315.930

22/1987

Studia Scientiarum Mathematicarum Hungarica

OLUME 22
UMBERS 1—4
37

AKADÉMIAI KIADÓ, BUDAPEST

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

A QUARTERLY OF THE HUNGARIAN ACADEMY OF SCIENCES

Studia Scientiarum Mathematicarum Hungarica publishes original papers on mathematics mainly in English, but also in German, French and Russian.

Studia Sciantiarum Mathematicarum Hungarica is published in yearly volumes of four issues (mostly double numbers published semiannually) by

AKADÉMIAI KIADÓ
Publishing House of the Hungarian Academy of Sciences
H-1054 Budapest, Alkotmány u. 21.

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor
P.O. Box 127
H—1364 Budapest

Subscription information

Orders should be addressed to

KULTURA Foreign Trading Company
P.O. Box 149
H—1389 Budapest

or to its representatives abroad.

Studia Scientiarum Mathematicarum Hungarica

Volume 22

Akadémiai Kiadó, Budapest
1987

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Volume 22

CONTENTS

86-3637—Szegedi Nyomda — Felelős vezető: Surányi Tibor igazgató

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 1—9

ON THE VIBRATION OF A STRING

I. JOÓ

Dedicated to Professor L. Fejes Tóth on the occasion of his 70th birthday

1. The necessity of the diophantine approximation in some problems of the control theory of distributed parameter systems was noted by A. G. Butkovskiĭ [1]. He investigated the system

$$\text{(1)}\qquad \begin{cases} \dfrac{\partial^2 y}{\partial t^2} = \dfrac{\partial^2 y}{\partial x^2} + \delta(x-a)\cdot u(t), \\ y(0, t) = y(1, t) = 0, \\ 0 < x < 1, \quad 0 < t < T, \quad 0 < a < 1. \end{cases}$$

It was pointed out in [1] by the help of diophantine approximations, which points $a \in (0, 1)$ are points of controllability of the system (1). This means that for any initial data (from some function space) we can obtain $y(\cdot, T) = y_t(\cdot, T) = 0$.

The aim of the present paper is to continue the work of A. G. Butkovskiĭ and describe the set of reachability of the system (1) for different T.

2. In this paper we consider the system more general than (1)

$$\text{(2)}\qquad \begin{cases} \varrho(x)\dfrac{\partial^2 y}{\partial t^2} = \dfrac{\partial^2 y}{\partial x^2} + \delta(x-a)u(t), \\ y(0, t) = y(1, t) = 0, \\ 0 < x < 1, \quad 0 < t < T, \quad 0 < a < 1, \end{cases}$$

with zero initial conditions

$$\text{(3)}\qquad y(\cdot, 0) = y_t(\cdot, 0) = 0,$$

where $\varrho \in C^2[0, 1]$, $\varrho(x) > 0$ $(0 \leqq x \leqq 1)$, $u \in L^2(0, T)$.

We say that $y \in L^2([0, 1] \times [0, T])$ is a solution of (2)—(3) if for any $z \in C^2([0, 1] \times [0, T])$ such that $z(0, t) = z(1, t) = 0$ $(0 \leqq t \leqq T)$, $z(\cdot, T) = z_t(\cdot, T) = 0$ the equation

$$\text{(4)}\qquad \int_0^1 \int_0^T y(\varrho z_{tt} - z_{xx})\, dx\, dt = \int_0^T z(a, t) u(t)\, dt$$

holds.

1980 *Mathematics Subject Classification*. Primary 93C20; Secondary 35C10.
Key words and phrases. String, control, reachability set, eigenfunction expansion.

THEOREM 1. *For any $u\in L^2(0,T)$ the system (2)—(3) has a unique solution y in $L^2([0,1]\times[0,T])$. This solution belongs to $H^1([0,1]\times[0,T])$ and has such a representation for which the phase trajectory $(y(\cdot,t), y_t(\cdot,t))$ of the system (2)—(3) belongs to $W_2^1(0,1)\oplus L^2(0,1)(=:Y)$ for every $t\in[0,T]$ and the mapping $t\mapsto(y(\cdot,t), y_t(\cdot,t))$: $[0,T]\to Y$ is continuous. The same assertions hold when (3) is replaced by $y(x,0)\in H^1(0,1)$, $y_t(x,0)\in L^2(0,1)$ and the same proof works.*

PROOF. We are looking for $y(x,t)$ in the form

(5) $$y(x,t)=\sum_{n=1}^{\infty} c_n(t)\,v_n(x),$$

where $\{v_n(x)\}_1^\infty$ are the solutions of the boundary value problem

(6) $$\begin{cases} v_n''(x)+\lambda_n\varrho(x)v_n(x)=0, & (0\leqq x\leqq 1),\\ v_n(0)=v_n(1)=0, \end{cases}$$

and $c_n(t)\in L^2(0,T)$.

According to [2, Chapter II, §5, No 2] the functions $\{v_n\}$ form[1] a complete orthonormal system in $L^2(0,1)$, further by [2, Chapter II, §4, No 11] we have

(7) $$0\leqq\lambda_n=\left(n\frac{\pi}{h}\right)^2+O(1),\qquad h:=\int_0^1\sqrt{\varrho(x)}\,dx.$$

Choose $z(x,t):=v_n(x)b(t)$ with $b(t)\in C^2[0,T]$,

(8) $$b(T)=b'(T)=0$$

in (4); substitute (5) into (4), we get

(9) $$\int_0^T c_n(t)b''(t)\,dt+\lambda_n\int_0^T c_n(t)b(t)\,dt=v_n(a)\int_0^T b(t)u(t)\,dt,$$

i.e. $c_n(t)$ is the solution of

(10) $$c_n''(t)+\lambda_n c_n(t)=v_n(a)u(t)$$

in the distribution sense (or "weak sense" [3]). Put $\omega_n:=\sqrt{\lambda_n}$, $\omega_n:=-\omega_n$, $b_n:=i\omega_n c_{|n|}+c_{|n|}'$ (the derivative is used in distribution sense). Then we have

$$b_n'=i\omega_n c_{|n|}'+(v_{|n|}(a)u-\omega_n^2 c_{|n|})=v_{|n|}(a)u+i\omega_n b_n.$$

This equation is solvable also in classical sense, namely

(11) $$b_n(t)=b_n(0)e^{i\omega_n t}+\int_0^t v_{|n|}(a)u(\tau)e^{i\omega_n(t-\tau)}\,d\tau,$$

[1] After normalization.

and it is clear that the functions

$$c_n(t) = \frac{b_n(t) - b_{-n}(t)}{2i\omega_n}, \quad c_n'(t) = \frac{b_n(t) + b_{-n}(t)}{2} \tag{12}$$

are absolutely continuous, thus, integrating by parts in (9) two times, we get

$$0 = [c_n b']_0^T - [c_n' b]_0^T + \int_0^T (c_n'' + \lambda_n c_n - v_n(a)u)b = c_n'(0)b(0) - c_n(0)b'(0)$$

for every b in (8), hence

$$c_n(0) = c_n'(0) = 0 \quad (\forall n \in \mathbf{Z}). \tag{13}$$

Then by (12) we have $b_n(0)=0$ and from (11) we get

$$c_n(t) = v_n(a) \int_0^t u(\tau) \frac{\sin \sqrt{\lambda_n}\,(t-\tau)}{\sqrt{\lambda_n}} dt, \tag{14}$$

$$c_n'(t) = v_n(a) \int_0^t u(\tau) \cos \sqrt{\lambda_n}\,(t-\tau)\, d\tau. \tag{15}$$

To complete the proof of the existence of the solution of (2)—(3) consider the expansion of an arbitrary $z(x, t)$ in (4) with respect to the system $\{v_n(x)\}_1^\infty$ for any fixed $t \in [0, T]$:

$$z(x, t) = \sum_{n=1}^{\infty} d_n(t) v_n(x), \qquad (\text{in } L_\varrho^2(0, 1)), \tag{16}$$

where

$$d_n(t) := \int_0^1 z(x, t) v_n(x) \varrho(x) dx, \tag{17}$$

hence $d_n(t) \in C^2[0, T]$ and $d_n(T) = d_n'(T) = 0$ $(n = 1, 2, \ldots)$. We know (4) for any finite sum of the form (14). We prove (4) for $z(x, t) \in C^2([0, 1] \times [0, T])$. For any fixed $t \in [0, T]$ consider the expansion of $z_{xx}(x, t)/\varrho(x)$ with respect to the system $\{v_n(x)\}_1^\infty$ in $L_\varrho^2(0, 1)$:

$$z_{xx}(x, t) \sim \sum_{n=1}^{\infty} \tilde{d}_n(t) v_n(x) \varrho(x). \tag{18}$$

Instead of (18) we can write by (6):

$$z_{xx}(x, t) \sim -\sum_{n=1}^{\infty} d_n^*(t) \lambda_n \varrho(x) v_n(x) = \sum_{n=1}^{\infty} d_n^*(t) v_n''(x) \tag{19}$$

in $L^2([0, 1] \times [0, T])$, where $d_n^*(t) := -\tilde{d}_n(t)/\lambda_n$.

Hence, the series $\sum d_n^*(t) v_n(x)$ converges uniformly[2] on $[0, 1]$ in x for any fixed $t \in [0, T]$ to a function of the form $z(x, t) + p(t) + xq(t)$. Now we show that

[2] We use $|v_n(x)| \leq C$ $(0 \leq x \leq 1;\ n = 1, 2, \ldots)$ (cf. [4]).

$p(t)=q(t)=0$ for every $t\in[0, T]$. Indeed, $0=z(0, t)=-p(t)$ further $0=z(1, t)=$ $=-p(t)-q(t)$. Hence

$$z(x, t) = \sum_{n=1}^{\infty} d_n^*(t)v_n(x)$$

uniformly in x on $[0, 1]$ for every fixed $t\in[0, T]$, so by (17) we have $d_n=d_n^*$.

Summarizing our considerations: we have proved that one can differentiate the series (16) term by term in x two times and the resulting series converges to $z_{xx}(x, t)$ in $L^2([0, 1]\times[0, T])$-norm (by the Cauchy-condition of convergence). On the other hand, (17) implies $d_n''(t)=\int_0^1 z_{tt}(x, t)v_n(x)\varrho(x)\,dx$, i.e. we have $z_{tt}(x, t)=$ $=\sum_{n=1}^{\infty} d_n''(t)v_n(x)$ in the norm of $L^2([0, 1]\times[0, T])$. Finally,

$$z(a, t) = \sum_{n=1}^{\infty} d_n(t)v_n(a). \tag{20}$$

We know that $v_n(a)=O(1)$ $(n=1, 2, \dots)$ (cf. [4]) and $\int_0^T \sum_{n=1}^{\infty} |d_n(t)|^2\,dt<\infty$, hence, the series (20) converges in $L^2(0, T)$ in t. From our results it follows that (4) holds for any $z\in C^2([0, 1]\times[0, T])$ with the properties before (4). Therefore we have proved the existence of the solution of (2)—(3) in $L^2([0, 1]\times[0, T])$. The uniqueness follows from the fact that any $y\in L^2([0, 1]\times[0, T])$ can be written in the form (5), where the coefficients $c_n(t)$ are uniquely determined by (10) (at $c_n(0)=c_n'(0)=0$).

Now we prove that $(y(\cdot, t), y_t(\cdot, t))\in Y$ for any fixed $t\in[0, T]$. To this first remark that $\{v_n'(x)/\sqrt{\lambda_n}\}$ is an orthonormal system in $L^2(0, 1)$ (cf. (6)), secondly we prove to this that $y(x, t)\in H^1([0, 1]\times[0, T])$. Indeed, it is easy to see from (14) that $\int_0^T \sum_{n=1}^{\infty} |c_n(t)|^2\lambda_n\,dt<\infty$, hence, the series $\sum c_n(t)v_n'(x)$ converges in $L^2([0, 1]\times[0, T])$, therefore $y_x=\sum c_n(t)v_n'(x)$ in the distribution sense, i.e., for any $\varphi(x, t)\in C_0^\infty([0, 1]\times[0, T])$

$$\int_0^1\int_0^T \sum_{n=1}^{N} c_n(t)v_n'(x)\varphi(x, t)\,dx\,dt = -\int_0^1\int_0^T \sum_{n=1}^{N} c_n(t)v_n(x)\frac{\partial}{\partial x}\varphi(x, t)\,dx\,dt$$

and both of the sums are convergent in $L^2([0, 1]\times[0, T])$. By (15) $\int_0^T \sum_{n=1}^{\infty} |c_n'(t)|^2\,dt<\infty$, hence the series $\sum c_n'(t)v_n(x)$ converges in $L^2([0, 1]\times[0, T])$ and we get the distribution-equality $\sum c_n'(t)v_n(x)=y_t(x, t)$ just as in the case of y_x. Thus we have proved $y\in H^1([0, 1]\times[0, T])$[3]. Using this fact and that we can differentiate (5) term by term in the sense that the equality remains valid in $L^2([0, 1]\times[0, T])$ (this was proved

[3] We have also proved the relations $y(\cdot, t)\in H^1(0, 1)$, $y(x, \cdot)\in H^1(0, T)$.

above), we get for any fixed $t\in[0, T]$:

$$\|(y(\cdot, t), y_t(\cdot, t))\|_Y^2 \asymp$$

$$\asymp \|y(\cdot, t)\|_{L^2(0,1)}^2 + \|y_x(\cdot, t)\|_{L^2(0,1)}^2 + \|y_t(\cdot, t)\|_{L^2(0,1)}^2 \asymp$$

$$\asymp \sum_{n=1}^{\infty} |c_n(t)|^2 + \sum_{n=1}^{\infty} \lambda_n |c_n(t)|^2 + \sum_{n=1}^{\infty} |c_n'(t)|^2.$$

On the other hand we obtain from (14) and (15):

$$\sum_{n=1}^{\infty} \lambda_n |c_n(t)| + \sum_{n=1}^{\infty} |c_n'(t)|^2 =$$

$$= \sum_{n=1}^{\infty} |v_n(a)|^2 \left\{ \left| \int_0^t u(\tau) \sin \sqrt{\lambda_n}\, (t-\tau)\, d\tau \right|^2 + \left| \int_0^t u(\tau) \cos \sqrt{\lambda_n}\, (t-\tau)\, d\tau \right|^2 \right\}$$

for every $t\in[0, T]$, (we used also $v_n(a)=O(1)$ $(n=1, 2, \ldots)$). Thus we have proved $(y(\cdot, t), y_t(\cdot, t))\in Y$ and that the representation of the solution y given by the series (17) has a phase trajectory depending continuously on t follows from:

$$\|(y(\cdot, t_1), y_t(\cdot, t_1)) - (y(\cdot, t_2), y_t(\cdot, t_2))\|_Y \asymp$$

$$\asymp \sum_{n=1}^{\infty} \lambda_n |c_n(t_1) - c_n(t_2)|^2 + \sum_{n=1}^{\infty} |c_n'(t_1) - c_n'(t_2)|^2 < \varepsilon$$

(if $|t_1 - t_2|$ is small enough). This is an easy consequence of (14) and (15). Theorem 1 is proved.

