Studia Scientiarum Mathematicarum Hungarica

AUXILIO
CONSILII INSTITUTI MATHEMATICI
ACADEMIAE SCIENTIARUM HUNGARICAE

REDIGIT
A. RÉNYI

ADIUVANTIBUS
M. ARATÓ, L. FEJES TÓTH, T. FREY,
G. FREUD, L. KALMÁR, A. PRÉKOPA,
K. TANDORI

TOMUS III.
FASC. 1–3.
1968

AKADÉMIAI KIADÓ, BUDAPEST

Studia Scientiarum Mathematicarum Hungarica

A Magyar Tudományos Akadémia matematikai folyóirata

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

Technikai szerkesztő: Katona Gy.

Kiadja az Akadémiai Kiadó, 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 (Budapest II., Fő u. 32).

Cserekapcsolatok felvétele ügyében kérjük az MTA Matematikai Kutató Intézete Könyvtárához (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: Budapest V., Reáltanoda u. 13—15, Hungary.
Technical Editor: Gy. Katona

Subscription rate: Ft 165 per volume. Orders may be placed with *Kultúra* Trading Co. for Books and Newpapers, Budapest 62, P.O.B. 149 or with its representatives abroad.

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

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

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 1—4.

LINEARE PARTIELLE DIFFERENTIALGLEICHUNGEN VON HOHER ORDNUNG

von

H. HORNICH

Wie kürzlich gezeigt [1], gibt es auf einem Intervall des $\mathfrak{R}_1$ Funktionen $y \in C^\infty$, für welche zu jedem $K>0$ eine Folge natürlicher Zahlen $(\nu_1, \nu_2, \ldots)$ existiert, so daß y keiner Differentialgleichung

$$y^{(n)} + \sum_{i=0}^{n-1} \sigma_i y^{(i)} = \varphi \tag{1}$$

mit integrierbaren Koeffizienten σ_i, φ und $|\sigma_i| < K$, $|\varphi| < K$ genügt, wenn n eine der Zahlen ν_k ist. Zum Beispiel sind alle im Intervall nicht regulären Funktionen C^∞ sogar für jede hinreichend große Ordnung n nicht Lösung einer solchen Differentialgleichung. Für reguläre Funktionen gibt es eine Größe, die darüber entscheidet, ob die Funktion die obige Eigenschaft besitzt oder nicht [2].

Im folgenden sollen diese Sätze auf Funktionen von mehreren Variablen, bzw. auf lineare partielle Differentialgleichungen übertragen werden, was wegen der Anzahl der dabei auftretenden Glieder einer solchen Differentialgleichung einiger Aufmerksamkeit bedarf.

In einem Gebiet G des $\mathfrak{R}_n$ sei $\mathfrak{R}$ die Klasse der linearen partiellen Differentialglichungen, die ein einziges Glied höchster Ordnung haben, dessen Koeffizient 1 sei. Sei $(\nu_1, \ldots, \nu_n) = (\nu)$ ein n-Tupel ganzer Zahlen $\nu_j \geqq 0$ und $\nu = \nu_1 + \nu_2 + \ldots + \nu_n$, so schreiben wir die Ableitung ν-ter Ordnung

$$\frac{\partial^\nu u}{\partial x_1^{\nu_1} \ldots \partial x_n^{\nu_n}} = \frac{\partial^\nu u}{\partial x^{(\nu)}}$$

und die Differentialgleichungen aus $\mathfrak{R}$

$$\frac{\partial^\nu u}{\partial x^{(\nu)}} + \sum_{k<\nu} \alpha_{(k)} \frac{\partial^k u}{\partial x^{(k)}} = \varphi \tag{2}$$

wobei die Summe über alle n-Tupel (k) mit Ordnungen $k < \nu$ erstreckt werden soll. Die Koeffizienten $\alpha_{(k)}$ und φ sind Funktionen in G, die wir als beschränkt annehmen.

Die Klasse der auf G beliebig differenzierbaren Funktionen sei wieder C^∞.

Sei für eine Funktion $u \in C^\infty$ und einen Punkt $p \in G$ und ein n-Tupel (ν)

$$\left| \frac{\partial^\nu u}{\partial x^{(\nu)}} \right|^{1/\nu} = S_{(\nu)};$$

wenn die Koeffizienten der Differentialgleichung (2) in p

$$|\alpha_{(k)}|, \quad |\varphi| < \frac{S^{\nu}_{(\nu)}}{1+\sum\limits_{k<\nu} S^{k}_{(k)}} = q_{(\nu)} \tag{3}$$

sind, dann ist

$$\left| \sum_{k<\nu} \alpha_{(k)} \frac{\partial^k u}{\partial x^{(k)}} - \varphi \right| < q_{(\nu)} \cdot \left| 1+\sum_{k<\nu} S^{k}_{(k)} \right| = S^{\nu}_{(\nu)} \tag{4}$$

und u genügt keiner partiellen Differentialgleichung aus $\mathfrak{R}$, deren Koeffizienten (3) genügen.

Wir bilden für eine Funktion $u \in C^{\infty}$

$$\|u\| = \sup_{p \in G} \overline{\lim_{r}} \frac{S_{(\nu)}}{\nu^{n-1}} \tag{5}$$

(über alle $(\nu)=(\nu_1, \nu_2, \ldots, \nu_n)$ zu nehmen). Zunächst einige Eigenschaften von $\|u\|$. Für ein Polynom P ist

$$\|P\| = 0. \tag{6}$$

Für zwei Funktionen $u_1, u_2 \in C^{\infty}$ ist

$$\|u_1 + u_2\| \leqq \max(\|u_1\|, \|u_2\|). \tag{7}$$

Für eine Ableitung gilt

$$\left\| \frac{\partial u}{\partial x_j} \right\| = \|u\|. \tag{8}$$

Wir zeigen weiter:

Ist $n>1$ und für eine Funktion u aus C^{∞}

$$\|u\| > 0, \tag{9}$$

dann ist in mindestens einem Punkt von G

$$\overline{\lim_{\nu}} \frac{S^{\nu}_{(\nu)}}{1+\sum\limits_{k<\nu} S^{k}_{(k)}} > 0. \tag{10}$$

Wäre das nicht der Fall, so ist in jedem Punkt $p \in G$

$$\frac{S^{\nu}_{(\nu)}}{1+\sum\limits_{k<\nu} S^{k}_{(k)}} \to 0. \tag{11}$$

In einem festen Punkt p ist dann für jedes $\alpha>0$ bei fast allen (ν)

$$S^{\nu}_{(\nu)} < \alpha \left(1+\sum_{k<\nu} S^{k}_{(k)}\right). \tag{12}$$

Sei (ν, n) die Anzahl der Darstellungen der natürlichen Zahl ν als Summe von n geordneten ganzen Zahlen $\geqq 0$; wegen

$$(\nu+1, n) = (\nu, n)+(\nu+1, n-1) \tag{13}$$

folgt durch Induktion bei festem n

$$(14) \qquad (\nu, n) = \frac{1}{(n-1)!}\nu^{n-1} + \sum_{j=2}^{n} \nu^{n-j} p_j(n),$$

wo die $p_j(n)$ Polynome in n sind. Aus (12) folgt für ein hinreichend großes N für alle n-Tupel $(N+m)$, $m \geqq 0$ (S. [2])

$$S^{N+m}_{(N+m)} < \alpha\left(1 + \sum_{k<N} S^k_{(k)}\right)(1+\alpha(N,n))(1+\alpha(N+1,n)) \ldots (1+\alpha(N+m-1,n)).$$

Wegen (14) kann man $\eta > 0$ so wählen, daß für alle j

$$(15) \qquad (N+j, n) < \frac{1+\eta}{(n-1)!}(N+j)^{n-1}$$

gilt, so daß

$$(16) \qquad (1+\alpha(N,n)) \ldots (1+\alpha(N+m-1,n)) \leqq \left(1+\alpha\frac{1+\eta}{(n-1)!}(N+m-1)^{n-1}\right)^m.$$

Also ist

$$S^{N+m}_{(N+m)} < \alpha\left(1+\sum_{k<N} S^k_{(k)}\right)\left(1+\alpha\frac{1+\eta}{(n-1)!}(N+m-1)^{n-1}\right)^m$$

und

$$(17) \qquad S_{(N+m)} < \alpha^{1/N+m}\left(1+\sum_{k<N} S^k_{(k)}\right)^{1/N+m}\left(1+\alpha\frac{1+\eta}{(n-1)!}(N+m-1)^{n-1}\right).$$

Für $n=1$ folgt wie früher [3]

$$\overline{\lim_{\nu}}\, S_{(\nu)} \leqq 1$$

und

$$\|u\| \leqq 1.$$

Sei nun $n>1$. Dann folgt aus (17)

$$\overline{\lim_{\nu}} \frac{S_{(\nu)}}{\nu^{n-1}} = 0$$

und $\|u\| = 0$ im Widerspruch zur Annahme.

Daraus folgt im Zusammenhang mit (3), (4) und dem anfangs Gezeigten:

Ist für eine Funktion $u \in C^\infty$ in G

$$\|u\| > 0,$$

so gibt es einen Punkt $p \in G$ und eine Folge von n-Tupeln ganzer Zahlen $(\nu^{(k)})$, so daß in p

$$\frac{S^{\nu^{(k)}}_{(\nu^{(k)})}}{1+\sum_{k<\nu^{(k)}} S^k_{(k)}} \to q > 0$$

ist; dann ist also u sicher nicht Lösung einer Differentialgleichung (2) mit $(\nu) = (\nu^{(k)})$ und Koeffizienten $|\alpha_{(k)}|$, $|\varphi| < q$.

Nun noch eine Bemerkung:
Ist für eine Funktion u aus C^∞ $\|u\|>0$, so ist für $n>2$ in mindestens einem Punkt

$$\overline{\lim_{\nu}} \sqrt[\nu]{\left|\frac{\partial^\nu u}{\partial x^{(\nu)}}\right| \frac{1}{(\nu!)^{n-1}}} > 0$$

also u in G nicht regulär.
Für $\|u\|=0$ ist überall in G

$$s_{(\nu)} = \sqrt[\nu]{\left|\frac{\partial^\nu u}{\partial x^{(\nu)}}\right| \frac{1}{(\nu!)^{n-1}}} \to 0,$$

für hinreichend kleine $s_{(\nu)}$ ist es möglich, daß u Lösung einer linearen partiellen Differentialgleichung (2) mit beschränkten Koeffizienten ist.

LITERATURVERZEICHNIS

[1] HORNICH, H.: A property of the real not regular functions C^∞, *Proc. Amer. Math. Soc.* **17** (1966) 321—324.
[2] HORNICH, H.: Zum Konvergenzverhalten der ganzen Funktionen, *Monatsh. Math.* **70** (1966) 330—336.
[3] l. c. [2] S. 335.

Technische Hochschule, Wien

(Eingegangen: 13. März, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 5—11.

ON THE ABSOLUTE RIESZ SUMMABILITY OF A SERIES RELATED TO A FOURIER SERIES

by

B. D. MALVIYA

1.1. Definitions. Let $\sum a_n$ be a given infinite series, and let $\lambda_n = \lambda(n)$ be a positive monotonic function of n tending to infinity with n. We write

$$A_\lambda(\omega) = A_\lambda^0(\omega) = \sum_{\lambda_n \leqq \omega} a_n;$$

$$A_\lambda^\gamma(\omega) = \sum_{\lambda_n \leqq \omega} (\omega - \lambda_n)^\gamma a_n, \qquad \gamma > 0.$$

The series $\sum a_n$ is said to be summable (R, λ, γ), $\gamma \geqq 0$, to sum s, if $A_\lambda^\gamma(\omega)/\omega^\gamma \to s$, as $\omega \to \infty$, and is said to be absolutely summable (R, λ, γ), or summable $|R, \lambda, \gamma|$, $\gamma \geqq 0$, if $A_\lambda^\gamma(\omega)/\omega^\gamma \in \mathrm{BV}(A, \infty)$,[1] where A is a finite positive number. [2]

The sequence $\{\lambda_n\}$ is called the 'type' and the number γ is called the 'order'.

An equivalent definition is obtained, as follows, by a suitable extension of the definition of the type $\lambda(x)$ at points other than those given by $x = n$ $(n = 1, 2, \ldots)$, and a corresponding change in the variable involved in the 'Riesz mean' $A_\lambda^\gamma(\omega)/\omega^\gamma$.

Let $\lambda = \lambda(\omega)$ be a differentiable, monotonic increasing function of ω in (A, ∞), where A is a positive constant, and let $\lambda(\omega)$ tend to infinity with ω. We write

$$C_\gamma(\omega) = \sum_{n \leqq \omega} \{\lambda(\omega) - \lambda(n)\}^\gamma a_n, \qquad \gamma > 0.$$

Then $\sum a_n$ is said to be summable $|R, \lambda, \gamma|$, $\gamma > 0$, if the integral

$$\int_A^\infty |d[C_\gamma(\omega)/\{\lambda(\omega)\}^\gamma]|$$

converges.

Now, for $\gamma > 0$, $m < \omega < m + 1$,

$$\frac{d}{d\omega}[C_\gamma(\omega)/\{\lambda(\omega)\}^\gamma] = \frac{\gamma\lambda'(\omega)}{\{\lambda(\omega)\}^{\gamma+1}} \sum_{n \leqq \omega} \{\lambda(\omega) - \lambda(n)\}^{\gamma-1} \lambda(n) a_n.$$

By definition, the summability $|R, \lambda, 0|$ is equivalent to absolute convergence, whatever be the type $\lambda(\omega)$.

For convenience, we shall adopt here the alternative definition given above.

[1] By '$f(x) \in BV(h, k)$' we mean that $f(x)$ is a function of bounded variation over the interval (h, k).

[2] Obrechkoff [2], [3].

1. 2. Let $\varphi(t)$ be an even function integrable in the sense of Lebesgue in $(0, \pi)$ and defined outside $(-\pi, \pi)$ by periodicity. We assume, without any loss of generality, that the constant term in the Fourier series of $\varphi(t)$ is zero and the special point to be considered is the origin. In these circumstances

$$\varphi(t) \sim \sum_{n=1}^{\infty} A_n \cos nt, \tag{1.21}$$

where

$$A_n = \frac{2}{\pi} \int_0^{\pi} \varphi(t) \cos nt \, dt, \tag{1.22}$$

and we are to consider the series $\sum A_n$.

1.3. We use the following notations:

$$\Phi_\alpha(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-u)^{\alpha-1} \varphi(u) \, du \qquad (t > 0 \quad \text{and} \quad 0 < \alpha < 1), \tag{1.31}$$

$$\Phi_0(t) = \varphi(t), \tag{1.32}$$

$$\varphi_\alpha(t) = \Gamma(1+\alpha) t^{-\alpha} \Phi_\alpha(t) \qquad 0 \leqq \alpha < 1, \tag{1.33}$$

$$p(\omega) = \sum_{n \leqq \omega} e^{(\lambda(n))^{\Delta}} \big(\lambda(n)\big)^{-1} A_n \qquad \Delta > 0. \tag{1.34}$$

2. Introduction. The following theorem concerning the absolute RIESZ summability of the series $\sum_2^{\infty} A_n/\log n$ has been demonstrated by MOHANTY and MISRA.

THEOREM A. [3] *If $\varphi_\alpha(t)$ is of bounded variation in $(0, \pi)$, then the series $\sum_2^{\infty} A_n/\log n$*

The object of the present paper is to replace the sequence of factors $\{1/\log n\}$ in this theorem by a generalised sequence of factors $\{1/\lambda(n)\}$ and prove that under certain conditions on $\lambda(n)$, one gets the summability $|R, e^{(\lambda(n))^{\Delta}}, 1|$, $\Delta = 1 + \frac{1}{\alpha}$, of $\sum A_n/\lambda(n)$ if, as before, $\varphi_\alpha(t) \in \mathrm{BV}(0, \pi)$.

More precisely we prove the following theorem.

THEOREM. *Let $\Delta = 1 + \frac{1}{\alpha}$, $0 < \alpha < 1$, and $\lambda(x)$ satisfy the following conditions:*

$$\begin{cases} \text{(i)} \ e^{(\lambda(x))^{\Delta}} \big(\lambda(x)\big)^{-1} x^{-1} \ \textit{is monotonic non-decreasing,} \\ \text{(ii)} \ \big(\lambda'(x)\big)^{-1} \big(\lambda(x)\big)^{1-\Delta} \ \textit{is monotonic non-decreasing,} \end{cases} \tag{2.1}$$

$$\lambda[(k/u)\{\lambda(k/u)\}^{1/\alpha}] = O\big(\lambda(k/u)\big), \quad \textit{as} \quad u \to 0, \tag{2.2}$$

is summable $|R, e^{(\log n)^{\Delta}}, 1|$, where $0 < \alpha < 1$ and $\Delta = 1 + \frac{1}{\alpha}$.

[3] MOHANTY and MISRA [1].

[4] k is a suitably chosen constant.

$$\int_A^\omega x^\alpha \frac{d}{dx}\left(\frac{1}{\lambda(x)}\right)dx = O\left(\frac{\omega^\alpha}{\lambda(\omega)}\right), \quad as \quad \omega \to \infty, \tag{2.3}$$

$$\begin{cases} \text{(i)} \int_\omega^\infty x^{\alpha-1}\frac{d}{dx}\{(\lambda(x))^{1/\alpha}\}\,dx = O\left[\left\{\frac{(\lambda(\omega))^{1/\alpha}}{\omega}\right\}^{1-\alpha}\right], \quad as \quad \omega \to \infty. \\ \text{(ii)} \int_A^\infty x^{\alpha-1}\frac{d}{dx}\{(\lambda(x))^{1/\alpha}\}\,dx = O(1), \text{ for finite } A, \end{cases} \tag{2.4}$$

$$\begin{cases} \text{(i)} \int_1^\omega \frac{1}{x\lambda'(x)(\lambda(x))^\Delta}\frac{d}{dx}(e^{(\lambda(x))^\Delta})\,dx = O\left\{\frac{e^{(\lambda(\omega))^\Delta}}{\omega\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \\ \text{(ii)} \int_1^\omega \frac{1}{\lambda'(x)(\lambda(x))^\Delta}\frac{d}{dx}(e^{(\lambda(x))^\Delta})\,dx = O\left\{\frac{e^{(\lambda(\omega))^\Delta}}{\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \quad as \quad \omega \to \infty.^5 \end{cases} \tag{2.5}$$

Then, for $0<\alpha<1$, *if* $\varphi_\alpha(t)\in \mathrm{BV}(0,\pi)$, *the series* $\sum A_n/\lambda(n)$ *is summable* $|R, e^{(\lambda(n))^\Delta}, 1|$.

NOTE: Throughout the sequel we use $\lambda(n)$ in the sense defined by the hypotheses (2. 1)—(2. 5).

3. We require the following order-estimates, as $\omega\to\infty$ and $t\to 0$, and give their proofs below:

$$\xi(\omega,t) = \sum_{n\leqq\omega} e^{(\lambda(n))^\Delta}(\lambda(n))^{-1}\cos nt = \begin{cases} O\left\{\frac{e^{(\lambda(\omega))^\Delta}}{\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \\ O\{e^{(\lambda(\omega))^\Delta}(\lambda(\omega))^{-1}t^{-1}\}; \end{cases} \tag{3.1, 3.2}$$

$$\eta(\omega,t) = \sum_{n\leqq\omega} e^{(\lambda(n))^\Delta}(\lambda(n))^{-1}n^{-1}\sin nt = \begin{cases} O\left\{\frac{e^{(\lambda(\omega))^\Delta}}{\omega\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \\ O\{e^{(\lambda(\omega))^\Delta}(\lambda(\omega))^{-1}\omega^{-1}t^{-1}\}; \end{cases} \tag{3.3, 3.4}$$

$$g(\omega,u) = \frac{1}{\Gamma(1-\alpha)}\int_u^\pi (t-u)^{-\alpha}\xi(\omega,t)\,dt = \begin{cases} O\left\{\frac{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}}{\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \\ O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}u^{-1}(\lambda(\omega))^{-1}\}; \end{cases} \tag{3.5, 3.6}$$

$$G(\omega,u) = \frac{1}{\Gamma(1+\alpha)}\int_0^u v^\alpha \frac{d}{dv}g(\omega,v)dv = \begin{cases} O\left\{\frac{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}u^\alpha}{\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}, \\ O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}u^{-1+\alpha}(\lambda(\omega))^{-1}\}. \end{cases} \tag{3.7, 3.8}$$

[5] We observe that (2. 5) (i) and (2. 5) (ii) may be replaced by (2. 5) (i) and the hypothesis

$$\text{(ii)}' \quad \int_1^\omega \frac{dx}{x\lambda'(x)} = O\left(\frac{1}{\lambda'(\omega)}\right), \tag{2.5}$$

since (2. 5) (i) and (2. 5) (ii) together imply (2. 5) (ii)′.

PROOF OF (3. 1).
Let $m \leqq \omega < m+1$. Using $|\cos nt| \leqq 1$,

$$\xi(\omega, t) \leqq \sum_{n=1}^{m} e^{(\lambda(n))^{\Delta}} (\lambda(n))^{-1} < \int_{1}^{\omega} \frac{e^{(\lambda(x))^{\Delta}}}{\lambda(x)}\, dx + \frac{e^{(\lambda(m))^{\Delta}}}{\lambda(m)} =$$

$$= O\left\{\int_{1}^{\omega} \frac{1}{\lambda'(x)(\lambda(x))^{\Delta}} \frac{d}{dx}\left(e^{(\lambda(x))^{\Delta}}\right) dx\right\} + O\left\{\frac{e^{(\lambda(\omega))^{\Delta}}}{\lambda(\omega)}\right\} =$$

$$= O\left\{\frac{e^{(\lambda(\omega))^{\Delta}}}{\lambda'(\omega)(\lambda(\omega))^{\Delta}}\right\}, \text{ by (2. 5) (ii) and (2. 1).}$$

PROOF OF (3. 2).
The estimate follows from (2. 1) (i) by ABEL's Lemma.

PROOF OF (3. 3).
It is similar to that of (3. 1) and follows by using (2. 1) and (2. 5) (i).

PROOF OF (3. 4).
The estimate follows from hypothesis (2. 1) (i) by ABEL's Lemma.

PROOF OF (3. 5) and (3. 6).
Let $u + \frac{1}{\omega} < \pi$. We write

$$\Gamma(1-\alpha) g(\omega, u) = \left\{\int_{u}^{u+\frac{1}{\omega}} + \int_{u+\frac{1}{\omega}}^{\pi}\right\} (t-u)^{-\alpha} \xi(\omega, t)\, dt = I_1 + I_2. \text{ By (3. 1) and (3. 2),}$$

$$|I_1| < K e^{(\lambda(\omega))^{\Delta}} \int_{u}^{u+\frac{1}{\omega}} (t-u)^{-\alpha} \min\left[\frac{1}{\lambda'(\omega)(\lambda(\omega))^{\Delta}}, (\lambda(\omega))^{-1} t^{-1}\right] dt <^{6}$$

$$< K e^{(\lambda(\omega))^{\Delta}} \min\left[\frac{1}{\lambda'(\omega)(\lambda(\omega))^{\Delta}}, u^{-1}(\lambda(\omega))^{-1}\right] \int_{u}^{u+\frac{1}{\omega}} (t-u)^{-\alpha}\, dt =$$

$$= O\left\{e^{(\lambda(\omega))^{\Delta}} \omega^{-1+\alpha}\right\} \min\left[\frac{1}{\lambda'(\omega)(\lambda(\omega))^{\Delta}}, u^{-1}(\lambda(\omega))^{-1}\right].$$

By the second mean value theorem, we have

$$I_2 = \omega^{\alpha} \int_{u+\frac{1}{\omega}}^{\zeta} \xi(\omega, t)\, dt = \omega^{\alpha} [\eta(\omega, t)]_{u+\frac{1}{\omega}}^{\zeta}, \qquad \left(u + \frac{1}{\omega} < \zeta < \pi\right).$$

[6] Throughout this paper K denotes a positive constant, not alway the same.

Hence

$$I_2 = O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}\}\min\left[\frac{1}{\lambda'(\omega)(\lambda(\omega))^\Delta}, u^{-1}(\lambda(\omega))^{-1}\right]$$

and

$$g(\omega,u) = O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}\}\min\left[\frac{1}{\lambda'(\omega)(\lambda(\omega))^\Delta}, u^{-1}(\lambda(\omega))^{-1}\right].$$

PROOF OF (3. 7).
We have, by (3. 5),

$$\Gamma(1+\alpha)G(\omega,u) = \int_0^u v^\alpha \frac{d}{dv} g(\omega,v)\,dv =$$

$$= [v^\alpha g(\omega,v)]_0^u - \alpha\int_0^u v^{\alpha-1} g(\omega,v)\,dv = O\left\{\frac{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}u^\alpha}{\lambda'(\omega)(\lambda(\omega))^\Delta}\right\}.$$

PROOF OF (3. 8). By (3. 6),

$$\Gamma(1+\alpha)G(\omega,u) = [v^\alpha g(\omega,v)]_0^u - \alpha\int_0^u v^{\alpha-1} g(\omega,v)\,dv =$$

$$= O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}u^{-1+\alpha}(\lambda(\omega))^{-1}\} + \alpha\int_0^u v^{\alpha-1} O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}v^{-1}(\lambda(\omega))^{-1}\}\,dv.$$

Hence the result.

4. Proof of Theorem. To prove the theorem, we have to show that when $\Delta = 1+\frac{1}{\alpha}$,

$$I = \int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1} e^{-(\lambda(\omega))^\Delta}|p(\omega)|\,d\omega < \infty.$$

We have

$$A_n = \frac{2}{\pi}\int_0^\pi \varphi(t)\cos nt\,dt = \frac{2}{\pi}\frac{1}{\Gamma(1-\alpha)}\int_0^\pi \cos nt\int_0^t (t-u)^{-\alpha}\,d\Phi_\alpha(u) =$$

$$= \frac{2}{\pi}\frac{1}{\Gamma(1-\alpha)}\int_0^\pi d\Phi_\alpha(u)\int_u^\pi (t-u)^{-\alpha}\cos nt\,dt$$

$$\frac{\pi}{2}p(\omega) = \frac{1}{\Gamma(1-\alpha)}\int_0^\pi d\Phi_\alpha(u)\int_u^\pi (t-u)^{-\alpha}\left\{\sum_{n\leqq\omega} e^{(\lambda(n))^\Delta}(\lambda(n))^{-1}\cos nt\right\}dt =$$

$$= \frac{1}{\Gamma(1-\alpha)}\int_0^\pi d\Phi_\alpha(u)\int_u^\pi (t-u)^{-\alpha}\xi(\omega,t)\,dt = \int_0^\pi g(\omega,u)\,d\Phi_\alpha(u) =$$

$$= [g(\omega,u)\Phi_\alpha(u)]_0^\pi - \int_0^\pi \Phi_\alpha(u)\frac{d}{du}g(\omega,u)\,du.$$

Further, since $\varphi_\alpha(+0)$ is finite,

$$\int_0^\pi \Phi_\alpha(u)\frac{d}{du}g(\omega,u)\,du = \frac{1}{\Gamma(1+\alpha)}\int_0^\pi \varphi_\alpha(u)u^\alpha\frac{d}{du}g(\omega,u)\,du =$$

$$= \frac{1}{\Gamma(1+\alpha)}\Big[\varphi_\alpha(u)\int_0^u v^\alpha\frac{d}{dv}g(\omega,v)\,dv\Big]_0^\pi - \frac{1}{\Gamma(1+\alpha)}\int_0^\pi\Big(\int_0^u v^\alpha\frac{d}{dv}g(\omega,v)\,dv\Big)d\varphi_\alpha(u) =$$

$$= \varphi_\alpha(\pi)G(\omega,\pi) - \int_0^\pi G(\omega,u)\,d\varphi_\alpha(u).$$

So, we have finally,

$$\frac{\pi}{2}p(\omega) = [g(\omega,u)\Phi_\alpha(u)]_0^\pi - \varphi_\alpha(\pi)G(\omega,\pi) + \int_0^\pi G(\omega,u)\,d\varphi_\alpha(u) =$$

$$= O\{e^{(\lambda(\omega))^\Delta}\omega^{-1+\alpha}(\lambda(\omega))^{-1}\} + \int_0^\pi G(\omega,u)\,d\varphi_\alpha(u).$$

Hence

$$I \leqq K\int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1}\omega^{-1+\alpha}(\lambda(\omega))^{-1}\,d\omega +$$

$$+K\int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1}e^{-(\lambda(\omega))^\Delta}\Big|\int_0^\pi G(\omega,u)\,d\varphi_\alpha(u)\Big|\,d\omega.$$

By (2. 4) (ii), the first integral

$$\int_A^\infty \lambda'(\omega)\omega^{-1+\alpha}((\lambda(\omega))^{\Delta-2}\,d\omega = O\Big[\int_A^\infty \omega^{-1+\alpha}\frac{d}{d\omega}\{(\lambda(\omega))^{\Delta-1}\}\,d\omega\Big] = O(1),$$

and the integral

$$\int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1}e^{-(\lambda(\omega))^\Delta}\Big|\int_0^\pi G(\omega,u)\,d\varphi_\alpha(u)\Big|\,d\omega \leqq$$

$$\leqq \int_0^\pi |d\Phi_\alpha(u)|\int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1}e^{-(\lambda(\omega))^\Delta}|G(\omega,u)|\,d\omega.$$

Since $\varphi_\alpha(t)$ is of bounded variation in $(0,\pi)$, to prove the theorem it will be sufficient to show that, uniformly in $0<u<\pi$,

$$J = \int_A^\infty \lambda'(\omega)(\lambda(\omega))^{\Delta-1}e^{-(\lambda(\omega))^\Delta}|G(\omega,u)|\,d\omega = O(1).$$

Writing

$$J = \int_A^\tau + \int_\tau^\infty = J_1 + J_2,$$

where $\tau = \frac{k}{u}\{\lambda(k/u)\}^{1/\alpha}$, we have by (3. 7) and (2. 3),

$$J_1 = \int_A^\tau \lambda'(\omega)\big(\lambda(\omega)\big)^{\Delta-1} e^{-(\lambda(\omega))^\Delta} O\{e^{(\lambda(\omega))^\Delta} \omega^{-1+\alpha} u^\alpha (\lambda'(\omega))^{-1}\big(\lambda(\omega)\big)^{-\Delta}\}\, d\omega \leqq$$

$$\leqq K\left\{u^\alpha \int_A^\tau \frac{\omega^{-1+\alpha}}{\lambda(\omega)}\, d\omega\right\} \leqq K\left[\frac{u^\alpha \omega^\alpha}{\lambda(\omega)}\right]_A^\tau + K\left\{u^\alpha \int_A^\tau \omega^\alpha \frac{d}{d\omega}\left(\frac{1}{\lambda(\omega)}\right) d\omega\right\} \leqq$$

$$\leqq K + K\left\{\frac{u^\alpha \tau^\alpha}{\lambda(\tau)}\right\} = O(1),$$

since $\lambda(\tau) > \lambda(k/u)$, as $\tau > k/u$ for sufficiently small u, and by (3. 8), (2. 4) (i) and (2. 2)

$$J_2 = \int_\tau^\infty \lambda'(\omega)\big(\lambda(\omega)\big)^{\Delta-1} e^{-(\lambda(\omega))^\Delta} O\{e^{(\lambda(\omega))^\Delta} \omega^{-1+\alpha} u^{-1+\alpha}\big(\lambda(\omega)\big)^{-1}\}\, d\omega =$$

$$= O\left\{u^{-1+\alpha} \int_\tau^\infty \omega^{\alpha-1} \frac{d}{d\omega}\big((\lambda(\omega))^{\Delta-1}\big)\, d\omega\right\} = O\left[\frac{(\lambda(\tau))^{1/\alpha}}{u\tau}\right]^{1-\alpha} = O(1).$$

Hence the theorem is proved.

The author is grateful to Dr. T. PATI, Head of the Department of Post-Graduate Studies and Research in Mathematics, University of Jabalpur, for his valuable guidance during the preparation of this paper, and to Professor G. FREUD for his kind suggestions for the improvement of the presentation of the paper.

REFERENCES

[1] MOHANTY, R. and MISRA, B.: On the absolute logarithmic summability of a sequence related to a Fourier series, *Tôhoku Math. J.* **6** (1954) 5—12.

[2] OBRECHKOFF, N.: Sur la sommation absolute de séries de Dirichlet, *C. R. Acad. Sci. Paris* **186** (1928), 215—217.

[3] OBRECHKOFF, N.: Über die absolute summierung der Dirichletschen Reihen, *Math. Z.* **30** (1929) 375—386.

Department of Mathematics University of Allahabad, Allahabad (India) and Department of Post-Graduate Studies and Research in Mathematics, University of Jabalpur, Jabalpur (India)

(Received May 21, 1963.)
(Revised August 25, 1964.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 13—18.

A METHOD FOR THE SOLUTION OF LINEAR EQUATION SYSTEMS

by

G. TEVAN

Since the solution of an inhomogeneous equation system may be reduced to the solution of a (one order higher) homogeneous equation system, here we consider homogeneous equation systems only. I.e. we have to determine those column vectors $\mathbf{x}$, for which

$$\mathbf{Bx} = \mathbf{O}. \tag{1}$$

It is well-known, that the solution-vectors of the equation system

$$\mathbf{Ax} \equiv \mathbf{B}^*\mathbf{Bx} = \mathbf{O} \tag{2}$$

coincide with the solution vectors of Eq. (1).

The eigenvalues of the symmetrical matrix $\mathbf{A} = \mathbf{B}^*\mathbf{B}$ are non-negative. The solution vectors of Eq. (2) and consequently those of Eq. (1) are the eigenvectors of matrix $\mathbf{A}$, having an eigenvalue zero. It is also well-known that the iteration

$$\mathbf{x}_{n+1} = \mathbf{Ax}_n \tag{3}$$

leads to an eigenvector with maximal eigenvalue if the initial columnvector $\mathbf{x}_0$ is general enough. Then, from this the maximal eigenvalue $\lambda_{\max}$ can be determined.

Let us consider the matrix

$$\mathbf{A}_1 = \mathbf{E} - \frac{1}{\lambda_{\max}}\mathbf{A} \tag{4}$$

where $\mathbf{E}$ is the unit matrix.

The eigenvectors of matrix $\mathbf{A}_1$ obviously coincide with the eigenvectors of $\mathbf{A}$; the eigenvalues of $\mathbf{A}_1$ are of the form

$$1 - \frac{\lambda_i}{\lambda_{\max}}, \quad \text{where}$$

$\lambda_i \geqq 0$ denotes the i^{th} eigenvalue of the matrix $\mathbf{A}$. Consequently, the eigenvalues of matrix $\mathbf{A}_1$ are non-negative as well and their maximum is

$$1 - \frac{\lambda_{\min}}{\lambda_{\max}}, \quad \text{where}$$

$\lambda_{\min}$ denotes the minimal eigenvalue of $\mathbf{A}$.

If Eq. (1) has a solution and thus Eq. (2) has a solution too, then $\lambda_{\min} = 0$, *the maximal eigenvalue of matrix* $\mathbf{A}_1$ *equals* 1, *and the eigenvectors belonging to this eigenvalue coincide with the solutionvectors of* Eq. (1).

(It suffices, however, to calculate by iteration (3) the value of $\lambda_{\max}$ for a few digits only and then rounding it up.)

Thus, the iteration

$$\mathbf{x}_{n+1} = \mathbf{A}_1 \mathbf{x}_n \tag{5}$$

starting with a sufficiently general $\mathbf{x}_0$, leads to an eigenvector with unit eigenvalue, which is a solution-vector of Eq. (1). Let us denote this by $\mathbf{u}_1$. Thus

$$\mathbf{A}_1 \mathbf{u}_1 = \mathbf{u}_1 \quad \text{and} \quad \mathbf{A}\mathbf{u}_1 = \mathbf{0}.$$

Now the symmetrical matrix

$$\mathbf{A}_2 = \mathbf{A}_1 - \frac{\mathbf{u}_1 \mathbf{u}_1^*}{\mathbf{u}_1^* \mathbf{u}_1} \tag{6}$$

may be considered. The vector $\mathbf{u}_1$ is an eigenvector of $\mathbf{A}_2$ too, but with an eigenvalue zero; namely

$$\mathbf{A}_2 \mathbf{u}_1 = \mathbf{A}_1 \mathbf{u}_1 - \mathbf{u}_1 = \mathbf{u}_1 - \mathbf{u}_1 = \mathbf{O}.$$

On the basis of Eq. (6) it is obvious that all the other eigenvectors of $\mathbf{A}_1$ and $\mathbf{A}_2$ are orthogonal to $\mathbf{u}_1$, and they belong to the same eigenvalues. Thus, the solution-vectors of $\mathbf{A}_1$ and $\mathbf{A}_2$ are orthogonal to $\mathbf{u}_1$, and they belong to the same eigenvalues. Thus, the solutionvectors of Eq. (1) orthogonal to $\mathbf{u}_1$ belong to the maximal (i.e. unit) eigenvalues of $\mathbf{A}_2$ and conversely, if such an eigenvalue exists. Thus the iteration

$$\mathbf{x}_{n+1} = \mathbf{A}_2 \mathbf{x}_n \tag{7}$$

with a sufficiently general initial vector leads to a solutionvector $\mathbf{u}_2$ orthogonal to $\mathbf{u}_1$ for which

$$\mathbf{A}_2 \mathbf{u}_2 = \mathbf{u}_2; \ \mathbf{A}_1 \mathbf{u}_2 = \mathbf{u}_2 \quad \text{and} \quad \mathbf{A}\mathbf{u}_2 = \mathbf{O}, \ \mathbf{u}_1^* \mathbf{u}_2 = 0.$$

After this, we have to determine the matrix

$$\mathbf{A}_3 = \mathbf{A}_2 - \frac{\mathbf{u}_2 \mathbf{u}_2^*}{\mathbf{u}_2^* \mathbf{u}_2} \tag{8}$$

with the properties

$$\mathbf{A}_3 \mathbf{u}_2 = \mathbf{O} \quad \text{and} \quad \mathbf{A}_3 \mathbf{u}_1 = \mathbf{O}.$$

The eigenvectors of the matrices $\mathbf{A}_2$ and $\mathbf{A}_3$, orthogonal to $\mathbf{u}_2$ coincide as well as their eigenvalues. Thus the iteration

$$\mathbf{x}_{n+1} = \mathbf{A}_3 \mathbf{x}_n \tag{9}$$

leads, if there exists a solutionvector, equally orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$, to an eigenvector $\mathbf{u}_3$ with unit eigenvalue for which

$$\mathbf{u}_1^* \mathbf{u}_3 = 0, \ \mathbf{u}_2^* \mathbf{u}_3 = 0,$$

$$\mathbf{A}_3 \mathbf{u}_3 = \mathbf{A}_2 \mathbf{u}_3 = \mathbf{A}_1 \mathbf{u}_3 = \mathbf{u}_3, \ \mathbf{A}\mathbf{u}_3 = 0.$$

Following this, the matrix

$$\mathbf{A}_4 = \mathbf{A}_3 - \frac{\mathbf{u}_3 \mathbf{u}_3^*}{\mathbf{u}_3^* \mathbf{u}_3} \tag{10}$$

is to be determined and so on.

If the iteration results for a sufficiently general initial vector $\mathbf{x}_0$ in an eigenvalue less than 1, this means that there are no further linearly independent solution vectors.

Thus by means of the iteration method described above, it is possible to determine an orthogonal vector system for the general solution of the homogeneous linear equation system.

If the iteration (5) converges slowly, then this is due to the fact that matrix $\mathbf{A}_1$ has — besides the solution vectors belonging to unit eigenvalue — eigenvectors belonging to eigenvalues slightly less than 1. Let us assume now that such an eigenvector $\mathbf{v}_1$ — slowing down the convergence — is already known. The eigenvectors of the matrix

$$\mathbf{A}_1' = \mathbf{A}_1 - \frac{\mathbf{v}_1^* \mathbf{A}_1 \mathbf{v}_1}{\mathbf{v}_1^* \mathbf{v}_1} \cdot \frac{\mathbf{v}_1 \mathbf{v}_1^*}{\mathbf{v}_1^* \mathbf{v}_1} \tag{11}$$

coincide with those of $\mathbf{A}_1$. The eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_1'$ resp. coincide as well, with the exception of the eigenvalue belonging to $\mathbf{v}_1$, which equals 0 for $\mathbf{A}_1'$. Thus the convergence of the iteration with $\mathbf{A}_1'$ is better. It is easy to see that if $\mathbf{v}_1$ is only an approximation of the eigenvector slowing down the convergence but the condition $\mathbf{v}_1 \mathbf{u} = 0$ is fulfilled for any solutionvector $\mathbf{u}$, then the eigenvectors belonging to the greatest — unit — eigenvalue are the solutionvectors $\mathbf{u}$. Namely, the expression

$$\frac{\mathbf{x}^* \mathbf{A}_1' \mathbf{x}}{\mathbf{x}^* \mathbf{x}} = \frac{\mathbf{x}^* \mathbf{A}_1 \mathbf{x}}{\mathbf{x}^* \mathbf{x}} - \frac{\mathbf{v}_1^* \mathbf{A}_1 \mathbf{v}_1}{\mathbf{v}_1^* \mathbf{v}_1} \cdot \frac{(\mathbf{v}_1^* \mathbf{x})^2}{\mathbf{v}_1^* \mathbf{v}_1}$$

takes on its maximal — unit — value for $\mathbf{x} = \mathbf{u}$, the first term of the right hand side being here maximal of value 1, the second term being minimal of value 0. Thus it suffices to determine such an approximation of $\mathbf{v}_1$ for which $\mathbf{v}_1^* \mathbf{u} = 0$ holds.

In order to obtain the vector $\mathbf{v}_1$ let us consider the difference of two subsequent vectors of the iteration (5):

$$\mathbf{w}_1 = \mathbf{x}_{r+1} - \mathbf{x}_r. \tag{12}$$

This difference vector $\mathbf{w}_1$ is orthogonal to the solution vectors $\mathbf{u}$, as

$$(\mathbf{x}_{r+1} - \mathbf{x}_r)^* \mathbf{u} = (\mathbf{A}_1 \mathbf{x}_r - \mathbf{x}_r)^* \mathbf{u} = \mathbf{x}_r^* (\mathbf{A}_1 \mathbf{u} - \mathbf{u}) = \mathbf{O},$$

and due to the substraction in $\mathbf{w}_1$ the weight of the approximative $\mathbf{v}_1$ is small in comparison to the weights of the other eigenvectors. In order to be able to separate the unknown vector from $\mathbf{w}_1$, we consider the iteration

$$\mathbf{w}_{n+1} = \mathbf{A}_1 \mathbf{w}_n \tag{13}$$

with the initial vector $\mathbf{w}_1$. In consequence of the errors of rounding off, after several steps in general $\mathbf{w}_n^* \mathbf{u} \neq 0$ holds. Therefore let us consider for sufficiently large value of q the exact value of the expression

$$\mathbf{v}_1 = \mathbf{A}_1 \mathbf{w}_q - \mathbf{w}_q \tag{14}$$

satisfying exactly equation $\mathbf{v}_1^* \mathbf{u} = 0$. The value of q is to be chosen such that not

only $\mathbf{w}_q$ but $\mathbf{v}_1$ as well should approximate the eigenvector belonging to the largest eigenvalue less than 1. This is the case if the quantity

$$\frac{\mathbf{v}_1^* \mathbf{A}_1 \mathbf{v}_1}{\mathbf{v}_1^* \mathbf{v}_1} = \frac{(\mathbf{A}_1 \mathbf{w}_q - \mathbf{w}_q)^* \mathbf{A}_1 (\mathbf{A}_1 \mathbf{w}_q - \mathbf{w}_q)}{(\mathbf{A}_1 \mathbf{w}_q - \mathbf{w}_q)^* (\mathbf{A}_1 \mathbf{w}_q - \mathbf{w}_q)} \approx \frac{(\mathbf{w}_{q+1} - \mathbf{w}_q)^* (\mathbf{w}_{q+2} - \mathbf{w}_{q+1})}{(\mathbf{w}_{q+1} - \mathbf{w}_q)^* (\mathbf{w}_{q+1} - \mathbf{w}_q)},$$

that is if the quantity

$$K_q = \frac{(\triangle \mathbf{w}_q)^* (\triangle \mathbf{w}_{q+1})}{(\triangle \mathbf{w}_q)^* (\triangle \mathbf{w}_q)} \tag{15}$$

has already approximated its maximal value.

In case the iteration with matrix $\mathbf{A}_1$ converges slowly, we have to determine the vector $\mathbf{w}_1$ according to (12), and apply this as initial vector to the iteration (13). After every s steps — for a fixed value of s, say $s = 10$ — from four subsequent values of $\mathbf{w}_q$ three subsequent values of $\triangle \mathbf{w}_q$ have to be calculated and from these two subsequent values of K_q. Iteration (13) has to be stopped if $K_q - K_{q-1} \leqq \varepsilon$ (ε denotes the accuracy required). Following this the exact value of $\mathbf{v}_1$ has to be determined according to (14) and from this matrix $\mathbf{A}_1'$ of (11) (a few digits of $\frac{\mathbf{v}_1^* \mathbf{A}_1 \mathbf{v}_1}{(\mathbf{v}_1^* \mathbf{v}_1)^2} \approx \frac{K_q}{\mathbf{v}_1^* \mathbf{v}_1}$ are sufficient). Finally the iteration with $\mathbf{A}_1'$ has to be continued with the same vector $\mathbf{v}_r$ for which the iteration with $\mathbf{A}_1$ has been stopped.

It may occur that even the latter iteration converges slowly. Than another vector $\mathbf{v}_2$ can be determined in the same way as described above and the iteration can be continued with the matrix

$$\mathbf{A}_1'' = \mathbf{A}_1' - \frac{K_{q2}}{\mathbf{v}_2^* \mathbf{v}_2} \mathbf{v}_2 \mathbf{v}_2^*$$

and so on.

We illustrate the method by means of the following example. Let us determine the solution of the homogeneous linear system of equations $\mathbf{Bx} = \mathbf{v}$, where

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 2 & -1 & 1 & -2 \\ -3 & 4 & -4 & 3 \\ -2 & 1 & -1 & 2 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{B}^* \mathbf{B} = \begin{bmatrix} 18 & -16 & 16 & -18 \\ -16 & 18 & -18 & 16 \\ 16 & -18 & 18 & -16 \\ -18 & 16 & -16 & 18 \end{bmatrix}$$

Applying iteration (3),

$$\begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_3' & \mathbf{x}_4 \end{matrix}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 18 \\ -16 \\ 16 \\ -18 \end{bmatrix} \begin{bmatrix} 580 \\ -576 \\ 576 \\ -580 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 68 \\ -68 \\ 68 \\ -68 \end{bmatrix}, \qquad \lambda_{\max} = 68, \qquad \frac{1}{\lambda_{\max}} \approx 0{,}015$$

Matrix (4) is the following:

$$\mathbf{A}_1 = \mathbf{E} - 0{,}015\mathbf{A} = \begin{bmatrix} 0{,}73 & 0{,}24 & -0{,}24 & 0{,}27 \\ 0{,}24 & 0{,}73 & 0{,}27 & -0{,}24 \\ -0{,}24 & 0{,}27 & 0{,}73 & 0{,}24 \\ 0{,}27 & -0{,}24 & 0{,}24 & 0{,}73 \end{bmatrix}.$$

Iteration (5):

$$\begin{matrix} \mathbf{x}_1 \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{x}_2 \\ \begin{bmatrix} 0{,}73 \\ 0{,}24 \\ -0{,}24 \\ 0{,}27 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{x}_3 \\ \begin{bmatrix} 0{,}721 \\ 0{,}221 \\ -0{,}221 \\ 0{,}279 \end{bmatrix} \end{matrix}$$

The convergence is slow, therefore we apply (12):

$$\begin{matrix} \mathbf{x}_3 - \mathbf{x}_2 \\ \begin{bmatrix} -0{,}009 \\ -0{,}019 \\ 0{,}019 \\ 0{,}009 \end{bmatrix} \end{matrix}$$

Iteration (13) with $s=5$:

$$\begin{matrix} \mathbf{w}_1 = 20(\mathbf{x}_3 - \mathbf{x}_2) \\ \begin{bmatrix} -0{,}18 \\ -0{,}38 \\ 0{,}38 \\ 0{,}18 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{w}_2 \\ \begin{bmatrix} -0{,}265 \\ -0{,}261 \\ 0{,}261 \\ 0{,}265 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{w}_3 \\ \begin{bmatrix} -0{,}247 \\ -0{,}247 \\ 0{,}247 \\ 0{,}247 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{w}_4 \\ \begin{bmatrix} -0{,}232 \\ -0{,}232 \\ 0{,}232 \\ 0{,}232 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{w}_5 \\ \begin{bmatrix} -0{,}218 \\ -0{,}218 \\ 0{,}218 \\ 0{,}218 \end{bmatrix} \end{matrix}$$

As $\Delta\mathbf{w}_4 = \Delta\mathbf{w}_3 \cdot \frac{14}{15}$, so according to (15) $K_3 = \frac{14}{15} \approx 1$ and the same value is obtained for K_4. According to (14)

$$\mathbf{v}_1 = \mathbf{A}_1\mathbf{w}_3 - \mathbf{w}_3 = \frac{2}{1000}\begin{bmatrix} 7{,}41 \\ 7{,}41 \\ -7{,}41 \\ -7{,}41 \end{bmatrix}, \quad K_3 \frac{\mathbf{v}_1\mathbf{v}_1^*}{\mathbf{v}_1^*\mathbf{v}_1} = 0{,}25\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

The matrix $\mathbf{A}_1'$ according to (11), and the iteration with vector $\mathbf{x}_3$:

$$\mathbf{A}_1' = \begin{bmatrix} 0{,}48 & -0{,}01 & 0{,}01 & 0{,}52 \\ -0{,}01 & 0{,}48 & 0{,}52 & 0{,}01 \\ 0{,}01 & 0{,}52 & 0{,}48 & -0{,}01 \\ 0{,}52 & 0{,}01 & -0{,}01 & 0{,}48 \end{bmatrix};$$

$$\begin{matrix} \mathbf{x}_3 \\ \begin{bmatrix} 0{,}721 \\ 0{,}221 \\ -0{,}221 \\ 0{,}279 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{x}_4 \\ \begin{bmatrix} 0{,}487 \\ -0{,}013 \\ 0{,}013 \\ 0{,}513 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{x}_5 \\ \begin{bmatrix} 0{,}501 \\ -0{,}001 \\ 0{,}001 \\ 0{,}499 \end{bmatrix} \end{matrix} \quad \begin{matrix} \mathbf{x}_6 \\ \begin{bmatrix} 0{,}5 \\ 0 \\ 0 \\ 0{,}5 \end{bmatrix} \end{matrix}$$

Thus the first solution vector is obtained:

$$\mathbf{u}_1^* = [1 \quad 0 \quad 0 \quad 1]$$

By applying equality (6):

$$\mathbf{A}_2 = \mathbf{A}_1' - \frac{\mathbf{u}_1 \mathbf{u}_1^*}{\mathbf{u}_1^* \mathbf{u}_1} = \begin{bmatrix} -0{,}02 & -0{,}01 & 0{,}01 & 0{,}02 \\ -0{,}01 & 0{,}48 & 0{,}52 & 0{,}01 \\ 0{,}01 & 0{,}52 & 0{,}48 & -0{,}01 \\ 0{,}02 & 0{,}01 & -0{,}01 & -0{,}02 \end{bmatrix}$$

Iteration (7):

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -0{,}01 \\ 0{,}48 \\ 0{,}52 \\ 0{,}01 \end{bmatrix} \begin{bmatrix} 0{,}001 \\ 0{,}501 \\ 0{,}499 \\ -0{,}001 \end{bmatrix} \begin{bmatrix} 0 \\ 0{,}5 \\ 0{,}5 \\ 0 \end{bmatrix}$$

The second solution vector is: $\mathbf{u}_2^* = [0 \ 1 \ 1 \ 0]$. According to (8), matrix $\mathbf{A}_3$:

$$\mathbf{A}_3 = \mathbf{A}_2 - \frac{\mathbf{u}_2 \mathbf{u}_2^*}{\mathbf{u}_2^* \mathbf{u}_2} = \begin{bmatrix} -0{,}02 & -0{,}01 & 0{,}01 & 0{,}02 \\ -0{,}01 & -0{,}02 & 0{,}02 & 0{,}01 \\ 0{,}01 & 0{,}02 & -0{,}02 & -0{,}01 \\ 0{,}02 & 0{,}01 & -0{,}01 & -0{,}02 \end{bmatrix}$$

The eigenvalues of this matrix are all less than 1, thus no further solution vector exists.

REFERENCES

[1] ZURMÜHL, R.: *Matrizen*, Springer-Verlag, 1958.
[2] Фадеева, В. Н. и Фадеев, Д. К.: *Вычислительные методы линейной алгебры*, Москва—Ленинград, 1963.

Technical University, Miskolc

(Received November 10, 1964.)
(Revised Juli 20, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 19—42.

BINOMIAL AND HYPERGEOMETRIC GROUP-TESTING

by

M. SOBEL

1. Introduction

In binominal and hypergeometric group-testing the basic problem is to classify each of a fixed number N of units as satisfactory or defective. Corresponding to each unit we assign a random variable with the values zero or one according as the unit is satisfactory or defective, and we assume that these random variables are all independent and identically distributed. We shall distinguish problems (using the symbols B and H) according as our a priori knowledge gives us the value of the probability p of a unit being defective or the actual number of defectives D present among the N units at the outset. In each individual test we can include any number x of units ($1 \leqq x \leqq N$) but we also distinguish problems according to the type of results obtained by a single test. In the B-type the individual test informs us that either all x are good or at least one of the x is defective (but we don't know which ones or how many); in the H-type it tells us the exact number of defectives d_x present among the x units (but we don't know which ones they are, unless $d_x = x$). Thus we obtain four problems which we denote by BB, HB, BH and HH, where the first symbol refers to the a priori knowledge and the second refers to the type of results on a single test; we shall also refer to them as problems 1, 2, 3, 4, respectively. Problem BB has been considered by SOBEL and GROLL [8] and by SOBEL [9] and is included in the present paper only for the purpose of making comparisons. We wish to show that these problems can be handled by a common technique which we call procedure R_1 and we are interested in the interrelations between these problems. For example, using the appropriate procedure R_1 (defined below) the expected number of tests required for problem BB is an upper bound for the expected number of tests for both problems HB and BH; and similarly HH provides a lower bound. Problem BH is closely related to HH since, if we perform a single test at the outset on all N units, we can determine the number D of defectives; the knowledge of p then becomes superfluous and the problem reduces to an HH problem from that point on. Problems BH and HH can be regarded as coin weighing problems in which all the defective coins weigh the same (too heavy or too light) provided the scale used is an ordinary weighing scale and not a two-arm balance. For the corresponding HH problem with a two-arm balance see CAIRNS [2] and BELLMAN and GLUSS [1].

It is also possible to replace the a priori information by a (weaker) a priori distribution or by no information at all. Letting B* denote the former and a blank the latter, this gives problems B*B, B, B*H and H. Problems B*B and B were considered by SOBEL and GROLL [11]. An information-theoretic analysis of problem H (without the assumption of independent, identically distributed random variables) has been under active investigation by several workers as indicated in the remark at the end of the paper by ERDŐS and RÉNYI [3]. In particular, LINDSTRÖM [5] has

shown that for the problem H with D unknown

$$\lim_{N\to\infty} \frac{A(N|H)\log_2 N}{N} = 2 \tag{1.1}$$

where $A(N|\mathrm{H})$ is the expected number of tests using the optimal "non-sequential" procedure, a non-sequential procedure being one which specifies which units should be tested in all the tests (or which questions should be asked throughout the interrogation) before any experimentation begins. The close relation between the H and HH problems stems from the fact that under H we could test all the units on the first test and from there on we have a HH problem. Hence if $A(N, D|\mathrm{HH})$ denotes the expected number of tests using the optimal procedure for the HH problem and $A_s(N|\mathrm{H})$ is the optimal sequential result for the H problem, then

$$A_s(N|\mathrm{H}) \leqq \operatorname*{Max}_{0\leqq D\leqq N} A(N, D|\mathrm{HH}) + 1, \tag{1.2}$$

and it is conjectured that the maximum in (1. 2) is attained for D equal to the integer (or integers) closest to $N/2$.

The question of optimality of the procedure R_1 for the HB problem is the main topic of this paper. In particular, it is proved that for problem HB this procedure is optimal for $D=2$ and an infinite sequence of starting values of N; this is unlike the results for BB in [8] (see also [9] and [10]) where it is shown that the procedure R_1 is not optimal for values of p close to zero. For $D=1$ and any N it is shown that the procedure R_1 for the HB problem is the same as the halving procedure which is known ([7], [12]) to be optimal. Since the HUFFMAN lower bound is not always attainable in group-testing problems and since the mixing of units does not improve the expected number of tests in the HB problem, R_1 is optimal for all starting values N and any D.

We assume throughout that tests are conducted one at a time and that the results of any test are available before any subsequent test is started. The proposed procedure R_1 is sequential in the sense that the present test can and does depend on the results of previous tests.

2. Notation and Preliminaries for Problem HB

The total number N of units to be classified, the number D of defectives among them, and hence the number $S = N-D$ of satisfactory units at the outset are all given. At any stage of the procedure let n, d, s, with $n=d+s$, denote respectively the number of units not yet classified, the number of defectives among these n units, and the number of satisfactory units among them.

The procedure R_1 is implicitly defined by two recursion formulas and some trivial boundary conditions. It has the property that at any stage all the relevant information about previous tests is retained by merely separating the n units not yet classified into at most 2 sets. If there are 2 sets present then one of these sets of size m (where $2\leqq m\leqq s$) is known to contain at least one defective unit and we refer to this set as the defective set; the other set of size $n-m$ will then be called the remainder (or hypergeometric) set if $d>1$. The terminology "H-situation"

when there is only 1 set present ($m=0$) and "G-situation" when there are 2 sets present ($m \geqq 2$) is a carry-over from the BB problem in [8].

In an H-situation with parameters n, d, s, the probability $Q_H(x|d, s)$ that a sample of size x (chosen at random from the n units) contains no defectives is

$$Q_H(x|d,s) = \binom{s}{x}\Big/\binom{n}{x} = \binom{n-x}{d}\Big/\binom{n}{d} \tag{2.1}$$

and the probability $P_H(x|d, s)$ that it contains at least 1 defective is simply $1-Q_H(x|d, s)$.

In a G-situation with parameters m, d, s, let Y denote the number of defectives in the defective set and let X denote the number of defectives in a sample of size x drawn (at random) from the defective set. Then

$$\mathrm{P}\{X \geqq 1|Y \geqq 1\} = \frac{\mathrm{P}\{X \geqq 1\}}{\mathrm{P}\{Y \geqq 1\}} = \frac{1-\left[\binom{s}{x}\Big/\binom{n}{x}\right]}{1-\left[\binom{s}{m}\Big/\binom{n}{m}\right]} = \frac{\binom{s+d}{d}-\binom{s+d-x}{d}}{\binom{s+d}{d}-\binom{s+d-m}{d}} \tag{2.2}$$

and we denote this by $P_G(x|m, d, s)$. The corresponding expression for $Q_G(x|m,d,s) = 1-P_G(x|m, d, s)$ is

$$\mathrm{P}\{X = 0|Y \geqq 1\} = \frac{\left[\binom{s}{x}\Big/\binom{n}{x}\right]-\left[\binom{s}{m}\Big/\binom{n}{m}\right]}{1-\left[\binom{s}{m}\Big/\binom{n}{m}\right]} = \frac{\binom{s+d-x}{d}-\binom{s+d-m}{d}}{\binom{s+d}{d}-\binom{s+d-m}{d}}, \tag{2.3}$$

Let $G(m; d, s) = G(m; d, s|R_1, \mathrm{HB})$ denote the expected number of tests required under procedure R_1 if we start with a G-situation with parameters m, d, s and let $H(d, s) = H(d, s|R_1, \mathrm{HB})$ denote the same quantity for an H-situation with parameters d and s.

Procedure R_1 for Problem HB

The recursion formulas defining procedure R_1 are for $d \geqq 1$ and $s \geqq 1$

$$H(d,s) = 1 + \operatorname*{Min}_{1 \leqq x \leqq s} [Q_H(x|d,s)H(d,s-x)+P_H(x|d,s)G(x;d,s)] \tag{2.4}$$

and for $m \geqq 2$, $d \geqq 2$, $s \geqq 1$

$$\begin{aligned} G(m;d,s) = 1 + \operatorname*{Min}_{1 \leqq x \leqq m-1} [&Q_G(x|m,d,s)G(m-x;d,s-x)+ \\ &+P_G(x|m,d,s)G(x;d,s)] \end{aligned} \tag{2.5}$$

where Q_H, $P_H = 1-Q_H$, P_G and Q_G are given by (2. 1), (2. 2) and (2. 3).

The boundary conditions for this recursion are

$$H(0,s) = H(d,0) = 0 \qquad \text{for all} \quad d \geqq 0,\ s \geqq 0. \tag{2.6}$$

$$G(1;d,s) = H(d-1,s) \qquad \text{for all} \quad d \geqq 2,\ s \geqq 0. \tag{2.7}$$

$$G(m;1,s) = H(1,m-1) \qquad \text{for all} \quad s+1 \geqq m \geqq 1. \tag{2.8}$$

In particular, the boundary conditions take care of the cases in which $m=1$ or where there is a change in the value of d.

REMARK 1. In writing (2. 5) it is assumed that in a G-situation all the x units are taken from the defective set of size m in a nested manner, i.e. without mixing units from the two sets. This assumption characterizes the procedure R_1. In the BB problem this assumption was responsible for the lack of optimality; we shall be interested below in seeing whether it also causes a lack of optimality in the HB problem.

REMARK 2. If at any stage in the HB problem we reach a situation with $d=1$ then we have only H-situations from that point on and every test identifies good units. From that point on, it is also an HH-problem with $D=1$ and, as mentioned earlier, it is known that the so-called "halving-procedure" is optimal for this case; we show below that the procedure R_1 is the halving procedure as soon as we get d equal to 1. Assuming this result, we can then rewrite (2. 4) for $d=1$ in this problem (and also for the HH-problem with $D=1$) in the following simpler form. If s is odd (say, $s=2t-1$, where $t\geqq 1$)

$$H(1, 2t-1) = 1+H(1, t-1) \tag{2. 9}$$

and if s is even (say, $s=2t$, where $t\geqq 1$)

$$H(1, 2t) = 1+\frac{t}{2t+1}H(1, t-1)+\frac{t+1}{2t+1}H(1, t), \tag{2. 10}$$

the boundary condition is $H(1, 0)=0$. An explicit expression for $H(1, s)$ is given in (2. 22) below.

REMARK 3. In writing $G(x; d, s)$ at the end of (2. 5) we used the following result. If x units taken (at random) from a defective set of size m contain at least one defective then the $m-x$ remaining units can be combined with the remainder set of size $s+d-m$ without any "loss of information". In other words, the conditional distribution is such that we again have a G-situation with only two sets to keep track of. A proof of this is given in the Appendix to this paper. This process of combining the $m-x$ and $s+d-m$ units will be called recombination.

For any $s\geqq 1$ we let $s'=s+1$ and define the integers $b=b(s')$ and $c=c(s')$ by

$$s'=2^b+c \qquad (0\leqq c<2^b). \tag{2. 11}$$

Consider the explicit nonnegative function

$$h_1(s') = bs'+2c = (b+2)s'-2^{b+1} \tag{2. 12}$$

which is clearly an increasing function of s; we define $h_1(1)$ to be 0.

If we set $d=1$ in (2. 4), use (2. 1), (2. 2) and (2. 3) for Q_H, P_H, Q_G and P_G, and set $xH(1, x-1) = h(x)$ then we obtain for $s'\geqq 2$

$$h(s') = s'+\underset{1\leqq x\leqq s'-1}{\text{Min}}\,[h(x)+h(s'-x)] \tag{2. 13}$$

with boundary condition $h(1)=0$. We now wish to show

LEMMA 1. The function $h(s')$ in (2.13) with $h(1)=0$ is identical with $h_1(s')$ given by (2.12) with $h_1(1)=0$. For $s' \geqq 1$ the set of x values that minimize the right hand side of (2.13) include the integer (or integers) closest to $s'/2$.

PROOF. We will show that $h_1(s')$ satisfies (2.13) and, since the boundary condition is the same and (2.13) then determines $h(s')$ for all integers $s' > 1$, the result will follow. Suppose first that $h_1(x)+h_1(s'-x)$ is minimized by taking x as close as possible to $s'/2$. Then for even $s' \geqq 2$ we put $h_1(s'/2)$ on the right side of (2.13) and obtain for $x=s'/2$

$$s'+2\left[(b-1)\frac{s'}{2}+2\left(\frac{s'}{2}-2^{b-1}\right)\right] = bs'+2(s'-2^b) = h_1(s'), \tag{2.14}$$

so that (2.13) with h replaced by h_1 is satisfied for s' even. For odd $s' \geqq 3$ we take $x = (s'-1)/2$ and it is easy to verify that $b(x) = b-1$ and if $s' \neq 2^{b+1}-1$ then $b(s'-x) = b\big((s'+1)/2\big) = b-1$. Then the right side of (2.13) gives

$$\begin{aligned} s'+(b-1)\frac{s}{2}+2\left(\frac{s}{2}-2^{b-1}\right)+\left[(b-1)\left(\frac{s'+1}{2}\right)+2\left(\frac{s'+1}{2}-2^{b-1}\right)\right] = \\ = bs'+2(s'-2^b) = h_1(s'). \end{aligned} \tag{2.15}$$

If $s' = 2^{b+1}-1$ then the bracketed part B of (2.15) becomes

$$\left(\frac{s'+1}{2}\right)b+2\left(\frac{s'+1}{2}-2^b\right) = (b-1)\left(\frac{s'+1}{2}\right)+2\left(\frac{s'+1}{2}-2^{b-1}\right) = B \tag{2.16}$$

and thus we obtain $h_1(s')$ again. Hence $h_1(s')$ satisfies the recursion (2.13) with $x=[s'/2]$ and if our above supposition is true it follows that $h_1(s')=h(s')$ for all $s' \geqq 1$.

We now show that the values of x closest to $s'/2$ minimize

$$h_2(x; s') = h_1(x)+h_1(s'-x) \qquad (1 \leqq x \leqq s), \tag{2.17}$$

where $h_1(x)$ is given by (2.12). Clearly, for any $s' \geqq 1$, $h_2(x; s')$ is symmetric about $x=s'/2$. Hence it suffices to show that $h_2(x; s')$ is convex on the domain of integers x where $1 \leqq x \leqq s'/2$. For this convexity it is sufficient to show that $h_1(x)$ is convex for $x \geqq 2$, i.e., that for any 3 consecutive integers $x-1$, x, $x+1$ with $x \geqq 2$ the second difference $\Delta^2 h_1(x) = h_1(x+1)-2h_1(x)+h_1(x-1) \geqq 0$. For $x \geqq 2$ three cases arise according as for some $z \geqq 0$

$$\text{(i) } x = 2^z; \quad \text{(ii) } x = 2^z-1; \quad \text{(iii) } 2^{z-1} < x < 2^z-1. \tag{2.18}$$

For case (i) we obtain

$$\begin{aligned} \Delta^2 h_1(x) = z(2^z+1)+\big(2(2^z+1-2^z)\big)-2[z2^z+2(2^z-2^z)]+ \\ +(z-1)(2^z-1)+2(2^z-1-2^{z-1}) = 1 > 0. \end{aligned} \tag{2.19}$$

Similarly for cases (ii) and (iii), respectively, we obtain

$$\Delta^2 h_1(x)=0 \quad \text{and} \quad \Delta^2 h_1(x)=6x>0 \quad \text{for} \quad x \geqq 2. \tag{2.20}$$

This proves the convexity of $h_1(x)$ for $1 \leqq x \leqq s$ and hence $h_1(x)$ assumes its minimum at the integer(s) closest to $s'/2$.

To complete the proof of lemma 1 we use induction and assume that $h(x) = h_1(x)$ for all integers $x < s'$. Then by (2.13)

$$(2.21) \quad h(s') - s' = \operatorname*{Min}_{1 \leqq x \leqq s} [h(x) + h(s'-x)] = \operatorname*{Min}_{1 \leqq x \leqq s} [h_1(x) + h_1(s'-x)] = h_1(s') - s'.$$

Since $h(1) = h_1(1) = 0$, this proves the lemma.

It follows from this lemma that for $d = 1$ the procedure R_0 (defined in the next section) is equivalent to the halving procedure which is known to be optimal for $d = 1$ (see [7] and [12]).

In terms of the original notation the above lemma tells us that for $s \geqq 1$ (and also for $s = 0$)

$$(2.22) \qquad H(1, s) = b + \frac{2(s+1-2^b)}{s+1} = b + 2 - \frac{2^{b+1}}{s+1}$$

where b and c are defined by (2.11) with $s' = s+1$.

3. A Characterization of Procedure R_1 for the HB-Problem

In this section we define a procedure R_0 in an explicit manner and it will be shown in theorem 4 (in section 5 below) that R_0 and R_1 have the same expected number of tests and if the sample sizes of R_1 are unique these two procedures are identical (it is conjectured that they are identical if we start with any H-situation). We shall consider here only the case in which the initial number of defectives $D = 2$; S is arbitrary. The procedure R_1 has already been characterized when the number of remaining defectives $d = 1$ and R_0 is defined to be the "halving procedure" for $d = 1$; hence we only need to define R_0 for $d = 2$. For more generality we consider any H-situation and replace the initial S by the number of remaining satisfactory units s.

For $s' = s+1 \geqq 2$ let $b = b(s')$ and $c = c(s')$ be non-negative integers defined exactly as in (2.11) Using (3.1) we define $x_H(s) = x_H(2, s)$ by setting $x_H(0) = 0$, $x_H(1) = 1$ and for $s \geqq 2$ by

$$(3.1) \qquad x_H(s) = \begin{cases} \left[2^{b-2} + \dfrac{c+1}{2}\right] & \text{for} \quad 2^b \leqq s' < 3 \cdot 2^{b-1} \\ 2^{b-1} & \text{for} \quad 3 \cdot 2^{b-1} \leqq s' < 2^{b+1} \end{cases}$$

where $[y]$ is the largest integer less than or equal to y. It is easy to verify that $x_H(s)$ is a nondecreasing function of s (in particular, it is the same for $s' = 3 \cdot 2^{b-1} - 1$ and $s' = 3 \cdot 2^{b-1}$), that $x_H(s)$ is the minimum of the two expressions in (3.1) and finally that for all $s \geqq 1$

$$(3.2) \qquad \frac{s+1}{4} \leqq x_H(s) \leqq \frac{s+2}{3} = \frac{n}{3}.$$

Under the proposed procedure R_0 the same size $x(s)$ for the H-situation with $d = 2$ and $n = 2+s$ is given by (3.1).

To describe the procedure R_0 for the G-situation with $d = 2$ and general m, s

with $s \geqq m \geqq 2$ we first note that a non-trivial G-situation, i.e., with $m \geqq 2$, can only arise from an H-situation with $s \geqq 4$; this is so because $x_H(s) = 1$ for $s < 4$. Define the integers $p = p(m)$ and $r = r(m)$ for $m \geqq 2$ by

$$m = 2^p - r \qquad 0 \leqq r < 2^{p-1}; \tag{3.3}$$

for $m = 1$ we can write $p = r = 0$ or $p = r = 1$. For $m \geqq 1$ (and hence $s \geqq 1$) we define the vector $\vec{v}_m = \vec{v}(m; 2) = (v_1, v_2, \ldots, v_m)$ by setting

$$v_\alpha = \begin{cases} p-1 & \text{for} \quad \alpha = 1, 2, \ldots, r \\ p & \text{for} \quad \alpha = r+1, r+2, \ldots, m. \end{cases} \tag{3.4}$$

We denote that this definition of $\vec{v}_m$ does not depend on s except possibly for $s \leqq 3$ when $m = 1$. Every such vector $\vec{v}_m$ has the property that

$$\sum_{\alpha=1}^{m} 2^{-v_\alpha} = \frac{r}{2^{p-1}} + \frac{m-r}{2^p} = 1 \tag{3.5}$$

and we shall call this sum the value of the vector $\vec{v}_m$.

It is clear from (3. 5) and the above that for $m \geqq 2$ we can always subdivide $\vec{v}_m$ into subvectors $\vec{v}_j$ consisting of the first j components and $\vec{v}_{m-j}$ consisting of the last $m-j$ components so that each has the value 1/2. This process can be repeated on any subvector that has at least 2 components and the smallest possible value is clearly 2^{-p}.

We now complete the explicit description of procedure R_0 for the G-situation in terms of these subdivisions. In the G-situation, i.e., after a test on x_i units has failed and we set $m = x_i$, we test j units from the defective set of size m where j is defined above; let $\vec{v}_j$ and $\vec{v}_{m-j}$ be the left and right subvectors formed. If the test on j units succeeds then these are classified and we consider only $\vec{v}_{m-j}$ for subsequent tests; if it fails then we label these j units as a new defective set and we consider $\vec{v}_j$ for the next test since the $m-j$ units in $\vec{v}_{m-j}$ are now eligible for recombination with all the unclassified units outside the new defective set. In either case the procedure R_0 is to continue subdividing the newly formed subvectors until the subvector to be considered for subsequent tests contains only one component. We then continue this process further (testing a single unit at a time) until a defective unit is found, either by a direct test or by inference. This part of the procedure always leads to the classification of exactly one defective unit.

After a defective unit is found we are back in an H-situation and the whole process (including the determination of x values from the new s-value) starts all over again with the number of defectives reduced by one and the number of satisfactory units reduced by a random integer.

4. Explicit Formulas for R_0

Although we are interested in evaluating R_0 for all values of s, we shall be particularly interested in a sequence s_j of s values in which $s_1 = 1$, $s_2 = 2$ and s_j for $j \geqq 3$ is defined recursively by

$$s_{j+2} = \begin{cases} 2s_j + 1 & \text{for} \quad j = 4i+1 \quad \text{and} \quad 4i+2 \\ 2s_j + 2 & \text{for} \quad j = 4i+3 \quad \text{and} \quad 4i+4 \end{cases} \tag{4.1}$$

by letting $i=0, 1, 2, \ldots$. If we break up the sequence into groups of 4 and write each subgroup vertically, then some numerical values of the s_j after $s_0=0$ are

$$\begin{array}{ccccc} 1, & 8, & 36, & 148, & 596, \ldots \\ 2, & 12, & 52, & 212, & 852, \ldots \\ 3, & 17, & 73, & 297, & 1193, \ldots \\ 5, & 25, & 105, & 425, & 1705, \ldots \end{array} \tag{4. 2}$$

Using (4. 1) it is easy to obtain explicit expressions for s_j for each of the 4 types of j-values above; for odd and even j, respectively, these can be written as

$$s_j = \begin{cases} 2^{(j+1)/2} + \left[\dfrac{2^{(j-1)/2}-4}{3}\right] & \text{for odd} \quad j \geqq 1 \\ 2^{j/2} + \left[\dfrac{2^{(j+2)/2}-4}{3}\right] & \text{for even} \quad j \geqq 0, \end{cases} \tag{4. 3}$$

where $[x]$ was already defined in (3.1). For $j=4i+1$ and $4i+2$ the bracket signs in (4. 3) can be removed without any change and for $j=4i+3$ and $4i+4$ the bracket signs can be removed if we replace the "4" by a "5". For $j \geqq 1$, the expression in (4. 3) gives the correct b and c-values (say, b_j and c_j) for $s'_j = s_j+1$ as defined in (2.11). In particular, we note that $b_j = b(s'_j) = [(j+1)/2]$ for $j \geqq 0$. It is easily verified that for $s_j \geqq 1$ (or $j \geqq 1$)

$$\begin{aligned} s'_j &< 3 \cdot 2^{(j-1)/2} \quad \text{for} \quad j \quad \text{odd} \\ s'_j &\geqq 3 \cdot 2^{(j-2)/2} \quad \text{for} \quad j \quad \text{even} \end{aligned} \tag{4. 4}$$

and hence $x(s_j)$ takes its value from the first line of (3.1) when j is odd and from the second line in (3.1) when j is even. Using (3.1) and (4.3) we then obtain for $j \geqq 2$

$$x_H(s_j) = \begin{cases} 2^{(j-1)/2} - \left[\dfrac{2^{(j-1)/2}+1}{3}\right] & \text{for odd} \quad j \\ 2^{(j-2)/2} & \text{for even} \quad j \end{cases} \tag{4. 5}$$

and we note that the same expression holds for $j=1$.

We wish to establish 3 properties of the numbers s_j; these properties will help to characterize the procedure R_0 (and hence also R_1 if R_1 is identical with R_0).

PROPERTY 1. *For every* $j \geqq 1$

$$s_j - x_H(s_j) = s_{j-1}. \tag{4. 6}$$

PROOF. For odd $j = 4i+1$ we get from (4. 3) and (4. 5) for the left side of (4. 6)

$$2^{(j-1)/2} + \left[\frac{2^{(j-1)/2}-4}{3}\right] + \left[\frac{2^{(j-1)/2}+1}{3}\right] = 2^{(j-1)/2} + \left(\frac{2^{(j+1)/2}-5}{3}\right), \tag{4. 7}$$

subtracting 2/3 to make the 2nd square bracket an integer; this is the same as the result for s_{j-1} for even $(j-1)$ in (4. 3). For odd $j = 4i+3$ we subtract 1/3 from the first square bracket argument (and 0 from the second) so that 1/3 is added to the

last expression in (4. 7); this gives the expression for s_{j-1} in (4. 3) for even $j-1$. For even $j \geqq 2$ we obtain for the left side of (4. 6)

$$2^{(j-2)/2} + \left[\frac{4 \cdot 2^{(j-2)/2} - 4}{3}\right] = 2^{j/2} + \left[\frac{2^{(j-2)/2} - 4}{3}\right] = s_{j-1}.$$

PROPERTY 2. *For every* $j \geqq 0$

$$2^j \leqq \binom{s_j + 2}{2} < 2^{j+1} \tag{4. 8}$$

i.e., the b-value for $\binom{s_j+2}{2}$ *is* j.

PROOF. For odd j, using (4. 3), the inequalities (4. 8) reduce to

$$2^j \leqq \frac{98y_1^2 \pm 7y_1 - 1}{9} < 2^{j+1} \tag{4. 9}$$

where the $+$ and $-$ correspond to $j = 4i+1$ and $4i+3$, respectively, and $y_1 = 2^{(j-3)/2}$. Using the $+$ sign with the upper inequality in (4. 9), dropping the -1 yields

$$7y_1 < 46y_1^2 \quad \text{or} \quad y_1 > 7/46,$$

which holds for all odd $j \geqq 1$. Using the negative sign with the lower inequality in (4. 9) yields

$$26y_1^2 - 7y_1 - 1 \geqq 0,$$

which holds for all odd $j \geqq 1$.

For even j the inequalities (4. 8) reduce to

$$2^j \leqq \frac{25 \cdot 2^{j-1} \pm 5 \cdot 2^{(j-2)/2} - 1}{9} < 2^{j+1} \tag{4. 10}$$

where the $+$ and $-$ sign correspond to $j = 4i+2$ and $4i+4$, respectively. Letting $y_2 = 2^{(j-2)/2}$ we obtain as above for the upper and lower inequalities in (4. 10), respectively

$$y_2 > \frac{5}{22} \quad \text{and} \quad 14y_2^2 - 5y_2 - 1 \geqq 0 \quad \text{or} \quad y_2 \geqq 1/2,$$

both of which hold for all even $j \geqq 0$. In fact, we note that equality is attained only for $j = 0$.

PROPERTY 3. *For* $j \geqq 2$

$$2^{j-1} \leqq \binom{s_j + 2}{2} - \binom{s_{j-1} + 2}{2} < 2^j. \tag{4. 11}$$

PROOF. Letting M_j denote the middle expression in (4. 11) and using the above notation, we have

$$M_j = \begin{cases} \dfrac{y_3^2 \pm y_3}{3} & \text{for odd } j \\ \dfrac{34y_4^2 \pm y_4}{3} & \text{for even } j, \end{cases} \tag{4. 12}$$

where $y_3 = 2^{(j+1)/2}$ and $y_4 = 2^{(j-4)/2}$. Using (4. 12) it is easy to verify that for odd j we need $j \geqq 3$ to satisfy (4.11) and for even j we need $j \geqq 2$ to satisfy (4.11). In fact the equality in (4.11) holds only for $j = 3$.

From the description of procedure R_0 in section 3 we now obtain explicit formulas for the test group size $x_G = x_G(m; 2, s)$ for any G-situation. Then we obtain explicit formulas for $G_0(m; 2, s)$ and for $H_0(2, s)$ under the procedure R_0. With the help of the above 3 properties we then show that the formula for $H_0(2, s)$ simplifies for $s = s_j$ in the infinite sequence (4.1).

Consider the vector $\vec{v}_m$ defined in terms of m in (3. 4). If $r/2^{p-1} \geqq 1/2$ (or $m \leqq 3 \cdot 2^{p-2}$) then by procedure R_0 the next test-group size x is the root of

$$\frac{x}{2^{p-1}} = 1/2 \quad \text{or} \quad x = 2^{p-2}.$$

If $r/2^{p-1} \leqq 1/2$ (or $m \geqq 3 \cdot 2^{p-2}$) then by procedure R_0 the next test group size x is the root of

$$\frac{m-x}{2^p} = 1/2 \quad \text{or} \quad x = m - 2^{p-1} > 0.$$

Thus we can write for procedure R_0

$$x_G(m; 2, s) = \begin{cases} 2^{p-2} & \text{for} \quad 2^{p-1} \leqq m < 3 \cdot 2^{p-2} \\ m - 2^{p-1} & \text{for} \quad 3 \cdot 2^{p-2} \leqq m < 2^p. \end{cases} \tag{4.13}$$

and we denote it by $x_G(m; 2)$ since it does not depend on s. From (4. 13) we note that

$$\frac{m}{3} \leqq x_G(m; 2) \leqq \frac{m}{2} \tag{4.14}$$

and moreover

$$\begin{aligned} 2^{p-2} \leqq m - x < 2^{p-1} \quad &\text{when} \quad x = 2^{p-2} \\ 2^{p-2} \leqq x < 2^{p-1} \quad &\text{when} \quad m - x = 2^{p-1}, \end{aligned} \tag{4.15}$$

so that the largest powers of 2 contained in x and $m - x$ are $p-2$ and $p-1$.

After some preliminaries we now develop an explicit expression for $H_0(2, s)$ under procedure R_0 for any $s = s_j$ in the infinite sequence (4.1). Let $m_j = x_H(2, s_j)$ and let $b_j = b(s_j + 1)$ as defined in (3.1) and (2.11). Let $\vec{v}_j = \vec{v}(m_j; 2, s_j) = (v_{1j}, v_{2j}, \ldots, v_{m_j j})$; we need to discuss the significance of the components $v_{\alpha j}$ in the procedure R_0. The component $v_{\alpha j}$ of $\vec{v}_j$ corresponds to the α-th unit ($\alpha = 1, 2, \ldots, m_j$), after the units have been randomized and put in an arbitrary fixed order. If the first defective is in position α then from the description of R_0 it will take exactly $v_{\alpha j}$ tests, starting with the G-situation denoted by $G(m_j; 2, s_j)$, to discover this defective unit and thus get back to an H-situation. The probability that the first defective is in the α-th position is easily seen to be $(s_j + 2 - \alpha) f_j^{-1}$ where $f_j = \binom{s_j + 2}{2}$. Hence using (4. 6) we can write for $j \geqq 4$

$$f_j H_0(2, s_j) = f_j + f_{j-1} H_0(2, s_{j-1}) + \sum_{\alpha=1}^{m_j} (s_j + 2 - \alpha)\{v_{\alpha j} + H_0(1, s_j + 1 - \alpha)\}. \tag{4.16}$$

Actually (4.16) also holds for $j = 1, 2$ and 3 if we define $v_{1j} = 0$ for $1 \leqq j \leqq 3$; this

is reasonable since $m_j = x(s_j) = 1$ for these j-values. If we now iterate (4.16) on j then, using the definition of $h(x)$ in (2.13), we have for $j \geqq 1$

$$f_j H_0(2, s_j) = \sum_{i=1}^{j} f_i + \sum_{i=1}^{j} \sum_{\alpha=1}^{m_i} (s_i + 2 - \alpha) v_{\alpha i} + \sum_{\beta=1}^{s_j} h(\beta+1). \tag{4.17}$$

Similarly, using the fact that $v_{\alpha i} = p_i - 1$ for $\alpha = 1, 2, \ldots, r_i$ and $v_{\alpha i} = p_i$ for the remaining $m_i - r_i$ values of i, we can combine the first single sum and the double sum in (4.16) to obtain for any $j \geqq 1$

$$f_j H_0(2, s_j) = \sum_{\alpha=1}^{j} \binom{t_\alpha + 2}{2} + \sum_{\beta=1}^{s_j} h(\beta+1) + \sum_{i=1}^{j} p_i (f_i - f_{i-1}) \tag{4.18}$$

where $t_\alpha = s_\alpha - r_\alpha$, $s_0 = p_1 = p_2 = 0$ and $p_3 = 1$. Comparing (4.3) and (4.5) we note that $p_i = b_i - 1$ for $i \geqq 4$ and we define $p_1 = p_2 = 0$ and $p_3 = 1$ so that this relation also holds for $i = 1, 2$ and 3. We note from (4.3) or (4.5) that p increases by one when (and only when) i changes from even to odd and that $t_\alpha = s_\alpha$ for even α; hence (4.18) can be written as

$$f_j H_0(2, s_j) = p_j f_j + \sum_{\alpha=1}^{j} \binom{t_\alpha + 2}{2} - \sum_{\alpha=1}^{\left[\frac{j-1}{2}\right]} f_{2\alpha} + \sum_{\beta=1}^{s_j} h(\beta+1) = \tag{4.19}$$

$$= \left[\frac{j}{2}\right] f_j + \sum_{\alpha=1}^{\left[\frac{j+1}{2}\right]} \binom{t_{2\alpha-1} + 2}{2} + \sum_{\beta=1}^{s_j} h(\beta+1).$$

By (4.3) we find that

$$t_{2\alpha-1} = \begin{cases} 2^\alpha - 1 & \text{for } \alpha \text{ odd} \\ 2^\alpha - 2 & \text{for } \alpha \text{ even} \end{cases} \tag{4.20}$$

and hence the first sum F_j in the last expression in (4.19) becomes for odd $j \geqq 1$

$$F_j = \frac{2^{j+2} \pm 2^{\frac{j+1}{2}} - 1}{3} \tag{4.21}$$

where the $+$ sign holds for $j = 4i+1$ and the $-$ sign holds for $j = 4i+3$. For even j we use (4.21) with j replaced by $j-1$, applying the above "sign rule" to $j-1$.

Using (2.12) we find after some straightforward summation of series that

$$\sum_{\beta=1}^{s_j} h(\beta+1) = (b_j + 2) f_j + \frac{2^{2b_j+1} + 1}{3} - (2s_j + 3) 2^{b_j} \qquad (j \geqq 1). \tag{4.22}$$

We now insert this result in (4.19); using (4.3) to substitute for s_j and b_j in terms of j we obtain

$$f_j H_0(2, s_j) = \begin{cases} (j+2) f_j + F_j - \left(\dfrac{2^{j+3} - 1}{3}\right) - 2^{(j+1)/2} \left[\dfrac{2^{(j+1)/2} + 1}{3}\right] & \text{for } j \text{ odd} \\ (j+2) f_j + F_{j-1} - \left(\dfrac{2^{j+2} - 1}{3}\right) - 2^{j/2} \left[\dfrac{2^{(j+4)/2} + 1}{3}\right] & \text{for } j \text{ even.} \end{cases} \tag{4.23}$$

If we now consider the 4 cases for j and substitute for F_j from (4. 21) then we obtain for $j \geqq 1$ in all 4 cases the same simple result

$$H_0(2, s_j) = j+2-\frac{2^{j+1}}{f_j}. \tag{4. 24}$$

We note from property 2 in (4. 8) that 2^j is the largest power of 2 contained in f_j and hence by (4. 24)

$$j = [\log_2 f_j] \leqq H_0(2, s_j) < 1+[\log_2 f_j] = j+1. \tag{4. 25}$$

Using the monotonicity of $H_0(2, s)$ as a function of s we have

$$\lim_{j\to\infty} H_0(2, s_j) = \lim_{s\to\infty} H_0(2, s) \approx 2\log_2 s. \tag{4. 26}$$

REMARK. Suppose that $F(x, s) = \binom{s+2}{2} - \binom{s+2-x}{2}$ and $F(m-x, s-x,)$ the numerators of P_G and Q_G in (2. 2) and (2. 3), respectively, have the same highest power of 2, say $\beta-1$, contained in them. If the formula

$$G_0(m_1, 2, s_1) = q+2-\frac{2^{q+1}}{F(m_1, s_1)}, \tag{4. 27}$$

which is analogous to (4. 24), holds for $s_1 < s$ and all $m_1 (2 \leqq m_1 \leqq s_1)$ with $q = [\log_2 F(m_1, s_1)]$, then, using the fact that $F(m, s) = F(x, s)+F(m-x, s-x)$ so that $2^\beta \leqq F(m, s) < 2^{\beta+1}$, we have from (2. 5)

$$\begin{aligned} G_0(m; 2, s) &= 1+\frac{F(m-x, s-x)}{F(m, s)}\left(\beta+1-\frac{2^\beta}{F(m-x, s-x)}\right)+ \\ &+\frac{F(x, s)}{F(m, s)}\left(\beta+1-\frac{2^\beta}{F(x, s)}\right) = \beta+2-\frac{2^{\beta+1}}{F(m, s)}. \end{aligned} \tag{4. 28}$$

If $[\log_2 F(x, s)] = \beta_1 < \beta_2 = [\log_2 F(m-x, s-x)]$ then $\beta = [\log_2 F(m, s)] = \beta_2$ and the result in (4. 28) becomes

$$\begin{aligned} &\beta+2-\frac{2^{\beta+1}}{F(m, s)}+\frac{F(m-x, s-x)-2^{\beta+1}-F(x, s)(\beta_2-\beta_1-1)}{F(m, s)} > \\ &> \beta+2-\frac{2^{\beta+1}}{F(m, s)}+\frac{\{2^{\beta_2-\beta_1-1}-(\beta_2-\beta_1)\}2^{\beta_1+1}}{F(m, s)} \end{aligned} \tag{4. 29}$$

where the last term is nonnegative since $x \leqq 2^{x-1}$ for any integer $x \geqq 1$. Moreover $F(m-x, s-x) = (m-x)(2s+3-m-x)/2$ cannot be a power of 2 since the 2 factors have different parity and hence the inequality in (4. 29) is strict. It follows that the result in (4. 28) is a lower bound under R_0 for any m, s. Similarly we find from (2. 4) and (2. 8) that for any s

$$H_0(2, s) \geqq \beta+2-\frac{2^{\beta+1}}{f} \tag{4. 30}$$

where $f = \binom{s+2}{2}$ and $\beta = [\log_2 f]$; strict inequality holds in (4. 30) if $\left[\log_2 \binom{s+2-x}{2}\right] \neq [\log_2 F(x, s)]$, where x is the next group test size. Comparing with (4. 24), we note that procedure R_0 meets this lower bound (4. 30) for every s_j in the infinite sequence (4. 1).

From the explicit results obtained in section 5 below for $G_0(m; 2, s)$ for any pair (m, s) it can be verified that (4. 28) holds for $s = s_j$ and $m = m_j$ for any j. It follows that (4. 28) must hold for any G-situation that is attainable if we start with an H-situation with $s = s_j$ in the infinite sequence (4.1) and a similar result holds for (4. 30). In fact equality in (4. 28) and (4. 30) appears to hold if and only if we start with an H-situation with s equal to some s_j in the infinite sequence (4.1), but this has not been shown.

On the other hand neither (4. 28) nor (4. 30) hold with equality in general. For example, they do not hold with equality for $G_0(3; 2, 9)$ and $H_0(2, 4)$, respectively.

5. Optimality of Procedures R_0 and R_1

In this section we investigate lower bounds for the expected number of tests for any group testing procedure for the HB problem. In particular, we show that for $n = d + s_j$ with $d = 2$ and s_j in the infinite sequence (4. 1), the result for $H_0(2, s)$ in (4. 24) is equal to the lower bound and hence the procedure R_0 must be optimal for $s = s_j$ for any j.

The main technique used is the HUFFMAN encoding scheme [4] which was originally devised to find an optimal code, i.e., a code that used the smallest expected number of letters in the encoding alphabet. In our application the encoding alphabet is binary (say, zeros and ones) since each test has 2 possible outcomes. Each sequence of test results (or zeros and ones) leading to a decision about the true state of nature is a code word. If the probabilities of the various possible states of nature are given then the HUFFMAN scheme gives us a plan for finding the true state of nature with a minimum expected number of tests. This plan may or may not correspond to an "allowable procedure", i.e., in our case to a group-testing procedure.

It is well known that the HUFFMAN scheme consists of (i) ordering the known probabilities (say $p_1, p_2, \ldots, p_t$) associated with the t states of nature (we shall refer to these as "old" or "original" numbers), (ii) adding the two smallest and replacing them by their sum (which is the first "new" number), (iii) reordering the set of $t-1$ numbers and (iv) repeating the first three steps until a single "new" number equal to 1 remains. It is also well known that the sum of the new numbers, say $H^*(p_1, p_2, \ldots, p_t)$ is the required HUFFMAN expected value for the optimal code or in our case the HUFFMAN lower bound to the expected number of group tests for any group-testing procedure. Since the HUFFMAN scheme may not yield a group-testing procedure this lower bound may not be attainable. It has been shown in [9] for the BB problem that the HUFFMAN scheme in general does not yield a group-testing procedure and that the HUFFMAN lower bound is in general not attainable.

It has also been pointed out, e.g., by C. PICARD in [6], that any scheme of adding the p_i's two at a time can be regarded as a questionnaire for finding the true state

of nature and that the sum of the probabilities associated with each question (or the sum of the "new" numbers) is its expected value. Hence we can regard the HUFFMAN lower bound $H^*(p_1, p_2, \ldots, p_t)$ as the minimum value over all such "combining procedures" for the given vector $\vec{p}=(p_1, p_2, \ldots, p_t)$. It is slightly more convenient to drop the condition that the arguments be probabilities and consider the set of schemes for combining any set of non-negative numbers $a_1, a_2, \ldots a_t$ two at a time. Then $H^*(a_1, a_2, \ldots, a_t)$ is the sum of the new numbers for the HUFFMAN scheme and gives the minimum sum over all such procedures since

$$\frac{1}{A} H^*(a_1, a_2, \ldots, a_t) = H^*\left(\frac{a_1}{A}, \frac{a_2}{A}, \ldots, \frac{a_t}{A}\right) \tag{5.1}$$

for any $A>0$ and, in particular, (5. 1) holds for $A = a_1+a_2+\ldots+a_t$.

In our problem we shall be concerned with two HUFFMAN lower bounds which we denote as HLB $(2, s)$ and $L_H(2, s)$ for $d=2$. The former is the usual lower bound over all group-testing procedures and the latter is the lower bound over those procedures which search out the defectives in a certain pattern, namely, by using units only from the defective set when it is not empty. In the latter case the $L_H(2, s)$ is found as the sum of 2 quantities, $L'_H+L''_H$, after assuming without any loss of generality that the units have been ordered. L'_H is the HLB for the initial subproblem of finding the position of the first defective and L''_H is the expected value of the conditional HLB for finding the second defective given the position of the first defective. The value of L''_H reduces to the usual HLB(d, s) with $d=1$. For our problem with $d=1$ the value of t above is the same as the number of units remaining n and the (a posteriori) probabilities p_i are all equal to $1/n$. In this case the HUFFMAN procedure is a group-testing procedure; namely it is the halving procedure (as well as R_0 and R_1). Hence $H^*\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = H_0(1, n-1) = H_1(1, n-1)$ (written as $H(1, n-1)$) as given in (2. 22) and, using (5. 1) and the definition of $h(n)$ in (2. 13), we can write this as

$$H^*(1, 1, \ldots, 1) = h(n) \tag{5.2}$$

where n is the number of components on the left side of (5. 2). For any d we have $f(d)=\binom{n}{d}$ states of nature with equal probabilities $1/f(d)$ for each and hence, writing f for $f(2)$,

$$\text{HLB}(2, s) = H^*\left(\frac{1}{f}, \frac{1}{f}, \ldots, \frac{1}{f}\right) = \frac{h(f)}{f}. \tag{5.3}$$

We now can prove a main result of this paper as

THEOREM 1. *For any integer $j \geqq 0$ the procedure R_0 is optimal for $d=2$ and $s=s_j$.*

PROOF. By property 2 in (4. 8) the b-value for $f_j=\binom{s_j+2}{2}$ is j and hence, using (2.12) and (5. 3), we obtain the required result

$$\text{HLB}(2, s_j) = j+2-\frac{2^{j+1}}{f_j} = H_0(2, s_j). \tag{5.4}$$

Since the HLB(2, s) is the "absolute" minimum it follows that for any integer s

$$\text{HLB}(2, s) \leqq L_{\text{H}}(2, s) = L'_{\text{H}}(2, s) + L''_{\text{H}}(2, s). \tag{5. 5}$$

We shall be interested to see whether equality holds in (5. 5) for any s and to study the relation between HLB(2, s) and $H_0(2, s)$. To derive $L_{\text{H}}(2, s)$ we note that the probability that the first defective is in the i-th position is $(s'+1-i)/f$ where $f = \binom{s'+1}{2}$ and hence

$$L_{\text{H}}(2, s) = H^*\left(\frac{1}{f}, \frac{2}{f}, \ldots, \frac{s'}{f}\right) + \sum_{i=1}^{s'} \left(\frac{s'+1-i}{f}\right) = H(1, s'-i) = \tag{5. 6}$$

$$= \frac{1}{f}\left\{H^*(1, 2, \ldots, s') + \sum_{\alpha=2}^{s'} h(\alpha)\right\}$$

since $h(1)=0$ and $s'=s+1$.

We now consider two lemmas for evaluating $H^*(1, 2, \ldots, s')$ for any integer s'.

LEMMA 1. *For* $s' \equiv 0$ *or* -1 (mod 3)

$$H^*(1, 2, \ldots, s') = \frac{3}{2}\left[\frac{2s'}{3}\right]\left(\left[\frac{2s'}{3}\right]+1\right) + H^*\left(3\left[\frac{s'}{3}\right]+3, 3\left[\frac{s'}{3}\right]+6, \ldots, 3\left[\frac{2s'}{3}\right]\right) \tag{5. 7}$$

and for $s' \equiv -2$ (mod 3)

$$H^*(1, 2, \ldots, s') = \frac{(s'-1)(2s'+1)}{3} + H^*(s', s'+2, s'+5, \ldots, 2s'-2); \tag{5. 8}$$

here (5. 8) *agrees with* (5. 7) *except for the "extra" component* s' *in the last term in* (5. 8).

PROOF. Using induction it is easy to show that the "new" numbers up to $2s'$ form the arithmetic progression 3, 6, 9, ..., $3\left[\frac{2s'}{3}\right]$. In fact, if this were true for $s'_0 = 3r_0$ then on the next step we would add $(s'_0+1)+(s'_0+2) = 2s'_0+3 = 3\left[\frac{2s'}{3}\right]+3$ and on the following step we would add an "old" s'_0+3 plus a "new" s'_0+3 giving $2s'_0+6$, which adds two "new" numbers to the same progression. Since this holds for $r_0=1$, yielding the new numbers 3 and 6, it holds for all $r_0 \geqq 1$. For s'_0+3r_0-1 and $3r_0+1$ it is easy to see that essentially the same proof holds but in the former case the last number s'_0 is "used up", i.e., combined with another, whereas in the latter case the last number s'_0 is not "used up" and has to be included in subsequent combinations.

In (5. 7) and (5. 8) the first term represents the sum of the arithmetic progression from 3 to $3\left[\frac{2s'}{3}\right]$ and the last term represents the subsequent combinations. From the discussion above the subsequent combinations start with $3\left[\frac{s'}{3}\right]+3$ for the cases in (5. 7) and with s' in the case of (5. 8), using these numbers up to $\left[\frac{2s'}{3}\right]$ as originals. This proves lemma 1.

LEMMA 2. *If $a_1, a_2, \ldots, a_u$ form an increasing arithmetic progression between any positive number and its double then*

$$H^*(a_1, a_2, \ldots, a_u) = (v+1)(a_1 + \ldots + a_{2w}) + v(a_{2w+1} + \ldots + a_u) \tag{5.9}$$

where v and w are defined by

$$u = 2^v + w; \qquad 0 < w \leqq 2^v. \tag{5.10}$$

PROOF. In this case the combining procedure forms a regular pattern and we can arrange the "new" sums in columns, starting a new column for each "new" sum containing a_1. If u is a power of 2 then every a_i appears exactly v times and (5.9) is clear. If it is not a power of 2 then we obtain $v+1$ columns and the number of sums in the c-th column is $[u/2^c]$ for $c = 1, 2, \ldots, v$ and one sum equal to $\sum_{i=1}^{t} a_i$ in the last column.

In the first column a_u is "eligible" for omission (it is omitted if and only if u is odd). If it is omitted then it combines with a_1 in the second column and appears in each subsequent column, and the "new" sum $a_{u-2} + a_{u-1}$ becomes eligible for omission in the second column. If a_u is not omitted in the first column then it combines with a_{u-1} and the "new" sum $a_{u-1} + a_u$ is eligible for omission in the second column. In general, if any "new" sum with 2^{c-1} consecutive terms is omitted in the c-th column then in the next column it combines with a_1 and appears in each subsequent column; either this "new" sum terminates with a_t or all subsequent terms have been omitted in previous columns (and are now combined with a_1).

It follows from the above that

(i) No a_i can be omitted in more than one column.

(ii) The set of a_i that are omitted exactly once must form a "tail interval" of the form $a_i, a_{i+1}, \ldots, a_u$.

To complete the proof of the lemma we have to show that the number of a's omitted is $u - 2w$ where w is given by (5.10). The number of a's included in the "new" sums or in $H^*(a_1, a_2, \ldots, a_u)$ is precisely $h(u) = H^*(1, 1, \ldots, 1)$ with u components, which equals $uv + 2w$. If we subtract this from $u(v+1)$ we get the desired result. This completes the proof of lemma 2.

COROLLARY 2. If $\Delta > 0$ is the common difference of successive a's of lemma 2 and $0 \leqq \varepsilon \leqq \Delta$ then

$$H^*(a_1 + \varepsilon, a_2, a_3, \ldots, a_t) = (v+1)\varepsilon + H^*(a_1, a_2, \ldots, a_t). \tag{5.11}$$

PROOF. Since $\varepsilon \leqq \Delta$ the pattern will not affected and hence ε has to appear exactly the same number of times as a_1, thus proving the corollary.

We are now ready to apply these results to the three cases in lemma 1. We replace t by s' and use the symbols b and c defined in (2.11) by writting $s' = 2^b + c$. Let u_i denote the value of u if $s' \equiv i \pmod 3$ for $i = 0, -1, -2$ and define v_i and w_i similarly. Then

$$u_i = \left[\frac{s'+2}{3}\right] = \frac{s'-1}{3} \quad \text{and} \quad v_i = \begin{cases} b-2 & \text{if} \quad s' \geqq 3 \cdot 2^{b-1} + i \\ b-1 & \text{if} \quad s' < 3 \cdot 2^{b-1} + i \end{cases} \tag{5.12}$$

and, of course $w_i + 2^{v_i} = (s'-i)/3$.

It is clear that the arithmetic progression on the right side of (5. 7) satisfies the conditions of lemma 2 since $t \leqq 3\left[\frac{t}{3}\right]+3$ and $3\left[\frac{2t}{3}\right] \leqq 2t$; also the progression on the right side of (5. 8) satisfies the condition in corollary 2 with $\varepsilon = 1$. Substituting these sequences for $a_1, a_2, \ldots, a_u$ in (5. 9), we now compute the last term in (5. 7) and (5. 8) or $H^*(a_1, a_2, \ldots, a_{u_i}) = H_i^*$ (say) for $i=0, -1, -2$ using (5. 9) (5.11) and (5.12). For $2^b < s' \leqq 3 \cdot 2^{b-1} + i$ we obtain for the last term in (5. 7) and (5. 8)

$$H_i^* = (b-1)\binom{s'+1}{2} + 3 \cdot 2^{2b-3} - 3(2s'+1)2^{b-2} + \frac{(s'-i)(5s'+3+i)}{6} \tag{5.13}$$

and for $3 \cdot 2^{b-1} + i < s' \leqq 2^{b-1}$ we similarly obtain

$$H_i^* = b\binom{s'+1}{2} + 3 \cdot 2^{2b-1} - 3(2s'+1)2^{b-1} + \frac{(s'-i)(5s'+3+i)}{6}. \tag{5.14}$$

If we now complete the evaluation of $H^*(1, 2, \ldots, s')$ in (5. 7) and (5. 8) by adding on the first term then we obtain for each i the same result which we now state as

THEOREM 2. *For any integer* $s' \geqq 2$

$$H^*(1, 2, \ldots, s') = \begin{cases} (b+2)f + 3 \cdot 2^{2b-3} - 3 \cdot 2^{b-2}(2s'+1) & \text{for} \quad 2^b \leqq s' < 3 \cdot 2^{b-1} \\ (b+3)f + 3 \cdot 2^{2b-1} - 3 \cdot 2^{b-1}(2s'+1) & \text{for} \quad 3 \cdot 2^{b-1} \leqq s' < 2^{b+1} \end{cases} \tag{5.15}$$

where $f = \binom{s'+1}{2}$ *and* $b = b(s')$ *is defined by* (2. 11).

REMARK. It should be noted that if we difined $b' = b'(s')$ by $3 \cdot 2^{b'-2} \leqq s' < 3 \cdot 2^{b'-1}$ then we can write (5.15), using a single expression for all $s' \geqq 2$, as

$$H^*(1, 2, \ldots, s) = (b'+2)f + 3 \cdot 2^{2b'-3} - 3 \cdot 2^{b'-2}(2s'+1) \tag{5.15a}$$

COROLLARY. Using the result (5.15) it can be verified that for any integer $t \geqq 2$

$$H^*(1, 2, \ldots, t) - H^*(1, 2, \ldots, t-1) = h(t) + x_G(t; 2) \tag{5.16}$$

and hence summing on t from 2 to s' gives

$$H^*(1, 2, \ldots, s') = \sum_{\beta=1}^{s} h(\beta+1) + \sum_{m=2}^{s'} x_G(m; 2). \tag{5.17}$$

The proof of (5.17) is omitted.

In the rest of this section we shall prove that $L_H(2, s)$ in (5. 6) with $H^*(1, 2, \ldots, s')$ given by (5.15) is equal to $H_0(2, s)$; simultaneously we show that another expression $L_G(m; 2, s)$, which we define below, is equal to $G_0(m; 2, s)$. A by-product of this will be another main result that procedures R_0 and R_1 are equivalent in the sense of having the same expected number of tests.

Consider a subclass of group-testing procedures which give preference in any G-situation to partitioning the defective set and testing nested subsets in the defective set until a defective unit is found. We call these "non mixing" or "nested procedures" and denote the set of such procedures by $\mathcal{N}$, since they never mix units from a

defective set and a remainder set. For example, the procedures R_1 and R_0 are both in $\mathcal{N}$.

The quantity $L_H(2, s)$ computed in (5. 6) is a lower bound for any procedure in $\mathcal{N}$ if we start with an H-situation. Let $L_G(m; 2, s)$ denote the corresponding lower bound if we start with a G-situation with parameters $(m; 2, s)$. We again write $L_G(m; 2, s)$ as the sum of two quantities $L'_G(m; 2, s)+L''_G(m; 2, s)$, where the former is the lower bound for the subproblem of finding the first defective in the defective set and the latter is the expected value of the conditional lower bound for finding the second defective given the position of the first defective. For a G-situation the probability that the first defective is in the i-th position $(i=1, 2, \ldots, m)$ is $(n-i)/F_m$ where $F_m = \binom{n}{2} - \binom{n-m}{2}$ and hence

$$L_G(m; 2, s) = H^*\left(\frac{n-m}{F_m}, \frac{n-m+1}{F_m}, \ldots, \frac{n-1}{F_m}\right) + \sum_{i=1}^{m}\left(\frac{n-i}{F_m}\right) H(1, n-i-1) = \tag{5.18}$$

$$= \frac{1}{F_m}\{H^*(n-m, n-m+1, \ldots, n-1) + \sum_{\alpha=s'-m+1}^{s'} h(\alpha)\}.$$

If the condition of lemma 2 is satisfied then we obtain

$$H^*(n-m, n-m+1, \ldots, n-1) = (v+1)\sum_{\alpha=1}^{2w}(s'-m+\alpha) + v\sum_{\alpha=2w+1}^{m}(s'-m+\alpha) = \tag{5.19}$$

$$= \frac{mv}{2}(2s'+1-m) + (m-2^v)(2s'+1-2^{v+1})$$

where v is defined by writing $m = 2^v + w$ and $0 < w \leqq 2^v$.

We now wish to show that the procedure R_0 satisfies both (5. 6) and (5. 18) and we do this by showing that $L_H(2, s)$ and $L_G(m; 2, s)$ satisfy the basic recursion formulas (2. 4) and (2. 5) with x_H and x_G given by (3.1) and (4.13), respectively, as well as the boundary conditions (2. 6), (2. 7) and (2. 8). With $b = b(s')$ defined as in (2.11), the right-side of (2. 4) gives

$$1 + \frac{\binom{n-x}{2}}{\binom{n}{2}} \frac{\left\{H^*(1, 2, \ldots, s'-x) + \sum_{\alpha=2}^{s'-x} h(x)\right\}}{\binom{n-x}{2}} + \tag{5.20}$$

$$+ \frac{F_x}{\binom{n}{2}} \frac{\left\{H^*(n-x, n-x+1, \ldots, s') + \sum_{\alpha=s'-x+1}^{s'} h(\alpha)\right\}}{F_x} =$$

$$= \frac{1}{\binom{n}{2}}\left\{H^*(1, 2, \ldots, s'-x) + H^*(s'+1-x, s'+2-x, \ldots, s') + \binom{n}{2} + \sum_{\alpha=2}^{s'} h(\alpha)\right\},$$

where we have used the fact that the condition of lemma 2 is satisfied for the second term above. Suppose first that $3\cdot 2^{b-1}\leqq s'<2^{b+1}$ so that $x_H=2^{b-1}$ and hence $2^b\leqq s'-x_H<3\cdot 2^{b-1}$. Then, using the top expression in (5.15) with s' replaced by $s'-x$ and (5.19) with m replaced by x, we obtain for the sum of the first three terms in the last expression in (5. 20)

$$\frac{1}{\binom{n}{2}}\left\{(b+2)\binom{s'+1-x}{2}+3\cdot 2^{2b-3}-3\cdot 2^{b-2}(2s'+1-2^b)+\right.$$

$$\left.+2^{b-2}(b-1)(2s'+1-2^{b-1})+\binom{n}{2}\right\}=$$

$$=\frac{1}{\binom{n}{2}}\left\{(b+3)\binom{s'+1}{2}+3\cdot 2^{2b-1}-3\cdot 2^{b-1}(2s'+1)\right\}=\frac{H^*(1,2,\ldots,s')}{\binom{n}{2}}$$

which agrees with the left side of (2. 4) using the 2-nd expression in (5.15). Similarly, if $2^b\leqq s'<3\cdot 2^{b-1}$ and $b\geqq 2$ then $2^{b-2}\leqq x_H=\left[\frac{s'+1}{2}\right]-2^{b-2}<2^{b-1}$ and $s'-x_H=\left[\frac{s'}{2}\right]+2^{b-2}$; hence $3\cdot 2^{b-2}\leqq s'-x_H<2^b$, so that the b-value for $s'-x_H$ is $b-1$ and the v-value for x_H is $b-2$. The algebra is quite similar to the above (we omit the details) except that we consider two cases according as s' is odd or even; in both cases we get agreement with the upper value in (5.15).

To check (2. 5) we note that the right side of (2. 5) gives

$$1+\left(\frac{F_m-F_x}{F_m}\right)\frac{\left\{H^*(s'-m+1,s'-m+2,\ldots,s'-x)+\sum_{\alpha=s'-m+1}^{s'-x}h(\alpha)\right\}}{F_m-F_x}+$$

(5.21)

$$+\frac{F_x}{F_m}\frac{\left\{H^*(s'-x+1,s'-x+2,\ldots,s')+\sum_{\alpha=s'-x+1}^{s'}h(\alpha)\right\}}{F_x}.$$

and clearly we have only to show that

(5. 22) $$H^*(s'-m+1,\ldots,s'-x)+H^*(s'-x+1,\ldots,s')+F_m=H^*(s'-m+1,\ldots,s')$$

for $x=x_G(m;\ 2,\ s)$ given by (4.13). It is easily verified that the condition of lemma 2 is satisfied in all three terms above.

Suppose first that $2^{p-1}\leqq m<3\cdot 2^{p-2}$ so that $x_G=2^{p-2}$ and $2^{p-2}\leqq m-x_G<2^{p-1}$ so that the v-value of $m-x_G$ is $p-2$. Then, applying (5.19) in (5. 22) gives

$$F_m+\frac{(p-2)}{2}(m-2^{p-2})(2s'+1-m-2^{p-1})+(m-2^{p-1})(2s'+1-2^p)+$$

$$+2^{p-3}(p-2)(2s'+1-2^{p-2})=\frac{m(p-1)}{2}(2s'+1-m)+(m-2^{p-1})(2s'+1-2^p),$$

which agrees with the right side of (5.19). In the second case $3 \cdot 2^{p-2} \leqq m < 2^p$ so that $x_G = m - 2^{p-1}$. Hence $2^{p-2} \leqq x_G < 2^{p-1}$. The algebra is again similar and is omitted.

To check the boundary conditions we find that (2. 6) is trivially satisfied, (2. 7) gives $h(s')/s'$ on both sides for $d=2$ by the definition of $h(x)$ in (2.13), and (2. 8) is concerned with the case $d=1$ where R_0 is known to be optimal. This completes the proof of

THEOREM 3. *Procedure R_0 is an optimal procedure in the subclass $\mathcal{N}$; $H_0(2, s)$ and $G_0(m; 2, s)$ are given by* (5. 6) *and* (5.18), *respectively, for all allowable values of m and s.*

By the definition of procedure R_1 in terms of recursion formulas and minimizing over the x-integers, it follows that R_1 must be the optimal procedure in the class $\mathcal{N}$. The only assumption that could prevent R_1 from being unconditionally optimal is that it is a non-mixing procedure. Hence, since procedure R_0 was shown to be optimal in $\mathcal{N}$, it follows that the same result must hold for R_1, i.e., we obtain our final major result

THEOREM 4. *Procedure R_1 is equivalent to procedure R_0 if we start with an* H-*situation in the sense that they have the same expected number of tests. Moreover if the x-values of R_1 are unique, then these two procedures are identical.*

REMARK. We could not conclude the identity of procedures R_1 and R_0 in theorem 4 because the x_G and x_H-values under R_1 have not been shown to be unique. It is conjectured that they will always be unique if we start with an H-situation (or with any situation attainable from an H-situation). However it should be pointed out that for some unattainable G-situations the x_H-values under R_1 may not be unique and may even be different from that given by R_0 in (4.13). For example, we find under R_1 that $x_G(7; 2, 7) = x_G(7; 2, 8) = 2$ and $x_G(7; 2, 9) = 2$ or 3 but under R_0 we have $x_G(7; 2,8) = x_G(7) = 3$. These situations are unattainable if we start with an H-situation since by (3. 2) we always have $m \leqq (s+2)/3$.

6. Description of Tables

Table I gives the values for $H_1(d, s)$ for $d=2$ and $s=1(1)25$. Fractional values as well as their decimal equivalents are included here. The column headed min, max gives the minimum and maximum number of tests required by R_1. It would be of interest to prove that these values never differ by more than d for $d=2$ and 3; empirical calculations show that the difference may be $d+1$ for $d=4$ and 5. For $d=2$ the fact that the minimum and maximum number of tests never differ by more than 2 in Table I also indicates that the variance of the number of tests required by R_1 will be small. The lower bound HLB for $d=2$ is also included in Table I.

Table II gives some similar information for $d=3$, 4, and 5. We note that equalities between $H_1(d, s)$ and HLB(d, s) do not occur in this table for $s=1(1)30$ as they did for $d=2$ in Table I.

It is curious that the $x_G(d, m)$-values at the bottom of Tables I and II do not show any change as a function of d; this has not been explained.

Table I

Binomial and Hypergeometric Group Testing—The HB-problem

The procedure# R_1, its min, max and expected number of tests and lower bounds to the expected number for any group-testing procedure.
(The number of defectives is 2 and the number of satisfactory units is s.)

H-SITUATION INFORMATION				LOWER BOUND INFORMATION	
s	Test sizes+ $x_H(2, s)$	Number of tests required		Huffman lower bound $HLB(2, s)$	$H_1(2, s) - HLB(2, s)$
		min, max	Expected value $H_1(2, s)$		
1*	1	1,2	5/3 = 1.667	5/3 = 1.667	0
2*	1	2,3	8/3 = 1.667	8/3 = 2.667	0
3*	1	3,4	17/5 = 3.400	17/5 = 3.400	0
4	2	3,5	4 = 4.000	59/15 = 3.933	1/15 = 0.067
5*	2	4,5	94/21 = 4.476	94/21 = 4.476	0
6	2	4,6	137/28 = 4.893	34/7 = 4.857	1/28 = 0.036
7	2	4,6	21/4 = 5.250	47/9 = 5.222	1/36 = 0.028
8*	3	5,6	251/45 = 5.578	251/45 = 5.578	0
9	3	5,7	323/55 = 5.873	321/55 = 5.836	2/55 = 0.037
10	4	5,7	135/22 = 6.136	200/33 = 6.061	5/66 = 0.075
11	4	5,7	497/78 = 6.372	248/39 = 6.359	1/78 = 0.013
12*	4	6,7	600/91 = 6.593	600/91 = 6.593	0
13	4	6,8	714/105 = 6.800	712/105 = 6.781	2/105 = 0.019
14	4	6,8	839/120 = 6.992	104/15 = 6.933	7/120 = 0.059
15	4	6,8	975/136 = 7.169	121/17 = 7.118	7/136 = 0.051
16	5	6,8	1123/153 = 7.340	1121/153 = 7.327	2/153 = 0.013
17*	5	7,8	1283/171 = 7.503	1283/171 = 7.503	0
18	6	7,9	1455/190 = 7.658	1454/190 = 7.653	1/190 = 0,005
19	6	7,9	1639/210 = 7.805	1634/210 = 7.781	1/42 = 0,024
20	7	7,9	1835/231 = 7.944	1823/231 = 7.892	4/77 = 0.052
21	7	7,9	2043/253 = 8.075	2021/253 = 7.988	2/23 = 0.087
22	8	7,9	2263/276 = 8.199	2248/276 = 8.145	5/92 = 0.054
23	8	7,9	2495/300 = 8.317	2488/300 = 8.293	7/300 = 0,023
24	8	7,9	2740/325 = 8.431	2738/325 = 8.425	2/325 = 0.006
25*	8	8,9	2998/351 = 8.541	2998/351 = 8.541	0

G-Situation test sizes*

m	1	2	3	4	5	6	7	8
$x_G(2; m)$	—	1	1	2	2	2	3	4

The results for procedure R_0 are the same as for procedure R_1.
+ The values shown are all unique and they satisfy (3.1) and (4.13), respectively.
* Starred s-values are those in the infinite sequence (4.1).

7. Acknowledgement

The author wishes to thank Mr. ELAINE FRANKOWSKI and Mr. WON JOON PARK for helping with the calculations in Table II.

Table II

Binomial and Hypergeometric Group Testing—The HB Problem

The procedure R_1, its expected value H_1 (d, s)
and a lower bound HLB over all group testing procedures

H-SITUATION INFORMATION

s	d = 3			d = 4			d = 5		
	x_H	$H_1(3,s)$	HLB	x_H	$H_1(4,s)$	HLB	x_H	$H_1(5,s)$	HLB
1	1	2.2500	2.0000	1	2.8000	2.4000	1	3.3333	2.6667
2	1	3.5000	3.4000	1	4.2667	3.9333	1	5.0000	4.4762
3	1	4.4500	4.4000	1	5.3714	5.1715	1	6.2321	5.8571
4	1	5.2571	5.1714	1	6.3143	6.1714	1	7.2778	6.9841
5	1	5.9643	5.8571	1	7.1587	6.9841	1	8.2183	7.9841
6	2	6.5238	6.4762	1	7.9048	7.7810	1	9.0758	8.8918
7	2	7.0333	6.9333	2	8.5697	8.4485	1	9.8649	9.7071
8	2	7.4788	7.4485	2	9.1273	8.9657	1	10.5812	10.4087
9	2	7.9000	7.8364	2	9.6503	9.5678	2	11.2103	10.9770
10	2	8.2832	8.2098	2	10.1219	9.9770	2	11,7779	11.6360
11	3	8.6346	8.5934	2	10.5692	10.4996	2	12.3022	12.1246
12	3	8.9538	8.8747	2	10.9824	10.8747	2	12.7920	12.6761
13	3	9.2554	9.1714	2	11.3777	11.2790	2	13.2553	13.0878
14	4	9.5397	9.4941	3	11.7454	11.6614	2	13.6950	13.5910
15	4	9.7929	9.7451	3	12.0797	11.9432	2	14.1107	13.9433
16	4	10.0330	9.9432	3	12.4014	12.3092	2	14.5079	14.3897
17	4	10.2658	10.2035	3	12.7071	12.6312	3	14.8843	14.7557
18	4	10.4910	10.4601	3	12.9955	12.8801	3	15.2336	15.0524
19	4	10.7006	10.6701	4	13.2683	13.1497	3	15.5699	15.4581
20	4	10.9029	10.8436	4	13.5250	13.4581	3	15.8918	15.7665
21	4	11.1003	10.9881	4	13.7739	13.7048	3	16.1970	16.0074
22	5	11.2861	11.2192	4	14.0130	13.9041	3	16.4929	16.3764
23	5	11.4650	11.4246	4	14.2435	14.1329	4	16.7790	16.6663
24	5	11.6356	11.5997	4	14.4644	14.3996	4	17.0452	16.8963
25	6	11.8013	11.7497	4	14.6796	14.6204	4	17.3023	17.1605
26	6	11.9570	11.8790	4	14.8880	14.8043	4	17.5535	17.4572
27	6	12.1106	11.9911	4	15.0905	14.9586	4	17.7988	17.6982
28	6	12.2585	12.1776	5	15.2840	15.1775	4	18.0326	17.8955
29	7	12.3988	12.3484	5	15.4697	15.3984	4	18.2598	18.1158
30	7	12.5348	12.4986	5	15.6513	15.5869	4	18.4822	18.3850

G-Situation Test Sizes $x_G = x_G$ (m; d, s)

d \ m	1	2	3	4	5	6	7
3	—	1	1	2	2	2	3
4	—	1	1	2	2		
5	—	1	1	2			

Appendix

From a set S of size $n=d+s$ a subset S' of size m is selected at random and a test on these shows that there is at least one defective present. Then a subset S'' of S', where S'' contains x units ($1 \leqq x < m$), is selected at random and a test on these shows that at least one them is defective. We now wish to prove the

LEMMA. *In the above situation the $n-m$ units of $S-S'$ and the $m-x$ units of $S'-S''$ can be mixed together without any loss of information. Moreover, if they are mixed together, we are back in the so called* G*-situation described by the parameters* $(x; d, s)$.

PROOF. We consider two different methods of arriving at the partition of S into three sets, $S-S'$, $S'-S''$, S'', and we wish to show that they lead to the same (conditional) distribution. Method one is as described above. In method two we sample x units at random from the original set S and a test indicates at least one defective among them; we then sample $m-x$ units from the remaining $n-x$ but we do not test this subset. Since there was no test in the last step of method two, it is clear that this separation of $n-x$ units into two subsets is "artifical" and contains no information. Hence we can disregard this step or mix the $m-x$ and $n-m$ units without the loss of any information. It is clear from method two that we now have two subsets, one of which contains at least one defective and this is a G-situation, by definition. Hence it remains only to show that method one and two lead to the same (conditional) distribution.

Let D_m and D_x denote the (random) number of defectives in the subset of size m and x, respectively. Using method one, the conditional probability that $D_m=d_m$ and $D_x=d_x$ given that $D_m \geqq 1$ and $D_x \geqq 1$ is

$$\text{(A1)} \qquad P_1(d_m, d_x) = \frac{\binom{d}{d_m}\binom{n-d}{m-d_m}}{\binom{n}{m}-\binom{n-d}{m}} \frac{\binom{d_m}{d_x}\binom{m-d_m}{x-d_x}}{\binom{m}{x} W}$$

where W is a constant (i.e., not depending on d_m or d_x) defined by setting the sum of the right side of (A1) over the set $1 \leqq d_x \leqq d_m \leqq d$ equal to unity. Summing over d_x first from the $\max(1, x+d_m-m)$ to $\min(x, d_m)$ gives

$$\text{(A2)} \qquad W\binom{m}{x}\left[\binom{n}{m}-\binom{n-d}{m}\right] = \sum_{d_m \geqq 1}\binom{d}{d_m}\binom{n-d}{m-d_m}\left[\binom{m}{x}-\binom{m-d_m}{x}\right] =$$

$$= \binom{m}{x}\left[\binom{n}{m}-\binom{n-d}{m}\right] - \frac{(n-d)!\,d!}{x!\,(m-x)!\,(n-m)!}\sum_{d_m \geqq 1}\binom{m-x}{d_m}\binom{n-m}{d-d_m} =$$

$$= \binom{m}{x}\left[\binom{n}{m}-\binom{n-d}{m}\right] - \binom{m}{x}\frac{d!\,(n-d)!}{m!\,(n-m)!}\left[\binom{n-x}{d}-\binom{n-m}{d}\right] =$$

$$= \binom{n}{m}\binom{m}{x}\left[1-\frac{(n-d)!\,(n-x)!}{n!\,(n-x-d)!}\right] = \binom{n-x}{m-x}\left[\binom{n}{x}-\binom{s}{x}\right].$$

By method two we sample x units directly from the original set S of size $n=s+d$ and obtain for the conditional distribution of D_m and D_x given that

$$\text{(A3)} \qquad P_2(d_m, dx) = \frac{\binom{d}{dx}\binom{n-d}{x-d_x}}{\binom{n}{x}-\binom{s}{x}} \cdot \frac{\binom{d-d_x}{d_m-d_x}\binom{s-x+d_x}{s-m+d_m}}{\binom{n-x}{m-x}}.$$

The eight factorials left after cancellation in the numerator of (A3) are exactly the same (and similarly situated) as the eight factorials left in the numerator of (Al) after cancellation. Moreover we have shown in (A2) that the demominators are equal and this completes the proof of the lemma.

REFERENCES

[1] BELLMAN, R.E. and GLUSS, B.: On various versions of the defective coin problem, *Information and Control* **4** (1961), 118—131.
[2] CAIRNS, S. S.: Balance scale sorting. *Amer. Math. Monthly* **70** (1963) 136—148.
[3] ERDŐS, P. and RÉNYI, A.: On two problems of information theory, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** (1963) 229—243.
[4] HUFFMAN, D. A.: A method for the construction of minimum redundancy codes, *Proceedings I. R. E.* **9** (1952) 1098—1101.
[5] LINDSTRÖM, B.: On a combinatorial problem in number theory, *Canad. Math. Bull.* **8** (1965) 477—490.
[6] PICARD, C.: *Théorie des Questionnaires*, Gauthers-Villars, Paris, 1965.
[7] SANDELIUS, M.: On an optimal search procedure. *Amer. Math. Monthly* **68** (1961) 138—154.
[8] SOBEL, M. and GROLL, P. A.: Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Tech. J.* **38** (1959) 1179—1252.
[9] SOBEL, M.: Group testing to classify all defectives in a binomial sample, A chapter in *Information and Decision Processes* 127—161. McGraw-Hill, New York, 1960.
[10] SOBEL, M.: Optimal group testing. *Stanford University Technical Report No. 72.* (1964) 1—57.
[11] SOBEL, M. and GROLL, P. A.: Binomial group-testing with an unknown proportion of defectives, *Technometrics*, **8** (1966) 631—656.
[12] ZIMMERMANN, S.: An optimal search procedure, *Amer. Math. Monthly* **66** (1959) 960—693.

University of Minnesota, Minneapolis, Minn. USA

(Received September 20, 1966)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 43—69.

MAJORATION NUMÉRIQUE DU GRADIENT DES FONCTIONS HARMONIQUES À L'AIDE DE LEURS DERIVÉES NORMALES

par

G. ADLER[1]

Introduction

Dans le travail [1] nous avons établi des majorations *numériques* pour la module du gradient des fonctions harmoniques, en termes des valeurs aux limites de ces fonctions. (Problème A.) La méthode s'est aussi montrée convenable, avec des modifications, pour la solution du problème analogue relatif à l'équation de la chaleur, c'est-à-dire pour la majoration de la module du gradient des fonctions caloriques, à l'aide de leurs valeurs initiales et aux limites. Ces résultats se trouvent dans le travail [2]. De plus, la même méthode, mutatis mutandis, a été efficace dans la majoration des tensions se produisant dans un corps élastique, majoration à travers les déplacements superficiels (voir [3]).

Nous avons déjà posé dans [1] le problème de comment majorer numériquement la module du gradient des fonctions harmoniques dans le cas où — contrairement au cas discuté dans [1] — ces sont les dérivées normales des fonctions qui sont données sur la frontière. (Problème B.)

Dans le cas à 2 dimensions la solution du problème B résulte immédiatement de la solution du problème A, à l'aide des propriétés bien connues des fonctions conjuguées.

Dans ce travail nous donnerons la solution du problème B. Dans l'intérêt de la simplicité nous nous limiterons au cas à 3 dimensions. (Naturellement, le cas à 2 dimensions serait encore plus simple, mais nous ne voulons pas effectuer les calculs concrèts pour obtenir des résultats que l'on peut également obtenir d'une manière différente.)

Nous remarquons que dans les travaux [1] et [2] nous avons imposé des conditions plus fortes sur la régularité des valeurs aux limites, qu'il n'aurait été nécessaire: nous avons exigé que les dérivées des valeurs aux limites satisfassent à une condition de Hölder. Cette condition peut être remplacée par une condition intégrale moins restrictive du type de DINI. Dans ce travail nous nous servirons de cette condition ultérieure.

La majoration du gradient d'une fonction harmonique se base naturellement sur la majoration du gradient aux points de la frontière du domaine. Des majorations de ce type ont été démontrées, pour un cas tout à fait général, par AGMON, DOUGLIS et NIRENBERG dans leur travail [6]. Leur méthode consiste dans les suivants: premièrement ils donnent une majoration pour les dérivées de la solution du problème aux limites relatif à un demi-espace à partir d'une formule explicite fournissant

[1] L'auteur a achevé cet ouvrage durant qu'il jouissait d'une bourse de l'Université de Rome.

la solution des équations à coefficients constants, ensuite ils ramènent le cas d'une frontière générale au cas du demi-espace en transformant un morceau convenablement petit de la frontière en un domaine plan.

Notre méthode est fondée sur la discussion de la solution dans une sphère osculatrice à la frontière; ainsi nous n'avons pas besoin de la transformation mentionnée. Étant donné que nous nous sommes proposés pour but d'établir une majoration *numérique*, nous nous sommes limités au cas le plus simple, celui de l'équation de Laplace. Une telle restriction de la généralité nous a permis de déterminer les valeurs numériques des constants intervenant dans la majoration. De cette façon on peut considérer notre travail comme le premier pas vers une formule utilisable dans la pratique du calcul numérique, quoique nos coefficients soient encore bien loin des valeurs optimales.

Nous pensons que notre idée pour majorer les dérivées tangentielles à la frontière en termes de la dérivée normale ait un certain intérêt en elle-même.

Le schéma du travail est comme suit. Le § 1 contient les définitions. Dans le § 2 nous avons énoncé les théorèmes auxiliaires, dont les démonstrations se trouvent dans le § 4. Ces démonstrations exigent des calculs assez longs, pour cette raison nous avons trouvé convenient de les remettre à la fin du travail. Le § 3 est consacré à la démonstration du théorème principal.

§ 1. Définitions

1. On considère l'équation de Laplace à 3 dimensions:

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=0, \tag{1}$$

dans le domaine (ensemble ouvert et connexe) borné Ω de l'espace euclidien (x, y, z); on indiquera par Σ la frontière de Ω et par ν la normale extérieure de Σ.

Supposons qu'il existe le plan tangent T_P pour chaque point P de Σ. Considérons un système de coordonnées cartésiennes ξ^P, η^P, ζ^P [2] ayant son originie dans P, dont les axes ξ^P, η^P sont situés dans le plan T_P et l'axe positif ζ_P a la direction de la normale extérieure de Σ. Soit $\Sigma_r(P)$ le voisinage du point P sur Σ, formé par la composante [3] de l'ensemble des points de Σ intérieurs au cylindre $\{(\xi^P)^2+(\eta^P)^2=r^2,\ -\infty<\zeta^P<+\infty\}$, composante qui contient P.

Nous dirons que le domaine Ω appartient à la classe $\mathfrak{A}(r, L)$ $(r>0,\ L\geqq 0)$

1° s'il existe, pour chaque point $P\in\Sigma$, deux sphères $\Gamma_i(P)$ et $\Gamma_e(P)$, de rayon r, dont les frontières contiennent P, et qui sont telles que

(i) l'intérieur de $\Gamma_i(P)$ est situé dans Ω,

(ii) $\Gamma_e(P)$ a dans son intérieur des points extérieurs à Ω, mais ne possède dans son intérieur aucun point de $\Sigma_r(P)$;

2° s'il existe, pour chaque paire de points $A\in\Sigma$ et $B\in\Sigma$ une courbe réctifiable l, intérieure à Ω, avec des extrémités C, D, telle que

(i) sa longueur ne surpasse pas L,

(ii) les sphères de rayon r centrées sur l se trouvent dans $\Omega+\Sigma$, et

[2] L'indice supérieur P sera supprimé si cela ne donne pas lieu à équivoque.

[3] Sous composante on entend un sous-ensemble maximum connexe de l'ensemble donné.

(iii) les sphères de rayon r avec les centres C resp. D contiennent les points A resp. B sur leurs frontières;

3° si pour chaque $P_0 \in \Sigma$ et $P \in \Sigma_r(P_0)$, en indiquant par ν_0 resp. ν les normales extérieures de Σ aux points P_0 resp. P et par P' la projection de P sur le support de ν_0, on a

$$|\sin(\nu_0, \nu)| \leqq \frac{\overline{PP'}}{r}.$$

On dira qu'un domaine satisfaisant seulement aux conditions 1° et 3° appertient à la classe $\mathfrak{A}(r)$.

2. Il suit de la condition 3° qu'une droite parallèle à ν_0 n'a, au plus, qu'un seul point commun avec $\Sigma_r(P_0)$. Dans les définitions ci-dessous on supposera toujours que le domaine Ω est de classe $\Sigma(r)$.

La portion de surface $\Sigma_r(P)$ peut être représentée dans le système de coordonnées locales sous la forme $\zeta^P = f_P(\xi^P, \eta^P)$, où la fonction f_P, selon ce qui précède, est univalente.

Une fonction $F(R)$ $(R \in \Sigma)$ définie sur la surface Σ, où

$$R = (\xi^P, \eta^P, \zeta^P = f_P(\xi^P, \eta^P))$$

est le point courant de Σ, sera représentée dans le voisinage $\Sigma_r(P)$ sous la forme

$$F(R) = F_P(\xi^P, \eta^P).^4$$

Nous dirons que la fonction F, définie sur Σ, est de classe $\mathcal{K}(r)$ $(r>0)$ si l'intégrale

$$\int_0^{2\pi} \int_0^{r} \frac{|F_P(t\cos\alpha, t\sin\alpha) - F_P(0,0)|}{t}\, dt\, d\alpha$$

est convergente pour chaque $P \in \Sigma$.

Nous dirons que la fonction F, définie sur Σ, appartient à la classe $\mathcal{H}(\alpha, K)$ $(0<\alpha\leqq 1,\ K>0)$, si

$$|F_P(\xi^P, \eta^P) - F_P(0,0)| \leqq K[(\xi^P)^2 + (\eta^P)^2]^{\frac{\alpha}{2}} \qquad ((\xi^P)^2 + (\eta^P)^2 \leqq r^2),$$

où α et K sont des constantes indépendantes de P.

3. Nous introduisons pour la fonction F, définie sur Σ, les notations suivantes:

$$M_0 \equiv M_0(F) \equiv \max_{\Sigma} |F|,$$

$$M_1(s) \equiv M_1(F, s) \equiv \max_{P \in \Sigma} \int_0^{2\pi} \int_0^{s} \frac{|F_P(t\cos\alpha, t\sin\alpha) - F_P(0,0)|}{t}\, dt\, d\alpha.$$

[4] Si cela ne donne lieu à aucune équivoque (quand il s'agit d'un voisinage $\Sigma_r(P)$ fixé), l'indice P sera omis: $F(R) = F(\xi, \eta)$.

Il est évident que $\mathscr{H}(\alpha, K) \subset \mathscr{K}(r)$, et si $F \in \mathscr{H}(\alpha, K)$, alors

$$M_1(F, s) \leqq 2\pi \frac{K}{\alpha} s^{\alpha}.$$

Nous entendrons les intégrales au sens de Riemann.

4. Finalement, pour abréger la locution, nous introduisons l'abréviation suivante:

$$\mathfrak{R}\{A_1, A_2, \ldots, A_n\} \equiv \text{(il résulte, en vertu de } A_1, A_2, \ldots, A_n, \text{ que)},$$

où le symbole A_i peut substituer le numéro d'une formule, un théorème ou une condition. Par exemple:

$$\mathfrak{R}\{(13), \text{Lemme 7, } 1^\circ \text{ (ii) de déf. de } \mathfrak{A}(r), (19)\}$$

équivaut à dire:

„il résulte, en vertu de la formule (13), du Lemme 7, de la condition 1° (ii) de la définition de la classe $\mathfrak{A}(r)$ et de la formule (19), que ..."

§ 2. Lemmes

1. LEMME 1. — *Soient*

1° Ω *un domaine (à 3 dimensions) de classe* $\mathfrak{A}(r, L)$;

2° u *une fonction, harmonique dans* Ω *et admettant des dérivées premières continues dans* $\Omega + \Sigma$.

Alors

$$\max_{\Omega+\Sigma} u - \min_{\Omega+\Sigma} u \leqq 4r\left(1 + e^{3\frac{L}{r}}\right) \max_{\Sigma} \left|\frac{\partial u}{\partial \nu}\right|.$$

La démonstration de ce théorème se trouve dans le travail [4], pour le cas où le nombre des dimensions est quelconque.

LEMME 2a. — *Soient*

1° Ω *un domaine (à 3 dimensions) de classe* $\mathfrak{A}(r)$;

2° u *une fonction, harmonique dans* Ω,

(i) *admettant des dérivées premières continues dans* $\Omega + \Sigma$, *pour laquelle*

(ii) *la dérivée normale* $\dfrac{\partial u}{\partial \nu}$ *est de classe* $\mathscr{K}(r)$ *et*

(iii)

$$\max_{\Omega+\Sigma} |\operatorname{grad} u| = M.$$

Alors, en indiquant par P *un point arbitraire de* Σ, *par* ν *la normale extérieure de* Σ *au point* P *et par* P' *un point arbitraire de* Ω $(\overline{PP'} \leqq \frac{1}{2}r)$ *sur le support de* ν,

on a $\left(avec\ q \equiv \overline{PP'},\ M_1(s) \equiv M_1\left(\frac{\partial u}{\partial v},\ s\right)\right)$:

$$\left|\frac{\partial u(P')}{\partial v} - \frac{\partial u(P)}{\partial v}\right| \leqq$$

$$\leqq \frac{q}{r}\left[A_1 + B_1 \ln\frac{r}{q}\right] M + C_1 \frac{q}{r} M_1(r) + D_1 q \int_q^r \frac{M_1(\tau)}{\tau^2}\, d\tau + E_1 M_1(\sqrt{2}\, q),$$

où A_1, B_1, C_1, D_1 *et* E_1 *sont des constantes universelles satisfaisant aux inégalités*

$$A_1 < 45;\quad B_1 < 83;\quad C_1 < 1{,}81;\quad D_1 < 1{,}28;\quad E_1 < 1{,}81.$$

La démonstration de ce théorème se trouve dans la note [5].

LEMME 2b. — *En supposant les conditions du Lemme* 2a *satisfaites, on a:*

$$\left|\frac{\partial u(P')}{\partial v} - \frac{\partial u(P)}{\partial v}\right| \leqq A_2 \sqrt{\frac{q}{r}}\, M + B_2 M_1(r),$$

où A_2 *et* B_2 *sont des constantes universelles satisfaisant aux inégalités*

$$A_2 < 128;\quad B_2 < 4.$$

LEMME 3. *Soient* Ω *un domaine et* u *une fonction satisfaisant aux conditions du Lemme* 2a. *Alors, avec les notations de la figure* 1 *(où* $\overline{OP^{**}} = \overline{OP^{*}}$*), on a:*

$$\left|\frac{\partial u(P^{**})}{\partial \varrho} - \frac{\partial u(P^{*})}{\partial \varrho}\right| < r\left[A_3 r^2 \left(\frac{\alpha}{r-\varrho}\right)^3 + B_3 r \left(\frac{\alpha}{r-\varrho}\right)^2 + C_3 \frac{\alpha}{r-\varrho}\right] M,$$

où A_3, B_3 *et* C_3 *sont des constantes universelles satisfaisant aux inégalités*

$$A_3 < 68;\quad B_3 < 88;\quad C_3 < 102.$$

2. Dans le reste de ce paragraphe soit Γ la sphère de l'espace (x, y, z) définie par l'inégalité $x^2 + y^2 + z^2 < r^2$; de plus $P(x, y, z)$ et $Q(\xi, \eta, \zeta)$ indiqueront les points de la sphère fermée $\Gamma + \sigma$ et les points de sa surface σ, respectivement.

