}{8}\right)^2} + \frac{A}{|\gamma|} \cdot \frac{c_{16}}{A} < \frac{c_{17}}{|\gamma|},$$

and the same holds for the corresponding integral in $\left[-\infty, \frac{A}{4}\right]$. From this we have analogously to (20.7)

$$|I_3(\omega)-I_2(\omega)| \leqq \sum_{|\gamma|\leqq e^{\lambda_0}} \frac{|\gamma|^{\frac{1}{8}}}{|\gamma|} \cdot \frac{2c_{17}}{|\gamma|} < c_{18}. \tag{20.9}$$

So using (20.7) and (20.9) we get (20.5) to be valid with the choice $c_{12}=c_{15}+c_{18}$.

21. Thus we must investigate now the integral given by (20.4) which can be written according to (20.1) in the form

$$I_3(\omega) = \sum_{|\gamma|<A} \frac{e^{(\delta+i\gamma)\omega}}{\frac{1}{2}+\delta+i\gamma} \cdot 2\pi\left(1-\frac{|\gamma|}{A}\right). \tag{21.1}$$

Using the fact that the zeros of $\zeta(s)$ lie symmetrically to the real axis easy computation shows that

$$I_3(\omega) = 2\pi \sum_{0<\gamma<A} \frac{e^{\delta\omega}\{(1+2\delta)\cos(\gamma\omega)+2\gamma\sin(\gamma\omega)\}}{\left(\frac{1}{2}+\delta\right)^2+\gamma^2}\left(1-\frac{\gamma}{A}\right). \tag{21.2}$$

Now if we restrict ω beyond the previous $\log c_{11}+1 \leqq \omega \leqq \frac{\lambda_0}{2}-1$ by

$$\log c_{11}+1 \leqq \omega \leqq \frac{\lambda_0}{2\log A} \tag{21.3}$$

then we have for the zeros with $0<\gamma<A$

$$|\delta|\omega \leqq \frac{\log A}{4\lambda_0}\cdot\frac{\lambda_0}{2\log A} = \frac{1}{8} \tag{21.4}$$

and so

$$0.8 < e^{\delta\omega} < 1.2. \tag{21.5}$$

Let us introduce the notation:

$$J_\omega(\eta) = 2\pi \sum_{0<\gamma<A} \frac{e^{\delta\omega}\{(1+2\delta)\cos(\gamma\eta)+2\gamma\sin(\gamma\eta)\}}{\left(\frac{1}{2}+\delta\right)^2+\gamma^2}\left(1-\frac{\gamma}{A}\right). \tag{21.6}$$

Then obviously we have

$$I_3(\omega) = J_\omega(\omega). \tag{21.7}$$

First we note that choosing in $J_\omega(\eta)$

$$\eta = \frac{1}{A} \quad \text{and} \quad \eta = -\frac{1}{A}, \quad \text{resp.,} \tag{21.8}$$

$J_\omega(\eta)$ can be made "big positive" and "big negative", resp., for any ω in (21.3), if we choose A sufficiently large.

Namely, using (21.5) and the well-known fact

$$N(T) \overset{\text{def}}{=} \sum_{0<\gamma<T} 1 = \frac{T}{2\pi}\log\frac{T}{2\pi e}+O(\log T) > \frac{T}{7}\log T \quad \text{for} \quad T > T_1 \tag{21.9}$$

(see, e.g., INGHAM [2], Theorem 25) further the inequality

$$\sin t \geqq \frac{2}{\pi}\cdot t \quad \text{for} \quad 0 \leqq t \leqq \frac{\pi}{2} \tag{21.10}$$

we get by

$$\gamma > 14 \tag{21.11}$$

for any $A>2T_1$ the inequality

$$J_\omega\left(\frac{1}{A}\right) > 2\pi \sum_{0<\gamma<A} \frac{0.8\cdot 2\gamma\cdot\frac{2}{\pi}\cdot\frac{\gamma}{A}}{1.01\gamma^2}\left(1-\frac{\gamma}{A}\right) >$$

$$> \frac{6}{A}\sum_{0<\gamma<A}\left(1-\frac{\gamma}{A}\right) \geqq \frac{6}{A}N\left(\frac{A}{2}\right)\cdot\frac{1}{2} > \tag{21.12}$$

$$> \frac{1}{2}\cdot\frac{6}{A}\cdot\frac{A}{2\cdot 7}\log\frac{A}{2} > \frac{1}{5}\log\frac{A}{2} > \frac{1}{10}\log A.$$

And analogously we have for $A>2T_1$

$$J_\omega\left(-\frac{1}{A}\right) < 2\pi\left\{\sum_{0<\gamma<A}\frac{1.2\cdot 2}{\gamma^2} - \sum_{0<\gamma<A}\frac{0.8\cdot 2\gamma\cdot\frac{2}{\pi}\cdot\frac{\gamma}{A}}{1.01\gamma^2}\left(1-\frac{\gamma}{A}\right)\right\} < \tag{21.13}$$

$$< 4.8\pi\sum_{\gamma>0}\frac{1}{\gamma^2} - \frac{6}{A}\sum_{0<\gamma<A}\left(1-\frac{\gamma}{A}\right) < c_{19} - \frac{1}{10}\log A.$$

22. Now we shall apply Theorem B of the Appendix for the numbers

$$\frac{\gamma}{2\pi} \quad \text{with} \quad 0<\gamma<A \tag{22.1}$$

their total number being

$$N \overset{\text{def}}{=} N(A) < A\log A \quad \text{for} \quad A > c_{20} \tag{22.2}$$

and owing to $A>20$

$$N(A) \geqq 1. \tag{22.3}$$

We choose in Theorem B of the Appendix

$$q = B \overset{\text{def}}{=} [\log^2 A] \tag{22.4}$$

and

$$M = \left[\frac{\frac{\lambda_0}{2\log A}-1}{B^{N(A)}}\right] \geqq c(A)\lambda_0 \tag{22.5}$$

where $c(A)$ is an effectively computable constant depending only on A.

Thus denoting the distance of a real number x from the nearest integer by $\|x\|$, we get the existence of positive integer n_ν's with

$$1 \leqq n_1 < n_2 < \ldots < n_M \leqq \frac{\lambda_0}{2\log A} - 1 \tag{22.6}$$

for which all the relations

$$\left\|\frac{\gamma_j}{2\pi}n_\nu\right\| \leqq \frac{1}{B} \qquad (1\leqq j\leqq N,\ 1\leqq \nu\leqq M) \tag{22.7}$$

hold.

This implies

$$\left|\sin\left\{\gamma_j\left(n_\nu\pm\frac{1}{A}\right)\right\}-\sin\left(\pm\frac{\gamma_j}{A}\right)\right| \leqq 2\left|\sin\frac{\gamma_j n_\nu}{2}\right| < \frac{2\pi}{B} \tag{22.8}$$

and

$$\left|\cos\left\{\gamma_j\left(n_\nu\pm\frac{1}{A}\right)\right\}-\cos\left(\pm\frac{\gamma_j}{A}\right)\right| \leqq 2\left|\sin\frac{\gamma_j n_\nu}{2}\right| < \frac{2\pi}{B} \tag{22.9}$$

for $1\leqq j\leqq N$, $1\leqq\nu\leqq M$, where in the above formulae always both the upper or both the lower signs are meant.

Choosing the numbers ω_ν' and ω_ν'' as

$$\omega_\nu^{(i)} = n_\nu+\frac{(-1)^i}{A} \tag{22.10}$$

we get from (22.8)—(22.9) (the inequality (21.3) is satisfied owing to (22.6)) for $\nu>\log c_{11}+2$

$$\left|J_{\omega_\nu^{(i)}}(\omega_\nu^{(i)})-J_{\omega_\nu^{(i)}}\left(\frac{(-1)^i}{A}\right)\right| < 2\pi\sum_{0<\gamma<A}\frac{1.2(2+2\gamma)\cdot\frac{2\pi}{B}}{\gamma^2} < \tag{22.11}$$

$$< \frac{110}{B}\sum_{0<\gamma<A}\frac{1}{\gamma} < \frac{110}{B}\cdot c_{21}\log^2 A < c_{22}$$

considering (22.4) and the relation

$$\sum_{0<\gamma<A}\frac{1}{\gamma} < c_{21}\log^2 A \tag{22.12}$$

which is an easy consequence of (21.9).

Thus we have using (22.12)—(22.13)

$$J_{\omega_\nu'}(\omega_\nu') = I_3(\omega_\nu') < -\frac{1}{10}\log A+c_{19}+c_{22} \tag{22.13}$$

and

$$J_{\omega_\nu''}(\omega_\nu'') = I_3(\omega_\nu'') > \frac{1}{10}\log A-c_{22}. \tag{22.14}$$

Combining this with (20.5) we get already the needed results for the average of $G(v)$ in $\left[\omega-\frac{1}{4},\omega+\frac{1}{4}\right]$, namely we have

$$I_1(\omega_\nu') < -\frac{1}{10}\log A+c_{19}+c_{22}+c_{12} \tag{22.15}$$

and analogously

$$I_1(\omega_\nu'') > \frac{1}{10}\log A-c_{22}-c_{12}. \tag{22.16}$$

23. Now fixing A as

$$A = \max \{e^{10(c_{22}+c_{19}+c_{12}+4\pi)}, 2T_1, c_{20}, 20\} \tag{23.1}$$

from (22.15) and (22.16) we get considering the definition of $I_1(\omega)$ in (20.2):

$$\int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 G\left(\omega_\nu' + \frac{y}{A}\right) dy < -4\pi \tag{23.2}$$

and

$$\int_{-\frac{A}{4}}^{\frac{A}{4}} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 G\left(\omega_\nu'' + \frac{y}{A}\right) dy > 4\pi. \tag{23.3}$$

Since by (20.1) we have

$$\int_{-\infty}^{\infty} \left(\frac{\sin \frac{y}{2}}{\frac{y}{2}} \right)^2 dy = 2\pi \tag{23.4}$$

(23.2)—(23.4) give immediately

$$\min_{\omega_\nu' - \frac{1}{4} \leqq v \leqq \omega_\nu' + \frac{1}{4}} G(v) < -2 \tag{23.5}$$

and

$$\max_{\omega_\nu'' - \frac{1}{4} \leqq v \leqq \omega_\nu'' + \frac{1}{4}} G(v) > 2. \tag{23.6}$$

Taking in account $\frac{1}{A} < \frac{1}{20}$ and (22.10) we get

$$\left[\omega_\nu^{(i)} - \frac{1}{4}, \omega_\nu^{(i)} + \frac{1}{4}\right] \subset \left[n_\nu - \frac{1}{3}, n_\nu + \frac{1}{3}\right]. \tag{23.7}$$

Thus by (19.6)—(19.8) $\Delta_1(r)$ has at least one sign change in every interval

$$I_\nu \overset{\text{def}}{=} \left[e^{n_\nu - \frac{1}{3}}, e^{n_\nu + \frac{1}{3}}\right] \subset \left[c_{11}, e^{\frac{\lambda_0}{2}}\right] \tag{23.8}$$

(if $\nu > \log c_{11} + 2$).

As the n_ν-s are positive integers, these intervals are all disjoint, further their total number is by (22.5) at least

$$\begin{aligned} M - (\log c_{11} + 2) &\geqq c(A)\lambda_0 - \log c_{11} - 2 \geqq \\ &\geqq c_{23}\lambda_0 - \log c_{11} - 2 \geqq c_7 \lambda_0 \end{aligned} \tag{23.9}$$

(since here A is already fixed and so $c(A)=c_{23}$ is an effectively computable absolute positive constant).

Thus we have in Case B in the interval $\left[2, e^{\frac{\lambda_0}{2}}\right]$ at least $c_7\lambda_0$ sign changes of $\Delta_1(x)$, and thus owing to (16.4) Case B is settled, too, so Theorem 2 is completely proved.

Now we describe what sort of changes are necessary in the course of proof of Theorem 1 to get an effective value for Y_2. Analogously, even simpler one can effectivize the proof in the case $i=4$.

In the formulation of Lemma 1 we assert (6.2) without any condition. In (6.10) in the definition of $f(x)$ we work with $x^{1/4}$ instead of $x^{1/2}$ and analogously in (6.11) we define $H(s)$ in the last term with $\dfrac{1}{4\left(s-\frac{1}{4}\right)^2}$ instead of $\dfrac{1}{2\left(s-\frac{1}{2}\right)^2}$. We do not distinguish Cases A and B, we follow the line of Case A. As to $\varrho_0=\beta_0+i\gamma_0$ we choose the zero with the minimal imaginary part, i.e., $\varrho_0=\frac{1}{2}+i\gamma_0$ ($\gamma_0\approx 14.3$). Then (7.1) is satisfied trivially, and we get after at most $\left[\frac{\lambda}{2}\right]$ steps also the zero ϱ_1 as in (7.3) but we have now only $\beta_1\geqq\frac{1}{2}$ instead of $\beta_1\geqq\frac{1}{2}+\frac{1}{\lambda}$.

The next change is only in (11.1)—(11.3) where we get for K_1 in (11.3)

$$K_1 = e^{\frac{k}{16}+\frac{\mu}{4}}+O(1). \tag{24.1}$$

The estimation of K_2 being unchanged valid we get instead of (12.8) for K

$$|K| = O\left(e^{-97M+\mu\beta_1}+e^{\frac{k}{16}+\frac{\mu}{4}}\right) = O\left(e^{-97M+\mu\beta_1}\right) \tag{24.2}$$

by (6.3)—(6.7).

From this we get immediately

$$|U| \leqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{96M}} \tag{24.3}$$

again by (6.3)—(6.7), (10.8) and (7.3); i.e., the upper estimate (12.9) for U is unchanged valid.

Further we get for the residue R in (13.1) even a better upper bound than in (13.2), i.e., the final estimate in (13.2) remains valid

$$|R| \leqq \frac{1}{4}\left(2k\left|\frac{1}{4}-i\gamma_1\right|+\mu\right)e^{\frac{k}{16}+\frac{\mu}{4}-k\gamma_1^2} \leqq \frac{e^{k\beta_1^2+\mu\beta_1}}{e^{99M}} \tag{24.4}$$

and so the final lower estimate for $|U|$ in (13.14) is again valid, and the contradiction proves our modified Lemma 1.

Appendix

The following theorem is a special case of the so called second main theorem of the powersum theory.

THEOREM (T. Sós—Turán). *For arbitrary complex numbers z_j, and for a natural number n*

$$\max_{m<\nu\leqq m+n} \frac{\left|\sum_{j=1}^{n} z_j^{\nu}\right|}{|z_1|^{\nu}} \geqq \left(\frac{1}{8e\left(\frac{m}{n}+1\right)}\right)^{n}.$$

For the proof see VERA T. Sós—P. TURÁN [17].

If we choose here $m=a\frac{n}{d}$, $z_j=e^{\alpha_j\frac{a}{m}}=e^{\alpha_j\frac{d}{n}}$ we get from this

$$\max_{\frac{n}{a}d<\nu\leqq(a+d)\frac{n}{d}} \frac{\left|\sum_{j=1}^{n} e^{\alpha_j\frac{d}{n}\nu}\right|}{\left|e^{\alpha_1\frac{d}{n}\nu}\right|} \geqq \left(\frac{1}{8e\left(\frac{a}{d}+1\right)}\right)^{n}.$$

The above inequality implies immediately the continuous form of the second main theorem:

THEOREM A (T. Sós—Turán). *For arbitrary complex numbers α_j, and for positive real numbers a and d*

$$\max_{a<t\leqq a+d} \frac{\left|\sum_{j=1}^{n} e^{\alpha_j t}\right|}{|e^{\alpha_1 t}|} \geqq \left(\frac{1}{8e\left(\frac{a}{d}+1\right)}\right)^{n}.$$

The following theorem is an extension of Dirichlet's classical theorem on simultaneous approximation.

THEOREM B. *If $q\geqq 2$ and M are integers, $\gamma_1, \ldots, \gamma_N$ arbitrary real numbers, then there exist integer n_j's with*

$$1 \leqq n_1 < n_2 < \ldots < n_M \leqq Mq^N$$

such that for $1\leqq\mu\leqq M$, $1\leqq\nu\leqq N$

$$\|n_\mu\gamma_\nu\| \leqq \frac{1}{q}$$

where $\|x\|$ denotes the distance of x from the nearest integer.

For the proof see TITCHMARSH [18], p. 153.

REFERENCES

[1] ELLISON, W. J. and MENDÈS FRANCE, M.: *Les nombres premiers*, Paris, Hermann, 1975.
[2] INGHAM, A. E.: *The distribution of prime numbers*, Cambridge University Press, 1932.
[3] INGHAM, A. E.: A note on the distribution of primes, *Acta Arith.* **1** (2) (1936), 201—211.
[4] KNAPOWSKI, S.: On the sign changes in the remainder term in the prime number formula, *Journ. Lond. Math. Soc.* **36** (1961), 451—460.
[5] KNAPOWSKI, S.: On the sign changes of the difference $(\pi(x)-\text{li } x)$, *Acta Arith.* **7** (2) (1962), 107—120.
[6] KNAPOWSKI, S. and TURÁN, P.: On the sign changes of $(\pi(x)-\text{li } x)$, I. *Topics in Number Theory, Coll. Math. Soc. János Bolyai* **13**., North-Holland., Amsterdam—Oxford—New York, 1976, 153—169.
[7] KNAPOWSKI, S. and TURÁN, P.: On the sign changes of $(\pi(x)-\text{li } x)$, II., *Monatshefte für Math.* **82** (1976), 163—175.
[8] SHERMAN LEHMAN, R.: On the difference $\pi(x)-\text{li } x$, *Acta Arith.*, **11** (1966), 397—410.
[9] LEHMER, D. N.: *List of primes from* 1 *to* 10 006 721, Carnegie Inst. Wash. Publ. No. **165**. Wash. D. C. 1914.
[10] LITTLEWOOD, J. E.: Sur la distribution des nombres premiers, *C. R. Acad. Sci. Paris.* **158** (1914), 1869—1872.
[11] PINTZ, J.: Bemerkungen zur Arbeit von S. Knapowski und P. Turán, *Monatshefte für Math.* **82** (1976), 199—206.
[12] PÓLYA, G.: Über das Vorzeichen des Restgliedes im Primzahlsatz, *Gött. Nachr.* 1930, 19—27.
[13] PRACHAR, K.: *Primzahlverteilung*, Berlin—Göttingen—Heidelberg, 1957.
[14] RIEMANN, B.: Über die Anzahl der Primzahlen unter einer gegebenen Grösse, *Monatsh. Preuss. Akad. Wiss.*, Berlin, 1859, 671—680.
[15] SCHMIDT, E.: Über die Anzahl der Primzahlen unter gegebener Grenze, *Math. Ann.* **57** (1903), 195—204.
[16] SKEWES, S.: On the difference $\pi(x)-\text{li } x$, II. *Proc. London Math. Soc.* **5** (1955), 48—70.
[17] VERA T. SÓS and TURÁN, P.: On some new theorems in the theory of diophantine approximation, *Acta Math. Hung.* **6** (1955), 241—255.
[18] TITCHMARSH, E. C.: *The theory of Riemann zeta-function*, Clarendon Press, Oxford, 1951.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received January 10, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 371—380.

ON THE ACCELERATED STOCHASTIC APPROXIMATION

by

NGUYEN HUU TIEN

§ 1. Summary

An improvement of the Robbins—Monro procedure is given. The essential feature of this modified method is that it employs both the Kesten's idea and Venter's idea in order to improve the rate of convergence of the Robbins—Monro procedure.

§ 2. Introduction

Let $M(x)$ $(-\infty<x<\infty)$ be an unknown function with $M(\theta)=0$ and $M(x)\neq 0$, if $x\neq\theta$. Suppose that for any x, we can measure the value of $M(x)$ only with some random error Y_x, i.e., if we denote the result of measurement by Z_x, then $Z_x=M(x)+Y_x$. Our aim is to find the root θ.

For the solution of this problem Robbins and Monro ([1]) constructed the following sequence: let X_1 be an arbitrary real number and define the sequence $\{X_n\}$ by recursion:

$$X_{n+1}=X_n-\frac{1}{an}Z_n, \tag{2.1}$$

where $n=1, 2, 3, \ldots$; $a>0$ and $Z_n=Z_{x_n}$. It was shown that X_n converges to θ in probability. This process will be called Robbins—Monro process (RMP).

BLUM [2] under some simple conditions proved that

$$P\{\lim_{n\to\infty} X_n=\theta\}=1, \tag{2.2}$$

(see also DVORETZKY [3], SACKS [4] and VENTER [5]).

From practical point of view it is important to find that stochastic approximation method, which have a good enough rate of convergence.

In order to solve this problem VENTER [6] proposed a modification of the RMP. Venter's idea is the following: for each step the slope of the regression function at the root is estimated by the results of measurements and this information will be used to improve the rate of convergence.

Another modification of the RMP was proposed by KESTEN [7]. The basic idea of Kesten is the following: if the results Z_n of our measurements are positive (or negative) in a long run, then we can conjecture that our points X_n are far away from the root θ. Therefore we can modify the values of X_n in a stronger way to get the better rate of convergence (see also RÉVÉSZ [8]).

The purpose of this paper is to give some results about the modified RMP, which will be based on the ideas of Kesten and Venter. Then we shall see that our

method can join the advantages of the Venter-method and those of the Kesten-method. Through this paper our process will be called Kesten—Venter process (KVP).

In Section 3 some conditions and preliminaries are given for later reference. Section 4 contains the results on the KVP.

§ 3. Preliminaries and conditions

In this section we shall formulate some lemmas and conditions, which will be used to study the KVP.

LEMMA 3.1. *Let* $\{V_n\}$ *be a sequence of random variables and* $\{\mathscr{B}_n\}$ *a sequence of* σ*-fields such that* $\{V_1, \ldots, V_{n-1}\}$ *is measurable with respect to* $\mathscr{B}_n$ *for* $n>1$.

(i) *If* $\sum_{n=1}^{\infty} EV_n^2 < \infty$ *and* $\sum_{n=1}^{\infty} E[V_n|\mathscr{B}_n]$ *converges a.s. then* $\sum_{n=1}^{\infty} V_n$ *converges a.s.*

(ii) *If* $\sum_{n=1}^{\infty} b_n^{-2} EV_n^2 < \infty$ *with* $b_n \uparrow \infty$, *then*

$$P\left\{\lim_{n\to\infty} b_n^{-1} \sum_{k=1}^{n} \{V_k - E[V_k|\mathscr{B}_k]\} = 0\right\} = 1.$$

LEMMA 3.2. *If* $\{\xi_n\}$ *is a real sequence for which*

$$\xi_{n+1} = (1-a_n)\xi_n + b_n,$$

where $a_n \geqq 0$, $a_n \to 0$, $\sum_{n=1}^{\infty} a_n = \infty$ *and* $\sum_{n=1}^{\infty} b_n$ *converges, then*

$$\xi_n \to 0 \quad as \quad n \to \infty.$$

LEMMA 3.3. *Let* u_{nk} $(n, k=1, 2, 3, \ldots)$ *be a double array such that*

(i) $$E\{u_{nk}|u_{n1}, \ldots, u_{n,k-1}\} = 0;$$

(ii) $$\lim_{n\to\infty} \sum_{k=1}^{k_n} E\left|E(u_{nk}^2|u_{n1}, \ldots, u_{n,k-1}) - E(u_{nk}^2)\right| = 0;$$

(iii) $$\lim_{n\to\infty} \sum_{k=1}^{k_n} E(u_{nk}^2) = s^2;$$

(iv) *For every* $\varepsilon > 0$

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} E(u_{nk}^2 I_{\{|u_{nk}|>\varepsilon\}}) = 0,$$

where I_A *is the indicator function of* A.

Then $S_n = \sum_{k=1}^{k_n} u_{nk}$ *is asymptotically normal with mean* 0 *and variance* s^2.

LEMMA 3.4. *Let* $\xi_1, \ldots, \xi_n, \ldots$ *be a sequence of random variables with the distribution functions (d.f.-s)* $F_1(x), \ldots, F_n(x), \ldots$. *Suppose that* $F_n(x)$ *tends to a d.f.* $F(x)$

as $n\to\infty$. *Let* $\eta_1, \ldots, \eta_n, \ldots$ *be another sequence of random variables and suppose that* η_n *converges in probability to zero. Put*

$$X_n = \xi_n + \eta_n.$$

Then the d.f. of X_n *tend to* $F(x)$ *as* $n\to\infty$.

REMARK 3.1. Lemma 3.1 is a simple extension of theorems D and E, p. 387 of LOÈVE [9]. For the proof of Lemma 3.2 see VENTER [6]. Lemma 3.3 is a special case of a result of SACKS [4]. Lemma 3.4 follows from a theorem of CRAMÉR (p. 254 [10]). Therefore we will not prove these lemmas here.

We will also need the following condition on $M(x)$ as referred to:

(M I) $(x-\theta)M(x)>0$ if $x\neq\theta$.

(M II) For every $\delta>0$ there exists $c_\delta>0$ such that

$$\inf_{\substack{x-\theta>\delta \\ 0<c<C_\delta}} \frac{M(x+c)+M(x-c)}{2} > 0 \quad \text{and} \quad \sup_{\substack{x-\theta<-\delta \\ 0<c<C_\delta}} \frac{M(x+c)+M(x-c)}{2} < 0.$$

(M III) For some positive constants c_1 and d_1 and for all x

$$|M(x)| \leqq c_1 + d_1|x-\theta|.$$

(M IV) For some $s\geqq 2$, $\varrho>0$ and for $|x-\theta|<\varrho$

$$M(x) = \alpha(x-\theta)+f(x)+\delta(x),$$

where α is a positive number,

$$f(x) = \sum_{i=2}^{s} \alpha_i (x-\theta)^i,$$

and

$$\delta(x) = o(|x-\theta|^s) \quad \text{as} \quad |x-\theta| \to 0.$$

Especially, if $s=\infty$ then ϱ is the radius of convergence of the power $f(x)$ while $\delta(x)\equiv 0$ for $|x-\theta|<\varrho$.

The next conditions are regarding to the random error Y_x as referred to:

(Y I) $$\lim_{x\to\theta} E|Y_x|^2 = E|Y_\theta|^2 = \sigma^2.$$

(Y II) $$\lim_{r\to\infty} \lim_{\varepsilon\to 0+} \sup_{0<|x-\theta|<\varepsilon} E(|Y_x|^2 I_{\{|Y_x|>r\}}) = 0.$$

§ 4. Modification of the "accelerated stochastic approximation method"

4.1. A modification of Kesten's general convergence theorem

In order to prove the convergence of the KVP we will need the following modification of Kesten's general convergence theorem. Let θ be a real number and $T_n = T_n(t_1, \ldots, t_{2n-1})$ be measurable transformations, $n=1, 2, 3, \ldots$. Let $\{a_n\}$ be a sequence of positive numbers and X_1, $\{Y_n\}$ and $\{U_n\}$ be random variables,

$n=1, 2, 3, \ldots$. We define the sequence $\{X_n\}$ as follows:

$$X_{n+1}(\omega) = T_n(X_1(\omega), \ldots, X_n(\omega), U_1(\omega), \ldots, U_{n-1}(\omega)) + b_n(\omega) Y_n(\omega), \tag{4.1}$$

where the sequence $\{b_n(\omega)\}$ is defined in the following way

$$b_n = \begin{cases} a_n & \text{if} \quad n = 1, 2, \\ a_{\varphi(n)} & \text{if} \quad n = 3, 4, 5, \ldots, \end{cases} \tag{4.2}$$

where

$$\varphi(n) = 2 + \sum_{i=3}^{n} l[(X_i - X_{i-1})(X_{i-1} - X_{i-2})] \tag{4.3}$$

and

$$l(x) = \begin{cases} 1 & \text{if} \quad x \leqq 0; \\ 0 & \text{if} \quad x > 0. \end{cases}$$

It means that we take another a_n when $(X_i - X_{i-1})$ differs in sign from $(X_{i-1} - X_{i-2})$.

Let $\alpha_n(x_1, \ldots, x_n)$, $\beta_n(x_1, \ldots, x_n)$, $\gamma_n(x_1, \ldots, x_n)$ be nonnegative functions $(n=1, 2, 3, \ldots)$ and put

$$\varepsilon_N = \sup_{\{x_k\}} \sum_{n=N}^{\infty} \beta_n(x_1, \ldots, x_n), \tag{4.4}$$

$$\varrho(\delta) = \inf_{\substack{n=1,2,3,\ldots \\ x_1,\ldots,x_{n-1} \text{ arbitrary}}} \inf_{|x_n - \theta| \geqq \delta} \frac{\gamma_n(x_1, \ldots, x_n)}{b_n}. \tag{4.5}$$

Now we can formulate our

THEOREM 4.1. *Let us suppose that the following conditions are satisfied:*

$$|T_n(x_1, \ldots, x_n, u_1, \ldots, u_{n-1}) - \theta| \leqq \begin{cases} (1+\beta_n(x_1, \ldots, x_n))\,|x_n - \theta| - \gamma_n(x_1, \ldots, x_n) \\ \qquad\qquad \text{if} \quad (T_n - \theta)(x_n - \theta) > 0; \\ \alpha_n(x_1, \ldots, x_n) \quad \text{if} \quad (T_n - \theta)(x_n - \theta) \leqq 0; \end{cases} \tag{4.6}$$

$$\lim_{\varphi(n) \to \infty} \alpha_n(x_1, \ldots, x_n) = 0 \quad \textit{uniformly, for all sequences} \quad x_1, \ldots, x_n, \ldots \quad \textit{with} \quad \varphi(n) \to \infty; \tag{4.7}$$

$$\lim_{n \to \infty} \frac{(x_n - \theta)\beta_n(x_1, \ldots, x_n)}{b_n} = 0 \quad \textit{uniformly, for all sequences} \quad x_1, \ldots, x_n, \ldots; \tag{4.8}$$

$$\lim_{N \to \infty} \varepsilon_N = 0; \tag{4.9}$$

$$\varrho(\delta) > 0 \quad \textit{for every} \quad \delta > 0; \tag{4.10}$$

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad a_{n+1} \leqq a_n; \tag{4.11}$$

(4.12) i) $E(Y_n | X_1, \ldots, X_n, U_1, \ldots, U_{n-1}) = 0$,

ii) $E(Y_n^2 | X_1, \ldots, X_n, U_1, \ldots, U_{n-1}) \leqq \sigma^2 < \infty$,

with probability one; and

$$(4.13)\quad \begin{cases} \liminf\limits_{n\to\infty} \liminf\limits_{r\to 0} \inf\limits_{\substack{0<|x_n-\theta|\leqq r \\ x_1,\ldots,x_n,u_1,\ldots,u_{n-1}\text{ arbitrary}}} P\Big\{T_n(X_1,\ldots,X_n,U_1,\ldots,U_{n-1})+b_nY_n \geqq \\ \qquad\qquad \geqq X_n \Big| \begin{matrix} X_1=x_1,\ldots,X_n=x_n \\ U_1=u_1,\ldots,U_{n-1}=u_{n-1}\end{matrix}\Big\} > 0, \\ \liminf\limits_{n\to\infty} \liminf\limits_{r\to 0} \inf\limits_{\substack{0<|x_n-\theta|\leqq r \\ x_1,\ldots,x_n,u_1,\ldots,u_{n-1}\text{ arbitrary}}} P\Big\{T_n(X_1,\ldots,X_n,U_1,\ldots,U_{n-1})+b_nY_n < \\ \qquad\qquad < X_n \Big| \begin{matrix} X_1=x_1,\ldots,X_n=x_n \\ U_1=u_1,\ldots,U_{n-1}=u_{n-1}\end{matrix}\Big\} > 0. \end{cases}$$

Then

$$P\{\lim_{n\to\infty} X_n = \theta\} = 1.$$

Since the proof of this theorem is based on the idea of KESTEN [7], we will not give the details here, see [11].

4.2. Kesten—Venter Process (KVP)

Let X_1 be an arbitrary random variable, and $\{c_n\}$ be a sequence of positive numbers such that

$$(4.14)\qquad c_n = cn^{-\gamma}(1+o(1)),$$

where

$$c>0;\quad 0<\gamma<\frac{1}{2};\quad n=1,2,3,\ldots.$$

Suppose that there exist known constants a and b such that

$$(4.15)\qquad 0<a<\alpha<b<\infty,\quad \text{where}\quad \alpha=M'(\theta).$$

Define the sequence $\{X_n\}$ by the following recursion:

$$(4.16)\qquad X_{n+1}=X_n-\frac{1}{\mu_n A_n^*}\frac{Z_n'+Z_n''}{2},$$

where $Z_n'=Z_{X_n+c_n}$, $Z_n''=Z_{X_n-c_n}$ are random variables with conditional distributions

$$P\{Z_n'<z|X_1,Z_k',Z_k'',\ k=1,\ldots,n-1\}=P\{Z_n'<z|X_n+c_n\}=H(z|X_n+c_n),$$

and

$$P\{Z_n''<z|X_1,Z_k',Z_k'',\ k=1,\ldots,n-1\}=P\{Z_n''<z|X_n-c_n\}=H(z|X_n-c_n);$$

further, A_n^* is an estimate of α defined in the following form:

$$(4.17)\qquad A_n^*=\begin{cases} C \text{ constant if } n=1;\ \text{where } a\leqq C\leqq b; \\ A_{n-1} \text{ if } n=2,3,4,\ldots; \end{cases}$$

where

$$A_n = \begin{cases} a & \text{if} \quad B_n < a; \\ B_n & \text{if} \quad a \leqq B_n \leqq b; \\ b & \text{if} \quad B_n > b \end{cases} \tag{4.18}$$

and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} \frac{Z'_k - Z''_k}{2c_k}; \tag{4.19}$$

finally

$$\mu_n = \begin{cases} n & \text{if} \quad n = 1, 2; \\ \mu_{n-1} & \text{if} \quad (Z'_{n-1}+Z''_{n-1})(Z'_{n-2}+Z''_{n-2}) > 0 \quad \text{and} \quad n = 3, 4, 5, \ldots; \\ \mu_{n-1}+1 & \text{if} \quad (Z'_{n-1}+Z''_{n-1})(Z'_{n-2}+Z''_{n-2}) \leqq 0 \quad \text{and} \quad n = 3, 4, 5, \ldots. \end{cases} \tag{4.20}$$

The random variables Z'_n and Z''_n may be considered as the results of the measurements at (X_n+c_n) and (X_n-c_n), respectively, i.e.,

$$\begin{cases} Z'_n = M'_n+Y'_n = M(X_n+c_n)+Y_{X_n+c_n} = Z_{X_n+c_n}, \\ Z''_n = M''_n+Y''_n = M(X_n-c_n)+Y_{X_n-c_n} = Z_{X_n-c_n}. \end{cases} \tag{4.21}$$

For this reason (4.16) can be written as follows:

$$X_{n+1} = X_n - \frac{1}{\mu_n A_n^*} \left[\frac{M(X_n+c_n)+M(X_n-c_n)}{2} - \frac{Y_{X_n+c_n}+Y_{X_n-c_n}}{2} \right]. \tag{4.22}$$

This $\{X_n\}$ process will be called Kesten—Venter Process (KVP).

REMARK 4.1. Let us consider the KVP $\{X_n\}$. Let $\{U_n\}$ be a sequence of random variables, which are defined as follows:

$$U_n = \frac{Z'_n - Z''_n}{2c_n}, \quad n = 1, 2, 3, \ldots.$$

Let $\mathscr{B}_n$ be the σ-field in the underlying probability space induced by $\{X_1, Z'_k, Z''_k, k=1, \ldots, n-1\}$, i.e.,

$$\mathscr{B}_n = \sigma(X_1, Z'_k, Z''_k, \ k = 1, \ldots, n-1).$$

Then the random variables $X_1, \ldots, X_n, U_1, \ldots, U_{n-1}$ are measurables with respect to $\mathscr{B}_n$, i.e.,

$$\sigma(X_1, \ldots, X_n, U_1, \ldots, U_{n-1}) \subset \sigma(X_1, Z'_k, Z''_k, \ k = 1, \ldots, n-1) = \mathscr{B}_n.$$

It is also easy to verify that

$$\sigma(X_1, Z'_k, Z''_k, \ k = 1, \ldots, n-1) \subset \sigma(X_1, \ldots, X_n, U_1, \ldots, U_{n-1}).$$

Hence

$$\mathscr{B}_n = \sigma(X_1, Z'_k, Z''_k, \ k = 1, \ldots, n-1) = \sigma(X_1, \ldots, X_n, U_1, \ldots, U_{n-1}).$$

After this we can prove our

THEOREM 4.2. *If the sequence* $\{X_n\}$ *is given by* (4.16), *the conditions (M I), (M II), (M III) hold and there exists a* $\sigma_0>0$ *such that for every* n

$$E\{Y^2_{X_n\pm c_n}|X_1, \ldots, X_n, U_1, \ldots, U_{n-1}\} \leqq \sigma_0 < \infty, \tag{4.23}$$

with probability one;

$$\begin{aligned} &\lim_{r\to 0} \inf_{\substack{0<|x-\theta|\leqq r \\ 0<c<C_2}} P\{M(x+c)+M(x-c)+Y_{x+c}+Y_{x-c}>0\}>0; \\ &\lim_{r\to 0} \inf_{\substack{0<|x-\theta|\leqq r \\ 0<c<C_2}} P\{M(x+c)+M(x-c)+Y_{x+c}+Y_{x-c}\leqq 0\}>0. \end{aligned} \tag{4.24}$$

Then

$$P\Big\{\lim_{n\to\infty} X_n = \theta\Big\} = 1.$$

PROOF. For sake of simplicity we assume that $\theta=0$. By the definitions of the KVP and the random sequence $\{U_n\}$ we have

$$A_n^*(U_1, \ldots, U_{n-1}) = \begin{cases} C \quad \text{constant if} \quad n=1 \quad \text{where} \quad a \leqq C \leqq b; \\ B_{n-1} = B_{n-1}(U_1, \ldots, U_{n-1}) = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n-1} U_k \\ \qquad \text{if} \quad n \geqq 2 \quad \text{and} \quad a \leqq B_{n-1} \leqq b; \\ a \quad \text{if} \quad n \geqq 2 \quad \text{and} \quad B_{n-1} < a; \\ b \quad \text{if} \quad n \geqq 2 \quad \text{and} \quad B_{n-1} > b. \end{cases}$$

Hence the equation (4.22) can be written in the following form:

$$X_{n+1}(\omega) = T_n\big(X_1(\omega), \ldots, X_n(\omega), U_1(\omega), \ldots, U_{n-1}(\omega)\big) + b_n(\omega) Y_n^*(\omega), \tag{4.25}$$

where

$$\begin{cases} T_n = T_n(X_1, \ldots, X_n, U_1, \ldots, U_{n-1}) = X_n - \dfrac{1}{\mu_n} \dfrac{M(X_n+c_n)+M(X_n-c_n)}{2A_n^*(U_1, \ldots, U_{n-1})}; \\ b_n = \dfrac{1}{\mu_n}; \\ Y_n^* = -\dfrac{1}{A_n^*(U_1, \ldots, U_{n-1})} \dfrac{Y_{X_n+c_n}+Y_{X_n-c_n}}{2}. \end{cases} \tag{4.26}$$

Let $\alpha_n=\alpha_n(x_1, \ldots, x_n)$, $\beta_n=\beta_n(x_1, \ldots, x_n)$, $\gamma_n=\gamma_n(x_1, \ldots, x_n)$ be nonnegative functions defined as

$$\begin{cases} \alpha_n(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\mu_n} \dfrac{c_1+d_1|x_1|}{a} & \text{if} \quad |x_n| > c_n, \\ -\dfrac{1}{\mu_n} \dfrac{c_1+d_1 c_n}{a} & \text{if} \quad |x_n| < c_n; \end{cases} \\ \beta_n(x_1, \ldots, x_n) \equiv 0; \\ \gamma_n(x_1, \ldots, x_n) = \dfrac{1}{\mu_n} \dfrac{|M(x_n+c_n)+M(x_n-c_n)|}{2b}. \end{cases} \tag{4.27}$$

Now we want to show that these nonnegative functions under the conditions of Theorem 4.2 satisfy the conditions of Theorem 4.1. Therefore the assertion of Theorem 4.2 follows from Theorem 4.1.

Since this proof is very simple, using the definition of KVP, Remark 4.1 and the conditions of Theorem 4.2, we do not give the details here.

In order to investigate the rate of convergence of KVP we will also need the following

LEMMA 4.1. *Suppose that the conditions of Theorem 4.2 and the following condition will hold:*

(4.28) *$M(x)$ is continuous at $x=0$;*

(4.29) *The distribution functions $H(t|x_n+c_n)$ and $H(t|x_n-c_n)$ of random errors $Y_{X_n+c_n}$ and $Y_{X_n-c_n}$ are continuous in t, symmetrical with respect to $t=0$ and equicontinuous in certain neighborhood of $t=0$.*

Then

$$P\left\{\lim_{n\to\infty}\frac{\mu_n-\frac{n}{2}}{n^\beta}=0\right\}=1,$$

where $\frac{1}{2}<\beta<1$.

REMARK 4.2. Lemma 4.1 is very simple. One can see it by an elementary argument, using Theorem 4.2, Lemma 3.1, conditions (4.28) and (4.29) and the law of large numbers. So the details will be omitted (see [11]).

In the following we shall give some further theorems characterizing the properties of KVP. The proofs of these theorems are similar to those of Venter's results (see VENTER [6]), using Lemmas 3.1—3.4, 4.1 and Theorem 4.2. Therefore we can omit them.

THEOREM 4.3. *If the conditions of Lemma 4.1, condition (M IV) and*

$$s\alpha > \gamma b \tag{4.30}$$

hold, then $A_n^*\to\alpha$ with probability one as $n\to\infty$. Condition (4.30) *can be omitted if $s=\infty$ in condition (M IV).*

THEOREM 4.4. *Suppose that the conditions of Theorem 4.3 hold. Then*

a) *for any*

$$0\leqq\lambda<\min\left(\frac{1}{2},\,2\gamma\right) \tag{4.31}$$

we have

$$X_n-\theta=o(n^{-\lambda}) \tag{4.32}$$

with probability one as $n\to\infty$;

b) *for any*

$$0<\mu<\frac{1}{2}-\gamma \tag{4.33}$$

we have

$$A_n^* - \alpha = o(n^{-\gamma}) + o(n^{-\mu}) \tag{4.34}$$

with probability one as $n \to \infty$.