Define

$$\mathcal{D}_a(T) := \{(y(\cdot, T), y_t(\cdot, T)):\ u\in L^2(0, T)\}.$$

This is the set of all possible states of the string at time T. Let

$$\mathcal{B}_a(T) := \{(\sqrt{\lambda_n}\, c_n(T), c_n'(T)):\ u\in L^2(0, T)\}.$$

THEOREM 2. *$\mathcal{B}_a(T)\subset l_2$. $\mathcal{D}_a(T)$ and $\mathcal{B}_a(T)$ are isomorphic with an isomorphism between l_2 and a closed subspace Z of Y.*

PROOF. We have proved above the first part of the statement, it remained only to show that $\mathcal{D}_a(T)$ and $\mathcal{B}_a(T)$ are isomorphic to a closed subspace of Y. Let

$$Z := \left\{\left(\sum_{n=1}^{\infty} d_n v_n(x),\ \sum_{n=1}^{\infty} \tilde{d}_n v_n(x)\right):\ \sum_{n=1}^{\infty} (\lambda_n d_n^2 + \tilde{d}_n^2) < \infty\right\}.$$

For $(f, g)\in Z$ we have $f'(x)=\sum d_n v_n'(x)$ in $L^2(0, 1)$ (derivatives are taken in distribution sense). Hence

$$\|(f, g)\|_Y^2 \asymp \|f\|_{L^2(0,1)}^2 + \|f'\|_{L^2(0,1)}^2 + \|g\|_{L^2(0,1)}^2 \asymp$$

$$\asymp \sum_{n=1}^{\infty} (\lambda_n d_n^2 + \tilde{d}_n^2),$$

which shows that Z and l_2 are isomorphic with the extended isomorphism between $\mathscr{D}_a(T)$ and $\mathscr{B}_a(T)$. It remained to prove that $Z=H\oplus L^2(0,1)$, where H is closed in $W_2^1(0,1)$[4]. Indeed, let $f_n=\sum_k c_k^{(n)}v_k$, $f=\sum_k c_k v_k$, $f_n \xrightarrow{W_2^1} f$, $f_n\in H$. Then

$$\sum_k c_k^{(n)} v_k \to \sum_k c_k v_k \quad (\text{in } L^2) \tag{21}$$

and

$$f_n' = \sum_k \sqrt{\lambda_k}\, c_k^{(n)} \frac{v_k'}{\sqrt{\lambda_k}} \to f' \quad (\text{in } L^2). \tag{22}$$

Hence f' belongs to the closed subspace generated by $\left\{\frac{v_k'}{\sqrt{\lambda_k}}\right\}$, thus $f'=$ $=\sum \tilde{d}_k v_k/\sqrt{\lambda_k}=:\sum \sqrt{\lambda_k}\, d_k \frac{v_k'}{\sqrt{\lambda_k}}$ (the sum converges in L^2). Consequently, by (22), we get $(f_n', v_k'/\sqrt{\lambda_k})=\sqrt{\lambda_k}\, c_k^{(n)}\to\sqrt{\lambda_k}\, d_k=\left(f', \frac{v_k'}{\sqrt{\lambda_k}}\right)$ and by (21) $(f_n, v_k\varrho)=c_k^{(n)}\to c_k=$ $=(f, v_k\varrho)$, $d_k=c_k$ and we proved that $f\in H$. Theorem 2 is proved.

We shall prove on the reachability set

THEOREM 3. *Let* $\varrho\equiv 1$, $a=p/q$, $(p,q)=1$, *then*

(1) $\mathscr{D}_a(T_1)=\mathscr{D}_a(T_2)$ *if* $2\frac{q-1}{q}\leqq T_1\leqq T_2$,

(2) $\mathscr{D}_a(T_1)\subsetneqq\mathscr{D}_a(T_2)$ *if* $T_1<T_2\leqq 2\frac{q-1}{q}$,

(3) $\mathscr{D}_a(T)$ *is closed and of infinite codimension in* $Z\oplus L^2$ *for every* $T<\infty$.

PROOF. By Theorem 2 it is enough to prove these properties for $\mathscr{B}_a(T)\subset l_2$.

(1) If $T\geqq 2$, then $\mathscr{B}_a(T)$ is maximal, i.e., according to $v_n(x)=\sin \pi n x$, $\lambda_n=n^2\pi$ we have for $q|n$ that $c_n'(T)=\pi n c_n(T)=0$ and for the coordinates $q\nmid n$ all system (c_n) can be obtained for which $\sum_{q\nmid n}(\pi^2 n^2 c_n^2+c_n'^2)<\infty$. Obviously, the sets $\mathscr{B}_a(T)$ are increasing with T, hence it is enough to prove

$$\mathscr{B}_a\left(2\frac{q-1}{q}\right)=\mathscr{B}_a(2).$$

We need the

LEMMA [Butkovskiĭ [5], Chapter 2, §3, Theorem 3]. *If* $g_i\in L^2(0,1)$ *are arbitrary real functions and* (a_i) *is an arbitrary real sequence, then there exists a* $u\in L^2(0,1)$ *with* $a_i=\int_0^1 g_i(x)\,u(x)\,dx$ *if and only if there is a constant* $c>0$ *such that for any*

[4] It is easy to see that $H=\{f\in H^1(0,1)\colon f(0)=f(1)=0\}$.

finite real sequence $\xi_1, \xi_2, \dots, \xi_n$ *the estimate*

$$\Big|\sum_{i=1}^{n} a_i\xi_i\Big| \leqq C\Big\|\sum_{i=1}^{n} g_i\xi_i\Big\|_{L^2(0,1)}$$

holds.

In our case the functions (g_i) runs over the set of functions $\{\sin \pi nt, \cos \pi nt\}_{q\nmid n}$ so it is enough to prove that

$$\int_0^2 \Big|\sum_{i=1}^{n} g_i\xi\Big|^2 dx \leqq c_q \int_0^{2(q-1)/q} \Big|\sum_{i=1}^{n} g_i\xi\Big|^2 dx,$$

because for any $(a_i)\in l_2$ we have

$$\Big|\sum_{i=1}^{n} a_i\xi_i\Big|^2 \leqq \|(a_i)\|_{l_2}^2\Big(\sum_{i=1}^{n} \xi_i^2\Big) =$$

$$= \|a_i\|_{l_2}^2 \int_0^2 \Big|\sum_{i=1}^{n} g_i\xi_i\Big|^2 \leqq c_q\|(a_i)\|_{l_2}^2 \int_0^{2(q-1)/q} \Big|\sum_{i=1}^{n} g_i\xi_i\Big|^2,$$

and the condition of the Lemma is fulfilled. It is easy to see that if $q\nmid k$, then

(23) $$\sin k\pi x + \sin k\pi\left(x+\frac{2}{q}\right) + \dots + \sin k\pi\left(x+2\,\frac{q-1}{q}\right) \equiv 0,$$

(24) $$\cos k\pi x + \cos k\pi\left(x+\frac{2}{q}\right) + \dots + \cos k\pi\left(x+2\,\frac{q-1}{q}\right) \equiv 0,$$

hence, for any finite sum we have

$$\left|\sum_{q\nmid k}\left[a_k \cos k\pi\left(x+2\,\frac{q-1}{q}\right) + b_k \sin k\pi\left(x+2\,\frac{q-1}{q}\right)\right]\right| \leqq$$

$$\leqq \Big|\sum_{q\nmid k}[a_k \cos k\pi x + b_k \sin k\pi x]\Big| + \dots$$

$$\dots + \left|\sum_{q\nmid k}\left[a_k \cos k\pi\left(x+2\,\frac{q-2}{q}\right) + b_k \sin k\pi\left(x+2\frac{q-2}{q}\right)\right]\right| \leqq$$

$$\leqq \sqrt{q-1}\,\Big\{\Big|\sum_{q\nmid k}[a_k \cos k\pi x + b_k \sin k\pi x]\Big|^2 + \dots$$

$$\dots + \left|\sum_{q\nmid k}\left[a_k \cos k\pi\left(x+2\,\frac{q-2}{q}\right) + b_k \sin k\pi\left(x+2\,\frac{q-2}{q}\right)\right]\right|^2\Big\},$$

therefore

$$\int_0^2 \Big|\sum_{q\nmid k}[a_k \cos k\pi x + b_k \sin k\pi x]\Big|^2 dx \leqq$$

$$\leqq q \int_0^{2(q-1/q)} \Big|\sum_{q\nmid k}[a_k \cos k\pi x + b_k \sin k\pi x]\Big|^2 dx,$$

and the statement (1) is proved.

We remark that the last inequality holds for infinite sums, too, i.e. if $u\in L^2(0,2)$ and $(u, \cos kq\pi\cdot)=(u, \sin kq\pi\cdot)=0$ $(k=1,2,\dots)$, then

$$\int_0^2 u^2(x)\,dx \le q\int_0^{2(q-1)/q} u^2(x)\,dx, \tag{25}$$

$\left(\text{and also } \le q\left(\int_0^{2(l/q)} + \int_{2(l+1)/q}^2\right)u^2(x)\,dx \text{ for arbitrary } l=0,1,\dots,q-1\right).$

(2) Let $T_1<T_2\le 2(q-1)/q$. We prove that the characteristic function $u_2\in L^2(0,T_2)$ of the segment (T_1,T_2) generates by (14), (15) an element of $\mathscr{B}_a(T_2)\setminus\mathscr{B}_a(T_1)$. Indirectly, suppose that there exists an $u_1\in L^2(0,T_1)$ which defines by (14), (15) at $t=T_1$ the same numbers as u_2 at $t=T_2$. Let $u\in L^2(0,T_2)\subset L^2(0,2)$ be a function, defined by

$$u(x):=\begin{cases}-u_1(x) & \text{for } x\in(0,T_1],\\ u_2(x)=1 & \text{for } x\in(T_1,T_2),\\ 0 & \text{for } x\in[T_2,2].\end{cases}$$

For this function $q\nmid n$ implies $(u,\cos n\pi\cdot)=(u,\sin n\pi\cdot)=0$, hence u is a $\frac{2}{q}$-periodic function, but $u\equiv 0$ on $\left(2\frac{q-1}{q},2\right)$ and this is a contradiction.

(3) It is trivial that $\mathscr{B}_a(T)$ is of infinite codimension. For $T\ge 2\frac{q-1}{q}$, $\mathscr{B}_a(T)$ is closed by (1). Let $T\le 2\frac{q-1}{q}$ and let $(u_n)\subset L^2(0,T)$ be such a sequence for which

$$\big((u_n,\cos k\pi\cdot),(u_n,\sin k\pi\cdot)\big)_{q\nmid k}\xrightarrow{l_2}\big((c'_{k,0}),(k\pi c_{k,0})\big)_{q\nmid k}. \tag{26}$$

We have to prove that the last sequence can be generated with some $u\in L^2(0,T)$ by (14), (15). It is enough to show that the sequence (u_n) is bounded in $L^2(0,T)$-norm, because in this case, taking any weak cluster point $u\in L^2(0,T)$ of (u_n) we have

$$(u,\cos k\pi\cdot)=c'_{k,0},\quad (u,\sin k\pi\cdot)=k\pi c_{k,0}$$

for every $q\nmid k$.

Let $\tilde u_n$ be the orthogonal projection of $u_n\in L^2(0,2)$ onto the closed subspace spanned by $\{\cos k\pi x, \sin k\pi x\}_{q\nmid k}$. Then $u_n-\tilde u_n$ is bounded in $L^2(0,2)$ by (26), further by (25) we have

$$\int_0^2 \tilde u_n^2(x)\,dx = q\int_{2(q-1)/q}^2 \tilde u_n(x)^2\,dx \le$$

$$\le q(q-1)\int_0^{2(q-1)/q}(u_n-\tilde u_n)^2\,dx \le q(q-1)\int_0^2 (u_n-\tilde u_n)^2$$

and therefore $(\tilde u_n)$ is bounded in $L^2(0,2)$, consequently (u_n) is bounded, too. Theorem 3 is proved.

REFERENCES

[1] Butkovskiĭ, A. G., Применение некоторых результатов теории чисел в проблеме финитного управления и управляемости в распределенных системах, (Application of some results of number theory to problems of compactly supported control and of controllability in distributed systems), *Dokl. Akad. Nauk SSSR* **227** (1976), No. 2, 309—311.

[2] Neumark, M. A., *Lineare Differentialoperatoren*, Mathematische Monographien, Band XI, Akademie-Verlag, Berlin, 1960. *MR* **35** # 6884.

[3] Brézis, H., *Analyse fonctionnelle*, Théorie et applications, Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris, 1983. *MR* **85a**: 46001.

[4] Joó, I., Upper estimates for the eigenfunctions of the Schrödinger operator, *Acta Sci. Math. (Szeged)* **44** (1982), 87—93. *MR* **84m**: 34027a.

[5] Butkovskiĭ, A. G., *Методы управления системами с распределенными параметрами*, Наука, Москва, 1975.

(Received December 28, 1984)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
POSTAFIÓK 323
H—1445 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 11—18

TORISCHE VARIETÄTEN UND KONVEXE POLYTOPE

GÜNTER EWALD

Herrn Prof. Dr. László Fejes Tóth zum 70. Geburtstag gewidmet

1. Einleitung

In den vergangenen zwölf Jahren ist eine neuartige Beziehung zwischen einer Klasse algebraischer Varietäten, genannt torische Varietäten, und konvexen Körpern, insbesondere Polytopen, verwendet worden und hat zu interessanten Einsichten und Ergebnissen für beide Bereiche geführt. Das Buch „Toroidal Embeddings" von Kempf, Knudsen, Mumford, St. Donat [15] von 1972 gewinnt sein Hauptresultat mit einem langwierigen kombinatorisch-geometrischen Beweis eines Satzes der kombinatorischen Konvexgeometrie, noch ohne Bezugnahme auf konvexgeometrische Arbeiten. (Möglicherweise läßt sich der Beweis vereinfachen.) Ein Ergebnis in Ehlers [8, 1.3] ist Wiederentdeckung eines Resultates von Bruggesser und Mani [5] über die Schälbarkeit sphärischer Komplexe, wenn auch in anderer Sprache. Und in Danilov [7, 10.7.1] wird ein Satz von Kind und Kleinschmidt [16] über Parametersysteme in Stanley—Reisner—Ringen erneut bewiesen.

Zu einem expliziten Zusammenwirken von algebraischen Geometern und Konvexgeometern ist es zuerst bei der Lösung der sogenannten upper bound conjecture gekommen, die aussagt, wieviele k-dimensionale Seiten ein konvexes n-Polytop maximal besitzen kann, $k=0, ..., n-1$, und welche Polytope maximale Seitenzahlen besitzen. Nachdem McMullen [19] 1970 die Vermutung mit konvexgeometrischen Methoden bewiesen hatte, verallgemeinerte Stanley [23] 1975 das Ergebnis auf beliebige sphärische Zellkomplexe und gab einen algebraisch-geometrischen Beweis an. (Vgl. hierzu auch Kleinschmidt [17].)