Nous introduisons les coordonnées sphériques $\varrho, \varphi, \vartheta$ et $\bar{\varrho}, \bar{\varphi}, \bar{\vartheta}$ par les transformations

$$x = \varrho \sin\vartheta \cos\varphi \qquad \xi = \bar{\varrho} \sin\bar{\vartheta} \cos\bar{\varphi}$$

$$y = \varrho \sin\vartheta \sin\varphi \qquad \eta = \bar{\varrho} \sin\bar{\vartheta} \sin\bar{\varphi}$$

$$z = \varrho \cos\vartheta \qquad \zeta = \bar{\varrho} \cos\bar{\vartheta}.$$

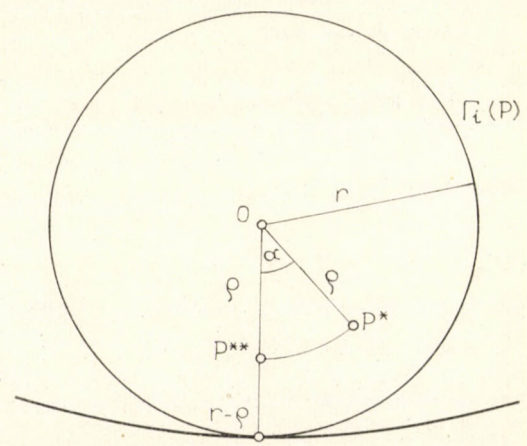


Fig. 1

Alors, comme il est connu, la solution (supposée assez régulière) de l'équation (1) dans Γ peut s'écrir, à l'aide de sa dérivée normale sur σ, sous la forme

$$u(P) = \int_{\sigma} N(P, Q)\, U(Q)\, d\sigma_Q, \tag{2}$$

où

$$U(Q) = \frac{\partial u(Q)}{\partial \nu}, \quad Q = Q(\bar{\varrho} = r, \bar{\varphi}, \bar{\vartheta}) = Q(\bar{\varphi}, \bar{\vartheta}),$$

$d\sigma_Q$ est l'élément de surface de σ au point Q, et

$$\begin{cases} N(P, Q) = \dfrac{1}{4\pi}\left\{\dfrac{2}{\varrho_1} - \displaystyle\int_0^r \frac{1}{l}\left(\frac{1}{r} - \frac{r}{\varrho}\frac{1}{\varrho_2}\right) dl\right\}, \\ q_1 = (r^2 + \varrho^2 - 2r\varrho\cos\psi)^{\frac{1}{2}}, \\ q_2 = \left(l^2 + \dfrac{r^4}{\varrho^2} - 2l\dfrac{r^2}{\varrho}\cos\psi\right)^{\frac{1}{2}}, \\ \cos\psi = \sin\vartheta\cos\varphi\sin\bar{\vartheta}\cos\bar{\varphi} + \sin\vartheta\sin\varphi\sin\bar{\vartheta}\sin\bar{\varphi} + \cos\vartheta\cos\bar{\vartheta}. \end{cases} \tag{3}$$

(C'est-à-dire $\psi = (POQ) \sphericalangle$.)

Soit $v(\varrho, \varphi, \vartheta)$ une fonction fournie par la formule (2), quand on y substitue au lieu des valeurs aux limites $U(Q)$ la fonction intégrable arbitraire $V(Q) = V(\bar{\varphi}, \bar{\vartheta})$.

Pour cette fonction v les majorations suivantes (Lemmes 4—8) sont valables:

LEMME 4. — *Soit*

$$|V(\varphi, \vartheta)| \leqq \begin{cases} 1, & si \quad 0 \leqq \vartheta \leqq \omega, \\ 0, & si \quad \omega < \vartheta \leqq \pi, \end{cases} \quad 0 \leqq \varphi \leqq 2\pi.$$

Alors

$$|v(r, \varphi, \vartheta)| \leqq A_4 r\omega \qquad (0 \leqq \vartheta \leqq \pi,\ 0 \leqq \varphi \leqq 2\pi),$$

où A_4 est une constante universelle satisfaisant à l'inégalité $A_4 < 15$.

LEMME 5. — *Soit*

$$\left.\begin{aligned} \left|\int_0^{\vartheta}\int_0^{\varphi} V(\bar{\varphi}, \bar{\vartheta})\sin\bar{\vartheta}\, d\bar{\varphi}\, d\bar{\vartheta}\right| &\leqq \vartheta^3, \quad si \quad 0 \leqq \vartheta \leqq \omega, \\ V(\varphi, \vartheta) &= 0, \quad si \quad \omega < \vartheta \leqq \pi, \end{aligned}\right\} 0 \leqq \varphi \leqq 2\pi \quad (\omega \leqq \pi/4).$$

Alors

$$\left|\frac{\partial v(r, \varphi, 0)}{\partial \vartheta}\right| \leqq A_5 r\omega,$$

où A_5 est une constante satisfaisant à l'inégalité $A_5 < 22$.

LEMME 6. — *Soient*

$$\left.\begin{aligned} \int_0^{2\pi}\int_0^{\omega} \frac{|V(\varphi, \vartheta)|}{\vartheta}\, d\vartheta\, d\varphi &\leqq 1, \\ V(\varphi, \vartheta) &= 0, \quad si \quad \omega < \vartheta \leqq \pi, \end{aligned}\right\} 0 \leqq \varphi \leqq 2\pi \quad \left(0 < \omega \leqq \frac{\pi}{2}\right)$$

et

Alors

$$\left|\frac{\partial v(r,\varphi,0)}{\partial\vartheta}\right| \leqq A_6 r,$$

où A_6 *est une constante universelle satisfaisant á l'inégalité* $A_6 < 1$.

LEMME 7. — *Soit u une fonction, harmonique dans la sphère* $0 \leqq \varrho < r$, *continue dans* $0 \leqq \varrho \leqq r$, *admettant la dérivée* $\partial u/\partial\varrho$ *continue dans* $\{0 < \varrho \leqq r, 0 \leqq \vartheta < \omega\}$ $(\omega \leqq \pi/4)$ *et satisfaisant aux conditions suivantes:*

$$\left.\begin{aligned} \frac{\partial u(r,\varphi,\vartheta)}{\partial\varrho} &= 0, \quad si \quad 0 \leqq \vartheta < \omega, \\ |u(r,\varphi,\vartheta)| &\leqq 1, \quad si \quad \omega \leqq \vartheta \leqq \pi, \end{aligned}\right\} 0 \leqq \varphi \leqq 2\pi.$$

Alors

$$\left|\frac{\partial u(r,\varphi,0)}{\partial\vartheta}\right| \leqq \frac{A_7}{\omega},$$

où A_7 *est une constante universelle satisfaisant á l'inégalité* $A_7 < 267$.

§ 3. Majoration

THÉORÈME. — *Soient*

1° Ω *un domaine (à 3 dimensions) de classe* $\mathfrak{A}(r, L)$;

2° *u une fonction, harmonique dans* Ω,

(i) *admettant des dérivées premières continues dans* $\Omega + \Sigma$ *et*

(ii) *la dérivée normale* $\partial u/\partial v$ *appartenant à la classe* $\mathscr{K}(r)$.

Alors

$$(*) \qquad \max_{\Omega+\Sigma} |\operatorname{grad} u| \leqq \frac{1+\left(A+Be^{3\frac{L}{r}}\right)\left(\frac{1}{1-\mu}\right)}{\mu} M_0\left(\frac{\partial u}{\partial v}\right) + \frac{C}{\mu} M_1\left(\frac{\partial u}{\partial v}, r\right),$$

où A, B et C sont des constantes universelles satisfaisant aux inégalités

$$A < 16 \cdot 10^{11}; \quad B < 23 \cdot 10^{10}; \quad C < 23 \cdot 10^3,$$

et μ *est un paramètre vérifiant l'inégalité* $0 < \mu < 1$.

REMARQUE 1. — Étant donné que $\mathfrak{A}(r; L) \subseteqq \mathfrak{A}(r', L+r-r')$ $(0 < r' \leqq r)$, on peut remplacer au second membre de l'inégalité $(*)$ r et L par r' et $L+r-r'$, respectivement. De cette manière on peut „contrebalancer" dans la majoration l'influence de la régularité de $\frac{\partial u}{\partial v}$ $\left(\text{qui intervient dans } M_1\left(\frac{\partial u}{\partial v}, r\right)\right)$ par $\max\left|\frac{\partial u}{\partial v}\right|$ et viceversa.

REMARQUE 2. — Pour simplifier les calculs, au cours de la démonstration, nous utiliserons le Lemme 2b, qui est une forme „faible" du Lemme 2a. En s'appuyant sur le Lemme 2a, le paramètre μ figurerait au second membre de l'inégalité $(*)$ sous une forme plus compliquée; on obtiendrait alors une possibilité ultérieure pour „contrebalancer" l'influence de M_0 par M_1 et viceversa.

REMARQUE 3. — — En substituant M dans l'inégalité du Lemme 2a par le second membre de l'inégalité (∗), on obtient la majoration numérique suivante concernant l'allure de la dérivée normale d'une fonction harmonique:

THÉORÈME. — *Soient*

1° Ω *un domaine (à 3 dimensions) de classe* $\mathfrak{A}(r, L)$;

2° u *une fonction, harmonique dans* Ω,

(i) *admettant des dérivées premières continues dans* $\Omega+\Sigma$ *et*

(ii) *la dérivée normale* $\partial u/\partial v$ *appartenant à la classe* $\mathcal{K}(r)$.

Alors on a, avec les notations du Lemme 2a,

$$\left|\frac{\partial u(P')}{\partial v}-\frac{\partial u(P)}{\partial v}\right| \leqq \frac{q}{r}\left(A_1+B_1\ln\frac{r}{q}\right)\frac{1+\left(A+Be^{3\frac{L}{r}}\right)\frac{1}{1-\mu}}{\mu}M_0+$$

$$+\frac{q}{r}\left[\left(A_1+B_1\ln\frac{r}{q}\right)\frac{C}{\mu}+C_1\right]M_1(r)+D_1 q\int_q^r\frac{M_1(\tau)}{\tau^2}d\tau+E_1 M_1(\sqrt{2}\,q).$$

DÉMONSTRATION.

1. Étant donné que les dérivées de la fonction u satisfont elles-mêmes à l'équation (1), le principe du maximum a aussi lieu pour la fonction $|\operatorname{grad} u|$, c'-est-à-dire $|\operatorname{grad} u|$ assume le maximum de ses valeurs prises dans le domaine fermé $\Omega+\Sigma$ sur la frontière Σ. Par cette raison il suffit de majorer $|\operatorname{grad} u|$ sur Σ. Posons

$$M=\max_{\Omega+\Sigma}|\operatorname{grad} u|. \tag{4}$$

2. Considérons un système de coordonnés ξ, η, ζ appartenant à un point arbitraire $P_0\in\Sigma$. (Nous rappelons que l'axe positif ζ a la direction de la normale extérieure de Σ à P_0.) Soient $\Gamma_i(P_0)$ et $\Gamma_i(P)$ $(P\in\Sigma_r(P_0))$ les sphères appartenant aux points P_0 et P respectivement, dont l'existence a été postulée dans la condition 1° (i) de la définition de la classe $\mathfrak{A}(r, L)$, et soit σ la frontière de $\Gamma_i(P_0)$. Soient $\varrho, \varphi, \vartheta$ respectivement $\varrho^*, \varphi^*, \vartheta^*$ des coordonnées sphériques dans $\Gamma_i(P_0)$ respectivement dans $\Gamma_i(P)$, définies par les transformations

$$\xi=\xi^{P_0}=\varrho\sin\vartheta\cos\varphi \qquad \xi^P=\varrho^*\sin\vartheta^*\cos\varphi^*$$

$$\eta=\eta^{P_0}=\varrho\sin\vartheta\sin\varphi \qquad \eta^P=\varrho^*\sin\vartheta^*\sin\varphi^*$$

$$r+\zeta=r+\zeta^{P_0}=\varrho\cos\vartheta \qquad r+\zeta^P=\varrho^*\cos\vartheta^*.$$

Les fonctions définies dans $\Gamma_i(P_0)$ ou bien sur $\Sigma_r(P_0)$ peuvent être exprimées soit par les coordonnées cartésiennes ξ, η, ζ, soit par les coordonnées sphériques; dans ce qui suit nous employeront tous les deux systèmes, quelquefois simultanément aussi dans la même formule, si cela simplifie l'écriture.

Soit $P^*=(\xi, \eta, \zeta^*)=(r, \varphi, \vartheta)$ $\left(\vartheta\leqq\frac{\pi}{2}\right)$ le point de σ, pour lequel $\overline{PP^*}\,||\,v_0$, $P=(\xi, \eta, \zeta)$ étant le point général de Σ (voir la figure 2a). Désignons par P^{**} le point du segment $\overline{OP}$, satisfaisant à la condition $\overline{OP^*}=\overline{OP^{**}}$ (voir la figure 2b).

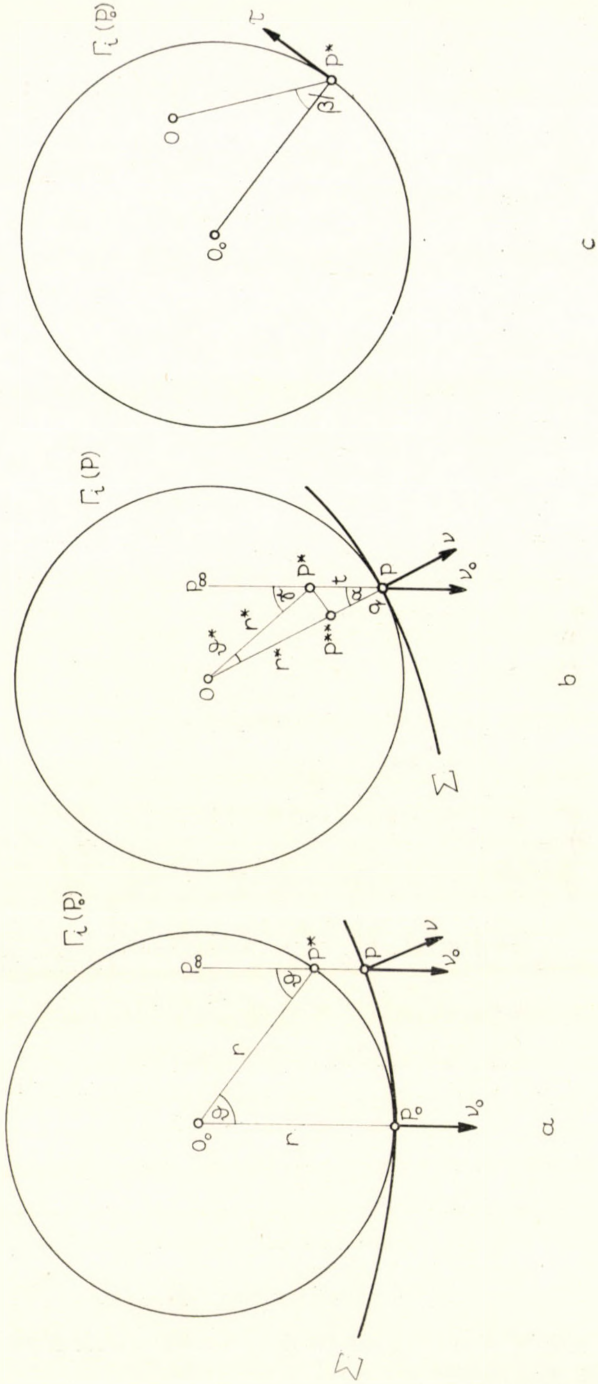


Fig. 2

Désignons par $v_j(\varrho, \varphi, \vartheta)$, $w_j(\varrho, \varphi, \vartheta)$, $z_j(\varrho, \varphi, \vartheta)$ les fonctions fournies par la formule (2), appliquée à la sphère $\Gamma_i(P_0)$ quand on y substitue au lieu des valeurs aux limites $U(Q)$ les valeurs suivantes V_j, W_j, Z_j, respectivement:

$$V_1(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P^*)}{\partial \varrho} - \dfrac{\partial u(P)}{\partial \nu}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$V_2(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P)}{\partial \nu} - \dfrac{\partial u(P_0)}{\partial \nu}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$V_3(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P_0)}{\partial \nu}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$W_1(\varphi, \vartheta) = \left\{\begin{array}{ll} 0, & 0 \leqq \vartheta \leqq \omega, \\ \dfrac{\partial u(r, \varphi, \vartheta)}{\partial \varrho} - \lambda, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$W_2(\varphi, \vartheta) = \left\{\begin{array}{ll} 0, & 0 \leqq \vartheta \leqq \omega, \\ \lambda, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi,$$

où

$$\lambda = \Big(\int\limits_{\omega \leqq \vartheta \leqq \pi} d\sigma \Big)^{-1} \int\limits_{\omega \leqq \vartheta \leqq \pi} \frac{\partial u(r, \varphi, \vartheta)}{\partial \varrho}\, d\sigma;$$

$$Z_1(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P^{**})}{\partial \varrho^*} - \dfrac{\partial u(P)}{\partial \nu}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$Z_2(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P^*)}{\partial \varrho^*} - \dfrac{\partial u(P^{**})}{\partial \varrho^*}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi;$$

$$Z_3(\varphi, \vartheta) = \left\{\begin{array}{ll} \dfrac{\partial u(P^*)}{\partial \varrho} - \dfrac{\partial u(P^*)}{\partial \varrho^*}, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{array}\right\} 0 \leqq \varphi \leqq 2\pi.$$

La fonction originelle u et les fonctions définies ci-dessus satisfont évidemment dans $\Gamma_i(P_0)$ aux relations suivantes:

$$\begin{aligned} u(\varrho, \varphi, \vartheta) &\equiv v_1 + v_2 + v_3 + w_1 + w_2, \\ V_1(\varphi, \vartheta) &\equiv Z_1 + Z_2 + Z_3. \end{aligned} \tag{5}$$

Nous allons déduire quelques majorations relatives à ces fonctions, au moyen des considérations géométriques.

3. Soient $\Sigma(\varphi, \vartheta)$, $\sigma(\varphi, \vartheta)$, $\Psi(\varphi, \vartheta)$ et $\Phi(\varphi, \vartheta)$ les portions de surface (P_0AB), (P_0CD), $(ABCD)$ et (P_0BC) respectivement, indiquées sur la figure 3 et $\Pi(\varphi, \vartheta)$ le domaine limité par ces surfaces et par $\Phi(0, \vartheta)$. Soit n la normale extérieure de la frontière de $\Pi(\varphi, \vartheta)$. ((Par les définitions déjà adoptées on a: $\frac{\partial}{\partial n} \equiv \frac{\partial}{\partial \nu}$ sur $\Sigma(\varphi, \vartheta)$ et $\frac{\partial}{\partial n} \equiv -\frac{\partial}{\partial \varrho}$ sur $\sigma(\varphi, \vartheta)$).

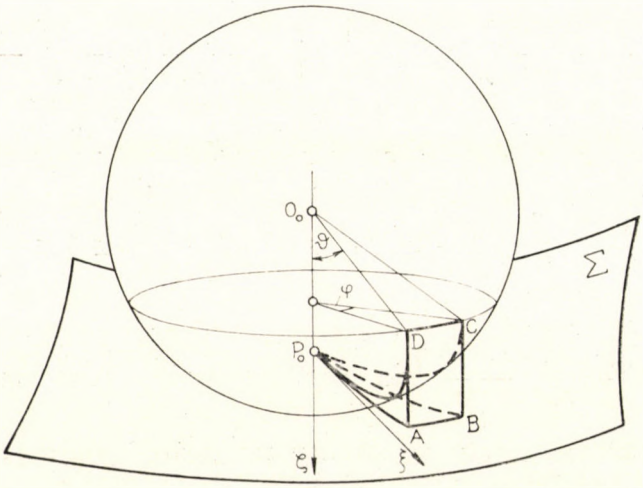


Fig. 3

Puisque la fonction u est harmonique dans $\Pi(\varphi, \vartheta)$ et régulière sur la frontière de $\Pi(\varphi, \vartheta)$, on a (en indiquant par $d\Sigma$, $d\sigma$, $d\Psi$, $d\Phi$ les éléments de surface):

$$\int_{\Sigma(\varphi,\,\vartheta)} \frac{\partial u}{\partial \nu}\, d\Sigma - \int_{\sigma(\varphi,\,\vartheta)} \frac{\partial u}{\partial \varrho}\, d\sigma + \int_{\Psi(\varphi,\,\vartheta)} \frac{\partial u}{\partial n}\, d\Psi + \int_{\Phi(0,\,\vartheta)} \frac{\partial u}{\partial n}\, d\Phi + \int_{\Phi(\varphi,\,\vartheta)} \frac{\partial u}{\partial n}\, d\Phi = 0. \tag{6}$$

Notons par $|\Psi(\varphi, \vartheta)|$ et $|\Phi(\varphi, \vartheta)|$ l'aire de $\Psi(\varphi, \vartheta)$ et de $\Phi(\varphi, \vartheta)$, respectivement. Puisque $\Omega \in \mathfrak{N}(r)$, on a les inégalités suivantes:

$$\left.\begin{aligned} |\Phi(\varphi, \vartheta)| &\leqq \int_0^{\vartheta} 2r(1-\cos\vartheta) r\cos\vartheta\, d\vartheta \leqq \frac{1}{3} r^2 \vartheta^3 \\ |\Psi(\varphi, \vartheta)| &\leqq \varphi \cdot r\sin\vartheta \cdot 2r(1-\cos\vartheta) \leqq 2\pi r^2 \vartheta^3 \end{aligned}\right\} \left(0 \leqq \vartheta \leqq \frac{\pi}{2}\right). \tag{7}$$

En nous rappelant la définition de M (voir (4)), il résulte des relations (6) et (7):

$$\left| \int_{\Sigma(\varphi,\,\vartheta)} \frac{\partial u}{\partial \nu}\, d\Sigma - \int_{\sigma(\varphi,\,\vartheta)} \frac{\partial u}{\partial \varrho}\, d\sigma \right| \leqq M\left(\frac{2}{3} + 2\pi\right) r^2 \vartheta^3 \leqq 7r^2 M\vartheta^3 \qquad \left(0 \leqq \vartheta \leqq \frac{\pi}{2}\right). \tag{8}$$

Soit $d\Sigma_P$ la projection sur Σ de l'élément $d\sigma_{P^*}$ au point P^*, la projection parallèle

à la normale v_0 au point P_0. Il découle de la condition 3° de la définition de $\mathfrak{A}(r, L)$ que

$$d\sigma_{P^*} \geqq d\Sigma_P \geqq d\sigma_{P^*} \cos\vartheta,$$

ainsi

$$0 \leqq d\sigma_{P^*} - d\Sigma_P \leqq d\sigma_{P^*}(1-\cos\vartheta) = \frac{1}{2} d\sigma_{P^*} \vartheta^2. \tag{9}$$

Il résulte des (8) et (9), avec des notations évidentes:

$$\left| r^2 \int_0^{\vartheta} \int_0^{\varphi} \left(\frac{\partial u(P)}{\partial v} - \frac{\partial u(P^*)}{\partial \varrho} \right) \sin\vartheta \, d\varphi \, d\vartheta \right| \equiv \left| \int_{\sigma(\varphi,\vartheta)} \frac{\partial u(P)}{\partial v} d\sigma_{P^*} - \int_{\sigma(\varphi,\vartheta)} \frac{\partial u(P^*)}{\partial \varrho} d\sigma_{P^*} \right| \leqq$$

$$\leqq \left| \int_{\sigma(\varphi,\vartheta)} \frac{\partial u(P)}{\partial v} (d\sigma_{P^*} - d\Sigma_P) \right| + \left| \int_{\Sigma(\varphi,\vartheta)} \frac{\partial u(P)}{\partial v} d\Sigma_P - \int_{\sigma(\varphi,\vartheta)} \frac{\partial u(P^*)}{\partial \varrho} d\sigma_{P^*} \right| \leqq$$

$$\leqq \frac{1}{2} M \int_{\sigma(\varphi,\vartheta)} \vartheta^2 \, d\sigma + 7r^2 M \vartheta^3 \leqq 9r^2 M \vartheta^3 \qquad \left(0 \leqq \vartheta \leqq \frac{\pi}{2}\right).$$

On a donc

$$\left| \int_0^{\vartheta} \int_0^{\varphi} \left(\frac{\partial u(P)}{\partial v} - \frac{\partial u(P^*)}{\partial \varrho} \right) \sin\vartheta \, d\varphi \, d\vartheta \right| \leqq 9M\vartheta^3 \qquad \left(0 \leqq \vartheta \leqq \frac{\pi}{2}\right). \tag{10}$$

Les majorations suivantes sont strictement géométriques et se réfèrent aux figures 2a, 2b et 2c. Dans ces majorations on admettra toujours que $\vartheta \leqq \frac{\pi}{4}$.

4. Il est évidente que

$$q \leqq t \equiv \overline{PP^*} \leqq (\mathfrak{R}\{1° \text{ de la déf. de } \mathfrak{A}(r)\}) \leqq 2r(1-\cos\vartheta) \leqq r\vartheta^2. \tag{11}$$

$\mathfrak{R}\{3° \text{ de la déf. de } \mathfrak{A}(r)\}$:

$$\alpha \leqq \vartheta. \tag{12}$$

On a, à partir du triangle $(PP^*O)_{\triangle}$ (voir la figure 2b):

$$\cos\gamma = \frac{-t + r\cos\alpha}{r^*} \geqq (\mathfrak{R}\{(11), (12)\}) \geqq$$

$$\geqq \frac{r\cos\vartheta - 2r(1-\cos\vartheta)}{r^*} = \frac{r}{r^*}(3\cos\vartheta - 2) > 0 \qquad \left(0 \leqq \vartheta \leqq \frac{\pi}{4}\right).$$

Il s'ensuit que $\gamma < \frac{\pi}{2}$, donc

$$\vartheta^* < \frac{\pi}{2}. \tag{13}$$

On obtient de (13) et de (12), compte tenu de l'inégalité

$$t \leqq 2r(1-\cos\vartheta) \leqq 2r\left(1 - \frac{1}{\sqrt{2}}\right) < \frac{3}{5} r, \tag{14}$$

que

$$\frac{2}{\pi}\vartheta^* \leqq \sin\vartheta^* = \frac{t}{r^*}\sin\alpha \leqq \frac{t\sin\alpha}{r-t} \leqq \frac{t\vartheta}{r-\frac{3}{5}r} = \frac{5}{2}\frac{t\vartheta}{r}. \tag{15}$$

Il résulte pour le triangle $(PP^*O)_\triangle$:

$$(r^*)^2 = r^2 + t^2 - 2rt\cos\alpha,$$

d'où

$$r - r^* = \frac{2rt\cos\alpha - t^2}{r+r^*} \geqq \frac{t}{2r}(2r\cos\alpha - t) \geqq \frac{t}{2r}\left(2r\frac{1}{\sqrt{2}} - \frac{3}{5}r\right) \geqq \frac{2}{5}t. \tag{16}$$

Я{(15), (16)}:

$$\frac{\vartheta^*}{r-r^*} \leqq \frac{\pi}{2}\frac{5}{2}\frac{t\vartheta}{r}\frac{1}{\frac{2}{5}t} \leqq 10\frac{\vartheta}{r}. \tag{17}$$

On a, à partir du triangle $(PP^*O)_\triangle$:

$$\sin\vartheta^* \leqq \frac{\sin\vartheta^*}{\sin(\alpha+\vartheta^*)} = \frac{t}{r} \leqq (\mathbf{Я}\{(11)\}) \leqq \vartheta^2; \tag{18}$$

Я{(13), (18)}:

$$\vartheta^* \leqq \frac{\pi}{2}\vartheta^2. \tag{19}$$

On aura donc:

$$(OP^*P_\infty)\sphericalangle = \alpha + \vartheta^* \leqq (\mathbf{Я}\{(12), (18)\}) \leqq \vartheta + \frac{\pi}{2}\vartheta^2 \leqq \frac{9}{4}\vartheta.$$

Il en résulte:

$$\beta = (OP^*O_0)\sphericalangle \leqq (OP^*P_\infty)\sphericalangle + (O_0P^*P_\infty)\sphericalangle \leqq \frac{9}{4}\vartheta + \vartheta = \frac{13}{4}\vartheta. \tag{20}$$

Nous passons maintenant aux majorations relatives aux fonctions auxiliaires v_j et z_j à l'aide des majorations obtenues dans les sections **3** et **4**. Dans ce qui suit nous supposerons que $\omega \leqq \frac{\pi}{4}$.

5. Compte tenu de l'inégalité $q \leqq t$ (voir la figure 2b), **Я**{(11), Lemme 2b}:

$$|Z_1(\varphi,\vartheta)| \leqq \left.\begin{cases} A_2\vartheta M + B_2 M_1(r), & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{cases}\right\} 0 \leqq \varphi \leqq 2\pi. \tag{21}$$

Я{(17), Lemme 3}:

$$\begin{aligned} |Z_2(\varphi,\vartheta)| &\leqq \left\{\begin{matrix}[A_3(10\vartheta)^3 + B_3(10\vartheta)^2 + C_3 10\vartheta]M \\ 0\end{matrix}\right\} \leqq \\ &\leqq \left.\begin{cases} D_3\vartheta M, & 0 \leqq \vartheta \leqq \omega, \\ 0, & \omega < \vartheta \leqq \pi, \end{cases}\right\} 0 \leqq \varphi \leqq 2\pi, \end{aligned} \tag{22}$$

où $D_3 < 5\cdot 10^4$.

On obtient par des considérations élémentaires et en vertu de (20) (voir la figure 2c):

$$\left|\frac{\partial u(P^*)}{\partial \varrho^*}-\frac{\partial u(P^*)}{\partial \varrho}\right| = \left|\left(\frac{\partial u(P^*)}{\partial \varrho}\cos\beta+\frac{\partial u(P^*)}{\partial \tau}\sin\beta\right)-\frac{\partial u(P^*)}{\partial \varrho}\right| \leqq$$

$$\leqq \left|\frac{\partial u(P^*)}{\partial \varrho}\right|(1-\cos\beta)+\left|\frac{\partial u(P^*)}{\partial \tau}\right|\sin\beta \leqq$$

$$\leqq M\left(\frac{\beta^2}{2}+\beta\right) = M\frac{13}{4}\vartheta\left(\frac{13}{8}\vartheta+1\right) \leqq 9M\vartheta \qquad \left(0\leqq\vartheta\leqq\frac{\pi}{4}\right).$$

Ainsi, on a pour Z_3:

$$|Z_3(\varphi,\vartheta)| \leqq \begin{Bmatrix} 9M\vartheta, & 0\leqq\vartheta\leqq\omega, \\ 0, & \omega<\vartheta\leqq\pi, \end{Bmatrix} 0\leqq\varphi\leqq 2\pi. \tag{23}$$

Я{(5), (21), (22), (23)}:

$$|V_1(\varphi,\vartheta)| \leqq \begin{Bmatrix} a_1\vartheta M+B_2M_1(r), & 0\leqq\vartheta\leqq\omega, \\ 0, & \omega<\vartheta\leqq\pi, \end{Bmatrix} 0\leqq\varphi\leqq 2\pi, \tag{24}$$

où

$$a_1 = A_2+D_3+9<51\cdot 10^3.$$

Я{(24), Lemme 4}:

$$|v_1(r,\varphi,\vartheta)| \leqq r\omega[a_2\omega M+a_3M_1(r)] \qquad (0\leqq\vartheta\leqq\pi,\ 0\leqq\varphi\leqq 2\pi), \tag{25}$$

où

$$a_2 = A_4a_1<765\cdot 10^3;$$

$$a_3 = A_4B_2<60.$$

Я{déf. de $V_2(\varphi,\vartheta)$ et $V_3(\varphi,\vartheta)$, Lemme 4}:

$$|v_2(r,\varphi,\vartheta)+v_3(r,\varphi,\vartheta)| \leqq A_4r\omega M_0 \qquad (0\leqq\vartheta\leqq\pi,\ 0\leqq\varphi\leqq 2\pi). \tag{26}$$

Я{déf. de λ}:

$$|\lambda| = \Big(\int\limits_{\omega\leqq\vartheta\leqq\pi} d\sigma\Big)^{-1}\left|\int\limits_{\omega\leqq\vartheta\leqq\pi}\frac{\partial u}{\partial \varrho}\,d\sigma\right| = \Big(\int\limits_{\omega\leqq\vartheta\leqq\pi} d\sigma\Big)^{-1}\left|\int\limits_{0\leqq\vartheta\leqq\omega}\frac{\partial u}{\partial \varrho}\,d\sigma\right| \leqq$$

$$\leqq M\frac{2r\pi\cdot r(1-\cos\omega)}{2r\pi\cdot r\left(1+\cos\frac{\pi}{4}\right)} \leqq \frac{3}{10}M\omega^2. \tag{27}$$

Я{(27), Lemme 4}:

$$|w_2(r,\varphi,\vartheta)| \leqq A_4r\pi\cdot\frac{3}{10}M\omega^2 \leqq a_4rM\omega^2 \qquad (0\leqq\vartheta\leqq\pi,\ 0\leqq\varphi\leqq 2\pi), \tag{28}$$

où $a_4<15$.

On peut admettre que

$$\max_{\Omega+\Sigma} u = -\min_{\Omega+\Sigma} u, \tag{29}$$

sans restreindre la généralité.

Я{(5), (25), (26), (28), (29), Lemme 1}:

$$|w_1(r,\varphi,\vartheta)| = |u(r,\varphi,\vartheta) - v_1(r,\varphi,\vartheta) - v_2(r,\varphi,\vartheta) - v_3(r,\varphi,\vartheta) - w_2(r,\varphi,\vartheta)| \leqq$$

$$\leqq 2r\left(1+e^{3\frac{L}{r}}\right) M_0 + r\omega[a_2\omega M + a_3 M_1(r)] + A_4 r\omega M_0 + a_4 r\omega^2 M = \tag{30}$$

$$= a_5 r\omega^2 M + r\left(a_6 + 2e^{3\frac{L}{r}}\right) M_0 + a_3 r\omega M_1(r) \qquad (0 \leqq \vartheta \leqq \pi,\ 0 \leqq \varphi \leqq 2\pi),$$

où

$$a_5 = a_2 + a_4 < 766 \cdot 10^3,$$

$$a_6 \leqq 2 + A_4 \frac{\pi}{4} < 14.$$

Я{déf. de $V_1(\varphi, \vartheta)$, (10), Lemme 5}:

$$\left|\frac{\partial v_1(r,\varphi,0)}{\partial\vartheta}\right| \leqq a_7 r\omega M, \tag{31}$$

où

$$a_7 = 9A_5 < 198.$$

Я{déf. de $V_2(\varphi,\vartheta)$, $\frac{\partial u}{\partial v} \in \mathcal{K}(r)$, Lemme 6}:

$$\left|\frac{\partial v_2(r,\varphi,0)}{\partial\vartheta}\right| \leqq A_6 r \int_0^{2\pi}\int_0^{\omega} \frac{|V_2(\varphi,\vartheta)|}{\vartheta}\, d\vartheta\, d\varphi \leqq \tag{32}$$

(en utilisant la substitution $s = r\sin\vartheta$ et les inégalités $r\sin\omega < r$, $\frac{d\vartheta}{\vartheta} \leqq \frac{ds}{s}\sqrt{2}$)

$$\leqq \sqrt{2} A_6 r \int_0^{2\pi}\int_0^{r} \frac{|V_2(\xi = s\cos\varphi, \eta = s\sin\varphi)|}{s}\, ds\, d\varphi \leqq \sqrt{2}\, A_6 r M_1(r).$$

Étant donné que $\frac{\partial u(P_0)}{\partial v}$ et λ sont constantes, on obtient pour les fonctions v_3 et w_2 (voir les définitions de $V_3(\varphi,\vartheta)$ et de $W_2(\varphi,\vartheta)$):

$$\frac{\partial v_3(r,\varphi,0)}{\partial\vartheta} = \frac{\partial w_2(r,\varphi,0)}{\partial\vartheta} = 0. \tag{33}$$

En considérant que la fonction $W_1(\varphi,\vartheta)$ satisfait à la condition de compatibilité relative à la dérivée normale d'une fonction harmonique, c'est que

$$\int_\sigma W_1(\varphi,\vartheta)\, d\sigma = 0,$$

$w_1(\varrho,\varphi,\vartheta)$ est une fonction, harmonique dans Γ, avec les valeurs aux limites

$$w_1(r,\varphi,\vartheta) = W_1(\varphi,\vartheta).$$

Donc, **Я**{(30), Lemme 7}:

$$(34)\qquad \left|\frac{\partial w_1(r,\varphi,0)}{\partial\vartheta}\right| \leqq \frac{A_7}{\omega}\left[a_5 r\omega^2 M + r\left(a_6 + 2e^{3\frac{L}{r}}\right)M_0 + a_3 r\omega M_1(r)\right] =$$

$$= \left[a_8\omega M + \left(a_9 + a_{10}e^{3\frac{L}{r}}\right)\frac{1}{\omega}M_0 + a_{11}M_1(r)\right]r,$$

où

$$a_8 = A_7 a_5 < 205\cdot 10^6, \quad a_9 = A_7 a_6 < 3740,$$

$$a_{10} = 2A_7 < 534, \qquad a_{11} = A_7 a_3 < 16\,020.$$

6. Choisissons le point P_0 de telle façon que

$$|\operatorname{grad} u(P_0)| = M.$$

On a

$$M \leqq \left|\frac{\partial u}{\partial\zeta^{P_0}}\right|_{\xi^{P_0}=\eta^{P_0}=\zeta^{P_0}=0} + \max_{0\leqq\alpha\leqq 2\pi}\left|\frac{\partial u}{\partial\xi^{P_0}}\cos\alpha + \frac{\partial u}{\partial\eta^{P_0}}\sin\alpha\right|_{\xi^{P_0}=\eta^{P_0}=\zeta^{P_0}=0} \leqq$$

$$\leqq M_0 + \sqrt{2}\,\frac{1}{r}\max_{0\leqq\varphi\leqq 2\pi}\left|\frac{\partial u(r,\varphi,0)}{\partial\vartheta}\right| \leqq$$

$$\leqq M_0 + \sqrt{2}\,\frac{1}{r}\max_{0\leqq\varphi\leqq 2\pi}\left|\frac{\partial v_1(r,\varphi,0)}{\partial\vartheta} + \frac{\partial v_2(r,\varphi,0)}{\partial\vartheta} + \frac{\partial w_1(r,\varphi,0)}{\partial\vartheta}\right| \leqq$$

(**Я**{(31), (32), (33), (34)})

$$\leqq M_0 + \sqrt{2}\,\frac{1}{r}\left\{a_7 r\omega M + \sqrt{2}\,A_6 r M_1(r) + \right.$$

$$\left. + r\left[a_8\omega M + \left(a_9 + a_{10}e^{3\frac{L}{r}}\right)\frac{1}{\omega}M_0 + a_{11}M_1(r)\right]\right\} =$$

$$= a_{12}\omega M + \left[1 + \left(a_{13} + a_{14}e^{3\frac{L}{r}}\right)\frac{1}{\omega}\right]M_0 + a_{15}M_1(r),$$

où

$$a_{12} = \sqrt{2}\,(a_7 + a_8) < 3\cdot 10^8, \quad a_{13} = \sqrt{2}\,a_9 < 53\cdot 10^2,$$

$$a_{14} = \sqrt{2}\,a_{10} < 760, \qquad a_{15} = \sqrt{2}\,(\sqrt{2}\,A_6 + a_{11}) < 23\cdot 10^3.$$

Il en résulte, avec $\mu = 1 - a_{12}\omega$ $\left(\omega < \frac{1}{a_{12}},\ 0<\mu<1\right)$:

$$M \leqq \frac{1 + \left(a_{13} + a_{14}e^{3\frac{L}{r}}\right)\frac{a_{12}}{1-\mu}}{\mu}M_0 + \frac{a_{15}}{\mu}M_1(r).$$

§ 4. Démonstration des lemmes

Comme nous l'avons déjà remarqué, les démonstrations des Lemmes 1 et 2a se trouvent dans les travaux [4] et [5], respectivement, ainsi nous ne les détaillerons pas.

Dans ce paragraphe nous utiliserons les notations du § 2, intervenant dans les formules (2) et (3).

DÉMONSTRATION DU LEMME 2b. Le Lemme 2b découle immédiatement du Lemme 2a, en vertu des inégalités suivantes:

$$\left(a+b\ln\frac{1}{x}\right)x \leqq (a+b)\sqrt{x} \qquad (a>0,\ b>0,\ 0\leqq x\leqq 1),$$

$$\sqrt{2}\,q \leqq \sqrt{2}\,\frac{r}{2} < r,$$

$$\frac{q}{r} \leqq \frac{1}{2},$$

et ($M_1(\tau)$ étant une fonction non-décroissante)

$$M_1(\sqrt{2}\,q) \leqq M_1(r),$$

$$q\int_q^r \frac{M_1(\tau)}{\tau^2}\,d\tau \leqq qM_1(r)\int_q^r \frac{d\tau}{\tau^2} < qM_1(r)\int_q^\infty \frac{d\tau}{\tau^2} = M_1(r).$$

DÉMONSTRATION DU LEMME 3. On peut supposer, sans restreindre la généralité, que dans le système de coordonnées sphériques ϱ, φ, ϑ on a

$$P^{**}=(\varrho,\ \varphi=0,\ \vartheta=0),\quad P^{*}=(\varrho,\ \varphi=0,\ \vartheta=\alpha).$$

On obtient de la formule (2):

$$\frac{\partial u(P^*)}{\partial\varrho} = \frac{r^2(r^2-\varrho^2)}{4\pi\varrho}\int_0^\pi\int_0^{2\pi}(r^2+\varrho^2-2r\varrho\cos\psi)^{-3/2}\,\frac{\partial u(\varrho=r,\bar\varphi,\bar\vartheta)}{\partial\varrho}\sin\bar\vartheta\,d\bar\varphi\,d\bar\vartheta, \tag{35}$$

où

$$\cos\psi = \sin\alpha\sin\bar\vartheta\cos\bar\varphi+\cos\alpha\cos\bar\vartheta.$$

On a ainsi pour la fonction $(\ldots)^{-3/2}$ qui figure derrière le signe d'intégrale:

$$\left|\frac{\partial}{\partial\alpha}(r^2+\varrho^2-2r\varrho\cos\psi)^{-3/2}\right| = \frac{3}{2}(r^2+\varrho^2-2r\varrho\cos\psi)^{-5/2}\cdot 2r\varrho\,|\cos\alpha\sin\bar\vartheta\cos\bar\varphi- \tag{36}$$

$$-\sin\alpha\cos\bar\vartheta| \leqq 3r\varrho[r^2+\varrho^2-2r\varrho(\sin\alpha\sin\bar\vartheta+\cos\alpha\cos\bar\vartheta)]^{-\frac{5}{2}}\cdot$$

$$\cdot(\cos\alpha\sin\bar\vartheta+\sin\alpha\cos\bar\vartheta) = 3r\varrho[r^2+\varrho^2-2r\varrho\cos(\alpha-\bar\vartheta)]^{-\frac{5}{2}}\sin(\alpha+\bar\vartheta) \leqq$$

$$\leqq 3r\varrho\left[(r-\varrho)^2+\frac{4}{\pi^2}r\varrho(\alpha-\bar\vartheta)^2\right]^{-5/2}(\alpha+\bar\vartheta).$$

En vertu du théorème de la moyenne de Lagrange, on a, pour une valeur convenable ϑ^* $(0<\vartheta^*<\alpha)$:

$$\left|\frac{\partial u(P^{**})}{\partial \varrho}-\frac{\partial u(P^*)}{\partial \varrho}\right| = \left|\frac{\partial^2 u(\varrho,\varphi=0,\vartheta=\vartheta^*)}{\partial\vartheta\,\partial\varrho}\right|\alpha \leqq$$

(par les formules (35) et (36))

$$\leqq \alpha\frac{r^2(r^2-\varrho^2)}{4\pi\varrho}\int_0^{2\pi}\int_0^{\pi} 3r\varrho\,\frac{\vartheta^*+\bar\vartheta}{\left[(r-\varrho)^2+\frac{4}{\pi^2}r\varrho(\vartheta^*-\bar\vartheta)^2\right]^{5/2}}\cdot$$

$$\cdot\left|\frac{\partial u(r,\bar\varphi,\bar\vartheta)}{\partial\varrho}\right|\sin\bar\vartheta\,d\bar\vartheta\,d\bar\varphi \leqq \alpha\frac{3r^3(r^2-\varrho^2)\varrho}{4\pi\varrho}\,2\pi\int_0^{\pi}\frac{\vartheta^*+\bar\vartheta}{[\cdots]^{5/2}}\,\bar\vartheta\,d\bar\vartheta\,M \leqq$$

(37)
$$\leqq 3r^4(r-\varrho)(\alpha^2 I_1+\alpha I_2)M,$$

où

$$I_1=\int_0^{\pi}\frac{\bar\vartheta\,d\bar\vartheta}{[\cdots]^{5/2}},\qquad I_2=\int_0^{\pi}\frac{\bar\vartheta^2\,d\bar\vartheta}{[\cdots]^{5/2}}.$$

Nous obtenons, avec la substitution

(38)
$$\bar\vartheta-\vartheta^*=\frac{\pi}{2}(r-\varrho)(r\varrho)^{-1/2}\tau:$$

(39)
$$I_1=\int_{-\frac{2}{\pi}\vartheta^*(r-\varrho)^{-1}(r\varrho)^{1/2}}^{\frac{2}{\pi}(\pi-\vartheta^*)(r-\varrho)^{-1}(r\varrho)^{1/2}}\frac{\left[\vartheta^*+\frac{\pi}{2}(r-\varrho)(r\varrho)^{-1/2}\tau\right]\frac{\pi}{2}(r-\varrho)(r\varrho)^{-1/2}}{(r-\varrho)^5(1+\tau^2)^{5/2}}\,d\tau =$$

$$=\vartheta^* I_1^{(1)}+I_1^{(2)},$$

où

$$I_1^{(1)}=\frac{\pi}{2}\int_{\cdots}^{\cdots}\frac{(r\varrho)^{-1/2}}{(r-\varrho)^4(1+\tau^2)^{5/2}}\,d\tau,$$

$$I_1^{(2)}=\frac{\pi^2}{4}\int_{\cdots}^{\cdots}\frac{(r\varrho)^{-1}\tau}{(r-\varrho)^3(1+\tau^2)^{5/2}}\,d\tau.$$

On aura pour $I_1^{(1)}$ et $I_1^{(2)}$, avec des majorations élémentaires:

(40)
$$I_1^{(1)}<\frac{4}{\sqrt{r\varrho}(r-\varrho)^4},\qquad I_1^{(2)}<\frac{10}{3}\frac{1}{r\varrho(r-\varrho)^3}.$$

On voit que, avec la même substitution (38):

$$I_2 = \frac{\pi}{2} \int_{-\frac{2}{\pi}\vartheta^*(r-\varrho)^{-1}\sqrt{r\varrho}}^{\frac{2}{\pi}(\pi-\vartheta^*)(r-\varrho)^{-1}\sqrt{r\varrho}} \frac{\left[\vartheta^* + \frac{\pi}{2}(r-\varrho)(r\varrho)^{-1/2}\tau\right]^2}{\sqrt{r\varrho}(r-\varrho)^4(1+\tau^2)^{5/2}}\, d\tau \leqq$$

$$\leqq \frac{\pi}{2}\frac{1}{\sqrt{r\varrho}}\left[\alpha^2 \frac{1}{(r-\varrho)^4}\int_{-\infty}^{+\infty}\frac{d\tau}{(1+\tau^2)^{5/2}} + 2\alpha\frac{\pi}{2}\frac{1}{\sqrt{r\varrho}(r-\varrho)^3}\int_0^{\infty}\frac{\tau\, d\tau}{(1+\tau^2)^{5/2}} + \right. \tag{41}$$

$$\left. + \frac{\pi}{4}\frac{1}{r\varrho(r-\varrho)^2}\int_{-\infty}^{+\infty}\frac{\tau^2\, d\tau}{(1+\tau^2)^{5/2}}\right] \leqq$$

$$\leqq 4\alpha^2\frac{1}{\sqrt{r\varrho}(r-\varrho)^4} + 7\alpha\frac{1}{r\varrho(r-\varrho)^3} + 12\frac{1}{(r\varrho)^{3/2}(r-\varrho)^2}.$$

Finalement il résulte des (37)—(41):

$$\left|\frac{\partial u(P^{**})}{\partial\varrho} - \frac{\partial u(P^*)}{\partial\varrho}\right| \leqq$$

$$\leqq 3\frac{r^4}{(r\varrho)^{3/2}}\left[8r\varrho\left(\frac{\alpha}{r-\varrho}\right)^3 + \frac{31}{3}\sqrt{r\varrho}\left(\frac{\alpha}{r-\varrho}\right)^2 + 12\frac{\alpha}{r-\varrho}\right]M \leqq$$

$$\leqq 3\frac{r^4}{\left(r\frac{r}{2}\right)^{3/2}}\left[8r^2\left(\frac{\alpha}{r-\varrho}\right)^3 + \frac{31}{3}r\left(\frac{\alpha}{r-\varrho}\right)^2 + 12\frac{\alpha}{r-\varrho}\right]M \leqq$$

$$\leqq r\left[68r^2\left(\frac{\alpha}{r-\varrho}\right)^3 + 88r\left(\frac{\alpha}{r-\varrho}\right)^2 + 102\frac{\alpha}{r-\varrho}\right]M.$$

Dans ce qui suit nous nous servirons de quelques majorations relatives au noyau $N(P, Q)$ (voir (3)).

Nous rappelons les notations

$$P = (\varrho, \varphi, \vartheta), \quad Q = (\bar{\varrho} = r, \bar{\varphi}, \bar{\vartheta}),$$

et celles de la (3).

Pour $\varrho = r$ on a:

$$\varrho_1|_{\varrho=r} = (2r^2 - 2r^2\cos\psi)^{1/2} = 2r\sin\frac{\psi}{2}, \tag{42}$$

et

$$\varrho_2|_{\varrho=r} = (l^2 + r^2 - 2lr\cos\psi)^{1/2} = \left[(l-r)^2 + 4lr\sin^2\frac{\psi}{2}\right]^{1/2} \geqq$$

$$\geqq \begin{cases} \frac{r}{2}, & \text{si} \quad 0 \leqq l \leqq \frac{r}{2}, \\ \sqrt{2}\, r\sin\frac{\psi}{2}, & \text{si} \quad \frac{r}{2} \leqq l \leqq r. \end{cases}$$

Il s'ensuit pour $0 \leqq l \leqq r$:

$$\varrho_2|_{\varrho=r} \geqq \frac{r}{2}\sin\frac{\psi}{2}. \tag{43}$$

En vertu de (43) on a pour la fonction à intégrer, qui figure dans (3):

$$\left|\frac{1}{l}\left(\frac{1}{r}-\frac{r}{\varrho}\frac{1}{\varrho_2}\right)\right|_{\varrho=r} = \frac{1}{l}\left|\frac{(l^2+r^2-2lr\cos\psi)^{1/2}-r}{r\varrho^2}\right|_{\varrho=r} = \tag{44}$$

$$= \left|\frac{l-2r\cos\psi}{r\varrho_2(r+\varrho_2)}\right|_{\varrho=r} \leqq \left[\frac{3r}{r^2\varrho_2}\right]_{\varrho=r} < \frac{6}{r^2}\frac{1}{\sin\frac{\psi}{2}}.$$

On obtient des (42) et (44)

$$|N(P,Q)|_{\varrho=r} \leqq \frac{1}{4\pi}\left(\frac{2}{2r\sin\frac{\psi}{2}}+r\frac{6}{r^2\sin\frac{\psi}{2}}\right) \leqq \frac{7}{12}\frac{1}{r\sin\frac{\psi}{2}}. \tag{45}$$

Il suit avec un calcul élémentaire:

$$\left[\frac{\partial}{\partial\vartheta}\frac{1}{\varrho_1}\right]_{\substack{\varrho=r\\ \varphi=\vartheta=0}} = -\frac{1}{2}(2r^2-2r^2\cos\bar{\vartheta})^{-3/2}(-2r^2)\sin\bar{\vartheta}\cos\bar{\varphi} =$$

$$= \frac{1}{8r\sin^3\frac{\bar{\vartheta}}{2}}\sin\bar{\vartheta}\cos\bar{\varphi},$$

$$\left[\frac{\partial}{\partial\vartheta}\frac{1}{\varrho_2}\right]_{\substack{\varrho=r\\ \varphi=\vartheta=0}} = \frac{lr}{(l^2+r^2-2lr\cos\bar{\vartheta})^{3/2}}\sin\bar{\vartheta}\cos\bar{\varphi};$$

par conséquent:

$$K(\bar{\varphi},\bar{\vartheta}) \equiv \left[\frac{\partial N}{\partial\vartheta}\right]_{\substack{\varrho=r\\ \varphi=\vartheta=0}} =$$

$$= \frac{1}{4\pi}\left[\frac{1}{4r}\frac{1}{\sin^3\frac{\bar{\vartheta}}{2}}+r\int_0^r\frac{dl}{(l^2+r^2-2lr\cos\bar{\vartheta})^{3/2}}\right]\sin\bar{\vartheta}\cos\bar{\varphi} =$$

$$= \frac{1}{4\pi}\left[\frac{1}{4r}\frac{1}{\sin^3\frac{\bar{\vartheta}}{2}}+r\int_0^r\frac{dl}{[(l-r\cos\bar{\vartheta})^2+r^2\sin^2\bar{\vartheta}]^{3/2}}\right]\sin\bar{\vartheta}\cos\bar{\varphi} =$$

(en utilisant la substitution $l - r\cos\bar{\vartheta} = r\sin\bar{\vartheta}\cdot\tau$)

$$= \frac{1}{4\pi}\left[\frac{1}{4\pi}\frac{1}{\sin^3\frac{\bar{\vartheta}}{2}} + r\int_{-\operatorname{ctg}\bar{\vartheta}}^{\operatorname{tg}\frac{\bar{\vartheta}}{2}} \frac{d\tau}{(1+\tau^2)^{3/2}}\frac{1}{r^2\sin^2\bar{\vartheta}}\right]\sin\bar{\vartheta}\cos\bar{\varphi} =$$

(46)

$$= \frac{1}{4\pi r}\left[\frac{1}{2}\frac{\cos\frac{\bar{\vartheta}}{2}}{\sin^2\frac{\bar{\vartheta}}{2}} + \frac{1}{\sin\bar{\vartheta}}\int_{-\operatorname{ctg}\bar{\vartheta}}^{\operatorname{tg}\frac{\bar{\vartheta}}{2}} \frac{d\tau}{(1+\tau^2)^{3/2}}\right]\cos\bar{\varphi}.$$

Il en résulte, pour $0 \leqq \bar{\vartheta} \leqq \frac{\pi}{2}$, que

$$|K(\bar{\varphi},\bar{\vartheta})| \leqq \frac{1}{4\pi r}\left[\frac{1}{2}\frac{1}{\left(\frac{2}{\pi}\frac{\bar{\vartheta}}{2}\right)^2} + \frac{1}{\frac{2}{\pi}\bar{\vartheta}}\int_{-\infty}^{+\infty}\frac{d\tau}{(1+\tau^2)^{3/2}}\right] \leqq$$

(47)

$$\leqq \frac{1}{4\pi r}\left[\frac{\pi^2}{2}\frac{1}{\bar{\vartheta}^2} + \pi\frac{\bar{\vartheta}}{\bar{\vartheta}^2}\right] < \frac{1}{r}\frac{1}{\bar{\vartheta}^2} \qquad \left(0 \leqq \bar{\vartheta} \leqq \frac{\pi}{2}\right).$$

On obtient, avec les mêmes majorations (47):

(48)
$$\left|\frac{\partial K}{\partial\bar{\varphi}}\right| \leqq \frac{1}{r}\frac{1}{\bar{\vartheta}^2} \qquad \left(0 \leqq \bar{\vartheta} \leqq \frac{\pi}{2}\right).$$

Il découle de (46):

$$\left|\frac{\partial K}{\partial\bar{\vartheta}}\right| \leqq \frac{1}{4\pi r}\left|-\frac{1+\cos^2\frac{\bar{\vartheta}}{2}}{4\sin^3\frac{\bar{\vartheta}}{2}} - \frac{\cos\bar{\vartheta}}{\sin^2\bar{\vartheta}}\int_{-\operatorname{ctg}\bar{\vartheta}}^{\operatorname{tg}\frac{\bar{\vartheta}}{2}}\frac{d\tau}{(1+\tau^2)^{3/2}} + \frac{1}{\sin\bar{\vartheta}}\left(\frac{1}{2}\cos\frac{\bar{\vartheta}}{2} - \sin\bar{\vartheta}\right)\right| \leqq$$

(49)

$$\leqq \frac{1}{4\pi r}\frac{21}{32}\frac{\pi^3}{\bar{\vartheta}^3} < \frac{2}{r}\frac{1}{\bar{\vartheta}^3} \qquad \left(0 < \vartheta \leqq \frac{\pi}{4}\right).$$

On a de même

(50)
$$\left|\frac{\partial^2 K}{\partial\bar{\varphi}\,\partial\bar{\vartheta}}\right| \leqq \frac{2}{r}\frac{1}{\bar{\vartheta}^3} \qquad \left(0 < \vartheta \leqq \frac{\pi}{4}\right).$$

DÉMONSTRATION DU LEMME 4. On a, pour $\vartheta = 0$: $\psi = \bar{\vartheta}$. Étant donné que dans (2)

$$d\sigma_Q = r^2\sin\bar{\vartheta}\,d\bar{\vartheta}\,d\bar{\varphi},$$

compte tenu de (45), on a pour $N(P, Q)d\sigma_Q$ dans (2):

$$(51) \qquad |[N(P, Q)]_{P=(r,\varphi,\vartheta=0)}\, d\sigma_Q| \leqq \frac{7}{12}\frac{1}{r\sin\frac{\bar{\vartheta}}{2}} r^2 \sin\bar{\vartheta}\, d\bar{\vartheta}\, d\bar{\varphi} \leqq \frac{7}{6} r\, d\bar{\vartheta}\, d\bar{\varphi}.$$

Soit $P(r, \varphi, \vartheta)$ un point arbitraire de σ. Choisissons un nouveau système de coordonnées $\varrho' = \varrho$, φ', ϑ', dont le pôle soit P:

$$(r, \varphi, \vartheta) = (r, \varphi', \vartheta' = 0).$$

Étant donné que $V(\varphi, \vartheta)$ s'annule sauf pour $0 \leqq \vartheta \leqq \omega$, $V \equiv 0$ en dehors de l'intervalle $\vartheta'_0 \leqq \vartheta' \leqq \vartheta'_1$, où

$$\vartheta'_0 = \max(0, \vartheta - \omega),$$

$$\vartheta'_1 = \min(\pi, \vartheta + \omega).$$

On obtient donc des (2) et (51):

$$|v(r, \varphi', \vartheta')| \leqq \int_0^{2\pi}\int_{\vartheta'_0}^{\vartheta'_1} 1 \cdot \frac{7}{6} r\, d\bar{\vartheta}\, d\bar{\varphi} \leqq \frac{28}{6}\pi r\omega < 15r\omega.$$

Démonstration du Lemme 5. On a, par la définition de $K(\bar{\varphi}, \bar{\vartheta})$:

$$\frac{\partial V(r, \varphi, \vartheta = 0)}{\partial \vartheta} = r^2 \int_0^{\omega}\int_0^{2\pi} K(\bar{\varphi}, \bar{\vartheta}) V(\bar{\varphi}, \bar{\vartheta}) \sin\bar{\vartheta}\, d\bar{\varphi}\, d\bar{\vartheta} =$$

(avex deux intégrations par partie)

$$= r^2 \left\{ K(2\pi, \omega) \int_0^{\omega}\int_0^{2\pi} V(\bar{\varphi}, \bar{\vartheta}) \sin\bar{\vartheta}\, d\bar{\varphi}\, d\bar{\vartheta} - \right.$$

$$- \int_0^{\omega} \frac{\partial K(2\pi, \bar{\vartheta})}{\partial \bar{\vartheta}} \left[\int_0^{\bar{\vartheta}}\int_0^{2\pi} V(\bar{\bar{\varphi}}, \bar{\bar{\vartheta}}) \sin\bar{\bar{\vartheta}}\, d\bar{\bar{\varphi}}\, d\bar{\bar{\vartheta}} \right] d\bar{\vartheta} -$$

$$- \int_0^{2\pi} \left[\frac{\partial K(\bar{\varphi}, \omega)}{\partial \bar{\varphi}} \int_0^{\omega}\int_0^{\bar{\varphi}} V(\bar{\bar{\varphi}}, \bar{\bar{\vartheta}}) \sin\bar{\bar{\vartheta}}\, d\bar{\bar{\varphi}}\, d\bar{\bar{\vartheta}} - \right.$$

$$\left. \left. - \int_0^{\omega} \frac{\partial^2 K(\bar{\varphi}, \bar{\vartheta})}{\partial \bar{\varphi}\, \partial \bar{\vartheta}} \left(\int_0^{\bar{\vartheta}}\int_0^{\bar{\varphi}} V(\bar{\bar{\varphi}}, \bar{\bar{\vartheta}}) \sin\bar{\bar{\vartheta}}\, d\bar{\bar{\varphi}}\, d\bar{\bar{\vartheta}} \right) d\bar{\vartheta} \right] d\bar{\varphi} \right\}.$$

Il en résulte, en utilisant les majorations (47)—(50) et l'hypothèse relative à $V(\varphi, \vartheta)$:

$$\left| \frac{\partial v(r, \varphi, \vartheta = 0)}{\partial \vartheta} \right| = r^2 \left\{ \frac{1}{r}\frac{1}{\omega^2}\omega^3 + \int_0^{\omega} \frac{2}{r}\frac{1}{\bar{\vartheta}^3}\bar{\vartheta}^3\, d\bar{\vartheta} + \right.$$

$$\left. + 2\pi \left[\frac{1}{r}\frac{1}{\omega^2}\omega^3 + \int_0^{\omega} \frac{2}{r}\frac{1}{\bar{\vartheta}^3}\bar{\vartheta}^3\, d\bar{\vartheta} \right] \right\} \leqq 22r\omega.$$

DÉMONSTRATION DU LEMME 6. Il vient de la formule (2), par la définition de $K(\varphi, \vartheta)$ et en vertu de (47):

$$\left|\frac{\partial v(r,\varphi,0)}{\partial \vartheta}\right| = r^2 \left| \int_0^{\omega}\int_0^{2\pi} K(\bar{\varphi},\bar{\vartheta})\, V(\bar{\varphi},\bar{\vartheta}) \sin\bar{\vartheta}\, d\bar{\varphi}\, d\bar{\vartheta} \right| \leqq$$

$$\leqq r^2 \int_0^{2\pi}\int_0^{\omega} \frac{1}{r}\,\frac{1}{\bar{\vartheta}^2}\, V(\varphi,\bar{\vartheta})\,\bar{\vartheta}\, d\bar{\vartheta}\, d\bar{\varphi} \leqq r\cdot 1 = r.$$

DÉMONSTRATION DU LEMME 7. On peut supposer, sans restreindre la généralité, que $\varphi = 0$. On peut supposer de plus, que la fonction $u(\varrho, \varphi, \vartheta) \equiv u(x, y, z)$ est antisymmétrique par raport au plan $x = 0$, parce que

$$\frac{\partial u(\varrho,\ \varphi=0,\ \vartheta=0)}{\partial \vartheta} = r\left[\frac{\partial u(x,y,z)}{\partial x}\right]_{\substack{x=y=0\\ z=\varrho}} =$$

$$= r\left[\frac{\partial}{\partial x}\left\{\frac{u(x,y,z)-u(-x,y,z)}{2}\right\}\right]_{\substack{x=y=0\\ z=\varrho}},$$

où la fonction entre les { } est antisymmétrique. C'est-à-dire qu'on peut supposer

$$u(x=0,\ y,\ z)=0.$$

Admettons l'existence d'une fonction $h(x, y, z) \equiv h(\varrho, \varphi, \vartheta)$, qui sera construite plus tard, satisfaisant aux conditions suivantes:

1. h est harmonique dans l'ensemble

$$\{0 \leqq \varrho < r,\ x \geqq 0\};$$

2. h est non-négative, continue et admette des dérivées continues dans l'ensemble

$$\{0 \leqq \varrho \leqq r,\ x \geqq 0\} - \{\varrho = r,\ x = 0,\ \omega \leqq \vartheta \leqq \pi\};$$

3. h s'ennule sur l'ensemble

$$\{0 \leqq \varrho \leqq r,\ x = 0\} - \{\varrho = r,\ x = 0,\ \omega \leqq \vartheta < \pi\},$$

4. $\partial h/\partial \varrho$ s'ennule sur l'ensemble

$$\{\varrho = r,\ x \geqq 0,\ 0 \leqq \vartheta < \omega\},$$

5. $h(P) \geqq \chi_1 > 0,\ P \in \{\varrho = r,\ x > 0,\ \omega \leqq \vartheta < \pi\}$,

6. $$0 < \left.\frac{\partial h}{\partial \vartheta}\right|_{\substack{\varrho=r\\ \varphi=\vartheta=0}} \leqq \chi_2.$$

Dans ces hypothèses nous obtenons les résultats suivants pour la fonction harmonique $v \equiv \frac{h}{\chi_1} - u$:

I. $\frac{\partial v}{\partial \varrho} = 0$ sur l'ensemble $\{\varrho = r,\ x \geqq 0,\ 0 \leqq \vartheta < \omega\}$,

II. $v \geqq 0$ sur l'ensemble $\{\varrho = r,\ x > 0,\ \omega \leqq \vartheta < \pi\}$,

III. $v = 0$ sur l'ensemble $\{0 \leqq \varrho \leqq r,\ x = 0\} - \{\varrho = r,\ x = 0,\ \omega \leqq \vartheta \leqq \pi\}$,

IV. v est bornée inférieurement sur l'ensemble $\{0 \leqq \varrho < r,\ x > 0\}$.

Il découle de ces propriétés (en vertu du principe du maximum), que $v \geqq 0$ sur l'ensemble $\{0 \leqq \varrho \leqq r,\ x \geqq 0\} - \{\varrho = r,\ x = 0,\ \omega \leqq \vartheta \leqq \pi\}$, donc, étant donné que $v(\varrho = r,\ \vartheta = 0) = 0$,

$$\left.\frac{\partial v}{\partial \vartheta}\right|_{\substack{\varrho = r \\ \varphi = \vartheta = 0}} \geqq 0,$$

c'est-à-dire

$$\left.\frac{\partial u}{\partial \vartheta}\right|_{\substack{\varrho = r \\ \varphi = \vartheta = 0}} \leqq \frac{\chi_2}{\chi_1}.$$

On obtient d'une façon analogue, à partir de la fonction $v \equiv \frac{h}{\chi_1} + u$, que

$$\left.\frac{\partial u}{\partial \vartheta}\right|_{\substack{\varrho = r \\ \varphi = \vartheta = 0}} \geqq -\frac{\chi_2}{\chi_1}.$$

On a ainsi

$$\left|\left.\frac{\partial u}{\partial \vartheta}\right|_{\substack{\varrho = r \\ \varphi = \vartheta = 0}}\right| \leqq \frac{\chi_2}{\chi_1}.$$

Il nous reste donc, pour compléter la démonstration, la construction d'une telle fonction h, et le calcul des constants χ_1, χ_2.

Nous rappelons les notations de (3), et affirmons que la fonction

$$h(\varrho, \varphi, \vartheta) = 4\pi \int_{\omega}^{\pi} \left.\frac{\partial N}{\partial \xi}\right|_{\substack{l = r \\ \xi = 0 \\ \bar{\varphi} = \frac{\pi}{2}}} \bar{\vartheta}\, d\bar{\vartheta} + 4\pi \int_{\varphi}^{\pi} \left.\frac{\partial N}{\partial \xi}\right|_{\substack{l = r \\ \xi = 0 \\ \bar{\varphi} = -\frac{\pi}{2}}} \bar{\vartheta}\, d\bar{\vartheta}$$

satisfaite aux conditions 1.—6. avec

$$\chi_1 = \frac{3}{16r^2}, \quad \chi_2 = \frac{50}{r^2 \omega}.$$

Les conditions 1.—4., compte tenu de la définition de $N(P, Q)$ et du fait que $\left.\frac{\partial N}{\partial \xi}\right|_{\xi = 0}$ est une fonction impaire en x, sont évidemment satisfaites.

On obtient avec un calcul élémentaire mais assez long:

$$h(r,\varphi,\vartheta) = r\sin\vartheta\cos\varphi\cdot$$

(52)

$$\cdot\left[2\int_{\omega}^{\pi}\frac{\bar{\vartheta}\,d\bar{\vartheta}}{\varrho_{11}^3}+2\int_{\omega}^{\pi}\frac{\bar{\vartheta}\,d\bar{\vartheta}}{\varrho_{12}^3}+\frac{1}{r}\int_{\omega}^{\pi}\bar{\vartheta}\int_{0}^{r}\frac{dl}{\varrho_{21}^3}\,d\bar{\vartheta}+\frac{1}{r}\int_{\omega}^{\pi}\bar{\vartheta}\int_{0}^{r}\frac{dl}{\varrho_{22}^3}\,d\bar{\vartheta}\right],$$

où

$$\varrho_{11}=\overline{PQ_1},\quad \varrho_{12}=\overline{PQ_2},$$

$$\varrho_{21}=\overline{PQ_1^*},\quad \varrho_{22}=\overline{PQ_2^*},$$

$$P=(\varrho=r,\varphi,\vartheta),$$

$$Q_1=\left(\bar{\varrho}=r,\ \bar{\varphi}=\frac{\pi}{2},\ \bar{\vartheta}\right),\quad Q_2=\left(\bar{\varrho}=r,\ \bar{\varphi}=-\frac{\pi}{2},\ \bar{\vartheta}\right),$$

$$Q_1^*=\left(\bar{\varrho}=l,\ \bar{\varphi}=\frac{\pi}{2},\ \bar{\vartheta}\right),\quad Q_2^*=\left(\bar{\varrho}=l,\ \bar{\varphi}=-\frac{\pi}{2},\ \bar{\vartheta}\right).$$

Posons

$$P'=\left(\varrho=r,\ \varphi=\frac{\pi}{2},\ \vartheta\right).$$

Pour démontrer 5. il suffit de nous limiter au cas $y\geqq 0$, c'est-à-dire $0\leqq\varphi\leqq\frac{\pi}{2}$.

De plus, il suffit de minorer le premier term de h, ce qui contient ϱ_{11}, parce que tous les termes sont évidemment positifs.

On obtient aisément:

$$\varrho_{11}\leqq\overline{PP'}+\overline{P'Q_1}=r\sqrt{2}\sin\vartheta(1-\sin\varphi)^{1/2}+2r\sin\frac{|\vartheta-\bar{\vartheta}|}{2}\leqq$$

$$\leqq r\sqrt{2}\sin\vartheta\cos\varphi+2r\sin\frac{|\vartheta-\bar{\vartheta}|}{2}\leqq\sqrt{2}\,x+r|\vartheta-\bar{\vartheta}|.$$

En vertu de cette inégalité, on obtient de (52):

$$h(\varrho=r)\geqq 2x\int_{\omega}^{\pi}\frac{\bar{\vartheta}\,d\bar{\vartheta}}{(\sqrt{2}\,x+r|\vartheta-\bar{\vartheta}|)^3}.$$

Nous distinguerons deux cas:

A) $\frac{\pi}{2}\leqq\vartheta\leqq\pi$.

On a, dans ce cas:

$$h(\varrho=r)\geqq 2x\int_{\frac{\pi}{4}}^{\vartheta}\frac{\bar{\vartheta}\,d\bar{\vartheta}}{[\sqrt{2}\,x+r(\vartheta-\bar{\vartheta})]^3}>2x\int_{\frac{\pi}{4}}^{\vartheta}\frac{\frac{\pi}{4}\,d\bar{\vartheta}}{[\cdots]^3}=$$

(par la substitution $r(\vartheta - \bar{\vartheta}) = \sqrt{2}x\tau$)

$$= 2x \int\limits_{\frac{r\left(\vartheta-\frac{\pi}{4}\right)}{\sqrt{2}x}}^{0} \frac{-\frac{\pi}{4}\frac{1}{r}\sqrt{2}\,x\,dt}{[\sqrt{2}\,x(1+\tau)]^3} > \frac{1}{5rx} \geqq \frac{1}{5r^2}.$$

B) $\omega \leqq \vartheta \leqq \dfrac{\pi}{2}$.

Dans ce cas:

$$h(\varrho = r) \geqq 2x \int\limits_{\vartheta}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{[\sqrt{2}\,x + r(\bar{\vartheta} - \vartheta)]^3} = \frac{1}{rx} \int\limits_{0}^{\frac{r(\pi-\vartheta)}{\sqrt{2}x}} \frac{\vartheta + \frac{\sqrt{2}\,x}{r}\tau}{(1+\tau)^3}\,d\tau >$$

$$> \frac{1}{rx} \int\limits_{0}^{\frac{r\frac{\pi}{2}}{\sqrt{2}r}} \frac{\frac{\sqrt{2}\,x}{r}\tau}{(1+\tau)^3}\,d\tau > \frac{3}{16r^2}.$$

En tenant compte de ces majorations, nous obtenons

$$h(\varrho = r) > \frac{3}{16r^2} = \chi_1 \qquad (\omega \leqq \vartheta \leqq \pi).$$

On a évidemment:

$$\left.\frac{\partial h}{\partial \vartheta}\right|_{\substack{\varrho=r \\ \varphi=\vartheta=0}} = r\left[4\int\limits_{\omega}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{\varrho_{11}^3} + 2\frac{1}{r}\int\limits_{\omega}^{\pi} \bar{\vartheta} \int\limits_{0}^{r} \frac{dl}{\varrho_{21}^3}\,d\bar{\vartheta}\right]_{\substack{\varrho=r \\ \varphi=\vartheta=0}}, \tag{53}$$

parce que pour $\varrho = r$: $\varrho_{11} = \varrho_{12}$, $\varrho_{21} = \varrho_{22}$; dans ce cas on obtient:

$$\varrho_{11} = r[\sin^2\bar{\vartheta} + (1 - \cos\bar{\vartheta})^2]^{1/2} = 2r\sin\frac{\bar{\vartheta}}{2} \geqq 2r\frac{\bar{\vartheta}}{\pi},$$

$$\varrho_{21} = (l^2 + r^2 - 2lr\cos\bar{\vartheta})^{1/2} = \left[(l-r)^2 + 4lr\sin^2\frac{\bar{\vartheta}}{2}\right]^{1/2} \geqq \left[(r-l)^2 + 4lr\frac{\vartheta^2}{\pi^2}\right]^{1/2}.$$

En utilisant ces inégalités, nous majorons séparément les intégrales dans (53):

$$\int\limits_{\omega}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{\varrho_{11}^3} < \int\limits_{\omega}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{8r^3\frac{\bar{\vartheta}^3}{\pi^3}} < \frac{4}{r^3}\frac{1}{\omega};$$

$$\int\limits_{\omega}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{\varrho_{21}^3} \leqq \int\limits_{\omega}^{\pi} \frac{\bar{\vartheta}\,d\bar{\vartheta}}{\left[(r-l)^2 + 4lr\frac{\vartheta^2}{\pi^2}\right]^{3/2}} =$$

$$\left(\text{par la substitution } 4lr\frac{\overline{\vartheta}^2}{\pi^2} = (r-l)^2\tau^2\right)$$

$$= \int_{\frac{2\sqrt{lr}\,\omega}{\pi(r-l)}}^{\frac{2\sqrt{lr}}{r-l}} \frac{\tau\,d\tau}{(1+\tau^2)^{3/2}}\,\frac{\pi^2}{4lr}\,\frac{1}{r-l} < \frac{17}{r^3}\,\frac{1}{\omega}.$$

Il en résulte que

$$\left.\frac{\partial h}{\partial \vartheta}\right|_{\substack{\varrho=r\\ \varphi=\vartheta=0}} < r\left(4\,\frac{4}{r^3}\,\frac{1}{\omega} + \frac{2}{r}\int_0^r \frac{17}{r^3}\,\frac{1}{\omega}\,dl\right) = \frac{50}{r^2}\,\frac{1}{\omega} = \chi_2.$$

C.q.f.d.

BIBLIOGRAPHIE

[1] ADLER, G.: Maggiorazione del gradiente delle funzioni armoniche mediante i loro valori al contorno, *Memorie dell'Accad. Naz. dei Lincei*, Serie VIII., **6** (1961) 185—201.

[2] ADLER, G.: Majoration du gradient des solutions de l'équation $\Delta u - au'_t = f$, I, II, *Acta Math. Acad. Sci. Hungar.* **15** (1964) 137—152, 259—283.

[3] ADLER, G.: Majoration des tensions dans un corps élastique à l'aide des déplacements superficiels, *Archive for Rational Mech. and Anal.* **16** (1964) 354—372.

[4] ADLER, G.: Maggiorazione delle funzioni armoniche con condizioni al contorno di tipo misto, *Atti della Acad. delle Sci. di Torino*, **95** (1960—61) 629—638.

[5] ADLER, G.: Analyse quantitative de l'allure de la dérivée normale de la solution du problème de Neumann, *Acta Math. Acad. Sci. Hungar.* **17** (1966) 369—378.

[6] AGMON, S., DOUGLIS, A., NIRENBERG, L.: Estimates Near the Boundary for Solutions of Elliptic Partial Differential Equations Satisfying General Boundary Conditions, I, II, *Comm. on Pure and Applied Mathematics*, **12** (1959) 623—727, **17** (1964) 35—92.

Institut de Mathématique de L'Académie des Sciences de Hongrie, Budapest

(Reçu le 17 octobre 1966.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 71—80.

ЗАДАЧА О ПОДСЧЕТЕ ЧИСЛА НЕКОТОРЫХ ДЕРЕВЬЕВ

G. OLÁH

Под графом понимается неориентированный граф без петель и кратных ребер. Вершины графа считаются различными, т. е. они снабжены числовыми индексами $(P_1, P_2, \ldots, P_n)$.

Связный граф циклов называется деревом. Обозначим t_n число различных деревьев с n данными вершинами. A. CAYLEY [5] впервые доказал, что

$$t_n = n^{n-2}. \tag{1}$$

Простейшим способом доказательства этой формулы является способ, предложенный A. PRÜFER-ом [6].

Дальнейшая задача о подсчете числа деревьев — следующая: Пусть дано n точек $P_1, P_2, \ldots, P_n$ из класса P и m точек $Q_1, Q_2, \ldots, Q_m$ из класса Q. Обозначим $t(n, m)$ число различных таких деревьев, вершинами которых являются $P_1, \ldots, P_n$ и $Q_1, \ldots, Q_m$, и каждое ребро которых соединяет одну вершину из класса P с одной вершиной из класса Q. H. I. SCOINS [7] впервые доказал, что

$$t(n, m) = n^{m-1} m^{n-1}. \tag{2}$$

Недавно RÉNYI A. [4] дал способом PRÜFER-а новое простейшее доказательство этой формулы (2).

Естественно следующее обобщение этой проблемы: дано n классов точек. Эти классы состоят из $l_1, l_2, \ldots, l_n$ точек соответственно. Пусть $t(l_1, l_2, \ldots, l_n)$ означает число деревьев, вершинами которых являются точки данных классов, и все ребра которых соединяют точки двух различных классов.

Методом PRÜFER-а мы докажем, что справедлива следующая

Теорема:

$$t(l_1, l_2, \ldots l_n) = \left(\sum_{i=1}^{n} l_i\right)^{n-2} \prod_{j=1}^{n} \left(\sum_{i=1}^{n} l_i - l_j\right)^{l_j - 1}. \tag{3}$$

Прежде чем приступить к доказательству теоремы, целесообразно познакомиться с неоднократно упоминавшимся методом PRÜFER-а. Поэтому мы докажем этим методом формулу CAYLEY (1).

Если удастся доказать, что множество деревьев с вершинами $1, 2, \ldots, n$ взаимно однозначно можно отобразить на множество последовательностей, составленных из чисел $1, 2, \ldots, n$ и состоящих из $n-2$ элементов, то (1) будет доказано, так как число таких последовательностей равно n^{n-2}.

Это отображение можно реализовать следующим образом: пусть G есть дерево, состоящее из n вершин, номер которых есть 1, 2, ..., n соответственно. Выберем из висячих вершин G ту, номер которой — наибольший, пусть это будет P_{y_1}. Пусть P_{x_1} есть та (единственная) вершина из G, которая соединена ребром с P_{y_1}. Обозначим через G' дерево, получающееся из G, если удалить вершину P_{y_1} и ребро $P_{y_1}P_{x_1}$. Пусть P_{y_2} есть та висячая вершина G', номер которой — наибольший, а P_{x_2} (единственная) вершина из G', соединенная ребром с P_{y_2}. Опуская вершину P_{y_2} и ребро $P_{y_2}P_{x_2}$, получаем из G' дерево G''. Продолжаем эту процедуру до тех пор, пока не получим дерево, состоящее из двух ребер. Сопоставим дереву G последовательность $(x_1, x_2, \ldots, x_{n-2})$. Покажем, что все последовательности, составленные из чисел 1, 2, ..., n и состоящие из $n-2$ элементов, могут быть получены таким образом, причем различным последовательностям соответствуют различные деревья и наоборот: различным деревьям соответствуют различные последовательности.

Первое утверждение доказывается так: для любой последовательности $(x_1, x_2, \ldots, x_{n-2})$, составленной из чисел 1, 2, ..., n, находится дерево, которому алгорифм PRÜFER-а сопоставляет эту последовательность. Пусть $z_1, z_2, \ldots, z_k$ суть расположенные в убывающем порядке те из чисел 1, 2, ..., n, которые не фигурируют в последовательности $(x_1, x_2, \ldots, x_{n-2})$. Очевидно $k \geqq 2$. Соединим ребром P_{z_1} и P_{x_1}. Затем рассмотрим последовательности $(x_2, \ldots, x_{n-2})$ и $(z_2, \ldots, z_k)$, полученные вычеркиванием x_1 и z_1. Если x_1 не встречается в последовательности $(x_2, \ldots, x_{n-2})$, то впишем его в последовательность $(z_2, \ldots, z_k)$ так, чтобы она осталась убывающей; если же x_1 фигурирует в последовательности $(x_2, \ldots, x_{n-2})$, то оставим $(z_2, \ldots, z_k)$ без изменений. Обозначим через $(z_1', z_2', \ldots, z_{k'}')$ последовательность, полученную таким образом из $(z_2, \ldots, z_k)$. Соединим ребром вершины $P_{z_1'}$ и P_{x_2} и продолжим этот процесс до тех пор, пока не будут использованы все числа x_i. Наконец соединим ребром вершины, соответствующие двум оставшимся числам z. Индукцией можно доказать, что так всегда получится дерево.

Второе утверждение, согласно которому различным последовательностям соответствуют различные деревья и различным деревьям соответствуют различные последовательности, очевидно следует из построения. Таким образом формула (1) доказана.

Переходим к доказательству теоремы (3). Пусть даны n классов точек, состоящие из $l_1, l_2, \ldots, l_n$ точек соответственно. Перенумеруем точки следующим образом: точкам первого класса сопоставим номера $n, 2n, \ldots, l_1 n$; точкам второго класса — номера $n-1, 2n-1, \ldots, l_2 n-1$; и т. д., наконец точкам последнего класса — номера $1, n+1, \ldots, l_n n-(n-1)$. Первый класс точек будем называть классом с остатком n, точки второго класса — классом с остатком $n-1$, и т. д., наконец последний класс назовем классом с остатком 1. Обозначим через G дерево, которое имеет $\sum_{i=1}^{n} l_i$ вершин, перенумерованные указанным способом, и все ребра которого соединяют пару вершин из классов с различным остатком. Дереву G сопоставляется последовательность вида

(4) $$(\ldots\ldots\ldots)\,(\ldots\ldots\ldots)\ \ldots\ (\ldots\ldots\ldots)\ \ldots,$$

$$l_1-1 \text{ мест} \quad l_2-1 \text{ мест} \qquad l_n-1 \text{ мест}$$

где на место l_1-1 точек в первой паре скобок будут вписываться числа, служащие номерами вершин дерева, остаток которых не равен n (число их равно $\sum_{i=1}^{n} l_i - l_1$); на место l_2-1 точек во второй паре скобок будут вписываться числа, служащие номерами вершин, остаток которых не равен $n-1$ (число их равно $\sum_{i=1}^{n} l_i - l_2$); и т. д., на место l_n-1 точек в последней паре скобок вписываются числа, служащие номерами вершин, остаток которых не равен 1 (число их равно $\sum_{i=1}^{n} l_i - l_n$); наконец, после последней скобки на место точек, число которых мы фиксируем позже, будут вписываться числа, служащие номерами вершин (число этих номеров равно $\sum_{i=1}^{n} l_i$).

Это сопоставление осуществляется следующим образом: из висячих вершин G выбирается та, номер которой — наибольший, пусть эта будет P_{y_1}, из класса с остатком k. Обозначим через P_{x_1} (единственную) вершину G, соединенную ребром с P_{y_1}. Запишем x_1 в (4) на первое место в k-той сзади паре скобок, если там есть место, в противном случае x_1 записывается на первое место после всех скобок. Обозначим через G' дерево, получаемое из G, удалением вершины P_{y_1} и ребра $P_{y_1}P_{x_1}$. Пусть P_{y_2} — та из висячих вершин G', номер которой — наибольший. Пусть P_{y_2} принадлежит классу с остатком k'. Обозначим через P_{x_2} (единственную) вершину из G', с которой P_{y_2} соединена ребром. Если в k'-той сзади паре скобок, соответствующей классу с остатком k', еще есть место, впишем на первое свободное место x_2. В противном случае запишем x_2 на первое свободное место после всех скобок. Обозначим через G'' дерево, получающееся из G', если из него удалить вершину P_{y_2} и ребро $P_{y_2}P_{x_2}$. Эта процедура продолжается до тех пор, пока не получается дерево с двумя вершинами. Они принадлежат двум разным классам, поэтому в ходе процедуры из каждого класса были вычеркнуты все вершины, кроме, быть может, одной. Очевидно в (4) были вписаны числа, соответствующие $\sum_{i=1}^{n} l_i - 2$ точкам, причем все скобки были заполнены. Поэтому после последней скобки фигурирует точно

$$\sum_{i=1}^{n} l_i - 2 - \sum_{j=1}^{n} (l_j - 1) = n-2$$

точек.

Из построения очевидно, что таким образом дереву указанного типа соответствует последовательность вида

(.....) (.....) ... (.....) (.....)

l_1-1 место l_2-1 место l_n-1 место $n-2$ места

Покажем, что наоборот, любой последовательности указанного типа можно сопоставить дерево, которому алгорифм PRÜFER-а сопоставляет именно эту последовательность, и что различным последовательностям сопоставляются различные деревья, а различным деревьям — различные последовательности.

Пусть дана любая последовательность

$$(4^*) \qquad \begin{aligned} &(u_{11}, u_{12}. \ldots u_{1,l_1-1})(u_{21}, u_{22}, \ldots, u_{2,l_2-1}) \ldots \\ &\ldots (u_{n1}, u_{n2}, \ldots, u_{n,l_n-1})(u_1, u_2, \ldots, u_{n-2}) \end{aligned}$$

в которой фигурирующие в последней паре скобок числа u_i выбираются из чисел

$$(5) \qquad \begin{aligned} &n, \quad 2n, \quad \ldots \quad l_1 n, \\ &n-1, \quad 2n-1, \quad \ldots \quad l_2 n-1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &1, \quad n+1, \quad \ldots \quad l_n n-(n-1), \end{aligned}$$

фигурирующие в первой паре скобок числа u_{1k} выбираются из тех же чисел, за исключением первой строки; числами во второй паре скобок могут быть числа из (5), кроме элементов второй строки; и т. д.; числами в предпоследней паре скобок могут служить числа из (5), кроме элементов последней строки. Пусть $(z_{11}, z_{12}, \ldots, z_{1r_1})$ суть расположенные в убывающем порядке числа из первой строки (5), не встречающиеся в (4*); $(z_{21}, z_{22}, \ldots, z_{2r_2})$ — расположенные в убывающем порядке числа из второй строки (5), не встречающиеся в (4*); и т. д.; $(z_{n1}, z_{n2}, \ldots, z_{nr_n})$ — расположенные в убывающем порядке числа из последней строки (5), не фигурирующие в (4*). Число чисел z не меньше чем 2. Обозначим через z_{k1} наибольшее из чисел z. Соединим ребром вершину $P_{z_{k1}}$ с вершиной $P_{u_{k1}}$, если u_{k1} существует, в противном случае соединим ее с P_{u_1}. Затем вычеркнем из последовательности (4*) число u_{k1} или u_1, а из множества чисел z — число z_{k1}. Если среди чисел u еще встречаются исключенные u_{k1} (или u_1), то продолжим указанным способом реконструкцию дерева. В противном случае проверим, в какой строке (5) фигурирует это число и, если это s-тая строка, то включим это число в последовательность чисел z_{si} так, чтобы она осталась убывающей. Затем продолжим реконструкцию дерева с полученной таким образом последовательностью z. Из конструкции очевидно, что каждое число из s-той строки (5) один и только один раз окажется включенным в множество чисел z_{si} $(s = 1, 2, \ldots, n)$. Поэтому вершину, соответствующую некоторому числу z_{si}, всегда можно соединить ребром согласно конструкции с некоторой вершиной $P_{u_{sj}}$, пока не останется единственное число z_{si}. Затем соответствующую ему вершину, возможно, нужно соединить ребром с некоторой точкой P_{u_j}. Но тогда число u_j уже не может фигурировать в s-той строке (5). Таким образом обеспечено то, что всегда соединяются вершины различных классов. В силу сказанного построение заканчивается тем, что останется пара чисел z, принадлежащих различным строкам из (5). В заключение соединим и соответствующие им вершины. То, что таким образом получается дерево, следует из приведенного выше простого построения PRÜFER-а. Это дерево, как мы видели, такое, что все ребра соединяют вершины двух различных классов. Остается доказать, что различным последовательностям соответствуют различные деревья, а различным деревьям — различные последовательности, но это очевидно из построения.

Таким образом доказано, что множество деревьев указанного типа и множество последовательностей вида (4*) можно привести во взаимно однозначное соответствие. Однако число последовательностей типа (4*) равно

$$\left(\sum_{i=1}^{n} l_i\right)^{n-2} \prod_{j=1}^{n} \left(\sum_{i=1}^{n} l_i - l_j\right)^{l_j - 1}$$

что и доказывает теорему (3).

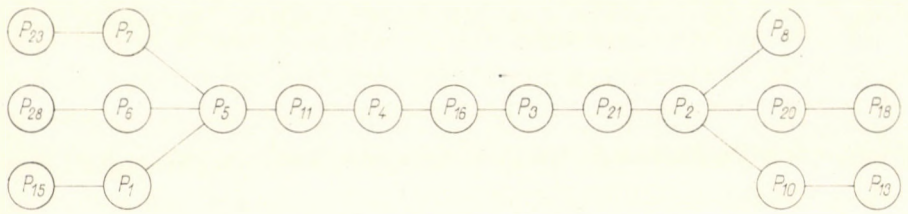
Из теоремы (3) очевидно следует результат SCOINS-а (2) и формула CAYLEY (1). В первом случае вершины дерева принадлежат минимальному числу классов — двум классам, во втором случае — максимальному числу классов, когда каждый класс содержит лишь по одной вершине.

В заключение в виде иллюстрации на одном примере покажем, каким образом можно сопоставить дереву последовательность и наоборот.

Пусть, например $l_1 = 4$, $l_2 = 1$, $l_3 = 6$, $l_4 = 2$, $l_5 = 5$, тогда точки нумеруются следующими числами:

$$\begin{array}{ll} & 20,\ 15,\ 10,\ \ 5, \\ & \ \ 4, \\ (6) & 28,\ 23,\ 18,\ 13,\ \ 8,\ \ 3, \\ & \ \ 7,\ \ 2, \\ & 21,\ 16,\ 11,\ \ 6,\ \ 1. \end{array}$$

Нарисуем дерево, 18 вершин которого перенумерованы этими числами и каждое ребро дерева соединяет пару вершин, номера которых принадлежат различным строкам из (6):



Этому дереву следует сопоставить последовательность следующего вида:

(. . .) ( ) (. . . . .) (.) (. . . .) (. . .)

Среди висячих вершин наибольший номер — 28 соответствует вершине P_{28}. Этот номер принадлежит третьему классу. P_{28} соединена с P_6, поэтому на первое место третьей пары скобок следует вписать 6. Вычеркнув вершину P_{28} и соответствующее ей ребро, получим дерево, среди висячих вершин которого наибольший номер соответствует вершине P_{23}. 23 также принадлежит третьему классу, но P_{23} соединена с вершиной P_7, поэтому на второе место третьей пары скобок следует вписать 7. После вычеркивания P_{23} и

$P_{23}P_7$, висячей вершиной с наибольшим номером окажется P_{18}. 18 также принадлежит третьему классу и P_{18} соединена с вершиной P_{20}, поэтому на третье место третьей пары скобок следует вписать 20. После вычеркивания вершины P_{18} и соответствующего ребра, висячей вершиной с наибольшим номером окажется вершина P_{20}; 20 принадлежит первому классу и P_{20} соединена ребром с вершиной P_2, поэтому на первое место первой пары скобок вписывается 2. После вычеркивания вершины P_{20}, наибольшей окажется P_{15} и так, на следующее место первой пары скобок вписывается 1. После вычеркивания вершины P_{15}, наибольшей окажется P_{13}, поэтому в третью пару скобок вписывается 10. После вычеркивания P_{13}, наибольшей будет P_{10}, поэтому в первую пару скобок вписывается 2. После вычеркивания P_{10}, наибольшей будет P_8, поэтому в третью пару скобок вписывается 2. После вычеркивания P_8, наибольшей будет P_7, поэтому в четвертую пару скобок вписывается 5. После вычеркивания P_7, наибольшей окажется вершина P_6, и так в пятую пару скобок вписывается 5. Вычеркнем и P_6. Дерево примет вид:

Соответствующая последовательность такова:

$$(2, 1, 2)\ (\ )\ (6, 7, 20, 10, 2)\ (5)\ (5, \ldots)\ (\ldots)$$

Висячая вершина с наибольшим номером здесь P_2, поэтому 21 следовало бы вписать в четвертую пару скобок, но там уже нет места, поэтому 21 вписывается в последнюю пару скобок. После вычеркивания P_2, наибольшей будет P_{21}, поэтому в пятую пару скобок вписывается 3. После вычеркивания P_{21}, наибольшей будет P_3, поэтому в третью пару скобок следовало бы вписать 16, но и там уже нет места, поэтому 16 попадает на второе место последней пары скобок. После вычеркивания P_3, наибольшей будет P_{16}, поэтому в пятую пару скобок вписывается 4. После вычеркивания P_{16}, наибольшей будет P_4, поэтому во вторую пару скобок следовало бы вписать 11, но там и не было места, поэтому 11 попадает в последнюю пару скобок. После вычеркивания P_4, наибольшей будет P_{11}, поэтому на последнее место пятой пары скобок вписывается 5. Так мы пришли к дереву с двумя вершинами:

в то же время все места последовательности уже заполнены:

$$(2, 1, 2)\ (\ )\ (6, 7, 20, 10, 2)\ (5)\ (5, 3, 4, 5)\ (21, 16, 11)$$

Таким образом исходному дереву однозначным образом соответствует эта последовательность.

Теперь, исходя из этой последовательности (чисел u) построим соответствующее ей дерево. Для этого определим соответствующие числа z. В первой группе чисел z окажется лишь 15 (так как 20, 10 и 5 фигурируют в последовательности), во второй группе нет ни одного числа (так как 4 фигу-

рирует в последовательности), в третью группу попадут в убывающем порядке 28, 23, 18, 13, 8 (так как в последовательности фигурирует лишь 3), наконец четвертая и пятая группа также не содержат чисел (так как 7, 2 и 21, 16, 11, 6, 1 фигурируют в последовательности). Поэтому последовательность соответствующих z такова:

(15) ( ) (28, 23, 18, 13, 8) ( ) ().

Наибольшим z является 28 в третьей паре скобок. Мы соединим ребром вершины P_{28} и P_6 (так как в последовательности u в третьей паре скобок на первом месте стоит 6) и вычеркнем из последовательности u числа 6, а из последовательности z — число 28. Так как в последовательности u 6 больше не встречается, впишем 6 в последовательность z в соответствующую группу на соответствующее место. 6 попадает в пятую группу и, так как она до сих пор не содержала чисел, то не нужно заботиться о порядке чисел. После этого полученные последовательности u и z таковы:

(2, 1, 2) ( ) (7, 20, 10, 2) (5) (5, 3, 4, 5) (21, 16, 11)

(15) ( ) (23, 18, 13, 8) ( )(6)

Теперь среди z наибольшим окажется 23, поэтому соединим ребром вершины P_{23} и P_7; вычеркнем в последовательности u число 7, стоящее на первом месте в третьей паре скобок, а в последовательности z число 23. 7 не встречается больше среди u, поэтому запишем 7 среди z на соответствующее место.

Новые последовательности таковы:

(2, 1, 2) ( ) (20, 10, 2) (5) (5, 3, 4, 5) (21, 16, 11)

(15) ( ) (18, 13, 8) (7) (6)

Теперь вершину P_{18} надо соединить с вершиной P_{20}, и из последовательностей надо вычеркнуть эти числа. Так как 20 больше не встречается среди u, запишем его в последовательность z на соответствующее место:

(2, 1, 2) ( ) (10, 2) (5) (5, 3, 4, 5) (21, 16, 11)

(20, 15) ( ) (13, 8) (7) (6)

Теперь вершину P_{20} надо соединить с вершиной P_2 и следует вычеркнуть стоящие на первом месте в первой паре скобок 2 и 20. Среди оставшихся u еще фигурирует 2, поэтому последовательность z не изменится:

(1, 2) ( ) (10, 2) (5) (5, 3, 4, 5) (21, 16, 11)

(15) ( ) (13, 8) (7) (6)

Теперь соединяются вершины P_{15} и P_1; среди u вычеркивается 1, среди z — 15. Среди u число 1 больше не встречается, поэтому 1 вписывается на соответствующее место среди z:

(2) ( ) (10, 2) (5) (5, 3, 4, 5) (21, 16, 11)

( ) ( ) (13, 8) (7) (6, 1)

Теперь соединяются вершины P_{13} и P_{10}; 10 вычеркивается среди u, а 13 — среди z. 10 следует вписать в последовательность z, после чего новые последовательности таковы:

(2) ( ) (2) (5) (5, 3, 4, 5) (21, 16, 11)

(10) ( ) (8) (7) (6, 1)

Соединим вершины P_{10} и P_2, и вычеркнем из последовательностей эти числа. 2 еще встречается среди u, поэтому последовательность z не меняется:

( ) ( ) (2) (5) (5, 3, 4, 5) (21, 16, 11)

( ) ( ) (8) (7) (6, 1)

Теперь следует соединить вершины P_8 и P_2 и вычеркнуть эти числа из последовательностей. Так как 2 больше не встречается среди u, его надо вписать в последовательность z:

( ) ( ) ( ) (5) (5, 3, 4, 5) (21, 16, 11)

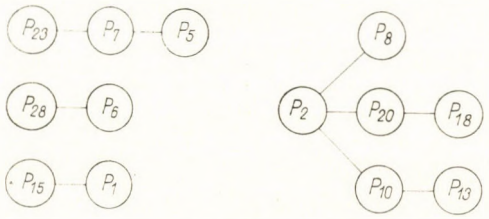
( ) ( ) ( ) (7, 2) (6, 1)

Соединим ребром вершины P_7 и P_5, исключим из последовательностей 5 и 7. Так как 5 еще встречается среди u, последовательность z не меняется:

( ) ( ) ( ) ( ) (5, 3, 4, 5) (21, 16, 11)

( ) ( ) ( ) (2) (6, 1)

До сих пор построен следующий граф:



Теперь надо соединить ребром вершины P_6 и P_5 и вычеркнуть 5 и 6 на первом месте пятых пар скобок. 5 еще встречается среди u, поэтому имеем:

( ) ( ) ( ) ( ) (3, 4, 5) (21, 16, 11)

( ) ( ) ( ) (2) (1)

Теперь найбольшее z равно 2, поэтому вершину P_2 надо было бы соединить с P_u, где u число из четвертой пары скобок, но так как там нет u, то вершина P_2 соединяется с вершиной P_{21}, так как в последней паре скобок на первом

месте стоит 21. Теперь следует вычеркнуть 21 и 2 и, так как 21 больше не встречается среди u, то его надо вписать на соответствующее место среди z:

$$( ) ( ) ( ) ( ) (3, 4, 5) (16, 11)$$
$$( ) ( ) ( ) ( ) (21, 1)$$

Соединим ребром вершины P_{21} и P_3, и исключим из последовательностей 3 и 21. Среди u число 3 больше не встречается, поэтому его надо вписать в последовательность z:

$$( ) ( ) ( ) ( ) (4, 5) (16, 11)$$
$$( ) ( ) (3) ( ) (1)$$

Соединим ребром вершины P_3 и P_{16} (так как в последовательности u в третьей паре скобок нет чисел), исключим эти числа из последовательностей и впишем 16 в последовательность z:

$$( ) ( ) ( ) ( ) (4, 5) (11)$$
$$( ) ( ) ( ) ( ) (16, 1)$$

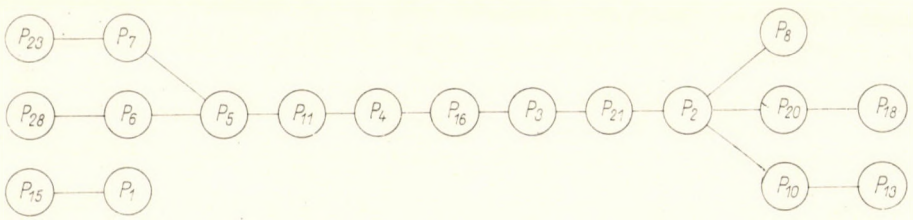
Теперь надо соединить вершины P_{16} и P_4, и исключить эти числа из последовательностей. 4 вписывается в последовательность z:

$$( ) ( ) ( ) ( ) (5) (11)$$
$$( ) (4) ( ) ( ) (1)$$

Теперь следует соединить вершины P_4 и P_{11}, исключить эти числа из последовательностей и вписать 11 в последовательность z:

$$( ) ( ) ( ) ( ) (5) ( )$$
$$( ) ( ) ( ) ( ) (11, 1)$$

Соединим ребром вершины P_{11} и P_5, исключим эти числа из последовательностей и впишем 5 в последовательность z. До сих пор был построен следующий граф:



Наконец, если соединить ребром вершины P_5 и P_1, соответствующие двум оставшимся z, мы получим дерево.

Исходной последовательности единственным образом соответствует это дерево.

БИБЛИОГРАФИЯ

[1] KÖNIG, D.: *Theorie der endlichen und unendlichen Graphen*, Akad. Verlagsgesellschaft, Leipzig, 1936.
[2] ORE, O.: *Theory of graphs*, Amer. Math. Soc. Coll. Publ., Vol. **38** 1962.
[3] BERGE, C.: *Théorie des graphes et ses applications*, Dunod. Paris, 1958.
Берж, К.: *Теория графов и ее применения*, Изд. Иностранной Литературы, М. 1962.
[4] RÉNYI, A.: Új módszerek és eredmények a kombinatorikus analízisben I. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **16** (1966) 77—105.
[5] CAYLEY, A.: *Collected papers*, Cambridge, 1897, Vol. 13. pp. 26—28.
[6] PRÜFER, A.: Neuer Beweis eines Satzes über Permutationen, *Archiv für Math. u. Phys.* **27** (1918) 142—144.
[7] SCOINS, H. I.: The number of trees with nodes of alternate parity *Proc. Cambridge Phylos. Soc.* **58** (1962) 12—16.

Университет имени Л. Этвеша, Будапешт

(Поступила 15-ого ноября 1966 г.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 81—91.

О НЕКОТОРЫХ СВОЙСТВАХ ГИПЕРГЕОМЕТРИЧЕСКИХ ФУНКЦИЙ ГУМБЕРТА

М. Б. КАПИЛЕВИЧ

§ 1. Теоремы сложения для функции Гумберта $\Phi_3(\beta, \gamma; x, y)$

Рассмотрим гипергеометрический ряд Гумберта [1]:

$$\Phi_3(\beta, \gamma; x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\beta)_m}{(\gamma)_{m+n} m!\, n!} x^m y^n, \tag{1.1}$$

сходящийся во всех точках плоскости (x, y), и будем полагать, что все встречающиеся ниже величины являются вещественными числами. В теории функции $\Phi_3(\beta, \gamma; x, y)$ важную роль играют теоремы сложения, характеризующие её зависимость от параметров (β, γ) и аргументов (x, y). Остановимся на одной такой теореме. Из (1.1) следует

$$\Phi_3(\beta, \gamma; x, y) = \sum_{m=0}^{\infty} \frac{(\beta)_m x^m}{(\gamma)_m m!} \bar{I}_{\gamma+m-1}(2\sqrt{y}), \tag{1.2}$$

где $\bar{I}_\nu(2\sqrt{z}) = {}_0F_1(\nu+1; z) = \Gamma(\nu+1) z^{-\nu/2} I_\nu(2\sqrt{z})$ — функция Бесселя—Клиффорда [1]. Воспользуемся вторым определенным интегралом Сонина, записав его в виде

$$\bar{I}_{\gamma_2}(2\sqrt{y_2}) = c_1 \int_0^{\frac{\pi}{2}} \bar{I}_{\gamma_1}(2\sqrt{y_1}\cos\varphi) \bar{I}_{\gamma_2-\gamma_1-1}(2\sqrt{y_2-y_1}\sin\varphi) \cdot$$

$$\cdot \cos^{2\gamma_1+1}\varphi \sin^{2(\gamma_2-\gamma_1)-1}\varphi d\varphi, \tag{1.3}$$

где $\gamma_2 > \gamma_1 > -1$, $c_1\Gamma(\gamma_1+1)\Gamma(\gamma_2-\gamma_1) = 2\Gamma(\gamma_2+1)$. Заменим в (1.3) γ_1 и γ_2 на γ_1+m-1 и γ_2+m-1, а затем подставим результат в (1.2), положив там $\gamma=\gamma_2$, $y=y_2$. Тогда получим

$$\Phi_3(\beta, \gamma_2; x, y_2) = c_2 \int_0^{\frac{\pi}{2}} \Phi_3(\beta, \gamma_1; x\cos^2\varphi, y_1\cos^2\varphi) \cdot$$

$$\cdot \bar{I}_{\gamma_2-\gamma_1-1}(2\sqrt{y_2-y_1}\sin\varphi) \cos^{2\gamma_1-1}\varphi \cdot \sin^{2(\gamma_2-\gamma_1)-1}\varphi d\varphi, \tag{1.4}$$

если $\gamma_2 > \gamma_1 > -1$, $c_2\gamma_2 = c_1\gamma_1$. Так как $\Phi_3(\beta, \gamma; x, 0) = {}_1F_1(\beta, \gamma; x)$, то при

$y_1=0$, $y_2=y$ (1. 4) дает интегральное представление функции Φ_3:

$$\Phi_3(\beta,\gamma_2;x,y)=c_2\int_0^{\frac{\pi}{2}}{}_1F_1(\beta,\gamma_1;x\cos^2\varphi)\bar{I}_{\gamma_2-\gamma_1-1}(2\sqrt{y}\sin\varphi)\cdot$$

$$\cdot\cos^{2\gamma_1-1}\varphi\cdot\sin^{2(\gamma_2-\gamma_1)-1}\varphi d\varphi, \tag{1.5}$$

которое упрощается, когда $\gamma_1=\beta$, ибо ${}_1F_1(\beta,\beta;z)=e^z$. Наоборот, принимая в (1. 4) $y_2=0$, $y_1=y$, найдем решение интегрального уравнения (1. 5) относительно ${}_1F_1(\beta,\gamma_1;x)$:

$${}_1F_1(\beta,\gamma_2;x)=c_2\int_0^{\frac{\pi}{2}}\Phi_3(\beta,\gamma_1;x\cos^2\varphi,y\cos^2\varphi)\bar{J}_{\gamma_2-\gamma_1-1}(2\sqrt{y}\sin\varphi)\cdot$$

$$\cdot\cos^{2\gamma_1-1}\varphi\sin^{2(\gamma_2-\gamma_1)-1}\varphi d\varphi. \tag{1.6}$$

Формула (1. 6) дает возможность связать $\Phi_3(x,y)$ с функциями Гумберта $\Phi_2(x,y)$, $\Psi_1(x,y)$, $\Xi_2(x,y)$, если в известных интегральных представлениях этих функций заменить ${}_1F_1$ выражением (1. 6). Подставим в (1. 4) степенной ряд $\bar{I}_\nu(2\sqrt{z})={}_0F_1(\nu+1;z)$, после чего используем снова (1. 4) при $y_2=y_1$, тогда придем к теореме сложения по аргументу y:

$$\Phi_3(\beta,\gamma;x,y_2)=\sum_{n=0}^{\infty}\frac{(y_2-y_1)^n}{n!\,(\gamma)_n}\Phi_3(\beta,\gamma+n;x,y_1), \tag{1.7}$$

из которой в случае $y_2=0$ возникает разложение интеграла (1. 6):

$${}_1F_1(\beta,\gamma;x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\,(\gamma)_n}y^n\Phi_3(\beta,y+n;x,y). \tag{1.8}$$

Наоборот, когда $y_1=0$, $y_2=y$, (1. 7) дает обращение ряда (1. 8):

$$\Phi_3(\beta,\gamma;x,y)=\sum_{n=0}^{\infty}\frac{y^n}{n!(\gamma)_n}{}_1F_1(\beta,\gamma+n;x). \tag{1.9}$$

§ 2. Несобственные интегралы с функциями Бесселя и Куммера

К рядам Гумберта $\Phi_1(x,y)$, $\Phi_3(x,y)$ и $\Xi_2(x,y)$ можно свести группу несобственных интегралов, содержащих функции Бесселя, Куммера и Макдональда. Приведем несколько примеров подобных интегралов.

1. Обозначим через $U(\alpha,\beta,\nu)$ выражение ($x>0$, $s>0$):

$$U(\alpha,\beta,\nu)=\int_0^{\infty}\xi^{2\beta-1}{}_1F_1(\alpha,\beta;-x\xi^2)\bar{J}_\nu(s\sqrt{b^2+\xi^2})d\xi. \tag{2.1}$$

Из асимптотических оценок $\bar{J}_\nu(z)=O(z^{-\nu-\frac{1}{2}})$ при $z\to\infty$,

$${}_1F_1(\alpha,\beta;z)=\frac{\Gamma(\beta)}{\Gamma(\beta-\alpha)}(-z)^{-\alpha}\left[1+o\left(\frac{1}{|z|}\right)\right],\ \text{когда}\ z\to-\infty, \tag{2.2}$$

следует, что (2.1) сходится, если $0<\beta<\alpha+\frac{\nu}{2}+\frac{3}{4}$. Разложим $\bar{J}_\nu(s\sqrt{b^2+\xi^2})$ в ряд Ломмеля [2]

$$\bar{J}_\nu(s\sqrt{b^2+\xi^2})=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\,(\nu+1)n}\left(\frac{bs}{2}\right)^{2n}\bar{J}_{\nu+n}(\xi s), \tag{2.3}$$

и вычислим возникающие обобщенные интегралы Ханкеля. В итоге, сравнивая результат с (1.9), получим:

$$U=f(\nu)\frac{s^{2(\alpha-\beta)}}{x^\alpha}\,\Phi_3\left(\alpha,\alpha-\beta+\nu+1;\,-\frac{s^2}{4x},\,-\frac{b^2s^2}{4}\right), \tag{2.4}$$

где $\Gamma(\alpha-\beta+\nu+1)f(\nu)=2^{2(\beta-\alpha)-1}\Gamma(\beta)\Gamma(\nu+1)$ С помощью (2.1), (2.4) можно исследовать более общий интеграл $U_m(\alpha,\beta,\nu)$ $(m=0,1,2,\ldots)$ вида

$$U_m=\int_0^\infty \xi^{2\beta-1}{}_1F_1(\alpha,\beta;-x\xi^2)\,(b^2+\xi^2)^m\bar{J}_{\gamma+m}(s\sqrt{b^2+\xi^2})\cdot \tag{2.5}$$

$$\cdot\,\bar{J}_{\nu+m}(\sigma\sqrt{b^2+\xi^2})\,d\xi,$$

если $x>0$, $s>0$, $0<\beta<\alpha+\nu+1$, $\nu>-\frac{1}{2}$. Действительно, заменим в (2.1) и (2.4) s на $\omega=\sqrt{s^2+\sigma^2-2s\sigma\cos\varphi}$, а затем, умножив на $C_m^\nu(\cos\varphi)\sin^{2\nu}\varphi d\varphi$, проинтегрируем результат по φ от $\varphi=0$ до $\varphi=\pi$. Используем, кроме того, формулу Гегенбауэра [см [2], стр. 400]:

$$\int_0^\pi \bar{G}_\nu(z\omega)C_m^\nu(\cos\varphi)\sin^{2\nu}\varphi d\varphi=D(s\sigma z^2)^m\bar{G}_{\nu+m}(sz)\bar{J}_{\nu+m}(\sigma z), \tag{2.6}$$

считая в ней $\nu>-\frac{1}{2}$, $z=\sqrt{b^2+\xi^2}$, $\bar{G}_\nu(z)=\bar{J}_\nu(z)$, $m!\,2^{2(\nu+m)-1}\Gamma^2(\nu+m+1)D=\pi\nu\Gamma(2\nu+m)$. Тогда при $DD_m=f(\nu)$ придем к выражению:

$$U_m=\frac{D_m}{x^\alpha(s\sigma)^m}\int_0^\pi \Phi_3\left(\alpha,\alpha-\beta+\nu+1;-\frac{\omega^2}{4x},-\frac{b^2\omega^2}{4}\right)\omega^{2(\alpha-\beta)}\cdot \tag{2.7}$$

$$\cdot\,C_m^\nu(\cos\varphi)\sin^{2\nu}\varphi d\varphi.$$

2. Рассмотрим теперь несобственный интеграл $V(\alpha,\beta,\mu,\nu)$

$$V=\int_0^\infty \xi^{2\beta-1}{}_1F_1(\alpha,\beta;-x\xi^2)\bar{J}_\mu(s\sqrt{b^2+\xi^2})\bar{J}_\nu(s\sqrt{b^2+\xi^2})\,d\xi, \tag{2.8}$$

содержащий произведение двух функций Бесселя с различными индексами, но одинаковыми аргументами. Если $x>0$, $s>0$, то (2.8) в силу (2.2) сходится при условии $0<\beta<\frac{\mu+\nu}{2}+\alpha+1$. Заменим в (2.1), (2.4) ν и s на $\mu+\nu$ и $2s\cos\varphi$, умножим результат на $\cos^{\mu+\nu}\varphi\cos(\mu-\nu)\varphi\,d\varphi$, а затем проинтегрируем по φ

от $\varphi=0$ до $\varphi=\frac{\pi}{2}$. Вычисляя появляющийся при этом интеграл по формуле Неймана [2]:

$$\bar{J}_\mu(z)\bar{J}_\nu(z) = \gamma_1 \int_0^{\frac{\pi}{2}} \bar{J}_{\mu+\nu}(2z\cos\varphi)\cos^{\mu+\nu}\varphi\cos(\mu-\nu)\varphi d\varphi, \tag{2.9}$$

$$\mu+\nu > -1, \quad \pi\Gamma(\mu+\nu+1)\gamma_1 = 2^{\mu+\nu+1}\Gamma(\mu+1)\Gamma(\nu+1),$$

получим для (2. 8) значение

$$V = \gamma_2 \frac{s^{2(\alpha-\beta)}}{x^\alpha} \int_0^{\frac{\pi}{2}} \Phi_3\left(\alpha, \alpha-\beta+\mu+\nu+1; -\frac{s^2}{x}\cos^2\varphi, -b^2 s^2\cos^2\varphi\right) \cdot$$

$$\cdot \cos^{2(\alpha-\beta)+\mu+\nu}\varphi\cos(\mu-\nu)\varphi d\varphi, \tag{2.10}$$

где $\gamma_2 = 2^{2(\alpha-\beta)}\gamma_1 f(\mu+\nu)$, $2(\alpha-\beta)+\mu+\nu > -1$. В частности, когда $\mu = \nu+n$, $(n=0, 1, 2, \ldots)$, (2. 10) можно записать в другой форме

$$V = \gamma_3 \frac{s^{2(\alpha-\beta)}}{x^\alpha} \int_0^{\frac{\pi}{2}} \Phi_3\left(\alpha, \alpha-\beta+\nu+n+1; -\frac{s^2}{x}\cos^2\varphi, -b^2 s^2\cos^2\varphi\right) \cdot$$

$$\cdot C_n^\nu(\cos\varphi)\cos^{2(\alpha-\beta+\nu)+n}\varphi\sin^{2\nu}\varphi\, d\varphi. \tag{2.11}$$

Здесь $0<\beta<\alpha+\nu+\frac{n+1}{2}$, $\nu > -\frac{1}{2}$, а γ_3 имеет вид:

$$\gamma_3 = \frac{n!\,2^{2\nu+n}\Gamma(\beta)\Gamma(\nu+1)\Gamma(\nu+n+1)}{\sqrt{\pi}\,(2\nu)_n\Gamma(\nu+\frac{1}{2})\Gamma(\alpha-\beta+\nu+n+1)}.$$

Особого внимания заслуживают случаи, когда ряд Куммера ${}_1F_1(\alpha, \beta; -x\xi^2)$ в формулах (2. 1), (2. 5), (2. 8) вырождается в более простые трансцендентные функции. Так, при $\alpha=\beta$ будет ${}_1F_1(\beta, \beta; -x\xi^2) = e^{-x\xi^2}$, и здесь (2.1), (2. 4) принимают вид:

$$U(\beta, \beta, \nu) = \int_0^\infty \xi^{2\beta-1} e^{-x\xi^2}\bar{J}_\nu\left(s\sqrt{b^2+\xi^2}\right)d\xi =$$

$$= \frac{\Gamma(\beta)}{2x^\beta}\Phi_3\left(\beta, \nu+1; -\frac{s^2}{4x}, -\frac{b^2 s^2}{4}\right), \quad (\beta>0). \tag{2.12}$$

В случае $\alpha=\beta$, $b=0$, $m=0$, когда $c_0^\nu(\cos\varphi)=1$, (2. 5), (2. 7) сводятся к известному обобщению второго экспоненциального интеграла Вебера [см. [2], стр. 432], а (2. 8) и (2.10) дают

$$V = \frac{\Gamma(\beta)}{2x^\beta}\,{}_3F_3\left(\frac{\mu+\nu+1}{2}, \frac{\mu+\nu+2}{2}, \beta; \mu+1, \nu+1, \mu+\nu+1; -\frac{s^2}{x}\right).$$

При $\alpha=\frac{\beta}{2}$ формулы (2. 1), (2. 5) и (2. 8) будут содержать выражение

$$(2.13a)\qquad {}_1F_1\left(\frac{\beta}{2}, \beta; -x\xi^2\right) = e^{-\frac{x\xi^2}{2}} \bar{I}_{\frac{\beta-1}{2}}\left(\frac{x\xi^2}{2}\right);$$

если $\alpha = \beta-1$, то там появится неполная гамма-функция Эйлера:

$$(2.13b)\qquad {}_1F_1(\beta-1, \beta; -x\xi^2) = \Gamma(\beta)\gamma^*(\beta-1; x\xi^2),$$

а при $\alpha = \beta+n$ $(n=0, 1, 2, \ldots)$ — полиномы Лагерра:

$$(2.13c)\qquad {}_1F_1(\beta+n, \beta; -x\xi^2) = \frac{n!}{(\beta)_n} e^{-x\xi^2} L_n^{\beta-1}(x\xi^2).$$

3. К функции Гумберта $\Phi_1(x, y)$ приводит интеграл вида:

$$(2.14)\qquad \bar{V}(\beta, \mu, \nu) = \int_0^\infty \xi^{2\beta-1} e^{-x\xi^2} {}_1F_1\left[\mu, \nu+1; -\frac{s^2}{4}(b^2+\xi^2)\right] d\xi,$$

где $x>0$, $\beta>0$. Действительно, по теореме сложения для ${}_1F_1$ имеем:

$$(2.15)\qquad {}_1F_1\left[\mu, \nu+1; -\frac{s^2}{4}(b^2+\xi^2)\right] = \sum_{n=0}^\infty \frac{(\mu)_\mu(-s^2\xi^2)^n}{n!2^{2n}(\nu+1)_n} {}_1F_1\left(\mu+n, \nu+n+1; -\frac{b^2s^2}{4}\right).$$

Подставим (2.15) в (2.14) и сопоставим результат с рядом

$$\Phi_1(\alpha, \beta, \gamma; x, y) = \sum_{n=0}^\infty \frac{(\alpha)_n(\beta)_n}{(\gamma)_n n!} x^n {}_1F_1(\alpha+n, \gamma+n; y).$$

Тогда получим

$$(2.16)\qquad \bar{V}(\beta, \mu, \nu) = \frac{\Gamma(\beta)}{2x^\beta} \Phi_1\left(\mu, \beta, \nu+1; -\frac{s^2}{4x}, -\frac{b^2s^2}{4}\right).$$

Нетрудно убедиться, что (2.12) является конфлюэнтным случаем формул (2.14), (2.16). Действительно, заменим в (2.14) и (2.16) s на $\frac{s}{\sqrt{\mu}}$ и перейдем к пределу при $\mu\to\infty$. В итоге, учитывая равенства

$$\lim_{\mu\to\infty} {}_1F_1\left(\mu, \nu+1; -\frac{z}{\mu}\right) = \bar{J}_\nu(2\sqrt{z}),$$

$$\lim_{\mu\to\infty} \Phi_1\left(\mu, \beta, \nu+1; \frac{X}{\mu}, \frac{Y}{\mu}\right) = \Phi_3(\beta, \nu+1; X, Y),$$

придем к формуле (2.12). Для (2.14) следует также особо отметить случаи вырождения (2.13).

4. Рассмотрим, наконец, пример интеграла, при изучении которого приходится привлечь ряд Гумберта $\Xi_2(x, y)$. А именно, обозначим через $W_m(\beta, \mu, \nu)$ $(m=0, 1, 2, \ldots)$ выражение:

$$W_m = \int_0^\infty \xi^{\beta-1}(b^2+\xi^2)^m \bar{I}_{\nu+m}(s\sqrt{b^2+\xi^2})\bar{J}_{\nu+m}(\sigma\sqrt{b^2+\xi^2})K_\mu(x\xi)\,d\xi, \tag{2.17}$$

где $x>s\geqq 0$, $\beta\pm\mu>0$. Исходя из (2. 3) и формулы (2. 6), взятой при $\bar{G}_\nu(z)=\bar{I}_\nu(z)$, получим

$$W_m = \frac{\varkappa_m}{x^\beta(s\sigma)^m}\int_0^\pi \Xi_2\left(\frac{\beta+\mu}{2}, \frac{\beta-\mu}{2}, \nu+1; \frac{\omega^2}{x^2}, \frac{b^2\omega^2}{4}\right) C_m^\nu(\cos\varphi)\sin^{2\nu}\varphi\,d\varphi. \tag{2.18}$$

Здесь $\nu>-\frac{1}{2}$, $\varkappa_m D=\gamma=2^{\beta-2}\Gamma\left(\frac{\beta+\mu}{2}\right)\Gamma\left(\frac{\beta-\mu}{2}\right)$, а ω и D имеют те же значения, что в (2. 6). Выражение (2.18) дает более простой результат, когда одна из величин s или σ равна нулю. Так, при $m=0$, $\sigma=0$ из (2.18) следует

$$\begin{aligned}\bar{W}_0(\beta, \mu, \nu) &= \int_0^\infty \xi^{\beta-1}\bar{I}_\nu(s\sqrt{b^2+\xi^2})K_\mu(x\xi)\,d\xi = \\ &= \frac{\gamma}{x^\beta}\Xi_2\left(\frac{\beta+\mu}{2}, \frac{\beta-\mu}{2}, \nu+1; \frac{s^2}{x^2}, \frac{b^2s^2}{4}\right).\end{aligned} \tag{2.19}$$

Если же $\sigma=0$, а $m=1, 2, 3, \ldots$, то в силу равенства $\int_0^\pi C_m^\nu(\cos\varphi)\sin^{2\nu}\varphi\,d\varphi=0$ будет $W_m=0$. В свою очередь (2.19) можно подвергнуть тем же операциям, которые выше были применены к (2.1), (2. 4) для получения формул (2.10), (2.11), причем с этой целью наряду с (2. 9) целесообразно использовать также интеграл Никольсона $(\mu<m+1,\ m=0, 1, 2, \ldots)$ [2]:

$$I_m(z)K_\mu(z) = \frac{2(-1)^m}{\pi}\int_0^{\frac{\pi}{2}} K_{\mu-m}(2z\cos\varphi)\cos(\mu+m)\varphi\,d\varphi.$$

§ 3. Примеры применения результатов § 1 и § 2.

Функция Гумберта Φ_3 играет важную роль при исследовании решений $z(x, s; a, b)$ и $\bar{z}(x, s; a, b)$ двух сингулярных задач Коши:

$$z(x, 0) = \tau(x), \quad z_s(x, 0) = 0, \tag{3.1}$$

$$\bar{z}(x, 0) = 0, \qquad \bar{z}_\eta(x, 0) = \nu(x), \quad \eta = -\left(\frac{s}{1-a}\right)^{1-a} \tag{3.2}$$

для уравнения параболического типа

$$z_x = z_{ss} + \frac{a}{s} z_s + b^2 z, \quad (a > 0,\ b = \text{const},\ s \geqq 0). \tag{3.3}$$

Покажем, что наиболее простому случаю $\tau(x) = x^\alpha$ и $\nu(x) = x^\alpha$ ($\alpha = \text{const} > 0$) отвечают обобщенные автомодельные интегралы уравнения (3. 3):

$$z^{(\alpha)}(x, s; a, b) = x^\alpha \Phi_3\left(-\alpha, \beta + \frac{1}{2}\,;\ -\frac{s^2}{4x}, -\frac{b^2 s^2}{4}\right), \tag{3.4a}$$

$$\bar{z}^{(\alpha)}(x, s; a, b) = x^\alpha \eta \Phi_3\left(-\alpha, \frac{3}{2} - \beta;\ -\frac{s^2}{4x}, -\frac{b^2 s^2}{4}\right). \tag{3.4b}$$

Для этого разделим в (3. 3) переменные, полагая $z(x, s) = T(x)R(s)$. Тогда получим $T' + \lambda^2 T = 0$, $R'' + \frac{a}{s} R' + (b^2 + \lambda^2) R = 0$, где $\lambda = \text{const}$. Этим уравнениям удовлетворяют функции ($2\beta = a$):

$$T_1 = e^{-\lambda^2 x}, \quad R_1 = \bar{J}_{\beta - \frac{1}{2}}\left(s\sqrt{b^2 + \lambda^2}\right), \quad R_2 = \eta \bar{J}_{\frac{1}{2} - \beta}\left(s\sqrt{b^2 + \lambda^2}\right). \tag{3.5}$$

Составим произведение $z = T_1 R_1$, а затем на его основе — интеграл

$$z^{(\alpha)} = A \int_0^\infty \lambda^{-2\alpha - 1} \bar{J}_{\beta - \frac{1}{2}}\left(s\sqrt{b^2 + \lambda^2}\right) e^{-\lambda^2 x} d\lambda, \qquad (A = \text{const}), \tag{3.6}$$

считая сначала, что $\alpha < 0$. Сопоставив (3. 6) с (2.12), придем при $A\Gamma(-\alpha) = 2$ к выражению (3. 4a). Однако непосредственной проверкой можно убедиться, что ряд (3. 4a) является решением уравнения (3. 3) не только когда $\alpha < 0$, но и при любых значениях $-\infty < \alpha < \infty$, причем $z^{(\alpha)}(x, 0) = x^\alpha$, $z_s^{(\alpha)}(x, 0) = 0$. Поступая аналогично с произведением $\bar{z} = T_1 R_2$, придем к решению (3. 4b) проблемы (3. 2) при $\nu(x) = x^\alpha$. Таким образом, если рассматривать для (3. 3) задачи Коши (3.1), (3. 2) с начальными данными, заданными в форме степенных рядов

$$\tau(x) = \sum_{m=0}^\infty A_m x^{m+\alpha}, \quad \nu(x) = \sum_{m=0}^\infty \bar{A}_m x^{m+\alpha}, \qquad (\alpha = \text{const} > 0), \tag{3.7}$$

то их решения определятся разложениями по функциям (3. 4) вида:

$$z(x, s; a, b) = \sum_{m=0}^\infty A_m z^{(m+\alpha)}(x, s; a, b), \quad \bar{z}(x, s; a, b) = \sum_{m=0}^\infty \bar{A}_m \bar{z}^{(m+\alpha)}(x, s; a, b). \tag{3.8}$$

Опираясь на (3. 8) и функциональные соотношения для ряда Гумберта $\Phi_3(x, y)$, можно найти соответствующие свойства функций (3. 4), (3. 8). Так, например, с помощью (1. 7) и (3. 4a) обнаруживаем теорему сложения по параметру b:

$$z^{(\alpha)}(x, s; a, b_2) = \sum_{n=0}^\infty g_n(\beta) s^{2n} (b_2^2 - b_1^2)^n z^{(\alpha)}(x, s; a + 2n, b_1), \tag{3.9}$$

где $n!\,(\beta+\frac{1}{2})_n\,2^{2n}g_n(\beta)=(-1)^n$. Воспользуемся далее известными формулами Берчнелла—Ченди [3]:

$$\Phi_3(\beta,\gamma;X,Y)=\sum_{n=0}^{\infty}\frac{(-1)^n(\beta)_nX^nY^n}{n!\,(\gamma+n-1)_n(\gamma)_{2n}}\,\bar{I}_{\gamma+2n-1}(2\sqrt{Y})\,{}_1F_1(\beta+n,\gamma+2n;X),$$

$$\bar{I}_{\gamma-1}(2\sqrt{Y})\,{}_1F_1(\beta,\gamma;X)=\sum_{n=0}^{\infty}\frac{(\beta)_nX^nY^n}{n!\,(\gamma)_n(\gamma)_{2n}}\,\Phi_3(\beta+n,\gamma+2n;X,Y),$$

и, имея в виду (3. 4a), положим в них $\beta=-\alpha$, $\gamma=\beta+\frac{1}{2}$, $X=-\frac{s^2}{4x}$, $Y=-\frac{b^2s^2}{4}$. Примем также во внимание значения

$$D_x^n z^{(\alpha)}(x,s;a,b)=(-1)^n(-\alpha)_n\,x^{\alpha-n}\Phi_3\left(n-\alpha,\beta+\frac{1}{2};-\frac{s^2}{4x},-\frac{b^2s^2}{4}\right),$$

$$D_x^n z^{(\alpha)}(x,s;a,0)=(-1)^n(-\alpha)_n x^{\alpha-n}\,{}_1F_1\left(n-\alpha,\beta+\frac{1}{2};-\frac{s^2}{4x}\right),$$

где $D_x=\dfrac{\partial}{\partial x}$, $n=0,1,2,\ldots$. Тогда придем к теореме сложения

$$z^{(\alpha)}(x,s;a,b)=\sum_{n=0}^{\infty}c_n(bs^2)^{2n}\bar{J}_{\nu+2n}(bs)D_x^n z^{(\alpha)}(x,s;a+4n,0), \tag{3.10}$$

а также к формуле, обращающей (3.10):

$$\bar{J}_\nu(bs)z^{(\alpha)}(x,s;a,0)=\sum_{n=0}^{\infty}\bar{c}_n(bs^2)^{2n}D_x^n z^{(\alpha)}(x,s;a+4n,b), \tag{3.11}$$

$$n!\,2^{4n}(\nu+n)_n(\nu+1)_{2n}c_n=1,\quad(\nu+1)_{2n}2^{2n}\bar{c}_n=g_n(\beta),\quad \nu=\beta-\frac{1}{2}.$$

Так как коэффициенты при $z^{(\alpha)}$ и $D_x^n z^{(\alpha)}$ в разложениях (3. 9), (3. 10), (3. 11) не зависят от α, то подставляя эти ряды в первую из формул (3. 8), убеждаемся, что равенства (3. 9), (3.10) и (3.11) выполняются не только для $z^{(\alpha)}(x,s;a,b)$, но также для решения $z(x,s;a,b)$ более общей проблемы (3.1), (3. 3), (3. 7). Построим теперь интеграл уравнения (3. 3) $Z(x,s;a,b)$ с особенностью в произвольной точке (x_0,s_0) полуплоскости $s>0$. Для этого образуем из множителей (3. 5) выражение

$$Z(x,s;x_0,s_0)=A_0(s_0)^{1-a}\int_0^{\infty}\lambda^{2\mu-1}\bar{J}_\nu\left(s\sqrt{b^2+\lambda^2}\right)\bar{J}_\nu\left(s_0\sqrt{b^2+\lambda^2}\right)e^{-\lambda^2|x-x_0|}d\lambda, \tag{3.12}$$

где $\mu>0$, $\nu=\frac{1}{2}-\beta$, $A_0=\text{const}$. Функция (3.12) симметрична относительно (x,s) и (x_0,s_0) удовлетворяет по обеим этим парам переменных уравнению

(3. 3) и обладает особенностью при $(x, s)\to(x_0, s_0)$. Преобразуя (3. 12) по формуле (2. 5), (2. 7), найдем:

$$(3.13)\qquad Z=\frac{A_0 D_0}{|x-x_0|^{\mu}}(s_0)^{1-a}\int_0^{\pi}\Phi_3\left(\mu, \nu+1; -\frac{\omega^2}{4|x-x_0|}, -\frac{b^2\omega^2}{4}\right)\sin^{2\nu}\varphi\, d\varphi,$$

где $a<2$, $\omega=\sqrt{s^2+s_0^2-2ss_0\cos\varphi}$. Особый интерес имеют значения $\mu=\nu+1$, $A_0 2^{2\nu}\Gamma^2(\nu+1)=1$, где (3.12), (3.13) упрощаются, а функция $V=s_0^a Z$, удовлетворяющая по (x_0, s_0) сопряженному с (3. 3) уравнению, дает фундаментальное (элементарное) решение.

Рассмотрим, наконец, задачу Коши (3.1), (3. 7) для более общего уравнения параболического типа

$$(3.14)\qquad z_x=z_{ss}+\frac{a}{s}z_s+B(s)z, \qquad \left(a=\frac{1}{3}\right),$$

полагая, что $B(s)$ вблизи $s=0$ определяется рядом по степеням переменной η:

$$(3.15)\qquad B(s)=b_{-1}s^{-\frac{2}{3}}+b_0+b_1 s^{\frac{2}{3}}+b_2 s^{\frac{4}{3}}+\ldots+b_n s^{\frac{2n}{3}}+\ldots$$

К форме (3.14) приводятся диффузионные уравнения Колмогорова—Фоккера—Планка для непрерывных вероятностных процессов [4], [5]; кроме того (3.14) встречается при изучении тепловых явлений в движущихся средах с вынужденной конвекцией теплоты [6], [7], [8], в проблемах турбулентной теплопроводности [9], и т. д. Будем искать $z(x, s)$ в форме ряда

$$(3.16)\qquad z(x, s)=z_0(x, s)+z_1(x, s)+\ldots+z_n(x, s)+\ldots,$$

и заменим (3.14) рекуррентной системой

$$(3.17_1)\qquad Q[z_0]=\frac{\partial z_0}{\partial x}-\frac{\partial^2 z_0}{\partial s^2}-\frac{a}{s}\frac{\partial z_0}{\partial s}-b_0 z_0=0, \qquad \left(a=\frac{1}{3}\right),$$

$$(3.17_2)\qquad Q[z_1]=b_{-1}s^{-\frac{2}{3}}z_0,$$

$$(3.17_3)\qquad Q[z_2]=b_{-1}s^{-\frac{2}{3}}z_1+b_1 s^{\frac{2}{3}}z_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$(3.17_n)\qquad Q[z_n]=b_{-1}s^{-\frac{2}{3}}z_{n-1}+b_1 s^{\frac{2}{3}}z_{n-2}+b_2 s^{\frac{4}{3}}z_{n-3}+\ldots+b_{n-1}s^{\frac{2n-2}{3}}z_0.$$

Чтобы обеспечить выполнение условий (3.1), положим также:

$$(3.18)\qquad z_0(x, 0)=\tau(x), \qquad z_{n+1}(x, 0)=z_{ns}(x, 0)=0, \quad (n=0, 1, 2, \ldots).$$

Решение $z_0(x, s; a, b)$ проблемы (3.17_1), (3.1), (3. 7) дается первой из формул (3. 8) при $b^2=b_0$. Можно далее показать, что $z_n(x, s)$, $(n=1, 2, \ldots)$, выражаются в простой форме через смежные функции $z^{(n)}=z_0(x, s; a+2n, b)$ $(n=0, 1, 2, \ldots)$

и их производные $z_s^{(n)}$. С этой целью достаточно воспользоваться двумя свойствами оператора $Q[z]$:

(3.19a) $$Q[s^\mu z^{(n)}] = 2(n-\mu)s^{\mu-1}z_s^{(n)} - \mu(a+\mu-1)s^{\mu-2}z^{(n)},$$

(3.19b) $$Q[s^\mu z_s^{(n)}] = 2(n-\mu)s^{\mu-1}\Lambda_x[z^{(n)}] + (\mu-2n-1)(a+2n-\mu)s^{\mu-2}z_s^{(n)},$$

где $\Lambda_x = \dfrac{\partial}{\partial x} - b_0$, а также рекуррентными формулами Дарбу—Вайнштейна [10]:

(3.20a) $$z_{0s}(x, s; a+2, b) = \frac{a+1}{s}[z_0(x, s; a, b) - z_0(x, s; a+2, b)],$$

(3.20b) $$\Lambda_x[z_0(x, s; a+2, b)] = \frac{a+1}{s} z_{0s}(x, s; a, b).$$

Действительно, чтобы обратить уравнение (3.17_2), положим в (3.19a) $n=1$, $\mu = a+1$, а затем к правой части полученного равенства применим (3.20a). Это дает

(3.21) $$Q[s^{1+a}z_0(a+2)] = -2a(a+1)s^{a-1}z_0(a).$$

Сравнивая теперь (3.17_2) с (3.21), находим

3.22) $$z_1(x, s; a) = -\frac{b_{-1}}{2a(a+1)} s^{1+a} z_0(x, s; a+2, b),$$

причем (3.22) удовлетворяет одновременно и требованиям (3.18). Если далее подставить (3.22) в (3.17_3) и снова привлечь (3.19) и (3.20), то удается в эксплицитной форме вычислить $z_2(x, s; a)$, а затем шаг за шагом найти $z_3, z_4, \ldots$. Таким образом, в силу (3.22) при построении обобщенных автомодельных решений $z^{(\alpha)}$ уравнения (3.14) второе приближение z_1 имеет вид:

(3.23) $$z_1^{(\alpha)}(x, s; a) = -\frac{b_{-1}}{2a(a+1)} x^\alpha s^{1+a} \Phi_3\left(-\alpha, \beta+\frac{3}{2}; -\frac{s^2}{4x}, -\frac{b^2 s^2}{4}\right),$$

а в случае регулярных данных (3.7) находим

$$z_1(x, s; a) = \sum_{m=0}^{\infty} A_m z_1^{(\alpha+m)}(x, s; a).$$

При этом последующие итерации $z_2^{(\alpha)}, z_3^{(\alpha)}, \ldots$ также выражаются через смежные конфлюэнтные гипергеометрические функции Гумберта. Простой пример разложения (3.16) даёт теорема сложения (3.9), где цепочка неоднородных уравнений типа (3.17) решается в замкнутой форме членами ряда (3.9). Если обозначить через $M(a, b) > 0$ максимум модуля функции $z(x, s; a, b)$ в прямоугольнике $\Omega(x_1 \leqq x \leqq x_2,\ 0 \leqq s \leqq s_1)$, то так как $|z(x, s; a, b)|$ при фиксированных x, s, b убывает с ростом a, то для всех $n = 0, 1, 2, \ldots$

(3.24) $$|z(x, s; a+2n, b)| \leqq M(a, b), \quad \text{когда} \quad (x, s) \in \Omega.$$

Поэтому разложение (3. 9) мажорируется в области Ω абсолютно и равномерно сходящимся рядом

$$|z(x, s; a, b_2)| \leqq M_1 \sum_{n=0}^{\infty} |g_n(\beta)|\,|b_2^2 - b_1^2|^n s^{2n} = M_1 \bar{I}_{\beta-\frac{1}{2}}[s\sqrt{|b_2^2 - b_1^2|}],$$

где $M_1 = M(a, b_1)$. Аналогичные оценки имеют место и в случае (3.14), (3.16). Например из (3. 22) и (3. 24) следует $|z_1(x, s; a)| \leqq \frac{|b_{-1}|}{2a(a+1)} M s^{1+a}$, когда $(x, s) \in \Omega$. В заключение отметим, что проблема (3.1), (3.14), после надлежащего преобразования уравнения (3.14), может быть решена также в сходных с (3.10) базисных рядах вида $z(x, s) = \sum_{n=0}^{\infty} A_n(s) D_x^n z(x, s; a + k \cdot n, b)$, ($k = 2$, или $k = 4$), коэффициенты которых $A_n(s)$ однозначно определяются из рекуррентной последовательности обыкновенных дифференциальных уравнений второго порядка. Эта рекуррентная последовательность решается в квадратурах, а в случае (3. 10) — в эксплицитной форме $A_n(s) = c_n(bs^2)^{2n} \bar{J}_{\nu+2n}(bs)$.

БИБЛИОГРАФИЯ

[1] Erdélyi, Magnus, Oberhettinger, Tricomi, *Higher Transcendental Functions*, Vol. 1—2, New York, 1953.

[2] Ватсон, Г. Н.: *Теория Бесселевых функций*, Москва, 1949.

[3] Burchnall, J. L. and Chaundy, T. W.: Expansions of Appell's double hypergeometric functions (II), *The Quarterly journal of mathematics*, Oxford series, Vol. 12, No 46, (1941) 112—128.

[4] Feller, W.: Two singular diffusion problems, *Annals of Mathematics*, Ser. 2, Vol. 54, No. 1, (1951) 173—182.

[5] Хасьминский, Р. З.: Распределение вероятностей для функционалов от траектории случайного процесса диффузионного типа, *Докл. Акад. Наук СССР*, Том 104, № 1, (1955) 22—25.

[6] Sutton, W. G. L.: On the equation of diffusion in a turbulent medium, *Proceedings of the Royal Society of London*. Ser. A, Vol., 182. No. 988, (1943) 48—75.

[7] Левич, В. Г.: *Физико-химическая гидродинамика*, Москва—Ленинград, 1959.

[8] Рубинштейн, Л. И.: Об интегральной величине тепловых потерь при нагнетании горячей жидкости в пласт, *Известия Высших учебных заведений СССР, Нефть и газ*, № 9, (1959) 41—48.

[9] Баренблатт, Г. И. и Левитан, Б. М.: О некоторых краевых задачах для уравнения турбулентной теплопроводности, *Изв. Акад. Наук СССР Сер. Мат.* Том 16, № 3, (1952) 253—280.

[10] Weinstein, A.: On a Cauchy problem with subharmonical initial values, *Ann. Mat. Pura. Appl.* Ser. 4, Vol. 43, (1957) 325—340.

Москва

(Поступила 23-его ноября 1966 г.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 93—126.