THEOREM 4.5. *Suppose that the conditions of Theorem 4.3 and conditions (Y I), (Y II) hold. Then for any*

$$\frac{1}{4} < \gamma < \frac{1}{2} \tag{4.35}$$

we have

$$n^{\frac{1}{2}}(X_n - \theta) \to \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha^2}\right) \tag{4.36}$$

and

$$n^{\frac{1}{2}} c_n (A_n^* - \alpha) \to \mathcal{N}\left(0, \frac{\sigma^2}{2(1+2\gamma)}\right) \quad as \quad n \to \infty. \tag{4.37}$$

If instead of condition (4.35) *we have*

$$\gamma = \frac{1}{4}, \tag{4.38}$$

then

$$n^{\frac{1}{2}}(X_n - \theta) \to \mathcal{N}\left(-\frac{2\alpha_2 c^2}{\alpha}, \frac{\sigma^2}{2\alpha^2}\right) \tag{4.39}$$

and

$$n^{\frac{1}{4}}(A_n^* - \alpha) \to \mathcal{N}\left(0, \frac{\sigma^2}{3c^2}\right) \tag{4.40}$$

as $n \to \infty$.

*

Acknowledgements. I should like to say thank to my teacher, Dr. P. RÉVÉSZ for numerous valuable advices, which meant a great help in my work.

REFERENCES

[1] ROBBINS, H.—MONRO, S.: A stochastic approximation method, *Ann. Math. Statist.* **22** (1951), 400—407.

[2] BLUM, J. R.: Approximation methods which converge with probability one, *Ann. Math. Statist.* **25** (1954), 382—386.

[3] DVORETZKY, A.: On stochastic approximation, *Proc. Third Berkeley Symp. Math. Statist. and Prob.* Vol. I (1956), 39—55.

[4] SACKS, J.: Asymptotic distribution of stochastic approximation procedures, *Ann. Math. Statist.* **29** (1958), 373—405.

[5] VENTER, J. H.: On Dvoretzky stochastic approximation theorems, *Ann. Math. Statist.* **37** (1966), 1534—1544.

[6] VENTER, J. H.: An extension of the Robbins—Monro procedure, *Ann. Math. Statist.* **38** (1967), 181—190.

[7] KESTEN, H.: Accelerated Stochastic Approximation, *Ann. Math. Statist.* **29** (1958), 41—59.

[8] RÉVÉSZ, P.: On the rate of convergence of Kesten's "accelerated stochastic approximation", *Studia Sci. Math. Hungar.* **9** (1974), 435—460.
[9] LOÈVE, M.: *Probability theory,* D. Van Nostrand Co., 2-nd ed., 1960.
[10] CRAMÉR, E.: *Mathematical methods of statistics,* Princeton Univ. Press, Princeton, 1946.
[11] NGUYEN HUU TIEN: *Accelerated stochastic approximation methods,* Dissertation, Budapest, 1977 (in Hungarian).

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received February 3, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 381—390

EXAMPLES FOR NON-ORDERLY SPACES

by

J. DEÁK

Is every completely regular space orderly (in the sense of [3], 2.8)? — E. DEÁK raised this problem in 1964 ([1], (7.1); see also [2], (18.22) and [3], 7.8). We shall answer this question in the negative. To make the difference between the two counter-examples clear, we introduce the notion of the *orderly directional dimension*. The problem of the characterization of orderly spaces remains open — a solution in terms of classical separation axioms cannot be expected for our examples are hereditarily normal compact spaces and, on the other hand, the Sorgenfrey-plane is hereditarily orderly but not normal ([2], p. 58). The last paragraph contains an example concerning the directional dimension and orderliness of the Stone-Čech compactification of orderly spaces.

§ 0

The terminology and notations of [3] will be used, with the exceptions below:

1) $A \subset B$ does not mean that A is a proper subset of B;
2) the closure of a set A is denoted by $\bar{A}$;
3) the axiom R_1) of [3], (1.5) is replaced by

$$\emptyset \in \mathscr{G}(\mathscr{R}), \quad X \in \mathscr{F}(\mathscr{R})$$

(this modification of the axioms of direction has no influence on the definition of directional dimension, orderliness &c and all the theorems in [3] remain valid, cf. [4], (0.1) and (0.5));

4) the natural order of a direction $\mathscr{R}$ (see [3], 1.7) is denoted by $<^{\mathscr{R}}$;
5) the elements of an ordered pair in a direction are separated by a semicolon: $(G;F)$; otherwise, ordered pairs are put into pointed brackets: $\langle a, b\rangle$.

Further notations. W is the class of all ordinal numbers;

$$\mathbf{L} = \{\alpha+t, -(\alpha+t): \alpha \in W,\ t \in [0, 1)\}$$

with $0+0 = -(0+0)$. Consider $\mathbf{L}$ with the linear order $<^{\mathbf{L}}$ defined by

$$\alpha < \beta \Rightarrow \alpha+t_1 <^{\mathbf{L}} \beta+t_2,$$

$$t_1 < t_2 \Rightarrow \alpha+t_1 <^{\mathbf{L}} \alpha+t_2,$$

$$\alpha+t_1 <^{\mathbf{L}} \beta+t_2 \Leftrightarrow -(\beta+t_2) <^{\mathbf{L}} -(\alpha+t_1)$$

where $\alpha, \beta \in W$ and $t_1, t_2 \in [0, 1)$. For $\alpha \in W$, $\alpha+0$ and $-(\alpha+0)$ will be denoted by α and $-\alpha$, respectively. For $p, q \in \mathbf{L}$, $\mathbf{L}(p, q)$ is the open interval (p, q) in $\mathbf{L}$ with the

topology induced by the order $<^{\mathbf{L}}|(p, q)$ (which will also be denoted by $<^{\mathbf{L}}$). The symbols $\mathbf{L}[p, q]$, $\mathbf{L}[p, q)$ and $\mathbf{L}(p, q]$ will be similarly used for closed and half-closed intervals of $\mathbf{L}$. Observe that for two intervals $A \subset B$ of $\mathbf{L}$, A is a subspace of B. Now $\mathbf{L}^1 = \mathbf{L}(-\omega_1, \omega_1)$ is the so-called *long line*; $\mathbf{H}^i = \mathbf{L}[O, \omega_i)$ $(1 \leqq i < \omega)$ is the ith *long halfline*. (Usually the space $\mathbf{H}^1$ is called the long line.)

For a feebly orderly (i.e., completely regular) space X, the *orderly directional dimension* of X, denoted by O-Dim X, is 0 iff X is indiscrete; otherwise, O-Dim X is the minimum of the cardinalities of the compatible *orderly* directional structures of X. In this terminology, the orderliness of a feebly orderly space X means Dim X = O-Dim X.

Let $\mathscr{R}$ be an orderly direction of a space X. For a point $x \in X$, there is a unique element $(G_x(\mathscr{R}); F_x(\mathscr{R})) = (G_x; F_x) \in \mathscr{R}$ with $x \in F_x - G_x$. $F_x - G_x$ is the *$\mathscr{R}$-plane* containing x. The order $<^{\mathscr{R}}$ induces a partial order $<^{(\mathscr{R})}$ on X:

$$x <^{(\mathscr{R})} y \Leftrightarrow (G_x; F_x) <^{\mathscr{R}} (G_y; F_y) \Leftrightarrow x \in G_y \Leftrightarrow y \notin F_x \Leftrightarrow F_x \subset G_y. \tag{0.a}$$

§ 1. Some lemmas

LEMMA 1. *Let $\mathfrak{R}$ be a compatible orderly directional structure of a space X and $q \in X$. Then there is another compatible orderly directional structure $\mathfrak{R}^*$ of X with $|\mathfrak{R}^*| \leqq 2|\mathfrak{R}|$ such that $F_q(\mathscr{R}^*) = X$ for each $\mathscr{R}^* \in \mathfrak{R}^*$.*

PROOF. For each $\mathscr{R} \in \mathfrak{R}$, take

$$\mathscr{R}^1 = \{(G; F) \in \mathscr{R}: q \notin F\} \cup \{(G_q(\mathscr{R}); X)\}$$

and

$$\mathscr{R}^2 = \{(G; F): (X - F; X - G) \in \mathscr{R}, q \notin F\} \cup \{(X - F_q(\mathscr{R}); X)\}.$$

Then the open $\{\mathscr{R}^1, \mathscr{R}^2\}$-halfspaces are just the open $\mathscr{R}$-halfspaces and $F_q(\mathscr{R}^1) = F_q(\mathscr{R}^2) = X$. Now

$$\mathfrak{R}^* = \{\mathscr{R}^1, \mathscr{R}^2: \mathscr{R} \in \mathfrak{R}\}$$

has the required properties.

LEMMA 2. *Let $\mathscr{R}$ be an orderly (but not necessarily compatible) direction of the space $X = \mathbf{L}[-\gamma, \omega_1]$, γ a countable ordinal number. Suppose that*

$$\omega_1 \in \overline{G_{\omega_1}}, \quad F_{\omega_1} = X. \tag{1.a}$$

Then

$$G_{\omega_1} = X - \{\omega_1\} \tag{1.b}$$

and for $\alpha < \omega_1$, there are sets G^α with

$$G^\alpha \in \mathscr{G}(\mathscr{R}), \quad G^\alpha \neq G_{\omega_1}, \quad G_{\omega_1} = \bigcup_{\alpha < \omega_1} G^\alpha. \tag{1.c}$$

PROOF. First of all we construct a series $y(\beta)$ satisfying

$$y(\beta) \in G_{\omega_1} \quad (\beta < \omega_1), \tag{1.d}$$

$$\beta < \beta' < \omega_1 \Rightarrow y(\beta) <^{\mathbf{L}} y(\beta') \tag{1.e}$$

and

(1.f) $$\beta < \beta' < \omega_1 \Rightarrow y(\beta) <^{(\mathscr{R})} y(\beta').$$

To begin with, let $y(0)$ be a point from G_{ω_1} (which is, according to (1.a), non-empty). Suppose that $0<\alpha<\omega_1$ and $y(\beta)$ has been defined for each $\beta<\alpha$ satisfying (1.d), (1.e) and (1.f) with $\beta, \beta'<\alpha$. From (1.d) and (0.a) we have $\omega_1 \notin F_{y(\beta)}$, thus the countable system of the closed sets $F_{y(\beta)}$ $(\beta<\alpha)$ has a common upper bound $y(\alpha)\neq\omega_1$ in the order $<^{\mathbf{L}}$. According to (1.a), this $y(\alpha)$ can be chosen from G_{ω_1}, so (1.d) is fulfilled for $\beta=\alpha$ as well. Since $y(\beta)\in F_{y(\beta)}$ and $y(\alpha)$ is an upper bound of $F_{y(\beta)}$ $(\beta<\alpha)$, we have (1.e) for $\beta'=\alpha$, too. Furthermore, $y(\alpha)\notin F_{y(\beta)}$ $(\beta<\alpha)$ means (1.f) with $\beta'=\alpha$ (see (0.a)).

Let now a point $x\in X-\{\omega_1\}$ be fixed. We are going to prove that $x\in G_{\omega_1}$. Choose a countable ordinal number α such that

(1.g) $$x <^{\mathbf{L}} y(\beta) \quad (\alpha < \beta < \omega_1)$$

(there is such an α, since $\{y(\beta)\}$ is an increasing series of the type ω_1, see (1.e)). If $\alpha+1<\beta<\omega_1$, we have $y(\alpha+1)<^{(\mathscr{R})}y(\beta)$, so $y(\alpha+1)\in G_{y(\beta)}$ (see (0.a)). On the other hand, (1.g) gives $x<^{\mathbf{L}}y(\alpha+1)$, thus

$$G_{y(\beta)} \cap \mathbf{L}[x, \omega_1) \neq \emptyset \quad (\alpha+1 < \beta < \omega_1).$$

Take now

$$z(\beta) = \inf_{<^{\mathbf{L}}} G_{y(\beta)} \cap \mathbf{L}[x, \omega_1) \quad (\alpha+1 < \beta < \omega_1).$$

Suppose that $\alpha+1<\beta<\beta'<\omega_1$ and $z(\beta)\neq x\neq z(\beta')$. From (1.f) we have $y(\beta)<^{(\mathscr{R})}$ $<^{(\mathscr{R})}y(\beta')$, i.e., $F_{y(\beta)}\subset G_{y(\beta')}$, so $\overline{G_{y(\beta)}}\subset G_{y(\beta')}$ and $z(\beta')<^{\mathbf{L}}z(\beta)$ (since $\mathbf{L}[x, \omega_1)$ is connected). But in the space X there is no decreasing series of the type ω_1, thus $z(\beta)=x$ for some $\beta<\omega_1$. Now

$$x\in\overline{G_{y(\beta)}} \subset G_{y(\beta+1)} \subset G_{\omega_1}$$

and (1.b) has been proved.

To show (1.c), put $G^\alpha=G_{y(\alpha)}$ $(\alpha<\omega_1)$. Here $\overline{G^\alpha}\subset G^{\alpha+1}\subset X-\{\omega_1\}$, thus $G^\alpha\neq$ $\neq G_{\omega_1}$. The lemma has been completely proved.

The proof above could have been told for ω_i instead of ω_1, thus we have (with $\gamma=0$):

LEMMA 2°. *Let $\mathscr{R}$ be an orderly direction of the space $X=\mathbf{L}[0, \omega_i]$, $1\leqq i<\omega$. Suppose that $\omega_i\in\overline{G_{\omega_i}}$, $F_{\omega_i}=X$. Then $G_{\omega_i}=X-\{\omega_i\}$ and $G_{\omega_i}=\cup\{G^\alpha\colon\ \alpha<\omega_i\}$ where $G^\alpha\in\mathscr{G}(\mathscr{R})$, $G^\alpha\neq G_{\omega_i}$ $(\alpha<\omega_i)$.*

§ 2. A space with Dim $=2$ and O-Dim uncountable

EXAMPLE 1. *Let Y be the one-point compactification of the long line $\mathbf{L}^1$. Then*

(2.a) $$\operatorname{Dim} Y = 2$$

and

(2.b) $$\text{O-Dim}\, Y > \omega.$$

The point in $Y-\mathbf{L}^1$ will be denoted by ω_1. If γ is a countable ordinal number, then $\mathbf{L}[-\gamma, \omega_1]$ is a subspace of Y.

PROOF OF (2.a). The directions

$$\mathscr{R} = \{(\mathbf{L}(-y, y); \mathbf{L}[-y, y]): 0 \leqq^{\mathbf{L}} y <^{\mathbf{L}} \omega_1\} \cup \{(\mathbf{L}^1; Y)\}$$

and

$$\mathscr{R}' = \{(\emptyset; \emptyset), (\mathbf{L}(0, \omega_1); \mathbf{L}[0, \omega_1]), (Y; Y)\}$$

form a compatible directional structure of Y, thus $\operatorname{Dim} Y \leqq 2$. On the other hand, $\operatorname{Dim} Y > 1$ follows from [3], (1.13).

PROOF OF (2.b). Suppose that there were a countable compatible orderly directional structure $\mathfrak{R}$ of Y. According to Lemma 1, we may suppose

$$F_{\omega_1}(\mathscr{R}) = Y \quad (\mathscr{R} \in \mathfrak{R}).$$

Now $\mathfrak{R}$ can be divided into two disjoint subsystems

$$\mathfrak{R}_1 = \{\mathscr{R} \in \mathfrak{R}: \omega_1 \in \overline{G_{\omega_1}(\mathscr{R})}\}$$

and

$$\mathfrak{R}_2 = \{\mathscr{R} \in \mathfrak{R}: \omega_1 \notin \overline{G_{\omega_1}(\mathscr{R})}\}.$$

If $\mathscr{R} \in \mathfrak{R}_1$, then

$$\omega_1 \in \overline{G_{\omega_1}(\mathscr{R}) \cap \mathbf{L}(0, \omega_1)} \tag{2.c}$$

or

$$\omega_1 \in \overline{G_{\omega_1}(\mathscr{R}) \cap \mathbf{L}(-\omega_1, 0)},$$

say (2.c) holds. Then Lemma 2 can be applied to the direction

$$\mathscr{R} | \mathbf{L}[-\gamma, \omega_1],$$

γ an arbitrary countable ordinal number. (1.b) means now

$$G_{\omega_1}(\mathscr{R}) \cap \mathbf{L}[-\gamma, \omega_1] = \mathbf{L}[-\gamma, \omega_1),$$

thus (as γ was arbitrary): $G_{\omega_1}(\mathscr{R}) = \mathbf{L}^1$. On the other hand, the closed sets $\overline{G_{\omega_1}(\mathscr{R})}$ $(\mathscr{R} \in \mathfrak{R}_2)$ have common lower and upper bounds in $\mathbf{L}^1$. To sum it up:

$$\begin{cases} \mathfrak{R} = \mathfrak{R}_1 \cup \mathfrak{R}_2, \\ \mathfrak{R}_1 = \{\mathscr{R} \in \mathfrak{R}: G_{\omega_1}(\mathscr{R}) = \mathbf{L}^1\}, \\ \mathfrak{R}_2 = \{\mathscr{R} \in \mathfrak{R}: G_{\omega_1}(\mathscr{R}) \subset \mathbf{L}(-\delta, \delta)\}, \quad \delta < \omega_1. \end{cases} \tag{2.d}$$

Let now α be an ordinal number, $\delta < \alpha < \omega_1$. As Y is a T_0-space, there is a direction $\mathscr{R}^\alpha \in \mathfrak{R}$ such that the points α and $-\alpha$ are in different $\mathscr{R}^\alpha$-planes, i.e., $-\alpha <^\alpha \alpha$ or $\alpha <^\alpha -\alpha$ where $<^\alpha$ means $<^{(\mathscr{R}^\alpha)}$. Because of (2.d), $\mathscr{R}^\alpha \in \mathfrak{R}_1$. Now we define a regressive function $f(\alpha)$ for $\delta < \alpha < \omega_1$:

$$f(\alpha) = \begin{cases} \sup\{\beta < \alpha: \beta <^\alpha -\alpha\} & \text{if} \quad -\alpha <^\alpha \alpha, \\ \sup\{\beta < \alpha: -\beta <^\alpha \alpha\} & \text{if} \quad \alpha <^\alpha -\alpha. \end{cases} \tag{2.e}$$

Evidently, $f(\alpha)\leqq\alpha$. To prove $f(\alpha)\neq\alpha$, suppose $-\alpha<^{\alpha}\alpha$ (the other case can be similarly dealt with). Then the βs in (2.e) are separated from α by the $\mathscr{R}^{\alpha}$-plane containing $-\alpha$, thus the ordinal numbers in an $\mathbf{L}^1$-neighbourhood of α are not considered when taking the supremum in (2.e), so $f(\alpha)<\alpha$, indeed. Further,

$$\text{(2.f)}\qquad \begin{cases} f(\alpha)+1\notin G_{-\alpha}(\mathscr{R}^{\alpha}) & \text{if} \quad -\alpha<^{\alpha}\alpha, \\ -(f(\alpha)+1)\notin G_{\alpha}(\mathscr{R}^{\alpha}) & \text{if} \quad \alpha<^{\alpha}-\alpha. \end{cases}$$

(Indeed: say $-\alpha<^{\alpha}\alpha$ and suppose (2.f) does not hold; then $f(\alpha)+1<^{\alpha}-\alpha$, and since $f(\alpha)+1$ was not taken into account in the supremum in (2.e), we have $f(\alpha)+1=\alpha$, contradicting the condition $-\alpha<^{\alpha}\alpha$.)

As $f(\alpha)$ is a regressive function for $\delta<\alpha<\omega_1$, there is a countable ordinal number γ such that $f(\alpha)=\gamma$ $(\alpha\in A)$ for an uncountable set A of countable ordinal numbers. Since $\mathfrak{R}_1$ is countable, we may suppose that there is a direction $\mathscr{R}\in\mathfrak{R}_1$ with $\mathscr{R}^{\alpha}=\mathscr{R}$ $(\alpha\in A)$ and, say, $\alpha<^{\alpha}-\alpha$ $(\alpha\in A)$. Now (2.f) means

$$-(\gamma+1)\notin G_{\alpha}(\mathscr{R}) \quad (\alpha\in A),$$

thus

$$\text{(2.g)}\qquad -(\gamma+1)\notin \bigcup_{\alpha\in A} G_{\alpha}(\mathscr{R}) = G\in\mathscr{G}(\mathscr{R}).$$

Take a set F such that $(G;F)\in\mathscr{R}$; if there are two such Fs, choose the larger one. Because of (2.d), $(\mathbf{L}^1;Y)\in\mathscr{R}$. (2.g) gives now

$$\big(G_{\alpha}(\mathscr{R});\ F_{\alpha}(\mathscr{R})\big)\leqq^{\mathscr{R}}(G;F)<^{\mathscr{R}}(\mathbf{L}^1;Y) \quad (\alpha\in A),$$

so

$$\alpha\in F_{\alpha}(\mathscr{R})\subset F\subset\mathbf{L}^1 \quad (\alpha\in A).$$

Since A is uncountable and F is closed, this means $\omega_1\in\mathbf{L}^1$, which is a contradiction. Thus (2.b) has been proved, too.

§ 3. A space with Dim = 3 and O-Dim = 4

EXAMPLE 2. *Take two copies each of the long halflines* $\mathbf{H}^1, \mathbf{H}^2, \mathbf{H}^3$ *and* $\mathbf{H}^4$. *Let* Z *be the one-point compactification of the topological sum of these eight spaces. Then*

$$\text{(3.a)}\qquad \operatorname{Dim} Z = 3$$

and

$$\text{(3.b)}\qquad \text{O-Dim}\, Z = 4.$$

Notations. The two copies of $\mathbf{H}^i$ $(1\leqq i\leqq 4)$ will be denoted by H^i and H^{-i}, $H^i\cup H^{-i}=V^i$. The points of H^i and H^{-i} corresponding to the point $y\in\mathbf{H}^i$ are $\langle y,i\rangle$ and $\langle y,-i\rangle$, respectively. The point added at the compactification is q. All the symbols $\langle\omega_i,i\rangle$ and $\langle\omega_i,-i\rangle$ $(1\leqq i\leqq 4)$ will mean the point q. Then the mapping

$$\begin{cases} f_{\varepsilon}\colon \mathbf{L}[0,\omega_{|\varepsilon|}]\to\overline{H^{\varepsilon}} \\ f_{\varepsilon}(y)=\langle y,\varepsilon\rangle \end{cases} \qquad (1\leqq|\varepsilon|\leqq 4)$$

(the closure being understood in the space Z) is a homeomorphism. Further notations:

$$H^\varepsilon(a, b) = f_\varepsilon(\mathbf{L}(a, b)) \qquad (a, b \in \mathbf{L}[0, \omega_{|\varepsilon|}],\ 1 \leqq |\varepsilon| \leqq 4)$$

and

$$V^i(a, b) = H^i(a, b) \cup H^{-i}(a, b) \qquad (a, b \in \mathbf{L}[0, \omega_i],\ 1 \leqq i \leqq 4).$$

Similar notations will be used for the f_ε-images of half-closed and closed intervals.

PROOF OF (3.a). With

$$\mathscr{R} = \Big\{(\emptyset; \emptyset), \Big(\bigcup_{1 \leqq i \leqq 4} H^i; \bigcup_{1 \leqq i \leqq 4} H^i\Big), \overline{(Z; Z)}\Big\},$$

$$\mathscr{R}_{ij} = \big\{(\emptyset; \emptyset)\big\} \cup \big\{(V^i[0, y); V^i[0, y]) : y \in \mathbf{L}(0, \omega_i)\big\} \cup$$

$$\cup \big\{(V^i; Z - V^j)\big\} \cup \big\{(Z - V^j[0, y]; Z - V^j[0, y)) : y \in \mathbf{L}(0, \omega_j)\big\} \cup \{(Z; Z)\},$$

$\{\mathscr{R}, \mathscr{R}_{12}, \mathscr{R}_{34}\}$ is a compatible directional structure of Z, thus $\operatorname{Dim} Z \leqq 3$. It is left to the reader to prove that the equality holds here.*

PROOF OF (3.b). Suppose that $\text{O-Dim}\, Z \leqq 3$. Then by Lemma 1, there is a compatible orderly directional structure $\mathfrak{R}$ of Z with

$$|\mathfrak{R}| \leqq 6, \quad F_q(\mathscr{R}) = Z \qquad (\mathscr{R} \in \mathfrak{R}). \tag{3.c}$$

Take now an H^ε $(1 \leqq |\varepsilon| \leqq 4)$ and a direction $\mathscr{R} \in \mathfrak{R}$. According to Lemma 2° (applied to $\mathscr{R} \mid \overline{H^\varepsilon}$):

$$H^\varepsilon \subset G_q(\mathscr{R}) \quad \text{or} \quad q \notin \overline{H^\varepsilon \cap G_q(\mathscr{R})}.$$

Since an arbitrary neighbourhood of q in Z contains another neighbourhood of q homeomorphic to Z, we may suppose without loss of generality that

$$H^\varepsilon \subset G_q(\mathscr{R}) \quad \text{or} \quad H^\varepsilon \cap G_q(\mathscr{R}) = \emptyset \quad (\mathscr{R} \in \mathfrak{R},\ 1 \leqq |\varepsilon| \leqq 4), \tag{3.d}$$

i.e., each $G_q(\mathscr{R})$ is the union of some H^εs.

Now we are going to prove that

$$[|\varepsilon_1| \neq |\varepsilon_2|,\ \mathscr{R} \in \mathfrak{R},\ H^{\varepsilon_1} \subset G_q(\mathscr{R})] \Rightarrow H^{\varepsilon_2} \cap G_q(\mathscr{R}) = \emptyset. \tag{3.e}$$

Suppose that (3.e) is not true. Then, by (3.d), there are $\varepsilon_1, \varepsilon_2$ and $\mathscr{R} \in \mathfrak{R}$ such that $1 \leqq |\varepsilon_1| < |\varepsilon_2| \leqq 4$ and

$$H^{\varepsilon_1} \cup H^{\varepsilon_2} \subset G_q(\mathscr{R}). \tag{3.f}$$

Lemma 2° applied to the direction $\mathscr{R} \mid \overline{H^{\varepsilon_1}}$ gives that there are sets $G^\alpha \in \mathscr{G}(\mathscr{R})$ $(\alpha < \omega_{|\varepsilon_1|})$ such that

$$G^\alpha \cap H^{\varepsilon_1} \neq H^{\varepsilon_1} \qquad (\alpha < \omega_{|\varepsilon_1|}) \tag{3.g}$$

and

$$H^{\varepsilon_1} = \bigcup_{\alpha < \omega_{|\varepsilon_1|}} G^\alpha \cap H^{\varepsilon_1}. \tag{3.h}$$

* From the point of view of the present example for a non-orderly space with O-Dim finite, it is irrelevant whether $\operatorname{Dim} Z = 3$ or less.

Now

$$\Gamma = \bigcup_{\alpha<\omega_{|\varepsilon_1|}} G^\alpha \in \mathcal{G}(\mathcal{R}). \tag{3.i}$$

(3.h) implies $q\in\bar{\Gamma}$, thus $\Gamma=G_q(\mathcal{R})$. From (3.f) and (3.g) we have

$$\overline{G^\alpha \cap H^{\varepsilon_2}} \subset H^{\varepsilon_2} \quad (\alpha < \omega_{|\varepsilon_1|}). \tag{3.j}$$

Further, (3.f), (3.h) and $\Gamma=G_q(\mathcal{R})$ give

$$H^{\varepsilon_2} = \bigcup_{\alpha<\omega_{|\varepsilon_1|}} G^\alpha \cap H^{\varepsilon_2}. \tag{3.k}$$

$\overline{H^{\varepsilon_2}}$ is homeomorphic to $\mathbf{L}[0, \omega_{|\varepsilon_2|}]$, and (3.j) means that the open subsets of $\mathbf{L}[0, \omega_{|\varepsilon_2|}]$ corresponding to the sets $G^\alpha\cap H^{\varepsilon_2}$ $(\alpha<\omega_{|\varepsilon_1|})$ do not intersect a neighbourhood of $\omega_{|\varepsilon_2|}$. Now (3.k) contradicts the inequality $|\varepsilon_1|<|\varepsilon_2|$, thus we have proved (3.e).

According to (3.e), each $G_q(\mathcal{R})$ $(\mathcal{R}\in\mathfrak{R})$ intersects at most one of the sets V^i $(1\leqq i\leqq 4)$. As $|\mathfrak{R}|\leqq 6$, there are a set V^{i_0} $(1\leqq i_0\leqq 4,\ i_0$ fixed) and a direction $\mathcal{R}_0\in\mathfrak{R}$ such that

$$V^{i_0}\cap G_q(\mathcal{R}) = \emptyset \qquad (\mathcal{R}\in\mathfrak{R},\ \mathcal{R}\neq\mathcal{R}_0). \tag{3.l}$$

Since $F_q(\mathcal{R})=Z$ $(\mathcal{R}\in\mathfrak{R})$, (3.l) means that $\mathcal{R}_0|\overline{V^{i_0}}$ is a compatible orderly direction of the subspace $\overline{V^{i_0}}$. Thus $\overline{V^{i_0}}$ is sub-orderable (so — as a compact space — orderable) *with q the last element* — this, however, is impossible, therefore O-Dim $Z>3$.

On the other hand, $\{\mathcal{R}^i\colon 1\leqq i\leqq 4\}$ is a compatible orderly directional structure of Z, where

$$\mathcal{R}^i = \{(\emptyset;\emptyset)\}\cup\{(H^i[0,y);\ H^i[0,y])\colon y\in\mathbf{L}(0,\omega_i)\}\cup$$
$$\cup\{(H^i;\ Z-H^{-i})\}\cup\{(Z-H^{-i}[0,y];\ Z-H^{-i}[0,y))\colon y\in\mathbf{L}(0,\omega_i)\}\cup\{(Z;Z)\},$$

so O-Dim $Z\leqq 4$, and (3.b) has been completely proved.

§ 4. On the directional dimension and orderliness of the Stone—Čech compactification

S. PURISCH [5] has proved that the STONE—ČECH compactification of a T_π-space X is (sub-)orderable iff X is pseudocompact and sub-orderable. Thus, in terms of the theory of directional structures: if the space X is pseudocompact and Dim $X\leqq 1$, then

i) Dim $\beta X\leqq 1$ and
ii) βX is orderly

(because of [3], 1.13). It is natural to ask if similar theorems hold for higher dimensions. Suppose that the space X is pseudocompact, orderly and Dim $X\leqq n$ (n a natural number) — is it true that

i) Dim $\beta X\leqq n$ and/or
ii) βX is orderly?

The answer is negative, there is a pseudocompact orderly T_2-space X with Dim $X=2$, Dim $\beta X=3$, O-Dim $\beta X=4$.

EXAMPLE 3. *Let A_i be the one-point compactification of the topological sum of the long halflines $\mathbf{H}^{2i-1}$ and $\mathbf{H}^{2i}$, a_i the point added at the compactification* $(i=1, 2)$, $A=A_1\times A_2$, *X a subspace of A,*

$$X=[A-((\{a_1\}\times A_2)\cup(A_1\times\{a_2\}))]\cup(\{a_1\}\times\{a_2\}).$$

Then X is pseudocompact,

$$\text{(4.a)} \qquad \text{Dim } X=\text{O-Dim } X=2$$

(thus X is orderly) and

$$\text{(4.b)} \qquad \text{Dim } \beta X=3, \quad \text{O-Dim } \beta X=4$$

(thus βX is not orderly).

PROOF OF (4.a). The spaces A_1 and A_2 are orderable, so they admit compatible orderly directions $\mathscr{R}_1$ and $\mathscr{R}_2$, respectively. These two directions induce in a natural way an *orderly* directional structure of two directions on the product space A (cf. the proof of [1], (3.1)). O-Dim is monotone, so O-Dim $X\leqq 2$. On the other hand, X is evidently not sub-orderable, so $2\leqq$ Dim X. Thus (4.a) has been proved (remember that Dim $X\leqq$ O-Dim X).

PROOF OF THE PSEUDOCOMPACTNESS. $\mathbf{H}^i$ $(1\leqq i\leqq 4)$ is sequentially compact and so is the product of two long half-lines. The space X is the union of four such products and one point, so it is sequentially compact as well, therefore it is pseudocompact.

PROOF OF (4.b). Let B be the topological sum of $\mathbf{L}[0, \omega_1]$ and $\mathbf{L}[0, \omega_2]$. C is the topological sum of $\mathbf{L}[0, \omega_3]$ and $\mathbf{L}[0, \omega_4]$. The copies of $\mathbf{L}[0, \omega_i]$ $(i=1, 2, 3, 4)$ used in the construction of B and C will be denoted by B^1, B^2, C^3 and C^4, respectively. The point of B^i or C^i $(1\leqq i\leqq 4)$ corresponding to $y\in\mathbf{L}[0, \omega_i]$ will be denoted by y^i. The subset of B^i $(i=1, 2)$ or C^i $(i=3, 4)$ corresponding to the interval $\mathbf{L}[p, q)\subset$ $\subset\mathbf{L}[0, \omega_i]$ is $B^i[p, q)$ or $C^i[p, q)$, respectively. Set

$$Q^{ij}=B^i[0, \omega_i)\times C^j[0, \omega_j) \qquad (i=1, 2;\ j=3, 4).$$

Take now $D=B\times C$ and construct another space E by identifying the four points

$$\langle\omega_i^i, \omega_j^j\rangle \qquad (i=1, 2;\ j=3, 4)$$

of D. This point will be denoted by e. (See Fig. 1.) One can readily see that

$$X\cong\{e\}\cup\bigcup_{\substack{i=1,2\\ j=3,4}} Q^{ij}=X'.$$

Any real-valued function defined on a long halfline is constant on a tail of it, thus a real-valued function defined on X' can be extended over the compact space E, so $E\cong\beta X$.

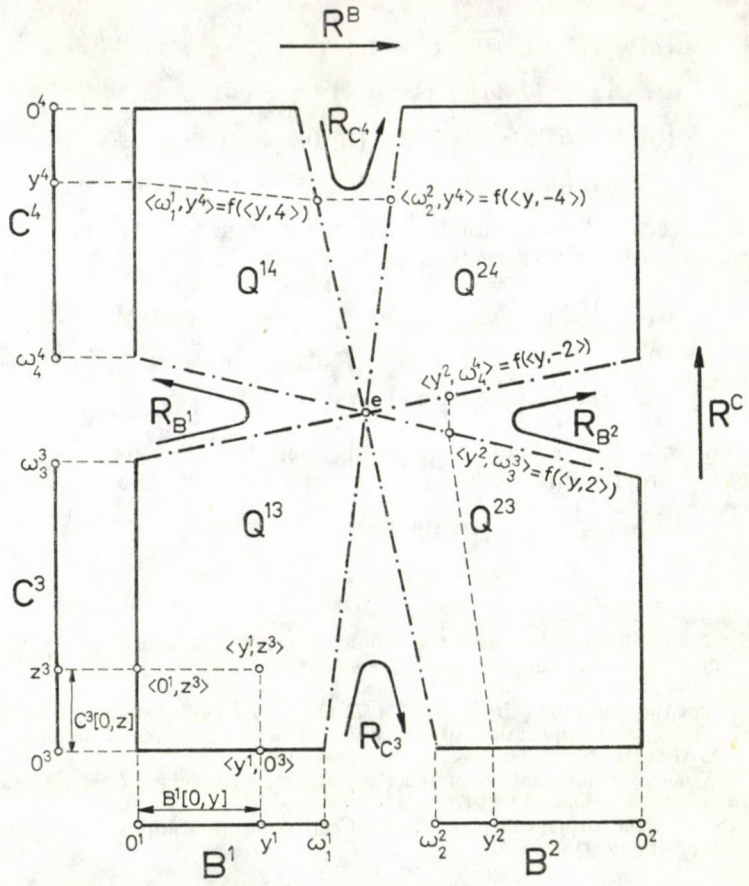


Fig. 1

Further, with Z the space from Example 2, the mapping

$$f(z)=\begin{cases} e & \text{if} \quad z=q, \\ \langle y^i, \omega_3^3\rangle & \text{if} \quad z=\langle y, i\rangle \in H^i, \\ \langle y^i, \omega_4^4\rangle & \text{if} \quad z=\langle y, -i\rangle \in H^{-i}, \end{cases} i=1,2,$$

$$\begin{cases} \langle \omega_1^1, y^i\rangle & \text{if} \quad z=\langle y, i\rangle \in H^i, \\ \langle \omega_2^2, y^i\rangle & \text{if} \quad z=\langle y, -i\rangle \in H^{-i}, \end{cases} i=3,4$$

is a homeomorphism from Z onto $Z'=\{e\}\cup(X-X')$ (Z' is the eight-pointed star in Fig. 1), so — according to (3.a) and (3.b) — Dim $E\geqq 3$ and O-Dim $E\geqq 4$ (as both dimensions are monotone).

To prove (4.b), we have to show that Dim $E\leqq 3$ and O-Dim $E\leqq 4$.

If

$$\mathscr{R} = \{(\emptyset; \emptyset), (\overline{Q^{13} \cup Q^{24}} - \{e\}; \overline{Q^{13} \cup Q^{24}}), (E; E)\},$$

$$\mathscr{R}^B = \{(B^1[0, y) \times C; B^1[0, y] \times C) : y \in \mathbf{L}(0, \omega_1)\} \cup$$

$$\cup \{(\emptyset; \emptyset), (B^1[0, \omega_1) \times C; E - (B^2[0, \omega_2) \times C)), (E; E)\} \cup$$

$$\cup \{(E - (B^2[0, y] \times C); E - (B^2[0, y) \times C)) : y \in \mathbf{L}(0, \omega_2)\}$$

and $\mathscr{R}^C$ is a direction defined analogously to $\mathscr{R}^B$ (see Fig. 1), then $\{\mathscr{R}, \mathscr{R}^B, \mathscr{R}^C\}$ is a compatible directional structure of E, so Dim $E \leqq 3$.

With

$$\mathscr{R}_{B^i} = \{(B^i[0, y) \times C^3; B^i[0, y] \times C^3) : y \in \mathbf{L}(0, \omega_i)\} \cup$$

$$\cup \{(\emptyset; \emptyset), (B^i[0, \omega_i) \times C^3; E - (B^i[0, \omega_i) \times C^4)), (E; E)\} \cup$$

$$\cup \{(E - (B^i[0, y] \times C^4); E - (B^i[0, y) \times C^4)) : y \in \mathbf{L}(0, \omega_i)\}$$

$(i = 1, 2)$ and $\mathscr{R}_{C^3}, \mathscr{R}_{C^4}$ defined analogously (see Fig. 1), $\{\mathscr{R}_{B^1}, \mathscr{R}_{B^2}, \mathscr{R}_{C^3}, \mathscr{R}_{C^4}\}$ is a compatible orderly directional structure of E, thus O-Dim $E \leqq 4$. This completes the proof of (4.b).

REFERENCES

[1] DEÁK, E.: Eine vollständige Charakterisierung der Teilräume eines euklidischen Raumes mittels der Richtungsdimension, *Publ. Math. Inst. Hung. Acad. Sci.* **9** Series A (1964), 437—464.

[2] DEÁK, E.: Dimension and convexity, IV., *MTA III. Oszt. Közl.* **18** (1968), 45—81 (in Hungarian).

[3] DEÁK, E.: Theory and application of directional structures, *Topics in topology*, edited by Á. Császár, North-Holland (1974), 187—211.

[4] DEÁK, J.: A new characterization of the class of subspaces of a Euclidean space, *Studia Sci. Math. Hung.* **11** (1976), 253—258.

[5] PURISCH, S.: On the orderability of Stone—Čech compactifications, *Proc. Amer. Math. Soc.* **41** (1973), 55—56.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13/15, 1053 Hungary

(Received March 21, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 391—398.

ÜBER HALBPRIMÄRE RINGE MIT KETTENBEDINGUNGEN FÜR IDEALE [1]

von

A. WIDIGER

1. Einleitung

Das Hauptziel dieser Arbeit ist es, einen Zerlegungssatz für halbprimäre Ringe zu beweisen, die der Minimal- bzw. der Maximalbedingung für zweiseitige Ideale genügen und deren Radikal den Nilpotenzgrad 2 hat. Genauer: Ein halbprimärer Ring A mit den genannten Eigenschaften ist die gruppendirekte Summe

$$A = B \oplus C$$

eines Ideales B von A und eines Unterringes C von A, wobei $B^2 = B$ ein Ring ist, für den $B/J(B)$ direkte Summe von Matrizenringen über unendlichen Schiefkörpern ist und C der strengen Minimal- bzw. Maximalbedingung genügt (siehe unten für die Definition). Dies verallgemeinert (allerdings nur für den Fall, daß $J(A)^2 = (0)$ ist) die Resultate von [5] und [6]. Ferner geben wir für einen halbprimären Ring mit nilpotentem Radikal mit Minimalbedingung für Ideale Spaltbarkeitskriterien an.

Das (Jacobson-) Radikal eines Ringes A bezeichnen wir mit $J(A)$. Ein *halbprimärer Ring* A sei ein Ring (im allgemeinen ohne Einselement), für den $A/J(A)$ artinsch ist (d. h. der Minimalbedingung für Rechtsideale genügt). $\oplus$ oder $\sum^{\oplus}$ mögen gruppentheoretische direkte Summen bezeichnen, $\boxplus$ oder $\sum^{\boxplus}$ ringtheoretische. Q sei der Körper der rationalen Zahlen, $Z(p^\infty)$ die Prüfersche p-Gruppe (p Primzahl).

Die additive Gruppe eines Ringes A werde mit $(A, +)$ bezeichnet. Genügt $(A, +)$ der Minimal- bzw. der Maximalbedingung für Untergruppen, so sagen wir, der Ring A genüge der *strengen Minimal-* bzw. *Maximalbedingung*.

2. Spaltbarkeitskriterien

Ein Ring wird *spaltbar* genannt, wenn sein maximales Torsionsideal ringdirekter Summand ist.

Sei A ein halbprimärer Ring mit nilpotentem Radikal

$$\bar{A} = A/J(A) = \bar{e}_1 \bar{A} \bar{e}_1 \oplus \ldots \oplus \bar{e}_n \bar{A} \bar{e}_n$$

eine Zerlegung von A als direkte Summe einfacher artinscher Ringe. Bekanntlich ([2], S.54) gibt es orthogonale Idempotente $\{e_1, \ldots, e_n\}$ von A mit $e_i \in \bar{e}_i$ $(i = 1, \ldots, n)$. Es seien die Idempotente nun so numeriert, daß $\bar{e}_1, \ldots, \bar{e}_t$ unendliche und $\bar{e}_{t+1}, \ldots, \bar{e}_n$ endliche Ordnung haben. Dann haben auch $e_{t+1}, \ldots, e_n$ endliche Ordnung,

[1] Diese Arbeit ist während eines Studienaufenthaltes in Budapest entstanden.

denn es gibt z. B. eine natürliche Zahl k mit $ke_{t+1}\in J(A)$, also gilt $k^m e_{t+1}=0$ wenn $J(A)^m=(0)$ ist.