Im Gefolge der hierbei entwickelten Methoden konnten Stanley [24] und Billera—Lee [4] dann die McMullensche „g-Vermutung" beweisen, die die Anzahlvektoren für Seiten simplizialer konvexer Polytope kennzeichnet.

Bernstein [3] entdeckte 1974 eine überraschende Beziehung zwischen gemischten Volumina konvexer Polytope und gewissen Schnittzahlen von Divisoren in torischen Varietäten. (Einen neuen Beweis hierfür hat Atiyah [2] angegeben; vgl. auch [6] und [18].) Hierauf aufbauend fand Teissier [25] einen neuen Beweis der Alexandroff—Fenchel-Ungleichungen über gemischte Volumina konvexer Körper. Ferner bewies er algebraisch-geometrische Sätze vom Brunn—Minkowskischen Typ [26].

Wir befassen uns im Folgenden vorwiegend mit Varianten der Hopfschen σ-Prozesse, wie sie in den Arbeiten von Kempf usw. [15], Ehlers [8] und Danilov [7] im Zusammenhang mit der Auflösung von Singularitäten sowie bei Oda und Miyake [21] bzw. Oda [20] beim Auf- und Niederblasen torischer Mannigfaltigkeiten auf-

1980 *Mathematics Subject Classification.* Primary 52A25; Secondary 14E15.
Key words and phrases. Convex polytopes, stellar subdivision, toroidal embedding, toric varieties, σ-process.

treten. Wir zeigen, daß jede kompakte projektive torische n-dimensionale Varietät ($n \geqq 2$) durch eine endliche Anzahl verallgemeinerter σ-Prozesse bzw. deren Inverse in einen projektiven Raum überführt werden kann; für $n=2$ auch ohne vorausgesetzte Projektivität (aufgrund des Ergebnisses [9]). Schließlich geben wir noch einen Algorithmus an, der durch lineares Rechnen mit den vorgegebenen erzeugenden Vektoren der Kegel des Kegelsystems einer torischen Varietät zu entscheiden gestattet, ob die Varietät projektiv ist oder nicht.

2. σ-Prozesse

Folgende Definition der torischen Varietäten ist für unsere Zwecke angemessen (vgl. [15], [8], [7]): Sei $\sigma := \mathbf{R}_+ a_1 + \ldots + \mathbf{R}_+ a_k$ ein konvexer Kegel im euklidischen $\mathbf{R}^n$ und sei $\check{\sigma}$ der zu σ duale Kegel. (Daß sich für Kegel auch der Buchstabe σ eingebürgert hat, ist zufällig.) Wir nehmen $a_1, \ldots, a_n$ als primitive Gitterpunkte in $Z^n \subset \mathbf{R}^n$ an. Dann sei R_σ der Ring der Laurent-Polynome $\sum a_j z^j$ mit $j=(j_1, \ldots, j_n) \in \check{\sigma} \cap \mathbf{Z}^n$, $z^j := z_1^{j_1} \ldots z_n^{j_n}$; $z_1, \ldots, z_n \in \mathbf{C}$ (wo $\mathbf{C}$ durch einen allgemeineren Körper ersetzt werden kann). Spec R_σ stellt eine affine algebraische Varietät dar. Sind $a_1, \ldots, a_k = a_n$ beispielweise die kanonischen Basisvektoren von $\mathbf{R}^n$, dann ist R_σ der Polynomring $\mathbf{C}[z_1, \ldots, z_n]$ und Spec R_σ kann als affiner $\mathbf{C}^n$ aufgefaßt werden.

Sei Σ ein endlicher Komplex von Kegeln σ der obengenannten Art, d. h. mit σ gehöre jede Seite von σ zu Σ und mit $\sigma, \sigma' \in \Sigma$ sei $\sigma \cap \sigma'$ Seite von σ und von σ'. Man nennt Σ einen *Fächer*. Verklebt man Spec R_σ und Spec $R_{\sigma'}$ durch Inklusion in Spec $R_{\sigma \cap \sigma'}$ für je zwei $\sigma, \sigma' \in \Sigma$, dann erhält man eine algebraische Varietät, genannt *torische Varietät* X_Σ. (Der Name rührt daher, daß X_Σ eine Torus- d. h. $\mathbf{C}^{*n}$-Wirkung auf einer offenen dichten Teilmenge gestattet und im wesentlichen durch diese Eigenschaft gekennzeichnet ist.)

X_Σ ist genau dann singularitätenfrei, wenn jeder n-dimensionale Kegel $\sigma \in \Sigma$ simplizial ist und $\sigma = \mathbf{R}_+ a_1 + \ldots + \mathbf{R}_+ a_n$, $\det(a_1, \ldots, a_n) = \pm 1$ gilt. Genau dann ist X_Σ kompakt, wenn Σ den $\mathbf{R}^n$ überdeckt (vgl. [8 Satz 1]).

Wir beschränken uns in diesem Abschnitt auf simpliziale Fächer Σ, d. h. wir setzen voraus, daß jeder Kegel $\sigma \in \Sigma$ simplizial ist. Ferner nehmen wir an, daß Σ den $\mathbf{R}^n$ überdeckt, X_Σ also kompakt ist.

Ist $\sigma = \mathbf{R}_+ a_1 + \ldots + \mathbf{R}_+ a_k$, dann sind die durch $a_1, \ldots, a_k$ dargestellten Punkte Ecken eines $(k-1)$-Simplex $\mathring{\sigma}$. Die Zuordnung $\sigma \to \mathring{\sigma}$ ergibt dann eine kombinatorische Isomorphie $\varphi \colon \Sigma \to \mathring{\Sigma}$ zwischen Σ und einem Simplizialkomplex $\mathring{\Sigma}$, dessen Trägermenge $|\mathring{\Sigma}|$ eine $(n-1)$-Sphäre darstellt. Man sieht, daß $\mathring{\Sigma}$ Randkomplex eines Sternkörpers mit 0 im Kern ist.

Aus $\mathring{\Sigma}$ erhält man Σ zurück, indem man jede Zelle $\mathring{\sigma} \in \mathring{\Sigma}$ durch den Kegel $\sigma = \{\lambda x | x \in \mathring{\sigma}, \lambda \in \mathbf{R}_+\}$ ersetzt. Falls es einen (zu $\mathring{\Sigma}$ kombinatorisch isomorphen) Simplizialkomplex Σ^1 gibt, so daß ebenfalls Σ aus Σ^1 durch $\sigma^1 \to \sigma$ mit $\sigma = \{\lambda x | x \in \sigma^1, \lambda \in \mathbf{R}_+\}$ entsteht und Σ^1 Randkomplex eines konvexen Polytops ist, dann nennen wir $\mathring{\Sigma}$ und auch Σ *stark polytopal*. Es gilt (vgl. [7], S. 118):

(1) *Σ ist genau dann stark polytopal, wenn X_Σ projektiv ist.*

Im Falle $n=2$ gibt es offensichtlich stets ein solches Σ^1; jede 2-dimensionale kompakte torische Varietät ist also projektiv. (Beispiele in Fig. 1.)

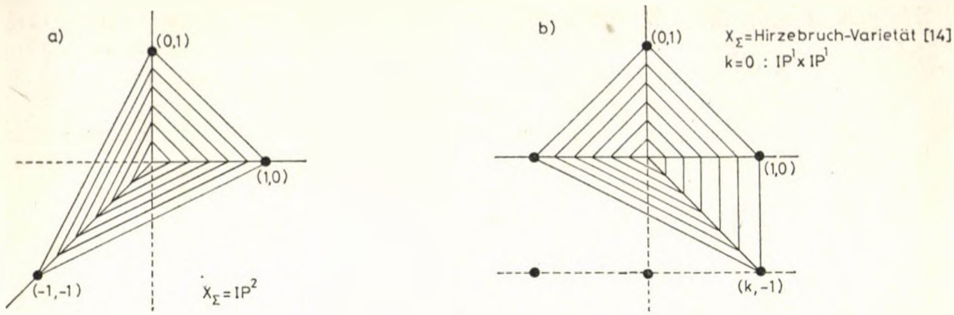


Fig. 1

(Über nicht stark polytopale Σ vgl. unten Abschnitt 3.)

Auf $\mathring{\Sigma}$ wenden wir eine *stellare Unterteilung* an, d. h. wir ersetzen den Stern st$(\mathring{\sigma}, \mathring{\Sigma})$ einer Zelle $\mathring{\sigma} \in \mathring{\Sigma}$ mit Dimension$\geqq 1$ durch die Verbindung eines relativ inneren Punktes p von $\mathring{\sigma}$ mit dem Verkettungskomplex link$(\mathring{\sigma}, \mathring{\Sigma})$ (vgl. Fig. 2).

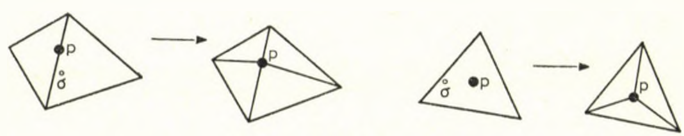


Fig. 2

Der neue Komplex sei mit $S(p; \mathring{\sigma})\mathring{\Sigma}$ bezeichnet. Ist p rational, dann sei $\tilde{p} = tp$ der dazu proportionale primitive Gitterpunkt $(t > 0)$. Durch φ^{-1} wird die Unterteilung auf Σ übertragen, wobei $\mathbf{R}_+ p$ der neue eindimensionale Kegel ist:

$$\Sigma \to S(\tilde{p}, \sigma)\Sigma.$$

Man hat (vgl. [8], [21]):

(2) *Ist X_Σ singularitätenfrei und wählt man für $\sigma = \mathbf{R}_+ a_1 + \ldots + \mathbf{R}_+ a_k$ den Gitterpunkt $\tilde{p}$ als $\tilde{p} = a_1 + \ldots + a_k$ (auch, was mißverständlich ist, „baryzentrische" Unterteilung genannt); dann ist*

$$X_{S(\tilde{p}, \sigma)\Sigma} \to X_\Sigma$$

ein Hopfscher σ-Prozeß, d. h. $X_{S(\tilde{p}; \sigma)\Sigma}$ wird in einem Punkt niedergeblasen.

Damit hat also die Anwendung von σ-Prozessen bzw. inversen σ-Prozessen auf eine torische Varietät ihr kombinatorisches Analogon in speziellen inversen stellaren Unterteilungen bzw. stellaren Unterteilungen von Σ. Beispielweise kann man leicht zeigen ([21], S. 377):

(3) *Jede kompakte torische 2-Mannigfaltigkeit läßt sich entweder auf $\mathbf{P}^2$ oder auf eine Hirzebruch-Varietät niederblasen. Durch Auf- und Niederblasen gelangt man stets zu $\mathbf{P}^2$.*

Für 3-Mannigfaltigkeiten weiß man schon viel weniger. So haben Oda und Miyake ([20], [21]) diejenigen nicht niederblasbaren kompakten torischen 3--Mannigfaltigkeiten gekennzeichnet, für die Σ höchstens 8 eindimensionale Seiten besitzt. Sie greifen dabei auf die kombinatorische Klassifikation konvexer 3-Polytope mit höchstens 8 Ecken zurück (vgl. Grünbaum [13]). Alle so gewonnenen Mannigfaltigkeiten können durch Auf- und Niederblasen wieder in $\mathbf{P}^3$ überführt werden. Hieran schließen Oda und Miyake die folgende Vermutung an:

(4) Vermutung: Jede kompakte torische 3-Mannigfaltigkeit läßt sich durch Auf- und Niederblasen in endlich vielen Schritten in $\mathbf{P}^3$ überführen.

Ein kombinatorisch-geometrischer Beweis hierfür dürfte äußerst schwierig sein, da man nur sehr spezielle („baryzentrische" s. o.) stellare Prozesse zur Verfügung hat. Wir fügen eine Abschwächung von (4) hinzu, die immer noch schwer zugänglich sein dürfte:

(5) Vermutung: Jede kompakte torische 3-Mannigfaltigkeit läßt sich durch Auf- und Niederblasen in endlich vielen Schritten in eine projektive torische 3-Mannigfaltigkeit überführen.

Analoga zu σ-Prozessen werden aber auch bei der Auflösung von Singularitäten herangezogen. Kommt beispielweise in einem Fächer Σ, der in $\mathbf{R}^2$ liegt, der Kegel $\sigma=\mathbf{R}_+ a_1+\mathbf{R}_+ a_2=\mathbf{R}_+(1, 0)+\mathbf{R}_+(1, 2)$ vor, dann wird durch den Übergang von X_Σ zu $X_{S(p,\sigma)\Sigma}$ mit $p=(1, 1)$ die in $\operatorname{Spec}\mathbf{C}[\check{\sigma}\cap\mathbf{Z}^2]$ liegende Singularität aufgelöst (vgl. [15], S. 36, [7], S. 100).

Umgekehrt tritt in $\operatorname{Spec}\mathbf{C}[\check{\sigma}\cap\mathbf{Z}^n]$ für $\sigma=\mathbf{R}_+a_1+\ldots+\mathbf{R}_+a_n$ in einem n-dimensionalen Fächer stets dann eine isolierte Singularität auf, wenn man den inversen σ-Prozeß, der $S(\bar{p}, \sigma)$ mit einem $\bar{p}\neq a_1+\ldots+a_n$ entspricht, anwendet, auch für singularitätenfreies $\operatorname{Spec}\mathbf{C}[\check{\sigma}\cap\mathbf{Z}^n]$. Ferner kann bei einer inversen stellaren Unterteilung eine Singularität zustandekommen, wie obiges Beispiel zeigt.

Sei allgemein der Morphismus

$$X_{S(\bar{p},\sigma)\Sigma} \to X_\Sigma$$

als *verallgemeinerten σ-Prozeß* bezeichnet, wenn $\bar{p}$ einen beliebigen primitiven Gitterpunkt im relativen Innern von σ darstellt.

Wir beweisen folgende schwächere Aussage als (3), lassen aber Singularitäten zu und dehnen sie im projektiven Fall auf höhere Dimensionen aus.

SATZ 1. *Sei X_Σ eine kompakte torische Varietät der Dimension $n\geqq 2$. Unter jeder der folgenden Bedingungen gibt es eine Folge $S_1, \ldots, S_r$ von verallgemeinerten σ-Prozessen bzw. hierzu inversen Prozessen, so daß*

$$S_r\ldots S_1 X_\Sigma = \mathbf{P}^n:$$

(a) $n\leqq 3$,
(b) $n>3$ *und X_Σ projektiv.*

BEWEIS. In die Sprache der Komplexe übersetzt, kann man die Behauptung im Fall (b) äquivalent so ausdrücken: Ist P ein konvexes n-Polytop, dann existieren konvexe n-Polytope $P=P_0, P_1, \ldots, P_r$, so daß

(I) P_r ein n-Simplex ist,

(II) der Randkomplex von P_j aus dem Randkomplex von P_{j-1} bis auf eine Isomorphie φ_j durch stellare Unterteilung oder inverse stellare Unterteilung entsteht, $j=1, \ldots, r$,

(III) unter φ_j stets die vom stellaren Prozeß nicht betroffenen Ecken von P_{j-1} festbleiben.