LÖSUNG VON GLEICHUNGEN DURCH SCHRITTWEISE STÖRUNG

von

T. FREY

§ 1. Einleitung

Die NEWTON—RAPHSONsche Methode, bzw. die neuen verfeinerten Varianten dieses Verfahrens wurden neuerdings immer öfter in der numerischen Praxis angewendet, uzw. nicht nur beim Aufsuchen von Wurzeln nichtlinearer Gleichungssysteme, sondern auch bei der Lösung von allgemeineren Operatorengleichungen (s. z.B. [1], [2]). Es ist bekannt, daß alle diese Verfahren sehr schnell konvergieren, falls nur die erste Annäherung genug gut ist. In der Praxis haben wir aber eben damit die größten Schwierigkeiten. Eben deswegen hatte man beim Aufsuchen von Wurzeln nichtlinearer Gleichungen auch solche — scheinbar mit sehr vielen arithmetischen Operationen arbeitende — Hilfsmitteln gebraucht, wie numerische Integration von Differentialgleichungen (s. z.B. [3]). Auch in unserem Rechenzentrum müßten wir bei der Lösung von algebraischen Gleichungen sehr hohen ($24 \sim 48$-ten) Grades mit sehr nahe liegenden Wurzelngruppen ein dem KIZNERschen (s. [3]) ähnliches Verfahren anwenden. Der Grundgedanke war wie folgt: wir suchten eine Funktionenschar $\varphi(z, t)$ mit der Eigenschaft, daß ein Element, $\varphi(z, t_0)$, ebenso viele und von uns genau bekannten Wurzeln, $z_{i,0}$ $(i=1, 2, \ldots, k)$ besitzt, wie die zu lösende Gleichung $f(z)=0$, ferner ein anderes Element dieser Schar, $\varphi(z, t_1) \equiv f(z)$ sei. Ist nun φ genügend oft nach z und t differenzierbar, so sind die Wurzeln der Elemente der angegebenen Schar im allgemeinen stetig differenzierbare Funktionen des Parameters, und genügen der Anfangswertaufgabe

$$\frac{dz_i}{dt} \cdot \frac{\partial \varphi}{\partial z} + \frac{\partial \varphi}{\partial t} \equiv 0; \quad z_i(t_0) = z_{i,0}. \tag{1}$$

Wir haben nun diese Gleichung mit einer RUNGE—KUTTA Methode — mit „automatisch variierter Schrittweite" — von t_0 zu t_1 integriert, um die Wurzeln von $f(z)$ anzugeben. Da wir aber viele Wurzelngruppen mit sehr nahe liegenden Wurzeln hatten, müßten wir sehr komplizierte Formen für φ wählen, um mit 10^3—10^4 RUNGE—KUTTA-Schritten die Lösungen mit befriedigender Genauigkeit zu bekommen; kommt man nämlich an einer Integralkurve $z(t)$ sehr nahe zu einer Nullstelle von $\partial\varphi/\partial z$, so wird in diesem Bereich die Schrittweite sehr klein. Da man aber nur sehr schwer eine passende Schar im Falle nahe liegender Wurzeln der zu lösenden Gleichung angeben kann, die solche Integralkurven besitzt, welche keine Nullstelle von $\partial\varphi/\partial z$ stark annähern, müßten wir eine andere numerische Integrationsmethode ausarbeiten, die auch in der Nähe von Singularitäten der Differentialgleichung praktisch brauchbar ist. Die so ausgearbeitete Methode hatten wir dann auch im Falle allgemeinerer Problemenklassen der numerischen Mathematik ausprobiert und mit Erfolg angewendet. In § 2 werden wir eben deswegen die numeri-

schen Integrationsmethoden des verallgemeinerten Runge—Kutta Typs im Falle Differentialgleichungen in Banachräumen theoretisch untersuchen. In § 3 werden wir einige konkrete Formeln angeben, u. zw. auch solche, die in einer Umgebung einer singulären Stelle anwendbar sind. Endlich in § 4 werden wir einige konkrete, wichtige Anwendungen betrachten.

§ 2. Formeln vom Runge-Kutta Typ für Differentialgleichungen in Banachräumen

Betrachten wir das Anfangswertproblem

$$\dot{x} = \frac{dx}{dt} = f(x, t); \qquad x(t_0) = x_0, \tag{2}$$

wo t eine reelle Veränderliche ist, ferner B ein Banachraum, $f(x, t)$ aber eine Funktion, die $B \times R$ in B — u. zw. genügend oft nach t bzw x im FRECHETschen Sinne stetig differenzierbar — abbildet. Es ist nun gut bekannt (s. z.B. [1], [4]), daß man eine Funktion $x(t)$ — welche R in B, u. zw. genügend oft nach t differenzierbar abbildet — so angeben kann, daß (2) erfüllt sei. Die Frage ist die Anwendbarkeit, bzw. die Fehlerabschätzung der Formel vom RUNGE—KUTTA Typ. Zu diesem Zweck betrachtet man den formalen Differentialoperator,

$$D = \frac{\partial}{\partial t} + \frac{\partial}{\partial x}(\ )f(x, t). \tag{3}$$

den man an einer beliebigen, nach beiden ihren Variablen stetig FRECHET-differenzierbaren und $B \times R$ in B abbildenden Funktion $u(x, t)$ anwenden kann, u. zw. folgendermaßen:

$$Du = \frac{\partial}{\partial t} u + \frac{\partial}{\partial x}(u) \cdot f(x, t), \tag{4}$$

wo $\frac{\partial}{\partial x}(u)$ ein linearer Operator — nämlich die partielle FRECHET-Ableitung von u des Banachraumes der linearen Operatoren $(B \to B)$ ist, den man in (4) auf $f(x, t)$ anwendet. D selbst ist also auch ein Element des Raumes $(B \to B)$. Es ist nun bekannt, daß im betrachteten Fall die Kettenregel der Differentialrechnung gültig ist (s. z.B. [2], [4]), so daß (4) die Ableitung der Funktion $u[x(t), t]$ nach t längs einer Integralkurve $x(t)$ von (2) darstellt. Man definiert nun auch die formalen höheren Potenzen von D durch die folgende Rekursionsformel:

$$D^n = D^{n-1} \frac{\partial}{\partial t} + D^{n-1} \frac{\partial}{\partial x}(\ ) \cdot f(x, t) \qquad (n = 2, 3, \ldots) \tag{5}$$

wo D^{n-1} ein schon definiertes Element des Banachraumes $(B \to B)$, und somit nach

(5) auch $D^n \in (B \to B)$ ist. Wendet man nun (5) an sich selbst an, so ergibt sich

$$D^n = D^{n-2}\frac{\partial^2}{\partial t^2} + D^{n-2}\frac{\partial^2}{\partial x\,\partial t}(\ )f(x,t) + D^{n-2}\frac{\partial^2}{\partial t\,\partial x}(\ )f(x,t) + \tag{6}$$
$$+ D^{n-2}\left(\frac{\partial^2}{\partial x^2}(\ )f(x,t)\right)f(x,t),$$

wo also $\dfrac{\partial^2(\ )}{\partial x^2}$ ein Element des Banachraumes $(B \to (B \to B))$ ist.

Nimmt man nun an, daß die Funktion u, an welche man D^n anwendet, genügend oft nach beiden ihren Variablen stetig differenzierbar ist, so kann man — wie es gut bekannt ist — die Reihenfolge der Differenzierung nach t und x vertauschen; daneben kann man formal das letzte Glied von (6) in der Form $D^{n-2}\dfrac{\partial^2}{\partial x^2}(\ )f^2$ schreiben. Somit ist

$$D^n = D^{n-2}\frac{\partial^2}{\partial t^2} + 2D^{n-2}\frac{\partial^2}{\partial x\,\partial t}(\ )f + D^{n-2}\frac{\partial^2}{\partial x^2}(\ )f^2. \tag{6a}$$

Wendet man nun (5) nochmals an, so bekommt man allgemein

$$D^n = D^{n-k}\frac{\partial^k}{\partial t^k} + \binom{k}{1}D^{n-k}\frac{\partial^k}{\partial x\,\partial t^{k-1}}(\ )f + \binom{k}{2}D^{n-k}\frac{\partial^k}{\partial x^2\,\partial t^{k-2}}(\ )f^2 + \tag{7}$$
$$+ \ldots + \binom{k}{k}D^{n-k}\frac{\partial^k}{\partial x^k}(\ )f^k$$

für $k = 1, 2, \ldots, n$, das man durch vollständige Induktion nach (5) gleich erhält, falls die gut bekannten Eigenschaften der Binomialkoeffizienten, und die angegebenen Voraussetzungen über stetige Differenzierbarkeit von u, und die angegebene Verkürzung benützt werden. $\dfrac{\partial^k}{\partial x^r \partial t^{k-r}}$ ist hier ein Element des Banachraumes $[{}_1B \to ({}_2B \to ({}_3 \ldots \to ({}_rB \to B))\ldots)]$. Formal kann man also D^n als formale n-te Potenz von D aufschreiben.

D^n gehört nach (5) zu $(B \to B)$, folglich ist für beliebige u und v

$$D^n(u+v) = D^n u + D^n v \tag{8}$$

gültig. Es sei nun $P(x, t)$ eine nach beiden ihren Argumenten genügend oft stetig differenzierbare Funktion, welche $B \times R$ in $(B \to B)$ abbildet. DP sei dann folgendermaßen definiert:

$$DP = \frac{\partial}{\partial t}P + \frac{\partial}{\partial x}P(\ )\cdot f(x,t), \tag{9}$$

wo also $DP \in (B \to B)$. Wir zeigen nun durch vollständige Induktion, daß

SATZ 1.

$$D^n(Pu) = \binom{n}{0}D^n P(u) + \binom{n}{1}D^{n-1}P(Du) + \binom{n}{2}D^{n-2}P(D^2u) + \tag{10}$$
$$+ \ldots + \binom{n}{n}D^0 P\cdot(D^n u)$$

gültig ist, wo $D^0P \equiv P$ *bedeutet.*

(10) ist nämlich für $n=1$ gültig, da Produkt eines linearen Operators und eines Elements nach den bekannten Regeln der Analysis zu differenzieren sind, folglich

$$D(Pu) = \frac{\partial}{\partial t}(Pu) + \frac{\partial}{\partial x}[Pu]\cdot(f) =$$

$$= \left[\frac{\partial}{\partial t}P\right](u) + P\left(\frac{\partial}{\partial t}u\right) + \left[\frac{\partial}{\partial x}P\right](u)(f) + P\frac{\partial}{\partial x}u(f) = DP(u) + P(Du).$$

In den weiteren werden wir die gut bekannte Tatsache benützen, daß die höheren Ableitungen nach x symmetrische Operatoren sind (s. z.B. [4]). Ist nun (10) für $n=N$ schon bewiesen, so gilt

$$D^{N+1}(Pu) = D^N\frac{\partial}{\partial t}(Pu) + D^N\frac{\partial}{\partial x}(Pu)f = D^N\left(\left[\frac{\partial}{\partial t}P\right]\cdot(u) + P\left(\frac{\partial}{\partial t}u\right)\right) +$$

$$+ D^N\left(\left[\frac{\partial}{\partial x}P\right](u) + P\frac{\partial}{\partial x}u\right)(f) = \binom{N}{0}D^N\left[\frac{\partial}{\partial t}P\right](u) + \binom{N}{1}D^{N-1}\left[\frac{\partial}{\partial t}P\right](Du) +$$

$$+ \binom{N}{2}D^{N-2}\left[\frac{\partial}{\partial t}P\right](D^2u) + \ldots + \binom{N}{N}\left[\frac{\partial}{\partial t}P\right](D^Nu) + \binom{N}{0}D^N\left[\frac{\partial}{\partial x}P\right](u)(f) +$$

$$+ \binom{N}{1}D^{N-1}\left[\frac{\partial}{\partial x}P\right](Du)(f) + \binom{N}{2}D^{N-2}\left[\frac{\partial}{\partial x}P\right](D^2u)(f) +$$

$$+ \ldots + \binom{N}{N}\left[\frac{\partial}{\partial x}P\right](D^Nu)(f) + \binom{N}{0}D^NP\left(\frac{\partial}{\partial t}u\right) + \binom{N}{1}D^{N-1}P\left(D\frac{\partial}{\partial t}u\right) +$$

$$+ \binom{N}{2}D^{N-2}P\left(D^2\frac{\partial}{\partial t}u\right) + \ldots + \binom{N}{N}P\left(D^N\frac{\partial}{\partial t}u\right) + \binom{N}{0}D^NP\left(\frac{\partial}{\partial x}u\right)f +$$

$$+ \binom{N}{1}D^{N-1}P\left(D\frac{\partial}{\partial x}u\right)f + \binom{N}{2}D^{N-2}P\left(D^2\frac{\partial}{\partial x}u\right)f + \ldots + \binom{N}{N}P\left(D^N\frac{\partial}{\partial x}u\right)f =$$

$$= \binom{N}{0}\left\{D^N\left[\frac{\partial}{\partial t}P\right](u) + D^N\left[\frac{\partial}{\partial x}P\right](u)(f)\right\} +$$

$$+ \left\{\binom{N}{1}\cdot\left[D^{N-1}\left[\frac{\partial}{\partial t}P\right](Du) + D^{N-1}\left[\frac{\partial}{\partial x}P\right](Du)(f)\right] +\right.$$

$$\left. + \binom{N}{0}\cdot\left[D^NP\left(\frac{\partial}{\partial t}u\right) + D^NP\left(\frac{\partial}{\partial x}u\right)(f)\right]\right\} +$$

$$+ \left\{\binom{N}{2}\cdot\left[D^{N-2}\left[\frac{\partial}{\partial t}P\right](D\ u) + D^{N-2}\left[\frac{\partial}{\partial x}P\right](D^2u)(f)\right] + \binom{N}{1}\cdot\left[D^{N-1}P\left(D\frac{\partial}{\partial t}u\right) +\right.\right.$$

$$\left.\left. + D^{N-1}P\left(D\frac{\partial}{\partial x}u\right)(f)\right]\right\} + \ldots + \binom{N}{N}\left\{P\left(D^N\frac{\partial}{\partial t}u\right) + P\left(D^N\frac{\partial}{\partial x}u\right)(f)\right\} =$$

$$= \binom{N+1}{0}D^{N+1}P(u) + \left\{\binom{N}{1} + \binom{N}{0}\right\}D^NP(Du) + \ldots + \binom{N+1}{N+1}P\,(D^{N+1}u),$$

d.h. (10) ist auch für $n=N+1$, also für jedes n gültig.

Wir beweisen nun die Gültigkeit von

SATZ 2.

$$D(D^n u) = D^{n+1}u + nD^{n-1}\left[\frac{\partial}{\partial x}u\right](Df), \tag{11}$$

u. zw. wieder durch vollständige Induktion. Ist nämlich $n=1$, so ist (mit der Abkürzung, eingeführt in § 6a):

$$D(Du) = D\left(\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}u(f)\right) = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}u(f)\right) + \frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}u(f)\right)(f) =$$

$$= \frac{\partial^2 u}{\partial t^2} + \frac{\partial^2}{\partial x\,\partial t}u(f) + \frac{\partial}{\partial x}u\left(\frac{\partial f}{\partial t}\right) + \frac{\partial^2 u}{\partial t\,\partial x}(f) + \frac{\partial^2}{\partial x^2}u(f)(f) +$$

$$+\left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial x}f\right)(f) = \left(\frac{\partial^2}{\partial t^2}u + 2\frac{\partial^2}{\partial t\,\partial x}u(f) + \frac{\partial^2}{\partial x^2}u(f^2)\right) +$$

$$\frac{\partial}{\partial x}u\left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}(f)\right) = D^2 u + \left[\frac{\partial}{\partial x}u\right](Df),$$

d.h. (11) ist gültig. Ist es schon für $n=N$ bewiesen, so gilt, falls man auch (10) benützt:

$$D(D^{N+1}u) = D\left[D^N\frac{\partial}{\partial t}u + D^N\frac{\partial}{\partial x}u(f)\right] = D\left(D^N\frac{\partial}{\partial t}u\right) + D\left(D^N\frac{\partial}{\partial x}u(f)\right) =$$

$$= D\left(D^N\frac{\partial}{\partial t}u\right) + D\left(D^N\frac{\partial}{\partial x}u\right)(f) + D^N\frac{\partial}{\partial x}u\,(Df) = D^{N+1}\frac{\partial}{\partial t}u +$$

$$+ND^{N-1}\frac{\partial^2}{\partial t\,\partial x}u(Df) + D^{N+1}\frac{\partial}{\partial x}u(f) + ND^{N-1}\frac{\partial^2}{\partial x^2}u(f)(Df) +$$

$$+D^N\frac{\partial}{\partial x}u(Df) = D^{N+1}\frac{\partial}{\partial t}u + D^{N+1}\frac{\partial}{\partial x}u(f) +$$

$$+N\left\{D^{N-1}\frac{\partial}{\partial t}\left(\frac{\partial}{\partial x}u\right) + D^{N-1}\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}u\right)(f)\right\}(Df) + D^N\frac{\partial}{\partial x}u(Df) =$$

$$= D^{N+2}u + (N+1)D^N\frac{\partial}{\partial x}u(Df),$$

d.h. (11) auch für $n=N+1$ gültig, w.z.b.w.

Wir werden auch die Gleichung

SATZ 3.

$$D^n(Du) = D^{n+1}u + \binom{n}{1}D^{n-1}\frac{\partial}{\partial x}u(Df) + \binom{n}{2}D^{n-2}\frac{\partial}{\partial x}u(D^2 f) + \ldots + \binom{n}{n}\frac{\partial}{\partial x}u(D^n f) \tag{12}$$

benützen. (12) ist ja für $n=1$ nach (11) gültig. Ist es schon für $n=N$ bewiesen, so gilt

$$D^{N+1}(Du) = D^N \frac{\partial}{\partial t}(Du) + D^N\left[\frac{\partial}{\partial x}(Du)\right](f).$$

Hier ist nun

$$\frac{\partial}{\partial t}(Du) = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}u(f)\right) = \frac{\partial^2}{\partial t^2}u + \frac{\partial^2}{\partial t\,\partial x}u(f) + \frac{\partial}{\partial x}u\left(\frac{\partial}{\partial t}f\right) =$$

$$= D\frac{\partial}{\partial t}u + \left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial t}f\right)$$

und $\dfrac{\partial}{\partial x}(Du) \in (B \to B)$:

$$\frac{\partial}{\partial x}(Du) = \frac{\partial^2}{\partial t\,\partial x}u + \frac{\partial^2}{\partial x^2}u(f) + \left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial x}f\right) = D\frac{\partial}{\partial x}u + \left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial x}f\right)$$

also nach (10)

$$D^{N+1}(Du) = D^N\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}u\right) + D^N\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}u\right)(f) +$$

$$+ D^N\left\{\left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial t}f\right) + \frac{\partial^2}{\partial t\,\partial x}u(f) + \frac{\partial^2}{\partial x^2}u(f)(f) + \left[\frac{\partial}{\partial x}u\right]\left(\frac{\partial}{\partial x}f\right)(f)\right\} =$$

$$= D^{N+1}\frac{\partial}{\partial t}u + D^N\left\{\frac{\partial}{\partial t}\left(\frac{\partial}{\partial x}u(f)\right) + \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}u(f)\right)(f)\right\} = D^{N+1}\frac{\partial}{\partial t}u +$$

$$+ D^{N+1}\left\{\frac{\partial}{\partial x}[u(f)]\right\} = D^{N+1}\frac{\partial}{\partial t}u + D^{N+1}\left(\frac{\partial}{\partial x}u\right)(f) + \binom{N+1}{1}D^N\left(\frac{\partial}{\partial x}u\right)(Df) +$$

$$+ \binom{N+1}{2}D^{N-1}\left(\frac{\partial}{\partial x}u\,(D^2f) + \ldots + \binom{N+1}{N+1}\frac{\partial}{\partial x}u\,(D^{N+1}f) = D^{N+2}u +\right.$$

$$+ \binom{N+1}{1}D^N\left(\frac{\partial}{\partial x}u\right)(Df) + \ldots + \binom{N+1}{N+1}\left(\frac{\partial}{\partial x}u\right)(D^{N+1}f).$$

(12) ist also auch für $n=N+1$, folglich für alle n gültig.
Wir werden daneben auch die Relation in

SATZ 5 benützen:

$$\frac{\partial^n}{\partial x^n}(Du) \in ({}_1B \to ({}_2B \to ({}_3\ldots \to ({}_{n+1}B \to B)\ldots))) = D\frac{\partial^n u}{\partial x^n} +$$

(13)

$$+ \binom{n}{1}\frac{\partial^n u}{\partial x^n}\frac{\partial f}{\partial x} + \binom{n}{2}\frac{\partial^{n-1}u}{\partial x^{n-1}}\frac{\partial^2 f}{\partial x^2} + \ldots + \binom{n}{n}\frac{\partial u}{\partial x}\frac{\partial^n f}{\partial x^n}.$$

Für $n=1$ ist nämlich

$$\frac{\partial}{\partial x}(Du)=\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}u+\frac{\partial}{\partial x}u(f)\right)=\frac{\partial}{\partial t}\frac{\partial u}{\partial x}+\frac{\partial}{\partial x}\frac{\partial u}{\partial x}(f)+$$

$$+\frac{\partial u}{\partial x}\frac{\partial f}{\partial x}=D\frac{\partial u}{\partial x}+\frac{\partial u}{\partial x}\frac{\partial f}{\partial x},$$

ist also (13) gültig. Ist es schon für $n=N$ bewiesen, so gilt

$$\frac{\partial^{N+1}}{\partial x^{N+1}}(Du)=\frac{\partial}{\partial x}\left[\frac{\partial^N}{\partial x^N}(Du)\right]=\frac{\partial}{\partial x}\left[D\frac{\partial^N u}{\partial x^N}+\binom{N}{1}\frac{\partial^N u}{\partial x^N}\frac{\partial f}{\partial x}+\ldots+\binom{N}{N}\frac{\partial u}{\partial x}\frac{\partial^N f}{\partial x^N}\right]=$$

$$=\frac{\partial}{\partial x}\left[\frac{\partial}{\partial t}\frac{\partial^N u}{\partial x^N}+\frac{\partial}{\partial x}\frac{\partial^N u}{\partial x^N}(f)+\binom{N}{1}\frac{\partial^N u}{\partial x^N}\frac{\partial f}{\partial x}+\binom{N}{2}\frac{\partial^{N-1}u}{\partial x^{N-1}}\frac{\partial^2 f}{\partial x^2}+\ldots\right]=$$

$$=\frac{\partial}{\partial t}\frac{\partial^{N+1}u}{\partial x^{N+1}}+\left[\frac{\partial}{\partial x}\frac{\partial^{N+1}u}{\partial x^{N+1}}\right](f)+\binom{N}{0}\frac{\partial^{N+1}u}{\partial x^{N+1}}\frac{\partial f}{x\partial}+\binom{N}{1}\frac{\partial^{N+1}u}{\partial x^{N+1}}\frac{\partial f}{\partial x}+$$

$$+\binom{N}{1}\frac{\partial^N u}{\partial x^N}\frac{\partial^2 f}{\partial x^2}+\binom{N}{2}\frac{\partial^N u}{\partial x^N}\frac{\partial^2 f}{\partial x^2}+\ldots+\binom{N}{N}\frac{\partial u}{\partial x}\frac{\partial^{N+1}f}{\partial x^{N+1}}=D\frac{\partial^{N+1}u}{\partial x^{N+1}}+$$

$$+\binom{N+1}{0}\frac{\partial^{N+1}u}{\partial x^{N+1}}\frac{\partial f}{\partial x}+\binom{N+1}{1}\frac{\partial^N u}{\partial x^N}\frac{\partial^2 f}{\partial x^2}+\ldots+\binom{N+1}{N+1}\frac{\partial u}{\partial x}\frac{\partial^{N+1}f}{\partial x^{N+1}},$$

also ist (13) wieder gültig, w.z.b.w.

Wir gebrauchen auch Formeln der Form $D^r\left[\frac{\partial^s}{\partial x^s}(Du)\right]$, welche wir aber in geschlossener Form nicht angeben, sondern mit Hilfe von (10)—(11)—(12)—(13) in den konkreten Fällen bilden werden.

Da der Operator D — wie wir es schon gesehen haben — die Ableitung längs der Integralkurven von (2) bildet, so können wir mit Hilfe von (10)—(11)—(13) die höheren Ableitungen der (2) genügenden Funktion $x(t)$ bilden, falls $f(x, t)$ genug oft differenzierbar ist. Es gelten nämlich die folgenden, rekursiv angebbaren Formeln:

$$\frac{dx}{dt}=f(x,t)$$

$$\frac{d^2x}{dt^2}=Df$$

$$\frac{d^3x}{dt^2}=D(Df)=D^2f+\left(\frac{\partial}{\partial x}f\right)(Df)$$

(14)
$$\frac{d^4x}{dt^3}=D\left(\frac{d^3x}{dt^3}\right)=D(D^2f)+D\left(\left(\frac{\partial}{\partial x}f\right)(Df)\right)=D^3f+2D\frac{\partial f}{\partial x}(Df)+$$

$$+D\frac{\partial f}{\partial x}(Df)+\frac{\partial f}{\partial x}(D(Df))=D^3f+3D\frac{\partial f}{\partial x}(Df)+\frac{\partial f}{\partial x}D^2f+\left(\frac{\partial f}{\partial x}\right)^2Df$$

$$\frac{d^5x}{dt^5} = D^4f+6D^2\frac{\partial f}{\partial x}(Df)+3\frac{\partial^2 f}{\partial x^2}(Df)(Df)+4D\frac{\partial f}{\partial x}(D^2f)+3\frac{\partial f}{\partial x}D\frac{\partial f}{\partial x}(Df)+$$

$$+4D\frac{\partial f}{\partial x}\cdot\frac{\partial f}{\partial x}(Df)+\frac{\partial f}{\partial x}(D^3f)+\left(\frac{\partial f}{\partial x}\right)^2(D^2f)+\left(\frac{\partial f}{\partial x}\right)^3(Df);\ \text{usf.}\ ^{1}$$

Die klassischen Hilfsgrößen des RUNGE—KUTTAschen Typs sind von der Form $\Delta t\cdot f=hf$. Hier braucht man also eine Taylor-Entwicklung von

$$f(t_0+\alpha h;\ x_0+\alpha hf_0+\Theta),$$

aufzuschreiben in bezug auf die Stelle (t_0, x_0), die wir im folgenden mit dem Index $_0$ verkürzen werden; Θ bedeutet hier ein Element von B, mindestens zweiten Grades in h. Es ist nun bekannt (s. z.B. [2]; [4]), daß

$$f(t_0+\alpha h; x_0+\alpha hf_0+\Theta) = f_0+\alpha h\left(\frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}(f_0)+\frac{\partial f}{\partial x}(\Theta)\right)+$$

$$+\frac{\alpha^2h^2}{2!}\left(\frac{\partial^2 f}{\partial t^2}+2\frac{\partial^2 f}{\partial t\,\partial x}(f_0)+\frac{\partial^2 f}{\partial x^2}(f_0)(f_0)\right)+\alpha h^2\frac{\partial^2 f}{\partial t\,\partial x}(\Theta)+$$

$$+\frac{1}{2}h^2\frac{\partial^2 f}{\partial x^2}[(\alpha f_0+\Theta)(\alpha f_0+\Theta)-\alpha^2(f_0)(f_0)]+\ldots = f_0+\alpha hDf_0+$$

(15)

$$+\left.\frac{\partial f}{\partial x}\right|_0\cdot(\Theta)+\frac{\alpha^2h^2}{2!}D^2f_0+\alpha hD\left.\frac{\partial f}{\partial x}\right|_0(\Theta)+\frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_0\cdot(\Theta^2)+\frac{\alpha^3h^3}{3!}D^3f_0+$$

$$+\frac{\alpha^2h^2}{2!}D^2\left.\frac{\partial f}{\partial x}\right|_0(\Theta)+\frac{\alpha h}{2!}D\left.\frac{\partial^2 f}{\partial x^2}\right|_0(\Theta^2)+\frac{1}{3!}\left.\frac{\partial^3 f}{\partial x^3}\right|_0(\Theta^3)+\ldots$$

(Hier haben wir wieder die Tatsache gebraucht, daß die stetigen höheren Ableitungen von f nach x symmetrische Operatoren sind, d.h., daß z.B. $\frac{\partial^2 f}{\partial x^2}(f_0)\cdot(\Theta)=$ $=\frac{\partial^2 f}{\partial x^2}(\Theta)\cdot(f_0)$ gültig ist). Jedoch werden wir auch solche neue Formeln angeben, wo auch $\frac{\partial f}{\partial x}$ bzw. Df explizit auftritt. Eben deswegen brauchen wir auch die Taylor-

[1] Es ist also interessant, daß von der fünften Ableitung an schon eine wichtige Differenz zwischen den RUNGE-KUTTA Formeln für eine Differentialgleichung mit einer reellen bzw. komplexen Veränderlichen, und für allgemeinere Veränderliche auftritt. Bei der ersten gilt $\frac{\partial f}{\partial x}D\frac{\partial f}{\partial x}(Df)=$ $=D\frac{\partial f}{\partial x}\cdot\frac{\partial f}{\partial x}(Df)$, bei der zweiten aber im allgemeinen nicht. Die bekannten RUNGE-KUTTA Formeln höherer als vierter Ordnung sind nur im ersten Fall anwendbar!

Entwicklung von $g(t_0+\alpha h; x_0+\alpha h f_0+\Theta)$. Nun ist

$$g(t_0+\alpha h; x_0+\alpha h f_0+\Theta) = g_0+\alpha h\left\{\frac{\partial g}{\partial t}+\frac{\partial g}{\partial x}(f_0)\right\}+\frac{\partial g}{\partial x}(\Theta)+$$

$$+\frac{\alpha^2 h^2}{2!}\left\{\frac{\partial^2 g}{\partial t^2}+2\frac{\partial^2 g}{\partial t\,\partial x}(f_0)+\frac{\partial^2 g}{\partial x^2}(f_0)(f_0)+\alpha h\frac{\partial^2 g}{\partial t\,\partial x}(\Theta)+\right.$$

$$\left.+\frac{1}{2}\frac{\partial^2 g}{\partial x^2}[(\alpha h f_0+\Theta)(\alpha h f_0+\Theta)-\alpha h^2 f_0^2]\right\}+\ldots = g_0+\alpha h D g_0+ \tag{16}$$

$$+\frac{\partial g}{\partial x}\bigg|_0(\Theta)+\frac{\alpha^2 h^2}{2}D^2 g_0+\alpha h D\frac{\partial g}{\partial x}\bigg|_0(\Theta)+\frac{1}{2}\frac{\partial^2 g}{\partial x^2}\bigg|_0(\Theta)(\Theta)+$$

$$+\frac{\alpha^3 h^3}{3!}D^3 g_0+\frac{\alpha^2 h^2}{2!}D^2\frac{\partial g}{\partial x}\bigg|_0(\Theta)+\frac{\alpha h}{2}D\frac{\partial^2 g}{\partial x^2}\bigg|_0(\Theta)(\Theta)+\ldots.$$

Wir betrachten zunächst die Fehlerabschätzung einer allgemeinen Form des RUNGE—KUTTAschen Typs. Hier soll man bedenken, daß alle diese Formeln mit einer gewichteten Summe $k_n(h)$ arbeiten, die eine Taylor Entwicklung (in h) besitzt, übereinstimmend mit derselben der Differenz $x(t_0+h)-x(t_0)=\Delta x_0$ bis zur n-ten Potenz von h. Es ist nun sehr leicht zu zeigen, daß eine Abschätzung der Form

$$\|\varepsilon(h)\| = \|k_n(h)-\Delta x_0\| \leqq C(k_n)\cdot\frac{h^{n+1}}{(n+1)!} \tag{17}$$

feststeht, jedoch die Konstante $C(k_n)$ stark von der Form von k_n abhängt. Um auch diese Konstante abzuschätzen, müssen wir einige Bemerkungen machen. Vor allem soll darauf hingewiesen werden, daß gemäß unserer Voraussetzungen in bezug auf f eine Schranke $F(h)$ für $\|f(t, x)\|$ so angegeben werden kann, daß in D: $\{|t-t_0|\leqq h;\ \|x-x_0\|\leqq|t-t_0|\cdot F\}\|f(t, x)\|\leqq F$ gültig sei. Dann gilt der

HILFSSATZ 1. Für $|t-t_0|\leqq h$ ist $(t, x(t))\in D$ gültig.

BEWEIS: Es ist gut bekannt (s. z.B. [4]), daß neben den angegebenen Voraussetzungen $x(t)$ durch die Approximationsfolge

$$x_0\equiv x_0;\quad x_{n+1}(t) = x_0+\int_{t_0}^{t} f(\xi; x_{(n)}(\xi))\,d\xi \tag{18}$$

gleichmäßig (u. zw. im „starken" Sinne) angenähert werden kann. $(t, x_{(0)}(t))\in D$ ist nun evident; ist $(t, x_{(n)}(t))\in D$ für $n=N$ schon bewiesen, so gilt

$$\|x_{(N+1)}(t)-x_0\| \leqq \left|\int_{t_0}^{t}\|f(\xi; x_{(N)}(\xi))\|\,d\xi\right| \leqq |t-t_0|\cdot F, \tag{19}$$

d.h. auch $(t, x_{N+1}(t))\in D$, w.z.B.w.

Einen ebensolchen Bereich $D(k_n)$ für $(t_0+\tau h; x_0+k_n(\tau h))$ anzugeben ist jedoch schon viel schwerer. Ist nämlich $k_n(h)$ eine gewichtete Summe von Werten

nur von $f(t, x)$, mit nichtnegativen Gewichten, so kann man $D(k_n)=D$ wählen. Ist aber k_n allgemeiner, besitzt z.B. Glieder, gebildet mit Hilfe von $\frac{\partial f}{\partial x}$ bzw. Df, so soll man die Schranken dieser Größen in $D(k_n)$ auch betrachten, um $D(k_n)$ anzugeben. Es sei also vorausgesetzt, daß $C(k_n)$ eine solche Konstante ist, daß auch $k_n(\tau h)$, auch alle Hilfsgrößen von k_n für jedes $0 \leqq \tau \leqq 1$ in $D(k_n):\{|t-t_0| \leqq h;$ $\|x-x_0\| \leqq |t-t_0| C(k_n) F\}$ bleiben. $C(k_n)$ kann man mit Hilfe der expliziten Form von $k_n(h)$ und mit Hilfe von Schranken für $\left\| \frac{\partial f}{\partial x} \right\|$, $\|Df\|$ usf. in $D(k_n)$ abschätzen.

Betrachte man nun $k_n(\tau h)$ für $0 \leqq \tau \leqq 1$. Man kann zunächst eine solche Funktion $\Delta(t; y)$ angeben, für welche die Lösung der Anfangswertaufgabe $\dot{y}=f(t, y)+$ $+\Delta(t, y)$; $y(t_0)=x_0$ kongruent $x_0+k_n(\tau h=t-t_0)$ ist. Wählt man nämlich $\Delta(t, y)$ beliebig, aber so, daß

$$\frac{dk_n(\tau h)}{d(\tau h)} \equiv f\big(t_0+\tau h;\, x_0+k_n(\tau h)\big)+\Delta\big(t_0+\tau h;\, x_0+k_n(\tau h)\big) \tag{20}$$

gültig ist, so entspricht uns Δ. Es folgt also aus unseren obigen Überlegungen, daß man eine solche Funktion Δ auch so wählen kann, daß $\left\| \frac{\partial^{\varrho}\Delta}{\partial t^{\varrho_1}\partial y^{\varrho_2}} \right\| \leqq$ $\leqq \big(1+C(k_n)\big) \left\| \frac{\partial^{\varrho} f}{\partial t^{\varrho_1}\partial y^{\varrho_2}} \right\|$ gültig sei $(\varrho_1+\varrho_2=\varrho \leqq n)$. Man kann somit eine obere Schranke für die Norm der n-ten Ableitungen von f bzw. Δ in D bzw. $D(k_n)$ angeben. Da der Definition von Δ bzw. $k_n(\tau h)$ gemäß alle die Ableitungen von Δ höchstens $(n-1)$-ter Ordnung in (t_0, x_0) verschwinden, kann man durch die Abschätzung der n-ten Ableitungen von f und mit Hilfe des verallgemeinerten Mittelwertsatzes (s. z.B. [2]) eine konstante $C_1(k_n)$ so angeben, daß

$$\|\Delta(t, y)\| \leqq C_1(k_n) \frac{|t-t_0|^n}{n!} \tag{21}$$

für $|t-t_0| \leqq h$; $\|y-x_0\| \leqq |t-t_0| C(k_n) F$ gültig ist.
Mit diesen Bezeichnungen gilt nun der

SATZ 6.

$$\|\varepsilon(h)\| \leqq C_1(k_n) \cdot \frac{|h|^{n+1}}{(n+1)!} \exp\{L[1+C(k_n)] \cdot |h|\}, \tag{22}$$

wo L die Lipschitz-Konstante von $f(t, x)$ bezeichnet.

BEWEIS: Betrachten wir die Integralgleichung

$$y(t) = x_0 + \int_{t_0}^{t} \{f(\xi; y(\xi)) + \Delta(\xi, y(\xi))\}\, d\xi, \tag{23}$$

welche eben die Lösung $k_n(\tau h=t-t_0)$ besitzt. Suchen wir nun die Lösung von (23)

mit Hilfe sukzessiver Approximation, und nimmt man $y_0(t)=x(t)$ als erste Annäherung, so bekommt man

$$(24)\qquad y_1(t)=x_0+\int_{t_0}^{t} f(\xi, x(\xi))d\xi+\int_{t_0}^{t} \Delta(\xi, x(\xi))d\xi = x(t)+\int_{t_0}^{t} \Delta(\xi, x(\xi))d\xi,$$

da x die Lösung der Anfangswertaufgabe $\dot{x}=f(t, x)$; $x(t_0)=x_0$ ist. Nun (21) gemäß

$$(25)\qquad \|y_1(\xi)-x(\xi)\| \leqq \left|\int_{t_0}^{t} C_1(k_n)\frac{|\xi-t_0|^n}{n!}d\xi\right| = C_1(k_n)\frac{|t-t_0|^{n+1}}{(n+1)!}.$$

Ebenso bekommen wir

$$\|y_2(\xi)-y_1(\xi)\| \leqq \left|\int_{t_0}^{t} \|f(\xi, y_1(\xi))-f(\xi, x(\xi))\|d\xi\right| +$$

$$+\left|\int_{t_0}^{t} \|\Delta(\xi, y_1(\xi))-\Delta(\xi, x(\xi))\|d\xi\right| \leqq \int_{t_0}^{t}\left\{L\cdot C_1(k_n)\frac{|\xi-t_0|^{n+1}}{(n+1)!}+\right.$$

$$\left.+(1+C(k_n))L\cdot C_1(k_n)\frac{|\xi-t_0|^{n+1}}{(n+1)!}\right\}d\xi = [2+C(k_n)]\cdot LC_1(k_n)\frac{|t-t_0|^{n+2}}{(n+2)!},$$

$$\|y_3(\xi)-y_2(\xi)\| \leqq \int_{t_0}^{t}\{L+(1+C(k_n))L\}[2+C(k_n)]LC_1(k_n)\frac{|\xi-t_0|^{n+2}}{(n+2)!}d\xi =$$

$$= \{L[2+C(k_n)]\}^2 C_1(k_n)\frac{|t-t_0|^{n+3}}{(n+3)!}, \text{ usf.}$$

Ist also für $m=N$

$$(26)\qquad \|y_{m+1}-y_m\| \leqq \{L[2+C(k_n)]\}^m C_1(k_n)\frac{|t-t_0|^{m+n+1}}{(m+n+1)!}$$

noch gültig, so bekommt man

$$\|y_{N+2}-y_{N+1}\| \leqq \int_{t_0}^{t}\{L+[1+C(k_n)L\}\cdot\{L[2+C(k_n)\}^N C_1(k_n)\frac{|\xi-t_0|^{N+n+1}}{(N+n+1)!}d\xi =$$

$$= \{L[2+C(k_n)]\}^{N+1} C_1(k_n)\frac{|t-t_0|^{N+n+2}}{(N+n+2)!},$$

d.h. (26) ist auch für $m=N+1$, d.h. für alle m gültig. Daraus folgt also einerseits, daß $y_n(t)$ stark gleichmäßig in $D(k_n)$ gegen $k_n(\tau h=t-t_0)$ strebt, ferner, daß

$$(27)\qquad \|k_n(\tau h)-x(t_0+\tau h)\| \leqq \sum_{m=1}^{\infty}\|y_m(\tau h)-y_{m-1}(\tau h)\| \leqq \sum_{m=1}^{\infty}\frac{\{L[2+C(k_n)]\}^m}{(m+n+1)!}\cdot$$

$$\cdot C_1(k_n)\cdot(\tau h)^{m+n+1} \leqq C_1(k_n)\frac{|\tau h|^{n+1}}{(n+1)!}\cdot\exp\{L[2+C(k_n)]\cdot|\tau h|\},$$

w.z.b.w.

Mit Hilfe dieses Satzes kann man jedoch praktisch keine Fehlerabschätzung angeben, da auch $C_1(k_n)$, auch $C(k_n)$ sehr schwer bestimmt werden kann. Man kann jedoch mit Hilfe (27) eine Formel

$$x(t_0+h)-x(t_0) = k_n(h)+\psi_n(t_0, h)\cdot h^{n+1} \tag{28}$$

angeben, wo ψ die Menge $R\times R$ in B abbildet, u. zw. stetig. Auch in unserem Fall kann man also die praktisch immer angewendete Methode mit der verdoppelten Schrittweite anwenden.

§ 3. Neue Formeln des Runge-Kuttaschen Typs

Wir werden die klassischen Runge—Kutta Formeln in drei Hinsichten verallgemeinern. Einerseits — um mit dem Rechenaufwand zu sparen, was bei Differentialgleichungen in allgemeineren Räumen vielleicht noch wichtiger als bei Vektor- oder Skalarräumen ist — werden wir solche Formeln angeben, welche der Struktur nach ebenso aufgebaut sind wie die klassischen, — jedoch benützen wir hier Hilfsgrößen aus vorigen Intervallen. Gleichzeitig können wir mit diesen Hilfsgrößen auch Formeln für die doppelte Intervallänge angeben, wodurch man ohne weiteres eine grobe, aber wichtige Fehlerabschätzung angeben kann. Wir brauchen dazu den

HILFSSATZ 2. Die Funktion $f(t, x)$ sei in der Umgebung $D(k_n)$ von (t_0, x_0) n-mal stetig differenzierbar, und setzen wir voraus, daß dem Satz 6 gemäß $\|x(t_0+h;\ x_0, t_0)-(x_0+k_n(h))\| = O(h^{n+1})$. Dann ist auch $\|x[t_0;\ x_0+k_n(h),\ t_0+h]-x_0\| = O(h^{n+1})$ gültig.

BEWEIS. Wie schon bemerkt wurde, gilt auf Grund unserer Voraussetzungen die Relation

$$x_1-x_0 = x(t_0+h;\ x_0, t_0)-x_0 = k_n(h)+\psi_n(t_0, h)\cdot h^{n+1}. \tag{29}$$

Daneben ist das Anfangswertproblem unseren Voraussetzungen gemäß eindeutig lösbar; folglich gilt

$$x(t_0;\ x_1,\ t_0+h)=x_0. \tag{30}$$

Da ferner f in $D(k_n)$ eine Lipschitz-Bedingung mit der Konstante L, u. zw. in t gleichmäßig erfüllt, so gilt

$$\|x(t_0;\ x_0+k_n(h),\ t_0+h)-x_0\| = \|x(t_0;\ x_0+k_n(h),\ t_0+h)-$$

$$-x(t_0;\ x_1,\ t_0+h)\| \leqq \psi_n(t_0, h)\cdot h^{n+1}\cdot e^{Lh} = O(h^{n+1}),$$

w.z.b.w.

Wir werden nun einige neue — in dieser Arbeit aber nur einfache — Formeln mit Hilfe von Hilfssatz 2 angeben. Und zwar erst eine einpunktige Formel dritter Ordnung, mit einer Hilfsformel für doppelte Schrittweite.

Es sei also

$$k_0 = hf(t_0+\alpha h;\ x_0+\beta k_{-1}+(\alpha-\beta)k_{-2})$$

und

$$k_3(h) = R_{-1}k_{-1}+R_0k_0,$$

wo die negativen Indexe Größen der vorigen Schritte bedeuten. Hier ist nun dem Hilfssatz 2 bzw. (14) gemäß

$$k_{-5} = hf_0 + O(h^2); \quad k_{-4} = hf_0 + O(h^2)$$

$$x_{-3} = x_0 - 3hf_0 + O(h^2)$$

$$k_{-3} = hf(t_{-3} + \alpha h;\, x_{-3} + \beta k_{-4} + (\alpha - \beta)k_{-5}) =$$

$$= hf(t_0 - (3-\alpha)h;\, x_{-3} - (3-\alpha)hf_0 + O(h^2)) = hf_0 - (3-\alpha)h^2 Df_0 + O(h^3)$$

$$x_{-2} = x_0 - 2hf_0 + O(h^2)$$

$$k_{-2} = hf(t_{-2} + \alpha h;\, x_{-2} + \beta k_{-3} + (\alpha - \beta)k_{-4}) = hf_0 - (2-\alpha)h^2 Df_0 + O(h^3)$$

$$x_{-1} = x_0 - hf_0 + \frac{1}{2} h^2 Df_0 + O(h^3)$$

$$k_{-1} = hf(t_{-1} + \alpha h;\, x_{-1} + \beta k_{-2} + (\alpha - \beta)k_{-3}) =$$

$$= hf\Big(t_0 - (1-\alpha)h;\, x_0 - (1-\alpha)hf_0 -$$

$$-\left[\frac{1}{2} + \beta(2-\alpha) + (\alpha-\beta)(3-\alpha)\right] h^2 Df_0 + O(h^3)\Big) = hf_0 - (1-\alpha)h^2 Df_0 +$$

$$+ \frac{(1-\alpha)^2}{2} h^3 D^2 f_0 + \left(\frac{1}{2} + 3\alpha - \beta - \alpha^2\right) h^3 f'_{x,0} Df_0 + O(h^4)$$

$$k_0 = hf(t_0 + \alpha h;\, x_0 + \alpha hf_0 - [\beta(1-\alpha) + (\alpha-\beta)(2-\alpha)]h^2 Df_0 + O(h^3)) =$$

$$= hf_0 + \alpha h^2 Df_0 + \frac{\alpha^2}{2} h^3 D^2 f_0 + (2\alpha - \beta - \alpha^2) h^3 f'_{x,0} Df_0 + O(h^4).$$

Unsere Gleichungen sind also:

$$R_{-1} + R_0 = 1$$

$$-(1-\alpha)R_{-1} + \alpha R_0 = \frac{1}{2}$$

$$\frac{(1-\alpha)^2}{2} R_{-1} + \frac{\alpha^2}{2} R_0 = \frac{1}{6}$$

$$\left(\frac{1}{2} + 3\alpha + \beta + \alpha^2\right) R_{-1} + (-2\alpha + \beta + \alpha^2) R_0 = \frac{1}{6},$$

d. h.: $R_1 = \frac{1}{2} - \sqrt{\frac{1}{6}}$, $R_0 = \frac{1}{2} + \sqrt{\frac{1}{6}}$, $\alpha = 1 - \sqrt{\frac{1}{6}}$, $\beta = \frac{17 - 2\sqrt{6}}{12}$.

Die Frage bleibt noch offen, wie man die Rechnung mit der Formel

$$k_0 = hf\left(t_0 + \left(1 - \frac{1}{\sqrt{6}}\right)h;\ x_0 + \frac{17 - 2\sqrt{6}}{12}k_{-1} - \frac{5}{12}k_{-2}\right); \tag{31}$$

$$k_3(h) = \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right)k_{-1} + \left(\frac{1}{2} + \frac{1}{\sqrt{6}}\right)k_0$$

anfangen kann. Unsere Formeln zeigen aber, daß für k_{-5} und k_{-4} uns eine Pünktlichkeit $O(h^2)$ genügt, d.h. man kann $k_{-5} \cong hf(t_{-5}, x_{-5})$; $k_{-4} \cong hf(t_{-4}, x_{-4})$ beachten. Man kann dann k_{-3} und k_{-2} mit einer Pünktlichkeit $O(h^3)$, endlich k_{-1} und k_0 mit einer Pünktlichkeit $O(h^4)$ berechnen. Es bedeutet aber, daß — ausgehend von „(t_{-5}, x_{-5})" — $x_{-4}, x_{-3}, x_{-2}, x_{-1}$, und x_0 mit Hilfe einer anderen Methode dritter Ordnung gerechnet werden soll, und nur dann von (31) Gebrauch gemacht werden kann.

Für die verdoppelte Schrittweite braucht man eine Formel

$$\hat{k}_3(2h) = S_{-3}k_{-3} + S_{-2}k_{-2} + S_{-1}k_{-1} + S_0k_0, \tag{32}$$

wo man aber den Punkt (t_{-1}, x_{-1}) als Basispunkt betrachten soll. Außerdem ist (32) nicht vollständig passend für unsere Zwecke, da in den Gliedern vierter Ordnung $\hat{k}_3(2h)$ nicht $k_3(h)$ entspricht. (Es bedeutet dann, daß $\tilde{\psi}_3$ in $\hat{k}_3$ sicher nicht in der Nähe von ψ_3 in k_3 liegt.) Um auch diese Glieder einander entsprechend zu wählen, braucht man eine Formel

$$k_3(2h) = R_{-7}k_{-7} + R_{-6}k_{-6} + R_{-5}k_{-5} + R_{-4}k_{-4} + R_{-3}k_{-3} + \tag{33}$$

$$+ R_{-2}k_{-2} + R_{-1}k_{-1},$$

im allgemeinen ist jedoch schon $\hat{k}_3$ genug grob, so daß praktisch $\frac{1}{14}\|k_3(2h; t)_{-1} - k_3(h; t_{-1}) - k_3(h; t_0)\|$ eine obere Schranke der Fehler $\|k_3(h) - \Delta x\|$ ist. In unserem Fall ist z.B.

$$\hat{k}_3(2h) = \left(\frac{1}{2} + \frac{1}{\sqrt{6}}\right)k_{-3} - \left(1 + \frac{4}{\sqrt{6}}\right)k_{-2} + \left(\frac{5}{2} + \frac{3}{\sqrt{6}}\right)k_{-1}. \tag{34}$$

Eine andere Formel dritter Ordnung bekommt man mit Hilfe zweier Hilfsgrößen in jedem Schritt, welche sehr stabil ist, da sie nur wenig die Hilfsgrößen des vorigen Teilintervalles benützt. Es sei

$$k_0 = hf(t_0;\ x_0);\ k_1 = hf(t_0 + \alpha h;\ x_0 + (\alpha - \beta)k_0 + \beta k_{-1});$$

$$k_3(h) = R_0k_0 + R_1k_1.$$

Um auch die Glieder vierter Ordnung in $k_3(h)$ zu betrachten, benötigt man k_{-1} mit einer Pünktlichkeit dritter Ordnung, d. h. k_{-3} in zweiter, d.h. k_{-5} in erster

Ordnung. Nun

$$k_{-5} = hf_0 + O(h^2)$$

$$x_{-2} = x_0 - 2hf_0 + O(h^2); \quad k_{-4} = hf_0 + O(h^2)$$

$$k_{-3} = hf\big(t_{-2} + \alpha h;\, x_{-2} + (\alpha-\beta)k_{-4} + \beta k_{-5}\big) =$$

$$= hf\big(t_0 - (2-\alpha)h;\, x_0 - (2-\alpha)hf_0 + O(h^2)\big) = hf_0 - (2-\alpha)h^2 Df_0 + O(h^3);$$

$$x_{-1} = x_0 - hf_0 + \frac{1}{2}h^2 Df_0 + O(h^3);$$

$$k_{-2} = hf(t_{-1}; x_{-1}) = hf_0 - h^2 Df_0 + \frac{h^3}{2}D^2 f_0 + \frac{h^3}{2}f'_{x,0}Df_0 + O(h^4)$$

$$x_{-1} = x_0 - hf_0 + \frac{h^2}{2}Df_0 + O(h^3);$$

$$k_{-1} = hf\big(t_{-1} + \alpha h;\, x_{-1} + (\alpha-\beta)k_{-2} + \beta k_{-3}\big) =$$

$$= hf\left(t_0 - (1-\alpha)h;\, x_0 - (1-\alpha)hf_0 + \left(\frac{1}{2} - \alpha - \beta + \alpha\beta\right)h^2 Df_0 + O(h^3)\right) =$$

$$= hf_0 - (1-\alpha)h^2 Df_0 + \frac{(1-\alpha)^2}{2!}h^3 D^2 f_0 +$$

$$+\left(\frac{1}{2} - \alpha - \beta + \alpha\beta\right)h^3 f'_{x,0}Df_0 + O(h^4)$$

$$k_0 = hf(t_0, x_0) = hf_0$$

$$k_1 = hf\left(t_0 + \alpha h;\, x_0 + \alpha hf_0 - \beta(1-\alpha)h^2 Df_0 + \beta\frac{(1-\alpha)^2}{2}h^3 D^2 f_0 +\right.$$

$$\left.+\left(\frac{1}{2}\beta + \alpha\beta^2 - \alpha\beta - \beta^2\right)h^3 f'_{x,0}Df_0 + O(h^4)\right) = hf_0 + \alpha h^2 Df_0 +$$

$$+\frac{\alpha^2}{2}h^3 D^2 f_0 - \beta(1-\alpha)h^3 f'_{x,0}Df_0 + \frac{\alpha^3}{6}h^4 D^3 f_0 + \beta\frac{(1-\alpha)^2}{2}h^4 f'_{x,0}D^2 f_0 +$$

$$+\left(\frac{1}{2}\beta + \alpha\beta^2 - \alpha\beta - \beta^2\right)h^4 f'^2_{x,0}Df_0 - \alpha\beta(1-\alpha)h^4 Df'_x Df + O(h^5).$$

Folglich

$$R_0 + R_1 = 1; \quad R_1\alpha = \frac{1}{2}; \quad R_1\frac{\alpha^2}{2} = \frac{1}{6}; \quad -R_1\beta(1-\alpha) = \frac{1}{6};$$

also

$$\alpha = \frac{2}{3}; \quad R_1 = \frac{3}{4}; \quad R_0 = \frac{1}{4}; \quad \beta = -\frac{2}{3}.$$

Unsere Formel ist also

$$k_3(h) = \frac{1}{4}k_0 + \frac{3}{4}k_1 = \frac{1}{4}hf(t_0, x_0) + \frac{3}{4}hf\left(t_0 + \frac{2}{3}h; x_0 + \frac{4}{3}k_0 - \frac{2}{3}k_{-1}\right) =$$

$$= hf_0 + \frac{1}{2}h^2 Df_0 + \frac{1}{6}h^3 D^2 f_0 + \frac{1}{6}h^3 f'_{x,0} Df_0 + \frac{1}{27}h^4 D^3 f_0 - \tag{35}$$

$$- \frac{1}{36}h^4 f'_{x,0} D^2 f_0 + \frac{1}{36}h^4 f'^2_{x,0} Df_0 + \frac{1}{9}h^4 Df'_{x,0} Df_0 + O(h^5).$$

Eine Formel dritter Ordnung für die doppelte Schrittweite (mit dem Basispunkt t_{-1}, x_{-1}) ist

$$\hat{k}_3(2h) = \frac{2}{5}k_{-2} + k_0 + \frac{3}{5}k_1, \tag{36}$$

welche jedoch nicht in den Gliedern vierter Ordnung mit deren in (35) übereinstimmt. Eben deswegen geben wir jetzt die Glieder $k_{-6}, k_{-5}, \ldots, k_0, k_1$ in einer Pünktlichkeit vierter Ordnung in bezug auf t_{-1}, x_{-1} an:

$$k_{-11} = hf_{-1} + O(h^2); \quad k_{-10} = hf_{-1} + O(h^2)$$

$$k_{-9} = hf\left(t_{-5} + \frac{2}{3}h, \quad x_{-5} + \frac{4}{3}k_{-10} - \frac{2}{3}k_{-11}\right) =$$

$$= hf\left(t_{-1} - \frac{10}{3}h; x_{-1} - \frac{10}{3}hf_{-1} + O(h^2)\right) = hf_{-1} - \frac{10}{3}h^2 Df_{-1} + O(h^3);$$

$$k_{-8} = hf(t_{-4}, x_{-4}) = hf(t_{-1} - 3h; x_{-1} - 3hf_{-1} + O(h^2)) = hf_{-1} - 3h^2 Df_{-1} + O(h^2)$$

$$k_{-7} = hf\left(t_{-4} + \frac{2}{3}h; x_{-4} + \frac{4}{3}k_{-8} - \frac{2}{3}k_{-9}\right) =$$

$$= hf\left(t_{-1} - \frac{7}{3}h; x_{-1} - \frac{7}{3}hf_{-1} + \frac{49}{18}h^2 Df_{-1} + O(h^3)\right) = hf_{-1} -$$

$$- \frac{7}{3}h^2 Df_{-1} + \frac{49}{18}h^3 D^2 f_{-1} + \frac{49}{18}h^3 f'_{x,-1} \cdot Df_{-1} + O(h^4)$$

$$k_{-6} = hf(t_{-3}, x_{-3}) = hf\left(t_{-1} - 2h; x_{-1} - 2hf_{-1} + 2h^2 Df_{-1} - \frac{4}{3}h^3 D^2 f_{-1} -\right.$$

$$\left. - \frac{4}{3}h^3 f'_{x,-1} Df_{-1} + O(h^4)\right) = hf_{-1} - 2h^2 Df_{-1} + 2h^3 D^2 f_{-1} + 2h^3 f'_{x,-1} Df_{-1} -$$

$$- \frac{4}{3}h^4 D^3 f_{-1} - \frac{4}{3}h^3 f'_{x,-1} D^2 f_{-1} - \frac{4}{3}h^4 f'^2_{x,-1} Df_{-1} + 4h^4 Df'_{x,-1} Df_{-1} + O(h^5)$$

$$k_{-5} = hf\left(t_{-3}+\frac{2}{3}h; x_{-3}+\frac{4}{3}k_{-6}-\frac{2}{3}k_{-7}\right) = hf\left(t_{-1}-\frac{4}{3}h; x_{-1}-\frac{4}{3}hf_{-1}+\right.$$

$$\left.+\frac{8}{9}h^2Df_{-1}-\frac{13}{27}h^3D^2f_{-1}-\frac{13}{27}h^3f'_{x,-1}Df_{-1}+O(h^4)\right) = hf_{-1}-\frac{4}{3}h^2Df_{-1}+$$

$$+\frac{8}{9}h^3D^2f_{-1}+\frac{8}{9}h^3f'_{x,-1}Df_{-1}-\frac{32}{81}h^4D^3f_{-1}-\frac{13}{27}h^4f'_{x,-1}D^2f_{-1}-$$

$$-\frac{13}{27}h^3f'^2_{x,-1}Df_{-1}-\frac{32}{27}h^4Df'_{x,-1}Df+O(h^5)$$

$$k_{-4} = hf(t_{-2}, x_{-2}) = hf\left(t_{-1}-h; x_{-1}-hf_{-1}+\frac{h^2}{2}Df_{-1}-\frac{h^3}{6}D^2f_{-1}-\right.$$

$$\left.-\frac{h^3}{6}f'_{x,-1}Df_{-1}+O(h^4)\right) = hf_{-1}-h^2Df_{-1}+\frac{h^3}{2}D^2f_{-1}+\frac{h^3}{2}f'_{x,-1}Df_{-1}-$$

$$-\frac{h^4}{6}D^3f_{-1}-\frac{h^4}{6}f'_{x,-1}D^2f_{-1}-\frac{h^3}{6}f'^2_{x,-1}Df_{-1}-\frac{h^4}{2}Df'_{x,-1}Df_{-1}+O(h^5);$$

$$k_{-3} = hf\left(t_{-2}+\frac{2}{3}h; x_{-2}+\frac{4}{3}k_{-4}-\frac{2}{3}k_{-5}\right) = hf\left(t_{-1}-\frac{1}{3}h; x_{-1}-\right.$$

$$\left.-\frac{1}{3}hf_{-1}+\frac{1}{18}h^2Df_{-1}-\frac{5}{54}h^3D^2f_{-1}-\frac{5}{54}h^3f'_{x,-1}Df_{-1}+O(h^4)\right) = hf_{-1}-$$

$$-\frac{1}{3}h^2Df_{-1}+\frac{1}{18}h^3D^2f_{-1}+\frac{1}{18}h^3f'_{x,-1}Df_{-1}-\frac{1}{162}h^4D^3f_{-1}-\frac{5}{54}h^4f'_{x,-1}D^2f_{-1}-$$

$$-\frac{5}{54}h^4f'^2_{x,-1}Df_{-1}-\frac{1}{54}h^4Df'_{x,-1}Df_{-1}+O(h^5)$$

$$k_{-2} = hf(t_{-1}, x_{-1}) = hf_{-1}$$

$$k_{-1} = hf\left(t_{-1}+\frac{2}{3}h; x_{-1}+\frac{2}{3}hf_{-1}+\frac{2}{9}h^2Df_{-1}-\frac{1}{27}h^3D^2f_{-1}-\right.$$

$$\left.-\frac{1}{27}h^3f'_{x,-1}Df_{-1}+O(h^4)\right) = hf_{-1}+\frac{2}{3}h^2Df_{-1}+\frac{2}{9}h^3D^2f_{-1}+\frac{2}{9}h^3f'_{x,-1}Df_{-1}+$$

$$+\frac{4}{81}h^4D^3f_{-1}-\frac{1}{27}h^4f'_{x,-1}Df_{-1}-\frac{1}{27}h^4f'^2_{x,-1}Df_{-1}+\frac{4}{27}h^4Df'_{x,-1}Df_{-1}+O(h^5)$$

$$k_0 = hf(t_0, x_0) = hf\left(t_0, +h; x_{-1}+hf_{-1}+\frac{h^2}{2}Df_{-1}+\frac{h^3}{6}D^2f_{-1}+\right.$$

$$\left.+\frac{h^3}{6}f'_{x,-1}Df_{-1}+O(h^4)\right) = hf_{-1}+h^2Df_{-1}+\frac{h^3}{2}D^2f_{-1}+\frac{h^3}{2}f'_{x,-1}Df_{-1}+$$

$$+\frac{h^4}{6}D^3f_{-1}+\frac{h^4}{6}f'_{x,-1}D^2f_{-1}+\frac{h^4}{6}f'^2_{x,-1}Df_{-1}+\frac{h^4}{2}Df'_{x,-1}Df_{-1}+O(h^5)$$

$$k_1 = hf\left(t_0+\frac{2}{3}h; x_0+\frac{4}{3}k_0-\frac{2}{3}k_{-1}\right) = hf\left(t_{-1}+\frac{5}{3}h; x_{-1}+\frac{5}{3}hf_{-1}+\right.$$

$$\left.+\frac{25}{18}h^2Df_{-1}+\frac{37}{54}h^3D^2f_{-1}+\frac{37}{54}h^3f'_{x,-1}Df_{-1}+O(h^4)\right) =$$

$$= hf_{-1}+\frac{5}{3}h^2Df_{-1}+\frac{25}{18}h^3D^2f_{-1}+\frac{25}{18}h^3f'_{x,-1}Df_{-1}+\frac{125}{162}h^4D^3f_{-1}+$$

$$+\frac{37}{54}h^4f'_{x,-1}D^2f_{-1}+\frac{37}{54}h^4f'^2_{x,-1}Df_{-1}+\frac{125}{54}h^4Df'_{x,-1}Df_{-1}+O(h^5).$$

Die gesuchte Formel hat die Gestalt

$$k_3(2h) = S_{-4}k_{-4}+S_{-3}k_{-3}+S_{-2}k_{-2}+S_{-1}k_{-1}+S_0k_0+S_1k_1,$$

mit

$$\begin{aligned}
&S_{-4}+ && S_{-3}+S_{-2}+ && S_{-1}+ && S_0+ && S_1 = 2\\
&-S_{-4}-\frac{1}{3} && S_{-3} && +\frac{2}{3} && S_{-1}+ && S_0+\frac{5}{3} && S_1 = 2\\
&\frac{1}{2}S_{-4}+\frac{1}{18} && S_{-3} && +\frac{2}{9} && S_{-1}+\frac{1}{2}S_0+\frac{25}{18} && S_1 = \frac{4}{3}\\
&-\frac{1}{6}S_{-4}-\frac{1}{162} && S_{-3} && +\frac{4}{81} && S_{-1}+\frac{1}{6}S_0+\frac{125}{162} && S_1 = \frac{16}{27}\\
&-\frac{1}{6}S_{-4}-\frac{5}{54} && S_{-3} && -\frac{1}{27} && S_{-1}+\frac{1}{6}S_0+\frac{37}{54} && S_1 = -\frac{4}{9}\\
&-\frac{1}{2}S_{-4}-\frac{1}{54} && S_{-3} && +\frac{4}{27} && S_{-1}+\frac{1}{2}S_0+\frac{125}{54} && S_1 = \frac{16}{9},
\end{aligned}$$

folglich

$$k_3(2h) = -\frac{25}{12}k_{-4}+\frac{51}{4}k_{-3}-\frac{40}{3}k_{-2}+0\cdot k_{-1}+\frac{65}{12}k_0-\frac{3}{4}k_1. \tag{37}$$

Wir betrachten jetzt eine andere Gruppe von Formeln, wo wir mit Hilfe von $\partial f/\partial x$ bzw. von Df eine solche Hilfsgrößengruppe aufbauen, welche alle, auch Glieder höherer Ordnung in h enthalten. Wir geben hier unten den Grundgedanken und einige Beispiele an, wie man diese Formeln im allgemeinen konstruieren kann. (Hier kann man natürlich auch diejenigen Modifikationen benützen, welche wir früher eingeführt haben, jedoch beschäftigen wir uns jetzt mit dieser Frage nicht.) Die erste und zweite Formel sind Formeln vierter, die dritte aber fünfter Ordnung.

Es sei also erst

$$\begin{aligned}
k_0 &= hf_0+a_{12}h^2Df_0+a_{13}h^3f'_{x,0}Df_0+a_{14}h^4f'^2_{x,0}Df_0\\
k_0^* &= E_1hf_0+b_{12}h^2Df_0+b_{13}h^3f'_{x,0}Df_0+b_{14}h^4f'_{x,0}Df_0\\
k_1 &= hf_1+a_{22}h^2Df_1+a_{23}h^3f'_{x,1}Df_1+a_{24}h^4f'^2_{x,1}Df_1\\
k &= R_0k_0+R_1k_1,
\end{aligned} \tag{38}$$

wo Index 1 die Stelle $(t_0+E_1h;\ x_0+k_0^*)$ bedeutet. Die Entwicklung von f bzw. von $\partial f/\partial x$ an Stelle 1 in bezug auf Stelle 0 ist schon bekannt, diejenige von Df kann man aber gleich mit Hilfe der Formel (16) bzw. mit Hilfe der Sätze 1—5 angeben. Es gilt nämlich

$$Df(t_0+E_1h;\ x_0+E_1hf_0+\Theta) = Df_0+E_1hD(Df)_0+\frac{\partial}{\partial x}(Df)_0\cdot(\Theta)+$$

$$+\frac{E_1^2h^2}{2}D^2(Df)_0+E_1hD\left(\frac{\partial}{\partial x}Df\right)_0\cdot(\Theta)+\frac{1}{2}\frac{\partial^2}{\partial x^2}(Df)_0(\Theta)(\Theta)+\dots$$

$$\dots = Df_0+E_1h(D^2f_0+f'_{x,0}Df_0)+(Df'_x+f'^2_x)\cdot(\Theta)+$$

$$+\frac{E_1^2h^2}{2}(D^3f_0+2Df'_{x,0}Df+f'_{x,0}D^2f_0)+E_1h[D^2f'_x+f''_{x,x}\cdot(Df)+f'_{x,0}Df'_{x,0}+$$

$$+Df'_{x,0}f'_{x,0}]\cdot(\Theta)+\frac{1}{2}[Df''_{xx}+2f''_{xx}\cdot f'_x+f'_x\cdot f''_{xx}]|_0\cdot(\Theta)\cdot(\Theta)+\dots$$

Man bekommt so die Formel vierter Ordnung

$$k_0 = hf_0+\frac{1}{6}h^2Df_0+\frac{1}{60}h^3f'_{x,0}Df_0+\frac{1}{15}h^4f'^2_{x,0}Df,$$

$$k_0^* = hf_0+\frac{2}{5}h^2Df_0,$$

$$(t_1,x_1) = (t_0+h;\ x_0+k_0^*), \tag{39}$$

$$k_1 = hf_1-\frac{1}{6}h^2Df_1+\frac{1}{12}h^3f'_{x,1}Df_1,$$

$$k_4(h) = \frac{1}{2}(k_0+k_1).$$

Man will jedoch $\partial f/\partial x$ bzw. Df so selten berechnen, wie nur möglich, da diese Termen im allgemeinen sehr schwer zu berechnen sind. Man kann nämlich eine Formel vierter Ordnung — und in jeder Hilfsgröße auch mindestens zweiter Ordnung in h — folgendermaßen aufbauen:

$$k_0 = hf_0+\frac{1}{18}h^2Df_0-\frac{3}{11}h^3f'_{x,0}Df_0;$$

$$k_0^* = \frac{3}{4}hf_0+\frac{9}{32}h^2Df_0;$$

$$(t_1,x_1) = \left(t_0+\frac{3}{4}h;\ x_0+k_0^*\right); \tag{40}$$

$$k_1 = hf_1+\frac{1}{18}h^2Df_0+\frac{1}{4}h^2f'_{x,0}(f_1-f_0);$$

$$k_4(h) = \frac{11}{27}k_0+\frac{16}{27}k.$$

Wir geben zuletzt eine Formel fünfter Ordnung auch an. Es sei also

$$k_0 = hf_0 + a_{12}h^2 Df_0 + a_{13}h^3 f'^2_{x,0} Df_0 + a_{14}h^4 f'^2_{x,0} Df_0 + a_{15}h^5 f'^3_{x,0} Df_0;$$

$$k_0^* = E_1 hf_0 + b_{12}h^2 Df_0 + b_{13}h^3 f'_{x,0} Df_0;$$

$$(t_1, x_1) = (t_0 + E_1 h; x_0 + k_0^*);$$

$$k_1 = hf_1 + a_{12}h^2 Df_0 + a_{23}h^2 f'_{x,0}(f_1 - f_0) + a_{24}h^3 f'^2_{x,0}(f_1 - f_0);$$

$$k_1^* = (E_{21}f_0 + E_{22}f_1)h + b_{22}h^2 Df_0 + b_{23}h^2 f'_{x,0} \cdot (f_1 - f_0);$$

$$(t_2, x_2) = (t_0 + (E_{21} + E_{22})h; x_0 + k_1^*);$$

$$k_2 = hf_2 + a_{12}h^2 Df_0 + a_{33}h^2 f'_{x,0}(f_1 - f_0) + c_{33}h^2 f'_{x,0}(f_2 - f_0);$$

$$k_5(h) = S_1 k_0 + S_2 k_1 + S_3 k_2.$$

Hier ist

$$hf_1 = hf_0 + E_1 h^2 Df_0 + (b_{12}h^3 f'_{x,0} Df_0 + b_{13}h^4 f'^2_{x,0} Df_0) + \frac{E_1^2}{2} h^3 D^2 f_0 +$$

$$+ (E_1 b_{12} h^4 Df'_{x,0} Df_0 + E_1 b_{13} h^5 Df'_{x,0} \cdot f'_{x,0} Df_0) + \frac{1}{2} b_{12}^2 f''_{xx,0} (Df_0)^2 +$$

$$+ \ldots + \frac{E_1^3}{6} h^4 D^3 f + \left(\frac{E_1^2}{2} b_{12} h^5 D^2 f'_{x,0} \cdot Df_0 + \ldots \right) + \frac{E_1^4}{24} h^5 D^4 f_0 + \ldots,$$

ferner (mit der Verkürzung $E_2 = E_{21} + E_{22}$):

$$x_2 = x_0 + E_2 hf_0 + h^2 (E_1 E_{22} + b_{22}) Df_0 + h^3 \frac{E_1^2}{2} E_{22} D^2 f_0 +$$

$$+ h^3 (b_{12} E_{22} + b_{23} E_1) f'_{x,0} Df_0 + h^4 \frac{E_1^3}{6} E_{22} D^3 f_0 +$$

$$+ h^4 (b_{13} E_{22} + b_{23} b_{12}) f'^2_{x,0} Df_0 + h^4 b_{23} \frac{E_1^2}{2} f'_{x,0} D^2 f_0 + h^4 E_1 b_{12} E_{22} Df'_x D + \ldots,$$

folglich

$$hf_2 = hf_0 + E_2 h^2 Df_0 + h^3 (E_1 E_{22} + b_{22}) f'_{x,0} Df_0 + h^4 \frac{E_1^2}{2} E_{22} f'_{x,0} D^2 f_0 +$$

$$+ h^4 (b_{12} E_{22} + b_{23} E_1) f'^2_{x,0} Df_0 + h^5 \frac{E_1^3}{6} E_{22} f'_{x,0} D^3 f + h^5 \frac{E_1^2}{2} b_{23} f'^2_{x,0} D^2 f_0 +$$

$$+ h^5 E_1 b_{12} E_{22} f'_{x,0} Df'_{x,0} Df_0 + h^5 (b_{13} E_{22} + b_{23} b_{12}) f'^3_{x,0} Df_0 + \frac{E_2^2}{2} h^3 D^2 f_0 +$$

$$+ h^4 (E_2 E_1 E_{22} + E_2 b_{22}) Df'_{x,0} Df_0 + h^5 \frac{E_1^2}{2} E_2 E_{22} Df'_{x,0} D^2 f_0 +$$

$$+ h^5 (b_{12} E_2 E_{22} + b_{23} E_1 E_2) Df'_{x,0} \cdot f'_{x,0} Df_0 + \ldots + h^5 \frac{(E_1 E_{22} + b_{22})^2}{2} f''_{xx,0} (Df)_0^2 +$$

$$+ h^4 \frac{E_2^3}{6} D^3 f_0 + h^5 \frac{E_2^2}{2} (E_1 E_{22} + b_{22}) D^2 f'_{x,0} Df_0 + \ldots + h^5 \frac{E_2^4}{24} D^4 f_0 + \ldots$$

Die erste Gruppe unserer Gleichungen, welche die Gleichheit der Koeffizienten von f, Df, D^2f, D^3f, D^4f, Df'_xDf, $D^2f'_xDf\,Df'_xD^2f$ und $f''_{xx}(Df)^2$ ausdrücken, zerfällt nun in zwei Teilgruppen. Die letzte 7 enthalten nämlich S_1 und a_{12} nicht. Diese Gleichungen sind:

$$\frac{1}{6} = S_2\frac{E_1^2}{2}+S_3\frac{E_2^2}{2},$$

$$\frac{1}{24} = S_2\frac{E_1^3}{6}+S_3\frac{E_2^3}{6},$$

$$\frac{1}{120} = S_2\frac{E_1^4}{24}+S_3\frac{E_2^4}{24},$$

$$\frac{1}{8} = S_2E_1b_{12}+S_3E_2(E_1E_{22}+b_{22}),$$

$$\frac{1}{20} = S_2\frac{E_1^2}{2}b_{12}+S_3\frac{E_2^2}{2}(E_1E_{22}+b_{22}),$$

$$\frac{1}{40} = S_2\frac{b_{12}^2}{2}+S_3\frac{(E_1E_{22}+b_{22})^2}{2},$$

$$\frac{1}{30} = S_3\frac{E_1^2}{2}E_{22}E_2$$

Betrachten wir die ersten zwei bzw. die vierte und fünfte Gleichung. Daraus bekommt man einerseits

$$S_2 = \frac{1}{12}\frac{4E_2-3}{E_1^2(E_2-E_1)};\quad S_3 = \frac{1}{12}\frac{3-4E_1}{E_2^2(E_2-E_1)},$$

anderseits, mit Hilfe der Verkürzung $w=E_1E_{22}+b_{22}$,

$$b_{12} = \frac{1}{40}\frac{5E_2-4}{S_2E_1(E_2-E_1)};\quad w = \frac{1}{40}\frac{4-5E_1}{S_3E_2(E_2-E_1)}.$$

Setzt man diese in dritte bzw. sechste Gleichung ein, so folgt einerseits

$$15(E_2+E_1)-20E_1E_2=12,$$

anderseits

$$300E_2^2+620E_1E_2+300E_1^2-945E_1-945E_2+732=0.$$

Kombiniert man diese Gleichungen, so folgt

$$10(E_1+E_2)^2-31(E_1+E_2)+24=0,$$

d. h. $E_1+E_2=\frac{8}{5}$ bzw. $E_1\cdot E_2=\frac{3}{5}$, d. h. $E_1=\frac{3}{5}$; $E_2=1$.

Somit gilt

$$S_2=\left(\frac{5}{6}\right)^3;\quad S_3=\left(\frac{3}{6}\right)^3 \quad \text{und} \quad S_1=\left(\frac{4}{6}\right)^3,$$

ferner

$$b_{12}=\frac{9}{50},\quad w=\frac{1}{2}.$$

Endlich der letzten Gleichung gemäß ist $E_{22}=\frac{40}{27}$, und so

$$E_{21}=-\frac{13}{27},\quad \text{und}\quad b_{22}=-\frac{7}{18}.$$

Betrachtet man die Gleichheit der Koeffizienten von Df, so folgt

$$a_{12}=\frac{1}{36}.$$

Betrachten wir jetzt die Koeffizienten von $f'_{x,0}D^2f_0, f'_{x,0}D^3f_0, f'_{x,0}Df'_{x,0}Df_0$ und $Df'_{x,0}\cdot f'_{x,0}Df_0$ (wählt man die beiden letzteren gleich, d.h. f'_x und Df'_x vertauschbar, so folgt $a_{33}=0$, $c_{33}=0$, $a_{23}=\frac{2}{25}$, $b_{23}=0$, $b_{13}=0$, $a_{24}=\frac{2}{25}$ und $a_{13}=-\frac{3}{32}$, $a_{14}=-\frac{3}{32}$, $a_{15}=0$):

$$\frac{1}{24}=S_2a_{23}\frac{E_1^2}{2}+S_3\left\{\frac{E_1^2}{2}E_{22}+a_{33}\frac{E_1^2}{2}+c_{33}\frac{E_2^2}{2}\right\}$$

$$\frac{1}{120}=S_2a_{23}\frac{E_1^3}{6}+S_3\left\{\frac{E_1^3}{6}E_{22}+a_{33}\frac{E_1^3}{6}+c_{33}\frac{E_2^3}{6}\right\}$$

$$\frac{1}{40}=S_2a_{23}E_1b_{12}+S_3\{E_1b_{12}E_{22}+a_{33}E_1b_{12}+c_{33}(E_1E_2E_{22}+E_2b_{22})\}$$

$$\frac{1}{30}=S_2E_1b_{13}+S_3\{b_{12}E_2E_{22}+b_{23}E_1E_2\}.$$

Daraus folgt

$$b_{23}=0,\quad a_{23}=0,\quad a_{33}=\frac{10}{9},\quad c_{33}=-\frac{4}{25}$$

und

$$b_{13}=0.$$

Zuletzt bekommt man durch die Koeffizienten von $f'_{x,0}{}^{2}_{0}D^2f_0$, $f'_{x,0}Df_0$, $f'_{x,0}{}^{2}_{0}Df_0$ und $f'_{x,0}{}^{3}_{0}Df_0$ die Gleichungen

$$\frac{1}{120} = S_2 a_{24}\frac{E_1^2}{2} + S_3\left(\frac{E_1^2}{2}b_{23} + c_{33}\frac{E_1^2}{2}E_{22}\right)$$

$$\frac{1}{6} = S_1 a_{13} + S_2\{b_{12} + a_{23}E_1\} + S_3\{(E_1E_{22} + b_{22}) + a_{33}E_1 + c_{33}E_2\}$$

$$\frac{1}{24} = S_1 a_{14} + S_2\{b_{13} + a_{23}b_{12} + a_{24}E_1\} + S_3\{(b_{12}E_{22} + b_{23}E_1) + a_{33}b_{12} + c_{33}\cdot(E_1E_{22} + b_{22})\}$$

$$\frac{1}{120} = S_1 a_{15} + S_2(a_{23}b_{13} + a_{24}b_{12}) + S_3\{b_{13}E_{22} + b_{23}b_{12} + a_{33}b_{13} + c_{33}\cdot(b_{12}E_{22} + b_{23}E_1)\},$$

woraus

$$a_{24} = \frac{82}{625}, \quad a_{13} = -\frac{171}{800}, \quad a_{14} = -\frac{141}{800}, \quad a_{15} = 0.$$

Unsere Formel ist also

$$k_0 = hf_0 + \frac{1}{36}h^2Df_0 - \frac{171}{800}h^3f'_{x,0}Df_0 - \frac{141}{800}h^4f'_{x,0}{}^{2}_{0}Df_0;$$

$$k_0^* = \frac{3}{5}hf_0 + \frac{9}{50}h^2Df_0;$$

$$(t_1, x_1) = \left(t_0 + \frac{3}{5}h; x_0 + k_0^*\right)$$

(41)

$$k_1 = hf_1 + \frac{1}{36}h^2Df_0 + 0\cdot h^2f'_{x,0}(f_1 - f_0) + \frac{82}{625}h^3f'_{x,0}{}^{2}_{0}(f_1 - f_0);$$

$$k_1^* = h\left(\frac{40}{27}f_1 - \frac{13}{27}f_0\right) - \frac{7}{18}h^2Df_0;$$

$$(t_2, x_2) = (t_0 + h; x_0 + k_1^*)$$

$$k_2 = hf_2 + \frac{1}{36}h^2Df_0 + \frac{10}{9}h^2f'_{x,0}(f_1 - f_0) - \frac{4}{25}h^2f'_{x,0}(f_2 - f_0);$$

$$k_5(h) = \frac{1}{216}(64k_0 + 125k_1 + 27k_2).$$

Die neuen Formeln sind in jenem Falle sehr gut, falls die erste Ableitung stark steigt.

Als letztes Problem beschäftigen wir uns mit der Umgebung singulärer Stellen, wo eben die in h nichtlinearen, oben angegebenen Formeln die Möglichkeit geben, brauchbare Methoden des Runge—Kuttaschen Typs herzuleiten. Wir beschäftigen

uns jedoch erst mit dem Fall, wo die gesuchte Funktion, $x(t)$ R in R, bzw. K in K abbildet, d.h. wenn $x(t)$ eine Inverse besitzt. Wir können dann die Rolle von x und t vertauschen, d.h. statt der Gleichung

$$\dot{x} = f(t, x) \quad \text{ja} \quad \frac{dt}{dx} = \frac{1}{f(t, x)} = g(x, t)$$

betrachten, mit dem Hilfsoperator $\tilde{D} = \frac{\partial}{\partial x} + g\frac{\partial}{\partial t}$. Ein direktes Vertauschen der Veränderlichen (falls $|f|$ zu groß wird) ist jedoch nicht erwünscht, da man dann $\Delta t = h$ nicht vorschreiben kann (genauer man arbeitet dann so, daß man in der ersten Relation $h_{(0)}$ annimmt, und k aus der Gleichung

$$h_{(0)} = kg_{(0)} + k^2 a_{12} \widetilde{Dg_0} + \dots$$

berechnet, in den weiteren Formeln muß man jedoch schon mit dieser k arbeiten, um $h^*_{(0)}$, h_1 usf. auszuwerten. Die endgültige h wird also eine gewichtete Summe, wo nur der erste Glied vorgeschrieben ist. Unsere Formeln zeigen jedoch schon, daß im Falle, wenn wir Hilfsgrößen erster Ordnung benützen, und $|f_0| = \infty$, d.h. $g_0 = 0$ ist, $h_{(0)}$ nicht vorgeschrieben werden kann. Wenn wir aber Hilfsgrößen höherer Ordnung benützen, und $|f_0| = \infty$, jedoch $|Df_0| < \infty$ ist, so kann man $h_{(0)}$ vorschreiben, und k wird dadurch definiert u. zw. ergibt sich $|k| = O(\sqrt{|h|})$; aus dieser Bemerkung folgt der zu beschreibende Weg.

Um diese Schwierigkeit zuüberwinden, gebrauchen wir entweder die triviale Formel zweiter Ordnung

$$h = kg_0 + \frac{1}{2} k^2 \widetilde{Dg_0} \tag{42}$$

$\left(\text{d.h. } k = -\frac{g_0}{2} + \sqrt{\frac{g_0^2}{4} - \frac{1}{2} h\widetilde{Dg_0}}\right)$, oder aber tauschen wir die Rolle von x und t in indirekter Weise auf. Es bedeutet, daß wir eine der angegebenen direkten Formeln höherer Ordnung betrachten, und dort erst die Ausdrücke von f mit jenen von g vertauschen, und danach die Formeln — jedoch mit Gliedern höherer Ordnung entsprechend ergänzt — invertieren. Hier gebraucht man also erst die Relationen

$$f_i = \frac{1}{g_i}; \quad \frac{\partial}{\partial x} f_i = -\frac{1}{g_i^2}\frac{\partial}{\partial x} g_i; \quad Df_i = -\frac{1}{g_i^3}\widetilde{Dg_i}. \tag{43}$$

Es bedeutet also, daß wir erst die Formel

$$k^*_{(i)} = \alpha_{i1} hf_i + \alpha_{i2} h^2 Df_i + \alpha_{i3} h^3 f'_{x,i} Df_i + \dots \tag{44}$$

$$\text{bzw.} = h(\alpha^{(0)}_{i1} f_0 + \alpha^{(1)}_{i1} f_1 + \dots + \alpha^{(j)}_{i1} f_j) + \alpha_{i2} h^2 Df_0 + \alpha^{(1)}_{i3} h^2 f'_{x,0}(f_1 - f_0) + \dots$$

in

$$k^*_{(i)} = \alpha_{i1} h \frac{1}{g_i} - \alpha_{i2} h^2 \frac{\widetilde{Dg_i}}{g_i^3} + \alpha_{i3} h^3 \frac{g'_{x,i}\widetilde{Dg_i}}{g_i^5} + \dots$$

$$\text{bzw.} = h\left(\alpha^{(0)}_{i1}\frac{1}{g_0} + \alpha^{(1)}_{i1}\frac{1}{g_1} + \dots + \alpha_{i2} h^2 \frac{\widetilde{Dg_0}}{g_0^3} + \alpha^{(1)}_{i3} h^2 \frac{g'_{x,0}}{g_0^2}\left(\frac{1}{g_1} - \frac{1}{g_0}\right) + \dots\right)$$

überführen. Im weiteren setzen wir voraus, daß im Fall, wenn wir solche Formeln benützen, wo verschiedene Indexe auftreten, die Singularität in, oder in der Nähe von (t_0, x_0) liegt; dann ordnen wir (44) so um, daß es die Form

$$\frac{k_i^*}{h} = a_1 \cdot \frac{1}{g^{(i)}} + a_2 \frac{1}{g^{(i)3}} h + a_3 \frac{1}{g^{(i)5}} h^2 + a_4 \frac{1}{g^{(i)7}}^3 + \dots \tag{45}$$

besitze, wo die Bedeutung von $g^{(i)}$: $g^{(i)} = g_0$, wenn Index 0 in der Formel vorkommt, und $g^{(i)} = g_i$, wenn nicht. Es sei ferner vorausgesetzt, daß die betrachtete Grundformel n-ter Ordnung ist. Die ersten zwei Glieder in (45) sind nun Anfangsglieder der Entwicklung nach h von $a_1 \left(g^{(i)2} - 2\frac{a_2}{a_1} h\right)^{-\frac{1}{2}}$. Folglich (für genügend kleine h):

$$\frac{k_i^*}{h} - \frac{a_1}{\sqrt{g^{(i)2} - 2\frac{a_2}{a_1} h}} = \left(a_3 - \frac{3a_2^2}{2a_1}\right) \frac{h^2}{g^{(i)5}} + \left(a_4 - \frac{5a_2^3}{2a_1^2}\right) \frac{h^3}{g^{(i)7}} +$$

$$+ \left(a_5 - \frac{35}{8} \frac{a_2^4}{a_1^3}\right) \frac{h^4}{g^{(i)9}} + \dots$$

Die ersten zwei Glieder sind nun wieder Anfangsglieder der Entwicklung von

$$\left(a_3 - \frac{3a_2^2}{2a_1}\right) h^2 \cdot \left(g^{(i)2} - \frac{2}{5} \cdot \frac{a_4 - \frac{5a_2^3}{2a_1^2}}{a_3 - \frac{3a_2^2}{2a_1}} h \right)^{-\frac{5}{2}}.$$

Damit

$$\frac{k_i^*}{h} - \frac{a_1}{\sqrt{g^{(i)2} - 2\frac{a_2}{a_1} h}} - \frac{\left(a_3 - \frac{3}{2}\frac{a_2^2}{a_1}\right) h^2}{\sqrt{\left(g^{(i)2} - \frac{2a_4 - 5\frac{a_2^3}{a_1^2}}{5a_3 - \frac{15}{2}\frac{a_2^2}{a_1}} h \right)^5}} = \dots .$$

Das Verfahren setzen wir so weit fort, daß die Glieder höchstens $(n-1)$-ten Grades in h rechts abfallen. Dann setzen wir die linke Seite gleich 0, woraus sich

$$k_i^* = \frac{a_1 h}{\sqrt{g^{(i)2} - 2\frac{a_2}{a_1} h}} + \frac{\frac{2a_1 a_3 - 3a_2^2}{2a_1} h^3}{\left(g^{(i)2} - 2\frac{2a_4 a_1^2 - 5a_2^3}{10a_3 a_1^2 - 15a_2^2 a_1} h \right)^{5/2}} + \dots \tag{46}$$

ergibt. Das Problem der Pünktlichkeit von (46) bleibt offen; (46) ist denn theoretisch nur für solche h mit (44) äquivalent, welche in Hinsicht von $|g^{(i)}|$ genug klein sind. Setzt man jedoch den Wert von $a_1, a_2, \dots$ in (46) ein, wählt man ferner $g^{(i)} = 0$, und betrachtet man nur das erste Glied an der rechten Seite von (46), so bekommt

man dieselbe Relation, die man auch durch direktes Vertauschen der Rolle von t und x mit Hilfe einer Formel zweiter Ordnung bekommen kann. (46) kann man also wie assymptotische Formel in der Nähe von singulären Stellen betrachten.

Es bleibt noch die Frage zu beantworten, wie man in der Nähe von singulären Stellen numerisch arbeiten kann, wenn x keine skalare Veränderliche ist, also wenn es keine Inverse Funktion $t=t(x)$ gibt. In solchen Fällen kann man jedoch ein Hilfsparameter τ mit Hilfe einer Relation

$$\frac{dt}{d\tau} = \gamma(\|f(x,t)\|) \tag{47}$$

so einführen, daß

$$\frac{dx}{d\tau} = \gamma(\|f(x,t)\|)\cdot f(x,t) \tag{48}$$

schon keine Singularität besitze. Dann existiert die inverse Funktion $\tau=\tau(t)$, und in (47) kann man die oben eingeführten Methoden anwenden.

§ 4. Anwendungen

Die Methode ist in allen Fällen, wo man die Newtonsche Methode anwendet, brauchbar, aber auch dann, wenn wir keine „erste Annäherung" kennen, da die Schrittweite der benützten Formel des RUNGE—KUTTAschen Typs sich automatisch durch die Pünktlichkeitserforderungen regulieren läßt. Wir weisen hier auf [2], wo man viele allgemeine, mit Hilfe der NEWTON—RAPHSONschen Methode behandelte Beispiele findet. Eben deshalb geben wir hier nur zwei neue und wichtige Anwendungsgebiete an.

Betrachten wir zuerst ein lineares Differentialgleichungssystem mit periodischen Koeffizientenmatrizen der Form

$$\dot{\mathbf{x}} = \mathbf{P}(t)\cdot\mathbf{x}. \tag{49}$$

Es ist wohlbekannt, daß man ein Grundsystem von (49) in der Form

$$\mathbf{X}(t) = \mathbf{Q}(t)\cdot e^{\mathbf{I}\cdot t} \tag{50}$$

darstellen kann, mit periodischer $\mathbf{Q}$ und konstanter $\mathbf{I}$. Die explizite Form für $\mathbf{Q}$ und $\mathbf{I}$ kann man jedoch fast nie darstellen. Wir betrachten eben deswegen erst das inverse Problem: es sei $\mathbf{Q}(t)$ und $\mathbf{I}$ angegeben, und man soll $\mathbf{P}(t)$ finden. Nun

$$\frac{d}{dt}\mathbf{X}(t) = \dot{\mathbf{X}}(t) = \dot{\mathbf{Q}}(t)\cdot e^{\mathbf{I}t} + \mathbf{Q}(t)\cdot\mathbf{I}e^{\mathbf{I}\cdot t} = [\dot{\mathbf{Q}}(t)\cdot\mathbf{Q}^{-1}(t) + \mathbf{Q}(t)\cdot\mathbf{I}\cdot\mathbf{Q}^{-1}(t)]\cdot\mathbf{Q}(t)\cdot e^{\mathbf{I}t}$$

$$= [\dot{\mathbf{Q}}(t)\cdot\mathbf{Q}^{-1}(t) + \mathbf{Q}(t)\cdot\mathbf{I}\cdot\mathbf{Q}^{-1}(t)]\cdot\mathbf{X}(t), \tag{51}$$

folglich (50) mit

$$\mathbf{P}(t) = [\dot{\mathbf{Q}}(t)\mathbf{Q}^{-1}(t) + \mathbf{Q}(t)\cdot\mathbf{I}\cdot\mathbf{Q}^{-1}(t)] \tag{52}$$

(49) genügt.

Wir zeigen nun, daß $\mathbf{Q}$ bzw. $\mathbf{I}$ differenzierbare Funktionen von $\mathbf{P}$ sind, und geben zugleich die Frechetsche Ableitung an. Zu diesem Zweck schreiben wir (52) in der Form

$$\dot{\mathbf{Q}}(t)+\mathbf{Q}(t)\cdot\mathbf{I}-\mathbf{P}(t)\cdot\mathbf{Q}(t)\equiv\mathbf{O} \tag{53}$$

bzw. die Veränderungen in der Form

$$\begin{aligned}\Delta\dot{\mathbf{Q}}(t)+\Delta\mathbf{Q}(t)\cdot\mathbf{I}-\mathbf{P}(t)\cdot\Delta\mathbf{Q}(t) &= -\mathbf{Q}(t)\cdot\Delta\mathbf{I}+\Delta\mathbf{P}(t)\cdot\mathbf{Q}(t)\\ &\quad -\Delta\mathbf{Q}(t)\cdot\Delta\mathbf{I}+\Delta\mathbf{P}(t)\cdot\Delta\mathbf{Q}(t).\end{aligned} \tag{54}$$

Wir zeigen nun, daß in (54) die zwei letzten Glieder eine Größenordnung $\|\Delta\mathbf{P}(t)\|_t^2$ besitzt, wo $\|\ \|_t$ mit Hilfe einer entsprechenden Matrixnorm $\|,\|$ die Größe

$$\sup_{t_0\leqq t\leqq t_0+T}\|\Delta\mathbf{P}(t)\| = \|\Delta\mathbf{P}(t)\|_t, \tag{55}$$

ferner T die Periodenlänge bedeutet. (54) ist nämlich eine lineare Gleichung für $\Delta\mathbf{Q}(t)$, so geordnet, daß der links stehende „homogene" Ausdruck mit (53) äquivalent ist; $\Delta\mathbf{I}$ ist ferner dadurch charakterisiert, daß die Lösung von (54) periodisch sein muß. Der Hauptteil von (54) besitzt also die Lösung

$$\begin{aligned}\Delta\mathbf{Q}^{(1)}(t) &= \int_{t_0}^{t}\mathbf{Q}(t)\mathbf{Q}^{-1}(\tau)\{-\mathbf{Q}(\tau)\Delta\mathbf{I}^{(1)}+\Delta\mathbf{P}(\tau)\cdot\mathbf{Q}(\tau)\}d\tau =\\ &= \mathbf{Q}(t)\int_{t_0}^{t}\{\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau)-\Delta\mathbf{I}^{(1)}\}d\tau.\end{aligned} \tag{56}$$

Da nun $\Delta\mathbf{Q}^{(1)}(t)$ auch periodisch sein muß, so folgt

$$\int_{t_0}^{t_0+T}\{\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau)-\Delta\mathbf{I}^{(1)}\}d\tau = 0,$$

d.h.

$$\Delta\mathbf{I}^{(1)} = \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau)d\tau. \tag{57}$$

Ist also die gewählte Matrixnorm gegen die Ähnlichkeitstransformation invariant, so gilt

$$\|\Delta\mathbf{I}^{(1)}\|\leqq\|\Delta\mathbf{P}(t)\|_t, \tag{58}$$

und somit

$$\|\Delta\mathbf{Q}^{(1)}(t)\|_t\leqq\|\mathbf{Q}(t)\|_t\cdot T\cdot 2\|\Delta\mathbf{P}(t)\|_t = K\cdot\|\Delta\mathbf{P}(t)\|_t. \tag{59}$$

Wählen wir nun eine zweite Annäherung von $\Delta\mathbf{Q}$ bzw. $\Delta\mathbf{I}$ durch

$$\begin{aligned}\Delta\mathbf{Q}^{(2)} = \mathbf{Q}(t)\cdot\int_{t_0}^{t}\{&\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau)-\mathbf{Q}^{-1}(\tau)\Delta\mathbf{Q}^{(1)}(\tau)\Delta\mathbf{I}^{(1)}-\\ &-\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^{(1)}(\tau)-\Delta\mathbf{I}^{(2)}\}d\tau,\end{aligned} \tag{60}$$

wo man $\Delta\mathbf{I}^{(2)}$ wieder so wählt, daß $\Delta\mathbf{Q}^{(2)}$ periodisch sei. Folglich

$$\Delta\mathbf{I}^{(2)} = \frac{1}{T}\int_{t_0}^{t_0+T} \{\mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau) - \mathbf{Q}^{-1}(\tau)\Delta\mathbf{Q}^{(1)}(\tau)\Delta\mathbf{I}^{(1)} - $$

$$+ \mathbf{Q}^{-1}(\tau)\Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^{(1)}(\tau)\}\,d\tau,$$

d.h.

$$\Delta\mathbf{I}^{(2)} - \Delta I^{(1)} = \frac{1}{T}\int_{t_0}^{t_0+T} \mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^{(1)}(\tau) - \Delta\mathbf{Q}^{(1)}(\tau)\Delta\mathbf{I}^{(1)}\}\,d\tau,$$

folglich

$$\|\Delta\mathbf{I}^{(2)} - \Delta\mathbf{I}^{(1)}\| \leqq \frac{1}{T}\cdot T\cdot\|\mathbf{Q}^{-1}(t)\|_t\cdot K\cdot\|\Delta\mathbf{P}(t)\|_t^2 = K_1\cdot\|\Delta P(t)\|_t^2. \tag{61}$$

Daraus folgt

$$\Delta\mathbf{Q}_2 - \Delta\mathbf{Q}_1 = \mathbf{Q}(t)\int_{t_0}^{t}\langle\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^{(1)}(\tau) - \Delta\mathbf{Q}^{(1)}(\tau)\Delta\mathbf{I}^{(1)}\} -$$

$$-\{\Delta\mathbf{I}^{(2)} - \Delta\mathbf{I}^{(1)}\}\rangle\,d\tau \tag{62}$$

folglich

$$\|\Delta\mathbf{Q}_2 - \Delta\mathbf{Q}_1\| \leqq T\cdot\|\mathbf{Q}(t)\|_t\cdot\|\mathbf{Q}^{-1}(t)\|_t\{2K + K_1\}\cdot\|\Delta\mathbf{P}(t)\|_t^2 \leqq K_1\{2K + K_1\}\,\|\Delta\mathbf{P}(t)\|_t^2.$$

Nun ebenso

$$\Delta\mathbf{Q}^{(3)}(t) = \mathbf{Q}(t)\int_{t_0}^{t}\langle\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau) - \Delta\mathbf{Q}^{(2)}(\tau)\Delta\mathbf{I}^{(2)} +$$

$$+\Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^{(2)}(\tau)\} - \Delta\mathbf{I}^{(3)}\rangle\,d\tau$$

und

$$\Delta\mathbf{I}^{(3)} = \int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau) - \Delta\mathbf{Q}^{(2)}(\tau)\Delta\mathbf{I}^{(2)} + \Delta\mathbf{P}(\tau)\Delta\mathbf{Q}^2(\tau)\}\,d\tau,$$

d. h.

$$\Delta\mathbf{I}^{(3)} - \Delta\mathbf{I}^{(2)} = \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{Q}^{(1)}(\tau)\Delta\mathbf{I}^{(1)} - \Delta\mathbf{Q}^{(2)}\Delta\mathbf{I}^{(2)} + \Delta\mathbf{P}(\tau)[\Delta\mathbf{Q}^{(2)}(\tau) -$$

$$-\Delta\mathbf{Q}^{(1)}(\tau)]\}\,d\tau = \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)[\Delta\mathbf{Q}^2(\tau) - \Delta\mathbf{Q}^{(1)}(\tau)] - [\Delta\mathbf{Q}^{(2)}(\tau) -$$

$$-\Delta\mathbf{Q}^{(1)}(\tau)]\Delta\mathbf{I}^{(1)} - \Delta\mathbf{Q}^{(2)}(\tau)\cdot[\Delta\mathbf{I}^{(2)} - \Delta\mathbf{I}^{(1)}]\}\,d\tau,$$

folglich

$$\|\Delta\mathbf{I}^{(3)} - \Delta\mathbf{I}^{(2)}\| \leqq \|\mathbf{Q}^{-1}(t)\|_t\cdot\|\Delta\mathbf{P}(t)\|_t^3\cdot\{K_1\{2K_2 + K_1\} + K_1\{2K + K_1\} +$$

$$+2K\cdot K_1\} \leqq 2\{2K + K_1\}^2\,\|\Delta\mathbf{P}(t)\|_t^3,$$

falls nur $\|\Delta\mathbf{P}(t)\|_t$ genug klein — in Hinsicht von K und K_1 — ist. Damit

$$\Delta\mathbf{Q}^{(3)} - \Delta\mathbf{Q}^{(2)} = \mathbf{Q}(t)\int_{t_0}^{t}\langle\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{Q}^{(1)}(\tau)\,\Delta\mathbf{I}^{(1)} - \Delta\mathbf{Q}^{(2)}(\tau)\,\Delta\mathbf{I}^{(2)} +$$

$$+\Delta\mathbf{P}(t)[\Delta\mathbf{Q}^{(2)}(\tau) - \Delta\mathbf{Q}^{(1)}(\tau)]\} - [\Delta\mathbf{I}^{(3)} - \Delta\mathbf{I}^{(2)}]\}\,d\tau,$$

folglich

$$\|\Delta\mathbf{Q}^{(3)} - \Delta\mathbf{Q}^{(2)}\|_t \leqq \|\mathbf{Q}(t)\|_t \cdot T \cdot \|\mathbf{Q}^{-1}(t)\| \cdot \|\Delta\mathbf{P}(t)\|_t^3\{K_1\cdot(2K+K_1) +$$

$$+2KK_1 + 2\{2K+K_1\}^2\} \leqq 2\{2K+K_1\}^3 \cdot \|\Delta\mathbf{P}(t)\|_t^3.$$

Es sei also vorausgesetzt, daß die Abschätzung

$$\|\Delta\mathbf{I}^{(n+1)} - \Delta\mathbf{I}^{(n)}\| \leqq n\{2K+K_1\}^n \cdot \|\Delta\mathbf{P}(t)\|_t^{n+1}, \tag{62}$$

bzw.

$$\|\Delta\mathbf{Q}^{(n+1)} - \Delta\mathbf{Q}^{(n)}\| \leqq n\{2K+K_1\}^{n+1} \cdot \|\Delta\mathbf{P}(t)\|_t^{n+1} \tag{63}$$

für $n = N-1$ schon bewiesen sind. Dann gilt

$$\Delta\mathbf{Q}^{(N+1)}(t) = \mathbf{Q}(t)\cdot\int_{t_0}^{t}\langle\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau) - \Delta\mathbf{Q}^{(N)}(\tau)\,\Delta\mathbf{I}^{(N)} +$$

$$+\Delta\mathbf{P}(\tau)\,\Delta\mathbf{Q}^{(N)}(\tau)\} - \Delta\mathbf{I}^{(N+1)}\rangle\,d\tau, \tag{64}$$

ferner um $\Delta\mathbf{Q}^{(N+1)}(t)$ periodisch zu machen

$$\Delta\mathbf{I}^{(N+1)} = \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)\mathbf{Q}(\tau) - \Delta\mathbf{Q}^{(N)}(\tau)\,\Delta\mathbf{I}^{(N)} + \Delta\mathbf{P}(\tau)\,\Delta\mathbf{Q}^{(N)}(\tau)\}\,d\tau. \tag{65}$$

Da nun hier

$$\Delta\mathbf{I}^{(N+1)} - \Delta\mathbf{I}^{(N)} = \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{Q}^{(N-1)}(\tau)\,\Delta\mathbf{I}^{(N-1)} - \Delta\mathbf{Q}^{(N)}(\tau)\,\Delta\mathbf{I}^{(N)} +$$

$$+\Delta\mathbf{P}(\tau)[\Delta\mathbf{Q}^{(N)}(\tau) - \Delta\mathbf{Q}^{(N-1)}(\tau)]\}\,d\tau =$$

$$= \frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{P}(\tau)[\Delta\mathbf{Q}^{(N)}(\tau) - \Delta\mathbf{Q}^{(N-1)}(\tau)] - [\Delta\mathbf{Q}^{(N)}(\tau) - \Delta\mathbf{Q}^{(N-1)}(\tau)]\cdot$$

$$\cdot\,\Delta\mathbf{I}^{(N-1)} - \Delta\mathbf{Q}^{(N)}(\tau)\cdot[\Delta\mathbf{I}^{(N)} - \Delta\mathbf{I}^{(N-1)}]\}\,d\tau,$$

so folgt

$$\|\Delta\mathbf{I}^{(N+1)} - \Delta\mathbf{I}^{(N)}\| \leqq \frac{1}{T}\cdot T\cdot\|\mathbf{Q}^{-1}(t)\|_t\cdot\|\Delta\mathbf{P}(t)\|_t^{(N+1)}\{(N-1)\{2K+K_1\}^N\cdot(1+2K) +$$

$$+2K_1(2K+K_1)^{N-1}\} \leqq N\{2K+K_1\}^{N+1}\cdot\|\Delta\mathbf{P}(t)\|_t^{N+1},$$

d.h. (62) ist auch für $n=N$ gültig. Ebenso

$$\Delta\mathbf{Q}^{(N+1)}(t)-\Delta\mathbf{Q}^{(N)}(t)=\mathbf{Q}(t)\cdot\int_{t_0}^{t}\langle\mathbf{Q}^{-1}(\tau)\{\Delta\mathbf{Q}^{(N-1)}(\tau)\,\Delta\mathbf{I}^{(N-1)}-$$

$$-\Delta\mathbf{Q}^{(N)}(\tau)\,\Delta\mathbf{I}^{(N)}+\Delta\mathbf{P}(\tau)[\Delta\mathbf{Q}^{(N)}(\tau)-\Delta\mathbf{Q}^{(N-1)}(\tau)]\}-[\Delta\mathbf{I}^{(N+1)}-\Delta\mathbf{I}^{(N)}]\rangle\,d\tau$$

folglich

$$\|\Delta\mathbf{Q}^{(N+1)}-\Delta\mathbf{Q}^{(N)}\|_t\leqq T\cdot\|\mathbf{Q}(t)\|_t\cdot\|\mathbf{Q}^{-1}(t)\|_t\cdot\|\Delta\mathbf{P}(t)\|_t^{N+1}\cdot\{\langle(N-1)2K(2K+K_1)+$$

$$+2K_1\rangle(2K+K_1)^{N-1}+(N-1)(2K+K_1)^N+N\{2K+K_1\}^N\leqq$$

$$\leqq N\{2K+K_1\}^{N+1}\|\Delta\mathbf{P}(t)\|_t^{N+1}.$$

also ist (63) auch für $n=N$ gültig. (62) und (63) sind also für alle n gültig, und diese Tatsache zeigt, daß die Folge $\{\Delta\mathbf{Q}^{(n)}(t)\}$ bzw. $\{\Delta\mathbf{I}^{(n)}\}$ stark gleichmäßig gegen eine periodische $\Delta\mathbf{Q}(t)$, bzw. eine konstante $\Delta\mathbf{I}$ streben, welche also (54) genügen. Daneben zeigen unsere Relationen, daß die (56) bzw. (57) genügende $\Delta\mathbf{Q}^{(1)}(t)$ bzw. $\Delta\mathbf{I}^{(1)}(t)$ nur in $O(\|\Delta P(t)\|_t^2)$ von $\Delta\mathbf{Q}$ bzw. $\Delta\mathbf{I}$ abweichen; man kann also die FRECHET-Ableitungen gleich (56) bzw. (57) entnehmen. Es gilt nämlich der

SATZ 7. $\mathbf{Q}(\mathbf{P})$ *und* $\mathbf{I}(\mathbf{P})$ *sind Frechet-differenzierbare Funktionen von* $\mathbf{P}$ *(falls die Mehrdeutigkeit von* $\mathbf{Q}(\mathbf{P};t)$ *bzw.* $\mathbf{I}(\mathbf{P})$ *durch die Forderung* $\mathbf{Q}(\mathbf{P};t_0)\equiv\mathbf{Q}(\mathbf{P}_0,t_0)$ *aufgelöst ist), u. zw. gelten die Relationen:*

$$\frac{d\mathbf{I}}{d\mathbf{P}}=\frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\tau)\cdot(\;)\cdot\mathbf{Q}(\tau)\,d\tau \tag{65}$$

und

$$\frac{d\mathbf{Q}}{d\mathbf{P}}=\mathbf{Q}(t)\cdot\int_{t_0}^{t}\left\{\mathbf{Q}^{-1}(\tau)\cdot(\;)\cdot\mathbf{Q}(\tau)-\frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\xi)\cdot(\;)\mathbf{Q}(\xi)\,d\xi\right\}d\xi. \tag{66}$$

Es sei hier gleich bemerkt, daß man ähnlich wie oben auch zeigen kann, daß $\mathbf{I}$ und $\mathbf{Q}$ analytische Funktionen von z in $\mathbf{P}+z\Delta\mathbf{P}$ sind, und die höheren Ableitungen eben den Formeln (64) bzw. (65) gemäß angeben kann.

Kennt man also ein Tripel von zueinander gehörenden $\mathbf{P}_0(t)$, $\mathbf{Q}_0(t)$ und $\mathbf{I}_0$, so genügen die zu $\mathbf{P}(t;z)=\mathbf{P}_0(t)+z[\mathbf{P}(t)-\mathbf{P}_0(t)]$ gehörenden $\mathbf{Q}(t,z)$ und $\mathbf{I}(z)$ die Differentialgleichungen bzw. Anfangswertaufgaben:

$$\begin{aligned}\frac{d\mathbf{Q}(t,z)}{dz}&=\mathbf{Q}(t,z)\cdot\int_{t_0}^{t}\left\{\mathbf{Q}^{-1}(\tau,z)\cdot[\mathbf{P}(\tau)-\mathbf{P}_0(\tau)]\mathbf{Q}(\tau,z)-\right.\\&\left.-\frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\xi,z)\cdot[\mathbf{P}(\xi)-\mathbf{P}_0(\xi)]\cdot\mathbf{Q}(\xi,z)\,d\xi\right\}d\tau\\\frac{d\mathbf{I}(z)}{dz}&=\frac{1}{T}\int_{t_0}^{t_0+T}\mathbf{Q}^{-1}(\xi;z)\cdot[\mathbf{P}(\xi)-\mathbf{P}_0(\xi)]\mathbf{Q}(\xi,z)\,d\xi,\\\mathbf{Q}(t,0)&=\mathbf{Q}_0;\quad\mathbf{I}(0)=I_0,\end{aligned} \tag{67}$$

und es gilt der

SATZ 8. *Um ein Grundsystem der Differentialgleichung* $\dot{\mathbf{x}}=\mathbf{P}(t)\cdot\mathbf{x}$ *mit einer beliebigen Pünktlichkeit in der Form* $\mathbf{X}(t)=\mathbf{Q}(t)\cdot e^{\mathbf{L}\cdot t}$ *anzugeben, genügt es, mit Hilfe einer der in § 3 angegebenen Formeln des verallgemeinerten Runge—Kutta Typs mit entsprechender Schrittweite das Differentialgleichungssystem in* (67) *numerisch zu integrieren.*

Betrachten wir nun eine andere interessante Frage: die optimale Steuerung eines Systems. Den matematischen Gestalt dieses Problems kann man in folgende Form gießen: Es seien B_1, B_2, B_3 lineare supermetrische Räume (bzw. Banachräume), $F_0(u_1, u_2)$ bilde $B_1\times B_2$ in B_3 ab, $G_0(u_1, u_2)$ sei aber reellwertiges Funktional in $B_1\times B_2$. Man soll dann dasjenige Element $(u_1^{(0)}, u_2^{(0)})\in B_1\times B_2$ suchen, die der Gleichung $F_0(u_1, u_2)=\Theta$ genügt, für welches ferner an der „Fläche" $F_0(u_1, u_2)=\Theta$ das Funktional $G_0(u_1, u_2)$ einen lokalen extremalen Wert annimmt, d.h. man soll die Aufgabe

$$G_0(u_1, u_2)=\text{extr.},\quad F_0(u_1, u_2)=\Theta \tag{68}$$

lösen. Es ist nun wohlbekannt (s.z.B. [4]), daß im Falle, wenn F_0 und G_0 stetige partielle Ableitungen in $(u_1^{(0)}, u_2^{(0)})$ haben, ein Element $w^{(0)}\in\overline{B}_3$ so angegeben werden kann ($\overline{B}_3$ ist der zu B_3 konjugierte Raum), daß

$$\frac{\partial[G_0+wF_0]}{\partial u_1}=\frac{\partial[G_0+wF_0]}{\partial u_2}=\frac{\partial[G_0+wF_0]}{\partial w}=\Theta \tag{69}$$

in $(u_1^{(0)}, u_2^{(0)}, w^{(0)})$ erfüllt seien.

Betrachten wir nun die Aufgabe

$$G_1(u_1, u_2)=\text{extr.},\quad F_1(u_1, u_2)=\Theta, \tag{70}$$

wo G_0 eine Annäherung von G_1, ferner F_0 eine Annäherung von F_1 ist. Es sei nun $\gamma(u_1, u_2, t)$ ein Funktional in $B_1\times B_2\times R$, ebenso $\varphi(u_1, u_2, t)$ bilde $B_1\times B_2\times R$ in B_3 ab, ferner sei

$$\begin{aligned}&\gamma(u_1, u_2, t_0)\equiv G_0(u_1, u_2);\quad \gamma(u_1, u_2, t_1)\equiv G_1(u_1, u_2),\\ &\varphi(u_1, u_2, t_0)\equiv F_0(u_1, u_2);\quad \varphi(u_1, u_2, t_1)\equiv F_1(u_1, u_2)\end{aligned} \tag{71}$$

und γ bzw. φ besitzen stetige partielle Ableitungen nach u_1, u_2 in einer entsprechend großen Umgebung von $(u_1^{(0)}, u_2^{(0)}, t_0)$. Setzen wir nun zuerst auch voraus, daß die Aufgabe

$$\gamma(u_1, u_2, t)=\text{extr.},\quad \varphi(u_1, u_2, t)=\Theta \tag{72}$$

für jede fixierte $t\in[t_0, t_1]$ eine Lösung $u_1(t)$, $u_2(t)$ besitzt, welche stetig von t abhängt. Dann gilt der

HILFSSATZ 3. Neben den angegebenen Bedingungen ist auch der verallgemeinerte Lagrangesche Multiplikator, $w(t)\in\overline{B}_3$, überall stetig, wo $\sum_{i=1}^{2}\|\varphi'_{u_i}(u_1(t), u_2(t)\|>0$ gilt.

BEWEIS. Die stetig vorausgesetzten Funktionen $u_1(t)$ und $u_2(t)$, zusammen mit $\omega(t)$ genügen dem Gleichungssystem

$$\begin{aligned} &\gamma'_{u_1}(u_1(t), u_2(t), t)+\omega(t)\cdot\varphi'_{u_1}(u_1(t), u_2(t), t)=\Theta \\ &\gamma'_{u_2}(u_1(t), u_2(t), t)+\omega(t)\cdot\varphi'_{u_2}(u_1(t), u_2(t), t)=\Theta \\ &\varphi(u_1(t), u_2(t), t)=\Theta. \end{aligned} \tag{73}$$

Betrachten wir nun einen Wert $t_2\in[t_0, t_1]$, wo $\varphi'_{u_i}(u_1(t), u_2(t))\neq\Theta$ ist ($i=1$ oder 2). Da $\varphi'_{u_i}(u_1(t), u_2(t))$ unseren Voraussetzungen nach eine stetige Funktion von t ist, kann man ein $\varepsilon(t_2)$ so angeben, daß $\|\varphi'_{u_i}(u_1(t), u_2(t), t)\|$ in $t_2-\varepsilon\leqq t\leqq t_2+\varepsilon\geqq$ $\geqq\frac{1}{2}\|\varphi'_{u_i}(u_1(t_2), u_2(t_2), t_2)\|$ sei. Wir zeigen nun, daß dann $\lim\limits_{\Delta t\to 0}\|\omega(t_2+\Delta t)-$ $-\omega(t_2)\|=0$ erfüllt ist. Es gilt nämlich (73) nach

$$\begin{aligned} &\gamma'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t), t_2+\Delta t)-\gamma'_{u_i}(u_1(t_2), u_2(t_2), t_2)= \\ &=-\omega(t_2+\Delta t)\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t), t_2+\Delta t)+ \\ &+\omega(t_2)\varphi'_{u_i}(u_1(t_2), u_2(t_2), t_2)=[\omega(t_2)-\omega(t_2+\Delta t)]\cdot \\ &\cdot\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t), t_2+\Delta t)- \\ &-\omega(t_2)[\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t), t_2+\Delta t)-\varphi'_{u_i}(u_1(t_2), u_2(t_2)t_2)], \end{aligned}$$

d.h.

$$\begin{aligned} &[\omega(t_2+\Delta t)-\omega(t_2)]\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t)t_2+\Delta t)= \\ &=-\{\gamma'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t)t_2+\Delta t)-\gamma'_{u_i}(u_1(t_2), u_2(t_2)t_2)+ \\ &+\omega(t_2)[\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t)t_2+\Delta t)-\varphi'_{u_i}(u_1(t_2), u_2(t_2)t_2)]\}, \end{aligned} \tag{74}$$

also

$$\begin{aligned} &\|\omega(t_2)-\omega(t_2+\Delta t)\|\leqq 2\frac{1}{\|\varphi'_{u_i}(u_1(t_2), u_2(t_2)t_2)\|}\cdot \\ &\cdot\{\|\gamma'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t)t_2+\Delta t)-\gamma'_{u_i}(u_1(t_2), u_2(t_2)t_2)\|+ \\ &+\|\omega(t_2)\|\cdot\|\varphi'_{u_i}(u_1(t_2+\Delta t), u_2(t_2+\Delta t)t_2+\Delta t)-\varphi'_{u_i}(u_1(t_2), u_2(t_2)t_2)\| \end{aligned}$$

gültig, falls $|\Delta t|\leqq\varepsilon(t_2)$ ist. Da nun hier die rechte Seite unseren Voraussetzungen nach gegen 0 strebt, falls $\Delta t\to 0$, unsere Behauptung ist also wahr.

Ist jedoch der Bereich $\varphi(u_1, u_2, t)=\Theta$ kompakt für $t_0\leqq t\leqq t_1$, und besitzt das Extremumproblem $\gamma(u_1, u_2 t)=\text{extr.}$, $\varphi(u_1, u_2, t)=\Theta$ für jede $t\in[t_0, t_1]$ eine eindeutig definierte strenge Extremallösung, sind ferner die partielle Ableitungen stetig, und ist in den Extremalstellen $\|\varphi'_{u_1}\|+\|\varphi'_{u_2}\|>0$ immer erfüllt, so kann man beweisen, daß auch $u_1(t)$, auch $u_2(t)$ stetig sind.

SATZ 9. *Neben den oben angegebenen Bedingungen ist auch der Extremalwert, auch die Extremalstelle* $(u_1(t), u_2(t))$, *auch der verallgemeinerte Lagrangesche Multiplikator* $\omega(t)$ *eine stetige Funktion.*

BEWEIS. Wäre $[u_1(t), u_2(t)]$ in $t_2\in[t_0, t_1]$ nicht stetig, so könnte man eine Folge $\{[u_1(t^{(i)}), u_2(t^{(i)})]\}$ von Extremalstellen mit $t^{(i)}\to t_2$ derart angeben, daß $\lim\limits_{i\to\infty}\{[u_1(t^{(i)}), u_2(t^{(i)})]\}=[v_1, v_2]\neq[u_1(t_2), u_2(t_2)]$, jedoch $\varphi(v_1, v_2, t_2)=\Theta$ fest-

stehe, da unseren Voraussetzungen nach der Bereich $\varphi(u_1, u_2, t) = \Theta$ kompakt ist. Es gilt also die Gleichung

$$\gamma(v_1, v_2, t_2) = \gamma(u_1(t_2), u_2(t_2), t_2) + \Delta$$

(wo $\Delta \gtrless 0$ ist, davon abhängend, ob wir minimisieren bzw. maximieren), da wir eine einzige strenge Extremallösung haben. Ist also i genügend groß, so liegt $\gamma(u_1(t^{(i)}), u_2(t^{(i)}), t^{(i)})$ genügend nahe zu $\gamma(v_1, v_2, t_2) + \frac{1}{4}\Delta$, während wir in der Nähe von $u_1(t_2), u_2(t_2)$ eine (v_1^*, v_2^*) so angeben können, daß $\varphi(v_1^*, v_2^*, t^{(i)}) = \Theta$, und $\gamma(v_1^*, v_2^*, t^{(i)})$ genügend nahe zu $\gamma(u_1(t_2), u_2(t_2), t_2) + \frac{3}{4}\Delta$ liege. Dies zeigt aber, daß $[u_1(t^{(i)}), u_2(t^{(i)})]$ keine Extremalstelle sein kann, im Gegensatz zu unseren Voraussetzungen, $[v_1, v_2] \neq [u_1(t_2), u_2(t_2)]$ kann also nicht bestehen. Damit haben wir aber unsere Behauptung bewiesen.

SATZ 10. *Es sei jetzt vorausgesetzt, daß φ und γ in einer genügend großen Umgebung von $u_1^{(0)}, u_2^{(0)}, t_0$ stetige zweite partielle Ableitungen besitzen, und die bilineare Form*

$$\frac{1}{2!}\left(\frac{\partial^2 G_0}{\partial u_1^2} + \omega \frac{\partial^2 F_0}{\partial u_1^2}\right)(\Delta u_1)(\Delta u_1) + \left(\frac{\partial^2 G_0}{\partial u_1 \partial u_2} + \omega \frac{\partial^2 F_0}{\partial u_1 \partial u_2}\right)(\Delta u_1)(\Delta u_2) +$$

$$+ \frac{1}{2!}\left(\frac{\partial^2 G_0}{\partial u_2^2} + \omega \frac{\partial^2 F_0}{\partial u_2^2}\right)(\Delta u_2)(\Delta u_2) \quad \text{in} \quad (u_1^{(0)}, u_2^{(0)}, \omega^{(0)}) \text{ definit}$$

(positiv, resp. negativ definit, dem entsprechend, ob wir γ minimieren, bzw. maximieren wollen) ist, ferner, daß $\sum_{i=1}^{2} \|\varphi'_{u_i}(u_1(t), u_2(t))\|$ keine Nullstelle besitzt.

Dann sind die (73) *genügende $u_1(t), u_2(t)$ und $\omega(t)$ stetig differenzierbar, genügen dem Differentialgleichungssystem*

$$\begin{aligned} &\gamma''_{u_1 u_1} \cdot \dot{u}_1 + \gamma''_{u_1 u_2} \cdot \dot{u}_2 + \gamma''_{u_1 t} + \dot{\omega} \cdot \varphi'_{u_1} + \omega(\varphi''_{u_1 u_1} \dot{u}_1 + \varphi''_{u_1 u_2} \dot{u}_2 + \varphi''_{u_1 t}) = \Theta \\ &\gamma''_{u_1 u_2} \dot{u}_1 + \gamma''_{u_2 u_2} \dot{u}_2 + \gamma''_{u_2 t} + \dot{\omega} \varphi'_{u_2} + \omega(\varphi''_{u_1 u_2} \dot{u}_1 + \varphi''_{u_2 u_2} \dot{u}_2 + \varphi''_{u_2 t}) = \Theta \\ &\varphi'_{u_1} \dot{u}_1 + \varphi'_{u_2} \dot{u}_2 + \varphi'_t = \Theta, \end{aligned} \tag{75}$$

und geben eine Lösung des Extremalproblems längs der Integralkurven von (75), *bis einer t_3, wo die bilineare Form*

$$\frac{1}{2!}(\gamma''_{u_1 u_1} + \omega\varphi''_{u_1 u_1})(\Delta u_1)(\Delta u_1) + (\gamma''_{u_1 u_2} + \omega\varphi''_{u_1 u_2})(\Delta u_1) \cdot (\Delta u_2) +$$

$$+ (\gamma''_{u_2 u_2} + \omega\varphi''_{u_2 u_2})(\Delta u_2)(\Delta u_2) \tag{76}$$

semidefinit wird.

BEWEIS. Erst zeigen wir, daß im Falle, wenn (73) erfüllt ist, und dort (76) entsprechend definit ist, unser Etremumproblem eine Lösung hat.

Es gilt nämlich hier, dass für hinreichend kleine Δu_1 und Δu_2 mit $\|\Delta u_1\| + \|\Delta u_2\| > 0$ die Größe $\Delta(\gamma + \omega\varphi)$ ein stabiles Vorzeichen besitzt, d.h. $\gamma + \omega\varphi$ ein Extremum besitzt. Daneben ist $\Delta(\gamma + \omega\varphi) = \Delta\gamma + \omega\Delta\varphi$, und für solche Paare

Δu_1 und Δu_2, welche zur Tangentialmannigfaltigkeit von φ gehören (und (73) läßt nur solche zu), ist auch $\|\Delta\varphi\| = o(\sqrt{\|\Delta u_1\|^2 + \|\Delta u_2\|^2}) = o(\|\Delta\gamma\|)$ gültig. Für diese Tangentialmannigfaltigkeit besitzt also auch $\Delta\gamma$ ein stabiles Vorzeichen, w.z.b.w.

Die Voraussetzung, daß (76) definit ist, und $\|\varphi'_{u_1}\| + \|\varphi'_{u_2}\| > 0$ ist, sichert die Auflösbarkeit von (75) nach $\dot{u}_1$, $\dot{u}_2$, und $\dot{\omega}$, welche auf Grund unserer Voraussetzungen stetige Funktionen ihrer Veränderlichen sind (s.z.B. [1] bzw. [4]). Man kann nun längs der Integralkurven von (75) die Bedingung über Definitheit von (76) bzw. die Gültigkeit von $\sum \|\varphi'_{u_i}\| > 0$ kontrollieren, und da alle partiellen Ableitungen unseren Voraussetzungen gemäß stetig sind, so ist auch die obere bzw. untere Grenze von (76) für $\sqrt{\|\Delta u_1\|^2 + \|\Delta u_2\|^2} = 1$ stetig, d.h. (76) ist in $t = t_2$ definit, und man kann daher eine $\varepsilon(t_2) > 0$ so angeben, daß für $t_2 - \varepsilon < t_2 + \varepsilon$ (76) noch definit bleibe. Wird diese Bedingung erfüllt, so gibt $u_1(t_2)$, $u_2(t_2)$, wie wir es oben gezeigt haben, eine Lösung des Extremumproblems.

Man kann also mit Hilfe der Gleichung (75), falls man sie numerisch von t_0 zu t_1 integriert, eine angenäherte Lösung des Extremumproblems

$$F_1(u_1, u_2) = \Theta, \quad G_1(u_1, u_2) = \text{extr.} \tag{77}$$

angeben, um so mehr vielleicht auch beweisen, daß (77) eine Lösung besitzt, falls man zeigen kann, daß (76) längs der Integralkurve von (75) definit, und $\|\varphi'_{u_1}\| + \|\varphi'_{u_2}\| > 0$ gültig bleibt.

LITERATURVERZEICHNIS

[1] KANTOROWITSCH, L. W. und AKILOW, G. P.: *Funktionalanalysis in normierten Räumen*, Akademie. V., Berlin, 1964.

[2] COLLATZ, L.: *Funktionalanalysis und numerische Mathematik*, Springer, Berlin—Göttingen—Heidelberg, 1964.

[3] KIZNER, W.: A Numerical Method for Findig Solutions of Nonlinear Equations, *Siam J. Appl. Math.* **12** (1964).

[4] LJUSTERNIK, L. A. und SOBOLEW, W. I.: *Elemente der Funktionalanalysis*, Akademie. V., Berlin, 1955.

Rechenzentrum der Ungarischen Akademie der Wissenschaften, Budapest

(Eingegangen: 10. Dezember, 1966.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 127—136.

A DUALITY RELATION FOR DISCRETE ORTHOGONAL SYSTEMS

by

G. K. EAGLESON

1. Introduction

Given a finite system of functions

$$\{\varphi_n(i)\}; \quad n, i=0, 1, \ldots, N$$

which are orthogonal with respect to the discrete distribution

$$\{P_i\}; \quad i=0, 1, \ldots, N; \quad P_i>0, \quad \text{each} \quad i,$$

and which have normalising constants $\{\pi_n\}$, the following matrix can be constructed

$$B = \begin{bmatrix} \varphi_0(0)\sqrt{\pi_0 P_0} & \ldots & \varphi_0(N)\sqrt{\pi_0 P_N} \\ \vdots & & \\ \varphi_N(0)\sqrt{\pi_N P_0} & \ldots & \varphi_N(N)\sqrt{\pi_N P_N} \end{bmatrix}.$$

The rows of B are orthogonal and normalised and as B is finite, this implies that the columns are also. Hence both

$$\sum_{i=0}^{N} \varphi_n(i)\varphi_m(i)\sqrt{\pi_n \pi_m} P_i = \delta_{nm}$$

and

$$\sum_{n=0}^{N} \varphi_n(i)\varphi_n(j)\pi_n \sqrt{P_i P_j} = \delta_{ij}$$

are satisfied.

So the functions $\{\varphi_n(i)\}$ could be considered as functions of i, for fixed n, which are orthogonal with respect to $\{P_i\}$ and with normalising constants $\{\pi_n\}$. Or, they could be considered as functions of n, for fixed i, orthogonal with respect to $\{\pi_n\}$ and with normalising constants $\{P_i\}$. It would seem reasonable to call the two systems arising from the $\{\varphi_n(i)\}$, dual.

In this chapter, the concept of a dual system is extended from the finite to the infinite case and those polynomial systems which are self-dual are determined. Finally, the concept of duality is used to show the connection between the integral expansions of the transition probabilities for birth-and-death processes obtained by KARLIN and MCGREGOR [3] and the canonical expansions of the same processes.

2. Dual Systems

We prove the following theorem [1]:

THEOREM 1. *If* $\{\varphi_n(i)\}_{n=0}^{\infty}$ *is a system of orthogonal functions defined on a discrete distribution* $\{p_i\}_{i=0}^{\infty}$, $p_i>0$, *whose normalising constants are* $\{\pi_n\}_{n=0}^{\infty}$ *then the dual orthogonality relation*

$$p_i \sum_{n=0}^{\infty} \varphi_n(i)\varphi_n(j)\pi_n = \delta_{ij} \tag{1}$$

holds, iff the $\{\varphi_n(i)\}$ *are a complete system.*

PROOF. Without loss of generality, we may take the set of points on which $\{p_i\}$ is non-zero to coincide with the set of indices of the $\{\varphi_n(i)\}$.

SUFFICIENCY. Let H be the Hilbert space of real sequences $\mathbf{x}=\{x_j\}$ $(0\leqq j<\infty)$ with

$$\sum_j x_j^2 p_j < \infty.$$

Then the inner product of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$(\mathbf{x},\mathbf{y}) = \sum_j x_j y_j p_j, \quad \mathbf{x},\mathbf{y}\in H.$$

Let $\mathbf{u}(i)=\{\delta_{ij}\}$, so that $\mathbf{u}(i)\in H$, all i.

As the set $\{\sqrt{\pi_n}\varphi_n(j)\}$ is orthonormal and complete, the $\mathbf{u}(i)$ will have an expansion in terms of the $\{\sqrt{\pi_n}\varphi_n(j)\}$ convergent in mean square. The coefficients in the expansion will be:

$$(\mathbf{u}(i), \sqrt{\pi_n}\boldsymbol{\varphi}_n) = \sqrt{\pi_n}\varphi_n(i)p_i$$

Thus

$$(\mathbf{u}(i),\mathbf{u}(k)) = \sum_j \delta_{ij}\delta_{kj}p = \delta_{ik}p_k = \sum_n (\mathbf{u}(i), \sqrt{\pi_n}\boldsymbol{\varphi}_n)(\mathbf{u}(k), \sqrt{\pi_n}\boldsymbol{\varphi}_n) =$$

$$= \sum_n \pi_n\varphi_n(i)\varphi_n(k)p_i p_k.$$

Q. e. d.

NECESSITY. Let d_l be a function square-summable with respect to $\{p_j\}$. Then

$$\sum_{i=0}^{\infty} d_i\delta_{ij} = d_j$$

where

$$\sum_{i=0}^{\infty} d_i^2 p_i < \infty.$$

[1] Theorem 1 could be deduced from more general theorems in the theory of Hilbert-space, e.g. from the theorem that if U is a unitary operator, then its adjoint is its inverse and is an unitary operator, too; we prefer however to give a direct and elementary proof.

As (1) holds,

$$d_j = \sum_{i=0}^{\infty} d_i \sum_{n=0}^{\infty} \varphi_n(i)\varphi_n(j)p_i\pi_n = \sum_{n=0}^{\infty} \varphi_n(j)\sqrt{\pi_n} \sum_{i=0}^{\infty} d_i\varphi_n(i)p_i\sqrt{\pi_n} =$$

$$= \sum_{n=0}^{\infty} \sqrt{\pi_n}\varphi_n(j)b_n$$

where

$$\sum_{n=0}^{\infty} b_n^2 = \sum_{n=0}^{\infty} \sum_{i,k=0}^{\infty} d_i d_k \varphi_n(i)\varphi_n(k)p_i p_k \pi_n = \sum_{i=0}^{\infty} d_i^2 p_i < \infty.$$

Hence d_j may be written as a linear combination of the $\{\varphi_n(j)\}$ where the sum of the squares of the coefficients is finite. That is, the system $\{\varphi_n(i)\}$ is complete.

DEFINITION. If $\{\varphi_n(i)\}$ is a complete system of orthogonal functions defined on a discrete distribution, then the set of functions defined by

$$D_i(n) \equiv \varphi_n(i)$$

is called the *dual orthogonal system* of the $\{\varphi_n(i)\}$.

COROLLARY. The dual orthogonal system is also complete.

Thus the theorem and its corollary state that every complete orthogonal system defined on a discrete distribution has an associated dual orthogonal system which is also complete and which is obtained from the original system by interchanging the roles of variable and index. In the dual system, the roles of normalising constants and weight function are also interchanged.

EXAMPLES

1. *The Hahn-polynomials.* (See KARLIN and MCGREGOR, [4]). The Hahn polynomials are defined by

$$Q_n(x) = \sum_{k=0}^{N-1} \frac{(-n)_k(-x)_k(n+\alpha+\beta+1)_k}{(\alpha+1)_k(-N+1)_k k!}; \qquad x = 0, 1, \ldots, N-1; \quad \alpha, \beta > -1.$$

They are orthogonal with respect to the weights

$$\varrho(x) = \binom{\alpha+x}{x}\binom{\beta+N-1-x}{N-1-x}\Big/\binom{N+\alpha+\beta}{N-1}; \qquad x = 0, 1, \ldots, N-1$$

with normalising constants

$$\pi_n = \binom{N-1}{n} \frac{\Gamma(\beta+1)}{\Gamma(\alpha+1)\Gamma(\alpha+\beta+1)} \frac{\Gamma(n+\alpha+1)\Gamma(n+\alpha+\beta+1)(2n+\alpha+\beta+1)}{\Gamma(n+\beta+1)\Gamma(n+1)(\alpha+\beta+1)}.$$

If $\lambda_n = n(n+\alpha+\beta+1)$, the functions defined by

$$R_k(\lambda_n) = Q_n(k)$$

are also polynomials and satisfy the orthogonality relation

$$\sum_{n=0}^{N-1} R_i(\lambda_n) R_j(\lambda_n) \pi_n \varrho(j) = \delta_{ij}.$$

So the Hahn polynomials are an example of an orthogonal polynomial system whose dual system (suitably transformed) is also polynomial.

2. *The Poisson—Charlier Polynomials.* (See ERDÉLYI [2]).
The Poisson—Charlier polynomials are defined by

$$Q_n(x; a) = \sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{x}{r} r!/a^r; \qquad a>0; \quad x=0, 1, \ldots.$$

They are orthogonal with respect to the distribution

$$\varrho(x) = e^{-a} a^x/x!; \qquad x = 0, 1, \ldots$$

with normalising constants

$$\pi_n = a^n/n!.$$

They satisfy the symmetry relation

$$Q_n(x; a) = Q_x(n; a)$$

and so are an example of an orthogonal polynomial system which is self-dual.

REMARKS. (i) Professor A. RÉNYI (personal communication) pointed out to me the possibility of using Hilbert space techniques to prove the sufficiency of Theorem 1.

(ii) The dual orthogonality relation for the Hahn polynomials was noted by KARLIN and MCGREGOR [4].

(iii) If a system of orthogonal functions satisfies a set of recurrence relations (difference equations) then its dual system satisfies a set of difference equations (recurrence relations).

(iv) Every limit theorem on a system of complete orthogonal functions will have a dual.

(v) In the case of a complete system of orthogonal functions defined on a continuous distribution, although an analogue of (1) holds a.e., the concept of duality loses its significance.

3. Self-dual Polynomial Systems

The Poisson—Charlier polynomials are an example of an orthogonal polynomial system which is self-dual. We shall determine the set of all orthogonal polynomial systems which, when transformed by suitable multiplying constants, are self-dual.

DEFINITION. Given a system of orthogonal polynomials $\{U_n(i)\}$, we say that they are *symmetrisable* if there exists constants w_n such that

$$w_n U_n(i) = w_n U_i(n), \quad \text{all} \quad n \quad \text{and} \quad i.$$

Then w_n may be normalised so that $w_0 = 1$. This is equivalent to saying that the system $\{U_n(i)\}$ is symmetrisable if

$$\frac{U_n(i)}{U_n(0)} = \frac{U_i(n)}{U_i(0)} \quad \text{all } n \text{ and } i.$$

Defining

$$V_n(i) = w_n U_n(i) = V_i(n),$$

we have

$$\begin{cases} \sum_{n=0}^{\infty} V_n(i) V_n(j) \pi_n' p_j' = \delta_{ij} \\ \sum_{i=0}^{\infty} V_n(i) V_m(i) \pi_n' p_i' = \delta_{nm}. \end{cases}$$

Assume the system of orthogonal polynomials, $\{U_n(i)\}$ is symmetrisable. Then $V_n(0) = V_0(i) = 1$ and the $\{V_n(i)\}$ must satisfy a second order recurrence relation of the form (see KARLIN and MCGREGOR [3])

$$(2) \qquad \begin{cases} \lambda_0 V_0(i) - \lambda_0 V_1(i) = -iV_0(i) \\ \lambda_n V_{n+1}(i) - (\lambda_n + \mu_n) V_n(i) + \mu_n V_{n-1}(i) = -iV_n(i), \qquad n \geqq 1. \end{cases}$$

The symmetry of the V's implies that they also satisfy second order difference equations of the form

$$(3) \qquad \begin{cases} \lambda_0 V_i(0) - \lambda_0 V_i(1) = -iV_i(0) \\ \lambda_n V_i(n+1) - (\lambda_n + \mu_n) V_i(n) + \mu_n V_i(n-1) = -iV_i(n), \qquad n \geqq 1. \end{cases}$$

LEMMA. The coefficients λ_n and μ_n appearing in the recurrence relations of a symmetrisable system of orthogonal polynomials are themselves polynomials in n, of degree at most one.

PROOF. Consider the expression

$$(4) \qquad P_{ij}(t) = p_j' \sum_{n=0}^{\infty} e^{-nt} V_n(i) V_n(j) \pi_n', \qquad 0 \leqq t \leqq \infty.$$

For fixed i and j, the series converges absolutely, uniformly with respect to t, for

$$p_j' \sum_{n=0}^{\infty} |e^{-nt} V_n(i) V_n(j) \pi_n'| \leqq p_j' \left\{ \sum_{n=0}^{\infty} \pi_n' V_n^2(i) \sum_{n=0}^{\infty} \pi_n' V_n^2(j) \right\}^{\frac{1}{2}}$$
$$= \{p_j' | p_i'\}^{1/2} < \infty.$$

Now (3) implies that

$$(5) \qquad \frac{dP(t)}{dt} = AP$$

where

$$A = \begin{bmatrix} \lambda_0 & \lambda_0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \\ \vdots & & & \end{bmatrix}.$$

Further, as $P(0)=I$, (5) implies that

$$a_{ij} = \lim_{t\to 0} \frac{P_{ij}(t)}{t} = P'_{ij}(0), \qquad i \neq j$$

and

$$a_{ii} = \lim_{t\to 0} \frac{P_{ij}(t)-1}{t}$$

Now $P_{ij}(t)$ is a continuous function of t which has a derivative at the origin. Hence, for t small enough,

$$\frac{|P_{ij}(t)-P_{ij}(0)|}{t} \leqq (1+\varepsilon)a_{ij}.$$

Thus

$$\left| \sum_{j=0}^{\infty} (j-i)^r \frac{P_{ij}(t)}{t} \right| \leqq \sum_{j\neq i} |j-i|^r (1+\varepsilon)a_{ij} = (1+\varepsilon)(\lambda_i+\mu_i)$$

and the sum converges uniformly with respect to t, for t small enough. So, if we consider the functions

$$A(i) = \lim_{t\to 0} \frac{1}{t} \sum_{j=0}^{\infty} (j-i)P_{ij}(t)$$

and

$$B(i) = \lim_{t\to 0} \frac{1}{t} \sum_{j=0}^{\infty} (j-i)^2 P_{ij}(t),$$

it is easy to see that

$$A(i) = \lambda_i - \mu_i$$

and

$$B(i) = \lambda_i + \mu_i.$$

In order to calculate $A(i)$ and $B(i)$, we use the expansion (4) and the fact that the $\{V_n(i)\}$ are a system of orthogonal polynomials (see Sarmanov [7]). It follows that

$$A(i) = -(i-\mu)$$

$$B(i) = i\left(\frac{\mu_3-\mu_2\mu_1}{\sigma^2}\right) - 2\mu_1 + \frac{\mu_1(\mu_3-\mu_2\mu_1)}{\sigma^2} + \frac{2(\mu_2^2-\mu_3\mu_1)}{\sigma^2}$$

where $\mu_r = E_{p_i}(i^r)$, $r=1, 2, 3$. So the Lemma is proved.

THEOREM 2. *A system of orthogonal polynomials is symmetrisable iff it is one of the following*

(i) *The Poisson—Charlier polynomials, orthogonal with respect to the Poisson distribution.*

(ii) *The Krawtchouk polynomials, orthogonal with respect to the binomial distribution.*

(iii) *The polynomials, orthogonal with respect to the negative binomial distribution.*

PROOF. NECESSITY. From the Lemma, a symmetrisable system of orthogonal polynomials satisfies a set of second order difference equations whose coefficients

are polynomials of degree at most one. It follows from the work of O. E. LANCASTER [6] that the only polynomial solutions to such equations are the three listed in the Theorem.

SUFFICIENCY. It is well known that each of the three systems listed above are symmetrisable (see ERDÉLYI [2]).

4. The Integral Representations of the Transition Probabilities of Birth-and-death Processes

KARLIN and MCGREGOR [3] have obtained an integral representation for the transition probabilities of a birth-and-death process. A short resumè of some of their work is given here.