Wir betrachten die zweiseitige Peircesche Zerlegung von A bezüglich der Idempotente $\{e_1, \ldots, e_n\}$:

$$A = \sum_{i,j=1}^{n}{}^{\oplus} e_i A e_j \oplus (1-e)Ae \oplus eA(1-e) \oplus (1-e)A(1-e).$$

Offenbar kann man schreiben

$$(1-e)Ae = \sum_{i=1}^{n}{}^{\oplus} (1-e)Ae_i, \quad eA(1-e) = \sum_{i=1}^{n}{}^{\oplus} e_i A(1-e).$$

Man setze

$$C = \sum_{i,j=t+1}^{n}{}^{\oplus} e_i A e_j \oplus \sum_{i=t+1}^{n}{}^{\oplus} (1-e)Ae_i \oplus \sum_{i=t+1}^{n}{}^{\oplus} e_i A(1-e) \oplus (1-e)A(1-e),$$

$$B = \sum_{\text{Rest}}{}^{\oplus} e_i A e_j \oplus \sum_{i=1}^{t}{}^{\oplus} (1-e)Ae_i \oplus \sum_{i=1}^{t}{}^{\oplus} e_i A(1-e),$$

wobei $\sum_{\text{Rest}}$ bedeuten möge, daß über alle Indexpaare i, j summiert wird, die in C nicht vorkommen.

Trivialerweise sind die Gruppen von $e_i Ae_j$, $(1-e)Ae_i$, $e_i A(1-e)$ für $i, j\leqq t+1$ beschränkte Torsionsgruppen.

LEMMA 1. *Für* $i>t$, $j\leqq t$ (*und* $j>t, i\leqq t$) *gilt* $e_i Ae_j=(0)$.

BEWEIS. Es ist $e_i Ae_j\subseteqq J(A)$. Angenommen, es gibt ein $l>1$ mit $e_i Ae_j\cap \cap J(A)^l=(0)$ und $e_i Ae_j\cap J(A)^{l-1}\neq(0)$.

$e_i Ae_j\cap J(A)^{l-1}$ ist ein $e_j Ae_j$-Rechtsmodul und kann wegen

$$(e_i Ae_j\cap J(A)^{l-1})e_j J(A)e_j \subseteqq e_i Ae_j\cap J(A)^l = (0)$$

als unitärer $e_j Ae_j/e_j J(A)e_j$-Modul aufgefaßt werden.

Als vollständig reduzibler Modul über $\bar{e}_j\bar{A}\bar{e}_j$ ist $e_i Ae_j\cap J(A)^{l-1}$ torsionsfrei wegen $j\leqq t$. Andererseits (wegen $i>t$) ist die additive Gruppe von $e_i Ae_j$ eine Torsionsgruppe. Unsere Annahme ist also falsch. Da $J(A)$ aber nilpotent ist, gilt $e_i Ae_j=(0)$.

LEMMA 2. $e_i Ae_j$, $(1-e)Ae_i$, $e_i A(1-e)$ *sind für* $i, j\leqq t$ *torsionsfrei und teilbar.*

BEWEIS. Der Beweis werde etwa für $e_i Ae_j$ geführt; in den anderen Fällen verläuft er analog. Wir schließen induktiv nach dem Nilpotenzgrad m des Radikales $J(A)$: Ist $J(A)=(0)$, so ist $e_i Ae_j=(0)$ für $i\neq j$, und für $i=j$ ist $e_i Ae_i$ ein voller Matrizenring über einem Schiefkörper der Charakteristik 0, also seine additive Gruppe torsionsfrei und teilbar.

Sei nun die Behauptung für Ringe mit einem Nilpotenzgrad $\leqq m$ des Radikales als richtig angenommen und sei $J(A)^{m+1}=(0)$, $J(A)^m\neq(0)$.

$e_i Ae_j\cap J(A)^m$ ist (wie im Beweis von Lemma 1) ein unitärer Rechts-$\bar{e}_j\bar{A}\bar{e}_j$-Modul, also vollständig reduzibel und wegen der Eigenschaften von $\bar{e}_j\bar{A}\bar{e}_j$ teilbar und torsionsfrei zu sein, auch teilbar und torsionsfrei.

$$e_i Ae_j/(e_i Ae_j\cap J(A)^m) \cong (e_i Ae_j + J(A)^m)/J(A)^m$$

ist torsionsfrei und teilbar nach Induktionsvoraussetzung, denn der Nilpotenzgrad des Radikales von $A/J(A)^m$ ist $\leqq m$. Folglich ist e_iAe_j torsionsfrei und teilbar, wie behauptet. Über die additive Struktur von $(1-e)A(1-e)$ läßt sich im allgemeinen natürlich nichts aussagen. Fehlt dieser Summand in der Zerlegung von A, so ist offenbar C das maximale Torsionsideal von A und offensichtlich gilt (wegen $(1-e)Ae_iA(1-e)\subseteq(1-e)A(1-e)=(0)$)

$$A = B \boxplus C.$$

SATZ 1. *Ein halbprimärer Ring mit nilpotentem Radikal und mit (Rechts-, Links-) Einselement ist spaltbar.*

BEWEIS. Hat A etwa ein Rechtseinselement e', so lassen sich die Idempotente $e_1, \ldots, e_n$ bekanntlich so wählen, daß $e_1+\ldots+e_n=e'$ gilt. Dann ist also $(1-e')A(1-e')=(0)$.

SATZ 2. *Ein halbprimärer Ring mit nilpotentem Radikal und Minimalbedingung für Ideale ist spaltbar genau dann, wenn* $(1-e)Ae_iA(1-e)=(0)$ *gilt für* $i=1, \ldots, t$.

Für den Beweis und auch noch später benötigen wir das

LEMMA 3. *Ist A ein halbprimärer Ring mit nilpotentem Radikal und Minimalbedingung (Maximalbedingung) für Ideale, so ist* $(1-e)A(1-e)$ *ein Ring mit strenger Minimalbedingung (Maximalbedingung).*

BEWEIS. Wir schließen wieder mit vollständiger Induktion nach dem Nilpotenzgrad m von $J(A)$.

Ist $J(A)=(0)$, so gilt $(1-e)A(1-e)\subseteq J(A)=(0)$.

Sei die Behauptung für Ringe mit einem Nilpotenzgrad $\leqq m$ des Radikales als richtig angenommen und sei $J(A)^{m+1}=(0)$, $J(A)^m\neq(0)$. Wir betrachten $X=(1-e)A(1-e)\cap J(A)^m$.

Sei $x\in X$, $a\in A$. Dann ist $a-ea\in J(A)$. Es sei

$$x = y-ey-ye+eye, \quad y\in A.$$

Wegen $xea=0$ gilt

$$x(a-ea) = xa\in J(A)^{m+1} = (0),$$

also $xa=0$. Genauso folgt $ax=0$.

Folglich ist jede Untergruppe von $(1-e)A(1-e)\cap J(A)^m$ ein Ideal von A. Wegen der Voraussetzung über A ist daher $(1-e)A(1-e)\cap J(A)^m$ ein Ring mit strenger Minimalbedingung (Maximalbedingung). Weiter gilt

$$(1-e)A(1-e)/\big((1-e)A(1-e)\cap J(A)^m\big) \cong (1-\tilde{e})\tilde{A}(1-\tilde{e})$$

($\tilde{x}$ das Bild von x beim natürlichen Homomorphismus von A auf $\tilde{A}=A/J(A)^m$). Nach Induktionsvoraussetzung genügt $(1-\tilde{e})\tilde{A}(1-\tilde{e})$ der strengen Minimalbedingung (Maximalbedingung). Hieraus folgt die Behauptung.

BEWEIS VON SATZ 2. Nach Lemma 3 hat $(1-e)A(1-e)$ Minimalbedingung für Untergruppen, ist also direkte Summe einer endlichen Gruppe und endlich vieler Exemplare von Prüferschen Gruppen. Ist $(1-e)Ae_iA(1-e)=(0)$ $(i=1, \ldots, t)$, so gilt offenbar wegen Lemma 1

$$A = B \boxplus C,$$

und C ist das maximale Torsionsideal von A.

Sei umgekehrt A spaltbar, $A=D\dotplus C$, D torsionsfrei. Da $(1-e)Ae_i$, $e_iA(1-e)$ teilbar sind für $i=1, \ldots, t$, ist $(1-e)Ae_iA(1-e)$ teilbar. Ferner ist $(1-e)Ae_iA(1-e)\subseteq \subseteq(1-e)A(1-e)$ eine Torsionsgruppe. Nach [3], Hilfssatz 10.3 liegt jedes $Z(p^\infty)$ im Annullator von A. $(C^2, +)$ ist daher eine beschränkte Torsionsgruppe, folglich $(1-e)Ae_iA(1-e)=(0)$, $i=1, \ldots, t$.

FOLGERUNG 1 *(vgl. hierzu [1]). Dann und nur dann ist der halbprimäre Ring mit nilpotentem Radikal und Minimalbedingung für Ideale spaltbar, wenn A^2 keine Untergruppe vom Typ $Z(p^\infty)$ enthält.*

Wir geben ein Beispiel für einen nicht spaltbaren halbprimären Ring mit Minimalbedingung für Ideale an: Es sei E die Gruppe aller rationalen Zahlen mit ungeradem Nenner und $\hat{Q}=(Q, +)/E\cong Z(2^\infty)$ und

$$A=\begin{bmatrix}0 & Q & \hat{Q}\\ 0 & Q & Q\\ 0 & 0 & 0\end{bmatrix}=\left\{\begin{bmatrix}0 & a & \hat{d}\\ 0 & b & c\\ 0 & 0 & 0\end{bmatrix}: a, b, c\in Q,\ \hat{d}\in\hat{Q}\right\}.$$

Die Addition sei die übliche Addition von Matrizen; die Multiplikation sei definiert durch

$$\begin{bmatrix}0 & a & \hat{d}\\ 0 & b & c\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & a_1 & \hat{d}_1\\ 0 & b_1 & c_1\\ 0 & 0 & 0\end{bmatrix}=\begin{bmatrix}0 & ab_1 & \widehat{ac_1}\\ 0 & bb_1 & bc_1\\ 0 & 0 & 0\end{bmatrix}.$$

Man überzeugt sich leicht, daß ein Ring vorliegt, der die verlangten Eigenschaften hat.

FOLGERUNG 2. *Ein halbprimärer Ring mit nilpotentem Radikal mit Minimal- und Maximalbedingung für Ideale ist spaltbar.*

BEWEIS. Wegen der Maximalbedingung enthält der Ring kein $Z(p^\infty)$.

FOLGERUNG 3. *Ein halbprimärer Ring A mit Minimalbedingung für Ideale, für den $J(A)^2=(0)$ gilt, ist spaltbar.*

BEWEIS. $(1-e)Ae_iA(1-e)\subseteq J(A)^2$.

FOLGERUNG 4 *(SZÁSZ [4]). Jeder artinsche Ring ist spaltbar.*

BEWEIS. Der Beweis von Lemma 3 zeigt, daß für einen artinschen Ring $e_iA(1-e)$ der strengen Minimalbedingung genügt. Wegen Lemma 2 ist daher $e_iA(1-e)=(0)$ für $i=1, \ldots, t$.

3. Zerlegungssatz

SATZ 3. *Seien A, B halbprimäre Ringe mit nilpotentem Radikal und der Eigenschaft, daß $A/J(A)$ und $B/J(B)$ endlich sind. Ist M ein $A-B$-Bimodul mit Minimalbedingung (Maximalbedingung) für Untermoduln, so hat $(M, +)$ Minimalbedingung (Maximalbedingung) für Untergruppen.*

BEWEIS. Wir beweisen den Satz mit vollständiger Induktion nach dem Minimum r der Nilpotenzgrade von $J(A)$ und $J(B)$. Es sei $r=1$. Ohne Beschränkung der

Allgemeinheit können wir annehmen, daß $J(B)=(0)$ ist. Wir schließen wiederum induktiv nach dem Nilpotenzgrad m von $J(A)$. Im Falle $m=1$ ist M also ein $A-B$-Bimodul über den halbeinfachen Ringen A und B. Als B-Rechtsmodul hat M etwa die Darstellung

$$M = M_0 \oplus M_1$$

mit einem trivialen B-Rechtsmodul M_0 und einem unitären B-Rechtsmodul M_1. Offenbar sind sowohl M_0 als auch M_1 Untermoduln des Bimoduls M. M_0 hat als A-Linksmodul die Zerlegung

$$M_0 = {}_0M_0 \oplus {}_1M_0$$

mit einem trivialen A-Linksmodul ${}_0M_0$ und einem unitären vollständig reduziblen A-Linksmodul ${}_1M_0$. Beide sind offenbar Untermoduln des Bimoduls M und genügen daher der Minimalbedingung (Maximalbedingung) für Untermoduln. Folglich hat $({}_0M_0, +)$ Minimalbedingung (Maximalbedingung) für Untergruppen. Aus dem gleichen Grunde ist ${}_1M_0$ die direkte Summe endlich vieler einfacher A-Linksmoduln. Wegen der Voraussetzung über A ist jeder einfache A-Linksmodul endlich. Daher ist ${}_1M_0$ endlich.

Entsprechend ist M_1 die direkte Summe

$$M_1 = {}_0M_1 \oplus {}_1M_1$$

eines trivialen A-Linksmoduls und eines unitären A-Linksmoduls. Sowohl ${}_0M_1$ als auch ${}_1M_1$ sind Untermoduln des Bimoduls M. Wie oben folgt, daß ${}_0M_1$ endlich ist. Um die Behauptung zu beweisen, genügt es zu zeigen, daß ${}_1M_1$ endlich ist. Wir können also annehmen, daß $M={}_1M_1$ ein unitärer $A-B$-Bimodul ist. Sei $x_1 \in M$. Wegen der Endlichkeit von A und B ist der von x_1 erzeugte Untermodul Ax_1B von M endlich. Wäre M unendlich, so gäbe es $x_2 \in M, x_2 \notin Ax_1B$. Dann ist Ax_2B endlich und folglich Ax_1B+Ax_2B endlich. Ferner ist $Ax_1B \subsetneqq Ax_1B+Ax_2B$. So kann man weiter schließen und erhält eine streng aufsteigende Kette von Untermoduln von M. Im Falle der Maximalbedingung ergibt sich also ein Widerspruch. Für den Fall, daß M die Minimalbedingung für Untermoduln erfüllt, beweisen wir zunächst eine

Zwischenbehauptung: Ist M unendlich, X ein endlicher Untermodul von M und $M=X \oplus Z$ eine Zerlegung von M als B-Rechtsmodul, so gibt es ein $0 \neq z \in Z$ mit $Az \subseteq Z$.

Denn sei $A=\{a_1, \ldots, a_l\}$. Für $z \in Z$ hat man

$$a_i z = x_i + y_i, \quad x_i \in X, \quad y_i \in Z, \quad i = 1, \ldots, l.$$

Da nun X endlich und Z unendlich ist, muß es $z, z' \in Z, z \neq z'$ geben mit $x_i = x_i'$ für jedes $i=1, \ldots, l$. Das bedeutet aber $a_i(z-z') \in Z, i=1, \ldots, l$, wie behauptet.

Sei jetzt etwa $M=Ax_1B \oplus Z_1$ eine Zerlegung von M als B-Rechtsmodul. Falls M unendlich ist, gibt es wegen der Zwischenbehauptung und wegen der Zornschen Lemmas einen Untermodul $U_1 \neq (0)$ von M, $U_1 \subseteq Z_1$, der maximal ist bezüglich der Eigenschaft $Ax_1B \cap U_1 = (0)$. Sei $0 \neq x_2 \in U_1$. Dann ist Ax_1B+Ax_2B endlich und man hat eine Zerlegung von M als B-Rechtsmodul:

$$M = Ax_1B \oplus Ax_2B \oplus Z_2,$$

$Z_2 \subseteq Z_1$. Wie oben folgt die Existenz eines Untermoduls U_2 von M, für den offenbar $U_1 \supsetneq U_2$ gilt. Da M unendlich angenommen war, kann man das Verfahren fortsetzen und erhält eine streng absteigende nicht abbrechende Kette von Untermoduln, im Widerspruch zur Voraussetzung.

Sei nun die Behauptung zunächst als richtig angenommen für Nilpotenzgrad von $J(A)<m$ (weiterhin $J(B)=(0)$), und sei $J(A)^m=(0), J(A)^{m-1}\neq(0)$. Dann ist $J(A)^{m-1}M$ ein Untermodul von M. Wegen $J(A)\big(J(A)^{m-1}M\big)=J(A)^m M=(0)$ kann $J(A)^{m-1}M$ als $A/J(A)-B$-Modul aufgefaßt werden, wobei die Untermoduln übereinstimmen. Nach dem ersten Teil des Beweises ist $J(A)^{m-1}M$ daher ein Modul mit strenger Minimalbedingung (Maximalbedingung). $M/J(A)^{m-1}M$ kann wegen $J(A)^{m-1}\big(M/J(A)^{m-1}M\big)$ als $A/J(A)^{m-1}-B$-Modul aufgefaßt werden, wobei die Untermoduln übereinstimmen. Nach unserer Induktionsannahme (Nilpotenzgrad des Radikales von $A/J(A)^{m-1}$ ist $<m$) genügt $M/J(A)^{m-1}M$ der strengen Minimalbedingung (Maximalbedingung). Wir haben also, daß sowohl $J(A)^{m-1}M$ als auch $M/J(A)^{m-1}M$ der strengen Minimalbedingung (Maximalbedingung) genügen, also auch M. Für $r=1$ ist damit Satz 3 bewiesen.

Angenommen nun, der Satz sei richtig, falls das Minimum der Nilpotenzgrade von $J(A)$ und $J(B)$ kleiner als r ist, und sei M ein $A-B$-Modul, wobei (ohne Beschränkung der Allgemeinheit) $J(A)^r=(0)$, $J(A)^{r-1}\neq(0)$, $J(B)^{r-1}\neq(0)$. Wir betrachten $J(A)^{r-1}M$. $J(A)^{r-1}M$ ist ein Untermodul von M und kann als $A/J(A)-B$-Modul aufgefaßt werden. Nach unserer Induktionsvoraussetzung ist $J(A)^{r-1}M$ ein Modul mit strenger Minimalbedingung (Maximalbedingung). $M/J(A)^{r-1}M$ kann wieder als $A/J(A)^{r-1}-B$-Modul aufgefaßt werden und ist folglich nach Induktionsvoraussetzung ein Modul mit strenger Minimalbedingung (Maximalbedingung). Dasselbe gilt also auch für M, und Satz 3 ist bewiesen.

FOLGERUNG 5. *Ist A ein halbprimärer Ring mit nilpotentem Radikal, so daß $A/J(A)$ endlich ist, und genügt A der Minimalbedingung (Maximalbedingung) für Ideale, so genügt A der strengen Minimalbedingung (Maximalbedingung).*

FOLGERUNG 6. *Besitzt ein Ring A wie in der letzten Folgerung ein (einseitiges) Einselement oder genügt A sowohl der Minimal- als auch der Maximalbedingung für Ideale, so ist A endlich.*

Jetzt geben wir den angekündigten Zerlegungssatz an.

SATZ 4. *Es sei A ein halbprimärer Ring, der der Minimalbedingung (Maximalbedingung) für Ideale genügt und für den $J(A)^2=(0)$ gilt. Dann ist*

$$A = B \oplus C,$$

und es gilt:

(1) *B ist ein Ideal und C ein Unterring von A.*

(2) *C ist ein Ring mit strenger Minimalbedingung (Maximalbedingung).*

(3) *$B/J(B)$ ist die direkte Summe von vollen Matrizenringen über unendlichen Schiefkörpern.*

(4) *$B^2=B$ und B enthält kein Ideal $\neq(0)$, das der strengen Minimalbedingung (Maximalbedingung) genügt.*

Ist A ein Ring mit Maximalbedingung für Ideale, so auch B.

BEWEIS. Wir setzen wieder $A=B\oplus C$ mit

$$C=\sum_{i,j=t+1}^{n}{}^{\oplus} e_iAe_j\oplus\sum_{i=t+1}^{n}{}^{\oplus}(1-e)Ae_i\oplus\sum_{i=t+1}^{n}{}^{\oplus}e_iA(1-e)\oplus(1-e)A(1-e),$$

$$(*)\qquad B=\sum_{\text{Rest}}{}^{\oplus}e_iAe_j\oplus\sum_{i=1}^{t}{}^{\oplus}(1-e)Ae_i\oplus\sum_{i=1}^{t}{}^{\oplus}e_iA(1-e)$$

wobei hier die e_i so numeriert sind, daß gerade $\bar{e}_1\bar{A}\bar{e}_1, \ldots, \bar{e}_t\bar{A}\bar{e}_t$ unendlich und $\bar{e}_{t+1}\bar{A}\bar{e}_{t+1}, \ldots, \bar{e}_n\bar{A}\bar{e}_n$ endlich sind.

Daß C ein Unterring von A und B ein Ideal von A ist, bestätigt man durch Nachrechnen, wenn man beachtet, daß $J(A)^2=(0)$ ist. Nach Lemma 3 genügt $(1-e)A(1-e)$ der strengen Minimalbedingung (Maximalbedingung). e_iAe_j $(i, j\geqq t+1)$ ist ein $e_iAe_i-e_jAe_j$-Modul. Gilt $U_1\supsetneqq U_2$ für zwei Untermoduln dieses Moduls, so gilt trivialerweise $AU_1A\supseteq AU_2A$. Sei $e_iae_j\in U_1, e_iae_j\notin U_2$. Wäre $AU_1A=AU_2A$, so müßte gelten

$$e_iae_j=\sum_l a_lu_la_l',\quad a_l, a_l'\in A,\quad u_l\in U_2,$$

also (mit $u_l=e_iu_le_j$)

$$e_iae_j=e_ie_iae_je_j=\sum_l(e_ia_le_i)(e_iu_le_j)(e_ja_l'e_j)\in U_2,$$

ein Widerspruch. Folglich gilt $AU_1A\supsetneqq AU_2A$. Es folgt, daß e_iAe_j der Minimalbedingung (Maximalbedingung) für Untermoduln genügt. Nach Satz 3 genügt e_iAe_j der strengen Minimalbedingung (Maximalbedingung).

$(1-e)Ae_i$ $(i\geqq t+1)$ ist ein $0-e_iAe_i$-Bimodul. Wegen $J(A)^2=(0)$ gilt $A(1-e)Ae_i=(0)$ (vgl. den Beweis von Lemma 3). Sind $U_1\supsetneqq U_2$ zwei Untermoduln von $(1-e)Ae_i$, so zeigt man wie oben $U_1A\supsetneqq U_2A$. Wegen der vorausgesetzten Minimalbedingung (Maximalbedingung) für Ideale ist also der $0-e_iAe_i$-Modul $(1-e)Ae_i$ ein Modul mit Minimalbedingung (Maximalbedingung) für Untermoduln. Also folgt nach Satz 3, daß auch $(1-e)Ae_i$ der strengen Minimalbedingung (Maximalbedingung) genügt.

Aus Symmetriegründen gilt das auch für $e_iA(1-e)$. (2) ist damit bewiesen. (3) gilt nach der Numerierung der e_i $B^2=B$ ergibt sich aus $(*)$. Sei D ein Ideal von B, das der strengen Minimalbedingung (Maximalbedingung) genügt. $DJ(B)$ $(\subseteq D\cap J(B))$ genügt auch der strengen Minimalbedingung (Maximalbedingung) und ist ein $B/J(B)$-Modul. Wegen (3) muß daher $(DJ(B))B=(0)$ sein. Analog folgt $B(DJ(B))=(0)$. Wegen $(*)$ ist daher $DJ(B)=(0)$ und analog $J(B)D=(0)$. Also ist D ein $B/J(B)$-Modul. Wieder folgt $DB=BD=(0)$, also wegen $(*)$ $D=(0)$.

Um die letzte Behauptung des Satzes zu beweisen, sei also A ein Ring mit Maximalbedingung für Ideale. Es genügt ersichtlich zu zeigen, daß e_iAe_j, $i\leqq t$, $j>t$, als e_iAe_i-Linksmodul der Maximalbedingung für Untermoduln genügt. Wir setzen der Kürze halber $\bar{e}_i\bar{A}\bar{e}_i=S$, $\bar{e}_j\bar{A}\bar{e}_j=K$.

Wegen $J(A)^2=(0)$ kann man e_iAe_j als S-Linksmodul auffassen. Angenommen, er genügte nicht der Maximalbedingung. Dann wäre e_iAe_j als vollständig reduzibler S-Modul direkte Summe unendlich vieler einfacher Untermoduln:

$$e_iAe_j=\sum_{\nu\in\Gamma}{}^{\oplus}X_\nu,\qquad \Gamma \text{ unendlich.}$$

Ferner ist e_iAe_j zugleich ein $S-K$-Bimodul mit Maximalbedingung für Untermoduln. Nun ist K endlich $(j>t)$. Sei $0\neq x_1\in X_{\nu_1}$. Dann ist x_1K endlich, also $x_1K\subseteq X_{\nu_1}\oplus\ldots\oplus X_{\nu_l}$, folglich $Sx_1K\subseteq X_{\nu_1}\oplus\ldots\oplus X_{\nu_l}$. Man wähle $0\neq x_2\in X_{\nu_{l+1}}$. Es folgt analog $Sx_1K+Sx_2K\subseteq X_{\nu_1}\oplus\ldots\oplus X_{\nu_l}\oplus\ldots\oplus X_{\nu_{l+k}}$ und $Sx_1K\subsetneq Sx_1K+Sx_2K$. Man kann also (so fortsetzend) eine streng aufsteigende Kette von Untermoduln des $e_iAe_i-e_jAe_j$-Moduls e_iAe_j angeben, im Widerspruch zur vorausgesetzten Maximalbedingung für Ideale von A. Damit ist der Beweis beendet.

Ob das Analogon der letzten Behauptung auch für die Minimalbedingung gilt, wissen wir nicht. Es gilt aber die

FOLGERUNG 7. *Es sei A ein halbprimärer Ring mit Minimal- und Maximalbedingung für Ideale und $J(A)^2=(0)$. Dann gilt*

$$A = B\oplus C,$$

(1) *B ist Ideal und C ist Unterring von A.*
(2) *C ist ein endlicher Ring.*
(3) *$B/J(B)$ ist die direkte Summe von vollen Matrizenringen über unendlichen Schiefkörpern.*
(4) *$B^2=B$, und B enthält kein Ideal $\neq(0)$, das der strengen Minimal- oder Maximalbedingung genügt.*
(5) *B ist ein Ring mit Minimal- und Maximalbedingung für Ideale.*

BEWEIS. (1), (3) und (4) folgen unmittelbar aus Satz 4. C ist einerseits ein Ring mit strenger Minimalbedingung, andererseits ein Ring mit strenger Maximalbedingung, also ist C endlich. Nach Satz 4 genügt B der Maximalbedingung für Ideale. Nach dem Beweis von Satz 4 haben die e_iAe_j, $i\leqq t$, $j>t$ endliche Dimension als $\bar{e}_i\bar{A}\bar{e}_i$-Linksmoduln, genügen also auch der Minimalbedingung. Da analoges für e_jAe_i gilt, folgt die Behauptung.

LITERATUR

[1] AYOUB, CH.: Conditions for a ring to be fissile, *Acta Math. Acad. Sci. Hungar.* 30 (1977), 233—237.
[2] JACOBSON, N.: *Structure of rings*, Providence, 1964.
[3] KERTÉSZ, A.: *Vorlesungen über artinsche Ringe*, Budapest—Leipzig, 1968.
[4] SZÁSZ, F.: Über artinsche Ringe, *Bull. Acad. Polon. Sci.* **11** (1963), 351—354.
[5] WIDIGER, A.: A general decomposition theorem for artinian rings, *Studia Sci. Math. Hungar.* **12** (1977), 29—36.
[6] WIDIGER, A.: Nichtprime Cohen-Ringe (im Druck).

Sektion Mathematik der Martin-Luther-Universität,
DDR—401 Halle, Universitätsplatz 6

(Eingegangen am 24. März, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 399—406.

S-SPECTRAL CAPACITIES AND CLOSED OPERATORS

by

B. NAGY

1. Preliminaries

Bounded decomposable operators in Banach spaces are generalizations of spectral and prespectral operators. C. APOSTOL [1] and C. FOIAŞ [6] have proved that an operator is decomposable if and only if it has a spectral capacity which is uniquely determined by the operator and which, to some extent, plays the role of the spectral measure. F.-H. VASILESCU [10], [11], [12] has defined and studied *S*-residually decomposable operators, an extension of the notion of closed spectral [4] and prespectral [7], [8] as well as of decomposable operators. The aim of this paper is to extend the result of C. Apostol and C. Foiaş to *S*-residually decomposable operators. We remark that a similar result was proved by I. BACALU [2], [3] for bounded *S*-decomposable operators. Our method of proof will be different in most parts, and will employ a technique, due to C. FOIAŞ [6], in a simplified form. We observe that a different definition of closed operators having a spectral capacity was employed by I. ERDELYI [5].

Let X be a complex Banach space and let $B(X)$ and $C(X)$ denote the class of bounded and closed linear operators in X, respectively. C will denote the complex plane and $\overline{C}$ is its compactification. Unless otherwise stated, all topological concepts will be understood in the topology of $\overline{C}$. $\overline{H}$ and H° will denote the closure and the interior, respectively, of a set $H\subset\overline{C}$, and $H^c:=\overline{C}\setminus H$. If $T\in C(X)$, then $D(T)$ denotes its domain and $\sigma(T)$ denotes its extended spectrum [9; p. 298], i.e. $\infty\in\sigma(T)$ if and only if $T\in C(X)\setminus B(X)$. Further, we set $\varrho(T):=\sigma(T)^c$. If Y is a subspace of X and $T(Y\cap D(T))\subset Y$, then $T|Y$ denotes the restriction of T to $Y\cap D(T)$.

Recall [10] that an open set $G\subset\overline{C}$ is of analytic uniqueness of $T\in C(X)$ if for any open set $H\subset G$ and any holomorphic function $f\colon H\to D(T)$ such that $(z-T)f(z)=0$ for $z\in H\cap C$, it follows $f(z)=0$ on H. For any $T\in C(X)$ there exists a unique maximal open set of analytic uniqueness Ω_T, and Ω_T^c is denoted by S_T. A holomorphic function $f_x\colon G\to D(T)$ for which $(z-T)f_x(z)=x$ if $z\in G\cap C$ is called a T-associated function of $x\in X$ on the open set G. $\delta_T(x)$ denotes the (open) set of points $z\in\overline{C}$ such that z has an open neighbourhood where a T-associated function of x exists. We put $\gamma_T(x):=\delta_T(x)^c$, $\varrho_T(x):=\delta_T(x)\cap\Omega_T$ and $\sigma_T(x):=$ $=\varrho_T(x)^c$. Hence, on $\varrho_T(x)$ there is a unique T-associated function of x, which will be denoted by $x(\cdot)$. For any $H\subset\overline{C}$ define

$$X_T(H) := \{x\in X;\ \sigma_T(x)\subset H\}$$

which is a (possibly void) linear subspace in X. A closed subspace Y in X belongs to the class I_T if $Y\subset D(T)$ and $TY\subset Y$. If F is a closed set in $\overline{C}$, define

$$I_{T,F} := \{Y\in I_T;\ \sigma(T|Y)\subset F\}.$$

If $I_{T,F}$ has an upper bound with respect to the relation $\subset$, which belongs to $I_{T,F}$, then it is denoted by $X_{T,F}$ and is called a maximal invariant space of T (on F).

Let S be a closed set in $\bar{C}$. A finite family of open sets $(G_1, G_2, \ldots, G_n; G_s)$ is an S-covering of the closed set $D \subset \bar{C}$ if $\bigcup_{j=1}^{n} G_j \cup G_s \supset D \cup S$ and $\bar{G}_j \cap S = \emptyset$ for $j=1, \ldots, n$. Suppose $T \in C(X)$ and $S \subset \sigma(T)$. T is called S-residually decomposable [10], if for each closed set $F \subset \bar{C}$ with $F \cap S = \emptyset$ there exists $X_{T,F}$, further for any S-covering $(G_1, \ldots, G_n; G_s)$ of $\sigma(T)$ there exist subspaces $X_1, X_2, \ldots$ $\ldots, X_n \in I_T$ such that $\sigma(T|X_j) \subset G_j$ $(j=1, \ldots, n)$ and $X = X_1 + \ldots + X_n + X_T(\bar{G}_s)$. Notice that [11; Proposition 3.1] implies that an S-residually decomposable operator T belongs to $B(X)$ if and only if the set S is bounded.

For a closed set $S \subset \bar{C}$ define

$$A := \{F \text{ is a closed subset of } \bar{C},\ F \cap S = \emptyset\},$$

$$B := \{F \text{ is a closed subset of } \bar{C},\ F \supset S\},$$

and let $Z := A \cup B$.

DEFINITION 1. An *S-spectral capacity* is a mapping E of Z into the family of the closed linear subspaces of X for which

(i) $E(\emptyset) = \{0\}$ and $E(\bar{C}) = X$,
(ii) for any sequence $F_n \in Z$ $(n=1, 2, \ldots)$,

$$E\left(\bigcap_{n=1}^{\infty} F_n\right) = \bigcap_{n=1}^{\infty} E(F_n),$$

(iii) for any S-covering $(G_1, \ldots, G_n; G_s)$ of $\bar{C}$

$$X = E(\bar{G}_1) + \ldots + E(\bar{G}_n) + E(\bar{G}_s).$$

Let $T \in C(X)$ and suppose S is a closed subset of $\sigma(T)$. T *has an S-spectral capacity* if there exists an S-spectral capacity E such that for all $F \in Z$

(iv) $TE(F) \subset E(F)$,

(v) $\sigma(T|E(F)) \subset F$,

and for all $F \in A$

(vi) $E(F) \subset D(T)$.

REMARK. Suppose T has an S-spectral capacity E. If S is bounded, we may choose an S-covering (G_1, G_s) of $\bar{C}$ such that $\bar{G}_s$ is bounded. By (v), then $T|E(\bar{G}_s) \in B(E(\bar{G}_s))$, and (vi) yields that $X = E(\bar{G}_1) + E(\bar{G}_s) \subset D(T)$, hence $T \in B(X)$. Conversely, $T \in B(X)$ clearly implies that S is bounded.

2. A characterization theorem

THEOREM. *$T\in C(X)$ has an S-spectral capacity E if and only if T is S-residually decomposable and $X_T(F)$ is closed in X for every $F\in B$. In this case*

$$E(F)=X_{T,F} \quad \text{for every} \quad F\in A,$$

$$E(F)=X_T(F) \quad \text{for every} \quad F\in B,$$

hence E is uniquely determined by T.

PROOF. I. Suppose T is S-residually decomposable and $X_T(F)$ is closed for every F in B. Define

$$E(F):=\begin{matrix} X_{T,F} & \text{for } F \text{ in } A \\ X_T(F) & \text{for } F \text{ in } B. \end{matrix}$$

Note that if $S=\emptyset$ (hence $A=B$), then [10; Proposition 4.1] yields that $T\in B(X)$ and T is decomposable in the sense of FOIAŞ [6]. Hence $X_T(F)=X_{T,F}$ for any F in Z, thus $E(F)$ is well-defined even in this case. Returning to the general case, it is easily seen that the mapping E satisfies properties (i), (iii) and (vi) of Definition 1 as well as properties (iv) and (v) for F in A, and property (ii) for F in B. We will show that E satisfies the remaining properties of Definition 1.

Suppose $F\in B$, $x\in X_T(F)\cap D(T)$, then $(z-T)x(z)=x$ for z in $\varrho_T(x)\cap C$, and $z\mapsto Tx(z)=zx(z)-x$ is analytic there, with values in $D(T)$. Hence $(z-T)Tx(z)=Tx$.

Further, if $\infty\in\varrho_T(x)$ and $x(z)=\sum_{n=-\infty}^{0} x_n z^n$ in a neighbourhood of infinity, then $\lim_{z\to\infty}\frac{x(z)}{z}=0$ and $\lim_{z\to\infty} T\frac{x(z)}{z}=x_0$. Since T is closed, $x_0=0$, thus $z\mapsto zx(z)$ is holomorphic at ∞. Hence $\sigma_T(Tx)\subset\sigma_T(x)$, proving (iv) for F in B.

If $F\in B$ and $z\in F^c\cap C$, then $x\in X_T(F)$ implies $x(z)\in X_T(F)$, by [10; Proposition 2.2]. If we define $T_z x:=x(z)$ on $X_T(F)$, then it can be shown as in [12; Lemma 2.8] that $T_z\in B(X_T(F))$. Further, for $x\in X_T(F)$

$$(z-T)T_z x=(z-T)x(z)=x,$$

and for $x\in X_T(F)\cap D(T)$

$$T_z(z-T)x=((z-T)x)(z)=x,$$

hence $T_z=(z-T|X_T(F))^{-1}$. Further, $F^c\ni\infty$ implies $T\in B(X)$, which proves (v) for F in B.

To prove (ii), first we show that $F\in A$, $H\in B$ (hence $F\cap H\in A$) imply $X_{T,F\cap H}=X_{T,F}\cap X_T(H)$. By definition, we have $X_{T,F\cap H}\subset X_{T,F}$. For $z\in(F^c\cup H^c)\cap C$ and $y\in X_{T,F\cap H}=:Y$ we have $(z-T)(z-T|Y)^{-1}y=y$, hence $\delta_T(y)\supset H^c$. [10; Proposition 4.2] implies $S_T\subset S\subset H$, therefore $\sigma_T(y)=\gamma_T(y)\cup S_T\subset H$. Thus we have obtained $Y\subset X_{T,F}\cap X_T(H)$, and now we prove the converse inclusion.

By (iv) and (vi), $V:=X_{T,F}\cap X_T(H)\in I_T$. We show that for every $z_0\in(F^c\cup H^c)\cap C$ the operator z_0-T is injective and surjective on V. Assume that $(z_0-T)v=0$ for some $v\in V$. If $z_0\in F^c$, then $v=0$, for z_0-T is injective on all of $X_{T,F}$. If $z_0\in H^c$, then the same is valid for $X_T(H)$. Thus z_0-T is injective on V.

Now let w be arbitrary in V. If $z_0 \in (F^c \cup H^c) \cap C$ belongs also to $\varrho(T|X_{T,F})$, then $w_0 := (z_0 - T|X_{T,F})^{-1} w \in X_{T,F}$, further $\sigma_T(w_0) = \sigma_T(w)$, by [10; Proposition 2.2], hence $w_0 \in V$ and $(z_0 - T)w_0 = w$. On the other hand, if $z_0 \in \sigma(T|X_{T,F}) \subset F$, then $z_0 \in H^c$, hence there exists $w(z_0) \in X_T(H)$ and $(z_0 - T)w(z_0) = w$. Further, by [10; Proposition 3.1], $X_{T,F}$ is a T-absorbing subspace of X, which implies $w(z_0) \in V$. Hence $z_0 - T$ is surjective on V. Finally, $V \in I_T$ implies that $T|V \in B(V)$, thus we obtain $X_{T,F\cap H} = X_{T,F} \cap X_T(H)$.

Now we prove that if the sets F_n $(n = 1, 2, \ldots)$ belong to A and $F := \bigcap_{n=1}^{\infty} F_n$, then $X_{T,F} = \bigcap_{n=1}^{\infty} X_{T,F_n}$. Notice that if some F_n is unbounded, then $F_n \cap S = \emptyset$ implies that S is bounded, hence $T \in B(X)$ and $\sigma(T)$ is bounded. Corollary 2 to Proposition 3.1 in [10] yields that $\sigma(T|X_{T,F_n}) \subset F_n \cap \sigma(T)$, hence $X_{T,F_n} = X_{T,F_n \cap \sigma(T)}$, thus we may and will assume that each F_n is bounded. Suppose first that $F \neq \emptyset$.

Set $G_n := \{z \in C : \operatorname{dist}(z, F) < n^{-1}\}$, then the sequence of the sets $\bar{G}_n$ is directed on the left by the relation $\subset$, and [10; Proposition 3.2] gives that $X_{T,F} = \bigcap_{n=k}^{\infty} X_{T,\bar{G}_n}$ where k is sufficiently large. Put $Y := \bigcap_{n=1}^{\infty} X_{T,F_n}$, then clearly $X_{T,F} \subset Y$. To prove the converse containment relation, we assume that, on the contrary, there exists y belonging to $Y \setminus X_{T,\bar{G}_r}$ for some $r \geq k$, for which $\bar{G}_r \in A$ also holds.

Now choose an open set H_s such that $G_r^c \subset H_s$ and $\bar{H}_s \subset F^c$, then (G_r, H_s) is an open S-covering of $\bar{C}$, hence $y = y_r + y_s$ with $y_r \in X_{T,\bar{G}_r}$ and $y_s \in X_T(\bar{H}_s)$. Setting $G := F^c \cap \bar{H}_s^c$, we have $\gamma_T(y) \subset \bigcap_{n=1}^{\infty} F_n = F$, hence $\varrho_T(y) \supset F^c \setminus S_T \supset G$, thus the functions $y(\cdot)$ and $y_s(\cdot)$ exist in G. Then $y_r(\cdot)$ also exists and we have for $z \in G$

$$y(z) = y_r(z) + y_s(z).$$

If D is a bounded Cauchy domain [9; p. 289] such that $F \subset D$ and $\bar{D} \subset \bar{H}_s^c$, then its positively oriented boundary $B(D)$ is contained in G and $\int_{B(D)} y_s(z)\,dz = 0$, since $y_s(\cdot)$ is holomorphic in $\bar{H}_s^c$. Further, $T|X_{T,F_n}$ (n is arbitrary but fixed) is a bounded linear operator on the subspace X_{T,F_n} with spectrum in F_n. Since $y \in X_{T,F_n}$, there exists a bounded Cauchy domain D_1 such that $F_n \subset D_1$ and $\bar{D}_1 \subset S^c$ and

$$(2\pi i)^{-1} \int_{B(D)} y(z)\,dz = (2\pi i)^{-1} \int_{B_1(D)} (z - T|X_{T,F_n})^{-1} y\,dz = y,$$

for $y(\cdot)$ is holomorphic in $F^c \cap S^c$ and coincides with the latter integrand on $B(D_1)$.