Dieses Ergebnis wurde aber in [10] bewiesen.

Fall (a) läßt sich wie (b) zeigen, wenn man zuerst durch stellare Unterteilungen Σ stark polytopal macht. Für $n=2$ ist dies ohnehin der Fall, für $n=3$ ist es in [9] bewiesen.

Aus [10] ergibt sich noch:

ZUSATZ. *Man kommt in Satz* 1(b), *falls Σ simplizial ist, stets mit höchstens* $2f_{n-1}-f_0-v_{\max}+n-1$ *Prozessen aus, wo f_j die Anzahl von j-dimensionalen Zellen von $\mathring{\Sigma}$ ist und $v_{\max}$ die maximale Anzahl von $(n-1)$-Zellen angibt, die eine Ecke gemeinsam haben.*

3. Welche torische Varietäten sind projektiv?

Eine notwendige Bedingung dafür, daß X_Σ projektiv, d. h. Σ stark polytopal ist (vgl. oben (1)) besteht darin, daß der Komplex $\mathring{\Sigma}$ *polytopal*, d. h. zum Randkomplex eines konvexen Gitterpolytops kombinatorisch isomorph ist. Da man inzwischen zahlreiche Beispiele nicht polytopaler Komplexe $\mathring{\Sigma}$ der Dimension $\geqq 3$ kennt (vgl. hierzu [13], [1], [12]), lassen sich Beispiele nicht-projektiver torischer Varietäten der Dimension $\geqq 4$ angeben; man muß allerdings solche aussuchen, die wenigstens als Randkomplex von Sternkörpern realisierbar sind (was nicht immer der Fall ist, ein Beispiel haben der Verf. und C. Schulz angegeben).

Für Dimension 2 von $\mathring{\Sigma}$, d. h. Varietätendimension 3 ist dagegen jedes $\mathring{\Sigma}$ polytopal (nach dem bekannten Satz von Steinitz, vgl. [13]), allerdings nicht notwendigerweise stark polytopal. Wir geben hierfür weiter unten ein Beispiel an, und zwar gewinnen wir dies aufgrund eines Kriteriums für starke Polytopalität von $\mathring{\Sigma}$ bzw. Σ.

Sei also Σ der Fächer einer beliebigen kompakten torischen Varietät. Ist $\sigma=\mathbf{R}_+a_{i_1}+\ldots+\mathbf{R}_+a_{i_n}+\mathbf{R}_+a_{i_{n+1}}+\ldots+\mathbf{R}_+a_{i_{n+r}}$ ein n-dimensionaler Kegel und ist $r>0$, dann ist σ nicht simplizial. Durch endlich viele Hyperebenen kann aber σ in simpliziale Kegel zerschnitten werden, deren erzeugende Vektoren Teilmengen von $\{a_{i_1}, \ldots, a_{i_{n+r}}\}$ bilden. Die dabei neu entstehenden Kegel der Dimension $n-1$ nennen wir kurz *neue* Kegel. Sind so alle n-dimensionale Kegel von Σ, die nicht schon simplizial sind, zerschnitten, dann entsteht ein neuer Fächer Σ'.

In jedem $\sigma'=\mathbf{R}_+a_{i_1}+\ldots+\mathbf{R}_+a_{i_k}\in\Sigma'$ fasse man wieder $a_{i_1}, \ldots, a_{i_k}$ als Punkte auf und betrachte das $(k-1)$-Simplex, dessen Ecken sie bilden (vgl. Abschnitt 2). Dann entsteht ein Simplizialkomplex $\mathring{\Sigma}'$, der einen Sternkörper (mit 0 im Innern) berandet. $\mathring{\Sigma}'$ ist orientierbar. Wir führen eine Orientierung ein und übertragen diese kanonisch auf Σ'.

Wir denken uns jetzt den $\mathbf{R}^n$, der von Σ bzw. Σ' überdeckt wird, zum $\mathbf{R}^{n+1}$, erweitert; der neue kanonische Basisvektor sei e. Genau dann ist Σ stark polytopal, wenn eine positiv homogene konvexe reelle Funktion

$$f\colon \mathbf{R}^n \to \mathbf{R}_+$$

mit folgenden Eigenschaften existiert:
(a) $f|\sigma$ ist für jedes $\sigma\in\Sigma$ linear,
(b) $f|M$ ist für jede Obermenge $M\notin\Sigma$ eines σ nicht linear.
Beispielsweise hat die „Distanzfunktion" eines konvexen Polytops (mit 0 im Innern) diese Eigenschaften.

Die Funktionen f kann man mit Hilfe ihrer Graphen wie folgt kennzeichnen. Jedem a_i, das als erzeugender Vektor in Σ auftritt, ordnen wir einen Punkt $b_i\colon a_i+t_ie$ mit $t_i>0$ zu und bilden jeden Kegel $\sigma'=\mathbf{R}_+a_{i_1}+\ldots+\mathbf{R}_+a_{i_k}\in\Sigma'$ auf einen Kegel $\tilde{\sigma}':=\mathbf{R}_+b_{i_1}+\ldots+\mathbf{R}_+b_{i_k}$ gleicher Dimension in $\mathbf{R}^{n+1}$ ab. Σ' geht dabei in ein simpliziales Kegelsystem $\tilde{\Sigma}$ über. $|\tilde{\Sigma}|$ kann dann als Graph einer positiv homogenen Funktion $f\colon \mathbf{R}^n\to\mathbf{R}_+$ aufgefaßt werden und man hat:

(1) Genau dann erfüllt f die Voraussetzungen (a), (b), wenn es zu jedem $\sigma\in\Sigma$ eine Linearfunktion L_σ gibt mit $f|\sigma=L_\sigma|\sigma$ und $f(x)>L_\sigma(x)$ für $x\notin\sigma$.

Seien $\sigma_1'=\mathbf{R}_+a_{i_1}+\ldots+\mathbf{R}_+a_{i_n}$ und $\sigma_2'=\mathbf{R}_+a_{i_2}+\ldots+\mathbf{R}_+a_{i_{n+1}}$ Kegel aus Σ' mit der gemeinsamen Seite $\sigma_1'\cap\sigma_2'=\mathbf{R}_+a_{i_2}+\ldots+\mathbf{R}_+a_{i_n}$. Sei die Orientierung der Kegel so gewählt, daß in $\mathbf{R}^n$ $\det(a_{i_1}\ldots a_{i_n})>0$ und $\det(a_{i_2}\ldots a_{i_{n+1}})>0$ gilt. Dann hat man nach (1):

(2) Genau dann erfüllt f die Voraussetzungen (a), (b), wenn in $\mathbf{R}^{n+1}$ für jedes Paar σ_1', σ_2' von n-Zellen mit gemeinsamer $(n-1)$-Seite gilt:

$$(*)\qquad \det(b_{i_1}\ldots b_{i_{n+1}})\begin{cases}>0 & \text{falls } \sigma_1\cap\sigma_2 \text{ kein neuer Kegel ist,}\\ =0 & \text{sonst.}\end{cases}$$

Setzt man in $(*)$ $b_i=a_i'+t_ie$ ein (a_i' jetzt als Vektor in $\mathbf{R}^{n+1}$ aufgefaßt), so ergibt sich:

SATZ 2. *Genau dann ist Σ stark polytopal, wenn das System von Ungleichungen bzw. Gleichungen*

$$(\sigma')\qquad \begin{gathered} t_{i_1}\det(ea_{i_2}\ldots a_{i_{n+1}})+\ldots+t_{i_{n+1}}\det(a_{i_1}a_{i_2}\ldots a_{i_n}e)\\ \begin{cases}>0, & \text{falls } \sigma'=\sigma_1'\cap\sigma_2' \text{ kein neuer Kegel ist,}\\ =0 & \text{sonst}\end{cases}\end{gathered}$$

eine Lösung besitzt, wo σ' alle $(n-1)$-dimensionalen Kegel von Σ' durchläuft.

ZUSATZ. *Die Lösbarkeit des Systems (σ') läßt sich mit den Methoden der linearen Optimierung entscheiden; und zwar mit Hilfe des Khachian-Algorithmus.* (Siehe hierzu etwa [22].)

Beispiel einer nicht projektiven torischen Varietät (vgl. [7], S. 119):

$a_1=(1,0,1)$, $a_2=(0,1,1)$, $a_3=(0,0,1)$, $a_4=(3,-1,1)$, $a_5=(-1,3,1)$, $a_6=(-1,-1,1)$, $a_7=(0,0,-1)$. Die Kegelstruktur ist aus Figur 3 ersichtlich. Stellt man die zu den 2-dimensionalen Kegeln mit Erzeugenden a_2, a_4 bzw. a_3, a_5 bzw. a_1, a_6 gehörigen

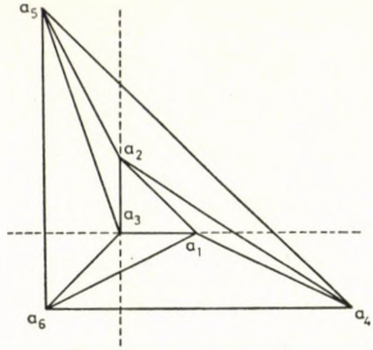


Fig. 3

Ungleichungen (σ') auf, dann erhält man

$$-4t_1+4t_3+t_4-t_6>0$$

$$4t_1-4t_2-t_4+t_5>0$$

$$4t_2-4t_3-t_5+t_6>0.$$

Addiert man diese Ungleichungen, so ergibt sich der Widerspruch $0>0$.

LITERATURVERZEICHNIS

[1] Altshuler, A. and Steinberg, L., The complete enumeration of the 4-polytopes and 3-spheres with 9 vertices, *Pacific J. Math.* **117** (1985), 1—16.

[2] Atiyah, M. F., Angular momentum, convex polyhedra and algebraic geometry, *Proc. Edinburgh Math. Soc.* (2) **26** (1983), 121—133. *MR* **85a**: 58027.

[3] Bernstein, D., The number of roots of a system of equations, *Funkcional. Anal. i Priložen.* **9** (1975), no. 3, 1—4 (in Russian). *MR* **55** #8034.

[4] Billera, L. J. and Lee, C. W., A proof of the sufficiency of McMullen's conditions for *f*-vectors of simplicial convex polytopes, *J. Combin. Theory Ser. A* **31** (1981), 237—255. *MR* **82m**: 52006.

[5] Bruggesser, H. und Mani, P., Shellable decompositions of cells and spheres, *Math. Scand.* **29** (1971), 197—205. *MR* **48** #7286.

[6] Burago, D. M. and Zalgaller, V. A., *Geometric inequalities*, Nauka, Leningrad Otdel., Leningrad, 1980 (in Russian). *MR* **82d**: 52009.

[7] Danilov, V. I., The geometry of toric varieties, *Uspehi Mat. Nauk* **33** (1978), no. 2, 85—134 (in Russian); *Russian Math. Surveys* **33** (1978), no. 2, 97—154. *MR* **80g**: 14001.

[8] Ehlers, F., Eine Klasse komplexer Mannigfaltigkeiten und die Auflösung einiger isolierter Singularitäten, *Math. Ann.* **218** (1975), 127—156. *MR* **58** #11502.

[9] Ewald, G., Über stellare Unterteilung von Simplizialkomplexen, *Arch. Math. (Basel)* **46** (1986), 153—158.

[10] Ewald, G., Über stellare Äquivalenz konvexer Polytope, *Resultate Math.* **1** (1978), 54—60. *MR* **80b**: 52012.

[11] Ewald, G. and Shephard, G. C., Stellar subdivisions of boundary complexes of convex polytopes, *Math. Ann.* **210** (1974), 7—16. *MR* **50** #3115.

[12] Ewald, G., Kleinschmidt, P., Pachner, U. und Schulz, C., Neuere Entwicklungen der kombinatorischen Konvexgeometrie, In: *Contributions to geometry* (Proc. Geom. Sympos., Siegen, 1978), Basel, 1979, 131—163.

[13] Grünbaum, B., *Convex polytopes*, Pure and applied mathematics, Vol. 16, Interscience Publishers [John Wiley and Sons, Inc.], New York, 1967. *MR* **37** #2085.

[14] HIRZEBRUCH, F., Über eine Klasse von einfach-zusammenhängenden komplexen Mannigfaltigkeiten, *Math. Ann.* **124** (1951), 77—86. *MR* **13**—574.
[15] KEMPF, G., KNUDSEN, F., MUMFORD, D. and SAINT-DONAT, B., *Toroidal embeddings*, I, Lecture notes in mathematics, Vol. 339, Springer-Verlag, Berlin, 1973. *MR* **49** #299.
[16] KIND, B. and KLEINSCHMIDT, P., Schälbare Cohen—Macauley-Komplexe und ihre Parametrisierung, *Math. Z.* **167** (1979), 173—179. *MR* **80k**: 13010.
[17] KLEINSCHMIDT, P., Über Hilbert-Funktionen graduierter Gorenstein-Algebren, *Arch. Math. (Basel)* **43** (1984), 501—506.
[18] KOUCHNIRENKO, A. G., Polyèdres de Newton et nombres de Milnor, *Invent. Math.* **32** (1976), 1—31. *MR* **54** #7454.
[19] MCMULLEN, P., The maximum numbers of faces of a convex polytope, *Mathematika* **17** (1970), 179—184. *MR* **44** #921.
[20] ODA, T., *Torus embeddings and applications*, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, 57, Published for the Tata Institute of Fundamental Research, Bombay by Springer-Verlag, Berlin—New York, 1978. *MR* **81e**: 14001.
[21] ODA, T. and MIYAKE, K., Almost homogeneous algebraic varieties under algebraic torus action, *Manifolds—Tokyo* 1973 (Proc. Internat. Conf., Tokyo, 1973), Univ. Tokyo Press, Tokyo, 1975, 373—381. *MR* **52** #406.
[22] PAPADIMITRIOU, C. H. and STEIGLITZ, K., *Combinatorial optimization: algorithms and complexity*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1982. *MR* **84k**: 90036.
[23] STANLEY, R. P., The upper bound conjecture and Cohen—Macaulay rings, *Studies in Appl. Math.* **54** (1975), 135—142. *MR* **56** #16640.
[24] STANLEY, R. P., The number of faces of a simplicial convex polytope, *Adv. in Math.* **35** (1980), 236—238. *MR* **81f**: 52014.
[25] TEISSIER, B., Bonnesen-type inequalities in algebraic geometry I. Introduction to the problem. *Seminar on Differential Geometry*, Ann. of Math. Studies, 102, Princeton Univ. Press, Princeton, N. J., 1982, 85—105. *MR* **83d**: 52010.
[26] TEISSIER, B., Sur une inégalité à la Minkowski pour les multiplicités, Anhang in: Eisenbud, D. and Levine, H. I., An algebraic formula for the degree of a C^∞ map germ, *Ann. Math.* (2) **106** (1977), 19—44. *MR* **57** #7651.