The transition probability matrix $P(t)=(P_{ij}(t))$, $t\geqq 0$, $i,j=0,1,\ldots$ of a birth-and-death process satisfies the differential equations

$$P'(t)=AP(t), \tag{6}$$

$$P'(t)=P(t)A, \tag{7}$$

the initial condition

$$P(0)=I \tag{8}$$

and has the properties

$$P_{ij}(t)\geqq 0, \tag{9}$$

$$\sum_{j=0}^{\infty} P_{ij}(t) \leqq 1 \tag{10}$$

$$P(t+s) = P(t)P(s). \tag{11}$$

The matrix $A=(a_{ij})$ is of the form

$$\begin{cases} a_{i,i+1} = \lambda_i \\ a_{i,i} = -(\lambda_i+\mu_i) \\ a_{i,i-1} = \mu_i \\ a_{ij}=0 \quad \text{if} \quad |i-j|>1 \end{cases} \tag{12}$$

KARLIN and MCGREGOR have shown that, associated with every such matrix A, there exists a system, $\{Q_n(x)\}$, of polynomials defined by the recurrence relations

$$\begin{cases} Q_0(x) \equiv 1 \\ -xQ_0(x) = -(\lambda_0+\mu_0)Q_0(x)+\lambda_0 Q_1(x) \\ -xQ_n(x) = -\mu_n Q_{n-1}(x)-(\lambda_n+\mu_n)Q_n(x)+\lambda_n Q_{n+1}(x), \qquad n\geqq 1. \end{cases} \tag{13}$$

They proved that there always exists at least one positive regular normed measure $\psi(x)$ on $0\leqq x\leqq\infty$ with respect to which the $\{Q_n(x)\}$ are orthogonal, i.e.

$$\int_0^{\infty} Q_i(x)Q_j(x)\,d\psi(x)\pi_j = \delta_{ij} \tag{14}$$

where

$$\pi_j = \lambda_0\lambda_1 \ldots, \lambda_{j-1}/\mu_1\mu_2 \ldots \mu_j. \tag{15}$$

Such a $\psi(x)$ they call a solution to the S-moment problem as the relations (13) and (14) determine the moments of $\psi(x)$. Further, any system of orthogonal polynomials satisfies a recurrence relation of the form (13) and so is associated with some matrix.

They proved that the transition probabilities of a birth-and-death process can always be expressed in the form

$$P_{ij}(t;\psi) = \pi_j \int_0^\infty e^{-xt} Q_i(x) Q_j(x)\, d\psi\,(x). \tag{16}$$

A solution, $\psi(x)$, of the S-moment problem is called extremal if the $\{Q_n(x)\}$ are a complete system in $L_2(\psi)$. If the solution of the S-moment problem is unique, it is extremal. If it is not unique, then there exists a one-parameter family of solutions of the S-moment problem which are extremal, in this case, there is an extremal solution with mass at the origin.

5. Canonical Expansions of Birth-and-death Processes

A bivariate distribution function $F(x, y)$ is said to be φ^2-bounded with respect to its marginal distributions $G(x)$ and $H(y)$ if the following integral is finite (see H. O. LANCASTER [5]).

$$\varphi^2 + 1 = \iint \left\{ \frac{dF(x,y)}{dG(x)\, dH(y)} \right\}^2 dG(x)\, dH(y) < \infty \tag{17}$$

If an honest birth-and-death process $(P_{ij}(t))$ has a normed invariant measure $\{\pi_i\}$, i.e.

$$\begin{cases} \sum_{i=0}^{\infty} \pi_i P_{ij}(t) = \pi_j \quad \text{all} \quad i,j \quad \text{and} \quad t \\ \pi_i \geqq 0, \quad \sum_{i=0}^{\infty} \pi_i = 1 \end{cases} \tag{18}$$

then the transition probabilities of the process can be used to construct a class of symmetric bivariate distributions, depending on the parameter t. If these distributions are φ^2-bounded with respect to $\{\pi_i\}$ and $\{\pi_j\}$, they will have a canonical expansion of the form

$$P_{ij}(t) = \pi_j \sum_{n=0} e^{-\alpha_n t} R_n(i) R_n(j) p_n \tag{19}$$

in mean square where $0 = \alpha_0 \leqq \alpha_1 \leqq \alpha_2 \leqq \ldots$.

The functions $\{\sqrt{p_n} R_n(i)\}$ are orthonormal and complete with respect to $\{\pi_i\}$ and are called the canonical variables of the bivariate distributions. The $\{e^{-\alpha_n t}\}$ are called the canonical correlations. Those particular cases when the canonical variables of a birth-and-death process are polynomials are investigated in [1]. The $\{\pi_i\}$ defined by (15) symmetrise A and hence $P(t)$ (see Lemma 6 of [3]). If the process is honest, then the $\{\pi_i\}$ is an invariant measure but it need not be totally finite.

It follows from (6) and (7) that the canonical variables satisfy the following difference equations

$$(20)\qquad \begin{cases} \lambda_0 R_n(1) - (\lambda_0 + \mu_0) R_n(0) = -\alpha_n R_n(0) \\ \lambda_i R_n(i+1) - (\lambda_i + \mu_i) R_n(i) + \mu_i R_n(i-1) = -\alpha_n R_n(i), \qquad i \geqq 1 \end{cases}$$

$$(21)\qquad \begin{cases} \pi_1 \mu_1 R_n(1) - \pi_0(\lambda_0 + \mu_0) R_n(0) = -\alpha_n \pi_0 R_n(0) \\ \pi_{i+1}\mu_{i+1} R_n(i+1) - \pi_i(\lambda_i + \mu_i) R_n(i) + \pi_{i-1}\lambda_{i-1} R_n(i-1) = -\alpha_n \pi_i R_n(i), \; i \geqq 1. \end{cases}$$

If $R_n(0) = 0$, the difference equation (20) implies that first $R_n(1) = 0$ and then $R_n(1) = 0$ all i. So without loss of generality, we may assume $R_n(0) \neq 0$. Now the $\{p_n\}$ can be so chosen to normalise the $R_n(i)$ at the origin, i.e. so that $R_n(0) = 1$, all n.

In general $\alpha_n \leqq \alpha_{n+1}$, but the above difference equations imply that the equality is impossible. If $\alpha_n = \alpha_{n+1}$, (20) would imply that

$$R_n(1) = R_{n+1}(1) \quad \text{since} \quad R_n(0) = R_{n+1}(0) = 1.$$

Thus $R_n(i)$ and $R_n(i+1)$ are both solutions of the same second-order difference equation and both satisfy identical boundary conditions. Hence $R_n(i) = R_{n+1}(i)$ all i. But this contradicts the orthogonality of the system $\{R_n(i)\}$.

It is clear that the canonical variables are the duals (suitably renumbered) of a polynomial system, defined on the set $\{0 = \alpha_0, \alpha_1, \ldots\}$ and the canonical expansion is identical with the KARLIN and MCGREGOR integral representation.

More precisely, if we restrict our attention to φ^2-bounded birth-and-death processes, we have

THEOREM 3. *The canonical expansion of an honest, φ^2-bounded, birth-and-death process, possessing a totally finite invariant measure, coincides with the Karlin and McGregor integral representation of the same process iff $\mu_0 = 0$ and the solution of the corresponding S-moment problem is extremal and is supported by a discrete set of points, containing the origin.*

PROOF. SUFFICIENCY. If the solution to the S-moment problem is extrema and has discrete support, then the $\{Q_i(n)\}$ are complete and have a dual system $\{R_n(i)\}$ which is complete and orthogonal with respect to $\{\pi_i\}$.

$\mu_0 = 0$ implies that $Q_j(0) = 1$, all j (p. 494 in [3]). Now if $P_n = \text{mass of } \psi(x)$ at α_n,

$$\sum_{j=0}^{\infty} P_{ij}(t) = \sum_{n=0}^{\infty} e^{-\alpha_n t} Q_i(\alpha_n) \sum_{j=0}^{\infty} \pi_j Q_j(\alpha_n) P_n = \sum_{n=0}^{\infty} e^{-\alpha_n t} Q_i(\alpha_n) \sum_{j=0}^{\infty} \pi_j Q_j(0) Q_j(\alpha_n) P_n =$$

$$= \sum_{n=0}^{\infty} e^{-\alpha_n t} Q_i(\alpha_n) \delta_{0\alpha_n} = e^{-\alpha_0 t} = 1.$$

Hence, the process is honest.

Also, the mass at the origin is

$$\left\{ \sum_{i=0}^{\infty} \pi_i Q_i^2(0) \right\}^{-1} = \left\{ \sum_{i=0}^{\infty} \pi_i \right\}^{-1} \neq 0.$$

So the invariant measure $\{\pi_i\}$ is totally finite.

NECESSITY. Honesty implies $\mu_0 = 0$. The remainder of the Theorem is obvious from the comments above.

Q. e. d.

ACKNOWLEDGEMENTS. I wish to acknowledge the encouragement and help of Professor A. RÉNYI in this work.

REFERENCES

[1] EAGLESON, G. K.: Канонические разложения процессов гибели и размножения (to appear in *Теор.Вероятност.иПримен.*)
[2] ERDÉLYI, A.: (editor): *Higher Transcendental Functions*, Vol. 11, McGraw-Hill, New York (1953).
[3] KARLIN, S. and MCGREGOR, J. L.: The differential equations of birth-and-death processes and the Stieltjes moment problem, *Trans. Amer. Math. Soc.* **85** (1957) 489—546.
[4] KARLIN, S., and MCGREGOR, J. L.: The Hahn polynomials, formulas and an application, *Scripta Math.* **26** (1961) 33—46.
[5] LANCASTER, H. O.: The structure of bivariate distributions, *Ann. Math. Statist.* **29** (1958) 719—736.
[6] LANCASTER, O. E.: Orthogonal polynomials defined by difference equations, *Amer. Math.* **63** (1941) 185.
[7] Сарманов, О. В.: Исследование стационарных марковских процессов методом разложения по собственным функциям, *Труды Мат. Инст. Стеклова* **60** (1961) 238—261.

The University of Sydney, Department of Mathematical Statistics

(Received January 2, 1967.)
(Revised August 2, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 137—152.

BIRECOUVREMENTS ET BIREVÊTEMENTS D'UN ENSEMBLE FINI

par

L. COMTET

Notre but est de calculer et d'estimer le nombre $c(n)$ des recouvrements d'un ensemble fini E à n éléments, tels que chaque point de E soit deux fois recouvert: en abrégé „birecouvrements"; ce nombre $c(n)$ généralise le nombre (de Bell) $b(n)$ des partitions de E ([1], [2]) puisqu'une partition est un recouvrement tel que chaque point de E soit une fois recouvert. Les blocs d'un birecouvrement étant, par définition, tous distincts, nous sommes amenés à étudier des systèmes de parties de E plus généraux que les birecouvrements et que nous appelons „birevêtements"; chaque point de E est encore deux fois recouvert mais certains blocs du système peuvent ne pas différer; leur nombre $v(n)$ est en relation simple avec $c(n)$. Au passage s'introduit le nombre $c(n, k)$ des birecouvrements à k blocs: il généralise le nombre de Stirling de seconde espèce $S(n, k)$ ([3], p. 32). Nous donnons enfin une estimation et une fonction génératrice des $c(n)$ et $v(n)$.

1. Introduction

Dans tout ce qui suit, le nombre d'éléments d'un ensemble M se note $|M|$. Soit E un ensemble fini ayant n éléments, $|E|=n$, et soit $\mathscr{P}'(E)$ l'ensemble de ses parties non Vides.

DÉFINITION 1. Un *système* $\mathscr{S}$ de E est un ensemble non vide (non ordonné) de parties non vides distinctes de E: $\mathscr{S} \subset \mathscr{P}'(E)$. Les *blocs* d'un système sont les parties de E dont il est constitué. Un *k-système* est un système constitué de k blocs.

DÉFINITION 1′. Un *agrégat* $\mathscr{A}$ de E est un ensemble non vide (non ordonné) de parties non vides de E, chacune pouvant apparaître plusieurs fois; la donnée d'un agrégat $\mathscr{A}$ équivaut donc à la donnée d'une fonction φ définie sur $\mathscr{P}'(E)$ et dont les valeurs sont des entiers $\geqq 0$, telle que $\varphi(A)$ soit le nombre de fois qu'apparaît dans $\mathscr{A}$ la partie $A \in \mathscr{P}'(E)$. Il pourra être utile de partager $\mathscr{P}'(E)$ en classes $\varepsilon_h(\mathscr{A})$, h entier $\geqq 0$, définies par:

$$\varepsilon_h(\mathscr{A}) = \varphi^{-1}(h) = \{A | A \in \mathscr{P}'(E), \varphi(A) = h\}.$$

Les *blocs* d'un agrégat sont encore les parties de E dont il est constitué. Un *k-agrégat* est un agrégat constitué de k blocs, distincts ou non:

$$\sum_{A \in \mathscr{P}'(E)} \varphi(A) = k;$$

en d'autre termes, c'est une k-combinaison avec répétition dans $\mathscr{P}'(E)$.

DÉFINITION 2. Un système $\mathscr{S}$ est un „*birecouvrement*" si chaque $x \in E$ appartient exactement à 2 blocs (distincts) de $\mathscr{S}$; on note $\mathbf{c}(E)$ l'ensemble des birecouvrements de E, et l'on pose $c(n) \equiv |\mathbf{c}(E)|$. Un *k-birecouvrement* est un birecouvrement constitué de k blocs; on note $\mathbf{c}(E, k)$ l'ensemble des k-birecouvrements de E, et l'on pose $c(n, k) \equiv |\mathbf{c}(E, k)|$. Evidemment:

$$c(n) = \sum_k c(n, k).$$

DÉFINITION 2′. Un agrégat $\mathscr{A}$ est un „*birevêtement*" si chaque $x \in E$ appartient exactement à 2 blocs, distincts ou non, de $\mathscr{A}$; on note $\mathbf{v}(E)$ l'ensemble des birevêtements de E, et l'on pose $v(n) \equiv |\mathbf{v}(E)|$. Un *k-birevêtement* est un birevêtement constitué de k blocs; on notera $\mathbf{v}(E, k)$ l'ensemble des k-birevêtements de E, et l'on pose $v(n, k) \equiv |\mathbf{v}(E, k)|$. Evidemment:

$$v(n) = \sum_k v(n, k).$$

Il est clair que $\mathbf{c}(E) \subset \mathbf{v}(E)$. Par exemple, pour $E \equiv \{1, 2, 3, 4\}$,

$$\mathscr{A}_1 \equiv \big\{\{1,2\}, \{2,3\}, \{2,4\}\big\}, \quad \mathscr{A}_2 \equiv \big\{\{1\}, \{1\}, \{2,3\}, \{4\}, \{2,3,4\}\big\},$$

$$\mathscr{A}_3 \equiv \big\{\{1,2,3\}, \{3,4\}, \{1,2,4\}\big\}, \quad \text{on a} \quad \mathscr{A}_1 \notin \mathbf{v}(E), \quad \mathscr{A}_2 \in \mathbf{v}(E) \quad \text{et} \quad \notin \mathbf{c}(E)$$

$$\mathscr{A}_3 \in \mathbf{c}(E).$$

Donnons un exemple de problème de dénombrement où interviennent les nombres $c(n, k)$ et $v(n, k)$. Soient deux jeux absolument identiques de 52 cartes à jouer; on les mélange ensemble et l'on obtient ainsi un jeu de 104 cartes deux par deux indiscernables. On répartit les 104 cartes en 5 tas tels que dans chaque tas, toutes les valeurs des cartes soient différentes, l'ordre des tas et des cartes n'intervenant pas. Il est facile de voir que le nombre des répartitions de cartes en 5 tas ayant la propriété requise est égal à $v(52,5)$; en effet, soit E l'ensemble des 52 valeurs possibles de cartes; chaque tas détermine un bloc de E et une répartition des cartes en 5 tas équivaut à la donnée d'un 5-birevêtement de E, puisque chacune des 52 valeurs apparait dans deux tas différents. Si l'on ajoute la condition que deux tas distincts ne sont pas composés de valeurs toutes égales, le nombre des répartitions possibles devient $c(52, 5)$.

2. Relation entre les $c(n, k)$ et les $v(n, k)$

A tout k-birevêtement ϑ de E, $|E| = n$, associons les 4 ensembles suivants:

(1) l'ensemble $\varepsilon_1(\vartheta)$ $(\in \mathscr{P}'(E)$, voir définition 1′).
(2) l'ensemble $\varepsilon_2(\vartheta)$.
(3) la partie $E_2(\vartheta) \equiv \bigcup_{B \in \varepsilon_2(\vartheta)} B$ de E; pour $a \equiv |E_2(\vartheta)|$, on a évidemment $0 \leqq a \leqq n$.
(4) la partie $E_1(\vartheta) \equiv E \backslash E_2(\vartheta)$ de E.

Si $\varepsilon_2(\vartheta)$ est vide, $\varepsilon_1(\vartheta)=\vartheta$ est un k-birecouvrement de E. Si $\varepsilon_1(\vartheta)$ est vide, $\varepsilon_2(\vartheta)$ est une $\frac{k}{2}$-partition de E. Sinon,

$$1 \leqq a = |E_2(\vartheta)| \leqq n-1$$

et, si l'on pose $u \equiv |\varepsilon_2(\vartheta)|$, il apparaît que $|\varepsilon_1(\vartheta)| = k-2u$; dans ces conditions, $\varepsilon_2(\vartheta)$ est une u-partition de $E_2(\vartheta)$ et $\varepsilon_1(\vartheta)$ est un $(k-2u)$-birecouvrement de $E_1(\vartheta)$, $1 \leqq u \leqq \{k/2, a\}$.

Réciproquement, la donnée des quatre ensembles E_1, E_2, $\mathbf{p}$, $\mathbf{q}$, définis ci-après, équivaut à la donnée d'un k-birevêtements:

(1) E_1 et E_2, parties de E, telles que

$$E_1 \cap E_2 = \emptyset, \quad E_1 \cup E_2 = E, \quad 0 \leqq |E_2| \equiv a \leqq n$$

(2) $\mathbf{p}$ qui est une u-partition de E_2 quand $E_2 \neq \emptyset$ avec $1 \leqq u \leqq \{k/2, a\}$, et qui est vide dans le cas contraire.

(3) $\mathbf{q}$ qui est un $(k-2u)$-birecouvrement de E_1 quand $E_1 \neq \emptyset$, et qui est vide autrement.

On obtiendra donc tous les k-birevêtements en faisant varier a et u indépendamment de manière convenable; ainsi, $S(a, u)$ désignant le nombre de u-partitions d'un ensemble à a éléments (nombres de Stirling de seconde espèce, [3] p. 32) prolongé par $S(0, 0) \equiv 1$, il vient, avec k et $n \geqq 1$:

$$v(n,k) = \sum_{\substack{0 \leqq a \leqq n \\ 0 \leqq u \leqq \{k/2,\, a\}}} \left\{ \binom{n}{a} \cdot S(a,u) \cdot c(n-a, k-2u) \right\},$$

le facteur $\binom{n}{a}$ correspondant au nombre de choix de $E_2 \subset E$, $|E_2| = a$. En définitive,

PROPOSITION 1. *Les nombres $c(n, k)$ de k-birecouvrements et $v(n, k)$ de k-birevêtements de E, $|E| = n \geqq 1$, $k \geqq 1$, sont liés par:*

$$(1) \qquad v(n,k) = c(n,k) + \sum_{1 \leqq a \leqq n} \left\{ \binom{n}{a} \sum_{1 \leqq u \leqq \{a,\, k/2)} S(a,u) \cdot c(n-a, k-2u) \right\}.$$

Cette formule (1) permet le calcul de proche en proche des $v(n, k)$ en fonction des $c(n, k)$. Pour inverser cette formule, faisons la

CONVENTION DE PROLONGEMENT. Prolongeons les suites doubles

$$S(n,k), \quad c(n,k), \quad v(n,k), \quad \binom{n}{k}$$

de la manière suivante:

$$S(n,k) \equiv 0 \text{ si } k > n \text{ ou si } k \leqq 0 \text{ ou si } n \leqq 0, \text{ sauf } S(0,0) \equiv 1.$$

$$c(n,k) \text{ et } v(n,k) = 0 \text{ si } k \leqq 0 \text{ ou si } n \leqq 0, \text{ sauf } c(0,0) \equiv v(0,0) \equiv 1.$$

$\binom{n}{k} = 0$ si $k > n$ ou si $k < 0$ ou si $n < 0$, avec

$$\binom{n}{0} \equiv \binom{n}{n} \equiv 1, \quad \text{si} \quad n \geqq 0.$$

Compte tenu de cette convention, la formule (1) s'écrit alors:

$$v(n,k)=\sum_{a,\,u\geqq 0}\left\{\binom{n}{a}\cdot S(a,u)\cdot c(n-a,k-2u)\right\},\qquad n,k\geqq 0. \tag{2}$$

Définissons alors les fonctions génératrices formelles $C(y,z)$ et $V(y,z)$ des $c(n,k)$ et $v(n,k)$ par:

$$C(y,z)\equiv\sum_{n,\,k\geqq 0}c(n,k)y^k\frac{z^n}{n!};\quad V(y,z)\equiv\sum_{n,\,k\geqq 0}v(n,k)y^k\frac{z^n}{n!}. \tag{3}$$

(2) et (3) entraînent:

$$V(y,z)=\sum_{n,\,k,\,a,\,u\geqq 0}\binom{n}{a}\cdot S(a,u)\cdot c(n-a,k-2u)\cdot y^k\frac{z^n}{n!}$$

ou encore, en faisant le changement de variable de sommation: $b\equiv n-a$, $w\equiv k-2u$:

$$V(y,z)=\sum_{a,\,b,\,u,\,w\geqq 0}S(a,u)\cdot c(b,w)\frac{z^{a+b}}{a!\,b!}\cdot y^{2u+w}.$$

Utilisant l'identité bien connue ([3], p. 43)

$$\sum_{a\geqq 0}S(a,u)\frac{z^a}{a!}=\frac{(e^z-1)^u}{u!} \tag{4}$$

il vient:

$$V(y,z)=\sum_{b,\,u,\,w\geqq 0}\frac{\{y^2(e^z-1)\}^u}{u!}\cdot c(b,w)\cdot y^w\cdot\frac{z^b}{b!}=$$

$$=\exp\{y^2(e^z-1)\}\sum_{b,\,w\geqq 0}c(b,w)\cdot y^w\cdot\frac{z^b}{b!}.$$

Proposition 2. *Les fonctions génératrices $C(y,z)$ et $V(y,z)$ des $c(n,k)$ et $v(n,k)$, définies en* (3) *sont liées par:*

$$V(y,z)=\exp\{y^2(e^z-1)\}\cdot C(y,z). \tag{5}$$

Il est alors facile d'inverser (1); en effet, d'après (4) et (5):

$$C(y,z)=\exp\{-y^2(e^z-1)\}\cdot\sum_{n,\,k\geqq 0}v(n,k)\cdot y^k\cdot\frac{z^n}{n!}=$$

$$=\sum_{n,\,k,\,s\geqq 0}(-1)^s\frac{y^{2s}(e^z-1)^s}{s!}v(n,k)\cdot y^k\cdot\frac{z^n}{n!}=$$

$$=\sum_{n,\,k,\,s,\,t\geqq 0}(-1)^s S(t,s)\cdot v(n,k)\cdot y^{2s+k}\cdot\frac{z^{n+t}}{n!\,t!}.$$

Faisons le changement de variables de sommations $N \equiv n+t$, $K \equiv 2s+k$ et identifions les coefficients de $y^K \cdot \frac{z^N}{N!}$ de chaque membre:

$$c(N, K) = \sum_{s,\, t \geqq 0} \binom{N}{t} (-1)^s \cdot S(t, s) \cdot v(N-t, K-2s),$$

ou encore, après le nouveau changement de variables $n \equiv N$, $k \equiv K$, $a \equiv t$, $u \equiv s$ et avec notre *convention:*

PROPOSITION 3. *On peut calculer les* $c(n, k)$ *à partir des* $v(n, k)$ *par la formule suivante, inverse de* (1), n *et* $k \geqq 1$:

$$(6) \qquad c(n, k) = v(n, k) + \sum_{1 \leqq a \leqq n} \left\{ \binom{n}{a} \sum_{1 \leqq u \leqq \{a,\, k/2\}} (-1)^u \cdot S(a, u) \cdot v(n-a, k-2u) \right\}.$$

3. Récurrence sur les nombres $c(n, k)$ de k-birecouvrements de E, $|E| = n$

Tout ce paragraphe n'utilise que des observations élémentaires sans le moindre recours à la théorie des fonctions génératrices.

(I) *Notations.* Il est important, dans ce qui suit, de ne pas oublier la *convention* faite plus haut. Une légère réflexion montre d'abord que: $c(i, 0) = c(i, 1) = c(i, 2) = 0$ pour $i \geqq 1$ et que $c(1, j) = 0$ si $j \geqq 1$. Adjoignons à E, $|E| = n \geqq 1$, un $(n+1)^{\text{ème}}$ élément x et soit $F \equiv E \cup \{x\}$. Considérons l'ensemble $\mathbf{c}(F, k+1)$, $k \geqq 3$ (définition 2) des birecouvrements de F, à $(k+1)$ blocs. Pour $\mathscr{R} \in \mathbf{c}(F, k+1)$, soient $A(\mathscr{R})$ et $B(\mathscr{R})$ les deux blocs (distincts) de $\mathscr{R}$, qui contiennent x, leur ordre n'intervenant évidemment pas; soient aussi $A_0(\mathscr{R}) \equiv A(\mathscr{R}) \setminus \{x\}$, $B_0(\mathscr{R}) \equiv B(\mathscr{R}) \setminus \{x\}$. Partageons alors $\mathbf{c}(F, k+1)$ en les cinq classes disjointes suivantes:

$$\mathbf{a} \equiv \{\mathscr{R} | \mathscr{R} \in \mathbf{c}(F, k+1);\ A_0(\mathscr{R}) \neq \emptyset,\ A_0(\mathscr{R}) \notin \mathscr{R};\ B_0(\mathscr{R}) = \emptyset\}$$

$$\mathbf{b} \equiv \{\mathscr{R} |\ d^\circ\ ;\ A_0(\mathscr{R}) \neq \emptyset,\ A_0(\mathscr{R}) \notin \mathscr{R};\ B_0(\mathscr{R}) \neq \emptyset,\ B_0(\mathscr{R}) \notin \mathscr{R}\}$$

$$\mathbf{d} \equiv \{\mathscr{R} |\ d^\circ\ ;\ A_0(\mathscr{R}) \in \mathscr{R};\ B_0(\mathscr{R}) = \emptyset\}$$

$$\mathbf{e} \equiv \{\mathscr{R} |\ d^\circ\ ;\ A_0(\mathscr{R}) \in \mathscr{R};\ B_0(\mathscr{R}) \neq \emptyset,\ B_0(\mathscr{R}) \notin \mathscr{R}\}$$

$$\mathbf{f} \equiv \{\mathscr{R} |\ d^\circ\ ;\ A_0(\mathscr{R}) \in \mathscr{R};\ B_0(\mathscr{R}) \in \mathscr{R}\}$$

L'ordre de $A_0(\mathscr{R})$ et $B_0(\mathscr{R})$ n'intervenant pas, il est clair que tous les cas possibles pour les $\mathscr{R}$ ont été épuisés, donc que:

$$(7) \qquad c(n+1, k+1) = |\mathbf{c}(F, k+1)| = |\mathbf{a}| + |\mathbf{b}| + |\mathbf{d}| + |\mathbf{e}| + |\mathbf{f}|.$$

(II) *Calcul de* $|\mathbf{a}|$. Soit α la fonction définie sur $\mathbf{a}$ et à valeurs dans $\mathbf{c}(E, k)$, telle que, pour $\mathscr{R} \in \mathbf{a}$: $\alpha(\mathscr{R}) \equiv$ trace de $\mathscr{R}$ sur $E = \{R \cap E | R \in \mathscr{R}\}$. α est visiblement surjective: tout $\mathscr{R}' \in \mathbf{c}(E, k)$ est atteint. Posons alors, comme d'habitude, pour $\mathscr{R}' \in \mathbf{c}(E, k)$:

$$\alpha^{-1}(\mathscr{R}') \equiv \{\mathscr{R} | \mathscr{R} \in \mathbf{a},\ \alpha(\mathscr{R}) = \mathscr{R}'\}$$

et cherchons le nombre d'éléments de $\alpha^{-1}(\mathscr{R}')$. Pour construire les $\mathscr{R} \in \alpha^{-1}(\mathscr{R}')$ à partir de $\mathscr{R}'$, il suffit de choisir un bloc $B \in \mathscr{R}'$, puis de le border par x, c'est-à-dire de le remplacer par $B \cup \{x\}$; comme $\mathscr{R}'$ possède k blocs, il y a k choix possibles du bloc B; donc $|\alpha^{-1}(\mathscr{R}')| = k$ pour tout $\mathscr{R}' \in \mathbf{c}(E, k)$. Ainsi:

$$|\mathbf{a}| = \sum_{\mathscr{R}' \in \mathbf{c}(E,\, k)} |\alpha^{-1}(\mathscr{R}')| - k \cdot c(n, k).$$

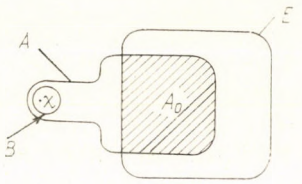


Abb. 1

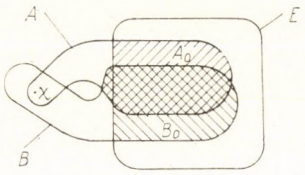


Abb. 2

(III) *Calcul de* $|\mathbf{b}|$. Définissons sur $\mathbf{b}$ la fonction β à valeurs dans $\mathbf{c}(E, k+1)$ par: $\beta(\mathscr{R}) \equiv$ trace de $\mathscr{R}$ sur $E = \{R \cap E | R \in \mathscr{R}\}$. β est surjective et, pour $\mathscr{R}' \in \mathbf{c}(E, k+1)$, on a $|\beta^{-1}(\mathscr{R}')| = \binom{k+1}{2}$; en effet, pour construire tous les $\mathscr{R} \in \mathbf{b}$ tels que $\beta(\mathscr{R}) = \mathscr{R}'$, il suffit de choisir une paire (non ordonnée) de blocs de $\mathscr{R}'$, puis de les border chacun par x. Ainsi:

$$|\mathbf{b}| = \sum_{\mathscr{R}' \in \mathbf{c}(E,\, k+1)} |\beta^{-1}(\mathscr{R}')| = \binom{k+1}{2} \cdot c(n, k+1).$$

(IV) *Calcul de* $|\mathbf{d}|$. Pour $A_0 \subset E$, $A_0 \neq \emptyset$, soit $\mathbf{d}(A_0)$ l'ensemble des recouvrements $\mathscr{R} \in \mathbf{d}$ tels que $A_0(\mathscr{R}) = A_0$ (et que $A_0 \in \mathscr{R}$). Définissons alors sur $\mathbf{d}(A_0)$ la fonction δ_{A_0}, à valeurs dans $\mathbf{c}(E \backslash A_0, k-2)$ par: $\delta_{A_0}(\mathscr{R}) \equiv$ trace de $\mathscr{R}$ sur $E \backslash A_0 = \ldots$; δ_{A_0} est surjective, et pour tout $\mathscr{R}' \in \mathbf{c}(E \backslash A_0, k-2)$, on a $|\delta_{A_0}^{-1}(\mathscr{R}')| = 1$ évidemment. Donc, en posant $u \equiv |A_0|$, $1 \leqq u \leqq n$, il vient:

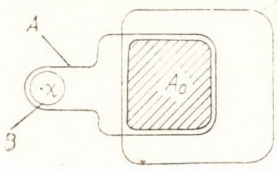


Abb. 3

$$|\mathbf{d}(A_0)| = \sum_{\mathscr{R}' \in \mathbf{c}(E \backslash A_0,\, k-2)} |\delta_{A_0}^{-1}(\mathscr{R}')| = c(n-u, k-2).$$

c'est-à-dire:

$$|\mathbf{d}| = \sum_{A_0 \in \mathscr{P}'(E)} |\mathbf{d}(A_0)| = \sum_{1 \leqq u \leqq n} \Big\{ \sum_{|A_0| = u} |\mathbf{d}(A_0)| \Big\} = \sum_{1 \leqq u \leqq n} \binom{n}{u} \cdot c(n-u, k-2).$$

(V) *Calcul de* $|\mathbf{e}|$. Dans ce cas, $A_0(\mathscr{R}) \cap B_0(\mathscr{R}) = \emptyset$, puisque, s'il existait un élément y commun à $A_0(\mathscr{R})$ et $B_0(\mathscr{R})$, il appartiendrait à $A(\mathscr{R})$: il serait donc trois fois recouvert et $\mathscr{R}$ ne serait plus un birecouvrement. Pour $A_0 \subset E$, $A_0 \neq \emptyset$, soit $\mathbf{e}(A_0)$ l'ensemble des birecouvrements $\mathscr{R} \in \mathbf{e}$ tels que $A_0(\mathscr{R}) = A_0$ (et que $A_0 \in \mathscr{R}$). Définissons alors sur $\mathbf{e}(A_0)$ la fonction ε_{A_0} à valeurs dans $\mathbf{c}(E \backslash A_0, k-1)$ par: $\varepsilon_{A_0}(\mathscr{R}) \equiv$ trace de $\mathscr{R}$ sur $E \backslash A_0 = \ldots$; ε_{A_0} est surjective, et pour tout $\mathscr{R}' \in \mathbf{c}(E \backslash A_0, k-1)$,

on a $|\varepsilon_{A_0}^{-1}(\mathscr{R}')| = k-1$, puisque chaque $\mathscr{R} \in \varepsilon_{A_0}^{-1}(\mathscr{R}')$ s'obtient en choisissant l'un des $(k-1)$ blocs de $\mathscr{R}'$, en le bordant par x, et en adjoignant au système ainsi obtenu les deux blocs A_0 et $A_0 \cup \{x\}$. Donc, en posant $|A_0| = u$, $1 \leqq u \leqq n$,

$$|\mathbf{e}(A_0)| = \sum_{\mathscr{R}' \in \mathbf{c}(E \setminus A_0,\, k-1)} |\varepsilon_{A_0}^{-1}(\mathscr{R}')| = (k-1) \cdot c(n-u, k-1),$$

c'est-à-dire:

$$|\mathbf{e}| = \sum_{A_0 \in \mathscr{P}'(E)} |\mathbf{e}(A_0)| = \sum_{1 \leqq u \leqq n} \sum_{|A_0| = u} |\mathbf{e}(A_0)| = \sum_{1 \leqq u \leqq n} \binom{n}{u} \cdot (k-1) \cdot c(n-u, k-1).$$

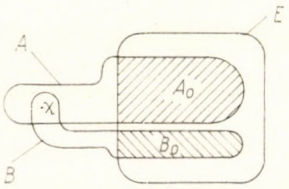


Abb. 4

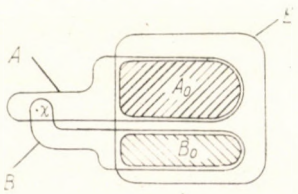


Abb. 5

(VI) *Calcul de* $|\mathbf{f}|$. On a encore $A_0(\mathscr{R}) \cap B_0(\mathscr{R}) = \emptyset$. Pour $A_0 \subset E$, $B_0 \subset E$, $A_0 \neq \emptyset$, $B_0 \neq \emptyset$, $A_0 \cap B_0 = \emptyset$, la paire (A_0, B_0) n'étant pas ordonnée, soit $\mathbf{f}(A_0, B_0)$ l'ensemble des $\mathscr{R} \in \mathbf{f}$ tels que $A_0(\mathscr{R}) = A_0$ et $B_0(\mathscr{R}) = B_0$ (alors $A_0, B_0 \in \mathscr{R}$). Définissons sur $\mathbf{f}(A_0, B_0)$ la fonction φ_{A_0, B_0} à valeurs dans $\mathbf{c}(E \setminus (A_0 \cup B_0), k-3)$ par:

$$\varphi_{A_0, B_0}(\mathscr{R}) \equiv \text{trace de } \mathscr{R} \text{ sur } E \setminus (A_0 \cup B_0) = \ldots$$

On voit aisément que $|\varphi_{A_0, B_0}^{-1}(\mathscr{R}')| = 1$, donc que $|\mathbf{f}(A_0, B_0)| = c(n-v, k-3)$ où l'on a posé: $v \equiv |A_0| + |B_0|$; il s'ensuit, après un calcul facile:

$$|\mathbf{f}| = \sum_{\substack{A_0, B_0 \in \mathscr{P}'(E) \\ (A_0, B_0) \text{ non ordonné};\ A_0 \cap B_0 \neq \emptyset}} c(n-v, k-3) = \sum_{1 \leqq v \leqq n} (2^{v-1} - 1) \cdot \binom{n}{v} \cdot c(n-v, k-3)$$

(VII) *Récapitulation.* (7) entraîne

$$c(n+1, k+1) = kc(n, k) + \frac{k(k+1)}{2} \cdot c(n, k+1) +$$

$$+ \sum_{1 \leqq u \leqq n} \binom{n}{u} \{(k-1) \cdot c(n-u, k-1) + c(n-u, k-2)\} +$$

$$+ \sum_{1 \leqq v \leqq n} (2^{v-1} - 1) \binom{n}{v} \cdot c(n-v, k-3),$$

ou encore, en réunissant les deux Σ en un seul:

PROPOSITION 4. *Le nombre $c(n, k)$ de k-birecouvrements de E, $|E| = n \geqq 1$ satisfait la relation de récurrence suivante, où $k \geqq 3$:*

$$c(n+1, k+1) = kc(n, k) + \frac{k(k+1)}{2} \cdot c(n, k+1) +$$

$$+ \sum_{0 \leqq v \leqq n-1} \binom{n}{v} \{(k-1) \cdot c(v, k-1) + c(v, k-2) + (2^{n-v-1} - 1) \cdot c(v, k-3)\}. \tag{8}$$

Observons enfin que la méthode employée permet aussi, de proche en proche, *l'énumération* des systèmes $\mathscr{R} \in \mathbf{c}(E, k)$.

4. Valeurs de $c(n, k)$

Soit $k(n)$ le plus grand entier k tel que $c(n, k) \neq 0$; montrons que $k(n) = [\frac{3}{2} n]$, ou $[z]$ désigne le plus grand entier $\leqq z$. Celà revient à prouver que pour tout birecouvrement $\mathscr{R}$ de E, $|E| = n$, on a $|\mathscr{R}| \leqq \frac{3}{2} n$, et qu'il existe un birecouvrement ayant $[\frac{3}{2} n]$ blocs. Pour celà, associons à tout $\mathscr{R} \in \mathbf{c}(E)$ et tout entier $h \geqq 0$, le système $\mathscr{R}_h$:

$$\mathscr{R}_h \equiv \{R | R \in \mathscr{R}, |R| = h\}.$$

Utilisant les deux égalités suivantes (la seconde provient de ce que chaque $x \in E$ est deux fois recouvert par $\mathscr{R}$):

$$|\mathscr{R}| = \sum_{h \geqq 1} |\mathscr{R}_h|, \quad 2n = \sum_{R \in \mathscr{R}} |R|,$$

il vient:

$$2n = \sum_{R \in \mathscr{R}} |R| = \sum_{h \geqq 1} \sum_{R \in \mathscr{R}_h} |R| = \sum_{h \geqq 1} h |\mathscr{R}_h| \geqq |\mathscr{R}_1| + 2 \sum_{h \geqq 2} |\mathscr{R}_h| =$$

$$= |\mathscr{R}_1| + 2(|\mathscr{R}| - |\mathscr{R}_1|) = 2|\mathscr{R}| - |\mathscr{R}_1|;$$

donc, puisque $|\mathscr{R}_1| \leqq n$:

$$2|\mathscr{R}| \leqq 2n + |\mathscr{R}_1| \leqq 3n, \qquad \text{q.e.d.}$$

La valeur $\left[\frac{3n}{2}\right]$ est atteinte par $|\mathscr{R}|$: il suffit pour celà d'envisager, pour $E = \{x_1, x_2, \ldots, x_n\}$, le recouvrement $\mathscr{R}_0$ suivant, défini par ses blocs:

$$\mathscr{R}_0 \equiv \begin{cases} \{x_1\}, \{x_2\}, \ldots, \{x_n\}; \{x_1, x_2\}, \{x_3, x_4\}, \ldots, \{x_{n-1}, x_n\} & \text{si } n \text{ est pair,} \\ d^\circ \quad ; \{x_1, x_2\}, \ldots, \{x_{n-4}, x_{n-3}\}, \{x_{n-2}, x_{n-1}, x_n\} & \text{si } n \text{ est impair.} \end{cases}$$

De ce résultat se déduit sans peine que $n(k)$, plus petit entiern tel que $c(n, k) \neq 0$, vaut $\lfloor \frac{2}{3} k \rfloor$, où $\lfloor z \rfloor$ désigne le plus petit entier $\geqq z$. Révélons les valeurs de $c(n, k)$ et celles de $c(n) = \sum_{k \geqq 3} c(n, k)$, $2 \leqq n \leqq 7$ (voir page 145).

Enfin, la formule (8) fournit facilement:

$$c(n, 3) = (1/2)(3^{n-1} - 1); \quad c(n, 4) = (1/2)(3^{n-1} - 1)(2^{n-2} - 1);$$

$$c(2v, 3v) = (2v-1)!!$$

n \ k	3	4	5	6	7	8	9	10...	$c(n)$
2	1	4							1
3	4	39	25	3					8
4	13	280	472	256	40				80
5	40	1815	6185	7255	3306	535	15		1088
6	121	11284	70700	149660	131876	51640	8456	420	19232
7	364								424400
⋮									

5. Calcul des nombres $v(n, k)$ de k-birevêtements de E, $|E| = n$, par le théoreme de Pólya—De Bruijn

(1) et (8) permettent le calcul de proche en proche des $v(n, k)$ à partir des $c(n, k)$. Nous allons cependant établir une formule donnant $v(n, k)$ sous forme compacte; nous aurons par là un procédé de vérification des valeurs de $c(n, k)$ déjà trouvées; de plus cette formule sera utilisée ultérieurement pour l'estimation de $c(n)$; enfin, elle fournira parmi tant d'autres un exemple d'application du grand théorème de Pólya—De Bruijn, que nous commençons par rappeler ([4] p. 162 et [5]):

Théorème de Pólya—De Bruijn. *Soient deux ensembles finis D et R; G (resp. H) un groupe de permutations de D (resp. de R); $\mathcal{F}$ l'ensemble des applications de D dans R; $\hat{\mathcal{F}}$ l'ensemble quotient de $\mathcal{F}$ par la relation d'équivalence:*

$$f_1 \sim f_2 \Leftrightarrow \exists g \in G,\ \exists h \in H \quad \text{telles que} \quad f_1 g = h f_2 .$$

Soit W une application qui á toute $f \in \mathcal{F}$ associe un entier $W(f)$ — le poids de f — et telle que:

$$f_1 \sim f_2 \Rightarrow W(f_1) = W(f_2), \tag{9}$$

ce qui légitime la définition du poids $W(F)$ d'une classe d'équivalence (ou modèle) $F \in \hat{\mathcal{F}}$ par:

$$W(F) \equiv W(f), \quad f \in F.$$

Soit aussi $i(g, h)$ la somme des poids des fonctions $f \in \mathcal{F}$ telles que $fg = hf$, ce que l'on note:

$$i(g, h) \equiv \sum_{f}^{(g,h)} W(f). \tag{10}$$

De toutes ces hypothèses s'ensuit que l'„inventaire" des modèles vaut:

$$I(\hat{\mathcal{F}}) \equiv \sum_{F \in \hat{\mathcal{F}}} W(F) = \frac{1}{|G| \cdot |H|} \sum_{g \in G,\, h \in H} i(g, h). \tag{11}$$

Spécialisons ce théorème de De Bruijn au problème qui nous préoccupe. Soit $D \equiv \{a_1, a_1', a_2, a_2', \ldots, \ldots, a_n, a_n'\}$ un ensemble à $2n$ éléments associés 2 à 2, a_i et

a_i' étant dits *homologues* ($1 \leqq i \leqq n$); soit G le groupe des permutations de D engendré par les n transpositions $(a_i a_i')$, $1 \leqq i \leqq n$; donc $|G| = 2^n$; soit $R \equiv \{1, 2, \ldots, k\}$; soit enfin H le groupe symétrique de R; donc $|H| = k!$. Il apparaît que la donnée de $F \in \hat{\mathscr{F}}$ équivaut à la donnée d'un agrégat de $E \equiv \{a_1, a_2, \ldots, a_n\}$ ayant *au plus* k *blocs*, et *au plus birecouvrant*, en ce sens que chaque a_i, $1 \leqq i \leqq n$ appartient à 1 ou 2 blocs; en effet, la donnée de $f \in \mathscr{F}$ définit un ensemble ordonné de blocs de D: $f^{-1}(1), f^{-1}(2), \ldots, f^{-1}(k)$ en nombre $\leqq k$, puisque certains $f^{-1}(i)$ peuvent être vides; le groupe G identifie a_1 et a_1', a_2 et a_2', ... donc transforme les blocs précédents en blocs de E, et le groupe H efface le numérotage de ces blocs. Introduisons la condition:

$$f(a_i) \neq f(a_i'),\ 1 \leqq i \leqq n, \tag{12}$$

et définissons le poids $W(f)$ comme étant égal à 1 si f satisfait cette condition (12) et à 0 dans le cas contraire. (On voit sans trop d'effort que W satisfait la condition (9)). Alors l'inventaire des modèles $I(\hat{\mathscr{F}}) = \Sigma\, W(F)$ vaut le nombre des birevêtements de E, ayant au plus k blocs, puisque tout agrégat défini par F et qui ne recouvre pas deux fois chaque $a_i \in E$ a une participation nulle dans l'inventaire: c'est le mérite de la définition de W. En d'autres termes, posant:

$$\bar{v}(n, k) \equiv v(n, 1) + v(n, 2) + \ldots + v(n, k),$$

on a, d'après le théorème de PÓLYA—DE BRUIJN:

$$I(\hat{\mathscr{F}}) = \frac{1}{|G| \cdot |H|} \sum_{g \in G,\, h \in H} i(g, h) = \bar{v}(n, k). \tag{13}$$

Passons maintenant au calcul effectif de $i(g, h)$ pour $g \in G$, $h \in H$. Supposons que g est du type $(1^{2s} 2^t)$, $2s + 2t = |D| = 2n$, $s \geqq 0$, $t \geqq 0$, et que g est du type $(1^{c_1}\, 2^{c_2}\, 3^{c_3}\, \ldots)$, $c_1 + 2c_2 + 3c_3 + \ldots = k$ ([3], p. 67); compte tenu de (10), (12), $i(g, h)$ est exactement le nombre des $f \in \mathscr{F}$ telles que, pour tout $d \in D$, on ait $fg(d) = hf(d)$ et $f(a_i) \neq f(a_i')$, $1 \leqq i \leqq n$. Or deux cas peuvent se présenter pour d:

(1°) d appartient à un 1-cycle de g: $g(d) = d$, et la condition $fg(d) = hf(d)$ implique que $f(d) = hf(d)$, ce qui prouve que $f(d)$ appartient lui aussi à un 1-cycle de h; par ailleurs, la condition (12) exige que, pour l'homologue d' de d, on ait $f(d) \neq f(d')$; en conséquence, l'image de tout couple (d, d') de points homologues invariants est un couple ordonné de 2 points (distincts) appartenant l'un et l'autre aux 1-cycles de h.

(2°) d appartient à un 2-cycle de h: $g(d) \neq d$, $g^2(d) = d$, et la condition $fg(d) = hf(d)$ combinée avec (12) implique que $hf(d) \neq f(d)$; par ailleurs $h^2 f(d) = h(hf)d = (hf)g(d) = fg^2(d) = f(d)$, ce qui prouve que $f(d)$ appartient à un 2-cycle de h.

Réciproquement, on voit facilement que toute fonction f transformant les 1-cycles de g en des 1-cycles de h avec la condition $f(d) \neq f(d')$, et transformant les 2-cycles de g en des 2-cycles de h, satisfait les conditions $fg = hf$ et $f(a_i) \neq f(a_i')$, $1 \leqq i \leqq n$. Compte tenu de l'ordre à choisir sur les 2-cycles de h, et de la condition exprimée à la fin du (1°) ci-dessus, le nombre des choix pour f est:

$$i(g, h) = \{c_1(c_1 - 1)\}^s \{2c_2\}^t.$$

Ainsi, d'après (13) et [3], p. 67:

$$\bar{v}(n,k) = \frac{1}{2^n \cdot k!} \sum_{\substack{g \in G \\ h \in H}} \{c_1(c_1-1)\}^s \{2c_2\}^t =$$

$$= \frac{1}{2^n \cdot k!} \sum_{h \in H} \sum_{s+t=n} \binom{n}{s} \{c_1(c_1-1)\}^s \{2c_2\}^t = \frac{1}{2^n \cdot k!} \sum_{h \in H} \{c_1(c_1-1)+2c_2\}^n =$$

$$= \frac{1}{2^n \cdot k!} \sum_{c_1+2c_2+\ldots+kc_k=k} \{c_1(c_1-1)+2c_2\}^n \frac{k!}{1^{c_1} c_1! \, 2^{c_2} c_2! \ldots k^{c_k} c_k!} =$$

$$= \frac{1}{2^n \cdot k!} \sum_{c_1+2c_2+v=k} \frac{x_v}{v!} \frac{\{c_1(c_1-1)+2c_2\}^n}{c_1! \, 2^{c_2} c_2!},$$

où l'on a posé

$$x_v \equiv \sum_{3c_3+4c_4+\ldots=k} \frac{v!}{3^{c_3} c_3! \, 4^{c_4} c_4! \ldots}.$$

x_v est précisément égal au nombre de permutations sans 1-cycle ni 2-cycle d'un ensemble à v éléments. En utilisant [3], p. 70, on trouve facilement que:

$$\sum_{v \geqq 0} x_v \frac{t^v}{v!} = (1-t)^{-1} \exp\left\{-t - \frac{t^2}{2}\right\}; \tag{14}$$

d'où une formule compacte pour $\bar{v}(n,k)$:

$$\bar{v}(n,k) = \frac{1}{2^n} \sum_{0 \leqq v \leqq k} \left\{ \frac{x_v}{v!} \sum_{c_1+2c_2=k-v} \frac{\{c_1(c_1-1)+2c_2\}^n}{2^{c_2} c_1! \, c_2!} \right\}, \tag{15}$$

les x_v étant fournis par (14).

Il s'en déduit une formule pour $v(n,k) = \bar{v}(n,k) - \bar{v}(n,k-1)$ après quelques manipulations simples:

PROPOSITION 5. *Le nombre $v(n,k)$ des k-birevêtements de E, $|E| = n \geqq 0$, est donné par la formule suivante, où $k \geqq 0$:*

$$v(n,k) = \frac{1}{2^n} \sum_{0 \leqq v \leqq k} \left\{ \frac{y_v}{v!} \sum_{c_1+2c_2=k-v} \frac{\{c_1(c_1-1)+2c_2\}^n}{2^{c_2} c_1! \, c_2!} \right\},$$

avec

$$\sum_{v \geqq 0} y_v \frac{t^v}{v!} = \exp\left\{-t - \frac{t^2}{2}\right\}.$$

Livrons les valeurs de $v(n,k)$ et celles de $v(n) = \sum_k v(n,k)$, $1 \leqq n \leqq 7$:

n \ k	2	3	4	5	6	7	8	9	10	11	12	13	14...	$v(n)$
1	1													1
2	1	1	1											3
3	1	4	7	3	1									16
4	1	13	46	47	25	6	1							139
5	1	40	295	587	516	235	65	10	1					1750
6	1	121	1846	6715	9690	7053	3006	800	140	15	1			29388
7	2	364	11347	73003	170051	189458	118231	46795	12201	2170	266	21	1	623909

6. Estimations de $c(n)$ et $v(n)$

Il est clair que:

$$c(n,k) \leqq v(n,k); \quad c(n) \leqq v(n) \qquad (n, k \geqq 0) . \tag{16}$$

Etablissons d'abord une minoration de $c(n)$. La formule (8) montre que:

$$c(n,k) \geqq \frac{k(k-1)}{2}\, c(n-1,k)$$

$$c(n(k)+1,1) \geqq \frac{k(k-1)}{2}\, c(n(k),k), \quad n \geqq 2,\ n \geqq n(k),\ k \geqq 3,$$

où $n(k)$ désigne le plus petit entier n tel que $c(n,k) \neq 0$, donc $\geqq 1$: d'après le paragraphe 4, $n(k) = \lfloor \frac{2}{3} k \rfloor$.

Multiplions membre à membre les $(n-n(k))$ inégalités ci-dessus; il vient

$$c(n,k) \geqq \left\{\frac{k(k-1)}{2}\right\}^{n-n(k)} \geqq \left\{\frac{k(k-1)}{2}\right\}^{n-\frac{2}{3}k-1} \equiv \gamma(n,k)$$

$$c(n) = \sum_{3 \leqq k \leqq \left\lfloor \frac{3}{2} n \right\rfloor} c(n,k) \geqq \sum \gamma(n,k) \geqq \gamma(n,k_n),$$

où $k_n \equiv \left\lfloor \dfrac{3n}{2 \log n} \right\rfloor$ est proche de l'abcisse du maximum de la fonction $\gamma(n,t)$ de la variable t. On trouve, après un calcul facile que

$$\log \gamma\,(n,k_n) = (2n \log n)(1 + o\,[1]);$$

donc

$$\log c(n) \geqq (2n \log n)(1 + o\,[1]) \qquad (n \to \infty). \tag{17}$$

Etablissons maintenant une majoration de $v(n)$. Tout birevêtement ayant au plus $2n$ blocs, on aura

$$v(n) = \bar{v}(n, 2n),$$

donc, d'après (15):

$$v(n) = \frac{1}{2^n} \sum_{0 \leqq v \leqq 2n} \left\{ \frac{x_v}{v!} \sum_{c_1 + 2c_2 = 2n - v} \frac{(c_1(c_1-1) + 2c_2)^n}{2^{c_2} c_1!\, c_2!} \right\}$$

Posons $\mu \equiv 2n - v$ et $\dfrac{h_\mu}{\mu!} \equiv \sum_{c_1 + 2c_2 = \mu} \dfrac{1}{2^{c_2} c_1!\, c_2!}$.

Un calcul simple prouve que

$$(1^\circ) \max_{c_1 + 2c_2 = \mu} \{c_1(c_1-1) + 2c_2\} \leqq \mu^2; \quad (2^\circ) \sum_{\mu \geqq 0} h_\mu \frac{t^\mu}{\mu!} = \exp\left(t + \frac{t^2}{2}\right). \tag{18}$$

En conséquence:

$$v(n) \leqq \frac{1}{2^n} \sum_{\mu + \nu = 2n} \frac{x_\nu h_\mu}{\nu!\, \mu!} \cdot \mu^{2n} \leqq \frac{(2n)^{2n}}{2^n} \sum_{\mu + \nu = 2n} \frac{x_\nu}{\nu!} \cdot \frac{h_\mu}{\mu!} .$$

Le dernier Σ est égal (voir (14), (18)) au coefficient de t^{2n} dans $(1-t)^{-1}$, soit 1; donc

$$v(n) \leqq 2^n \cdot n^{2n};$$

d'où résulte, d'après (16):

$$\log c(n) \leqq \log v(n) \leqq (2n \log n)(1 + o\,[1]) \qquad (n \to \infty). \tag{19}$$

(17) et (19) impliquent en définitive:

PROPOSITION 6. *Le nombre $c(n)$ de birecouvrements de E et le nombre $v(n)$ de birevêtements de E, $|E| = n$, sont tels que:*

$$\log c(n) \sim \log v(n) \sim 2n \log n \qquad (n \to \infty). \tag{20}$$

7. Fonctions génératrices des $c(n)$ et $v(n)$

(I) *Recouvrement $\varrho(\mathscr{S})$ associé à un birecouvrement $\mathscr{S}$; les nombres $c(n, k, a)$.*

Soit $\mathscr{S}$ un birecouvrement de E, $|E| = n$, ayant k blocs $(\mathscr{S} \in \mathbf{c}(E, k))$ que nous numérotons de 1 à k; $\mathscr{S}$ est ainsi transformé en un birecouvrement ordonné que nous désignons par $\bar{\mathscr{S}}$:

$$\bar{\mathscr{S}} \equiv \{S_1, S_2, \ldots, S_k\}.$$

Associons à $\mathscr{S}$ la partition $\pi(\mathscr{S})$ de E dont les blocs sont les classes de l'équivalence $\alpha_{\mathscr{S}}$ sur E, définie pour $x, x' \in E$ par:

$$x \overset{\alpha}{\approx} x' \Leftrightarrow x' \in \alpha_{\mathscr{S}}(x) \equiv \Big(\bigcap_{x \in M \in \mathscr{S}} M\Big) \cap \Big(\bigcap_{x \notin N \in \mathscr{S}} (E \backslash N)\Big).$$

Appelons *atome* chaque bloc de $\pi(\mathscr{S})$ et supposons qu'il y en ait le nombre a:

$$1 \leqq a = |\pi(\mathscr{S})| \leqq n.$$

Soit aussi K l'ensemble des k premiers nombres entiers:

$$K \equiv \{1, 2, \ldots, k\}.$$

Nous pouvons maintenant associer au système ordonné $\bar{\mathscr{S}}$ de E un certain système (non ordonné) $\varrho(\bar{\mathscr{S}})$ de K, constitué uniquement des *paires* (i.e. blocs à 2 éléments) définies ainsi:

i et j appartiennent à une paire de $\varrho(\bar{\mathscr{S}})$ $(i \neq j) \Leftrightarrow S_i \cap S_j \neq \emptyset$.

Ce nouveau système $\varrho(\bar{\mathscr{S}})$ de K a les 3 propriétés suivantes:

(1) $\varrho(\bar{\mathscr{S}})$ est un recouvrement en paires de K: ceci résulte de ce que $\mathscr{S}$ est un birecouvrement de E.

(2) Il y a correspondance biunivoque entre les paires qui constituent $\varrho(\bar{\mathscr{S}})$ et les atomes de $\pi(\mathscr{S})$; donc $|\varrho(\bar{\mathscr{S}})| = |\pi(\mathscr{S})| = a$.

(3) Aucune paire n'est isolée en ce sens que toute paire en rencontre au moins une autre: ceci résulte de ce que tous les blocs du birecouvrement $\mathscr{S}$ sont distincts.

Observons d'ailleurs que $\varrho(\overline{\mathscr{S}})$ pourrait être considérée comme un graphe de K.

Soit $\mathbf{r}(K, a)$ l'ensemble des recouvrements $\mathscr{R}$ de K ayant les 3 propriétés ci-dessus, et soit $r(k, a) = |\mathbf{r}(K, a)|$. Tout $\mathscr{R} \in \mathbf{r}(K, a)$ est l'image d'un nombre de k-birecouvrements $\mathscr{S} \in \mathbf{c}(E, k)$ égal à:

$$c(n, k, a) \equiv \frac{1}{k!} \cdot S(n, a) \cdot r(k, a) \cdot a! ; \tag{21}$$

en effet, il y a $S(n, a)$ choix possibles pour la partition $\pi(\mathscr{S})$, puis $r(k, a)$ choix pour le recouvrement $\mathscr{R} = \varrho(\overline{\mathscr{S}}) \in \mathbf{r}(K, a)$ de K, et $a!$ choix pour la bijection entre $\pi(\mathscr{S})$ et $\varrho(\overline{\mathscr{S}})$; enfin le terme $1/k!$ provient de l'effacement du numérotage des blocs de $\mathscr{S}$. En définitive, les choix précédents étant indépendants, *le nombre* $c(n, k, a)$ *des* k*-birecouvrements de* E*, ayant* a *atomes,* $|E| = n$, vaut bien la valeur indiquée en (21). En faisant une convention analogue à celle du paragraphe 2, il en résulte que:

$$c(n, k) = \sum_{a \geqq 0} c(n, k, a) = \sum_{a \geqq 0} \frac{a!}{k!} S(n, a) \cdot r(k, a)$$

(II) *Calcul de* $r(k, a)$. Désignons par $\mathbf{r}'(K, a)$ l'ensemble des recouvrements de K avec a paires non nécessairement isolées, et posons $r'(k, a) \equiv |\mathbf{r}'(K, a)|$, $|K| = k$. Ce nombre $r'(k, a)$ se calcule facilement par la méthode que nous avons donnée en [6]; on trouve:

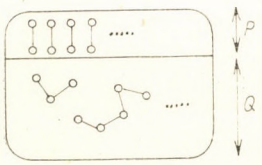


Abb. 6

$$r'(k, a) = \sum_{0 \leqq s \leqq k} (-1)^{k-s} \cdot \left(\binom{s}{2} \right) \binom{k}{s}, \qquad k, a \geqq 0. \tag{22}$$

Dans ces conditions, pour $\mathscr{R}' \in \mathbf{r}'(K, a)$, posons $P \equiv \bigcup_{M \in \mathscr{R}'} M$, les paires M étant toutes *isolées*, $p \equiv |P|$, et $Q \equiv \bigcup_{N \in \mathscr{R}'} N$, aucune des paires N n'étant isolées, $q \equiv |Q|$. Un calcul facile prouve que:

$$r'(k, a) = \sum_{\substack{2p+q=k \\ p+r=a}} \frac{k!}{2^p \cdot p!\, q!} r(q, r), \qquad k, a \geqq 0$$

où les $r(q, r)$ ont été définis dans le paragraphe (7. (I)).

Inversant cette formule par les techniques habituelles de fonction génératrice, et compte tenu de (22), il vient:

$$r(k, a) = \sum_{\substack{2p+q=k \\ p+r=a}} (-1)^p \cdot \frac{k!}{2^p \cdot p!\, q!} r'(q, r) =$$

$$= k! \cdot \sum_{\substack{2p+q=k \\ p+r=a \\ 0 \leqq s \leqq k}} \frac{(-1)^{p+q-s}}{2^p \cdot p!\, q!} \cdot \binom{q}{s} \cdot \left(\binom{s}{2} \right), \qquad k, a \geqq 0. \tag{23}$$

(21) et (23) impliquent finalement:

$$(24)\qquad c(n,k,a)=\sum_{\substack{2p+q=k\\ p+r=a\\ 0\leqq s\leqq k}} a!\,S(n,a)\frac{(-1)^{p+q-s}}{2^p\cdot p!\,q!}\cdot\binom{q}{s}\cdot\left(\!\!\binom{s}{2}\!\!\right)\cdot\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)\qquad n,k,a\geqq 0.$$

(III) *Valeur de la fonction génératrice* $\Gamma(x,y,z)$ *des* $c(n,k,a)$. Posons

$$(25)\qquad \Gamma(x,y,z)\equiv\sum_{n,k,a\geqq 0} c(n,k,a)\cdot x^a\cdot y^k\cdot\frac{z^n}{n!}.$$

On a successivement, d'après (24), et en prenant pour nouvelle variable de sommation $t\equiv q-s$, p, r, s, n, (auquel cas $a=p+r$, $q=s+t$, $k=2p+s+t$):

$$\Gamma(x,y,z)=\sum_{\substack{n,k,a\geqq 0\\ 2p+q=k,\ p+r=a,\ 0\leqq s\leqq q}} a!\,S(n,a)\frac{(-1)^{p+q+s}}{2^p\cdot p!\,q!}\cdot\binom{q}{s}\cdot\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)\cdot x^a y^k\cdot\frac{z^n}{n!}=$$

$$=\sum_{p,r,s,t,n\geqq 0}(p+r)!\,S(n,p+r)\frac{(-1)^{p+t}}{2^p\cdot p!\,s!\,t!}\cdot\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)\cdot x^{p+r}y^{2p+s+t}\cdot\frac{z^n}{n!}=$$

$$=e^{-y}\sum_{p,r,s\geqq 0}\left\{\frac{(-1)^p}{2^p\cdot p!\,s!}\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)y^{2p+s}x^{p+r}\sum_{n\geqq 0}(p+r)!\,S(n,p+r)\frac{z^n}{n!}\right\}.$$

Le dernier Σ vaut, d'après (4), $(e^z-1)^{p+r}$, d'où

$$\Gamma(x,y,z)=e^{-y}\sum_{p,r,s\geqq 0}\frac{1}{p!\,s!}\cdot\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)\cdot\left\{-\frac{xy^2}{2}(e^z-1)\right\}^p\{x(e^z-1)\}^r\cdot y^s=$$

$$=e^{-y}\cdot\exp\left\{-\frac{xy^2}{2}(e^z-1)\right\}\sum_{s\geqq 0}\left\{\frac{y^s}{s!}\sum_{r\geqq 0}\left(\!\!\binom{\binom{s}{2}}{r}\!\!\right)\{x(e^z-1)\}^r\right\}.$$

Donc formellement:

$$(26)\qquad \Gamma(x,y,z)=\sum_{n,k,a\geqq 0} c(n,k,a)x^a y^k\frac{z^n}{n!}=$$

$$=\exp\left\{-y-\frac{xy^2}{2}(e^z-1)\right\}\sum_{s\geqq 0}\frac{y^s}{s!}\{1+x(e^z-1)\}^{\frac{s(s-1)}{2}}.$$

(IV) *Valeur des fonctions génératrices des* $c(n,k)$, $v(n,k)$, $c(n)$, $v(n)$. Compte tenu de ce que $c(n,k)=\sum_{a\geqq 0} c(n,k,a)$, il vient, d'après (3), (5) et (25):

$$C(y,z)=\sum_{n,\,k\geqq 0} c(n,k)\cdot y^k\frac{z^n}{n!}=\sum_{n,\,k,\,a>0} c(n,k,a)\cdot y^k\frac{z^n}{n!}=\Gamma(1,y,z)$$

$$V(y,z)=\sum_{n,\,k\geqq 0} v(n,k)y^k\frac{z^n}{n!}=\exp\{y^2(e^z-1)\}\cdot\Gamma(1,y,z)$$

Introduisons les fonctions génératrices $\mathscr{C}(z)$ et $\mathscr{V}(z)$ des $c(n)$ et $v(n)$:

$$\mathscr{C}(z)\equiv\sum_{n\geqq 0} c(n)\cdot\frac{z^n}{n!};\qquad \mathscr{V}(z)\equiv\sum_{n\geqq 0} v(n)\frac{z^n}{n!}.$$

Compte tenu de $c(n)=\sum_{k\geqq 0} c(n,k)$ et de $v(n)=\sum_{k\geqq 0} v(n,k)$, il vient:

$$\mathscr{C}(z)=\sum_{n,\,k\geqq 0} c(n,k)\frac{z^n}{n!}=C(1,z)=\Gamma(1,1,z)$$

$$\mathscr{V}(z)=\sum_{n,\,k\geqq 0} v(n,k)\frac{z^n}{n!}=V(1,z)=\exp(e^z-1)\cdot\Gamma(1,1,z)=\exp(e^z-1)\cdot\mathscr{C}(z).$$

Donc, par utilisation de (26):

PROPOSITION 7. *Les fonctions* $C(y,z)$, $V(y,z)$, $\mathscr{C}(z)$, $\mathscr{V}(z)$, *génératrices respectivement des nombres* $c(n,k)$, $v(n,k)$, $c(n)$, $v(n)$, *satisfont les identités formelles suivantes:*

$$C(y,z)=\sum_{n,\,k\geqq 0} c(n,k)\cdot y^k\frac{z^n}{n!}=\exp\left\{-y-\frac{y^2}{2}(e^z-1)\right\}\sum_{s\geqq 0}\frac{y^s}{s!}\cdot\exp\left\{\frac{s(s-1)}{2}z\right\}$$

$$V(y,z)=\sum_{n,\,k\geqq 0} v(n,k)\cdot y^k\frac{z^n}{n!}=\exp\left\{-y+\frac{y^2}{2}(e^z-1)\right\}\sum_{s\geqq 0}\frac{y^s}{s!}\cdot\exp\left\{\frac{s(s-1)}{2}z\right\}$$

$$\mathscr{C}(z)=\sum_{n\geqq 0} c(n)\cdot\frac{z^n}{n!}=\exp\left\{-1-\frac{1}{2}(e^z-1)\right\}\sum_{s\geqq 0}\frac{1}{s!}\exp\left\{\frac{s(s-1)}{2}z\right\}$$

$$\mathscr{V}(z)=\sum_{n\geqq 0} v(n)\cdot\frac{z^n}{n!}=\exp\left\{-1+\frac{1}{2}(e^z-1)\right\}\sum_{s\geqq 0}\frac{1}{s!}\exp\left\{\frac{s(s-1)}{2}z\right\}.$$

BIBLIOGRAPHIE

[1] ROTA, G. C.: The number of partitions of a set, *Amer. Math. Monthly* **71** (1964) 498—504.
[2] DE BRUIJN, N. G.: *Asymptotic Method in Analysis*, 2ème éd. North Holland, Amsterdam, 1961.
[3] RIORDAN, J.: *An introduction to combinatorial Analysis*, John Wiley, N. Y. 1958.
[4] DE BRUIJN, N. G.: *Pólya's theory of counting*, in "Applied Combinatorial Mathematics", John Wiley 1964.
[5] DE BRUIJN, N. G.: Generalization of Pólya's fundamental theorem, *Indag. Math.* **21** (1959) 59—69.
[6] COMTET, L.: Recouvrements, bases de filtre et topologies d'un ensemble fini, *C. R. Acad. Sci. Paris* **262** (1966) 1091—1094.

Faculté des Sciences d'Orsay, Département des Mathématiques, France

(Reçu le 10 février 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 153—158.

НЕСМЕЩЕННЫЕ ОЦЕНКИ ПАРАМЕТРА КОМПЛЕКСНОГО СТАЦИОНАРНОГО ГАУССОВСКОГО МАРКОВСКОГО ПРОЦЕССА. ПРИБЛИЖЕННЫЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

M. ARATÓ

Рассматривается комплексный стационарный гауссовский марковский процесс $\zeta(t)$ с математическим ожиданием $\mathsf{M}\zeta(t)=0$ и функцией ковариации

$$\mathsf{M}\zeta(t+s)\overline{\zeta(t)} = \frac{1}{\lambda} e^{-\lambda|s|-i\omega s}$$

(где ω известна). В данной заметке рассматриваются оценки параметра „затухания" λ с помощью статистик

$$s_2^1 = \frac{1}{2}[|\zeta(0)|^2+|\zeta(T)|^2 \quad \text{и} \quad s_2^2 = \frac{1}{T}\int_0^T |\xi(t)|^2 dt$$

отдельно. Как легко показать, ни s_2^1 ни s_2^2 не являются допустимыми[1] оценками $1/\lambda$ (см. статью [4], где это доказывается для одномерного процесса), тем не менее эти оценки представляют интерес. Кроме того рассматривается приближение характеристической функции оценки максимального правдоподобия и точность приближения с функцией нормального распределения.

Напомним совместную характеристическую функцию случайных величин s_1^2 и s_2^2 (см. [2] или [3]).

$$(1)\quad \varphi_{s_1^2, Ts_2^2}(\alpha_1,\alpha_2) = \frac{4\lambda(\lambda^2-2i\alpha_2)^{\frac{1}{2}} e^{T\lambda - T\sqrt{\lambda^2-2i\alpha_2}}}{(\lambda-i\alpha_1+\sqrt{\lambda^2-2i\alpha_2})^2-(\lambda-i\alpha_1-\sqrt{\lambda^2-2i\alpha_2})^2 e^{-2T\sqrt{\lambda^2-2i\alpha_2}}}.$$

В дальнейшем предполагается, что $T=1$.

1. Оценка s_1^2, для значений $\lambda \ll 1$

Из данных предположений легко вывести, что $Ms_1^2 = 1/\lambda$, т. е. s_1^2 можно использовать для оценки $1/\lambda$. С другой стороны, из (1) следует для характеристической функции λs_1^2

$$(1.1)\qquad f_{\lambda s_1^2}(\alpha) = \frac{1}{1-i\alpha-\alpha^2\dfrac{1-e^{-2\lambda}}{4}},$$

[1] Несмещенная оценка ξ параметра $f(\lambda)(=\mathsf{M}_\lambda\xi)$ называется допустимой на компакте Λ_0, если нет такой оценки нуля χ, $\mathsf{M}_\lambda\chi=0$ (при $\lambda\in\Lambda_0$), что $\mathsf{D}_\lambda^2(\xi+\chi)\leqq\mathsf{D}_\lambda^2(\xi)$ при всех $\lambda\in\Lambda_0$, причем для одного λ имеет место знак неравенства (в противном случае ξ называется недопустимой).

и отсюда получаем

$$\mathsf{P}_\lambda\left\{\frac{1}{s_1^2}<\lambda\cdot y\right\}=\mathsf{P}_\lambda\left\{\lambda s_1^2>\frac{1}{y}\right\}=\frac{e^\lambda-e^{-\lambda}}{2(1-e^{-\lambda})}e^{-2\frac{1-e^{-\lambda}}{1-e^{-2\lambda}}\cdot\frac{1}{y}}-\frac{e^\lambda-e^{-\lambda}}{2(1+e^{-\lambda})}e^{-2\frac{1+e^{-\lambda}}{1+e^{-2\lambda}}\cdot\frac{1}{y}}. \tag{1.2}$$

Так как статистика s_1^2 легко вычисляется (и при $\lambda\to 0$ получается χ^2 распределение с двумя степенями свободы), представляется интересным показать, при каких λ можно вместо оценки максимального правдоподобия $\hat{\lambda}$, взять s_1^2. В следующей таблице 1. даются значения y при разных p ($=P\{$„оценка" $<\lambda\cdot y\}$), и λ, для оценки максимального правдоподобия и оценки $1/s_1^2$. Легко убедиться, что при $\lambda<0\cdot 1$ оценки $\hat{\lambda}$ и $1/s_1^2$ являются приближенно эквивалентными.

Таблица 1

	λ \ p	0,1	0,05	0,025	0,01	0,001	0,9	0,95	0,975	0,99	0,999
$\hat{\lambda}$	0	0,951	19,52	39,60	99,9	1000	0,4352	0,3351	0,2620	0,2165	0,1460
	0,1	6,79	10,92	16,20	25,0	53,3	0,443	0,343	0,271	0,225	0,154
s_1^2		6,82	11,15	17,30	29,5	101,4	0,446	0,345	0,281	0,225	0,151
$\hat{\lambda}$	0,5	4,08	5,59	7,24	9,58	16,36	0,477	0,378	0,308	0,257	0,185
s_1^2		4,52	6,86	10,17	16,72	55,2	0,482	0,380	0,314	0,254	0,173

2. Оценка s_2^2, для значения $\lambda\gg 1$

При наших предположениях $Ms_2^2=1/\lambda$, т. е. s_2^2 является несмещенной оценкой $1/\lambda$. Статистику s_2^2 для больших значений λ ($\lambda\gg 1$) легче использовать, чем оценку максимального правдоподобия. Характеристическая функция имеет вид (см. (1)):

$$f_{s_2^2}(\alpha)=\frac{4\lambda(\lambda^2-2i\alpha)^{\frac{1}{4}}e^{\lambda-\sqrt{\lambda^2-2i\alpha}}}{(\lambda+\sqrt{\lambda^2-2i\alpha})^2-(\lambda-\sqrt{\lambda^2-2i\alpha})^2e^{-2\sqrt{\lambda^2-2i\alpha}}}. \tag{2.1}$$

Для определения значений y, при которых имеет место соотношение $\mathsf{P}\{\lambda^2 s_2^2<\lambda\cdot y\}=p$ (где число p данное). Мы использовали вычислительную машину УРАЛ-2. Определив значение интеграла

$$\frac{2e^y}{\pi}\int_{-\infty}^{\infty}\frac{\sqrt{r}\,e^{\lambda-\lambda\sqrt{r}\cos\varphi/2}\{\alpha_1[\sigma\cos\gamma+s\sin\gamma]+\alpha_2[\sigma\sin\gamma-s\cos\gamma]\}}{(\sigma^2+s^2)(\alpha_1^2+\alpha_2^2)}\,ds \tag{2.2}$$

при разных y, и последовательным приближением по y можно найти искомое

значение. Функцию распределения случайной величины s_2^2 не удалось найти в явном виде. Величины в интеграле (2. 2) имеют следующие значения:

$$\alpha_1 = A_1^2 - A_2^2 + \{[A_2^2 - B_1^2]\cos(2\lambda\sqrt{r}\sin\varphi/2) + 2B_1A_2\sin(2\lambda\sqrt{r}\sin\varphi/2)\}e^{-2\lambda\sqrt{r}\cos\varphi/2},$$

$$\alpha_2 = 2A_1A_2 + \{2B_1A_2\cos(2\lambda\sqrt{r}\sin\varphi/2) + (B_1^2 - A_2^2)\sin(2\lambda\sqrt{r}\sin\varphi/2)\}e^{-2\lambda\sqrt{r}\cos\varphi/2},$$

$$\gamma = \lambda ys + \varphi/2 - \lambda\sqrt{r}\sin\varphi/2, \quad A_1 = 1 + \sqrt{r}\cos\varphi/2, \quad A_2 = \sqrt{r}\sin\varphi/2,$$

$$B_1 = (1 - \sqrt{r}\cos\varphi/2),$$

$$\varphi = \operatorname{arc\,tg}\frac{2s}{1+2\sigma}, \qquad r^2 = 4s^2 + (1+2\sigma)^2, \qquad \sigma = 1/\lambda.$$

Определение одного интеграла при данном y требует 10—15 минут для значений $\lambda \sim 10$, и 5 минут для значений $\lambda \sim 100$ (при точности 10^{-4}). Вычисление интеграла (2. 2), если $\lambda < 10$, с данным методом происходит очень медленно. В нижеследующей таблице даются значения y при разных p и λ для $1/s_2^2$, максимального правдоподобия ($\hat{\lambda}$) и для нормального приближения (н. п.)

Таблица 2

	λ \ p	0,1	0,05	0,025	0,01	0,90	0,95	0,975	0,99
Н. п.		1,1281	1,1645	1,1960	1,2326	0,8719	0,8355	0,8040	0,7674
$\hat{\lambda}$	100	1,1413	1,1832	1,2205	1,2654	0,8847	0,8533	0,8269	0,7972
s_2^2		1,1305	1,1720	1,2096	1,253	0,8760	0,8449	0,8188	0,7895
Н. п.		1,403	1,516	1,620	1,734	0,597	0,484	0,380	0,266
$\hat{\lambda}$	10	1,530	1,701	1,867	2,073	0,714	0,641	0,588	0,527
s_2^2		1,414	1,558	1,713	2,03	0,648	0,562	0,534	0,49
		—	—	—	—	—	—	—	—
$\hat{\lambda}$	5	1,809	2,090	2,354	2,710	0,647	0,562	0,497	0,432
s_2^2						0,535	0,47		

Интересно заметить, что при больших значениях λ оценка $1/s_2^2$ является более „симметричной", чем оценка максимального правдоподобия. Даже при $\lambda \sim 10$ доверительные границы более „короткие" по оценке $1/s_2^2$, чем по оценке максимального правдоподобия.

3. Оценки максимального правдоподобия, приближения для значений $\lambda \gg 1$

а) Оценка максимального правдоподобия

$$\hat{\lambda} = \frac{-(s_1^2 - T) + \sqrt{(s_1^2 - T)^2 + 4Ts_2^2}}{2Ts_2^2}$$

зависит от двух статистик s_1^2 и s_2^2. Для определения функции распределения при данном λ, $\mathsf{P}_\lambda\{\hat{\lambda} < \lambda y\}$, нам достаточно вычислить функцию распределения случайной величины $\zeta_y = \lambda y s_1^2 + \lambda^2 y^2 s_2^2$. Преобразование Лапласа функции распределения ζ_y имеет вид

$$F^*(p) = \frac{4(1+2y^2p)^{\frac{1}{4}} e^{\lambda - \lambda\sqrt{1+2y^2p}}}{p\left\{\left[1+yp+\sqrt{1+2y^2p}\right]^2 - \left[1+y\,p-\sqrt{1+2y^2p}\right]^2 e^{-2\lambda\sqrt{1+2y^2p}}\right\}}. \tag{3.1}$$

Возникает вопрос, нельзя ли пренебречь вторым членом в знаменателе, если $\lambda \gg 1$? Этот вопрос интересен потому, что преобразование Лапласа функции

$$\tilde{F}(p) = \frac{4(1+2y^2)^{\frac{1}{4}} e^{\lambda - \lambda\sqrt{1+2y^2p}}}{p\left[1+yp+\sqrt{1+2y^2p}\right]^2} \tag{3.2}$$

дается в явном виде (хотя использовать его для вычислений пока невозможно).

С помощью функции (3. 2) для определения y при данном $p\,(=\mathsf{P}\{\hat{\lambda} > \lambda \cdot y\})$ мы вычисляем интегралы

$$\frac{2e^{\sigma(\lambda y+1)}}{\pi} \int_{-\infty}^{\infty} \frac{\sqrt{r}\, e^{\lambda(1-\sqrt{r}\cos\varphi/2)}\{[\sigma\cos\gamma + s\sin\gamma](A_1^2 - A_2^2) + 2A_1A_2[\sigma\sin\gamma - s\cos\gamma]\}}{(\sigma^2+s^2)[A_1^2+A_2^2]^2}\, ds \tag{3.3}$$

с достаточной точностью (10^{-4}), где

$$A_1 = (1+y\sigma+\sqrt{r}\cos\varphi/2), \quad A_2 = (ys+\sqrt{r}\sin\varphi/2), \quad r^2 = (1+2y^2\sigma)^2 + (2y^2s)^2,$$

$$\varphi = \operatorname{arc\,tg}\frac{2y^2 s}{1+2y^2\sigma}, \quad \gamma = \{(\lambda y+1)s + \varphi/2 - \lambda\sqrt{r}\sin\varphi/2\}, \quad \sigma = 1/\lambda$$

Любопытно заметить, что для $\lambda < 10$ в вычислении интеграла (3. 3) возникают трудности, в то же время в точных вычислениях такого явления не было (см. [3]).

Из таблицы 3 видно, что приближение (3. 2) является вполне удовлетворительным даже при $\lambda \sim 5$ (особенно при значениях $y > 1$).

б) Еще в статье [1] заметили, что для больших λ оценка $\hat{\lambda}$ имеет приближенно нормальное распределение.

$$\mathsf{P}\{\hat{\lambda} < y \cdot \lambda\} = \mathsf{P}\{\hat{\lambda} < \lambda + z\sqrt{\lambda}\} \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt, \tag{3.4}$$

Таблица 3

	λ \ p	0,1	0,05	0,025	0,01	0,001	0,9	0,95	0,975	0,99	0,999
$\hat{\lambda}$	100	1,1413	1,1832	1,2205	1,2654	1,3641	0,8847	0,8533	0,8269	0,7972	0,732
приб.		1,1413	1,1834	1,2211	1,2654	1,362	0,8853	0,8535	0,8273	0,797	0,732
$\hat{\lambda}$	10	1,527	1,701	1,867	2,073	2,575	0,714	0,641	0,588	0,527	0,422
приб.		1,527	1,700	1,867	2,073	2,579	0,714	0,641	0,590	0,527	0,422
$\hat{\lambda}$	5	1,809	2,090	2,354	2,710	3,583	0,647	0,562	0,497	0,432	0,319
приб.		1,81	2,1	2,4	2,	3,	0,648	0,561	0,49		
$\hat{\lambda}$	3	2,107	2,510	2,911	3,443	4,752	0,600	0,506	0,439	0,373	0,268
приб.		2,17					0,59	0,504			

и отсюда для y_p (при данном p) получается следующее соотношение

$$y_p = 1 + \frac{z_p}{\sqrt{\lambda}}$$

(в таблице 2 значения для нормального приближения по этой формуле даются). Так как нормальное приближение даже при $\lambda \sim 100$ не действует с достаточной точностью, можно предполагать, что y_p имеет вид

$$y_p = 1 + \frac{z_p}{\sqrt{\lambda}} + \frac{c_p}{\lambda}, \tag{3. 5}$$

где c_p вычисляется по данным таблицы статьи [3].

В следующей таблице мы даем значения z_p и c_p при разных p.

Приближение (3. 5) для $\lambda > 50$ дает достаточную точность (три верных знака для y_p) и намного облегчает вычислительную работу.

Таблица 4

p	0,1	0,05	0,025	0,01	0,001	0,9	0,95	0,965	0,99	0,999
z_p	−1,2815	−1,6449	−1,9600	−2,3264	−3,0900	1,2815	1,6449	1,9600	2,3264	3,0900
c_p	1,30	1,78	2,29	2,98	4,13	1,31	1,87	2,46	3,29	5,51

БИБЛИОГРАФИЯ

[1] Арато, М., Колмогоров А. Н. и Синай, Я. Г. Об оценке параметров комплексного стационарного гауссовского марковского процесса, *Докл. Акад. Наук СССР* 146 (1962) 747—750.

[2] Arató M.: Folytonos állapotú Markov-folyamatok statisztikai vizsgálatáról, *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **14** (1964) 317—330.

[3] Арато, М.: Вычисление доверительных границ для параметра „затухания" комплексного стационарного гауссовского марковского процесса, (в печати в журнале *Теор. Вероятност. и Применен*).

[4] Арато, М.: О подобных критериях и допустимых оценках стационарного гауссовского марковского процесса, *Studia Sci. Math. Hungar.* **3** (1968)

Вычислительный Центр Академии Наук Венгрии, Будапешт

(Поступила 21-ого марта 1967 г.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 159—166.

О ПОДОБНЫХ КРИТЕРИЯХ И ДОПУСТИМЫХ ОЦЕНКАХ СТАЦИОНАРНОГО ГАУССОВСКОГО МАРКОВСКОГО ПРОЦЕССА

M. ARATÓ

1. Во многих важных задачах математической статистики, подлежащих проверке гипотезы, требуются критерии $\varphi(X)$, для которых

$$\mathsf{M}_\theta \varphi(X) = \alpha \tag{1}$$

при всех распределениях X (где X обозначает „наблюдение"), принадлежащих заданному семейству $\mathscr{P} = \{\mathsf{P}_\theta^X, \theta \in \Omega\}$. По этому поводу см. книги [5] и [6]. Критерии удовлетворяющие условию (1) называются *подобными* по отношению к семейству распределений P_θ^X, $(\theta \in \Omega)$, или, короче, по отношению к параметрическому множеству Ω[1].

Если T достаточная статистика для параметра θ (или для семейства $\mathscr{P}$) и $\mathscr{P}^T$ означает семейство $\{\mathsf{P}_\theta^T, \theta \in \Omega\}$ распределений T, тогда любой критерий φ, для которого

$$\mathsf{M}[\varphi(X)|T=t] = \alpha, \tag{2}$$

с точностью до P^T-меры нуль, оказывается подбным, так как

$$\mathsf{M}_\theta \varphi(X) = \mathsf{M}_\theta\big(M[\varphi(X)|T=t]\big) = \alpha, \quad \text{при} \quad \theta \in \Omega. \tag{3}$$

О критериях, удовлетворяющих (2), говорят, что они имеют неймановскую структуру относительно T.

Ещё напомним, что семейство вероятностных мер $\mathscr{P} = \{\mathsf{P}_\theta, \theta \in \Omega\}$ называется *полным* (см. LEHMANN [5] стр. 182, или Линник [6] стр. 81), если из соотношения

$$\mathsf{M}_\theta f(x) = 0, \quad \text{при} \quad \theta \in \Omega \tag{4}$$

следует, что измеримая функция

$$f(x) = 0 \quad \text{почти всюду } \mathsf{P}_\theta, \ \theta \in \Omega.$$

Семейство называется *ограниченно полным*, если любая измеримая ограниченная функция $f(x)$ при условии (4) почти всюду равна 0. Нам нужно следующее утверждение, принадлежащее Леману и Шеффе:

Теорема 1. *Пусть дано семейство распределений $\mathscr{P} = \{P_\theta, \theta \in \Omega\}$ и T достаточная статистика для $\mathscr{P}$. Тогда для того, чтобы все подобные критерии*

[1] Здесь и в дальнейшем рассматриваются рандомизованные критерии $\varphi(X)$, т. е. если X принимает значение x, то производится случайный эксперимент с двумя возможными исходами: H и $\bar{H}$, вероятности которых равны $\varphi(x)$ и $1-\varphi(x)$. В случае H гипотеза отвергается, а в противном случае принимается. $\varphi(X)$ называется *критической функцией*, $0 \leqq \varphi(X) \leqq 1$ при всех x.

имели найма новскую структуру относительно T, необходимо и достаточно, чтобы семейство $\mathscr{P}^T$, распределений T, было ограниченно полным.

Доказательство можно найти в [5] стр. 185 или в [6] стр. 82. В этой заметке рассматривается стационарный гауссовский марковский процесс с непрерывным и дискретным временем. В непрерывном случае, при предположении $\mathsf{M}\xi(t)=0$ и $\mathsf{M}\xi(t)\xi(s)=\frac{1}{2\lambda}\exp(-\lambda|t-s|)$, мы имеем для процесса $\xi(t)$, $0\leqq t\leqq T$,

$$(5)\qquad \frac{dP}{dV}=\sqrt{\frac{\lambda}{\pi}}\exp\left\{-\lambda\left[\frac{\xi^2(0)+\xi^2(T)}{2}-\frac{T}{2}+\frac{\lambda}{2}\int_0^T\xi^2(t)\,dt\right]\right\},\qquad \lambda>0,$$

(см. [1] и [8]), где $V=L\times W$, L-обычная лебеговская мера на прямой, а W — известная условная мера Винера. „Наблюдение" X в этом случае является реализацией процесса $\xi(t)$ в промежутке времени $0\leqq t\leqq T$.

В случае с дискретным временем, при предположении $\mathsf{M}\xi(k)=0$, $\mathsf{M}\xi^2(k)=1$ плотность вероятностных величин $\zeta(1),\zeta(2),\ldots,\zeta(n)$ имеет вид

$$(6)\qquad (2\pi)^{-\frac{n}{2}}(1-\varrho^2)^{-\frac{n-1}{2}}\exp-\frac{1}{2(1-\varrho^2)}\left\{x_1^2+x_n^2+(1+\varrho^2)\sum_2^{n-1}x_i^2-2\varrho\sum_1^{n-1}x_ix_{i+1}\right\},$$

где $\varrho=\mathsf{M}\xi(k)\xi(k+1)$, $(-1<\varrho<1)$.

Семейства $\{\mathsf{P}_\varrho,\ -1<\varrho<1\}$ и $\{\mathsf{P}_\lambda,\ \lambda>0\}$ являются *экспоненциальными*. Из теоремы факторизации следует, что в случае непрерывного времени существует 2-мерная достаточная статистика

$$(7)\qquad \chi_1=\frac{\xi^2(0)+\xi^2(T)}{2},\qquad \chi_2=\frac{1}{T}\int_0^T\xi^2(t)\,dt,$$

а в случае дискретного времени 3-мерная достаточная статистика

$$(8)\qquad T_1=\frac{\xi^2(1)+\xi^2(n)}{2},\qquad T_2=2\sum_1^{n-1}\xi(i)\xi(i+1),\qquad T_3=\sum_2^{n-1}\xi^2(i).$$

Известно (см. [5] стр. 78), что распределения величин $\chi=(\chi_1,\chi_2)$ соответсвенно $T=(T_2,T_2,T_3)$ принадлежат экспоненциальному семейству и имеют вид (плотности вероятности):

$$(9)\qquad p_\lambda(t_1,t_2)=\sqrt{\frac{\lambda}{\pi}}\exp-\{\lambda t_1+\lambda^2t_2\}\cdot h(t_1,t_2)$$

при $t_1\cdot t_2>0$, соотвественно

$$(10)\ p_\varrho(t_1,t_2,t_3)=(2\pi)^{-\frac{n}{2}}(1-\varrho^2)^{-\frac{n-1}{2}}\exp-\frac{1}{1-\varrho^2}\{t_1-\varrho t_2+(1+\varrho^2)t_3\}h(t_1,t_2,t_3)$$

(где $t_1>0$, $t_3>0$).

Функция $h(t_1, t_2)$ в явном виде неизвестна, но знаем характеристическую функцию плотности вероятности $p_\lambda(t_1, t_2)$, которая имеет вид

$$f_\lambda(\alpha_1, \alpha_2) = \frac{2\sqrt{\lambda T}(\lambda^2 T^2 - 2i\alpha_2 T)^{1/4} e^{\frac{\lambda T}{2}}}{[(\lambda T - i\alpha_1 T + \sqrt{\lambda^2 T^2 - 2i\alpha_2 T})^2 e^{\sqrt{\lambda^2 T^2 - 2i\alpha_2 T}} - (\lambda T - i\alpha_1 T - \sqrt{\lambda^2 T^2 - 2i\alpha_2 T})^2 e^{-\sqrt{\lambda^2 T^2 - 2i\alpha_2 T}}]^{1/2}} \tag{11}$$

Из этого выражения видно, что $h(t_1, t_2) > 0$ при $t_1, t_2 > 0$.

А. Н. Колмогоров ещё в 1960 г. поставил задачу найти в этих случаях подобные множества (подобные критерии). Я в своей диссертации [2] в 1962 г. показал, что в данном случае можно использовать интересный метод R. A. Wijsman-а. Статья Вийсмана вскрывала интересные связи между статистикой и теорией уравнений в частных производных. Этот метод позже получил широкое признание, о котором существует красивая книга Ю. В. Линника [6]. В моей диссертации я показал, что семейства (9) и (10) не являются ограниченно полными. Из этого факта и из теоремы 1. следует, что все подобные критерии, не имеющие неймановскую структуру, представляются в виде

$$g(T) + \alpha, \tag{12}$$

где $\mathsf{M}_\theta g(T) \equiv 0$ $(\theta \in \Omega)$. (Здесь θ используется вместо ϱ и λ). На самом деле, если $\Phi(X)$ является подобным критерием, то функция $f(X) = \mathsf{M}(\Phi | T) - \alpha$ такая, что $\mathsf{M}_\theta g(T) = \mathsf{M}_\theta f(x) = 0$ при всех $\theta \in \Omega$, где по определению $g(T(x)) = f(x)$. Отсюда видно, что все подобные критерии имеют вид (12). Ниже мы докажем утверждение, что семейства (9) и (10) не являются ограниченно полными, но докажем ещё и теорему о представлении подобных критериев.

2. Имеет место следующее утверждение

Теорема 2. 1. *Экспоненциальное семейство* (9) *не является ограниченно полным.*

Доказательство. Для доказательства строим функцию $F(t_1, t_2) = P(t)$, которая ограничена, исчезает вне куба R $(R \subset T_1 \times T_2,\ t_1, t_2 > 0)$, и для которой преобразование Лапласа $\mathscr{L}(F)$ исчезает при $\lambda > 0$

$$\int_0^\infty \int_0^\infty F(t_1, t_2) e^{-(\lambda t_1 + \lambda^2 t_2)}\, dt_1\, dt_2 \equiv 0 \qquad (\text{при } \lambda > 0). \tag{2.1}$$

Возьмем для этой цели дважды непрерывно дифференцируемую функцию $G(t_1, t_2)$ внутри R, исчезающую вне R, у которой все частные производные непрерывны на границе R. Введем дифференциальный оператор

$$D = \frac{\partial^2}{\partial t_1^2} + \frac{\partial}{\partial t_2}, \tag{2.2}$$

то в силу свойства $G(t_1, t_2)$

$$\mathscr{L}(DG) = (\lambda^2 - \lambda^2)\mathscr{L}(G) \equiv 0. \tag{2.3}$$

Таким образом функция $F(t_1, t_2) = DG(t_1, t_2)$ удовлетворяет условию (2. 1), а функцию

$$\frac{F(\chi_1, \chi_2)}{h(\chi_1, \chi_2)} = \frac{DG(\chi_1, \chi_2)}{h(\chi_1, \chi_2)} \tag{2. 4}$$

можно использовать для построения подобного критерия, не имеющего неймановскую структуру:

$$\Phi(\chi) = \alpha + \frac{F(\chi)}{h(\chi)}. \tag{2. 5}$$

В роль $G(t_1, t_2)$ можно взять, например, следующую функцию (или линейную комбинацию таких функций): пусть куб R задан условиями

$$R: (0 <) a_1 < t_1 < a_1 + l; \quad (0 <) a_2 < t_2 < a_2 + l$$

и возьмем $G(t) = \prod_1^2 (t_i - a_i)^2 (a_i + l - t_i)^2$ при $(t_1, t_2) \in R$ и $G(t) = 0$ при $t \in R$. Естественно возникает вопрос, можно ли таким образом, т. е. с помощью метода Вийсмана, получить все подобные критерии? Полное описание подобных критерий ясно связана задачей отысканием всех предкотестов (или котестов) в смысле Линника (см. [6] стр. 107, 137). Если $\varphi(X)$ рандомизованный подобный критерий (тест) уровня α $(0 < \alpha < 1)$, то *котестом* называется любая статистика вида $\psi = A(\varphi - \alpha)$, где A константа. *Предкотестом* уровня α будем называть любую статистику $\xi(X)$ с конечным математическим ожиданием и

$$M_\theta \xi(x) \equiv 0, \quad \text{при} \quad \theta \in \Omega.$$

Нахождение всех котестов ψ можно задать в форме

$$\left(\frac{\partial^2}{\partial t_1^2} + \frac{\partial}{\partial t_2} \right) G = h\psi, \tag{2. 6}$$

где ψ котест уровня α и G решение уравнения (2. 6). Уравнение является параболическим уравнением, похоже на уравнение теплопроводности в одномерном пространстве, где t_2 является временем, t_1 местом, G температурой, а $\psi \cdot h$ источником тепла. Если бы имели дело с обычной задачей теплопроводности, то решение можно было бы записать с помощью соответсвующей функции Грина. Так как дифференциальный оператор D является оператором с постоянными коэффициентами, можно использовать общие результаты теории дифференциальных уравнений (см. [7]).

Имеет место следующее утверждение.

Теорема 2. 2. *Любой котест ψ (где $\psi h \in L_2$), уровня α, представляется в виде*

$$\psi = \frac{1}{h} DG,$$

где G принадлежит L_2 вместе со своим градиентом (в смысле обобщенных функций) и G обращается в нуль вне некоторого компакта $T_1 \times T_2$ $(t_1, t_2 > 0)$.

Доказательство, в таком виде, сразу следует из теоремы 2. 1 и леммы 1. 7 Хермандера [10].

Замечание. Более глубокое утверждение о представлении всех котестов ψ (уже не принадлежащих L_2) также имеет место, но для этого надо использовать более тонкие рассуждения Паламадова [7], кто первым обратил внимание на результаты Хермандера.

3. В случае с дискретным временем рассматривается дифференциальный оператор

$$D^* = \frac{\partial^2}{\partial t_2^2} - \frac{\partial^2}{\partial t_1 \partial t_3} + \frac{\partial^2}{\partial t_1^2} \tag{3.1}$$

и легко доказывается утверждение:

Теорема 3. 1. *Экспоненциальное семейство* (10) *не является ограниченно полным.*

Точно так же, как в пункте 2, доказывается

Теорема 3. 2. *Любой котест* $\psi(\psi h \in L_2)$, *уровня* α, *представляется в виде*

$$\psi = \frac{1}{h} D^* G,$$

где G *принадлежит* L_2, *с компактным носителем, принадлежащим* $\operatorname{int}(t_1, t_3 > 0)$.

4. В этом пункте рассматриваем несмещенные оценки λ и ϱ экспоненциальных семейств (5) и (6). Хорошо известно (см. Колмогоров [4] и Линник [6] стр. 56—57), что для широкого круга вопросов возможно ограничиться лишь несмещенными оценками, зависящими от достаточных статистик.

В нашем случае семейство мер достаточных статистик $\chi = (\chi_1, \chi_2)$ и $T = (T_1, T_2, T_3)$ не является ограниченно полным, поэтому любая статистика $\xi(\chi)$ или $\xi(T)$, для которой существуют математическое ожидание

$$\mathsf{M}_\theta \xi = f(\theta)$$

и дисперсия

$$\mathsf{D}_\theta^2 \xi = \mathsf{M}_\theta (\xi - f(\theta))^2,$$

при $\theta \in \Omega$, является несмещенной оценкой функции $f(\theta)$, но не является единственной оценкой.

Как известно, для описания всех несмещенных оценок $f(\theta)$ достаточно найти одну и, кроме того, все несмещенные оценки нуля (Н. О. Н). Напомним, что несмещенная оценка ξ называется допустимой на компакте Ω_0, если нет такой Н. О. Н. η, что $\mathsf{D}_\theta^2(\xi + \eta) \leqq \mathsf{D}_\theta^2(\xi)$ при $\theta \in \Omega_0$, причем хотя бы для одного θ имеет место неравенство. Оценка ξ называется наилучшей на Ω_0, если $\mathsf{D}_\theta^2(\xi + \eta) \geqq \mathsf{D}_\theta^2(\xi)$ для любой Н. О. Н. η.

Докажем следующее утверждение.

Теорема 4. 1. *Для любого полинома* $Q(\chi_1, \chi_2)$ ($\neq$ const) *найдется такой компакт* Λ_0 *значений параметра* λ, *что на нем* Q *является недопустимой оценкой функции* $\mathsf{M}_\lambda Q(\chi_1, \chi_2) = f(\lambda)$.

Доказательство. Мы будем применять некоторые соображения типа метода Вийсмана, которые использовались Каганом для непольных экспоненциальных семейств, при некоторых дополнительных ограничений (см. [6] гл. VII. 3.). Пусть $\psi_0(\chi_1, \chi_2)$ функция, удовлетворяющая условиям

1. $\psi_0 > 0$ при $\chi \in R$

2. $\psi_0 = 0$ при $\chi \in R$

3. ψ_0 всюду непрерывна и имеет не менее $2k_1 + 3k_2$ частных производных по каждому аргументу, и все производные обращаются в нуль на границе R.

Здесь R означает в пространстве χ ограниченный замкнутый куб, в котором $h(\chi) \geqq \varepsilon_0 > 0$ (такой куб существует). $m \geqq 1$ степень полинома Q и $a_{k_1 k_2} \chi_1^{k_1} \chi_2^{k_2}$ один из его старших членов ($a_{k_1 k_2} \neq 0$, $k_1 + k_2 = m$). В дальнейшем вместо λ пишем λ_1, а вместо λ^2 пишем λ_2 ($\lambda_2 = \lambda_1^2$ условие рассматривается позже).

Пусть

$$\psi(t_1, t_2) = w\psi_0(t_1, t_2), \tag{4. 1}$$

где дифференциальный оператор

$$w = \left(\frac{\partial^2}{\partial t_1^2} + \frac{\partial}{\partial t_2}\right)^m \tag{4. 2}$$

применим к функции $\psi_0(t_1, t_2)$ в силу условия 3). Тогда преобразование Лапласа функции ψ равно

$$\iint \psi(t_1, t_2)\, e^{(\lambda_1 t_1 + \lambda_2 t_2)}\, dt_1\, dt_2 = \iint w\psi_0\, e^{(\lambda_1 t_1 + \lambda_2 t_2)}\, dt_1\, dt_2 = (\lambda_1^2 - \lambda_2), \tag{4. 3}$$

где

$$V(\lambda_1, \lambda_2) = \iint \psi_0(t_1, t_2)\, e^{(\lambda_1 t_1 + \lambda_2 t_2)}\, dt_1\, dt_2.$$

Если

$$\eta = \begin{cases} \dfrac{\psi(\chi)}{h(\chi)}, & \text{при} \quad \chi \in R \\ 0, & \text{при} \quad \chi \in R, \end{cases} \tag{4. 4}$$

то при условии $\lambda_1^2 = \lambda_2$ η является Н. О. Н.:

$$\mathsf{M}_{\lambda_1, \lambda_2}(\eta) = 0. \tag{4. 5}$$

Далее для оценки Q

$$\mathsf{M}_{\lambda_1, \lambda_2}(\eta \cdot Q) = C(\lambda) \iint \psi \cdot Q e^{(\lambda_1 t_1 + \lambda_2 t_2)}\, dt_1\, dt_2 = C(\lambda) D\left((\lambda_1^2 - \lambda_2)^m V(\lambda_1, \lambda_2)\right) \tag{4. 6}$$

где дифференциальный оператор

$$D = \sum a_{i_1 i_2} \frac{\partial^{i_1 + i_2}}{\partial \lambda_1^{i_1} \partial \lambda_2^{i_2}}.$$

Ясно, что при условии $\lambda_1^2 = \lambda_2$ из (4. 6)

$$\mathsf{M}_{\lambda_1}(\eta \cdot Q) = V(\lambda_1, \lambda_1^2)[P(\lambda_1)], \tag{4. 7}$$

где многочлен $P(\lambda_1)$ не равен тождественно нулю и можно найти такой компакт Λ_0, который не содержит ни одного нуля многочлена $P(\lambda)$. На этом компакте

$$\mathsf{M}_\lambda(\eta \cdot Q) \neq 0 \qquad \text{(можно считать, что} > 0). \tag{4. 8}$$

Далее ясно, что

$$\mathsf{M}_\lambda(\eta^2) \leqq K_0 < \infty \tag{4. 9}$$

и если рассматривать статистику

$$\tilde{Q} = Q - \gamma \cdot \eta \tag{4. 10}$$

то

$$\mathsf{M}_\lambda(\tilde{Q})^2 = \mathsf{M}_\lambda(Q)^2 - 2\gamma \mathsf{M}_\lambda(\eta \cdot Q) + \gamma^2 \mathsf{M}_\lambda(\eta^2) \tag{4. 11}$$

и константу γ можно выбрать так, чтобы

$$-2\gamma \mathsf{M}_\lambda(\eta Q) + \gamma^2 \mathsf{M}_\lambda(\eta^2) < 0, \tag{4. 12}$$

на $\lambda \in \Lambda_0$, откуда и следует недопустимость оценки $Q(\chi_1, \chi_2)$ на Λ_0. Тем самым доказана наша теорема.

Заметим еще, что при $m = 1$, т. е. когда Q имеет вид

$$Q(\chi_1, \chi_2) = a_1 \chi_1 + a_2 + \chi_2 \qquad (a_i \neq 0,\ a_i > 0,\ i = 1, 2),$$

то

$$\mathsf{M}_\lambda(\eta \cdot Q) = V(\lambda_1, \lambda_1^2)[a_2 - 2a_1 \lambda_1]$$

и

$$\mathsf{M}_\lambda(\eta \cdot Q) = \begin{cases} > 0 & \text{при} \quad \lambda > \dfrac{a_2}{2a_1} \\ < 0 & \text{при} \quad \lambda > \dfrac{a_2}{2a_1} \end{cases}$$

Следствие 1. Ни χ_1 ни χ_2 отдельно не является допустимой оценкой $1/\lambda$ на любом компакте Λ_0.

Следствие 2. Полиномы $Q_1(\chi_1)$ и $Q_2(\chi_2)$ не являются допустимыми оценками на любом компакте Λ_0 функций

$$f_1(\lambda) = \mathsf{M}_\lambda Q_1, \quad f_i(\lambda) = \mathsf{M}_\lambda Q_2.$$

Эти следствия получаются дословным повторением предыдущего доказательства.

В случае дискретного времени такое же положение имеет место.

БИБЛИОГРАФИЯ

[1] ARATÓ, M.: Оценка параметров стационарного гауссовского марковского процесса, *Докл. Акад. Наук. СССР* **145** (1962) № 1, 13—16.
[2] ARATÓ, M.: Некоторые статистические вопросы стационарных гауссовских марковских процессов, Диссертация, Москва М. Г. У. 1962.
[3] ARATÓ, M.: Вычисление доверительных границ для параметра „затухания" комплексного стационарного гауссовского процесса. (В печати в журнале Теор. Вероятност. и Применен.).
[4] Колмогоров, А. Н.: Несмещенные оценки, *Изв. Акад. Наук. СССР* Сер. Мат. № 4 (1950) 303—326.
[5] Леманн Э.: *Проверка статистических гипотез,* (Английское издание в 1959) New York
[6] Лиинык, Ю. В.: *Статистические задачи с мешающими параметрами,* Москва, 1966.
[7] Паламодов, В. П.: О проверке многомерной полиномиальной гипотезы, *Докл. Акаб. Наук. СССР* **172** (1966) № 2, 291—293.
[8] STRIEBEL, CH.: Densities for Stochastic Processes, *Ann. Math. Stat.* **30** (1959) 559—567
[9] WIJSMAN, R. A.: Incomplete suffcient statistics and similar tests, *Ann. Math. Stat.* **29** (1958), 1028—1045.
[10] HÖRMANDER, L.: On the theory of general partial differential operators. *Acta Math.* **94** (1955) 161—248.

Вычислительный Центр Академии Наук Венгрии, Будапешт

(Поступила 21-ого марта 1967 г.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 167—179.

СИСТЕМА ОБСЛУЖИВАНИЯ С ПЕРЕКЛЮЧЕНИЕМ

J. GERGELY

1. Описание системы

В работе [1] Гавера рассматривается следующая задача обслуживания: К одному прибору поступают потоки требований типа A и B. Длительности обслуживания требований прибором пусть будут T_1 и T_2. В том случае, когда прибор заканчивает обслуживание одного требования типа A и начинает обслуживать требование типа B, или обратно, для переключения прибора нужно случайное время T_{12} и T_{21}. T_1, T_2, T_{12} и T_{21} — независимые с функциями распределений $B_1(t)$, $B_2(t)$, $C_{12}(t)$ и $C_{21}(t)$. Входящие потоки — пуассоновские с параметрами a_1 и a_2, независимы между собой и от длительности обслуживания требований и от времени переключения прибора.

Гавер исследовал систему в следующих случаях:

1. Обслуживание требований производится по порядку их поступлений.
2. Требования типа B имеют приоритет и рассматриваются различные формы приоритета.

Процесс, происходящий в системе обслуживания, зависит и от примененной стратегии, если система свободна. В работе [1] можно найти следующие два случая:

а) Если система обслуживания становится пустой и последнее обслуженное требование было из типа B, немедленно начинается переключение прибора к обслуживанию требований типа A. ($a_1 > a_2$).

б) В том случае, когда система обслуживания свободна, переключение начинается только тогда, когда поступает требование и нужно переключить прибор.

Мы здесь исходим из следующего предположения:
требования типа A и B становятся в отдельную очередь. Если прибор обслуживает требование типа A, переключение происходит в том случае, когда в очереди типа A больше нет требований, а в очереди типа B ожидают, по крайней мере, n_2 требований, и обратно, если в очереди типа B требования нет, а в очереди типа A ожидают по крайней мере n_1 требований, где $n_1 \geqq 1$, $n_2 \geqq 1$.

В пункте II выводим соотношение для производящих функций вероятностей состояний системы обслуживания. Докажем, что система характеризуется эргодической цепью Маркова. В пункте III дадим прием на вычисление стационарных вероятностей. В конце статьи покажем пример и полученные результаты.

2. Формулировка соотношений

Обозначим через $t_1, t_2, \ldots$ моменты окончаний обслуживаний. Состояние системы в точке t_N, $N \geqq 1$, будет характеризоваться вектором $\xi_N = (i, k_1, k_2)$, где $i=1$, если в точке t_N закончилось обслуживание требования типа A, а $i=2$, если закончилось обслуживание требования типа B; после окончания обслуживания остается в системе k_1 требований типа A и k_2 — типа B.

Поскольку промежутки $\tau^{(N)} = t_{N+1} - t_N$ состоят из времен обслуживаний, переключений и ожиданий, на прибывание новых требований, предполагается, что функция распределения случайных величин $\tau^{(N)}$ — известна. Так как входящий поток — пуассоновский и в промежутке $\tau^{(N)}$ обслуживается одно требование, то имеется возможность определения вероятности изменений вектора ξ_N во время $\tau^{(N)}$. Это означает, что зная только вектор ξ_N, можно определить вероятность того, каким будет вектор ξ_{N+1}. Другими словами, состояния системы в точках t_N, $N \geqq 1$ образуют цепь Маркова. По структуре системы обслуживания видно, что из любого состояния можно перейти в любое другое состояние за конечное число шагов, т. е. цепь неприводимая. Легко убедиться в том, что цепь является апериодичной.

Пусть $p_{i,k_1,k_2,N}$ вероятность того, что после окончания N-ого обслуживания состояние системы (i, k_1, k_2). Введем следующие функции:

$$P_{i,N}(z_1, z_2) = \sum_{k_1, k_2 \geqq 0} p_{i,k_1,k_2,N} z_1^{k_1} z_2^{k_2}, \tag{1}$$

$$\beta_i(z) = \int_0^\infty e^{-zt}\, dB_i(t), \quad \gamma_{ij}(z) = \int_0^\infty e^{-zt}\, dC_{ij}(t), \tag{2}$$

где $i=1, 2$; $j=1, 2$; $i \neq j$ и

$$0 \leqq z_1 \leqq 1; \quad 0 \leqq z_2 \leqq 1; \quad 0 \leqq z. \tag{3}$$

Докажем, что связь между функциями (1) и (2) при (3) устанавливает следующее соотношение:

$$\begin{aligned} z_i P_{i,N+1}(z_1, z_2) = &\Bigg\{ P_{i,N}(z_1, z_2) - P_{i,N}(0, z_2) + \\ &+ \frac{a_i}{a_1 + a_2} z_i \sum_{k_j=0}^{n_j - 1} \left[p_{i,k_j} z_j^{k_j} \sum_{l=0}^{n_j - k_j - 1} \left(\frac{a_j}{a_1 + a_2} z_j \right)^l \right] + \left[P_{j,N}(z_1, 0) - \right. \\ &\left. - \sum_{k_i=0}^{n_i - 1} \left\{ p_{j,k_i} z_i^{k_i} \left[1 - \left(\frac{a_i}{a_1 + a_2} z_i \right)^{n_i - k_i} \right] \right\} \right] \gamma_{ji}(a_1 + a_2 - a_1 z_1 - a_2 z_2) \Bigg\} \cdot \\ &\cdot \beta_i(a_1 + a_2 - a_1 z_1 - a_2 z_2). \end{aligned} \tag{4}$$

где

$$p_{1,k_2} = p_{1,0,k_2,N} \quad \text{и} \quad p_{2,k_1} = p_{2,k_1,0,N}.$$

Доказательство. Введем следующие дополнительные события (см. [2]): пусть требования будут красными или синими. Обозначим через z_1 или z_2 вероятность того, что требования, прибывающие в очередь типа A или B,

красные, тогда $p_{i,k_1,k_2,N} z_1^{k_1} z_2^{k_2}$ вероятность того, что после окончания обслуживания N-ого требования состояние системы (i, k_1, k_2) и все оставшиеся требования в системе красные. А $P_{i,N}(z_1, z_2)$ — вероятность того, что после обслуживания N-ого требования в системе синих требований не остается, и последнее обслуженное требование было типа A, если $i=1$, а типа B, если $i=2$. Вычислим вероятность того, что за время обслуживания одного требования (T_1 или T_2) и переключения прибора (T_{12} или T_{21}) синие требования не прибывают. Эти вероятности для времени обслуживания (при $i=1, 2$):

$$\sum_{k_1, k_2 \geqq 0} z_1^{k_1} z_2^{k_2} \int_0^\infty \frac{(a_1 t)^{k_1}}{k_1!} e^{-a_1 t} \frac{(a_2 t)^{k_2}}{k_2!} e^{-a_2 t} \, dB_i(t) = \beta_i(a_1 + a_2 - a_1 z_1 - a_2 z_2),$$

и подобным образом для времени переключения (при $i=1, 2$; $j=1, 2$; $i \neq j$) получим

$$\gamma_{ij}(a_1 + a_2 - a_1 z_1 - a_2 z_2).$$

Пусть будет $i=1, j=2$.

Левая сторона уравнения (4) выражает вероятность того события, что $N+1$-ое обслуженное требование было типа A и красное, после заканчивания обслуживания этого требования в системе не остаются синие требования.

С другой стороны это событие создается как дизьюнктное соединение следующих двух событий:

а) N-ое обслуженное требование было типа A, после заканчивания обслуживания этого требования в системе не остаются синие требования. Если $k_1 \geqq 1$, немедленно начинается обслижувание следующего требования типа A, если $k_1 = 0$ и $k_2 < n_2$, прибор ожидает на поступление следующего красного требования типа A. За это время поступают не более $n_2 - k_2 - 1$ красное требование типа B. За время $N+1$-ого обслуживания синие требования не прибывают.

б) N-ое обслуженное требование было типа B, после заканчивания обслуживания этого требования, в системе не остаются синие требования и никаких требований типа B. Если $k_1 \geqq n_1$ немедленно, а в случае $k_1 < n_1$ после поступления $n_1 - k_1$ красных требований типа A начинается переключение прибора. За время переключения и $N+1$-ого обслуживания синие требования не прибывают.

В правой стороне уравнения (4) стоит сумма вероятностей событий *а)* и *б)*.

Так же при $i=2$, $j=1$.

Обозначим среднее время обслуживания и переключения прибора при $i=1, 2$; $j=1, 2$; $i \neq j$ через

$$\beta_{i1} = \int_0^\infty t \, dB_i(t) \quad \text{и} \quad \gamma_1^{ij} = \int_0^\infty t \, dC_{ij}(t). \tag{5}$$

Раньше уже показывали, что состояния системы $\xi_N = (i, k_1, k_2)$, $N \geqq 1$, в моментах окончаний обслуживаний создают неприводимую, непериодическую цепь Маркова. Докажем, что эта цепь Маркова — эргодическая в том

смысле, что существует стационарное распределение вероятностей $p_{i,k_1,k_2,N}$ то есть имеет место следующая теорема.

Теорема. *Если*

$$a_1\beta_{11}+a_2\beta_{21}<1, \tag{6}$$

тогда существуют пределы (*при* $i=1,2;\ k_1\geqq 0;\ k_2\geqq 0$)

$$p_{i,k_1,k_2}=\lim_{N\to\infty} p_{i,k_1,k_2,N}. \tag{7}$$

Доказательство. Упорядочим состояние системы каким-нибуть образом; а так, что в начале выступают все такие состояния, для которых $k_1\leqq n_1$ и $k_2\leqq n_2$, затем следуют другие состояния. Обозначим порядковый номер состояния (i,k_1,k_2) через $s=s(i,k_1,k_2)$ и пусть

$$s_0=\max_{k_1\leqq n_1,k_2\leqq n_2} s(i,k_1,k_2).$$

Обозначим через q_{st} — переходную вероятность за один шаг из состояния с порядковым номером s в состояние с порядковым номером t. Пусть y_s означает среднее время, необходимое для заканчивания обслуживания требований, которые находятся в системе при $s=s(i,k_1,k_2)$, предполагая, что новые требования не прибывают.

Покажем, что введенные значения удовлетворяют условия следующей леммы.

Лемма. *Для того, чтобы неприводимая, непериодическая цепь Маркова имела стационарное распределение, достаточно существования* $\varepsilon>0$, *натурального числа* s_0 *и набора неотрицательных чисел* $y_0, y_1, y_2, \ldots$, *таких, что:*

$$\sum_{t\geqq 0} q_{st}y_t\leqq y_s-\varepsilon \quad \textit{для} \quad s>s_0, \tag{8}$$

$$\sum_{t\geqq 0} q_{st}y_t<+\infty \quad \textit{для} \quad s\leqq s_0. \tag{9}$$

Доказательство леммы можно найти в книге Климова [2] (стр. 229).

В нашем случае y_s определяется следующим образом (при $i=1,2;\ j=1,2;\ i\neq j$):

$$y_s=\begin{cases} k_i\beta_{i1}+\gamma_1^{ij}+k_j\beta_{j1}, & \text{если} \quad k_i>0,\ k_j\geqq n_j,\\ k_i\beta_{i1}, & \text{если} \quad k_i>0,\ k_j<n_j,\\ \gamma_1^{ij}+k_j\beta_{j1}, & \text{если} \quad k_i=0,\ k_j\geqq n_j. \end{cases}$$

В лемме выражение $\sum_{t\geqq 0} q_{st}y_t$ означает среднее время, необходимое от окончания одного обслуживания до полного обслуживания системы при условии, что в начале система находилась в состоянии, у которого $s=s(i,k_1,k_2)$, предполагая, что новые требования не прибывают. Если $t\leqq s_0$, то-есть $k_1\leqq n_1, k_2\leqq n_2$, удовлетворяются (9), конечностью слагаемых. В противном случае, используя определение y_s:

а) При $k_i>0$, $k_j\geqq n_j$ получим

$$\sum_t q_{st}y_t = (k_i-1)\beta_{i1}+a_i\beta_{i1}\beta_{i1}+\gamma_1^{ij}+k_j\beta_{j1}+a_j\beta_{i1}\beta_{j1} = \tag{10}$$
$$= y_s-\beta_{i1}(1-a_i\beta_{i1}-a_j\beta_{j1}) = y_s-\varepsilon_1,$$

б) При $k_i>0$, $k_j<n_j$

$$\sum_t q_{st}y_t = (k_i-1)\beta_{i1}+a_i\beta_{i1}\beta_{i1} = y_s-\beta_{i1}(1-a_i\beta_{i1}) = y_s-\varepsilon_2, \tag{11}$$

в) При $k_i=0$, $k_j\geqq n_j$

$$\sum_t q_{st}y_t = (k_j-1)\beta_{j1}+(\gamma_1^{ij}+\beta_{j1})a_j\beta_{j1} = y_s-(\beta_{j1}+\gamma_1^{ij})(1-a_j\beta_{j1}) = y_s-\varepsilon_3. \tag{12}$$

(Если $k_i=0$ и $k_j<n_j$, то $t\leqq s_0$.) Из условия (6) следует, что в формулах (10), (11) и (12)

$$\varepsilon_1 = \beta_{i1}(1-a_i\beta_{i1}-a_j\beta_{j1})>0$$
$$\varepsilon_2 = \beta_{i1}(1-a_i\beta_{i1})>0$$
$$\varepsilon_3 = (\beta_{j1}+\gamma_1^{ij})1-a_j\beta_{j1})>0,$$

поэтому при $\varepsilon=\min(\varepsilon_1,\varepsilon_2,\varepsilon_3)$ сбывается (8). По лемме существует стационарное распределение, то-есть справедливо (7). Таким образом наше утверждение доказано.

Продолжим аналитически определение функции $P_{i,N}(z_1,z_2)$ на комплексную область

$$|z_1|\leqq 1, \quad |z_2|\leqq 1. \tag{13}$$

Функция $P_{i.N}(z_1,z_2)$ при (13) является производящей функцией. Обозначим еще функции (2) в комплексной области $\mathrm{Re}\, z\geqq 0$ где они являются преобразованиями Лапласа—Стилтьеса функцией $B_i(t)$ и $C_{ij}(t)$ (при $i=1,2$; $j=1,2$; $i\neq j$). Выполняя эти продолжения и используя принцип аналитического продолжения, получаем, что соотношение (4) справедливое для области (13).

Из известных теорем, которые относятся производящим функциям, по (7) следует, что в области (13), при $N\to\infty$ существуют пределы

$$P_{i,N}(z_1,z_2)\to P_i(z_1,z_2)=\sum_{k_1,k_2\geqq 0} p_{i,k_1,k_2}z_1^{k_1}z_2^{k_2}. \tag{14}$$

Уже показали, что уравнение (4) справедливо при любом N, в области (13). Совершая предел $N\to\infty$ тоже будет справедливо, то-есть

$$z_iP_i(z_1,z_2) = \left\{P_i(z_1,z_2)-P_i(0,z_2)+\frac{a_i}{a_1+a_2}z_i\sum_{k_j=0}^{n_j-1}\cdot\left[p_{i,k_j}z_j^{k_j}\sum_{l=0}^{n_j-k_j-1}\left(\frac{a_j}{a_1+a_2}z_j\right)^l\right]+\right.$$

$$+\left[P_j(z_1,0)-\sum_{k_i=0}^{n_i-1}\cdot\left\{p_{j,k_i}z_i^{k_i}\left[1-\left(\frac{a_i}{a_1+a_2}z_i\right)^{n_i-k_i}\right]\right\}\right]\cdot \tag{15}$$

$$\left.\cdot\gamma_{ji}(a_1+a_2-a_1z_1-a_2z_2)\right\}\beta_j(a_1+a_2-a_1z_1-a_2z_2)$$

где

$$p_{1,k_2}=p_{1,0,k_2} \quad \text{и} \quad p_{2,k_1}=p_{2,k_1,0}.$$

3. Определение стационарных вероятностей

Пусть будет в уравнении (15) $z_1=z_2=1$, $i=1$ *и* $j=2$ получим:

$$P_1(1,1)=\left\{P_1(1,1)-P_1(0,1)+\frac{a_1}{a_1+a_2}\sum_{m=0}^{n_2-1}\left[p_{1,0,m}\sum_{l=0}^{n_2-m-1}\left(\frac{a_2}{a_1+a_2}\right)^l\right]+\right.$$

(16)

$$\left.+\left[P_2(1,0)-\sum_{m=0}^{n_1-1}p_{2,m,0}\left(1-\left[\frac{a_1}{a_1+a_2}\right]^{n_1-m}\right)\right]\gamma_{21}(0)\right\}\beta_1(0).$$

Так как $\gamma_{21}(0)=\beta_1(0)=1$, из (16) получаем, что

$$P_1(0,1)-P_2(1,0)=\sum_{m=0}^{n_2-1}p_{1,0,m}\left[1-\left(\frac{a_2}{a_1+a_2}\right]^{n_2-m}\right]-$$

(17)

$$-\sum_{m=0}^{n_1-1}p_{2,m,0}\left[1-\left(\frac{a_1}{a_1+a_2}\right)^{n_1-m}\right].$$

(То же самое можно получить, исходя из уравнения (15) при $z_1=z_2=1$, $i=2$ и $j=1$.)

Дифференцируя уравнение (15) по z, в точке $z_1=z_2=1$ при $i=1$, $j=2$ и $i=2, j=1$, получим два уравнения, с помощью которых, (используя уравнение (17) и обозначение (5)), получим:

$$P_1(0,1)\gamma_1^{12}+P_2(1,0)\gamma_1^{21}=\frac{1-a_1\beta_{11}-a_2\beta_{21}}{a_1+a_2}-$$

$$-\frac{1-a_1\gamma_1^{12}}{a_1}\sum_{m=0}^{n_2-1}p_{1,0,m}\left[1-\left(\frac{a_2}{a_1+a_2}\right)^{n_2-m}\right]- \tag{18}$$

$$-\frac{1-a_2\gamma_1^{21}}{a_2}\sum_{m=0}^{n_1-1}p_{2,m,0}\left[1-\left(\frac{a_1}{a_1+a_1}\right)^{n_1-m}\right].$$

Уравнения (17) и (18) являются системой линейных уравнений для $P_1(0, 1)$ и $P_2(1, 0)$. Определитель системы $\gamma_1^{21}+\gamma_1^{12}\neq 0$, поэтому уравнения (17) и (18) определяют $P_1(0, 1)$ и $P_2(1, 0)$ как линейныйе функции вероятностей $p_{1,0,m}$ при $m=0, 1, \ldots, n_2$ и $p_{2,m,0}$ при $m=0, 1, \ldots, n_1$.

Из уравнения (15) выражая функцию $P_i(z_1z_2)$ в знаменателе правой части будет функция

$$1-\frac{z_i}{\beta_i(a_1+a_2-a_1z_1-a_2z_2)}, \qquad i=1,2. \tag{19}$$

Пусть функция (19) при $z_i=z$ и $z_j=f_i(z)$ $(i=1, j=2$ и $i=2, j=1)$, обращается в 0, то-есть

$$\beta_i(a_1+a_2-a_iz-a_jf_i(z))\equiv 0, \qquad (i=1,\ j=2 \quad \text{и} \quad i=2,\ j=1),$$

или

$$\int_0^\infty e^{-[a_1+a_2-a_iz-a_jf_i(z)]t}\,dB_i(t)\equiv z, \qquad (i=1,\ j=2 \quad \text{и} \quad i=2,\ j=1). \tag{20}$$

Дифференцируя функцию (20) по z при $z=1$ получим

$$f_i'(1)=\frac{1-a_i\beta_{i1}}{a_j\beta_{j1}}>0, \qquad (i=1,\ j=2 \quad \text{и} \quad i=2,\ j=1). \tag{21}$$

Из (21) следует, что в окрестности точки $z=1$ (при $|z|\leqq 1$) $|f_i(z)|\leqq 1$ $(i=1, 2)$. Так как функция $P_i(z_1, z_2)$ аналитична в области (13), то выражая её из уравнения (15), необходимо, чтобы при $z_i=z$, $z_j=f_i(z)$ числитель был равен 0, то-есть

$$\begin{aligned} &P_i[f_i(z)]-P_j[z]\gamma_{ji}(a_1+a_2-a_iz-a_jf(z))- \\ &-\frac{a_i}{a_1+a_2}z\sum_{m=0}^{n_j-1}p_{i,m}f_i(z)^m\sum_{l=0}^{n_j-m-1}\left[\frac{a_j}{a_1+a_2}f_i(z)\right]^l+ \\ &+\sum_{m=0}^{n_i-1}p_{j,m}z^m\left[1-\left(\frac{a_i}{a_1+a_2}z\right)^{n_i-m}\right]\gamma_{ji}(a_1+a_2-a_iz-a_jf_i(z))\equiv 0 \\ &\qquad i=1,\ j=2 \quad \text{и} \quad i=2,\ j=1, \end{aligned} \tag{22}$$

где

$$P_1[z]=P_1(0,z), \quad P_2[z]=P_2(z,0), \quad p_{1,m}=p_{1,0,m}, \quad p_{2,m}=p_{2,m,0}. \tag{23}$$

Дифференцируем равенство (22) в точке $z=1$. Обозначим n-ую производную функции $P_i[z]$ через $P_i^{(n)}[z]$, $(P_i^{(0)}[z]=P_i[z])$. Из равенства (22) при $i=1$, $j=2$ и $i=2$, $j=1$ $P_i^{(n)}[1]$, $n=1, 2, \ldots$; $i=1, 2$, получается как линейное выражение значений $P_i^{(k)}[1]$, $i=1, 2$; $k=0, 1, \ldots, n-1$ и вероятностей $p_{1,m}$, $m=0, 1, \ldots, n_2-1$ и $p_{2,m}$; $m=0, 1, \ldots, n_1-1$. Здесь определитель

$$\begin{vmatrix} f_1'(1)^n & -1 \\ -1 & f_2'(1)^n \end{vmatrix}=\left[\frac{1-a_1\beta_{11}}{a_2\beta_{11}}\frac{1-a_1\beta_{21}}{a_1\beta_{21}}\right]^n-1>0.$$

С помощью значений $P_i^{(n)}[1]$, $i=1, 2$, $n=0, 1, \ldots$ построим степенной ряд функции $P_i[z]$, $i=1, 2$ при $z=1$:

$$\sum_{k=0}^{\infty}\frac{P_i^{(k)}[1]}{k!}(z-1)^k, \qquad (i=1, 2). \tag{24}$$

Дифференцируем l раз функцию (24) при $z=0$, получим

$$p_{i,l}=\sum_{k=0}^{\infty}\frac{(-1)^kP_i^{(k+l)}[1]}{k!}, \qquad i=1, 2,\ l=0, 1, 2, \ldots \tag{25}$$

Как уже отмечали $P_i^{(n)}[1]$, $i=1, 2$ линейно выражается через $P_i^{(k)}[1]$, $k=0, 1\ldots$, $i=1, 2$; $p_{1,m}$, $m=0, 1, \ldots, n_2-1$ и $p_{2,m}$, $m=0, 1, \ldots, n_1-1$, поэтому правая сторона равенства (25) является линейным выражением $p_{1,m}$, $m=0, 1, \ldots, n_2-1$ и $p_{2,m}$, $m=0, 1, \ldots, n_1-1$. Это выражение по формуле (18) является неоднородным и имеет следующий вид:

$$p_{i,l}=\sum_{k=0}^{n_2-1}a_{i,k}^{(l)}p_{1,k}+\sum_{k=0}^{n_1-1}b_{i,k}^{(l)}p_{2,k}+c_i^{(l)} \qquad (i=1, 2; \quad l=0, 1, 2, \ldots) \tag{26}$$

где $a_{i,k}^{(l)}$, $b_{i,k}^{(l)}$ и $c_i^{(l)}$ неизвестные коэффициенты.

Пусть $q_{i,k}^{(m)}$, $i=1, 2$; $k=0, 1, \ldots, n_j-1$; $j=1, 2$; $i\neq j$; $m=0, 1, \ldots, n_1+n_2$ такие числа, для которых

$$q_{i,k}^{(m)}>0, \quad \sum_{i,k} q_{i,k}^{(m)}<1.$$

При фиксированном значении l подставим в правую часть равенства (26) числа $p_{i,k}=q_{i,k}^{(m)}$ и этими же числами вычисляем правую часть (25). Сравнивая их при $m=0, 1, \ldots, n_1+n_2$ получим систему линейных уравнений

$$\sum_{k=0}^{n_2-1} q_{i,k}^{(m)} a_{i,k}^{(l)}+\sum_{k=0}^{n_1-1} q_{2,k}^{(m)} b_{i,k}^{(l)}+c_i^{(l)}=\sum_{k=0}^{\infty}\frac{(-1)^k P_i^{(k+l)}[1]}{k!} \tag{27}$$

$$i=1, 2, \quad m=0, 1, \ldots, n_1+n_2$$

для определения коэффициентов $a_{i,k}^{(l)}$, $b_{i,k}^{(l)}$ и $c_i^{(l)}$. Необходимо выбрать числа $q_{i,k}^{(m)}$ так, чтобы определитель системы (27) не становился 0.

Вычисление правой части (27) происходит следующим образом: при значении $p_{i,k}=q_{i,k}^{(m)}$ из (17) и (18) выражаем значения $P_i[1]$, потом изложенным способом определяем $P_i^{(k)}[1]$, $k\geqq 1$, и подставим их в правую часть (25). Целесообразно выбирать числа $q_{i,k}^{(m)}$ так, чтобы они являлись грубыми оценками вероятностей $p_{i,k}$.

Разрешая (27) при $i=1, 2$; $l=0, 1$; $\ldots, n_j-1$; $j=1, 2$; $i\neq j$, из (26) получим значения $p_{i,l}$ такими же индексами, а из (25) при $i=1, 2$; $l=n_j, n_j+1, \ldots$; $j=1, 2$; $i\neq j$.

Функции $\beta_i(a_1+a_2-a_1z_1-a_2z_2)$, $\gamma_{ij}(a_1+a_2-a_1z_1-a_2z_2)$ при $i=1$, $j=2$ и $i=2$, $j=1$ аналитичные в области (13), поэтому их можно разложить в степенные ряды. Подставим $z_j=0$ ($i=1$, $j=2$ и $i=2$, $j=1$) в (15). Пусть для функций $\beta_i(a_1+a_2-a_iz_i)$ и $\beta_i(a_1+a_2-a_iz_i)\gamma_{ij}(a_1+a_2-a_iz_i)$ по z_i в точке $z_i=0$ при $i=1$, $j=2$ и $i=2$, $j=1$ степенными рядами будут

$$\beta_i(a_1+a_2-a_iz_i)=\sum_{k=0}^{\infty} b_{i,k}z_i^k,$$

$$\beta_i(a_1+a_2-a_iz_i)\gamma_{ji}(a_1+a_2-a_iz_i)=\sum_{k=0}^{\infty} d_{i,k}z_i^k.$$

Подставляя эти выражения и выражение (14) в уравнение (15) при $i=1$, $j=2$ и $i=2$, $j=1$, $z_j=0$ с помощью сравнения коэффициентов z_i^k, получим:

$$P_{i,k}\cdot b_{i,0}=p_{i,k-1}+p_{i,0}\left(b_{i,k}-\frac{a_i}{a_1+a_2}b_{i,k-1}\right)- \tag{28}$$

$$-\sum_{l=0}^{k-1} p_{i,l}b_{i,k-l}+\sum_{l=0}^{k} p_{j,l}d_{i,k-l}-D_{i,k},$$

где

$$D_{i,k}=\begin{cases}\sum\limits_{l=0}^{n_i-1} p_{j,l}d_{i,k-l}-d_{i,k-n_1+1}, & \text{если} \quad k\geqq n_i-1\\ \sum\limits_{l=0}^{k} p_{j,l}d_{i,k-l}, & \text{если} \quad k<n_i-1,\end{cases}$$

$$p_{1,k}=p_{1,k,0}, \quad p_{2,k}=p_{2,0,k}.$$

Из уравнения (28) можно по очереди определить вероятности $p_{i,k}$ при $i=1, 2,\ k=1, 2, \ldots$.

Введем следующие обозначения:

$$m_{i,k}^{(1)} = \sum_{l=0}^{\infty} p_{i,k,l}, \quad m_{i,k}^{(2)} = \sum_{l=0}^{\infty} p_{i,l,k}, \qquad i=1, 2. \tag{29}$$

Мы уже определили значения $m_{1,0}^{(1)} = P_1(0, 1)$ и $m_{2,0}^{(2)} = P_2(1, 0)$. Используя (23) и введя обозначения $\bar{P}_1[1] = P_1(1, 0)$ и $\bar{P}_2[1] = P_2(0, 1)$ значения $m_{1,0}^{(2)} = P_1(1, 0)$ и $m_{2,0}^{(1)} = P_2(0, 1)$ определяются из следующих уравнений при $i=1,\ j=2$ и $i=2,\ j=1$:

$$\bar{P}_i[1]\left[1-\frac{1}{\beta_i(a_j)}\right]+\bar{P}_j[1]\gamma_{ji}(a_j) = \tag{30}$$

$$= \frac{a_j}{a_1+a_2} p_{i,0} + \gamma_{ji}(a_j) \sum_{l=0}^{n_i-1} p_{j,l}\left[1-\left(\frac{a_i}{a_1+a_2}\right)^{n_i-l}\right],$$

которые получаются из (15), при $z_t=1$ и $z_t=0$.

При $i=1,\ j=2$ и $i=2,\ j=1$ в (15) поочередно подставляем $z_j=1$ и $z_i=1$ и выпишем степенные ряды обеих частей этих равенств по z_i, потом по z_j. Сравнивая коэффициенты z_i^k, а также z_j^k получим

$$m_{i,k}^{(1)} \bar{b}_{i,0} = m_{i,k-1}^{(1)} + m_{i,0}^{(1)} \bar{b}_{i,k} - \sum_{l=0}^{k-1} m_{i,l}^{(1)} \bar{b}_{i,k-l} - \tag{31}$$

$$- \sum_{l=0}^{n_j-1} p_{i,l}\left[1-\left(\frac{a_j}{a_1+a_2}\right)^{n_j-l}\right] \bar{b}_{i,k-1} - \sum_{l=0}^{k} p_{j,l} d_{i,k-l} - \bar{D}_{i,k};$$

где

$$\bar{D}_{i,k} = \begin{cases} \sum_{l=0}^{n_i-1} p_{j,l} \bar{d}_{i,k-l} - d_{i,k-n_i-1}, & \text{если} \quad k \geqq n_i-1 \\ \sum_{l=0}^{k} p_{j,l} \bar{d}_{i,k-l}, & \text{если} \quad k < n_i-1, \end{cases}$$

$$m_{i,k}^{(2)}(1-\bar{b}_{j,0}) = \sum_{l=0}^{k-1} m_{i,l}^{(2)}(\bar{b}_{j,k-l}-1) - \sum_{l=0}^{k} p_{i,l} \bar{b}_{j,k-l} + m_{2,0}^{(2)} \bar{d}_{i,k} - \tag{32}$$

$$- \sum_{l=0}^{n_i-1} p_{j,l}\left[1-\left(\frac{a_i}{a_1+a_2}\right)^{n_i-l}\right] \bar{d}_{j,k} + \frac{a_i}{a_1+a_2} \sum_{r=0}^{r_0} p_{i,r} \sum_{l=0}^{n_j-r-1} \left(\frac{a_j}{a_1+a_2}\right)^{l} \bar{b}_{j,k-r-l},$$

где $r_0 = \min(k, n_j-1)$, $i=1, j=2$ или $i=2, j=1$, $k=1, 2, \ldots$, $\bar{b}_{i,k}$ и $\bar{d}_{i,k}$ коэффициенты степенных рядов функций $\beta_i(a_i-a_i z_i)$ и $\beta_i(a_i-a_i z_i)\gamma_{ji}(a_i-a_i z_i)$. Из равенств (31) и (32) можно поочередно вычислить значения $m_{i,k}^{(1)}$ и $m_{i,k}^{(2)}$, $i=1, 2$.

Вероятностные толкования величин $m_{i,k}^{(1)}$ и $m_{i,k}^{(2)}$ по формулам (29) известны.

Для контроля правильности вычислений применяются формулы

$$\sum_{k=0}^{\infty} m_{i,k}^{(1)} = \sum_{k=0}^{\infty} m_{i,k}^{(2)} = \frac{a_i}{a_1+a_2}, \qquad i=1, 2.$$

4. Пример

Определяются значения $p_{1,0,0}$ и $p_{2,0,0}$ (в дальнейшем p_1 и p_2) для следующего случая: $n_1=n_2=1$,

$$B_i(t)=1-e^{-\lambda_i t}, \quad C_{ij}(t)=1-e^{-\mu_i t}, \qquad i=1,\ j=2 \quad \text{и} \quad i=2,\ j=1,$$

где $\lambda_i, \mu_i>0$. Тогда

$$\beta_i(z)=\frac{\lambda_i}{\lambda_i+z}, \quad \gamma_{ij}(z)=\frac{\mu_i}{\mu_i+z},$$

а из равенства (20)

$$f_i(z)=-\frac{\lambda_i}{a_j}\frac{1}{z}-\frac{a_i}{a_j}z+\frac{a_1+a_2+\lambda_i}{a_j}$$

Уравнения (17) и (18) в нашем примере (используя (23)):

$$P_1[1]-P_2[1]=\frac{a_1}{a_1+a_2}p_1-\frac{a_2}{a_1+a_2}p_2;$$

(33)

$$\frac{1}{\mu_1}P_1[1]+\frac{1}{\mu_2}P_2[1]=\frac{1}{a_1+a_2}\left\{1-\frac{a_1}{\lambda_1}-\frac{a_2}{\lambda_2}-\left(1-\frac{a_1}{\mu_1}\right)p_1-\left(1-\frac{a_2}{\lambda_2}\right)p_2\right\},$$

Дифференцируем n-раз уравнение (22) в точке $z=1$, мы получим, при $i=1$, $j=2$ и $i=2$, $j=1$, что при $n=1$

$$\frac{\lambda_i-a_i}{a_j}P_i^{(1)}[1]-P_j^{(1)}[1]=\frac{a_i}{a_1+a_2}(p_1+p_2)-\frac{\lambda_i}{\mu_2}\frac{a_j}{a_1+a_2}p_j-\frac{\lambda_i}{\mu_j}P_j[1], \tag{34}$$

а при $n>1$

$$\left(\frac{\lambda_i-a_i}{a_j}\right)^n P_i^{(n)}[1]-P_j^{(n)}[1]=\frac{\lambda_i}{\mu_j}\sum_{k=1}^{n}\binom{n}{k}\left(\frac{\lambda_i}{\mu_j}-1\right)^{k-1}P_j^{(n-k)}[1]+$$

(35)

$$+n!\frac{\lambda_i}{\mu_j}\left(\frac{\lambda_i}{\mu_j}-1\right)^{n-2}\left[\frac{a_i}{a_1+a_2}-\left(\frac{\lambda_i}{\mu_j}-1\right)\frac{a_j}{a_1+a_2}\right]p_j-Q_{i,n},$$

где

$$Q_{i,n}=\{P_i[f_i(1)]\}^{(n)}-P_i^{(n)}[1]\{f_i'(1)\}^n.$$

$Q_{i,n}$ являются линейными выражениями значений $P_i^{(k)}[1]$, при $i=1,2$ и $k=2, 3, \ldots, n-1$. Например:

$$Q_{i,2}=-2\frac{\lambda_i}{a_j}P_i^{(1)}[1],$$

$$Q_{i,3}=6\frac{\lambda_i}{a_j}P_i^{(1)}[1]-6\frac{\lambda_i-a_i}{a_j}\frac{\lambda_i}{a_j}P_i^{(2)}[1],$$

и т. д.