For any fixed z in G we decompose $y_r(z)$ according to the covering (G_r, H_s) and obtain

$$y_r(z) = g_r + g_s$$

where $y_r(z) \in D(T)$ and $g_r \in X_{T,\bar{G}_r} \cap D(T)$, hence $g_s \in X_T(\bar{H}_s)$ belongs also to $D(T)$. Applying $(z - T)$, we get, by what has been proved earlier,

$$y_r - (z - T)g_r = (z - T)g_s \in X_{T,\bar{G}_r} \cap X_T(\bar{H}_s) = X_{T,\bar{G}_r \cap \bar{H}_s} =: X_{rs}.$$

Since $z\in G\subset \overline{H}_s^c$, there exists

$$g=(z-T|X_{rs})^{-1}(z-T)\,g_s\in X_{rs}\subset X_T(\overline{H}_s).$$

Thus $g-g_s\in X_T(\overline{H}_s)$ and $(z-T)(g-g_s)=0$. Since $z-T$ is injective on $X_T(\overline{H}_s)$, we get $g_s=g\in X_{T,\overline{G}_r}$, hence $y_r(z)\in X_{T,\overline{G}_r}$ for any z in G. But this implies

$$y=(2\pi i)^{-1}\int\limits_{B(D)} y(z)\,dz=(2\pi i)^{-1}\int\limits_{B(D)} y_r(z)\,dz\in X_{T,\overline{G}_r},$$

a contradiction.

Suppose now that $F:=\bigcap\limits_{n=1}^{\infty} F_n$ is void, then $X_{T,F}=\{0\}\subset\bigcap\limits_{n=1}^{\infty} X_{T,F_n}$. When we prove the converse containment relation, we may assume that no F_n is void. Since $\overline{C}$ is compact, we may also assume (rearranging the indices, if necessary) that for some positive integer r the set $K_1:=\bigcap\limits_{n=1}^{r-1} F_n$ is nonvoid and $\bigcap\limits_{n=1}^{r} F_n=\emptyset=F$. Putting $K_2:=F_r$, by [10; Proposition 4.2] we have that $K_i\cap S_T=\emptyset$ $(i=1,2)$. Hence the open set

$$L:=\Omega_T\cap K_1^c\cap K_2^c$$

is nonvoid. By the preceding paragraphs $X_{T,K_1}=\bigcap\limits_{n=1}^{r-1} X_{T,F_n}$. Take any x in $\bigcap\limits_{n=1}^{\infty} X_{T,F_n}$, then $x\in X_{T,K_i}$ $(i=1,2)$. Define

$$f_i(z):=(z-T|X_{T,K_i})^{-1}x\quad\text{for}\quad z\in K_i^c\cap C\quad(i=1,2).$$

Since L is nonvoid, the functions f_i have a common holomorphic continuation on all of C, which we denote by f. Moreover, $f(\infty)=\lim\limits_{z\to\infty}(z-T|X_{T,K_i})^{-1}x=0$, hence Liouville's theorem yields that $f(z)\equiv 0$. Thus for $z\in K_i^c\cap C$ we have

$$x=(z-T)(z-T|X_{T,K_i})^{-1}x=0,$$

therefore $\bigcap\limits_{n=1}^{\infty} X_{T,F_n}=\{0\}=X_{T,F}$. Thus we have proved that E has all the properties stated in Definition 1.

II. Now suppose that T has an S-spectral capacity E. The proof of the fact that T is S-residually decomposable with the stated properties will be divided into several steps.

1° If $z\in C\setminus S$ and $(z-T)x=0$, then $x\in E(\{z\})$. Indeed, if n is a sufficiently large positive integer, then $F:=\{w\in C:\ |w-z|\leqq n^{-1}\}\in A$, $H:=\{w\in C:\ |w-z|\geqq(2n)^{-1}\}\cup\{\infty\}\in B$, and (F°, H°) is an open S-covering of $\overline{C}$, hence (iii) gives $x=x_1+x_2$, where $x_1\in E(F)$, $x_2\in E(H)$ and $x_1, x_2\in D(T)$, by (vi). Applying $z-T$ we obtain, by assumption, $0=(z-T)x_1+(z-T)x_2$. (iv) and (ii) yield $(z-T)x_2\in E(F\cap H)$, thus (v) implies

$$u:=(z-T|E(F\cap H))^{-1}(z-T)x_2\in E(F\cap H).$$

Hence $u-x_2\in E(H)$ and $(z-T)(u-x_2)=0$. By (v), $z-T$ is injective on $E(H)$, thus $x_2=u\in E(F)$, which implies $x\in E(F)$ for all n large enough. But then (ii) yields $x\in E(\{z\})$.

2° $S_T\subset S$. Indeed, we may assume that $S\neq\overline{C}$. Suppose that G is a nonvoid connected open subset of $S^c\cap C$ and $(z-T)f(z)=0$ for $z\in G$, where $f: G\to D(T)$ is holomorphic. If D_1 and D_2 are two disjoint closed disks contained in G, then 1°

gives $f(z)\in E(\{z\})\subset E(D_i)$ for $z\in D_i$. The Hahn—Banach theorem implies that $f(z)\in E(D_1)$ for $z\in G$. For $z\in D_2$ we obtain $f(z)\in E(D_1)\cap E(D_2)=\{0\}$, hence, by analyticity, $f(z)=0$ on G, which implies $S_T\subset S$.

3° For $F\in B$ we have $E(F)\subset X_T(F)$. Indeed, we may assume that $F\neq\bar{C}$. By (v), for $z\in F^c\cap C$ there exists $(z-T|E(F))^{-1}\in B(E(F))$, and 2° implies the statement.

4° If the open set G contains $S\cup\sigma_T(x)$, then $x\in E(\bar{G})$.

To prove this put $G_1:=G$ and let G_2 be an open subset of $\bar{C}$ such that $G_1\cup G_2=\bar{C}$ and $S\cup\sigma_T(x)\subset\bar{G}_2^c$. Then, by property (iii), $x=x_1+x_2$ with $x_i\in E(\bar{G}_i)$ $(i=1,2)$. Define the set G' by $S^c\cap\varrho_T(x)\cap\bar{G}_2^c$, then for z in G' there exists $(z-T|E(\bar{G}_2))^{-1}\in B(E(\bar{G}_2))$. Hence $G'\subset\varrho_T(x)\cap\varrho_T(x_2)\subset\varrho_T(x_1)$, and for z in G'

$$x_1(z)=x(z)-x_2(z),\quad\text{and}\quad x_2(z)=(z-T|E(\bar{G}_2))^{-1}x_2.$$

Now we show that for every z in G', $x_1(z)\in E(\bar{G}_1)$. For a fixed z we obtain, by (iii)

$$x_1(z)=g_1+g_2,\quad\text{where}\quad g_i\in E(\bar{G}_i)\quad(i=1,2).$$

Further, $\bar{G}_2\in A$ implies $g_2\in D(T)$, thus $x_1(z)\in D(T)$ implies also $g_1\in D(T)$, and an application of $z-T$ gives

$$x_1-(z-T)g_1=(z-T)g_2\in E(\bar{G}_1\cap\bar{G}_2).$$

Since $z\in G'\subset\bar{G}_2^c$, there exists

$$g=(z-T|E(\bar{G}_1\cap\bar{G}_2))^{-1}(z-T)g_2\in E(\bar{G}_1\cap\bar{G}_2)\subset E(\bar{G}_2).$$

Hence $g-g_2\in E(\bar{G}_2)$ and $(z-T)(g-g_2)=0$. But $z-T$ is injective on $E(\bar{G}_2)$, hence $g_2=g\in E(\bar{G}_1)$, thus $x_1(z)\in E(\bar{G}_1)$.

Now we distinguish between two cases. If $T\in B(X)$, then $S\subset C$ and $\sigma_T(x)\subset$ $\subset\sigma(T)\subset C$ imply that $S\cup\sigma_T(x)$ is compact in the topology of C, hence there exists a bounded Cauchy domain D such that $S\cup\sigma_T(x)\subset D$ and $\bar{D}\subset\bar{G}_2^c\cap C$, thus its positively oriented boundary, $B(D)\subset G'$. Then

$$(2\pi i)^{-1}\int_{B(D)}x(z)\,dz=(2\pi i)^{-1}\int_{|z|=|T|+1}(z-T)^{-1}x\,dz=x.$$

Further, $\bar{G}_2^c\subset\varrho(T|E(\bar{G}_2))$ implies

$$\int_{B(D)}x_2(z)\,dz=\int_{B(D)}(z-T|E(\bar{G}_2))^{-1}x_2\,dz=0,$$

hence

$$x=(2\pi i)^{-1}\int_{B(D)}x_1(z)\,dz\in E(\bar{G}).$$

If $T\in C(X)\setminus B(X)$, then $\infty\in S\cup\sigma_T(x)$, and this set is compact in the topology of $\bar{C}$, and we may assume that it is a proper subset of $\bar{C}$ (otherwise $E(\bar{G})=X$). Then there exists an unbounded Cauchy domain D [9; p. 293] such that $S\cup\sigma_T(x)\subset D$, $\bar{D}\subset\bar{G}_2^c$, with positively oriented boundary $B(D)\subset G'$. The set $\bar{D}^c$ is a bounded Cauchy domain with positively oriented boundary $-B(D)$, and $x(\cdot)$

is holomorphic in $\overline{D}^c$, hence $\int\limits_{B(D)} x(z)dz=0$. Further, $\sigma(T|E(\overline{G}_2))\subset\overline{D}^c$, thus

$$(2\pi i)^{-1}\int\limits_{B(D)} x_2(z)\,dz=(2\pi i)^{-1}\int\limits_{B(D)}(z-T|E(\overline{G}_2))^{-1}x_2\,dz=-x_2,$$

hence

$$x_2=(2\pi i)^{-1}\int\limits_{B(D)} x_1(z)\,dz\in E(\overline{G}).$$

From this $x=x_1+x_2\in E(\overline{G})$, and 4° is proved.

5° For $F\in B$ we have $E(F)=X_T(F)$. Indeed, we clearly can construct open sets G_n such that $F\subset G_n$ $(n=1, 2, \ldots)$ and $F=\bigcap\limits_{n=1}^{\infty}\overline{G}_n$. If $x\in X_T(F)$, then $x\in E(\overline{G}_n)$ for every n, by 4°, hence $X_T(F)\subset E(F)$, and 3° proves the statement.

6° For $F\in A$ we have $E(F)=X_{T,F}$.

By (iv)—(vi) of Definition 1 we obtain that $E(F)$ belongs to the class $I_{T,F}$. We will show that $Y\in I_{T,F}$ implies $Y\subset E(F)$.

Suppose, on the contrary, that $Y\in I_{T,F}$ and there exists $y\in Y\setminus E(F)$. Let H_n be a sequence of open sets such that $F\subset H_n$, $\overline{H}_n\cap S=\emptyset$ $(n=1, 2, \ldots)$ and $F=\bigcap\limits_{n=1}^{\infty}\overline{H}_n$, then for some r we have $y\notin E(\overline{H}_r)$. Put $G_r:=H_r$ and choose an open set G_s with the following properties: $G_r^c\subset G_s$ and $\overline{G}_s\subset F^c$. Then (G_r, G_s) is an open S-covering of $\overline{C}$, hence $y=y_r+y_s$, where $y_r\in E(\overline{G}_r)$ and $y_s\in E(\overline{G}_s)=X_T(\overline{G}_s)$ in view of $\overline{G}_s\in B$ and 5°. Put now $G:=\overline{G}_s^c\cap F^c$. The function $y_s(\cdot)$ exists on $\overline{G}_s^c$, whereas $(z-T|Y)^{-1}\in B(Y)$ for z in F^c, hence $y(z)=(z-T|Y)^{-1}y$ exists for $z\in F^c\cap S_T^c\supset G$. Thus $y_r(\cdot)$ exists on G and

$$y(z)=y_r(z)+y_s(z)\quad\text{for}\quad z\in G.$$

For a fixed z in G we decompose $y_r(z)$ as follows:

$$y_r(z)=g_r+g_s\quad\text{where}\quad g_i\in E(\overline{G}_i)\quad(i=r, s).$$

Since $z\in G\subset\overline{G}_s^c$, the same technique as in 4° yields $g_s\in E(\overline{G}_r\cap\overline{G}_s)\subset E(\overline{G}_r)$, hence $y_r(z)\in E(\overline{G}_r)$ for z in G.

We make a distinction depending on the boundedness of F. If F is bounded, then there exists a bounded Cauchy domain D such that $F\subset D$ and $\overline{D}\subset\overline{G}_s^c$, thus $B(D)\subset G$. We obtain

$$(2\pi i)^{-1}\int\limits_{B(D)} y(z)\,dz=y\quad\text{and}\quad\int\limits_{B(D)} y_s(z)\,dz=0,$$

for $y_s(\cdot)$ is holomorphic in $\overline{G}_s^c$. Hence

$$y=(2\pi i)^{-1}\int\limits_{B(D)} y_r(z)\,dz\in E(\overline{G}_r),$$

a contradiction. If F is unbounded, then $\infty\in F$, hence S is bounded and $T\in B(X)$. $y\in Y\setminus E(F)$ implies that F is a proper subset of $\overline{C}$. Then there exists an unbounded Cauchy domain D satisfying the inclusion relations above, and $\overline{D}^c$ is a bounded

Cauchy domain with positively oriented boundary $-B(D)$. Then

$$\int_{B(D)} y(z)\,dz = \int_{B(D)} (z-T|Y)^{-1} y\,dz = 0,$$

for the latter integrand is holomorphic in F^c. On the other hand, $y_s(\cdot)$ is holomorphic in $\bar{G}_s^c$ and $y_s(z)=(z-T)^{-1}y_s$ for $|z|>|T|$, hence for K large enough

$$(2\pi i)^{-1} \int_{B(D)} y_s(z)\,dz = -(2\pi i)^{-1} \int_{|z|=K} (z-T)^{-1} y_s\,dz = -y_s.$$

Thus we obtain

$$y_s = (2\pi i)^{-1} \int_{B(D)} y_r(z)\,dz \in E(\bar{G}_r),$$

which implies $y=y_r+y_s\in E(\bar{G}_r)$, a contradiction. Notice that we have proved 6° and the uniqueness of the spectral capacity.

7° T is S-residually decomposable.

Indeed, for F in A, $X_{T,F}$ exists, by 6°. Further, for every open S-covering $\left((G_j)_{j=1}^n, G_s\right)$ of $\sigma(T)$ there exists an open S-covering $\left((H_j)_{j=1}^n, G_s\right)$ of $\sigma(T)$ such that $\bar{H}_j\subset G_j$ $(j=1, 2, \ldots, n)$. Also, there exists an open set H_0 such that $\bar{H}_0\cap\sigma(T)$ is void and $\bigcup_{j=0}^{n} H_j\cup G_s=\bar{C}$. By (iii), we have $X=\sum_{j=0}^{n} E(\bar{H}_j)+E(\bar{G}_s)$. Further, $\sigma(T)\in B$, hence $E(\sigma(T))=X_T(\sigma(T))=X$, since $\sigma_T(x)\subset\sigma(T)$ for each x in X. Then $E(\bar{H}_0)=E(\bar{H}_0\cap\sigma(T))=\{0\}$, thus $X=\sum_{j=1}^{n} X_{T,\bar{H}_j}+X_T(\bar{G}_s)$, and 7° is proved.

8° For F in B, $X_T(F)=E(F)$ is closed, and the proof is complete.

REFERENCES

[1] APOSTOL, C.: Spectral decompositions and functional calculus, *Rev. Roum. Math. Pures Appl.* **13** (1968), 1481—1528.

[2] BACALU, I.: S-spectral capacities, *Studii Cerc. Mat.* **26** (1974), 1189—1195 (in Rumanian).

[3] BACALU, I.: The uniqueness of the multidimensional S-spectral capacity, *Studii Cerc. Mat.* **28** (1976), 131—134 (in Rumanian).

[4] BADE, W. G.: Unbounded spectral operators, *Pacific J. Math.* **4** (1954), 373—392.

[5] ERDELYI, I.: Unbounded operators with spectral capacities, *J. Math. Analysis Appl.* **52** (1975), 404—414.

[6] FOIAŞ, C.: Spectral capacities and decomposable operators, *Rev. Roum. Math. Pures Appl.* **13** (1968), 1539—1545.

[7] NAGY, B.: Unbounded prespectral operators, *Periodica Math. Hung.* **9** (1978), 277—283.

[8] NAGY, B.: Analytic functions of prespectral operators, *Acta Math. Acad. Sci. Hung.* **31** (1978), 157—171.

[9] TAYLOR. A. E.: *Introduction to functional analysis*, Wiley, New York, 1958.

[10] VASILESCU, F.-H.: Residually decomposable operators in Banach spaces, *Tôhoku Math. J.* **21** (1969), 509—522.

[11] VASILESCU, F.-H.: Residual properties for closed operators on Fréchet spaces, *Illinois J. Math.* **15** (1971), 377—386.

[12] VASILESCU, F.-H.: On the residual decomposability in dual spaces, *Rev. Roum. Math. Pures Appl.* **16** (1971), 1573—1587.

Department of Mathematics, Faculty of Chemistry, University of Technology
Stoczek u. 2, H—1111 Budapest, Hungary

(Received 29 April, 1978; revised October 8, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 407—408.

A CHARACTERIZATION OF THE CLASS OF COMPACT HAUSDORFF SPACES

by

D. PETZ

Abstract. The category of compact Hausdorff spaces is characterized as the only full, non-trivial, epireflective subcategory of the category of Hausdorff spaces and continuous maps which is preserved by epimorphisms.

Fixing a category of topological spaces, for example the category T_2 of Hausdorff spaces and continuous maps, classes of spaces can be characterized by means of the global behaviour of the objects and morphisms. Such kinds of characterizations were obtained in [1] for various familiar subcategories of T_2, among others for the full subcategory CT_2 of compact Hausdorff spaces: CT_2 is the only non-trivial, left-fitting, productive subcategory of T_2 preserved by shrinks, in TOP, of T_2 extremal monos. A very algebraic characterization is due to Herrlich and Strecker: CT_2 is the only non-trivial, epireflective subcategory of T_2 which is varietal ([3]).

In this paper a subcategory is assumed to be full and isomorphism-closed, and a subcategory is called non-trivial provided that it is proper and contains a space with at least two points. With this terminology we state

THEOREM. *CT_2 is the only non-trivial, epireflective subcategory of T_2 which is preserved by epimorphisms.*

PROOF. It is well-known that in T_2 an epireflective subcategory is a productive and closed-hereditary class and the epimorphisms are precisely the continuous dense maps. So the theorem can be reformulated in the form: CT_2 is the only non-trivial subcategory $\mathscr{C}$ of T_2 which has the following properties

(i) productive (i.e., if $X_i \in \mathscr{C}$ $(i \in I)$ then $\prod \{X_i : i \in I\} \in \mathscr{C}$);
(ii) closed-hereditary (i.e., if $X \in \mathscr{C}$ and Y is a closed subspace of X then $Y \in \mathscr{C}$);
(iii) preserved by continuous onto maps (i.e., if $f: X \to Y$ is a continuous onto map, $X \in \mathscr{C}$ and $Y \in T_2$ then $Y \in \mathscr{C}$);
(iv) extensional (i.e., if $X \in \mathscr{C}$ and Y is a Hausdorff extension of X then $Y \in \mathscr{C}$).

Let $\mathscr{C}$ be a subcategory of T_2 which satisfies the conditions (i)—(iv) and contains a non-singleton space. Then it contains the discrete dyad D and the Cantor set D^ω. Since the continuous image $[0, 1]$ of D^ω is in $\mathscr{C}$, hence according to the condition (i) and (ii) $\mathscr{C}$ contains all compact spaces.

Let us suppose that there is a non-compact space X in $\mathscr{C}$. Since a Hausdorff space is compact if and only if its every closed subspace is H-closed (see [4]), hence there is a closed subspace Y of X which is not H-closed and Y is contained in $\mathscr{C}$. Let $Z = Y \cup \{p\}$ be a one-point extension of Y and $\varkappa$ be a cardinal. $W_\varkappa = Z^\varkappa \setminus \{(p)\}$

is an extension of $Y^\varkappa$, so $W_\varkappa$ is in $\mathscr{C}$. Let us define

$$D_\varkappa = \{x \in Z^\varkappa : x_i \text{ equals to } p \text{ or to } q \text{ for any } i \in \varkappa \text{ and } x_i \text{ equals to } q \text{ for only one } i \in \varkappa\}$$

where $q \in Y$ is a fixed point. $D_\varkappa$ is a closed discrete subspace of $W_\varkappa$ with cardinality $\varkappa$. Hence we obtain that a subcategory $\mathscr{C}$ satisfying the conditions (i)—(iv) and containing a non-compact space must contain all discrete spaces and so by virtue of the condition (iii) $\mathscr{C}$ is identical with the whole T_2.

Finally, we show some classes of Hausdorff spaces which satisfy three of the four conditions.

EXAMPLE 1. The class of H-closed spaces is productive and preserved by epimorphisms but it is not closed-hereditary.

EXAMPLE 2. The class of finite discrete spaces is closed-hereditary and preserved by epimorphisms but it is not productive.

EXAMPLE 3. A Hausdorff space is said to be ω-bounded provided the closure of any countable subspace is compact. The class of ω-bounded spaces is productive, preserved by continuous onto maps and closed-hereditary but it is not extensional.

EXAMPLE 4. The class of compact 0-dimensional spaces is productive, closed-hereditary and extensional but it is not preserved by onto maps.

BIBLIOGRAPHY

[1] FRANKLIN, S. P., LUTZER, D. J. and THOMAS, B. V. S.: On subcategories of TOP, *Trans. Amer. Math. Soc.* **225** (1977), 267—278.

[2] HERRLICH, H.: Topologische Reflexionen und Coreflexionen, *Lecture Notes in Math.*, no. **78**, Springer-Verlag, Berlin and New York, 1968.

[3] HERRLICH, H. and STRECKER, G. E.: Algebra $\cap$ Topology=compactness, *General Topology and Appl.* **1** (1971), 283—287.

[4] ILIADIS, S. D. and FOMIN, S. V.: The method of centered systems in the theory of topological spaces, *Uspehi Mat. Nauk* **21** 4. (1966), 47—76 (in Russian).

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received May 22, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 409—412.

A REMARK ON THE HADWIGER NUMBERS OF A CONVEX DISC

by

L. FEJES TÓTH and A. HEPPES

A set of closed convex bodies is said to form a packing if no two bodies have interior points in common. In a packing the bodies a and b are said to be neighbours of order k or kth neighbours if there are in the packing $k+1$ bodies, but not less than $k+1$ bodies, containing a and b and having a connected point-set-union. The kth Hadwiger number H_k of a body b is defined as the maximal number of its neighbours of order $\leqq k$ extended over all packings of translates of b.

It is known [1, 2, 3] that in n-space we have for any convex body

$$H_k \leqq (2k+1)^n - 1$$

with equality only for parallelohedra. What is the minimum of H_k for all n-dimensional convex bodies? We shall see that this question is even for $n=2$ hopeless.

In the plane we have for any convex disc

$$3k(k+1) \leqq H_k \leqq 4k(k+1).$$

The upper bound is attained for parallelograms. The lower bound follows from the fact that in a lattice-packing of translates of a convex disc in which each disc has six first neighbours, a disc has 2.6 second neighbours, 3.6 third neighbours, and so on, and thus $1.6+2.6+\ldots+k\cdot 6=3k(k+1)$ neighbours of order $\leqq k$. For a strictly convex disc d the above lattice-packing is a special case of a more general packing of translates of d which arises by adding to d the maximal number of first neighbours, then the maximal number of second neighbours and so on. In each step of this process we can choose the position of a new neighbour of a certain order arbitrarily by which the rest of the neighbours of this order is uniquely determined [4].

It is known [5] that this "greedy" construction is for each disc d only up to a certain value $k_0=k_0(d)$ renumerative, so that for $k>k_0$ we have $H_k>3k(k+1)$. But it is imaginable that there is a set of discs $d_1, d_2, \ldots$ such that the sequence $k_0(d_1), k_0(d_2), \ldots$ tends to infinity. We shall rule out this possibility by proving the following

THEOREM. *For $k>24$ we have for any convex disc $H_k>3k(k+1)$.*

Since the kth Hadwiger number of a convex disc d is equal to the kth Hadwiger number of the disc arising from d by central symmetrisation we can restrict ourselves to a centro-symmetric convex disc c. According to a theorem of DOWKER [6], there is among the hexagons circumscribed about c of least area a hexagon H symmetric about the centre C of c. The side-midpoints of H determine an affinely

regular hexagon h inscribed into c. It is easy to see that $|H|/|h|=4/3$, where $|x|$ denotes the area of the domain x.

Let $\bar{H}$ be the hexagon of area $|H|$ concentric with and homothetic to h (Fig. 1). The disc c must intersect a side s of $\bar{H}$, since otherwise there would be a hexagon containing c of smaller area than $|\bar{H}|=|H|$, contrary to the choice of H. We suppose that s is in a horizontal position above C. Let P be the highest point of c, and Q the intersection of h and the segment CP. We have $CP/CQ \geqq 2/\sqrt{3}$.

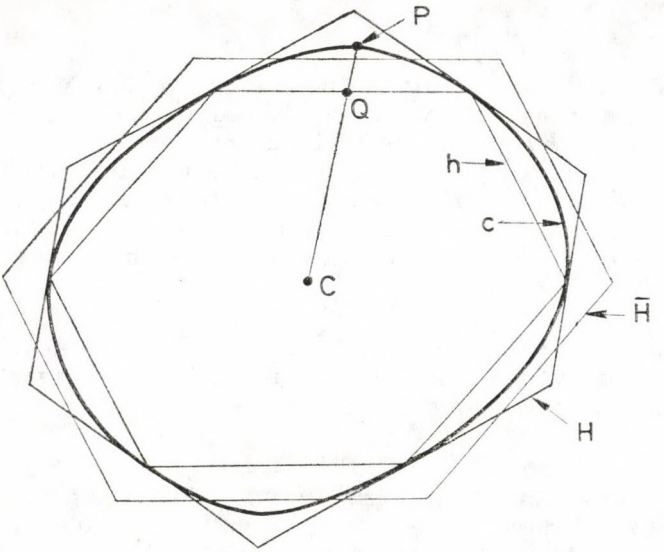


Fig. 1

Let L be the lattice-packing of translates of c which arises by tiling the plane with translates of H and inscribing into each hexagon a translate of c. In L let A and B be the centres of two discs a and b, both touching c, as well as each other, such that the triangle ABC contains P. Let r_i be the horizontal row of discs which contains those ith neighbours of c in L whose centres lie in the closed angular region ACB. Let c_i be the disc arising from c by the translation $2i\overrightarrow{CP}$. In the sequence $c_1, c_2, \ldots$ let c_j be the first disc which has a point in common with a disc c' of r_{j+2}. Since

$$7 \cdot 2\overline{CP} \geqq 7 \cdot 2\frac{2}{\sqrt{3}}\overline{CQ} > 8 \cdot 2\overline{CQ},$$

the centre of c_7 lies above the central line of the row r_8, showing that $j \leqq 7$.

Let us first suppose that besides c' c_j has no interior point in common with another disc of r_{j+2}.

Remove from L all discs which have interior points in common with any of $c_1, \ldots, c_j$. We claim that the number of these discs is at most $3j+1$. To see this observe that any row is overlapped by at most two members of the sequence and any pair of touching translates intersect at most three discs of a row. Thus our

sequence intersects at most three discs of a row. But from r_1 we loose only a and b. The row r_2 is intersected only by c_1 and c_2, but the only disc in r_2 intersected by c_1 (which arises from c by the translation $\overrightarrow{CA}+\overrightarrow{CB}$) is also intersected by c_2. Thus we loose also from r_2 only two discs. The same is true for r_{j+1} because this row is intersected only by c_j. By supposition, we loose from r_{j+2} only c' and from r_{j+3} no disc. Thus the number of discs in question is at most $2+2+2+1+3(j-2)=3j+1$, as claimed.

By the removal of a and b some discs which were $(j+2)$nd neighbours of c in L become $(j+3)$rd neighbours of c in the mutilated packing. Therefore we start, instead of c, with the disc c_0 which arises from c by the translation $\overrightarrow{AC}+\overrightarrow{BC}$. It is easy to check that each neighbour of c_0 of order $\leqq j+4$ in L which has not been removed remains a neighbour of c_0 of order $\leqq j+4$ in the mutilated packing. Adding to the mutilated packing the discs $c_1, \ldots, c_j$ we obtain a new packing in which c_0 has fewer neighbours of order $\leqq j+4$ than in L. However, the number of the neighbours lost is not more than $2j+1$.

Now we translate c' upwards parallel to those sides of H which are not intersected by the lines AC and BC so as to obtain a new disc c_{j+1} which has a boundary point but no interior point in common with c_j. Consider all discs in L which arise from c' by a translation $\overrightarrow{CX}$ where the point X lies in the closed angular region ABC. Together with c' translate these discs in the same direction and through the same distance as c' and replace the original discs by their translates. In the packing L' obtained in this way c_{j+1} is a $(j+3)$ rd neighbour of c_0 and the two first neighbours of c_{j+1} above it are $(j+4)$ th neighbours of c_0. Thus in L' the defect in neighbours of order $\leqq j+4$ compared with L is at most $2j+1-3=2j-2$. But since for $k \geqq j+4$ c_0 has in L' one more kth neighbours than in L, the defect in neighbours of order $\leqq k$ decreases by one when we go over from k to $k+1$. Thus c_0 has in L' at least as many neighbours of order $\leqq 3j+2$ as in L and more neighbours of order $\leqq 3j+3$ in L' than in L.

We still have to discuss the case when c_j intersects the interior of two discs c' and c'' of the row r_{j+2}. Now we consider all discs in L which can be obtained either from c' or from c'' by a translation in a direction $\overrightarrow{CX}$. We translate simultaneously c' and c'' along with these discs so that they should not interfere with other discs of L until c_j has neither with c' nor with c'' interior points in common but touches one of them. This construction also guarantees more neighbours of order $\leqq 3j+3$ of c_0 in the new packing than in L. Since $j \leqq 7$, we have $3j+3 \leqq 24$.

This completes the proof of the theorem.

By a modification of the above construction, the number 24 in the theorem can be reduced to 23. First we consider the case when $CP/CQ \leqq 6/5$. Let c' be a disc in r_9 intersected by c_7. Consider the discs c_7', c_6' and c_5' which arise from c' by the translations $2\overrightarrow{PC}$, $4\overrightarrow{PC}$ and $6\overrightarrow{PC}$ and connect c_3 and c_5' by a disc c_4' touching both c_3 and c_5'. Now the discs $c_1, c_2, c_3, c_4', c_5', c_6', c_7'$ intersect at most two discs from r_1, r_2, r_7 and r_8, and it can be shown that they intersect at most three discs from $r_3, \ldots, r_6$. Thus we have to remove at most 20 discs.

In the case when $CP/CQ > 6/5$ we have, in the original proof, $j \leqq 5$ so that $3j+3 \leqq 18$.

REFERENCES

[1] HADWIGER, H., DEBRUNNER, H. and KLEE, V.: *Combinatorial geometry in the plane*, New York, 1964, Theorem 43, p. 18.
[2] GROEMER, H.: Abschätzungen für die Anzahl der konvexen Körper, die einen konvexen Körper berühren, *Monatsh. Math.* **65** (1961), 74—81.
[3] GRÜNBAUM, B.: On a conjecture of Hadwiger, *Pacific J. Math.* **11** (1964), 215—219.
[4] FEJES TÓTH, L.: Egy záródási tétel, *Mat. Lapok* **23** (1972), 9—12.
[5] FEJES TÓTH, L.: On Hadwiger numbers and Newton numbers of a convex body, *Studia Sci. Math. Hungar.* **10** (1975), 111—115.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received July 7, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 413—417.

ON THE RATE OF CONVERGENCE IN THE MARTINGALE CENTRAL LIMIT THEOREM

by

T. MÓRI

Abstract. Heyde and Brown have given an estimate of the departure from normality of a certain class of martingales. The proof of their theorem is based on the martingale version of the Skorokhod embedding theorem. The aim of the present paper is to give a more general estimate of the rate of convergence in the martingale CLT. The method to be applied is based on the Esseen inequality and on the proof of the martingale CLT due to Brown.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and

$$\{X_{nk}, \mathcal{F}_{nk}\colon\ n = 1, 2, \ldots;\ k = 1, 2, \ldots, k_n\}$$

be a martingale difference array (MDA) defined on it, i.e., for each $n \geqq 1$ and $1 \leqq k \leqq k_n$ we suppose that $\mathcal{F}_{n,k-1} \subset \mathcal{F}_{nk}$, X_{nk} is $\mathcal{F}_{nk}$-measurable and $E(X_{nk}|\mathcal{F}_{n,k-1}) = 0$ a.s ($\mathcal{F}_{n0}$ is the trivial field).

We shall denote the conditional expectation $E(\cdot\,|\mathcal{F}_{nk})$ by $E_k(\cdot)$, if hereby no ambiguity can arise. For the sake of convenience we shall write $\sum_k$ instead of $\sum_{k=1}^{k_n}$.

Assume that $EX_{nk}^2 < \infty$ and define $S_{nk} = \sum_{j=1}^{k} X_{nj}$, $S_n = S_{n,k_n}$, $\sigma_{nk}^2 = E_{k-1}X_{nk}^2$, $V_{nk}^2 = \sum_{j=1}^{k} \sigma_{nj}^2$, $V_n^2 = V_{n,k_n}^2$, $A_n = E|V_n^2 - 1|$. Denote by F_n the distribution function of the variable S_n. The standard normal distribution function will be denoted, as usual, by Φ.

We shall need some simple properties of the Lévy metric $\mathcal{L}$ defined on the set of all distribution functions. Let X, Y be random variables with distribution functions F and G, resp. Then

$$\mathcal{L}(F, G) \leqq [E(X-Y)^2]^{1/3} \tag{1}$$

and

$$\mathcal{L}(F, G) \leqq \sup_x |F(x) - G(x)|. \tag{2}$$

If G is totally continuous, then

$$\sup_x |F(x) - G(x)| \leqq (1 + g)\mathcal{L}(F, G) \tag{3}$$

where $g = \sup_x |G'(x)|$.

By the help of the term $L(n, \varepsilon) = \sum_k E\left(X_{nk}^2 I(|X_{nk}| > \varepsilon)\right)$ (where I denotes the

indicator function) the (strong) Lindeberg condition can be formulated as follows:

$$\lim_{n\to\infty} L(n,\varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0.$$

It is clear that the Lindeberg condition is equivalent to

$$W_n = \int_0^1 L(n,\varepsilon)\,d\varepsilon \to 0 \quad \text{as} \quad n \to \infty,$$

supposing that the sequence A_n is bounded.

Now we can formulate our

THEOREM.

$$\sup_x |F_n(x) - \Phi(x)| \leqq 7(W_n^{1/4} + A_n^{1/3}). \tag{4}$$

This theorem seems to compare with the result of HEYDE and BROWN (1970) in a similar manner as the Lindeberg theorem does with the Ljapunov version of the CLT. We mention, however, that applying a suitable refinement of the method due to Heyde and Brown one can estimate the departure from normality of the row sum S_n by $C(W_n^{2/9} + A_n^{1/3})$.

In the case of independent summands HERTZ (1969) has used W_n for estimating the rate of convergence in the CLT. As a consequence of her results the following estimate can be obtained:

$$\sup_x |F_n(x) - \Phi(x)| \leqq C(W_n + A_n),$$

and this is sharp in the sense that neither the power of W_n nor that of A_n can be increased. We should remark, that (4) seems to be far from being sharp even in the martingale case.

PROOF OF THE THEOREM. First we deal with the case

$$V_n^2 \leqq 1 \quad \text{a.s.} \tag{5}$$

By partial summation we get

$$\begin{aligned} &\left|E\left(\exp\left\{itS_n + \frac{1}{2}t^2V_n^2\right\}\right) - 1\right| = \\ &= \left|\sum_k E\left(\exp\left\{itS_{nk} + \frac{1}{2}t^2V_{nk}^2\right\} - \exp\left\{itS_{n,k-1} + \frac{1}{2}t^2V_{n,k-1}^2\right\}\right)\right| \leqq \\ &\leqq \sum_k \left|E\left(\exp\left\{itS_{n,k-1} + \frac{1}{2}t^2V_{nk}^2\right\} E_{k-1}\left(\exp\{itX_{nk}\} - \exp\left\{-\frac{1}{2}t^2\sigma_{nk}^2\right\}\right)\right)\right| \leqq \\ &\leqq e^{t^2/2} \sum_k E\left|E_{k-1}\left(\exp\{itX_{nk}\} - \exp\left\{-\frac{1}{2}t^2\sigma_{nk}^2\right\}\right)\right| \end{aligned} \tag{6}$$

for every real t.

It is easy to see that for real x

$$|e^{ix}-1-ix+\frac{1}{2}x^2| \leqq \frac{1}{2}x^2\mu(x) \tag{7}$$

where $\mu(x)=\min\{2, \frac{1}{3}|x|\}$, and for positive x

$$|e^{-x}-1+x| \leqq v(x) \tag{8}$$

where $v(x)=\frac{1}{2}x^2$ if $x\leqq 2$ and $2x-2$ otherwise.

Now using (7) and (8) then applying the Jensen inequality to the convex function $v(x)$ we get

$$\begin{aligned}
&\sum_k E\left|E_{k-1}(\exp\{itX_{nk}\}-\exp\{-\tfrac{1}{2}t^2\sigma_{nk}^2\})\right| = \\
&= \sum_k E\left|E_{k-1}(\exp\{itX_{nk}\}-1-itX_{nk}+\tfrac{1}{2}t^2X_{nk}^2-\exp\{-\tfrac{1}{2}t^2\sigma_{nk}^2\}+1-\tfrac{1}{2}t^2\sigma_{nk}^2)\right| \leqq \\
&\leqq \sum_k E\left(\tfrac{1}{2}t^2X_{nk}^2\mu(tX_{nk})\right)+\sum_k Ev\left(\tfrac{1}{2}t^2\sigma_{nk}^2\right) \leqq \\
&\leqq \frac{1}{2}t^2\sum_k E\left(X_{nk}^2\mu(tX_{nk})\right)+\sum_k Ev\left(\tfrac{1}{2}t^2X_{nk}^2\right).
\end{aligned} \tag{9}$$

Since we may assume that $t\neq 0$, the right side of (9) can be treated as follows:

$$\begin{aligned}
&\frac{1}{2}t^2\sum_k E\left(X_{nk}^2\mu(tX_{nk})\right) = \\
&= \frac{1}{2}t^2\sum_k E\left(X_{nk}^2\int_0^\infty I(y<|tX_{nk}|)\right)d\mu(y) = \\
&= \frac{1}{2}t^2\int_0^\infty \sum_k E\left(X_{nk}^2 I(y<|tX_{nk}|)\right)d\mu(y) = \\
&= \frac{1}{6}t^2\int_0^6 L\left(n, \frac{y}{|t|}\right)dy = \frac{1}{6}|t|^3\int_0^{6/|t|} L(n,\varepsilon)\,d\varepsilon \leqq \\
&\leqq \frac{1}{6}|t|^3\left(1+\frac{6}{|t|}\right)W_n = \left(t^2+\frac{1}{6}|t|^3\right)W_n.
\end{aligned} \tag{10}$$

By similar arguments

$$\sum_k Ev\left(\tfrac{1}{2}t^2X_{nk}^2\right) \leqq \left(t^2+\frac{1}{2}|t^3|\right)W_n. \tag{11}$$

From (6), (9), (10) and (11) it follows that

$$\left|E\left(\exp\left\{itS_n+\frac{1}{2}t^2V_n^2\right\}\right)-1\right| \leqq e^{t^2/2}\left(2t^2+\frac{2}{3}|t|^3\right)W_n.$$

Since by (5) we have

$$E\left|e^{\frac{1}{2}t^2}-e^{\frac{1}{2}V_n^2}\right| \leqq E\left|e^{t^2/2}\frac{1}{2}t^2(1-V_n^2)\right| = \frac{1}{2}t^2e^{t^2/2}A_n,$$

the following inequality is valid for the difference between the characteristic function of S_n and that of the standard normal distribution:

$$(12)\quad \left|E(\exp\{itS_n\})-\exp\left\{-\frac{1}{2}t^2\right\}\right| =$$

$$=\exp\left\{-\frac{1}{2}t^2\right\}\left|E\left(\exp\left\{itS_n+\frac{1}{2}t^2V_n^2\right\}\right)-1+E\left[\exp\{itS_n\}\left(\exp\left\{\frac{1}{2}t^2\right\}-\exp\left\{\frac{1}{2}t^2V_n^2\right\}\right)\right]\right| \leqq$$

$$\leqq \left(2t^2+\frac{2}{3}|t|^3\right)W_n+\frac{1}{2}t^2A_n.$$

Now it is time to apply the Esseen theorem (see LOÈVE (1955)). With the notation $Z_n=W_n^{1/4}+A_n^{1/3}$

$$\sup_x|F_n(x)-\Phi(x)| \leqq \frac{2}{\pi}\int_0^{1/Z_n}\left[\left(2t+\frac{2}{3}t^2\right)W_n+\frac{1}{2}tA_n\right]dt+\frac{24}{\pi}(2\pi)^{-1/2}Z_n \leqq$$

$$\leqq \frac{1}{\pi}\left[2Z_n^2+\frac{4}{9}Z_n+\frac{1}{2}Z_n+24(2\pi)^{-1/2}Z_n\right] \leqq 4Z_n$$

because we may assume that $Z_n \leqq \frac{1}{4}$. This completes the first step of the proof.

The general case is fairly easy to reduce to the special case proved above. We have only to apply the familiar technique to be found e.g. in BILLINGSLEY (1961). Let

$$\tau_n = k_n \quad \text{if} \quad V_n^2 < 1, \quad \text{and} \quad \tau_n = k \quad \text{if} \quad V_{n,k-1}^2 < 1 \leqq V_{nk}^2,$$

$$c_n = 1 \quad \text{if} \quad V_n^2 < 1, \quad \text{and} \quad 0 < c_n \leqq 1, \quad V_{n,\tau_n-1}^2+c_n^2\sigma_{n\tau_n}^2 = 1 \quad \text{if} \quad V_n^2 \geqq 1,$$

$$\tilde{X}_{nk} = X_{nk}[I(k<\tau_n)+c_nI(k=\tau_n)].$$

Since $I(k<\tau_n)$ and $c_nI(k=\tau_n)$ are $\mathcal{F}_{n,k-1}$-measurable random variables, $\{\tilde{X}_{nk}, \mathcal{F}_{nk}\}$ is also MDA. Quantities to be defined analogously to those related to the original MDA are to be distinguished by tilde. Then

$$\tilde{V}_n^2 = \sum_k \tilde{\sigma}_{nk}^2 = \sum_k \sigma_{nk}^2[I(k<\tau_n)+c_n^2I(k=\tau_n)] = \min\{V_n^2, 1\}$$

thus $\tilde{A}_n \leqq A_n$, and since $|\tilde{X}_{nk}| \leqq |X_{nk}|$, therefore $\tilde{L}(n,\varepsilon) \leqq L(n,\varepsilon)$, thus $\tilde{W}_n \leqq W_n$.