(Eingegangen am 3. Januar 1985)

MATHEMATISCHES INSTITUT
RUHR-UNIVERSITÄT BOCHUM
GEBÄUDE NA
UNIVERSITÄTSSTRASSE 150
D—4630 BOCHUM 1
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 19—51

MINIMAL PRESENTATION OF THE 10 COMPACT EUCLIDEAN SPACE FORMS BY FUNDAMENTAL DOMAINS

EMIL MOLNÁR

To Professor László Fejes Tóth on occasion of his 70th birthday

1. Introduction

A complete connected Riemannian n-dimensional manifold of constant sectional curvature will be called a space form. It is well-known that each space form can be represented as an orbit space M/G. Here M is one of the classical n-dimensional simply connected spaces of constant curvature k, i.e., M is either a spherical $(k>0)$ or the euclidean $(k=0)$ or a hyperbolic n-space $(k<0)$. The isometry group G acts discontinuously and freely on M, i.e., there is a nonempty open set V in M so that no two distinct points of V are equivalent under G, moreover, the identity 1 is the only element of G which has fixed points. Then G is isomorphic to the fundamental group of the manifold M/G.

Intuitively each space form is locally isometric to one of the classical n-spaces.

Concerning the general theory we refer to the monograph [23] of J. A. Wolf. For hyperbolic space forms see [2, 4, 13] and the report [21] of W. P. Thurston. Geometric constructions of some hyperbolic space forms are presented in [1, 2, 9, 16, 17, 18, 22].

In this paper the 10 classical compact 3-dimensional euclidean space forms will be described by means of a geometric method presenting each of them minimally by a fundamental domain.

For each group G from among the 10 non-isomorphic (affinely non equivalent) crystallographic space groups acting freely on the euclidean 3-space [3, 5, 10, 23], there is a compact fundamental polyhedron $\mathscr{P}$ (e.g. a Dirichlet polyhedron) endowed with a face identification. The identifying isometries generate the group G. The so-called cycle relations, belonging to the edge segment equivalence classes of $\mathscr{P}$, give us a presentation of G in the sense of Poincaré. Such a relation shows how the G-images of $\mathscr{P}$ (in the fundamental tiling) surround an edge segment of the corresponding equivalence class [2, 13, 14]. For a given group G we have more possible fundamental polyhedra, hence more presentations as well, respectively.

Our first purpose is to generalize this geometric presentation for those compact fundamental domains, which are topological polyhedra with some minimality conditions [14, 15]. Such a fundamental domain $\mathscr{F}$ for the group G shall have the following properties:

1) The interior of the compact fundamental domain $\mathscr{F}$ is homeomorphic to an open 3-dimensional simplex. The G-images of $\mathscr{F}$ tile the space without common inner points.

1980 *Mathematics Subject Classification.* Primary 51M20; Secondary 20F05, 20H15, 57S30, 05C25.

Key words and phrases. Euclidean space form, minimally presenting fundamental polyhedron, generators and relations for fundamental group of a space form.

2) The boundary of $\mathscr{F}$ consists of finitely many side faces. Each side face is the common part of $\mathscr{F}$ and exactly one neighbouring G-image in the corresponding fundamental space tiling. The side faces may be curved, but each of them has a relative interior homeomorphic to an open 2-simplex.

3) If the domain $\mathscr{F}$ and one of its G-images $\mathscr{F}^{g}$, for certain $g \in G$, have the face f_g in common, then $\mathscr{F}$ and its image $\mathscr{F}^{g^{-1}}$ also have exactly one common face denoted by $f_{g^{-1}}$. We say that $f_{g^{-1}}$ is identified with f_g by the isometry g, or f_g is identified with $f_{g^{-1}}$ by g^{-1}. These identifying isometries generate the group G.

4) The boundary of each side face f of the domain $\mathscr{F}$ consists of finitely many edge segments along each of which f joins exactly one neighbouring side face. The edge segments may be curved but each of them has a relative interior homeomorphic to an open interval. The side face identifications induce the G-equivalence of the edge segments. We require that a segment shall not have any G-equivalent points in its interior. To each edge segment equivalence class there belongs a cycle relation of the identifying isometries, which expresses how the G-images of $\mathscr{F}$ (in the fundamental tiling) surround an edge segment of the equivalence class (in the sense of Poincaré as before). This will be shown concretely on examples later.

5) The boundary of each edge segment consists of two vertices. For each vertex A of $\mathscr{F}$ the edge segments and the faces, containing A, form cycles in which the consecutive edge segments are contained in exactly one common face, the consecutive faces have exactly one common edge segment.

6) The fundamental topological polyhedron $\mathscr{F}$ is said to be minimally presenting the isometry group G above, iff $\mathscr{F}$ has the minimum number of faces, moreover, it has the minimum number of edge segments divided into equivalence classes, whose number is also minimal.

The requirements 1)—6) guarantee that the fundamental domain $\mathscr{F}$ serves an algebraic presentation for G with the smallest number of generators with minimum number of relations, moreover, the sum of the relation lengths is also minimal.

We remark that neither involutive elements (reflections) nor elements of finite order occur in G because of the free action of G.

The general strategy for finding a fundamental domain $\mathscr{F}$, presenting the group G minimally, will be the following.

i) We start with giving the group G by the usual matrix-vector presentation in the conventional basis, which spans the translation lattice $\mathbf{L}_G$ belonging to G. This is well-known together with the typization of the transformations in G [3, 5, 10, 19].

ii) We select in all possible ways collections of typical generators which generate the translational lattice $\mathbf{L}_G$ and the so-called point group G_0 of the group G. We reduce the collection if it is possible. Equivalent collections will be related by means of the affine normalizer N_G of G.

This follows by a theorem of Bieberbach and Frobenius, which states, that isomorphic space groups G and G', acting on the euclidean space E^n, are affinely conjugate; i.e. there is an affine bijection φ of E^n such that $G' = \varphi^{-1} G \varphi$ holds [3, 11, 12, 19, 20, 23]. In the case $G' = G$ this affinity φ, belonging to N_G, provides affine fundamental domains, giving equivalent presentations for G.

iii) For a given generator collection we determine a complete set of defining relations [6]. We reduce the number and the length of the relations by Tietze transformations. We select all the minimal sets of defining relations with minimal length sums up to certain algebraic and affine equivalence. The proof of the minimality needs some calculations checking finitely many cases. The method is straightforward but lengthy and tedious. We shall give an illustration here for the group $G = \mathbf{P4_1}$.

iv) We construct a typical fundamental domain $\mathcal{F}$ which corresponds to such a minimal presentation according to the criteria in 1)—6). The combinatorial structure of $\mathcal{F}$ [7] will be uniquely determined by the presentation.

The resulting fundamental domains and the corresponding presentations will be described in the figures. We use consequently the notations of the crystallography [10].

To summarize the results we formulate the following

THEOREM. *The 10 compact euclidean space forms have 18 minimally presenting fundamental domains: the group* $\mathbf{P1}$ *has 1 topological parallelepipedon (brick); the group* $\mathbf{P2_1}$ *has 2 topological bricks; the group* $\mathbf{Pb}$ *has 3 topological bricks; the group* $\mathbf{Bb}$ *has 1 topological "simplex" (with additional edge segments and vertices); the group* $\mathbf{P2_12_12_1}$ *has 1 "simplex"; the group* $\mathbf{Pca2_1}$ *has 4 bricks; the group* $\mathbf{Pna2_1}$ *has 1 "simplex"; the group* $\mathbf{P4_1}$ *has 1 "simplex"; the group* $\mathbf{P3_1}$ *has 3 minimally presenting domains: the most interesting one has 4 side faces, two of them have not common edge, the other two domains are "simplices"; the group* $\mathbf{P6_1}$ *has 1 "simplex".*

2. The space form $E^3/\mathbf{P4_1}$ as a typical example detailed

We start the crystallographic group $G = \mathbf{P4_1}$ which acts discontinuously and freely on the euclidean 3-space $M = E^3$. Therefore, the orbit space $E^3/\mathbf{P4_1}$ is a (compact) euclidean space form.

We choose an origin O and a base $\overrightarrow{OE_1} = \mathbf{e}_1$, $\overrightarrow{OE_2} = \mathbf{e}_2$, $\overrightarrow{OE_3} = \mathbf{e}_3$ for the lattice $\mathbf{L}_G$ with the Gram matrix

$$((\mathbf{e}_i;\, \mathbf{e}_j)) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix} \tag{1}$$

where $c > 0$ is an arbitrary constant as usual, (;) denotes the euclidean scalar product in the vector space $\mathbf{E}^3$.

The generating screw motion $\mathrm{s}\,(s, \mathbf{s})$ for $G = \mathbf{P4_1}$, will be given by the linear transformation $s\colon \mathbf{E}^3 \to \mathbf{E}^3$

$$\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix} s := \begin{pmatrix} \mathbf{e}_1\, s \\ \mathbf{e}_2\, s \\ \mathbf{e}_3\, s \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix} \tag{2}$$

and by the vector $\mathbf{s} := \overrightarrow{OO^s}$,

$$\mathbf{s} = \begin{pmatrix} 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}. \tag{3}$$

Hence the s-image $Y := X^s$ of a point X will be expressed by the position vectors $\overrightarrow{OX} = \mathbf{x} = x^i \mathbf{e}_i$, $y^j \mathbf{e}_j = \mathbf{y} = \overrightarrow{OY} = \overrightarrow{OX^s} = \mathbf{x}^s = \mathbf{x}s + \mathbf{s} = x^i s_i^j \mathbf{e}_j + s^j \mathbf{e}_j$. In components:

$$(4) \qquad (y^1\ y^2\ y^3) = (x^1\ x^2\ x^3)\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

We use the row-column multiplication convention, but we prefer row matrices for vector components, moreover, transformations and matrices operate on the right. The summation convention is understood for the same upper and lower indices.

We can see that the generating translations for $\mathbf{L}_G$ are

$$(5) \qquad \mathrm{p}_1(I, \mathbf{e}_1), \quad \mathrm{p}_2(I, \mathbf{e}_2) = \mathrm{s}^{-1}\mathrm{p}_1\mathrm{s}, \quad \mathrm{p}_3(I, \mathbf{e}_3) = \mathrm{s}^4.$$

For illustration we construct the Dirichlet polyhedron $\mathcal{D}_O$ belonging to the origin O. From this one can imagine the whole fundamental tiling by Dirichlet polyhedra according to the G-orbit of O:

$$O^G := \{O^g \in E^3 : g \in G\}.$$

In Figure 1 the Dirichlet domain

$$\mathcal{D}_O := \{X \in E^3 : d(X, O) \leqq d(X, O^g) \text{ for every } g \in G\}$$

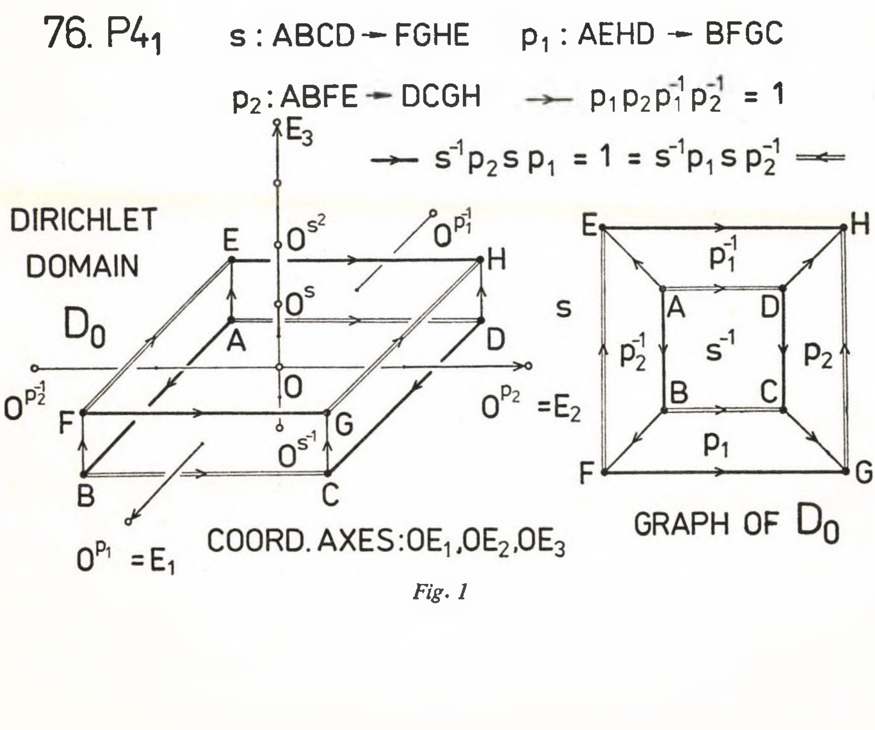


Fig. 1

is a quadratic parallelepipedon (column) whose opposite side faces are paired by the 4_1 screw motion s, by the translations p_1 and p_2, respectively. For example, the face $f_{s^{-1}}:=ABCD$, lying on the symmetry plane of O and $O^{s^{-1}}$, is mapped by s onto the face $f_s:=FGHE$, lying on the symmetry plane of O^s, and O; moreover, the domain $\mathcal{D}_O$ is mapped by s onto the domain $\mathcal{D}_{O^s}$ joining $\mathcal{D}_O$ along the face f_s. Of course, the inverse screw motion s^{-1} maps the face $f_s:=FGHE$ onto the face $f_{s^{-1}}:=ABCD$ and the domain $\mathcal{D}_O$ is mapped by s^{-1} onto the domain $\mathcal{D}_{O^{s^{-1}}}$ joining $\mathcal{D}_O$ along the face $f_{s^{-1}}$. In the Schlegel diagram of $\mathcal{D}_O$ we omit the f's from the face symbols. This diagram will be referred to in the following as the graph of the domain [14, 15, 16, 17].

The pairing isometries, or shortly identifications, generate the group $G=\mathbf{P4}_1$. These identifications induce a partition of the edges into classes of oriented segments such that a segment does not contain two G-equivalent points in its interior. Consider, e.g., the segment equivalence class

$$\Rightarrow= \quad \{FE;\, AD;\, BC;\, GH\}.$$

Now, form a so-called cycle transformation and deduce the corresponding cycle relation as follows.