Ход вычисления следующий: Выберем изложенным способом три значения для p_i, обозначим эти через $q_i^{(m)}$, $m=1, 2, 3$, $i=1, 2$. При $p_i=q_i^{(m)}$ решаем

системы линейных уравнений (33), (34) и при $n=2, 3, 4, 5, 6$ (35) и вычислим следующие значения

$$R_i(q_1^{(m)}, q_2^{(m)}) = \sum_{k=0}^{k_0} (-1)^k \frac{P_i^{(k)}[1]}{k!}, \qquad i=1,2,\ m=1,2,3, \tag{36}$$

где k_0 такое большое число, для которого значение

$$\sum_{k=k_0}^{\infty} (-1)^k \frac{P_i^{(k)}[1]}{k!}$$

достаточно малое. ((36) приближенно соответствует правой части (25) при $l=0$). По выражению (27) при $i=1,2$:

$$a_i q_1^{(m)} + b_i q_2^{(m)} + c_i = R_i(q_1^{(m)}, q_2^{(m)}), \qquad m=1,2,3. \tag{37}$$

Решим систему линейных уравнений (37) для a_i, b_i и c_i при $i=1, 2$. Значения p_1 и p_2 получим из уравнений

$$(1-a_1)p_1 - b_1 p_2 = c_1$$
$$-a_2 p_1 + (1-b_2)p_2 = c_2.$$

Вычисления производились на вычислительной машине Урал-2.

Некоторые численные результаты для примера показывают рис. 1., рис. 2. и рис. 3.

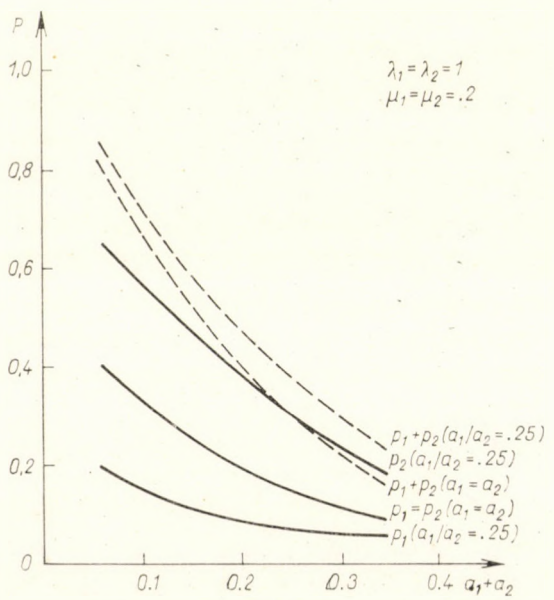
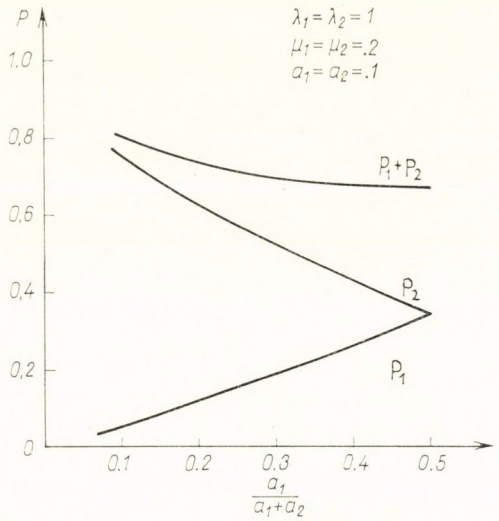
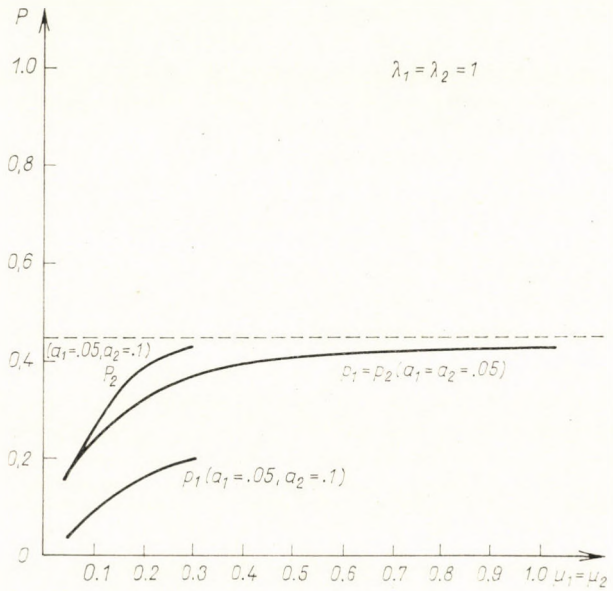


Рис 1


P
1.0
0,8
0,6
0,4
0,2
0
0.1
0.2
0.3
0.4
0.5
$\frac{a_1}{a_1+a_2}$
$\lambda_1=\lambda_2=1$
$\mu_1=\mu_2=.2$
$a_1=a_2=.1$
P_1+P_2
P_2
P_1

Puc 2


P
1.0
0,8
0,6
0,4
0,2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0 $\mu_1=\mu_2$
$\lambda_1=\lambda_2=1$
$(a_1=.05, a_2=.1)$
P_2
$p_1=p_2\,(a_1=a_2=.05)$
$p_1\,(a_1=.05,\ a_2=.1)$

Puc 3

Замечание. Если $\lambda=\lambda_1=\lambda_2$, $\mu_1\to\infty$, $\mu_2\to\infty$, тогда система (в случае нашего примера) переходит в модель Эрланга, где входящий параметр a_1+a_2, а параметр обслуживания λ. В этом случае вероятность того, что система свободна

$$p=p_1+p_2=1-\frac{a_1+a_2}{\lambda}.$$

Отсюда, если $\lambda=1$, $a_1+a_2=0{,}1$, тогда $p=0.9$ и $p_1=p_2=0.45$. Рис. 3. и дальнейшие вычисления показывают что, p_1 и p_2 приближаются к уровню 0.45. Например если $\mu_1=\mu_2=1.$, 1.5, 2, 4, тогда $p_1=p_2=0.425$, 0.433, 0.437, 0.444.

БИБЛИОГРАФИЯ

[1] GAVER, D. P.: Competetive queueing: idleness probabilities under priority disciplines, *J. R. Statist. Soc. B*, **25** (1963) 489—499.

[2] КЛИМОВ, Г. П.: *Стохастические системы обслуживания,* Москва, Издательство Наука, 1966.

Вычислительный Центр Академии Наук Венгрии, Будапешт

(Поступила 21-ого марта 1967 г.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 181—194.

ZUR ZWEISTUFIGEN SATZSTRUKTUR-GRAMMATIK, II[1]

von

R. PÉTER

Bereits in I habe ich darüber bereichtet, daß — wie ich von L. KALMÁR erfahren habe — unter den Vorschlägen bezüglich der Weiterbildung der algorithmischen Sprache ALGOL 60 auch eine solche Sprache aufgeworfen wurde, welche durch unendlich viele Satzstruktur-Produktionen definiert wird, derart, daß diese Produktionen selber durch eine besondere Metasprache generiert werden; ferner habe ich die Lösung des sich hier natürlicherweise erhebenden Problems angedeutet: ob bzw. unter welchen Bedingungen eine derart zweistufig generierte Sprache auch einstufig (durch endlich viele Produktionen) definiert werden kann. Die genauen Definitionen und Beweise gebe ich in vorliegender Arbeit an.

1. Die zweistufige Definition einer Sprache

Nach unwesentlichen Abänderungen erhält man die folgende Abstraktion für eine allgemeine zweistufig definierte Sprache. Diese wird durch ein geordnetes Quintupel

$$(Z, M, P, V, K)$$

endlicher disjunkter und die Zeichen „ , " und „:" nicht enthaltender Mengen angegeben, wobei die Elemente von Z **Zeichen**, die Elemente von M **Metazeichen,** die Elemente von P **Metaproduktionen,** die Elemente von V **primitive Vorproduktionen** und die Elemente von K **Kategorienamen** genannt werden. Dabei gehört hier die **Metasprache,** als die erste Stufe, zur zweistufigen Definition, hat also einen ganz anderen Sinn, als die früher übliche „Metaalgol-Sprache", welche hier nach CURRY „Episprache" genannt werden soll.

Mit Hilfe der Episprache werde ich die Begriffe **Zeichenkette, Zeichenkettenliste, potentielle Produktion, Mischkette, Mischkettenliste, potentielle Metaproduktion, potentielle Vorproduktion** definieren, und diese Begriffe dann zur näheren Angabe der Mengen P, V, K benutzen.

Zuerst kommt

< Zeichen > :: = ...

[1] Siehe: PÉTER, R., „Zur zweistufigen Satzstruktur-Grammatik I", *Diese Zeitschrift* **2** (1967) S. 455—456; kurz als "I" zitiert.

wobei auf der rechten Seite von „::=" sämtliche Elemente von Z stehen, durch Zeichen „|" von einander getrennt. Ebenso

< Metazeichen > ::= . . . . . .

mit den Elementen von M auf der rechten Seite, durch Zeichen „|" getrennt. Dann folgen

< Zeichenkette > ::= < Zeichen > | < Zeichen-
kette > < Zeichen >

< Zeichenkettenliste > ::= < Zeichenkette > | < Zeichen-
kettenliste > < Komma > < Zeichenkette >

< potentielle Produktion > ::= < Zeichenkette > < Doppel-
punkt > < Zeichenkettenliste >

< Mischkette > ::= < Zeichenkette > | < Metazei-
chen > | < Mischkette > < Zeichenkette > | < Misch-
kette > < Metazeichen >

< Mischkettenliste > ::= < Mischkette > | < Mischketten-
liste > < Komma > < Mischkette >

< potentielle Metaproduktion > ::= < Metazeichen > < Dop-
pelpunkt > < Mischkette >

< potentielle Vorproduktion > ::= < Mischkette > < Dop-
pelpunkt > < Mischkettenliste >

< Komma > ::= ,

< Doppelpunkt > ::= :

Wie man sieht, wird das Zeichen „::=" der Definitionsgleichheit der Episprache in den Produktionen unserer zweistufigen Definition durch „:" ersetzt; ferner wird jede an keine andere Spitzklammer stossende Spitzklammer weggelassen, und die an einander stossenden „><" werden durch „ ," ersetzt (dazu gehört jedenfalls, daß erst auch die „terminalen Begriffe" — deren genaue Angabe später folgt — zwischen Spitzklammern gesetzt gedacht sind); endlich treten in diesen Produktionen keine Zeichen „|" für „oder" auf. Auch das letzte ist unwesentlich; der Gebrauch des Zeichens „|" bedeutet ja nur eine kürzere Schreibweise. Wird „::=" als „ist nach einer der möglichen Definitionen" gelesen, so ist z.B. die in der Episprache angegebene Definition der < Zeichenkette > nur eine Vereinigung von

< Zeichenkette > ::= < Zeichen >

und

< Zeichenkette > ::= < Zeichenkette > < Zeichen >

Die Menge P ist nun eine (endliche) Teilmenge der Menge der potentiellen Metaproduktionen, und V ist eine (endliche) Teilmenge der potentiellen Vorproduktionen.

Zur näheren Bestimmung von K gehören weitere Definitionen.

Unter einer **unmittelbaren Entwicklung eines Metazeichens** $m \in M$ wird eine beliebige Mischkette μ verstanden, für welche

$$m : \mu \in P$$

gilt. Eine Mischkette μ heisst eine **Entwicklung eines Metazeichens** $m \in M$, falls

a) μ eine unmittelbare Entwicklung von m ist,

b) oder eine solche Entwicklung $\mu_1 m' \mu_2$ von m und eine solche unmittelbare Entwicklung μ' von m' existiert, daß

$$\mu = \mu_1 \mu' \mu_2$$

ist (wobei natürlich m' ein Metazeichen, μ' eine Mischkette und sowohl μ_1 als auch μ_2 entweder leer oder eine Mischkette ist).

Eine Entwicklung eines Metazeichens wird **terminal** genannt, falls sie keine Metazeichen enthält. Man kann sich auf solche Fälle beschränken, wobei jedes Metazeichen als linke Seite von Metaproduktionen auftritt, und wobei kein Metazeichen seine eigene Entwicklung ist; dann besitzt ein beliebiges Metazeichen m allgemein unendlich viele terminalen Entwicklungen; ich werde diese kurz die **Werte** von m nennen, und die Zeichenketten, die aus einer Mischkette durch Ersetzen jedes darin enthaltenen Metazeichens durch einen seiner Werte (in jedem Vorkommen eines Metazeichens durch denselben Wert) entstehen, werde ich die Werte der Mischkette nennen.

Durch Einsetzen ihrer Werte für die Metazeichen werden wie folgt weitere **Vorproduktionen** aus den primitiven generiert. (Im Begriff „Vorproduktion" sind auch die primitiven Vorproduktionen mitenthalten.)

Betrachten wir eine (primitive oder bereits generierte) Vorproduktion, und nehmen wir an, dass sie eine potentielle Vorproduktion, d.h. der Form

$$\mu : \Lambda$$

ist, wobei μ eine Mischkette und Λ eine Mischkettenliste ist. Kommen darin keine Metazeichen vor, ist also μ eine Zeichenkette und Λ eine Zeichenkettenliste (und so die betrachtete Vorproduktion eine potentielle Produktion), dann sagen wir, daß dies eine **Produktion** ist, und generieren daraus keine weitere Vorproduktionen.

Andernfalls sei m ein in $\mu : \Lambda$ auftretendes Metazeichen, und sei λ eine beliebige terminale Entwicklung von m, ferner sei μ' jene Mischkette und Λ' jene Mischkettenliste, welche aus μ bzw. aus Λ entsteht, wenn in diesen jedes Vorkommen von m durch λ ersetzt wird. Dann ist auch

$$\mu' : \Lambda'$$

eine potentielle Vorproduktion; und auch diese soll als Vorproduktion gelten. Andere Vorproduktionen bzw. Produktionen als die derart generierten sollen nicht zugelassen werden.

So ist die (allgemein unendliche) Menge R der Produktionen eine Teilmenge der Menge der potentiellen Produktionen.

Die Zeichenketten, die als linke Seiten oder als Glieder der rechten Seiten (bezüglich des Zeichens „:") der Produktionen auftreten, sollen **Begriffe** genannt werden (dabei lautet die exakte Definition der **Glieder** einer Zeichenkettenliste Λ wie folgt: ist Λ eine Zeichenkette, so ist es selber sein einziges Glied; sonst ist Λ der Form $\Lambda = \Lambda', \lambda$ wobei Λ' eine Zeichenkettenliste und λ' eine Zeichenkette ist; und die Glieder von Λ sind die Glieder von Λ' und noch λ). Unter den Begriffen werden **potentielle Kategorienamen** und **terminale Begriffe** unterschieden, je nachdem sie unter den linken Seiten der Produktionen vorkommen oder nicht.

Nun ist die Menge K eine (endliche) Teilmenge der Menge der potentiellen Kategorienamen.

Unter einer **unmittelbaren Entwicklung eines potentiellen Kategorienamens** λ wird eine beliebige Zeichenkettenliste Λ verstanden, für welche

$$\lambda : \Lambda \in R$$

gilt. Eine Zeichenkettenliste Λ heisst eine **Entwicklung eines potentiellen Kategorienamens** λ, falls

a) Λ eine unmittelbare Entwicklung von λ ist,

b) oder eine solche Entwicklung $\Lambda_1^* \lambda' {}^*\Lambda_2$ von λ und eine solche unmittelbare Entwicklung Λ' von λ' existiert, dass

$$\Lambda = \Lambda_1^* \Lambda' {}^*\Lambda_2$$

ist, wobei λ' ein potentieller Kategoriename, Λ' eine Zeichenkettenliste, Λ_1^* leer oder eine Zeichenkettenliste mit einem danach gesetzten Komma und ${}^*\Lambda_2$ leer oder eine Zeichenkettenliste mit einem davor gesetzten Komma ist.

Eine Entwicklung eines potentiellen Kategorienamens wird **terminal** genannt, falls sie nur terminale Begriffe als Glieder enthält. Die Menge sämtlicher terminalen Entwicklungen eines potentiellen Kategorienamens ergibt die durch diesen Namen bezeichnete **potentielle Kategorie**. Die durch die Elemente von K bezeichneten potentiellen Kategorien heissen **Kategorien**. Unter der betrachteten **Sprache** wird die Zuordnung der Kategorien zu ihren Namen verstanden.

Diese Sprache wurde eigentlich durch zwei kontextunabhängigen Satzstruktur-Grammatiken[2] generiert. Zur ersten gehört Z als terminales Vokabular, M als Hilfsvokabular, P als Produktionenmenge, und sämtliche Elemente von M gelten als ausgezeichnete. Die darüber gebaute zweite Grammatik ist eine Satzstruktur-Grammatik in einem verallgemeinerten Sinn: die dazu gehörigen Mengen sind nämlich unendlich. Die Elemente des terminalen Vokabulars sind hier die (allgemein unendlich vielen) terminalen Begriffe, die Elemente des Hilfsvokabulars die (allgemein ebenfalls unendlich vielen) potentiellen Kategorienamen, und so ist die dazu gehörige Produktionenmenge R allgemein auch unendlich. Jedoch gibt es auch hier nur endlich viele ausgezeichneten Hilfsbegriffe: diese sind nämlich die Elemente von K. Die Sprache wird durch die Mengen der terminalen Entwicklungen dieser endlich vielen ausgezeichneten Hilfsbegriffe bestimmt.

So erhebt sich die Frage, ob diese Sprache etwa auch durch endlich viele Produktionen, also einstufig, durch eine einzige Satzstruktur-Grammatik (im ursprünglichen Sinn) generiert werden kann.

[2] Dieser Begriff wurde von CHOMSKY eingeführt; siehe z.B. CHOMSKY, N., *Syntactic Structures* ('S-Gravenhage, 1957).

2. Beispiel einer zweistufig generierten Sprache, die sich einstufig nicht generieren läßt

Betrachten wir die durch den Mengenquintupel

$$(Z;\ M;\ P;\ V;\ K)$$

zweistufig generierte Sprache, wo

$$Z = \{z_1; z_2; z_3\};\quad N = \{m\};\quad P = \{m: z_1 m z_1;\ m: z_2\};$$

$$V = \{z_3: m z_2 m\};\quad K = \{z_3\}$$

ist. (Bei Aufzählungen gebrauche ich Strichpunkte statt Kommata, da hier das Komma eine besondere Rolle erhalten hat.)

Man sieht leicht, dass hier sämtliche Werte (d.h. terminale Entwicklungen) des einzigen Metazeichens m die folgenden sind:

$$\underbrace{z_1 z_1 \dots z_1}_{n\text{-mal}}\, z_2\, \underbrace{z_1 z_1 \dots z_1}_{n\text{-mal}} \quad \text{für} \quad n = 0; 1; 2; \dots$$

und daraus ergeben sich sämtliche terminale Entwicklungen des einzigen Kategorienamens z_3 als

$$(*) \qquad \underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n} \quad \text{für} \quad n = 0; 1; 2; \dots$$

Wie man sieht, kommen unter diesen Zeichenketten auch beliebig lange vor (wobei unter der **Länge** einer Zeichenkette die Anzahl der darin enthaltenen Zeichen verstanden wird, jedes sovielmal gerechnet, wievielmal es auftritt). Die Frage ist, ob diese Zeichenketten auch durch endlich viele Produktionen, genauer durch eine kontextunabhängige Satzstruktur-Grammatik im ursprünglichen Sinn (wobei der ausgezeichnete Hilfsbegriff z_3 ist) generiert werden können.

Es gelten nun folgende Sätze von Y. BAR-HILLEL—M. PERLES—E. SHAMIR[3]:

I. Zu einer kontextunabhängigen Satzstruktur-Grammatik gibt es immer eine andere, mit demselben terminalen Vokabular, Hilfsvokabular und ausgezeichneten Hilfsbegriff, welche — mit Ausnahme einer endlichen Teilmenge der Sprache — dieselbe Sprache erzeugt, und so beschaffen ist, dass die rechten Seiten ihrer Produktionen mindestens der Länge 2 sind.

II. Durch eine kontextunabhängige Satzstruktur-Grammatik, deren Produktionen Zeichenketten mindestens der Länge 2 als rechte Seiten enthalten, sei eine Sprache S generiert. Dann gibt es eine natürliche Zahl q, so dass jede zu S gehörige Zeichenkette λ grösserer Länge als q in der Form

$$\lambda = \lambda' \lambda_1 \lambda'' \lambda_2 \lambda'''$$

[3] BAR-HILLEL, Y., GAIFMAN, C., SHAMIR, E.: "On formal properties of simple phrase structure grammars", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **14** (1961) S. 143—172. — Ich sage hier nur die benützten Teile der betreffenden Sätze aus.

geschrieben werden kann, wo unter den leeren oder nicht leeren Zeichenketten λ'; λ_1; λ''; λ_2; λ''' die Zeichenketten λ_1 und λ_2 nicht beide leer sind, und auch

$$\lambda'\lambda_1\lambda_1\lambda''\lambda_2\lambda_2\lambda'''$$

zu S gehört.

In unserem Fall handelt es sich um die Sprache, die aus den zu (*) gehörigen Zeichenketten besteht. Nehmen wir an, daß diese durch eine kontextunabhängige Satzstruktur-Grammatik generiert werden kann. Dann können nach I die zu (*) gehörigen Zeichenketten von einem genügend großen n an auch durch eine solche kontextunabhängige Satzstruktur-Grammatik generiert werden, deren Produktionen als rechte Seiten Zeichenketten mindestens der Länge 2 enthalten. So lässt sich II auf die aus diesen Zeichenketten bestehende Sprache anwenden: eine zu (*) gehörige Zeichenkette mit genügend großem n kann in der Form

$$\lambda'\lambda_1\lambda''\lambda_2\lambda'''$$

geschrieben werden, wo die Zeichenketten λ_1 und λ_2 nicht beide leer sind, und auch

$$\lambda'\lambda_1\lambda_1\lambda''\lambda_2\lambda_2\lambda'''$$

zu (*) gehört. Also gilt mit einem genügend großen n und einem $n'>n$

$$\underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n}\, z_2\, \underbrace{z_1 \dots z_1}_{n} = \lambda'\lambda_1\lambda''\lambda_2\lambda'''$$

und

$$\underbrace{z_1 \dots z_1}_{n'}\, z_2\, \underbrace{z_1 \dots z_1}_{n'}\, z_2\, \underbrace{z_1 \dots z_1}_{n'}\, z_2\, \underbrace{z_1 \dots z_1}_{n'} = \lambda'\lambda_1\lambda_1\lambda''\lambda_2\lambda_2\lambda'''.$$

Auf beiden linken Seiten kommt das Zeichen z_2 genau 3-mal vor. Die Verdoppelung von λ_1 und λ_2 würde aber auch das Vorkommen von z_2 vermehren, falls eines von diesen Zeichenketten das Zeichen z_2 enthalten würde. Demnach sind λ_1 und λ_2 Teile je einer aus n z_1-Zeichen bestehenden Teilkette (die auch für beide dieselbe sein kann). Durch die Verdoppelung von λ_1 und λ_2 vermehrt sich daher die Anzahl der z_1-Zeichen in zwei Teilketten dieser Art; in den übrigen bleibt diese Anzahl unverändert n, kann also nicht zu $n'>n$ anwachsen.[4]

So sind wir in Widerspruch geraten mit der Annahme, daß die betrachtete zweistufig generierte Sprache auch einstufig, durch eine kontextunabhängige Satzstruktur-Grammatik generiert werden kann. Die Einführung der zweistufigen Definitionen ist also bereits in einfachen Fällen eine prinzipielle Notwendigkeit.

Die Einfachheit der betrachteten Sprache zeigt sich auch darin, dass **die Menge ihrer terminalen Begriffe (die hier mit der Sprache übereinstimmt) in der Wortemenge über das Alphabet $\{z_1; z_2\}$ primitiv-rekursiv ist.**[5]

[4] Das Beispiel ist mit unwesentlicher Abweichung dasselbe als ein für andere Zwecke schon in der in Fussnote[3] zitierten Arbeit diskutiertes Beispiel.

[5] Zu diesen Begriffen siehe z.B. PÉTER, R., „Über die Rekursivität der Begriffe der mathematischen Grammatiken", *Publications of the Math. Inst. of the Hungarian Acad. of Sciences* **8** (1963) S. 213—228.

Sogar mit endlich vielen terminalen Begriffen ergibt sich ein ähnliches Gegenbeispiel, falls auch in den Metaproduktionen nicht nur Mischketten, sondern auch Mischkettenlisten als rechte Seiten zugelassen werden. Es genügt dazu zwischen die Metazeichen und die Zeichenketten in den Mischketten, die in den Elementen von P und V auftreten, Kommata zu setzen, also statt diese Mengen die folgenden aufzunehmen:

$$P' = \{m: z_1, m, z_1; m: z_2\}$$

$$V' = \{z_3: m, z_2, m\}.$$

Dann sind die Werte von m die Zeichenkettenlisten

$$\underbrace{z_1, \ldots, z_1}_{n}, z_2, \underbrace{z_1, \ldots, z_1}_{n} \qquad (n = 0; 1; 2; \ldots)$$

und daher sämtliche Produktionen:

$$z_3 : \underbrace{z_1, \ldots, z_1}_{n}, z_2, \underbrace{z_1, \ldots, z_1}_{n}, z_2, \underbrace{z_1, \ldots, z_1}_{n}, z_2, \underbrace{z_1, \ldots, z_1}_{n} \qquad (n = 0; 1; 2; \ldots)$$

wo die Menge der rechten Seiten die Sprache ergibt.

Wie man sieht, besitzt diese Sprache nur zwei terminalen Begriffe: z_1 und z_2. Trotzdem kann sie nicht durch endlich viele Produktionen generiert werden. Denn sonst könnten, nach Hinzunahme des Zeichens "," zum terminalen Vokabular, die Zeichenkettenlisten auf den rechten Seiten der Produktionen auch als Zeichenketten betrachtet werden, und so hätte man es mit einer kontextunabhängigen Satzstruktur-Grammatik im üblichen Sinn zu tun. Doch genau so wie vorhin kann man einsehen, daß diese Sprache durch keine solche Grammatik generiert werden kann.

3. Die Existenz einer einstufigen Definition für eine zweistufig definierte Sprache im Fall endlich vieler terminalen Begriffe

Von nun an halten wir uns an die ursprüngliche Definition, **wobei die rechten Seiten der Metaproduktionen Mischketten sind,** und dabei **beschränken wir uns auf solche zweistufig generierte Sprachen, welche nur endlich viele terminalen Begriffe enthalten** (möglicherweise wird die Weiterbildung vom ALGOL 60 zu einer solchen Sprache führen). Ich werde beweisen, daß **sich eine solche Sprache auch durch endlich viele Produktionen generieren lässt.**

Die Grundgedanke des Beweises ist: Gibt es nur endlich viele terminalen Begriffe, so können die von diesen und von den (ebenfalls endlich vielen) Kategorienamen verschiedenen Werte der Glieder der rechten Seiten einer primitiven Vorproduktion nur eine „ordnende" Rolle erhalten: durch sie wird nur das bestimmt, welche primitive Vorproduktionen (nach geeigneten Einsetzungen für ihre Metazeichen) auf den betrachteten Wert des betreffenden Gliedes anwendbar sind — wodurch bestimmt wird, in welcher Anordnung die terminalen Begriffe in die terminalen Entwicklungen der Kategorienamen hineinkommen. Da es nur endlich viele Kombinationen der (endlich vielen) primitiven Vorproduktionen gibt, können

in dieser Hinsicht die Werte der Glieder der rechten Seiten der primitiven Vorproduktionen in endlich viele Mengen verteilt werden, und es ist nur ihre Zugehörigkeit zu diesen Mengen maßgebend, nicht ihre konkrete Form.

Zur Ausführung dieses Leitgedankens sei eine Reihenfolge der (endlich vielen) Metazeichen fixiert; die Folgen, welche aus dieser Folge so entstehen, daß jedes Glied durch einen seiner Werte ersetzt wird, werde ich **Stellen** nennen. Die Menge M^* der Stellen sei unendlich (nur in diesem Fall hat unser Problem einen Sinn).

Seien sämtliche Kategorienamen (die zu K gehörigen Zeichenketten):

$$\lambda_1; \lambda_2; \ldots; \lambda_e.$$

Die Werte der linken Seite oder eines Gliedes der rechten Seite einer primitiven Vorproduktion können nur durch endlich viele Ersetzungen ihrer Metazeichen durch gewisser ihrer Werte (kurz: durch endlich viele „Einsetzungen") einem Kategorienamen oder einem terminalen Begriff gleich werden. (Denn die Länge nur endlich vieler Zeichenketten übersteigt nicht die Höchstlänge der Kategorienamen und der terminalen Begriffe, und durch Einsetzen einer Zeichenkette erhält man eine mindestens so lange Kette als die Eingesetzte.) So gibt es eine natürliche Zahl q mit folgender Eigenschaft: Sei V_1 die (jedenfalls endliche) Menge aller Vorproduktionen, die aus den primitiven Vorproduktionen so entstehen, dass in diesen für je eine mögliche Kombination (je einer möglichen Klasse) der Metazeichen eine entsprechende Kombination von Werten der entsprechenden Metazeichen höchstens der Länge q eingesetzt wird (die Kombination 0-ter Klasse inbegriffen, wonach V_1 eine Erweiterung von V ist). Sei ferner $M_1^* \subseteq M^*$ die Menge derjenigen Stellen, zu welchen Werte von Metazeichen größerer Länge als q gehören. Dann erscheinen alle Kategorienamen bestimmt auch als linke Seiten gewisser zu V_1 gehörigen Vorproduktionen; und an einer Stelle aus M_1^* nehmen die von den Kategorienamen verschiedenen linken Seiten und die von den Kategorienamen und von den terminalen Begriffen verschiedenen Glieder der rechten Seiten keiner zu V_1 gehörigen Vorproduktion Kategorienamen oder terminale Begriffe als Werte an. Von nun an sollen V_1 bzw. M_1^* statt V bzw. M^* angewandt werden; so kommt dieselbe Produktionenmenge R zustande. Aus den Produktionen sollen aber nur jene beibehalten werden, die in der Bildung der terminalen Entwicklungen der Kategorienamen eine Rolle erhalten; die Menge dieser Produktionen sei R_1. Unter den „Werten" der linken Seite einer zu V_1 gehörigen Vorproduktion werde ich von nun an die linken Seiten der aus dieser Vorproduktion durch Einsetzungen entstehenden und zu R_1 gehörigen Produktionen verstehen. Für die „Werte" der Glieder der entsprechenden rechten Seiten gilt das Analoge. (Geht man von einer Zeichenkette aus, so ist unter einer aus ihr durch Einsetzung entstehenden Zeichenkette sie selbst zu verstehen.)

Nach diesen Vorbereitungen ordnen wir die zu V_1 gehörigen Vorproduktionen in eine Folge, in der zuerst die von den Kategorienamen verschiedenen linken Seiten besitzenden auftreten (sei ihre Anzahl r), dann die λ_1 als linke Seite enthaltenden (der Anzahl i_1), dann die λ_2 als linke Seite enthaltenden (der Anzahl i_2), usw., zuletzt die λ_e als linke Seite enthaltenden (der Anzahl i_e). Sei

$$r + i_1 + i_2 + \ldots + i_e = r_e.$$

Bezeichne $\gamma_{i,j}$ für $i = 1; 2; \ldots; r_e$ das j-te, von den terminalen Begriffen und von

den Kategorienamen verschiedene Glied der rechten Seite der i-ten Vorproduktion in dieser Folge, und β_i für $i=1; 2; \ldots; r$ die linke Seite der i-ten Vorproduktion. Seien die Mengen der Werte von diesen Mischketten der Reihe nach:

Die Mengen der Werte der linken Seiten:	Die Mengen der Werte der von den terminalen Begriffen und von den Kategorienamen verschiedenen Glieder der rechten Seiten:		
L_1	$G_{1,1}$	$G_{1,2}$	$\ldots\ G_{1s_1}$
L_2	$G_{2,1}$	$G_{2,2}$	$\ldots\ G_{2,s_2}$
. . . . .			
L_r	$G_{r,1}$	$G_{r,2}$	$\ldots\ G_{r,s_r}$
$\{\lambda_1\}$	$G_{r+1,1}$	$G_{r+1,2}$	$\ldots\ G_{r+1,s_{r+1}}$
. . . . .			
$\{\lambda_1\}$	$G_{r+i_1,1}$	$G_{r+i_1,2}$	$\ldots\ G_{r+i_1,s_{r+i_1}}$
. . . . .			
. . . . .			
$\{\lambda_e\}$	$G_{r_e-i_e+1,1}$	$G_{r_e-i_e+1,2}$	$\ldots\ G_{r_e-i_e+1,s_{r_e-i_e+1}}$
. . . . .			
$\{\lambda_e\}$	$G_{r_e,1}$	$G_{r_e,2}$	$\ldots\ G_{r_e,s_{r_e}}$

Da für $i=1; 2; \ldots; r_e$; $j=1; 2; \ldots; s_i$ jeder zu $G_{i,j}$ gehörige Wert von den terminalen Begriffen und von $\lambda_1; \ldots; \lambda_e$ verschieden ist, so muß jeder solche Wert in einigen der L_k ($k=1; 2; \ldots; r$) enthalten sein. Ordnen wir sämtliche Kombinationen (1-ter, 2-ter, ..., r-ter Klasse) der Zahlen $1; 2; \ldots; r$ in eine Folge

$$C_1; C_2; \ldots C_{2^r-1}$$

und sei $C_{f(k,u)}$ für $k=1; 2; \ldots; r$; $u=1; 2; \ldots; 2^{r-1}$ unter diesen die u-te solche Kombination, welche auch die Zahl k enthält. Wenn zu einer Kombination C_l ($l=1; 2; \ldots; 2^r-1$) die Zahlen $k_1; k_2; \ldots; k_v$ gehören, so bezeichne für $i=1; 2; \ldots$ $\ldots; r_e$; $j=1; 2; \ldots; s_i$

$$G_{i,j,l} \subseteq G_{i,j}$$

die Menge derjenigen zu $G_{i,j}$ gehörigen Zeichenketten, die aus den Mengen $L_1; L_2; \ldots; L_r$ in jedem der $L_{k_1}; L_{k_2}; \ldots; L_{k_v}$ und nur in diesen enthalten sind; und die Teilmenge von M_1^* der Stellen, an welchen $\gamma_{i,j}$ Werte aus $G_{i,j,l}$ annimmt, sei durch $M_{i,j,l}$ bezeichnet. (Natürlich kann $G_{i,j,l}$ und damit $M_{i,j,l}$ für gewisse l auch leer sein, doch nicht für alle.)

So gilt (da die überflüssigen Produktionen weggelassen wurden) für $k=1; 2; \ldots; r$

$$(**) \qquad L_k = \bigcup_{i=1}^{r_e} \bigcup_{j=1}^{s_i} \bigcup_{u=1}^{2^{r-1}} G_{i,j,f(k,u)}.$$

Bezeichne $N_{k,i,j,l} \subseteq M_1^*$ die Menge derjenigen Stellen, wo β_k Werte aus $G_{i,j,l}$ annimmt.

Betrachtet man also eine Produktion aus R_1, deren linke Seite zu L_k gehört, so gehört diese linke Seite zu einem der nicht leeren $G_{i,j,f(k,u)}$, und wird von

β_k an einer zu $N_{k,i,j,f(k,u)}$ gehörigen Stelle angenommen. Auf der rechten Seite der betrachteten Produktion kommen an der selben Stelle angenommenen Werte von $\gamma_{k,1}; \gamma_{k,2}; \ldots; \gamma_{k,s_k}$ vor. In welcher Kombination der Wertemengen der linken Seiten dann ein solcher Wert von $\gamma_{k,g}$ $(g=1;2;\ldots;s_k)$ auftritt, das hängt davon ab, bei welchen l die betreffende Stelle zu $M_{k,g,l}$ gehört.

Bezeichne

$$G_{k,g,l,i,j,f(k,u)} \subseteq G_{k,g,l}$$

die Menge jener Werte von $\gamma_{k,g}$, welche an Stellen aus

$$M_{k,g,l} \cap N_{k,i,j,f(k,u)}$$

angenommen werden. $G_{k,g,l}$ setzt sich aus solchen Teilmengen zusammen: es ist für $k=1;2;\ldots;r$; $g=1;2;\ldots;s_k$; $l=1;2;\ldots;2^r-1$

$$G_{k,g,l} = \bigcup_{i=1}^{r_e} \bigcup_{j=1}^{s_i} \bigcup_{u=1}^{2^{r-1}} G_{k,g,l,i,j,f(k,u)}.$$

Für $i=1;2;\ldots;r$ können auch die in (**) vorkommenden Glieder auf diese Form gebracht werden; so ergibt sich aus (**) für $k=1;2;\ldots;r$

$$L_k = \bigcup_{i=1}^{r} \bigcup_{j=1}^{s_i} \bigcup_{u=1}^{2^{r-1}} \bigcup_{i'=1}^{r_e} \bigcup_{j'=1}^{s_{i'}} \bigcup_{u'=1}^{2^{r-1}} G_{i,j,f(k,u),i',j',f(i,u')} \cup$$

$$\cup \bigcup_{i=r+1}^{r_e} \bigcup_{j=1}^{s_i} \bigcup_{u=1}^{2^{r-1}} G_{i,j,f(k,u)}.$$

Nach den vorherigen ist jede aus der k-ten zu V_1 gehörigen Vorproduktion durch Einsetzen entstehende, zu R_1 gehörige Produktion so beschaffen, daß

a) falls $1 \leqq k \leqq r$ gilt, dann die linke Seite entweder bei $1 \leqq i \leqq r$ zu einem nicht leeren $G_{i,j,f(k,u),i',j',f(i,u')}$ oder bei $r+1 \leqq i \leqq r_e$ zu einem nicht leeren $G_{i,j,f(k,u)}$ gehört, und für $g=1;2;\ldots;s_k$ das aus $\gamma_{k,g}$ entstehende Glied der rechten Seite zu einem nicht leeren $G_{k,g,l_g,i,j,f(k,u)}$ mit denselben i, j und u;

b) falls $r+i_1+\ldots+i_{t-1}+1 \leqq k \leqq r+i_1+\ldots+i_{t-1}+i_t$ bei einem $1 \leqq t \leqq e$ gilt, dann die linke Seite λ_t ist, und für $g=1;2;\ldots;s_k$ das durch Einsetzen aus $\gamma_{k,g}$ entstehende Glied der rechten Seite zu einem nicht leeren G_{k,g,l_g} gehört,

wobei beidenfalls

$$1 \leqq j \leqq s_i; \quad 1 \leqq u \leqq 2^{r-1}; \quad 1 \leqq i' \leqq r_e; \quad 1 \leqq j' \leqq s_{i'};$$

$$1 \leqq u' \leqq 2^{r-1}; \quad 1 \leqq l_g \leqq 2^r - 1$$

bestehen.

Nur die Zugehörigkeit zu den genannten Mengen und nicht die konkrete Form der als linke Seiten und als die betrachteten Glieder der rechten Seiten der Produktionen auftretenden Zeichenketten entscheidet, in welcher Reihenfolge die Produktionen angewendet werden können, und in welcher Reihenfolge die bisher als rechtsseitige Glieder ausser Acht gelassenen terminalen Begriffe und auch als rechtsseitige Glieder auftretende Kategorienamen dadurch in die Entwicklungen hineinkommen; und allein dies ist maßgebend dafür, welche Zeichenkettenlisten als terminale Entwicklungen der Kategorienamen zustandekommen, da aus den terminalen Entwicklungen schon sämtliche Werte der $\gamma_{i,j}$ verschwinden müssen.

Daher können alle Zeichenketten, die zu je einem der unter a) und b) aufgezählten, mit gewissen Indizes versehenen nicht leeren G-Mengen gehören, durch je eine einzige solche Zeichenkette vertreten werden, und die gewählten Repräsentanten können genau so bezeichnet werden, wie die durch sie repräsentierten Mengen. Diese Bezeichnungen als Elemente eines Hilfsvokabulars aufgefasst, zu welchem auch $\lambda_1; \ldots; \lambda_e$ gehören, besitzen wir — in den aus den unter a) und b) beschriebenen Produktionen durch die genannten Repräsentanten Ausgewählten — bereits endlich viele Produktionen, welche sämtliche terminale Entwicklungen der Kategorienamen generieren.

4. Ein Beispiel

Die in der allgemeinen Behandlung vielleicht nicht genügend übersichtliche Angabe der endlich vielen Produktionen, durch welche eine zweistufig definierte, aber nur endlich viele terminalen Begriffe besitzende Sprache generiert werden kann, soll nun durch ein verhältnissmässig einfaches Beispiel klargestellt werden, in welchem die Mengen, deren Existenz im allgemeinen Fall benutzt wurde, effektiv konstruiert werden können.

Die Sprache sei durch den Mengenquintupel

$$(Z; M; P; V; K)$$

angegeben, wobei

$$Z = \{z_1; z_2; z_3; t_1; t_2\}$$

$$M = \{m\}$$

$$P = \{m: z_1 m z_1; m: z_2\}$$

$$V = \{m: z_1 m z_1, t_1; z_1 m z_1: t_2; z_3: m\}$$

$$K = \{z_3\}$$

ist. Hier ist M und P dasselbe wie im ersten Beispiel des Kapitels **2,** woraus sich für sämtliche terminalen Entwicklungen von m

$$\underbrace{z_1 z_1 \ldots z_1}_{n\text{-mal}}\, z_2\, \underbrace{z_1 z_1 \ldots z_1}_{n\text{-mal}} \qquad (n = 0; 1; 2; \ldots)$$

ergibt. Diese Zeichenketten werde ich kurz durch a_n bezeichnen. So kann die Stellenmenge M^* kurz durch

$$M^* = \{a_n\}$$

bezeichnet werden, wo dies und ähnliche Bezeichnungen immer so zu verstehen sind, dass die Elemente der betreffenden Menge durch Einsetzen sämtlicher nichtnegativen Zahlen für n entstehen.

Da

$$z_1 m z_1 = a_{n+1}$$

ist, so ergeben sich durch Einsetzungen aus V die folgenden Produktionen für $n=0; 1; 2; \ldots$

$$(***) \qquad \begin{aligned} a_n &: a_{n+1}, t_1 \\ a_{n+1} &: t_2 \\ z_3 &: a_n. \end{aligned}$$

(Hier entstehen weder die beiden terminalen Begriffe $t_1; t_2$ noch der einzige Kategoriename z_3 durch Einsetzung aus einer von ihnen verschiedenen rechtsseitigen Glied oder linken Seite der Elemente von V, daher bleibt hier der Übergang von M^* bzw. V auf M_1^* bzw. V_1 weg.)

Da hier die $\gamma_{i,j}$ nur mit dem zweiten Index 1 auftreten, können diese „Kolumnenindizes" überall, wo sie eine Rolle erhalten würden, weggelassen werden. So ist hier

$$\beta_1 = m \qquad \gamma_1 = z_1 m z_1$$

$$\beta_2 = z_1 m z_1$$

$$\gamma_3 = m.$$

Es gibt ein Wert von β_1 (bei $m=a_0$), die von β_2 nicht angenommen wird, doch alle Werte von β_2 werden auch von β_1 angenommen. Daher kommen hier nur die folgenden Kombinationen der linksseitigen Indizes in Frage:

$$C_1 = 1; \quad C_2 = 1\,2$$

und für $f(k, u)$ (für den Index jener Kombination, in welcher die Zahl k zum u-tenmal auftritt) nur die folgenden Möglichkeiten:

$$1 = f(1, 1); \quad 2 = f(1, 2) = f(2, 1).$$

Da $G_{i,l}$ die Menge jener Werte von γ_i bedeutet, die genau von jenen β_j angenommen werden, für welche j zur Kombination C_l gehört, sind die folgenden Mengen $G_{i,l}$ nicht leer:

$$G_{1,2} = \{a_{n+1}\}; \quad G_{3,1} = \{a_0\}; \quad G_{3,2} = \{a_{n+1}\}.$$

Die Menge der Stellen, an welchen die Elemente von $G_{1,2}$ von γ_1 angenommen werden, ist die ganze Stellenmenge M^*:

$$M_{1,2} = \{a_n\}.$$

Da $N_{1,i,l}$ die Menge jener Stellen bedeutet, wo β_1 die zu $G_{i,l}$ gehörigen Werte annimmt (was nur für $l=f(1, 1)$ oder $l=f(1, 2)$ der Fall sein kann), sind nur die folgenden Mengen $N_{1,i,l}$ nicht leer:

$$N_{1,1,2} = N_{1,1,f(1,2)} = \{a_{n+1}\}$$

$$N_{1,3,1} = N_{1,3,f(1,1)} = \{a_0\}$$

$$N_{1,3,2} = N_{1,3,f(1,2)} = \{a_{n+1}\}.$$

Wir haben die gemeinsamen Teile der $M_{1,l}$ und $N_{1,i,f(1,u)}$ zu bilden. Die folgenden von diesen sind nicht leer:

$$M_{1,2} \cap N_{1,1,f(1,2)} = \{a_{n+1}\}$$
$$M_{1,2} \cap N_{1,3,f(1,1)} = \{a_0\}$$
$$M_{1,2} \cap N_{1,3,f(1,2)} = \{a_{n+1}\}.$$

Es sind endlich die Mengen $G_{1,l,i,f(1,u)}$ der an den Stellen $M_{1,l} \cap N_{1,i,f(1,u)}$ angenommenen Werte von γ_1 zu bestimmen. Diese sind (die übereinstimmenden durch einen gemeinsamen Zeichen bezeichnet):

$$H_1 = G_{1,2,1,f(1,2)} = G_{1,f(1,2),1,f(1,2)} = G_{1,f(2,1),1,f(1,2)} =$$
$$= G_{1,2,3,f(1,2)} = G_{1,f(1,2),3,f(1,2)} = G_{1,f(2,1),3,f(1,2)} =$$
$$= \{a_{n+2}\}$$

und

$$H_2 = G_{1,2,3,f(1,1)} = G_{1,f(1,2),3,f(1,1)} = G_{1,f(2,1),3,f(1,1)} =$$
$$= \{a_1\}.$$

Ferner sei noch

$$H_3 = G_{3,1} = G_{3,f(1,1)} = \{a_0\}$$

und

$$H_4 = G_{3,2} = G_{3,f(1,2)} = G_{3,f(2,1)} = \{a_{n+1}\}.$$

Nach a) und b) des Kapitels **3** kann die betrachtete Sprache (mit Werten von u; i'; u'; l, bei welchen die entsprechenden Mengen nicht leer sind) durch Produktionen folgender Form generiert werden:

$$G_{1,f(1,u),i',f(1,u')} : G_{1,l,1,f(1,u)},\ t_1$$
$$G_{3,f(1,u)} : G_{1,l,3,f(1,u)},\ t_1$$
$$G_{1,f(2,u),i',f(1,u')} : t_2$$
$$G_{3,f(2,u)} : t_2$$
$$z_3 : G_{3,l}$$

wobei die mit Indizes versehenen Zeichen G nicht mehr als Mengenbezeichnungen, sondern als Hilfszeichen zu betrachten sind; die bisher als Zeichen übereinstimmender Mengen gegoltenen unter diesen können aber auch durch die für sie eingeführten gemeinsamen Zeichen vertreten werden. So ergeben sich die folgenden neun Produktionen:

1. $H_1 : H_1, t_1$ **5.** $H_1 : t_2$ 8. $z_3 : H_3$

2. $H_2 : H_1, t_1$ 6. $H_2 : t_2$ **9.** $z_3 : H_4$

3. $H_3 : H_2, t_1$ **7.** $H_4 : t_2$

4. $H_4 : H_1, t_1$

Die terminalen Entwicklungen von z_3 sind Listen mit Gliedern t_1; t_2. Man kann von den erhaltenen Produktionen ablesen, welche Listen solcher Art zu diesen

Entwicklungen gehören. Bei der Bildung der Entwicklungen hat man von einem der beiden letzten Produktionen auszugehen. Durch Anwendung von **7** auf die rechte Seite von **9** erscheint t_2 als einziges Glied einer terminalen Entwicklung; die Bildung (von rechts nach links) einer terminalen Entwicklung nimmt auch sonst dann und nur dann ein Ende, sobald ein Glied t_2 erscheint. Durch Anwendung von **4** auf **9** tritt t_1 als letztes Glied einer terminalen Entwicklung auf; und durch wiederholter Anwendung von **1** erscheinen dann von rechts nach links fortschreitend Glieder t_1 beliebiger Anzahl. Zum Abschluss durch ein Glied t_2 kommt man durch Anwendung von **5.** So kann man sich auf die fünf Produktionen

1; 4; 5; 7; 9

beschränken; die übrigen vier können weggelassen werden.

Geht man von der inzwischen klargewordenen Tatsache aus, daß sämtliche terminale Entwicklungen von z_3 mit t_2 beginnen, worauf Glieder t_1 beliebiger (auch 0-ter) Anzahl folgen, so sieht man leicht, daß die betrachtete Sprache auch durch die folgenden zwei Produktionen generiert werden kann:

$$z_3 : z_3, t_1$$

$$z_3 : t_2.$$

Dies hätte man in diesem einfachen Beispiel bereits von den unter (***) angegebenen unendlich vielen Produktionen ablesen können; doch das Zweck dieses Beispiels war die allgemein eingeführten Begriffe klar zu stellen.

Eötvös Loránd Universität, Budapest

(Eingegangen: 29. März, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 195—200.

ON THE PERMEABILITY OF A LAYER OF PARALLELOGRAMS

by

L. FEJES TÓTH

A set of non-overlapping, nowhere accumulating open convex discs contained in a parallel strip is said to form a *layer*. Let w be the width of the strip and l the length of a path connecting one edge of the strip with the other in such a way as to evade all discs. The *permeability* p of the layer is defined by

$$p = w/\inf l,$$

where the infimum extends over all paths of the above kind.

We remind of the following results [1]: The permeability of a layer of congruent circles is always greater than $\sqrt{27}/2\pi$, but there is a layer of incongruent circles with a permeability less than $\sqrt{27}/2\pi$.

In this paper we want to point out that for layers consisting of squares a similar proposition does not hold. We shall show that the infimum of the permeabilities of all layers of congruent homothetic squares is equal to 2/3 and that this constant is simultaneously the infimum of the permeabilities of all layers of squares of any size and orientation.

We shall prove a little more. We start with the following

THEOREM 1. *Let Π be a parallelogram and p the permeability of a layer of parallelograms similar to Π. Then there is a layer of translated replicas of Π with a permeability less than p.*

PROOF. We name X a boundary point of Π such that the halfline issuing from X vertically downwards intersects Π. We call Y a boundary point of Π other than a point X. Let λ be the length of the shortest path joining a point X to a point Y through boundary points of Π. Let δ be the level difference between X and Y, i.e. the distance between the horizontal lines through X and Y.

Write $Q(X, Y) = \lambda/\delta$ and consider $\min_Y Q(X, Y)$ for a fixed point X when Y ranges over all points defined above. If all sides of Π have the same slope ω (angle included with a horizontal line), we have

$$\min_Y Q(X, Y) = \csc \omega.$$

In the opposite case the minimum is attained for one of the lower extremities of the two sides with the greater slope or for both.

Now consider

$$q = \max_X \min_Y Q(X, Y),$$

where X ranges over all points defined above. If each side of Π has the same slope ω, we have $q=\csc\omega$. Otherwise the maximum is attained for a point X lying on the side with the smaller slope in such a way that the minimum is attained for two positions of Y.

Finally, consider the maximum

$$\bar{q}=\max_{\Pi} q$$

of q for all orientations of the rigidly moving parallelogram Π. Since a rotation of Π through a small angle implies a small variation of q, this maximum also exists.

We claim that the permeability of a layer of parallelograms similar to Π is $>1/\bar{q}$. To see this we shall construct a path of length $l<\bar{q}w$ crossing the layer.

We suppose the layer to lie in a vertical plane in a horizontal position. We start from a point U of the upper edge of the strip and go vertically downwards until we arrive to a boundary point X of a parallelogram Π. The point X may coincide with U. Now we proceed along the boundary of Π so as to get to a point Y which minimizes $Q(X, Y)$. We choose Y as our new starting point and proceed in the same way as above, and so on, until we arrive to a point L of the lower edge of the strip. The path constructed in this way consists of a finite number of portions of length $\lambda\leqq\bar{q}\delta$, where δ is the level difference between the extremities of the respective portion. We choose U so that for the first portion we have $\lambda<\bar{q}\delta$. If the first portion is a vertical segment UX, we automatically have $\lambda=\delta<\bar{q}\delta$. If, on the other hand, $U\equiv X$ for any choice of U, so that the first portion is XY, we choose U so as to be a boundary point of Π other than the point X at which $\min_{Y} Q(X, Y)$ attains its maximum. Summing up the inequalities $\lambda\leqq\bar{q}\delta$, we obtain the inequality $l<\bar{q}w$ for the length l of the path UL, as stated.

We continue to construct a layer of translated replicas of Π with a permeability arbitrarily close to $1/\bar{q}$. Let $\overline{\Pi}$ be the parallelogram which yields the maximum $\bar{q}$ of q among all parallelograms congruent to Π. Let $\overline{X}$ be the boundary point of $\overline{\Pi}$ for which $\min_{Y} Q(X, Y)$ attains its maximum. Finally, let $\overline{Y}_1$ and $\overline{Y}_2$ be the two points at which $\min_{Y} Q(\overline{X}, Y)$ is attained, i.e. the two lower vertices of $\overline{\Pi}$.

Consider the lattice generated from $\overline{\Pi}$ by the vectors $\overrightarrow{\overline{X}\overline{Y}_1}$ and $\overrightarrow{\overline{X}\overline{Y}_2}$. Let Λ be the layer consisting of those lattice-parallelograms which are contained in a horizontal strip of width w. A path crossing Λ is divided by the points homologous in the lattice to $\overline{X}$ (Fig. 1) into, say, n portions. Let $\lambda_1, \ldots, \lambda_n$ be the lengths of these portions and $\delta_1, \ldots, \delta_n$ the respective level differences. Observing that

$$\lambda_i\leqq\bar{q}\delta_i, \qquad i=2, \ldots, n-1,$$

we obtain for the total length l of the path the inequality

$$l\geqq\bar{q}w-(\bar{q}-1)(\delta_1+\delta_n).$$

On the other hand, we have $\delta_1<b$ and $\delta_n<b$, where b is the breadth of $\overline{\Pi}$ in the vertical direction. Thus

$$l>\bar{q}w-2(\bar{q}-1)b,$$

whence

$$\frac{w}{l} < \frac{1}{\bar{q} - \dfrac{2(\bar{q}-1)b}{w}},$$

showing that for sufficiently big values of w the permeability of Λ will get arbitrarily close to $1/\bar{q}$. This completes the proof of Theorem 1.

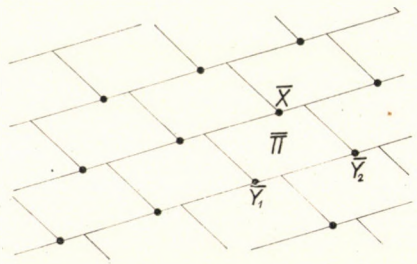


Fig. 1

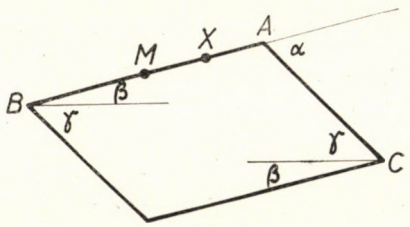


Fig. 2

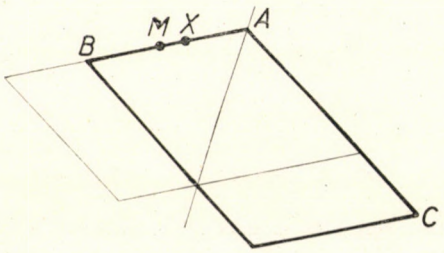


Fig. 3

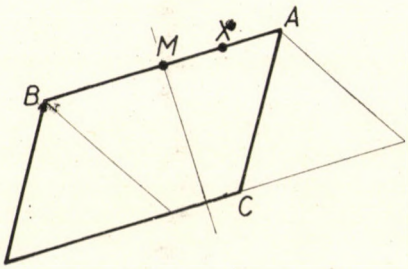


Fig. 4

We proceed to give some hints to compute $\bar{q}$. Let A be the highest vertex of Π, B and C the adjacent vertices of Π, β and γ the slopes of the sides AB and AC and $\alpha = \beta + \gamma$ the outer angle of Π at A (Fig. 2). Suppose that $\beta \leqq \gamma$. Then the point X may be supposed to lie on the segment AM, where M is the point halfway between A and B. We also may suppose that $\overline{AB} \geqq \overline{AC}$. Otherwise reflect Π in the line bisecting the inner angle at A, obtaining a parallelogram which has a greater value of $\min_Y Q(X, Y)$ than Π (Fig. 3). Again, we may suppose that $\alpha \leqq \pi/2$. Otherwise reflect Π in the orthogonal bisector of the side AB (Fig. 4), once more increasing by this operation $\min_Y Q(X, Y)$. Write $2s = \overline{AB} \geqq \overline{AC} = 1$ and $q = q(\beta)$. Since

$$\bar{q} = \max q(\beta) \geqq q(0) = \frac{s+1}{\sin \alpha},$$

we, finally, may suppose that

$$\csc \beta \geqq (s+1) \csc \alpha.$$

The position of X at which $\max_X \min_Y Q(X, Y)$ is attained is given by the equations

$$\frac{s+1+x}{\sin(\alpha-\beta)+(s+x)\sin\beta} = \frac{s+1-x}{\sin(\alpha-\beta)-(s-x)\sin\beta} = q,$$

where $x = \overline{MX}$. Eliminating x, we obtain

$$q^2 s \sin^2\beta + q\sin(\alpha-\beta) - s - 1 = 0.$$

Thus *for a parallelogram with a ratio of the sides* $2s:1 \geqq 1$ *and a non-obtuse angle* α *the value of* q *is given by the maximum of the function* $q(\beta)$ *defined for*

$$0 \leqq \beta \leqq \min\left(\arcsin\frac{\sin\alpha}{s+1}, \frac{\alpha}{2}\right)$$

by

$$q(\beta) = \begin{cases} \dfrac{s+1}{\sin\alpha}, & \beta = 0 \\ \dfrac{-\sin(\alpha-\beta)+[\sin^2(\alpha-\beta)+4s(s+1)\sin^2\beta]^{\frac{1}{2}}}{2s\sin^2\beta}, & \beta > 0. \end{cases}$$

It will be interesting to consider some special cases.

We claim that for a rectangle $\bar{q} = q(0) = s+1$. Thus in order to construct a horizontal layer with a small permeability we must arrange the rectangles like the bricks in the wall (Fig. 5). We shall prove this by showing that

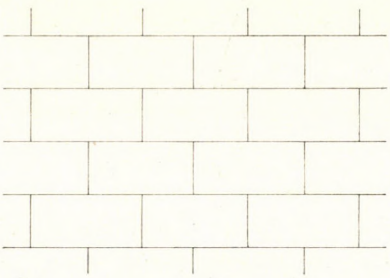


Fig. 5

$$\frac{-\cos\beta+[\cos^2\beta+4s(s+1)\sin^2\beta]^{\frac{1}{2}}}{2s\sin^2\beta} < s+1,$$

$$0 \leqq \beta \leqq \pi/4.$$

Transform this inequality as follows:

$$\cos^2\beta+4s(s+1)\sin^2\beta < [\cos\beta+2s(s+1)\sin^2\beta]^2,$$

$$4s(s+1)\sin^2\beta < $$

$$< 4s(s+1)\cos\beta\sin^2\beta + 4s^2(s+1)^2\sin^2\beta,$$

$$1 < \cos\beta + s(s+1)\sin^2\beta,$$

$$1-\cos\beta < s(s+1)(1-\cos\beta)(1+\cos\beta),$$

$$1 < s(s+1)(1+\cos\beta).$$

Since $s \geqq 1/2$, the last inequality is certainly satisfied if $\cos\beta > 1/3$, which is for $0 < \beta < \pi/4$ obviously true.

Combining this result with the proof of Theorem 1, we obtain

THEOREM 2. *The lower bound of the permeabilities of all layers of rectangles with a side-ratio not exceeding* r *is equal to* $2/(2+r)$.

Now we consider a rhombus $(2s=1)$ with an angle $\alpha \leqq \alpha_0 = \arccos 1/3 \approx$ $\approx 70°31'44''$. Note that α_0 equals the dihedral angle of a regular tetrahedron. We shall show that $\bar{q} = q(\alpha/2) = \csc\alpha/2$, which means that the rhombi will put up the greatest resistance in the direction of the shorter diagonal (Fig. 6).

By a suitable transformation of the inequality

$$\frac{-\sin(\alpha-\beta)+[\sin^2(\alpha-\beta)+3\sin^2\beta]^{\frac{1}{2}}}{\sin^2\beta} \leqq \csc\alpha/2,$$

$$0 \leqq \beta \leqq \alpha/2 \leqq \alpha_0/2$$

we obtain

$$G(\alpha, \beta) =$$

$$= \sin^2\beta + 2\sin\frac{\alpha}{2}\sin(\alpha-\beta) - 3\sin^2\frac{\alpha}{2} \geqq 0.$$

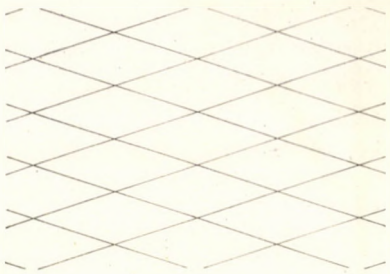


Fig. 6

Since $G(\alpha, \alpha/2)=0$, the inequality will be proved by showing that

$$f(\beta) = \sin^2\beta + 2\sin\frac{\alpha}{2}\sin(\alpha-\beta),\ 0 \leqq \beta \leqq \frac{\alpha}{2}$$

is a decreasing function. In order to prove that

$$f'(\beta) = \sin 2\beta - 2\sin\frac{\alpha}{2}\cos(\alpha-\beta) \leqq 0,$$

observe that $f'(\alpha/2)=0$ and

$$g(\beta) = \frac{1}{2}f''(\beta) = \cos 2\beta - \sin\frac{\alpha}{2}\sin(\alpha-\beta) \geqq 0.$$

The last inequality follows from $g(0)>0$, $g\left(\frac{\alpha}{2}\right) = 1-3\sin^2\frac{\alpha}{2} \geqq 0$ and

$$g''(\beta) = -4\cos 2\beta + \sin\frac{\alpha}{2}\sin(\alpha-\beta) < -4\cos\alpha_0 + 1 = -\frac{1}{3} < 0.$$

We recapitulate the result obtained in

THEOREM 3. *The lower bound of the permeabilities of all layers of rhombi having an angle not greater than* arc cos 1/3 *and not less than* α *equals* $\sin\alpha/2$.

We conclude with two remarks.

REMARK 1. Apart from a rectangle, $\overline{\Pi}$ has no horizontal side.

Otherwise we would have, for some $\alpha<\pi/2$ and all values of β such that $0 \leqq \beta \leqq \leqq \min\left(\text{arc}\sin\frac{\sin\alpha}{s+1}, \frac{\alpha}{2}\right)$, $q(\beta) \leqq q(0)$. But this inequality is equivalent with $h(\beta) = s(s+1)\sin^2\beta + \sin\alpha\sin(\alpha-\beta) - \sin^2\alpha \geqq 0$, which, because of $h(0)=0$ and $h'(0) = -\frac{1}{2}\sin 2\alpha < 0$, is not true.

REMARK 2. Apart from a rhombus with an angle $\leqq$ arc cos 1/3, adjacent sides of $\overline{\Pi}$ have different slopes.

First we consider a rhombus with an acute angle $\alpha>\alpha_0$. Using the above notations, we have

$$\tfrac{1}{2}f''(\alpha/2) = 1-3\sin^2\alpha/2 < 0,$$

showing that at $\beta=\alpha/2$ the function $f'(\beta)$ decreases. Since $f'(\alpha/2)=0$, $f'(\beta)$ assumes positive values at the neighbourhood of $\beta=\alpha/2$. Therefore $f(\beta)$ increases here, which, in view of $G(\alpha, \alpha/2)=0$, implies that near to $\beta=\alpha/2$ we have

$$G(\alpha, \beta) = f(\beta) - 3\sin^2 \alpha/2 < 0,$$

i.e. $q(\beta) > q(\alpha/2)$.

We still have to deal with the case when $\beta=\gamma=\alpha/2$ and $\overline{AB} > \overline{AC}$. Let Z be the vertex of Π opposite to A. Choosing X to be on the side AB due above Z, we have $Q(X, C) > Q(X, Z) = \csc \alpha/2$. Turn Π continuously about X in a direction such as to decrease $Q(X, C)$. Since $Q(X, Z)$ increases by a rotation in both directions, there will be a position of Π such that $Q(X, C) = Q(X, Z) > \csc \alpha/2$, showing that originally Π was not in an extremal position.

REFERENCE

[1] FEJES TÓTH, L.: On the permeability of a circle-layer, *Studia Sci. Math. Hungar.* **1** (1966) 5—10.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received March 31, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 201—209.

RATIONAL APPROXIMATION ON THE WHOLE REAL AXIS

by

G. FREUD and J. SZABADOS

§ 1. Introduction

In the last three years the theory of rational approximation showed a great development. A lot of results were published concerning finite intervals. But there is a possibility of investigations on the whole real axis. Namely, as one of us proved [1], *there exist rational functions $Q_n(x)$ of degree n at most such that*

$$(1.1) \qquad ||x| - Q_n(x)| \leqq \frac{3}{2}(1+x^2)e^{-\sqrt{\frac{n-3}{2}}} \qquad (-\infty < x < +\infty,\ n = 4, 5, \ldots).$$

Our results will be based mostly on this result. We shall frequently use the ideas of [2] also.

§ 2. Estimations by Certain Modules of Continuity

As ČEBYŠEV proved, if $f(x)$ is continuous on the whole real axis and the limits

$$(2.1) \qquad \lim_{x\to-\infty} f(x), \quad \lim_{x\to+\infty} f(x)$$

exist, are equal and finite, then $f(x)$ can be uniformly approximated by rational functions on the whole real axis. But he did not deal with the question of rapidity of this approximation. The convergence depends on the structural properties of $f(x)$. By ACHIESER's method (see [3], p. 76) we can easily prove the following

THEOREM 1. *If the limits* (2.1) *exist, are equal and finite then there exist rational functions $R_n(x)$ of degree n at most such that*

$$(2.2) \qquad |f(x) - R_n(x)| \leqq 48\omega\left(\frac{1}{n}\right) \qquad (-\infty < x < +\infty,\ n = 1, 2, 3, \ldots)$$

where $\omega(\delta)$ is the module of continuity of the function $f\left(\operatorname{tg}\frac{t}{2}\right)$ on the interval $[-\pi, +\pi]$.

PROOF. By assumption, $f\left(\operatorname{tg}\frac{t}{2}\right)$ is a continuous, 2π-periodical function. Therefore, by the classical Jackson's theorem, there exist trigonometrical polynomials $T_{\left[\frac{n}{2}\right]}(t)$ of degree $\left[\frac{n}{2}\right]$ at most such that

$$\left|f\left(\operatorname{tg}\frac{t}{2}\right) - T_{\left[\frac{n}{2}\right]}(t)\right| \leqq 12\omega\left(\frac{1}{\left[\frac{n}{2}\right]}\right) \leqq 48\omega\left(\frac{1}{n}\right) \qquad (-\infty < t < +\infty,\ n = 2, 3, \ldots),$$

Substituting

$$\operatorname{tg}\frac{t}{2}=x, \quad \sin t=\frac{2x}{1+x^2}, \qquad \cos t=\frac{1-x^2}{1+x^2}$$

we have (2. 2), where $R_n(x)$ is a rational function with denominator $(1+x^2)^{\left[\frac{n}{2}\right]}$, qu. e. d.

THEOREM 2. *Let $f(x)$ be continuous and of bounded variation on the whole real axis. Then there exist rational functions $R_N(x)$ of degree N at most such that*

$$(2.3)\quad |f(x)-R_N(x)| = (1+x^2)\,O\left(\omega\left(\delta_{\left[\frac{N}{3}\right]}\right)\right) \qquad (-\infty < x < +\infty,\ N=0,1,\ldots)$$

where $\omega(\delta)$ is the module of continuity of the function $f\left(\frac{x}{1-|x|}\right)$ on the interval $[-1, +1]$ and δ_n is the unique solution of the equation

$$(2.4)\qquad \frac{4\log^2\delta_n}{\omega(\delta_n)}=n \qquad (n=0,1,2,\ldots).$$

REMARKS. Being $f(x)$ of bounded variation, the finite limits (2.1) must exist and therefore $f\left(\frac{x}{1-|x|}\right)$ is continuous in $[-1, +1]$. (2. 3) is generally a better estimation than (2. 2) (see [2], Satz 1) apart from the weight-function $1+x^2$. Of course, if the limits (2.1) are different then certain weight-functions must occur in the estimations.

PROOF. We may assume that the total variation of $f(x)$ is $\leqq 1$. Let N be sufficiently large and $n=\left[\frac{N}{3}\right]$. According to Satz 1 in [2], there exist rational functions

$$(2.5)\qquad r_n^{\{i\}}(x)=\sum_{j=1}^{s_i} a_{ij}\bar{R}_m(x-\xi_{ij})+A_i, \quad r_n^{\{i\}}(0)=f(0) \qquad (i=1,2)$$

of degree n at most such that

$$(2.6)\qquad \max_{-1\leqq x\leqq 0}\left|f\left(\frac{x}{1+x}\right)-r_n^{\{1\}}(x)\right| = O(\omega(\delta_n)), \quad \max_{0\leqq x\leqq +1}\left|f\left(\frac{x}{1-x}\right)-r_n^{\{2\}}(x)\right| = O(\omega(\delta_n))$$

where

$$(2.7)\qquad s_i\leqq\frac{1}{\omega(\delta_n)}, \quad |a_{ij}|\leqq\frac{2\omega(\delta_n)}{\delta_n} \qquad (j=1,2,\ldots,s_i;\ i=1,2),$$

$$(2.8)\qquad -1=\xi_{11}<\xi_{12}<\ldots<\xi_{1s_1}<0=\xi_{21}<\xi_{22}<\ldots<\xi_{2s_2}<+1,$$

$$(2.9)\qquad m=4\,[\log^2\delta_n]$$

and A_i $(i=1, 2)$ are constants,

$$\bar{R}_m(x) = x\frac{\prod_{k=0}^{m-1}\left(x+e^{-\frac{k}{\sqrt{m}}}\right)-\prod_{k=0}^{m-1}\left(-x+e^{-\frac{k}{\sqrt{m}}}\right)}{\prod_{k=0}^{m-1}\left(x+e^{-\frac{k}{\sqrt{m}}}\right)+\prod_{k=0}^{m-1}\left(-x+e^{-\frac{k}{\sqrt{m}}}\right)} \tag{2.10}$$

is the NEWMAN's rational function (see [4]). Being m an even number, we have by (2.10) and (2. 9)

$$|\bar{R}_m(x)| \leqq x\frac{x^m e^{\frac{2\sqrt{m}}{x}}-x^m e^{-\frac{\sqrt{m}}{x}}}{x^m e^{\frac{\sqrt{m}}{4x}}} = xe^{\frac{7\sqrt{m}}{4x}}\left(1-e^{-\frac{3\sqrt{m}}{x}}\right) \leqq e^{\frac{7\sqrt{m}}{4}}\cdot 3\sqrt{m} \leqq$$

$$\leqq \frac{1}{\delta_n^{7/2}}\cdot 6|\log \delta_n| \leqq \frac{6}{\delta_n^4} \qquad (x \geqq 1).$$

But $|\bar{R}_m(x)| = O(1)$ for $|x| \leqq 1$ (see [4]), and $\bar{R}_m(-x) = \bar{R}_m(x)$ for all real x. Therefore

$$|\bar{R}_m(x)| = O(\delta_n^{-4}) \qquad (-\infty < x < +\infty).$$

Hence and from (2. 5), (2. 7) we get

$$|r_n^{\{i\}}(x)| \leqq \frac{1}{\omega(\delta_n)}\cdot\frac{2\omega(\delta_n)}{\delta_n}O(\delta_n^{-4})+$$
$$+|A_i| = O(\delta_n^{-5}) \qquad (i=1,2;\ -\infty < x < +\infty). \tag{2.11}$$

P. SZÜSZ and P. TURÁN [5] showed that the poles of $\bar{R}_m(x)$ are pure imaginary numbers $\pm b_k i$ $(b_k > 0)$ satisfying the equations

$$\sum_{\nu=0}^{m-1} \operatorname{arc\,tg}\left(b_k e^{\frac{\nu}{\sqrt{m}}}\right) = \frac{(2k-1)\pi}{2} \qquad \left(k=1,2,\ldots,\frac{m}{2}\right).$$

From this easy to see that

$$\frac{\pi}{4m}e^{-\sqrt{m}} \leqq b_k \leqq m \qquad \left(k=1,2,\ldots,\frac{m}{2}\right). \tag{2.12}$$

The poles of $\bar{R}_m(x-\xi_{1j})$ are $\xi_{1j}\pm b_k i$, thus the poles of $r_n^{\{1\}}\left(\frac{x}{1-x}\right)$ are (see (2. 5))

$$\frac{\xi_{1j}+\xi_{1j}^2+b_k^2 \pm b_k i}{(1+\xi_{1j})^2+b_k^2}.$$

But by (2. 8) $-1 \leqq \xi_{1j} < 0$, and we can see by (2.12) and (2. 9) that the poles of $r_n^{\{1\}}\left(\frac{x}{1-x}\right)$ are not nearer to the real axis than

$$\frac{|b_k|}{(1+\xi_{1j})^2+b_k^2} \geqq \frac{|b_k|}{4+b_k^2} \geqq \frac{\pi}{8m^3}e^{-\sqrt{m}} \geqq \delta_n^8$$

for sufficiently large n's. Applying Lemma 1 from [6], we have by (2. 11)

$$\left|\frac{d}{dx} r_n^{\{1\}}\left(\frac{x}{1-x}\right)\right| \leqq \frac{2nO(\delta_n^{-5})}{\delta_n^8} = O(n\delta_n^{-13}) \qquad (-\infty < x < +\infty) \tag{2.13}$$

and analogously

$$\left|\frac{d}{dx} r_n^{\{2\}}\left(\frac{x}{1+x}\right)\right| = O(n\delta_n^{-13}) \qquad (-\infty < x < +\infty). \tag{2.14}$$

Now let

$$T(x) = \frac{r_n^{\{1\}}\left(\frac{x}{1-x}\right) + r_n^{\{2\}}\left(\frac{x}{1+x}\right)}{2} - \frac{|x|}{2x}\left[r_n^{\{1\}}\left(\frac{x}{1-x}\right) - r_n^{\{2\}}\left(\frac{x}{1+x}\right)\right] \tag{2.15}$$

then

$$T(x) = \begin{cases} r_n^{\{1\}}\left(\frac{x}{1-x}\right) & \text{if} \quad -\infty < x \leqq 0 \\ r_n^{\{2\}}\left(\frac{x}{1+x}\right) & \text{if} \quad 0 \leqq x < +\infty \end{cases}$$

and so by (2. 6)

$$|f(x) - T(x)| = O(\omega(\delta_n)) \qquad (-\infty < x < +\infty). \tag{2.16}$$

Further let

$$R_N(x) = \frac{r_n^{\{1\}}\left(\frac{x}{1-x}\right) + r_n^{\{2\}}\left(\frac{x}{1+x}\right)}{2} - \frac{Q_{N-2n}(x)}{2x}\left[r_n^{\{1\}}\left(\frac{x}{1-x}\right) - r_n^{\{2\}}\left(\frac{x}{1+x}\right)\right]$$

a rational function of degree N at most. We get by (2.15), (1.1), (2.5), (2.13) and (2. 14)

$$\begin{aligned} |T(x) - R_N(x)| &\leqq \frac{3}{4}(1+x^2)e^{-\sqrt{\frac{N-2n-3}{2}}}\left[\left|\frac{r_n^{\{1\}}\left(\frac{x}{1-x}\right) - r_n^{\{1\}}(0)}{x}\right| + \right. \\ &\left. + \left|\frac{r_n^{\{2\}}\left(\frac{x}{1+x}\right) - r_n^{\{2\}}(0)}{x}\right|\right] = (1+x^2)O(n\delta_n^{-13})e^{-\sqrt{\frac{N-2n-3}{2}}} = \\ &= (1+x^2)O(\omega(\delta_n)) \qquad (-\infty < x < +\infty), \end{aligned} \tag{2.17}$$

namely, by (2. 4) and $\delta_n = O(\omega(\delta_n))$ we have for sufficiently large n's the inequalities

$$N \geqq 3n > 2{,}5n + 3 + 2\log^2 \frac{\delta_n^{14}}{n} > 2n + 3 + 2\log^2 \frac{\delta_n^{13}\omega(\delta_n)}{n}.$$

From (2.16) and (2.17) we have our theorem.

§ 3. A Localization Theorem

Let $\varphi(x)$ be continuous in a finite interval $[a, b]$, and $p_n(x)$ its best approximating polynomial of degree n at most. We shall use the notation

$$E_n(\varphi(x); a, b) = \max_{a \leqq x \leqq b} |\varphi(x) - p_n(x)|.$$

THEOREM 3. *Assume that $f(x)$ is continuous on the whole real axis and the finite limits* (2.1) *exist. Further let*

$$-\infty < \xi_1 < \xi_2 < \ldots < \xi_s < +\infty, \quad \xi_1 < 0 < \xi_s \qquad (s \geqq 2),$$

$$\varepsilon_n = \max\left[\max_{1 \leqq i \leqq s-1} E_n(f(x); \xi_i, \xi_{i+1}); E_n\left(f\left(\frac{1}{x}\right); \frac{1}{\xi_1}, 0\right); E_n\left(f\left(\frac{1}{x}\right); 0, \frac{1}{\xi_s}\right)\right].$$

Then there exist rational functions $R_N(x)$ of degree N at most such that

$$|f(x) - R_N(x)| = O\left(\varepsilon_n + (1+x^2)e^{-\sqrt{\frac{n}{12}}}\right) \qquad \left(-\infty < x < +\infty;\ n = 2\left[\frac{N}{6s+4}\right]\right).$$

PROOF. By assumption, there exist polynomials $p_n^{\{i\}}(x)$ of degree n at most such that

$$\max_{\xi_i \leqq x \leqq \xi_{i+1}} |f(x) - p_n^{\{i\}}(x)| \leqq \varepsilon_n \qquad (i = 1, \ldots, s-1),$$

$$\max_{\frac{1}{\xi_1} \leqq x \leqq 0} \left|f\left(\frac{1}{x}\right) - p_n^{\{0\}}(x)\right| \leqq \varepsilon_n, \qquad \max_{0 \leqq x \leqq \frac{1}{\xi_s}} \left|f\left(\frac{1}{x}\right) - p_n^{\{s\}}(x)\right| \leqq \varepsilon_n.$$

By adding a suitable chosen linear function to $p_n^{\{i\}}(x)$ we get polynomials $P_n^{\{i\}}(x)$ for which

(3.1)
$$\max_{\xi_i \leqq x \leqq \xi_{i+1}} |f(x) - P_n^{\{i\}}(x)| \leqq \varepsilon_n \qquad (i = 1, \ldots, s-1),$$

$$\max_{\frac{1}{\xi_1} \leqq x \leqq 0} \left|f\left(\frac{1}{x}\right) - P_n^{\{0\}}(x)\right| \leqq \varepsilon_n, \qquad \max_{0 \leqq x \leqq \frac{1}{\xi_s}} \left|f\left(\frac{1}{x}\right) - P_n^{\{s\}}(x)\right| \leqq \varepsilon_n,$$

and

(3.2)
$$P_n^{\{i\}}(\xi_i) = f(\xi_i), \quad P_n^{\{i\}}(\xi_{i+1}) = f(\xi_{i+1}) \qquad (i = 1, \ldots, s-1),$$
$$P_n^{\{0\}}\left(\frac{1}{\xi_1}\right) = f(\xi_1), \quad P_n^{\{s\}}\left(\frac{1}{\xi_s}\right) = f(\xi_s).$$

Now let $T_n(x^{\{i\}})$ be the Chebishev polynomial of degree n, transformated to that interval on which $P_n^{\{i\}}(x)$ approximates, and

$$r_n^{\{i\}}(x) = \frac{(1+\eta_n)P_n^{\{i\}}(x)}{1+\eta_n T_n(x^{\{i\}})} \quad (i = 0, 1, \ldots, s), \qquad \eta_n = e^{-\sqrt{\frac{n}{12}}}.$$

Being n an even number, we have by (3. 2)

$$(3.3)\qquad r_n^{\{i\}}(\xi_i)=f(\xi_i),\quad r_n^{\{i\}}(\xi_{i+1})=f(\xi_{i+1})\qquad (i=1,\dots,s-1),$$

$$r_n^{\{0\}}\left(\frac{1}{\xi_1}\right)=f(\xi_1),\quad r_n^{\{s\}}\left(\frac{1}{\xi_s}\right)=f(\xi_s),$$

and by (3. 1)

$$(3.4)\qquad \max_{\xi_i\leqq x\leqq\xi_{i+1}}|f(x)-r_n^{\{i\}}(x)|=O(\varepsilon_n+\eta_n)\qquad (i=1,\dots,s-1),$$

$$(3.5)\qquad \max_{\frac{1}{\xi_1}\leqq x\leqq 0}\left|f\left(\frac{1}{x}\right)-r_n^{\{0\}}(x)\right|=O(\varepsilon_n+\eta_n),\qquad \max_{0\leqq x\leqq\frac{1}{\xi_s}}\left|f\left(\frac{1}{x}\right)-r_n^{\{s\}}(x)\right|=O(\varepsilon_n+\eta_n).$$

It is easy to verify (see [2]) that

$$(3.6)\qquad \sup_{-\infty<x<+\infty}|r_n^{\{i\}}(x)|=O(\eta_n^{-1})\qquad (i=0,1,\dots,s)$$

and

$$(3.7)\qquad \sup_{-\infty<x<+\infty}\left|\frac{d}{dx}r_n^{\{i\}}(x)\right|=O\left(\frac{n^2}{\eta_n}\right)\qquad (i=0,1,\dots,s).$$

We may write (3. 5) in the form

$$(3.8)\qquad \sup_{-\infty<x\leqq\xi_1}\left|f(x)-r_n^{\{0\}}\left(\frac{1}{x}\right)\right|=O(\varepsilon_n+\eta_n),$$

$$\sup_{\xi_s\leqq x<+\infty}\left|f(x)-r_n^{\{s\}}\left(\frac{1}{x}\right)\right|=O(\varepsilon_n+\eta_n).$$

Let

$$(3.9)\qquad T(x)=\frac{r_n^{\{0\}}\left(\frac{1}{x}\right)+r_n^{\{s\}}\left(\frac{1}{x}\right)}{2}+\frac{1}{2}\sum_{i=2}^{s-1}|x-\xi_i|\,\frac{r_n^{\{i\}}(x)-r_n^{\{i-1\}}(x)}{x-\xi_i}+$$

$$+\frac{|x-\xi_1|}{2(x-\xi_1)}\left[r_n^{\{1\}}(x)-r_n^{\{0\}}\left(\frac{1}{x}\right)\right]+\frac{|x-\xi_s|}{2(x-\xi_s)}\left[r_n^{\{s\}}\left(\frac{1}{x}\right)-r_n^{\{s-1\}}(x)\right].$$

Clearly

$$T(x)=\begin{cases} r_n^{\{i\}}(x) & \text{if}\quad \xi_i\leqq x\leqq\xi_{i+1}\qquad (i=1,\dots,s-1)\\ r_n^{\{0\}}\left(\frac{1}{x}\right) & \text{if}\quad -\infty<x\leqq\xi_1\\ r_n^{\{s\}}\left(\frac{1}{x}\right) & \text{if}\quad \xi_s\leqq x<+\infty\end{cases}$$

therefore by (3. 4) and (3. 8)

$$|f(x)-T(x)|=O(\varepsilon_n+\eta_n)\qquad (-\infty<x<+\infty).$$

Consider the rational function (see (1. 1))

$$R_N(x) = \frac{r_n^{\{0\}}\left(\frac{1}{x}\right) + r_n^{\{s\}}\left(\frac{1}{x}\right)}{2} + \frac{1}{2}\sum_{i=2}^{s-1} Q_n(x-\xi_i)\frac{r_n^{\{i\}}(x) - r_n^{\{i-1\}}(x)}{x-\xi_i} + \tag{3.10}$$

$$+ \frac{Q_n(x-\xi_1)}{2(x-\xi_1)}\left[r_n^{\{1\}}(x) - r_n^{\{0\}}\left(\frac{1}{x}\right)\right] + \frac{Q_n(x-\xi_s)}{2(x-\xi_s)}\left[r_n^{\{s\}}\left(\frac{1}{x}\right) - r_n^{\{s-1\}}(x)\right].$$

This is of degree not greater than $2n+3ns \leqq N$.

Now we have by (3. 3) and (3. 7)

$$\left|\frac{r_n^{\{i\}}(x) - r_n^{\{i-1\}}(x)}{x-\xi_i}\right| \leqq \left|\frac{r_n^{\{i\}}(x) - r_n^{\{i\}}(\xi_i)}{x-\xi_i}\right| + \left|\frac{r_n^{\{i-1\}}(x) - r_n^{\{i-1\}}(\xi_i)}{x-\xi_i}\right| = O\left(\frac{n^2}{\eta_n}\right) \tag{3.11}$$

$$(-\infty < x < +\infty,\ i = 1, \ldots, s-1)$$

and

$$\left|\frac{r_n^{\{1\}}(x) - r_n^{\{0\}}\left(\frac{1}{x}\right)}{x-\xi_1}\right| \leqq \left|\frac{r_n^{\{1\}}(x) - r_n^{\{1\}}(\xi_1)}{x-\xi_1}\right| + \left|\frac{r_n^{\{0\}}\left(\frac{1}{x}\right) - r_n^{\{0\}}\left(\frac{1}{\xi_1}\right)}{x-\xi_1}\right|.$$

Here the first term can be estimated as above. As regards to the second term, let first of all $x \in \left[\frac{3}{2}\xi_1, \frac{1}{2}\xi_1\right]$. Then by (3. 7)

$$\left|\frac{r_n^{\{0\}}\left(\frac{1}{x}\right) - r_n^{\{0\}}\left(\frac{1}{\xi_1}\right)}{x-\xi_1}\right| = \frac{1}{|x\xi_1|}\left|\frac{r_n^{\{0\}}\left(\frac{1}{x}\right) - r_n^{\{0\}}\left(\frac{1}{\xi_1}\right)}{\frac{1}{x} - \frac{1}{\xi_1}}\right| \leqq$$

$$\leqq \frac{2}{\xi_1^2}\sup_{-\infty<x<+\infty}\left|\frac{d}{dx}r_n^{\{0\}}(x)\right| = O\left(\frac{n^2}{\eta_n}\right).$$

On the other hand, if $x \notin \left[\frac{3}{2}\xi_1, \frac{1}{2}\xi_1\right]$ then by (3. 6)

$$\left|\frac{r_n^{\{0\}}\left(\frac{1}{x}\right) - r_n^{\{0\}}\left(\frac{1}{\xi_1}\right)}{x-\xi_1}\right| = \frac{2\sup\limits_{-\infty<x<+\infty}|r_n^{\{0\}}(x)|}{\frac{1}{2}|\xi_1|} = O(\eta_n^{-1}).$$

Thus

$$\left|\frac{r_n^{\{1\}}(x) - r_n^{\{0\}}\left(\frac{1}{x}\right)}{x-\xi_1}\right| = O\left(\frac{n^2}{\eta_n}\right) \qquad (-\infty < x < +\infty), \tag{3.12}$$

and analogously

$$\left|\frac{r_n^{\{s\}}\left(\frac{1}{x}\right) - r_n^{\{s-1\}}(x)}{x-\xi_s}\right| = O\left(\frac{n^2}{\eta_n}\right) \qquad (-\infty < x < +\infty). \tag{3.13}$$

We obtain by (3.9), (3.10), (1.1), (3.11), (3.12) and (3.13)

$$|T(x)-R_N(x)| \leqq \frac{3}{4}\left[1+\max_{1\leqq i\leqq s}(x-\xi_i)^2\right] e^{-\sqrt{\frac{n-3}{2}}} \cdot s\frac{n^2}{\eta_n} =$$

$$= (1+x^2)O\left(e^{-\sqrt{\frac{n}{3}}+\frac{1}{2}\sqrt{\frac{n}{3}}}\right) = (1+x^2)O\left(e^{-\sqrt{\frac{n}{12}}}\right) \qquad (-\infty < x < +\infty)$$

which proves Theorem 3.

COROLLARY. If $f(x)$ and $f\left(\frac{1}{x}\right)$ are analytic on the corresponding intervals then

$$|f(x)-R_N(x)| = (1+x^2)O\left(e^{-\sqrt{\frac{1}{6}\left[\frac{N}{6s+4}\right]}}\right) \qquad (-\infty < x < +\infty).$$

REMARK. In connection with Theorems 1—3, when (2. 1) does not hold, we may apply in certain cases these theorems. Namely, if with a certain integer $p \geqq 0$ the function $f(x)(1+x^2)^{-p}$ has finite limits in $-\infty$ and $+\infty$ and the further conditions hold for this function, we obtain the above estimations for $f(x)$, multiplying by $(1+x^2)^p$, and with $n-2p$ instead of n.

§ 4. An ad hoc Method for Approximation of arc tg x

The previous Corollary gives a weak estimation in case of the function arc tg x. We prove the stronger

THEOREM 4. *There exist rational functions $R_n(x)$ of degree n at most such that*

$$(4.1) \qquad |\operatorname{arc\,tg} x - R_n(x)| \leqq \frac{|x|^3}{1+x^2}\cdot\left(\frac{|x|}{1+\sqrt{1+x^2}}\right)^{n-3} \qquad (-\infty < x < +\infty).$$

PROOF. Consider

$$\operatorname{arc\,tg} x = \frac{1}{2x}\int_{-1}^{+1}\frac{dy}{y^2+\frac{1}{x^2}} \qquad (x \neq 0)$$

and approximate the right hand integral by the Gaussian quadrature of order n. It is well-known that with arbitrary but fixed $x \neq 0$

$$\left|\int_{-1}^{+1}\frac{dy}{y^2+\frac{1}{x^2}} - \sum_{k=1}^{n}\frac{A_{kn}}{\xi_{kn}^2+\frac{1}{x^2}}\right| \leqq 4E_{2n-1}\left(\frac{1}{y^2+\frac{1}{x^2}};\ -1,\ +1\right)$$

where ξ_{kn} are the roots of the Legendre polynomial of degree n and A_{kn} are some coefficients. Here the rational function

$$r_{2n}(x) = x^2\sum_{k=1}^{n}\frac{A_{kn}}{\xi_{kn}^2x^2+1}$$

is of degree $2n$, and (see [7])

$$E_{2n-1}\left(\frac{1}{y^2+\frac{1}{x^2}};\ -1,\ +1\right) < \frac{x^4}{2(x^2+1)\left(\frac{1}{|x|}+\sqrt{\frac{1}{x^2}+1}\right)^{2n-2}} \qquad (x\neq 0),$$

i.e. (4. 1) holds for $x\neq 0$. But it is true for $x=0$, too.

REFERENCES

[1] FREUD, G.: A remark concerning the rational approximation of $|x|$, *Studia Sci. Math. Hung.* **2** (1967) 115—118.

[2] FREUD, G.: Über die Approximation reeller Funktionen durch rationale gebrochene Funktionen, *Acta Math. Acad. Sci. Hungar.* **17** (1966) 313—324.

[3] ACHIESER, N. I.: *Lectures on approximation theory* (Russian) Moscow, 1965.

[4] NEWMAN, D.: Rational approximation to $|x|$, *Michigan Math. J.* **11** (1964) 11—14.

[5] SZÜSZ, P., TURÁN, P.: A konstruktiv függvénytan egy új irányáról, (Hungarian), *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **16** (1966) 33—45.

[6] SZABADOS, J.: Structural properties of continuous functions connected with the order of rational approximation, I, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **10** (1967) 95—102.

[7] BERNSTEIN, S.: On best approximation of the function $\int_0^\infty |y|^s d\psi(s)$ on the interval $[-1, +1]$ (Russian), *Collected Works* (Moscow, 1955), Vol. 2, pp. 361—370.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest
and *Eötvös Loránd University, Budapest*

(Received April 2, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 211—218.

ONE-SIDED SPLINE APPROXIMATION

by

A. MEIR and A. SHARMA

1. Recently G. FREUD [2] has obtained estimates for the degree of one-sided approximations to a given function by algebraic polynomials in the L_1-norm. An application of his estimates leads him to obtain a refinement of the remainder in a Tauberian theorem. An important role of one-sided polynomial approximation is played in KARAMATA's proof of the famous LITTLEWOOD 0-Tauberian theorem. More precisely, Freud shows that for a function $f(x) \in C^{\nu-1}[-1, 1]$ where $f^{(\nu-1)}(x)$ is the integral of a function with total variation V_ν there exist polynomials $p_n(x)$ and $P_n(x)$ of degree n satisfying

$$p_n(x) \leqq f(x) \leqq P_n(x) \qquad (-1 \leqq x \leqq 1)$$

$$\int_{-1}^{+1} \{P_n(x) - p_n(x)\} \frac{dx}{\sqrt{1-x^2}} \leqq \frac{AV_\nu}{n^{\nu+1}}, \text{ A constant.}$$

During the last decade, spline functions, their properties of best fit and their degree of approximation have been studied by several authors ([1], [4], [5], [7]). As for the degree of approximation by cubic splines, we have shown recently [6] that for $f \in C^2[0, 1]$ with modulus of continuity $\omega_2(\delta)$ of f'', we have

$$\|f^{(r)}(x) - \Phi_n^{(r)}(x)\|_\infty \leqq 5h_n^{2-r}\omega_2(h_n), \qquad r = 0, 1, 2 \tag{1.1}$$

where Φ_n is the interpolatory cubic spline on $n+1$ nodes with guage h_n. This suggests that with respect to the degree of approximation, the number of nodes of a spline plays essentially the same role as the degree of a polynomial in polynomial approximation. In the present note we obtain estimates for one-sided L_1-approximation by linear and cubic splines, and also by first order trigonometric splines. It appears that these results can be extended to higher order splines (algebraic and trigonometric). In § 2 and 3, we give the proof of our result for cubic splines on equidistant nodes and in § 4 and 5 we sketch an outline of the proofs in the case of linear algebraic and trigonometric splines on arbitrarily given nodes.

2. One-Sided Cubic Splines. We formulate

THEOREM 1. *Suppose* $f(x) \in C^2\,[0, 1]$ *and* $f''(x)$ *is the integral of* $f_3(x) \in BV\,[0, 1]$. *Then for* $n \geqq 2$ *there exist cubic splines* $\varphi_n(x)$ *and* $\Phi_n(x)$ *with nodes* $x_k = \dfrac{k}{n}$ $(0 \leqq k \leqq n)$ *such that*

$$\varphi_n(x) \leqq f(x) \leqq \Phi_n(x), \qquad 0 \leqq x \leqq 1 \tag{2.1}$$

and

$$\int_0^1 \{\Phi_n(x) - \varphi_n(x)\}\, dx \leqq \frac{A \cdot V_3}{n^4} \tag{2.2}$$

where A *is a constant* $\leqq \dfrac{1}{12}$ *independent of* n *and*

$$V_3 = \int_0^1 |df_3|.$$

The proof of the theorem is based essentially on the following:

LEMMA. *For any* ξ, $0 \leqq \xi \leqq 1$, *there exist cubic splines* $\gamma(x, \xi)$ *and* $\Gamma(x, \xi)$ *with nodes* $x_k = \dfrac{k}{n}$ $(0 \leqq k \leqq n)$ *such that*

$$\gamma(x, \xi) \leqq (x - \xi)_+^3 \leqq \Gamma(x, \xi) \tag{2.3}$$

and

$$\int_0^1 \{\Gamma(x, \xi) - \gamma(x, \xi)\}\, dx \leqq \frac{1}{2n^4}. \tag{2.4}$$

PROOF OF THE LEMMA. For any given ξ, there exists an integer μ, $0 \leqq \mu \leqq n$ such that $\xi = x_\mu + \dfrac{\theta}{n}$, $0 \leqq \theta < 1$. We now set [See Figure 1]

$$\gamma(x, \xi) = \sum_{j=0}^{3} A_j (x - x_{\mu+j-1})_+^3 \tag{2.5}$$

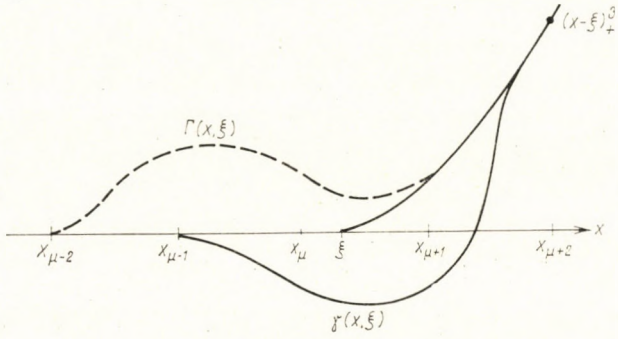


Fig. 1

and require that for $x \geqq x_{\mu+2}$, $\gamma(x, \xi) \equiv (x-\xi)^3$. This identity is equivalent to the system of equations

$$\sum_{j=0}^{3} A_j x_{\mu+j-1}^k = \xi^k, \qquad (k = 0, 1, 2, 3).$$

Simple computation shows that

$$A_0 = -\frac{\theta(1-\theta)(2-\theta)}{6} < 0, \quad A_1 = \frac{(1+\theta)(1-\theta)(2-\theta)}{2} > 0$$

(2. 6)

$$A_2 = \frac{\theta(1+\theta)(2-\theta)}{2} > 0, \quad A_3 = -\frac{\theta(1-\theta)(1+\theta)}{6} < 0.$$

Since $A_0 < 0$ and $A_3 < 0$ we have

(2. 7) $$\gamma(x, \xi) \leqq 0 \quad \text{for} \quad x \leqq x_\mu,$$

and

(2. 8) $$\gamma(x, \xi) \leqq (x-\xi)^3 \quad \text{for} \quad x \geqq x_{\mu+1}.$$

If $x_\mu \leqq x \leqq x_{\mu+1}$, setting $x = x_\mu + \frac{t}{n}$, we have for $0 \leqq t \leqq 1$

(2. 9) $$\gamma(x, \xi) - (x-\xi)_+^3 = A_0 \left(\frac{1+t}{n}\right)^3 + A_1 \left(\frac{t}{n}\right)^3 - \left(\frac{t-\theta}{n}\right)_+^3 \overset{\text{def}}{=} \frac{1}{n^3} g(t, \theta).$$

Now, if $0 \leqq t \leqq \theta$, we have on using (2. 6)

$$\operatorname{sgn} g(t, \theta) = \operatorname{sgn} \left\{ -\left(\frac{1+t}{t}\right)^3 + 3 \cdot \frac{1+\theta}{\theta} \right\} \leqq$$

(2. 10)

$$\leqq \operatorname{sgn} \left\{ -\left(\frac{1+t}{t}\right)^3 + 3 \cdot \frac{1+t}{t} \right\} = -1.$$

On the other hand if $\theta \leqq t \leqq 1$, it is easily seen that

$$g(t, 0) = 0, \; g(t, t) \leqq 0$$

and

$$\left.\frac{\partial^2 g}{\partial \theta^2}\right]_{\theta=0} \geqq 0, \quad \left.\frac{\partial^2 g}{\partial \theta^2}\right]_{\theta=t} > 0$$

so that from the linearity of $\frac{\partial^2 g}{\partial \theta^2}$, we get $\frac{\partial^2 g}{\partial \theta^2} > 0$ for $0 \leqq \theta \leqq t$. Thus, $g(t, \theta)$ being a convex function of θ in $[0, t]$,

(2. 11) $$g(t, \theta) \leqq 0, \; 0 \leqq \theta \leqq t.$$

Combining (2. 7)—(2. 11), we have

(2. 12) $$\gamma(x, \xi) \leqq (x-\xi)_+^3, \quad \text{for all} \quad x.$$

Similarly, setting

$$\Gamma(x, \xi) = \sum_{j=0}^{3} B_j (x - x_{\mu+j-2})_+^3$$

with $\Gamma(x, \xi) \equiv (x-\xi)^3$ for $x \geqq x_{\mu+1}$, we have

$$B_0 = \frac{\theta(1-\theta)(1+\theta)}{6}, \quad B_1 = -\frac{\theta(1-\theta)(2+\theta)}{2}$$
$$B_2 = \frac{(1-\theta)(1-\theta)(2+\theta)}{2}, \quad B_3 = \frac{\theta(1+\theta)(2+\theta)}{6}. \tag{2.13}$$

By suitable modifications of the previous argument we show that

$$\Gamma(x, \xi) \geqq (x-\xi)_+^3 \quad \text{for all} \quad x. \tag{2.14}$$

This completes the proof of (2. 3). Now using (2. 6) and (2.13), we have by straight forward computation

$$\int_0^1 \{\Gamma(x, \xi) - \gamma(x, \xi)\}\, dx = \frac{\theta(1-\theta^2)}{n^4} \leqq \frac{1}{2n^4}$$

which proves (2. 4).

3. Proof of Theorem 1. By MacLaurin's formula, we have for all x, $0 \leqq x \leqq 1$,

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f_3(0) + \int_0^1 \frac{(x-\xi)_+^3}{3!}\, df_3(\xi). \tag{3. 1}$$

Since $f_3(x)$ is of BV, $f_3(x) = f_3^+(x) - f_3^-(x)$, where $f_3^+(x)$ and $f_3^-(x)$ are non-decreasing and

$$|df_3(x)| = df_3^+(x) + df_3^-(x). \tag{3. 2}$$

We define now $\Phi_n(x)$ and $\varphi_n(x)$ as follows:

$$\Phi_n(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f_3(0) +$$
$$+ \frac{1}{6} \int_0^1 \Gamma(x, \xi)\, df_3^+(\xi) - \frac{1}{6} \int_0^1 \gamma(x, \xi)\, df_3^-(\xi). \tag{3. 3}$$

$$\varphi_n(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f_3(0) +$$
$$+ \frac{1}{6} \int_0^1 \gamma(x, \xi)\, df_3^+(\xi) - \frac{1}{6} \int_0^1 \Gamma(x, \xi)\, df_3^-(\xi). \tag{3. 4}$$

It is easy to verify that $\Phi_n(x)$, $\varphi_n(x)$ are cubic splines with nodes $x_k = \frac{k}{n}$, $0 \leqq k \leqq n$. Also from (2. 3), (3. 1), (3. 3) and (3. 4) it follows that (2. 1) holds. Now from (3. 2),

(3. 3) and (3. 4),

$$\int_0^1 \{\Phi_n(x)-\varphi_n(x)\}\,dx = \frac{1}{6}\int_0^1 dx \int_0^1 \{\Gamma(x,\xi)-\gamma(x,\xi)\}\,|df_3(\xi)| =$$

$$= \frac{1}{6}\int_0^1 |df_3(\xi)| \int_0^1 \{\Gamma(x,\xi)-\gamma(x,\xi)\}\,dx \leqq \frac{1}{12n^4}\int_0^1 |df_3(\xi)| = \frac{V_3}{12n^4}$$

on the basis of (2. 4) which completes the proof of the theorem.

4. Linear Splines. Let $0 = x_0 < x_1 < \ldots < x_n = 1$, be given nodes with $\max_i (x_{i+1}-x_i) = \Delta$. Then we can formulate

THEOREM 2. *Suppose $f(x) \in C[0, 1]$ and that $f(x)$ is the integral of $f_1(x) \in BV[0, 1]$. Then there exist first order splines $l_n(x)$ and $L_n(x)$ with the nodes $\{x_j\}_0^n$ such that*

$$l_n(x) \leqq f(x) \leqq L_n(x), \qquad 0 \leqq x \leqq 1 \tag{4.1}$$

and

$$\int_0^1 \{L_n(x)-l_n(x)\}\,dx \leqq V_1 \cdot \Delta^2 \tag{4.2}$$

where V_1 is the variation of $f_1(x)$.

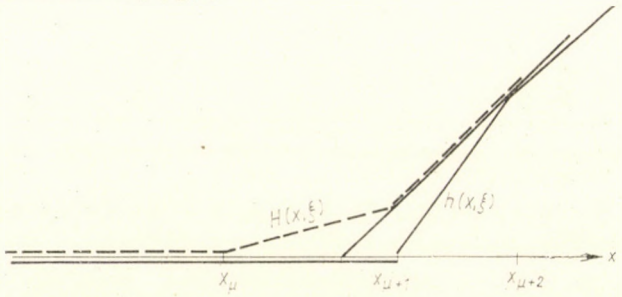


Fig. 2

The proof of this theorem follows the lines of Theorem 1 (and is indeed much simpler). It is enough to observe that in this case for fixed ξ, we may determine the following first order splines $H(x, \xi)$ and $h(x, \xi)$ (corresponding to Γ and γ in the above lemma) for the curve $(x-\xi)_+$ [see Figure 2.]:

$$H(x,\xi) \equiv 0, \qquad x \leqq x_\mu$$

$$= \frac{x_{\mu+1}-\xi}{x_{\mu+1}-x_\mu}(x-x_\mu), \qquad x_\mu < x \leqq x_{\mu+1}$$

$$= x-\xi, \qquad x > x_{\mu+1}.$$

$$h(x,\xi) \equiv 0, \qquad x \leqq x_{\mu+1}$$

$$= \frac{x_{\mu+2}-\xi}{x_{\mu+2}-x_{\mu+1}} \cdot (x-x_{\mu+1}), \quad x_{\mu+1} < x \leqq x_{\mu+2}$$

$$= x-\xi, \qquad x > x_{\mu+2}$$

5. Linear Trigonometric Splines. Trigonometric splines have been introduced by SCHOENBERG [5] and their properties of best fit have also been obtained. We shall use a recent result of *M.* PICONE [3] to obtain approximation by one-sided trigonometric splines. Let $0=x_0<x_1<\ldots<x_n=2\pi$ be any given nodes with $\max_i (x_{i+1}-x_i) = \Delta < \frac{\pi}{2}$ and $\min_i (x_{i+1}-x_i) = \delta$. Then we have the following

THEOREM 3. *Suppose* $f\in C^1[0, 2\pi]$ *and* $f'(x)$ *is the integral of a function* $f_2(x)\in BV[0, 2\pi]$. *Then there exist first order trigonometric splines* $\lambda_n(x)$ *and* $\Lambda_n(x)$ *with nodes* $\{x_i\}_0^n$ *such that*

$$\lambda_n(x)\leqq f(x)\leqq \Lambda_n(x), \quad 0\leqq x\leqq 2\pi \tag{5. 1}$$

and

$$\int_0^{2\pi} \{\Lambda_n(x)-\lambda_n(x)\}\,dx \leqq \frac{2}{3}\frac{\Delta^4}{\delta} V \tag{5. 2}$$

where V *is the total variation of* $f(x)+f''(x)$.

PROOF. From the result of PICONE [3] or by direct computation, we see that

$$f(x) = \tau(x)+2\int_0^{2\pi} \sin^2\left(\frac{x-\xi}{2}\right)_+ d(f_2(\xi)+f(\xi)) \tag{5. 3}$$

where

$$\tau(x) = f(0)+f'(0)\sin x+f_2(0)(1-\cos x).$$

We now approximate $2\sin^2\left(\frac{x-\xi}{2}\right)_+$ from above and below by first order trigonometric splines. In fact if $x_\mu\leqq\xi<x_{\mu+1}$ then defining constants $\alpha_0, \alpha_1, \alpha_2$ by the identity

$$\sum_{i=0}^{2} \alpha_i \sin^2\frac{x-x_{\mu-1+i}}{2} \equiv \sin^2\frac{x-\xi}{2} \quad \text{for} \quad x > x_{\mu+1}$$

we see by elementary computations that

$$\alpha_0 = -\frac{\sin\frac{\xi-x_\mu}{2}\sin\frac{x_{\mu+1}-\xi}{2}}{\sin\frac{x_\mu-x_{\mu-1}}{2}\sin\frac{x_{\mu+1}-x_{\mu-1}}{2}}$$

and

$$\alpha_2 = \frac{\sin\frac{\xi-x_\mu}{2}\sin\frac{\xi-x_{\mu-1}}{2}}{\sin\frac{x_{\mu+1}-x_\mu}{2}\sin\frac{x_{\mu+1}-x_{\mu-1}}{2}}$$

so that $\alpha_0 < 0$ and $0 < \alpha_2 < 1$. Also observing that $\sin\alpha \sin\beta \leqq \sin^2(\alpha+\beta)$ for $\alpha, \beta \leqq \frac{\pi}{4}$, and $\frac{\sin t}{t}$ is decreasing for $0 < t < \frac{\pi}{2}$, we see that

$$|\alpha_0| \leqq \frac{\sin\frac{x_{\mu+1}-x_\mu}{2}}{\sin\frac{x_\mu - x_{\mu-1}}{2}} \leqq \frac{\sin\frac{\Delta}{2}}{\sin\frac{\delta}{2}} \leqq \frac{\Delta}{\delta}. \tag{5.4}$$

Consider now the spline

$$t(x,\xi) \equiv 2\sum_{i=0}^{2} \alpha_i \sin^2\left(\frac{x - x_{\mu-1+i}}{2}\right)_+$$

Since $\alpha_0 < 0$ and $\alpha_2 > 0$, it is immediate that

$$t(x,\xi) \leqq 2\sin^2\left(\frac{x-\xi}{2}\right)_+, \tag{5.5}$$

for $x \leqq x_\mu$ and $x \geqq \xi$. Also it is easily seen that $\frac{\partial^2 t(x,\xi)}{\partial x^2} > 0$ for $x^\mu < x < \xi$, so that $t(x, \xi)$ is a convex function of x. Hence (5. 5) is valid also for $x_\mu < x \leqq \xi$.

Similarly, we can show that

$$T(x,\xi) \equiv 2\sum_{i=0}^{2} \beta_i \sin^2\left(\frac{x - x_{\mu+i}}{2}\right)_+$$

satisfies for all x in $[0, 2\pi]$,

$$T(x,\xi) \geqq 2\sin^2\left(\frac{x-\xi}{2}\right)_+. \tag{5.6}$$

Computations completely analogous to the above show that

$$0 < \beta_0 \leqq 1, \quad \beta_2 < 0 \quad \text{and} \quad |\beta_2| < \frac{\Delta}{\delta}. \tag{5.7}$$

Since $f_2 \in BV[0, 2\pi]$, so is $F(x) \equiv f_2(x) + f(x)$; hence $F(x) = F^+(x) - F^-(x)$, where F^+, F^- are non-decreasing functions. We now set

$$\lambda_n(x) = \tau(x) + \int_0^{2\pi} t(x,\xi)\, dF^+(\xi) - \int_0^{2\pi} T(x,\xi)\, dF^-(\xi)$$

$$\Lambda_n(x) = \tau(x) + \int_0^{2\pi} T(x,\xi)\, dF^+(\xi) - \int_0^{2\pi} t(x,\xi)\, dF^-(\xi).$$

Then it is easy to see that $\lambda_n(x)$ and $\Lambda_n(x)$ are first order trigonometric splines with the prescribed nodes and that

$$\int_0^{2\pi} \{\Lambda_n(x) - \lambda_n(x)\}\, dx = \int_0^{2\pi} dx \int_0^{2\pi} \{T(x,\xi) - t(x,\xi)\}\left(dF^+(\xi) + dF^-(\xi)\right) = \tag{5.8}$$

$$= \int_0^{2\pi} |dF(\xi)| \int_0^{2\pi} \{T(x,\xi) - t(x,\xi)\}\, dx.$$

Now it is easy to check that uniformly in ξ

$$\int_0^{2\pi} \{T(x,\xi)-t(x,\xi)\}\,dx \leqq \frac{\Delta^3}{6}\{\alpha_2-\alpha_0+\beta_0-\beta_2\} \tag{5.9}$$

which by (5. 4) and (5. 7),

$$\leqq \frac{2}{3}\frac{\Delta^4}{\delta}.$$

Now (5. 2) follows from (5. 8) and (5. 9).

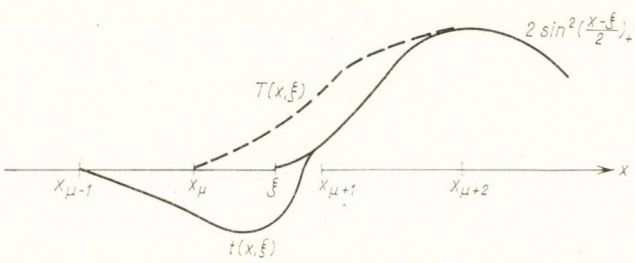


Fig. 3

REMARK 1. If $x_k=\frac{2k\pi}{n}$, $0\leqq k\leqq n$, then we see from (5. 2) that the degree of one-sided approximation by linear trigonometric splines is $O\left(\frac{1}{n^3}\right)$.

REMARK 2. A comparison of our earlier result (1. 1) with (2. 2) is in order. If in (1. 1), $h_n=\frac{1}{n}$ and if f satisfies conditions of Theorem 1 above, we have $\omega_2\left(\frac{1}{n}\right)=$ $=O\left(\frac{1}{n}\right)$ and so the cubic spline approximation is uniformly $O\left(\frac{1}{n^3}\right)$, whereas the one-sided approximation by cubic splines in the L_1 norm is $O\left(\frac{1}{n^4}\right)$.

REFERENCES

[1] BIRKHOFF, G. and DE BOOR, C.: Error bounds for spline interpolation, *J. Math. Mech.* **13** (1966) 827—836.
[2] FREUD, G.: Über einseitige Approximation durch Polynome, I, *Acta Sci. Math. (Szeged)* **16** (1955) 12—28.
[3] PICONE, M.: Sull'approssimazione tayloriana di una funzione, *Accad. Naz. Lincei. Rend.* **38** (4) (1965) 441—447.
[4] SCHOENBERG, I. J. and WHITNEY, A.: On Pólya frequency functions, III, *Trans. Amer. Math. Soc.* **74** (1953) 246—259.
[5] SCHOENBERG, I. J.: On trigonometric spline interpolation, *J. Math. Mech.* **13** (1964) 795—826.
[6] SHARMA, A. and MEIR, A.: Degree of approximation of spline interpolation, *J. Math. Mech.* **15** (1966) 759—767.
[7] WALSH, J. L., AHLBERG, J. H. and NILSON, E. N.: Best approximation properties of spline fit, *J. Math. Mech.* **11** (1962) 225—234.

University of Alberta, Edmonton, Alberta, Canada

(Received April 21, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 219—224.

SOME EXAMPLES OF INFINITELY DIVISIBLE POINT PROCESSES

by

P. M. LEE

1. Introduction

In this paper the theory of infinitely divisible point processes developed independently by the author and others (see [1], [6], [7], [8], [9], [10]) is applied to obtain results about certain special types of point process. We take as our starting point the notion of a point process as a random measure introduced in [13], using the following definitions and theorems from the papers referred to above:

DEFITION 1. *A point process on the real line is a stochastic process* $X(C)$ $(C \in \mathcal{B})$ *defined on the class* $\mathcal{B}$ *of Borel sets in* R *whose sample functions are non-negative integer-valued (except possibly for the value* $+\infty$ *on unbounded sets) and* σ*-additive with probability* 1.

DEFINITION 2. *A point process is said to be completely random (cf.* [5]), *or without after-effects, if for every class* $\{C_a\}$ *of disjoint Borel sets,* $\{X(C_a)\}$ *is a family of independent random variables.*

DEFINITION 3. *For each m-tuple* $\mathbf{C} = (C_1, \ldots, C_m)$ *of Borel sets we define the function* $h(\mathbf{s}; \mathbf{C})$ *(called the h-function) of a point process* X *by the equation*

$$h(\mathbf{s}; \mathbf{C}) = -\log \mathsf{E} \exp(-\mathbf{s} \cdot \mathbf{X}(\mathbf{C})) \qquad (\mathbf{s} \in R_+^m, \text{ where } R_+ = [0, \infty)),$$

where $\mathbf{X}(\mathbf{C})$ *denotes the vector* $(X(C_1), \ldots, X(C_m))$.

DEFINITION 4. *The superposition (sum)* $X = \sum_n X_n$ *of two or more point processes* X_n *is defined as the process* $X(C) = \sum_n X_n(C)$.

DEFINITION 5. *A point process* $X(C)$ *is said to be infinitely divisible if there exists a set* $X_{r,k}$ $(r = 1, 2, \ldots; k = 1, 2, \ldots, r)$ *of point processes such that for any fixed* r *the* $X_{r,k}$ *are independently and identically distributed and*

$$X \doteq \sum_{k=1}^{r} X_{r,k}$$

i.e. X *has the same distribution as the sum on the right hand side.*

THEOREM 1. *A point process* X *is infinitely divisible if and only if its h-functions can be put in the form*

$$h(\mathbf{s}; \mathbf{C}) = \sum_{\mathbf{n} > 0} [1 - \exp(-\mathbf{s} \cdot \mathbf{n})] \Lambda\{\mathbf{s}(\mathbf{C}) = \mathbf{n}\}$$

where Λ is a measure on the set $M_P(\mathscr{B})$ of measures x on $\mathscr{B}$ taking values $0, 1, 2, \ldots, +\infty$ only (or strictly on the σ-ring of subsets of $M_P(\mathscr{B})$ generated by the subsets occurring in the above expression) and $\Lambda(X(C)>0)<+\infty$ for every bounded Borel set C. Conversely, to each such Λ measure there corresponds an infinitely divisible point process.

2. Generalized Poisson Process

The first, and in some ways the most interesting, particular point process we shall examine will be the generalized Poisson process. For any non-negative measure μ on the Borel sets of the real line the corresponding generalized Poisson process $q(\mu)$ is defined as a completely random point process in which, for each bounded Borel set C, $X(C)$ is a Poisson variable of mean $\mu(C)$. If μ is λ times Lebesgue measure, $q(\mu)$ is an ordinary Poisson process of rate λ. In general

$$h(\mathbf{s};\mathbf{C}) = \sum_{i=1}^{m} [1-\exp(-s_i)]\mu(C_i)$$

for any m-tuple $\mathbf{C}$ of disjoint Borel sets. It is easy to deduce that

$$\Lambda\{x(C_i) = \delta_j^i \text{ for all } i\} = \mu(C_j)$$

$$\Lambda\left\{\sum_{j=1}^{m} x(C_i) \neq 1\right\} = 0$$

and thus Λ is concentrated on

$$D_1 = \{\delta_t;\ t \in R\}$$

(where δ_t is the DIRAC measure, so that $\delta_t(A)=1$ if $t\in A$, $\delta_t(A)=0$ if $t\notin A$) in the sense that

$$\Lambda(M_P \backslash D_1) = 0.$$

Conversely, let X be an infinitely divisible point process for which $\Lambda(M_P\backslash D_1)=0$. If

$$\mu(C) = \Lambda\{x(C)=1\}$$

then X is distributed as the generalized Poisson process $q(\mu)$. Weakening these hypotheses we can obtain:

THEOREM 2. *An infinitely divisible point process X is a generalized Poisson process if and only if there exists a sequence $K_s \uparrow R$ of bounded Borel sets for each of which $X(K_s)$ is a Poisson variable.*

PROOF. The necessity of the condition is evident.

To prove that the condition is sufficient, suppose that it holds for a point process which is not a generalized Poisson process. Then $\Lambda(M_P\backslash D_1)>0$ and hence

$$\lim_{s\to\infty} \Lambda\{x(K_s) > 1\} > 0$$

so that there exists s and $n>1$ such that $\Lambda\{x(K_s)=n\}>0$. It follows that $X(K_s)$ is *not* a Poisson variable which is a contradiction.

COROLLARY. *A stationary infinitely divisible point process X is a Poisson process if and only if there exists a sequence (K_s) of Borel sets such that $X(K)$ is a Poisson variable and either $K_s \uparrow R$ or each K_s is an interval and the length of K_s tends to infinity.*

These results no longer hold without further assumptions (such as complete randomness, which is evidently a sufficient condition) if the assumption that X is infinitely divisible be suppressed. A stationary point process in which the counts in all intervals have Poisson distributions without the process being Poisson can be constructed as follows.

We first consider four Poisson variables X_1, X_2, X_3, X_4. Starting from the probabilities in the case of independent events we modify slightly the probabilities of some simple events, increasing the probability of those to which a $+$ is annexed by an amount δ and decreasing that of those to which a $-$ is annexed by a similar amount, leaving other probabilities constant.

X_1	X_2	X_3	X_4	
0	0	1	0	$+$
0	0	1	1	$-$
0	1	0	0	$-$
0	1	0	1	$+$
1	0	1	0	$-$
1	0	1	1	$+$
1	1	0	0	$+$
1	1	0	1	$-$

Probabilities of compound events can be deduced by addition. In particular it can be deduced that any two or three consecutive random variables are stochastically independent. Moreover the sum of all four clearly has a Poisson distribution.

Joining independent copies of this set of random variables gives a stochastic process in which any two or three consecutive components are stochastically independent and in which the sum of any set of n consecutive components has a Poisson distribution. By letting X_n be the number of calls in $[n, n+1]$ and assigning calls occurring in such an interval to independently uniformly distributed times we can modify this process to give a continuous time process such that the number of calls in any interval has a Poisson distribution. Finally mixing the continuous time process with their translates gives strictly stationary processes with the same properties.

RÉNYI [12] has recently shown that if $X(I)$ is a Poisson variable not merely when I is an interval but also when it is a finite union of intervals then X is necessarily a Poisson process. In a remark added to the published version of his paper he refers to the existence of counter-example with the same properties as the above due to L. SHEPP which is to be published in a forthcoming paper of J. GOLDMAN.

3. Bulk Poisson Processes

A point process considered as a random measure is necessarily purely atomic, the mass of each atom being a (non-negative) integer. Clearly if μ has no atoms then with probability 1 all the atoms of $q(\mu)$ have mass 1. A bulk Poisson process is defined as a process obtained by replacing each atom of such a process by an atom with a positive integer valued random mass, the distribution of this mass depending solely on the position (time) of the atom and being independent of the masses of the other atoms. An example of this sort of situation is obtained by analysing the process consisting of the times at which particles are emitted from a sample of radio-active material to record the number emitted at each time.

Alternatively, a bulk Poisson process may be defined as a completely random point process X in which for each C, $X(C)$ is of extended Poisson type, i.e. has the same distribution as $\sum_{n<0} \Pi(\lambda_n)n$ where the Π's are independent Poisson variables with means λ_n such that $\sum\lambda_n<\infty$. From this formulation, which is easily seen to be equivalent to the former, it follows that an infinitely divisible point process without fixed atoms is a bulk Poisson process if and only if it is completely random (the case where the process *has* fixed atoms can be dealt with, but it is of no great interest). It is known (see [2], [4], [11], [14]) that, even without the assumption of infinite divisibility, a point process which is decomposable and fulfils a mild condition must be a bulk Poisson process. A rather more general version of this result than those given to date is proved by a new method below:

THEOREM 3. *Let X be a point process which is completely random and such that either*

(i) *$A(t)$ being defined as the expectation of $X[0, t)$ for $t>0$ and of $X[t, 0)$ for $t<0$ $(A(0)=0)$, $A(t)$ is a continuous finite-valued function of t or*

(ii) *p_t being defined as $\mathrm{P}\{X[0, t)>0\}$ for $t>0$ and as $\mathrm{P}\{X[t, 0)>0\}$ for $t<0$ $(p_0=1)$, p_t is a continuous non-vanishing function of t.*

Then X is a bulk Poisson process.

PROOF. Let $I(r, k; a) = [ra+r^{-1}k, ra+r^{-1}\{k+1\})$ *for* $r=1, 2, \ldots; k=1, 2, \ldots, \ldots, r^2; a=\ldots, -1, 0, 1, \ldots$. Then we specify a set $X_{r,k}$ of point processes by setting

$$\mathrm{P}\{X_{r,k}(C)=X(C)\}=1$$

if $C\subset I(r, k; a)$ for some a, and

$$\mathrm{P}\{X_{r,k}(C)=0\}=1$$

if $C\cap\left(\bigcup_a I(r, k; a)\right)=\emptyset$. It follows from complete randomness that the $X_{r,k}$ are independent and that for all r

$$X=\sum_{k=1}^{r^2} X_{r,k}$$

We now show that for any c, $\mathrm{P}\{X_{r,k}(-c, c)>0\}\to 0$ as r tends to infinity, uniformly with respect to k. For, if $r>2c$

$$\mathrm{P}\{X_{r,k}(-c, c)>0\}\leqq\mathrm{P}\{X(I(r, k; 0))>0\}.$$

Now if $p_{s,t}$ denotes $\mathsf{P}\{X[s, t)=0\}$ for $s<t\leqq 0$ or $0\leqq s<t$, under condition (i)

$$1-p_{s,t}\leqq \mathsf{E}X[s, t)=|A(t)-A(s)|\to 0 \quad \text{as} \quad s\to t$$

while under condition (ii)

$$1-p_{s,t}\leqq 1-\min\{p_s/p_t, p_t/p_s\}=(\max\{p_s^{-1}, p_t^{-1}\})|p_s-p_t|\to 0$$

as $s\to t$. Moreover, the convergences are uniform on $(-c, c)$. Hence, given $\varepsilon>0$ if r is large enough, and hence $I(r, k; 0)$ small enough, we obtain

$$\mathsf{P}\{X_{r,k}(-c, c) > 0\} \leqq \varepsilon.$$

We may deduce that for all m-tuples C, the $X_{r,t}(C)$ are uniformly asymptotically negligible, and so (from the central limit theorem) that X is a bulk Poisson process. The theorem is proved.

To conclude this section, we note that in terms of the measure Λ, bulk Poisson processes are characterized by having $\Lambda(M_P\backslash\Delta_1)=0$ where

$$\Delta_1=\{n\delta_t;\ n \quad \text{a non-negative integer}, \quad t\in R\},$$

a result which is proved in the same way as the characterization of the generalized Poisson process given earlier. An alternative formulation (from complete randomness) is that $\Lambda\{x(C)>0\}$ forms a measure on Borel sets.

4. A Curiosity

A further example of an infinitely divisible point process, whicht might be called a squared Poisson process, arises from the consideration of a point process X whose Λ measure is a constant multiple of the probability measure P' corresponding to a Poisson process of rate λ; thus $\Lambda=kP'$. This is evidently a possible choice for Λ and indeed is one of the most obvious cases where Λ is totally finite.

Let $\mathbf{C}$ be an m-tuple of disjoint Borel sets, the Lebesgue measure of C_i being $|C_i|$, and let $|\mathbf{C}|^{\mathbf{n}}=|C_1|^{n_1}\ldots|C_m|^{n_m}$, $\sum_{i=1}^{m}|C_i| = c$, $\|\mathbf{n}\| = n_1+\ldots+n_m$, $\mathbf{n}! = n_1!\ldots n_m!$. Then the h-functions of $\mathbf{X}(\mathbf{C})$ are given by

$$h(\mathbf{s};\mathbf{C}) = \sum_{\mathbf{n}>0}[1-\exp(-\mathbf{s}\cdot\mathbf{n})]\Lambda\{x(\mathbf{C})=n\} =$$

$$= \sum_{\mathbf{n}>0}\frac{k\lambda^{\|\mathbf{n}\|}}{\mathbf{n}!}|\mathbf{C}|^{\mathbf{n}}e^{-\lambda c}[1-\exp(-\mathbf{s}\cdot\mathbf{n})] =$$

$$= k-k\exp\left(-c\lambda+\lambda\sum_{i=1}^{m}|C_i|e^{-s_i}\right)$$

$$\mathsf{E}\exp(-\mathbf{s}\cdot\mathbf{X}(\mathbf{C}) = \sum_{r=1}^{\infty}\frac{k^r}{r!}e^{-k}\exp\left[-r\lambda\sum_{i=1}^{m}|C_i|(1-e^{-s_i})\right]$$

which is the same as for a Poisson process whose rate is a random variable taking the value $r\lambda$ with probability $(r!)^{-1}k^re^{-k}$.

This example suggests the more general case of a Poisson process whose rate is a non-negative infinitely divisible random variable with an h-function of the form

$$h(r) = \int (1 - e^{-ru})\, L(du) < \infty.$$

The h-function of $X(C)$ then takes the form

$$h(\mathbf{s}; \mathbf{C}) = \sum_{\mathbf{n}>0} (\mathbf{n}!)^{-1} |\mathbf{C}|^{\mathbf{n}} [1 - \exp(-\mathbf{s} \cdot \mathbf{n})] \int u^{||\mathbf{n}||} e^{-uc} L(du)$$

provided that this expression is finite (otherwise no ordinary point process can be defined since infinite measure must be given to unbounded sets with positive probability). The Λ measure corresponding to this is the Stieltjes integral with respect to $L(du)$ of the probability measures corresponding to Poisson processes of rate u.

REFERENCES

[1] KERSTAN, J. and MATTHES, K.: Stationäre zufällige Punktfolgen, II, *Jahresbericht der Deuts. Math. Vereinigung* **66** (1964) 106.

[2] KHINCHIN, A. YA.: *Mathematical Methods in the theory of queueing*, London, Griffin, 1960.

[3] KHINCHIN, A. YA.: Sequences of chance events without after-effects, *Teor. Veroyatnost. i Primenen.* **1** (1956) 3.

[4] KHINCHIN, A. YA.: On Poisson sequences of chance events, *Teor. Veroyatnost. i Primenen.* **1** (1956) 320.

[5] KINGMAN, J. F. C.: Completely random measures, *Pacific. J. Math.* **21** (1967) 59.

[6] LEE, P. M.: A structure theorem for infinitely divisible point processes, Unpublished address to I. A. S. P. S., Berne, 1964.

[7] LEE, P. M.: Infinitely divisible stochastic processes, Cambridge Ph. D. Dissertation, 1966.

[8] LEE, P. M.: Infinitely divisible stochastic processes, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **7** (1967) 147.

[9] MATTHES, K.: Unbeschränkt teilbäre Verteilungegesetze stationärer zufälliger Punktfolgen, *Wissen. Zeits. der Hochschule für Electrotechnik Ilmenau* **9** (1963) 235.

[10] MATTHES, K.: Stationäre zufällige Punktfolgen, I, *Jahresbericht der Deuts. Math. Vereinigung* **66** (1963) 66.

[11] REDHEFFER, R. M.: A note on the Poisson law, *Math. Magazine* **26** (1953) 185.

[12] RÉNYI, A.: Remarks on the Poisson process, *Studia Sci. Math. Hung.* **2** (1967) 119—123.

[13] RYLL-NARDZEWSKI, C.: Remarks on processes of calls, *Proceedings of the IV th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California U. P. 1961. Vol. 2, 455.

[14] RÉNYI, A.: On composed Poisson distributions, II, *Acta Math. Acad. Sci. Hungar.* **2** (1951) 83—98.

Peterhouse, Cambridge

(Received April 25, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 225—241.

ON SOME GENERALIZATION OF A RESTRICTED RANDOM WALK

by

S. G. MOHANTY

1. Introduction

Besides a detailed review work on random sequences given by BARTON and MALLOWS [1], a systematic development of some of the basic questions connected with random sequence has been presented mainly in chapter 3 of FELLER's book [6], by providing the necessary historical background. Of these, results relating to one-dimensional random walk, on the distributions of (i) number of zeros and (ii) number of positive steps (leading to arc sine law) are very significant. In recent years, several authors have made their contributions, by extending and generalizing these problems in various directions. The purpose of this paper is to further extend a few of these results.

Without giving a full account of all recent developments, it is necessary to briefly mention those which are pertinent to the results of this paper. Keeping this in mind, we consider a random walk in which particle initially at the origin, moves at any stage, either one unit (or step) to the right with probability p or one unit to the left with probability $q(=1-p)$. For the symmetric random walk where $p=q=\frac{1}{2}$, CSÁKI and VINCZE have obtained the distribution of the number of steps greater than α (an integer) and the joint distribution of the number of times the particle crosses (or crossings) the origin and the number of positive steps in both cases (*a*) when the particle returns to the origin at the end [3], and (*b*) when it ends somewhere else [4]. Another related result is the distribution of the number of times the particle crosses α, which is given by CSÁKI [2] in case (*b*). In his paper [9], KANWAR SEN has further determined the joint distributions of number of times the particle crosses α and the number of steps greater than α, both for (*b*) and the case (*c*) where the particle ends at a fixed point. While a part of ENGLEBERG's results [5] that coincides with a few in [9], is the derivation of the distribution of the number of times the particle crosses the origin, the other part deals with the determination of the distribution of the number of zeros in case (*c*).

In this paper, we focus our attention to the one-dimensional restricted random walk (to be denoted by $R(m, n; \mu)$), where a particle initially at the origin, moves at any stage either one step to the right or μ (a positive integer) steps to the left, and reaches the point $m-\mu n$ in $m+n$ steps. It may be noted that no assumption is made regarding the probabilities of moving to the right or to the left, because our interest is limited to the derivation of the conditional distributions, subject to the last restriction (namely, that the particle reaches the fixed point $m-\mu n$ at the end of $m+n$ steps). The generalized case with μ, has already been dealt by TAKÁCS, in obtaining the distribution of number of positive steps for $m>\mu n$ in [12] and for general m and n in [13]. Here, though our aim is mainly to treat $R(m, n; \mu)$, for presenting results, similar to those mentioned in the last pragraph, the ensuing discussion is very helpful to view the problems from a wider perspective.

2. Preliminaries

One schematic way of dealing with problems on the random walk is to represent each movement of the particle to the right or to the left by a horizontal unit or a vertical unit, so that in the new system, the restricted random walk in $R(m, n; \mu)$ corresponds to the minimal lattice paths the particle describes from the origin to (m, n). Any such path can be represented by a non-negative nondecreasing vector $(a_1, \ldots, a_n)$, $0 \leqq a_1 \leqq \ldots \leqq a_n \leqq m$, where a_i is the horizontal distance of the path from$(m, n-i)$, $i=1, \ldots, n$ (see [8]). Let $A=(a_i, \ldots, a_n)$ and $B=(b_1, \ldots, b_n)$ be two paths subject to the restriction $a_i \geqq b_i$ for all i. Denote by $N(A)=N(a_1, \ldots, a_n)$, the number of paths from $(0, 0)$ to (m, n), not crossing A. Then, it can be readily seen that

$$
\begin{aligned}
N(A) &= 1 \cdot N(a_1-b_1, a_2-b_1-1, \ldots, a_n-b_1-1) \\
&\quad + N(b_1) \cdot N(a_2-b_2-1, a_3-b_2-1, \ldots, a_n-b_2-1) \\
&\quad + N(b_1, b_2) \cdot N(a_3-b_3-1, \ldots, a_n-b_3-1) \\
&\quad + \ldots \\
&\quad + N(b_1, \ldots, b_{n-1}) \cdot N(a_n-b_n-1) \\
&\quad + N(b_1, \ldots, b_n) \cdot 1,
\end{aligned} \tag{1}
$$

where $N(A)=0$ if any $a_i<0$. By setting $a_1=a_2=\ldots=a_n=m$ and $b_n=m$ in (1), we can express the following recurrence relation (see [10]):

$$
\begin{aligned}
N(b_1, \ldots, b_n) = \binom{m+n}{n} - \sum_{i=1}^{n-2} \binom{m+n-b_{i+1}-i-1}{n-i} N(b_1, \ldots, b_i) - \\
- \binom{m+n-b_1-1}{n}.
\end{aligned} \tag{2}
$$

From (2) it can be shown without difficulty that

$$
N(b_1, \ldots, b_n) = \begin{vmatrix}
\binom{m+n}{n} & \binom{m-b_{n-1}+1}{2} & \binom{m-b_{n-2}+2}{3} & \ldots & \binom{m-b_1+n-1}{n} \\
\binom{b_{n-2}+n-2}{n-2} & 1 & \binom{b_{n-2}-b_{n-2}}{1} & \ldots & \binom{b_{n-2}-b_1+n-3}{n-2} \\
\binom{b_{n-3}+n-3}{n-3} & 0 & 1 & \ldots & \binom{b_{n-3}-b_1+n-4}{n-3} \\
\vdots & \vdots & \vdots & \ldots & \vdots \\
1 & 0 & 0 & \ldots & 1
\end{vmatrix} \tag{3}
$$

with $N(b_1) = \binom{m+1}{1}$.

Again, as a special case, letting $a_1 = m - \mu n\ (m > \mu n)$, $a_{i+1} = a_i + \mu$, $i = 1, \ldots, n-1$, we see that $N(A)$ gives the number of paths from $(0, 0)$ to (m, n), not crossing the line $x = \mu y$, the expression for which is known [8], [11], to be

$$\frac{m - \mu n + 1}{m + n + 1} \binom{m + n + 1}{n}. \tag{4}$$

(It can be observed that (4) is also the expression for the number of paths from $(0, 0)$ to $(m+1, n)$, not touching the line $x = \mu y$.) Furthermore, if we set $b_{i+1} = b_i + \mu$, $i = 1, \ldots, n-1$, the special case of (1) gives rise to the identity

$$\frac{m - \mu n + 1}{m + n + 1} \binom{m + n + 1}{n} = \sum_{k=0}^{n} \frac{b_1 + 1}{b_1 + (\mu + 1)k + 1} \binom{b_1 + (\mu + 1)k + 1}{k} \cdot$$
$$\cdot \frac{m - \mu n - b_1}{m - \mu n - b_1 + (\mu + 1)(n - k)} \binom{m - \mu n - b_1 + (\mu + 1)(n_k)}{n - k} \tag{5}$$

which in another form, becomes

$$\frac{a}{a + (\mu + 1)n} \binom{a + (\mu + 1)n}{n} = \sum_{k=0}^{n} \frac{b}{b + (\mu + 1)k} \binom{b + (\mu + 1)k}{k} \cdot$$
$$\cdot \frac{a - b}{a - b + (\mu + 1)(n - k)} \binom{a - b + (\mu + 1)(n - k)}{n - k}. \tag{6}$$

This is similar to VANDERMONDE's convolution formula [7] or CAUCHY's summation formula. In the sequel, the above schematic representation together with (1) (its special cases (4), (5) or (6)) play a vital role.

Because of the above discussion we have stated results interchangeably either in terms of random walk or that of lattice paths which should not cause any ambiguity. A final remark is that it is obviously known that the total number of paths from $(0, 0)$ to (m, n) is $\binom{m+n}{n}$ and therefore the probability of any event under consideration would only involve the calculation of the number of paths corresponding to the particular event, which would concern the rest part of the paper.

3. Distribution of Touches

Consider a boundary defined by the path $C = (c_1, \ldots, c_n)$ where $c_n = m$. We say that a path $A = (a_1, \ldots, a_n)$ touches C at $(m - c_j, n - j + 1)$, $j = 1, \ldots, n$, when $a_j = c_j$ and $a_i < c_i$ for all $i \neq j$. Let $(m - c_j, n - j + 1)$ $j = 1, \ldots, n$ be called the jth node of C. Denote by $D(r, s; C)$ the number of paths touching C exactly at r $(r = 1, \ldots, s)$ nodes among the first s nodes $(s = 1, \ldots, n)$ of C.

THEOREM 1

$$D(r,s;C)=\begin{cases}N(c_{r+1}-1,\ldots,c_s-1,\ c_{s+1},\ldots,c_n)-N(c_r-1,\ldots,c_s-1,\ c_{s+2},\ldots,c_n), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{when} \quad r<s<n-1; \\ N(c_{r+1}-1,\ldots,c_{n-1}-1,c_n)-N(c_{r-1},\ldots,c_{n-1}-1), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{when} \quad r<s=n-1; \\ N(c_{r+1}-1,\ldots,c_n-1), \quad \text{when} \quad r<s=n; \\ N(c_{r+1}-c_r,\ c_{r+2}-c_r,\ldots,c_n-c_r), \quad \text{when} \quad r=s<n; \\ 1 \quad \text{when} \quad r=s=n.\end{cases} \tag{7}$$

PROOF. Trivially, the last two expressions are true. Now we shall check (7) for $r=1<s$. Clearly,

$$\begin{aligned} D(1,s;C) &= 1\cdot N(c_2-c_1-1,\ldots,c_s-c_1-1,\ c_{s+1}-c_1,\ldots,c_n-c_1) \\ &\quad +N(c_1-1)\cdot N(c_3-c_2-1,\ldots,c_s-c_2-1,\ c_{s+1}-c_2,\ldots,c_n-c_2) \\ &\quad +N(c_1-1,c_2-1)\cdot N(c_4-c_3-1,\ldots,c_s-c_3-1,\ c_{s+1}-c_3,\ldots,c_n-c_3) \\ &\quad +\ldots \\ &\quad +N(c_1-1,\ldots,c_{s-1}-1)\cdot N(c_{s+1}-c_s,\ldots,c_n-c_s), \end{aligned} \tag{8}$$

where the second factor of the last term on the right hand side is equal to 1 if $s=n$. Substituting $a_i=c_{i+1}-1$ $(1\leqq i\leqq s-1)$, $a_s=c_{s+1}$, $b_i=c_i-1$ $(1\leqq i\leqq s)$ and $a_i=b_i=c_{i+1}$ $(s+1\leqq i\leqq n-1)$ in (1), we obtain,

$$D(1,s;C)=\begin{cases}N(c_2-1,\ldots,c_s-1,\ c_{s+1},\ldots,c_n)-N(c_1-1,\ldots,c_s-1,\ c_{s+2},\ldots,c_n), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for} \quad s<n-1; \\ N(c_2-1,\ldots,c_{n-1},c_n)-N(c_1-1,\ldots,c_{n-1}-1), \quad \text{for} \quad s=n-1; \\ N(c_2-1,\ldots,c_n-1), \quad \text{for} \quad s=n.\end{cases}$$

By induction hypothesis, we can show that,

$$\begin{aligned} D(r,s;C) &= 1\cdot N(c_{r+1}-c_r-1,\ldots,c_s-c_r-1,\ c_{s+1}-c_r,\ldots,c_n-c_r) \\ &\quad +\sum_{i=r}^{s-2} N(c_r-1,\ldots,c_i-1)\cdot N(c_{i+2}-c_{i+1}-1,\ldots,c_s-c_{i+1}-1, \\ &\quad c_s-c_{i+1},\ldots,c_n-c_{i+1})+ \\ &\quad +N(c_r-1,\ldots,c_{s-1}-1)\cdot N(c_{s+1}-c_s,\ldots,c_n-c_s), \end{aligned} \tag{9}$$

where the middle summation is zero if $s=r+1$. Again, the application of (1) for $a_i=c_{i+r}-1$ $(1\leqq i\leqq s-r)$, $a_{s-r+1}=c_{s+1}$, $b_i=b_{i+r-1}-1$ $(1\leqq i\leqq s-r+1)$ and $a_i=b_i=c_{i+r}$ $(s-r+2\leqq i\leqq n-r)$, simplifies (9) to (7). This completes the proof for Theorem 1.