Referring to the special case already proved we have

$$\sup_x |\tilde{F}_n(x)-\Phi(x)| \leqq 4(W_n^{1/4}+A_n^{1/3}). \tag{13}$$

In virtue of (1)

$$\begin{aligned}\mathscr{L}(F_n, \tilde{F}_n) &\leqq \left\{E\left[\sum_k (X_{nk}-\tilde{X}_{nk})\right]^2\right\}^{1/3} = \\ &= \left\{E\left[\sum_k (X_{nk}-\tilde{X}_{nk})^2\right]\right\}^{1/3} = \left\{E\left[\sum_k E_{k-1}(X_{nk}-\tilde{X}_{nk})^2\right]\right\}^{1/3} = \\ &= \left\{E\sum_k \left(\sigma_{nk}^2-\tilde{\sigma}_{nk}^2-2E_{k-1}[\tilde{X}_{nk}(X_{nk}-\tilde{X}_{nk})]\right)\right\}^{1/3} = \\ &= \{E[(V_n^2-1)^+-2c_n(1-c_n)\sigma_{n\tau_n}^2]\}^{1/3} \leqq A_n^{1/3}.\end{aligned} \tag{14}$$

Finally, (2), (3), (13) and (14) imply

$$\begin{aligned}\sup_x |F_n(x)-\Phi(x)| &\leqq (1+(2\pi)^{-1/2})\mathscr{L}(F_n, \Phi) \leqq \\ &\leqq (1+(2\pi)^{-1/2})5(W_n^{1/4}+A_n^{1/3}) < 7(W_n^{1/4}+A_n^{1/3})\end{aligned}$$

which was to be proved.

REFERENCES

[1] BILLINGSLEY, P.: The Lindeberg—Lévy theorem for martingales, *Proc. Amer. Math. Soc.* **12** (1961), 788—792.
[2] BROWN, B. M.: Martingale central limit theorems, *Ann. Math. Statist.* **42** (1971), 59—66.
[3] HERTZ, E.: On convergence rates in the central limit theorem, *Ann. Math. Statist.* **40** (1969), 475—479.
[4] HEYDE, C. C. and BROWN, B. M.: On the departure from normality of a certain class of martingales, *Ann. Math. Statist.* **41** (1970), 2161—2165.
[5] LOÈVE, M.: *Probability Theory*, Van Nostrand, Princeton, 1955.

Department of Probability Theory, Eötvös Loránd University,
Múzeum krt. 6—8, H—1088 Budapest, Hungary

(Received July 13, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 419—424.

ON THE PERMEABILITY PROBLEM

by

J. PACH

A set of non-overlapping open domains lying in the plane in a parallel strip of width w, is said to form a *layer*. The *permeability* p of the layer was defined by L. FEJES TÓTH [2], as follows:

$$p = \frac{w}{\inf l} \tag{1}$$

where l is the length of a path connecting one edge of the strip to the other, evading all domains of the layer, and the infimum extends over all paths of this kind.

In [3] the following theorem is proved.

THEOREM 1. *The lower bound of the permeabilities of all layers of squares equals* $\frac{2}{3}$.

We note that the tiles on the wall can form a horizontal layer of squares of permeability arbitrarily close to $\frac{2}{3}$ (Fig. 1). In [1] B. BOLLOBÁS gave another proof of Theorem 1.

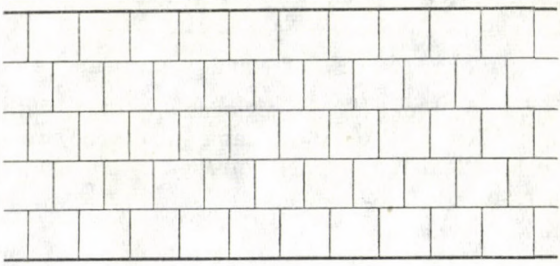


Fig. 1

L. FEJES TÓTH made the conjecture that in any arrangement of non-overlapping open unit squares every two points being at a distance d from each other, can be joined by a path of length $\leqq \frac{3}{2}d+o(d)$ evading all squares. However, the methods used in [3] and [1] do not work for this problem. The aim of the present note is to verify this conjecture. Actually, we prove the following

THEOREM 2. *Let $\mathscr{S}$ be any arrangement of non-overlapping open squares with sides $\leqq 1$. Let y and z be two points outside the squares, satisfying $d(y, z)=d$. Then y and z can be connected by a path of length $\leqq \frac{3}{2}d+4\sqrt{d}+1$ avoiding the members of $\mathscr{S}$. (Here and in what follows $d(y, z)$ denotes the distance of y and z.)*

Before we turn to the proof of Theorem 2, let us show that Theorem 1 can easily be deduced from it. Let L be a layer of squares with width w, having permeability $p=\frac{2}{3}-\varepsilon$, $(\varepsilon>0)$. We may assume without loss of generality that every square belonging to L has sides $\leqq 1$. (Otherwise we apply a similarity transformation.) Let us fit together n congruent replicas of the layer L. The obtained layer L^n has width nw and its permeability is at least $\frac{2}{3}-\varepsilon$. Let f be an arbitrary straight line orthogonal to the edges of L^n. The intersection points of f and the two edges are denoted by y and z. Then, using Theorem 2 for $\mathscr{S}=L^n$, there exists a path of length $\leqq \frac{3}{2}nw+4\sqrt{nw}+1$ connecting y and z. By the definition of the permeability we get

$$\frac{nw}{\frac{3}{2}nw+4\sqrt{nw}+1} \leqq \frac{2}{3}-\varepsilon \tag{2}$$

which is a contradiction if n is large enough. This proves Theorem 1.

Let $\mathscr{S}=\{S_i | i\in I\}$ be an arbitrary arrangement of non-overlapping squares, y and z be two points contained in no S_i. We say that $P_i(y, z)$ is a *simple path avoiding* S_i if $P_i(y, z)$ can be obtained from the straight segment yz in the following way. If yz does not intersect S_i then $P_i(y, z)=yz$. Suppose that yz intersects S_i in the segment u_iv_i. The points u_i, v_i divide the boundary of S_i into two arcs, let $\overset{\frown}{u_iv_i}$ denote the shortest one. (In case of equality $\overset{\frown}{u_iv_i}$ can be chosen arbitrarily.) Now we obtain $P_i(y, z)$ by replacing the segment u_iv_i of S_i with $\overset{\frown}{u_iv_i}$. If we carry out these replacements for all $i\in I$ at the same time then we get a *simple path avoiding* $\mathscr{S}$, denoted by $P(y, z)$. The length of $P_i(y, z)$ and $P(y, z)$ are denoted by $l_i(y, z)$ and $l(y, z)$, resp.

REMARK 1. $l(y, z)\leqq 2d(y, z)$.

To prove Theorem 1 we need a lemma. Let $R=x_1x_2x_3x_4$ be a rectangle and $\mathscr{S}_R=\{S_i | i\in I\}$ be any arrangement of non-overlapping squares inside R. Let $x\in x_1x_2$, then x' denotes the only point of x_3x_4 for which xx' is orthogonal to x_1x_2. Let $l(x)=l(x, x')$ denote the length of $P(x, x')$, that is the length of a simple path from x to x', avoiding $\mathscr{S}_R$.

LEMMA. *Let R, $\mathscr{S}_R$, $l(x)$ be as above. Then the following inequality holds:*

$$\int_{x_1}^{x_2} l(x)\,dx \leqq \frac{3}{2}A(R), \tag{3}$$

where $A(R)$ denotes the area of the rectangle R.

PROOF. Let $d(x_1, x_2)=a$, $d(x_1, x_4)=b$. First we show that

$$\int_{x_1}^{x_2} [l_i(x)-b]\,dx \leqq \frac{A(S_i)}{2} \tag{4}$$

holds for any $i\in I$, where $l_i(x)=l_i(x, x')$ denotes the length of a simple path from x to x' avoiding S_i. Assume that the straight segment xx' intersects S_i in the segment $u_i(x)v_i(x)$. The length of the arc $\overset{\frown}{u_i(x)v_i(x)}$ will be denoted by $\lambda_i(x)$. Obviously, we have

$$\int_{x_1}^{x_2} [l_i(x)-b]\,dx = \int [\lambda_i(x)-d(u_i(x), v_i(x))]\,dx = \int \lambda_i(x)\,dx - A(S_i), \tag{5}$$

where the integral extends over all values x for which xx' intersects S_i. Let $\alpha_i \leqq \frac{\pi}{4}$ denote the angle of a side of S_i and the line x_1x_2 (Fig. 2). An easy computation shows that

$$\int \lambda_i(x)\,dx = r_i \sin\alpha_i(r_i + r_i \operatorname{tg}\alpha_i) + \sqrt{2}r_i \cos\left(\frac{\pi}{4}+\alpha_i\right)\cdot\left[2r_i - \frac{\sqrt{2}r_i\cos\left(\frac{\pi}{4}+\alpha_i\right)}{2\cos\alpha_i}\right] = \tag{6}$$

$$= \frac{3}{2}r_i^2 \frac{1+2\cos^2\alpha_i}{3\cos\alpha_i} \leqq \frac{3}{2}A(S_i).$$

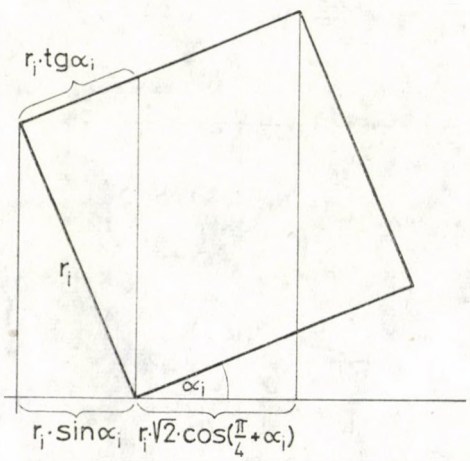


Fig. 2

(Here r_i is the length of a side of S_i.) Combining (5) and (6) we obtain (4). Since

$$\int_{x_1}^{x_2} l(x)\,dx - A(R) = \int_{x_1}^{x_2} [l(x)-b]\,dx = \sum_{i\in I} \int_{x_1}^{x_2} [l_i(x)-b]\,dx, \tag{7}$$

using (4) we get

$$\int_{x_1}^{x_2} l(x)\,dx - A(R) \leqq \sum_{i\in I} \frac{A(S_i)}{2} \leqq \frac{A(R)}{2}$$

which completes the proof of the Lemma.

Now we are in the position to prove THEOREM 2. The proof is indirect. Suppose that $\mathscr{S}$ is an arrangement of open squares with sides $\leqq 1$, y and z are two points outside the squares such that $d(y, z)=d$ and there is no path of length $\leqq \frac{3}{2}d+4\sqrt{d}+1$ from y to z evading the members of $\mathscr{S}$. Clearly, $d>64$ must hold, otherwise taking into consideration Remark 1, y and z can be joined by a simple path of length $\leqq 2d<\frac{3}{2}d+4\sqrt{d}+1$. Let e and f be the straight lines orthogonal to yz, passing through y and z, resp. Furthermore, let $x_1, x_2 \in e$ and $x_3, x_4 \in f$ such that $d(x_1, y)=d(x_2, y)=d(x_3, z)=d(x_4, z)=\frac{\sqrt{d}}{2}+\sqrt{2}$. $\mathscr{S}_R$ denotes the set of all squares belonging to $\mathscr{S}$, which are contained in the rectangle $R=x_1x_2x_3x_4$ (Fig. 3).

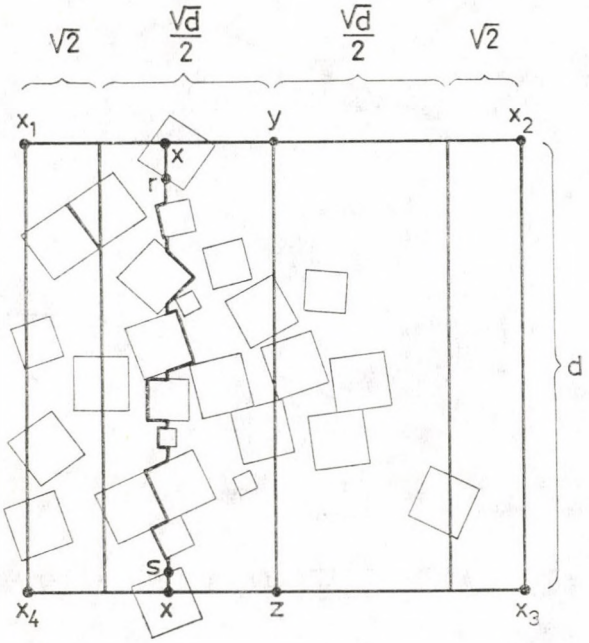


Fig. 3

Let x be an arbitrary point of the segment x_1x_2 and $l(x)$ denote the same as in the Lemma. First we prove that

$$l(x) \geqq \begin{cases} \frac{3}{2}d+2\sqrt{d}+1-4\sqrt{2} & \text{if} \quad d(x, y) \leqq \frac{\sqrt{d}}{2}, \\ d & \text{if} \quad d(x, y) > \frac{\sqrt{d}}{2}. \end{cases} \tag{8}$$

We have to show only the first part of (8), because the second one is an immediate consequence of the definition of $l(x)$. Suppose $d(x, y) \leqq \frac{\sqrt{d}}{2}$ and $P(x, x')$ is a simple path avoiding $\mathscr{S}_R$. Using our assumption that any square belonging to $\mathscr{S}$ has sides $\leqq 1$, we can find two points $r, s \in P(x, x')$ such that $d(r, x), d(s, x') \leqq \sqrt{2}$ and the arc of $P(x, x')$ between r and s does not intersect any member of $\mathscr{S}$. Taking into account Remark 1, we can proceed from y to r on a path of length $\leqq 2d(y, r)$. Then, from r to s we can follow up the arc of $P(x, x')$ and finally, we can reach z using a simple path of length $\leqq 2d(s, z)$ which evades the members of $\mathscr{S}$. The length of this path from y to z is at most

$$l(x)+2[d(x, y)+d(r, x)]+2[d(s, x')+d(x', z)] \leqq$$

$$\leqq l(x)+4\left(\frac{\sqrt{d}}{2}+\sqrt{2}\right).$$

Using our indirect hypothesis, this expression is greater than $\frac{3}{2}d+4\sqrt{d}+1$ which proves the first inequality of (8). From (8)

$$\int_{x_1}^{x_2} l(x)\,dx \geqq \sqrt{d}\left[\frac{3}{2}d+2\sqrt{d}+1-4\sqrt{2}\right]+2\sqrt{2}d \tag{9}$$

follows immediately. On the other hand, by the Lemma we have

$$\int_{x_1}^{x_2} l(x)\,dx \leqq \frac{3}{2}d\left(\sqrt{d}+2\sqrt{2}\right) \tag{10}$$

which leads to a contradiction if $d>64$. This proves Theorem 2.

By the same method as used above, one can prove the following theorems which are generalizations of the corresponding results of [3].

THEOREM 3. *Let $\mathscr{R}$ be any arrangement of non-overlapping open rectangles with sides $\leqq 1$ and with side-ratio $\leqq r$. Let y and z be two points outside the rectangles, satisfying $d(y, z)=d$. Then y and z can be connected by a path of length $\leqq \frac{2+r}{r}d+$ $+O(\sqrt{d})$ evading the members of $\mathscr{R}$.*

THEOREM 4. *Let $\mathscr{R}$ be any arrangement of non-overlapping open rhombi with sides $\leqq 1$ and with acute angles $\geqq \alpha$. Let y and z be two points outside the rhombi,*

satisfying $d(y, z)=d$. *Then* y *and* z *can be connected by a path of length*

a) $\leqq \dfrac{1}{\sin\dfrac{\alpha}{2}} d+O(\sqrt{d})$ *if* $\alpha \leqq \arccos\dfrac{1}{3}$,

b) $\leqq 3\sqrt{\sin^6\dfrac{\alpha}{2}+\cos^6\dfrac{\alpha}{2}}\, d+O(\sqrt{d})$ *if* $\alpha > \arccos\dfrac{1}{3}$,

avoiding the members of $\mathscr{R}$.

Finally, we note that L. Fejes Tóth made the stronger conjecture that instead of $\frac{3}{2}d+4\sqrt{d}+1$ in Theorem 2 one can write $\frac{3}{2}d+\frac{1}{2}$. For a similar open problem concerning circle-layers see [4].

REFERENCES

[1] BOLLOBÁS, B.: Remarks to a paper of L. Fejes Tóth, *Studia Sci. Math. Hung.* **3** (1968), 373—379.

[2] FEJES TÓTH, L.: On the permeability of a circle-layer, *Studia Sci. Math. Hung.* **1** (1966), 5—10.

[3] FEJES TÓTH, L.: On the permeability of a layer of parallelograms, *Studia Sci. Math. Hung.* **3** (1968), 195—200.

[4] FEJES TÓTH, L.: Research problem № 24, *Periodica Math. Hung.* **9** (1978), 173—174.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received August 25, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 425—428.

ON THE TRIANGULARIZABILITY OF PLANAR DIFFERENTIAL SYSTEMS WITHOUT CRITICAL POINTS

by

G. TÓTH

1. Preliminary Concepts and Results

In this note we continue the examinations of the paper [3] to get further results about the triangularizability of the two-dimensional differential system

$$\dot{x}_1 = f(x_2), \quad \dot{x}_2 = g(x_1, x_2) \tag{1}$$

where the functions f and g are continuous and

$$f^2(x_2) + g^2(x_1, x_2) > 0 \quad \text{for every} \quad (x_1, x_2) \in \mathbf{R}^2$$

i.e. the differential system (1) has no critical points. We shall suppose moreover that the functions f and g are n-times continuously differentiable for some $n \in \mathbf{N}$.

We call the differential system (1) triangularizable (in diffeomorphic sense), if there exists a C^n-diffeomorphism $U=(u_1, u_2)$ of $\mathbf{R}^2$ which transforms the system (1) to a triangularized form:

$$\dot{u}_1 = a(u_1), \quad \dot{u}_2 = b(u_1, u_2)$$

(see [4]).

Throughout this note we shall use the definitions and notions of the above mentioned papers [3] and [4] and we refer also to the book [1].

In [3] we gave a sufficient condition for the triangularizability of differential system (1) in homeomorphic case where our examinations were based upon the total topological description of unstable triangularizable dynamical systems on the plane.

In [4] we proved the validity of this description in diffeomorphic case, too, therefore we may state the diffeomorphic version of Theorem 2 of [3] as follows:

THEOREM 1. *The differential system* (1) *(with the above conditions) is triangularizable provided that the following assumptions are valid:*

A. *The function f has at most two zeros;*

B. $g(x_1, x_2) \neq 0$ *for every* $|x_2| > K$, *where* K *is a sufficiently large number.*

2. New Results

Our aim now is to give two generalizations of Theorem 1. To do this, first of all we mention that in the proof of Theorem 1 Condition B ensured the parallelizability of the dynamical system induced by (1) outside a sufficiently wide

strip parallel with the first axis. So, if we introduce the property:

C_L. The dynamical system
$\varphi|\varphi(\mathbf{R}^2-\mathbf{R}\times(-L, L);\ \mathbf{R})$
is parallelizable for some $L>0$,

then our problem is reduced to find analytic assumptions ensuring the condition C_L.

First we give a simple generalization as follows:

THEOREM 2. *If Condition A holds and for each finite interval $I\subset\mathbf{R}$ the set*

$$Z_g\cap(I\times\mathbf{R})$$

is bounded, where $Z_g=\{(x_1, x_2)\in\mathbf{R}^2:\ g(x_1, x_2)=0\}$, then the differential system (1) *is triangularizable.*

PROOF. We must show that our assumption implies C_L for some $L>0$. Let $L>\max(|\mu|, |\nu|)$ where μ and ν are the zeros of f, if any. Without loss of generality we may restrict ourselves to the upper half-plane $\mathbf{R}\times(L, \infty)$ since otherwise we perform a reflection to the first axis. By similar reasons we may suppose that $f>0$ on (L, ∞) since otherwise we can introduce a new independent variable $\hat{t}=-t\in\mathbf{R}$.

If φ is not parallelizable on the upper half-plane $\mathbf{R}\times(L, \infty)$ then — because of the fundamental theorem of topological dynamics (see [1]) — there is an improper saddle point on it which has limit half-trajectories $\varphi(p;\ \mathbf{R}_+)$, $\varphi(q;\ \mathbf{R}_-)\subset$ $\subset\mathbf{R}\times(L, \infty)$ with $p_1<q_1$. Since the function f does not change its sign on (L, ∞) the functions

$$\varphi_1(p, \cdot),\quad \varphi_1(q, \cdot):\ \mathbf{R}\to\mathbf{R}$$

are strictly increasing and so $\varphi_1(p, t)<\varphi_1(q, s)$ for each $t\in\mathbf{R}_+$ and $s\in\mathbf{R}_-$. Therefore

$$p^*=\lim_{t\to\infty}\varphi_1(p, t)\leqq\lim_{t\to-\infty}\varphi_1(q, t)=q^*$$

and

$$\lim_{t\to\infty}\varphi_2(p, t)=\lim_{t\to-\infty}\varphi_2(q, t)=\infty \tag{2}$$

(see Figure 1). Let $I=[p_1, q_1]$ and consider the restrictions:

$$\varphi_2(r^n, \cdot)|\{t\in\mathbf{R}:\ \varphi_1(r^n, t)\in I\},$$

where $r_1^n=p_1$ with $r_2^n\nearrow p_2$. Since $\varphi(q;\ \mathbf{R})$ is a limit trajectory there exists a divergent sequence $(t_n)_{n\in\mathbf{N}}\subset\mathbf{R}_+$ so that $\varphi_1(p, t_n)=q_1$ with $\varphi_2(p, t_n)\nearrow q_2$ for each sufficiently large $n\geqq N\in\mathbf{N}$. Because of the relations (2) there exists a maximum point z^n on $\varphi_2(r^n, \cdot)|\{t\in\mathbf{R}:\ \varphi_1(r^n, t)\in I\}$ from an index $n\geqq N_0\in\mathbf{N}$ such that $z_1^n\in I$. But $z^n\in Z_g$ and $z_2^n\to\infty$ which is contradiction to the boundedness of $Z_g\cap(I\times\mathbf{R})$.

The reasonings of the previous proof show that the situation is in close relation with the continuations of solutions of ordinary differential equations. This idea leads us to the following:

THEOREM 3. *If Condition A holds and for each finite interval $I\subset\mathbf{R}$ there exists an integrable function*

$$M_I:\ \mathbf{R}-(-L, L)\to\mathbf{R}_+$$

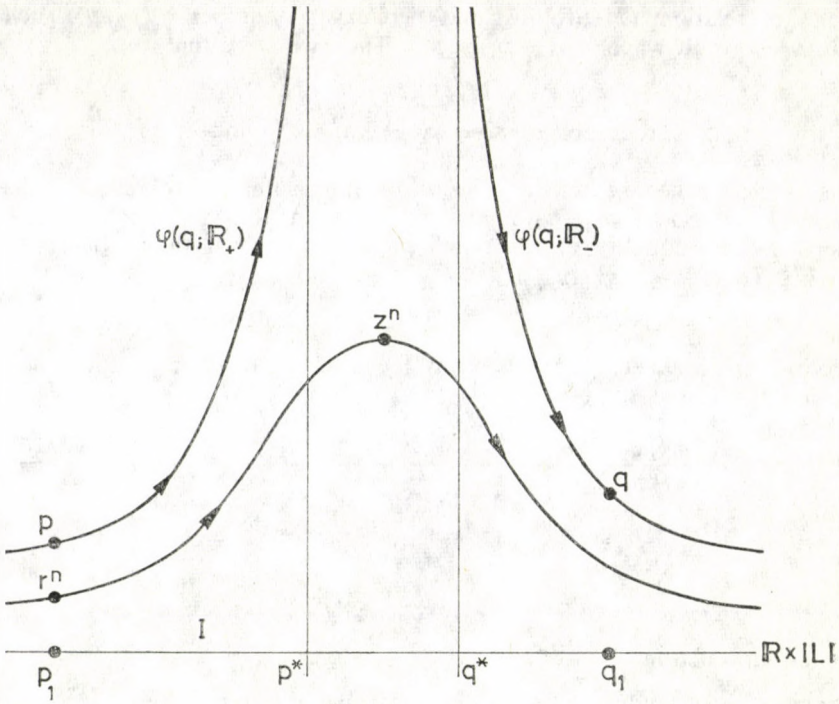


Fig. 1

($L > \max(|\mu|, |\nu|)$ *as above*) *for which*

$$\int_L^{\infty} \frac{1}{M_I} = \int_{-\infty}^{-L} \frac{1}{M_I} = \infty$$

and

(3) $$\max_{s,t \in I} |g(s, x_2) - g(t, x_2)| \leqq M_I(x_2)|f(x_2)| \quad \textit{for each} \quad |x_2| \geqq L,$$

then the differential system (1) *is triangularizable.*

PROOF. To prove the validity of C_L we proceed again the argumentation as in the proof of Theorem 2 and we have two trajectories $\varphi(p; \mathbf{R})$ and $\varphi(q; \mathbf{R})$ satisfying (2).

Now let $I=[p_1, q_1]$ again and let h: $[p_1, p^*) \to \mathbf{R}$ be defined by

$$h(\varphi_1(p, t)) = \min_{\tau \geqq t} \varphi_2(p, \tau), \quad t \geqq 0.$$

Obviously, h is nondecreasing and $\lim_{s \to p^*} h(s) = \infty$.

From the relations (2) it follows that there exists a number $x\in[p_1, p^*)$ such that $\varphi_2(p, t)\geqq q_2$ is valid when $\varphi_1(p, t)\in[x, p^*)$. Then we state that

$$M_I\circ h \geqq \dot{h}$$

on the interval $[x, p^*)$ except possibly a countable point-set on which $\dot{h}$ does not exist.

This is clear when $\dot{h}$ vanishes. Therefore it is sufficient to restrict ourselves to the case $\dot{h}>0$. So let $\varphi(p, t_0)$ be an arbitrary point for which $\varphi_1(p, t_0)\geqq x$ and let us prove the inequality at the point $\varphi_1(p, t_0)$ provided that $\dot{h}(\varphi_1(p, t_0))>0$. Clearly $h(\varphi_1(p, t_0))=\varphi_2(p, t_0)$. If we set

$$s_0 = \sup pr_1(\mathbf{R}\times\{\varphi_2(p, t_0)\}\cap\varphi(q; \mathbf{R}))\in I$$

(the section in brackets is not empty because $\varphi_2(p, t)\geqq q_2$ and (2)) then $g(s_0, \varphi_2(p, t_0))\leqq 0$ and therefore

$$\dot{h}(\varphi_1(p, t_0)) = \frac{g(\varphi_1(p, t_0), \varphi_2(p, t_0))}{f(\varphi_2(p, t_0))} \leqq \frac{g(\varphi_1(p, t_0), \varphi_2(p, t_0))-g(s_0, \varphi_2(p, t_0))}{f(\varphi_2(p, t_0))} \leqq$$

$$\leqq M_I(\varphi_2(p, t_0)) = (M_I\circ h)(\varphi_1(p, t_0)).$$

So

$$p^*-x \geqq \int_x^{p^*} \frac{\dot{h}}{M_I\circ h} = \int_{h(x)}^{\infty} \frac{1}{M_I}$$

which is a contradiction.

REMARK. If we study the triangularization problem of unstable second order differential equation

$$\ddot{x}+g(x, \dot{x}) = 0$$

we see that Condition A is automatically satisfied because

$$f(x_2) = -x_2, \quad x_2\in\mathbf{R}.$$

In this way we can get some results to the second order equations, too, as consequences of Theorem 2 and 3.

REFERENCES

[1] NIEMITZKII, V. V. and STEPANOV, V. V.: *Qualitative Theory of Differential Equations*, Gos. Izd. Techn.-Theor. Lit., Moscow—Leningrad (1969) (in Russian).

[2] SAMOVOL, V. S.: On the Reduction of Dynamical Systems into Triangular Form, *Diff. Ur.* (1969), № 6, 1076—1182 (in Russian).

[3] TÓTH, G.: On the Triangularizability of Planar Orthogonal Differential Equations, *Period. Math. Hung.* **8** (1977), 243—251.

[4] TÓTH, G.: On the Global Triangularizability of Planar Differentiable Dynamical Systems, *Studia Sci. Math. Hungar.* **11** (1976), 211—228.

Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13—15, Hungary 1053

(Received September 12, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 429—443

ON THE STRONG SUMMABILITY BY THE (C, α)-MEANS OF FOURIER SERIES

by

V. TOTIK

1. Let f be a 2π-periodic integrable function with Fourier series

$$S[f] = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \sin \nu x) = \sum_{\nu=0}^{\infty} A_\nu(x),$$

and let

$$\tilde{S}[f] = \sum_{\nu=1}^{\infty} (a_\nu \sin \nu x - b_\nu \cos \nu x)$$

be the conjugate series to $S[f]$. Let us denote by $\sigma_k^\alpha(f;x) = \sigma_k^\alpha(x)$ and $\tilde{\sigma}_k^\alpha(f;x) = \tilde{\sigma}_k^\alpha(x)$ the (C, α)-means of $S[f]$, and $\tilde{S}[f]$, respectively, thus, e.g.,

$$\sigma_k^\alpha(x) = \frac{1}{A_k^\alpha} \sum_{\nu=0}^{k} A_{k-\nu}^\alpha A_\nu(x) \qquad (\alpha > -1,\ k = 0, 1, \ldots),$$

where $A_k^\alpha = \binom{k+\alpha}{k}$. Specially, $s_k(x) = \sigma_k^0(x)$ and $\tilde{s}_k(x) = \tilde{\sigma}_k^0(x)$ are the partial sums of $S[f]$ and $\tilde{S}[f]$.

By the well-known result of ZYGMUND [6] we have for any $p > 0$

$$\frac{1}{n+1} \sum_{k=0}^{n} |s_k(x) - f(x)|^p = o_x(1) \tag{1.1}$$

and

$$\frac{1}{n+1} \sum_{k=0}^{n} |\tilde{s}_k(x) - \tilde{f}(x)|^p = o_x(1) \tag{1.2}$$

almost everywhere (a.e.). Zygmund proved also that (1.1) is valid at every point of continuity.

LEINDLER [4] raised the problem: Can we replace the partial sums s_k and $\tilde{s}_k$ in (1.1) and (1.2) by (C, α)-means with negative α? In the present paper we deal with these questions. Our main results are[1]

THEOREM 1. *Let* $0 \geqq \alpha > -\frac{1}{2}$, $p > 0$ *and* $p\alpha > -1$. *If* $f \in L^{\frac{1}{1+\alpha}}$ *then we have*

$$\frac{1}{n+1} \sum_{k=0}^{n} |\sigma_k^\alpha(x) - f(x)|^p = o_x(1) \tag{1.3}$$

[1] It has turned out that G. SUNOUCHY proved also this result (see *Tôhoku Math. J.* **6** (1954), 220—225). Our proof, using a result from [7] is much shorter.

and

$$\frac{1}{n+1}\sum_{k=0}^{n}|\tilde{\sigma}_k^{\alpha}(x)-\tilde{f}(x)|^p = o_x(1) \tag{1.4}$$

almost everywhere.

From this, using the method of [2, Theorem 8] we get easily the

COROLLARY. *With the assumptions of Theorem 1 we have for* $\gamma > -p\alpha$ *and for* $\beta > 0$

$$\lim_{n\to\infty}\frac{1}{A_n^{\gamma}}\sum_{k=0}^{n}A_{n-k}^{\gamma-1}|\sigma_k^{\alpha}(x)-f(x)|^p = 0 \quad \text{(a.e.)}$$

and

$$\lim_{n\to\infty}\frac{1}{(n+1)^{\beta}}\sum_{k=0}^{n}(k+1)^{\beta-1}|\sigma_k^{\alpha}(x)-f(x)|^p = 0 \quad \text{(a.e.)},$$

respectively.

Theorem 2 will show that the assumption $f\in L^{\frac{1}{1+\alpha}}$ cannot be reduced in Theorem 1.

THEOREM 2. *If* $0 > \alpha > -1$ *and* $p > 0$, *then there are functions* f *such that* $f\in L^{\beta}$ *for every* $\beta < \frac{1}{1+\alpha}$ *but*

$$\limsup_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}|\sigma_k^{\alpha}(x)-f(x)|^p = \infty \tag{1.5}$$

almost everywhere.

The existence of such an f for which $f\in L^{\beta}$ for any $\beta < \frac{1}{1+\alpha}$ and

$$\limsup_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}|\tilde{\sigma}_k^{\alpha}(x)-\tilde{f}(x)|^p = \infty \quad \text{(a.e.)}$$

could be proved similarly.

We have mentioned that (1.1) is valid at every point of continuity. As regards (1.3) and (1.4) the corresponding statement is

THEOREM 3. *Let us suppose that the assumptions of Theorem 1 are fulfilled. If* f *is continuous at* x *then* (1.3) *is true. If, in addition, there exists* $\tilde{f}(x)$, *then* (1.4) *is valid, too.*

Furthermore, there is a function f *such that* $f\in L^{\beta}$ *for every* $\beta < \frac{1}{1+\alpha}$, f *is continuous at the point* π, *but*

$$\lim_{k\to\infty}|\sigma_k^{\alpha}(\pi)| = \infty. \tag{1.6}$$

COROLLARY. *Under the assumptions of Theorem 1* (1.3) *is true uniformly in* x, *if* f *is a continuous function.*

The following theorem shows that the condition $p\alpha > -1$ was essential in Theorem 1.

THEOREM 4. *If* $0>\alpha>-1$ *and* $p\alpha<-1$ *then there are continuous functions for which* (1.5) *is true almost everywhere.*

2. To prove our theorems we require some lemmas.

LEMMA 1. *Let* $\alpha>-\frac{1}{2}$ *and*

$$F(z)=\frac{1}{2}a_0+\sum_{\nu=1}^{\infty}(a_\nu-ib_\nu)z^\nu \tag{2.1}$$

which for $z=e^{ix}$ *reduces to* $S[f]+i\tilde{S}[f]$. *If* $\tau_k^\alpha(x)$ *is the k-th* (C,α)*-mean of* (2.1) *for* $z=e^{ix}$, *then we have for* $-\frac{1}{\alpha}>p\geqq 2$

$$\left\{\sum_{k=0}^{\infty}k^{(1+\alpha)p}|\tau_k^\alpha(x)-\tau_k^{\alpha+1}(x)|^p\varrho^{kp}\right\}^{\frac{1}{p}}\leqq K\left\{\int_{-\pi}^{\pi}\frac{|F'(ze^{ix})|^q}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi\right\}^{\frac{1}{q}}$$

$$\left(q=\frac{p}{p-1},\ z=\varrho e^{i\varphi},\ \varrho<1\right).$$

The proof can be found in [3, pp. 158—159].

LEMMA 2. *Let us suppose that* $\gamma>1$, f *is non-negative and vanishes in a perfect set* $E\subseteqq(0,2\pi)$. *If* $U(\varrho,\varphi)$ *is the Poisson-integral of* f *then*

$$\int_{-\pi}^{\pi}\frac{U^\gamma(\varrho,\varphi+x)}{|1-\varrho e^{i\varphi}|^\gamma}\,d\varphi=o_x\left((1-\varrho)^{1-\gamma}\right)\qquad(\varrho\to 1-0)$$

for almost every $x\in E$.

The proof is given in [7, II, pp. 184—188].

LEMMA 3. *If* $\{\varphi_k\}_{k=1}^{\infty}$ *is an orthogonal sequence on* $[a,b]$ *with* $|\varphi_k|\leqq 1$ $(k=1,2,\ldots)$, *then*

$$\lim_{n\to\infty}\frac{1}{2^n}\sum_{k=1}^{2^n}\varphi_k(t)=0$$

a.e. in $[a,b]$.

This is an easy consequence of [1, 1.5.1.].

3. PROOF OF THEOREM 1. We may suppose $-\frac{1}{2}<\alpha<0$. But then $f\in L^{\frac{1}{1+\alpha}}$ implies $\tilde{f}\in L^{\frac{1}{1+\alpha}}$ (see [7, I, p. 253.]), and so we have to prove only (1.3).

It is enough to prove that

$$\frac{1}{n}\sum_{k=n+1}^{2n}|\sigma_k^\alpha(x)-f(x)|^p=o_x(1) \tag{3.1}$$

a.e. for any $p<-\frac{1}{\alpha}$, namely (3.1) implies for $2^{m-1}<n\leqq 2^m$

$$\frac{1}{n+1}\sum_{k=0}^{n}|\sigma_k^\alpha(x)-f(x)|^p \leqq \frac{1}{n+1}\left(O_x(1)+\sum_{\nu=0}^{m-1}\sum_{k=2^\nu+1}^{2^{\nu+1}}|\sigma_k^\alpha(x)-f(x)|^p\right)=$$

$$=o_x\left(\frac{1}{n+1}\sum_{\nu=0}^{m-1}2^\nu\right)=o_x(1)\quad\text{(a.e.)}.$$

First we show (3.1) when $f\in L^r$ for some $2\geqq r>\frac{1}{1+\alpha}$.

By Hölder inequality we may suppose $-\frac{1}{\alpha}>p\geqq\frac{r}{r-1}$ $\left(\text{namely } r>\frac{1}{1+\alpha}\right.$ implies $\left.\frac{r}{r-1}\alpha>-1\right)$.

With the notation $\varphi_x(t)=\frac{1}{2}(f(x+t)+f(x-t)-2f(x))$ we have

$$\sigma_k^\alpha(x)-f(x)=\frac{2}{\pi}\int_0^\pi\varphi_x(t)K_k^\alpha(t)\,dt$$

where $K_k^\alpha(t)$ denotes the (C,α)-kernel (see [7, I, p. 94.]). This gives

$$\left\{\frac{1}{n}\sum_{k=n+1}^{2n}|\sigma_k^\alpha(x)-f(x)|^p\right\}^{\frac{1}{p}}\leqq\left\{\frac{1}{n}\sum_{k=n+1}^{2n}\left|\frac{2}{\pi}\int_0^{1/n}\varphi_x(t)K_k^\alpha(t)\,dt\right|^p\right\}^{\frac{1}{p}}+$$

$$+\left\{\frac{1}{n}\sum_{k=n+1}^{2n}\left|\frac{2}{\pi}\int_{1/n}^{\pi}\varphi_x(t)K_k^\alpha(t)\,dt\right|^p\right\}^{\frac{1}{p}}=I_1(x)+I_2(x).$$

If x is a Lebesgue point of f then by $|K_k^\alpha(t)|\leqq 2k$ ([7, I, p. 95]) we have

$$\left|\int_0^{1/n}\varphi_x(t)K_k^\alpha(t)\,dt\right|\leqq 2k\int_0^{1/n}|\varphi_x(t)|\,dt=2ko_x\left(\frac{1}{n}\right)=o_x(1)\quad(k\leqq 2n)$$

and so $I_1(x)=o_x(1)$.

To estimate $I_2(x)$ we use the formula

$$K_k^\alpha(t)=\frac{1}{A_k^\alpha}\frac{\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}+\frac{2\theta(t)\alpha}{k\left(2\sin\frac{t}{2}\right)^2}\quad\left(|\theta|\leqq 1,\ \frac{1}{k}\leqq t\leqq\pi\right)$$

(see [7, I, p. 95]). Writing this in the place of $K_k^\alpha(t)$ we obtain

$$I_2(x)\leqq\left\{\frac{1}{n}\sum_{k=n+1}^{2n}\left|\frac{2}{\pi}\frac{1}{A_k^\alpha}\int_{1/n}^{\pi}\frac{\varphi_x(t)\cos\left[\frac{1+\alpha}{2}t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\sin kt\,dt\right|^p\right\}^{\frac{1}{p}}+$$

$$+\left\{\frac{1}{n}\sum_{k=n+1}^{2n}\left|\frac{2}{\pi}\frac{1}{A_k^\alpha}\int_{1/n}^{\pi}\frac{\varphi_x(t)\sin\left[\frac{1+\alpha}{2}t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\cos kt\,dt\right|^p\right\}^{\frac{1}{p}}+$$

$$+\left\{\frac{1}{n}\sum_{k=n+1}^{2n}\left|\frac{2}{\pi}\int_{1/n}^{\pi}\frac{\varphi_x(t)2\theta(t)\alpha}{k\left(2\sin\frac{t}{2}\right)^2}dt\right|^p\right\}^{\frac{1}{p}}=I_{21}(x)+I_{22}(x)+I_{23}(x).$$

Let $\Phi_{x,r}(h)=\int_0^h|\varphi_x(t)|^r\,dt$. At every Lebesgue point $\Phi_{x,1}(h)=o_x(h)$, and so we get by partial integration

$$\left|\int_{1/n}^{\pi}\frac{\varphi_x(t)2\theta(t)\alpha}{k\left(2\sin\frac{t}{2}\right)^2}dt\right|\leqq\frac{8}{n}\int_{1/n}^{\pi}\frac{|\varphi_x(t)|}{t^2}dt=\frac{8}{n}\frac{\Phi_{x,1}(t)}{t^2}\Bigg|_{\frac{1}{n}}^{\pi}+\frac{16}{n}\int_{1/n}^{\pi}\frac{\Phi_{x,1}(t)}{t^3}dt=o_x(1),$$

and together with this $I_{23}(x)=o_x(1)$.

The estimations of $I_{21}(x)$ and $I_{22}(x)$ are similar, so we shall deal only with the former.

Let $q=\frac{p}{p-1}$. Because of $p\geqq\frac{r}{r-1}$ and $r\leqq2$ we have $p\geqq2$ and $q\leqq r$ by which $f\in L^q$. Thus we can apply the Hausdorff—Young theorem ([7, II, p. 101]) to the function

$$h_x(t)=\begin{cases}\dfrac{\varphi_x(t)\sin\left[\frac{1+\alpha}{2}t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}} & \text{if}\quad t\in\left(\frac{1}{n};\,\pi\right]\\ 0 & \text{otherwise,}\end{cases}$$

and we obtain

$$I_{21}(x)=O\left(n^{-\alpha-\frac{1}{p}}\left(\int_{1/n}^{\pi}\frac{|\varphi_x(t)|^q}{t^{(1+\alpha)q}}dt\right)^{\frac{1}{q}}\right),$$

since $c_\alpha k^\alpha\leqq A_k^\alpha\leqq b_\alpha k^\alpha$ $(\alpha>-1,\ c_\alpha>0,\ k=1,2,\ldots)$.