Choose an edge segment, say FE, and consider one of the faces, say $FEHG$, whose boundary contains FE. The isometry s^{-1} maps FE and $FEHG$ onto AD and $ADCB$. There exists exactly one other face $ADHE$, with AD on its boundary, furthermore, an isometry p_1 which maps AD and $ADHE$ onto BC and $BCGF$. We obtain a cycle of generating isometries s^{-1}, p_1, s, p_2^{-1} according to the scheme

$$(FE;\, f_s = FEHG) \xrightarrow{s^{-1}} (AD;\, f_{s^{-1}} = ADCB);$$

$$(AD;\, f_{p_1^{-1}} = ADHE) \xrightarrow{p_1} (BC;\, f_{p_1} = BCGF);$$

$$(BC;\, f_{s^{-1}} = BCDA) \xrightarrow{s} (GH;\, f_s = GHEF);$$

$$(GH;\, f_{p_2} = GHDC) \xrightarrow{p_2^{-1}} (FE;\, f_{p_2^{-1}} = FEAB).$$

Forming the inverse transformations these facts can also be formulated as follows. The edge segment FE is consecutively surrounded by domains

$$\mathcal{D}_O,\ \mathcal{D}_O^{s},\ \mathcal{D}_O^{p_1^{-1}s},\ \mathcal{D}_O^{s^{-1}p_1^{-1}s},\ \mathcal{D}_O^{p_2s^{-1}p_1^{-1}s}.$$

The first four ones fill an angular region of measure

$$\varepsilon(FE)+\varepsilon(AD)+\varepsilon(BC)+\varepsilon(GH) = 2\pi,$$

therefore

$$\mathcal{D}_O^{p_2s^{-1}p_1^{-1}s} = \mathcal{D}_O,$$

and we get the relation $p_2s^{-1}p_1^{-1}s=1$ or the so-called cycle relation

(6) $$s^{-1}p_1sp_2^{-1} = 1$$

in the sense of Poincaré. Choosing now the face $f_{p_2^{-1}}=FEAB$ to FE as the first step of the process above, we obtain the inverse cycle and the inverse relation

$$p_2 s^{-1} p_1^{-1} s = 1;$$

and starting, e.g., with AD and $f_{p_1^{-1}}=ADHE$, we arrive at

$$p_1 s p_2^{-1} s^{-1} = s(s^{-1} p_1 s p_2^{-1}) s^{-1} = 1$$

as cycle relation. These lead to relations equivalent to (6).

In Figure 1 all the three segment equivalence classes are described and the corresponding cycle relations are similarly deduced. As a special case of the well-known Poincaré theorem [13, 14], we could reformulate that the polyhedron $\mathcal{D}_O$ constructed is in fact a fundamental domain for the group G generated by the identifying isometries, furthermore the three cycle relations form a complete set of relations for G, i.e., any relation in G can be algebraically derived from these. What is more, the fundamental domain $\mathcal{D}_O$ can be so deformed that it will be an equivalent topological polyhedron described in the introduction. Equivalence means that the deformed polyhedron provides the same presentation as $\mathcal{D}_O$ before. We can start with an arbitrary point A; then we form its translates $B:=A^{p_1}$, $D:=A^{p_2}$, $C:=A^{p_1 p_2}=A^{p_2 p_1}$; then we can choose a face (it may be curved), connecting these points with great freedom; then the face $FGHE$ has to be $A^s B^s C^s D^s$. The equalities $DCGH=A^{p_2}B^{p_2}F^{p_2}D^{p_2}$, $BCGF=A^{p_1}D^{p_1}H^{p_1}E^{p_1}$ are required for the other (curved) faces in order to get finally a topological polyhedron $\mathcal{D}$. Now, the relations for G guarantee that the G-images of $\mathcal{D}$ form a fundamental tiling in which a (curved) edge segment, say FE, is surrounded as before by fundamental domains in the tiling constructed by means of $\mathcal{D}_O$.

This latter assertion gives us the basic idea for the generalization.

3. Presentations for $\mathbf{P4_1}$ with two generators

From the presentation in Figure 1 we conclude that the group $G=\mathbf{P4_1}$ can be generated by two isometries as well. There are three essential cases.

1) The screw motion s described in formulas 2 (2), (3) and the translation $p:=p_1$ in 2 (5) generate $\mathbf{P4_1}$. We have already learned that the translations p_1 and $p_2=s^{-1}p_1 s$ commute, moreover, the additional relation

$$p_1^{-1} = s^{-1} p_2 s = s^{-2} p_1 s^2$$

holds. We get the presentation

$$\text{(1)} \qquad \mathbf{P4_1} = (s, p - 1 = ps^{-1}psp^{-1}s^{-1}p^{-1}s = s^{-2}ps^2p).$$

Using $s^{-1}ps=sp^{-1}s^{-1}$ from the second relation, the first one can be written also in another form, but we assert that the length sum of the relation system cannot be reduced. It would be interesting to decide whether these presentations are geometrically realizable or not, but we shall find a shorter one in the following.

2) The 4_1 screw motion $s_0:=s$ and the 2_1 screw motion $s_2=s^2p_1$ also gen-

erate **P4$_1$**. As before we can easily deduce

(2) $$\mathbf{P4_1} = (s_0, s_2 - 1 = s_0^{-4} s_2^2 = s_0 s_2 s_0 s_2^{-1} s_0 s_2^{-1} s_0 s_2^{-1})$$

which is geometrically realizable but it is not the shortest presentation.

3) The 4_1 screw motions $s_0 := s$ and $s_1 := sp_1$ give us the minimal presentation desired. From the presentation in Figure 1 we have $p_1 = s_0^{-1} s_1$, $p_2 = s_0^{-1} p_1 s_0 = s_0^{-2} s_1 s_0$ which commute:

$$1 = s_0^{-1} s_1 s_0^{-2} s_1 s_0 s_1^{-1} s_0 s_0^{-1} s_1^{-1} s_0^2 = s_0 s_1 s_0^{-2} s_1 s_0 s_1^{-2}.$$

Moreover, we have seen that $1 = s^{-2} p_1 s^2 p_1 = s_0^{-3} s_1 s_0 s_1$, hence $s_0^{-2} s_1 s_0 = s_0 s_1^{-1}$. Thus the previous equality is more simply: $1 = s_0 s_1 s_0 s_1^{-3}$.

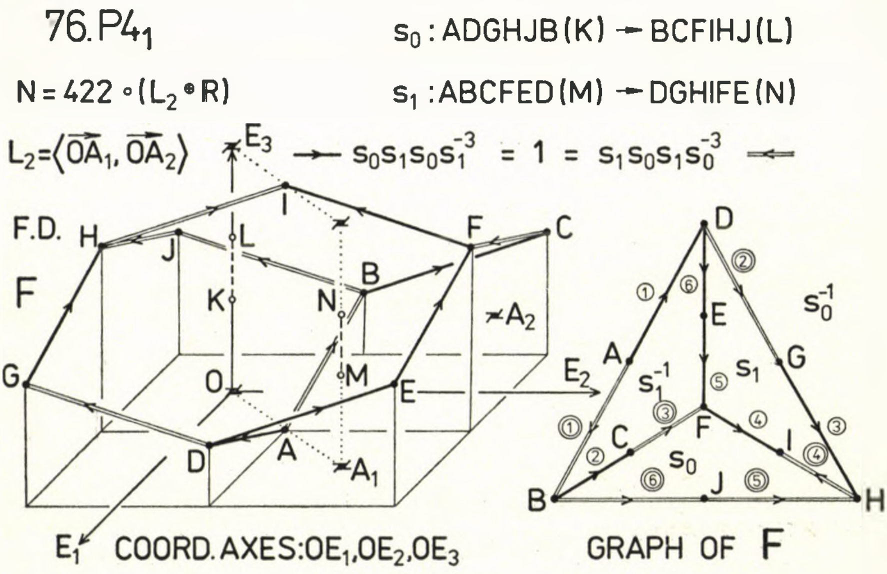


Fig. 2

Finally, we choose the presentation illustrated in Fig. 2. Here the starting point is $A\left(\frac{1}{4}; \frac{1}{4}; 0\right)$. The other points are:

$$B := A^{s_0}\left(-\frac{1}{4}; \frac{1}{4}; \frac{1}{4}\right), \quad J := A^{s_0^2}\left(-\frac{1}{4}; -\frac{1}{4}; \frac{1}{2}\right), \quad H := A^{s_0^3}\left(\frac{1}{4}; -\frac{1}{4}; \frac{3}{4}\right),$$

$$I := A^{s_0^4} = A^{s_1^4}\left(\frac{1}{4}; \frac{1}{4}; 1\right), \quad D := A^{s_1}\left(\frac{3}{4}; \frac{1}{4}; \frac{1}{4}\right),$$

$$E := A^{s_1^2}\left(\frac{3}{4}; \frac{3}{4}; \frac{1}{2}\right), \quad F := A^{s_1^3}\left(\frac{1}{4}; \frac{3}{4}; \frac{3}{4}\right),$$

$$C := D^{s_0} = A^{s_1 s_0}\left(-\frac{1}{4}; \frac{3}{4}; \frac{1}{2}\right), \quad G := B^{s_1} = A^{s_0 s_1}\left(\frac{3}{4}; -\frac{1}{4}; \frac{1}{2}\right).$$

The (curved) face $f_{s_0^{-1}}$ is defined to be the cone with vertex $K\left(0; 0; \frac{3}{8}\right)$ erected on the closed polygon $ADGHJB$. Analogously, $L:=K^{s_0}\left(0; 0; \frac{5}{8}\right)$ and $BCFIHJ$ define the face f_{s_0}. In the same manner will be defined the faces $f_{s_1^{-1}}$ and f_{s_1} indicated in Figure 2. We can check that the domain obtained is a topological polyhedron $\mathscr{F}$ whose graph as a Schlegel diagram is also described in Figure 2, but the f's are omitted from the face symbols. We have two equivalence classes of edge segments according to the two relations indicated. We illustrate the Poincaré cycles as follows:

$\rightarrow\!\!-$

$$
(3)\qquad
\begin{tikzcd}
(AD = f_{s_1^{-1}} \cap f_{s_0^{-1}}) \arrow[r, "s_0"] & (BC = f_{s_0} \cap f_{s_1^{-1}}) \arrow[r, "s_1"] & (GH = f_{s_1} \cap f_{s_0^{-1}}) \arrow[d, "s_0"] \\
(f_{s_1} \cap f_{s_1^{-1}} =: ED) \arrow[u, "s_1^{-1}"] & (f_{s_1} \cap f_{s_1^{-1}} =: FE) \arrow[l, "s_1^{-1}"'] & (f_{s_1} \cap f_{s_0} =: IF) \arrow[l, "s_1^{-1}"']
\end{tikzcd}
$$

$\Rightarrow\!\!=$

$$
(4)\qquad
\begin{tikzcd}
(AB = f_{s_0^{-1}} \cap f_{s_1^{-1}}) \arrow[r, "s_1"] & (DG = f_{s_1} \cap f_{s_0^{-1}}) \arrow[r, "s_0"] & (CF = f_{s_0} \cap f_{s_1^{-1}}) \arrow[d, "s_1"] \\
(f_{s_0} \cap f_{s_0^{-1}} =: JB) \arrow[u, "s_0^{-1}"] & (f_{s_0} \cap f_{s_0^{-1}} =: HJ) \arrow[l, "s_0^{-1}"'] & (f_{s_0} \cap f_{s_1} =: IH). \arrow[l, "s_0^{-1}"']
\end{tikzcd}
$$

(Imagine both written along circles, hence the name: cycle relations.)

Now, we deduce, e.g., that the edge segment AD, in the fundamental tiling is surrounded by the domain $\mathscr{F}$ and its images $\mathscr{F}s_0^{-1}$, $\mathscr{F}s_1^{-1}s_0^{-1}$, $\mathscr{F}s_0^{-1}s_1^{-1}s_0^{-1}$, $\mathscr{F}s_1s_0^{-1}s_1^{-1}s_0^{-1}$, $\mathscr{F}s_1s_1s_0^{-1}s_1^{-1}s_0^{-1}$. Finally $\mathscr{F}s_1s_1s_1s_0^{-1}s_1^{-1}s_0^{-1}=\mathscr{F}$ holds, since we know that $s_0s_1s_0s_1^{-3}=1$ is fulfilled now for $\mathbf{P4}_1$, but a shorter relation does not hold. Thus we get the domain $\mathscr{F}$ as a fundamental domain for the group $\mathbf{P4}_1$.

Let us remark that we have some freedom in choosing the starting point A, but the combinatorial structure of $\mathscr{F}$ is determined by the relation system selected and by the prescriptions 1)—6) for $\mathscr{F}$ in the introduction. In Figure 3 the combinatorial structure of $\mathscr{F}$ is illustrated in the usual way [7]. The generators determine the level of the 2-dimensional faces. The relations prescribe the level of the 1-dimensional edge segments and the incidences between faces and segments. For example, the Poincaré cycles (3), (4) require two common edge segments in $f_{s_0^{-1}} \cap f_{s_1^{-1}}$, which have a common vertex A, denote them by AD and AB. The isometry s_0 maps AD into $f_{s_0} \cap f_{s_1^{-1}}$ and AB into $f_{s_0} \cap f_{s_0^{-1}}$, therefore the s_0-image of A must be contained in the face cycle $(f_{s_0}, f_{s_0^{-1}}, f_{s_1^{-1}})$ by requirement 5). Then we may identify this image by B and we get the vertex figure of B as described in Figures 2 and 3. In the same way we obtain that s_1 maps AD into $f_{s_1} \cap f_{s_1^{-1}}$ and AB into $f_{s_1} \cap f_{s_0^{-1}}$ and we may identify the s_1-image of A by D as described in Figures 2 and 3, and so on. This process uniquely yields a combinatorial construction of the domain $\mathscr{F}$.

Of course, we will prove algebraically that the presentation in Figure 2 is in fact the shortest one. The proof will not be difficult but lengthy. It is based on a

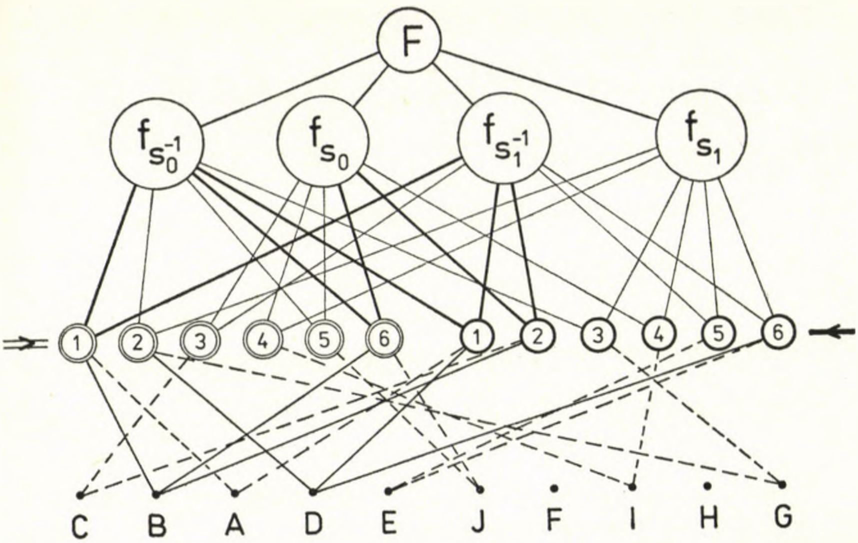


Fig. 3

general method, which works also in cases of the other groups considered. Now, we proceed some calculations in the group $G = \mathbf{P4_1}$ by using the matrix-vector presentation in **2**.