Results similar to (7) may be established for any consecutive s nodes of C.

An interesting interpretation of Theorem 1 in terms of a generalized random walk is as follows:

In a random walk, the particle starting from the origin moves at any stage, either a unit to the right or $c_{i+1}-c_i$ $(i=1, \ldots, n-1)$ units to the left, if the particle has moved $(i-1)$ times to the left preceeding to the present one. By duality, it can be shown that $D(r, s; C)$ is equal to the number of ways in which the particle reaches (but never crosses) c_1 exactly r times, in $m+n$ steps. Note that the particle to stay at a given point is implied by $c_{i+1}-c_i = 0$.

Now, let us put $c_i = m-\mu(s-1)$, $c_i = c_1+(i-1)\mu$, $2 \leqq i \leqq s$, $c_s = c_{s+1} = \ldots = c_n$ and finally $t = n-s+1$. Then with the help of Theorem 2 in [8], we get the following corollary.

COROLLARY 1 (i). The number of paths from $(0, 0)$ to (m, n), touching the line $x=\mu(y-t)$, exactly r times is given by

$$\sum_{k=0}^{t-1} \frac{m-\mu(n-t-r)}{m+n-(\mu+1)k-r+\mu t} \binom{m+n-(\mu+1)k-r+\mu t}{n-r-k} \binom{(\mu+1)k-\mu t}{k}$$

$$-\sum_{k=0}^{t-2} \frac{m-\mu(n-t-r+1)}{m+n-(\mu+1)k-r+\mu(t-1)} \cdot \tag{10}$$

$$\cdot \binom{m+n-(\mu+1)k-r+\mu(t-1)}{n-r-k} \binom{(\mu+1)k-\mu(t-1)}{k}.$$

Expression (10) is also the same for the number of paths from $(0, 0)$ to (m, n), touching the line $x=m+\mu(y-n+t)$, r times.

COROLLARY 1 (ii). The number of paths from $(0, 0)$ to (m, n) touching the line $x=\mu y$, exactly r times is

$$\frac{m-\mu(n-r)}{m+n-r} \binom{m+n-r}{n-r}, \qquad m > \mu n. \tag{11}$$

PROOF. This follows from (7) and (4).

COROLLARY 1 (iii). The number of paths from $(0, 0)$ to (m, n) touching the line $x=\mu y$, at least r times is

$$\frac{m+\mu+1-\mu(n-r+1)}{m+n-r+1} \binom{m+n-r+1}{n-r}. \tag{12}$$

PROOF. It can be checked by simple induction.

COROLLARY 1 (iv). The number of paths from $(0, 0)$ to (m, n) touching the line $x = y-t$, exactly r times is

$$\frac{m-n+2t+r-1}{m+n-r+1} \binom{m+n-r+1}{n-t-r+1}. \tag{13}$$

(This has been obtained in [5] for $t=0$.)

PROOF. Putting $\mu=1$ in (10), we get

$$\sum_{k=0}^{t-1}\frac{m-n+t+r}{m+n+t-r-2k}\binom{m+n+t-r-2k}{n-r-k}\binom{2k-t}{k}$$

$$-\sum_{k=0}^{t-2}\frac{m-n+t+r-1}{m+n+t-r-2k-1}\binom{m+n+t-r-2k-1}{n-r-k}\binom{2k-t+1}{k},$$

where the two sums can be written, with the help of (11) in [7], as

$$\binom{m+n-r}{n-r}-\binom{m+n-r}{n-r-t}$$

and

$$\binom{m+n-r}{n-r}-\binom{m+n-r}{n-r-t+1}$$

respectively. From this, (13) follows immediately.

The interpretation of these corollaries in the context of the random walk is quite obvious.

Next, we would be interested in finding the number of paths which touch the given boundary C at least r times. Even though results analogous to (7) can be derived without much difficulty, we mention only for the particular case when $c_s=c_{s+1}=\ldots=c_n$ and $r\leqq s$. (Corollary 1 (iii) typifies one such result.)

From the fact that

$$\sum_{k=r}^{s-1}N(c_{k+1}-1,\ldots,c_s-1,\underbrace{c_s,\ldots,c_s}_{n-s})+\sum_{k=s}^{n-1}N(\underbrace{c_s,\ldots,c_s}_{n-k})+1=$$

$$=N(c_{r+1},\ldots,c_s,\underbrace{c_s+1,\ldots,c_s+1}_{n-s}),$$

$$\sum_{k=r}^{s-1}N(c_{k-1},\ldots,c_s-1,\underbrace{c_s,\ldots,c_s}_{n-s-1})+\sum_{k=s}^{n-1}N(\underbrace{c_s,\ldots,c_s}_{n-k})+1=$$

$$=N(c_r,\ldots,c_s,\underbrace{c_s+1,\ldots,c_{s+1}}_{n-s-1}),$$

and by the application of Theorem 1, we get

$$\sum_{k=r}^{s}D(k,s;C)=N(c_{r+1},\ldots,c_s,\underbrace{c_s+1,\ldots,c_s+1}_{n-s})- \tag{14}$$

$$-N(c_r,\ldots,c_s,\underbrace{c_s+1,\ldots,c_s+1}_{n-s-1}).$$

Clearly $D(s;\ C) = \sum_{k=1}^{s} D(k,\ s;\ C)$ represents the number of ways in which the particle describing the random walk corresponding to Theorem 1, reaches c_1 (but never crosses it).

It is not difficult to realise that (1) cannot be utilized fruitfully to answer every type of problem associated with the random walk in the general case and therefore, we shall henceforth limit our discussions only to $R(m, n;\ \mu)$.

4. Distribution of Crossings

Let $N_r\ (m, n;\ \mu,\ \alpha)$ denote the number of cases in which the particle in $R(m, n;\ \mu)$ crosses (may or may not reach) a given point $\alpha \geqq 0$, r times. $(N_r(m, n;\ \mu,\ \alpha)$ also represents the number of paths from $(0, 0)$ to (m, n), each of which crosses the line $x = \mu y + \alpha$, r times). Furthermore, let $N_r(m, n;\ \mu,\ 0)_i\ i = 1, 2$, be the number of cases where the particle in $R(\mu n, n;\ \mu)$ crosses the origin r times, given that the first step is to the left or to the right according as $i = 1$ or 2.

THEOREM 2

(a) *For* $i = 1, 2$,

$$
(15)\qquad N_r(\mu n, n;\ \mu, 0)_i = \begin{cases} \sum_{k=0}^{j} (-1)^k \mu^{j-k} \dfrac{(\mu+2)(j+1)-k-1}{(\mu+1)n+j-k} \begin{pmatrix} (\mu+1)n+j-k \\ n-j-1 \end{pmatrix} \begin{pmatrix} j \\ k \end{pmatrix}, \\ \qquad\qquad \text{when} \quad r = 2j,\ j = 0, 1, \ldots; \\ \sum_{k=0}^{j+i-1} (-1)^k \mu^{j-k+i-1} \left(\dfrac{(\mu+2)(j+2)-k-(i-1)\mu-2}{(\mu+1)n+j+i-k-1} \right) \cdot \\ \cdot \begin{pmatrix} (\mu+1)n+j+i-k-1 \\ n+i-j-3 \end{pmatrix} \begin{pmatrix} j+i-1 \\ k \end{pmatrix} \quad \text{when} \quad r = 2j+1, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 0, 1, \ldots . \end{cases}
$$

(b) *For* $\alpha = 0$ *and* $m > \mu n$,

$$
(16)\qquad N_r(m, n;\ \mu, 0) = \begin{cases} \sum_{k=0}^{j} (-1)^k \mu^{j-k} \dfrac{m-\mu(n-j-1)+2j+2-k}{m+n+j+1-k} \begin{pmatrix} m+n+j+1-k \\ n-j-1 \end{pmatrix} \begin{pmatrix} j \\ k \end{pmatrix} \\ \qquad\qquad \text{when} \quad r = 2j+1,\ j = 0, 1, \ldots . \\ \sum_{k=0}^{j} (-1)^k \mu^{j-k} \dfrac{m-\mu(n-j)+2j+1-k}{m+n+j+1-k} \begin{pmatrix} m+n+j+1-k \\ n-j \end{pmatrix} \begin{pmatrix} j \\ k \end{pmatrix} \\ \qquad\qquad \text{when} \quad r = 2j,\ j = 0, 1, \ldots . \end{cases}
$$

(c) *For* $\alpha > 0$ *and* $m > \mu n + \alpha$,

$$
(17)\qquad N_r(m, n;\ \mu, \alpha) = \begin{cases} \sum_{k=0}^{j} (-1)^k \mu^{j-k} \dfrac{m-\mu(n-j)+2j+2-k}{m+n+j+2-k} \begin{pmatrix} m+n+j+2-k \\ n-j \end{pmatrix} \begin{pmatrix} j \\ k \end{pmatrix} \\ \qquad\qquad \text{when} \quad r = 2j+1,\ j = 0, 1, \ldots; \\ 0 \qquad\qquad \text{otherwise} \end{cases}
$$

Before proceeding further, we present below another result in the form of a lemma, which would be quite helpful in proving the above theorem.

LEMMA. *The number of ways in which the particle starting from the origin, returns to the origin in* $2n$ *steps such that*

(i) *the first step is to the right,*

and

(ii) *it crosses the origin only once, from the position* β $(0<\beta<\mu)$,

is given by

$$\frac{\mu+2}{(\mu+1)n+1}\binom{(\mu+1)n+1}{n-1} \tag{18}$$

if the particle is allowed to reach the origin before the end, and by

$$\frac{\mu}{(\mu+1)n-1}\binom{(\mu+1)n-1}{n-1} \tag{19}$$

if the particle is not allowed to reach the origin except at the end. [In terms of lattice paths, (18) is equal to the number of paths from $(0, 0)$ to $(\mu n, n)$ such that (i) the first step is a horizontal unit and (ii) each path crosses the line $x=\mu y$ once from one of the positions $\{(\mu y+\beta, y): 0\leqq y\leqq n-1, 0<\beta<\mu\}$ and (iii) it may touch the line $x=\mu y$. A typical path of this nature is shown in the Figure 1.]

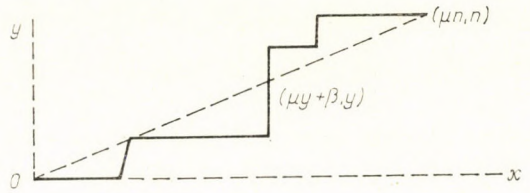


Fig. 1

PROOF. It is evident from Figure 1 and (4) that the desired number in the first part is expressed as

$$\sum_{y=0}^{n-1}\frac{\beta+1}{(\mu+1)y+\beta+1}\binom{(\mu+1)y+\beta+1}{y}\frac{\mu-\beta+1}{(\mu+1)(n-y-1)+\mu-\beta+1}\cdot$$
$$\cdot\binom{(\mu+1)(n-y-1)+\mu-\beta+1}{n-y-1}$$

whereas the number in the second part as

$$\sum_{y=0}^{n-1}\frac{\beta}{(\mu+1)y+\beta}\binom{(\mu+1)y+\beta}{y}\frac{\mu-\beta}{(\mu+1)(n-y-1)+\mu-\beta}\cdot\binom{(\mu+1)(n-y-1)+\mu-\beta}{n-y-1}$$

which with the help of (5) (or (6)) simplify to (18) and (19) respectively.

We offer a few remarks on the lemma. Interesting enough is that (18) and (19) are independent of β. However, this property in general, does not exist for the wider class of problems discussed in the previous section. Also for $\beta = \mu$, it is not difficult to obtain the expression corresponding to (18) as

$$(20) \qquad \frac{\mu+2}{(\mu+1)n+1}\binom{(\mu+1)n+1}{n-1} - \frac{1}{(\mu+1)n+1}\binom{(\mu+1)n+1}{n}.$$

Because of these, we would be in a position to obtain quite a few results in closed form. While application of (18) and (20) will be made in this section, (19) will be needed in the next section.

PROOF OF THEOREM 2. Firstly, we proceed to prove (a). Besides the fact that the above lemma leads directly to the expression for $N_1(\mu n, n; \mu, 0)_2$, it is also known that

$$(21) \qquad N_{2j+1}(\mu n, n; \mu, 0)_2 = \sum_{y=1}^{n-j} N_1(\mu y, y; \mu, 0)_2 \cdot N_{2j-1}(\mu(n-y), n-y; \mu, 0)_2.$$

By induction, we get

$$N_{2j+1}(\mu n, n; \mu, 0) = \sum_{y=1}^{n-j}\left[\frac{\mu(\mu+2)}{(\mu+1)y+1}\binom{(\mu+1)y+1}{y-1} - \frac{\mu+1}{(\mu+1)y}\binom{(\mu+1)y}{y-1}\right]$$

$$\left[\sum_{k=0}^{j}(-1)^k\mu^{j-k}\frac{(\mu+2)j-k}{(\mu+1)(n-y)+j-k}\binom{(\mu+1)(n-y)+j-k}{n-y-j}\binom{j}{k}\right] =$$

$$= \sum_{k=0}^{j}(-1)^k\mu^{j-k}\binom{j}{k}\sum_{y=0}^{n-j-1}\frac{(\mu+2)j-k}{(\mu+1)(n-y-1)+j-k}\binom{(\mu+1)(n-y-1)+j-k}{n-y-j-1}$$

$$\left[\frac{\mu(\mu+2)}{(\mu+1)(y+1)+1}\binom{(\mu+1)(y+1)+1}{y} - \frac{\mu+1}{(\mu+1)(y+1)}\binom{(\mu+1)(y+1)}{y}\right] =$$

$$= \sum_{k=0}^{j}(-1)^k\mu^{j-k}\binom{j}{k}\left[\frac{\mu((\mu+2)(j+1)-k)}{(\mu+1)n+j+1-k}\binom{(\mu+1)n+j+1-k}{n-j-1} - \right.$$

$$\left. - \frac{(\mu+1)(j+1)+j-k}{(\mu+1)n+j-k}\binom{(\mu+1)n+j-k}{n-j-1}\right] =$$

by (6)

$$= \mu^{j+1}\frac{(\mu+2)(j+1)-k}{(\mu+1)n+j+1}\binom{(\mu+1)n+j+1}{n-j-1} + (-1)^{j+1}\frac{(\mu+1)(j+1)}{(\mu+1)n}\binom{(\mu+1)n}{n-j-1} +$$

$$+ \sum_{k=1}^{j}(-1)^k\mu^{j-k+1}\binom{j}{k}\frac{(\mu+2)(j+1)-k}{(\mu+1)n+j+1-k}\binom{(\mu+1)n+j+1-k}{n-j-1} +$$

$$+ \sum_{k=0}^{j-1}(-1)^{k+1}\mu^{j-k}\binom{j}{k}\frac{(\mu+1)(j+1)+j-k}{(\mu+1)+j-k}\binom{(\mu+1)n+j-k}{n-j-1}.$$

But the last two terms can be combined to give,

$$\sum_{k=1}^{j}(-1)^k\mu^{j-k+1}\binom{j+1}{k}\frac{(\mu+2)(j+1)-k}{(\mu+1)n+j+1-k}\binom{(\mu+1)n+j+1-k}{n-j-1}$$

and thus the desired expression for $N_{2j+1}(\mu n, n; \mu, 0)_2$ is derived.

Using induction again, we get

$$(22)\qquad N_{2j}(\mu n, n; \mu, 0)_2 = \sum_{y=r}^{n-1} N_{2j-1}(\mu y, y; \mu, 0)_2 \cdot N_0(\mu(n-y), n-y; \mu, 0)_2 =$$

$$= \sum_{y=j}^{n-1}\left[\sum_{k=0}^{j}(-1)^k\mu^{j-k}\frac{(\mu+2)j-k}{(\mu+1)y+j-k}\binom{(\mu+1)y+j-k}{y-j}\binom{j}{k}\right]\frac{\mu+1}{(\mu+1)(n-y)}\cdot$$

$$\cdot\binom{(\mu+1)(n-y)}{n-y-1} = \sum_{k=0}^{j}(-1)^k\mu^{j-k}\binom{j}{k}\sum_{y=0}^{n-j-1}\frac{(\mu+2)j-k}{(\mu+1)(y+j)+j-k}\cdot$$

$$\cdot\binom{(\mu+1)(y+j)+j-k}{y}\frac{\mu+1}{(\mu+1)(n-y-j)}\binom{(\mu+1)(n-y-j)}{n-y-j-1}$$

which by (6) yields the required result.

Obviously,

$$N_{2j}(\mu n, n; \mu, 0)_2 = N_{2j}(\mu n, n; \mu, 0)_1 .$$

Finally, expression for $N_{2j+1}(\mu n, n; \mu, 0)_1$, is obtainable from

$$(23)\qquad N_{2j+1}(\mu n, n; \mu, 0)_1 = \sum_{y=1}^{n-j-1} N_0(\mu y, y; \mu, 0)_1 \cdot N_{2j}(\mu(n-y), n-y; \mu, 0)_2$$

and (6).

Furthermore, for $m > \mu n + \alpha$ and $\alpha \geqq 0$, we have

$$N_0(m, n; \mu, 0) = \frac{m-\mu n+1}{m+n+1}\binom{m+n+1}{n};$$

$$N_{2j}(m, n; \mu, 0) = \sum_{y=j}^{n} N_{2j-1}(\mu y, y; \mu, 0)_2 \cdot N_0(m-\mu y; n-y; \mu, 0), \qquad j = 1, 2, \ldots;$$

$$N_{2j+1}(m, n; \mu, 0) = \sum_{y=1}^{n-j} N_0(\mu y, y; \mu, 0)_1 \cdot N_{2j}(m-\mu y, n-y; \mu, 0), \qquad j = 0, 1, \ldots;$$

$$N_{2j+1}(m, n; \mu, \alpha) = \sum_{y=0}^{n-j} N_0(\mu y+\alpha, y; \mu, 0) \cdot N_{2j}(m-\mu y-\alpha, n-y; \mu, 0),$$

$$\alpha > 0,\ j = 0, 1, \ldots .$$

An inductive argument with the aid of (6) would as above, prove (b) and (c).

When $\mu=1$, (17) becomes

$$N_{2j+1}(m,n;1,\alpha)=\sum_{k=0}^{j}(-1)^k\binom{j}{k}\frac{m-n+3j+2-k}{m+n+j+2-k}\binom{m+n+j+2-k}{n-j}=$$

$$=\sum_{k=0}^{j-1}(-1)^k\binom{j-1}{k}\left[\frac{m-n+3j+2-k}{m+n+j+2-k}\binom{m+n+j+2-k}{n-j}-\right.$$

$$\left.-\frac{m-n+3j+1-k}{m+n+j+1-k}\binom{m+n+j+1-k}{n-j}\right]=\sum_{k=0}^{j-1}(-1)^k\binom{j-1}{k}\frac{m-n+3j+3-k}{m+n+j+1-k}\cdot$$

$$\cdot\binom{m+n+j+1-k}{n-j-1}.$$

Repeating this kind of reduction, we obtain

$$N_{2j+1}(m,n;1,\alpha)=\frac{m-n+4j+2}{m+n+2}\binom{m+n+2}{n-2j}.$$

Similar simplifications of (15) and (16) are possible. A corollary to this effect is given below.

COROLLARY 2.

(a) $N_r(\mu n,n;1,0)=N_r(\mu n,n;1,0)+N_r(\mu n,n;1,0)=\dfrac{2(r+1)}{n}\binom{2n}{n-r-1}$;

(b) For $m>n$,

$$N_r(m,n;1,0)=\frac{m-n+2r+1}{m+n+1}\binom{m+n+1}{n-r};$$

(c) For $\alpha>0$ and $m>n+\alpha$,

$$N_r(m,n;1,\alpha)=\begin{cases}\dfrac{m-n+2r}{m+n+2}\binom{m+n+2}{n-r+1} & \text{when } r \text{ is odd;}\\ 0 & \text{otherwise.}\end{cases}$$

Note that ENGLEBERG's Theorem 2.1 [5] checks with corollary 2(a) and (b), whereas SEN's Theorem 1.1 and Theorem 1.2 [9] are the same as corollary 2(b) and (c) respectively.

5. Distribution of Arrivals

In this section, we would be concerned with the number of cases in which the particle reaches a given point, a fixed number of times. To be more specific, let $N_r^*(m,n;\mu,\alpha)$ denote the number of cases in which the particle in $R(m,n;\mu)$ reaches $\alpha\geqq 0$, r times. By convention, in counting for r, we include the starting point (that is, the origin) in all cases belonging to $N_r^*(m,n;\mu,0)$ or to $N_r^*(\mu n,n;\mu,0)$ and include the end point in all cases belonging to $N_r^*(\mu n,n;\mu,0)$ or $N_r^*(\mu n+\alpha,n;\mu,\alpha)$.

THEOREM 3. *For* $\alpha \geqq 0$, *and* $m \geqq \mu n + \alpha$,

$$N^*_{r+1}(m, n; \mu, \alpha) = \begin{cases} (\mu+1)^r \dfrac{m-\mu(n-r)}{m+n-r} \binom{m+n=r}{n-r}, & r = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

(Evidently, $N^*_r(m, n; \mu, \alpha)$ is the number of paths from $(0, 0)$ to (m, n), such that each path reaches r points among $\{(\mu y + \alpha, y): y = 0, 1, \ldots, n\}$).

PROOF. Consider the case when $\alpha = 0$, and $m = \mu n$. Trivially (24) is true for $r = 0$. A moment's reflection shows that the particle can reach the origin once (i) either by moving first to the right, then crossing the origin once from the point $\beta (0 < \beta < \mu)$ and finally reaching the origin, (ii) or by reaching the origin at the end, without crossing it earlier. It follows from the lemma of the previous section that the number of cases in (i) is given by

$$\frac{(\mu-1)\mu}{(\mu+1)n-1} \binom{(\mu+1)n-1}{n-1},$$

and that in (ii) by

$$\frac{2\mu}{(\mu+1)n-1} \binom{(\mu+1)n-1}{n-1}.$$

This fact checks (24) for $r = 1$. Moreover,

$$N^*_{r+1}(\mu n, n; \mu, 0) = \sum_{y=1}^{n-r+1} N^*_2(\mu y, y; \mu, 0) \cdot N^*_r(\mu(n-y), n-y; \mu, 0), \tag{25}$$

for $r \geqq 2$.

Induction and relation (25) would give rise to (24). Finally the proof is complete, when we observe the following:

$$N^*_1(m, n; \mu, 0) = \frac{m-\mu n}{m+n} \binom{m+n}{n}, \qquad m > \mu n;$$

$$N^*_{r+1}(m, n; \mu, 0) = \sum_{y=1}^{n-r+1} N^*_2(\mu y, y; \mu, 0) \cdot N^*_r(m-\mu y, n-y; \mu, 0), \qquad m > \mu n, \ r \geqq 1;$$

$$N^*_{r+1}(\mu n + \alpha, n; \mu, \alpha) = N^*_{r+1}(\mu n + \alpha, n; \mu, 0);$$

and

$$N^*_{r+1}(m, n; \mu, \alpha) = \sum_{y=0}^{n-r} N^*_1(\mu n + \alpha, y; \mu, 0) \cdot N^*_{r+1}(m - \mu y - \alpha, n-y; \mu, 0)$$

$$m > \mu n + \alpha, \ \alpha > 0, \ r \geqq 0.$$

As it is seen, special case of Theorem 4, with $\alpha = 0$ and $\mu = 1$, becomes Theorem 3.1 in [5].

6. Joint Distribution of the Number of Times and the Number of Steps in α^+

First, we define a set α^+ (α^-) to consist of all points greater (less) than α. We say that the particle is in α^+, if it has neither reached nor crossed α from the right after its arrival in α^+. A step in α^+ implies that the destination point is in α^+.

Next, let us denote by $M_r(m, n; \mu, \alpha, s)$ the number of cases in $R(m, n; \mu)$ such that:

(a) the particle moves to α^+ as soon as it reaches α from α^-;

(b) it reaches r times the point $\alpha+1$, from $(\alpha+1)^-$;

(c) in $(\alpha+1)^+$, it moves s times to the left;

(d) after its i th arrival at $\alpha+1$ from $(\alpha+1)^-$, it reaches or crosses α for the first time from the point $\alpha+v_i$ ($v_i=1, \ldots, \mu$; $i=1, \ldots, r-1$) in one step.

THEOREM 4. *For* $1 \leqq r \leqq n$, $m \geqq \mu n+\alpha$, $\alpha \geqq 0$, $0 \leqq s \leqq n-r$ *and* $1 \leqq v_i \leqq \mu$,

$$M_{r+1}(m, n; \mu, \alpha, s) = \left[\frac{m-\mu n+v-\alpha}{(\mu+1)s+m-\mu n+v-\alpha}\binom{(\mu+1)s+m-\mu n+v-\alpha}{s}\right] \tag{26}$$

$$\left[\frac{\mu r+\alpha-v}{(\mu+1)(n-s)+\alpha-v-r}\binom{(\mu+1)(n-s)+\alpha-v-r}{n-r-s}\right]$$

where $v = v_1+\ldots+v_r$.

PROOF. Let $M_r^*(m, n; \mu, 0, s)$ denote the number of cases in $R(m, n; \mu)$ each of which in addition to (a), (b), (c), (d) above with $\alpha=0$, satisfies that (e) the particle in the first step, moves to 0^+. We assert that

$$M_r^*(\mu n, n; \mu, 0, s) = \tag{27}$$

$$= \frac{v}{(\mu+1)s+v}\binom{(\mu+1)s+v}{s}\frac{\mu r-v}{(\mu+1)(n-s)-v-r}\binom{(\mu+1)(n-s)-v-r}{n-r-s}$$

and

$$M_{r+1}^*(m, n; \mu, 0, s) = \tag{28}$$

$$= \frac{m-\mu n+v}{(\mu+1)s+m-\mu n+v}\binom{(\mu+1)s+m-\mu n+v}{s}\frac{\mu r-v}{(\mu+1)(n-s)-v-r}\cdot$$

$$\cdot\binom{(\mu+1)(n-s)-v-r}{n-r-s}.$$

When $r=1$, $\alpha=0$, $m=\mu n$, a typical path corresponding to the random walk in $R(\mu n, n; \mu)$ having restrictions (a), (b), (c) (d) and (e) is given in Figure 2.

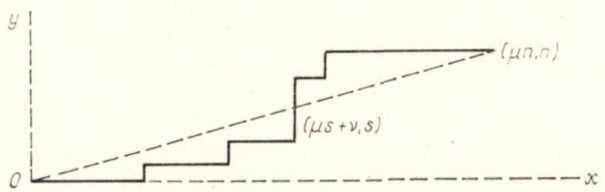


Fig. 2

Clearly, the path is not allowed to touch the line $x=\mu y$, except at the beginning and at the end. Thus

$$M_1^*(\mu n, n; \mu, 0, s) = \frac{v}{(\mu+1)s+v}\binom{(\mu+1)s+v}{s}\frac{\mu-v}{(\mu+1)(n-s)-v-1}\cdot$$
$$\cdot\binom{(\mu+1)(n-s)-v-1}{n-s-1}.$$

But

$$M_r^*(\mu n, n; \mu, 0, s) = \sum_{s_1=0}^{s}\sum_{y=s_1+1}^{n-s+s_1-r+1} M_1(\mu y, y; \mu, 0, s)\cdot$$
$$\cdot M_{r-1}(\mu(n-y), n-y; \mu, 0, s-s_1)$$

which by induction, yields

$$\left[\sum_{s_1=0}^{s}\frac{v_1}{(\mu+1)s_1+v_1}\binom{(\mu+1)s_1+v_1}{s_1}\frac{v-v_1}{(\mu+1)(s-s_1)+v-v_1}\binom{(\mu+1)(s-s_1)+v-v_1}{s-s_1}\right]\cdot$$
$$\cdot\left[\sum_{y=0}^{n-r-s}\frac{\mu-v_1}{(\mu+1)(y+1)-v_1-1}\binom{(\mu+1)(y+1)-v_1-1}{y}\right]\cdot$$
$$\cdot\frac{\mu(r-1)-v+v_1}{(\mu+1)(n-y-s-1)-v+v_1-r+1}\binom{(\mu+1)(n-y-s-1)-v+v_1-r+1}{n-r-s-y}.$$

We obtain (27) by the application of (6) to both sums. Furthermore,

$$M_{r+1}^*(m, n; \mu, 0, s) =$$
$$= \sum_{s_1=0}^{s} M_r(\mu(n-s+s_1), n-s+s_1; \mu, 0, s_1)\frac{m-\mu n}{(\mu+1)(s-s_1)+m-\mu n}\cdot$$
$$\cdot\binom{(\mu+1)(s-s_1)+m-\mu n}{s-s_1}$$

which can also readily verified to be equal to (28). The proof becomes complete when the following relation is observed:

$$M_{r+1}(m, n; \mu, \alpha, s) = \sum_{y=0}^{n-r-s}\frac{\alpha}{(\mu+1)y+\alpha}\cdot$$
$$\cdot\binom{(\mu+1)y+\alpha}{y} M_{r+1}^*(m-\mu y-\alpha, n-y; \mu, 0, s).$$

Suppose $M_{r+1}(m, n; \mu, \alpha, s, v)$ represents the number of cases in $R(m, n; \mu)$ with conditions (a), (b), (c), (d) and (f) that in α^+, the particle moves $m-\mu n-\alpha+$ $+\mu s+v$ steps to the right (i.e. v, which is equal to $\sum_{i=1}^{r} v_i$ is fixed). Theorem 4 suggests that we need to find the distribution of sums of identically distributed independent random variables, each having discrete uniform distribution.

THEOREM 5. *Let $X_1, X_2, \ldots, X_r$ be independent random variables, with the probability distribution as follows:*

$$P(X_i=a)=\frac{1}{\mu},\ a=0,1,\ldots,\mu-1;\ i=1,\ldots,r.$$

Then, the probability distribution of $Y=X_1+\ldots+X_r$ is given by

$$P(Y=a)=\begin{cases}\dfrac{\sum_{j=0}^{k-1}(-1)^j\binom{r}{j}\binom{a+r-1-\mu j}{r-1}}{\mu^r} & \text{when } (k-1)(\mu-1)\leqq a<\\ & <k(\mu-1),\ k=1,\ldots,r;\\ \dfrac{1}{\mu^r} & \text{when } a=r(\mu-1);\\ 0 & \text{otherwise.}\end{cases} \tag{29}$$

PROOF. The result is trivially true for the last part. Also it is obvious that the denominator should be μ^r. Therefore, it remains to verify the expression in the numerator, which is equal to the number of ways of obtaining a as the sum of r integers, each taking values from 0 to $\mu-1$. Denote this number by $(r, a; \mu)$. Then $(r, a; \mu)$ satisfies the following:

$$(1,a;\mu)=\begin{cases}1 & a=0,1,\ldots,\mu-1;\\ 0 & \text{otherwise;}\end{cases} \tag{30}$$

$$(r, r(\mu-1);\mu)=1 \quad \text{for all} \quad r; \tag{31}$$

and

$$(r,a;\mu)=\begin{cases}\sum_{i=a-\mu+1}^{a}(r-1,i;\mu), & r\geqq 1,\ 0\leqq a\leqq r(\mu-1),\\ 0 & \text{otherwise.}\end{cases} \tag{32}$$

Evidently,

$$(2,a;\mu)=\begin{cases}\binom{a+1}{1}, & 0\leqq a<\mu-1,\\ \binom{a+1}{1}-2\binom{a+1-\mu}{1}, & \mu-1\leqq a<2(\mu-1).\end{cases}$$

So using induction, we can write for $(k-1)(\mu-1)\leqq a<k(\mu-1)$,

$$(r,a;\mu)=\sum_{i=a-\mu+1}^{a}(r-1,i;\mu)=$$

$$=\sum_{i=a-\mu+1}^{a}\sum_{j=0}^{k-2}(-1)^j\binom{r-1}{j}\binom{i+r-2-\mu j}{r-2}+$$

$$+(-1)^{k-1}\binom{r-1}{k-1}\sum_{i=(k-1)\mu}^{a}\binom{i+r-2-(k-1)\mu}{r-2}=$$

$$=\sum_{j=0}^{k-2}(-1)^j\binom{r-1}{j}\sum_{i=a-\mu(j+1)+1}^{a-\mu j}\binom{i+r-2}{r-2}+(-1)^{k-1}\binom{r-1}{k-1}\sum_{i=0}^{a-(k-1)\mu}\binom{i+r-2}{r-2}=$$

$$= \sum_{j=0}^{k-2} (-1)^j \binom{r-1}{j} \left[\sum_{i=0}^{a-\mu j} \binom{i+r-2}{r-2} - \sum_{i=0}^{a-\mu(j+1)} \binom{i+r-2}{r-2} \right] +$$

$$+ (-1)^{k-1} \binom{r-1}{k-1} \sum_{i=0}^{a-(k-1)\mu} \binom{i+r-2}{r-2} = \sum_{i=0}^{a} \binom{i+r-2}{r-2} +$$

$$+ \sum_{j=1}^{k-2} (-1)^j \left[\binom{r-1}{j} + \binom{r-1}{j-1} \right] \sum_{i=0}^{a-\mu j} \binom{i+r-2}{r-2} + (-1)^{k-1} \left[\binom{r-1}{k-2} + \binom{r-1}{k-1} \right] \cdot$$

$$\cdot \sum_{i=0}^{a-(k-1)\mu} \binom{i+r-2}{r-2} = \binom{a+r-1}{r-1} + \sum_{j=1}^{k-2} (-1)^j \binom{r}{j} \binom{a+r-1-\mu j}{r-1} +$$

$$+ (-1)^{k-1} \binom{r}{k-1} \binom{a+r-1-\mu(k-1)}{r-1} = \sum_{j=0}^{k-1} (-1)^j \binom{r}{j} \binom{a+r-1-\mu j}{r-1}.$$

Hence the theorem is proved.

Now, it is not difficult to find an expression for $M_{r+1}(m, n; \mu, \alpha, s, v)$ which is stated as a corollary.

COROLLARY 2.

$$M_{r+1}(m, n; \mu, \alpha, s, v) = M_{r+1}(m, n; \mu, \alpha, s) \cdot (r, v-r; \mu). \tag{33}$$

Let $S_r(m, n; \mu, \alpha, J)$ be the number of cases in $R(m, n; \mu)$ such that the particle moves J steps in α^+, $J = m-\mu n-\alpha, \ldots, m+n-\alpha$ and reaches r times the point $\alpha+1$ from the left. When r, α, s and $v (0 \leqq v \leqq \mu r,\ 0 \leqq s \leqq n-r)$ are fixed, the number J of steps in α^+ is known to be equal to $m-\mu n-\alpha+(\mu+1)s+v$. On the other hand, fixing J and r, it can be seen that

$$s_1^* = \max\left(\frac{J-m+\mu n-\alpha-\mu r}{\mu+1}, 0 \right) \leqq s \leqq \frac{J-m+\mu n+\alpha}{\mu+1} = s_2^*.$$

Moreover, for given $J \neq m+n-\alpha$, we get $1 \leqq r \max(m+n-\alpha-J, n) = r^*$, whereas $J = m+n-\alpha$ leads to $r=0$, $s=n$, and $v=0$. Combining these facts, the next result gives the joint distribution of number of times and number of steps in α^+.

COROLLARY 3.

$$S_{r+1}(m, n; \mu, \alpha, J) = \sum_{s_1^* \leqq s \leqq s_2^*} M_{r+1}(m, n; \mu, \alpha, s, v) \tag{34}$$

$$\text{for} \quad J = m-\mu n-\alpha, \ldots, m+n-\alpha-1 \quad \text{and} \quad 1 \leqq r \leqq r^*,$$

and

$$S_r(m, n; \mu, \alpha, m+n-\alpha) = \begin{cases} \dfrac{m-\mu n-\alpha}{m+n-\alpha} \binom{m+n-\alpha}{n}, & r = 0; \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

Corollary 3 is analogous to Theorem 4.1 of [9].

The last part can be shown to be a direct consequence of the definitions.

By summing over r in (34) one can obtain $S(m, n; \mu, \alpha, J)$, the number of cases in $R(m, n; \mu)$ such that the particle moves J steps in α^+. When $\alpha = 0$, $m > \mu n$, an expression for the same is given in [12] as Theorem 1. However, a direct proof of the equality of two expressions is not quite obvious.

We may conclude with the remark that expressions, say for the number of cases in $R(m, n; \mu)$ such that the number of steps in α^+ and the number of times the particle reaches α are fixed, can be derived in a similar manner.

REFERENCES

[1] BARTON, D. E. and MALLOWS, C. L.: Some aspects of the random sequence, *Ann. Math. Statist.* **36** (1965) 236—260.

[2] CSÁKI, E.: On the number of intersections in the one-dimensional random walk, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961) 281—286.

[3] CSÁKI, E. and VINCZE, I.: On some problems connected with the Galton test, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961) 97—109.

[4] CSÁKI, E. and VINCZE, I.: On some distributions connected with the arc-sine law, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** (1963) 281—291.

[5] ENGLEBERG, O.: On some problems concerning a restricted random walk, *J. Appl. Probl.* **2** (1965) 396—404.

[6] FELLER, W.: *An Introduction to Probability Theory and its Applications*, 1 (2nd edition) John Wiley, New York.

[7] GOULD, H.: Some generalizations of Vandermonde's convolution, *Amer. Math. Montly* **58** (1956) 84—91.

[8] MOHANTY, S. G. and NARAYANA, T. V.: Some properties of compositions and their application to probability and statistics, I, *Biometrische Zeitschrieft* **3** (1961) 252—258.

[9] SEN, K.: On some combinatorial relations concerning the symmetric random walk, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **9** (1964) 335—357.

[10] SWITZER, P.: Significance probability bounds for rank orderings, *Ann. Math. Statist*, **35** (1964) 891—894.

[11] TAKÁCS, L.: A generalization of the ballot problem and its application in the theory of queues, *J. Amer. Statist. Assoc.* **57** (1962) 327—337.

[12] TAKÁCS, L.: The distribution of majority times in a ballot, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** (1963) 118—121.

[13] TAKÁCS, L.: Fluctuations in the ratio of scores in counting a ballot, *J. Appl. Prob.* **1** (1964) 393—396.

University of Bonn and Indian Institute of Technology, Delhi

(Received April 26, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 243—248.

ÜBER EINE MITTELWERTABSCHÄTZUNG VON E. LANDAU UND O. TOEPLITZ

von

K. SZILÁRD

Es ist sechzig Jahre her, daß E. LANDAU und O. TOEPLITZ folgende Abschätzung des Absolutwertes der Ableitung einer analytischen Funktion (bzw. Abschätzung der grössten Schwankung einer analytischen Funktion für $|z| \leqq R$ mit Hilfe des Absolutwertes der Ableitung für $z=0$) gefunden haben [1] [1]. Es sei $f(z)$ eine analytische Funktion der komplexen Veränderlichen z, die für $|z| \leqq R$ (also auch auf der Kreislinie $|z|=R$) regulär ist, a und b seien irgendwelche z-Werte für $|z|=R$, dann gilt die Abschätzung

$$|f'(0)| \leqq \operatorname*{Max}_{a,\,b} \frac{|f(b)-f(a)|}{2R} \tag{1}$$

wobei sich die positive Zahl 2 durch keine größere ersetzen läßt. LANDAU und TOEPLITZ haben sogar gezeigt, daß es auf der Kreislinie $|z|=R$ zwei, einander diametral gegenüberliegende Punkte z_1 und $-z_1$ gibt so, daß

$$|f'(0)| \leqq \frac{|f(z_1)-f(-z_1)|}{2R} \tag{2}$$

gilt.

Ich will zeigen erstens, daß die Ungleichung (2) ein Spezialfall einer anderen Ungleichung ist, die sich auf irgendwelche endlich viele Punkte auf der Kreislinie $|z|=R$ bezieht und zweitens, daß im Falle $f'(0) \neq 0$ wir über $f'(0)$ mehr als eine Abschätzung von oben aussagen können. Zuerst will ich eine Verallgemeinerung der Ungleichung (2) herleiten.

SATZ 1. *Es seien $ze^{i\alpha_1}, ze^{i\alpha_2}, \ldots, ze^{i\alpha_n}$ irgendwelche, nicht notwendigerweise von einander verschiedene n Punkte, deren gegenseitige Lage auf der Kreislinie $|z|=R$ durch die reellen Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_n$ bestimmt ist und die Funktion $w=f(z)$ sei für $|z| \leqq R$ stetig und für $|z|<R$ regulär analytisch. Dann läßt sich die komplexe Zahl z_1 vom absoluten Betrage R so wählen (d.h. dieses, als starr gedachtes Punktsystem auf der Kreislinie $|z|=R$ läßt sich um den Punkt $z=0$ als Mittelpunkt so verdrehen), daß die Ungleichung*

$$|f'(0)| \leqq \frac{1}{nR} \left| \sum_{\nu=1}^{n} [f(z_1 e^{i\alpha_\nu}) - f(0)] \cdot e^{-i\alpha_\nu} \right| \tag{3}$$

gilt. Hierbei läßt sich die positive Zahl n durch keine grössere ersetzen. (Setzt man $n=2$ und $\alpha_1=0$, $\alpha_2=\pi$ und nimmt man an, daß $f(z)$ für $|z| \leqq R$ regulär sei, so geht die Ungleichung (3) in die Ungleichung (2) über.)

[1] Die Ziffern in den eckigen Klammern beziehen sich auf das Literaturverzeichnis.

BEWEIS. Der Einfachheit der Schreibweise halber nehmen wir an, daß $f(0)=0$ sei (wäre dies nicht der Fall, so würden wir statt $f(z)$ die Funktion $[f(z)-f(0)]$ betrachten). Es sei z ein innerer Punkt, des betrachteten Kreisgebietes, d.h. es sei $|z|<R$, dann haben wir, wenn $f(z) = a_1 z + a_2 z^2 + \dots$:

$$f(ze^{i\alpha_\nu}) = a_1 ze^{i\alpha_\nu} + a_2 z^2 e^{i2\alpha_\nu} + \dots$$

$$f(ze^{i\alpha_\nu}) \cdot e^{-i\alpha_\nu} = a_1 z + a_2 z^2 e^{i\alpha_\nu} + a_3 z^3 e^{2i\alpha_\nu} + \dots,$$

$$\sum_{\nu=1}^{n} f(ze^{i\alpha_\nu}) e^{-i\alpha_\nu} = na_1 z + a_2 z^2 \sum_{\nu=1}^{n} e^{i\alpha_\nu} + a_3 z^3 \sum_{\nu=1}^{n} e^{2i\alpha_\nu} + \dots .$$

Dividieren wir beide Seiten der letzteren Gleichung durch nz so erhalten wir eine Gleichung für die Darstellung der im Inneren des Kreises $|z| \leqq R$ regulären analytischen Funktion

$$F(z) = \frac{1}{nz} \sum_{\nu=1}^{n} f(ze^{i\alpha_\nu}) e^{-i\alpha_\nu},$$

$$F(z) = a_1 + \left(\frac{a_2}{n} \sum_{\nu=1}^{n} e^{i\alpha_\nu} \right) z + \left(\frac{a_3}{n} \sum_{\nu=1}^{n} e^{i2\alpha_\nu} \right) z^2 + \dots \tag{4}$$

Die Funktion $F(z)$ ist ihrer Herleitung nach in den inneren Punkten z, $(|z|<R)$ analytisch, doch auf der abgeschlossenen Kreisfläche $|z| \leqq R$ stetig; nimmt also ihr absoluter Betrag den maximalen Wert in einem Randpunkte z_1, $(|z_1|=R)$ an, und da $F(0)=f'(0)=a_1$ ist, so haben wir:

$$|a_1| = |f'(0)| \leqq \frac{1}{nR} \left| \sum_{\nu=1}^{n} f(z_1 e^{i\alpha_\nu}) e^{-i\alpha_\nu} \right| = |F(z_1)|,$$

womit die Ungleichung (3) bewiesen ist.

Die positive Zahl n in dieser Ungleichung läßt sich durch keine größere ersetzen, da für die Funktion

$$f(z) = a_1 z$$

in bezug auf beliebige n Punkte $ze^{i\alpha_1}, ze^{i\alpha_2}, \dots, ze^{i\alpha_n}$ $(|z|=R)$ gilt:

$$a_1 = \frac{1}{nz} \sum_{\nu=1}^{n} f(ze^{i\alpha_\nu}) e^{-i\alpha_\nu}.$$

Damit ist die Behauptung des Satzes 1 bewiesen.

Natürlich hätten wir in ähnlicher Weise eine Abschätzung des Absolutwertes der Ableitung einer analytischen Funktion in einem beliebigen Punkte z_0 ihres Definitionsgebietes bekommen können, dann hätten wir für einen gewissen Wert z_1 die Ungleichung

$$|f'(z_0)| \leqq \frac{1}{nR} \left| \sum_{\nu=1}^{n} f[z_0 + (z_1 - z_0) e^{i\alpha_\nu}] e^{-i\alpha_\nu} \right|$$

bekommen (wo $f(z)$ im Kreise $|z-z_0| \leqq R$ stetig, für $|z-z_0|<R$ analytisch, $|z_1-z_0| = R$ und $f(z_0)=0$ ist). Betrachten wir jetzt den Fall: $f(0)=a_0=0$ und

$f'(0)=a_1\neq 0$, wobei wir auch annehmen, daß die n Punkte $ze^{i\alpha_1}, ze^{i\alpha_2}, \ldots, ze^{i\alpha_n}$ auf der Kreislinie $|z|=R$ regelmäßig verteilt seien. Wir nehmen an, daß

$$\alpha_1 = 0, \quad \alpha_2 = \frac{2\pi}{n}, \quad \alpha_3 = \frac{4\pi}{n}, \ldots, \quad \alpha_n = \frac{n-1}{n}2\pi \quad \text{ist.}$$

In diesem Falle nimmt die Funktion $F(z)$, die wir nunmehr auch durch den Index n kenntlich machen, also $F_n(z)$ an Stelle von $F(z)$ schreiben wollen, eine besonders einfache Gestalt an, es ist nämlich der in der Formel (4) vorkommende Faktor

$$\sum_{\nu=0}^{n-1} e^{\varkappa i\frac{2\pi}{n}\nu} = 0$$

wenn $\varkappa$ kein Vielfaches von n, und ist gleich n, wenn $\varkappa$ ein Vielfaches von n ist, woraus man einfach berechnet:

$$F_n(z) = a_1 + a_{n+1}z^n + a_{2n+1}z^{2n} + \ldots \tag{5}$$

Da die Potenzreihe $f(z) = a_1 z + a_2 z^2 + \ldots$ in einem Kreise $|z|\leqq R_1 < R$ absolut und gleichmäßig konvergent ist, so folgt aus der Gleichung (5), daß

$$\lim_{n\to\infty} F_n(z) = a_1 \tag{6}$$

gleichmäßig im Kreise $|z|\leqq R_1$.

Die Richtigkeit der Beziehung (6) hätte man auch auf einem anderen Wege einsehen können, da nämlich die rechte Seite der Gleichung

$$F_n(z) = \frac{1}{nz}\sum_{\nu=0}^{n-1} f\left(ze^{i\nu\frac{2\pi}{n}}\right)e^{-i\nu\frac{2\pi}{n}}, \qquad (|z| = R_0 \leqq R)$$

eine Näherungssume des Integrals

$$\frac{1}{2\pi i}\int_{|\zeta|=R_0}\frac{f(\zeta)}{\zeta^2}\,d\zeta = f'(0)$$

ist. In der vorigen Näherungssumme kann z eine beliebige fixierte komplexe Zahl im Kreise $|z|\leqq R$ bedeuten.

Die Gleichungen (5) und (6) besagen zusammen folgendes:

Wenn man in einer w-Ebene die Menge der Bildpunkte

$$w = F_n(z)$$

betrachtet und gleichzeitig eine (z.B. kreisförmige) Umgebung des Punktes a_1, die Menge der Punkte w mit $|w-a_1|<\varrho$ ($\varrho>0$, eine Konstante) ins Auge faßt, so sieht man, daß für genügend große n die Menge der Bildpunkte $w=F_n(z)$ wo $|z|\leqq R_1 < R$, zu der Umgebung $|w-a_1|<\varrho$ gehört. Ist z_1 ein fixierter Punkt im Kreise $|z|\leqq R$, so haben die Punkte $z_1, z_1 e^{i\frac{2\pi}{n}}, z_1 e^{i\frac{4\pi}{n}}, \ldots, z_1 e^{i\frac{n-1}{n}2\pi}$, dasselbe Bild $w_1 = F_n(z_1) = F_n\left(z_1 e^{i\frac{2\pi}{n}}\right) = \ldots = F_n\left(z_1 e^{i\frac{n-1}{n}2\pi}\right)$. Mithin nimmt die im Kreise $|z|\leqq R$ definierte stetige und in $|z|<R$ analytische Funktion $w=F_n(z)$ bereits im

Sektor $0 \leqq \varphi < \frac{2\pi}{n}$, $(z = re^{i\varphi})$ sämtliche ihre Werte an, sie ist automorf in bezug auf die Transformation $z' = ze^{i\frac{2\pi}{n}}$. Wenn man den trivialen Fall $F_n(z) = a_1$ (der natürlich auch vorkommen kann) außer Acht läßt, so konstatiert man, daß der Punkt a_1 ein innerer Punkt der Menge der Bildpunkte $w = F_n(z)$ ist und daß die Kreislinie $|w| = a_1$ (wir haben doch $a_1 \neq 0$ vorausgesetzt) die Grenze der Bildpunkte (für $|z| \leqq R_1$) mindestens zweimal schneidet (siehe Fig. 1). Es gibt dann sowohl Bildpunkte w mit $|w| > |a_1|$, als auch mit $|w| < |a_1|$ und $|w| = |a_1|$, wobei ihre Urbilder auf der Kreislinie $|z| = R_1$ periodisch verteilt sind (d.h. mit z_1 zusammen auch die Punkte $z_1 e^{iv\frac{2\pi}{n}}$ Urbilder eines und desselben Punktes w_1 sind). Für $n = 2$ hat man den Fall von LANDAU und TOEPLITZ.

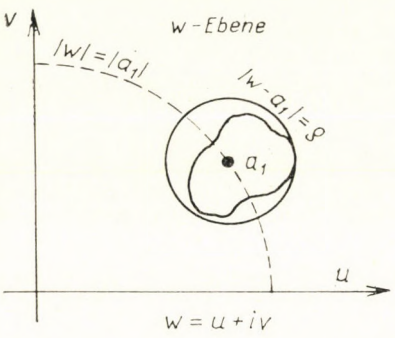


Fig. 1

Auf die Funktionen $F_n(z)$ bin ich bei dem Versuche einen „rein funktionentheoretischen" Beweis folgenden Satzes von PÓLYA und SZEGŐ [2] zu finden, gestoßen:

Die Funktion $w = a_1 z + a_2 z^2 + \ldots$ bilde den Kreis $|z| < R$ auf ein einfach zusammenhängendes Gebiet $\mathfrak{G}$ schlicht ab und $\mathfrak{G}^*$ sei ein Gebiet in der w-Ebene, das wir durch Symmetrisierung von $\mathfrak{G}$ in bezug auf die u-Achse $(w = u + iv)$ nach STEINER, oder in bezug auf eine Halbgerade die durch den Punkt $w = 0$ geht, nach PÓLYA, erhalten haben. Fernerhin sei $w^* = f^*(z) = a_1^* z + a_2^* z + \ldots$ eine Funktion, die eine schlichte Abbildung des Kreises $|z| < R$ auf das Gebiet $\mathfrak{G}^*$ (also mit $f^*(0) = 0$) verwirklicht. Dann ist

$$|a_1| \leqq |a_1^*|.$$

Bei HAYMAN [3] ist dieser Satz in einer allgemeineren Form ausgesprochen, doch für die Ausarbeitung einer Beweismethode habe ich es versucht den Satz unter weiteren einschränkenden Voraussetzungen zu beweisen.

Es handelt sich darum, zu zeigen, daß wenn man für einen beliebigen Wert von n die beiden Funktionen

$$F_n(z) = a_1 + a_{n+1} z^n + a_{2n+1} z^{2n} + \ldots$$

und

$$F_n^*(z) = a_1^* + a_{n+1}^* z^n + a_{2n+1}^* z^{2n} + \ldots$$

miteinander vergleicht, so ist es immer möglich solche z_1 und z_1^* Werte zu finden, daß

$$|F_n(z_1)| \leqq |F_n^*(z_1^*)|, \quad (|z_1| \leqq R_1,\ |z_1^*| \leqq R_1,\ R_1 < R)$$

ist, woraus man nach (6) für $n \to \infty$ die Ungleichung

$$|a_1| \leqq |a_1^*|$$

erhalten würde. Daß das Gelingen eines solchen Beweises nicht aussichtslos ist, soll folgendes Beispiel zeigen. Betrachten wir das symmetrisierte Gebiet $\mathfrak{G}^*$. Der Einfachheit halber nehmen wir an, daß wir nach der Steinerschen Methode in

bezug auf die u-Achse symmetrisiert haben. Nehmen wir an, daß wir die Abbildungen des abgeschlossenen Kreises $|z| \leqq R$, sowohl für $w=f(z)$ als auch für $w^*=f^*(z)$ ins Auge gefaßt haben und auch, daß die Funktion $w^*=f^*(z)$ so ausgefallen sei, daß in ihrer Potenzreihenentwicklung ausser des Gliedes $a_1^* z$ die ungeraden Potenzen fehlen (z.b. sei $\mathfrak{G}^*$ eine abgeschlossene Kreisfläche). Für $n=2$ erhalten wir dann

$$w^* = F_2^*(z) = \frac{1}{2z}[f^*(z)-f^*(-z)] = a_1^*$$

und

$$w = F_2(z) = \frac{1}{2z}[f(z)-f(-z)] = a_1 + a_3 z^2 + a_5 z^4 + \dots .$$

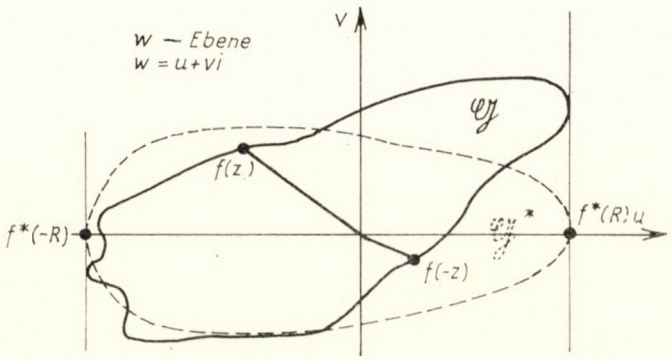


Fig. 2

Wir können auch annehmen, daß a^* und a_1 reelle positive Zahlen seien und setzen wir voraus, daß $|f(z)-f(-z)|$ für $|z|=R$ niemals größer als $|f^*(z)-f^*(-z)|$ ausfällt. Betrachten wir das Gebiet $\mathfrak{G}$ und sein symmetrisiertes $\mathfrak{G}^*$ und setzen wir $z=R$. Es ist dann leicht einzusehen, daß $|f^*(R)-f^*(-R)|$ gerade gleich der Länge desjenigen Teiles der u-Achse ist, welcher in das Gebiet $\mathfrak{G}^*$ fällt, und, daß die absoluten Beträge der Realteile der Größen $[f(z)-f(-z)]$ niemals größer als $|f^*(R)-f^*(-R)|$ sein können (siehe Fig. 2). Nun ist a_1 eine reelle Zahl (>0) und der entsprechende Punkt a_1 ist ein innerer Punkt der Menge der Bildpunkte $w=F_2(z)$. Auf einer geschlossenen Kurve in der w-Ebene die den Punkt a_1 endlich oft umkreist (auf der Kurve $w=F_2(z)$ wo $|z|=R$ ist) sind die Werte der Realteile von w nicht größer als

$$w^* = \frac{1}{2R}[f^*(R)-f^*(-R)] = a_1^*,$$

woraus die Ungleichung

$$0 < a_1 \leqq a_1^*$$

folgt.

LITERATURVERZEICHNIS

[1] LANDAU, E. und TOEPLITZ, O.: Über die grösste Schwankung einer analytischen Funktion in einem Kreise, *Archiv der Mathematik und Physik* **11** (1907) 302—307.
[2] PÓLYA, G. and SZEGŐ, G.: *Isoperimetric Inequalities in Mathematical Physics*, Annals of Mathematics Studies, No. 27, Princeton, 1951, p. 187.
[3] HAYMAN, W. K.: *Multivalent Functions*, Cambridge Tracts..., No. 48. Cambridge, 1958. p. 84.

Mathematisches Institut der Ungarischen Akademie der Wissenschaften, Budapest

(Eingegangen: 26. April, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 249—225.

ÜBER EINE KLASSE LAGRANGESCHER INTERPOLATIONSVERFAHREN

von

G. FREUD

1. Formulierung des Satzes

Es sei $\alpha(x)$ eine in $[-1, +1]$ definierte nicht abnehmende Funktion, $p_n(d\alpha; x)$ sei das Orthogonalpolynom n-ten Grades bezüglich $d\alpha(x)$, d.h. es sei $p_n(d\alpha; x) = \gamma_n(d\alpha)x^n + \ldots$, $\gamma_n(d\alpha) > 0$ und

$$\int_{-1}^{+1} p_n(d\alpha; x)p_m(d\alpha; x)\, d\alpha(x) = \begin{cases} 1 & \text{für } n = m \\ 0 & \text{für } n \neq m \end{cases} \tag{1}$$

Die Nullstellen von $p_n(d\alpha; x)$ liegen dann in $(-1, +1)$; sie seien in abnehmender Folge

$$x_{1n}(d\alpha) > x_{2n}(d\alpha) > \ldots > x_{nn}(d\alpha). \tag{2}$$

Wo kein Missverständniss zu befürchten ist, schreiben wir x_{kn} am Stelle von $x_{kn}(d\alpha)$.

Die LANGRANGEsche Interpolationsformel bezüglich des Grundpunktsystems (2) sei

$$L_n(d\alpha; x) = \sum_{k=1}^{n} l_n(d\alpha; x_{kn}; x)f(x_{kn}). \tag{3}$$

Es sei $f(x)$ eine in $[-1, +1]$ definierte Funktion und sei

$$F(x) = f(x) - \frac{1-x}{2}f(-1) - \frac{1+x}{2}f(+1). \tag{4}$$

Dann ist das Lagrangesche Interpolationspolynom über die Grundpunkte

$$1, x_{1n}(d\alpha), x_{2n}(d\alpha), \ldots, x_{nn}(d\alpha), -1 \tag{5}$$

— wie man leicht nachrechnet — durch die Formel

$$L_n^*(d\alpha; f; x) = \frac{1+x}{2}f(1) + \frac{1-x}{2}f(-1) + \\ + (1-x^2)L_n\left(d\alpha;\ \frac{F(t)}{1-t^2};\ x\right) \tag{6}$$

dargestellt. Über die Konvergenz der Interpolationsfolge (3) sind u.a. folgende recht allgemeine Sätze bekannt:

a) Aus

$$\alpha'(x) \geqq m \geqq 0 \tag{7}$$

fast überall in $x\in[-1, +1]$ folgt, daß für jede stetig differenzierbare Funktion $f(x)$ $L_n(d\alpha; f; x)$ glichmäßig in $[-1, +1]$ gegen $f(x)$ konvergiert.

b) Aus (7) folgt auch, daß für jede Funktion $f(x)$, welche die Lipschitz-Bedingung

$$f(x+h)-f(x) = o(|h|^{1/2}) \tag{8}$$

in $[-1, +1]$ gleichmäßig befriedigt, $L_n(d\alpha; f; x)$ gleichmäßig in jedem inneren Teilintervall von $[-1, +1]$ gegen $f(x)$ strebt.

c) Aus $m>0$,

$$\alpha'(x)\geqq m(1-x^2)^{-1/2} \tag{9}$$

fast überall in $x\in[-1, +1]$ folgt, dass für jede Funktion $f(x)$, welche (8) gleichmäßig in $[-1, +1]$ befriedigt ist, $L_n(d\alpha; f; x)$ gleichmäßig in $[-1, +1]$ gegen $f(x)$ strebt.

Diese Sätze wurden von J. SHOHAT [5] entdeckt; ein anderer Beweis wurde von G. GRÜNWALD und P. TURÁN [3] gegeben. [1] Später zeigte G. ALEXITS [1], dass es im Satze b) genügt wenn man (7) nur fast überall in einem Teilintervall $[a, b]\subset$ $\subset(-1, +1)$ voraussetzt. Es folgt dann, daß $L_n(d\alpha; f; x)$ in jedem echten Teilintervall $[a+\delta, b-\delta]$ gegen $f(x)$ strebt. *Die Interesse dieser Sätze liegt in den Umstand, dass es keine Voraussetzung bezüglich der Grössenordnung von* $p_n(d\alpha; x)$ *enthält.*

In vorliegender Arbeit geben wir einen änlichen Satz bezüglich $L_n^*(d\alpha; f; x)$:

SATZ. *Es sei für ein* $m>0$

$$\alpha'(x)\geqq m(1-x^2)^{1/2} \tag{10}$$

fast überall in $x\in[-1, +1]$ *gültig, ferner sei*

$$\int_{-1}^{+1}\frac{d\alpha(x)}{\sqrt{1-x^2}}<\infty; \tag{11}$$

dann gilt für jede Funktion $f(x)$ *welche* (8) *gleichmäßig in* $[-1, +1]$ *befriedigt,*

$$\lim_{n\to\infty} L_n^*(d\alpha; f; x) = f(x) \tag{12}$$

und zwar gilt (12) *gleichmäßig in* $x\in[-1, +1]$.

Verglichen mit den änlichen Satz b), hat zunächst unser Satz den Vorteil, daß wir für die gleiche Funktionenklasse die gleichmäßige Konvergenz der Interpolationspolynome in dem ganzen Orthogonalitätsintervall $[-1, +1]$ behaupten können. Was die Bedingungen bezüglich der Belegungsfunktion $\alpha(x)$ betrifft, kann auch diese als milder betrachtet werden, da es Belegungen umfaßt, für welche $\alpha'(x)$ für $x\to+1$, bzw. $x\to-1$ gegen Null strebt. Gegenüber diesen Gewinn erscheint der Verlust an zugelassenen $\alpha(x)$, welche durch die neue Bedingung (11) entsteht, gering zu sein. Das gilt noch mehr im Vergleich mit Satz c).

[1] Die Sätze wurden ursprünglich für absolut stetige $\alpha(x)$ angekündet; die Verallgemeinerung bietet aber keine Schwierigkeiten.

2. Einige Hilfssätze

Es ist aus der Theorie der Orthogonalpolynome bekannt, daß das Minimum des Integrals

$$\int_{-1}^{+1} \pi_{n-1}^2(t)\, d\alpha(t),$$

falls $\pi_{n-1}(t)$ alle die Polynome höchstens $n-1$-ten Grades durchläuft, welche $\pi_{n-1}(x)=1$ befriedigen, gleich

$$\lambda_n(d\alpha; x) = \left\{\sum_{\nu=0}^{n-1} p_\nu^2(d\alpha; x)\right\}^{-1} \tag{13}$$

ist (Vgl. M. RIESZ [4]). Gilt für zwei nichtabnehmende Funktionen $\alpha_1(x)$ und $\alpha_2(x)$

$$\alpha_2(x+h)-\alpha_2(x) \geqq \alpha_1(x+h)-\alpha_1(x)$$

für jedes $-1 \leqq x < x+h \leqq +1$, dann sagen wir, $d\alpha_2$ ist eine Maiorante von $d\alpha_1$, im Zeichen $d\alpha_2 \geqq d\alpha_1$. Aus der Minimumeigenschaft von $\lambda_n(d\alpha; x)$ folgt

HILFSSATZ 1. *Aus* $d\alpha_2 \geqq d\alpha_1$ *folgt*

$$\lambda_n(d\alpha_2; x) \geqq \lambda_n(d\alpha_1; x) \qquad (-1 \leqq x \leqq +1).$$

Dieser wohlbekannte Hilfssatz bildete die Grundlage der Beweise der Sätze a), b) und c).

Wir führen neben $\alpha(x)$ die weiteren Belegungsfunktionen

$$\alpha^*(x) = \int_{-1}^{x} (1-t)\, d\alpha(t), \quad \alpha^{**}(x) = \int_{-1}^{x} (1+t)\, d\alpha(t) \tag{14}$$

ein.

HILFSSATZ 2. *Es gilt die Formel*

$$\sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})} = \lambda_n^{-1}(d\alpha^*; x) \tag{15}$$

und

$$\sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1+x_{kn})\lambda_n(d\alpha; x_{kn})} = \lambda_n^{-1}(d\alpha^{**}; x). \tag{16}$$

BEWEIS. Es gilt für beliebige Polynome höchstens $2n-1$-ten Grades $\pi_{2n-1}(x)$ die Gauss—Jacobische Quadraturformel [2]

$$\int_{-1}^{+1} \pi_{2n-1}(t)\, d\alpha(t) = \sum_{k=1}^{n} \lambda_n(d\alpha; x_{kn})\pi_{2n-1}(x_{kn}).$$

[2] Es ist aus der Formel selbst ersichtlich, daß die Koeffizienten

$$\lambda_n(d\alpha; x_{kn}) = \int_{-1}^{+1} l_n^2(d\alpha; x_{kn}; x)\, d\alpha(x)$$

das Minimum des Integrals $\int_{-1}^{+1} \pi_{n-1}^2(t)\, d\alpha(t)$ unter den Polynomen höchstens n-1-ten Grades mit $\pi_{n-1}(x_{kn})=1$ ergeben.

Wir setzen in die Quadraturformel $\pi_{2n-1}(t) = (1-t)\pi_{n-1}^2(t)$, wo $\pi_{n-1}(t)$ ein Polynom höchstens $n-1$-ten Grades mit $\pi_{n-1}(x)=1$ ist.

Es folgt unter Beachtung der ersten Hälfte von (14)

$$\int_{-1}^{+1} \pi_{n-1}^2(t)\, d\alpha^*(t) = \sum_{k=1}^{n} \lambda_n(d\alpha; x_{kn})(1-x_{kn})\pi_{n-1}^2(x_{kn}),$$

und mit Hilfe der Cauchy-schen Ungleichung

$$1 = \pi_{n-1}^2(x) = \left[\sum_{k=1}^{n} l_n(d\alpha; x_{kn}; x)\pi_{n-1}(x_{kn})\right]^2 \leqq$$

$$\leqq \sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})} \sum_{k=1}^{n} \lambda_n(d\alpha; x_{kn})(1-x_{kn})\pi_{n-1}^2(x_{kn}) =$$

$$= \sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})} \int_{-1}^{+1} \pi_{n-1}^2(t)\, d\alpha^*(t)$$

so daß

$$\int_{-1}^{+1} \pi_{n-1}^2(t)\, d\alpha^*(t) \geqq \frac{1}{\sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})}}$$

gültig ist, und in dieser Ungleichung tretet für

$$\pi_{n-1}(t) = \frac{\sum_{k=1}^{n} \frac{l_n(d\alpha; x_{kn}; x)\, l_n(d\alpha; x_{kn}; t)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})}}{\sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})}}$$

das Gleichheitszeichen ein. Aus der zu Anfang dieses Teiles erwähnten Minimumeigenschaft von $\lambda_n(d\alpha; x)$ folgt also (15); in analoger Weise erhält man (16), w.z.b.w.

HILFSSATZ 3. *Es sei*

$$\beta_1(x) = \int_{-1}^{x} (1-t)\sqrt{1-t^2}\, dt, \quad \beta_2(x) = \int_{-1}^{x} (1+t)\sqrt{1-t^2}\, dt$$

dann gilt

$$\lambda_n^{-1}(d\beta_1; x) \leqq \frac{4n}{(1-x)^2(1+x)} \tag{17}$$

und

$$\lambda_n^{-1}(d\beta_2; x) \leqq \frac{4n}{(1+x)^2(1-x)}. \tag{18}$$

BEWEIS. Wir zeigen die Gültigkeit von (17); die Gültigkeit von (18) folgt dann aus Symmetriegründen. Eine direkte Rechnung zeigt, daß

$$p_n(d\beta_1; \cos\theta) = [\pi(n+1)(n+2)]^{-1/2}(1-\cos\theta)^{-1} \cdot$$

$$\cdot\left\{(n+2)\frac{\sin(n+1)\theta}{\sin\theta} - (n+1)\frac{\sin(n+2)\theta}{\sin\theta}\right\} =$$

$$= [\pi(n+1)(n+2)]^{-1/2}(1-\cos\theta)^{-1}\left\{-(n+1)\frac{\cos(n+\frac{3}{2})\theta}{\cos\frac{\theta}{2}} + \frac{\sin(n+1)\theta}{\sin\theta}\right\}$$

ist. Aus dieser Formel ergibt sich

$$|p_n(d\beta_1; x)| \leqq \frac{2}{(1-x)\sqrt{1+x}}, \tag{19}$$

und aus (13) und (19) ergibt sich (17), w.z.b.w.

HILFSSATZ 4. *Aus* (10) *folgt*

$$\lambda_n^{-1}(d\alpha^*; x) \leqq \frac{4n}{m(1-x)^2(1+x)} \tag{20}$$

und

$$\lambda_n^{-1}(d\alpha^{**}; x) \leqq \frac{4n}{m(1-x)(1+x)^2}. \tag{21}$$

BEWEIS. Aus (10) und (14) folgt $d\alpha^* \leqq m d\beta_1$, bzw. $d\alpha^{**} \leqq m d\beta_2$. Infolge Hilfssatz 1 ist dann

$$\lambda_n(d\alpha^*; x) \leqq \lambda_n(m d\beta_1; x) = m\lambda_n(d\beta_1; x)$$

bzw.

$$\lambda_n(d\alpha^{**}; x) \leqq \lambda_n(m d\beta_2; x) = m\lambda_n(d\beta_2; x).$$

Es folgt jetzt (20) bzw. (21) aus Hilfssatz 3, w.z.b.w.

HILFSSATZ 5. *Aus* (11) *folgt*

$$\sum_{k=1}^{n} \lambda_n(d\alpha; x_{kn})(1-x_{kn}^2)^{-1/2} \leqq \int_{-1}^{+1} \frac{d\alpha(t)}{\sqrt{1-t^2}}. \tag{22}$$

BEWEIS. Wir setzen $F(t) = (1-t^2)^{-1/2}$, und es sei $h_n(F; t)$ das Hermitesche Interpolationspolynom höchstens $2n-1$-ten Grades mit

$$h_n(F; x_{kn}) = F(x_{kn}),\ h_n'(F; x_{kn}) = F'(x_{kn}) \qquad (k = 1, 2, \ldots n).$$

Nach der Restgliedformel der Interpolation ist

$$F(x) - h_n(F; x) = (x-x_{1n})^2 \ldots (x-x_{nn})^2 \frac{F^{(2n)}(\xi)}{(2n)!} \geqq 0.$$

Es folgt mit Hilfe der Quadraturformel

$$\int_{-1}^{+1} F(t)\,d\alpha(t) \leqq \int_{-1}^{+1} h_n(F;t)\,d\alpha(t) = \sum_{k=1}^{n} \lambda_n(d\alpha; x_{kn})F(x_{kn})$$

w.z.b.w.

3. Beweis des Satzes

Aus den Hilfssätzen 2 und 4 folgt, dass unter Voraussetzungen des Satzes

$$\sum_{k=1}^{n} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn}^2)\lambda_n(d\alpha; x_{kn})} \leqq \sum_{x_{kn}\leqq 0} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn})\lambda_n(d\alpha; x_{kn})} +$$

$$+ \sum_{x_{kn}\leqq 0} \frac{l_n^2(d\alpha; x_{kn}; x)}{(1+x_{kn})\lambda_n(d\alpha; x_{kn})} \leqq \lambda_n^{-1}(d\alpha^*; x) + \lambda_n^{-1}(d\alpha^{**}; x) \leqq \tag{23}$$

$$\leqq \frac{4n}{m}\left[\frac{1}{(1-x)^2(1+x)} + \frac{1}{(1+x)^2(1-x)}\right] \leqq \frac{16n}{m(1-x^2)^2}$$

gültig ist.

Aus der Lipschitz-Bedingung folgt, daß für jedes ganze n ein Polynom höchstens n-ten Grades $\psi_n(x)$ gibt, so daß

$$|f(x)-\psi_n(x)| \leqq \varepsilon_n \frac{\sqrt[4]{1-x^2}}{n^{1/2}}; \qquad \varepsilon_n \to 0 \tag{24}$$

gültig ist (sehe I. E. GOPENGAUS [2] und S. A. TELJAKOVSKI [6]).

Es ist dann

$$f(x) - L_n^*(d\alpha; f; x) = f(x) - \psi_n(x) - L_n^*(d\alpha; f-\psi_n; x) \tag{25}$$

und unter Beachtung von $\psi_n(\pm 1) = f(\pm 1)$

$$L_n^*(d\alpha; f-\psi_n; x) = (1-x^2)\sum_{k=1}^{n} \frac{l_n(d\alpha; x_{kn}; x)}{1-x_{kn}^2}\cdot[f(x_{kn})-\psi_n(x_{kn})],$$

und weiter aus (24):

$$|L_n^*(d\alpha; f-\psi_n; x)| \leqq \frac{\varepsilon_n}{n^{1/2}}(1-x^2)\sum_{k=1}^{n}\left|\frac{l_n(d\alpha; x_{kn}; x)}{(1-x_{kn}^2)^{3/4}}\right| \leqq$$

$$\leqq \frac{\varepsilon_n}{n^{1/2}}(1-x^2)\sqrt{\sum_{k=1}^{n}\frac{l_n^2(d\alpha; x_{kn}; x)}{(1-x_{kn}^2)\lambda_n(d\alpha; x_{kn})}}\cdot\sqrt{\sum_{k=1}^{n}\lambda_n(d\alpha; x_{kn})(1-x_{kn}^2)^{-1/2}},$$

also unter Beachtung von (23) und Hilfssatz 5

$$|L_n^*(d\alpha;\, f-\psi_n;\, x)| \leqq \frac{\varepsilon_n}{n^{\frac{1}{2}}}(1-x^2)\sqrt{\frac{16n}{m(1-x^2)^2}}\sqrt{\int_{-1}^{+1}\frac{d\alpha(x)}{\sqrt{1-x^2}}} =$$

(26)

$$= \frac{4}{\sqrt{m}}\varepsilon_n\sqrt{\int_{-1}^{+1}\frac{d\alpha(x)}{\sqrt{1-x^2}}} \to 0.$$

Infolge (24), (25) und (26) strebt $L_n^*(d\alpha; f; x)$ in $x\in[-1, +1]$ gleichmäßig gegen $f(x)$, w.z.b.w.

LITERATURVERZEICHNIS

[1] ALEXITS, G.: Eine Bemerkung zur Konvergenzfrage des Lagrangeschen Interpolationsverfahrens, *Acta Math. Acad. Sci. Hungar.* **4** (1953) 233—236.

[2] Гопенгауз И. Е.: К теореме А. Ф. Тимана о приближении функций многочленами на конечном отрезке, *Математические Записки* **I** (1967) 163—172.

[3] GRÜNWALD, G. und TURÁN, P.: Über Interpolation, *Annali di Scoula Norm. Sup. di Pisa* **7** (1938) 137—146.

[4] RIESZ, M.: Sur le probléme des momentes, III, *Arkiv för Matem. Astr. och. Fys.* **17/16** (1923).

[5] SHOHAT, J.: On interpolation, *Ann. Math.* **34** (1933) 130—146.

[6] Теляковски, С. А. Две теоремы о приближении функций алгебраическими многочленами, *Мат. Сбор.* **70** (1966) 252—265.

Mathematisches Institut der Ungarischen Akademie der Wissenschaften, Budapest

(Eingegangen: 30. April, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 257—277.

ON A FIRST ORDER NONLINEAR DIFFERENTIAL EQUATION SYSTEM

by

I. BIHARI and T. FÉNYES

1. The system in question arisen in applications on chemical-biological processes in meats is as follows

$$\begin{aligned} x' + a(xy)' &= -k_1 x \\ y' + a(xy)' &= -k_2 y \end{aligned} \qquad \left(x = x(t),\ y = y(t),\ ' = \frac{d}{dt} \right) \tag{1}$$

where a is a *large* parameter $a \geqq 100$, $0, 1 \leqq k_i \leqq 0, 3$ $(i = 1, 2)$ and $x(0) > 0$, $y(0) > 0$.

It has been suspected that $x \to 0$, $y \to 0$ as $t \to +\infty$. The proof of this conjecture and determination of the solution will be proposed here.

There is no difficulty to obtain a power series expansion for x and y of the form

$$x = \sum_{n=0}^{\infty} a_n t^n, \quad y = \sum_{n=0}^{\infty} b_n t^n$$

or an expansion in decreasing powers of the large parameter a as

$$x = \sum_{n=0}^{\infty} a^{-n} f_n(t), \quad y = \sum_{n=0}^{\infty} a^{-n} g_n(t)$$

and to determine subsequently $f_n(t)$, $g_n(t)$ in the form of power series in t. However the series so received — as power series in general — converge slowly for large t, thus just the behaviour in infinity we are interested in, cannot be decided by them.

First let us investigate the influence of the increase of a. Equation (1) seems to give $(xy)' = 0$, $xy = \text{const}$ for $a = \infty$. In fact, by solving (1) for x and y we have with $\varepsilon = \dfrac{1}{a}$.

$$\begin{aligned} x' &= -k_1 x + (k_1 + k_2) \frac{xy}{\varepsilon + x + y} = x f(x, y) \\ y' &= -k_2 y + (k_1 + k_2) \frac{xy}{\varepsilon + x + y} = y g(x, y) \end{aligned} \tag{2}$$

which is completely equivalent to (1) provided $\varepsilon \neq 0$, while for $\varepsilon = 0$ this gives $x'y + y'x = (xy)' = 0$, $xy = c^2$, i.e. in this limit case the origin is a *saddle point* with respect to the autonomous system (2). In the same time (2) can be integrated

now in closed form. Namely by separation of the variables

$$\frac{\left(x^2-\frac{1}{\nu}c^2\right)^{\frac{\nu+1}{2\nu}}}{x}=Ce^{-k_2t},\quad C=\text{const},\quad \nu=\frac{k_1}{k_2} \tag{3}$$

and $y=\frac{c^2}{x}$. In the sequel the following facts must be taken into account.

1. $x(t)>0$, $y(t)>0$ for $t>0$ provided $x(0)>0$, $y(0)>0$. For if first $x(t)$ vanished at a $t=t_0$, then $y(t_0)\neq 0$, $x'(t)<0$ for $t<t_0$ near t_0 and assuming $y(t_0)\neq v\varepsilon$ (viz. $f(0, v\varepsilon)=0$) the following two cases are possible:

a) $x(t)$ conserves its positive sign, then being $f(x, y)\neq 0$ (moreover $f(x, y)<0$) $x'<0$ for $t>t_0$ near t_0, consequently $x(t)<0$ in the same time,

b) $x(t)$ changes its sign, x' does the same, i.e. $x(t)$ has a minimum $x(t_0)=0$ at $t=t_0$ and so $x(t)>0$ for $t>t_0$ near t_0. On the other hand if $y(t_0)=v\varepsilon$, then $y'(t_0)=-k_1\varepsilon$ implies the decrease of $y(t)$ at $t=t_0$ and $f(x, y)$ continues to be negative and the argument of a) and b) hold now too. (Contradiction in both cases.)

2. The origin is a stable *non proper node* if $k_1\neq k_2$ and a *proper node* if $k_1=k_2$ with respect to the linear system

$$x'=-k_1x,\ y'=-k_2y$$

and according to a well-known theorem concerning system (2) too, provided $\varepsilon\neq 0$, since the second terms in (2) are small of second order in the neighbourhood of the origin. — Having no other critical point in the first quadrant, $x\to 0$, $y\to 0$ as $t\to\infty$.

3. The same theorem asserts nothing if $\varepsilon=0$, being the terms in question small of first order only, but the above immediate integration reveals that the origin is a saddle point now. The dependence of the solution on the parameter ε is *not continuous* for $\varepsilon=0$.

Therefore (3) gives a good approximation for small ε too provided t is not too large, viz. as long as $\varepsilon\ll x+y$.

On the other hand, if already $x\ll\varepsilon$, $y\ll\varepsilon$, then $x+y$ can be neglected beside ε (if $\varepsilon\neq 0$), giving

$$x'=-k_1x+(k_1+k_2)axy \tag{4}$$

$$y'=-k_2y+(k_1+k_2)axy$$

whence e.g. y can be eliminated and a second order nonlinear equation will be obtained for x. However it is not worthy for the present purposes. — By the substitution (variation of the constants)

$$x=A(t)e^{-k_1t},\ y=B(t)e^{-k_2t} \tag{5}$$

we have

$$A'=(k_1+k_2)a\,ABe^{-k_2t} \tag{6}$$

$$B'=(k_1+k_2)a\,ABe^{-k_1t}$$

or in integral equation form

$$A(t) = A(t_0) + a(k_1+k_2)\int_{t_0}^{t} ABe^{-k_2 t}\,dt$$

$$B(t) = B(t_0) + a(k_1+k_2)\int_{t_0}^{t} ABe^{-k_1 t}\,dt$$

Let us apply the successive approximations

$$A_{n+1}(t) = A(t_0) + a(k_1+k_2)\int_{t_0}^{t} A_n(\tau)B_n(\tau)e^{-k_2\tau}\,dt$$

(6a)

$$B_{n+1}(t) = B(t_0) + a(k_1+k_2)\int_{t_0}^{t} A_n(\tau)B_n(\tau)e^{-k_1\tau}\,d\tau$$

with $A_0(t) = A(t_0) = A_0$, $B_0(t) = B(t_0) = B_0$. Obtaining

$$A_1(t) = A_0\left[\left(1+\frac{k_1+k_2}{k_2}aB_0\right)e^{-k_2 t_0} - \frac{k_1+k_2}{k_2}aB_0 e^{-k_2 t}\right]$$

$$B_1(t) = B_0\left[\left(1+\frac{k_1+k_2}{k_1}aA_0\right)e^{-k_1 t_0} - \frac{k_1+k_2}{k_1}aA_0 e^{-k_1 t}\right]$$

whence

$$x_1(t) = A_0\left[\left(1+\frac{k_1+k_2}{k_2}aB_0\right)e^{-(k_2 t_0+k_1 t)} - \frac{k_1+k_2}{k_2}aB_0 e^{-(k_1+k_2)t}\right]$$

$$y_1(t) = B_0\left[\left(1+\frac{k_1+k_2}{k_1}aA_0\right)e^{-(k_1 t_0+k_2 t)} - \frac{k_1+k_2}{k_1}aA_0 e^{-(k_1+k_2)t}\right]$$

etc.

2.1. This last result gives the idea of looking for the asymptotic expression of the actual solution of (1) in the following form of exponential double series (to be precised in paragraph 3) provided $v = \dfrac{k_1}{k_2}$ is irrational (suppose e.g. $0 < v < 1$) and $a < \infty$

$$x \sim \sum_{\substack{m,n=0\\ m+n>0}}^{\infty} a_{mn}e^{-(mk_1+nk_2)t}, \quad y \sim \sum_{\substack{m,n=0\\ m+n>0}}^{\infty} b_{mn}e^{-(mk_1+nk_2)t} \tag{7}$$

The restriction $m+n>0$ (i.e. $a_{00}=b_{00}=0$) is necessary to satisfy conditions $x\to 0$, $y\to 0$ as $t\to\infty$.

First some simplifications on (1) will be carried out. Namely let us consider that

1° Having a solution $x(t), y(t)$ of (1) $X(t)=ax(t)$, $Y(t)=ay(t)$ satisfy equation (1) with $a=1$. Therefore it can and will be assumed $a=1$ in (1) conserving the denotations x and y instead of X and Y respectively.

2° With the new independent variable $\tau = k_2 t$ (1) will have the form (using again t in place of τ)

(1′)
$$x' + (xy)' = -\nu x$$
$$y' + (xy)' = -y.$$

Then (7) can be replaced by

(7′)
$$x \sim \sum_{\substack{m,n=0 \\ m+n>0}}^{\infty} a_{mn} e^{-(m\nu+n)t}, \quad y \sim \sum_{\substack{m,n=0 \\ m+n>0}}^{\infty} b_{mn} e^{-(m\nu+n)t}$$

It is important to remark that $m_1\nu + n_1 = m_2\nu + n_2$ involves $m_1 = m_2$, $n_1 = n_2$, since ν is irrational. — These asymptotic expansions are unique and may be obtained by putting (7′) in (1′) and comparing on both sides the coefficients of equal powers (s. paragraph 3 devoted to these questions). The series so received can converge or not, but in all events the partial sums of them and their formal derivatives (obtained by term by term differentiation) furnish arbitrary good approximations for large t of the actual solution and its derivative. — Now

(8)
$$xy \sim \sum_{m,n=0}^{\infty} c_{mn} e^{-(m\nu+n)t}, \quad c_{mn} = \sum_{p=0}^{m} \sum_{q=0}^{n} a_{pq} b_{m-p,n-q}$$

where c_{mn} does not depend on a_{mn}, b_{mn}, because $a_{00} = b_{00} = 0$.

The said comparison of the coefficients gives

(9)
$$\begin{aligned} (a_{mn} + c_{mn})(m\nu + n) &= \nu a_{mn} \\ (b_{mn} + c_{mn})(m\nu + n) &= b_{mn} \end{aligned} \qquad m+n > 0$$

or

(9′)
$$\begin{aligned} a_{mn} &= -\frac{m\nu+n}{(m-1)\nu+n} c_{mn} \\ b_{mn} &= -\frac{m\nu+n}{m\nu+n-1} c_{mn} \end{aligned} \qquad m+n > 1$$

Let us see the first few steps of the recursion. If

1. $m+n = 1$ and

a) $m=0$, $n=1$, then $c_{01} = a_{00}b_{01} + a_{01}b_{00} = 0$ and so $a_{01} = 0$ while the second line of (9) reduces to $b_{01} = b_{01}$, i.e. $b_{01} = b_{01} = \mu$ remains undetermined (arbitrary).

b) $m=1$, $n=0$ gives $b_{10} = 0$, $a_{10} = \lambda$ arbitrary,

2. $m+n = 2$,

a) $m=0$, $n=2$ gives $c_{02} = 0$, hence $a_{02} = b_{02} = 0$,

b) $m=2$, $n=0$ involves evenso $a_{20} = b_{20} = 0$,

c) $m=1$, $n=1$, then $c_{11} = \lambda\mu$ and

$$a_{11} = -(1+\nu)\lambda\mu, \quad b_{11} = -\left(1 + \frac{1}{\nu}\right)\lambda\mu.$$

In the same way

$$a_{03} = b_{03} = a_{30} = b_{30} = 0$$

$$a_{12} = \frac{1}{2}(1+\nu)(2+\nu)\lambda\mu^2, \quad b_{12} = (2+\nu)\lambda\mu^2$$

$$a_{21} = \left(2+\frac{1}{\nu}\right)\lambda^2\mu, \quad b_{21} = \frac{1}{2}\left(1+\frac{1}{\nu}\right)\left(2+\frac{1}{\nu}\right)\lambda^2\mu$$

The irrationality of ν excludes the possibility of a relation $a_{mn}k_1 q = a_{mn}k_2 p$, where $p \neq q$, except $a_{mn}=0$. On the other hand if ν is rational the expansion (7′) is not unique. — Therefore we have

$$x \sim \lambda e^{-\nu t} - (1+\nu)\lambda\mu e^{-(\nu+1)t} + \frac{1}{2}(1+\nu)(2+\nu)\lambda\mu^2 e^{-(\nu+2)t} +$$

$$+\left(2+\frac{1}{\nu}\right)\lambda^2\mu e^{-(2\nu+1)t} + \ldots$$

$$y \sim \mu e^{-t} - \left(1+\frac{1}{\nu}\right)\lambda\mu e^{-(\nu+1)t} + (2+\nu)\lambda\mu^2 e^{-(\nu+2)t} +$$

$$+\frac{1}{2}\left(1+\frac{1}{\nu}\right)\left(2+\frac{1}{\nu}\right)\lambda^2\mu e^{-(2\nu+1)t} + \ldots$$

2. 2. If $\nu = \frac{k_1}{k_2}$ is rational $\nu = \frac{p}{q} < 1$, $(p, q) = 1$, then instead of (7′) the asymp totic forms of the actual solutions are

$$x \sim \sum_{\substack{m=0 \\ m+n>0}}^{q-1} \sum_{n=0}^{\infty} a_{mn} e^{-\left(\frac{m}{q}+n\right)t}, \quad y \sim \sum_{\substack{m=0 \\ m+n>0}}^{q-1} \sum_{n=0}^{\infty} b_{mn} e^{-\left(\frac{m}{q}+n\right)t} \tag{10}$$

Then

$$xy \sim \sum_{m=0}^{q-1} \sum_{n=0}^{\infty} c_{mn} e^{-\left(\frac{m}{q}+n\right)t}, \qquad c_{mn} = \sum_{\substack{r=0 \\ 0<r+s<m+n}}^{m} \sum_{s=0}^{n} a_{rs} b_{m-r,n-s} \tag{11}$$

Substitution in (1′) and comparison gives now

$$\begin{aligned} \left(\frac{m}{q}+n\right)(a_{mn}+c_{mn}) &= \frac{p}{q} a_{mn} \\ \left(\frac{m}{q}+n\right)(b_{mn}+c_{mn}) &= b_{mn} \end{aligned} \qquad (m+n>0) \tag{12}$$

whence a_{mn}, b_{mn} may be determined recursively.

EXAMPLE. If $\frac{p}{q} = \frac{2}{3}$, then $b_{01} = \mu$, $a_{20} = \lambda$ are arbitrary. Correspondingly

$$x \sim \lambda e^{-\frac{2}{3}t} - \frac{5}{3}\lambda\mu e^{-\frac{5}{3}t} + \frac{20}{9}\lambda\mu^2 e^{-\frac{8}{3}t} - \frac{220}{81}\lambda\mu^3 e^{-\frac{11}{3}t} + \dots$$

$$y \sim \mu e^{-t} - \frac{5}{2}\lambda\mu e^{-\frac{5}{3}t} + \frac{8}{3}\lambda\mu^2 e^{-\frac{8}{3}t} - \frac{220}{72}\lambda\mu^3 e^{-\frac{11}{3}t} + \dots$$

2. 3. Finally in the case $v=1$ one of x and y may be eliminated from (1′) leading to a nonlinear first order equation unsolvable in closed form. However it is not useful here. — The asymptotic series of the actual solution is now

$$x \sim \sum_{n=1}^{\infty} a_n e^{-nt}, \quad y \sim \sum_{n=1}^{\infty} b_n e^{-nt}$$

Then

$$xy \sim \sum_{n=2}^{\infty} c_n e^{-nt}, \quad c_n = \sum_{l=1}^{n-1} a_l b_{n-l}$$

Substitution and comparison gives

$$(n-1)a_n + nc_n = 0 \tag{13}$$

$$(n-1)b_n + nc_n = 0$$

whence $a_n = b_n$ for $n > 1$ while $a_1 = \lambda$, $b_1 = \mu$ remain undetermined free parameters. The recursion gives

$$a_2 = b_2 = -2\lambda\mu, \; a_3 = b_3 = 3\lambda\mu(\lambda+\mu),$$

etc. Thus

$$x \sim \lambda e^{-t} - 2\lambda\mu e^{-2t} + 3\lambda\mu(\lambda+\mu)e^{-3t} + \dots \tag{14}$$

$$y \sim \mu e^{-t} - 2\lambda\mu e^{-2t} + 3\lambda\mu(\lambda+\mu)e^{-3t} + \dots$$

3. 1. Now we have to prove: the formal series found in 2. are asymptotic series of the actual solutions. — Let us begin with case $v=1$. Then the system reads as

$$x' = -x + f(x,y), \quad y' = -y + f(x,y), \quad f(x,y) = \frac{2xy}{1+x+y} \tag{15}$$

whence $y - x = Ce^{-t}$ and it can be assumed $C \geqq 0$. In eliminating y

$$x' = -x + f(x,t), \quad f(x,t) = \frac{2x^2 + 2Cxe^{-t}}{1+2x+Ce^{-t}} \tag{16}$$

which is equivalent to the integral equation

$$x(t) = x(0)e^{-t} + e^{-t}\int_0^t e^{\tau} f(x(\tau),\tau)\,d\tau \tag{17}$$

which will be replaced by the equation

$$x(t) = Ae^{-t} - e^{-t}\int_t^\infty e^\tau f(x(\tau), \tau)\, d\tau, \qquad A > 0 \tag{18}$$

the solution of which — if exists — satisfies (16) and concerning the solutions vanishing in infinity equation (18) is equivalent to (16). — Let (18) be solved by successive approximations as follows

$$\begin{aligned} x_0(t) &= Ae^{-t} \\ x_{n+1}(t) &= Ae^{-t} - e^{-t}\int_t^\infty e^\tau f(x_n(\tau), \tau)\, d\tau, \qquad (n = 0, 1, 2, \ldots) \end{aligned} \tag{19}$$

the form of which involves that $x_n(t) > 0$ for sufficient large t provided it exists. — The partial derivative

$$f_x = \frac{\partial f}{\partial x} = \frac{4x(1+x) + 2Ce^{-t}(1+2x+Ce^{-t})}{(1+2x+Ce^{-t})^2}$$

is positive for positive x, $f(x, t)$ is increasing in x, consequently for $x(t)$ satisfying $0 < x(t) \leqq Ke^{-t}$ the integral in (18) exists. In fact

$$\int_t^\infty e^\tau f(x(\tau), \tau)\, d\tau \leqq \int_t^\infty e^\tau f(Ke^{-\tau}, \tau)\, d\tau \leqq \int_t^\infty e^\tau (2K^2 e^{-2\tau} + 2KCe^{-2\tau})\, d\tau =$$

$$= 2K(K+C)e^{-t}$$

In the same time $0 < x_n(t) \leqq Ke^{-t}$ $(t \geqq t_0)$ implies

$$0 < x_{n+1}(t) \leqq Ae^{-t} + 2K(K+C)e^{-2t} \leqq Ke^{-t} \qquad (t \geqq t_0)$$

provided $2K(K+C)e^{-t_0} < A \leqq K - 2K(K+C)e^{-t_0}$ (what can be satisfied for some $t_0 > 0$) and also implies the validity of $f_x \leqq Le^{-t}$ $(t \geqq t_0)$ with some $L = \text{const.}$ — Letting $\Delta_0 = x_0$, $\Delta_n = |x_n - x_{n-1}|$ $(n = 1, 2, \ldots)$ we have

$$\Delta_{n+1} \leqq e^{-\tau}\int_t^\infty e^\tau |f(x_n(\tau), \tau) - f(x_{n-1}(\tau), \tau)|\, d\tau$$

But

$$|f(x_n(\tau), \tau) - f(x_{n-1}(\tau), \tau| = \Delta_n f_x(\xi, \tau) \leqq \Delta_n Le^{-\tau}$$

$$\xi = x_{n-1}(\tau) + \theta(x_n(\tau) - x_{n-1}(\tau)), \quad 0 < \theta < 1$$

thus

$$\Delta_{n+1} \leqq Le^{-t}\int_t^\infty e^\tau \Delta_n e^{-\tau}\, d\tau$$

Suppose $\Delta_n \leqq Aq^n e^{-t}$, where $q = \text{const}$, $0 < q < 1$ which is valid for $n = 0$, then

$$\Delta_{n+1} \leqq Aq^n e^{-t} Le^{-t}$$

and $\Delta_{n+1} \leqq Aq^{n+1}e^{-t}$ provided $Le^{-t} \leqq q$ what holds for t large enough. So $S = \sum_{n=0}^{\infty} \Delta_n$ converges and $S \leqq 2Ae^{-t}$ provided $q \leqq \frac{1}{2}$. Therefore $\{x_n(t)\}$ $(n=0, 1, 2, \ldots)$ converges uniformly for $t \geqq t_0$ to a limit function $x(t)$ and $0 \leqq x(t) < 2Ae^{-t}$ (moreover $0 < x(t)$; s. paragraph 1). This function satisfies (18) and (16). — To prove the uniqueness of the solution (of (18)) denote $|z(t) - x_n(t)|$ by Δ_n, then provided z is a solution of (18)

$$z(t) - x_n(t) = -e^{-t} \int_t^{\infty} e^{\tau} [f(y(\tau), \tau) - f(x_n(\tau), \tau)] \, dt$$

which implies

$$\Delta_{n+1} \leqq e^{-t} \int_t^{\infty} e^{\tau} L e^{-\tau} \Delta_n \, d\tau$$

and assumption $\Delta_n \leqq Aq^n e^{-t}$ results in $\Delta_{n+1} \leqq Aq^{n+1}e^{-t}$. Therefore this assumption holds for $n=0, 1, 2, \ldots$ and $\Delta_n \Rightarrow 0$ as $n \to \infty$. This means that if $z(t)$ is a solution of (18), $z(t) \equiv x(t)$, $t \geqq t_0$.

The next step will show that $x(t)$ and $y(t)$ can be expanded in asymptotic series as follows

$$x \sim \sum_{n=1}^{\infty} a_n e^{-nt}, \quad y \sim \sum_{n=1}^{\infty} b_n e^{-nt} \tag{20}$$

where

$$a_1 = A, \quad b_1 = B = A + C, \quad a_k = b_k = -\frac{k}{k-1} d_k,$$
$$d_k = \sum_{l=1}^{k-1} a_l b_{k-l}, \qquad k = 2, 3, \ldots \tag{21}$$

This means that

$$\lim_{t=\infty} e^{(n-1)t} \left(x - \sum_{k=1}^{n-1} a_k e^{-kt} \right) = 0, \quad \text{but} \quad \lim_{t=\infty} e^{nt} \left(x - \sum_{k=1}^{n-1} a_k e^{-kt} \right) =$$
$$= a_n, \qquad n = 1, 2, \ldots$$

and a similar relation holds concerning $y(t)$, i.e. these limits exists and have the said values.

The proof proceeds by induction. Denote

$$e^{(n+1)t} \left(x - \sum_{k=1}^{n} a_k e^{-kt} \right)$$

by $I_n(t)$ and suppose

$$\lim_{t=\infty} I_{n-1} = a_n, \quad a_1 = A, \quad b_1 = A + C, \quad a_k = b_k =$$
$$= -\frac{k}{k-1} d_k, \qquad k = 2, 3, \ldots, n \tag{22}$$

then we have to show

$$\lim_{t=\infty} I_n(t) = a_{n+1}, \quad a_{n+1} = -\frac{n+1}{n} d_{n+1}, \quad d_{n+1} = \sum_{l=1}^{n} a_l b_{n+1-l} \tag{23}$$

and a similar limit relation concerning $y(t)$. —

Equation (22) says

$$x = \sum_{k=1}^{n} a_k e^{-kt} + o(e^{-nt}) \tag{24}$$

and so

$$x^2 = \sum_{k=2}^{n+1} C_k e^{-kt} + o(e^{-(n+1)t}), \quad C_k = \sum_{l=1}^{k-1} a_l a_{k-l}$$

$$1 + 2x = 1 + 2\sum_{k=1}^{n} a_k e^{-kt} + o(e^{-nt}) \tag{25}$$

whence (s. (16))

$$f(x,t) = \frac{2\sum_{k=2}^{n+1} C_k e^{-kt} + 2C\sum_{k=1}^{n} a_k e^{-(k+1)t} + o(e^{-(n+1)t})}{1 + 2\sum_{k=1}^{n} a_k e^{-kt} + Ce^{-t} + o(e^{-nt})} =$$

$$= \frac{2\sum_{k=2}^{n+1} (C_k + Ca_{k-1}) e^{-kt} + o(e^{-(n+1)t})}{1 + (2a_1 + C)e^{-t} + 2\sum_{k=2}^{n} a_k e^{-kt} + o(e^{-nt})}$$

This may be brought on the form

$$f(x,t) = \sum_{k=2}^{n+1} \alpha_k e^{-kt} + o(e^{-(n+1)t})$$

In equating the two expressions of $f(x, t)$ and comparing the coefficients of equal powers we get $d_2 = 2(c_2 + Ca_1) = 2(a_1^2 + Ca_1)$ and

$$\alpha_k + \alpha_{k-1}(2a_1 + C) + 2\alpha_{k-2}a_2 + 2\alpha_{k-3}a_3 + \ldots + 2\alpha_2 a_{k-2} = \tag{26}$$

$$= 2(C_k + Ca_{k-1}), \qquad (k = 2, 3, \ldots, n+1)$$

which uniquely determines the α_k's. Asserted

$$\alpha_k = -(k-1)a_k = kd_k, \qquad k = 2, 3, \ldots, n+1 \ ^1 \tag{27}$$

what is valid for $k = 2$, since $\alpha_2 = 2(a_1^2 + Ca_1) = 2a_1(a_1 + C) = -a_2$. Really, the values (27) satisfy (26) as it will now be shown. — Denote the difference of the

[1] For $k = n+1$ the central term is defined by the right-hand term.

left and right members of (26) by D. Then replacing in (26) α_k by (27)

$$D = \alpha_k + \alpha_{k-1}(2a_1 + C) + 2\sum_{l=2}^{k-2} \alpha_{k-l} a_l - 2(C_k + Ca_{k-1}) = -(k-1)a_k -$$

$$-(k-2)a_{k-1}(2a_1 + C) - 2\sum_{l=2}^{k-2}(k-l-1)a_{k-l}a_l - 2\sum_{l=1}^{k-1} a_l a_{k-l} - 2Ca_{k-1} =$$

$$= -(k-1)a_k - kCa_{k-1} - 2\sum_{l=1}^{k-1}(k-l)a_l a_{k-l}, \qquad (k = 2, 3, \ldots, n)$$

But $-(k-1)a_k = kd_k$, $C = b_1 - a_1$, thus

$$k(d_k - b_1 a_{k-1} + a_1 a_{k-1}) = kC_k = k\sum_{l=1}^{k-1} a_l a_{k-l}$$

and so $D = \sum_{l=1}^{k-1}(2l-k)a_l a_{k-l}$. By interchanging l and $k-l$

$$D = \sum_{l=1}^{k-1}(k-2l)a_l a_{k-l} = -D, \qquad D = 0.$$

The last equation of (27) gives $\alpha_{n+1} = (n+1)d_{n+1}$.

PROOF of (23). With respect to (18) we have

$$I_n(t) = e^{(n+1)t}\left[Ae^{-t} - e^{-t}\int_t^\infty e^\tau f(x(\tau), \tau)\, d\tau - \sum_{k=1}^{n} a_k e^{-kt}\right]$$

Here

$$-e^{-t}\int_t^\infty e^\tau f(x(\tau), \tau)\, d\tau = -e^{-t}\int_t^\infty \left[e^\tau \sum_{k=2}^{n+1} \alpha_k e^{-k\tau} + o(e^{-(n+1)\tau})\right] d\tau =$$

$$= -\frac{\alpha_{n+1}}{n} e^{-(n+1)t} - \sum_{k=2}^{n} \frac{\alpha_k}{k-1} e^{-kt} + o(e^{-(n+1)t}).$$

Consequently,

$$I_n(t) = -\frac{\alpha_{n+1}}{n} - e^{(n+1)t}\sum_{k=2}^{n}\left(\frac{\alpha_k}{k-1} + a_k\right) e^{-kt} + o(1)$$

whence by (27) $\lim_{t=\infty} I_n(t) = -\frac{\alpha_{n+1}}{n} = a_{n+1}$. Having $y = x + Ce^{-t}$ the assertion concerning y follows at once.

3. 2. Take now the general case ($v \neq 1$)

(28) $$x' = -vx + f(x, y), \quad y' = -y + f(x, y), \quad f(x, y) = (v+1)\frac{xy}{1+x+y}$$

Letting

$$x = x_1, \quad y = x_2, \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \bar{x}, \quad \begin{bmatrix} -v & 0 \\ 0 & -1 \end{bmatrix} = \bar{A}, \quad \begin{bmatrix} f \\ f \end{bmatrix} = \bar{f}(\bar{x})$$

system (28) reads as

(28′) $$\bar{x}' = \bar{A}\bar{x} + \bar{f}(\bar{x})$$

which is equivalent to

$$\bar{x}(t) = e^{\bar{A}t}\bar{x}(0) + \int_0^t e^{\bar{A}(t-\tau)}\bar{f}(\bar{x}(\tau))\,d\tau, \quad e^{\bar{A}t} = \begin{bmatrix} e^{-\nu t} & 0 \\ 0 & e^{-t} \end{bmatrix} \tag{29}$$

Decompose $e^{\bar{A}t}$ as $e^{\bar{A}t} = \bar{B}_1(t) + \bar{B}_2(t)$ where

$$\bar{B}_1(t) = \begin{bmatrix} 0 & 0 \\ 0 & e^{-t} \end{bmatrix}, \quad \bar{B}_2(t) = \begin{bmatrix} e^{-\nu t} & 0 \\ 0 & 0 \end{bmatrix}$$

and replace (29) by

$$\bar{x}(t) = e^{\bar{A}t}\bar{a} + \int_0^t \bar{B}_1(t)e^{-\bar{A}\tau}\bar{f}(\bar{x}(\tau))\,d\tau - \int_t^\infty \bar{B}_2(t)e^{-\bar{A}\tau}\bar{f}(\bar{x}(\tau))d\tau \tag{30}$$

where for a while $\bar{a}$ means an arbitrary constant vector with *positive* coordinates a_1, a_2 and

$$\bar{B}_1(t)e^{-\bar{A}\tau} = \begin{bmatrix} 0 & 0 \\ 0 & e^{-t+\tau} \end{bmatrix}, \quad \bar{B}_2(t)e^{-\bar{A}\tau} = \begin{bmatrix} e^{-\nu(t-\tau)} & 0 \\ 0 & 0 \end{bmatrix}$$

The solution of (30) satisfies (28′) and conversely — at least for the solutions in question. — However it must be shown that (30) has a (unique) solution. Let us apply successive approximations for this end as follows

$$\bar{x}_0 = e^{\bar{A}t}\bar{a} \tag{31}$$

$$\bar{x}_{n+1} = e^{\bar{A}t}\bar{a} + \int_0^t \bar{B}_1(t)e^{-\bar{A}\tau}\bar{f}(\bar{x}_n(\tau))\,d\tau - \int_t^\infty \bar{B}_2(t)e^{-\bar{A}\tau}\bar{f}(\bar{x}_n(\tau))\,d\tau \qquad (n=0, 1, 2, \ldots)$$

The argument of the existence and "positivity" for large t of $\bar{x}_n(t)$ proceeds as follows. — Decomposed in components equation (31) reads as

$$x_0(t) = a_1 e^{-\nu t}, \quad y_0 = a_2 e^{-t}$$

$$x_{n+1} = e^{-\nu t}\left(a_1 - \int_t^\infty e^{\nu\tau} f(x_n(\tau), y_n(\tau))\,d\tau\right) \qquad (n = 0, 1, 2, \ldots) \tag{32}$$

$$y_{n+1} = e^{-t}\left(a_2 + \int_0^t e^{\tau} f(x_n(\tau), y_n(\tau))\,dt\right)$$

Suppose $0 < x_n \leqq Ke^{-\nu t}$, $0 < y_n \leqq Ke^{-t}$ for $t \geqq t_0$, (what is satisfied by x_0 and y_0 respectively provided $a_1, a_2 \leqq K$), then making use of $f(x, y) \leqq (\nu+1)xy$ $(x>0, y>0)$ we have

$$I_n = \int_t^\infty e^{\nu\tau} f(x_n, y_n)\,d\tau \leqq (\nu+1)K^2 \int_t^\infty e^{\nu\tau}e^{-(\nu+1)\tau}\,d\tau = (1+\nu)K^2 e^{-t}$$

i.e. I_n (and x_{n+1}) exists and $0 < x_{n+1} \leqq Ke^{-\nu t}$ for $t \geqq t_0$ provided

$$(1+\nu)K^2 e^{-t_0} < a_1 \leqq K - (1+\nu)K^2 e^{-t_0}.$$

Even as

$$J_n = \int_0^t e^{\tau} f(x_n, y_n)\,d\tau \leqq \int_0^t e^{\tau}(1+\nu)K^2 e^{-(\nu+1)\tau}\,d\tau = \frac{1+\nu}{\nu}K^2(1-e^{-\nu t})$$

and

$$0 < y_{n+1} \leqq e^{-t}\left(a_2 + \frac{1+v}{v} K^2(1-e^{-vt})\right) \leqq Ke^{-t}$$

provided $a_2 + \frac{1+v}{v} K^2 \leqq K$ what is satisfied for certain K when a_2 is small enough. Let us introduce the denotations

$$\Delta_0^1 = x_0, \quad \Delta_0^2 = y_0, \quad \Delta_n^1 = |x_n - x_{n-1}|, \quad \Delta_n^2 = |y_n - y_{n-1}|, \quad \Delta_n = \Delta_n^1 + \Delta_n^2$$

and regard the difference

$$\delta_n = f(x_n, y_n) - f(x_{n-1}, y_{n-1}) = f_x(\xi, \eta)(x_n - x_{n-1}) + f_y(\xi, \eta)(y_n - y_{n-1})$$

$$\xi = x_{n-1} + \theta(x_n - x_{n-1}), \; \eta = y_{n-1} + \theta(y_n - y_{n-1}), \quad 0 < \theta < 1$$

$$0 < \xi + \eta < 1$$

then $|\delta_n| \leqq \max(|f_x|, |f_y|)\Delta_n$, but

$$f_x = (1+v)\frac{y(1+y)}{(1+x+y)^2}, \quad f_y = (1+v)\frac{x(1+x)}{(1+x+y)^2}$$

and $\max(|f_x|, |f_y|) \leqq (1+v)(x+y)$. Consequently

$$|\delta_n| \leqq (1+v)K(e^{-vt} + e^{-t})\Delta_n$$

and by (32)

$$\Delta_{n+1}^1 \leqq (1+v)Ke^{-vt}\int_t^\infty (1 + e^{(v-1)\tau})\Delta_n d\tau$$

$$\Delta_{n+1}^2 \leqq (1+v)Ke^{-t}\int_0^t (1 + e^{(1-v)\tau})\Delta_n d\tau$$

Suppose now that $\Delta_n \leqq Lq^n(e^{-vt} + e^{-t})$, L, q constants, $0 < q < 1$ then

$$\Delta_{n+1}^1 \leqq (1+v)Ke^{-vt}Lq^n\int_t^\infty (1 + e^{(v-1)\tau})(e^{-v\tau} + e^{-\tau})\,d\tau =$$

$$= (1+v)KLq^n e^{-vt}\left(\frac{e^{-vt}}{v} + 2e^{-t} - \frac{e^{(v-2)t}}{v-2}\right)$$

(the expression in the last paranthesis is less than 1 for $t \geqq t_0$)

$$\Delta_{n+1}^2 \leqq (1+v)Ke^{-t}Lq^n\int_0^t (1 + e^{(1-v)\tau})(e^{-v\tau} + e^{-\tau})\,d\tau =$$

$$= 2(1+v)Ke^{-t}Lq^n\int_0^t [2e^{-v\tau} + e^{-\tau} + e^{(1-2v)\tau}]d\tau =$$

$$= (1+v)KLq^n e^{-t}\left[\frac{2}{v}(1-e^{-vt}) + \frac{e^{(1-2v)t} - 1}{1-2v} + (1-e^{-t})\right], \qquad \left(v \neq \frac{1}{2}\right)$$

(if $v=\frac{1}{2}$ the second term in the bracket is t)

$$\Delta_{n+1} \leqq (1+v)KLq^n\left(e^{-vt}+\frac{2}{v}e^{-t}+e^{-t}+\frac{1}{|1-2v|}e^{-2vt}+e^{-t}\right) =$$

$$= (1+v)KLq^n\left[e^{-vt}\left(1+\frac{1}{|1-2v|}e^{-vt}\right)+2\frac{1+v}{v}e^{-t}\right] \leqq$$

$$\leqq (1+v)KLq^n 2\frac{1+v}{v}(e^{-vt}+e^{-t}) = 2\frac{(1+v)^2}{v}KLq^n(e^{-vt}+e^{-t})$$

and so

$$\Delta_{n+1} \leqq Lq^{n+1}(e^{-vt}+e^{-t})$$

provided $2\dfrac{(1+v)^2}{v}K \leqq q$ (if $v=\dfrac{1}{2}$, then $te^{-t}<e^{-\frac{1}{2}t}$ and the same estimate holds for Δ_{n+1}) and really K may be chosen so if $|\bar{a}|$ is sufficiently small. Therefore $\Sigma\Delta_n$ is uniformly convergent for $t \geqq t_0$, and so are $\Sigma\Delta_n^1$ and $\Sigma\Delta_n^2$ too, i.e. $\lim\limits_{n=\infty} x_n = x$, $\lim\limits_{n=\infty} y_n = y$ exist, etc. and $|\bar{x}| \leqq 2C(e^{-vt}+e^{-t})$, moreover $0<x \leqq Ke^{-vt}$, $0<y \leqq Ke^{-t}$ (s. the rows following (32)). By (32)

$$x(t) = e^{-vt}\left(a_1 - \int_t^\infty e^{v\tau} f(x(\tau), y(\tau))\,d\tau\right)$$

(33)

$$y(t) = e^{-t}\left(a_2 + \int_0^t e^{\tau} f(x(\tau), y(\tau))\,d\tau\right)$$

which exhibits the fact that

(34) $$\lim_{t=\infty} x(t)e^{vt} = a_1$$

and it will be easily shown that

(35) $$\lim_{t=\infty} y(t)e^{t}$$

exists too, namely

$$a_2' = \int_0^\infty e^{\tau} f(x(\tau), y(\tau))\,d\tau \leqq \int_0^\infty e^{\tau}(1+v)x(\tau)y(\tau)\,d\tau \leqq (1+v)K^2\int_0^\infty e^{\tau}e^{-v\tau}e^{-\tau}\,d\tau =$$

$$= \frac{1+v}{v}K^2$$

So the value of the last limit is a_2+a_2'.

Now introduce the denotations $a_1 = a_{10}$, $a_2+a_2' = b_{01}$ and suppose 1° v is irrational $(0<v<1)$. Then asserted

(36) $$x(t) \sim \sum_{\substack{m,n=0\\ m+n>0}}^{\infty} a_{mn}e^{-(mv+n)t}, \quad y(t) \sim \sum_{\substack{m,n=0\\ m+n>0}}^{\infty} b_{mn}e^{-(mv+n)t}$$

the meaning of which is that the limits

(37) $$\lim_{t=\infty} I_{mn} = a_{mn}, \quad \lim_{t=\infty} J_{mn} = b_{mn} \qquad (m+n>0)$$

where

$$I_{mn} = e^{(mv+n)t}\left[x(t) - \sum_{kv+l<mv+n} a_{kl}e^{-(kv+l)t}\right]$$

(38)

$$J_{mn} = e^{(mv+n)t}\left[y(t) - \sum_{kv+l<mn+n} b_{kl}e^{-(kv+l)t}\right]$$

exist and

$$[(m-1)v+n]a_{mn} + (mv+n)c_{mn} = 0$$

(39)

$$[mv+n-1]a_{mn} + (mv+n)c_{mn} = 0 \qquad (m+n>0$$

$$c_{mn} = \sum_{p=0,q=0}^{m,n} a_{pq}b_{m-p,n-q}$$

hold. The last sum does not depend on a_{mn}, b_{mn}.

The proof proceeds by induction. In fact as it has been shown a_{10}, b_{01} exist and (34)—(35) imply that a_{01} and b_{10} exist and are zero. Namely assume $(k-1)v<$ $<1<kv$ ($k>0$ an integer) and e.g. (for the sake of simplicity) $k=2$. Then only the existence of the limit

$$\lim_{t=\infty} e^t(x - a_1e^{-vt}) = a_{01}, \qquad a_1 = a_{10}$$

must be shown (relation $b_{10} = \lim_{t=\infty} e^{vt}y = 0$ is clear by (35)). But

$$I = e^t(x - a_1e^{-vt}) = -e^{(1-v)t}\int_t^\infty e^{v\tau}f(x(\tau), y(\tau))\,d\tau$$

$$|I| \leqq e^{(1-v)t}\int_t^\infty e^{v\tau}(1+v)K^2e^{-v\tau}e^{-\tau}\,d\tau = (1+v)K^2e^{-vt} \to 0 \quad \text{as} \quad t \to \infty$$

Suppose (37) hold for $mv+n<m_0v+n_0$ and e.g. $m_0v<1<(m_0+1)v$. Let us prove the existence of a_{m_0+1,n_0}, b_{m_0+1,n_0}. (In the sequel the index 0 will be omitted.) By (37)

$$x = \sum_{kv+l\leqq mv+l} a_{kl}e^{-(kv+l)t} + o(e^{-(mv+n)t})$$

(40)

$$y = \sum_{kv+l\leqq mv+l} b_{kl}e^{-(kv+l)t} + o(e^{-(mv+n)t})$$

and by (33)

$$I_{m+1,n} = e^{[(m+1)v+n]t}\left[a_1e^{-vt} - e^{-vt}\int_t^\infty e^{v\tau}f(x(\tau), y(\tau))\,d\tau - \right.$$

$$\left. - \sum_{kv+l<(m+1)v+n} a_{kl}e^{-(kv+l)t}\right]$$

(41)

$$J_{m+1,n} = e^{[(m+1)v+n]t}\left[a_2e^{-t} + e^{-t}\int_0^t e^{\tau}f(x(\tau), y(\tau))\,d\tau - \right.$$

$$\left. - \sum_{kv+l<(m+1)v+n} b_{kl}e^{-(kv+l)t}\right]$$

Here

$$f(x, y) = (1+v)\frac{xy}{1+x+y}$$

and

$$xy = \sum_{1+v \leqq kv+l \leqq (m+1)v+n} c_{kl} e^{-(kv+l)t} + o(e^{-[(m+1)v+n]t})$$

$$1+x+y = 1+\sum_{0<kv+l \leqq mv+n} (a_{kl}+b_{kl}) e^{-(kv+l)t} + o(e^{-(mv+n)t})$$

whence

$$f(x, y) = \sum_{1+v \leqq kv+l \leqq (m+1)v+n} \alpha_{kl} e^{-(kv+l)t} + o(e^{-[(m+1)v+n]t}), \tag{42}$$

where the coefficients α_{kl} are uniquely determined by the following recursive formulae

$$(1+v)c_{kl} = \alpha_{kl} + \sum_{\substack{p,q=0 \\ 0<p+q \leqq k+l-2}}^{k-1,\,l-1} (a_{pq}+b_{pq})\alpha_{k-p,\,l-q} \tag{43}$$

Asserted

$$\alpha_{kl} = (kv+l)c_{kl} = -[(k-1)v+l]a_{kl}, \quad 2 \leqq k+l \leqq (m+1)+n. \tag{44}$$

(For $k = m+1$, $l=n$, the right-hand member is defined just by the central member.) This value of α_{kl} must satisfy (43). Denote the difference of the left and right members of (43) by D, then we have

$$D = (1+v)c_{kl} - (kv+l)c_{kl} + \sum_{0<p+q<k+l} (a_{pq}+b_{pq})[(k-p)v+l-q]c_{k-p,l-q} =$$

$$= [(1-k)v+1-l]c_{kl} + \sum_{0<p+q<k+l} [(k-p)v+l-q]\left[\frac{(k-p)v+l-q-1}{(k-p)v+l-q} a_{pq} b_{k-p,\,l-q} + \right.$$

$$\left. + \frac{(k-p-1)v+l-q}{(k-p)v+l-q} b_{pq} a_{k-p,\,l-q}\right] = [(1-k)v+1-l]c_{kl} +$$

$$+ \sum_{0<p+q<k+l} [(k-p)v+l-q-1]a_{pq} b_{k-p,\,l-q} +$$

$$+ \sum_{0<p+q<k+l} [(k-p-1)v+l-q] b_{pq} a_{k-p,\,l-q}$$

By interchanging in the last sum p with $k-p$ and q with $l-q$ respectively

$$D = [(1-k)v+1-l]c_{kl} + \sum \underbrace{[(k-p)v+l-q-1+(p-1)v+q]}_{(k-1)v+l-1} a_{pq} b_{k-p,\,l-q} =$$

$$= [(1-k)v+1-l]c_{kl} + [(k-1)v+l-1]c_{kl} = 0$$

what proves (44). — Replace $f(x, y)$ in (41) by (42). Obtaining

$$I_{m+1,n} = e^{[(m+1)\nu+n]t}\left[a_{10}e^{-\nu t} - e^{-\nu t}\int_t^\infty e^{\nu\tau}\alpha_{m+1,n}e^{-[(m+1)\nu+n]\tau}\,d\tau - \right.$$

$$-e^{-\nu t}\int_t^\infty e^{\nu\tau}\sum_{2\leqq k+l\leqq(m+1)+n}\alpha_{kl}e^{-(k\nu+l)t}\,d\tau - \sum_{1\leqq k+l\leqq(m+1)+n}a_{kl}e^{-(k\nu+l)t} +$$

$$\left. + o(e^{-[(m+1)\nu+n]t})\right] = -\frac{\alpha_{m+1,n}}{m\nu+n} -$$

$$-e^{[(m+1)\nu+n]t}\sum_{2\leqq k+l<(m+1)+n}\left(\frac{\alpha_{kl}}{(k-1)\nu+l}+a_{kl}\right)e^{-(k\nu+l)t}+o(1)$$

and by (44) $\lim\limits_{t=\infty} I_{m+1,n} = a_{m+1,n}$. — Analogously

$$J_{m+1,n} = e^{[(m+1)\nu+n]t}\left[e^{-t}\underbrace{\left(a_2+\int_0^\infty e^{\tau}f(x(\tau),y(\tau))\,d\tau\right)}_{b_{01}} - \right.$$

$$-e^{-t}\int_0^\infty e^{\tau}\alpha_{m+1,n}e^{-[(m+1)\nu+n]\tau}\,d\tau - e^{-t}\int_t^\infty e^{\tau}\sum_{2\leqq k+l<m+1+n}\alpha_{kl}e^{-(k\nu+l)\tau}\,d\tau -$$

$$\left. - \sum_{1\leqq k+l<m+1+n}b_{kl}e^{-(k\nu+l)t} + o(e^{-[(m+1)\nu+n]t}\right] = -\frac{\alpha_{m+1,n}}{(m+1)\nu+n-1} -$$

$$-e^{[(m+1)\nu+n]t}\sum_{1\leqq k+l<m+1+n}\left(\frac{\alpha_{kl}}{k\nu+l-1}+b_{kl}\right)e^{-(k\nu+l)t}+o(1)$$

whence by (44) $\lim\limits_{t=\infty} J_{m+1,n} = b_{m+1,n}$.

It can be proved in the same way

$$\lim_{t=\infty} I_{m,n+1} = a_{m,n+1}, \qquad \lim_{t=\infty} J_{m,n+1} = b_{m,n+1}$$

2° If $0<\nu<1$ is *rational* $\nu=\frac{p}{q}$, $(p, q)=1$, then the following asymptotic expansions hold

$$x \sim \sum_{\substack{m=0\\ m+n>0}}^{q-1}\sum_{n=0}^{\infty} a_{mn}e^{-\left(\frac{m}{q}+n\right)t}, \qquad y \sim \sum_{\substack{m=0\\ m+n>0}}^{q-1}\sum_{n=0}^{\infty} b_{mn}e^{-\left(\frac{m}{q}+n\right)t} \tag{45}$$

where a_{mn}, b_{mn} are defined by (11)—(12), i.e. by

$$(m-p+nq)a_{mn}+(m+nq)c_{mn} = 0$$

$$[m+(n-1)q]b_{mn}+(m+nq)c_{mn} = 0 \qquad (m+n>0) \tag{46}$$

$$c_{mn} = \sum_{0<r+s<m+n}\sum a_{rs}b_{m-r,\,n-s}$$

The proof is inductive now too. Equations (34)—(35) assure the existences of a_{p0} and b_{01} and imply that

$$a_{k0} = 0,\ k < p,\ b_{k0} = 0,\ k \leqq q-1.$$

Suppose the existences of the limits

$$\lim_{t=\infty} I_{mn} = \lim_{t=\infty} e^{\left(\frac{m}{q}+n\right)t}\left(x - \sum_{\frac{k}{q}+l<\frac{m}{q}+n} a_{kl}e^{-\left(\frac{k}{q}+l\right)t}\right) = a_{mn}$$

(47)

$$\lim_{t=\infty} J_{mn} = \lim_{t=\infty} e^{\left(\frac{m}{q}+n\right)t}\left(y - \sum_{\frac{k}{q}+l<\frac{m}{q}+n} b_{kl}e^{-\left(\frac{k}{q}+l\right)t}\right) = b_{mn}$$

$$m \leqq m_0, \quad n \leqq n_0$$

with a_{kl}, b_{kl} given by (46) the validity of which is assumed for $m \leqq m_0$, $n \leqq n_0$. Then we have to show (index zero omitted)

a) the existences of $\lim_{t=\infty} I_{m+1,n} = a_{m+1,n}$, $\lim_{t=\infty} J_{m+1,n} = b_{m+1,n}$ provided $m < q-1$,

b) that of $\lim_{t=\infty} I_{0,n+1} = a_{0,n+1}$, $\lim_{t=\infty} J_{0,n+1} = b_{0,n+1}$ provided $m = q-1$.

See first the case

a) then

(48)

$$I_{m+1,n} =$$

$$= e^{\left[\frac{m+1}{q}+n\right]t}\left[a_{10}e^{-\nu t} - e^{-\nu t}\int_t^\infty e^{\nu\tau} f(x(\tau), y(\tau))\,d\tau - \sum_{\frac{k}{q}+l<\frac{m+1}{q}+n} a_{kl}e^{-\left[\frac{k}{q}+l\right]t}\right]$$

now by (47)

$$x = \sum_{\frac{k}{q}+l\leqq\frac{m}{q}+n} a_{kl}e^{-\left[\frac{k}{q}+l\right]t} + o\left(e^{-\left[\frac{m}{q}+n\right]t}\right),$$

(49)

$$y = \sum_{\frac{k}{q}+l\leqq\frac{m}{q}+n} b_{kl}e^{-\left[\frac{k}{q}+l\right]t} + o\left(e^{-\left[\frac{m}{q}+n\right]t}\right)$$

whence

$$xy = \sum_{\nu+1\leqq\frac{k}{q}+l\leqq\frac{m+1}{q}+l} c_{kl}e^{-\left[\frac{k}{q}+l\right]t} + o\left(e^{-\left[\frac{m+1}{q}+n\right]t}\right)$$

(50)

$$1+x+y = 1+\sum_{\frac{k}{q}+l\leqq\frac{m}{q}+n} (a_{kl}+b_{kl})e^{-\left[\frac{k}{q}+l\right]t} + o\left(e^{-\left[\frac{m}{q}+n\right]t}\right)$$

and

(51)

$$f(x,y) = \sum_{\nu+1\leqq\frac{k}{q}+l\leqq\frac{m+1}{q}+n} \alpha_{kl}e^{-\left[\frac{k}{q}+l\right]t} + o\left(e^{-\left[\frac{m+1}{q}+n\right]t}\right)$$

By comparison of the coefficients

$$(52)\qquad (\nu+1)c_{kl}=\alpha_{kl}+\sum_{0<r+s<k+l-2}(a_{rs}+b_{rs})\alpha_{k-r,l-s},\qquad 2\leqq k+l\leqq(m+1)+n$$

Asserted

$$(53)\qquad \alpha_{kl}=\frac{k+lq}{q}c_{kl}=-\frac{k-p+lq}{q}a_{kl}=-\frac{k+(l-1)q}{q}b_{kl}$$

(for $k=m+1$, $l=n$ the a_{kl}, b_{kl} are defined by c_{kl}.) The difference of the left and right sides of (52) be denoted by D. Then making use of (53)

$$D=(\nu+1)c_{kl}-\frac{k+lq}{q}c_{kl}+\sum_{0<r+s\leqq k+l-2}\left(\frac{k-r+(l-s-1)q}{q}a_{rs}b_{k-r,l-s}+\right.$$

$$\left.+\frac{r-p+sq}{q}a_{rs}b_{k-r,\,l-s}\right)=\frac{p-k+(1-l)q}{q}c_{kl}+\frac{k-p+(l-1)q}{q}c_{kl}=0$$

Putting (51) in (48)

$$I_{m+1,n}=e^{\left[\frac{m+1}{q}+n\right]t}\left[a_{10}e^{-\nu t}-e^{-\nu t}\int_t^\infty e^{\nu t}\alpha_{m+1,n}e^{-\left[\frac{m+1}{q}+n\right]\tau}d\tau-\right.$$

$$-e^{-\nu t}\int_t^\infty e^{\nu\tau}\sum_{\nu+1\leqq\frac{k}{q}+l<\frac{m+1}{q}+l}\alpha_{kl}e^{-\left[\frac{k}{q}+l\right]\tau}d\tau-\sum_{\frac{k}{q}+l<\frac{m+1}{q}+n}a_{kl}e^{-\left[\frac{k}{q}+l\right]t}+$$

$$\left.+o\left(e^{-\left[\frac{m+1}{q}+n\right]t}\right)\right]=-\frac{q}{m-p+nq}\alpha_{m+1,n}-$$

$$-e^{\left[\frac{m+1}{q}+n\right]t}\sum_{\nu+1\leqq\frac{k}{q}+l<\frac{m+1}{q}+n}\left(\frac{q}{k-p+lq}\alpha_{kl}+a_{kl}\right)e^{-\left[\frac{k}{q}+l\right]t}+o(1)$$

By (53) $I_{m+1,n}\to a_{m+1,n}$ as $t\to\infty$ and even as $J_{m+1,n}\to b_{m+1,n}$ as $t\to\infty$.

Case b).

$$I_{0,n+1}=e^{(n+1)t}\left[a_{10}e^{-\nu t}-e^{-\nu t}\int_t^\infty e^{\nu\tau}\sum_{\nu+1<\frac{k}{q}+l\leqq n+1}\alpha_{kl}e^{-\left[\frac{k}{q}+l\right]t}d\tau-\right.$$

$$\left.-\sum_{\frac{k}{q}+l<n+1}a_{kl}e^{-\left[\frac{k}{q}+l\right]t}+o(e^{-(n+1)t})\right]=$$

$$=e^{(n+1)t}\left[a_{10}e^{-\nu t}-e^{-\nu t}\int_t^\infty\alpha_{0,n+1}e^{-(n+1)\tau}d\tau-\right.$$

$$\left.-e^{-\nu t}\int_t^\infty e^{\nu\tau}\sum_{\nu+1<\frac{k}{q}+l<n+1}\alpha_{kl}e^{-\left[\frac{k}{q}+l\right]\tau}d\tau-\sum_{\frac{k}{q}+l<n+1}a_{kl}e^{-\left[\frac{k}{q}+l\right]t}\right]+o(1)=$$

and in the same way $J_{0,n+1}\to b_{0,n+1}$.

Now the proof is complete. The formal process of paragraph 2 gives really the unique asymptotic series of the actual solutions. Assertion concerning their derivatives can be proved in the usual way or by the differential equation.

The question of convergence of these series is not yet decided.

4. We raise the problem of finding the parameters λ, μ from the "initial values" $x(t), y(t)$ if t is sufficiently large.

First a lemma is needed:

LEMMA. Let $f(\lambda, \mu), g(\lambda, \mu)$ be defined and differentiable in a domain D of the λ, μ plan and let the partial derivatives $f_\lambda, f_\mu, g_\lambda, g_\mu$ satisfy

$$|f_\lambda|, |f_\mu|, |g_\lambda|, |g_\mu| \leqq \varrho, \ \lambda, \mu \in D, \ \varrho = \text{const} < \tfrac{1}{2}.$$

Then there exist a pair of values λ^*, μ^* satisfying

$$\lambda = f(\lambda, \mu), \ \mu = g(\lambda, \mu) \tag{54}$$

and λ^*, μ^* may be obtained by the iterations

$$\lambda_{n+1} = f(\lambda_n, \mu_n) \qquad (n = 0, 1, 2, \ldots)$$
$$\mu_{n+1} = g(\lambda_n, \mu_n)$$

where $\lambda_0, \mu_0 \in D$ are arbitrary values.

PROOF. By the denotations

$$\Delta_0 = |\lambda_0| + |\mu_0|, \ \Delta_n = |\lambda_n - \lambda_{n-1}| + |\mu_n - \mu_{n-1}| \qquad (n = 1, 2, \ldots)$$

we have

$$\lambda_{n+1} - \lambda_n = f(\lambda_n, \mu_n) - f(\lambda_{n-1}, \mu_{n-1}) =$$
$$= f_\lambda(\xi, \eta)(\lambda_n - \lambda_{n-1}) + f_\mu(\xi, \eta)(\mu_n - \mu_{n-1})$$

which and its analogous concerning $\mu_{n+1} - \mu_n$ give

$$|\lambda_{n+1} - \lambda_n| \leqq \varrho \Delta_n, \ |\mu_{n+1} - \mu_n| \leqq \varrho \Delta_n$$
$$\Delta_{n+1} \leqq 2\varrho \Delta_n$$

i.e. the series $\sum_{n=1}^{\infty} \Delta_n$ is convergent and the series

$$\lambda_0 + \sum_{n=1}^{\infty} (\lambda_n - \lambda_{n-1}), \quad \mu_0 + \sum_{n=1}^{\infty} (\mu_n - \mu_{n-1})$$

are absolute convergent, the limits $\lim_{n=\infty} \lambda_n = \lambda^*$, $\lim_{n=\infty} \mu_n = \mu^*$ exist and satisfy (54).

Now according paragraph 3.1 equation (24) says

$$x(t) = \sum_{k=1}^{n} a_k e^{-kt} + \varepsilon_n' e^{-nt} \quad a_1 = \lambda, \ b_1 = \mu, \ a_k = a_k(\lambda, \mu)$$
$$y(t) = \sum_{k=1}^{n} b_k e^{-kt} + \eta_n' e^{-nt} \quad b_k = b_k(\lambda, \mu), \ \varepsilon_n' = \varepsilon_n'(t, \lambda, \mu) \tag{55}$$
$$\eta_n' = \eta_n'(t, \lambda, \mu)$$

where

$$\lim_{t\to\infty} \varepsilon_n' = \lim_{t\to\infty} \eta_n' = 0$$

whence

$$a_1 = \lambda = x(t)e^t - \sum_{k=2}^{n} a_k e^{-(k-1)t} + \varepsilon_n(t, \lambda, \mu)e^{-(n-1)t}, \qquad \varepsilon_n \to 0,\ t \to \infty,\ \varepsilon_n = -\varepsilon_n'$$

which and the analogous formula have the forms

$$\lambda = f_n(t, \lambda, \mu) + \varepsilon_n(t, \lambda, \mu)e^{-(n-1)t} \tag{56}$$

$$\mu = g_n(t, \lambda, \mu) + \eta_n(t, \lambda, \mu)e^{-(n-1)t} \qquad \varepsilon_n, \eta_n \to 0,\ t \to \infty.$$

The f_n and g_n are polynomials in λ, μ of the form

$$\sum_{i,k} A_{ik}\lambda^{p_i}\mu^{q_k}, \quad 2 \leqq p_i + q_k \leqq n-1$$

having partial derivatives $f_{n\lambda}$, ... which are arbitrary small in absolute value if $|\lambda|$ and $|\mu|$ are small enough.

To prove that ε_n and η_n have (bounded) partial derivatives in D an induction will be applied. — Assertion holds for $n=1$. Make use of the denotations

$$I_{n-1} = e^{nt}\left(x - \sum_{k=1}^{n-1} a_k e^{-kt}\right) = a_n - \varepsilon_n \tag{57}$$

of 3.1 and suppose a_n has the said derivatives. Then in

$$I_n = a_{n+1} - \varepsilon_{n+1}$$

ε_{n+1} is built up by entire rational operations from a_i $(i=1, 2, \ldots, n)$ and ε_n, consequently it has the required derivatives.

Therefore conditions of the Lemma are satisfied by (56). — Nevertheless the problem arises whether are the parameters λ, μ obtainable with some precision using instead of (56) the truncated system

$$\lambda = f_n(t, \lambda, \mu),\ \mu = g_n(t, \lambda, \mu). \tag{58}$$

The answer is affirmative. — Namely equations

$$\lambda_{N+1} = f_n(t, \lambda_N, \mu_N) + \varepsilon_n(t, \lambda_N, \mu_N)e^{-(n-1)t} \qquad (N = 0, 1, 2, \ldots)$$

$$\lambda'_{N+1} = f_n(t, \lambda'_N, \mu'_N) \qquad (\lambda'_0 = \lambda_0,\ \mu'_0 = \mu_0)$$

imply

$$\lambda_{N+1} - \lambda'_{N+1} = f_n(t, \lambda_N, \mu_N) - f_n(t, \lambda'_N, \mu'_N) + \varepsilon_n(t, \lambda_N, \mu_N)e^{-(n-1)t} = \tag{59}$$

$$= f_{n\lambda}(t, \xi, \eta)(\lambda_N - \lambda'_N) + f_{n\mu}(t, \xi, \eta)(\mu_N - \mu'_N) + \varepsilon_n(t, \lambda_N, \mu_N)e^{-(n-1)t}$$

where ξ and η are suitable intermediate values. If

$$|f_{n\lambda}|,\ |f_{n\mu}|,\ |g_{n\lambda}|,\ |g_{n\mu}| \leqq \varrho < \frac{1}{2}$$

and

$$|\lambda_N - \lambda'_N| + |\mu_N - \mu'_N| = \Delta_N$$

then by (59) and its analogous for μ

$$\Delta_{N+1} \leqq 2\varrho\Delta_n + |\varepsilon_n(t, \lambda_N, \mu_N)|e^{-(n-1)t}$$

which involves

$$\Delta_{N+1} \leqq (2\varrho)^{N+1}\Delta_0 + (|\varepsilon_n^N| + 2\varrho\,|\varepsilon_n^{N-1}| + \ldots + (2\varrho)^N\,|\varepsilon_n^0|)e^{-(n-1)t}$$

But $\Delta_0 = 0$ and $|\varepsilon_n^i| \leqq \varepsilon_n$ $(i = 0, 1, 2, \ldots, N)$ where $\varepsilon_n \to 0$ as $t \to \infty$. So

$$\Delta_{N+1} \leqq \varepsilon_n \frac{1}{1-2\varrho} e^{-(n-1)t}$$

that is

$$|\lambda - \lambda'|,\ |\mu - \mu'| = o(e^{-(n-1)t}).$$

This shows that system (58) can be applied determining λ and μ provided t and n are sufficiently large.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received May 20, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 279—286.

SUR UNE CLASSE PARTICULIÈRE DE SÉRIES DE FOURIER À CERTAINES PUISSANCES ABSOLUMENT CONVERGENTES

par

L. ALPÁR

§ 1. Introduction

Nous généralisons dans cette note deux résultats de J. P. KAHANE ([2], Théorèmes IV et V, pp. 253—254).

Soit $\mathscr{A}$ l'ensemble des fonctions

$$F(t) = \sum_{n=-\infty}^{\infty} a_n e^{int}, \quad \text{avec} \quad \|F\| = \sum_{n=-\infty}^{\infty} |a_n| < \infty.$$

Les propositions mentionnées concernent des fonctions qui appartiennent à $\mathscr{A}$ et sont en outre de la forme

$$e^{ivf(t)} = \sum_{n=-\infty}^{\infty} a_{nv} e^{int} \qquad (v = 0, \pm 1, \pm 2, \ldots)$$

où t et $f(t)$ sont réelles. $e^{if} \in \mathscr{A}$ entraîne évidemment $e^{ivf} \in \mathscr{A}$. Avec ces notations les théorèmes en question s'énoncent comme suit.

THÉORÈME IV. — *Si, sur un segment I arbitrairement petit, f est deux fois dérivable avec $f''(t) > \varkappa > 0$, et si $e^{if} \in \mathscr{A}$, il existe une constante positive $\lambda = \lambda(f)$ telle que $\|e^{ivf}\| > \lambda |v|^{1/2}$ ($-\infty < v$ entier $< \infty$).*

REMARQUE 1. — La condition imposée à $f''(t)$ ($t \in I$) assure qu'il existe au moins un intervalle où $f(t)$ n'est pas linéaire. Cette restriction est inévitable, étant donné que selon le théorème III de KAHANE ([2], p. 251) *si $f(t)$ est continue et linéaire par intervalle, et si $f(t) \equiv f(t+2\pi) \pmod{2\pi}$, on a $\|e^{ivf}\| = O(\log |v|)$*, et dans ces conditions le théorème IV ne peut pas subsister.

THÉORÈME V. — *Si f est réelle, analytique, périodique et de période 2π, et non constante, il existe deux constantes positives $\lambda_1 = \lambda_1(f)$ et $\lambda_2 = \lambda_2(f)$ telles que $\lambda_1 |v|^{1/2} < \|e^{ivf}\| < \lambda_2 |v|^{1/2}$ ($-\infty < v$ entier $< \infty$).*

En tenant compte de ces théorèmes nous allons étudier les problèmes suivants.

1° Il découle de ces résultats que dans les deux cas envisagés

$$\lim_{|v| \to \infty} \sum_{n=-\infty}^{\infty} |a_{nv}| = \infty,$$

tandis que selon la formule de Parseval

$$\sum_{n=-\infty}^{\infty} |a_{nv}|^2 = 1 \qquad (v = 0, \pm 1, \pm 2, \ldots).$$

Il y a donc lieu de soulever la question: Quelles sont les bornes inférieures et supérieures des sommes

$$\sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} \tag{1.1}$$

si $0<\alpha<1$ et f remplit certaines conditions qui seront précisées? Plus généralement, que peut-on dire sur les sommes (1. 1) lorsque $0<\alpha<2$? Comment se change la majoration de (1. 1) quand $\alpha<0$? Cette somme étant alors bornée, on cherchera à déterminer sa borne supérieure en fonction de v.

2° Est-il nécessaire d'admettre la stipulation $e^{if}\in\mathscr{A}$ intervenant dans le théorème IV?

3° Peut-on remplacer dans le théorème V l'analyticité de f par une condition plus faible?

Dans cet ordre d'idées le § 2 est consacré à la généralisation du théorème IV et le § 3 à celle du théorème V. Le § 4 traite enfin le cas où $\alpha<0$, mais seulement pour des $f(t)$ analytiques le long de l'axe réel.

*

Nous désignerons par c, c', c_i, C, C_i $(i=0, 1, 2, \ldots)$ des constantes numériques positives ou des quantités variant entre deux limites positives bornées. Les symboles O et o seront employés, sauf avis contraire, au sens de $|v|\to\infty$. C^k $(k=1, 2, \ldots)$ resp. C^∞ dénote la classe des fonctions k fois continûment resp. indéfiniment dérivables.

§ 2. Fonctions dérivables sur un segment

THÉORÈME 1. — *Si, sur un segment I arbitrairement petit, f est deux fois dérivable avec $f''(t)>\varkappa>0$ et α $(0<\alpha<2)$ est un nombre donné, il existe une constante positive $\lambda=\lambda(f, \alpha)$ telle que*

$$\lambda|v|^{\alpha/2} < \sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} \qquad (v = 0, \pm1, \pm2, \ldots). \tag{2.1}$$

DÉMONSTRATION. — Lorsque

$$\sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} = \infty, \tag{2.2}$$

la relation (2.1) est automatiquement vérifiée. Les cas intéressants sont donc ceux où (2. 2) n'a pas lieu, c'est ce que nous admettons pour la suite.