$f\in L^q$ implies that $\Phi_{x,q}(h)=o_x(h)$ a.e. ([7, I, p. 65]), and so

$$I_{21}(x)=O\left(n^{-\alpha-\frac{1}{p}}\left(\frac{\Phi_{x,q}(t)}{t^{(1+\alpha)q}}\Bigg|_{\frac{1}{n}}^{\pi}\right)^{\frac{1}{q}}+n^{-\alpha-\frac{1}{p}}\left(\int_{1/n}^{\pi}\frac{\Phi_{x,q}(t)}{t^{(1+\alpha)q+1}}dt\right)^{\frac{1}{q}}\right)=$$

$$=O\left(n^{-\alpha-\frac{1}{p}}+n^{-\alpha-\frac{1}{p}+1+\alpha}\left(\Phi_{x,q}\left(\frac{1}{n}\right)\right)^{\frac{1}{q}}\right)+o_x\left(n^{-\alpha-\frac{1}{p}}\left(\int_{1/n}^{\pi}\frac{dt}{t^{(1+\alpha)q}}\right)^{\frac{1}{q}}\right)=o_x(1)$$

a.e., where we used that $\alpha+\frac{1}{p}>0$ and $(1+\alpha)q>1$.

Collecting the above estimations we obtain (3.1).

After that we turn to the case $f\in L^{\frac{1}{1+\alpha}}$.

We may suppose that $f\geqq 0$. Let f_1 be a bounded function coinciding with f on a perfect set $E\subseteq(0;\ 2\pi)$ and equal to 0 elsewhere. By the above case $\left(f\in L^r \text{ for some } r>\frac{1}{1+\alpha}\right)$, (3.1) is true for f_1 a.e., and so we only have to prove that it is also true for $f-f_1$ a.e. in E, since the measure of E may be arbitrarily close to 2π.

Thus we have reduced the problem to proving that if an $f\in L^{\frac{1}{1+\alpha}}$ is non-negative and vanishes in a perfect set E, then (3.1) is true a.e. in E.

Because of $\sigma_k^{\alpha+1}\to f(x)$ $(k\to\infty)$ a.e., it is enough to show that for $p<-\frac{1}{\alpha}$

$$\frac{1}{n}\sum_{k=n+1}^{2n}|\sigma_k^{\alpha+1}-\sigma_k^{\alpha}(x)|^p=o_x(1) \tag{3.2}$$

a.e. in E.

Since $\alpha>-\frac{1}{2}$ and $p\alpha>-1$, we can give a number p' having the properties $p'\geqq\max(2,p)$ and $p'\alpha>-1$, and so, by Hölder inequality, we can suppose in the proof of (3.2) that $2\leqq p<-\frac{1}{\alpha}$.

Using the notations of Lemma 1 we have

$$\left\{\sum_{k=0}^{\infty}k^{(1+\alpha)p}|\tau_k^{\alpha}(x)-\tau_k^{\alpha+1}(x)|^p\varrho^{kp}\right\}^{\frac{1}{p}}\leqq K\left\{\int_{-\pi}^{\pi}\frac{|F'(ze^{ix})|^q}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi\right\}^{\frac{1}{q}}\quad(z=\varrho e^{i\varphi}). \tag{3.3}$$

Let $\delta=1-\varrho$. If we show that

$$\int_{-\pi}^{\pi}\frac{|F'(ze^{i\varphi})|^q}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi=o_x(\delta^{1-(2+\alpha)q}) \tag{3.4}$$

a.e. in E, then setting $\delta=\frac{1}{n}$ and using the inequality

$$\left(1-\frac{1}{n}\right)^k\geqq\left(1-\frac{1}{n}\right)^{2n}>e^{-2}\quad(k\leqq 2n)$$

we obtain (3.2) from (3.3) (take into account also that $\sigma_k^{\alpha}(x)-\sigma_k^{\alpha+1}(x)$ is the real part of $\tau_k^{\alpha}(x)-\tau_k^{\alpha+1}(x)$).

Finally, if $f(\varrho,\varphi)$ is the Poisson integral of f then

$$|F'(ze^{ix})|\leqq\frac{2}{\delta}f(\varrho,\varphi+x)$$

(see [7, I, p. 258]), thus to prove (3.4) it is enough to show that

$$I_x(\varrho)=\int_{-\pi}^{\pi}\frac{f^q(\varrho,\varphi+x)}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi=o_x(\delta^{1-(1+\alpha)q})$$

a.e. in E.

Let $P(\varrho, \varphi)$ and $U(\varrho, \varphi)$ be the Poisson kernel and the Poisson integral of $f^{\frac{1}{1+\alpha}}$, respectively. By Jensen inequality we have

$$f^{\frac{1}{1+\alpha}}(\varrho, \varphi) = \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(u)P(\varrho, \varphi-u)\,du\right)^{\frac{1}{1+\alpha}} \leqq \frac{1}{\pi}\int_{-\pi}^{\pi} (f(u))^{\frac{1}{1+\alpha}} P(\varrho, \varphi-u)\,du = U(\varrho, \varphi)$$

and so by Lemma 2

$$I_x(\varrho) = \int_{-\pi}^{\pi} \frac{(f^{\frac{1}{1+\alpha}}(\varrho, \varphi+x))^{(1+\alpha)q}}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi \leqq \int_{-\pi}^{\pi} \frac{U^{(1+\alpha)q}(\varrho, \varphi+x)}{|1-\varrho e^{i\varphi}|^{(1+\alpha)q}}\,d\varphi = o_x(\delta^{1-(1+\alpha)q})$$

a.e. in E, and thus the proof of Theorem 1 is completed.

PROOF OF THEOREM 2. We shall give a function which is in L^β for every $\beta < \frac{1}{1+\alpha}$ and for which (1.5) is true almost everywhere in $\left(-\frac{\pi}{2}; 0\right)$. At the end of this proof we shall sketch how could be constructed a function satisfying (1.5) a.e.

For the 2π-periodic function f let

$$f(x) = \begin{cases} x^{-(1+\alpha)} \log\frac{1}{x} & \text{if} \quad 0 < x \leqq 1, \\ 0 & \text{if} \quad x \in (-\pi; 0] \cup (1; \pi]. \end{cases}$$

It is clear that $f \in L^\beta$ for every $\beta < \frac{1}{1+\alpha}$.

Since f vanishes in $(-\pi; 0)$, it is enough to show that if

$$C_k(x) = \frac{1}{A_k^\alpha}\int_{1/k}^{\pi} \frac{f(x+t)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)t - \frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\,dt$$

then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=n+1}^{2n} |C_k(x)|^p = \infty \tag{3.5}$$

a.e. in $\left(-\frac{\pi}{2}; 0\right)$ (see the proof of Theorem 1).

Let ε be a fixed positive number which we shall choose later, and let $x \in \left(-\frac{\pi}{2}; 0\right)$. For large enough k we have

$$C_k(x) = \frac{1}{A_k^\alpha}\int_0^1 \frac{f(\tau)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)(\tau-x) - \frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{\tau-x}{2}\right)^{1+\alpha}}\,d\tau,$$

and so with the notations

$$a_k(x) \equiv \left(k+\frac{1}{2}+\frac{\alpha}{2}\right)(-x)-\frac{\pi\alpha}{2} \pmod{2\pi} \quad (a_k(x)\in(0;\,2\pi])$$

and

$$b_k(x)=\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)^{-1}(2\pi-a_k(x)) \quad (k=1,\,2,\,\ldots)$$

we obtain

$$C_k(x)=\frac{1}{A_k^\alpha}\left(\int_0^{b_k(x)}+\int_{b_k(x)}^{1}\right)\frac{f(\tau)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)\tau+a_k(x)\right]}{\left(2\sin\frac{\tau-x}{2}\right)^{1+\alpha}}\,d\tau=I_{k1}(x)+I_{k2}(x).$$

Now let us suppose that

$$a_k(x)\in(0;\,\varepsilon). \tag{3.6}$$

By the second mean value theorem we get $I_{k2}(x)\geqq 0$, thus if we show that with the suitable choice of ε it is attainable that (3.6) should imply $I_{k1}(x)\geqq c(x)\log k$ with a $c(x)>0$ constant independent from k, then

$$C_k(x)\geqq c(x)\log k \tag{3.7}$$

will be satisfied, too. But

$$I_{k1}(x)\geqq\left[2\sin\frac{1}{2}\left(\frac{\pi-a_k(x)}{k+\frac{1}{2}+\frac{\alpha}{2}}-x\right)\right]^{-(1+\alpha)}\frac{1}{A_k^\alpha}\int_0^{b_k(x)}f(\tau)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)\tau+a_k(x)\right]d\tau\geqq$$

$$\geqq\frac{1}{A_k^\alpha}\int_0^{b_k(x)}f(\tau)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)\tau+a_k(x)\right]d\tau. \tag{3.8}$$

Here

$$\int_0^{b_k(x)}f(\tau)\sin\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)\tau+a_k(x)\right]d\tau= \tag{3.9}$$

$$=\frac{1}{k+\frac{1}{2}+\frac{\alpha}{2}}\int_0^{2\pi-a_k(x)}\left(\frac{u}{k+\frac{1}{2}+\frac{\alpha}{2}}\right)^{-(1+\alpha)}\left(\log\frac{k+\frac{1}{2}+\frac{\alpha}{2}}{u}\right)\sin\left(u+a_k(x)\right)du=$$

$$=\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)^{\alpha}\left(\log\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)\right)g(a_k(x))+\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)^{\alpha}h(a_k(x))$$

where

$$g(y)=\int_0^{2\pi-y}u^{-(1+\alpha)}\sin(u+y)\,du \quad\text{and}\quad h(y)=\int_0^{2\pi-y}u^{-(1+\alpha)}\left(\log\frac{1}{u}\right)\sin(u+y)\,du.$$

Clearly, $g(y)$ and $h(y)$ are continuous functions and $g(0)>0, h(0)>0$, thus we can give $\varepsilon>0$ so that $g(a_k(x))>\frac{1}{2}g(0)$ and $h(a_k(x))>0$ should follow from (3.6).

If we choose ε in accordance with the previous condition then (3.8) and (3.9) give (3.7) for every k satisfying (3.6). Thus we have

$$\frac{1}{n}\sum_{k=n+1}^{2n}|C_k(x)|^p \geqq (c(x))^p(\log n)^p\frac{1}{n}\sum_{\substack{n<k\leqq 2n\\ a_k(x)\in(0;\,\varepsilon)}}1. \tag{3.10}$$

Let now $x=\vartheta\pi$ $\left(x\in\left(-\frac{\pi}{2};\,0\right)\right)$ where ϑ is irrational. It is known (see [5, p. 71]) that the sequence $\{k\vartheta\}_{k=1}^{\infty}$ is mod 1 of uniform distribution in $(0;\,1]$, and this implies that $\{a_k(x)\}_{k=1}^{\infty}$ is of uniform distribution in $(0;\,2\pi]$. Thus to every $\eta>0$ there exists N_η such that for $n>N_\eta$ we have

$$\left|\left(\frac{1}{n}\sum_{\substack{1\leqq k\leqq n\\ a_k(x)\in(0;\,\varepsilon)}}1\right)-\frac{\varepsilon}{2\pi}\right|<\eta.$$

By this we have for $n>N_{\frac{\varepsilon}{12\pi}}$

$$\frac{1}{n}\sum_{\substack{n<k\leqq 2n\\ a_k(x)\in(0;\,\varepsilon)}}1>\frac{1}{n}\left[\left(\frac{\varepsilon}{2\pi}-\frac{\varepsilon}{12\pi}\right)2n-\left(\frac{\varepsilon}{2\pi}+\frac{\varepsilon}{12\pi}\right)n\right]=\frac{\varepsilon}{4\pi}$$

and so (3.10) gives

$$\frac{1}{n}\sum_{k=n+1}^{2n}|C_k(x)|^p>(c(x)\log n)^p\frac{\varepsilon}{4\pi},$$

from which (3.5) already follows.

Finally, we sketch in a few words how we can prove the proposition of Theorem 2 with the aid of the above argument.

For $0<\eta<\varepsilon<1$ let

$$f(\eta,\varepsilon;x)=\begin{cases}x^{-(1+\alpha)}\log\dfrac{1}{x} & \text{if } x\in(\eta;\,\varepsilon),\\ 0 & \text{if } x\in[-\pi;\,\eta]\cup[\varepsilon;\,\pi).\end{cases}$$

The function

$$f(x)=\sum_{n=1}^{\infty}\sum_{j=1}^{4}f\left(\eta_{4n+j},\varepsilon_{4n+j};\,x-j\frac{\pi}{2}\right)$$

is in L^β for every $\beta<\frac{1}{1+\alpha}$ provided that $\varepsilon_{4(n+1)+j}<\eta_{4n+j}$ $(n=1,2,\ldots;\ j=1,2,3,4)$. Now taking into account that (1.3) is true a.e. for bounded functions, we can determine the sequences $\{\eta_k\}_{k=1}^{\infty}$ and $\{\varepsilon_k\}_{k=1}^{\infty}$ step by step so that f will satisfy (1.5) a.e. We do not go into the details.

PROOF OF THEOREM 3. Let $\varepsilon>0$ such that f be bounded in $(x-\varepsilon;\ x+\varepsilon)$, let f_1 coinciding with f in $(x-\varepsilon;\ x+\varepsilon)$ and 0 elsewhere and $f_2=f-f_1$. The first part of the proof of Theorem 1 $\left(f\in L^r \text{ for some } r>\frac{1}{1+\alpha}\right)$ shows that (1.3) is true for f_1, thus we only have to prove the same for f_2.

$f_2=f_2^+-f_2^-$ where f_2^+ and f_2^- are the positive and negative part of f_2, respectively. But f_2^+ is non-negative and vanishes in $(x-\varepsilon; x+\varepsilon)$, thus $\sigma_k^{\alpha+1}(f_2^+;\ x)\to 0=$ $=f_2^+(x)$, $\tilde{\sigma}_k^{\alpha+1}(f_2^+;\ x)\to\tilde{f}_2^+(x)$ (see [5, I, p. 94]), and the proof of (3.4) shows that it is true for f_2^+ at x. By the proof of Theorem 1 these imply that (1.3) is valid for f_2^+. Similar argument shows that (1.3) is true for f_2^-, too, and so it is also true for f_2, which was to be proved.

The proof of (1.4) when $\tilde{f}(x)$ is defined and f is continuous at x follows the same line, only we have to use the formulas

$$\tilde{K}_k^\alpha(t)=\frac{1}{2}\operatorname{cotg}\frac{t}{2}-H_k^\alpha(t)$$

and

$$H_k^\alpha(t)=\frac{1}{A_k^\alpha}\frac{\cos\left[\left(k+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}+\frac{2\theta(t)\alpha}{k\left(2\sin\frac{t}{2}\right)^2}$$

$$\left(-1<\alpha<1,\ |\theta|\leqq 1,\ \frac{1}{k}\leqq t\leqq\pi\right)$$

instead of the one for $K_k^\alpha(t)$.

Finally, to prove the last statement of Theorem 3 let $g(x)$ be 2π-periodic and

$$g(x)=|x|^{-(1+\alpha)}\log\frac{1}{|x|}\quad\text{if}\quad x\in[-\pi;\pi],$$

and let

$$f(x)=\frac{g(x)\left(2\sin\frac{\pi-|x|}{2}\right)^{1+\alpha}}{\cos\frac{1+\alpha}{2}x}\qquad(x\in[-\pi;\pi]).$$

We shall show that f satisfies the requirements.

It is clear that $f\in L^\beta$ for every $\beta<\frac{1}{1+\alpha}$ and that f is continuous at π. We have

$$|\sigma_n^\alpha(\pi)|=|\sigma_n^\alpha(\pi)-f(\pi)|\geqq\frac{2}{\pi}\frac{1}{A_n^\alpha}\left|\int_{1/n}^{\pi}\frac{\varphi_\pi(t)\sin\left[\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}dt\right|-$$

$$-\frac{2}{\pi}\left|\int_0^{1/n}\varphi_\pi(t)K_n^\alpha(t)\,dt\right|-\frac{2}{\pi}\left|\int_{1/n}^{\pi}\frac{\varphi_\pi(t)2\theta(t)\alpha}{n\left(2\sin\frac{t}{2}\right)^2}dt\right|$$

(see the proof of Theorem 1). Here the second and third member are $o(1)$ since f is continuous at π.

Now $\varphi_\pi(t)=2f(\pi-t)$, thus

$$\int_{1/n}^{\pi} \frac{\varphi_\pi(t)\sin\left[\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\,dt = \int_0^{\pi-\frac{1}{n}} \frac{2f(\tau)(-1)^n\cos\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)\tau}{\left(2\sin\frac{\pi-\tau}{2}\right)^{1+\alpha}}\,d\tau =$$

$$= (-1)^n \int_{-\pi}^{\pi} g(\tau)\cos n\tau\,d\tau + (-1)^{n+1}\int_{-\pi}^{\pi} g(\tau)\,\mathrm{tg}\left(\frac{1+\alpha}{2}\tau\right)\sin n\tau\,d\tau + O\left(\frac{1}{n}\right).$$

The function $g(\tau)\,\mathrm{tg}\left(\frac{1+\alpha}{2}\tau\right)$ is of bounded variation on $[-\pi;\ \pi]$, thus its Fourier coefficients are $O\left(\frac{1}{n}\right)$. It is known (see [7, I, p. 190]) that the cosine coefficients of $g(\tau)$ are asymptotically $n^\alpha\log n$, and so we get that

$$|\sigma_n^\alpha(\pi)| \geqq c\frac{1}{A_n^\alpha}n^\alpha\log n \geqq c_1\log n \quad (c, c_1>0,\ n\geqq n_0)$$

which proves (1.6).

The proof of Theorem 3 is thus completed.

PROOF OF THEOREM 4. To simplify the computation we assume that $0>\alpha\geqq\frac{1}{2}$.

We shall construct a continuous function f for which (1.5) is true a.e. in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$. On the basis of our argument the proof of the general case is very easy but technically much more complicated.

It is enough to show a continuous f for which

$$\limsup_{n\to\infty}\frac{1}{n}|\sigma_n^\alpha(x)-f(x)|^p=\infty$$

a.e. in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$.

Taking into account that

$$\sigma_n^\alpha(x)-f(x)=\frac{2}{\pi}\int_0^{1/n}\varphi_x(t)K_n^\alpha(t)\,dt+\frac{2}{\pi}\frac{1}{A_n^\alpha}\int_{1/n}^{\pi}\frac{\varphi_x(t)\sin\left[\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\,dt+$$

$$+\frac{2}{\pi}\int_{1/n}^{\pi}\frac{\varphi_x(t)2\theta(t)\alpha}{n\left(2\sin\frac{t}{2}\right)^2}\,dt \quad (|\theta|\leqq 1)$$

and that the first and third member on the right are $o(1)$ whenever f is continuous,

our problem is reduced to constructing a continuous f for which

$$\limsup_{n\to\infty}\frac{1}{n}\frac{1}{(A_n^{(\alpha)})^p}\left|\int_{1/n}^{\pi}\frac{\varphi_x(t)\sin\left[\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}dt\right|^p=\infty \tag{3.11}$$

a.e. in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$.

Let n be a fixed natural number divisible by 3, and for the 2π-periodic $g_n(t)$ let

$$g_n(t)=\operatorname{sign} t\sin nt\quad(t\in[-\pi;\pi]).$$

With the notations $x_j=\frac{j\pi}{n}\left(|j|\leqq\frac{n}{3}\right)$ and $\psi_x(t)=\frac{1}{2}\left(g_n(x+t)+g_n(x-t)-2g_n(x)\right)$ we obtain easily that

$$\psi_{x_j}(t)=\begin{cases}0 & \text{if}\quad 0\leqq t\leqq\frac{|j|\pi}{n}\quad\text{or}\quad\pi-\frac{|j|\pi}{n}\leqq t\leqq\pi,\\ (-1)^j\sin nt & \text{if}\quad\frac{|j|\pi}{n}\leqq t\leqq\pi-\frac{|j|\pi}{n}.\end{cases} \tag{3.12}$$

Let

$$Q_n(f;x)=\int_{1/n}^{\pi}\frac{\frac{1}{2}\left(f(x+t)+f(x-t)-2f(x)\right)\sin\left[\left(n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}dt.$$

(3.12) gives

$$Q_n(g_n;x_j)=\frac{1}{2}(-1)^j\int_{\frac{|j|\pi}{n}}^{\pi-\frac{|j|\pi}{n}}\frac{\cos\left[\left(\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}dt+$$

$$+\frac{1}{2}(-1)^{j+1}\int_{\frac{|j|\pi}{n}}^{\pi-\frac{|j|\pi}{n}}\frac{\cos\left[\left(2n+\frac{1}{2}+\frac{\alpha}{2}\right)t-\frac{\pi\alpha}{2}\right]}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}dt.$$

By the second mean value theorem the second member on the right is not greater in absolute value than

$$\frac{1}{2}\frac{1}{\left(2\sin\frac{|j|\pi}{2n}\right)^{1+\alpha}}\frac{2}{2n+\frac{1}{2}+\frac{\alpha}{2}}\leqq n^{\alpha}$$

thus if $n \geqq n_0$ and $|j| \leqq \frac{n}{3}$ then we have

$$|Q_n(g_n; x_j)| \geqq \frac{1}{2} \int_{\pi/3}^{2\pi/3} \cos\left(\frac{1+\alpha}{2} t - \frac{\pi\alpha}{2}\right) dt \geqq \frac{1}{8}. \tag{3.13}$$

Using the inequality

$$|\psi_{x_j}(t) - \psi_{x_j+\frac{1}{n}y}(t)| \leqq 2n \frac{|y|}{n} = 2|y|$$

we obtain from (3.13) that if $0 \leqq y \leqq \frac{1}{2} 10^{-2}(-\alpha)$ then

$$\left|Q_n\left(g_n; x_j + \frac{1}{n} y\right)\right| \geqq Q_n(g_n; x_j) - \int_{1/n}^{\pi} \frac{2y}{\left(2 \sin \frac{t}{2}\right)^{1+\alpha}} dt \geqq \frac{1}{8} + \frac{1}{\alpha} y \cdot 2^{-\alpha} \pi \geqq 10^{-1}.$$

We have got that if $n \geqq n_0$, $10^{-m_0} \leqq \frac{10^{-2}(-\alpha)}{2\pi}$ (m_0 integer) and

$$x \in \left(\frac{j}{n}\pi; \frac{j}{n}\pi + 10^{-m_0}\frac{\pi}{n}\right) = I_j(n) \quad \left(j = -\frac{n}{3}, \ldots, \frac{n}{3} - 1\right)$$

then

$$Q_n(g_n; x) \geqq 10^{-1}. \tag{3.14}$$

Let $$\delta = \frac{1}{2}\frac{1}{p}(-1 - p\alpha).$$

We shall define a sequence $\{n_k\}_{k=1}^{\infty}$ of natural numbers as follows: Let us suppose that $n_0 \leqq n_1 < \ldots < n_{k-1}$ are already defined. The function $h_k(t) = \sum_{i=1}^{k-1} n_i^{-\delta} g_{n_i}(t)$ belongs to the class Lip 1, which implies (as one can see easily) that $Q_n(h_k; x) = O(n^{\alpha})$ uniformly in x. Let now n_k be such that

(a) $$|Q_{n_k}(h_k; x)| < 10^{-2} n_k^{-\delta} \quad (x \in [-\pi; \pi]),$$

(b) $$10^{m_0} n_{k-1} | n_k$$

and

(c) $$n_k^{-\delta} < \min_{1 \leqq i \leqq k-1} \left\{\frac{10^{-2}}{2^k}(-\alpha) n_i^{-\delta}\right\}$$

should be satisfied. Thus we have defined the sequence $\{n_k\}$.

We claim that the function

$$f(x) = \sum_{k=1}^{\infty} \frac{g_{n_k}(x)}{n_k^{\delta}}$$

satisfies our requirements.

(c) shows that $n_k^{-\delta}<\frac{1}{2^k}$, and so f as the sum of a uniformly convergent series of continuous functions is continuous itself.

If now $x\in I_j(n_k)$ for some k and $|j|\leqq\frac{n_k}{3}$, then (3.14), (a) and (c) imply

$$|Q_{n_k}(f;\,x)|\geqq|Q_{n_k}(g_{n_k}n_k^{-\delta};\,x)|-|Q_{n_k}(h_k;\,x)|-$$

$$-2\max_{-\pi\leqq t\leqq\pi}\Big|\sum_{l=k+1}^{\infty}n_l^{-\delta}g_{n_l}(t)\Big|\int_{1/n}^{\pi}\frac{1}{\left(2\sin\frac{t}{2}\right)^{1+\alpha}}\,dt\geqq$$

$$\geqq 10^{-1}n_k^{-\delta}-10^{-2}n_k^{-\delta}-2\frac{\pi}{(-\alpha)}n_k^{-\delta}\sum_{l=k+1}^{\infty}\frac{10^{-2}}{2^l}(-\alpha)\geqq\frac{1}{20}n_k^{-\delta}$$

and so

$$\frac{1}{n_k}\frac{1}{(A_{n_k}^{\alpha})^p}|Q_{n_k}(f;\,x)|^p\geqq cn_k^{-1-\alpha p-\delta p}=cn_k^{\frac{1}{2}(-\alpha p-1)}.$$

If $k\to\infty$ the right-hand side tends to infinity, thus it remains to show that if

$$H_k=\bigcup_{j=-\frac{n_k}{3}}^{\frac{n_k}{3}-1}I_j(n_k)$$

then almost every x in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$ is contained in infinitely many H_k.

Let $\chi_k(t)=1$ if $t\in H_k$ and 0 elsewhere. We have to show that $\sum_{k=1}^{\infty}\chi_k(t)=\infty$ a.e. in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$, and this will follow if we prove the existence of such a sequence $\{m_n\}$ for which

$$\lim_{n\to\infty}\frac{1}{m_n}\sum_{k=1}^{m_n}(\chi_k(t)-10^{-m_0})=0 \tag{3.15}$$

a.e. in $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$. By (b) and the structure of H_k the sequence $\{\chi_k(t)-10^{-m_0}\}_{k=1}^{\infty}$ is orthogonal on $\left(-\frac{\pi}{3};\frac{\pi}{3}\right)$, and so the existence of a sequence $\{m_n\}$ satisfying (3.15) follows from Lemma 3.

The proof of Theorem 4 is completed.

REFERENCES

[1] ALEXITS, G.: *Konvergenzprobleme der Orthogonalreihen*, Akadémiai Kiadó, Budapest, 1960.
[2] LEINDLER, L.: On summability of Fourier series, *Acta Sci. Math.* **29** (1968), 147—162.
[3] LEINDLER, L.: On strong approximation of Fourier series, *Periodica Math. Hung.* **1** (1971), 157—162.
[4] LEINDLER, L.: On the strong summability and approximation of Fourier series, *Approximation Theory* (Banach Center Publications, to appear).
[5] PÓLYA, G. und SZEGŐ, G.: *Aufgaben und Lehrsätze aus der Analysis, I*, Berlin, 1925.
[6] ZYGMUND, A.: On the convergence and summability of power series on the circle of convergence, *Proc. London Math. Soc.* **47** (1941), 326—350.
[7] ZYGMUND, A.: *Trigonometric Series I, II*, Cambridge, 1959.

Bolyai Institute of the József Attila University,
Szeged, Aradi vértanúk tere 1, Hungary 6720

(Received September 25, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 445—446.

IF $L\left(\frac{1}{2}, \chi\right) > 0$, THEN $L\left(\frac{1}{2}, \chi\right)$ CANNOT BE A MINIMUM

by

A. MALLIK

INTRODUCTION. If $L(s, \chi)$ is the L-function belonging to the real primitive character $\chi \pmod{|D|}$, for a fundamental discriminant D, then it is an outstanding conjecture that

$$L(s, \chi) \neq 0 \quad \text{for} \quad s \in [0, 1]. \tag{1}$$

This conjecture seems totally unapproachable at the moment, but we mention here a result of LOW [2], who proved that

$$L(s, \chi) \neq 0 \quad \text{for} \quad s \in [0, 1] \quad \text{if} \quad 0 > D > -593\,000, \quad D \neq -115\,147. \tag{2}$$

Values for $L\left(\frac{1}{2}, \chi\right)$ for $D < 0$ have been used by MONTGOMERY and WEINBERGER [3], in connection with the class number two problem. For a selection of values of the class number $h(D)$, they selected the largest $|D|$ known, e.g., $h(-163) = 1$, $h(-427) = 2$, and amongst other things computed $L\left(\frac{1}{2}, \chi\right)$ for that discriminant. The values found were extremely close to zero, e.g., they found $L\left(\frac{1}{2}, \chi\right) = 6.0362 \times 10^{-5}$, $D = -115\,147$. This numerical evidence might lead one to suspect that $L\left(\frac{1}{2}, \chi\right)$ is a minimum of $L(s, \chi)$ for $s \in [0, 1]$. To prove that this is not the case we prove:

THEOREM. *Let χ be a real primitive character* $\bmod |D|$, *with* $|D| > 320$, *such that* $L\left(\frac{1}{2}, \chi\right) > 0$, *then* $L\left(\frac{1}{2}, \chi\right)$ *cannot be a minimum of* $L(s, \chi)$ *for* $s \in [0, 1]$.

COROLLARY. *If* $D \neq -115\,147$, *and* $-320 > D > -593\,000$, *then* $L'\left(\frac{1}{2}, \chi\right) < 0$.

PROOFS. The functional equation for L-functions belonging to real primitive characters is given by

$$G(s, \chi) = \left(\frac{|D|}{\pi}\right)^{\frac{s+a}{2}} \Gamma\left(\frac{s+a}{2}\right) L(s, \chi) = G(1-s, \chi), \tag{3}$$

where

$$a = \frac{1}{2}(1 - \chi(-1)) \in \{0, 1\}.$$

Thus from (3) we have,

$$L(1-s, \chi) = \left(\frac{|D|}{\pi}\right)^{s-\frac{1}{2}} \frac{\Gamma\left(\frac{s+a}{2}\right)}{\Gamma\left(\frac{1-s+a}{2}\right)} L(s, \chi). \tag{4}$$

Differentiating (4) with respect to s and then setting $s=\frac{1}{2}$, gives

$$-2\frac{L'}{L}\left(\frac{1}{2}, \chi\right) = \log\left(\frac{|D|}{\pi}\right) + \frac{\Gamma'}{\Gamma}\left(\frac{1}{4}+\frac{a}{2}\right). \tag{5}$$

We now show that for $|D|$ sufficiently large the r.h.s. of (5) is positive. For $a=1$,

$$\frac{\Gamma'}{\Gamma}\left(\frac{1}{4}+\frac{a}{2}\right) = \frac{\Gamma'}{\Gamma}\left(\frac{3}{4}\right) > -1.08;$$

and for $a=0$,

$$\frac{\Gamma'}{\Gamma}\left(\frac{1}{4}+\frac{a}{2}\right) = \frac{\Gamma'}{\Gamma}\left(\frac{1}{4}\right) > -4.21.$$

The lower bounds are got from Table 2 on p. 15 of [1].

Thus the r.h.s. of (5) is positive for $\log\left(\frac{|D|}{\pi}\right) > 4.22$, i.e., $|D| > \pi e^{4.6} \approx 314$. This proves the theorem and the corollary follows immediately from Low's result [2], on noting that $L(1, \chi) > 0$.

REFERENCES

[1] JAHNKE—EMDE—LOSCH: *Tables of higher functions*, 6-th edition, McGraw—Hill, 1960.
[2] LOW, M.: Real zeros of the Dedekind zeta function of an imaginary quadratic field, *Acta Arith.* **14** (1968), 117—140.
[3] MONTGOMERY, H. and WEINBERGER, P. J.: Notes on small class numbers, *Acta Arith.* **24** (1974), 529—542.

University of Port Harcourt,
POB 5323, Nigeria

(Received October 9, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 447—456.

A PROBLEM CONNECTED WITH MULTIPLE CIRCLE-PACKINGS AND CIRCLE-COVERINGS

by

G. FEJES TÓTH

Dedicated to Professor J. Molnár on his 60th birthday

A system of circles is said to form a *k-fold packing* if each point of the plane belongs to at most k circles. Analogously, a system of circles is said to form a *k-fold covering* if each point of the plane belongs to at least k circles. Let δ_k be the supremum of the densities of all k-fold packings of equal circles and let Δ_k be the infimum of the densities of all k-fold coverings with equal circles. The longstanding problems of finding the densest k-fold packing and the thinnest k-fold covering of the plane with equal circles are solved only for $k=1$, i.e., for ordinary packings and coverings, and seem to be extremely difficult for $k \geqq 2$. Therefore the rather vast literature devoted to the study of multiple packings and coverings with equal circles [1, 2, 3, 4, 5, 6, 7, 10, 12, 14, 18, 19] deals only with the following two simpler problems:

Find the densest k-fold lattice-packing and the thinnest k-fold lattice-covering with circles.

Give possibly good lower and upper bounds for the quantities δ_k and Δ_k.

As to the first problem, the densest k-fold lattice-packing of equal circles is known for $1 \leqq k \leqq 7$ [2, 7, 18]; the thinnest k-fold lattice-covering with circles is known for $1 \leqq k \leqq 4$ [1]. Denoting with $\bar{\delta}_k$ the density of the densest k-fold lattice-packing of circles and with $\bar{\Delta}_k$ the density of the thinnest k-fold lattice-covering with circles, we have

$$\bar{\delta}_1 = \frac{\pi}{2\sqrt{3}} = 0.906\ldots, \quad \bar{\delta}_2 = 2\bar{\delta}_1 = 1.813\ldots,$$

$$\bar{\delta}_3 = 3\bar{\delta}_1 = 2.720\ldots, \quad \bar{\delta}_4 = 4\bar{\delta}_1 = 3.627\ldots,$$

$$\bar{\delta}_5 = \frac{4\pi}{\sqrt{7}} = 4.749\ldots, \quad \bar{\delta}_6 = \frac{35\pi}{8\sqrt{6}} = 5.611\ldots,$$

$$\bar{\delta}_7 = \frac{8\pi}{\sqrt{15}} = 6.489\ldots;$$

$$\bar{\Delta}_1 = \frac{2\pi}{\sqrt{27}} = 1.209\ldots, \quad \bar{\Delta}_2 = 2\bar{\Delta}_1 = 2.418\ldots,$$

$$\bar{\Delta}_3 = \frac{\pi\sqrt{27\,138 + 2910\sqrt{97}}}{216} = 3.435\ldots,$$

$$\bar{\Delta}_4 = \frac{25\pi}{18} = 4.363\ldots.$$

Several attempts were made to construct dense k-fold lattice-packings of circles for $k>7$ and thin k-fold lattice-coverings with circles for a few values of $k>4$ [3, 4, 5, 14] but almost all of these constructions have been improved by U. BOLLE [7]. Bolle also gave asymptotic bounds for $\bar{\delta}_k$ and $\bar{\Delta}_k$. He proved that there are positive constants c_1, c_2, c_3 and c_4 such that

$$k-c_1k^{2/5} \leqq \bar{\delta}_k \leqq k-c_2k^{1/4}$$

and

$$k+c_3k^{1/4} \leqq \bar{\Delta}_k \leqq k+c_4k^{2/5}.$$

It is known that $\delta_2 \geqq 1.854\ldots > \bar{\delta}_2$, $\delta_3 \geqq 2.781\ldots > \bar{\delta}_3$, $\delta_4 \geqq 3.708\ldots > \bar{\delta}_4$ [19], $\Delta_2 \leqq 2.347\ldots < \bar{\Delta}_2$ [10] and $\Delta_3 \leqq 3.298\ldots < \bar{\Delta}_3$ [6]. Obviously, we have $\delta_k \geqq \bar{\delta}_k$, and for $k>4$ no better lower bound for δ_k is known. Similarly, we have $\Delta_k \leqq \bar{\Delta}_k$, and for $k>3$ no better upper bound for Δ_k is known. As to the opposite bounds for δ_k and Δ_k we have obviously $\delta_k \leqq k \leqq \Delta_k$. Independently from one another L. FEW and K. BÖRÖCZKY proved bounds of the form $\delta_2 \leqq c < 2$ with $c > 1.999$ (oral communication). But it is surprising that apart of these unpublished bounds no better upper bounds for δ_k and no better lower bounds for Δ_k were known than the above trivial bounds. In a recent paper [12] I proved that

$$\delta_k \leqq \frac{\pi}{6}\cot\frac{\pi}{6k} \quad \text{and} \quad \Delta_k \geqq \frac{\pi}{3}\operatorname{cosec}\frac{\pi}{3k}.$$

The fact that these bounds are better than the trivial bounds becomes obvious by observing that they are equal to the density of k coincident circles with respect to the circumscribed and inscribed regular $6k$-gon, respectively.

As a starting point of our investigations we mention the following problem (see [15] p. 80). Distribute in the plane equal circles with a prescribed density d so as to cover the greatest possible part of the plane, more precisely, so as to maximize the density s_1 of the part of the plane covered by the circles. If $d \leqq \delta_1$ the circles can be arranged so as to form a packing. Now we have $s_1 = d$. On the other hand, if $d \geqq \Delta_1$ the circles can be arranged so as to cover the plane completely, i.e., we can attain $s_1 = 1$. For $d > \delta_1$ we have $s_1 < d$ and for $d < \Delta_1$ we have $s_1 < 1$. As we see, the problem under consideration is a generalization of the problems of the densest packing and thinnest covering of the plane with equal circles. For $\delta_1 \leqq d \leqq \Delta_1$ the solution of the problem is given, roughly speaking, by placing the circles in a lattice generated by an equilateral triangle [15]. We shall study a natural generalization of this problem which connects the problems of the densest k-fold packing and the thinnest k-fold covering of the plane with equal circles.

Distribute in the plane equal circles with a given density d. Let s_i be the density of the set of those points of the plane which are covered by at least i circles. How should the circles be arranged so as to maximize the sum $\sum_{i=1}^{k} s_i$?

Let $h_k(d)$ be the supremum of $\sum_{i=1}^{k} s_i$ taken over all systems of equal circles with density d. It is easily seen that $\sum_{i=1}^{\infty} s_i = d$. Thus we have $h_k(d) \leqq d$ and equality holds only if there is an arrangement of equal circles with density d such that $s_i = 0$

for $i>k$. It is not difficult to deduce that then there exists also a k-fold packing of equal circles with density d. Hence we obtain $\delta_k = \sup_{h_k(d)=d} d$. In a similar way we see that $\Delta_k = \inf_{h_k(d)=k} d$.

Of course this problem is hopelessly difficult for $k \geqq 2$. Therefore we shall deal only with the more modest problem of finding a non-trivial upper bound for $h_k(d)$. We shall also study the above problem for $k=2$ under the restriction that the circles form a lattice.

Let $\bar{h}_k(d)$ be the supremum of the sum $\sum_{i=1}^{k} s_i$ extended over all lattice arrangements of circles with density d. Let $f_n(x)$ denote the area of the intersection of a circle of unit area and a concentric regular n-gon of area x. We shall write $|X|$ for the area (Lebesgue measure) of a point set X and $\|Y\|$ for the cardinality of the set Y. Using these notations our results can be formulated as follows:

THEOREM 1. *Let $\mathcal{S}$ be a finite system of circles of unit area, H a convex polygon with at most six sides and $F(i)$ the set of those points of H which are covered by at least i circles of $\mathcal{S}$. Then we have*

$$\sum_{i=1}^{k} |F(i)| \leqq \|\mathcal{S}\| f_{6k}(k|H|/\|\mathcal{S}\|).$$

THEOREM 2. *We have*

$$\bar{h}_2(d) = d f_6(2/d).$$

If $\bar{\delta}_2 \leqq d \leqq \bar{\Delta}_2$ then the lattice of circles of density d for which s_1+s_2 is maximal is generated by two orthogonal vectors whose lengths are in the proportion $\sqrt{3}:1$.

In the case when $k=1$ Theorem 1 reduces to a theorem of L. FEJES TÓTH (see [15] p. 80). Theorem 2 is a generalization of the above mentioned results which state that $\bar{\delta}_2 = 2\bar{\delta}_1$ and $\bar{\Delta}_2 = 2\bar{\Delta}_1$. As a corollary of Theorem 1 we obtain the following

THEOREM 3. *We have*

$$h_k(d) \leqq d f_{6k}(k/d).$$

Theorem 3 contains the bounds for δ_k and Δ_k mentioned in the introduction:

$$\delta_k = \sup_{h_k(d)=d} d \leqq \sup_{d f_{6k}(k/d)=d} d = \frac{\pi}{6} \cot \frac{\pi}{6k}$$

and

$$\Delta_k = \inf_{h_k(d)=k} d \geqq \inf_{d f_{6k}(k/d)=k} d = \frac{\pi}{3} \operatorname{cosec} \frac{\pi}{3k}.$$

The proof of the Theorems 1 and 2 rests on the investigation of the k'th Dirichlet cells. Let $\mathcal{S}$ be a system of different equal circles and H a domain. The k'th *Dirichlet cell* D_C^k of the circle $C \in \mathcal{S}$ with respect to H is the set of all points $P \in H$ such that there are at most $k-1$ centres of the circles of $\mathcal{S}$ nearer to P than the centre of C. It will not make any confusion that we do not indicate in our notations the dependence of the k'th Dirichlet cell on $\mathcal{S}$ and H. We shall use the following simple properties of the k'th Dirichlet cells:

(i) *If $k \leqq \|\mathcal{S}\|$ then the k'th Dirichlet cells cover H exactly k times.* More precisely, each point of H which is not a boundary point of a k'th Dirichlet cell lies in the interior of exactly k k'th Dirichlet cells. Indeed, if P is an arbitrary point of H and $O_1, O_2, \ldots$ are the centres of the circles of $\mathcal{S}$, choosing the notations so that $PO_1 \leqq PO_2 \leqq \ldots \leqq PO_k \leqq PO_{k+1} \leqq \ldots$, then P lies in the interior of the k'th Dirichlet cell of the circles centred at $O_1, \ldots, O_k$ except when $PO_k = PO_{k+1}$. In this latter case P is a common boundary point of the k'th Dirichlet cells of the circles centred at O_k and O_{k+1}.