We have already learned from the theory of crystallographic groups the importance of the affine normalizer for a given space group [11, 12, 20]. This leads to the typization of the group elements and we can deduce that there are three types of two-generator presentations for the space group $\mathbf{P4_1}$. The normalizer of $\mathbf{P4_1}$ is

$$N = \mathbf{422} \circ (\mathbf{L}_2 \oplus \mathbf{R}), \tag{5}$$

a semi-direct product. The linear part **422** is generated by the 4-rotation s and the half-turn h defined by

$$\begin{pmatrix} \mathbf{e}_1 s \\ \mathbf{e}_2 s \\ \mathbf{e}_3 s \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{e}_1 h \\ \mathbf{e}_2 h \\ \mathbf{e}_3 h \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{6}$$

the translational part is a direct sum where $\mathbf{L}_2$ is a Z-lattice generated by $\overrightarrow{OA_1} := \mathbf{a}_1 = \frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2$ and $\overrightarrow{OA_2} := \mathbf{a}_2 = -\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2$ (Fig. 2):

$$\mathbf{L}_2 = \{l^1 \mathbf{a}_1 + l^2 \mathbf{a}_2 \colon l^1, l^2 \in \mathrm{Z} \text{ (integers)}\}, \tag{7}$$

$$\mathbf{R} = \{c\mathbf{e}_3 \colon c \in \mathrm{R} \text{ (real numbers)}\}. \tag{8}$$

i) At the first type we choose a screw motion and a translation:

$$\mathrm{s}'\left(s, \left(x+\frac{1}{4}\right) \mathbf{e}_3\right), \quad \mathrm{p}'(I, \mathbf{e}_1+y\mathbf{e}_3), \tag{9}$$

where the 4-rotation s is that of (6) or of **2** (2), x and y are integers for which $4x+1$ and y are relative primes:

$$(4x+1, y) = 1. \tag{10}$$

This last assertion and the translational component $\mathbf{e}_1$ appearing in p′ guarantee that s′ and p′ generate $\mathbf{P4}_1$, in fact, as we shall see. Using the affine normalizer N of $\mathbf{P4}_1$ in (5), we can easily see that every screw motion with 4-rotational part is equivalent to s′, moreover, an additional generating translation is affinely equivalent to p′ for appropriate x and y.

Now, we shall calculate the transformation g of the form

$$\mathrm{g} = \mathrm{s}'^{u_1}\mathrm{p}'^{v_1}\mathrm{s}'^{u_2}\mathrm{p}'^{v_2}\dots\mathrm{s}'^{u_i}\mathrm{p}'^{v_i} \tag{11}$$

with integer exponents, the natural index i will be chosen later. We apply the multiplication rule of affine transformations:

$$(a, \mathbf{a})(b, \mathbf{b}) = (ab, \mathbf{a}b+\mathbf{b}), \quad (a, \mathbf{a})^{-1} = (a^{-1}, -\mathbf{a}a^{-1}),$$
$$(a, \mathbf{a})^{u} = (a^{u}, \mathbf{a}a^{u-1}+\mathbf{a}a^{u-2}+\dots+\mathbf{a}a+\mathbf{a}), \tag{12}$$

if u positive integer,

$$(a, \mathbf{a})^{-u} = ((a, \mathbf{a})^{-1})^{u} = (a^{-u}, -\mathbf{a}a^{-u}-\dots-\mathbf{a}a^{-1}).$$

We assert that the equations

$$(\mathbf{a}a^{u-1}+\dots+\mathbf{a}a+\mathbf{a})(a-I) = \mathbf{a}(a^{u}-I),$$
$$(-\mathbf{a}a^{-u}-\dots-\mathbf{a}a^{-2}-\mathbf{a}a^{-1})(a-I) = \mathbf{a}(a^{-u}-I) \tag{13}$$

uniformly hold for the translational parts. This will be useful later. Now, we turn to calculation of g in (11):

$$\mathrm{g}\left(s^{u_1+u_2+\dots+u_i}, \left((u_1+u_2+\dots+u_i)\left(x+\frac{1}{4}\right)+(v_1+\dots+v_i)y\right)\mathbf{e}_3+\right.$$
$$\left.+\mathbf{e}_1(v_1 s^{u_2+\dots+u_i}+v_2 s^{u_3+\dots+u_i}+\dots+v_{i-1}s^{u_i}+v_i I)\right). \tag{14}$$

First let $i=1$. By the assumption (10) we can reach

$$u_1(4x+1)+v_1(4y) = 1.$$

Then $\mathrm{g}\left(s, v_1\mathbf{e}_1+\frac{1}{4}\mathbf{e}_3\right)$ is a 4_1 screw motion and $\mathrm{g}^{-4y}\mathrm{p}'$ is the translation p_1, hence s′ and p′ in (9) generate $\mathbf{P4}_1$, in fact.

Secondly, we want to solve the equation $\mathrm{g}=1$ for $u_1, \dots, u_i, v_1, \dots, v_i$, each of them is different from zero, and to minimalize the length of g

$$|u_1|+|v_1|+|u_2|+|v_2|+\dots+|u_i|+|v_i| \tag{15}$$

for every index i. This would be a difficult task, in general. Fortunately, we have known a presentation (1) corresponding to our case $x=0$ and $y=0$. To show that this is minimal in the first type, we can restrict ourselves up to $i=4$. From (14) it follows that

$$g = 1 \quad \text{iff} \tag{16}$$

$$u_1+u_2+\dots+u_i = 4k \quad (k \text{ integer})$$

and

$$k(4x+1)+(v_1+v_2+\dots+v_i)y = 0$$

and

$$\mathbf{e}_1(v_1 s^{-u_1}+v_2 s^{-u_1-u_2}+\dots+v_{i-1}s^{-u_1-u_2\dots-u_{i-1}}+v_i I) = \mathbf{0}$$

hold under the assumptions mentioned.

The most tedious task is to check the last matrix equation but now we have to try only finite many cases, a lot of them are equivalent from the viewpoint of the problem (which would lead to equivalent relations). Indeed, we get $(u_1, u_2, u_3, u_4)=$ $=(1, -1, 1, -1)$, $(v_1, v_2, v_3, v_4)=(1, 1, -1, -1)$ or $=(1, -1, -1, 1)$ and $(u_1, u_2)=$ $=(2, -2)\sim(-2, 2)$, $(v_1, v_2)=(1, 1)\sim(-1, -1)$ providing the presentation (1) as a minimal one in the first type, when the generators in (9) are chosen.

ii) For second type of two generator presentation of $\mathbf{P4_1}$ we consider the extended case of 2):

$$s_0'\left(s, \left(x+\frac{1}{4}\right)\mathbf{e}_3\right), \quad s_2'\left(s^2, \mathbf{e}_1+\left(y+\frac{1}{2}\right)\mathbf{e}_3\right) \tag{17}$$

with the 4-rotation s of (6) and with appropriate integers x and y:

$$(4x+1, 2y+1) = 1. \tag{18}$$

Now, calculate the transformation

$$g = s_0'^{u_1} s_2'^{v_1} s_0'^{u_2} s_2'^{v_2} \dots s_0'^{u_i} s_2'^{v_i} \tag{19}$$

with integer exponents, again.

$$\begin{aligned} g\Big(s^{u_1+2v_1+u_2+2v_2+\dots+u_i+2v_i}, &\left(\left(x+\frac{1}{4}\right)(u_1+u_2+\dots+u_i)+\left(y+\frac{1}{2}\right)\cdot\right.\\ \cdot(v_1+v_2+\dots+v_i)\Big)\mathbf{e}_3+\mathbf{e}_1&\big((s^{2(v_1-1)}+\dots+s^2+I)s^{u_2}s^{2v_2}\dots s^{u_i}s^{2v_i}+\\ &+(s^{2(v_2-1)}+\dots+s^2+I)s^{u_3}s^{2v_3}\dots s^{u_i}s^{2v_i}+\dots\\ &\dots+(s^{2(v_{i-1}-1)}+\dots+s^2+I)s^{u_i}s^{2v_i}+(s^{2(v_i-1)}+\dots+s^2+I)\big)\Big) \end{aligned} \tag{20}$$

holds if $v_1, v_2, \dots, v_i$ are positive integers. If not, the sums of linear transformations, acting on the vector $\mathbf{e}_1$, will change according to (12). Fortunately, this $\mathbf{e}_1$-component will be the zero vector, iff its image at the linear transformation (s^2-I) vanishes, and then we can apply (13). Then this $\mathbf{e}_1$-component will be

$$\begin{aligned} &\mathbf{e}_1\big((s^{2v_1}-I)s^{u_2}s^{2v_2}\dots s^{u_i}s^{2v_i}+(s^{2v_2}-I)s^{u_3}s^{2v_3}\dots\\ &\dots s^{u_i}s^{2v_i}+\dots+(s^{2v_{i-1}}-I)s^{u_i}s^{2v_i}+(s^{2v_i}-I)\big) =\\ &= \mathbf{e}_1(s^{2v_1+u_2+2v_2+\dots+u_i+2v_i}-s^{u_2+2v_2+\dots+u_i+2v_i}+s^{2v_2+\dots+u_i+2v_i}-\\ &-s^{u_3+2v_3+\dots+u_i+2v_i}+\dots+s^{2v_{i-1}+u_i+2v_i}-s^{u_i+2v_i}+s^{2v_i}-I). \end{aligned} \tag{21}$$

Our purpose is again to solve the equation $g=1$ for non-zero integers $u_1, \ldots, u_i$ and $v_1, \ldots, v_i$, and to minimalize the length of g as in (15) for every index i. From (20), (21) we have the criterion

$$g = 1 \quad \text{iff} \tag{22}$$

$$u_1+2v_1+u_2+2v_2+\ldots+u_i+2v_i = 4k \quad (k \text{ integer})$$

and

$$k(4x+1)+(y-2x)(\ _1+v_2+\ldots+v_i) = 0$$

and

$$\mathbf{e}_1(s^{-u_1}-s^{-u_1-2v_2}+s^{-u_1-2v_2-u_2}-s^{-u_1-2v_2-u_2-2v_2}+\ldots$$
$$\ldots+s^{-u_1-2v_1-u_2-2v_2-\cdots-u_i}-I) = \mathbf{0}$$

hold under the assumption mentioned. We can check that $(u_1, v_1)=(-4, 2)$ and $(u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4)=(1, 1, 1, -1, 1, -1, 1, -1)$ are solutions according to the presentation (2). The systematic procedure up to $i=4$ shows that this is the minimal one in the type ii) of the form (17). We can show again that every presentation of $\mathbf{P4}_1$ with 2 screw motion generators of rotational components s and s^2, respectively, will be equivalent to (17) by the affine normalizer N in (5) for appropriate x and y satisfying (18).

iii) The third type of two generator presentation for $\mathbf{P4}_1$ is the extension of the case 3):

$$s_0'\left(s, \left(x+\frac{1}{4}\right)\mathbf{e}_3\right), \quad s_1'\left(s, \mathbf{e}_1+\left(y+\frac{1}{4}\right)\mathbf{e}_3\right) \tag{23}$$

with the 4-rotation s of (6) and integers x, y satisfying

$$(4x+1, 4y+1) = 1. \tag{24}$$

Let us calculate the transformation

$$g = s_0'^{u_1} s_1'^{v_1} s_0'^{u_2} s_1'^{v_2} \ldots s_0'^{u_i} s_1'^{v_i} \tag{25}$$

with integer non-zero exponents, again.

$$\begin{aligned} g\Bigg(&s^{u_1+v_1+u_2+v_2+\cdots+u_i+v_i}, \left(\left(x+\frac{1}{4}\right)(u_1+u_2+\ldots+u_i)+\left(y+\frac{1}{4}\right)\cdot\right. \\ &\cdot(v_1+v_2+\ldots+v_i)\Big)\mathbf{e}_3+\mathbf{e}_1\big((s^{(v_1-1)}+\ldots+s+I)s^{u_2}s^{v_2}\ldots s^{u_i}s^{v_i}+ \\ &+(s^{(v_2-1)}+\ldots+s+I)s^{u_3}s^{v_3}\ldots s^{u_i}s^{v_i}+\ldots+(s^{(v_{i-1}-1)}+\ldots+s+I)s^{u_i}s^{v_i}+ \\ &+(s^{(v_i-1)}+\ldots+s+I)\big)\Bigg) \end{aligned} \tag{26}$$

holds if $v_1, v_2, \ldots, v_i$ are positive integers. Otherwise the linear transformation acting on $\mathbf{e}_1$ changes according to (12). This $\mathbf{e}_1$-component is the zero vector, iff its $(s-I)$-image vanishes and we apply (13) again. Then this $\mathbf{e}_1$-component will be

$$\begin{aligned} &\mathbf{e}_1\big((s^{v_1}-I)s^{u_2}s^{v_2}\ldots s^{u_i}s^{v_i}+(s^{v_2}-I)s^{u_3}s^{v_3}\ldots s^{u_i}s^{v_i}+\ldots \\ &\ldots+(s^{v_{i-1}}-I)s^{u_i}s^{v_i}+(s^{v_i}-I)\big) = \mathbf{e}_1(s^{v_1+u_2+v_2+\cdots+u_i+v_i}-s^{u_2+v_2+\cdots+u_i+v_i}+ \\ &+s^{v_2+\cdots+u_i+v_i}-s^{u_3+v_3+\cdots+u_i+v_i}+\ldots+s^{v_{i-1}+u_i+v_i}-s^{u_i+v_i}+s^{v_i}-I). \end{aligned} \tag{27}$$

The purpose is again to solve the equation $g=1$ for non-zero integers $u_1, \ldots, u_i$ and $v_1, \ldots, v_i$ and to minimalize the length of g as in (15) for every index i. From (26) and (27) we calculate

(28) $$g = 1 \quad \text{iff}$$

$$u_1+v_1+u_2+v_2+\ldots+u_i+v_i = 4k \quad (k \text{ integer})$$

and

$$k(4x+1)+(y-x)(v_1+v_2+\ldots+v_i) = 0$$

and

$$e_1(s^{-u_1}-s^{-u_1-v_1}+s^{-u_1-v_1-u_2}-s^{-u_1-v_1-u_2-v_2}+\ldots+s^{-u_1-v_1-u_2-v_2-\cdots-u_i}-I) = 0.$$

We can check that $(u_1, v_1, u_2, v_2)=(1, 1, 1, -3)$ and $=(-3, 1, 1, 1)$ are solutions according to the presentation in Figure 2. We have not other essentially different solution up to $i=3$. That means this is the absolute minimal presentation for the group $\mathbf{P4_1}$. Of course we can show again that every presentation of $\mathbf{P4_1}$ with 2 screw motion generators, both of 4-rotational component s, is affinely equivalent to (23) for integers x and y satisfying (24).