Nous allons utiliser une méthode due à Z. L. LEIBENZON [3], employée aussi par KAHANE. Soit

$$a_{nv} = b_{nv} + c_{nv}, \quad b_{nv} = \frac{1}{2\pi}\int_I e^{ivf(t)-int}\,dt.$$

D'après un lemme de VAN DER CORPUT ([4], I., p. 197) $|b_{nv}|<(2/\pi)(\varkappa|v|)^{-1/2}=\varepsilon$ pour tout n. Si $K>1$ est une constante, il résulte de $|a_{nv}|>K\varepsilon$ que $|c_{nv}|>\dfrac{K-1}{K}|a_{nv}|$ et, par suite,

$$\sum_{|a_{nv}|>K\varepsilon} |a_{nv}|^2 \leqq \left(\frac{K}{K-1}\right)^2 \sum_{n=-\infty}^{\infty} |c_{nv}|^2 = \left(\frac{K}{K-1}\right)^2 \left(1-\frac{|I|}{2\pi}\right) = 1-\theta.$$

Choisissons K pour que l'on ait $\theta>0$. On a alors

$$\sum_{n=-\infty}^{\infty}|a_{nv}|^{2-\alpha}\geqq\sum_{|a_{nv}|\leqq K\varepsilon}|a_{nv}|^{2-\alpha}\geqq\frac{1}{(K\varepsilon)^{\alpha}}\sum_{|a_{nv}|\leqq K\varepsilon}|a_{nv}|^{2}\geqq\frac{\theta}{(K\varepsilon)^{\alpha}}$$

et c'est l'inégalité (2. 1).

REMARQUE 2. — En faisant abstraction du cas trivial signalé sous (2. 2), $e^{if}\notin\mathscr{A}$ même si la série (2.1) converge, mais avec un exposant de convergence supérieur à 1. (2.1) a lieu alors avec $2-\alpha>1$ et avec la borne inférieure $\lambda|v|^{\alpha/2}$ où $\alpha/2<1/2$.

§ 3. Fonctions dérivables sur l'axe rèel

Nous allons maintenant prouver que la condition d'analyticité du théorème V peut être essentiellement attenuée.

Soit $\mathscr{A}^{\beta}$ resp. $\mathscr{A}^{\beta+}$ $(\beta>0)$ l'ensemble des fonctions

$$h(t)=\sum_{n=-\infty}^{\infty}b_n e^{int},\quad\text{avec}\quad\sum_{n=-\infty}^{\infty}|b_n|^{\beta}<\infty\quad\text{resp.}\quad\sum_{n=-\infty}^{\infty}|b_n|^{\beta+\eta}<\infty$$

où $\eta>0$ est aussi petit que l'on veut. On écrira $\mathscr{A}$ au lieu de $\mathscr{A}^1$. On peut définir de même, avec $\beta=0$, l'ensemble de fonctions $\mathscr{A}^{0+}$. Il est simple de voir que, pour $0<\beta\leqq 1$, $h\in\mathscr{A}^{\beta}$ resp. $\mathscr{A}^{\beta+}$ entraîne $h^{v}\in\mathscr{A}^{\beta}$ resp. $\mathscr{A}^{\beta+}$. Lorsque $\beta>1$ ce n'est pas toujours le cas. Néanmois si $h(t)\neq 0$ est à variation bornée, on a $h^{v}\in\mathscr{A}^{1+}$ pour tout v. Dans ce qui suit nous allons considérer des fonctions $h(t)=e^{if(t)}$ où $f(t)$ est à variation bornée et, par suite, $e^{ivf(t)}\in\mathscr{A}^{1+}$.

THÉORÈME 2. — *Si $f\in C^1$, et si f' est à variation bornée sur l'intervalle $[-\pi,\pi]$, de plus si $f(t)\equiv f(t+2\pi)\pmod{2\pi}$, alors $e^{if}\in\mathscr{A}^{\frac{1}{2}+}$ et, pour $1/2<2-\alpha\leqq 2$, il existe une constante positive $\lambda_0=\lambda_0(f,\alpha)$ telle que*

$$\sum_{n=-\infty}^{\infty}|a_{nv}|^{2-\alpha}<\lambda_0|v|^{\alpha/2}\qquad(v=0,\pm1,\pm2,\ldots).\tag{3.1}$$

DÉMONSTRATION. — Il découle des conditions imposées à $f(t)$ que e^{ivf} est périodique de période 2π et que

$$f(t)=g(t)+\gamma t\tag{3.2}$$

où $g(t)\in C^1$, est périodique de période 2π, $g'(t)$ est à variation bornée sur $[-\pi,\pi]$ et γ $(-\infty<\gamma<\infty)$ est un entier. En effet, les hypothèses faites sur f montrent que $[f(t)-f(t+2\pi)]/2\pi$ est d'une part une fonction continue d'autre part un entier. C'est donc une constante à valeur entière qu'on note par γ. Il s'ensuit que $f'(t)-f'(t+2\pi)=0$ c'est-à-dire que $f'(t)$ est périodique de période 2π. C'est qui vérifie notre assertion.

Lorsque $n\neq 0$ et $vn^{-1}f'(t)-1\neq 0$ quel que soit t, on a

$$a_{nv}=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{ivf(t)-int}\,dt=\frac{1}{2\pi in}\int_{-\pi}^{\pi}\frac{de^{ivf(t)-int}}{vn^{-1}f'(t)-1}.\tag{3.3}$$

Or, pour chaque v fixé on peut déterminer un entier $n_0>0$ tel que l'on ait 1) $n_0=l|v|$, $l>1$ étant une constante, et 2) $|vn^{-1}f'(t)-1|>c_1>0$ pour $|n|\geqq n_0$. En tenant compte des propriétés énumérées de e^{ivf}, f et f' et en intégrant par parties, on obtient à partir de (3. 3), pour $|n|\geqq n_0$,

$$|a_{nv}| \leqq \frac{1}{2\pi|n|}\left|\int_{-\pi}^{\pi} e^{iv f(t)-int}\,d\big(vn^{-1}f'(t)-1\big)^{-1}\right| \leqq \frac{|v|}{2\pi|n|^2}\int_{-\pi}^{\pi}\frac{|df'(t)|}{[vn^{-1}f'(t)-1]^2} \leqq c_2\frac{|v|}{n^2}.$$

Par conséquent, la série $\sum_{|n|\geqq n_0}|a_{nv}|^{2-\alpha}$ converge si $1/2<2-\alpha$. Cette condition étant supposée remplie, on a

$$\sum_{|n|\geqq n_0}|a_{nv}|^{2-\alpha} = O(|v|^{-1+\alpha}) \tag{3. 4}$$

et $-1+\alpha<\alpha/2$ si $\alpha<2$, ce qui est assuré par l'inégalité $\alpha<3/2$.

Soit maintenant $\varepsilon=|v|^{-\frac{1}{2}}$. Un raisonnement analogue à celui que nous avons employé dans la démonstration du théorème 1 permet d'écrire

$$\sum_{|n|<n_0}|a_{nv}|^{2-\alpha} \leqq \frac{1}{(K\varepsilon)^{\alpha}}\sum_{\substack{|a_{nv}|>K\varepsilon\\ |n|<n_0}}|a_{nv}|^2+\sum_{\substack{|a_{nv}|\leqq K\varepsilon\\ |n|<n_0}}|a_{nv}|^{2-\alpha}=S_1+S_2, \tag{3. 5}$$

Il est clair que $S_1=O(|v|^{\alpha/2})$ et que la somme S_2 contient moins que $2n_0$ termes. On a donc $S_2=O(|v|^{\alpha/2})$. L'inégalité (3. 1) est, par suite, la conséquence de (3. 4) et (3. 5).

Les théorèmes 1 et 2 laissent prévoir l'existence d'une proposition précisant simultanément les bornes supérieures et inférieures de la somme considérée.

THÉORÈME 3. — *Si* $f\in C^k$ $(k\geqq 2)$, *si* f *est non linéaire, et si* $f(t)\equiv f(t+2\pi)$ (mod 2π), *alors* $e^{if}\in\mathscr{A}^{1/k+}$ *et, pour* $k^{-1}<2-\alpha\leqq 2$, *il existe deux constantes positives* $\lambda_1=\lambda_1(f,\alpha)$ *et* $\lambda_2=\lambda_2(f,\alpha)$ *telles que*

$$\lambda_1|v|^{\alpha/2}<\sum_{n=-\infty}^{\infty}|a_{nv}|^{2-\alpha}<\lambda_2|v|^{\alpha/2} \qquad (v=0,\pm1,\pm2,\ldots). \tag{3. 6}$$

Si de plus $f^{(k)}(t)$ *est à variation bornée sur l'intervalle* $[-\pi,\pi]$, *alors* $f\in\mathscr{A}^{1/(k+1)+}$ *et* (3. 6) *a lieu pour tout* α *vérifiant l'inégalité* $(k+1)^{-1}<2-\alpha\leqq 2$.

DÉMONSTRATION. — Pour $1/2<2-\alpha\leqq 2$ le théorème 3 n'est autre chose qu'une combinaison des théorèmes 1 et 2. Il reste à prouver que la borne inférieur de $2-\alpha$ dépend de k et s'abaisse avec k^{-1}. On établira d'abord la seconde inégalité (3. 6), la première sera ensuite une conséquence du théorème 1.

Les propriétés de $f(t)$ indiquées déjà au cours de la démonstration du théorème 2 subsistent encore. Il faut y ajouter pourtant que les dérivées $f^{(j)}(t)$ $(2\leqq j\leqq k)$ sont aussi périodiques de période 2π, et que $g(t)$ n'est pas constante. En effet, si $g(t)$ se réduisait à une constante, c'est que $f(t)$ était linéaire (cf. (3. 2)), on aurait $a_{nv}=0$ pour $n\neq-\gamma$ et $|a_{-\gamma v}|=1$, et la première inégalité (3. 6) n'aurait pas lieu (voir encore la remarque 1, p. 279).

Dès lors la démonstration suit à peu près la même voie que celle du théorème 2. Notons encore par n_0 l'entier précédemment introduit. En utilisant les propriétés

invoquées de e^{ivf}, f et $f^{(j)}$ $(1 \leqq j \leqq k)$ et en intégrant, de la manière signalée dans (3. 3), $(k-1)$ fois par parties, on obtient, pour $|n| \geqq n_0$,

$$(3.7) \qquad |a_{nv}| \leqq \frac{1}{2\pi} \left| \frac{v}{n^k} \int_{-\pi}^{\pi} e^{ivf(t)-int} d \frac{H_{k-1}(t, vn^{-1})}{[vn^{-1}f'(t)-1]^{2k-3}} \right| \leqq c_3 \left| \frac{v}{n^k} \right|$$

où $H_{k-1}(t, vn^{-1}) = O(1)$ est un polynome des $k-1$ premières dérivées de $f(t)$ et des puissances de vn^{-1} avec $|vn^{-1}| < l^{-1} < 1$. La série $\sum_{|n| \geqq n_0} |a_{nv}|^{2-\alpha}$ converge donc pour $k(2-\alpha) > 1$ ou $k^{-1} < 2-\alpha$. Cela étant, on a

$$(3.8) \qquad \sum_{|n| \geqq n_0} |a_{nv}|^{2-\alpha} = O(|v|^{1-(k-1)(2-\alpha)})$$

et $1-(k-1)(2-\alpha) \leqq \alpha/2$ si $\alpha \leqq 2$, ce qui est garantit par l'hypothèse $\alpha < 2-k^{-1}$.

Il est à voir ensuite que les dérivées de f n'interviennent pas dans la relation (3. 5), elle reste donc valable même lorsque $f \in C^k$ $(k > 1)$, ce qui revient à dire que

$$(3.9) \qquad \sum_{|n| < n_0} |a_{nv}|^{2-\alpha} = O(|v|^{\alpha/2}).$$

(3. 8) et (3. 9) entraînent la seconde inégalité (3. 6).

Quand $f^{(k)}(t)$ est en outre à variation bornée, on peut remplacer dans (3. 7) k par $k+1$. $H_k(t, vn^{-1})$ substituer ainsi à $H_{k-1}(t, vn^{-1})$, n'est plus alors nécessairement dérivable mais seulement une fonction à variation bornée. Il en résulte que $|a_{nv}| \leqq c_4 |vn^{-(k+1)}|$ pour $|n| \geqq n_0$ et que la somme envisagée converge pour $(k+1)^{-1} < 2-\alpha$.

La seconde inégalité (3. 6) et la condition $k^{-1} < 2-\alpha \leqq 2$ resp. $(k+1)^{-1} < 2-\alpha \leqq 2$ signifient que $e^{if} \in \mathscr{A}^{1/k+}$ resp. $\mathscr{A}^{1/(k+1)+}$. Par conséquent, en vertu du théorème 1, la première inégalité (3. 6) est également vérifiée.

REMARQUE 3. — Quand $\alpha = 1$, le théorème 3 fournit le résultat cité de KAHANE, toutefois avec la différence que la fonction analytique qui figure dans l'énoncé du théorème V se réduit dans le théorème 3 à une fonction deux fois continûment dérivable.

COROLLAIRE. — *Si $f \in C^\infty$, ou en particulier si f est analytique, non linéaire, et si $f(t) \equiv f(t+2\pi) \pmod{2\pi}$, alors $e^{if} \in \mathscr{A}^{0+}$ et, pour $0 < 2-\alpha \leqq 2$, il existe deux constantes $\lambda_1 = \lambda_1(f, \alpha)$ et $\lambda_2 = \lambda_2(f, \alpha)$ telles que*

$$\lambda_1 |v|^{\alpha/2} < \sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} < \lambda_2 |v|^{\alpha/2} \qquad (v = 0, \pm 1, \pm 2, \ldots).$$

Quand $f(t)$ est analytique et satisfait aux autres conditions mentionnées, le corollaire peut être démontré d'une manière directe en appliquant un procédé analogue à celui de KAHANE.

§ 4. Fonctions analytiques

Lorsque $\alpha<0$ la formule de Parseval ou celle de Hausdorff—Young donne seulement le résultat trivial que la somme en question est inférieure ou égale à 1. Les méthodes que nous venons d'utiliser dans l'étude des cas précédents ne s'appliquent non plus si $\alpha<0$. De cette raison pour $\alpha<0$ nous allons prouver un théorème moins général que ceux qui ont été obtenus pour $0<\alpha<2$. Nous nous bornons à l'examen des $f(t)$ réelles et analytiques sur l'axe réel, non linéaires, remplissant en outre la condition $f(t)\equiv f(t+2\pi) \pmod{2\pi}$. Pour les fonctions de ce genre nous avons établi deux lemmes ([1], pp. 191, 193) sur lesquels nous nous appuyerons par la suite. Introduisons quelques notations pour pouvoir formuler ces lemmes.

Soient $t_1<t_2<\ldots<t_h$ les racines de l'équation $f''(t)=0$ contenues dans l'intervalles $[-\pi, \pi]$ avec les multiplicités respectives $p_1-2, p_2-2, \ldots, p_h-2$ telles que $f^{(s)}(t_j)=0$ pour $2\leqq s<p_j$ et $f^{(p_j)}(t_j)\neq 0$ $(j=1, 2, \ldots, h)$. Il est manifeste que $\min_j p_j \geqq 3$.

LEMME. 1. — *Soit* $p=\max p_j$. *Alors*

$$a_{nv} = O(|v|^{-1/p}) \tag{4.1}$$

uniformément en n.

Quant à l'autre lemme, nous avons déjà montré dans [1] qu'il suffit de l'établir pour $v>0, n\geqq 0$ et $f'(t)>c>0$, les autres cas se ramènent facilement à celui-là ([1], p. 190). Dans cette hypothèse, les résultats obtenus dans [1] ((2. 7) p. 192 et p. 198) permettent la conclusion immédiate que

$$\sum_{n=-\infty}^{-1} |a_{nv}|^{2-\alpha} = O(e^{-c'(2-\alpha)v}) \qquad (\alpha<2). \tag{4.2}$$

LEMME 2. — *Soit donné le nombre* nv^{-1} $(n\geqq 0, v>0)$. *Lorsque*

$$cv^{-(p-1)/p} < \min_j |f'(t_j)-nv^{-1}| < c, \tag{4.3}$$

posons

$$\min_j |f'(t_j)-nv^{-1}| = cv^{(-1/p+\eta)(p-1)}, \tag{4.4}$$

avec $0<\eta=\eta(nv^{-1})<1/p$. *Alors*

$$|a_{nv}| \leqq Cv^{-1/p-(p-2)\eta/2} \tag{4.5}$$

où la constante $C=C(f)$ *ne dépend ni de* n *ni de* v. *Si par contre*

$$\min_j |f'(t_j)-nv^{-1}| \geqq c, \tag{4.6}$$

on a

$$|a_{nv}| = O(v^{-1/2}) \tag{4.7}$$

uniformément en n.

Nous fairons usage du lemme 1 quand

$$\min_j |f'(t_j)-nv^{-1}| \leqq cv^{-(p-1)/p}, \tag{4.8}$$

tandis que nous aurons besoin du lemme 2 lorsque (4. 3) ou (4. 6) a lieu. Nous pouvons énoncer maintenant la proposition suivante.

THÉORÈME 4. — *Si f est réelle, analytique, et non linéaire, si de plus $f(t) \equiv f(t+2\pi)$* (mod 2π), *alors*

$$\sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} = \begin{cases} O(|v|^{\alpha/2}) & \text{pour} \quad -2/(p-2) \leqq \alpha < 2 \\ O(|v|^{(\alpha-1)/p}) & \text{pour} \quad \alpha \leqq -2/(p-2). \end{cases} \tag{4. 9}$$

DÉMONSTRATION. — Soit encore $n_0 = lv$ $(v>0)$. On a, en vertu de (3. 8) et du corollaire du théorème 4,

$$\sum_{n=n_0}^{\infty} |a_{nv}|^{2-\alpha} = O(v^{1-k(2-\alpha)}) = O(v^{\alpha/2}) \tag{4. 10}$$

pour tout $\alpha<2$ et avec un k tel que $(k+1)^{-1}<2-\alpha$ d'ailleurs aussi grand que l'on veut. On a de plus

$$\sum_{n=0}^{n_0-1} |a_{nv}|^{2-\alpha} = \sum_{\substack{|a_{nv}|>K\varepsilon \\ 0\leqq n<n_0}} |a_{nv}|^{2-\alpha} + \sum_{\substack{|a_{nv}|\leqq K\varepsilon \\ 0\leqq n<n_0}} |a_{nv}|^{2-\alpha} = S_1 + S_2 \tag{4. 11}$$

où $\varepsilon = v^{-1/2}$ et K est une constante. Il est évident que

$$S_2 = O(v^{\alpha/2}). \tag{4. 12}$$

L'évaluation de S_1 exige un raisonnement plus détaillé. Les a_{nv} indiqués sous (4. 1), (4. 5) et en particulier ceux qui sont signalé sous (4. 7) n'interviennent que dans la somme S_1 qui ne contient que des termes de tels ordres de grandeur.

1° Soit t_i une valeur particulière de t_j qui, avec un v fixé, satisfait à la relation (4. 8). En général il y a alors plusieurs n vérifiant (4. 8) avec le même v et le même t_i. Soient n_{i1} le plus petit et n_{i2} le plus grand de ces n. Ainsi, v et t_i étant donnés, les n qu'on peut envisager sont déterminés, vu (4. 8), par l'inégalité

$$vf'(t_i) - cv^{1/p} \leqq n_{i1} \leqq n \leqq n_{i2} \leqq vf'(t_i) + cv^{1/p}, \quad n_{i2} - n_{i1} \leqq 2cv^{1/p}.$$

En notant par $\mathcal{N}_1$ l'ensemble des entiers contenus dans les intervalles $[n_{i1}, n_{i2}]$ $(i=1, 2, \ldots, h)$, on déduit à partir de (4.1)

$$S_1' = \sum_{n\in\mathcal{N}_1} |a_{nv}|^{2-\alpha} = O(v^{1/p} v^{-(2-\alpha)/p}) = O(v^{(\alpha-1)/p}). \tag{4. 13}$$

2° Désignons maintenant par t_i la valeur particulière de t_j qui, avec un v fixé, réalise (4. 3). Les n vérifiant (4. 3) avec un v et t_i donné se trouvent parmi ceux qui remplissent les inégalités

$$vf'(t_i) - cv \leqq n_{i3} \leqq n \leqq n_{i4} \leqq vf'(t_i) - cv^{1/p}, \quad n_{i4} - n_{i3} = O(v),$$

$$vf'(t_i) + cv^{1/p} \leqq n_{i5} \leqq n \leqq n_{i6} \leqq vf'(t_i) + cv, \quad n_{i6} - n_{i5} = O(v).$$

n_{i3} et n_{i4} resp. n_{i5} et n_{i6} sont les plus petits et les plus grands des n qu'on doit considérer.

v étant fixé, on a $\eta = \eta(n)$. Exprimons η à l'aide de (4. 4) et posons son expression trouvée dans (4. 5), nous obtenons

$$|a_{nv}| \leqq C_1 v^{-1/2} |f'(t_i) - nv^{-1}|^{-(p-2)/2(p-1)}.$$

Qand n croît de n_{i3} à n_{i4}, les $|a_{nv}|$ forment une suite monotone décroissante et, si

$$b = \frac{p-2}{2(p-1)}(2-\alpha) \neq 1 \quad \text{ou} \quad \alpha \neq -\frac{2}{p-2} = \alpha^*,$$

la somme des $|a_{nv}|^{2-\alpha}$ correspondants peut être majorée par une intégrale. Nous avons ainsi

$$S_1^*(\alpha) = \sum_{n=n_{i3}}^{n_{i4}} |a_{nv}|^{2-\alpha} = O(v^{\alpha/2}|1-v^{-(p-1)(1-b)/p}|) \qquad (b \neq 1). \tag{4.14}$$

Si $1-b>0$, c'est que $\alpha>\alpha^*$, on a $S_1^*(\alpha)=O(v^{\alpha/2})$; si au contraire $1-b<0$, donc $\alpha<\alpha^*$, on a $S_1^*(\alpha)=O(v^{(\alpha-1)/p})$. Quand $b=1$, $\alpha=\alpha^*$, on a $\alpha^*/2=(\alpha^*-1)/p$. Comme $S_1^*(\alpha)$ est une fonction continue de α pour $\alpha<2-\delta$ ($\delta>0$), il vient $S_1^*(\alpha^*)=$ $=O(v^{\alpha^*/2})=O(v^{(\alpha^*-1)/p})$. On trouve un résultat analogue lorsque n varie de n_{i5} à n_{i6}. Soit $\mathcal{N}_2$ l'ensemble des entiers appartenants aux intervalles $[n_{i3}, n_{i4}]$ et $[n_{i5}, n_{i6}]$ ($i=1, 2, \ldots, h$), on tire alors de (4. 14)

$$\begin{aligned} S_1'' &= \sum_{n\in\mathcal{N}_2} |a_{nv}|^{2-\alpha} = O(v^{\alpha/2}) \quad \text{pour} \quad -2/(p-2) \leqq \alpha < 2, \\ S_1''' &= \sum_{n\in\mathcal{N}_2} |a_{nv}|^{2-\alpha} = O(v^{(\alpha-1)/p}) \quad \text{pour} \quad \alpha = -2/(p-2). \end{aligned} \tag{4.15}$$

3° Finalement, le nombre des $|a_{nv}|>K\varepsilon$ et satisfaisant à (4. 7) est inférieur à n_0. Par conséquent, en désignant par $\mathcal{N}_3$ l'ensemble des indices n de ces a_{nv}, on peut écrire

$$S_1^{IV} = \sum_{n\in\mathcal{N}_3} |a_{nv}|^{2-\alpha} = O(v^{\alpha/2}). \tag{4.16}$$

Il découle ainsi de (4. 2), (4.10)—(4.13), (4.15), (4.16) que

$$\sum_{n=-\infty}^{\infty} |a_{nv}|^{2-\alpha} = O(v^{\alpha/2}) + O(v^{(\alpha-1)/p})$$

d'où la relation (4. 9).

BIBLIOGRAPHIE

[1] ALPÁR, L.: Sur une classe particulière de séries de Fourier ayant de sommes partielles bornées, *Studia Sci. Math. Hungar.* **1** (1966) 189—204.

[2] KAHANE, J. P.: Sur certaines classes de séries de Fourier absolument convergentes, *J. de Mathématiques Pures et Appliquées.* **35** (1956) 249—259.

[3] Лейбензон, З. Л.: О кольце функции с абсолютно сходящимися рядами Фурье, *Успехи Мат. Наук.* **9** (1954) № 3, 157—162.

[4] ZYGMUND, A.: *Trigonometric Series*, I—II. University Press, Cambridge, 1959.

Institut de Mathématique de l'Académie des Sciences de Hongrie, Budapest

(Reçu le 29 mai 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 287—292.

AN ITERATED LOGARITHM LAW FOR SEMIMARTINGALES AND ITS APPLICATION TO EMPIRICAL DISTRIBUTION FUNCTION

by

E. CSÁKI

Introduction

The discovery of the iterated logarithm law is due to KHINTCHINE [1], who proved it for the difference between the numbers of figures 1 and 0 in the dyadic expansion of real numbers, i.e. for partial sums of RADEMACHER functions. Since then in a series of papers it has been generalized to other cases, e.g. to the sum of independent random variables. In this paper the iterated logarithm law will be extended to semi-martingales satisfying an additional condition. In the second part of the paper this will be applied for some functionals of empirical distribution functions. The paper by SMIRNOV [2] is to be mentioned here in which the iterated logarithm law is proved for the KOLMOGOROV distance by direct method. The same theorem appears in the paper by K. L. CHUNG [3].

1. The Iterated Logarithm Law for Semi-martingales

Let $\xi_1, \xi_2, \ldots, \xi_n \ldots$ be a sequence of random variables, satisfying the semi-martingale inequality, i.e.

$$\mathsf{E}(\xi_n|\xi_1, \ldots \xi_{n-1}) \geqq \xi_{n-1} \qquad n=2, 3, \ldots . \tag{1}$$

Let the random variables ξ_n satisfy the additional condition

$$\mathsf{E}(e^{t\xi_n}) \leqq (\Psi(t))^n \tag{2}$$

for $t>0$, where $\Psi(t)$ is a function such that

$$\Psi(t) = 1+\frac{A^2}{2}t^2+O(t^3) \qquad t \to +0 \tag{3}$$

with some constant $A>0$.

THEOREM 1.1. *If the sequence* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *satisfies the conditions* (1) *and* (2) *with* (3), *then*

$$\mathsf{P}\left(\limsup_{n\to\infty} \frac{\xi_n}{\sqrt{n \log\log n}} \leqq A\sqrt{2}\right) = 1. \tag{4}$$

PROOF. By virtue of the Borel—Cantelli lemma it suffices to prove that the series

$$\sum_{(k)} \mathsf{P}\left(\max_{n_k < i \leqq n_{k+1}} \xi_i \geqq (A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k}\right) = \sum_{(k)} P_k \tag{5}$$

is convergent for some appropriate sequence n_k and $\varepsilon>0$. As the sequence

$e^{t\xi_1}, e^{t\xi_2}, \ldots, e^{t\xi_n}, \ldots$ is also a semi-martingale (see [4], pp. 295. Theorem 1.1) applying KOLMOGOROV's inequality ([4] pp. 314, Theorem 3. 2) and the condition (2), we get

$$P_k = \mathsf{P}\left(\max_{n_k < i \leqq n_{k+1}} \xi_i \geqq (A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k}\right) =$$

$$= \mathsf{P}\left(\max_{n_k < i \leqq n_{k+1}} e^{t\xi_i} \geqq e^{t(A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k}}\right) \leqq$$

$$\leqq \frac{\mathsf{E}(e^{t\xi_{n_{k+1}}})}{e^{t(A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k}}} \leqq \frac{(\psi(t))^{n_{k+1}}}{e^{t(A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k}}}.$$

Now put $n_k = \left[\left(1+\frac{\varepsilon}{A\sqrt{2}}\right)^k\right]$ and $t = \frac{\sqrt{2}}{A}\sqrt{1+\frac{\varepsilon}{A\sqrt{2}}}\sqrt{\frac{\log\log n_{k+1}}{n_{k+1}}}$, then

$$n_{k+1}\log\Psi(t) = n_{k+1}\log\left(1+\frac{A^2}{2}t^2+O(t^3)\right) = n_{k+1}\frac{A^2t^2}{2}+o(1) =$$

$$= \left(1+\frac{\varepsilon}{A\sqrt{2}}\right)\log\log n_{k+1}+o(1) \qquad (k\to\infty)$$

and

$$t(A\sqrt{2}+\varepsilon)\sqrt{n_k \log\log n_k} = (A\sqrt{2}+\varepsilon)\frac{\sqrt{2}}{A}\sqrt{1+\frac{\varepsilon}{A\sqrt{2}}}\sqrt{\frac{n_k}{n_{k+1}}}\cdot$$

$$\cdot\sqrt{\log\log n_k \log\log n_{k+1}} = 2\left(1+\frac{\varepsilon}{A\sqrt{2}}\right)\log\log n_{k+1}+o(1) \qquad (k\to\infty),$$

hence

$$P_k \leqq c\frac{1}{(\log n_{k+1})^{1+\frac{\varepsilon}{A\sqrt{2}}}} \leqq \frac{c}{(k+1)^{1+\frac{\varepsilon}{A\sqrt{2}}}},$$

so the series (5) converges, which proves Theorem 1.1. Some corollaries to this theorem is to be mentioned here.

COROLLARY 1.1. The condition (2) can be weakened by

$$\mathsf{E}(e^{t\xi_n}) \leqq C_n(t)(\Psi(t))^n, \tag{2/a}$$

where $C_n(t)$ is a factor not disturbing the convergence of the series (5). E.g. $C_n(t) = (t\sqrt{n})^K$, where K is a positive constant.

COROLLARY 1. 2. If the sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ is monotone non-decreasing (being trivially a semi-martingale) and (2) or (2/a) holds, then (4) is true.

COROLLARY 1. 3. If the sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ is arbitrary (not necessarily a semi-martingale) and (2) or (2/a) holds for $\xi_n' = \max(\xi_1, \ldots, \xi_n)$, then (4) is true.

COROLLARY 1. 4. If (1) and (2) or (2/a) holds for the sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ then the iterated logarithm law is true not only for $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ but also for the sequence $\xi_n' = \max(\xi_1, \ldots, \xi_n)$ $(n=1, 2, \ldots)$ with the same constant, i.e.

$$\mathsf{P}\left(\limsup_{n\to\infty} \frac{\xi_n'}{\sqrt{n \log\log n}} \leqq A\sqrt{2}\right) = 1.$$

The corollaries 1. 1—1. 2—1. 3 are trivial, but Corollary 1. 4 needs some explanation. The sequence $e^{\frac{t}{2}\xi_1}, \ldots, e^{\frac{t}{2}\xi_n}$ is a semi-martingale for $t>0$ and as $e^{\frac{t}{2}\xi_n'} = \max_{1\leqq i\leqq n} e^{\frac{t}{2}\xi_i}$, using Theorem 3. 4 in [4] pp. 317 with $\alpha=2$, we obtain

$$\mathsf{E}(e^{t\xi_n'}) \leqq 4\mathsf{E}(e^{t\xi_n}) \leqq 4C_n(t)(\Psi(t))^n.$$

2. Application to the Empirical Distribution Function

Let $F_n(x)$ be the empirical distribution function of the sample $(\eta_1, \eta_2, \ldots, \eta_n)$. Assume that the theoretical distribution function $F(x)$ is continuous. We shall first consider the KOLMOGOROV distance

$$D_n = \sup_{(x)} |F_n(x) - F(x)|.$$

THEOREM 2. 1. (SMIRNOV)

$$\mathsf{P}\left(\limsup_{n\to\infty} D_n \sqrt{\frac{n}{\log\log n}} = \frac{1}{\sqrt{2}}\right) = 1. \tag{6}$$

PROOF. Put $\xi_n = n \max_{(x)} (F_n(x) - F(x))$. We shall show that the conditions (1) and (2/a) holds for ξ_n. Let ϱ_n denote the x-point, where the maximum takes place. It can be seen that

$$\mathsf{E}(\xi_{n+1} - \xi_n | \xi_1, \ldots, \xi_n, \varrho_n) \geqq F(\varrho_n)(1 - F(\varrho_n)) - (1 - F(\varrho_n))F(\varrho_n) = 0,$$

and hence

$$\mathsf{E}(\xi_{n+1} | \xi_1, \ldots, \xi_n) \geqq \xi_n$$

so the sequence $\xi_1, \ldots, \xi_n$ is a semi-martingale.

To prove that the condition (2/a) holds, we shall use the distribution of ξ_n, first derived by SMIRNOV [2]:

$$H(z) = \mathsf{P}(\xi_n < z) = 1 - \frac{z}{n^n} \sum_{k=0}^{[n-z]}{}' \binom{n}{k} (n-z-k)^{n-k}(z+k)^{k-1} \qquad 0 < z < n \tag{7}$$

It is easy to see that

$$\mathsf{E}(e^{t\xi_n}) = \int_0^n e^{tz}\, dH(z) = 1 + t\mathsf{E}(\xi_n) + t\int_0^n (e^{tz} - 1)(1 - H(z))\, dz. \tag{8}$$

Now we shall estimate the third term of the above expression.

$$\int_0^n (e^{tz}-1)(1-H(z))\,dz = \sum_{j=1}^{n} \int_{j-1}^{j} (e^{tz}-1)(1-H(z))\,dz =$$

$$= \sum_{k=0}^{n} \binom{n}{k} \int_{\frac{k}{n}}^{1} (e^{t(ny-k)}-1)(ny-k)(1-y)^{n-k}y^{k-1}\,dy \leqq$$

$$\leqq \sum_{k=0}^{n} \binom{n}{k} \int_{0}^{1} (e^{t(ny-k)}-1)(ny-k)(1-y)^{n-k}y^{k-1}\,dy =$$

$$= n\frac{e^t-1}{e^t} \int_0^1 e^{tny}(1-y+ye^{-t})^{n-1}(1-y)\,dy \leqq$$

$$\leqq n\frac{e^t-1}{e^t} \int_0^1 e^{tny}(1-y+ye^{-t})^{n}\,dy \leqq n\frac{e^t-1}{e^t}(\Psi(t))^n,$$

where

$$\Psi(t) = \frac{e^t-1}{t}\exp\left\{\frac{t}{e^t-1}-1\right\}.$$

It can be seen that

$$\Psi(t) = 1+\frac{t^2}{8}+O(t^3) \qquad t \to +0$$

and the factor $nt\dfrac{e^t-1}{e^t}$ does not disturb the convergence of the series (5), because

$$nt\frac{e^t-1}{e^t} \leqq nt^2 \quad \text{if} \quad t \geqq 0.$$

The first two terms in (8) are of smaller order than the third one, owing to the fact that

$$1+tE(\xi_n) \leqq ct\sqrt{n}.$$

Repeating the same argument for $\bar{\xi}_n = n \sup_{(x)} (F(x)-F_n(x))$ one can see by Theorem 1.1, that

$$\mathsf{P}\left(\limsup_{n\to\infty} D_n \sqrt{\frac{n}{\log\log n}} \leqq \frac{1}{\sqrt{2}}\right) = 1.$$

Applying the classical iterated logarithm law for the distance at the median i.e. for $(F_n(m)-F(m))$, where $F(m)=\frac{1}{2}$, it can be seen that $1/\sqrt{2}$ is the best possible constant, so Theorem 2.1 is true.

Now we shall consider the two-sample case.

Let $F_n(x)$ and $G_n(x)$ be two empirical distribution functions and

$$D_{n,n} = \sup_{(x)} |F_n(x)-G_n(x)|.$$

THEOREM 2. 2. If the respective distribution functions are equal and continuous, then

$$\mathsf{P}\left(\limsup_{n\to\infty} D_{n,n}\sqrt{\frac{n}{\log\log n}}=1\right)=1. \tag{9}$$

PROOF. Now put $\xi_n=n\max_{(x)}\left(F_n(x)-G_n(x)\right)$, then it can be seen similarly to the idea of Theorem 2.1 that the sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ is a semi-martingale. The distribution of ξ_n was given by GNEDENKO and KOROLYUK [5]:

$$\mathsf{P}(\xi_n \geqq k)=\frac{\binom{2n}{n+k}}{\binom{2n}{n}} \qquad k=0, 1, 2, \ldots, n. \tag{10}$$

$$\mathsf{E}(e^{t\xi_n})=\sum_{k=0}^{n} e^{tk}\mathsf{P}(\xi_n=k)=\frac{1-e^{-t}}{\binom{2n}{n}}\sum_{k=0}^{n} e^{tk}\binom{2n}{n+k}+e^{-t}\leqq$$

$$\leqq \frac{1-e^{-t}}{\binom{2n}{n}e^{tn}}(1+e^t)^{2n}+e^{-t}\leqq ct\sqrt{n}\left(\frac{1}{4}e^{-t}+\frac{1}{2}+\frac{1}{4}e^t\right)^n+1.$$

Here

$$\Psi(t)=\frac{1}{2}\left(1+\frac{e^t+e^{-t}}{2}\right)=1+\frac{t^2}{4}+O(t^3), \qquad t\to 0.$$

The same argument as in the proof of Theorem 1.1 shows that (9) holds, thus Theorem 2. 2 is proved.

Now let us consider the more general statistics:

$$D_n(\tau)=\sup_{(x)}\{\tau(F(x))|F_n(x)-F(x)|\}$$

where $\tau(y)$ is a positive weight function. If $\tau(y)$ is bounded from above and $\sup_{0\leqq y\leqq 1}\tau(y)=a$, then it can be stated that

$$\mathsf{P}\left(\limsup_{n\to\infty} D_n(\tau)\sqrt{\frac{n}{\log\log n}}\leqq\frac{a}{\sqrt{2}}\right)=1 \tag{11}$$

and similar statement holds for the two sample case.

For the CRAMÉR-VON MISES statistics

$$\omega_n^2=\int_{-\infty}^{\infty}\left[(F_n(x)-F(x))\tau(F(x))\right]^2 dF(x)$$

the trivial estimation $\omega_n\leqq aD_n$ is true, thus

$$\mathsf{P}\left(\limsup_{n\to\infty}\omega_n\sqrt{\frac{n}{\log\log n}}\leqq\frac{a}{\sqrt{2}}\right)=1. \tag{12}$$

In the latter two cases, however, the constant $a/\sqrt{2}$ seems to be not the best possible one, but we cannot say anything about it.

REFERENCES

[1] KHINTCHINE, A. JA.: Über dyadische Brüche, *Math. Z.* **18** (1923) 109—118.
[2] СМИРНОВ, Н. В.: Приближение законов распределения случайных величин по эмпирическим данным, *Успехи Мат. Наук*, **10** (1944) 179—206.
[3] CHUNG, K. L.: An estimate concerning the Kolmogorov limit distribution, *Trans. Amer. Math. Soc.* **67** (1949) 36—50.
[4] DOOB, J. L.: *Stochastic processes*, New York, John Wiley & Sons, 1953.
[5] ГНЕДЕНКО, Б. В. и КОРОЛЮК, В. С.: О максимальном расхождении двух эмпирических распределений, *Докл. Акад. Наук СССР* **80** (1951) 525—528.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received May 30, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 293—297.

SUR LA CARACTERISATION DES FONCTIONS DERIVABLES PAR LEUR APPROXIMATION TRIGONOMETRIQUE

A la mémoire de Jean Favard

par

G. ALEXITS

1. Désignons par $f(x)$ une fonction intégrable, 2π-périodique de période 2π, par $\tilde{f}(x)$ sa fonction conjuguée; soit en plus $\sigma_n^\alpha(x)$ la n-ième moyenne de Cesàro d'ordre α de la série de Fourier

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1}$$

et $\tilde{\sigma}_n^\alpha(x)$ la moyenne conjuguée correspondante. (Pour $\alpha = 1$ nous allons écrire tout simplement $\sigma_n(x)$ resp. $\tilde{\sigma}_n(x)$).

Nous nous sommes occupés, à plusieurs reprises, de la caractérisation des fonctions ayant presque partout une dérivée bornée en certain sens. Le premier résultat en était que $f \in \text{Lip } 1$ équivaut à

$$\|\tilde{f}(x) - \tilde{\sigma}_n(x)\|_C = O\left(\frac{1}{n}\right) \tag{2}$$

où $\|\cdot\|_C$ désigne la norme de l'espace C des fonctions continues ([1] et [4]). Plus tard, nous avons étendu la validité de ce théorème au cas où la dérivée $f'(x)$ existante presque partout n'est plus bornée, comme dans le cas $f \in \text{Lip } 1$, mais seulement intégrable, c'est-à-dire que la fonction $f(x)$ est à variation bornée. Même dans ce cas, la condition nécessaire et suffisante est donnée par la relation (2), si l'on y remplace la norme $\|\cdot\|_C$ par la norme $\|\cdot\|_L$ de l'espace des fonctions intégrables.

Lorsqu'on cherche à améliorer la méthode de l'approximation telle qu'elle caractérise des classes de plus en plus fines, le problème le plus proche de ceux que nous venons d'indiquer est de trouver la caractérisation des fonctions absolument continues et des fonctions à dérivée continue. Le résultat respectif est très simple et suit les mêmes lignes que nous venons de signaler.

Soit $\{\lambda_n\}$ une suite de nombres positifs, concave par en bas et tendant vers l'infini. Multiplions la série de Fourier (1) terme à terme par $\{\lambda_n\}$ et désignons par $\sigma_n^\alpha(\lambda, x)$ resp. $\tilde{\sigma}_n^\alpha(\lambda, x)$ la n-ième moyenne (C, α) de la série

$$\frac{a_0 \lambda_0}{2} + \sum_{n=1}^{\infty} \lambda_n (a_n \cos nx + b_n \sin nx)$$

resp. de la série conjuguée

$$\sum_{n=1}^{\infty} \lambda_n (a_n \sin nx - b_n \cos nx).$$

Les théorèmes correspondant aux problèmes mentionnés ci-dessus sont les suivants:

THÉORÈME 1. *Pour que la fonction $f(x)$ soit absolument continue, il faut et il suffit l'existence d'une suite $\{\lambda_n\}$ concave par en bas et tendant vers l'infini telle que l'on ait*

$$\|\tilde{f}_\lambda(x) - \tilde{\sigma}_n^\alpha(\lambda, x)\|_L = O\left(\frac{1}{n}\right) \qquad (0 < \alpha)$$

où $\tilde{f}_\lambda(x)$ désigne la limite de $\{\tilde{\sigma}_n^\alpha(\lambda, x)\}$ dans l'espace L.

THÉORÈME 2. *Pour que $f(x)$ ait une dérivée continue partout, il faut et il suffit l'existence d'une suite $\{\lambda_n\}$ concave par en bas et tendant vers l'infini telle que*

$$\|\tilde{f}_\lambda(x) - \tilde{\sigma}_n^\alpha(\lambda, x)\|_C = O\left(\frac{1}{n}\right) \qquad (0 < \alpha).$$

2. La démonstration est basée sur trois lemmes élémentaires concernant les moyennes de Cesàro des séries arbitraires. Soit Σx_ν une série formée d'éléments d'un espace de Banach dont la norme sera désignée par $\|\cdot\|$. Soit en plus σ_n^α la n-ième moyenne (C, α) de Σx_ν et $\bar{\sigma}_n^\alpha$ celle de $\Sigma \frac{x_\nu}{\nu}$. Les trois lemmes en question sont les suivants:

LEMME 1. *Si l'on a $\|\sigma_n^\alpha\| \leqq K$ pour $n = 1, 2, \ldots$, alors $\{\bar{\sigma}_n^\alpha\}$ tend vers un élément $\bar{\sigma}^\alpha$ et*

$$\|\bar{\sigma}^\alpha - \bar{\sigma}_n^\alpha\| \leqq \frac{K(\alpha)}{n}$$

où $K(\alpha)$ est une constante ne dépendant que de α et K.

LEMME 2. *Si l'on a*

$$\|\bar{\sigma}^\alpha - \bar{\sigma}_n^\alpha\| \leqq \frac{K}{n},$$

alors

$$\|\sigma_n^\alpha\| \leqq K(\alpha)$$

où la constante $K(\alpha)$ ne dépend que de α et K.

LEMME 3. *Désignons par σ_n la n-ième moyenne $(C, 1)$ de Σx_ν et par $\sigma_n(\lambda)$ celle de $\Sigma \lambda_\nu x_\nu$. Si*

$$\|\sigma - \sigma_n\| \to 0,$$

il existe une suite $\{\lambda_n\}$ concave par en bas et tendant vers l'infini telle qu'il existe un $\sigma(\lambda)$ pour lequel

$$\|\sigma(\lambda) - \sigma_n(\lambda)\| \to 0.$$

3. Pour la démonstration du lemme 1, posons $A_\nu^\alpha = \binom{\nu+\alpha}{\nu}$, et $x_0 = 0$, alors

$$\bar\sigma_m^\alpha - \bar\sigma_n^\alpha = \sum_{\nu=n+1}^{m} (\bar\sigma_\nu^\alpha - \bar\sigma_{\nu-1}^\alpha) = \sum_{\nu=n+1}^{m} \frac{1}{A_\nu^\alpha \nu} \sum_{k=0}^{\nu} A_{\nu-k}^{\alpha-1} x_k =$$

$$= \sum_{\nu=n+1}^{m-1} \left(\frac{1}{A_\nu^\alpha \nu} - \frac{1}{A_{\nu+1}^\alpha (\nu+1)} \right) \sum_{k=0}^{\nu} A_k^{\alpha-1} \sum_{l=0}^{k} \frac{A_{k-l}^{\alpha-1}}{A_k^{\alpha-1}} x_l +$$

$$+ \frac{1}{A_m^\alpha m} \sum_{k=0}^{m} A_k^{\alpha-1} \sum_{l=0}^{k} \frac{A_{k-l}^{\alpha-1}}{A_k^{\alpha-1}} x_l - \frac{1}{A_{n+1}^\alpha (n+1)} \sum_{k=0}^{n} A_k^{\alpha-1} \sum_{l=0}^{k} \frac{A_{k-l}^{\alpha-1}}{A_k^{\alpha-1}} x_l =$$

$$= \sum_{\nu=n+1}^{m-1} \frac{\alpha+1}{(\alpha+\nu+1)\nu} \sum_{k=0}^{\nu} \frac{A_k^{\alpha-1}}{A_\nu^\alpha} \sigma_k^{\alpha-1} + \frac{1}{m} \sum_{k=0}^{m} \frac{A_k^{\alpha-1}}{A_m^\alpha} \sigma_k^{\alpha-1} -$$

$$- \frac{1}{n+1} \sum_{k=0}^{n} \frac{A_k^{\alpha-1}}{A_{n+1}^\alpha} \sigma_k^{\alpha-1} =$$

$$= \sum_{\nu=n+1}^{m-1} \frac{\alpha+1}{(\alpha+\nu+1)\nu} \sigma_\nu^\alpha + \frac{\sigma_m^\alpha}{m} - \frac{\sigma_n^\alpha}{\alpha+n+1}.$$

En tenant compte de la condition $\|\sigma_\nu^\alpha\| \leqq K$, on obtient

$$\|\bar\sigma_m^\alpha - \bar\sigma_n^\alpha\| \leqq K \left(\frac{\alpha+1}{n+1} + \frac{1}{m} + \frac{1}{\alpha+n+1} \right).$$

Il s'ensuit d'abord la convergence de $\{\bar\sigma_n^\alpha\}$ vers un élément $\bar\sigma_n^\alpha$ et puis

$$\|\bar\sigma^\alpha - \bar\sigma_n^\alpha\| \leqq \frac{(\alpha+2)K}{n},$$

ce qui était la proposition du lemme 1.

Quant au lemme 2, remarquons d'abord que l'on conclut par un calcul immédiat que

$$\sigma_n^\alpha = (\alpha+n+1)(\bar\sigma_n^\alpha - \bar\sigma_n^{\alpha+1}),$$

par conséquent

$$\|\sigma_n^\alpha\| \leqq (\alpha+n+1)\{\|\bar\sigma_n^\alpha - \bar\sigma^\alpha\| + \|\bar\sigma^\alpha - \bar\sigma_n^{\alpha+1}\|\}.$$

Or, il suit de l'hypothèse du lemme 2 que l'on a de même

$$\|\bar\sigma^\alpha - \bar\sigma_n^{\alpha+1}\| \leqq \frac{K_1}{n}$$

où K_1 ne dépend que de K et α, donc

$$\|\sigma_n^\alpha\| \leqq \frac{(K+K_1)(\alpha+n+1)}{n}$$

et le lemme 2 est également prouvé.

En ce qui concerne la démonstration du lemme 3, elle est aussi élémentaire, mais nous ne la reproduisons pas, puisque nous l'avons déja fait dans notre note [2]. (Remarquons qu'au fond, nous avons démontré les lemmes 1 et 2 déja dans [1], mais ils n'étaient formulés que pour un cas spécial et la démonstration était inutilement compliquée; c'est pour cela que nous avons préféré de la détailler ici.)

4. Passons à la démonstration du théorème 1 et soit, dans ce but, $f(x)$ une fonction absolument continue. Alors la dérivée $\sigma_n^{\alpha\prime}(x)$ de $\sigma_n^{\alpha}(x)$ est la n-ième moyenne (C, α) de la série de Fourier de $f'(x)$ et, par conséquent,

$$\|f'(x)-\sigma_n^{\alpha}(x)\|_L \to 0 \qquad (0<\alpha).$$

En désignant donc par $\sigma_n^{\alpha\prime}(\lambda, x)$ la n-ième moyenne (C, α) de la série

$$\sum_{k=1}^{\infty} \lambda_k k(a_k \sin kx - b_k \cos kx), \tag{3}$$

il existe, en vertu du lemme 3, une suite $\{\lambda_n\}$ concave d'en bas, tendant vers l'infini et telle que, pour $\alpha=1$, on ait

$$\|g_\lambda(x)-\sigma_n^{\alpha\prime}(\lambda, x)\|_L \to 0. \tag{4}$$

Il s'ensuit que (3) est la série de Fourier de $g_\lambda(x)$ et, par suite, (4) subsiste pour tout $0<\alpha$ (cf. [5], p. 148). Nous pouvons donc appliquer notre lemme 1 à la série (3) et nous obtenons

$$\|\tilde{f}_\lambda(x)-\tilde{\sigma}_n^{\alpha}(\lambda, x)\|_L = O\left(\frac{1}{n}\right), \tag{5}$$

ce qui prouve la nécessité de notre condition.

Supposons inversement que la condition (5) soit satisfaite. Il résulte alors, d'après le lemme 2,

$$\|\sigma_n^{\alpha\prime}(\lambda, x)\|_L = O(1).$$

La série (3) est donc une série de Fourier—Stieltjes ([5], p. 137). Mais la série

$$\sum_{k=1}^{\infty} k(a_k \sin kx - b_k \cos kx) \tag{6}$$

s'obtient de (3) par multiplication terme à terme avec $\{1/\lambda_n\}$. Cette dernière suite étant convexe et tendant vers zéro, elle transforme toute série de Fourier—Stieltjes en une série de Fourier ([5], p. 179), (6) est donc la série de Fourier d'une fonction intégrable $g(x)$. On a donc

$$\int_0^x g(t)\,dt \sim C+\sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx)$$

où le second membre est la série de Fourier de $f(x)+C-a_0/2$; $f(x)$ est donc en effet une fonction absolument continue et le théorème 1 est entièrement démontré.

5. Il est inutile de détailler la démonstration du théorème 2, parce qu'elle n'est qu'une variante de celle du théorème 1. L'unique différence consiste en ce que nous avons à prendre la norme $\|\cdot\|_C$ au lieu de $\|\cdot\|_L$ et nous devons tenir compte du fait que $\|f'(x)-\sigma_n^{\alpha'}(x)\|_C \to 0$ est la condition nécessaire est suffisante pour que $f'(x)$ soit continue.

BIBLIOGRAPHIE

[1] ALEXITS, G.: Sur l'ordre de grandeur de l'approximation d'une fonction par les moyennes de sa série de Fourier, *Mat. Fiz. Lapok* **48** (1941) 410—421 (hongrois) et 421—422 (français).
[2] ALEXITS, G.: Über die Transformierten der arithmetischen Mittel von Orthogonalreihen, *Acta Math. Acad. Sci. Hung.* **2** (1951) 1—17.
[3] ALEXITS, G.: Sur l'ordre de grandeur de l'approximation d'une fonction périodique par les sommes de Fejér, *Acta Math. Acad. Sci. Hung.* **3** (1952) 29—40.
[4] ZAMANSKY, M.: Classes de saturation de certains procédés d'approximation des séries de Fourier des fonctions continues et applications à quelques problèmes d'approximation, *Ann. Sci. Ecole Norm. Sup.* (3) **66** (1949) 19—93.
[5] ZYGMUND, A.: *Trigonometric Series*, Cambridge, 1959, vol. I.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received June 2, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 299—324.

ÜBER EINE VERMUTUNG VON ERDŐS BETREFFS POLYNOME, II.

von

Á. ELBERT

Die Vermutung von ERDŐS, um die es sich in [1] handelt, lautet folgendermassen:

Durch F_n sei die Gesamtheit der Polynome $f(x)$ von der Gestalt $f(x)=\prod_{i=1}^{n}(x-x_i)$ mit $-1\leqq x\leqq 1$ bezeichnet und $E(f)$ sei die Menge der $x\in(-\infty,\infty)$ mit $|f(x)|\leqq 1$. $E(f)$ besteht aus abgeschlossenen Intervallen, deren Gesamtlänge durch $|E(f)|$ bezeichnet sei. Führt man die Größe

$$E_n=\sup_{f(x)\in F_n}|E(f)|$$

ein, so lautet die Vermutung von ERDŐS:

$$E_n\leqq 2\sqrt{2}\qquad (n=1,2,\ldots). \tag{0.1}$$

In der vorliegenden Arbeit werde ich die Richtigkeit dieser Vermutung völlig beweisen, den Beweis habe ich schon in einer früheren Arbeit [2] begonnen. Die Arbeit gliedert sich in drei Teile; in dem ersten leite ich eine Integralgleichung ab und mit Hilfe dieser Integralgleichung beweise ich im zweiten Teil die Richtigkeit der Ungleichung (0, 1). Im dritten Teil werden einige Probleme, die mit dieser Vermutung verwandt sind, betrachtet. Zwei Folgerungen werden noch bewiesen:

FOLGERUNG 1. Es sei $f(x)=(x+1)^m\prod_{i=1}^{n-m}(x-x_i)$ mit $-1<x_i\leqq 1$, $i=1,2,\ldots$ $\ldots,n-m$, ferner sei $m>\alpha_0 n$, wo α_0 eine positive Wurzel der Gleichung

$$(1+\alpha)^{1+\alpha}(1-\alpha)^{1-\alpha}=2\qquad (0<\alpha<1)$$

ist, dann gibt es eine reelle Zahl $\xi\in(-1,1]$ mit $|f(\xi)|>1$. ($\alpha_0=0{,}7799\ldots$).

FOLGERUNG 2. Es sei $f(x)=(x+1)^{n_1}(x-1)^{n_2}\prod_{i=1}^{n_3}(x-x_i)$ mit $-1<x_i<1$, $i=1,2,\ldots,n_3$, $n=n_1+n_2+n_3$ ferner

$$\left(1+\frac{n_1}{n}+\frac{n_2}{n}\right)^{1+\frac{n_1}{n}+\frac{n_2}{n}}\cdot\left(1+\frac{n_1}{n}-\frac{n_2}{n}\right)^{1+\frac{n_1}{n}-\frac{n_2}{n}}\cdot$$

$$\cdot\left(1-\frac{n_1}{n}+\frac{n_2}{n}\right)^{1-\frac{n_1}{n}+\frac{n_2}{n}}\cdot\left(1-\frac{n_1}{n}-\frac{n_2}{n}\right)^{1-\frac{n_1}{n}-\frac{n_2}{n}}>4, \tag{0.2}$$

dann gibt es eine reelle Zahl $\xi\in(-1,1)$ mit $|f(\xi)|>1$.

1. Das Herleiten der Integralgleichung

In der Arbeit [2] handelte es sich um extremale Polynome $f^*(x)$ n-ten Grades, d.h. um Polynome, die der Gleichung $|E(f^*)| = E_n$ genügen. Es wurde dort gezeigt, daß es genügt nur Polynome von geradem Grade in Betracht zu ziehen. Für ein extremales Polynom $f^*(x)$ geraden Grades gilt entweder

$$f^*(x) = (x+1)^{\frac{n}{2}}(x-1)^{\frac{n}{2}} \tag{1.1}$$

oder

$$f^*(x) = (x+1)^{n_1}(x-1)^{n_2}\prod_{i=1}^{n_3}(x-x_i) \tag{1.2}$$

$$\left(n = n_1+n_2+n_3,\ n_1 > \frac{n}{2},\ n_2 > 0,\ n_3 > 0\right)$$

wobei

$$0 \leqq x_2 < x_3 < \ldots < x_{n_3-1} < d\left(\frac{n_1}{n_2}\right), \qquad -1 < x_1 < x_2,\ x_{n_3-1} < x_{n_3} < 1, \tag{1.3}$$

$$|f^*(x)| \leqq 1 \quad \text{für} \quad -1 \leqq x \leqq 1, \tag{1.4}$$

und es gibt $n_3 - 1$ Werte ξ_i $(i = 1, 2, \ldots, n_3 - 1)$ mit

$$|f^*(\xi_i)| = 1 \quad \text{und} \quad x_i < \xi_i < x_{i+1}. \tag{1.5}$$

Die Funktion $d(x)$ ist hier die Inverse von

$$x = x(d) = \frac{-\log(1-d)}{\log(1+d)}, \qquad 0 < d < 1. \tag{1.6}$$

$x(d)$ wächst von 1 bis ∞, wenn d von 0 bis 1 wächst.

Für das Weitere werden wir folgendes Lemma nötig haben:

LEMMA. Es sei $f(x)$ eine beliebige stetige Funktion, die für die x-Werte im Intervall $(-\infty, \infty)$ definiert ist und ξ, η sollen reelle Zahlen sein: $\xi \leqq \eta$. Für die Funktion $g(x) = (x-\xi)(x-\eta)f(x)$ soll im Intervall $[\xi, \eta]$ die Ungleichung $|g(x)| < 1$ gelten. Dann existiert eine reelle Zahl $\varepsilon_0 > 0$, so dass für alle Funktionen $g^*(x) = (x-\xi^*)(x-\eta^*)f(x)$ mit $\xi^* = \xi - \varepsilon$, $\eta^* = \eta + \varepsilon$ und $0 < \varepsilon < \varepsilon_0$ die Ungleichung

$$|E(g)| < |E(g^*)|$$

gültig ist.

Den Beweis dieses Lemmas kann man in [2] finden (Lemma 1). In dieser Arbeit setze ich mir zum Ziele zu beweisen, daß für ein extremales Polynom nur (1.1) gilt, offenbar genügt es hierzu zu beweisen, daß kein extremales Polynom mit den Eigenschaften (1.2)—(1.5) existiert. Für das Weitere werden nur Polynome mit den Eigenschaften (1.2)—(1.5) betrachtet.

Es sei also $f^*(x)$ ein solches Polynom. Betrachten wir das Polynom $[f^*(x)]^2$. Es besitzt zweifache Wurzeln im Interwall $(-1, 1)$, deshalb kann man die Größe $|E([f^*(x)]^2)|$ nach dem vorigen Lemma vergrößern. Es sei $F_{2n}(f^*)$ die Gesamtheit sämtlicher Polynome $2n$-ten Grades, welche wir aus $[f^*(x)]^2$ durch endlichmalige

Anwendung des Lemmas erhalten können. Ähnlich wie bei dem Beweise der Existenz des extremalen Polynoms in [2] kann man die Existenz eines Polynoms $f^{(1)}(x) \in F_{2n}$, welches der Gleichung

$$|E(f^{(1)})| = \sup_{f \in F_{2n}(f^*)} |E(f)|$$

genügt, einsehen, und für das Polynom $f^{(1)}(x)$ gilt:

$$|E(f^*)| < |E(f^{(1)})|, \quad \text{grad } f^{(1)}(x) = 2n. \tag{1.6}$$

Das Lemma wird nur auf die Wurzeln, welche im Intervall $(-1, 1)$ liegen, angewendet und wir erhalten:

$$f^{(1)}(x) = (x+1)^{m_1^{(1)}}(x-1)^{m_2^{(1)}} \prod_{i=n}^{m_3^{(1)}} (x - x_i^{(1)}), \tag{1.7}$$

$$(m_1^{(1)} + m_2^{(1)} + m_3^{(1)} = 2n,\ m_1^{(1)} \geqq 2n_1,\ m_2^{(1)} \geqq 2n_2,\ m_3^{(1)} > 0)$$

$$0 < x_2^{(1)} < x_3^{(1)} < \ldots < x_{m_3^{(1)}-1}^{(1)} < d\left(\frac{m_1^{(1)}}{m_2^{(1)}}\right), \tag{1.8}$$

$$-1 < x_1^{(1)} < x_2^{(1)},\ x_{m_3^{(1)}-1}^{(1)} < x_{m_3^{(1)}}^{(1)} < 1,$$

$$|f^{(1)}(x)| \leqq 1 \qquad -1 \leqq x \leqq 1, \tag{1.9}$$

und es gibt $m_3^{(1)} - 1$ Werte $\xi_i^{(1)}$ $(i = 1, 2, \ldots, m_3^{(1)} - 1)$ mit

$$|f^{(1)}(\xi_i^{(1)})| = 1 \quad \text{und} \quad x_i^{(1)} < \xi_i^{(1)} < x_{i+1}^{(1)}. \tag{1.10}$$

In analoger Weise erhält man das Polynom $f^{(2)}(x)$ $4n$-ten Grades aus $f^{(1)}(x)$, und so weiter. Es sei $m = m^{(\nu)} = 2^\nu n$ $\nu = 2, 3, \ldots$, so gelten für das Polynom $f^{(\nu)}(x)$ folgende Beziehungen:

$$|E(f^{(\nu-1)})| < |E(f^{(\nu)})|, \tag{1.11}$$

$$\text{grad} f^{(\nu)}(x) = 2^\nu n = m^{(\nu)} \tag{1.12}$$

$$f^{(\nu)}(x) = (x+1)^{m_1^{(\nu)}}(x-1)^{m_2^{(\nu)}} \prod_{i=1}^{m_3^{(\nu)}} (x - x_i^{(\nu)}), \tag{1.13}$$

$$(m_1^{(\nu)} + m_2^{(\nu)} + m_3^{(\nu)} = m^{(\nu)},\ m_1^{(\nu)} \geqq 2m_1^{(\nu-1)},\ m_2^{(\nu)} \geqq 2m_2^{(\nu-1)},\ m_3^{(\nu)} > 0)$$

$$0 \leqq x_2^{(\nu)} < x_3^{(\nu)} < \ldots < x_{m_3^{(\nu)}-1}^{(\nu)} < d\left(\frac{m_1^{(\nu)}}{m_2^{(\nu)}}\right), \tag{1.14}$$

$$-1 < x_1^{(\nu)} < x_2^{(\nu)},\ x_{m_3^{(\nu)}-1}^{(\nu)} < x_{m_3^{(\nu)}}^{(\nu)} < 1,$$

$$|f^{(\nu)}(x)| \leqq 1 \qquad -1 \leqq x \leqq 1, \tag{1.15}$$

und es gibt $m_3^{(\nu)} - 1$ Werte $\xi_i^{(\nu)}$ $(i = 1, 2, \ldots, m_3^{(\nu)} - 1)$ mit

$$|f^{(\nu)}(\xi_i^{(\nu)})| = 1 \qquad \text{und} \qquad x_i^{(\nu)} < \xi_i^{(\nu)} < x_{i+1}^{(\nu)}. \tag{1.16}$$

Definiert man die Größen $\alpha^{(\nu)}$, $\beta^{(\nu)}$, $\gamma^{(\nu)}$ durch

$$\alpha^{(\nu)} = \frac{m_1^{(\nu)}}{m^{(\nu)}}, \quad \beta^{(\nu)} = \frac{m_2^{(\nu)}}{m^{(\nu)}}, \quad \gamma^{(\nu)} = \frac{m_3^{(\nu)}}{m^{(\nu)}}, \tag{1.17}$$

so erhält man aus (1.13):

$$\alpha^{(\nu)} \geqq \alpha^{(\nu-1)}, \quad \beta^{(\nu)} \geqq \beta^{(\nu-1)}, \quad \alpha^{(\nu)}+\beta^{(\nu)}+\gamma^{(\nu)} = 1,$$

hieraus folgt unmittelbar die Konvergenz der Folgen $\{\alpha^{(\nu)}\}$, $\{\beta^{(\nu)}\}$ bzw. $\{\gamma^{(\nu)}\}$:

$$\alpha = \lim_{\nu\to\infty} \alpha^{(\nu)}, \quad \beta = \lim_{\nu\to\infty} \beta^{(\nu)}, \quad \gamma = \lim_{\nu\to\infty} \gamma^{(\nu)} = 1-\alpha-\beta. \tag{1.18}$$

Es ist offenbar aus (1. 2):

$$\alpha \geqq \frac{n_1}{n} > \frac{1}{2}, \quad \beta \geqq \frac{n_2}{n} > 0. \tag{1.19}$$

Es seien die Funktionen $\Gamma_\nu(x)$ definiert wie folgt:

$$\Gamma_\nu(x) = \begin{cases} 0 & x \leqq x_1^{(\nu)} \\ \dfrac{i-1}{m^{(\nu)}} + \dfrac{x-x_i^{(\nu)}}{m^{(\nu)}(x_{i+1}^{(\nu)}-x_i^{(\nu)})} & x_i^{(\nu)} \leqq x \leqq x_{i+1}^{(\nu)},\ i = 1, 2, \ldots, m_3^{(\nu)}-1 \\ \dfrac{m_3^{(\nu)}-1}{m^{(\nu)}} & x \geqq x_{m_3^{(\nu)}}^{(\nu)}. \end{cases} \tag{1.20}$$

Diese Funktionen sind nichtabnehmende Funktionen von x und es gelten folgende Hilfssätze.

HILFSSATZ 1. Wenn $x'<x''$ und $m^{(\nu)}[\Gamma_\nu(x'')-\Gamma_\nu(x')]>2$ ist, so existieren mindestens zwei Wurzeln vom Polynom $f^{(\nu)}(x)$ im Intervall $[x', x'']$.

BEWEIS. Auf Grund der Definition (1. 20) ist dies trivial.

HILFSSATZ 2. Die Funktionenfolge $\{\Gamma_\nu(x)\}$ ist gleichgradig stetig im Intervall $[0, d]$, wo d eine beliebige reelle Zahl ist, die der Ungleichung $d\left(\frac{\alpha}{\beta}\right)<d<1$ genügt.

BEWEIS. Zuerst werden wir die Richtigkeit der Ungleichung

$$x_{i+1}^{(\nu)} - x_i^{(\nu)} \geqq \frac{2\sqrt{1-d^2}}{m^{(\nu)}} \qquad (i = 1, 2, \ldots, m_3^{(\nu)}-1) \tag{1.21}$$

für $\nu>\nu_0$ beweisen, wo ν_0 eine, nur von d abhängige, genügend grosse Zahl ist. Nach (1.17) und (1.18) gilt

$$\lim_{\nu\to\infty} d\left(\frac{m_1^{(\nu)}}{m_2^{(\nu)}}\right) = \lim_{\nu\to\infty} d\left(\frac{\alpha^{(\nu)}}{\beta^{(\nu)}}\right) = d\left(\frac{\alpha}{\beta}\right), \tag{1.22}$$

so gilt auch

$$d\left(\frac{m_1^{(\nu)}}{m_2^{(\nu)}}\right) < \frac{1}{2}\left[d+d\left(\frac{\alpha}{\beta}\right)\right] \quad \text{für} \quad \nu > \nu'.$$

Es sei

$$\nu_0 = \max\left\{\nu', \frac{4\sqrt{1-d^2}}{d-d\left(\frac{\alpha}{\beta}\right)}, \frac{2\sqrt{1-d^2}}{d}\right\}.$$

Aus einem bekannten Satz von MARKOFF [3] und aus (1.15) folgt

$$|(f^{(\nu)}(x))'| \leqq \frac{m^{(\nu)}}{\sqrt{1-x^2}} \leqq \frac{m^{(\nu)}}{\sqrt{1-d^2}} \quad \text{für} \quad |x| \leqq d,$$

und aus (1.14) bzw. (1.16) für $i=2, 3, \ldots, m_3^{(\nu)}-2$ und $\nu > \nu_0$

$$1 = |f^{(\nu)}(\xi_i^{(\nu)}) - f^{(\nu)}(x_i^{(\nu)})| \leqq \frac{m^{(\nu)}}{\sqrt{1-d^2}}(\xi_i^{(\nu)} - x_i^{(\nu)}),$$

$$1 = |f^{(\nu)}(\xi_i^{(\nu)}) - f^{(\nu)}(x_{i+1}^{(\nu)})| \leqq \frac{m^{(\nu)}}{\sqrt{1-d^2}}(x_{i+1}^{(\nu)} - \xi_i^{(\nu)}),$$

hieraus folgt aber

$$\xi_i^{(\nu)} - x_i^{(\nu)} \geqq \frac{\sqrt{1-d^2}}{m^{(\nu)}}$$

(1.23)

$$x_{i+1}^{(\nu)} - \xi_i^{(\nu)} \geqq \frac{\sqrt{1-d^2}}{m^{(\nu)}}.$$

Die Summen der rechten und linken Seiten dieser Ungleichungen liefern die Ungleichung (1.21) für $i=2, 3, \ldots, m_3^{(\nu)}-2$. Wenn $-d \leqq x_1^{(\nu)}$ ist, so erhält man die gewünschte Ungleichung (1.21) für $i=1$ wie im Falle $i=2, 3, \ldots, m_3^{(\nu)}-2$, wenn aber $-1 < x_1^{(\nu)} < -d$ ist, so ist $x_2^{(\nu)} - x_1^{(\nu)} > d \geqq \frac{2\sqrt{1-d^2}}{\nu_0} > \frac{2\sqrt{1-d^2}}{m^{\nu}}$. Der Beweis für $i=m_3^{(\nu)}-1$ geht auch auf diese Weise. Wenn $x_{m_3^{(\nu)}}^{(\nu)} \leqq d$ ist, so erhält man wie im Falle $i=2, 3, \ldots, m_3^{(\nu)}-2$: $x_{m_3^{(\nu)}}^{(\nu)} - x_{m_3^{(\nu)}-1}^{(\nu)} \geqq \frac{2\sqrt{1-d^2}}{m^{(\nu)}}$, und wenn $x_{m_3^{(\nu)}}^{(\nu)} > d$ ist, so haben wir:

$$x_{m_3^{(\nu)}}^{(\nu)} - x_{m_3^{(\nu)}-1}^{(\nu)} > d - d\left(\frac{m_1^{(\nu)}}{m_2^{(\nu)}}\right) > \frac{1}{2}\left[d - d\left(\frac{\alpha}{\beta}\right)\right] \geqq \frac{2\sqrt{1-d^2}}{\nu_0} > \frac{2\sqrt{1-d^2}}{m^{(\nu)}}.$$

Damit ist die Richtigkeit der Ungleichung (1.21) für $\nu \geqq \nu_0$ völlig bewiesen.

Aus der Definition von $\Gamma_\nu(x)$ durch (1.20) und aus (1.21) folgt unmittelbar:

$$|\Gamma_\nu(x') - \Gamma_\nu(x'')| \leqq \frac{|x'-x''|}{2\sqrt{1-d^2}} \quad \text{für} \quad \nu \geqq \nu_0, \tag{1.24}$$

d.h. daß die Funktionenfolge $\{\Gamma_\nu(x)\}$ gleichgradig stetig ist. Damit ist der Hilfssatz 2 bewiesen.

Nach einem bekannten Satz von ARZELÀ kann man eine unendliche, gleichmäßig konvergente Teilfolge $\{\Gamma_{\nu_k}(x)\}$ aus der Folge $\{\Gamma_\nu(x)\}$ auswählen:

$$\lim_{k\to\infty} \Gamma_{\nu_k}(x) = \Gamma(x) \qquad 0 \leqq x \leqq d. \tag{1.25}$$

Die Grenzfunktion $\Gamma(x)$ ist auch eine nicht abnehmende Funktion von x und genügt nach (1.24) der Ungleichung

$$|\Gamma(x') - \Gamma(x'')| \leqq \frac{|x'-x''|}{2\sqrt{1-d^2}}. \tag{1.26}$$

Aus (1.14) erhalten wir die Ungleichung

$$0 \leqq \Gamma_\nu(0) \leqq \frac{1}{m^{(\nu)}}$$

und hieraus:

(1. 27) $$\Gamma(0)=0.$$

Für beliebige reelle d' mit $d\left(\frac{\alpha}{\beta}\right) < d' < d$ existiert nach (1. 22) eine natürliche Zahl $\nu = \nu(d')$ so, daß für alle $\nu \geqq \nu(d')$ die Ungleichung

$$d\left(\frac{m_1^{(\nu)}}{m_2^{(\nu)}}\right) < d'$$

gilt. Aus (1.14) erhalten wir

$$\frac{m_3^{(\nu)}-2}{m^{(\nu)}} \leqq \Gamma_\nu(d') \leqq \frac{m_3^{(\nu)}-1}{m^{(\nu)}} \qquad (\nu \geqq \nu(d')),$$

d.h. nach (1.17)

$$\gamma^{(\nu)} - \frac{2}{m^{(\nu)}} \leqq \Gamma_\nu(d') \leqq \gamma^{(\nu)} - \frac{1}{m^{(\nu)}},$$

und nach (1.18) bzw. (1. 25)

$$\Gamma(d') = \gamma \quad \text{für} \quad d\left(\frac{\alpha}{\beta}\right) < d' < d.$$

Die Funktion $\Gamma(x)$ ist stetig, deshalb folgt: $\Gamma\left(d\left(\frac{\alpha}{\beta}\right)\right) = \gamma$.

Es sei

(1.28)
$$a = \max\{x;\ \Gamma(x) = 0,\ x \geqq 0\}$$
$$b = \min\left\{x;\ \Gamma(x) = \gamma,\ x \leqq d\left(\frac{\alpha}{\beta}\right)\right\}.$$

Aus Stetigkeitsgründen folgt unmittelbar:

(1. 29) $$\Gamma(a)=0,\ \Gamma(b)=\gamma.$$

Es seien die folgenden Funktionen für alle reelen Zahlen x definiert:

(1. 30) $$\psi_\nu(x) = \int_a^b \log|x-t|\, d\Gamma_\nu(t) \qquad (\nu = 1, 2, \ldots)$$

(1. 31) $$F_\nu(x) = \frac{1}{m^{(\nu)}} \sum_{i=1}^{m_3^{(\nu)}} \log|x - x_i^{(\nu)}| \qquad (\nu = 1, 2, \ldots).$$

HILFSATZ 3. Die Funktionenfolge $\{\psi_\nu(x)\}$ ist gleichgradig stetig im Intervall $(-\infty, \infty)$. Hier gilt nämlich die Ungleichung

(1. 32) $$|\psi_\nu(x+h) - \psi_\nu(x)| < \eta(h) \quad \text{für} \quad 0 < h < 3-2\sqrt{2} \quad \text{und} \quad \nu > \nu_0$$

mit

$$\eta(h) = \sqrt{h} + \frac{2}{\sqrt{1-d^2}} \int_0^{h+2\sqrt{h}} \log \frac{1}{t}\, dt, \tag{1.33}$$

wo d und v_0 die schon früher definierten Größen sind.

(Die Zahl $3-2\sqrt{2}$ ist dadurch gerechtfertigt, dass die Ungleichung $h+2\sqrt{h} \leqq 1$ gelten muss).

BEWEIS. Wir können uns auf den Fall $a < x-\sqrt{h} < x+h+\sqrt{h} < b$ beschränken, da der Beweis für andere Werte x in analoger Weise geht. Aus (1. 30) folgt

$$|\psi_v(x+h)-\psi_v(x)| \leqq \left| \int_a^{x-\sqrt{h}} [\log(x+h-t)-\log(x-t)]\, d\Gamma_v(t) \right| +$$

$$+ \left| \int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x-t|\, d\Gamma_v(t) \right| + \left| \int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x+h-t|\, d\Gamma_v(t) \right| +$$

$$+ \left| \int_{x+h+\sqrt{h}}^{b} [\log(t-x-h)-\log(t-x)\, d\Gamma_v(t) \right|.$$

Hieraus erhalten wir durch Anwendung der Ungleichung $\log(1+x) \leqq x$ und aus (1. 24)

$$|\psi_v(x+h)-\psi_v(x)| \leqq \int_a^{x-\sqrt{h}} \frac{h}{x-t}\, d\Gamma_v(t) +$$

$$+ \frac{2}{\sqrt{1-d^2}} \int_0^{h+2\sqrt{h}} \log \frac{1}{t}\, dt + \int_{x+h+\sqrt{h}}^{b} \frac{h}{t-x-h}\, d\Gamma_v(t) <$$

$$< \sqrt{h} \int_a^{x-\sqrt{h}} d\Gamma_v(t) + \frac{2}{\sqrt{1-d^2}} \int_0^{h+2\sqrt{h}} \log \frac{1}{t}\, dt +$$

$$+ \sqrt{h} \int_{x+h+\sqrt{h}}^{b} d\Gamma_v(t) < \sqrt{h} + \frac{2}{\sqrt{1-d^2}} \int_0^{h+2\sqrt{h}} \log \frac{1}{t}\, dt = \eta(h),$$

wie wir es behaupteten.

HILFSSAFZ 4. Für die Teilfolge $\{\Gamma_{v_k}(x)\}$ gilt

$$F_{v_k}(x) - \psi_{v_k}(x) \Rightarrow 0 \quad \text{wenn } k \to \infty \tag{1.34}$$

im Intervall $(-\infty, \infty)$, wo das Zeichen $\Rightarrow$ die Konvergenz nach Mass bedeutet.

BEWEIS. Es sei c eine reelle Zahl mit $0 < c < \sqrt{1-d^2}$ und

$$I_c^{(v)} = \bigcup_{i=1}^{m^{(v)}} \left(x_i^{(v)} - \frac{c}{m^{(v)}}, x_i^{(v)} + \frac{c}{m^{(v)}} \right).$$

Es ist offenbar, daß $|I_c^{(\nu)}| = \frac{2c}{m^{(\nu)}} m_3^{(\nu)} = 2c\gamma^{(\nu)} < 2c$ ist. Es sei $x \in [a, b] - I_c^{(\nu)}$. Wir nehmen an, daß es einen Index $r^{(\nu)}$ mit $x_{r^{(\nu)}}^{(\nu)} < x < x_{r^{(\nu)}+1}^{(\nu)}$ gibt, und es seien $p^{(\nu)}$ und $q^{(\nu)}$ diejenigen Indices, für welche die Relationen $p^{(\nu)} = \min\{i;\ x_i^{(\nu)} \geqq a\}$ bzw. $q^{(\nu)} = \max\{i;\ x_i^{(\nu)} \leqq b\}$ gelten so, daß

$$|F_\nu(x) - \psi_\nu(x)| \leqq \left| \frac{1}{m^{(\nu)}} \sum_{i=1}^{p^{(\nu)}} \log|x - x_i^{(\nu)}| \right| +$$

$$+ \left| \int_a^{x_{p^{(\nu)}}^{(\nu)}} \log|x-t|\, d\Gamma_\nu(t) \right| + \left| \frac{1}{m^{(\nu)}} \sum_{i=p^{(\nu)}+1}^{r^{(\nu)}} \log|x - x_i^{(\nu)}| - \right.$$

$$\left. - \int_{x_{p^{(\nu)}}^{(\nu)}}^{x_{r^{(\nu)}}^{(\nu)}} \log|x-t|\, d\Gamma_\nu(t) \right| + \left| \int_{x_{r^{(\nu)}}^{(\nu)}}^{x_{r^{(\nu)}+1}^{(\nu)}} \log|x-t|\, d\Gamma_\nu(t) \right| +$$

$$+ \left| \frac{1}{m^{(\nu)}} \sum_{i=r^{(\nu)}+1}^{q^{(\nu)}-1} \log|x - x_i^{(\nu)}| - \int_{x_{r^{(\nu)}+1}^{(\nu)}}^{x_{q^{(\nu)}}^{(\nu)}} \log|x-t|\, d\Gamma_\nu(t) \right| +$$

$$+ \left| \frac{1}{m^{(\nu)}} \sum_{i=q^{(\nu)}}^{m_3^{(\nu)}} \log|x - x_i^{(\nu)}| \right| + \left| \int_{x_{q^{(\nu)}}^{(\nu)}}^{b} \log|x-t|\, d\Gamma_\nu(t) \right|.$$

Aus der Definition (1. 20) erhalten wir:

$$\left| \frac{1}{m^{(\nu)}} \sum_{i=p+1}^{r} \log(x - x_i^{(\nu)}) - \int_{x_p^{(\nu)}}^{x_r^{(\nu)}} \log(x-t)\, d\Gamma_\nu(t) \right| =$$

$$= \frac{1}{m^{(\nu)}} \left| \sum_{i=p+1}^{r} \left[\log(x - x_i^{(\nu)}) - \int_{x_{i-1}^{(\nu)}}^{x_i^{(\nu)}} \log(x-t) \frac{dt}{x_i^{(\nu)} - x_{c-1}^{(\nu)}} \right] \right| <$$

$$< \frac{1}{m^{(\nu)}} \sum_{i=p+1}^{r} \log \frac{x - x_{i-1}^{(\nu)}}{x - x_i^{(\nu)}} = \frac{1}{m^{(\nu)}} \log \frac{x - x_p^{(\nu)}}{x - x_r^{(\nu)}} \leqq$$

$$\leqq \frac{1}{m^{(\nu)}} \log \frac{\frac{c}{m^{(\nu)}} + x_r^{(\nu)} - x_p^{(\nu)}}{\frac{c}{m^{(\nu)}}} < \frac{1}{m^{(\nu)}} \log \frac{3m^{(\nu)}}{c},$$

$p = p^{(\nu)}$, $r = r^{(\nu)}$ ist, und analog

$$\left| \frac{1}{m^{(\nu)}} \sum_{i=r+1}^{q-1} \log(x_i^{(\nu)} - x) - \int_{x_{r+1}^{(\nu)}}^{x_q^{(\nu)}} \log(t-x)\, d\Gamma_\nu(t) \right| < \frac{1}{m^{(\nu)}} \log \frac{3m^{(\nu)}}{c}.$$

Ferner ist

$$\left|\int_{x_r^{(\nu)}}^{x_{r+1}^{(\nu)}} \log|x-t|\,d\Gamma_\nu(t)\right| \leqq \frac{1}{m^{(\nu)}(x_{r+1}^{(\nu)}-x_r^{(\nu)})}\left|\int_{x_r^{(\nu)}}^{x}\log(x-t)\,dt+\int_{x}^{x_{r+1}^{(\nu)}}\log(t-x)\,dt\right| =$$

$$= \frac{1}{m^{(\nu)}(x_{r+1}^{(\nu)}-x_r^{(\nu)})}\left[x_{r+1}^{(\nu)}-x_r^{(\nu)}+(x-x_r^{(\nu)})\log\frac{1}{x-x_r^{(\nu)}}+(x_{r+1}^{(\nu)}-x)\log\frac{1}{x_{r+1}^{(\nu)}-x}\right] <$$

$$< \frac{1}{m^{(\nu)}}\log\frac{em^{(\nu)}}{c}.$$

Ähnlich erhalten wir:

$$\left|\int_a^{x_p^{(\nu)}}\log(x-t)\,d\Gamma_\nu(t)\right| \begin{cases} = 0 & \text{wenn} \quad p=1 \\ < \left|\int_{x_{p-1}^{(\nu)}}^{x_p^{(\nu)}}\log(x-t)\,d\Gamma_\nu(t)\right| < \frac{1}{m^{(\nu)}}\log\frac{em^{(\nu)}}{c} & \text{wenn} \quad p \geqq 2, \end{cases}$$

d.h.

$$\left|\int_a^{x_p^{(\nu)}}\log|x-t|\,d\Gamma_\nu(t)\right| < \frac{1}{m^{(\nu)}}\log\frac{em^{(\nu)}}{c}$$

und

$$\left|\int_{x_q^{(\nu)}}^{b}\log|x-t|\,d\Gamma_\nu(t)\right| < \frac{1}{m^{(\nu)}}\log\frac{em^{(\nu)}}{c}.$$

Aus der Voraussetzung $x \notin I_c^{(\nu)}$ und (1. 21) erhalten wir

$$\frac{1}{m^{(\nu)}}\left|\sum_{i=1}^{p}\log|x-x_i^{(\nu)}|\right| < \frac{1}{m^{(\nu)}}\left|\sum_{i=1}^{p}\log\frac{c}{m^{(\nu)}}i\right| < \frac{1}{c}\int_0^{\frac{pc}{m^{(\nu)}}}\log\frac{1}{t}\,dt,$$

es ist jedoch $\Gamma_\nu(x_p^{(\nu)}) = \dfrac{p-1}{m^{(\nu)}}$ und $\Gamma_\nu(x_p^{(\nu)}) < \Gamma_\nu(a)+\dfrac{1}{m^{(\nu)}}$, daher

$$\frac{1}{m^{(\nu)}}\left|\sum_{i=1}^{p}\log|x-x_i^{(\nu)}|\right| < \frac{1}{c}\int_0^{c\Gamma_\nu(a)+\frac{2c}{m^{(\nu)}}}\log\frac{1}{t}\,dt,$$

und ganz analog

$$\frac{1}{m^{(\nu)}}\left|\sum_{i=q}^{m_3^{(\nu)}}\log|x-x_i^{(\nu)}|\right| < \frac{1}{c}\int_0^{c[\gamma^{(\nu)}-\Gamma_\nu(b)]+\frac{c}{m^{(\nu)}}}\log\frac{1}{t}\,dt.$$

Wir erhalten also die Ungleichung für $v \geqq v_0$

$$|F_v(x)-\psi_v(x)| < \frac{1}{m^{(v)}} \log \frac{9e^3[m^{(v)}]^5}{c^5} + \tag{1.35}$$

$$+\frac{1}{c} \int_0^{c\Gamma_v(a)+\frac{2c}{m^{(v)}}} \log \frac{1}{t}\, dt + \frac{1}{c} \int_0^{c[\gamma^{(v)}-\Gamma_v(b)]+\frac{c}{m^{(v)}}} \log \frac{1}{t}\, dt.$$

Diese Ungleichung gilt nicht nur für $x \in [a, b] - I_c^{(v)}$, sondern auch für alle $x \notin I_c^{(v)}$ der Beweis geht in einzelnen speziällen Fällen ganz anlog wie oben. Weiterhin gilt es für beliebige $\varepsilon > 0$ nach (1. 25) bzw. (1. 29):

$$|F_{v_k}(x)-\psi_{v_k}(x)| < \varepsilon \quad \text{für} \quad k > K_c(\varepsilon) \quad \text{und} \quad x \notin I_c^{(v_k)},$$

wo $K_c(\varepsilon)$ eine nur von c und ε abhängige Konstante ist, daher ist:

$$\lambda(\{x;\ |F_{v_k}(x)-\psi_{v_k}(x)| \geqq \varepsilon,\ -\infty < x < \infty\}) < 2c \quad \text{für} \quad k > K_c(\varepsilon).$$

Da der Wert von c beliebig klein sein kann, erhalten wir:

$$\lim_{k\to\infty} \lambda(\{x;\ |F_{v_k}(x)-\psi_{v_k}(x)| \geqq \varepsilon,\ -\infty < x < \infty\}) = 0,$$

d.h. die Behauptung (1. 34) ist richtig, womit der Hilfssatz 4. bewiesen ist.

Wenden wir den Satz von BRAY [4] auf die Funktionenfolgen $\{\Gamma_{v_k}(x)\}$ und $\{\psi_{v_k}(x)\}$ an:

$$\lim_{k\to\infty} \psi_{v_k}(x) = \lim_{k\to\infty} \int_a^b \log|x-t|\, d\Gamma_{v_k}(t) = \int_a^b \log|x-t|\, d\Gamma(t) \overset{\text{def}}{=} \psi(x). \tag{1.36}$$

Nach (1. 26) ist die Funktion $\psi(x)$ stetig und nach dem Hilfssatz 4. ist

$$F_{v_k}(x) \rightrightarrows \psi(x) \quad \text{wenn} \quad k \to \infty \quad \text{im Intervall} \quad (-\infty, \infty). \tag{1.37}$$

Aus (1. 13), (1. 15) und (1. 31) erhalten wir für $-1 \leqq x \leqq 1$

$$\frac{1}{m^{(v)}} \log|f^{(v)}(x)| = \alpha^{(v)} \log|x+1| + \beta^{(v)} \log|x-1| + F_v(x) \leqq 0, \tag{1.38}$$

aus (1.18) und (1.37) folgt aber für $-1 \leqq x \leqq 1$

$$\varphi(x) \overset{\text{def}}{=} \alpha \log|x+1| + \beta \log|x-1| + \int_a^b \log|x-t|\, d\Gamma(t) \leqq 0. \tag{1.39}$$

HILFSSATZ 5. Im Intervall $[a, b]$ gilt die Gleichung $\varphi(x) = 0$.

BEWEIS. Zuerst beschäftigen wir uns mit x_0-Werten, für welche eine der Beziehungen:

$$\Gamma(x) - \Gamma(x_0) > 0 \quad \text{für alle } x \text{ mit} \quad x > x_0 \tag{1.40}$$

oder

$$\Gamma(x_0) - \Gamma(x) > 0 \quad \text{für alle } x \text{ mit} \quad x < x_0 \tag{1.41}$$

gilt. Wir werden für diese x_0 die Richtigkeit der Gleichung $\varphi(x_0)=0$ zeigen. Es sei vorausgesetzt, da (1. 40) erfüllt ist, so folgt aus (1. 25) für beliebige festgelegte x' mit $x'>x_0$:

$$\Gamma_{\nu_k}(x')-\Gamma_{\nu_k}(x_0)>\tfrac{1}{2}[\Gamma(x')-\Gamma(x_0)] \quad \text{für} \quad k>K_1,$$

und auch

$$m^{(\nu_k)}[\Gamma_{\nu_k}(x')-\Gamma_{\nu_k}(x_0)]>2 \quad \text{für} \quad k>K_2(\geqq K_1),$$

wo K_1 und K_2 genügend große reelle Zahlen sind. Nach dem Hilfssatz 1 gibt es mindestens zwei Wurzeln von $f^{(\nu_k)}(x)$ im Intervall $[x_0, x']$: $x_j^{(\nu)}$ und $x_{j+1}^{(\nu_k)}$, wo $j=j^{(\nu_k)}$ ist. Nach (1.16) ist $|f^{(\nu_k)}(\xi_j^{(\nu_k)})|=1$, daher

$$|\varphi(x_0)| = \left|\varphi(x_0)-\frac{1}{m^{(\nu_k)}}\log|f^{(\nu_k)}(\xi_j^{(\nu_k)})|\right| = |\alpha\log|x_0+1|+\beta\log|x_0-1|+\psi(x_0)-$$

$$-\alpha^{(\nu_k)}\log|\xi_j^{(\nu_k)}+1|-\beta^{(\nu_k)}\log|\xi_j^{(\nu_k)}-1|-F_{\nu_k}(\xi_j^{(\nu_k)})| \leqq$$

$$\leqq |(\alpha-\alpha^{(\nu_k)})\log|x_0+1|+(\beta-\beta^{(\nu_k)})\log|x_0-1||+\alpha^{(\nu_k)}|\log|x_0+1|-\log|\xi_j^{(\nu_k)}+1||+$$

$$+\beta^{(\nu_k)}|\log|x_0-1|-\log|\xi_j^{(\nu_k)}-1||+|\psi(x_0)-\psi_{\nu_k}(x_0)|+|\psi_{\nu_k}(x_0)-\psi_{\nu_k}(\xi_j^{(\nu_k)})|+$$

$$+|\psi_{\nu_k}(\xi_j^{(\nu_k)})-F_{\nu_k}(\xi_j^{(\nu_k)})|.$$

Da $\log x$ eine monoton wachsende Funktion von x ist, so erhalten wir

$$\alpha^{(\nu_k)}|\log|x_0+1|-\log|\xi_j^{(\nu_k)}+1|| < \log\frac{1+x'}{1+x_0},$$

und

$$\beta^{(\nu_k)}|\log|x_0-1|-\log|\xi_j^{(\nu_k)}-1|| < \log\frac{1-x'}{1-x_0},$$

und für beliebige, festgelegte Werte c mit $0<c<\sqrt{1-d^2}$ gilt es nach (1. 23) $\xi_j^{(\nu_k)}\notin I_c^{(\nu_k)}$, hieraus und aus (1. 35) folgt

$$\lim_{k\to\infty}|\psi_{\nu_k}(\xi_{j^{(\nu_k)}}^{(\nu_k)})-F_{\nu_k}(\xi_{j^{(\nu_k)}}^{(\nu_k)}| = 0.$$

Aus (1.18), (1. 32) und (1.6) folgt

$$|\varphi(x_0)| < \log\frac{1+x'}{1+x_0}+\log\frac{1-x_0}{1-x'}+\eta(x'-x_0)$$

für alle x' mit $x'>x_0$, also ist $\varphi(x_0)=0$. Der Beweis für der Fall $x<x_0$ geht analog. Wenden wir dieses Ergebnis für $x_0=a$ bzw. $x_0=b$ an. Nach (1. 28) genügt a der Beziehung (1. 40) und b der Beziehung (1. 41), also $\varphi(a)=\varphi(b)=0$. Es genügt zu beweisen, daß die Funktion $\Gamma(x)$ streng monoton wachsend ist. Wäre dies falsch, so gäbe es ein Intervall $[u, v]\subset(a, b)$, wo $\Gamma(x)$ konstant wäre. Wir können annehmen, daß $u=\min\{x;\ \Gamma(x)=\Gamma(u),\ a\leqq x\leqq v\}$ und $v=\max\{x;\ \Gamma(x)=\Gamma(u),\ u\leqq x\leqq b\}$ ist. Für diese u und v wäre nach dem obigen Resultat auch $\varphi(u)=0$, $\varphi(v)=0$ ferner

$$\varphi(x) = \alpha\log|x+1|+\beta\log|x-1|+\int_a^u\log|x-t|\,d\Gamma(t)+\int_v^b\log|x-t|\,d\Gamma(t).$$

Die Funktion $\varphi(x)$ wäre konkav im Intervall (u, v), hieraus wäre

$$\varphi\left(\frac{u+v}{2}\right) > \frac{1}{2}[\varphi(u)+\varphi(v)] = 0,$$

aber dies widerspräche der Ungleichung (1. 39), also ist $\varphi(x)=0$ im Intervall $[a, b]$, womit der Hilfssatz 5 bewiesen ist.

Die Funktion $\Gamma(x)$ ist monoton wachsend im Intervall $[a, b]$, deshalb existiert ihre Ableitung $\gamma(x)$ fast überall:

$$\gamma(x) = \frac{d}{dx}\Gamma(x),$$

und aus (1. 26) folgt

$$0 \leqq \gamma(x) \leqq \frac{1}{2\sqrt{1-d^2}}, \tag{1.42}$$

aus (1.29)

$$\int_a^b \gamma(x)\,dx = \gamma, \tag{1.43}$$

und

$$\varphi(x) = \alpha \log|x+1| + \beta \log|x-1| + \int_a^b \log|x-t|\gamma(t)\,dt \qquad -\infty < x < \infty \tag{1.44}$$

und speziell

$$\alpha \log(1+x) + \beta \log(1-x) + \int_a^b \log|x-t|\gamma(t)\,dt = 0 \qquad a \leqq x \leqq b. \tag{1.45}$$

Diese Gleichung ist eine Integralgleichung, welche leicht lösbar [5] ist:

$$\gamma(x) = \frac{\gamma}{\pi\sqrt{(b-x)(x-a)}} - \frac{1}{\pi^2\sqrt{(b-x)(x-a)}} \int_a^b \left(\frac{\beta}{1-t} - \frac{\alpha}{1+t}\right) \frac{\sqrt{(b-t)(t-a)}}{x-t}\,dt,$$

wo das Integral als ein Integral im Sinne des Cauchy'-schen Hauptwertes zu verstehen ist.

Durch Integration erhalten wir

$$\gamma(x) = \frac{1-x^2-\alpha(1-x)\sqrt{(1+a)(1+b)}-\beta(1+x)\sqrt{(1-a)(1-b)}}{\pi(1-x^2)\sqrt{(b-x)(x-a)}} \qquad (a \leqq x \leqq b).$$

Der Zähler muss Nullstellen in $x=a$ und $x=b$ nach (1. 42) haben, d.h. es gelten die Beziehungen:

$$\gamma(x) = \frac{1}{\pi}\frac{\sqrt{(b-x)(x-a)}}{1-x^2} \qquad a \leqq x \leqq b \tag{1.46}$$

$$\alpha = \frac{\sqrt{(1+a)(1+b)}}{2}, \qquad \beta = \frac{\sqrt{(1-a)(1-b)}}{2}. \tag{1.47}$$

Multipliziert man die Integralgleichung (1. 45) mit $\frac{1}{\sqrt{(b-x)(x-a)}}$ und integriert man nach x von a bis b, so findet man

$$\int_a^b [\alpha \log(1+x)+\beta \log(1-x)] \frac{dx}{\sqrt{(b-x)(x-a)}} + \int_a^b \gamma(t) \int_a^b \frac{\log|x-t|}{\sqrt{(b-x)(x-a)}} dx\, dt = 0,$$

da aber

$$\int_a^b \frac{\log|x-t|}{\sqrt{(b-x)(x-a)}} dx = \pi \log \frac{b-a}{4}$$

ist, so haben wir:

$$\gamma = \frac{1}{\pi \log \frac{4}{b-a}} \int_a^b [\alpha \log(1+x)+\beta \log(1-x)] \frac{dx}{\sqrt{(b-x)(x-a)}}.$$

Hieraus und aus (1. 47) folgt unmittelbar

$$(1.48) \quad (1+\alpha+\beta)^{1+\alpha+\beta}(1+\alpha-\beta)^{1+\alpha-\beta}(1-\alpha+\beta)^{1-\alpha+\beta}(1-\alpha-\beta)^{1-\alpha-\beta} = 4.$$

Diese Gleichung verbindet die beiden Größen α und β. Die Funktion $(1+x)^{1+x}(1-x)^{1-x}$ wächst von 1 bis 4, wenn x von 0 bis 1 wächst, es gibt also einen gewissen Wert α_0 mit $0<\alpha_0<1$ und

$$(1+\alpha_0)^{1+\alpha_0}(1-\alpha_0)^{1-\alpha_0} = 2.$$

Es ist $\alpha_0 = 0{,}7799\ldots$. Auch die Funktion $(y+x)^{y+x}(y-x)^{y-x}$ ist eine monoton wachsende Funktion von x, wenn x von 0 bis y wächst, also die linke Seite der Gleichung (1. 48) ist eine monoton wachsende Funktion von β für beliebig festgelegte α mit $\frac{1}{2}<\alpha<\alpha_0$, und für $\beta=0$ ist sie kleiner, für $\beta = 1-\alpha$ ist sie größer als 4, also für genau einen Wert $\beta=\beta(\alpha)$ ist die Gleichung (1. 48) erfüllt.

Mit Hilfe von α und $\beta(\alpha)$ können wir die Größen a und b aus (1. 47) als Funktion von α bestimmen: $a=a(\alpha)$, $b=b(\alpha)$. Dasselbe gilt für $\gamma(x)$ nach (1. 46) und für $\varphi(x)$ nach (1. 44). Für das Weitere werden wir die Größe α als Parameter betrachten und die entsprechenden Funktionen $\gamma(x)$ bzw. $\varphi(x)$ durch $\gamma_\alpha(x)$ bzw. $\varphi_\alpha(x)$ bezeichnen.

Aus (1. 37), (1. 38) und (1. 39) folgt also

$$\frac{1}{m^{(\nu_k)}} \log|f^{(\nu_k)}(x)| \Rightarrow \varphi_\alpha(x) \qquad -\infty<x<\infty$$

mit einem bestimmten α, wobei $\frac{1}{2}<\alpha<\alpha_0$. Aus (1. 6), (1.11) folgt:

$$(1.49) \qquad E_n = |E(f^*)| < |E(f^{(1)})| < \ldots < |E(\exp \varphi_\alpha(x))| \quad \text{mit} \quad \tfrac{1}{2}<\alpha<\alpha_0.$$

In dem zweiten Teil dieser Arbeit werden wir die Richtigkeit der Ungleichung

$$(1.50) \qquad |E(\exp \varphi_\alpha(x))| < 2\sqrt{2} \quad \text{für} \quad \tfrac{1}{2}<\alpha<\alpha_0$$

beweisen. Die Beziehungen (1. 49) und (1. 50) zeigen uns, dass ausser $f=(x^2-1)^{n/2}$ kein extremales Polynom geraden Grades existiert.

Ehe wir an den Beweis der Richtigkeit der Ungleichung (1. 50) herangehen, werden wir die anfangs angeführten zwei Folgerungen beweisen. Die Beweise führen wir indirekt.

Zuerst beschäftigen wir uns mit der Folgerung 2. Nehmen wir an, dass diese Folgerung falsch wäre, d.h. $|f(x)|\leqq 1$ im Intervall $[-1, 1]$. Dann erhalten wir eine Menge $F(f)$ von Polynomen aus dem Polynom $[f(x)]^2$ durch endlichmalige Anwendungen des Lemmas. In diesem Falle gäbe es also ein extremales Polynom $g^{(1)}(x)\in F_{2n}$ mit $|E(g^{(1)})|=\sup_{g\in F(f)}|E(g)|$. Hier hat das Polynom $g^{(1)}(x)$ eine ähnliche Gestalt wie $f^{(1)}(x)$; z.B. es gilt:

$$g^{(1)}(x)=(x+1)^{\bar{n}_1}(x-1)^{\bar{n}_2}\prod_{i=1}^{\bar{n}_3}(x-\bar{x}_i),$$

$$(\bar{n}_1+\bar{n}_2+\bar{n}_3=2n,\ \bar{n}_1\geqq 2n_1,\ \bar{n}_2\geqq 2n_2,\ \bar{n}_3>0).$$

Diese Formel ist analog zu (1. 7), und es gelten auch die Analoga der Beziehungen (1. 8)—(1. 10). Aus $g^{(1)}(x)$ erhalten wir jetzt das Polynom $g^{(2)}(x)$ $4n$-ten Grades, u.s.w. Ganz analog wie oben, erhalten wir auch die Integralgleichung (1. 45) und damit auch (1. 48) mit $\alpha\geqq\frac{n_1}{n}$ bzw. $\beta\geqq\frac{n_2}{n}$. Dies steht aber im Widerspruch zu der Ungleichung (0. 2), womit die Folgerung 2 bewiesen ist.

Wäre nun die Folgerung 1 falsch, so wäre $|f(x)|\leqq 1$ im Intervall $[-1, 1]$. Wir wiederholen den obigen Beweis, welcher zur Integralgleichung (1. 45) führt. Aus dem Polynom $f(x)$ können wir wieder, wie oben, ein extremales Polynom $g^{(1)}(x)$ $2n$-ten Grades mit $|g^{(1)}(x)|\leqq 1$ im Intervall $[-1, 1]$ erhalten. Für dieses gilt entweder

$$g^{(1)}(x)=(x+1)^{m_1^{(1)}}\prod_{i=2}^{m_3^{(1)}}(x-x_i^{(1)})\qquad(m_1^{(1)}+m_3^{(1)}=2n,\ m_1^{(1)}\geqq 2m,\ m_3^{(1)}>0)$$

oder

$$g^{(1)}(x)=(x+1)^{m_1^{(1)}}(x-1)^{m_2^{(1)}}\prod_{i=1}^{m_3^{(1)}}(x-x_i^{(1)})$$

$$(m_1^{(1)}+m_2^{(1)}+m_3^{(1)}=2n,\ m_1^{(1)}\geqq 2m,\ m_2^{(1)}>0,\ m_3^{(1)}>0),$$

und wiederum gelten die entsprechenden Beziehungen (1. 8)—(1.10), wo der Wert d im Falle $m_2=0$ gleich 1 zu nehmen ist. Aber ein Polynom von der zweiten Gestalt genügt den Bedingungen der Folgerung 2, da $\frac{m_1^{(1)}}{2n}>\alpha_0$ ist, was zu der Ungleichung $|g^{(1)}(x)|\leqq 1$ im Widerspruch steht, also $g^{(1)}(x)$ kann nur die erste Gestalt haben.

Aus dem Polynom $g^{(1)}(x)$ erhalten wir durch die mehrere Mal angewandte Methode das Polynom $g^{(2)}(x)$, welches wiederum keine Wurzel in $x=1$ besitzt

und aus $g^{(2)}(x)$ erhalten wir das Polynom $g^{(3)}(x)$, u.s.w. Für das Polynom $g^{(\nu)}(x)$ gilt:

$$(1.13') \qquad g^{(\nu)}(x) = (x+1)^{m_1^{(\nu)}} \prod_{i=1}^{m_3^{(\nu)}} (x - x_i^{(\nu)})$$

$$(m_1^{(\nu)} + m_3^{(\nu)} = 2^\nu n = m^{(\nu)},\ m_1^{(\nu)} \geqq 2m_1^{(\nu-1)},\ m_3^{(\nu)} > 0)$$

$$(1.14') \qquad 0 \leqq x_2^{(\nu)} < x_3^{(\nu)} < \ldots < x_{m_3^{(\nu)}}^{(\nu)} < 1, \qquad -1 < x_1^{(\nu)} < x_2^{(\nu)}$$

$$(1.15') \qquad |g^{(\nu)}(x)| \leqq 1 \quad \text{für} \quad x \in [-1, 1],$$

und es gibt $m_3^{(\nu)} - 1$ Werte $\xi_i^{(\nu)} (i = 1, 2, \ldots, m_3^{(\nu)} - 1)$ mit

$$(1.16') \qquad |g^{(\nu)}(\xi_i^{(\nu)})| = 1 \quad \text{und} \quad x_i^{(\nu)} < \xi_i^{(\nu)} < x_{i+1}^{(\nu)}.$$

Es sei

$$(1.17') \qquad \alpha^{(\nu)} = \frac{m_1^{(\nu)}}{m^{(\nu)}}, \quad \gamma^{(\nu)} = \frac{m_3^{(\nu)}}{m^{(\nu)}},$$

wobei $\alpha^{(\nu)} \geqq \alpha^{(\nu-1)}$ und $\alpha^{(\nu)} + \gamma^{(\nu)} = 1$. Wir können auch die Funktion $\Gamma_\nu(x)$ durch (1. 20) definieren. Der Hilfssatz 2 ist auch jetzt richtig. Nämlich aus dem schon erwähnten Satz von MARKOFF folgt:

$$1 = |g^{(\nu)}(x_i^{(\nu)}) - g^{(\nu)}(\xi_i^{(\nu)})| = \left| \int_{x_i^{(\nu)}}^{\xi_i^{(\nu)}} [g^{(\nu)}(x)]'\, dx \right| < \int_{x_i^{(\nu)}}^{\xi_i^{(\nu)}} \frac{m^{(\nu)}}{\sqrt{1-x^2}}\, dx$$

und

$$1 = |g^{(\nu)}(x_{i+1}^{(\nu)}) - g^{(\nu)}(\xi_i^{(\nu)})| < \int_{\xi_i^{(\nu)}}^{x_{i+1}^{(\nu)}} \frac{m^{(\nu)}}{\sqrt{1-x^2}}\, dx,$$

also

$$\frac{2}{m^{(\nu)}} < \int_{x_i^{(\nu)}}^{x_{i+1}^{(\nu)}} \frac{dx}{\sqrt{1-x^2}} = \arcsin x_{i+1}^{(\nu)} - \arcsin x_i^{(\nu)}.$$

Hieraus folgt es einerseits für $x' < x''$

$$\Gamma_\nu(x'') - \Gamma_\nu(x') < \frac{2}{m^{(\nu)}} + \frac{1}{2} \int_{x'}^{x''} \frac{dx}{\sqrt{1-x^2}},$$

d.h. daß die Funktionenfolge $\{\Gamma_\nu(x)\}$ gleichgradig stetig ist, deshalb können wir eine konvergente Teilfolge $\{\Gamma_{\nu_k}(x)\}$ mit einer Grenzfunktion $\Gamma(x)$ auswählen, wo die Ungleichung

$$(1.26') \qquad \Gamma(x'') - \Gamma(x') \leqq \frac{1}{2} \int_{x'}^{x''} \frac{dx}{\sqrt{1-x^2}} \qquad (x' < x'')$$

gilt; anderseits ist

$$x_i^{(\nu)} < \sin\left(\arcsin x_{i+1}^{(\nu)} - \frac{2}{m^{(\nu)}}\right) \leqq \cos\frac{2}{m^{(\nu)}}$$

also

$$x_{i+1}^{(\nu)} > \sin\left(\arcsin x_i^{(\nu)} + \frac{2}{m^{(\nu)}}\right) = x_i^{(\nu)}\cos\frac{2}{m^{(\nu)}} + \sqrt{1-[x_i^{(\nu)}]^2}\sin\frac{2}{m^{(\nu)}} =$$

$$= x_i^{(\nu)} + \sin\frac{1}{m^{(\nu)}}\cdot\sqrt{1-[x_i^{(\nu)}]^2} + \sqrt{1-[x_i^{(\nu)}]^2}\left(\sin\frac{2}{m^{(\nu)}} - \sin\frac{1}{m^{(\nu)}}\right) -$$

$$- x_i^{(\nu)}\left(1-\cos\frac{2}{m^{(\nu)}}\right) \geqq x_i^{(\nu)} + \sin\frac{1}{m^{(\nu)}}\cdot\sqrt{1-[x_i^{(\nu)}]^2} + \sin\frac{2}{m^{(\nu)}}\left(\sin\frac{2}{m^{(\nu)}} - \sin\frac{1}{m^{(\nu)}}\right) -$$

$$- \cos\frac{2}{m^{(\nu)}}\left(1-\cos\frac{2}{m^{(\nu)}}\right) > x_i^{(\nu)} + \sin\frac{1}{m^{(\nu)}}\cdot\sqrt{1-[x_i^{(\nu)}]^2},$$

d.h.

$$(1.21') \qquad x_{i+1}^{(\nu)} - x_i^{(\nu)} > \sin\frac{1}{m^{(\nu)}}\cdot\sqrt{1-[x_i^{(\nu)}]^2} \geqq \sin\frac{1}{m^{(\nu)}}\cdot\sqrt{1-x^2} \qquad (x_i^{(\nu)} \leqq x \leqq x_{i+1}^{(\nu)})$$

und es gelten auch die Ungleichungen

$$(1.23') \qquad \begin{aligned} \xi_i^{(\nu)} - x_i^{(\nu)} &> \sin\frac{1}{2m^{(\nu)}}\cdot\sqrt{1-[\xi_i^{(\nu)}]^2} \\ x_{i+1}^{(\nu)} - \xi_i^{(\nu)} &> \sin\frac{1}{2m^{(\nu)}}\cdot\sqrt{1-[\xi_i^{(\nu)}]^2}. \end{aligned}$$

Der Beweis der Ungleichungen (1. 23′) geht analog zu dem (1. 21′). Mit Hilfe von (1. 21′) können wir auch den Hilfssatz 3 mit modifiziertem $\eta(h)$ beweisen. Es gilt nämlich für die Funktion $\psi_\nu(x)$ mit $0<h<3-2\sqrt{2}$ und z.b. für $a+\sqrt{h}<x<$ $<b-h-\sqrt{h}$:

$$|\psi_\nu(x+h)-\psi_\nu(x)| \leqq \left|\int_a^{x-\sqrt{h}} \log[(x+h-t)-\log(x-t)]\,d\Gamma_\nu(t)\right| +$$

$$+\left|\int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x-t|\,d\Gamma_\nu(t)\right| + \left|\int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x+h-t|\,d\Gamma_\nu(t)\right| +$$

$$+\left|\int_{x+h+\sqrt{h}}^{b} [\log(t-x-h)-\log(t-x)]\,d\Gamma_\nu(t)\right| <$$

$$< \sqrt{h} + \frac{1}{m^{(\nu)}\sin\frac{1}{m^{(\nu)}}}\left\{\left|\int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x-t|\,\frac{dt}{\sqrt{1-t^2}}\right| +\right.$$

$$\left. +\left|\int_{x-\sqrt{h}}^{x+h+\sqrt{h}} \log|x+h-t|\,\frac{dt}{\sqrt{1-t^2}}\right|\right\}.$$

Hieraus erhalten wir durch einfache Abschätzungen:

$$|\psi_\nu(x+h)-\psi_\nu(x)|<\bar{\eta}(h) \tag{1.32'}$$

mit

$$\bar{\eta}(h)=\sqrt{h}+8\int_{1-h-\sqrt{h}}^{1}\log\frac{1}{1-t}\frac{dt}{\sqrt{1-t^2}}. \tag{1.33'}$$

Diese Ungleichungen liefern die Integralgleichung (1. 45) mit $\beta=0$ und auch die Gleichung (1. 48). Es gilt auch die Ungleichung (1. 39), hiermit ist es offenbar, daß $b=1$ ist, da es im Falle $b<1$ $\varphi(1)>\varphi(b)=0$ wäre, und dies stände im Widerspruch zu (1. 39).

Aus der Voraussetzung der Folgerung 1 erhalten wir

$$\alpha=\lim_{\nu\to\infty}\alpha^{(\nu)}\geqq\alpha^{(1)}=\frac{m_1^{(\nu)}}{2n}\geqq\frac{m}{n}>\alpha_0.$$

Die linke Seite der Gleichung (1. 48) ist aber grösser als 4 für $\alpha>\alpha_0$ und $\beta=0$, d.h. die Negation der Folgerung 1 führt zu einem Widerspruch, womit der Beweis der Folgerung 1 beendet ist.

2. Der Beweis der Ungleichung (0.1)

Bezeichnen wir durch $\delta(f)$ den Durchmesser der Menge $E(f)$, wo $f=f(x)$ eine stetige Funktion ist. Dann gilt die Ungleichung

$$|E(f)|\leqq\delta(f). \tag{2.1}$$

Betrachten wir die Größe $\delta(|x+1|^m|x-1|)$! Durch den Gedankengang, welcher in der Arbeit [3] angewandt ist, werden wir zeigen

$$\delta(|x+1|^m|x-1|)<2\sqrt{2}\quad\text{für}\quad 1<m<m_0, \tag{2.2}$$

wo m_0 eine später gegebene Zahl ist.

Es seien $A(m)$ bzw. $-B(m)$ die Abscissen des rechts- bzw. linksseitigen Endpunktes der Menge $E(|x+1|^m|x-1|)$, und ferner

$$A(m)=\sqrt{2}-s(m).$$
$$B(m)=\sqrt{2}+t(m).$$

Der Definition nach gilt es

$$\begin{aligned}(R-s)^m(R^{-1}-s)&=1\\(R^{-1}+t)^m(R+t)&=1,\end{aligned} \tag{2.3}$$

wo $R=1+\sqrt{2}$ ist.

$s=s(m)$ und $t=t(m)$ sind monoton wachsende Funktionen von m. Da $\delta(|x+1|^m|x+1|)=A(m)+B(m)=2\sqrt{2}+t(m)-s(m)$ ist, so genügt es um die Richtigkeit von (2. 2) zu beweisen:

$$t(m)<s(m)\quad\text{für}\quad 1<m<m_0. \tag{2.4}$$

Es sei $s(m)$ die Inverse von $m(s)$ und $n(s)$ sei durch die Gleichheit:

$$(R^{-1}+s)^{n(s)}(R+s)=1,$$

gegeben also

$$m(s)=\frac{-\log(R^{-1}-s)}{\log(R-s)}; \qquad 0<s<R^{-1}=\sqrt{2}-1$$

(2. 5)

$$n(s)=\frac{\log(R+s)}{-\log(R^{-1}+s)}; \qquad 0<s<2-\sqrt{2}.$$

Die Beziehung (2. 4) ist der Beziehung

(2. 6) $$m(s)<n(s) \quad \text{für} \quad 0<s<s_0=s(m_0)$$

gleichwertig.

Es sei

(2. 7) $$\theta(s) = -\log(R^{-1}+s)\log(R-s)[n(s)-m(s)] =$$

$$= \log(R+s)\log(R-s)-\log(R^{-1}+s)\log(R^{-1}-s).$$

Da

$$\log(1+u)\log(1-u) = -\sum_{i=1}^{\infty}\frac{p_i}{i}u^{2i},$$

wo

$$p_i = 1-\frac{1}{2}+\frac{1}{3}-\ldots+\frac{1}{2i-1}$$

ist, und

(2. 8) $$\log(1+u)+\log(1-u) = -\sum_{i=1}^{\infty}\frac{u^{2i}}{i},$$

so erhalten wir die Reihenentwicklung

(2. 9) $$\theta(s) = \sum_{i=1}^{\infty}\frac{P_i}{i}s^{2i}$$

mit

(2. 10) $$P_i = p_i(R^{2i}-R^{-2i})-(R^{2i}+R^{-2i})\log R \qquad (i=1,2,\ldots).$$

Es gilt auch

$$P_1 = (\sqrt{2}+1)^2-(\sqrt{2}-1)^2-[(\sqrt{2}+1)^2+(\sqrt{2}-1)^2]\log(1+\sqrt{2}) =$$

$$= 6\left[\frac{2\sqrt{2}}{3}-\log(1+\sqrt{2})\right] > 0,$$

da $\log(1+\sqrt{2})=0{,}881\ldots$ und $\frac{2\sqrt{2}}{3}=0{,}942\ldots$ ist.

$$P_i = (R^{2i}+R^{-2i})\left[p_i\frac{R^{2i}-R^{-2i}}{R^{2i}+R^{-2i}}-\log R\right] \leqq$$

da $$\leqq (R^{2i}+R^{-2i})(p_2-\log R)<0 \quad \text{für} \quad i=2,3,\ldots,$$

$$p_{i+1} = p_i-\frac{1}{2i}+\frac{1}{2i+1} = p_i-\frac{1}{2i(2i+1)} < p_i < p_{i-1} < \ldots < p_2 = \frac{5}{6} = 0{,}833\ldots$$

ist, und wir erhalten:

$$P_1 > 0,\ P_i < 0 \qquad i = 2, 3\ldots, .$$

Hieraus folgt aber, daß $\theta(s)$ eine und nur eine Nullstelle im Intervall $(0, \sqrt{2}-1)$ hat. Wir bezeichnen diese Nullstelle durch s_0, und erhalten: $\theta(s) > 0$ im Intervall $(0, s_0)$. Durch numerisches Rechnen finden wir aus (2. 7)

$$0{,}3481 < s_0 < 0{,}3482$$

und aus (2. 5)

$$3{,}7437 < m_0 = m(s_0) < 3{,}7457. \tag{2.11}$$

Aus der Konkavität der Funktion $\log t$ und aus der Ungleichung $\gamma_\alpha(t) \geqq 0$ folgt für beliebige u und v mit $-1 \leqq u \leqq a(\alpha) < b(\alpha) \leqq v \leqq 1$ und für $x \notin (u, v)$:

$$\varphi_\alpha(x) = \alpha \log|x+1| + \beta(\alpha)\log|x-1| + \int_{a(\alpha)}^{b(\alpha)} \log|x-t|\,\gamma_\alpha(t)\,dt >$$

$$> \alpha \log|x+1| + \beta(\alpha)\log|x-1| +$$

$$+ \int_{a(\alpha)}^{b(\alpha)} \left[\frac{v-t}{v-u}\log|x-u| + \frac{t-u}{v-u}\log|x-v|\right]\gamma_\alpha(t)\,dt.$$

Aus (1. 46) und (1. 47) erhalten wir durch Integrieren:

$$\int_{a(\alpha)}^{b(\alpha)} t\gamma_\alpha(t)\,dt = [1-\alpha-\beta(\alpha)][\alpha-\beta(\alpha)],$$

also

$$\varphi_\alpha(x) > \alpha\log|x+1| + \beta(\alpha)\log|x-1| + [1-\alpha-\beta(\alpha)]\frac{v-\alpha+\beta(\alpha)}{v-u}\log|x-u| +$$

$$+ [1-\alpha-\beta(\alpha)]\frac{\alpha-\beta(\alpha)-u}{v-u}\log|x-v|. \tag{2.12}$$

Von hier folgt es im Falle $u = -1$, $v = 1$

$$\varphi_\alpha(x) > \left(\frac{1}{2} + \frac{\alpha^2-\beta^2(\alpha)}{2}\right)\log|x+1| + \left(\frac{1}{2} - \frac{\alpha^2-\beta^2(\alpha)}{2}\right)\log|x-1| \quad \text{für} \quad |x| > 1,$$

und

$$|E(\exp\varphi_\alpha(x))| = \delta(\exp\varphi_\alpha(x)) < \delta\left(|x+1|^{\frac{1}{2}+\frac{\alpha^2-\beta^2(\alpha)}{2}}|x-1|^{\frac{1}{2}-\frac{\alpha^2-\beta^2(\alpha)}{2}}\right). \tag{2.13}$$

Nach (2. 2) gilt die Ungleichung $|E(\exp\varphi_\alpha(x))| < 2\sqrt{2}$, wenn

$$1 < \frac{\frac{1}{2}+\frac{\alpha^2-\beta^2(\alpha)}{2}}{\frac{1}{2}-\frac{\alpha^2-\beta^2(\alpha)}{2}} < m_0$$

ist d.h. wenn

$$0 < \alpha^2 - \beta^2(\alpha) < \frac{m_0 - 1}{m_0 + 1} \tag{2.14}$$

ist. Die Funktion $\alpha^2 - \beta^2(\alpha)$ wächst von 0 bis α_0^2, wenn α von $\frac{1}{2}$ bis α_0 wächst, also ist die Ungelichung (2. 14)

$$|E(\exp \varphi_\alpha(x))| < 2\sqrt{2} \quad \text{für} \quad \tfrac{1}{2} < \alpha < \alpha_1, \tag{2.15}$$

erfüllt, wo für die Zahl α_1 gilt:

$$\alpha_1^2 - \beta^2(\alpha_1) = \frac{m_0 - 1}{m_0 + 1}. \tag{2.16}$$

Es ist offenbar, daß $\alpha_1 < \alpha_0$ ist, da $\alpha_1^2 - \beta^2(\alpha_1) = \frac{m_0 - 1}{m_0 + 1} = 0{,}578\ldots < \alpha_0^2 - \beta^2(\alpha_0) = \alpha_0^2 = 0{,}608\ldots$.

Die übrigbleibenden Fälle $\alpha_1 \leqq \alpha < \alpha_0$ fordern feinere Überlegungen.

Zuerst können wir leicht zeigen:

$$\varphi_\alpha(-1{,}7713) > 0 \quad \text{für} \quad \tfrac{1}{2} \leqq \alpha \leqq \alpha_0. \tag{2.17}$$

Aus der Ungleichung (2.12) mit $u = 0$, $v = 1$ erhalten wir für $x = -1{,}7713$

$$\varphi_\alpha(-1{,}7713) > \alpha \log 0{,}7713 + [(1-\alpha)^2 - \beta^2(\alpha)] \log 1{,}7713 +$$

$$+ (\alpha - \alpha^2 + \beta^2(\alpha)) \log 2{,}7713 > \alpha \log 0{,}7713 + (1-\alpha)^2 \log 1{,}7713 +$$

$$+ (\alpha - \alpha^2) \log 2{,}7713 = \log 1{,}7713 - \alpha \log \frac{(1{,}7713)^2}{0{,}7713 \cdot 2{,}7713} -$$

$$- \alpha^2 \log \frac{2{,}7713}{1{,}7713} > \log 1{,}7713 - 0{,}78 \log \frac{(1{,}7713)^2}{0{,}7713 \cdot 2{,}7713} -$$

$$- 0{,}78^2 \log \frac{2{,}7713}{1{,}7713} > 0,$$

d.h. (2. 17) ist richtig.

Nach (1. 45) ist $\varphi_\alpha(b(\alpha)) = 0$, deshalb

$$\varphi_\alpha(2 - b(\alpha)) = \alpha \log(3 - b(\alpha)) + \beta(\alpha) \log(1 - b(\alpha)) + \int_{a(\alpha)}^{b(\alpha)} \log(2 - b(\alpha) - t)\, \gamma_\alpha(t)\, dt >$$

$$> \alpha \log(b(\alpha) + 1) + \beta(\alpha) \log|b(\alpha) - 1| + \int_{a(\alpha)}^{b(\alpha)} \log(b(\alpha) - t)\, \gamma_\alpha(t)\, dt = \varphi_\alpha(b(\alpha)) = 0,$$

hieraus folgt

$$\varphi_\alpha(2 - b(\alpha)) > 0. \tag{2.18}$$

Aus (2.17) und (2.18) erhalten wir

$$\delta(\exp \varphi_\alpha(x)) < 3{,}7713 - b(\alpha), \tag{2.19}$$

deshalb genügt es nur die Ungleichung

$$3{,}7713 - b(\alpha) < 2\sqrt{2} \quad \text{für} \quad \alpha_1 \leqq \alpha < \alpha_0 \tag{2.20}$$

zu beweisen. Wir definieren den Wert α_2 durch

$$\beta^2(\alpha_2) = \frac{1-0{,}78^2}{2}\left(2\sqrt{2} - 2{,}7713\right). \tag{2.21}$$

Nach (1. 48) gilt

$$\alpha_2 < 0{,}767, \tag{2.22}$$

da die linke Seite von (1. 48) für $\alpha = 0{,}767$ und $\beta = \beta(\alpha_2)$ $(= 0{,}1057 \ldots)$ größer als 4 ist.

Die Ungleichung (2. 20) gilt für $\alpha_2 < \alpha < \alpha_0$, es ist nämlich nach (1. 47):

$$1 - b(\alpha) = \frac{4\beta^2(\alpha)}{1-a(\alpha)}$$

$$a(\alpha_0) = 2\alpha_0^2 - 1,$$

und es gilt

$$a(\alpha) < a(\alpha_0) \quad \text{für} \quad 0{,}75 < \alpha < \alpha_0 \tag{2.23}$$

(auf den Beweis dieser Ungleichung kehren wir bald zurück), hieraus erhalten wir

$$1 - b(\alpha) = \frac{4\beta^2(\alpha)}{1-a(\alpha)} < \frac{4\beta^2(\alpha)}{1-a(\alpha_0)} = \frac{2\beta^2}{1-\alpha_0^2} \leqq \frac{2\beta^2(\alpha_2)}{1-\alpha_0^2} =$$

$$= \frac{1-0{,}78^2}{1-\alpha_0^2}\left(2\sqrt{2} - 2{,}7713\right) < 2\sqrt{2} - 2{,}7713 \quad \text{für} \quad \alpha_2 \leqq \alpha \leqq \alpha_0,$$

also gilt die Ungleichung (2. 20) für $\alpha_2 \leqq \alpha \leqq \alpha_0$.

Es genügt zu zeigen

$$\alpha_2 < \alpha_1. \tag{2.24}$$

In der Tat, die Funktion $\alpha^2 - \beta^2(\alpha)$ ist eine monoton wachsende Funktion von α, und deshalb folgt aus (2.16), (2. 21) und (2. 22)

$$\alpha_2^2 - \beta^2(\alpha_2) < 0{,}767^2 - \beta^2(\alpha_2) < \frac{m_0 - 1}{m_0 + 1} = \alpha_1^2 - \beta^2(\alpha_1),$$

also ist $\alpha_2 < \alpha_1$. Aus (2. 15), (2. 19), (2. 20) bzw. (2. 24) ergibt sich schließlich

$$\left|E\left(\exp \varphi_\alpha(x)\right)\right| = \delta\left(\exp \varphi_\alpha(x)\right) < 2\sqrt{2} \quad \text{für} \quad \tfrac{1}{2} < \alpha \leqq \alpha_0. \tag{2.25}$$

Es bleibt nur der Beweis der Ungleichung (2. 23) übrig. Löst man das Gleichungssystem (1. 47) auf, so ergibt sich für a:

$$a = a(\alpha) = \alpha^2 - \beta^2(\alpha) - \sqrt{[\alpha^2 - \beta^2(\alpha)]^2 - 2[\alpha^2 + \beta^2(\alpha)] + 1}.$$

Die Ungleichung (2. 23) ist folgender gleichwertig:

$$\alpha^2 - \beta^2(\alpha) - 2\alpha_0^2 + 1 < \sqrt{[\alpha^2 - \beta^2(\alpha)]^2 - 2[\alpha^2 + \beta^2(\alpha)] + 1}.$$

Quadriert man die beiden Seiten, so bekommt man die Ungleichung

$$\frac{\alpha^2}{\alpha_0^2}+\frac{\beta^2}{1-\alpha_0^2}<1 \quad \text{für} \quad 0,75<\alpha<\alpha_0 \text{ und } \beta=\beta(\alpha). \tag{2.26}$$

Wir betrachten die Gleichung (1. 48). Durch Logarithmieren erhalten wir

$$\begin{aligned}&(1+\alpha+\beta)\log(1+\alpha+\beta)+(1+\alpha-\beta)\log(1+\alpha-\beta)+\\&+(1-\alpha+\beta)\log(1-\alpha+\beta)+(1-\alpha-\beta)\log(1-\alpha-\beta)=\\&=\log 4=2(1+\alpha_0)\log(1+\alpha_0)+2(1-\alpha_0)\log(1-\alpha_0),\end{aligned}$$

also

$$(1+\alpha)\log\left(1-\frac{\beta^2}{(1+\alpha)^2}\right)+(1-\alpha)\log\left(1-\frac{\beta^2}{(1-\alpha)^2}\right)+\beta\log\frac{1+\frac{\beta}{1+\alpha}}{1-\frac{\beta}{1+\alpha}}+$$

$$+\beta\log\frac{1+\frac{\beta}{1-\alpha}}{1-\frac{\beta}{1-\alpha}}=2(1+\alpha_0)\log(1+\alpha_0)+2(1-\alpha_0)\log(1-\alpha_0)-2(1+\alpha)\log(1+\alpha)-$$

$$-2(1-\alpha)\log(1-\alpha).$$

Die linke Seite ist:

$$2\sum_{i=1}^{\infty}\frac{\beta^{2i}}{2i(2i-1)}\left[\frac{1}{(1+\alpha)^{2i-1}}+\frac{1}{(1-\alpha)^{2i-1}}\right]>2\frac{\beta^2}{1-\alpha^2},$$

und so genügt es die Ungleichung

$$\frac{\alpha^2}{\alpha_0^2}+\frac{1-\alpha^2}{1-\alpha_0^2}[(1+\alpha_0)\log(1+\alpha_0)+(1-\alpha_0)\log(1-\alpha_0)-(1+\alpha)\log(1+\alpha)-$$

$$-(1-\alpha)\log(1-\alpha)]<1$$

zu beweisen, welche dieselbe ist, wie die Ungleichung

$$\begin{aligned}&\frac{(1-\alpha_0^2)^2}{\alpha_0^2}\frac{1}{1-\alpha^2}-(1+\alpha)\log(1+\alpha)-(1-\alpha)\log(1-\alpha)<\\&<\frac{1-\alpha_0^2}{\alpha_0^2}-(1+\alpha_0)\log(1+\alpha_0)-(1-\alpha_0)\log(1-\alpha_0).\end{aligned}$$

Die linke Seite ist eine konvexe Funktion von α im Intervall $[0{,}75, \alpha_0]$, und für $\alpha=\alpha_0$ gilt das Zeichen $=$, und für $\alpha=0{,}75$ gilt das Zeichen $<$, also gilt die Ungleichung (2. 26) bzw. (2. 23).

Aus den Ungleichungen (1. 49) und (2. 25) ergibt sich für gerade n: $E_n=2\sqrt{2}$, und das einzige extremale Polynom ist $(x^2-1)^{n/2}$; für ungerade n ist $E_n<E_{2n}$ nach einem Resultat in [2], d.h. es ist $E_n<2\sqrt{2}$, hieraus folgt aber die Richtigkeit der Vermutung von ERDŐS.

3. Schlussbemerkungen

In diesem Teile befassen wir uns mit einigen Problemen, welche mit der Vermutung von ERDŐS verwandt sind.

Betrachten wir die folgende Aufgabe: Bezeichnet man durch $\mathscr{F}_r$ die Menge der Polynome von der Gestalt $f(x)=\prod_{i=1}^{n}(x-x_i)$ mit $-r\leqq x_i\leqq r$ für $i=1, 2, \ldots, n$, wo n eine beliebige natürliche Zahl und r eine nichtnegative reelle Zahl ist. Es sei

$$e(r)=\sup_{f(x)\in\mathscr{F}_r}|E(f)|.$$

Aus [6] ist uns bekannt: $e(r)\leqq 4$ für alle $r\geqq 0$ und zwar $e(r)=4$ für $r\geqq 2$. Aus [1] ist es bekannt, dass $e(r)=2\sqrt{1+r^2}$ für $0\leqq r\leqq\frac{3}{4}$ ist. Nach der soeben bewiesenen Vermutung von ERDŐS gilt es: $e(1)=2\sqrt{2}$. Im Lichte dieser Tatsache können wir auch die Vermutung aussprechen: $e(r)=2\sqrt{1+r^2}$ für $\frac{3}{4}\leqq r\leqq 1$. Ein ganz offenes Problem ist aber der Fall $1<r<2$. Die obigen Fälle können wir in der Formel zusammenfassen:

$$e(r)=\begin{cases}2\sqrt{1+r^2} & 0\leqq r\leqq\frac{3}{4}\\ 2\sqrt{1+r^2} \quad (?) & \frac{3}{4}<r<1\\ 2\sqrt{2} & r=1\\ ? & 1<r<2\\ 4 & r\geqq 2.\end{cases}$$

Wir können jedoch eine andere Größe untersuchen. Nämlich, es sei

$$\varepsilon(r)=\inf_{f(x)\in\mathscr{F}_r}|E(f)|.$$

Im folgenden werden wir die Beziehungen

$$\varepsilon(r)=\begin{cases}2 & 0\leqq r\leqq\frac{\sqrt{2}}{2}\\ <2 & \frac{\sqrt{2}}{2}<r<2\\ 0 & r\geqq 2\end{cases}\tag{3.1}$$

betrachten. Die Gleichheit $\varepsilon(r)=0$ für $r\geqq 2$ ist schon bekannt (s. in [1], Seite 132). Die Ungleichung $\varepsilon(r)<2$ für $r>\frac{\sqrt{2}}{2}$ stammt von ERDŐS, er hat mich darauf aufmerksam gemacht, daß der Wert $r=\frac{\sqrt{2}}{2}$ interessant in Hinsicht auf $\varepsilon(r)$ sein könnte, da es ihm gelingt die Ungleichung $\varepsilon(r)<2$ für $r>\frac{\sqrt{2}}{2}$ zu beweisen.

Zuerst werden wir den Fall $r \leqq \frac{\sqrt{2}}{2}$ betrachten. Es sei $f(x)$ ein Element aus $\mathcal{F}_r$: $f(x) = \prod_{i=1}^{n}(x - x_i)$ mit $-r \leqq x_i \leqq r$. Wir definieren die Funktion $F(x)$ und die Größe σ:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n}|x - x_i|,$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Wir können annehmen, daß $\sigma \leqq 0$ ist. Die Funktion $F(x)$ ist konvex, und es gilt: $F(0) < 1$, $F(-r) = r + \sigma \leqq r < 1$, also es gilt: $F(x) < 1$ für $-r \leqq x \leqq 0$. Wenn die Menge $E(f)$ zusammenhängend ist, so ist $|E(f)| \geqq 2$, denn der Punkt $x = \sigma - 1$ gehört zu $E(f)$, da $\sigma - 1 < -r$ ist, und

$$|f(\sigma - 1)| \leqq |F(\sigma - 1)|^n = 1,$$

dies gilt auch für $x = \sigma + 1$, und falls $\sigma + 1$ größer als jeder x ist, so ist

$$|f(\sigma + 1)| \leqq |F(\sigma + 1)|^n = 1,$$

wenn es aber eine Wurzel x_j mit $x_j \geqq \sigma + 1$ gibt, dann gehört x_j selbst zu $E(f)$, also dann erhalten wir in beiden Fällen: $[\sigma - 1, \sigma + 1] \subset E(f)$ bzw. $[\sigma - 1, \sigma + 1] \subseteqq [\sigma - 1, x_j] \subset E(f)$, deshalb ist $|E(f)| \geqq 2$.

Im Falle $0 \leqq r \leqq \frac{1}{2}$ gilt: $|f(x)| \leqq 1$ im Intervall $[-r, r]$, also ist $E(f)$ zusammenhängend, folglich $|E(f)| \geqq 2$.

Für das Weitere beschränken wir uns nur auf den Fall $\frac{1}{2} < r \leqq \frac{\sqrt{2}}{2}$, und wir können annehmen, dass die Menge $E(f)$ aus mehreren disjunkten Teilen besteht. Wir bezeichnen nun durch $E^*(f)$ den Teil von $E(f)$, welcher den Punkt $x = 0$ enthält. Nach obigen Überlegungen gilt: $[-r, 0] \subset E^*(f)$. Nach der Voraussetzung $\sigma \leqq 0$ gibt es Wurzeln, welche in $E^*(f)$ fallen, und sie seien: $x_1, x_2, \ldots, x_k$. Die Wurzeln $x_{k+1}, \ldots, x_n$ liegen also ausserhalb $E^*(f)$. Die Wurzeln seien der Größe nach geordnet: $-r \leqq x_1 \leqq x_2 \leqq x_3 \leqq \ldots \leqq x_n \leqq r$. Es gilt nun die Ungleichung $k > \frac{n}{2}$, denn $F'(x) = \frac{2i - n}{n}$ im Intervall (x_i, x_{i+1}) und die Funktion $F(x)$ nimmt ihr Minimum im Intervall $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$ für gerade n, bzw. im Punkte $x = x_{\frac{n+1}{2}}$ für ungerade n an, hieraus ergibt sich sofort, daß die Wurzel $x_{\frac{n}{2}+1}$ bzw. $x_{\frac{n+1}{2}}$ zu $E^*(f)$ gehört, also $k \geqq \min\left\{\frac{n}{2} + 1, \frac{n+1}{2}\right\} > \frac{n}{2}$. Es gilt auch: $k < n$, da $E(f)$ im Falle $k = n$ zusammenhängend wäre.

Wenn wir die Wurzeln $x_{k+1}, \ldots, x_n$ nach $x = r$ bewegen lassen, so erhalten wir das Polynom

$$f_1(x) = (x - r)^{n-k}\prod_{i=1}^{k}(x - x_i).$$

Es ist offenbar $E^*(f_1)\subseteqq E^*(f)$. Es sei $\sigma_1=\frac{1}{n}\left(\sum_{i=1}^{k}x_i+(n-k)r\right)$ und

$$f_2(x) = (x-\sigma_1)^k(x-r)^{n-k}.$$

Es gilt $|f_1(x)|\leqq|f_2(x)|$ für $x\leqq x_1$ bzw. $x\geqq x_k$, deshalb ist $E^*(f_2)\subseteqq E^*(f_1)$. Da $-r<\sigma_1<r$ ist, darum ergibt sich wieder für $f_3(x) = (x+r)^k(x-r)^{n-k}$ die Relation $E^*(f_3)\subset E^*(f_2)$.

Es sei $2r=a$, also ist $1<a\leqq\sqrt{2}$. Nun genügt es für die Menge $E(|x|^m|x-a|)$ mit $m=\frac{k}{n-k}>1$ zu zeigen, daß die Länge der Komponente, welche den Punkt $x=0$ enthält, mindestens den Wert 2 hat. Es seien die Abscissen der Endpunkte dieser Komponente durch $-1+s(m)$ bzw. $1+t(m)$ bezeichnet, dann gilt es:

$$(3.2)\qquad (1-s)^m(a+1-s) = 1 \qquad 0<s<1$$

$$(1+t)^m(a-1-t) = 1 \qquad 0<t<a-1.$$

Man muss die Richtigkeit der Ungleichung $t(m)>s(m)$ zeigen, welche äquivalent der Ungleichung $m(t)>n(t)$ ist, wo

$$m(t) = \frac{-\log(a-1-t)}{\log(1+t)}; \qquad 0<t<a-1$$

$$n(t) = \frac{\log(a+1-t)}{-\log(1-t)}; \qquad 0<t<a-1$$

ist. Es sei

$$\Theta(t) = -\log(1-t)\log(1+t)[m(t)-n(t)] =$$

$$=\log(1-t)\log(a-1-t)-\log(1+t)\log(a+1-t).$$

Die Potenzreihen-entwicklung von $\Theta(t)$ nach t ist:

$$\Theta(t) = \log\frac{1}{a^2-1}\sum_{i=1}^{\infty}\frac{t^{2i-1}}{2i-1}+\log\frac{a+1}{a-1}\sum_{i=1}^{\infty}\frac{t^{2i}}{2i}+\sum_{i=2}^{\infty}(B_i-C_i)t^i,$$

wo

$$B_i = \sum_{j=1}^{i-1}\frac{1}{j(i-j)(a-1)^j} \qquad i=2,3,\ldots$$

$$C_i = \sum_{j=1}^{i-1}\frac{(-1)^{i-j}}{j(i-j)(a+1)^j} \qquad i=2,3,\ldots.$$

Es ist offenbar, daß $B_i-C_i>0$ für $i=2,3,\ldots$ und $a^2-1\leqq 1$ für $1<a\leqq\sqrt{2}$ ist, d.h. die Koeffizienten sind in der Potenzreihenentwicklung von $\Theta(t)$ nicht negativ, es ist $\Theta(t)>0$ für $0<t<a-1$, womit die Behauptung: $t(m)>s(m)$ bewiesen ist, woraus Folgt: $|E^*(|x|^m|x-a|)| = 2+t(m)-s(m)>2$.

Vollständigkeitshalber werden wir den Beweis der Ungleichung $\varepsilon(r)<2$ für $r<\frac{\sqrt{2}}{2}$ skizzieren. In diesem Falle ist $a = 2r>\sqrt{2}$. Für große m sei $E(|x^m|\,|x-a|) =$

$= [-1+s(m),\ 1+t(m)] \cup [a-\delta_1(m),\ a+\delta_2(m)]$, hieraus folgt

$$s(m) = \frac{\log(a+1)+o(1)}{m}$$

$$t(m) = \frac{-\log(a-1)+o(1)}{m}$$

$$\delta_1(m) = \frac{1+o(1)}{a^m}$$

$$\delta_2(m) = \frac{1+o(1)}{a^m},$$

wo $o(1)$ eine von m abhängige Größe mit $\lim_{m\to\infty} o(1)=0$ bedeutet. Nun ergibt sich $|E(|x|^m|x-a|)| = 2+t(m)-s(m)+\delta_1(m)+\delta_2(m) = 2-\frac{\log(a^2-1)+o(1)}{m}$ also es gilt für eine genügend große natürliche Zahl m im Falle $a = 2r > \sqrt{2}$: $|E(x^m(x-a))| < 2$, womit die Relationen (3.1) bewiesen sind.