(ii) *Let $F(i)$ be the set of those points of H which are covered by at least i circles of $\mathcal{S}$. Write $F_C^k = D_C^k \cap C$. Then each point of H which is not a boundary point of a k'th Dirichlet cell is covered with the same multiplicity by the sets $F(i)$, $i=1, \ldots, k$, as by the sets F_C^k, $C \in \mathcal{S}$*

Let P be an arbitrary point of H which is not a boundary point of a k'th Dirichlet cell and belongs to exactly j of the sets F_C^k, $C \in \mathcal{S}$. By property (i) we have $j \leqq k$. If $j=k$ then the statement is obvious. Suppose that $j<k$. Let $C_1, \ldots, C_j$ be the circles for which $P \in F_{C_l}^k$, $l=1, \ldots, j$. We have to show that there is no further circle of $\mathcal{S}$ which contains P. Let C_{j+1} be the circle of $\mathcal{S}$ other than $C_1, \ldots, C_j$ whose centre lies nearest to P. If there are more than j circles of $\mathcal{S}$ containing P then obviously $P \in C_{j+1}$. But this is impossible, since, by definition, we have $P \in D_{C_{j+1}}^k$. This proves property (ii).

Let now H be a convex polygon with at most six sides. Then D_C^k is a simple polygon. To see this we observe that D_C^k can be constructed in the following way: For each circle $C' \in \mathcal{S}$, $C' \neq C$ we draw the radical line of C and C'. Then D_C^k is the set of all points $P \in H$ for which the segment joining P with the centre of C intersects at most $k-1$ of these lines. Let n_C^k be the number of angles of D_C^k less than π. We shall need the following

LEMMA. *Let $\mathcal{S}$ be a finite set of different equal circles such that the centre of any circle of $\mathcal{S}$ lies in H. Suppose that there are no four centres of the circles of $\mathcal{S}$ lying on a circle or on a straight line, no three centres have equal distances from a point of the boundary of H and no two centres have equal distances from a vertex of H. Then we have*

$$\sum_{C \in \mathcal{S}} n_C^k \leqq 6k \|\mathcal{S}\|.$$

As to the proof of the Lemma which requires a deeper study of the combinatorial structure of the k'th Dirichlet cells we refer to [12].

In the proof of Theorem 1 we shall make use of a further notion. Let A and B be two domains and p and q two positive numbers. We define the weighted area deviation, in short the *deviation of B from A* by

$$a(A, B) = p|A-B| + q|B-A|.$$

In various investigations it is convenient to impose certain conditions upon the weights. In [13], e.g., it was supposed that $\frac{1}{p} + \frac{1}{q} = 1$. In the present paper the condition $p+q=1$ will be more convenient. Note that for $p \neq q$ the deviation is not symmetric in A and B.

Let $a(n)$ be the minimum of the deviation of an n-gon from a circle C of unit area. Let B be an n-gon for which $a(C, B) = a(n)$. It is easily seen that B has the

property that each side of B is divided by the boundary of C into three segments in the ratio $p:2q:p$. It follows that B is a regular n-gon concentric with C.

It is obvious that the sequence $a(n)$, $n=3, \ldots$ is decreasing and it follows from a general theorem [13] that it is convex:

$$a(n-1)+a(n+1) \geqq 2a(n), \quad n = 4, 5, \ldots .$$

This can be checked also directly by a simple computation [11] using the above mentioned property of the n-gon with minimal deviation.

After this preliminaries the proof of Theorem 1 will be rather simple. Let u_n be the area of a regular n-gon inscribed into a circle of unit area. Let v_n be the area of a regular n-gon circumscribed about a circle of unit area. If $x \leqq u_n$ then $f_n(x)=x$ and if $x \geqq v_n$ then $f_n(x)=1$. Thus for $k|H|/\|\mathcal{S}\| \leqq u_{6k}$ and $k|H|/\|\mathcal{S}\| \geqq v_{6k}$ Theorem 1 yields for $\sum_{i=1}^{k} |F(i)|$ the trivial upper bounds $k|H|$ and $\|\mathcal{S}\|$, respectively. Therefore we suppose that $u_{6k}<k|H|/\|\mathcal{S}\|<v_{6k}$. By property (ii) we have $\sum_{i=1}^{k} |F(i)| = \sum_{C \in \mathcal{S}} |F_C^k|$. This shows that we may restrict ourselves to the case when $k \leqq \|\mathcal{S}\|$. For, if $k>\|\mathcal{S}\|$ then we have obviously

$$\sum_{i=1}^{k} |F(i)| = \sum_{C \in \mathcal{S}} |F_C^k| = \sum_{C \in \mathcal{S}} |C \cap H| \leqq$$

$$\leqq \|\mathcal{S}\| f_6(|H|) \leqq \|\mathcal{S}\| f_{6k}(k|H|/\|\mathcal{S}\|).$$

Further, we may assume without loss of generality that $\mathcal{S}$ satisfies all the conditions of the Lemma.

We have, on the one hand, $|F_C^k|=|C|-|C-D_C^k|$, on the other hand, $|F_C^k|= =|D_C^k|-|D_C^k-C|$. Thus if p and q are positive numbers for which $p+q=1$ then

$$\sum_{i=1}^{k} |F(i)| = \sum_{C \in \mathcal{S}} F_C^k = p \sum_{C \in \mathcal{S}} |C| + q \sum_{C \in \mathcal{S}} |D_C^k| - \sum_{C \in \mathcal{S}} a(C, D_C^k). \tag{1}$$

By property (i) and the supposition that $k \leqq \|\mathcal{S}\|$, we have

$$\sum_{C \in \mathcal{S}} |D_C^k| = k|H|. \tag{2}$$

Further, since the circles of $\mathcal{S}$ have unit area,

$$\sum_{C \in \mathcal{S}} |C| = \|\mathcal{S}\|. \tag{3}$$

We continue to investigate the sum $\sum_{C \in \mathcal{S}} a(C, D_C^k)$. We shall show that $a(C, D_C^k) \geqq \geqq a(n_C^k)$, where, in accordance with the notations of the Lemma, n_C^k denotes the number of angles of D_C^k less than π.

Let $\overline{F}_C^k$ be the convex hull of F_C^k. We say that the segment XY is an edge of $\overline{F}_C^k$ if XY belongs to the boundary of $\overline{F}_C^k$ and no line-segment belonging to the boundary of $\overline{F}_C^k$ contains XY as a proper part. Let XY be an edge of $\overline{F}_C^k$. We associate with XY the maximal segment $X'Y'$ of the line XY with the property that X' and

Y' are boundary points of D_C^k and the segment $X'Y'$ is contained in $C \cup D_C^k$. If $X'Y'$ is not identical with the arc XY of the boundary of D_C^k then we construct a new polygon by replacing the arc $X'Y'$ of the boundary of D_C^k by the segment $X'Y'$. By this transformation the number of convex angles of D_C^k does not increase while the number of concave angles and the deviation from C certainly decreases. Iterations of this transformation provide in finitely many steps a polygon $\tilde{D}_C^k$ which does not have vertices with concave angle in C. Let Z be a vertex of $\tilde{D}_C^k$ such that the angle at Z is concave. We observe that then the edges emanating from Z do not intersect C. Let UV be a segment contained in $\tilde{D}_C^k - C$ such that U and V are boundary points of $\tilde{D}_C^k$ and $Z \in UV$. Replacing the arc UV of the boundary of $\tilde{D}_C^k$ by the segment UV the number of concave angles and the deviation from C decreases and the number of convex angles does not increase. Repeating this operation we finally obtain a convex polygon $\hat{D}_C^k$ with at most n_C^k vertices such that $a(C, D_C^k) \geqq a(C, \hat{D}_C^k) \geqq a(n_C^k)$.

Using now the Lemma and the fact that the sequence $a(n)$, $n=3, 4, \ldots$ is decreasing and convex, we see that

$$\sum_{C \in \mathcal{S}} a(C, D_C^k) \geqq \sum_{C \in \mathcal{S}} a(n_C^k) \geqq \|\mathcal{S}\| a(6k). \tag{4}$$

Combining the relations (1) to (4), we obtain

$$\sum_{i=1}^{k} |F(i)| \leqq p\|\mathcal{S}\| + qk|H| - \|\mathcal{S}\| a(6k). \tag{5}$$

The inequality (5) is valid for any positive weights p and q for which $p+q=1$. We claim that p and q can be chosen so that the right-hand side of (5) becomes equal to $\|\mathcal{S}\| f_{6k}(k|H|/\|\mathcal{S}\|)$.

Let C be a circle of unit area and B a $6k$-gon for which $a(C, B) = a(6k)$. We remind of the fact that B is a regular $6k$-gon concentric with C such that each side of B is divided by the boundary of C into three segments in the ratio $p:2q:p$. It immediately follows that varying the weights p and q $(p+q=1)$ so that p increases from 0 to 1, $|B|$ increases continuously from u_{6k} to v_{6k}. By the supposition that $u_{6k} < k|H|/\|\mathcal{S}\| < v_{6k}$ we can choose the weights p and q so that $|B| = k|H|/\|\mathcal{S}\|$. Then we have $a(6k) = a(C, B) = p|C| + q|B| - f_{6k}(|B|) = p + qk|H|/\|\mathcal{S}\| - f_{6k}(k|H|/\|\mathcal{S}\|)$. Substituting this value of $a(6k)$ into (5) we obtain the inequality to be proved.

Now we turn to the proof of Theorem 2. Let $\mathcal{S}$ be a lattice-arrangement of circles of unit area with density d. Let C be an arbitrary circle of $\mathcal{S}$ and $D = D_C^2$ the second Dirichlet cell of C. By the homogeneity of the arrangement and the properties (i) and (ii) we have $s_1 + s_2 = d|C \cap D|$. Thus the inequality $s_1 + s_2 \leqq \leqq d f_6(2/d)$ is equivalent with the inequality $|C \cap D| \leqq f_6(2/d)$.

Let O denote the centre of C. Let O_1 be one of the points of the lattice Λ of the centres of the circles lying nearest to O. Let O_2 be one of the points of Λ not collinear with O and O_1 lying nearest to O such that $\sphericalangle O_1 O O_2 \leqq \pi/2$. We think O to be the origin of a coordinate system and shall denote a point P of the plane with the same symbol as the vector pointing from O to P. Write $O_3 = O_2 - O_1$, $O_4 = -O_1$, $O_5 = -O_2$, $O_6 = O_1 - O_2$, $O_7 = O_1 + O_2$, $O_8 = -O_1 - O_2$, $O_9 = 2O_1$ and $O_{10} = -2O_1$. We introduce the notation $C(XYZ)$ to denote the centre of the circum-

circle of the triangle XYZ and define the points E_i, F_i and G_i, $i=1, \ldots, 6$ as follows:

$$E_1 = C(OO_1O_2), \quad E_2 = C(OO_2O_3), \quad E_3 = C(OO_3O_4),$$
$$E_4 = C(OO_4O_5), \quad E_5 = C(OO_5O_6), \quad E_6 = C(OO_6O_1),$$
$$F_1 = C(OO_1O_7), \quad F_2 = C(OO_2O_4), \quad F_3 = C(OO_5O_8),$$
$$F_4 = C(OO_4O_8), \quad F_5 = C(OO_1O_5), \quad F_6 = C(OO_2O_7),$$
$$G_1 = C(OO_3O_7), \quad G_2 = C(OO_3O_{10}), \quad G_3 = C(OO_8O_{10}),$$
$$G_4 = C(OO_6O_8), \quad G_5 = C(OO_6O_9), \quad G_6 = C(OO_7O_9).$$

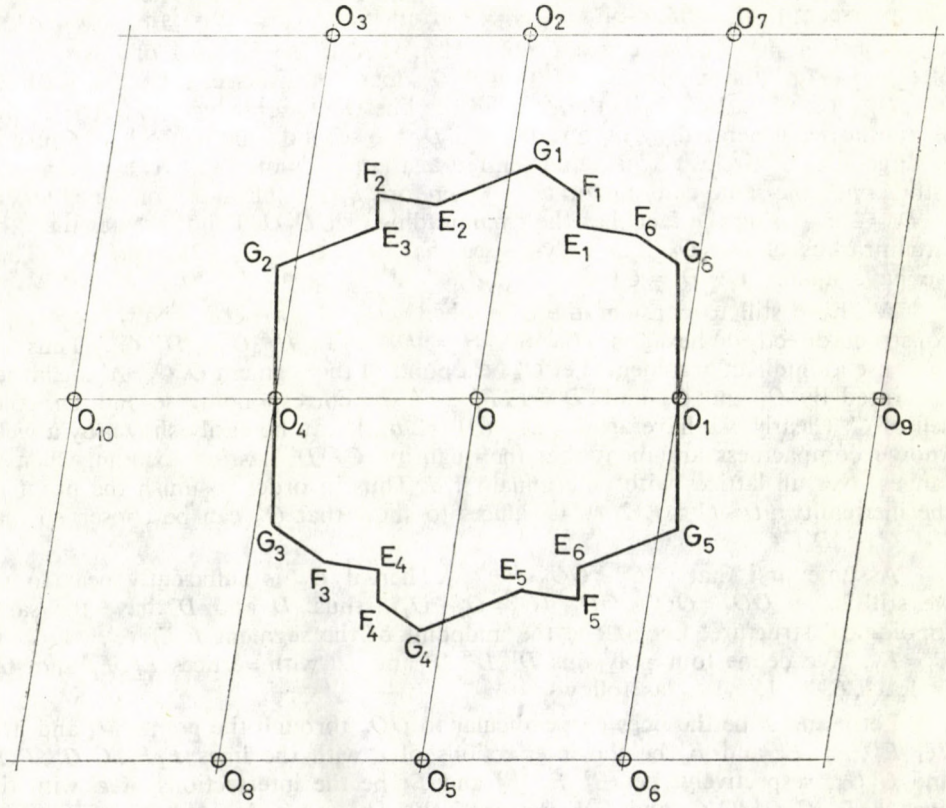


Fig. 1

It is easy to check that D is nothing else but the 18-gon $E_1F_1G_1E_2F_2E_3G_2G_3F_3E_4F_4G_4E_5F_5E_6G_5G_6F_6$. Obviously, all of the four quadruples of points (G_1, E_2, E_3, G_2), (G_3, F_3, F_4, G_4), (G_4, E_5, E_6, G_5) and (G_6, F_6, F_1, G_1) are collinear. Also it is clear that $OE_i = OE_j \leqq OF_i = OF_j \leqq OG_i = OG_j$ for any $i, j=1, \ldots, 6$. Further we observe that the triangles $E_1F_1F_6$, $E_2F_2E_3$, $F_3E_4F_4$ and

$E_5F_5E_6$ have the same area. Of course, some of the points E_i, F_i, G_i, $i=1, \ldots, 6$, may coincide. For $\sphericalangle O_1OO_2<\pi/2$ D is a 16-gon if $OO_1=OO_2<OO_3$, a 14-gon if $OO_1<OO_2=OO_3$ and a dodecagon if $OO_1=OO_2=OO_3$. If $\sphericalangle O_1OO_2=\pi/2$ then $E_1\equiv F_1\equiv F_6$, $E_2\equiv E_3\equiv F_2$, $E_4\equiv F_3\equiv F_4$ and $E_5\equiv E_6\equiv F_5$. In this case D is a hexagon which degenerates to a quadrangle if $OO_1=OO_2$.

We prove the inequality $|C\cap D|\leqq f_6(2/d)$ first in the case when $OF_i\leqq 1/\sqrt{\pi}=$ the radius of C. Now the triangles $E_1F_1F_6$, $E_2F_2E_3$, $F_3E_4F_4$ and $E_5F_5E_6$ are all contained in C. Let H denote the hexagon $G_1G_2G_3G_4G_5G_6$. Since H arises from D by cutting off the triangles $E_2F_2E_3$ and $E_5F_5E_6$ and adding the triangles $E_1F_1F_6$ and $F_3E_4F_4$, we have $|H|=|D|$ and $|C\cap D|=|C\cap H|$. We also observe that by property (i) we have $|D|=2/d$. Now the inequality $|C\cap D|=|C\cap H|\leqq f_6(2/d)$ follows immediately from the known fact (see e.g. [16] or [17]) that the area of the intersection of a circle of unit area and an n-gon of area x is at most $f_n(x)$.

Consider now the case when $OE_i\geqq 1/\sqrt{\pi}$. Let w be the area of the segment of C cut off by the line G_2G_3 ($w=0$ if G_2G_3 does not intersect C). Then we have $|C\cap D|=|C|-2w$. Let O_2' be the point of the line O_2O_3 for which $\sphericalangle O_1OO_2'=\pi/2$, Λ' the lattice generated by O_1 and O_2' and D' the second Dirichlet cell of C in the arrangement of circles of unit area centred at the points of Λ'. Λ' is a rectangular lattice with the same determinant as Λ. Therefore D' is a hexagon for which $|D'|=$ $=|D|=2/d$. Using the fact that the circumradius of OO_1O_2 is not greater than the circumradius of OO_1O_2' it is easy to see that $|C\cap D'|=|C|-2w=|C\cap D|$. Thus we have again $|C\cap D|=|C\cap D'|\leqq f_6(2/d)$.

We have still to consider the case when $OE_i<1/\sqrt{\pi}<OF_i$. Now we cannot construct directly a hexagon H with $|H|=|D|$ and $|H\cap C|>|D\cap C|$. Thus we shall use an indirect argument. Let O_2' be a point of the segment O_2O_3, Λ' the lattice generated by O_1 and O_2' and $D'=E_1'F_1'\ldots F_6'$ the corresponding second Dirichlet cell of C. Clearly, we have again $|D'|=|D|=2/d$. It can be easily shown by a well-known compactness argument that the quantity $|C\cap D|$ has a maximum when Λ ranges over all lattices with determinant $1/d$. Thus in order to finish the proof of the inequality $|D\cap C|\leqq f_6(2/d)$ it suffices to show that O_2' can be chosen so that $|D\cap C|<|D'\cap C|$.

Assume first that $OO_1<OO_2<OO_3$. Then if O_2' is sufficiently near to O_2 we still have $OO_1<OO_2'<OO_3'$ ($O_3'=O_2'-O_1$), thus D and D' have the same topological structure. Let M_i be the midpoint of the segment E_iE_{i+1}, $i=1, \ldots, 6$, $E_7=E_1$. We define four polygons D^1, D^2, D^3 and D^4 with vertices E_i^j, F_i^j and G_i^j, $i=1, \ldots, 6$, $j=1, \ldots, 4$, as follows:

Let e and e' be the lines perpendicular to OO_2' through the points M_1 and M_4. Let E_1^1, E_2^1, F_2^1 and F_6^1 be the intersections of e with the lines F_1F_5, G_1G_2, F_2F_4 and G_1G_6, respectively. Let E_4^1, E_5^1, F_3^1 and F_5^1 be the intersections of e' with the lines F_2F_4, G_4G_5, G_3G_4 and F_1F_6, respectively. The remaining vertices of D^1 coincide with the corresponding vertices of D.

Let f and f' be the lines perpendicular to OO_2' through the points $\frac{1}{2}O_2'$ and $-\frac{1}{2}O_2'$. We define D^2 in the same way as D^1 with f and f' instead of e and e'.

Let g and g' be the lines perpendicular to OO_3' through M_2 and M_5. Let E_2^3, E_3^3, G_1^3 and G_2^3 be the intersections of g with the lines $F_2^2F_6^2$, $F_2^2F_4^2$, $G_1^2G_6^2$ and

$G_2^2G_3^2$, respectively. Let E_5^3, E_6^3, G_4^3 and G_5^3 be the intersections of g' with the lines $F_3^2F_5^2$, $F_1^2F_5^2$, $G_3^2G_4^2$ and $G_5^2G_6^2$, respectively. Let the remaining vertices of D^3 coincide with the corresponding vertices of D^2.

Finally, we define D^4 in the same way as D^3 with the lines h and h' instead of g and g', where h and h' are the lines perpendicular to OO_3' through $\frac{1}{2}O_3'$ and $-\frac{1}{2}O_3'$.

Let the boundary of C intersect the line E_iE_{i+1} in the points $\underline{E}_i$ and $\bar{E}_{i+1}$ $(i=1, \ldots, 6, E_7=E_1, \bar{E}_7=\bar{E}_1)$ choosing the notations so that the order of the points on the line E_iE_{i+1} should be $\underline{E}_i, E_i, E_{i+1}, \bar{E}_{i+1}$. Write $E_1\underline{E}_1=E_2\bar{E}_2=r$, $E_2\underline{E}_2=$ $=E_3\bar{E}_3=s$, $O_2O_2'=x$, $\sphericalangle O_2OO_1=\alpha$ and $\sphericalangle O_2O_1O=\beta$. Then it is easily seen that

$$
\begin{aligned}
|C\cap D^1| &= |C\cap D|+O(x^2),\\
|C\cap D^2| &= |C\cap D^1|-2rx\cos\alpha+O(x^2),\\
|C\cap D^3| &= |C\cap D^2|+O(x^2),\\
|C\cap D^4| &= |C\cap D^3|+2sx\cos\beta+O(x^2),\\
|C\cap D'| &= |C\cap D^4|.
\end{aligned}
$$

Combining these relations we see that for sufficiently small values of x we have

$$|C\cap D'| = |C\cap D|+2x(s\cos\beta-r\cos\alpha)+O(x^2). \tag{6}$$

If $OO_1=OO_2$ or $OO_2=OO_3$ then displacing O_2 on the segment O_2O_3 new edges of D arise changing thereby the topological structure of D. (Even the order of the distances OO_1, OO_2 and OO_3 may change.) But if the displacement of O_2 is small, the new edges do not intersect C. Repeating the above argument we see that (6) remains valid also in this case.

If $OO_2<OO_3$ then we have $r<s$ and $\cos\alpha<\cos\beta$, i.e.,

$$s\cos\beta-r\cos\alpha>0.$$

In the case when $OO_2=OO_3$ we have $s\cos\beta-r\cos\alpha=0$. But if O_2 moves on the segment O_2O_3 in the direction of O_3 then OO_2 decreases and OO_3 increases so that $s\cos\beta-r\cos\alpha$ becomes positive. Thus (6) implies that if O_2' is sufficiently near to O_2 we have $|C\cap D'|>|C\cap D|$.

This completes the proof of the inequality $|C\cap D|\leqq f_6(2/d)$ and simultaneously the proof of the inequality $s_1+s_2\leqq df_6(2/d)$. The condition for the equality is that $G_1\ldots G_6$ is a regular hexagon which is equivalent with the condition described in Theorem 2.

To finish we mention some unsolved problems. As already noted Theorem 2 is a generalization of the result of A. HEPPES that $\bar{\delta}_2=2\bar{\delta}_1$ and the result of W. J. BLUNDON that $\bar{\Delta}_2=2\bar{\Delta}_1$. These results were generalized by DUMIR and HANS—GILL [8, 9] to arbitrary centro-symmetric convex domains: The density of a lattice double-packing of a centro-symmetric convex domain C cannot be greater than twice the density of the densest simple lattice-packing of C. Analogously, the density of a lattice double covering of C cannot be less than twice the density of the thinnest simple lattice-covering of C. It may be conjectured that these theorems can be united in a single theorem in the same way as the results of Heppes and Blundon are united in Theorem 2. However, the method used to deduce Theorem 2 does

not enable us to prove such a generalization of the theorems of Dumir and Hans-Gill.

Let ε_i, $i=1, 2, \ldots$ be a sequence of numbers from which, for sake of simplicity, we suppose that $\varepsilon_i=-1$, 0 or 1 and only finitely many ε_i's differ from O. What is the supremum of the sum $\sum_{i=1}^{\infty} \varepsilon_i s_i$ for all arrangements of equal circles with density d? We emphasize the following special case: $\varepsilon_1=1, \varepsilon_2=-1, \varepsilon_i=0$ for $i=3, 4, \ldots$. The quantity s_1-s_2 is nothing else but the density of the part of the plane covered exactly once. The problem of finding the maximum of this density for all arrangements of congruent circles is a longstanding problem ([15] p. 97). It is solved only under the additional condition that $s_2=s_3$ ([15] p. 98). With suitable choice of the sequence ε_i we can obtain other generalizations of known problems. So, e.g., it is easily seen that the problem of finding the supremum of the sum $\sum_{i=2}^{k+1} s_i$ for all arrangements of equal circles with density d unites the problem of the densest simple packing and the problem of the thinnest k-fold covering with equal circles.

REFERENCES

[1] BLUNDON, W. J.: Multiple covering of the plane by circles, *Mathematika* **4** (1957), 7—16.

[2] BLUNDON, W. J.: Multiple packing of circles in the plane, *J. London Math. Soc.* **38** (1963), 176—182.

[3] BLUNDON, W. J.: Note on a paper of A. Heppes, *Acta Math. Acad. Sci. Hungar.* **14** (1963), 317.

[4] BLUNDON, W. J.: Some lower bounds for density of multiple packing, *Canad. Math. Bull.* **7** (1964), 565—572.

[5] BLUNDON, W. J.: A three-fold non-lattice covering, *Canad. Math. Bull.* **20** (1977), 29—31.

[6] BLUNDON, W. J.: A nine-fold packing, *Acta Math. Acad. Sci. Hungar.* **32** (1978), 293—294.

[7] BOLLE, U.: *Mehrfache Kreisanordnungen in der Euklidischen Ebene*, Dissertation, Universität Dortmund (1976).

[8] DUMIR, V. C. and HANS-GILL, R. J.: Lattice double packings in the plane, *Indian J. Pure Appl. Math.* **3** (1972), 481—487.

[9] DUMIR, V. C. and HANS-GILL, R. J.: Lattice double coverings in the plane, *Indian J. Pure Appl. Math.* **3** (1972), 466—480.

[10] DANZER, L.: Drei Beispiele zu Lagerungsproblemen, *Arch. Math.* **11** (1960), 159—165.

[11] FEJES TÓTH, G.: Covering the plane by convex discs, *Acta Math. Acad. Sci. Hungar.* **23** (1972), 263—270.

[12] FEJES TÓTH, G.: Multiple packing and covering of the plane with circles, *Acta Math. Acad. Sci. Hungar.* **27** (1976), 135—140.

[13] FEJES TÓTH, G.: On a Dowker-type theorem of Eggleston, *Acta Math. Acad. Sci. Hungar.* **29** (1977), 131—148.

[14] FEJES TÓTH, G. and FLORIAN, A.: Mehrfache gitterförmige Kreis- und Kugelanordnungen, *Monatsh. Math.* **79** (1975), 13—20.

[15] FEJES TÓTH, L.: *Lagerungen in der Ebene, auf der Kugel und im Raum* (zweite Auflage), Springer-Verlag, Berlin—Heidelberg—New York, 1972.

[16] FEJES TÓTH, L.: On the isoperimetric property of the regular hyperbolic tetrahedra, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** A (1963), 53—57.

[17] HAJÓS, G.: Über den Durchschnitt eines Kreises und eines Polygons, *Anu. Univ. Sci. Budapest. Eötvös Sect. Math.* **11** (1968), 137—144.

[18] HEPPES, A.: Über mehrfache Kreislagerungen, *Elem. Math.* **10** (1955), 125—127.

[19] HEPPES, A.: Mehrfache gitterförmige Kreislagerungen in der Ebene, *Acta Math. Acad. Sci. Hungar.* **10** (1959), 141—148.

Mathematical Institute of the Hungarian Academy of Sciences

(Received October 31, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 457—469.

PRIMZAHLEN IN ARITHMETISCHEN PROGRESSIONEN UND EXPLIZITE FORMELN

von

H.-J. BESENFELDER

1. Einleitung

Man zeigt ganz elementar — analog dem euklidischen Beweis von der Unendlichkeit der Primzahlenmenge —, daß z. B. in den beiden arithmetischen Progressionen $4n-1$ und $4n+1$ jeweils unendlich viele Primzahlen liegen. Die entsprechende Aussage für eine beliebige Progression $an+b$, mit teilerfremden a und b, liegt dagegen sehr tief. Sie wurde zuerst von Dirichlet bewiesen und heißt seit dem „Dirichletscher Satz von den Primzahlen in arithmetischen Progressionen."

Der analytische Grund für sein Bestehen liegt in der Tatsache, daß die (zur jeweiligen Progression gehörende) Dirichletreihe $L(s, \chi)$ an der Stelle $s=1$ von Null verschieden ist. Die allermeisten der inzwischen gängigen Beweise dieser Eigenschaft von $L(s, \chi)$ haben die methodische Schwäche, zwischen „reellen" und „nicht reellen" Charakteren χ unterscheiden zu müssen. Dies gilt für den originalen Beweis von Dirichlet, der Hilfsmittel aus der algebraischen Zahlentheorie heranzieht (siehe etwa [8]), wie für mehr analytisch orientierte Beweise, siehe etwa [1], [2], [7].

Die schärfere Aussage, daß in jeder der $\varphi(a)$ primen Restklassen mod a grob gesagt „gleich viele" Primzahlen liegen, was z. B. durch

$$\lim_{s\to 1+0} \sum_{p\equiv b_1(a)} p^{-s} \Big/ \sum_{p\equiv b_2(a)} p^{-s} = 1$$

oder durch

$$\lim_{x\to\infty} \pi(x, b \bmod a)/\pi(x) = \frac{1}{\varphi(a)}$$

oder auch durch

$$\lim_{x\to\infty} \frac{\pi(x, b_1 \bmod a)}{\pi(x, b_2 \bmod a)} = 1$$

präziseren Sinn bekommt, erfordert noch weitergehende Untersuchungen und beruht auf der Tatsache, daß $L(s, \chi)$ nirgendwo auf der Geraden $\mathrm{Re}\,(s)=1$ verschwindet (vergl. [5], S. 107).

In der vorliegenden Arbeit wird gezeigt, wie man den Dirichletschen Satz — inclusive der Erkenntnis über die gleichartige Verteilung der Primzahlen über alle primen Restklassen — mit Hilfe von Expliziten Formeln methodisch einheitlich erhalten kann. Eine Trennung von reellen und nicht-reellen Charakteren fällt also nicht ins Gewicht. Zudem ist bemerkenswert, daß der Nachweis von $L(1+it, \chi)\neq 0$ (zum Zwecke einer schärferen Verteilungsaussage) mittels einer simplen Zusatzargumentation umgangen werden kann!

2. Voraussetzungen

2.1. Charaktere

Um den Dirichletschen Satz zu beweisen, genügt es z. B. zu zeigen, daß die Reihe

$$\sum_{\substack{p\equiv b \bmod a \\ p\leqq y}} \frac{1}{p}$$

über alle Primzahlen in der Restklasse $b \bmod a$ für wachsendes y divergiert — und dies für alle zu a primen b.

Eine erste Schwierigkeit hierbei liegt in der Aussonderung der Primzahlen p mit $p\equiv b \bmod a$. Man erreicht dieses durch Einführung einer speziellen zahlentheoretische Funktion, des „Charakters" $\chi \pmod a$. Ein Charakter läßt sich auffassen als Homomorphismus von der primen Restklassengruppe $\mathcal{R}_a^\times \bmod a$ in die komplexen Zahlen mit dem Betrag 1. Die Menge X der Charaktere wird selbst zu einer Gruppe, wenn man die Multiplikation in X durch

$$(\chi_1 \cdot \chi_2)(b) := \chi_1(b) \cdot \chi_2(b)$$

definiert. Das Einselement χ_0 in X mit der Eigenschaft $\chi_0(b)=1$, für alle $b\in\mathcal{R}_a^\times$, heißt Hauptcharakter. Die Werte dieser Charaktere sind stets a-te Einheitswurzeln, woraus sich schließen läßt, daß die Gruppe X ebenso wie $\mathcal{R}_a^\times$ genau $\varphi(a)$ (Eulersche φ-Funktion) verschiedene Elemente enthält. Wir benötigen für später die leicht zu gewinnenden Beziehungen:

$$\text{(A)} \quad \sum_{b\in\mathcal{R}_a^\times} \chi(b) = \begin{cases} \varphi(a), & \text{für} \quad \chi = \chi_0 \\ 0 & \text{sonst} \end{cases}$$

$$\text{(B)} \quad \sum_{\chi\in X} \chi(n) = \begin{cases} \varphi(a), & \text{für} \quad n \equiv 1 \bmod a \\ 0 & \text{sonst} \end{cases}$$

$$\text{(C)} \quad \sum_{\chi\in X} \chi(n)\cdot\bar{\chi}(m) = \begin{cases} \varphi(a), & \text{für} \quad n \equiv m \bmod a \quad \text{und} \quad n, m\in\mathcal{R}_a^\times \\ 0 & \text{sonst} \end{cases}$$

$\bar{\chi}$ ist hierbei der zu χ konjugiert komplexe Charakter.

Um die Isolierung der PZ vom Typ $p\equiv b \bmod a$ zu erreichen, wählt man in der Beziehung (C) $n=p$ und $m=b$, versieht dann beide Seiten mit dem Faktor $\frac{1}{p}$ und summiert schließlich über alle $p\leqq y$. Das liefert

$$\text{(2)} \qquad \sum_{p\leqq y} \sum_{\chi\in X} \chi(p)\cdot\bar{\chi}(b)\cdot\frac{1}{p} = \varphi(a)\cdot \sum_{\substack{p\leqq y \\ p\equiv b \bmod a}} \frac{1}{p},$$

also eine Gleichung, bei der die rechte Seite bis auf den Faktor $\varphi(a)$ genau der Summe in (1) entspricht. Wir werden eine ähnlich gebaute Summe untersuchen.

2.2. Die Dirichletschen L-Reihen

Erweitert man den Argumentbereich $\mathscr{R}_a^\times$ der Charaktere auf das vollständige Restsystem $\mathscr{R}_a$ durch die Festsetzung $\chi(b)=0$, für $b \notin \mathscr{R}_a^\times$ und definiert

$$\chi(n) := \chi(n \bmod a), \quad \text{für} \quad n \in \mathbf{N},$$

so gelangt man zu den „Dirichletschen Charakteren", mit denen sich die „Dirichletschen L-Reihen"

$$L(s, \chi) = \sum_{n \in \mathbf{N}} \frac{\chi(n)}{n^s}, \quad s = \sigma + it, \tag{3}$$

ergeben. $L(s, \chi)$ ist ersichtlich eine Verallgemeinerung der Riemannschen ζ-Funktion und bei beliebigen χ für $\mathrm{Re}\,(s)>1$ durch (3) definiert. Analog zur Eulerschen Identität

$$\sum_{n \in \mathbf{N}} \frac{1}{n^s} = \prod_{p \in \mathbf{P}} \frac{1}{1-p^{-s}}, \quad \text{für} \quad \mathrm{Re}\,(s) > 1,$$

hat man wegen der Multiplikativität von χ

$$L(s, \chi) = \sum_{n \in \mathbf{N}} \frac{\chi(n)}{n^s} = \prod_{p \in \mathbf{P}} \frac{1}{1-\chi(p)\,p^{-s}},$$

für $\mathrm{Re}\,(s)>1$. Es ist hier, wegen $\chi(p)=0$ für alle primen p, die den Modul a teilen, gleichgültig, ob das Produkt über alle $p \in \mathbf{P}$ läuft, oder nur über solche p, die relativ prim zu a sind. Für den Hauptcharakter $\chi_0 \pmod a$ hat man

$$L(s, \chi_0) = \prod_{\substack{p \in \mathbf{P} \\ p \nmid a}} (1-\chi(p)\,p^{-s})^{-1} = \zeta(s) \cdot \prod_{p \mid a} (1-p^{-s}),$$

woraus man entnimmt, daß $L(s, \chi_0)$ zur Zetafunktion analoge analytische Eigenschaften besitzt.

2.3. Explizite Formeln

Wir gehen aus von der Expliziten Formel

$$\begin{aligned} \lim_{T \to \infty} \sum_{\substack{\varrho = \sigma + i\gamma \\ |\gamma| < T}} M(\varrho) &= \varepsilon_0 \{M(0)+M(1)\} + F(0) \log \frac{f}{\pi} + \\ &+ \sum_p \sum_{n=1}^{\infty} \log p \cdot p^{-\frac{n}{2}} \{\chi(p^n) F(\log p^n) + \chi(p^{-n}) F(\log p^{-n})\} + \\ &- CF(0) + \text{v.p.} \int_{-\infty}^{\infty} \frac{F(x) \cdot e^{\left(\frac{3}{2}-\delta\right)|x|} - F(0)}{1-e^{2|x|}}\, dx, \end{aligned} \tag{4}$$

welche für eine gewisse Klasse von Funktionen F und deren Mellin-Transformierte M gilt [3]. In dieser Formel wird auf der linken Seite über die im kritischen Streifen liegenden Nullstellen ϱ von $L(s, \chi)$ summiert (nach der angezeigten Vorschrift); die Doppelsumme auf der rechten Seite läuft über alle Primzahlen p und über die

natürlichen Zahlen $\mathbf{N}$; die Konstante C ist die ‚Eulersche Konstante'; weiter ist

$$\varepsilon_0 = \varepsilon_0(\chi) = \begin{cases} 1 & \text{für } \chi = \chi_0 \\ 0 & \text{für } \chi \neq \chi_0 \end{cases}$$

und

$$\delta = \delta(\chi) = \begin{cases} 1 & \text{für } \chi(-1) = -1 \\ 0 & \text{für } \chi(-1) = 1; \end{cases}$$

f ist der „Führer" des Charakters χ. Da der Term $F(0)\log\frac{f}{\pi}$ in den folgenden Betrachtungen aus Gründen seiner Größenordnung von unbedeutendem Einfluß ist, kann auf eine weitere Erläuterung der Eigenschaften des Führers von χ hier verzichtet werden. Das Kürzel „v. p." (valor principalis) deutet an, daß im Falle einer Sprungstelle von F bei 0 der Hadamardsche Hauptwert des Integrals zu nehmen ist. In unserem Falle wird jedoch die in (4) konkret einzurechnende Funktion F bei 0 stetig sein und daher wird das „v. p." einfach fortgelassen werden.

2.4. Die Nullstellen von $L(s, \chi)$

Die Lage der Nullstellen ϱ von $L(s, \chi)$, über die auf der linken Seite der Formel (4) summiert wird, ist von entscheidender Bedeutung für das Dirichletsche Resultat.

In Analogie zur Riemannschen ζ-Funktion und ihrer Funktionalgleichung, welche z. B. in der symmetrischen Form

$$\xi(s) = \xi(1-s)$$

mit

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s)$$

gegeben werden kann, haben auch die L-Reihen eine Funktionalgleichung, und zwar gilt für (primitives) $\chi \neq \chi_0$,

$$\xi(s, \chi) = \eta_\chi \xi(1-s, \bar{\chi}),$$

wobei

$$\xi(s, \chi) = \left(\frac{f}{\pi}\right)^{(s+\delta)/2}\Gamma\left(\frac{s+\delta}{2}\right)L(s, \chi)$$

ist und η_χ eine nur vom Charakter χ abhängende Konstante vom Betrag 1 ist [6]; f und δ haben die Bedeutung wie in 2.3. erläutert. Für die Nullstellen von $\xi(s, \chi)$ gelten analoge Aussagen wie für jene von $\xi(s)$, jedoch mit dem Unterschied, daß sie im kritischen Streifen $(0 \leqq \mathrm{Re}\,(s) \leqq 1)$ zwar symmetrisch zu $\sigma = \frac{1}{2}$ sind, aber nicht unbedingt symmetrisch zur reellen Achse liegen müssen. Leider wissen wir i.A. wegen der Vielzahl der Charaktere über diese Nullstellen viel weniger als über jene der Zetafunktion, weshalb wir feinere Untersuchungen über $\sum_{\varrho} M(\varrho)$

vornehmen werden müssen. Dabei werden wir folgende Tatsache über ihre Verteilung verwenden:

(D) Die Anzahl der Nullstellen von $\xi(s, \chi)$ im Rechteck $0 \leqq \sigma \leqq 1$ und $T \leqq |\gamma| \leqq T+1$ ist von der Größenordnung

$$O(\log T) \quad \text{für} \quad T \to \infty \quad [6].$$

3. Der Dirichletsche Satz

3.1. Eine spezielle explizite Formel

Wir gehen mit der (zulässigen) Funktion

$$F(x) = e^{(c+1/2)x - x^2/4y}$$

mit den Parametern $y>0$ reell, c komplex, in die Formel (4) und erhalten zu jedem $\chi \in X$ speziell

$$\begin{aligned} 2\sqrt{\pi y} \sum_{\varrho(\chi)}^{*} e^{y(\varrho+c)^2} = \varepsilon_0 \left\{2\sqrt{\pi y}\left(e^{yc^2} + e^{y(1+c)^2}\right)\right\} + \log\frac{f}{\pi} + \\ - \sum_{p,n} \log p \cdot p^{nc} e^{-\log^2(p^n)/4y} \chi(p^n) + \\ - \sum_{p,n} \log p \cdot p^{-n(1+c)} e^{-\log^2(p^n)/4y} \chi(p^{-n}) + W_{\delta(\chi)} \end{aligned} \tag{5}$$

mit

$$W_{\delta(\chi)} = -C + \int_0^\infty \frac{e^{-x^2/4y + (1-\delta-c)x} - 1}{1-e^{2x}}\, dx + \int_0^\infty \frac{e^{-x^2/4y + (2-\delta-c)x} - 1}{1-e^{2x}}\, dx.$$

Der Stern * am Summenzeichen erinnert an die Summationsvorschrift. Nun multiplizieren wir (5) mit $\bar{\chi}(b)$ und summieren alle diese Gleichungen über die Charaktere $\chi \in X$ auf, das ergibt

$$\begin{aligned} 2\sqrt{\pi y} \sum_{\chi \in X} \left(\bar{\chi}(b) \sum_{\varrho(\chi)}^{*} e^{y(\varrho+c)^2}\right) = 2\sqrt{\pi y}\left\{e^{yc^2} + e^{y(1+c)^2}\right\} + \sum_{\chi \in X} \bar{\chi}(b) \log\frac{f}{\pi} + \\ - \varphi(a) \sum_{p^n \equiv b \bmod a} \log p \cdot p^{nc} e^{-\log^2(p^n)/4y} + \\ - \varphi(a) \sum_{p^n \equiv \frac{1}{b} \bmod a} \log p \cdot p^{-n(1+c)} e^{-\log^2(p^n)/4y} + \sum_{\chi} \bar{\chi}(b) W_{\delta(\chi)}. \end{aligned} \tag{6}$$

Wir werden nun das Wachstumsverhalten der beiden Seiten von (6) für $y \to \infty$ untersuchen. Dabei wird sich zeigen, daß — bei geeigneter Wahl des Parameters c — das Verhalten der linken Seite von (6) die Divergenz der Summen $\sum_{p^n \equiv b}$, $\sum_{p^n \equiv \frac{1}{b}}$ erzwingt, woraus wir die Unendlichkeit der Primzahlenmenge $p \equiv b \bmod a$ für jedes b aus $\mathcal{R}_a^\times$ ableiten werden. Insbesondere wird bei diesem Vorgehen der Zusammenhang dieser Primzahlverteilung mit der Tatsache $L(1, \chi) \neq 0$ für alle χ deutlich hervortreten.