LITERATURVERZEICHNIS

[1] ERDŐS, P., HERZOG, F. and PIRANIAN, G.: Metric Properties of Polynomials, *J. d'Analyse Math.* **6** (1958) 125—148.
[2] ELBERT, Á.: Über eine Vermutung von Erdős betreffs Polynome, I, *Studia Sci. Math. Hungar.* **1** (1966) 119—128.
[3] MARKOFF, W.: Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen, *Math. Ann.* **77** (1916) 213—258.
[4] BRAY, H. E.: Elementary properties of the Stieltjes Integral, *Ann. Math.* **29** (1919).
[5] CARLEMAN, T.: Über die Abelsche Integralgleichung mit konstanten Integrationsgrenzen, *Math. Z.* **15** (1922) 111—120.
[6] PÓLYA, G.: Beitrag zur Verallgemeinerung des Verzerrungssatzes auf mehrfach zusammenhängende Gebiete, *Sitzunsgsberichte d. Akad. Wiss. Berlin* (1928) 280—282.

Mathematisches Institut der Ungarischen Akademie der Wissenschaften, Budapest

(Eingegangen: 18. Januar, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 325—336.

ABSCHÄTZUNGEN VOM MENCHOFF-RADEMACHERSCHEN TYP FÜR DIE SUMMEN VON ORTHOGONALEN FUNKTIONEN

von

K. TANDORI

1. Einleitung

Für eine Folge $\{a_n\}_1^N$ setzen wir

$$I(a_1, \ldots, a_N) = \sup \int_0^1 \left(\max_{1 \leqq i \leqq j \leqq N} |a_i \varphi_i(x) + \ldots + a_j \varphi_j(x)| \right)^2 dx,$$

wobei das Supremum über alle im Grundintervall [0, 1] gebildeten orthonormierten Funktionensysteme $\{\varphi_n(x)\}_1^N$ zu nehmen ist.

Die Größe $I(a_1, \ldots, a_N)$ spielt in der Konvergenztheorie der Orthogonalreihen eine entscheidende Rolle. (Siehe z. B. K. TANDORI [4], [5].) Es ist ein Grundproblem, wie die Funktion $I(a_1, \ldots, a_N)$ von den Variablen $a_1, \ldots, a_N$ abhängt. Dieses Problem ist sehr schwer und noch ungelöst. Es gibt nur verschiedene Abschätzungen für $I(a_1, \ldots, a_N)$.

Im Punkt 2 dieser Note werden wir die bisherigen Abschätzungen und ihre Verhältnisse zueinander besprechen. Im Punkt 3 werden wir mit einer kleinen Modifizierung des Grundgedanken von D. E. MENCHOFF [1] und H. RADEMACHER [3] eine Vorschrift für die Abschätzung von $I(a_1, \ldots, a_N)$ angeben. Im Punkt 4 werden wir zeigen, daß die so erhaltene Abschätzung besser als alle bisherigen ist. Es ist wahrscheinlich, daß die so erhaltene Abschätzung schon genau ist; diese Vermutung zu beweisen scheint aber ein schweres Problem zu sein.

2. Verschiedene bekannte Abschätzungen für $I(a_1, \ldots, a_N)$

Es seien $F(a_1, \ldots, a_N)$, $G(a_1, \ldots, a_N)$ gegebene Funktionen und nehmen wir an, daß die Abschätzungen

$$I(a_1, \ldots, a_N) \leqq F(a_1, \ldots, a_N), \quad I(a_1, \ldots, a_N) \leqq G(a_1, \ldots, a_N)$$

für jede Folge $\{a_n\}_1^N$ bestehen. Die erste Abschätzung werden wir besser als die zweite nennen, wenn $F(a_1, \ldots, a_N) \leqq c_1 G(a_1, \ldots, a_N)$ für jede Folge $\{a_n\}_1^N$ gilt. (Im folgenden bezeichnen $c_1, c_2, \ldots$ positive, von N und a_n unabhängige Konstanten.) Diese zwei Abschätzungen sind äquivalent, wenn auch $G(a_1, \ldots, a_N) \leqq c_2 F(a_1, \ldots, a_N)$ für jede Folge $\{a_n\}_1^N$ besteht. Die Abschätzung $I(a_1, \ldots, a_N) \leqq F(a_1, \ldots, a_N)$ gültig für jede Folge $\{a_n\}_1^N$ nennen wir für eine Klasse K der Folgen $\{a_n\}_1^N$ genau, wenn auch $F(a_1, \ldots, a_N) \leqq c_3 I(a_1, \ldots, a_N)$ $(\{a_n\}_1^N \in K)$ gilt. Endlich nennen wir diese Abschätzung genau schlechthin, wenn die letzte Ungleichung für jede Folge $\{a_n\}_1^N$ besteht.

a) Die erste feinere Abschätzung für $I(a_1, \ldots, a_N)$ stammt von D. E. MENCHOFF [1] und H. RADEMACHER [2]. Sie haben

$$I(a_1, \ldots, a_N) \leqq c_4(1+\log N)^2 \sum_{n=1}^{N} a_n^2 \tag{1}$$

bewiesen; D. E. MENCHOFF [1] hat auch gezeigt, dass diese Abschätzung im Falle $a_n = a$ $(n=1, \ldots, N)$ genau ist.

b) Aus (1) ergibt sich durch einfache Rechnung

$$I_1(a_1, \ldots, a_N) \leqq c_5 \left(a_1^2 + \sum_{n=2}^{N} a_n^2 \log^2 n \right); \tag{2}$$

im Falle $a_n^2 \geqq a_{n+1}^2$ $(n=1, \ldots, N-1)$ ist diese Abschätzung genau. (Siehe A. ZYGMUND [8], Vol, II., S. 193.) Offensichtlich ist (2) besser als (1).

c) Neuerdings hat K. TANDORI [6] mit einer Modifizierung des Grundgedanken von D. E. MENCHOFF und H. RADEMACHER

$$I(a_1, \ldots, a_N) \leqq c_6(\log N + 1) \sum_{n=1}^{N} a_n^2 \log_+ \frac{a_1^2 + \ldots + a_N^2}{a_n^2} \tag{3}$$

gezeigt, wobei $\log_+ \dfrac{a_1^2 + \ldots + a_N^2}{a_n^2} = \log \dfrac{a_1^2 + \ldots + a_N^2}{a_n^2}$ für $\dfrac{a_1^2 + \ldots + a_N^2}{a_n^2} \geqq 2$,

und $\log_+ \dfrac{a_1^2 + \ldots + a_N^2}{a_n^2} = 1$ für $\dfrac{a_1^2 + \ldots + a_N^2}{a_n^2} < 2$ oder für $a_n = 0$

ist. (Im folgenden bedeutet $\log x$ den Logarithmus von x mit der Basis 2.)

d) Durch einfache Rechnung kann auch die bessere Abschätzung

$$I(a_1, \ldots, a_N) \leqq c_7 \left(a_1^2 + \sum_{n=2}^{N} \log n \cdot a_n^2 \log_+ \frac{a_1^2 + \ldots + a_N^2}{a_n^2} \right) \tag{4}$$

gezeigt werden.

Zum Beweis von (4) können wir ohne Beschränkung der Allgemeinheit $N = 2^\nu$ annehmen. Es sei $\{\varphi_n(x)\}_1^{2^\nu}$ ein orthonormiertes System in [0, 1]. Mit der Bezeichnung $s_n(x) = a_1\varphi_1(x) + \ldots + a_n\varphi_n(x)$ gilt

$$s_n^2(x) \leqq 3\left((s_n(x) - s_{2^m}(x))^2 + (s_{2^m}(x) - s_{2^\nu}(x))^2 + s_{2^\nu}^2(x) \right)$$

für $2^m < n \leqq 2^{m+1}$, und somit ist

$$\max_{1 \leqq n \leqq 2^\nu} s_n^2(x) \leqq 3 \left(s_1^2(x) + \sum_{m=0}^{\nu-1} \max_{2^m < n \leqq 2^{m+1}} (s_n(x) - s_{2^m}(x))^2 + \right.$$

$$\left. + \sum_{m=0}^{\nu-1} (s_{2^m}(x) - s_{2^\nu}(x))^2 + s_{2^\nu}^2(x) \right).$$

Daraus ergibt sich durch Anwendung von (3)

$$I(a_1, \ldots, a_N) \leqq 6\left(a_1^2 + c_6 \sum_{m=0}^{\nu-1} (1+\log 2^m) \sum_{n=2^m+1}^{2^{m+1}} a_n^2 \log_+ \frac{a_1^2+\ldots+a_N^2}{a_n^2} + \right.$$

$$\left. + \sum_{m=1}^{\nu-1} (a_{2^m+1}^2 + \ldots + a_{2^\nu}^2) + (a_1^2 + \ldots + a_{2^\nu}^2)\right),$$

woraus man (4) leicht bekommt.

Es kann gezeigt werden, daß die Abschätzung (4) besser als (2) ist. Zum Beweis können wir $a_1^2+\ldots+a_N^2=1$ ohne Beschränkung der Allgemeinheit annehmen. Wir bezeichnen mit I' bzw. mit I'' die Menge der Indizes $n\,(2\leqq n\leqq N)$, für die $0<a_n^2<1/n^4$ bzw. $a_n^2\geqq 1/n^4$ besteht. Dann gilt

$$\sum_{n=2}^{N} \log n \cdot a_n^2 \log_+ \frac{1}{a_n^2} = \sum_{I'} + \sum_{I''} \leqq$$

$$\leqq 2 \sum_{I'} \log n \cdot \frac{1}{n^2}\left(|a_n| \log \frac{1}{|a_n|}\right) + \sum_{I''} \log n \cdot a_n^2 \log n^4 \leqq$$

$$\leqq c_8 \sum_{n=2}^{\infty} \frac{\log n}{n^2} + c_9 \sum_{n=2}^{N} a_n^2 \log^2 n \leqq c_{10} \sum_{n=2}^{N} a_n^2 \log^2 n \qquad (c_{10}\geqq 1).$$

Daraus folgt

$$a_1^2 + \sum_{n=2}^{N} \log n \cdot a_n^2 \log_+ \frac{a_1^2+\ldots+a_N^2}{a_n^2} \leqq c_{10}\left(a_1^2 + \sum_{n=2}^{N} a_n^2 \log^2 n\right).$$

Also ist die Abschätzung (4) besser als (2). Man kann leicht einsehen, daß (2) und (4) im Falle $a_n^2 \geqq a_{n+1}^2$ $(n=1, \ldots, N-1)$ äquivalent sind.

e) Für $I(a_1, \ldots, a_N)$ gilt weiterhin die triviale Abschätzung

$$I(a_1, \ldots a_N) \leqq \left(\sum_{n=1}^{N} |a_n|\right)^2. \tag{5}$$

f) Durch Modifizierung des Grundgedanken von W. ORLICZ [2] hat K. TANDORI [7] noch eine Abschätzung für $I(a_1, \ldots, a_N)$ bewiesen. Zur Abschätzung von $I(a_1, \ldots, a_N)$ können wir $N=2^{2^\nu}$ $(\nu\geqq 0)$ ohne Beschränkung der Allgemeinheit voraussetzen. Es sei $\{a_{n_i}\}$ $(i=1, \ldots, 2^{2^\nu})$ die Anordnung der Folge $\{a_n\}_1^{2^{2^\nu}}$ mit $|a_{n_1}|\geqq|a_{n_2}|\geqq\ldots$. (Ist $|a_{n_i}|=\ldots=|a_{n_{i+j}}|$, dann sei $n_i<\ldots<n_{i+j}$.) Mit dieser Bezeichnung gilt

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq c_{11}\left(|a_{n_1}| + |a_{n_2}| + \sum_{m=0}^{\nu-1} \sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l}\right)^2. \tag{6}$$

(Wir bemerken, dass auch

$$\max I(a_1, \ldots, a_N) \leqq \left(\sum_{n=1}^{N} |a_n|\right)^2,$$

$$\max I(a_1, \ldots, a_N) \leqq c_{11}\left(|a_{n_1}| + |a_{n_2}| + \sum_{m=0}^{\nu-1} \sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l}\right.$$

bestehen, wobei das Maximum für alle Permutationen der Folge $\{a_n\}_1^{2^{2^\nu}}$ gebildet ist.)

Man kann zeigen, daß die Abschätzung (6) besser als (5) ist. Zum Beweis nehmen wir $N=2^{2^\nu}$ und $a_1^2+\ldots+a_{2^{2^\nu}}^2=1$ an. Dann gilt $|a_{n_l}|\leqq 1/\sqrt{l}$ $(l=1,\ldots,2^{2^\nu})$, und so ist

$$\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2\log^2 l \leqq 4|a_{2^{2^m}}|^2\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}}|a_{n_l}| \leqq 4\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}}|a_{n_l}|.$$

Daraus folgt

$$|a_{n_1}|+|a_{n_2}|+\sum_{m=0}^{\nu-1}\sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2\log^2 l} \leqq c_{12}\sum_{n=1}^{2^{2^\nu}}|a_n|$$

für jede Folge $\{a_n\}_1^{2^{2^\nu}}$. D.h. (6) ist besser als (5).

Wir bemerken noch, daß die Abschätzungen (4) und (6) unvergleichbar sind. Ist z.B. $a_n=1/\sqrt{n\log^3(n+1)(\log\log(n+2))^2}$ $(n=1,\ldots,2^{2^\nu})$, dann gilt

$$a_1^2+\sum_{n=2}^{2^{2^\nu}}\log n\cdot a_n^2\log_+\frac{a_1^2+\ldots+a_{2^{2^\nu}}^2}{a_n^2} \leqq c_{13}\left(1+\sum_{n=2}^{\infty}\frac{1}{n\log n(\log\log(n+1))^2}\right)\leqq c_{14}$$

$(\nu\geqq 0)$, jedoch

$$|a_{n_1}|+|a_{n_2}|+\sum_{m=0}^{\nu-1}\sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2\log^2 l} \geqq c_{14}\sum_{m=1}^{\nu-1}\frac{1}{m}\to\infty \qquad (\nu\to\infty).$$

Also ist die Abschätzung (6) nicht für jede Folge $\{a_n\}$ besser als (4).

Es sei nun $a_n=1/2^{2^\nu}$ $(n=1,\ldots,2^{2^\nu}-1)$ und $a_{2^{2^\nu}}=1$. Dann gelten

$$a_1^2+\sum_{n=2}^{2^{2^\nu}}\log n\cdot a_n^2\log_+\frac{a_1^2+\ldots+a_{2^{2^\nu}}^2}{a_n^2}\geqq 2^\nu\to\infty \qquad (\nu\to\infty),$$

$$|a_{n_1}|+|a_{n_2}|+\sum_{m=0}^{\nu-1}\sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2\log^2 l} \leqq c_{12}\sum_{n=1}^{2^{2^\nu}}|a_n|\leqq 2c_{12}<\infty.$$

Also ist die Abschätzung (4) nicht für jede Folge $\{a_n\}$ besser als (6).

Unter den Abschätzungen (1), (2), (3), (4), (5) und (6) sind also (4) und (6) die besten.

3. Eine neue Abschätzung für $I(a_1,\ldots,a_N)$

Zur Abschätzung von $I(a_1,\ldots,a_N)$ können wir $N=2^{2^\nu}$ $(\nu\geqq 0)$ und $a_1^2+\ldots+a_{2^{2^\nu}}^2=1$ annehmen. Für die Folge $\{a_n\}_1^{2^{2^\nu}}$ setzen wir die Anordnung $\{a_{n_i}\}_1^{2^{2^\nu}}$ definiert wie im Punkt 2. Es sei

$$I_m=\{n_i: 2^{2^m}<i\leqq 2^{2^{m+1}}\} \qquad (m=0,\ldots,\nu-1).$$

Die Elemente von I_m in wachsender Anordnung bezeichnen wir mit $n_i(m)$ $(i=1,\ldots,2^{2^{m+1}}-2^{2^m}=\nu(m))$; $n_1(m)<n_2(m)<\ldots$. Es ist zweckmäßig $a_{n_{\nu(m)+j}}=0$ $(j=1,\ldots,2^{2^m})$ zu setzen.

Es sei $\{\varphi_n(x)\}_1^{2^{2^\nu}}$ ein orthonormiertes System im Grundintervall $[0, 1]$. Wir setzen $\varphi_{n_{\nu(m)}+j}(x) \equiv 0$ für $j = 1, \ldots, 2^{2^m}$. Die Klasse der Summen

$$\sum_{i=1}^{2^{2^{m+1}}} a_{n_i(m)} \varphi_{n_i(m)}(x),$$

$$\sum_{i=1}^{2^{2^{m+1}-1}} a_{n_i(m)} \varphi_{n_i(m)}(x), \qquad \sum_{i=2^{2^{m+1}-1}+1}^{2^{2^{m+1}}} a_{n_i(m)} \varphi_{n_i(m)}(x),$$

$$\sum_{i=j2^{2^{m+1}-2}+1}^{(j+1)2^{2^{m+1}-2}} a_{n_i(m)} \varphi_{n_i(m)}(x) \qquad (j = 0, \ldots, 3),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_{n_1(m)}\varphi_{n_1(m)}(x), \quad a_{n_2(m)}\varphi_{n_2(m)}(x), \ldots, a_{n_{2^{2^{m+1}}}(m)}\varphi_{n_{2^{2^{m+1}}}(m)}(x)$$

bezeichnen wir mit $\Sigma(m)$. Die Summen $\sigma \equiv 0$ sollen aus $\Sigma(m)$ weggelassen werden; weiterhin im Falle $\sigma_1 \equiv \ldots \equiv \sigma_\mu$ sollen $\sigma_1, \ldots, \sigma_\mu$ als dasselbe Element von $\Sigma(m)$ betrachtet werden. Σ bezeichnet die Klasse der Summen eingliedrigen $a_{n_1}\varphi_{n_1}(x)$, $a_{n_2}\varphi_{n_2}$ und (x)aller Summen σ aus $\Sigma(m)$ $(m = 0, \ldots, \nu - 1)$.

Für jedes n $(1 \leqq n \leqq 2^{2^\nu})$ existieren Indizes $r(n)$, $r_m(n)$ $(m = 0, \ldots, \nu - 1)$ $(0 \leqq r(n) \leqq 2;\ 0 \leqq r_m(n) \leqq 2^{2^{m+1}})$ mit

$$\begin{aligned} s_n(x) &= a_1\varphi_1(x) + \ldots + a_n\varphi_n(x) = \\ &= \sum_{i=1}^{r(n)} a_{n_i}\varphi_{n_i}(x) + \sum_{m=0}^{\nu-1}\sum_{i=1}^{r_m(n)} a_{n_i(m)}\varphi_{n_i(m)}(x). \end{aligned} \tag{7}$$

(Ist $r(n) = 0$, bzw. $r_m(n) = 0$, dann soll man unter der entsprechenden Summe $\sum_{i=1}^{r(n)} a_{n_i}\varphi_{n_i}(x)$ bzw. $\sum_{i=1}^{r_m(n)} a_{n_i(m)}\varphi_{n_i(m)}(x)$ 0 verstehen.) Jede Summe $\sum_{i=1}^{r_m(n)} a_{n_i(m)}\varphi_{n_i(m)}(x)$ können wir in der Form

$$\sum_{i=1}^{r_m(n)} a_{n_i(m)}\varphi_{n_i(m)}(x) = \sum_{l=1}^{\varrho(n,m)} \sigma_l(n, m) \qquad (m = 0, \ldots, \nu - 1) \tag{8}$$

schreiben, wobei alle $\sigma_l(n, m)$ zu $\Sigma(m)$ gehören, und $\sigma_k(n, m)$ $\sigma_l(n, m)$ im Falle $k \neq l$ kein gemeinsames Glied haben. Auf Grund von (7) und (8) kann man für $s_n(x)$ ein System $\mathcal{F}_n$ der Summen σ aus Σ mit

$$s_n(x) = \sum_{\sigma \in \mathcal{F}_n} \sigma \tag{9}$$

angeben, wobei die verschiedenen σ aus $\mathcal{F}_n$ kein gemeinsames Glied haben. Eine solche Zerlegung nennen wir zulässig. (Wir bemerken, daß es für ein vorgegebenes n mehrere zulässige Zerlegungen gibt.)

Für eine zulässige Zerlegung $\mathcal{F}_n$ seien $\delta_\sigma(\mathcal{F}_n)$ $(\sigma \in \mathcal{F}_n)$ positive Zahlen mit

$$\sum_{\sigma \in \mathcal{F}_n} 1/\delta_\sigma^2(\mathcal{F}_n) \leqq 1. \tag{10}$$

Das System von den Zahlen $\delta_\sigma(\mathcal{F}_n)$ $(\sigma \in \mathcal{F}_n)$ bezeichnen wir mit $D(\mathcal{F}_n)$.

Aus (9) und (10) erhalten wir durch Anwendung der Cauchyschen Ungleichung

$$s_n^2(x) \leqq \sum_{\sigma \in \mathscr{F}_n} \sigma^2 \delta_\sigma^2(\mathscr{F}_n). \tag{11}$$

Wir bilden für jedes n eine zulässige Zerlegung $\mathscr{F}_n$ und für die Zerlegung $\mathscr{F}_n$ nehmen wir ein System $D(\mathscr{F}_n)$ mit (10). Ist $\sigma \notin \mathscr{F}_n$ für ein $\sigma \in \Sigma$, dann setzen wir $\delta_\sigma(\mathscr{F}_n) = 1$. Für jede Summe $\sigma \in \Sigma$ bilden wir

$$d_\sigma = d_\sigma(\mathscr{F}_1, \ldots, \mathscr{F}_{2^{2^\nu}}, D(\mathscr{F}_1), \ldots, D(\mathscr{F}_{2^{2^\nu}})) = \max(\delta_\sigma(\mathscr{F}_1), \ldots, \delta_\sigma(\mathscr{F}_{2^{2^\nu}})).$$

Aus (11) folgt

$$\max_{1 \leqq n \leqq 2^{2^\nu}} s_n^2(x) \leqq \sum_{\sigma \in \Sigma} d_\sigma^2 \sigma^2.$$

Daraus ergibt sich durch Benützung der Orthonormalität der Funktionen $\varphi_n(x)$

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq 2 \sum_{n=1}^{2^{2^\nu}} \vartheta_n a_n^2 \tag{12}$$

mit

$$\vartheta_n = \sum_\sigma{}' d_\sigma^2 \qquad (n = 1, \ldots, 2^{2^\nu}), \tag{13}$$

wobei Σ' bedeutet, daß es nur für solche $\sigma (\in \Sigma)$ zu summieren ist, die $a_n \varphi_n(x)$ als Glied besitzen.

Die Faktoren ϑ_n hängen von den Zerlegungen $\mathscr{F}_n$ $(n = 1, \ldots, 2^{2^\nu})$ und von den Systemen $D(\mathscr{F}_n)$ $(n = 1, \ldots, 2^{2^\nu})$ ab. Wir setzen

$$F(a_1, \ldots, a_{2^{2^\nu}}) = \inf 2 \sum_{n=1}^{2^{2^\nu}} \vartheta_n a_n^2,$$

wobei das Infimum für alle Systeme von zulässigen Zerlegungen $\mathscr{F}_n$ $(n = 1, \ldots, 2^{2^\nu})$ und für alle Systeme von $D(\mathscr{F}_n)$ $(n = 1, \ldots, 2^{2^\nu})$ mit (10) gebildet ist.

Wir werden zeigen, dass

$$F(a_1, \ldots, a_{2^{2^\nu}}) = \sum_{n=1}^{2^{2^\nu}} \theta_n a_n^2 \tag{14}$$

mit gewissen θ_n $(\theta_n \geqq 1;\ n = 1, \ldots, 2^{2^\nu})$ gilt. Auf Grund der Definition von $F(a_1, \ldots, a_{2^{2^\nu}})$ gibt es eine Folge von Systemen $\mathscr{F}_1(r), \ldots, \mathscr{F}_{2^{2^\nu}}(r)$ $(r = 1, 2, \ldots)$ und eine Folge von Systemen $D(\mathscr{F}_1(r)), \ldots, D(\mathscr{F}_{2^{2^\nu}}(r))$ $(r = 1, 2, \ldots)$ (wobei (10) für jedes n und r erfüllt wird) derart, daß mit den entsprechenden $\vartheta_n(r)$

$$2 \sum_{n=1}^{2^{2^\nu}} \vartheta_n(r) a_n^2 \to F(a_1, \ldots, a_{2^{2^\nu}}) \qquad (r \to \infty)$$

besteht. Offensichtlich können wir annehmen, daß für jedes n mit $a_n \neq 0$ die Folge $\{\vartheta_n(r)\}_1^\infty$ beschränkt is. Durch 2^{2^ν}-malige Anwendung des Satzes von Bolzano—Weierstrass können wir eine Teilfolge $\{r_i\}$ der natürlichen Zahlen auswählen, für

die $2\,\vartheta_n(r_i) \to \theta_n$ ist $(n=1, \ldots, 2^{2^\nu};\ i \to \infty)$. Nach obigen ergibt sich (14), da die Summe $\sum_{n=1}^{2^{2^\nu}} \vartheta_n a_n^2$ von den Faktoren ϑ_n stetig abhängt. Aus (12) folgt dann die Abschätzung

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq \sum_{i=1}^{2^{2^\nu}} \theta_n a_n^2. \tag{15}$$

Die Faktoren θ_n sind durch die Folge $\{a_n\}_1^{2^{2^\nu}}$ bestimmt. Sie hängen davon ab, in welchem Maß die angegebene Anordnung $a_1, \ldots, a_{2^{2^\nu}}$ von der dem absoluten Wert nach monoton abnehmenden Anordnung $a_{n_1}, \ldots, a_{n_{2^{2^\nu}}}$ verschieden ist; für die Folgen $\{a_n\}_1^{2^{2^\nu}}$ und $\{a \cdot a_n\}_1^{2^{2^\nu}}$ sind $(a \neq 0)$ die Faktoren θ_n dieselben. Das Problem, die θ_n als explizite Funktionen der Folge $\{a_n\}$ anzugeben, scheint sehr schwer zu sein.

Da $d_\sigma \geqq 1$ definitionsgemäss gilt, folgt aus (13)

$$\theta_n = \bar{\theta}_n 2^m \qquad (\bar{\theta}_n \geqq 1;\ n \in I_m;\ m=0, \ldots, \nu-1). \tag{16}$$

Da nach der Annahme $a_1^2 + \ldots + a_{2^{2^\nu}}^2 = 1$ und der Definition von I_m die Abschätzungen

$$a_n^2 \leqq a_{n_1(m)}^2 \ (n \in I_m), \quad 2^{2m} a_{n_1(m)}^2 \leqq 1$$

bestehen, gilt

$$2^m \leqq \log \frac{1}{a_n^2} \qquad (n \in I_m).$$

So ergibt sich aus (15) und (16)

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq \sum_{i=1}^{2^{2^\nu}} \bar{\theta}_n a_n^2 \log_+ \frac{1}{a_n^2} \qquad (\bar{\theta}_n \geqq 1;\ n=1, \ldots, 2^{2^\nu}). \tag{17}$$

Wahrscheinlich sind die Abschätzungen (15) und (17) äquivalent, und (15) ist schon genau. Diese Probleme scheinen jedoch recht schwer zu sein.

4. Verhältnisse der verschiedenen Abschätzungen

Wir werden zeigen, daß die Abschätzung (15) besser als (4) und (6) ist. Zum Beweis können wir $N = 2^{2^\nu}$ und $a_1^2 + \ldots + a_{2^{2^\nu}}^2 = 1$ annehmen. Im folgenden gebrauchen wir die Bezeichnungen des vorigen Punktes.

a) Wir beschäftigen uns erstens mit dem Verhältnis der Abschätzungen (4) und (15).

Für jedes m $(0 \leqq m \leqq \nu-1)$ und s $(0 \leqq s \leqq \nu-1)$ sei

$$I(m, s) = \{n : n \in I_m;\ 2^{2^s} < n \leqq 2^{2^{s+1}}\}.$$

Die Elemente von $I(m, s)$ bezeichnen wir in wachsender Anordnung mit $n_i(m, s)$ $(i=1, \ldots, R(m, s))$.

Ist $I(m, s)$ nicht leer, dann können wir für die Summe $\sum_{n \in I_m(s)} a_n \varphi_n(x)$ ein System $\mathscr{F}(m, s)$ von σ aus $\Sigma(m)$ eindeutig mit folgenden Bedingungen angeben: Es gilt

$$\sum_{n \in I(m, s)} a_n \varphi_n(x) = \sum_{\sigma \in \mathscr{F}(m, s)} \sigma, \tag{18}$$

wo die verschiedenen σ aus $\mathcal{F}(m, s)$ kein gemeinsames Glied haben, und die Anzahl der Elemente von $\mathcal{F}(m, s)$ minimal ist. Wenn $\overline{\overline{\mathcal{F}(m, s)}}$ die Mächtigkeit von $\mathcal{F}(m, s)$ bezeichnet, dann gilt offensichtlich

$$\overline{\overline{\mathcal{F}(m, s)}} \leqq 2 \cdot 2^s. \tag{19}$$

Wir bezeichnen mit $\mathcal{F}$ die Vereinigung der Systeme $\mathcal{F}(m, s)$ $(m, s = 0, \ldots, \nu - 1)$. Offensichtlich gehört $a_{n_i}\varphi_{n_i}(x)$ $(i \geqq 3)$ nur zu einer einzigen Summe σ aus $\mathcal{F}$.

Ist $I(m, s)$ nicht leer und $1 \leqq \varrho < R(m, s)$, dann können wir ein System $\mathcal{F}(m, s, \varrho)$ der Summen aus $\Sigma(m)$ mit folgenden Bedingungen eindeutig angeben: Es gilt

$$\sum_{i=1}^{\varrho} a_{n_i(m,s)} \varphi_{n_i(m,s)}(x) = \sum_{\sigma \in \mathcal{F}(m, s, \varrho)} \sigma, \tag{20}$$

wo die verschiedenen σ aus $\mathcal{F}(m, s, \varrho)$ kein gemeinsames Glied haben, und die Anzahl von σ aus $\mathcal{F}(m, s, \varrho)$ minimal ist. Offensichtlich gilt

$$\overline{\overline{\mathcal{F}(m, s, \varrho)}} \leqq 2 \cdot 2^s \qquad (\varrho = 1, \ldots, R(m, s) - 1). \tag{21}$$

Es sei $2^{2^{m_0}} < n \leqq 2^{2^{m_0+1}}$ $(0 \leqq m_0 \leqq \nu - 1)$. Dann gibt es Indizes $r(n)$, $\varrho(m, n)$ $(m = 0, \ldots, \nu - 1)$ $(0 \leqq r(n) \leqq 2,\ 0 \leqq \varrho(m, n) \leqq R(m, m_0) - 1)$ mit

$$\begin{aligned} s_n(x) = &\sum_{i=1}^{r(n)} a_{n_i}\varphi_{n_i}(x) + \sum_{m=0}^{\nu-1} \sum_{s=0}^{m_0-1} \sum_{n \in I(m,s)} a_n \varphi_n(x) + \\ &+ \sum_{m=0}^{\nu-1} \sum_{i=1}^{\varrho(m,n)} a_{n_i(m,m_0)} \varphi_{n_i(m,m_0)}(x). \end{aligned} \tag{22}$$

Es sei $\mathcal{F}_n^*$ die Vereinigung von $\mathcal{F}(m, s)$ $(m = 0, \ldots, \nu - 1;\ s = 0, \ldots, m_0 - 1)$, von $\mathcal{F}(m, m_0, \varrho(m, n))$ $(m = 0, \ldots, \nu - 1)$ und von dem System bestehend aus den eingliedrigen Summen $a_{n_i}\varphi_{n_i}(x)$ $(i = 1, 2)$. Dann folgt

$$s_n(x) = \sum_{\sigma \in \mathcal{F}_n^*} \sigma \tag{23}$$

aus (18), (20) und (22). Diese Zerlegung ist offensichtlich zulässig. Wir definieren die positiven Faktoren $\delta_\sigma(\mathcal{F}_n^*)$ für $\sigma \in \Sigma$ folgenderweise: Es sei

$$\delta_\sigma^2(\mathcal{F}_n^*) = 3$$

für $\sigma = a_{n_1}\varphi_{n_1}(x)$, bzw. $\sigma = a_{n_2}\varphi_{n_2}(x)$,

$$\delta_\sigma^2(\mathcal{F}_n^*) = 9 \cdot 2^{m_0+1} \cdot 2^r$$

für $\sigma \in \mathcal{F}(m_0 + r, m_0, \varrho(m_0 + r, n))$ $(r = 0, \ldots, \nu - 1 - m_0)$ oder für

$$\sigma \in \mathcal{F}(m_0 - r, m_0, \varrho(m_0 - r, n)) \quad (r = 1, \ldots, m_0),$$

ferner

$$\delta_\sigma^2(\mathcal{F}_n^*) = 12 \cdot 2^{m+1} 2^m \cdot 2^r$$

für $\sigma \in \mathcal{F}(m, \mu)$ $(m = 0, \ldots, \nu - 1;\ \mu = 0, \ldots, \nu - 1;\ \mu \neq m_0;\ \mu = m \pm r;\ r \geqq 1)$. Dann gilt

$$\sum_{\sigma \in \mathcal{F}_n^*} 1/\delta_\sigma^2(\mathcal{F}_n^*) \leqq 1 \tag{24}$$

auf Grund von (19) und (21).

Bilden wir das System $\mathcal{F}_n^*$ mit obiger Methode für jedes n $(1 \leqq n \leqq 2^{2^\nu})$, und wählen für jedes $\mathcal{F}_n^*$ ein System $D(\mathcal{F}_n^*)$ von positiven Zahlen $\delta_\sigma(\mathcal{F}_n^*)$ $(\sigma \in \mathcal{F}_n^*)$ wie oben. Dann gilt (24) für jedes n. Ist für ein $\sigma \in \Sigma$ $\delta_\sigma(\mathcal{F}_n^*)$ nicht definiert, dann setzen wir $\delta_\sigma(\mathcal{F}_n^*) = 1$. Es sei weiterhin

$$\bar{d}_\sigma = \bar{d}_\sigma\left(D(\mathcal{F}_1^*), \ldots, D(\mathcal{F}_{2^{2^\nu}}^*)\right) = \max\left(\delta_\sigma(\mathcal{F}_1^*), \ldots, \delta_\sigma(\mathcal{F}_{2^{2^\nu}}^*)\right),$$

für jedes $\sigma \in \Sigma$. Es sei endlich

$$\bar{\vartheta}_n = \sum_{\sigma}{}' \bar{d}_\sigma^2, \tag{25}$$

wobei Σ' bedeutet, daß man nur für solche $\sigma (\in \Sigma)$ zu summieren hat, die $a_n \varphi_n(x)$ als Glied besitzen. Auf Grund von (24) ergibt sich die Abschätzung

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq 2 \sum_{n=1}^{2^{2^\nu}} \bar{\vartheta}_n a_n^2.$$

Nun werden wir die Faktoren $\bar{\vartheta}_n$ abschätzen. Nach der Definition von $\delta_\sigma(\mathcal{F}_n^*)$, $\bar{d}_\sigma$ und $\bar{\vartheta}_n$ gilt

$$\bar{\vartheta}_n \leqq c_{13} \max\left(1, \log n \log \frac{1}{a_n^2}\right) \quad (\text{für } n = n_1, \quad \text{oder} \quad n = n_2). \tag{26}$$

Es sei $n \in I(m, m - r)$ $(r = 0, \ldots, m)$, oder $n \in I(m, m + r)$ $(r = 1, \ldots, \nu - 1 - m;$ $m = 0, \ldots, \nu - 1)$. Die Summen σ in (25) bezeichnen wir mit $\sigma_1, \ldots, \sigma_{p(n)}$. Laut Definition von $\mathcal{F}_n^*$ gehören diese σ_i zu $\Sigma(m)$ und es gilt $p(n) \leqq 2^{m+1}$. Nach der Definition von $\mathcal{F}$ gibt es unter σ_i nur eine einzige Summe σ_{i_0}, die $a_n \varphi_n(x)$ als Glied besitzt. Nach Definition von $\delta_\sigma(\mathcal{F}_n^*)$ und $\bar{d}_\sigma$ gilt $\bar{d}_{\sigma_{i_0}} = 12 \cdot 2^{2m+1} \cdot 2^r$ und $\bar{d}_{\sigma_i} = 9 \cdot 2^{m+1} \cdot 2^r$ für $i \neq i_0$. Daraus folgt

$$\bar{\vartheta}_n \leqq c_{14} 2^{2m} \cdot 2^r. \tag{27}$$

Aus (25), (26) und (27) erhalten wir

$$\begin{aligned} 2 \sum_{n=1}^{2^{2^\nu}} \bar{\vartheta}_n a_n^2 \leqq c_{15} \Bigg(& a_{n_1}^2 \max\left(1, \log n_1 \log \frac{1}{a_{n_1}^2}\right) + a_{n_2}^2 \max\left(1, \log^2 n_2\right) + \\ & + \sum_{m=1}^{\nu-1} \sum_{r=1}^{m} \sum_{n \in I(m, m-r)} a_n^2 2^{2m} \cdot 2^r + \sum_{m=0}^{\nu-1} \sum_{r=0}^{\nu-1-m} \sum_{n \in I(m, m+r)} a_n^2 \cdot 2^{2m} \cdot 2^r \Bigg). \end{aligned} \tag{28}$$

Auf Grund der Definition von $I(m, s)$ gilt

$$2^{2m} \cdot 2^r < \log n_2 \log \frac{1}{a_{n_2}^2} \quad (n \in I(m, m + r)),$$

und somit ist

$$\sum_{m=0}^{\nu-1} \sum_{r=0}^{\nu-1-m} \sum_{n \in I(m, m+r)} a_n^2 \cdot 2^{2m} \cdot 2^r \leqq \sum_{n=2}^{2^{2^\nu}} a_n^2 \log n \log \frac{1}{a_n^2}. \tag{29}$$

Für ein $m(0<m\leqq v-1)$ gilt weiterhin

$$\sum_{r=1}^{m}\sum_{n\in I(m,\,m-r)}a_n^2\,2^{2m}\cdot 2^r\leqq\sum_{r=1}^{m}(2^{2(m-r)+1}-2^{2m-r})\max_{n\in I(m,\,m-r)}a_n^2\,2^{2m}\cdot 2^r\leqq$$

$$\leqq\sum_{r=1}^{m}(2^{2(m-r)+1}-2^{2m-r})\min_{n\in J(m-1)}a_n^2\cdot 2^{2m}\cdot 2^r\leqq$$

$$\leqq\frac{2^{2m}-2^{2m-1}}{2}\min_{n\in J(m-1)}a_n^2\cdot 2^{2m}\cdot 2\sum_{r=1}^{m}\frac{2^{2m-r+1}-2^{2m-r}}{2^{2m}-2^{2m-1}}2^r\leqq$$

$$\leqq c_{16}2^{2m}\frac{2^{2m}-2^{2m-1}}{2}\min_{n\in J(m-1)}a_n^2,$$

wobei $J(m-1)$ die Vereinigung von $I(m-1,s)$ $(s=m-1,\ldots,v-1)$ ist. Offensichtlich ist die Mächtigkeit von $J(m-1)$ größer als $(2^{2m}-2^{2m-1})/2$ und es gilt $2^m\leqq 2\log n$ für $n\in J(m-1)$. Sodann ist

$$\sum_{r=1}^{m}\sum_{n\in I(m,\,m-r)}a_n^2\,2^{2m}\cdot 2^r\leqq c_{17}\sum_{n\in I(m-1)}a_n^2\log n\log\frac{1}{a_n^2}\qquad(m=1,\ldots,v-1).\tag{30}$$

Aus (28), (29), und (30) folgt

$$2\sum_{n=1}^{2^{2^v}}\bar{\vartheta}_n a_n^2\leqq c_{18}\left(a_1^2+\sum_{n=2}^{2^{2^v}}a_n^2\log n\log\frac{1}{a_n^2}\right).\tag{31}$$

Da die durch $\mathscr{F}_n^*$ definierten Zerlegungen zulässig sind, und für die Systeme $D(\mathscr{F}_n^*)$ (24) erfüllt ist, erhalten wir auf Grund der Definition von θ_n und (31)

$$2\sum_{n=1}^{2^{2^v}}\theta_n a_n^2\leqq c_{18}\left(a_1^2+\sum_{n=2}^{2^{2^v}}a_n^2\log n\log\frac{a_1^2+\ldots+a_{2^{2^v}}^2}{a_n^2}\right)$$

für jede Folge $\{a_n\}_1^{2^{2^v}}$.

Die Abschätzung (15) ist also besser als (4).

b) Endlich werden wir zeigen, daß die Abschätzung (15) besser als die Abschätzung (6) ist.

Für jedes $n(1\leqq n\leqq 2^{2^v})$ können wir

$$s_n(x)=\sum_{i=1}^{r(n)}a_{n_i}\varphi_{n_i}(x)+\sum_{m=0}^{v-1}\sum_{i=1}^{r_m(n)}a_{n_i(m)}\varphi_{n_i(m)}(x)\tag{32}$$

schreiben, wobei $0\leqq r(n)\leqq 2$, $0\leqq r_m(n)\leqq 2^{2m+1}$ $(m=0,\ldots,v-1)$ sind. Für jedes m können wir ein System $\mathscr{F}^*(n,m)$ von $\sigma\big(\in\Sigma(m)\big)$ eindeutig mit folgenden Eigenschaften angeben: Es gilt

$$\sum_{i=1}^{r_m(n)}a_{n_i(m)}\varphi_{n_i(m)}(x)=\sum_{\sigma\in\mathscr{F}^*(n,\,m)}\sigma,\tag{33}$$

wo die verschiedenen σ aus $\mathscr{F}^*(n, m)$ kein gemeinsames Glied haben und die Anzahl von σ aus $\mathscr{F}^*(n, m)$ minimal ist. Offensichtlich gilt

$$\overline{\overline{\mathscr{F}^*(n, m)}} \leqq 2 \cdot 2^m. \tag{34}$$

Es sei $\tilde{\mathscr{F}}_n$ die Vereinigung der Systeme $\mathscr{F}^*(n, m)$ $(m = 0, \ldots, v-1)$ und des Systems bestehend aus den eingliedrigen Summen $a_{n_i}\varphi_{n_i}(x)$ $(1 \leqq i \leqq r(n))$. Dann gilt

$$s_n(x) = \sum_{\sigma \in \tilde{\mathscr{F}}_n} \sigma, \tag{35}$$

und für jedes n ist dies eine zulässige Zerlegung von $s_n(x)$.

Es sei ϱ_i $(i = -2, -1, 0, \ldots, v-1)$ eine Folge von positiven Zahlen mit

$$\sum_{i=-2}^{v-1} \frac{1}{\varrho_i^2} \leqq 1. \tag{36}$$

Wir setzen unabhängig von n

$$\tilde{\delta}_\sigma^2 = \varrho_{-2}^2 \text{ bzw. } \tilde{\delta}_\sigma^2 = \varrho_{-1}^2 \text{ (für } \sigma = a_{n_1}\varphi_{n_1}(x) \text{ bzw. } \sigma = a_{n_2}\varphi_{n_2}(x)),$$
$$\tilde{\delta}_\sigma^2 = \varrho_m^2 2^{m+1} \text{ (für } \sigma \in \sum(m); \quad m = 0, \ldots, v-1).$$

Dann besteht auf Grund von (34) und (36)

$$\sum_{\sigma \in \tilde{\mathscr{F}}_n} 1/\tilde{\delta}_\sigma^2 \leqq 1 \tag{37}$$

für jedes n. Auf Grund von (37) gilt

$$\max_{1 \leqq n \leqq 2^{2^v}} s_n^2(x) \leqq \sum_{\sigma \in \mathscr{F}} \tilde{\delta}_\sigma^2 \sigma^2,$$

und somit ist

$$I(a_1, \ldots, a_{2^{2^v}}) \leqq 2 \sum_{n=1}^{2^{2^v}} \tilde{\vartheta}_n a_n^2, \tag{38}$$

wobei

$$\tilde{\vartheta}_n = \sum_{\sigma}{}' \tilde{\delta}_\sigma^2$$

ist, und Σ' bedeutet, dass man nur für solche $\sigma (\in \Sigma)$ zu summieren hat, die $a_n\varphi_n(x)$ als Glied besitzen. Die Anzahl solcher σ ist 2^{m+1} für $n \in I(m)$, und daher folgt

$$\tilde{\vartheta}_n = \varrho_m^2 2^{2(m+1)} \qquad (n \in I(m); \quad m = 0, \ldots, v-1)$$

auf Grund der Definition von $\tilde{\delta}_\sigma$. Aus (38) erhalten wir

$$I(a_1, \ldots, a_{2^{2^v}}) \leqq c_{19}\left(\varrho_{-2}^2 a_{n_1}^2 + \varrho_{-1}^2 a_{n_2}^2 + \sum_{m=0}^{v-1} \varrho_m^2 \sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l\right). \tag{39}$$

Durch einfache Rechnung ergibt sich

$$\inf\left(\varrho_{-2}^2 a_{n_1}^2 + \varrho_{-1}^2 a_{n_2}^2 + \sum_{m=0}^{v-1} \varrho_m^2 \sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l\right) \leqq$$
$$\leqq \left(|a_{n_1}| + |a_{n_2}| + \sum_{m=0}^{v-1} \sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l}\right)^2,$$

wobei das Infimum für alle Folgen $\{\varrho_m\}$ mit der Nebenbedingung (36) gebildet ist. Aus (39) folgt endlich die Abschätzung

$$I(a_1, \ldots, a_{2^{2^\nu}}) \leqq c_{19}\left(|a_{n_1}| + |a_{n_2}| + \sum_{m=0}^{\nu-1} \sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l}\right)^2.$$

Da die durch $\tilde{\mathcal{F}}_n$ definierte Zerlegungen zulässig sind und die Ungleichung (37) im Falle (36) für jedes n besteht, ergibt sich auf Grund der Definition von θ_n

$$2\sum_{n=1}^{2^{2^\nu}} \theta_n a_n^2 \leqq c_{20}\left(|a_{n_1}| + |a_{n_2}| + \sum_{m=0}^{\nu-1} \sqrt{\sum_{l=2^{2^m}+1}^{2^{2^{m+1}}} a_{n_l}^2 \log^2 l}\right)^2.$$

Die Abschätzung (15) ist also besser als (6).

LITERATURVERZEICHNIS

[1] MENCHOFF, D. E.: Sur les séries de fonctions orthogonales, I, *Fund. Math.* **4** (1923) 82—105.
[2] ORLICZ, W.: Zur Theorie der Orthogonalreihen, *Bulletin Intern. Acad. Sci. Polonaise Cracovie*, (1927), 81—115.
[3] RADEMACHER, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen, *Math. Ann.* **87** (1922) 112—138.
[4] TANDORI, K.: Über die Konvergenz der Orthogonalreihen, *Acta Sci. Math. (Szeged)* **24** (1963) 139—151.
[5] TANDORI, K.: Über die Konvergenz der Orthogonalreihen, II, *Acta Sci. Math. (Szeged)* **25** (1964) 219—232.
[6] TANDORI, K.: Bemerkung zur Konvergenz der Orthogonalreihen, *Acta Sci. Math. (Szeged)* **26** (1965) 249—251.
[7] TANDORI, K.: Über die orthogonalen Funktionen. X (Unbedingte Konvergenz), *Acta Sci. Math. (Szeged)* **23** (1962) 185—221.
[8] ZYGMUND, A.: *Trigonometric Series*, Cambridge, 1959.

Bolyai Institut, Szeged

(Eingegangen: 19. Juni, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 337—343.

ON THE ORDER OF CONVERGENCE OF FINITE-DIFFERENCE APPROXIMATIONS TO THE SOLUTION OF THE DIRICHLET PROBLEM IN A DOMAIN WITH CORNERS

by

L. VEIDINGER

The order of convergence of finite-difference approximations to the solutions of the Dirichlet problems for the two-dimensional Laplace and Poisson equations in a domain with corners has been investigated by many authors (see, for example, [1]—[5]). In this paper we shall discuss the order of convergence of a simple finite-difference approximation to the solution of the Dirichlet problem for the general second-order self-adjoint elliptic differential equation with two independent variables.

1. Let R be a bounded open plane region whose boundary C consists of a finite number of piecewise-analytic simple closed curves. Denote by $A_i\,(i=1, 2, \ldots, n)$ the corners of C, i.e. those points on C where distinct analytic curves meet.

We consider the boundary value problem

$$Lu(x, y)=g(x, y), \qquad (x, y)\in R, \tag{1}$$

$$u(x, y)=\varphi(x, y), \qquad (x, y)\in C,$$

where

$$Lu=\frac{\partial}{\partial x}\left[a(x, y)\frac{\partial u}{\partial x}\right]+\frac{\partial}{\partial x}\left[b(x, y)\frac{\partial u}{\partial y}\right]+\frac{\partial}{\partial y}\left[b(x, y)\frac{\partial u}{\partial x}\right]+$$

$$+\frac{\partial}{\partial y}\left[c(x, y)\frac{\partial u}{\partial y}\right]-f(x, y)u.$$

Let the coefficients $a(x, y)$, $b(x, y)$, $c(x, y)$, $f(x, y)$ and the right-hand side $g(x, y)$ be analytic in an open region G containing the closure of R in its interior. Let $\varphi(x, y)$ be continuous on C and analytic on every analytic portion of C. Suppose that at all points of R

$$a\xi^2+2b\,\xi\eta+c\eta^2\geqq\alpha(\xi^2+\eta^2) \qquad (\alpha=\text{const}>0)$$

for all real ξ, η. Moreover, we assume that $f(x, y)\geqq 0$.

Suppose the infinite plane of the region R is subdivided by two families of parallel lines into a square net. Let the lines of the net be $x=mh$ and $y=nh$ $(m, n=0, \pm 1, \pm 2, \ldots)$. The points (mh, nh) will be called the nodes of the net. The smallest squares bounded by four lines of the net are called meshes of the net. Denote by S_h the set of all nodes of the plane.

Let R^* be the union of all meshes contained in R and let C^* be the boundary of R^*. Let R_h^* consist of all the interior nodes of R^* and let C_h^* be the net boundary of R_h^*. Let $\overline{R}_h^*=R_h^*\cup C_h^*$. The points of intersection of the net lines with the boundary C form the set C_h.

We define

$$V_x(P) = h^{-1}[V(E)-V(P)],\ V_{\bar{x}}(P) = h^{-1}[V(P)-V(W)],$$
$$V_y(P) = h^{-1}[V(N)-V(P)],\ V_{\bar{y}}(P) = h^{-1}[V(P)-V(S)],$$

where $V=V(P)$ is any real-valued function defined on $\bar{R}_h^*$, $E=(x_P+h, y_P)$, $N=(x_P, y_P+h)$, $W=(x_P-h, y_P)$, $S=(x_P, y_P-h)$ are the four neighbors of the node $P=(x_P, y_P)$.

The solution u of the problem (1) is approximated by the solution U of the finite-difference problem (see [6], p. 139 and [11])

$$L_hU(P)=g(P), \quad P\in R_h^*, \tag{2}$$
$$U(P)=\varphi(P'), \quad P\in C_h^*,$$

where

$$L_hU = 0{,}5[(aU_x)_{\bar{x}}+(aU_{\bar{x}})_x+(bU_y)_{\bar{x}}+(bU_{\bar{y}})_x+(bU_x)_{\bar{y}}+(bU_{\bar{x}})_y+(cU_{\bar{y}})_y+(cU_y)_{\bar{y}}]-fU$$

and P' is the nearest point of C_h.

Consider for any function V defined on $\bar{R}_h^*$ the quadratic form

$$H_h(V) = 0{,}5h^2 \sum_{P\in\bar{R}_h^*}\{a(P)[(V_x(P))^2+(V_{\bar{x}}(P))^2]+$$
$$+2b(P)[V_x(P)V_y(P)+V_{\bar{x}}(P)V_{\bar{y}}(P)]+c(P)[(V_y(P))^2+(V_{\bar{y}}(P))^2]+2f(P)[V(P)]^2\}.$$

Here we have put

$$\begin{aligned} &V_x(P)=0, \quad \text{if} \quad P\in C_h^*,\ E\notin\bar{R}_h^*, \quad V_{\bar{x}}(P)=0, \quad \text{if} \quad P\in C_h^*,\ W\notin\bar{R}_h^*,\\ &V_y(P)=0, \quad \text{if} \quad P\in C_h^*,\ N\notin\bar{R}_h^*, \quad V_{\bar{y}}(P)=0, \quad \text{if} \quad P\in C_h^*,\ S\notin\bar{R}_h^*, \end{aligned} \tag{3}$$

We define

$$\|\delta V\| = \Big\{h^2 \sum_{P\in\bar{R}_h^*}[(V_x(P))^2+(V_y(P))^2]\Big\}^{\frac{1}{2}},$$

where the conventions (3) are valid. It is easy to see that

$$\|\delta V\|^2 = O(H_h(V)). \tag{4}$$

Let $W=W(P)$ be any function defined at the nodes which vanishes outside $\bar{R}_h^*$. Then we define

$$\|\delta W\|_1 = \Big\{h^2 \sum_{P\in S_h}[(W_x(P))^2+(W_y(P))^2]\Big\}^{\frac{1}{2}}$$

2. LEMMA 1. *u is an analytic function of x as well as of y in R.*
For a PROOF see [7], p. 179.

LEMMA 2. *u is analytic on C, excluding the corners.*
For a PROOF of a more general result see [8].

LEMMA 3. Let $A_i=(x_{A_i}, y_{A_i})$ be a corner of C, with interior angle $\pi\alpha_i$ $(0<\alpha_i<2)$. Let

$$x^* = k_{A_i}x+l_{A_i}y, \quad y^* = m_{A_i}x+n_{A_i}y \tag{5}$$

be the linear transformation which transforms the operator L into the normal form at the point A_i. Let $r_{A_i}=[(x-x_{A_i})^2+(y-y_{A_i})^2]^{\frac{1}{2}}$. If the transformation (5) transforms the angle $\pi\alpha_i$ into an angle $\pi\alpha_i^*$ $(0<\alpha_i^*<2)$ then for $\alpha_i^* \neq \frac{1}{m}$ (m an integer)

$$u(x,y) = u_1(x,y) + O\left(r_{A_i}^{\frac{1}{\alpha_i^*}}\right). \tag{6}$$

where $u_1(x, y)$ and its partial derivatives of all orders remain bounded when $(x, y) \to A_i$ in R while

$$u(x,y) = u_1(x,y) + O\left(r_{A_i}^{\frac{1}{\alpha_i^*}} |\log r_{A_i}|\right) \tag{7}$$

when $\alpha_i^* = \frac{1}{m}$ $(m=1, 2, \ldots)$. These relations may be indefinitely formally differentiated.

This lemma follows from the results of V. A. KONDRAT'EV (see [9], [10]).

LEMMA 4. Let W be any function defined at the nodes which vanishes outside $\bar{R}_h^*$. Then for h sufficiently small

$$\max_{P \in R_h^*} |W(P)| < c_1 |\log h|^{\frac{1}{2}} \|\delta W\|_1, \tag{8}$$

where c_1 is a positive constant depending only on the region R.

For a PROOF see [12], p. 239.

3. THEOREM 1. *For h sufficiently small the truncation error $z(P) = u(P) - U(P)$ satisfies the inequality*

$$\max_{P \in R_h^*} |z(P)| < c_2 h^{\frac{1}{2}} |\log h|^{\frac{1}{2}}, \tag{9}$$

where c_2 is a positive constant independent of h.

PROOF. The truncation error $z(P)$ is the solution of the problem

$$L_h z(P) = \Phi(P), \quad P \in R_h^*,$$

$$z(P) = u(P) - u(P'), \quad P \in C_h^*,$$

where $\Phi = L_h u - Lu$.

It is easy to see that $z = v + w$, where v is the solution of the problem

$$L_h v(P) = 0, \quad P \in R_h^*,$$

$$v(P) = u(P) - u(P'), \quad P \in C_h^*$$

and w is the solution of the problem

$$L_h w(P) = \Phi(P), \quad P \in R_h^*,$$

$$w(P) = 0, \quad P \in C_h^*.$$

From the finite-difference analogue of Dirichlet's principle (see [13]) it follows that

$$H_h(v) \leqq H_h(\bar{v}), \tag{10}$$

where

$$\bar{v}(P) = \begin{cases} u(P)-u(P'), & P \in C_h^*, \\ 0, & P \in R_h^*. \end{cases}$$

By definition of $\bar{v}$

$$H_h(\bar{v}) = O\left(\sum_{P \in C_h^*} [u(P)-u(P')]^2\right). \tag{11}$$

Let A_i be a corner of C with interior angle $\pi\alpha_i$ $(0<\alpha_i<2)$ and let $P \in C_h^*$. Denote by $r(P, A_i)$ the distance from the point P to the point A_i. If r_1 is a sufficiently small positive real number and $3h<r(P, A_i)<r_1$, then, using (6) and (7) respectively, we get

$$u(P)-u(P') = (x_P-x_{P'})u_x(P'')+(y_P-y_{P'})u_y(P'') = O\left([r(P, A_i)]^{\frac{1}{\alpha_i^*}-1-\varepsilon}h+h\right). \tag{12}$$

where P'' is a point in the interval PP' and ε is any positive real number. From (12) it follows that

$$\sum_{P \in C_h^*,\ 3h<r(P, A_i)<r_1} [u(P)-u(P')]^2 = O\left(\sum_{P \in C_h^*,\ 3h<r(P, A_i)<r_1} [r(P, A_i)]^{\frac{2}{\alpha_i^*}-2-2\varepsilon}h^2+h\right) =$$

$$= O\left(h^{\frac{2}{\alpha_i^*}-2\varepsilon} \sum_{1 \leqq n \leqq \frac{r_1}{h}} \frac{1}{n^{2-\frac{2}{\alpha_i^*}+2\varepsilon}}+h\right) = O(h) \tag{13}$$

On the other hand, using (6) and (7) we have

$$\sum_{P \in C_h^*,\ r(P, A_i)<3h} [u(P)-u(P')]^2 = O\left(h^{\frac{2}{\alpha_i^*}-2\varepsilon}+h^2\right) = O(h). \tag{14}$$

Summing (13) and (14) over all corners of C and applying Lemmas 1 and 2 we obtain

$$\sum_{P \in C_h^*} [u(P)-u(P')]^2 = O(h). \tag{15}$$

Combining (4), (10), (11) and (15) we have

$$\|\delta v\| = O(h^{\frac{1}{2}}). \tag{16}$$

Let

$$\overline{V}(P) = \begin{cases} v(P), & P \in \overline{R}_h^*, \\ 0, & P \notin \overline{R}_h^*. \end{cases}$$

Then by (15) we get

$$\|\delta \overline{V}\|_1^2 = \|\delta v\|^2 + O(h).$$

Hence, using (16) and (8) we obtain

$$\|\delta \overline{V}\|_1^2 = O(h)$$

and

$$\max_{P \in \overline{R}_h^*} |v(P)| = O(h^{\frac{1}{2}}|\log h|^{\frac{1}{2}}). \tag{17}$$

From the finite-difference analogue of GREEN's first identity (see [13]) it follows that

$$H_h(w) = -h^2 \sum_{P \in R_h^*} w(P) L_h w(P) = -h^2 \sum_{P \in R_h^*} w(P) \Phi(P).$$

Hence we get

$$H_h(w) \leqq S \max_{P \in R_h^*} |w(P)| \leqq \frac{1}{2\alpha} \left(\max_{P \in R_h^*} |w(P)| \right)^2 + \frac{\alpha}{2} S^2 \tag{18}$$

where

$$S = \sum_{P \in R_h^*} h^2 |\Phi(P)|$$

and α is any positive real number. On the other hand, from (4) and (8) it follows that

$$H_h(w) \geqq c_3 \left(\max_{P \in R_h^*} |w(P)| \right)^2 |\log h|. \tag{19}$$

Let $\alpha = c_4 |\log h|$, where $c_4 > \frac{1}{2c_3}$. Then from (18) and (19) we obtain

$$\max_{P \in R_h^*} |w(P)| < c_5 |\log h| S. \tag{20}$$

Our next aim is to estimate the sum S. Using Taylor's theorem it is easy to show that

$$|\Phi(P)| = O\left(h^2 \max_{1 \leqq q \leqq 4,\ (x,y) \in \Sigma} \left| \frac{\partial^q u(x,y)}{\partial x^\varrho \, \partial y^\sigma} \right| \right), \quad \varrho + \sigma = q, \tag{21}$$

where Σ is the closed square defined by the inequalities $x_P - h \leqq x \leqq x_P + h$, $y_P - h \leqq y \leqq y_P + h$.

Let r_2 be a sufficiently small positive real number. Then, using (6), (7) and (21) we get

$$\sum_{P \in R_h^*,\ 3h < r(P, A_i) < r_2} h^2 |\Phi(P)| = O\left(\sum_{P \in R_h^*,\ 3h < r(P, A_i) < r_2} [r(P, A_i)]^{\frac{1}{\alpha_i^*} - 4 - \varepsilon} h^4 + h^2 \right) =$$

$$= O\left(h^{\frac{1}{\alpha_i^*} - \varepsilon} \sum_{1 \leqq m^2 + n^2 \leqq \left(\frac{r_2}{h}\right)^2} \frac{1}{(m^2 + n^2)^{2 - \frac{1}{2\alpha_i^*} + \frac{\varepsilon}{2}}} + h^2 \right)$$

and, consequently

$$\sum_{P \in R_h^*,\ 3h < r(P, A_i) < r_2} h^2 |\Phi(P)| = \begin{cases} O(h^2), & \text{if } \alpha_i^* < \frac{1}{2}, \\ O\left(h^{\frac{1}{\alpha_i^*} - \varepsilon} \right), & \text{if } \alpha_i^* \geqq \frac{1}{2}. \end{cases} \tag{22}$$

On the other hand, by (6) and (7) we have

$$\sum_{P \in R_h^*,\ r(P, A_i) < 3h} h^2 |\Phi(P)| = \begin{cases} O(h^2), & \text{if } \alpha_i^* < \frac{1}{2}, \\ O\left(h^{\frac{1}{\alpha_i^*} - \varepsilon} \right), & \text{if } \alpha_i^* \geqq \frac{1}{2}. \end{cases} \tag{23}$$

Summing (22) and (23) over all corners of C and applying Lemmas 1 and 2 we obtain

$$S = O(h^{\beta^*}), \tag{24}$$

where

$$\beta^* = \begin{cases} \dfrac{1}{\max\limits_{i=1,\dots,n} \alpha_i^*} - \varepsilon, & \text{if } \max\limits_{i=1,\dots,n} \alpha_i^* \geqq \frac{1}{2}, \\ 2, & \text{if } \max\limits_{i=1,\dots,n} \alpha_i^* < \frac{1}{2} \text{ or if} \\ & \text{there are no corners.} \end{cases}$$

Inserting (24) into (20) we get

$$\max_{P \in R_h^*} |w(P)| = O(h^{\beta^*} |\log h|). \tag{25}$$

Combining (17) and (25) we obtain (9). This completes the proof of Theorem 1.

If $b(x, y) \equiv 0$, then the matrix corresponding to the problem (2) is diagonally dominant and of non-negative type and, consequently, the finite-difference analogue of the maximum principle is valid (see, for example, [14]). Let

$$\gamma^* = \begin{cases} \dfrac{1}{\max\limits_{i=1,\dots,n} \alpha_i^*} - \varepsilon, & \text{if } \max\limits_{i=1,\dots,n} \alpha_i^* \geqq 1, \\ 1, & \text{if } \max\limits_{i=1,\dots,n} \alpha_i^* < 1 \text{ or if} \\ & \text{there are no corners.} \end{cases}$$

(Note that $\alpha_i^* < 1$ if and only if $\alpha_i < 1$.) Since by Lemma 3 $u(P) - u(P') = O(h^{\gamma^*})$ for $P \in C_h^*$, we have instead of (17) the sharper estimate

$$\max_{P \in R_h^*} |v(P)| = O(h^{\gamma^*}). \tag{26}$$

Combining (25) and (26) we get

$$\max_{P \in R_h^*} |z(P)| = O(h^{\gamma^*}). \tag{27}$$

Thus we have proved the following theorem

THEOREM 2. *Assume that* $b(x, y) \equiv 0$. *Then the truncation error* $z(P)$ *satisfies* (27).

REFERENCES

[1] WALSH, J. L., YOUNG, D.: On the degree of convergence of solutions of difference equations to the solution of the Dirichlet problem, *J. Math. Phys.* **33** (1954) 80—93.

[2] LAASONEN, P.: On the degree of convergence of discrete approximations for the solutions of the Dirichlet problem, *Ann. Acad. Sci. Fenn. A. I.*, **246** (1957) 1—19.

[3] LAASONEN, P.: On the truncation error of discrete approximations to the solutions of Dirichlet problems in a domain with corners, *J. Assoc. Comput. Mach.* **5** (1958) 32—38.

[4] ВОЛКОВ, Е. А.: Эффективные оценки погрешности решений методом сеток задачи Дирихле для уравнения Лапласа на многоугольниках, *Докл. Акад. Наук СССР* **155** (1964) 735—738.

[5] HUBBARD, B.: Remarks on the order of convergence in the discrete Dirichlet problem, In the collection "Numerical solution of partial differential equations" edit. by J. H. Bramble, Academic Press, New York—London, 1966. 21—34.

[6] Саульев, В. К.: К вопросу решения задачи о собственных значениях методом конечных разностей, В. сб. „*Вычисл. матем. и вычисл. техн.*", 2, Изд-во АН СССР, Москва, 1955, 116—144.

[7] BERNSTEIN, D. L.: Existence theorems in partial differential equations, *Ann. Math. Studies*, **23**, Princeton Univ, Press, 1950.

[8] MORREY, C. B. and NIRENBERG, L.: On the analyticity of linear elliptic systems of partial differential equations, *Communs Pure and Appl. Math.* **10** (1957) 271—290.

[9] Кондратьев, В. А.: Краевые задачи для эллиптических уравнений в конических областях, *Докл. Акад. Наук* СССР **153** (1963) 27—29.

[10] Конгратьев, В. А.: Краевые задачи для эллиптических уравнений высших порядков при наличии особенностей границы, *Материалы к Совместному советско-американскому симпозиуму по уравнениям с частными производыми*, Изд-во Сибирского отделенни АН СССР, Новосибирск, 1963.

[11] Вейдингер, Л.: Об оценке погрещности при нахождении собственных значений методом конечных разностей, *Ж. Вычисл. Мат.* и *Мат. Физ.* **5** (1965) 806—815.

[12] BRAMBLE, J. H.: A second order finite difference analog of the first biharmonic boundary value problem, *Num. Math.* **9** (1966) 236—249.

[13] COURANT, R., FRIEDRICHS, K. und LEWY, H.: Über die partiellen Differenzengleichungen der mathematischen Physik, *Math. Ann.* **100** (1928) 32—74.

[14] BRAMBLE, J. H. and HUBBARD, B.: A theorem on error estimation for finite difference analogues of the Dirichlet problem for elliptic equations, *Contributions to Differential Equations*, 2, Wiley, New York, 1963.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received July 11, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 345—349.

THE COVERING OF GRAPHS BY CLIQUES

by

L. SURÁNYI

1. The investigation of the chromatic number of graphs is of great importance. The chromatic number of a graph gives us the minimal number of the mutually disjoint cliques[1] of the complementary graph "covering" all the vertices. In the last decade a number of authors dealt with the connection of this minimal covering number and the maximal number of the independent vertices of a graph. We denote by τ_G the minimal number of the mutually disjoint cliques of G covering all the vertices of the non-empty graph G and by φ_G the maximal number of the independent vertices of G. Since a clique cannot have more than one vertex of an independent set, the inequality $\varphi_G \leqq \tau_G$ holds for every graph. For the "most" of the graphs $\varphi_G < \tau_G$ holds and TUTTE and ZYKOV proved ([3], [8]) that for τ_G can be arbitrary large even for $\varphi_G = 2$.

However, for some special types of graphs the equality $\varphi_G = \tau_G$ is valid. DÉNES KÖNIG was the first who gave a class of such graphs, a theorem of his, found in 1932 (cf. [5]) being equivalent with the following

THEOREM 1. *If G is a bipartite graph, then $\varphi_G = \tau_G$.*

(This and the next theorems are stated only for non-empty graphs).

In 1959 A. HAJNAL and J. SURÁNYI [7] found the following

THEOREM 2. *If every polygon in G with more than 3 vertices has a diagonal*[2] *belonging to G, then $\varphi_G = \tau_G$.*

Having Theorems 1 and 2 one could guess that the equality $\varphi_G = \tau_G$ holds for every graph G in which every odd polygon with more than 3 vertices has a diagonal belonging to G. However, this is not the case as shown by the graph consisting of a 7-gon and of its 7 shortest diagonals. But the following generalization of the theorems 1 and 2, found by T. GALLAI [6] is true:

[1] We consider only graphs without loops and multiple edges. A graph G is called a *clique* if each two vertices of G are connected by an edge. A graph with just one vertex will be considered as a clique. A set S of mutually disjoint cliques of a graph G *covers* G if each vertex of G belongs to a clique of S. The vertices $x_1, \ldots, x_j$ ($j \geqq 1$) of G are *independent* (in G) if no two of them are connected by an edge (of G). A set of independent vertices is briefly called an *independent set.*

[2] $xy (= yx)$ will denote the edge connecting the vertices x and y. The diagonals of the n-gon $P = (x_1 \ldots x_n x_1)$ ($n \geqq 4$) are the edges $x_i x_{i+\alpha}$ ($i = 1, \ldots n-2$; $2 \leqq \alpha \leqq n-i$). The diagonals $x_i x_{i+2}$ ($i = 1, \ldots n-2$) and $x_{n-1}x_1$, $x_n x_2$ are the *shortest* diagonals of P. We say thath the diagonals $x_i x_j$ and $x_r x_s$ with $i<j$, $r<s$ *cross* if they satisfy one of the inequalities $i<r<j<s$, $r<i<s<j$.

THEOREM 3. *If every odd polygon of G can be triangulated in G* [3], *then* $\varphi_G = \tau_G$.

The proof of GALLAI is based on the characterization of the minimal separating sets of the graphs satisfying the condition of the theorem. These examinations are based on a lot of lemmas. In what follows, we prove this theorem by induction. This proof is much shorter but does not give any information on the structure of the examined graphs.

A common feature of the conditions of the three theorems is, that if a graph G fulfills these conditions, then every spanned [4] subgraph of G fulfills them as well. We can restate this property in the following form: if a graph G satisfies the conditions and x is an arbitrary vertex of G, then $G-x$ [5] satisfies them, too.

2. First we reformulate Theorem 3. It is apparent (and is easy to prove) that a triangulation of an odd polygon with more than 3 vertices always contains two non-crossing shortest diagonals, further if we consider two non-crossing shortest diagonals in an odd polygon P and omit from P the two triangles cut by these diagonals, we get an odd polygon again. A simple consequence of these facts is the following statement:

LEMMA. Every odd polygon of a graph G can be triangulated in G if and only if every odd polygon of G with more than 3 vertices has two non-crossing shortest diagonals in G.

In virtue of this lemma Theorem 3 is equivalent to the following

THEOREM 4. *If every odd polygon of G with more than 3 vertices has two non-crossing shortest diagonals in G, then* $\varphi_G = \tau_G$.

We shall prove this statement by induction with respect to the number of vertices of G.

In what follows a covering of G will always be given by a partition $F=(E_1, \ldots, E_t)$ of the set $V(G)$, in which the vertices of each class E_i span a clique in G. We shall call those partitions of this kind, where the number t of classes has its minimal value τ_G, *minimal partitions* of $V(G)$. If F is a partition of $V(G)$ and $x \in V(G)$, we denote the class of F which contains x by $F(x)$.

Theorem 4 is trivially true for $n=1$. We assume now that the statement of Theorem 4 is true for graphs with $n-1$ vertices ($n>1$) and consider a graph G with n vertices which satisfies the condition of the theorem. We make the assumption

$$\varphi_G < \tau_G \tag{1}$$

and we are going to show that this leads to a contradiction.

Let us choose an independent set of G with $r=\varphi_G$ vertices. We denote it by A and fix it for the following. By (1) G must have a vertex not belonging to A,

[3] A *triangulation* of the n-gon P consists of the vertices and edges of P together with n—3 pairwise non-crossing diagonals of P. The polygon P of G can be *triangulated in Z* if it has a triangulation which is a subgraph of G.

[4] $V(G)$ will always denote the set of vertices of the graph G. Let $S \subseteq V(G)$. The subgraph *spanned* by S(in G) is the subgraph of G whose vertex-set is S and whose edges are all the edges of G which have both vertices in S.

[5] If $x \in V(G)$ we denote by $G-x$ the graph which we obtain by omitting from G the vertex x and all edges incident to x.

otherwise the vertices of G, as cliques, would give a covering of G by φ_G cliques. We choose an arbitrary vertex b_0 of $V(G)-A$. Since the graph $H = G-b_0$ satisfies the conditions of Theorem 4 and has $n-1$ vertices, we have $\varphi_H=\tau_H$. The set A is obviously a maximal independent set of H too, so $\varphi_H=\varphi_G$, and the minimal partitions of the set $V(H)$ consist of $\tau_H=\varphi_G=r$ classes. Every class of such a partition contains exactly one vertex of A.

Let $\mathscr{F}_0$ denote the family of the minimal partitions of $V(H)$. b_0 is connected (by an edge of G) to at least one vertex of A otherwise $A\cup\{b_0\}$ would be an independent set of G with $r+1$ vertices. Let $A_1=\{a_0, \ldots, a_{\omega_1}\}$ $(\omega_1 \geqq 1)$ be the set of those vertices of A which are connected to b_0.

Let us choose a partition F_1 belonging to $\mathscr{F}_0$ for which the sum[6]

$$\sum_{i=1}^{\omega_1} |F_1(a_i)|$$

is minimal. Put $F_1(a_i)=E_i$ $(i=1, \ldots, \omega_1)$. Every E_i contains at least one vertex not connected with b_0. Namely, if b_0 were connected for some i to every vertex of E_i, then $E_i'=E_i\cup\{b_0\}$ would span a clique in G and so, replacing E_i with E_i' in F_1 we should obtain a minimal partition of $V(G)$ with r classes, in contradiction to (1). Let us choose for each i with $1\leqq i\leqq\omega_1$ a vertex b_i of E_i not connected with b_0 and let $B_1=\{b_1, \ldots b_{\omega_1}\}$.

In case there are some vertices in $A-A_1$ which are connected to some vertex of B_1, we denote the set of these vertices by $A_2=\{a_{\omega_1+1}, \ldots a_{\omega_2}\}$ anda ssign to every a_i $(\omega_1<i\leqq\omega_2)$ a vertex $b_{\sigma(i)}$ of B_1 that is connected to it. $(1\leqq\sigma(i)\leqq\omega_1)$. Naturally, $\sigma(i)=\sigma(j)$ can hold for $i\neq j$ too. Further, let us denote by $\mathscr{F}_1$ the family of those partitions belonging to $\mathscr{F}_0$, in which the classes containing $a_1, \ldots a_{\omega_1}$, are $E_1, \ldots E_{\omega_1}$, respectively.

Now we choose a partition F_2 belonging to $\mathscr{F}_1$ for which the sum

$$\sum_{i=\omega_1+1}^{\omega_2} |F_2(a_i)|$$

is minimal and put $F_2(a_i)=E_i$ $(i=\omega_1+1, \ldots \omega_2)$. In every $E_i(\omega_1<i\leqq\omega_2)$ there must exist a vertex not connected to $b_{\sigma(i)}$. If namely for some i all vertices of E_i are connected to $b_{\sigma(i)}$, then $E_i'=E_i\cup\{b_{\sigma(i)}\}$ spans a clique in G, and replacing E_i with E_i' and $E_{\sigma(i)}$ with $E_{\sigma(i)}'=E_{\sigma(i)}-b_{\sigma(i)}$ in F_2 we obtain a partition F_2' belonging to $\mathscr{F}_0$ for which the sum

$$\sum_{i=1}^{\omega_1} |F_2'(a_i)| < \sum_{i=1}^{\omega_1} |F_1(a_i)|$$

in contradiction to the minimal property of F_1.

Let us choose a vertex b_i in every E_i $(\omega_1<i\leqq\omega_2)$ not connected to $b_{\sigma(i)}$ and let $B_2=\{b_{\omega_1+1}, \ldots, b_{\omega_2}\}$.

Similarly now we define by induction the sets of vertices

$$A_j=\{a_{\omega_{j-1}+1}, \ldots, a_{\omega_j}\}, \quad B_j=\{b_{\omega_{j-1}+1}, \ldots, b_{\omega_j}\} \qquad (\omega_{j-1}<\omega_j,\ \omega_0=0),$$

[6] $|S|$ denote the number of elements of the finite set S.

also the family of partitions $\mathscr{F}_{j-1}$, the partition F_j for $j=1, \ldots, l$, further the class of vertices E_i and the function $\sigma(i)$ for $i=1, \ldots, \omega_l$, where $l(\geqq 1)$ will be defined later on. For the easier expression we use $\sigma(i)=0$ for $\omega_0 < i \leqq \omega_1$.

Suppose that these symbols are already defined for $j \leqq q$, $i \leqq \omega_q$, $(q \geqq 1)$ and put $A'_q = \bigcup_{j=1}^{q} A_j$.

If no vertex of $A - A'_q$ is connected to any vertex of B_q then $l=q$ and the procedure is finished. Otherwise let $A_{q+1} = \{a_{\omega_q+1}, \ldots a_{\omega+1}\}$ be the set of those vertices of $A - A'_q$ which are connected to some vertex of B_q.

For each i with $\omega_q < i \leqq \omega_{q+1}$ we choose a value $\sigma(i)$ satisfying $\omega_{q-1} < \sigma(i) \leqq \omega_q$ for which the edge $a_i b_{\sigma(i)}$ belongs to G. We denote by $\mathscr{F}_q$ the family of those partitions belonging to $\mathscr{F}_{q-1}$ for which the classes of vertices containing $a_1, \ldots, a_{\omega_q}$ are $E_1, \ldots, E_{\omega_q}$, respectively.

We choose a member F_{q+1} of $\mathscr{F}_q$ for which the sum

$$\sum_{i=\omega_q+1}^{\omega_{q+1}} |F_{q+1}(a_i)|$$

is minimal and put $F_{q+1}(a_i) = E_i$ $(i = \omega_q + 1, \ldots, \omega_{q+1})$.

We choose for each i with $\omega_q < i \leqq \omega_{q+1}$ a vertex b_i of the class E_i not connected with $b_{\sigma(i)}$. The existence of such vertices can be shown exactly on the same way as for $\omega_1 < i \leqq \omega_2$. Let $B_{q+1} = \{b_{\omega_q+1}, \ldots, b_{\omega_{q+1}}\}$. The procedure must stop, since the set A is finite.

From our definition it follows that the sets $A_1, \ldots, A_l, B_1, \ldots, B_l$ are mutually disjoint and no vertex of the set $A - A'_l$ is connected to any vertex of the set $B' = \bigcup_{i=0}^{l} B_i$ (where $B_0 = \{b_0\}$). The classes $E_1, \ldots, E_{\omega_l}$ are also mutually disjoint, and each of these classes spans a clique in G. Further

1. all the edges $a_i b_{\sigma(i)}$ and $a_i b_i$ $(i=1, \ldots \omega_l)$ belong to G;
2. none of the edges $b_i b_{\sigma(i)}$ $(i=1, \ldots, \omega_l)$ belongs to G.

We shall prove that B' is an independent set of G. By the above considerations this implies that $B' \cup (A - A'_l)$ is also an independent set of G, however B', containing b_0, has exactly one vertex more than A'_l, i.e. $B' \cup (A - A'_l)$ is an independent set of G consisting of $r+1$ vertices. This contradicts to $\varphi_G = r$ and so our theorem shall be proved. Hence, to finish our proof we have only to show, that B' is an independent set of G.

For this purpose we consider the subgraph K of G consisting of the vertices of the set $A'_l \cup B'$ and of the edges mentioned in 1. K has the following properties:

α) it is a connected graph,

β) it is a bipartite graph, which contains only edges connecting vertices of A'_l to vertices of B',

γ) each vertex a_i of A'_l is connected (in K) to just two vertices, namely to b_i and $b_{\sigma(i)}$, and these two vertices are not connected in G.

Properties β) and γ) can be seen immediately from the definition of A'_l and B'. α) is the consequence of the following: We define for $1 \leqq i \leqq \omega_l$ $i_1 = i$, $i_{j+1} = \sigma(i_j)$ if $i_j \neq 0$. There exists an integer $h \geqq 1$ with $i_{h+1} = 0$ and $(b_{i_1} a_{i_2} \ldots b_{i_h} a_{i_h} b_{i_{h+1}})$ is a simple path of K connecting b_i to b_0, and also a_i to b_0. We just mention that it can be easily verified — but we do not need it — that K is a tree.

Now we assume that contrary to our statement there exist two vertices of B': x' and x'', which are connected by an edge in G. Because of α) there exists a simple path S in K with the ends x' and x''. Because of β) $S=(x_1 y_1 \dots x_{m-1} y_{m-1} x_m)$ where $x_1 = x'$, $x_m = x''$, $m \geqq 2$, x_i belongs to B', y_i belongs to A_i' for every $i=1, \dots, \dots, m-1$. By γ) $m=2$ cannot occur. S and the edge $x'x''$ form an odd polygon P of G. The shortest diagonals of P are the edges

$$x_1x_2,\ y_1y_2,\ x_2x_3,\ \dots,\ x_{m-1}x_m,\ y_{m-1}x_1,\ x_my_1.$$

The edges x_ix_{i+1} $(i=1, \dots m-1)$ and the edges y_iy_{i+1} $(i=1, \dots, m-2)$ do not belong to G either because of γ, and because of the independence of A, respectively. So P can have only two shortest diagonals in G: $y_{m-1}x_1$ and x_my_1. However, these diagonals cross, so P cannot have two non-crossing shortest diagonals in G. This contradicts to our condition, and so B' is an independent set of G, as we stated.

REFERENCES

[1] BERGE, C.: Les problèmes de coloration en théorie des graphes, *Publ. Inst. Statist. Univ. Paris* **9** (1960) 123—160.

[2] BERGE, C.: Some classes of perfect graphs, *Six papers on graph theory*, Indian Statist, Inst., 1963.

[3] DESCARTES, B.: Solution of advanced problem 4526, *Amer. Math. Montly* **61** (1954) 352—353.

[4] DIRAC, G.: On rigid circuit graphs, *Abhandlungen aus dem Math. Seminar der Univ. Hamburg* **25** (1961) 71—76.

[5] GALLAI, T.: Über extreme Punkt- und Kantenmengen, *Ann. Univ. Sci. Budapest*, **2** (1959) 133—138.

[6] GALLAI, T.: Graphen mit triangulierbaren ungeraden Vielecken, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962) A. 3—36.

[7] HAJNAL, A. und SURÁNYI, J.: Über die Auflösung von Graphen in vollständigen Teilgraphen, *Ann. Univ. Sci. Budapest*, **1** (1958) 113—121.

[8] ZYKOV, A. A.: On some properties of linear complexes, (Russian) *Mat. Sbor.* **24** (1966) 1949, 163—188.

Eötvös Loránd University, Budapest

(Received July 31, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 351—358.

ON QUADRATIC INEQUALITIES IN PROBABILITY THEORY

by

J. GALAMBOS and A. RÉNYI

Summary

In this paper quadratic inequalities in the probabilities of Boolean functions of n variable events are considered. For a special class of such inequalities — called exact inequalities — a necessary and sufficient condition is given; this general theorem is applied to deduce certain special inequalities. Generalization to inequalities of degree higher than 2 is also considered.

§ 0. Notations

Let $S=(\Omega, \mathscr{A}, \mathsf{P})$ denote a probability space, i.e. let Ω be an arbitrary non-empty set, $\mathscr{A}$ a σ-algebra [1] of subsets of Ω and P a measure on $\mathscr{A}$ such that $\mathsf{P}(\Omega)=1$. We call the elements of $\mathscr{A}$ events and denote them by capital letters. We denote by $A+B$ the union and by AB the intersection of the sets A and B, and by $\bar{A}$ the complement of the set A with respect to Ω. As usual, $\bar{A}$ is interpreted as the event consisting in the non-occurrence of the event A, while $A+B$ and AB respectively, are interpreted as the event that at least one of the events A, B occurs, resp. that both the events A, B occur.

Let $p_1, p_2, \ldots, p_r$ be any set of positive numbers such that

$$\sum_{j=1}^{r} p_j = 1$$

We shall denote by $S_r(p_1, \ldots, p_r)$ that (finite) probability space in which the set Ω consists of r elements $\omega_1, \omega_2, \ldots, \omega_r$. $\mathscr{A}$ is the set of all 2^r subsets of Ω, and P is defined by

$$\mathsf{P}(A)=\sum_{\omega_j \in A} p_j \tag{0.1}$$

Especially $S_1(1)$ is the trivial probability space which contains only two events: the "certain event" Ω and the "impossible event" $\emptyset$ (the empty set). Further $S_2(\frac{1}{2}, \frac{1}{2})$ is the probability space (describing e.g. the throw of a fair coin) which contains only four events: $\Omega, \emptyset, \alpha=\{\omega_1\}$ and $\beta=\{\omega_2\}$ and $\mathsf{P}(\alpha)=\mathsf{P}(\beta)=\frac{1}{2}$.

A Boolean function $F=F(A_1, A_2, \ldots, A_n)$ of n variable events $A_1, \ldots, A_n$ is a function of these events which can be expressed by means of the variables

[1] All results of this paper are valid also if $\mathscr{A}$ is only an algebra of subsets of Ω and P a finitely additive nonnegative set function on $\mathscr{A}$ for which $\mathsf{P}(\Omega)=1$.

$A_1, \ldots, A_n$ and a finite number of Boolean operations, i.e. the operations $A+B$, AB, $\bar{A}$. We introduce the notation

$$A^1 = A, \quad A^{-1} = \bar{A}.$$

Let us denote by $\delta_k(m)$ the k-th digit of the binary representation of the non-negative integer m, i.e. we put

$$m = \sum_{k \geqq 0} \delta_k(m) 2^k \tag{0. 2}$$

Let us put further

$$\varepsilon_k(m) = 2\delta_{k-1}(m) - 1 \qquad (k = 1, 2, \ldots). \tag{0. 3}$$

Clearly $\varepsilon_k(m) = \pm 1$, and if m runs over the integers $0, 1, \ldots, 2^n - 1$, the n-tuple $\{\varepsilon_1(m), \ldots, \varepsilon_n(m)\}$ runs over all 2^n possible n-tuples of the signs $+1$ and -1.

Let us put

$$B_n(m) = A_1^{\varepsilon_1(m)} A_2^{\varepsilon_2(m)} \ldots A_n^{\varepsilon_n(m)} \qquad (0 \leqq m \leqq 2^n - 1) \tag{0. 4}$$

We call the $B_n(m)$ the basic Boolean functions of the variables $A_1, \ldots, A_n$. Clearly

$$B_n(m_1) B_n(m_2) = \emptyset \quad \text{if} \quad m_1 \neq m_2 \tag{0. 5}$$

and

$$\sum_{m=0}^{2^n - 1} B_n(m) = \Omega \tag{0. 6}$$

It is well known that every Boolean function $F(A_1, \ldots, A_n)$ can be uniquely represented in a „canonical form" as the sum of certain basic functions $B_n(m)$; thus there are only 2^{2^n} different Boolean functions of n variable events.

§ 1. Introduction

Some time ago, the second named author has proved ([1], see also [2]) the following

THEOREM 1. *Let $F_j = F_j(A_1, A_2, \ldots, A_n)$ $(j = 1, 2, \ldots, N)$ be arbitrary Boolean functions of the n variable events $A_1, \ldots, A_n$. The linear inequality*

$$\sum_{j=1}^{N} c_j \mathrm{P}(F_j) \geqq 0 \tag{1. 1}$$

(where $c_1, \ldots, c_N$ are real constants) is valid in every probability space S if it is valid in the trivial probability space $S_1(1)$.

This simple theorem is useful because it makes it possible to reduce the proof of any linear inequality among probabilities of Boolean functions to a corresponding combinatorial inequality.

To make this paper self-contained we reproduce here the proof of Theorem 1, especially as the proof is very short.

PROOF OF THEOREM 1. Let the expression of the functions $F_1, \ldots, F_N$ in canonical form be

$$F_j = \sum_{m \in E_j} B_n(m) \qquad (j = 1, 2, \ldots, N) \tag{1. 2}$$

where E_j is some subset of the set $\{0, 1, \ldots, 2^n-1\}$. It follows from (0. 5) that

$$\mathsf{P}(F_j)=\sum_{m\in E_j}\mathsf{P}(B_n(m)) \tag{1.3}$$

and thus

$$\sum_{j=1}^{N} c_j\mathsf{P}(F_j)=\sum_{m=0}^{2^n-1} d_m\mathsf{P}(B_n(m)) \tag{1.4}$$

where

$$d_m=\sum_{m\in E_j} c_j \tag{1.5}$$

Now evidently if $A_k=\Omega$ if $\varepsilon_k(m)=1$ and $A_k=\emptyset$ if $\varepsilon_k(m)=-1$, then $B_n(m)=\Omega$ and $B_n(l)=\emptyset$ for $l\neq m$, $0\leqq l\leqq 2^n-1$, thus for this special choice of the values of the variables $A_1, \ldots, A_n$ we have

$$\sum_{j=1}^{N} c_j\mathsf{P}(F_j)=d_m \tag{1.6}$$

Thus if (1.1) holds on $S_1(1)$ we have $d_m\geqq 0$ for $m=0, 1, \ldots, 2^n-1$ and thus it follows from (1. 4) that (1.1) holds for every choice of the values of the events $A_1, \ldots, A_n$ in every probability space S. Thus Theorem 1 is proved.

It is evident that Theorem 1 can be used also to prove identities. To prove that a relation

$$\sum_{j=1}^{N} c_j\mathsf{P}(F_j)=0 \tag{1.7}$$

is valid, according to Theorem 1 it is sufficient to verify that (1. 7) holds if all A_k are equal either to Ω or to $\emptyset$.

A typical example of an inequality which can be obtained as a special case of Theorem 1 is the following inequality, due to GUMBEL ([3]): Putting

$$\sigma_k^{(n)}=\sum_{1\leqq i_1<i_2<\ldots<i_k\leqq n}\mathsf{P}(A_{i_1}A_{i_2}\ldots A_{i_k}) \qquad (k=1, 2, \ldots, n) \tag{1.8}$$

one has for $2\leqq k\leqq n$

$$(n-k+1)\sigma_{k-1}^{(n)}\leqq\binom{n}{k}+(k-1)\sigma_k^{(n)}. \tag{1.9}$$

By means of Theorem 1 the proof of (1. 9) is reduced to a simple inequality between binomial coefficients (see [2], p. 30).

The aim of this paper is to prove a theorem similar to Theorem 1 for *quadratic* (instead of linear) inequalities. This will be done in § 2. In § 3 we give some applications of the general theorem of § 2. In § 4 we discuss the possibility of generalizing the result of § 2 to polynomial inequalities of the third and still higher degrees.

§ 2. A General Theorem on Quadratic Inequalities

In this § we consider quadratic inequalities of the form

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}\mathsf{P}(F_i)\mathsf{P}(F_j)\geqq 0 \tag{2.1}$$

where the $c_{i,j}$ are real constants, and $F_1, F_2, \ldots, F_N$ are Boolean functions of the variable events $A_1, \ldots, A_n$.

Note that it is no restriction that in (2. 1) no linear terms occur, because one of the F_i may be equal to Ω (which is also a Boolean function, namely a constant function) and thus inequalities which contain both quadratic and linear terms can be also written in the form (2.1).

We shall call an inequality (2.1) *exact*, if in (2.1) the equality sign is valid every time when each A_k is equal either to Ω or to $\emptyset$. By other words (2.1) is exact if equality is valid in (2.1) when the variables $A_1, \ldots, A_n$ are restricted to events in the trivial probability space $S_1(1)$.

We shall prove now the following

THEOREM 2. *Let* (2.1) *be an exact inequality. In order that* (2.1) *should be valid on every probability space S it is sufficient (and of course also necessary) that it should be valid on the probability space* $S_2(\frac{1}{2}, \frac{1}{2})$.

PROOF OF THEOREM 2. Let again (1. 2) be the expression of the function $F_j (1 \leqq j \leqq N)$ in canonical form. In view of (1. 3) we get

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i,j} \mathsf{P}(F_i)\mathsf{P}(F_j) = \sum_{r=0}^{2^n-1} \sum_{s=0}^{2^n-1} d_{r,s} \mathsf{P}\big(B_n(r)\big)\mathsf{P}\big(B_n(s)\big), \tag{2. 2}$$

where

$$d_{r,s} = \sum_{\substack{r \in E_i \\ s \in E_j}} c_{i,j} \tag{2. 3}$$

Now let us choose $A_k = \Omega$ if $\varepsilon_k(r) = 1$ and $A_k = \emptyset$ if $\varepsilon_k(r) = -1$ $(k = 1, 2, \ldots, n)$.

It follows that $\mathsf{P}\big(B_n(r)\big) = 1$ and $\mathsf{P}\big(B_n(s)\big) = 0$ if $s \neq r$; thus for this special choice of the values of the variables $A_1, \ldots, A_n$ we have

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i,j} \mathsf{P}(F_i)\mathsf{P}(F_j) = d_{r,r} \tag{2. 4}$$

As we have supposed that the inequality (2.1) is exact, it follows that

$$d_{r,r} = 0 \quad \text{for} \quad 0 \leqq r \leqq 2^n - 1. \tag{2. 5}$$

Putting

$$D_{r,s} = d_{r,s} + d_{s,r} \quad \text{for} \quad r \neq s \tag{2. 6}$$

we obtain

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i,j} \mathsf{P}(F_i)\mathsf{P}(F_j) = \sum_{0 \leqq r < s \leqq 2^n - 1} D_{r,s} \mathsf{P}\big(B_n(r)\big)\mathsf{P}\big(B_n(s)\big) \tag{2. 7}$$

Now let us choose an arbitrary pair (r, s) of integers, $0 \leqq r < s \leqq 2^n - 1$, and let us choose the values of the events A_k as follows:

$$\begin{aligned} A_k &= \Omega \quad \text{if} \quad \varepsilon_k(r) = \varepsilon_k(s) = 1 \\ A_k &= \alpha \quad \text{if} \quad \varepsilon_k(r) = 1 \quad \text{and} \quad \varepsilon_k(s) = -1 \\ A_k &= \beta \quad \text{if} \quad \varepsilon_k(r) = -1 \quad \text{and} \quad \varepsilon_k(s) = +1 \\ A_k &= \emptyset \quad \text{if} \quad \varepsilon_k(r) = \varepsilon_k(s) = -1 \end{aligned} \tag{2. 8}$$

where α and β are the events $\alpha=\{\omega_1\}$, $\beta=\{\omega_2\}$ of the probability space $S_2(\frac{1}{2}, \frac{1}{2})$. For this special choice of the values of the variables A_k we have clearly

$$B_n(r)=\alpha,\ B_n(s)=\beta \quad \text{and} \quad B_n(t)=\emptyset \quad \text{for} \quad t\neq r,\ t\neq s. \tag{2.9}$$

Thus we obtain for this choice of the values of the A_k

$$\mathsf{P}(B_n(r))=\mathsf{P}(B_n(s))=\tfrac{1}{2}, \qquad \mathsf{P}(B_n(t))=0 \quad \text{for} \quad t\neq r,\ t\neq s, \tag{2.10}$$

and therefore

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}\mathsf{P}(F_i)\mathsf{P}(F_j)=\tfrac{1}{4}D_{r,s} \tag{2.11}$$

Thus if (2.1) is valid on $S_2(\frac{1}{2}, \frac{1}{2})$, then we must have $D_{r,s}\geqq 0$ for all pairs (r, s) and thus in view of (2.7) it follows that (2.1) is valid on every probability space S and for every choice of the value of the variables A_k.

Thus Theorem 2 is proved.

Similarly as Theorem 1, Theorem 2 can be used also to prove identities. As a matter of fact we obtain from Theorem 2 the following

COROLLARY. *If*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}P(F_i)P(F_j)=0 \tag{2.12}$$

holds on $S_1(1)$ *and on* $S_2(\frac{1}{2}, \frac{1}{2})$, *then it holds identically on every probability space.*

§ 3. Some Applications of the General Theorem of § 2

In this § we consider some examples of quadratic inequalities which can be easily proved by means of Theorem 2.

EXAMPLE 1. Let us put $\sigma_0^{(n)}=1$ and

$$\sigma_k^{(n)}=\sum_{1\leqq i_1<i_2<\ldots<i_k\leqq n}\mathsf{P}(A_{i_1}A_{i_2}\ldots A_{i_k}) \tag{3.1}$$

We shall prove that the inequality

$$k\sigma_k^{(n)}\geqq\sigma_{k-1}^{(n)}(\sigma_1^{(n)}-k+1) \qquad (k=1, 2, \ldots, n) \tag{3.2}$$

is valid.

To prove (3.2) we first remark that it is a quadratic inequality of type (2.1). Further it is easy to see that (3.2) is an exact inequality. As a matter of fact if l among the events $A_1, \ldots, A_n$ are equal to Ω and the other $n-l$ to $\emptyset$, then three cases are possible:

a) either $l\leqq k-2$, in which case $\sigma_k^{(n)}=\sigma_{k-1}^{(n)}=0$ and thus both sides of (3.2) are equal to 0,

b) or $l=k-1$ in which case $\sigma_k^{(n)}=0$ and $\sigma_1^{(n)}-k+1=0$ and thus again both sides of (3.2) are equal to 0,

c) or $l\geqq k$, in which case $\sigma_k^{(n)}=\binom{l}{k}$, $\sigma_{k-1}^{(n)}=\binom{l}{k-1}$ and $\sigma_1^{(n)}=l$. As however

$$k\binom{l}{k}=\binom{l}{k-1}(l-k+1)$$

we have equality in (3. 2) in this case too. Thus (3. 2) is exact. Now let us check that (3. 2) holds for $S_2(\frac{1}{2}, \frac{1}{2})$. Suppose that among the events $A_1, \ldots, A_n$ l_1 are equal to Ω, l_2 to α, l_3 to β $(l_1+l_2+l_3 \leqq n)$ and the remaining $n-l_1-l_2-l_3$ to $\emptyset$. In this case

$$\sigma_j^{(n)} = \frac{1}{2}\left[\binom{l_1+l_2}{j}+\binom{l_1+l_3}{j}\right] \quad \text{for} \quad 1 \leqq j \leqq n$$

and thus

$$k\sigma_k^{(n)} - \sigma_{k-1}^{(n)}(\sigma_1^{(n)}-k+1) = \tfrac{1}{4}(l_2-l_3)\left[\binom{l_1+l_2}{k-1}-\binom{l_1+l_3}{k-1}\right] \geqq 0 \tag{3. 3}$$

Thus by Theorem 2 (3. 2) holds on every probability space S for any choice of the events $A_1, \ldots, A_n$.

It is interesting to compare (3. 2) with GUMBEL's inequality (1. 9). The fact that (3. 2) is exact, while in GUMBEL's inequality we have equality (as seen from the proof) on $S_1(1)$ only if $l=n$ or $l=n-1$, shows that, (3. 2) gives sometimes a better estimate than (1. 9). Another such instance is when the events all have probability $\frac{1}{2}$, and $k=2$. In this case (1. 9) gives for $\sigma_2^{(n)}$ only the trivial lower estimate 0, while (3. 2) gives the non-trivial (in fact, asymptotically best possible) lower estimate $\sigma_2^{(n)} \geqq \dfrac{n(n-2)}{8}$.

For $k=2$ we obtain as a special case of (3. 2) the well known inequality

$$\sigma_2^{(n)} \geqq \binom{\sigma_1^{(n)}}{2}. \tag{3. 4}$$

It follows from (3. 2) by induction that

$$\sigma_k^{(n)} \geqq \binom{\sigma_1^{(n)}}{k}. \tag{3. 5}$$

It should be noted that one can deduce from (3. 4) the following inequality:

$$\text{If} \quad \sigma_2^{(n)} \leqq \binom{n}{2}p^2 \quad \text{then} \quad \sigma_1^{(n)} \leqq np+\frac{1}{2}(1-p)+\frac{1-p^2}{4p(2n-1)}$$

As a matter of fact, it follows from (3. 4) and the inequality $\sqrt{1+x} \leqq 1+\dfrac{x}{2}$ that

$$\sigma_1^{(n)} \leqq \frac{1+\sqrt{1+8\sigma_2^{(n)}}}{2} \leqq \frac{1}{2}+\frac{1}{2}(2np-p)\sqrt{1+\frac{1-p^2}{(2np-p)^2}}$$

and thus that

$$\sigma_1^{(n)} \leqq np+\frac{1-p}{2}+\frac{1-p^2}{4p(2n-1)}$$

REMARK. The exact maximum of $\sigma_1^{(n)}$ under condition $\sigma_2^{(n)} \leqq \binom{n}{p}p^2$ was determined in [4].

EXAMPLE 2. Let us consider the quadratic relation

$$\mathrm{P}^2(A+B)+\mathrm{P}^2(AB) = \mathrm{P}^2(A)+\mathrm{P}^2(B)+2\mathrm{P}(A\bar{B})\mathrm{P}(\bar{A}B) \tag{3. 6}$$

It is evidently valid on $S_1(1)$ and also on $S_2(\frac{1}{2}, \frac{1}{2})$, thus it holds identically.

§ 4. Cubic Inequalities

Theorem 2 can be generalized for cubic inequalities

$$\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N} c_{i_1,i_2,i_3}\mathsf{P}(F_{i_1})\mathsf{P}(F_{i_2})\mathsf{P}(F_{i_3}) \geqq 0 \tag{4.1}$$

where $F_1, \ldots, F_N$ are Boolean functions of the variable events $A_1, \ldots, A_n$. The inequality (4.1) is called *exact of order* 2 if for every p $(0 \leqq p \leqq 1)$ equality stands in (4.1), if $A_1, \ldots, A_n$ are all events of $S_2(p, 1-p)$. (Clearly an inequality which is exact of order 2 is exact.)

We prove the following

THEOREM 3. *Let* (4.1) *be an inequality which is exact of order* 2. *If* (4.1) *holds on* $S_3(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, *it holds on every probability space.*

PROOF. If (1. 2) is the canonical form of F_j we have

$$\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N} c_{i_1,i_2,i_3}\mathsf{P}(F_{i_1})\mathsf{P}(F_{i_2})\mathsf{P}(F_{i_3}) = \tag{4.2}$$

$$= \sum_{r_1=0}^{2^n-1}\sum_{r_2=0}^{2^n-1}\sum_{r_3=0}^{2^n-1} d(r_1, r_2, r_3)\mathsf{P}(B_n(r_1))\mathsf{P}(B_n(r_2))\mathsf{P}(B_n(r_3))$$

where

$$d(r_1, r_2, r_3) = \sum_{r_h \in E_{i_h}(h=1,2,3)} c_{i_1,i_2,i_3} (h = 1, 2, 3) \tag{4.3}$$

Clearly (4.1) being exact implies that

$$d(r, r, r) = 0 \qquad (0 \leqq r \leqq 2^n - 1).$$

Let us put for $r \neq s$

$$D(r, s) = d(r, r, s) + d(r, s, r) + d(s, r, r).$$

Now from the supposition that in (4.1) equality holds on $S_2(p, q)$ $(q = 1-p)$ it follows that for any pair of numbers r, s $(r \neq s)$

$$D(r, s)p + D(s, r)q = 0. \tag{4.4}$$

By supposition (4. 4) holds for $p = \frac{1}{2}$ and also for some p for which $0 < p < \frac{1}{2}$; it follows that

$$D(r, s) = 0 \quad \text{if} \quad s \neq r. \tag{4.5}$$

Thus we obtain, putting

$$D(r_1, r_2, r_3) = d(r_1, r_2, r_3) + d(r_1, r_3, r_2) + d(r_2, r_1, r_3) +$$
$$+ d(r_2, r_3, r_1) + d(r_3, r_1, r_2) + d(r_3, r_2, r_1)$$

that

$$\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N} c_{i_1,i_2,i_3}\mathsf{P}(F_{i_1})\mathsf{P}(F_{i_2})\mathsf{P}(F_{i_3}) = \tag{4.6}$$

$$= \sum_{0 \leqq r_1 < r_2 < r_3 \leqq 2^n - 1} D(r_1, r_2, r_3)\mathsf{P}(B_n(r_1))\mathsf{P}(B_n(r_2))\mathsf{P}(B_n(r_3))$$

Now let r_1, r_2, r_3 be any three different numbers, $0 \leqq r_1 < r_2 < r_3 \leqq 2^n - 1$. Let us denote the atoms of $S_3(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ by α_1, α_2 and α_3. Let us put

$$A_k = \sum_{\varepsilon_k(r_i)=1} \alpha_i$$

It is easy to show that for this choice of the values of the variable events A_k we have

$$B_n(r_i) = \alpha_i \qquad (i = 1, 2, 3). \tag{4. 7}$$

As a matter of fact $A_k^{\varepsilon_k(r_i)} \supseteq \alpha_i$ $(k = 1, 2, \ldots, n)$ thus

$$B_n(r_i) = \prod_{k=1}^{n} A_k^{\varepsilon_k(r_i)} \supseteq \alpha_i \tag{4. 8}$$

As however the events $B_n(r_1)$, $B_n(r_2)$, $B_n(r_3)$ are disjoint, (4. 8) implies (4. 7).

Clearly (4. 7) implies that for any s, different from each of r_1, r_2, r_3, one has $B_n(s) = \emptyset$. Thus for the above choice of the values of the variables $A_1, \ldots, A_n$ we have

$$\sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} c_{i_1,i_2,i_3} \mathrm{P}(F_{i_1}) \mathrm{P}(F_{i_2}) \mathrm{P}(F_{i_3}) = \frac{1}{27} D(r_1, r_2, r_3) \tag{4. 9}$$

As by supposition (4.1) holds on $S_3(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, we obtain from (4. 9)

$$D(r_1, r_2, r_3) \geqq 0 \quad \text{for} \quad 0 \leqq r_1 < r_2 < r_3 \leqq 2^n - 1. \tag{4.10}$$

In view of (4. 6) it follows that (4.1) holds for every probability space S.

As an example consider the following cubic inequality

$$\mathrm{P}(AB)\mathrm{P}(BC)\mathrm{P}(AC) \geqq \mathrm{P}^2(ABC)[\mathrm{P}(AB) + \mathrm{P}(AC) + \mathrm{P}(BC) - 2\mathrm{P}(ABC)] \tag{4. 11}$$

Clearly (4.11) is exact of order two. Thus we have to check only that (4.11) holds on $S_3(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, which is easily done.

Theorem 3 could be generalized also for polynomial inequalities of degree greater than 3.

REFERENCES

[1] RÉNYI, A.: Quelques remarques sur les probabilités d'événements dépendantes, *Journal de Math.* **37** (1958) 393—398.

[2] RÉNYI, A.: *Wahrscheinlichkeitsrechnung mit einer Anhang über Informationstheorie*, Deutscher Verlag der Wissenschaften, Berlin, 1962.

[3] FRÉCHET, M.: *Les probabilités associées a un systeme d'événements compatibles et dépendants, I, II*, Hermann, Paris, 1940, 1943.

[4] RÉNYI, A., ERDŐS, P., NEVEU, J.: An elementary inequality between the probability of events, *Math. Scandinavica* **13** (1963) 99—104.

University of Ghana, Legon (Accra) and Eötvös Loránd University, Budapest

(Received August 10, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 359—363.

BEMERKUNG ZU EINER ARBEIT VON L. FEJES TÓTH

von

A. FLORIAN

In einer kürzlich erschienenen Arbeit untersucht L. FEJES TÓTH [1] das Problem der dichtesten Packung von Parallelbereichen eines gegebenen Rechtecks mit den Seiten $a \leqq b$ im Abstand 1. Nach allgemeinen Sätzen (vgl. [2]) erhält man eine dichteste Packung kongruenter, konvexer, zentralsymmetrischer Bereiche s durch folgende Konstruktion einer gitterartigen Packung: Man ermittle ein s enthaltendes, konvexes Sechseck h von minimalem Flächeninhalt (das stets existiert und zentralsymmetrisch angenommen werden kann), und parkettiere damit die Ebene. Die entsprechenden Exemplare von s bilden eine dichteste Packung, deren Dichte gleich s/h ist (wir bezeichnen einen Bereich und seinen Flächeninhalt mit demselben Symbol). FEJES TÓTH führt in [1] die Bestimmung von h für den Parallelbereich eines Rechtecks auf das entsprechende Problem für ein Quadrat der Seitenlänge $a=2x$ zurück und beweist das folgende Resultat: Es ist

$$(1)\qquad \frac{h}{2} = \begin{cases} \operatorname{Min}[f_0(x),\ f_{11}(x)] & (0 < x \leqq 2-\sqrt{3}) \\ \operatorname{Min}[f_0(x),\ f_{11}(x),\ f_{12}(x)] & \left(2-\sqrt{3} < x < 1-\dfrac{1}{\sqrt{2}}\right) \\ \operatorname{Min}[f_{11}(x),\ f_2(x)] & \left(1-\dfrac{1}{\sqrt{2}} \leqq x < 1\right) \\ f_2(x) & (1 \leqq x) \end{cases}$$

Dabei bedeuten

$$(2)\qquad f_0(x) = (1+\sqrt{2}\,x)\left(\sqrt{3-\sqrt{8}\,x-2x^2}+\sqrt{8}\,x\right),$$

$$(2')\qquad f_{11}(x) = (1+x)\left(\sqrt{3+2x-x^2}+2x\right),$$

$$(2'')\qquad f_{12}(x) = 8\,(\sin^3\alpha\cos\alpha+\sin^3\beta\cos\beta)+2x(1+x)$$

mit
$$\frac{\sin 3\alpha}{\cos\alpha} = \frac{\sin 3\beta}{\cos\beta} \qquad \left(\frac{\pi}{8} < \alpha < \beta < \frac{\pi}{4}\right)$$

und
$$x = 2\,(\sin^2\alpha+\sin^2\beta)-1,$$

$$(2''')\qquad f_2(x) = 2x^2+4x+\sqrt{8}-1.$$

Diese Funktionen haben einfache geometrische Bedeutungen (vgl. [1]). Die Minima in (1) werden nur durch numerische Rechnungen bestimmt.

Im folgenden geben wir an ihrer Stelle die die numerischen Untersuchungen bestätigenden exakten Beweise und haben damit den

SATZ: *Es sei s der Parallelbereich des Quadrates der Seitenlänge* $2x$ *im Abstand* 1. *Dann gilt für das kleinste, s enthaltende, konvexe, zentralsymmetrische Sechseck h*

$$\frac{h}{2} = \begin{cases} f_{11}(x) & (0 < x \leqq 2-\sqrt{3}) \\ f_{12}(x) & \left(2-\sqrt{3} < x < 1-\frac{1}{\sqrt{2}}\right) \\ f_2(x) & \left(1-\frac{1}{\sqrt{2}} \leqq x\right). \end{cases} \tag{3}$$

Wie wir zeigen werden, läßt sich $f_{12}(x)$ explizit in der Form

$$f_{12}(x) = 2\sqrt{-x^4+6x^3-12x^2+8x}+2x(1+x)$$

schreiben.

BEMERKUNG. Die Überlegungen am Schluß von [1] ergeben, daß das kleinste zentralsymmetrische, konvexe Sechseck, das den Parallelbereich des Rechtecks mit den Seiten $a \leqq b$ im Abstand 1 enthält, den Inhalt $h+(b-a)(2+a)$ hat, wenn h der entsprechende minimale Inhalt für das Quadrat mit der Seitenlänge a ist.

1. Es sei $0 < x \leqq 2-\sqrt{3}$. Die Gleichung

$$f_0(x)-f_{11}(x) = 0 \tag{4}$$

führt nach zweimaligem Quadrieren auf

$$\begin{aligned} &65x^6+16(-9+16\sqrt{2})x^5+4(209-100\sqrt{2})x^4+8(-86+73\sqrt{2})x^3+ \\ &+4(37-24\sqrt{2})x^2+16(12-9\sqrt{2})x+16(-3+2\sqrt{2}) = 0. \end{aligned} \tag{5}$$

Die linke Seite dieser Gleichung ist, wie man sofort sieht, eine konvexe Funktion, die für $x=0$ und $x=2-\sqrt{3}$ negative Werte annimmt. Folglich ist sie im ganzen Intervall $(0, 2-\sqrt{3}]$ negativ, weshalb (5) und daher auch (4) darin keine Lösung hat. Aus Stetigkeitsgründen und wegen $f_0(0) = f_{11}(0) = \sqrt{3}$, $f_0'(0) = \sqrt{8}+\sqrt{\frac{8}{3}} > 2+\frac{4}{\sqrt{3}} = f_{11}'(0)$ ist also

$$f_0(x) > f_{11}(x) \qquad (0 < x \leqq 2-\sqrt{3}). \tag{6}$$

2. Es sei nun $2-\sqrt{3} < x < 1-\frac{1}{\sqrt{2}}$. Die Funktion

$$\varphi(\alpha) = \frac{\sin 3\alpha}{\cos\alpha} = \frac{\sin\alpha(4\cos^2\alpha-1)}{\cos\alpha}$$

nimmt die Werte $\varphi\left(\frac{\pi}{8}\right) = \varphi\left(\frac{\pi}{4}\right) = 1$ an und besitzt bei γ, bestimmt durch

$$\cos^2\gamma = \frac{1+\sqrt{3}}{4} \qquad \left(\frac{\pi}{8} < \gamma < \frac{\pi}{4}\right)$$

ein Maximum; in $\left[\frac{\pi}{8}, \gamma\right]$ und in $\left[\gamma, \frac{\pi}{4}\right]$ verläuft sie monoton. Die Gleichung

$$\varphi(\alpha) = \varphi(\beta) \qquad \left(\frac{\pi}{8} < \alpha < \beta < \frac{\pi}{4}\right) \tag{7}$$

hat daher zu jedem β aus $\left(\gamma, \frac{\pi}{4}\right]$ genau eine Lösung α aus $\left[\frac{\pi}{8}, \gamma\right)$. Läuft β wachsend von γ bis $\frac{\pi}{4}$, so läuft α abnehmend von γ bis $\frac{\pi}{8}$. Man erhält den Zusammenhang explizit, indem man (7) auf die gleichwertige Form bringt

$$\cos\alpha\cos\beta\cos(\alpha+\beta) = \tfrac{1}{4} \tag{8}$$

und nach $\sin^2\alpha = u$ mit $\sin^2\beta = v$ auflöst:

$$u = \frac{1}{4}\left[3 - 2v - \sqrt{\frac{3v-4v^3}{1-v}}\right]. \tag{9}$$

Daraus ersieht man leicht, daß

$$x = 2(u+v) - 1 \tag{10}$$

monoton von $2-\sqrt{3}$ bis $1-\frac{1}{\sqrt{2}}$ läuft, wenn β das Intervall $\left[\gamma, \frac{\pi}{4}\right]$ durchläuft. Aus (9) und (10) folgt

$$u+v = \frac{x+1}{2}, \qquad uv = \frac{(2x-1)^2}{8x}. \tag{11}$$

Verwendet man (8), so ergibt sich

$$(\sin^3\alpha\cos\alpha + \sin^3\beta\cos\beta)^2 = \tag{12}$$

$$= u^3 + v^3 - (u^4+v^4) + 2uv\,(uv - u - v + \tfrac{3}{4}).$$

Zieht man schließlich (12) heran, so erhält man f_{12} explizit durch x ausgedrückt:

$$f_{12}(x) = 2\sqrt{-x^4+6x^3-12x^2+8x} + 2x(1+x). \tag{13}$$

a)

$$f_{11}(x) > f_{12}(x) \left(2-\sqrt{3} < x \leqq 1-\frac{1}{\sqrt{2}}\right) \tag{14}$$

ist nach (2′) und (13) gleichwertig mit

$$3x^4-24x^3+54x^2-24x+3 = 3(x^2-4x+1)^2 \geqq 0,$$

worin Gleichheit nur für $x=2\pm\sqrt{3}$ gilt.

b)

(15) $$f_0(x)=f_{12}(x)$$

ist nach (2) und (13) gleichwertig mit

(16) $$(1+\sqrt{2}\,x)\sqrt{3-\sqrt{8}\,x-2x^2}+2x^2+2(\sqrt{2}-1)x = 2\sqrt{-x^4+6x^3-12x^2+8x}$$

und daher auch mit

(17) $$\begin{aligned} &4(1+\sqrt{2}\,x)[x^2+(\sqrt{2}-1)x]\sqrt{3-\sqrt{8}\,x-2x^2} = \\ &= -4x^4+32x^3+8(-7+\sqrt{2})x^2+4(8-\sqrt{2})x-3. \end{aligned}$$

Durch Quadrieren folgt

(18) $$\begin{aligned} g(x) \equiv\ & -80x^8+128(3-2\sqrt{2})x^7+64(-35+7\sqrt{2})x^6+32(140-31\sqrt{2})x^5+ \\ & +8(-653+144\sqrt{2})x^4+32(111-21\sqrt{2})x^3+16(-78+13\sqrt{2})x^2+ \\ & +24(8-\sqrt{2})x-9 = 0. \end{aligned}$$

Bildet man die Ableitungen von $g(x)$, so sieht man, daß

$$g^{(6)}(x) < 0, \quad g^{(5)}\left(1-\frac{1}{\sqrt{2}}\right) > 0, \quad g^{(4)}\left(1-\frac{1}{\sqrt{2}}\right) < 0,$$

$$g'''\left(1-\frac{1}{\sqrt{2}}\right) > 0, \quad g''(2-\sqrt{3}) > 0, \quad g'\left(1-\frac{1}{\sqrt{2}}\right) = g\left(1-\frac{1}{\sqrt{2}}\right) = 0$$

und somit $g^{(5)}(x)>0$, $g^{(4)}(x)<0$, $g'''(x)>0$, $g''(x)>0$, $g'(x)<0$ und schließlich $g(x)>0$ in $\left(2-\sqrt{3}\,,\ 1-\frac{1}{\sqrt{2}}\right)$ ist. Daraus folgert man, daß (18) und daher auch (15) in diesem Intervall nicht gelten kann. Aus Stetigkeitsgründen und wegen $f_0(2-\sqrt{3}) > f_{11}(2-\sqrt{3}) = f_{12}(2-\sqrt{3})$ ist daher

(19) $$f_0(x) > f_{12}(x) \quad \left(2-\sqrt{3} \leqq x < 1-\frac{1}{\sqrt{2}}\right).$$

3. Es sei $1-\frac{1}{\sqrt{2}} \leqq x < 1$.

(20) $$f_{11}(x) > f_2(x)$$

ist gleichwertig mit

(21) $$\omega(x) \equiv -x^4+2x^2+4(3-2\sqrt{2})x-6+4\sqrt{2} > 0.$$

Wegen

$$\omega'(x) = 4(-x^3 + x + 3 - 2\sqrt{2}) > 0$$

und

$$\omega\left(1 - \frac{1}{\sqrt{2}}\right) = \frac{51}{4} - 9\sqrt{2} > 0.$$

ist (21) und damit auch (20) richtig.

(6), (14), (19) und (20) zeigen, daß aus (1) tatsächlich der behauptete Satz folgt.

LITERATURVERZEICHNIS

[1] FEJES TÓTH, L.: On the arrangement of houses in a housing estate, *Studia Sci. Math. Hungar.* **2** (1967) 37—42.

[2] FEJES TÓTH, L.: *Regular Figures*, Pergamon Press, Oxford 1964.

2. Institut für Mathematik der Technischen Hochschule, Wien.

(Eingegangen: 20. August, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 365—370.

ON APPROXIMATION BY POSITIVE LINEAR METHODS, II.

by

G. FREUD

1. Formulation of the Theorem

We are generalising the result of our previous paper[1] ("Part I") using as test functions an arbitrary ČEBYŠEV system in place of $\{1, \cos x, \sin x\}$. This paper is readable without Part I. Let $\{A_n\}$ be a sequence of bounded[2] linear operators transforming the space C of functions continuous on $[0, 1]$ into itself.

The norm of C is — as usual —

$$\|f\| = \max_{x \in [0,1]} |f(x)|.$$

We suppose the A_n's to be positive in KOROVKIN's notation, i.e. from $f(x) \geqq 0$ $(0 \leqq x \leqq 1)$ follows $(A_n f)(x) \geqq 0$ $(0 \leqq x \leqq 1)$. Further let $u_0(x)$, $u_1(x)$, $u_2(x)$ be two times continuously differentiable functions forming a strong ČEBYŠEV system, i.e. no u-polynomial $a_0u_0(x)+a_1u_1(x)+a_2u_2(x)$ has more than two zeros in $[0, 1]$ even if zeros are counted with multiplicity.

We denote by $m \geqq 2$ an integer and by $\{\lambda_n\}$ a decreasing sequence of positive numbers with $\lambda_1 \leqq 1$. In what follows let $c_1, c_2, \ldots$ be positive numbers depending on m, on the sequences $\{A_n\}$ and $\{\lambda_n\}$ and on the system $\{u_0, u_1, u_2\}$ only. Our aim is to prove the following

THEOREM. *From the assumption*

$$\|A_n u_i - u_i\| \leqq \lambda_n^m \qquad (i=0, 1, 2;\ n=1, 2, \ldots) \tag{1}$$

[1] On approximation by positive linear methods, I, *Studia Sci. Math. Hungar.* **2** (1967) 63—66.

As to comments on the history of the subject (including quotations), we refer to Part I. A correction must be made in the quotation [1]: H. BOHMAN *Arkiv för Matematik* **2** (1952) 43—52. In addition to the publications of H. BOHMAN and P. P. KOROVKIN we mention here the paper T. POPOVICIU: Asupra demonstratiei teoremei lui Weierstrass cu ajutorul polinoamelor de interpolare, *Lucrările Sesiuni Gen. Sci. Ac. R. P. R.* 1950, 1664—1667.

In this paper — as well as in BOHMAN's paper — positive operators of the form

$$(A_n f)(x) = \sum_{i=0}^{n} P_{n,i}(x) f(x_{ni})$$

are studied. It is proved that

$$\|A_n f - f\| \leqq 2\omega(A_n)$$

where

$$A_n^2 = \sup_x \sum_{i=0}^{n} P_{n,i}(x)(x - x_{ni})^2.$$

From this statement BOHMAN's theorem and even important cases of KOROVKIN's theorems are easy consequences. Nevertheless it remains the merit of H. BOHMAN to have formulated and proved first a "finite test" condition.

[2] The bound may *a priori* depend on n (see Lemma 5).

follows for any arbitrary $f \in C$

$$\|A_n f - f\| \leqq c_1 \omega_m(f; \lambda_n) + c_2 \|f\| \lambda_n^m. \tag{2}$$

REMARK. Let us replace C by the space $C_{2\pi}$ of 2π-periodic continuous functions. Let $\{A_n\}$ be a sequence of bounded positive linear operators transforming $C_{2\pi}$ into itself, $\{u_i\,;\, i=0, 1, 2\}$ be a strong Čebyšev system of twice differentiable 2π-periodic functions (i.e. for no u-polynomial does the number of zeros in $[0, 2\pi)$ counted with multiplicity exceed two). Even in this modified case (2) is a consequence of (1). There is no need for any alteration in the proof. It is in this form that our theorem is an extension of the result in Part I (see [1]).

2. Lemmata on u-polynomials

We consider the u-polynomials in x

$$U_1(x_0, x) = \begin{pmatrix} u_0(x_0) & u_1(x_0) & u_2(x_0) \\ u_0'(x_0) & u_1'(x_0) & u_2'(x_0) \\ u_0''(x_0) & u_1''(x_0) & u_2''(x_0) \end{pmatrix}^{-1} \begin{vmatrix} u_0(x) & u_1(x) & u_2(x) \\ u_0''(x_0) & u_1''(x_0) & u_2''(x_0) \\ u_0(x_0) & u_1(x_0) & u_2(x_0) \end{vmatrix} \tag{3}$$

and

$$U_2(x_0, x) = \pm \begin{vmatrix} u_0(x) & u_1(x) & u_2(x) \\ u_0(x_0) & u_1(x_0) & u_2(x_0) \\ u_0'(x_0) & u_1'(x_0) & u_2'(x_0) \end{vmatrix} \tag{4}$$

where the sign "$\pm$" is chosen (not depending on x) so that $U_2(x_0, x)$ is positive for at least one value of x. We regard x_0 as a parameter, the symbol $\ll' \gg$ means derivation with respect to x.

LEMMA 1. *The coefficients* $a_i(x_0)$ $(i=0, 1, 2)$ *of*

$$U_1(x_0, x) = a_0(x_0)u_0(x) + a_1(x_0)u_1(x) + a_2(x_0)u_2(x)$$

have bounds not depending on x_0. *Further, we have*

$$U_1(x_0, x_0) = 0, \quad U_1'(x_0, x_0) = 1 \tag{5}$$

and

$$|U_1''(x_0, x)| \leqq c_5. \tag{6}$$

PROOF. Formulas (5) follow from the definition.

If the derivate of the u-polynomial

$$V_1(x_0, x) = \begin{vmatrix} u_0(x) & u_1(x) & u_2(x) \\ u_0''(x_0) & u_1''(x_0) & u_2''(x_0) \\ u_0(x_0) & u_1(x_0) & u_2(x_0) \end{vmatrix}$$

would vanish at the point $x = x_0$, then this u-polynomial would have against our assumption a triple zero at $x = x_0$. So we have $V_1'(x_0, x_0) > 0$ and by a continuity argument $V_1'(x_0, x_0) \geqq c_4 > 0$, so that the coefficients of $U_1(x_0, x) = [V_1'(x_0, x_0)]^{-1} V_1(x_0, x)$ are bounded. It follows that (6) is satisfied, q.e.d.

LEMMA 2. The coefficients $b_i(x_0)$ $(i=0, 1, 2)$ *of* $U_2(x_0, x)=b_0(x_0)u_0(x)+$ $+b_1(x_0)u_1(x)+b_2(x_0)u_2(x)$ *have bounds not depending on* x_0. *We have further*

$$U_2(x_0, x_0)=U_2'(x_0, x_0)=0 \tag{7}$$

and

$$U_2(x_0, x)\geqq c_6(x-x_0)^2. \tag{8}$$

PROOF. (7) is a consequence of the definition, as well as the boundedness of the $b_i(x_0)$ $(i=0, 1, 2)$.

As a consequence of (7) $U_2(x_0, x)$ has constant sign for $x\neq x_0$, so that for $x\neq x_0$ $U_2(x_0, x)>0$. Now by a continuity argument, the choice of the sign $\pm$ in (4) depends neither on x_0 nor on x.

Again by the Čebyšev condition we have $U_2''(x_0, x_0)>0$ and by continuity $U_2''(x_0, x_0)>2c_7>0$. By a proper choice of $\delta>0$ we have

$$U_2''(x_0, x)\geqq 2c_7>0 \quad \text{for} \quad |x-x_0|\leqq\delta. \tag{9}$$

Outside this strip we have as a consequence of continuity

$$U_2(x_0, x)\geqq c_8>0 \quad \text{for} \quad |x-x_0|\geqq\delta. \tag{10}$$

From (7), (9) and (10) we conclude that (8) is satisfied with $c_6=\min(c_7, c_8)$, q.e.d.

LEMMA 3. *There exists an u-polynomial*

$$U_0(x) = \alpha_0u_0(x)+\alpha_1u_1(x)+\alpha_2u_2(x)$$

for which

$$U_0(x)\geqq 1 \qquad (0\leqq x\leqq 1). \tag{11}$$

PROOF. We take

$$U_0(x) = 8c_6^{-1}\,[U_2(\tfrac{1}{4}, x)+U_2(\tfrac{3}{4}, x)]$$

and apply (8).

3. Reduction of the Proof to a Special Case

LEMMA 4. *If our Theorem is true for* $m=2$, *then it is also valid for every integer* $m\geqq 2$.

PROOF. From assumption (1) and the validity of our theorem for $m=2$ we conclude (inserting $m=2$ and replacing λ_n by $\lambda_n^{m/2}$)

$$\|A_nf-f\|\leqq c_9\omega_2(f;\lambda_n^{m/2})+c_{10}\|f\|\lambda_n^m. \tag{12}$$

Now by MARCHAUD's inequality (see A. MARCHAUD [2] or A. F. TIMAN [3], III. 3. 3)

$$\omega_2(f;\lambda_n^{m/2})\leqq c_{10}\omega_m(f;\lambda_n)+c_{11}\|f\|\lambda_n^m. \tag{13}$$

From (12) and (13) we obtain (2), q.e.d.

LEMMA 5. *From* (1) *follows that the sequence* $\{A_n\}$ *is bounded in norm.*

PROOF. Let $\|f\|\leqq 1$. It follows

$$-U_0(x)\leqq f(x)\leqq U_0(x)$$

so that

$$-(A_n U_0)(x) \leqq (A_n f)(x) \leqq (A_n U_0)(x)$$

and by (1) and Lemma 1

$$\|A_n f\| \leqq \|A_n U_0\| \leqq \|U_0\| + c\lambda_n^m \leqq \|U_0\| + \lambda_1^m \leqq \|U_0\| + c$$

i.e.

$$\|A_n\| \leqq c_{12} \qquad (n = 1, 2, \ldots) \tag{14}$$

q.e.d.

LEMMA 6. *Our Theorem is valid, if the following statement is true:*

If (1) *is true with* $m = 2$ *then for every function* $f(x)$ *admitting a continuous second derivative* $f''(x)$ *in* [0, 1] *we have*

$$\|A_n f - f\| \leqq c_{13}(\|f\| + \|f''\|)\lambda_n^2. \tag{15}$$

PROOF. Let us assume that the statement is already proved and let $f \in C$. As was proved in the previous paper [1] of the author, we can find for every positive integer v, a twice continuously differentiable function $\gamma_v(x)$ so that

$$\|f - \gamma_v\| \leqq 2\omega_2(f; v^{-1}) \quad \text{and} \quad \|\gamma_v''\| \leqq 10v^2\omega_2(f; v^{-1}). \tag{16}$$

From the statement — which we assumed to hold — we obtain using (16)

$$\|A_n\gamma_v - \gamma_v\| \leqq c_{13}[\|f\| + 2\omega_2(f; v^{-1}) + 10v^2\omega_2(f; v^{-1})]\lambda_n^2. \tag{17}$$

Now let us choose v so that

$$(2v)^{-1} \leqq \lambda_n < v^{-1}. \tag{18}$$

From (14), (16), (17) and (18) we conclude

$$\|A_n f - f\| < \|A_n(f - \gamma_v)\| + \|A_n\gamma_v - \gamma_v\| + \|\gamma_v - f\| \leqq 2c_{12}\omega_2(f; v^{-1}) +$$

$$+ c_{13}\|f\|\lambda_n^2 + 12c_{13}v^2\lambda_n^2\omega_2(f; v^{-1}) + 2\omega_2(f; v^{-1}) \leqq c_{14}\omega_2(f; \lambda_n) + c_{13}\|f\|\lambda_n^2,$$

i.e. our Theorem is valid for $m = 2$. Then by Lemma 4 it is also valid for every integer $m \geqq 2$, q.e.d.

4. Proof of the Theorem

Taking lemma 4 and 6 into account we see that the only thing which remained to prove is the "statement" in Lemma 6. Now let $f(x)$ $(0 \leqq x \leqq 1)$ admit a second continuous derivative.

LEMMA 7. *We have*

$$\|f'\| \leqq 2(\|f\| + \|f''\|). \tag{19}$$

PROOF. By Lagrange's mean value theorem there exists a $\xi \in (0, 1)$ with $f(1) - f(0) = f'(\xi)$. Then we have

$$|f'(x)| = \left|f'(\xi) + \int_{\xi}^{x} f''(t)\,dt\right| \leqq |f(0)| + |f(1)| + |x - \xi|\,\|f''\| \leqq 2(\|f\| + \|f''\|),$$

q.e.d.

Let us consider the u-polynomial

$$U(f;x_0,x)=\frac{f(x_0)}{U_0(x_0)}U_0(x)+\left[f'(x_0)-\frac{f(x_0)}{U_0(x_0)}U_0'(x_0)\right]U_1(x_0,x)\equiv$$

(20)

$$\equiv\beta_0(f;x_0)u_0(x)+\beta_1(f;x_0)u_1(x)+\beta_2(f;x)u_2(x).$$

By Lemma 1, Lemma 3 and (19) we have

$$|\beta_0(f;x_0)|+|\beta_1(f;x_0)|+|\beta_2(f;x_0)|\leqq c_{15}(\|f\|+\|f''\|) \tag{21}$$

so that

$$|U''(f;x_0,x)|\leqq c_{16}(\|f\|+\|f''\|) \tag{22}$$

and by direct calculation

$$U(f;x_0,x_0)=f(x_0)\qquad U'(f;x_0,x_0)=f'(x_0). \tag{23}$$

It follows from (22) and (23)

$$|f(x)-U(f;x_0,x)|\leqq c_{17}(\|f\|+\|f''\|)(x-x_0)^2. \tag{24}$$

We obtain from (8)

$$U(f;x_0,x)-c_6^{-1}c_{17}(\|f\|+\|f''\|)U_2(x_0,x)\leqq f(x)\leqq \tag{25}$$

$$\leqq U(f;x_0,x)+c_6^{-1}c_{17}(\|f\|+\|f''\|)U_2(x_0,x).$$

We have from (1) (with $m=2$), (20), (21) and (23)

$$|[A_nU(f;x_0)](x_0)-f(x_0)|=$$

$$=|[A_nU(f;x_0)](x_0)-U(f;x_0,x_0)|\leqq \tag{26}$$

$$\leqq c_{15}(\|f\|+\|f''\|)\lambda_n^2$$

and by Lemma 2 and (1) (with $m=2$) we have

$$|[A_nU_2(x_0)](x_0)|=|[A_nU_2(x_0)](x_0)-U_2(x_0,x_0)|\leqq \tag{27}$$

$$\leqq(|b_0(x_0)|+|b_1(x_0)|+|b_2(x_0)|)\lambda_n^2\leqq c_{18}\lambda_n^2.$$

On behalf of the positivity of the operator A_n, we can apply it to the inequality (25). We insert $x=x_0$ after this application, and rearrange the terms; in this way we obtain

$$|(A_nf)(x_0)-[A_nU_n(f;x_0)]|\leqq c_6^{-1}c_{17}(\|f\|+\|f''\|)[A_nU_2(x_0)](x_0). \tag{28}$$

Finally from (26), (27) and (28) we have

$$|(A_nf)(x_0)-f(x_0)|\leqq c_{19}(\|f\|+\|f''\|)\lambda_n^2 \tag{29}$$

q.e.d.

REFERENCES

[1] FREUD, G.: Sui procidimenti lineari d'approssimazione, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* **26** (1959) 641—643.
[2] MARCHAUD, A.: Sur les derivées et sur les différences des fonctions de variables réelles, *Journ. Math. Pures et Appl.* **6** (1927) 337—425.
[3] ТИМАН А. Ф. *Теория приближения функций действительно переменного*, Гос, Изд. Физ.—Мат. Лит. Москва, 1960.

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received September 9, 1967.)

Studia Scientiarum Mathematicarum Hungarica 3 (1968) 371—372.

CORRIGENDA TO MY PAPER "SOME NEW RESULTS IN THE THEORY OF STABILITY" IN T. 2. F. 3—4. (1967) PP. 363—383.

by

T. FREY

In the paper mentioned above there is a very serious defect in connection with Theorems 1. and 2.: e.g. matrix norms invariant with respect to similarity transformations do not exist at all.[1] Consequently, in the relations and conditions figuring in § 2, i.e. in Theorems 1 and 2 as well, $\|\mathbf{B}(\tau)\|$ should be replaced by the so called transformed norm $\|\mathbf{B}(\tau)\|_{\mathbf{Y}(\tau)} = \|\mathbf{Y}^{-1}(\tau)\mathbf{B}(\tau)\mathbf{Y}(\tau)\|$, where $\|,\|$ may denote any arbitrary matrix-norm, e.g. the spectral norm. This modification, however, definitely weakens these two theorems. E.g. Theorem 3 usually cannot be proved by Theorem 2, what is more, we can prove it only in a slightly weakened form.

Acccording to the modified Theorem 3. — if λ_j is an eigen-value of matrix $\mathbf{A}$ with a multiplicity of $l(j)$ — we can prove only, that for an arbitrary $0 \leqq i \leqq l(j)-1$ and for the main vector $\mathbf{s}_j^{(i)}$ of matrix A we can find a solution $\mathbf{z}_j^{(i)}(t)$ of equation (28) for which (30)

$$\lim_{t\to\infty} \exp(-\varepsilon t)\left\{\mathbf{z}_j^{(i)}(t)\cdot\exp\left[-\int_{t_0}^{t}\lambda_j^{(i)}(s)ds\right]-\mathbf{s}_j^{(i)}\right\}=\mathbf{0}$$

is valid for any $\varepsilon>0$, also in the case $\mathbf{V}(t)\to\mathbf{O}$ if $t\to\infty$, but Var $(\|\mathbf{V}(t)\|)=o(t)$.

The proof follows the usual line of argument with the following modification: for the final estimations instead of Theorem 2 we have to follow the methods of [4] and [7] resp. So we mention first of all that from the estimation in (33) follows, that the condition of Theorem 8.1. in [4] concerning $D_{ki}=\operatorname{Re}[\lambda_k(t)-\lambda_i(t)]$ is a necessary consequence of the boundedness of Var $(\|\mathbf{V}(t)\|)$, for eigenvalues $\operatorname{Re}\lambda_k \neq \operatorname{Re}\lambda_i$ of $\mathbf{A}_i$: moreover $\int_{t_0}^{t} D_{ki}(\tau)\,d\tau \geqq Ct$ is valid, if $\int_{t_0}^{t}\operatorname{Re}\lambda_k > \operatorname{Re}\lambda_i$, whereas $\left|\int_{t_0}^{t} D_{ki}(\tau)\,d\tau\right| = o(t)$, for multiple eigen-values, and eigen-values with the same real part of $\mathbf{A}$.

For proving our statement we can follow the line of argument of the proof in [4]. (In

[1] Here I should like to express my gratitude to Prof. *Juan Jorge Schäffer* (Universidad de Montevideo), who was so kind as to send me a letter with the following counterexample:

$$\mathbf{A}(t)=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \mathbf{B}(t)=e^{-t}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix},\quad \mathbf{Y}(t)=\begin{pmatrix}e^{t} & 0\\ 0 & e^{-t}\end{pmatrix}$$

$$\mathbf{X}(t)=\begin{pmatrix}e^{t} & 0\\ 1-e^{-t} & e^{-t}\end{pmatrix},\quad \text{e.g.}\quad \mathbf{Y}^{-1}(t)\cdot\mathbf{X}(t)-\mathbf{E}=(e^{t}-1)\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}.$$

the following we will use also the notations of this book.) The modifications are as follows: for the multiple eigen-values of $\mathbf{A}$ (e.g. for λ_j) we divide $\Psi(t)$ into 3 parts: $\Psi_1(t)$ contains the fundamental solutions belonging to the eigen-values $\lambda(t)$ for which $\int_{t_0}^{t} \mathrm{Re}\,[\lambda(s)-\lambda_j^{(i)}(s)]\,ds = O(t)$, $\Psi_2(t)$ contains those for which $\lambda(t) \nrightarrow \lambda_j$, but $\int_{t_0}^{t} \mathrm{Re}\,[\lambda(s)-\lambda_j^{(i)}]\,ds = O(-t)$ is valid. Finally, $\Psi_3(t)$ contains the fundamental solutions belonging to the elements of the third block. $\Psi_1(t)$ and $\Psi_3(t)$ determine the integral over (t_0, t) and $\Psi_2(t)$ over (t, ∞). Namely, in our case $\Psi(t)$ and $\Psi^{-1}(\tau)$ are not diagonal matrices, nor is that their product — they are matrices of the type of upper triagonal, whose elements in the main diagonal are of the form $\exp\left(\int_{\tau}^{t} \lambda(s)\,ds\right)$, and the upper elements can be estimated by expressions of the form $e^K \cdot \frac{(t-\tau)^\nu}{\nu!} \cdot \exp\left(\int_{\tau}^{t} \lambda(s)\,ds\right)$. So when proving the existence and convergence of the sequences $\varphi^{(j)}(t)$, the parts generated by $\Psi_1(t)$ and $\Psi_3(t)$ can be handled together, but when discussing the asymptotic behaviour of the part generated by $\Psi_3(t)$ we need the factor $\exp(-\varepsilon t)$, too. The line of argument can be followed exactly.

We also mention that we have used the non-existent matrix-norm invariant with respect to similarity transformations for proving Theorems 4, 5, 6 in § 3.

But the correction of these theorems needs very strong modifications of the formulation and proof, which will be presented in a following paper (Corrigenda II).

Computing Centre of the Hungarian Academy of Sciences, Budapest

(Received January 15, 1968.)

INDEX

Printed in Hungary

A kiadásért felel az Akadémiai Kiadó igazgatója — Műszaki szerkesztő: Farkas Sándor
A kézirat nyomdába érkezett: 1968. II. 16. — Terjedelem: 32,75 (A/5) ív, 30 ábra

68-5869 — Szegedi Nyomda

Die *Studia Scientiarum Mathematicarum Hungarica* ist eine Halbjahrsschrift der Ungarischen Akademie der Wissenschaften. Sie veröffentlicht Originalbeiträge aus dem Bereiche der Mathematik in deutscher, englischer, französischer oder russischer Sprache. Es erscheint jährlich ein Band.

Adresse der Redaktion: Budapest V., Reáltanoda u. 13—15, Ungarn.
Technischer Redaktor: Gy. Katona.

Abonnementspreis pro Band (pro Jahr): 165.— Ft. 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 (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: Budapest V. Reáltanoda u. 13—15, Hongrie.
Rédacteur téchnique: Gy. Katona

Le prix de l'abonnement: 165 Forints par an (volume). On s'abonne chez *Kultúra*, Société pour le Commerce de Livres et Journaux (Budapest 62, P.O.B. 149) ou chez ses représentantsàt l.étranger.
Pour établir des relations d'échange on est prié de s'adresser à la Biblioth que de l'Institut de Mathématique (Budapest V., Reáltanoda u. 13—15).
Onfest 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 немецком, английском, французском и русском языках. Отдельные выпуски составляют ежегодно один том.

Адрес редакции: Budapest V., Reáltanoda u. 13—15, Венгрия.
Технический редактор: Gy. Katona.

Подписная цена на год (за один том): 165 форинтов. Подписка на журнал принимается Внешнеторговым предприятием „Культура" (Budapest 62, P. O. B. 149) или его представительствами за границей.
По поводу отношения обмена просим обращаться к Библиотеке Института Математики (Budapest V., Reáltanoda u. 13—15).
Работы, предназначенные для опубликования в журнале следует направлять по адресу редакции в двух экземплярах.

Ára: 141,— Ft

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 Libirairie
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.

CHECHOSLOVAKIA

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 Buch-export und Import
Leninstraße 16.
Leipzig C. I.
Zeitungsvertriebsamt
Clara Zetkin Straße 62.
Berlin N. W.

GERMAN FEDERAL REPUBLIC

Kunst und Wissen
Erich Bieber
Postfach 46.
7 Stuttgart S.

GREAT BRITAIN

Collet's' Subscription Dept.
44—45 Museum Streetf
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 Pffice:
India House OPP.
GPO Post Box 1374f
Bombay I.

ITALY

Santo Vanasia
71 Via M. Macchi
Milano
Libreria Commissionaria Sansoni
Via La Marmora 45
Firenze

JAPAN

Nauka Ltd.
2 Kanada-Zimbocho 2-chome
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
Importu Proizvedenij Pechati
Phenjan

NORWAY

Johan Grundt 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

VIII. 10. 1968. **Index: 26 748**