Wir wählen zunächst $c=-\frac{1}{2}$ und $b=1$, dann wird aus (6):

$$2\sqrt{\pi y}\sum_{\chi}\sum_{\varrho(\chi)}^{*} e^{y\left(\varrho-\frac{1}{2}\right)^2}=4\sqrt{\pi y}e^{y/4}+\sum_{\chi}\log\frac{f}{\pi}+$$

$$-2\varphi(a)\sum_{p^n\equiv 1}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y}+\sum_{\chi}W_{\delta(\chi)}. \tag{7}$$

Zur Vereinfachung dieser Formel (7) stellen wir folgendes fest:

(a) Es ist jedes $W_{\delta(\chi)}$ $\left(\text{bei } c=-\frac{1}{2}\right)$ beschränkt für $y\to\infty$. Man weist dies mit elementarer Abschätzung nach; vergleiche dazu etwa [4], wo dieser Formeltyp zu anderem Zweck benutzt worden ist.

(b) $\sum_{\chi}\log\frac{f}{\pi}$ ist eine feste Zahl, also für $y\to\infty$ beschränkt.

(c) Die Imaginärteile sind auf beiden Seiten identisch Null.

Es wird daher aus (7) für großes y:

$$2\sqrt{\pi y}\sum_{\chi}\sum_{\varrho(\chi)}^{*} e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}}=4\sqrt{\pi y}e^{y/4}+$$

$$-2\varphi(a)\sum_{p^n\equiv 1 \bmod a}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y}+O(1). \tag{8}$$

Wir zeigen nun, daß die linke Seite ein geringeres Wachstum besitzt als $4\sqrt{\pi y}\,e^{y/4}$, woraus sich zunächst auf die Unendlichkeit der Primzahlenmenge von Typ 1 mod a schließen läßt.

3.2. Das Nichtverschwinden von $L(1,\chi)$ für $\chi\neq\chi_0$

Wir erinnern daran, daß $L(1,\chi_0)$ unbeschränkt ist, daher sei $\chi\neq\chi_0$. Ein Wachstum der Summe

$$2\sqrt{\pi y}\sum_{\chi}\sum_{\varrho(\chi)}^{*} e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}} \tag{9}$$

von der Größenordnung $4\sqrt{\pi y}\,e^{y/4}$ für großes y ist wegen der Dichte der Nullstellen ϱ (siehe (D) in Abschnitt 2.4) nur möglich, falls es ein ϱ mit $\varrho=1$ gibt. Die Annahme $L(1,\chi)=0$ für einen komplexen Charakter können wir angesichts der Gleichung (8) als „trivial" abhaken. Mit $L(1,\chi)=0$ ist nämlich auch $L(1,\bar{\chi})=0$ und da wegen der Symmetrie der Nullstellen dann auch $L(0,\chi)=0$ bzw. $L(0,\bar{\chi})=0$ ist, wird das Wachstum von (9) mindestens gleich $8\sqrt{\pi y}e^{y/4}$, was von der rechten Seite nicht erbracht werden kann. Daher ist für komplexe Charaktere stets $L(1,\chi)\neq 0$.

Ein entsprechendes Argument zeigt sogleich, daß es unter allen $\chi\in X$ überhaupt nur einen (reellen) Charakter geben kann mit einer derartigen, dann einfachen Nullstelle. Diese Einsicht verwenden wir nun weiter.

Annahme: Es gebe ein reelles χ mit $L(1, \chi)=0$.

Die Formel (8) entstand aus (6) durch Wahl des Parameters $c=-\frac{1}{2}$ bei $b=1$. Da nach unserer bisherigen Kenntnis die als existent angenommene Nullstelle bei 1 (bzw. bei 0) einfach ist, erscheint in (6) links der Term

$$2\sqrt{\pi y}\{e^{yc^2}+e^{y(1+c)^2}\},$$

der sich ohne Einschränkung bezüglich c stets gegen den ersten Term auf der rechten Seite von (6) weghebt. Wir erhalten demnach bei der Wahl $c=-\frac{1}{2}+it$, $t\in\mathbf{R}$, aus (8):

$$2\sqrt{\pi y}\sum_{\chi}\sum_{\substack{\varrho(\chi)\\ \varrho\neq 1,0}}{}^{*} e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-(\gamma-t)^2\right\}}\cos\left\{2y\left(\sigma-\frac{1}{2}\right)(\gamma-t)\right\}= \tag{10}$$

$$=-2\varphi(a)\sum_{p^n\equiv 1 \bmod a}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y}\cos\{t\log p^n\}+O(1).$$

Die Abschätzungen, welche die übrigen Terme betreffen, bleiben dieselben wie bei $c=-\frac{1}{2}$ und sind in $O(1)$ für $y\to\infty$ zusammengefaßt.

Die Summe links spalten wir nun in drei Summen $\sum_1^*$, $\sum_2^*$, $\sum_3^*$ gemäß folgender Überlegung auf:

Wir betrachten eine „Fliege" $\mathcal{F}$ im kritischen Streifen, wie die Abb. 1 zeigt Für jede Nullstelle ϱ, die *außerhalb* dieser Fliege liegt, gilt

$$|\mathrm{Im}\,(\varrho)|>\left|\mathrm{Re}\,(\varrho)-\frac{1}{2}\right|,$$

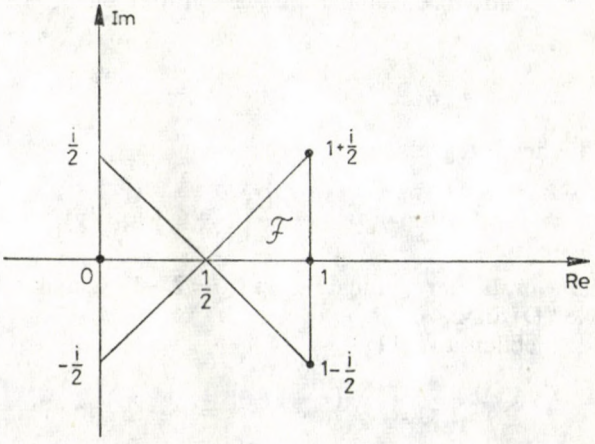


Abb. 1

woraus wir entnehmen, daß im Fall $t=0$ $\left(\text{also } c=-\frac{1}{2}\right)$ der zugehörige einzelne Summand

$$2\sqrt{\pi y}e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}}\cos\left\{2y\left(\sigma-\frac{1}{2}\right)\gamma\right\} \quad \text{für wachsendes } y$$

„exponentiell" fällt. Alle von solchen Nullstellen herkommenden Terme werden in $\sum_1^*$ aufsummiert.

LEMMA 1. *Es gilt* $\sum_1^* \to 0$ *für* $y \to \infty$.

BEWEIS.

$$\left|\sum_1^* 2\sqrt{\pi y}e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}}\cos\left\{2y\left(\sigma-\frac{1}{2}\right)\gamma\right\}\right| \leqq \sum_1^* 2\sqrt{\pi y}e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}}.$$

Sei $d>0$ der minimale Abstand dieser Nullstellen zum Rand von $\mathscr{F}$. Dann gilt für diese ϱ die Beziehung

$$|\gamma| \geqq \left|\sigma-\frac{1}{2}\right|+d$$

woraus

$$\gamma^2 > \left(\sigma-\frac{1}{2}\right)^2+d^2$$

folgt. Demnach ist

$$\sum_1^* 2\sqrt{\pi y}e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}} = e^{-\frac{d^2}{2}y}\sum_1^* 2\sqrt{\pi y}e^{y\left\{\frac{d^2}{2}+\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}} \tag{11}$$

mit negativen Exponenten der Terme in $\sum_1^*$. Es soll nun die Summe rechts unabhängig von σ abgeschätzt werden. Dazu ersetzen wir durchweg σ durch 1 und lassen die endlich vielen Nullstellen ϱ mit $\gamma^2 \leqq \frac{d^2}{2}+\frac{1}{4}$ unberücksichtigt, da ihr Term im Betrag in y monoton fällt, ihre Summe also beschränkt bleibt. Den Rest

$$\sum_{\gamma^2 > \frac{d^2}{2}+\frac{1}{4}}^{*}{}_1 2\sqrt{\pi y}e^{y\left\{\frac{d^2}{2}+\frac{1}{4}-\gamma^2\right\}}$$

dürfen wir mit Hilfe des Integrals

$$\int_0^\infty e^{-yw^2}\log w\,dw = -\frac{1}{4}\sqrt{\frac{\pi}{y}}\{C+\log 4y\}$$

abschätzen, da die Anzahl der ϱ mit $T \leqq \mathrm{Im}(\varrho) < T+1$ von der Größenordnung $O(\log T)$ ist (siehe (D) in 2.4).

Es wird also schließlich aus (11)

$$e^{-\frac{d^2}{2}y}\sum_1^* 2\sqrt{\pi y}e^{y\left\{\frac{d^2}{2}+\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}} = e^{-\frac{d^2}{2}y}O(\log y).$$

Für $y \to \infty$ verschwindet daher $\sum_1^*$, wie behauptet.

Die $\sum_2^*$ enthalte alle diejenigen Summanden, die von Nullstellen ϱ auf dem Rand der Fliege $\mathscr{F}$ herrühren. Für diese endlich vielen gilt bei $t=0$:

$$|\mathrm{Im}(\varrho)| = \mathrm{Re}(\varrho) - \frac{1}{2}$$

und es wird

$$\sum_2^* = \sum_{\substack{\varrho \text{ auf den Rand} \\ \text{von } \mathscr{F},\ \sigma \geqq 1/2}} 2\sqrt{\pi y}\cos(2y\gamma^2).$$

Diese Summe ist entweder identisch Null oder wächst wie $\sqrt{y}$, eventuell mit wechselnden Vorzeichen, die der Kosinusterm regelt.

In der Summe $\sum_3^*$ schließlich stecken alle Terme von Nullstellen $\varrho \neq 1$ innerhalb $\mathscr{F}$. Sie ist entweder identisch Null oder wächst — eventuell auch mit wechselnden Vorzeichen — höchstens wie $e^{\varepsilon y}$, wobei $0 < \varepsilon < \frac{1}{4}$ ist, da im Exponenten von

$$e^{y\left\{\left(\sigma-\frac{1}{2}\right)^2-\gamma^2\right\}}$$

der Ausdruck $\left(\sigma-\frac{1}{2}\right)^2$ echt größer ist als γ^2, aber unter $\frac{1}{4}$ bleibt.

Fassen wir diese Untersuchungen zusammen, dann bleibt aus (10) für $t=0$ stehen

$$\sum_1^* + \sum_2^* + \sum_3^* = -2\varphi(a) \sum_{p^n \equiv 1 \bmod a} \log p \cdot p^{-\frac{n}{2}} e^{-\log^2(p^n)/4y} + O(1). \tag{12}$$

Die linke Seite haben wir „fest im Griff", daher wenden wir uns der rechten Seite zu. Wir nehmen an, daß die Summe $\sum_{p^n \equiv 1 \bmod a}$ für $y \to \infty$ divergiert. Dann wird die rechte Seite von (12) für genügend großes y negativ bleiben und gegen $-\infty$ gehen. Da $\sum_1^*$ verschwindet und die überhaupt als divergend in Frage kommenden Summen $\sum_2^*$ und $\sum_3^*$ fortwährend ihr Vorzeichen ändern, oder aber gegen $+\infty$ gehen (falls es in $\sum_2^*$, $\sum_3^*$ ein entsprechendes ϱ mit $\gamma = 0$ gibt), widerspricht dieses Verhalten dem der rechten Seite.

Nehmen wir andererseits an, die Summe rechts konvergiert für $y \to \infty$, dann gilt dies auch für die linke Seite und es kann überhaupt keine Nullstelle in und auf dem Rand der Fliege geben! Wegen

$$\left|-2\varphi(a) \sum_{p^n \equiv 1 \bmod a} \log p \cdot p^{-\frac{n}{2}} e^{-\log^2(p^n)/4y} \cos(t \log p^n)\right| \leqq$$

$$\leqq 2\varphi(a) \sum_{p^n \equiv 1} \log p \cdot p^{-\frac{n}{2}} e^{-\log^2(p^n)/4y}$$

bleibt die gesamte rechte Seite von (12) bei beliebigen t beschränkt. Mit dieser Kenntnis wählen wir dann t derart, daß es mit der kleinsten überhaupt auftretenden Ordinate $\tilde{\gamma} \neq 0$ einer Nullstelle $\tilde{\varrho} = \tilde{\sigma} + i\tilde{\gamma}$ zusammenfällt:

$$t = \tilde{\gamma}.$$

Der Term

$$2\sqrt{\pi y}e^{y\left\{\left(\tilde{\sigma}-\frac{1}{2}\right)^2-(\tilde{\gamma}-t)^2\right\}}\cdot\cos\left\{2y\left(\tilde{\sigma}-\frac{1}{2}\right)(\tilde{\gamma}-t)\right\}$$

wird dann gleich

$$2\sqrt{\pi y}e^{y\left(\tilde{\sigma}-\frac{1}{2}\right)^2}$$

und geht mit wachsendem y nach unendlich.

Der „Rest" der $\sum^*$ geht wieder gegen 0 oder wächst sogar exponentiell (eventuell mit wechselnden Vorzeichen), falls etwa ganz in der „Nähe" von $\tilde{\varrho}$ noch eine oder mehrere Nullstellen liegen, für die

$$\left(\sigma-\frac{1}{2}\right)^2>(\gamma-\tilde{\gamma})^2$$

gilt. Es führt also die Annahme der Konvergenz von $\sum_{p^n\equiv 1 \bmod a}$ als auch die Annahme ihrer Divergenz zu Widersprüchen. Wir können daher zu dem Schluß kommen, daß die Annahme

$$L(1,\chi)=0$$

für kein $\chi \bmod a$ möglich ist.

3.3. Die Menge der Primzahlen vom Typ $p\equiv 1 \bmod a$

Aus dem eben abgeleiteten, folgt sogleich aus (8) mit $t=0$, daß für großes y

$$o(e^{y/4})=4\sqrt{\pi y}e^{y/4}-2\varphi(a)\sum_{p^n\equiv 1 \bmod a}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y} \tag{13}$$

ist, also

$$\varphi(a)\sum_{p^n\equiv 1 \bmod a}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y}\sim 2\sqrt{\pi y}e^{y/4} \tag{14}$$

gilt.

Um uns von den höheren Primzahlpotenzen zu befreien, schätzen wir die Summe $\sum_{\substack{p^n\equiv 1 \bmod a\\ n\geqq 2}}$ folgendermaßen ab.

Die elementaren Beziehungen

$$\sum_{p\leqq x}\frac{\log p}{p}=\log x+O(1)\quad\text{für großes } x \tag{15}$$

und

$$\sum_{p,m}\frac{\log p}{p^{ms}}=-\frac{\zeta'}{\zeta}(s)<\infty\quad\text{für}\quad \mathrm{Re}\,(s)>1 \tag{16}$$

seien dazu als bekannt vorausgesetzt ([7], S. 36, 31).

Für $n\geqq 3$ ist aus (7)

$$2\varphi(a)\sum_{\substack{p^n\equiv 1 \bmod a\\ n\geqq 3}}\log p\cdot p^{-\frac{n}{2}}e^{-\log^2(p^n)/4y}\leqq k\sum_{\text{alle } p,n}\frac{\log p}{p^{\frac{3}{2}n}}<\infty$$

nach (16) mit einer geeigneten Konstanten k bei großem y. Dieser Teil der Summe bleibt also $O(1)$ für $y \to \infty$.

Für $n=2$ wird das Endstück

$$\sum_{\substack{p^2 \equiv 1 \bmod a \\ p^2 \geqq e^{y/2}}} \frac{\log p}{p} e^{-\log^2(p^2)/4y}$$

nach oben durch die feste Zahl

$$\sum_{\text{alle } p^2} \frac{\log p}{p} p^{-\frac{1}{4}} < \infty,$$

nach (16), abgeschätzt. Das Anfangsstück

$$\sum_{\substack{p^2 \equiv 1 \bmod a \\ p^2 < e^{y/2}}} \frac{\log p}{p} e^{-\log^2(p^2)/4y}$$

wächst nach (15) höchstens so schnell wie $\frac{y}{2}$, denn es ist

$$\sum_{\text{alle } p < e^{y/2}} \frac{\log p}{p} = \frac{y}{2} + O(1).$$

Zusammengenommen hat die Teilsumme über die höheren Primzahlpotenzen höchstens das Wachstum $O(y)$ für $y \to \infty$. Es ist daher (14) gleichwertig mit der Beziehung

$$\varphi(a) \sum_{p \equiv 1 \bmod a} \log p \cdot p^{-\frac{1}{2}} e^{-\log^2 p/4y} \sim 2\sqrt{\pi y} e^{y/4}, \tag{17}$$

woraus sich die Unendlichkeit der Primzahlenmenge in der Restklasse $1 \bmod a$ ablesen läßt.

3.4. Die Primzahlen in den übrigen Restklassen

Das 3.3. entsprechende Resultat für eine beliebige Restklasse $b \bmod a$ in $\mathscr{R}_a^\times$ erhalten wir aus der Formel (6). Zunächst ergibt sich gemäß den vorangegangenen Überlegungen eine Beziehung der Art

$$\begin{aligned} o(e^{y/4}) = 4\sqrt{\pi y} e^{y/4} - \\ -\varphi(a) \sum_{p \equiv b \bmod a} \log p \cdot p^{-\frac{1}{2}} e^{-\log^2 p/4y} - \\ -\varphi(a) \sum_{p \equiv \frac{1}{b} \bmod a} \log p \cdot p^{-\frac{1}{2}} e^{-\log^2 p/4y} \end{aligned}$$

(entsprechend (13)).

Hieraus entnehmen wir sofort, daß in wenigstens einer der beiden Restklassen $b \bmod a$, $\frac{1}{b} \bmod a$ unendlich viele Primzahlen liegen müssen. Dieses „symmet-

rische" Resultat beruht darauf, daß wir den komplexen Parameter c speziell gleich $-\frac{1}{2}$ gewählt haben. Nun sind aber in (6) alle Terme bis auf die fraglichen Summen $\sum_{p\equiv b}$, $\sum_{p\equiv\frac{1}{b}}$ invariant unter der Substitution $c\to -1-c$. Wir wählen jetzt z.B. $c=-\frac{1}{4}$. Wären in einer der beiden Summen, also etwa in $\sum_{p\equiv b}$ nur endlich viele Glieder, so müßte die andere Summe, hier also $\sum_{p\equiv\frac{1}{b}\bmod a} \log p\cdot p^{-(1+c)}e^{-\log^2 p/4y}$, für die Werte $c_1=-\frac{1}{4}$ und $c_2=-1-c_1=-\frac{3}{4}$ dasselbe (exponentielle) Wachstum haben, was unmöglich ist.

Diese Zusatzüberlegung zeigt, daß in jeder der beiden Restklassen $b \bmod a$ bzw. $\frac{1}{b} \bmod a$ unendlich viele Primzahlen liegen. Da b beliebig aus $\mathscr{R}_a^\times$ wählbar ist, folgt das Dirichletsche Resultat in voller Allgemeinheit.

4. Die Verteilung der Primzahlen auf die Restklassen

Die weitergehende Frage, „wieviele" PZ in einer jeden primen Restklasse mod a, gemessen an der Anzahl aller PZ, liegen, läßt sich aus unseren Untersuchungen fast unmittelbar entnehmen.

Aus (5) erhalten wir bei $\chi=\chi_0$ und $c=-\frac{1}{2}$, entsprechend (17), die asymptotische Beziehung

$$\sum_{\text{alle } p} \log p\cdot p^{-1/2}e^{-\log^2 p/4y} \sim 2\sqrt{\pi y}e^{y/4}$$

und aus (6) bei $c=-\frac{1}{2}$ (mit obiger Zusatzüberlegung), ebenfalls nach (17), die Beziehung

$$\varphi(a)\sum_{p\equiv b \bmod a} \log p\cdot p^{-1/2}e^{-\log^2 p/4y} \sim 2\sqrt{\pi y}e^{y/4}.$$

Über den Quotienten

$$\lim_{y\to\infty} \frac{\sum_{p\equiv b \bmod a} \log p\cdot p^{-1/2}e^{-\log^2 p/4y}}{\sum_{\text{alle } p} \log p\cdot p^{-1/2}e^{-\log^2 p/4y}} = \frac{1}{\varphi(a)}$$

erreichen wir die Einsicht, daß die Primzahlen über alle Restklassen in $\mathscr{R}_a^\times$ gleichartig verteilt sind. Grob gesagt, es befinden sich in jeder primen Restklasse $b \bmod a$ gleichviele PZ, ihr Anteil beträgt jeweils $\frac{1}{\varphi(a)}$ von allen.

BEMERKUNG. Obwohl diese Summen über alle, bzw. gewisse Primzahlen bis unendlich laufen, muß man de facto nur bis $p=\exp(2y)$ summieren, um eine

Größenordnung der Art $2\sqrt{\pi y}\exp(y/4)$ zu erreichen. Für $p>\exp(2y)$ läßt sich das Endstück der Summe mit

$$\sum_p \frac{\log p}{p^{1+\varepsilon}} < \infty, \quad \varepsilon > 0,$$

vergleichen.

LITERATUR

[1] APOSTOL, M.: *Introduction to Analytic Number Theory*, Berlin/Heidelberg, 1976.
[2] AYOUB, R.: *Introduction to The Analytic Theory of Numbers*, Providence, R. I., 1963.
[3] BESENFELDER, H.-J.: Die Weilsche Explizite Formel und temperierte Distrlbutionen, *J. reine angew. Math.* **293/294** (1977), 228—257.
[4] BESENFELDER, H.-J., PALM, G.: Einige Äquivalenzen zur Riemannschen Vermutung, *J. reine angew. Math.* **293/294** (1977), 109—115.
[5] INGHAM, A. E.: *Distribution of Prime Numbers*, New York, 1971.
[6] PRACHAR, K.: *Primzahlverteilung*, Berlin, Springer, 1957.
[7] SCHWARZ, W.: *Einführung in Methoden und Ergebnisse der Primzahltheorie*, Mannheim, Bibliographisches Inst., 1969.
[8] WEYL, H.: *Algebraische Zahlentheorie*, Mannheim, Bibliographisches Inst., 1966.

FB 6 Mathematik/Philosophie der Universität
4500 Osnabrück, West Germany

(Eingegangen am 12. Dezember, 1978)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 471—475.

ON THE ESTIMATION OF REGRESSION COEFFICIENT IN CASE OF AN AUTOREGRESSIVE NOISE PROCESS

by

I. H. GAUDI

Abstract

There is given a simple example illustrated by numerical results which points out the defect of the conditional maximum likelihood estimation method.

Introduction

In statistical time series analysis one of the most frequently discussed problem has the following formulation: a time series of the form

$$y(t) = m(t)+x(t), \qquad t = 1, 2, \ldots, N$$

is observed, where $m(t)$ is an unknown deterministic function and $x(t)$ is a stochastic process with 0 mean and known spectrum. The purpose is to draw some conclusions for $m(t)$ from the observed process $y(t)$. In the practice we seek the function $m(t)$ in the form

$$m(t) = \sum_{v=1}^{k} a_v \varphi^{(v)}(t),$$

where a_v are unknown coefficients and $\varphi^{(v)}(t)$ are known functions (usually polynomials or trigonometric polynomials). We have to estimate the coefficients a_v. The most natural way is the method of least squares.

Making use of the following notations

$$\alpha = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}, \quad y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}, \quad \varphi^{(j)} = \begin{pmatrix} \varphi^{(j)}(1) \\ \varphi^{(j)}(2) \\ \vdots \\ \varphi^{(j)}(N) \end{pmatrix}$$

and

$$\Phi = (\varphi^{(1)}, \varphi^{(2)}, \ldots, \varphi^{(k)}),$$

the least square estimator $\hat{\alpha}$ of the vector α takes the form

$$\hat{\alpha} = (\Phi^* \Phi)^{-1} \Phi^* y.$$

In the case of normal white noise the estimator $\hat{\alpha}$ coincides with the maximum likelihood estimator of the vector α. If we suppose, that the noise process $x(t)$ is normal, but not white and it has known correlation matrix R, we have the maximum likelihood estimator α_0 of the vector α in the form

$$\alpha_0 = (\Phi^* R)^{-1} \Phi^* R^{-1} y.$$

It is well-known, that α_0 has minimal dispersion among the linear unbiased estimators of α. From the point of view of computational technics of the inversion of the matrix R for enormously large N is a difficult problem. Both the estimators $\hat{\alpha}$ and α_0 are normally distributed (as linear combinations of Gaussian variables) with expectations and variances

$$E\hat{\alpha} = (\Phi^*\Phi)^{-1}\Phi^* Ey = (\Phi^*\Phi)^{-1}\Phi^*\Phi\alpha = \alpha,$$

$$E(\hat{\alpha}-\alpha)(\hat{\alpha}-\alpha)^* = (\Phi^*\Phi)^{-1}\Phi^* R\Phi(\Phi^*\Phi)^{-1},$$

$$E\alpha_0 = (\Phi^* R^{-1}\Phi)^{-1}\Phi^* R^{-1}\Phi\alpha = \alpha,$$

$$E(\alpha_0-\alpha)(\alpha_0-\alpha)^* = (\Phi^* R^{-1}\Phi)^{-1}.$$

If the noise process is of autoregressive type with known parameters then another simple method — the so called conditional maximum likelihood estimation — is widely used (see ARATÓ, 1978; BRILLINGER, 1973). The purpose of our comparative study is to point out the limits of the applicability of the last mentioned method.

1. The conditional maximum likelihood estimator

Let us consider the process

$$y(t) = a\cos\omega t + x(t)$$

where the frequency ω is a given constant, a is the unknown parameter and $x(t)$ is a discrete time parameter second order autoregressive process, i.e., $x(t)$ satisfies the difference equation

$$x(t) = \alpha x(t-1) + \beta x(t-2) + \varepsilon(t).$$

The coefficients α and β are known real numbers satisfying the condition $\alpha^2+4\beta<0$, the process $\varepsilon(t)$ is a standard discrete time parameter white noise. The hidden period of this scheme is $2\pi/\omega_1$, where

$$\omega_1 = \arccos\frac{|\alpha|}{2\sqrt{-\beta}}.$$

Further on the damping parameter φ is equal to $\sqrt{-\beta}$.

If the frequency ω varies then this example becomes pathological in the neighbourhood of the hidden frequency ω_1 of the noise process.

The least square estimator $\hat{a}$ has the form

$$\hat{a} = \frac{\sum_{t=1}^{N} y(t)\cos\omega t}{\sum_{t=1}^{N}\cos^2\omega t}.$$

(The estimator $\hat{a}$ is the maximum likelihood estimator of a under the false hypothesis that the noise is white.)

The conditional maximum likelihood estimation $a_{M,c}$ can be calculated from the conditional density function

$$f_c = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\sum (x(t)-\alpha x(t-1)-\beta x(t-2))^2\right\}$$

of the process $x(t)=y(t)-a\cos\omega t$, $t=1, 2, \ldots, N$, under the condition that $x(0)=$ $=x_0$, $x(-1)=x_{-1}$. The solution of the likelihood equation

$$\frac{d \ln f_c}{da} = 0$$

can be written in the form

$$a_{M,c} = \frac{B_c}{A_c}$$

where

$$A_c = \sum_{t=1} (\cos\omega t-\alpha\cos\omega(t-1)-\beta\cos\omega(t-2))^2$$

and

$$B_c = \sum_{t=1} (y(t)-\alpha y(t-1)-\beta y(t-2))(\cos\omega t-\alpha\cos\omega(t-1)-\beta\cos\omega(t-2)).$$

If the noise process $x(t)$ is stationary then its unconditional density function f has the form

$$f = f_{st}\{x(-1), x(0)\} f_c,$$

where $f_{st}\{x(-1), x(0)\}$ is the density function of the two dimensional normal distribution with zero mean and covariance matrix

$$R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{21} \end{pmatrix}$$

with

$$r_{11} = \frac{1-\beta}{1+\beta} \cdot \frac{\sigma_\varepsilon^2}{(1-\beta)^2-\alpha^2}, \quad r_{12} = \frac{\alpha}{1-\beta} r_{11}$$

(see BOX and JENKINS, 1970). So the unconditional maximum likelihood estimator a_M can be written as

$$a_M = \frac{B}{A},$$

where

$$A = \cos^2(-\omega)+[\gamma\cos(-\omega)+\delta]^2+\sum_{t=3} (\cos\omega t-\alpha\cos\omega(t-1)-\beta\cos\omega(t-2))^2$$

and

$$B = y(-1)\cos(-\omega)+[\gamma y(-1)+\delta y(0)][\gamma\cos(-\omega)+\delta]+$$

$$+\sum_{t=3} [y(t)-\alpha y(t-1)-\beta y(t-2)][\cos\omega t-\alpha\cos\omega(t-1)-\beta\cos\omega(t-2)].$$

Here (γ, δ) is one of the solutions of the system of equations

$$0 = r_{11}\gamma+r_{12}\delta,$$

$$1 = r_{11}(\gamma^2+\delta^2)+2\gamma\delta r_{12}.$$

Observe that the terms in the expressions for B_c and B are independent random variables.

All of these estimators are unbiased if the noise process $x(-1), x(0), x(1), \ldots$ is stationary. Under this assumption their variance can easily be determined:

$$\hat{\sigma}^2 = E(\hat{a}-a)^2 = \frac{\sum_{i,j=1} r_{|i-j|} \cos \omega i \cos \omega j}{\sum_{i,j=1} \cos^2 \omega i \cos^2 \omega j},$$

where

$$r_{|i-j|} = Ex(i)x(0); \quad \sigma^2_{M,c} = E(a_{M,c}-a)^2 = \frac{1}{A_c}; \quad \sigma^2_M = E(a_M-a)^2 = \frac{1}{A}.$$

2. A numerical example

In our concrete example the parameters were chosen as follows:

$$a = 6.8, \quad \alpha = 1.83, \quad \beta = -0.98, \quad \omega = \frac{2\pi}{10},$$

$$r_{11} = 173.23, \quad r_{12} = 160.10.$$

So the period of the noise is 16.04 and the damping parameter is close to 1 $(\sqrt{0.98})$.

Table I shows the dependence of $\hat{\sigma}$, $\sigma_{M,c}$ and σ_M on the number N of observations for fixed ω. Tables II, III and IV show the dependence of $\hat{\sigma}$, $\sigma_{M,c}$ and σ_M on the frequency ω for $N=40$, $N=4\pi/\omega$ and $N=2\pi/\omega$, respectively.

TABLE I

The dependence of $\hat{\sigma}, \sigma_{M,c}$ and σ_M on the number N of observations for fixed $\omega(\omega = 2\pi/10)$

N	5	10	15	20	30	40	50	60
$\hat{\sigma}$	10.08	8.34	5.99	3.57	1.83	2.29	1.32	1.33
$\sigma_{M,c}$	2.77	1.96	1.60	1.38	1.13	0.98	0.88	0.80
σ_M	1.20	1.10	1.02	0.96	0.86	0.79	0.73	0.68

TABLE II

The dependence of $\hat{\sigma}, \sigma_{M,c}$ and σ_M on the frequency ω for $N = 40$

$T = \frac{2\pi}{\omega}$	40	30	20	18	16	14	12	10	8	5
$\hat{\sigma}$	4.06	2.99	9.34	11.19	12.31	9.83	3.76	2.29	1.32	0.73
$\hat{\sigma}_{M,c}$	1.78	2.14	4.18	7.11	29.08	4.75	1.93	0.98	0.52	0.18
σ_M	0.89	0.93	1.02	1.05	1.08	1.08	1.00	0.79	0.50	0.19

TABLE III

The dependence of $\hat{\sigma}, \sigma_{M,c}$ and σ_M on the frequency ω for $N = 4\pi/\omega$

$T=\frac{2\pi}{\omega}$	60	40	36	32	28	24	20	16	12	10	8	6
$\hat{\sigma}$	2.88	9.34	11.97	12.47	10.65	7.26	3.57	1.22	2.15	2.84	3.50	4.13
$\sigma_{M,c}$	1.71	4.18	7.52	32.52	5.70	2.49	1.38	0.82	0.49	0.37	0.27	0.20
σ_M	0.88	1.02	1.05	1.08	1.09	1.06	0.96	0.77	0.53	0.41	0.31	0.26

TABLE IV

The dependence of $\hat{\sigma}, \sigma_{M,c}$ and σ_M on the frequency ω for $N = 2\pi/\omega$

$T=\frac{2\pi}{\omega}$	50	40	30	20	18	16	14	12	10	8	6	5	4	3
$\hat{\sigma}$	2.13	4.06	4.09	12.79	13.16	12.79	11.74	10.17	8.34	6.62	5.43	5.12	5.00	5.01
$\sigma_{M,c}$	1.49	1.78	2.42	5.92	10.63	45.99	8.06	3.53	1.96	1.16	0.69	0.52	0.39	0.29
σ_M	0.85	0.89	0.95	1.04	1.06	1.08	1.10	1.11	1.10	1.02	0.82	0.68	0.54	0.65

If $\omega=\omega_1$ then the least square estimator is better than the conditional maximum likelihood one. Naturally, σ_M is always minimal. This phenomenon can be explained as follows. The function

$$z(t) = 0.98^{t/2} \cos \omega_1 t$$

is a solution of the difference equation

$$z(t) = \alpha z(t-1)+\beta z(t-2),$$

so the denominator $A_c \approx 0$.

The pathological behaviour of the conditional maximum likelihood estimator is the consequence of the false assumption that $x(-1)$ and $x(0)$ are known.

*

Acknowledgement. I would like to acknowledge greatfully the valuable advice of A. Krámli.

BIBLIOGRAPHY

[1] ARATÓ, M.: On the statistical examination of continuous state Markov processes, *Selected Transl. in Math. Statist. and Probability* **14** (1978), 203—288.

[2] BRILLINGER, D. R.: An empirical investigation of the Chandler wobble and two proposed excitation processes, *ISI Kongress*, Wien, 1973.

[3] BOX, G. E. P. and JENKINS, G. M.: *Time series analysis forecasting and control*, Holden Day, INC., San Francisco, 1970.

Computer and Automation Institute of the Hungarian Academy of Sciences, Budapest, Kende u. 13—17, Hungary 1111

(Received February 1, 1979)

Studia Scientiarum Mathematicarum Hungarica 12 (1977), 477.

BOOK REVIEW

Dodson, C. T. J., Categories, bundles and spacetime topology, Shiva Mathematics Series No. 1, Shiva Publishing Limited, Orpington, 1980, 223 pp.

This interesting book gives a good survey of the mathematical concepts and tools required by such rapidly developing fields of the theoretical physics as gauge-field theory and the study of various spacetime structures of relativity. The content reflects very well the diversity of notions and theorems, which help the beginner to look into this "new world" of spacetime geometry and to make himself master of handling the sophisticated analytical machinery of mathematics. By now the coordinate-free global method gained ground not only in the differential geometry but also in its applications to physics, and prevails both in the new view of the problems and their mathematical formulations. This book, which is the first volume of a series to be published by Shiva Publishing Limited at post-graduate or research level, is written in this style and it is very suitable to arouse interest or to introduce to the research work those who are interested in this important and fascinating chapter of science. Now let us glance through the book.

The first part (chapters I—III) is an introduction to the naive category theory and topology. We find here the basic concepts of category theory (morphisms, functors, diagrams, limits of diagrams, products, pullback, equalizer, complete categories, limit preserving functors, adjoint functors) and topology (topological space, continuous maps, separation axioms, compactness, connectedness, partition of unity, homotopy, covering space; partial ordering, sup and inf topologies, coinduced and induced topologies, product and coproduct topologies, projective limit and inductive limit topologies). The main part of the book is chapter IV on manifolds and bundles. Here the definitions of manifolds and Lie groups, then that of vector bundles, exact sequences, differentials, jets, fibre bundles, Lie algebras are given. The reader can get acquainted here with the notions and results of connection theory as well as those of the theory of curvature. Riemann and pseudo-Riemann structures are considered and, as an example, the fundamental quantities of the Schwarzschild geometry are determined. Chapter V is devoted to the spacetime structures. Beginning at the Lorentz structures and treating the consequences of the orientability of the space, the author shows the connection between the parallelizability of a manifold and the existence of spinor structures on it. The last section of this chapter deals with the singularities of the spacetime structures.

There are many examples in the book. If the proof of a theorem has been omitted — and this occurred very often — then hints and reference books have always been given to help the reader to reconstruct or find the verification of the statement.

J. Merza

INDEX

Printed in Hungary

A kiadásért felel az Akadémiai Kiadó igazgatója — Műszaki szerkesztő: Botyánszky Pál
A kézirat a nyomdába érkezett: 1980 I. 20. — Terjedelem: 17,25 (A/5) ív, 19 ábra

---

80-293 — Szegedi Nyomda — F. v.: Dobó József igazgató

Die *Studia Scientiarum Mathematicarum Hungarica* ist eine Halbjahrsschrift der Ungarischen Akademie der Wissenschaften. Sie veröffentlicht Originalbeiträge aus dem Bereich der Mathematik in deutscher, englischer, französischer oder russischer Sprache. Es erscheint jährlich ein Band.

Adresse der Redaktion: 1053 Budapest V., Reáltanoda u. 13—15, Ungarn.
Technischer Redaktor: E. Deák

Bestellbar bei Buch- und Zeitungs-Aussenhandelsunternehmen *Kultúra* (Budapest 62, P. O. B. 149), oder bei den Vertretungen im Ausland.

Austauschabmachungen können mit der Bibliothek des Mathematischen Instituts (1053 Budapest V., Reáltanoda u. 13—15) getroffen werden.

Die zur Veröffentlichung bestimmten Manuskripte sind in zwei Exemplaren an die Redaktion zu schicken.

Studia Scientiarum Mathematicarum Hungarica est une revue biannuelle de l'Académie Hongroise des Sciences publiant des essais originaux, en francais, anglais, allemand ou russe, du domaine des mathématiques.

Rédaction: 1053 Budapest V., Reáltanoda u. 13—15, Hongrie.
Rédacteur technique: E. Deák

On s'abonne chez *Kultúra*, Société pour le Commerce de Livres et Journaux (Budapest 62, P. O. B. 149) ou chez ses représentants à l'étranger.

Pour établir des relations d'échange on est prié de s'adresser à la Bibliothèque de l'Institut de Mathématique (1053 Budapest V., Reáltanoda u. 13—15).

On est prié d'envoyer les articles destinés à la publication en deux exemplaires à l'adresse de la Rédaction.

Studia Scientiarum Mathematicarum Hungarica — выходит два раза в год в Издании Академии Наук Венгрии. Журнал публикует оригинальные исследования в области математики на русском, немецком, английском, и французском языках. Отдельные выпуски составляют ежегодно один том.

Адрес редакции: 1053 Budapest V., Reáltanoda u. 13—15, Венгрия.
Технический редактор: E. Deák

Подписка на журнал принимается Внешнеторговым предприятием „Култура" (Budapest 62, P. O. B. 149) или его представительствами за границей.

По поводу отношения обмена просим обращаться к Библиотеке Инзтитута Математики (1053 Budapest V., Reáltanoda u. 13—15).

Работы, преднасначенные для опубликования в журнале следует направлять по адресу редакции в двух экземплярах.

All the reviews of the Hungarian Academy of Sciences may be obtained among others from the following bookshops:

ALBANIA

Ndermarja Shtetnore e Botimeve
Tirana

AUSTRALIA

A. Keesing
Box 4886, GPO
Sidney

AUSTRIA

Globus Buchvertrieb
Salzgries 16
Wien I.

BELGIUM

Office International de Librairie
30, Avenue Marnix
Bruxelles 5
Du Monde Entier
5, Place St. Jean
Bruxelles

BULGARIA

Raznoiznos
1 Tzar Assen
Sofia

CANADA

Pannonia Books
2 Spadina Road
Toronto 4, Ont.

CHINA

Waiwen Shudian
Peking
P.O.B. Nr. 88.

CZECHOSLOVAKIA

Artia A. G.
Ve Smeckách 30
Praha II.
Postova Novinova Sluzba
Dovoz tisku
Vinohradska 46
Praha 2
Postova Novinova Sluzba
Dovoz tlace
Leningradska 14
Bratislava

DENMARK

Ejnar Munksgaard
Nörregade 6
Kopenhagen

FINLAND

Akateeminen Kirjakauppa
Keskuskatu 2
Helsinki

FRANCE

Office International de Documentation et Libraire
48, rue Gay Lussac
Paris 5

GERMAN DEMOCRATIC REPUBLIC

Deutscher Buchexport und Import
Leninstraße 16.
Leipzig C. I.
Zeitungvertriebsamt
Clara Zetkin Straße 62.
Berlin N. W.

GERMAN FEDERAL REPUBLIC

Kunst und Wissen
Eich Bieber
Postfach 46.
7 Stuttgart S.

GREAT BRITAIN

Collet's' Subscription Dept.
44—45 Museum Street
London W. C. I.
Robert Maxwell and Co. Ltd.
Waynflete Bldg. The Plain
Oxford

HOLLAND

Swetz and Zeitlinger
Keizersgracht 471—487
Amsterdam C.
Martinus Nijhof
Lange Voorhout 9
The Hague

INDIA

Current Technical Literature
Co. Private Ltd.
Head Office:
India House OPP.
GPO Post Box 1374
Bombay I.

ITALY

Santo Vanasia
71 Via M. Macchi
Milano
Libreria Commissionaria Sanson
Via La Marmora 45
Firenze

JAPAN

Nauka Ltd.
2 Kanada-Zimbocho 2-ehome
Chiyoda-ku
Tokyo
Maruzen and Co. Ltd.
P.O. Box 605
Tokyo
Far Eastern Booksellers
Kanada P.O. Box 72
Tokyo

KOREA

Chulpanmul
Korejskoje Obschestvo po Exportu Importu Proizvedenij Pechati
Phenjan

NORWAY

Johan Grund Tanum
Karl Johansgatan 43
Oslo

POLAND

Export- und Import-Unternehmen
RUCH
ul. Wilcza 46.
Warszawa

ROUMANIA

Cartimex
Str. Aristide Briand 14—18.
Bucuresti

SOVIET UNION

Mezhdunarodnaja Kniga
Moscow
G—200

SWEDEN

Almquist and Wiksell
Gamla Brogatan 26
Stockholm

USA

Stechert Hafner Inc.
31 East 10th Street
New York 3 N. Y.
Walter J. Johnson
111 Fifth Avenue
New York 3 N. Y.

VIETNAM

Xunhasaba
Service d'Export et d'Import des Livres et Périodiques
19. Tran Quoc Toan
Hanoi

YUGOSLAVIA

Forum
Vojvode Misiva broj 1.
Novi Sad
Jugoslovenska Kniga
Terazije 27.
Beograd

21. XII. 1980.

Index: 26 748

HU ISSN 0081—6906