We have no more possibility. For example, a transformation

$$\bar{s}\left(s^3, xe_1+ye_2+\left(z+\frac{3}{4}\right)e_3\right)$$

belonging to $\mathbf{P4_1}$ has its inverse in the form $\bar{s}^{-1}\left(s, -x(e_1s)-y(e_2s)-\left(z+\frac{3}{4}\right)\cdot(e_3s)\right)$ with respect to (12) and to $s^4=I$. By definition **2** (2) we get

$$\bar{s}^{-1}\left(s, -xe_2+ye_1+\left(\frac{1}{4}-(z+1)\right)e_3\right),$$

and this is affinely equivalent (by an element of the affine normalizer N of $\mathbf{P4_1}$) to s′ in (9) with $x=-(z+1)$. The computation will not be detailed anymore.

After this relatively detailed treatment we shall discuss the other space forms more sketchily because we can follow the same sequence of thought.

4. The space form $E^3/\mathbf{P1}$

The usual fundamental parallelepipedon is described in Figure 4. The opposite faces are identified by parallel translations

$$p_1(I, e_1), \quad p_2(I, e_2), \quad p_3(I, e_3)$$

and we have 3 equivalence classes of edge segments, each of them contains 4 segments. The relations belonging to the Poincaré cycles express the commutativity of translations. The combinatorial structure of the domain $\mathcal{F}$ is unique. Affinely equivalent generator system can be obtained by means of the affine normalizer N of the space group **P1**:

(1) $$N = SL(3, Z)\circ\mathbf{R}^3$$

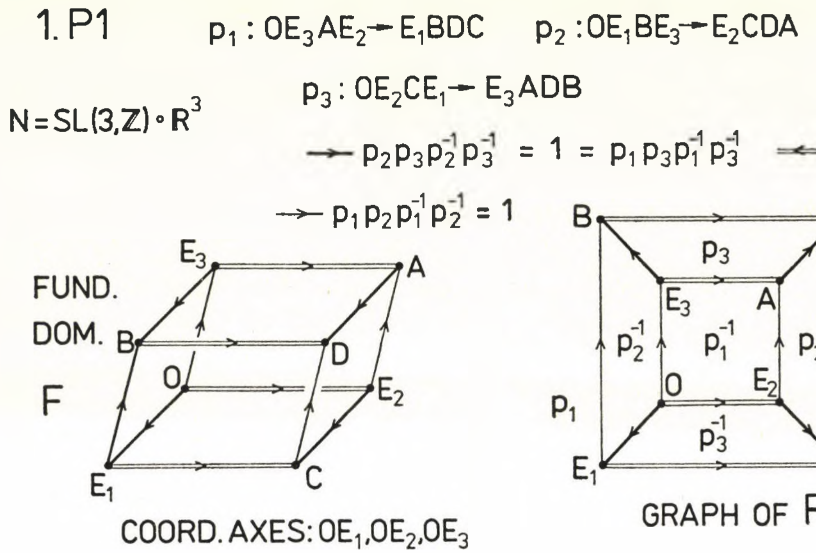


Fig. 4

as a usual semi-direct product. Here $SL(3, Z)$ denotes the group of integer unimodular transformations leaving the lattice $\mathbf{L}_{P1}=\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\rangle$ invariant. The translational part of N is

(2) $$\mathbf{R}^3=\{x^1\mathbf{e}_1+x^2\mathbf{e}_2+x^3\mathbf{e}_3\colon\ x^1, x^2, x^3\in R\}.$$

5. The space form $E^3/\mathbf{P}2_1$

The monoclinic lattice $\mathbf{L}_G$ for the space group $G=\mathbf{P}2_1$ is generated by $\overrightarrow{OE_i}:=\mathbf{e}_i$ as in Figures 5 and 6, where

(1) $$(\mathbf{e}_1\,;\,\mathbf{e}_3)=(\mathbf{e}_2\,;\,\mathbf{e}_3)=0.$$

The generating 2_1 screw motion $\mathrm{s}(s, \mathbf{s})$ is given by

(2) $$\begin{pmatrix}\mathbf{e}_1 s\\ \mathbf{e}_2 s\\ \mathbf{e}_3 s\end{pmatrix}=\begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \mathbf{e}_3\end{pmatrix}$$

and by the vector $\mathbf{s}:=\overrightarrow{OO^s}$

(3) $$\mathbf{s}=\left(1\ \ 1\ \ \frac{1}{2}\right)\begin{pmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \mathbf{e}_3\end{pmatrix}.$$

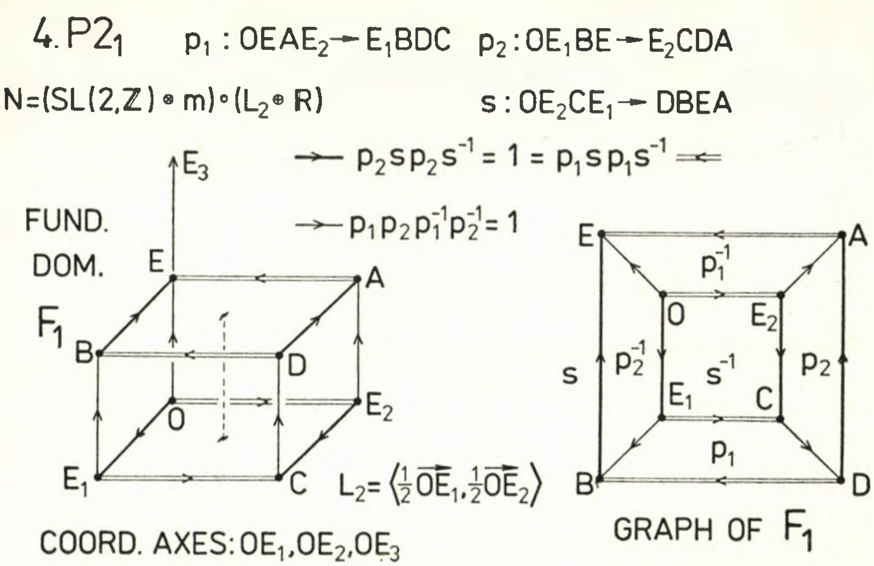

4. $P2_1$
$p_1 : OEAE_2 \to E_1BDC$
$p_2 : OE_1BE \to E_2CDA$
$N = (SL(2,\mathbb{Z}) \circledast m) \circ (L_2 \circledast R)$
$s : OE_2CE_1 \to DBEA$
$\to p_2 s p_2 s^{-1} = 1 = p_1 s p_1 s^{-1} \leftarrow$
$\to p_1 p_2 p_1^{-1} p_2^{-1} = 1$
FUND.
DOM.
F_1
E_3
E
A
B
D
O
E_2
E_1
C
$L_2 = \langle \frac{1}{2}\overrightarrow{OE_1}, \frac{1}{2}\overrightarrow{OE_2} \rangle$
COORD. AXES: OE_1, OE_2, OE_3
p_1^{-1}
p_2^{-1}
s^{-1}
p_2
s
p_1
GRAPH OF F_1

Fig. 5

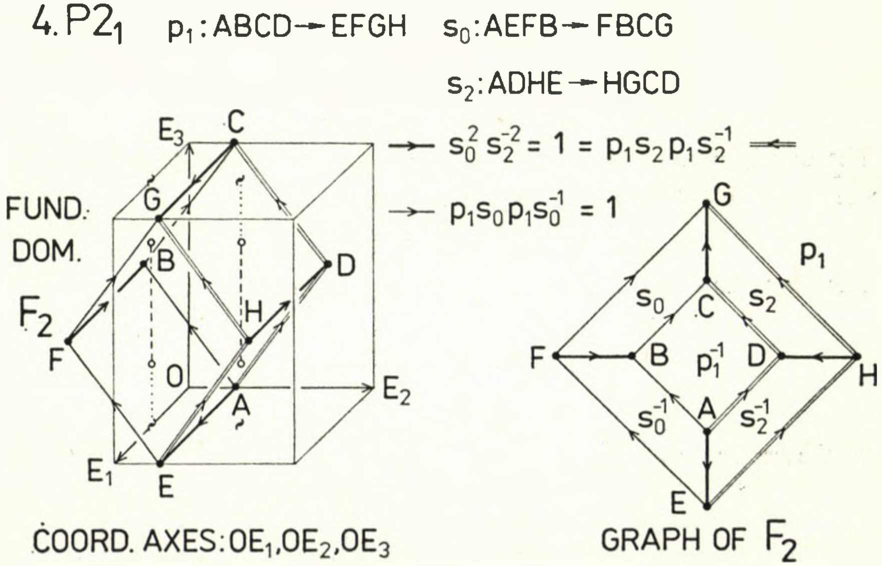

4. $P2_1$
$p_1 : ABCD \to EFGH$
$s_0 : AEFB \to FBCG$
$s_2 : ADHE \to HGCD$
$\to s_0^2 s_2^{-2} = 1 = p_1 s_2 p_1 s_2^{-1} \leftarrow$
$\to p_1 s_0 p_1 s_0^{-1} = 1$
FUND.
DOM.
F_2
E_3
C
G
B
D
H
F
O
A
E_2
E_1
E
COORD. AXES: OE_1, OE_2, OE_3
p_1
s_0
s_2
p_1^{-1}
s_0^{-1}
s_2^{-1}
GRAPH OF F_2

Fig. 6

Hence $p_3(I, e_3)=s^2$ and we have a 3-generator presentation in Figure 5 according to a minimal domain $\mathcal{F}_1$ with unique combinatorial structure.

Now let $s_2:=s$, $s_0:=sp_2^{-1}$ and p_1 be a new generator system. Then $p_2=s_0^{-1}s_2$ and we get a new presentation according to the domain $\mathcal{F}_2$ in Figure 6, which is also minimal in the sense defined earlier.

As a third possibility we can choose $s_1:=s_0p_1$ instead of p_1 to get the presentation

(4) $$\mathbf{P2}_1 = (s_0, s_1, s_2 - 1 = s_0^{-2}s_1^2 = s_0^{-2}s_2^2 = s_0^{-1}s_1s_2^{-1}s_0s_1^{-1}s_2),$$

which can also be written into other forms but these are not shorter. Thus we have two minimal presentations for the group $\mathbf{P2}_1$, only. Other minimal generator collections are equivalent to these ones by the affine normalizer N of $\mathbf{P2}_1$. We have

(5) $$N = (SL(2, Z)\otimes m)\circ(\mathbf{L}_2\oplus\mathbf{R})$$

as a semi direct product, where the translational part is a direct sum of the Z-lattice

(6) $$\mathbf{L}_2 := \left\langle \frac{1}{2}\mathbf{e}_1, \frac{1}{2}\mathbf{e}_2 \right\rangle$$

generated by $\mathbf{a}_1=\frac{1}{2}\mathbf{e}_1$ and $\mathbf{a}_2=\frac{1}{2}\mathbf{e}_2$ and the R-lattice

(7) $$\mathbf{R} := \{c\mathbf{e}_3 \colon c\in R\};$$

the linear part is a direct product where $SL(2, Z)$ denotes the group of integer unimodular transformations leaving the lattice $\mathbf{L}_2$ and the vector $\mathbf{e}_3$ invariant, m denotes the group generated by the linear map m defined by $\mathbf{e}_1m=\mathbf{e}_1$ and $\mathbf{e}_2m=\mathbf{e}_2$, $\mathbf{e}_3m=-\mathbf{e}_3$.

6. The space form $E^3/\mathbf{Pb}$

The primitive monoclinic lattice $\mathbf{L}_G$ for the space group $G=\mathbf{Pb}$ is generated by $\overrightarrow{OE_i}:=\mathbf{e}_i$ as before (see Figure 7—9), where:

(1) $$(\mathbf{e}_1; \mathbf{e}_3) = (\mathbf{e}_2; \mathbf{e}_3) = 0.$$

The generating glide reflection $b(b, \mathbf{b})$ is given by

(2) $$\begin{pmatrix} \mathbf{e}_1 b \\ \mathbf{e}_2 b \\ \mathbf{e}_3 b \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}$$

and by the vector $\mathbf{b}:=\overrightarrow{OO^b}$:

$$\mathbf{b} = \begin{pmatrix} 0 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}.$$

7. Pb $\quad p_1 : OE_3AE \to E_1BDC \qquad p_3 : OECE_1 \to E_3ADB$

$N = (FL(2,\mathbf{Z}) \bullet m) \circ (R^2 \bullet L_1) \qquad b : OE_1BE_3 \to ADCE$

$$p_3 b p_3 b^{-1} = 1 = p_1 p_3 p_1^{-1} p_3^{-1}$$

$$p_1 b p_1^{-1} b^{-1} = 1$$

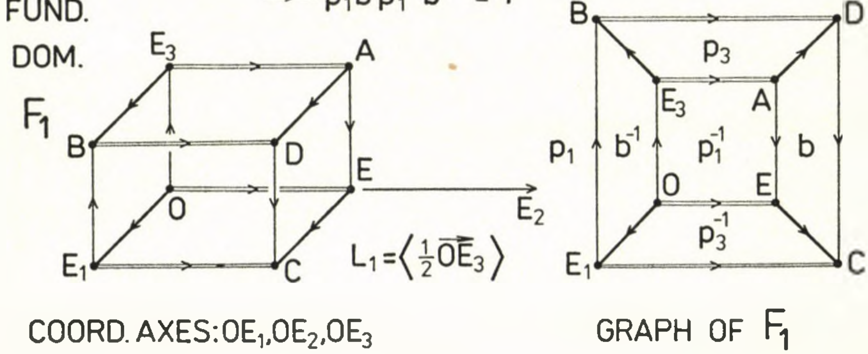


Fig. 7

7. Pb $\quad p_1 : OAE_2C \to E_1BED \qquad b_0 : OCDE_1 \to CE_2ED$

$b_3 : OE_1BA \to ABEE_2$

$$b_0^2 b_3^{-2} = 1 = p_1 b_0 p_1^{-1} b_0^{-1}$$

$$p_1 b_3 p_1^{-1} b_3^{-1} = 1$$

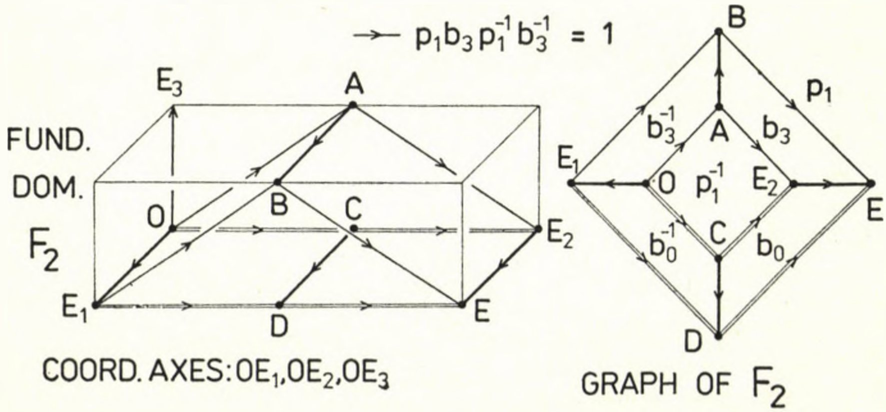


Fig. 8

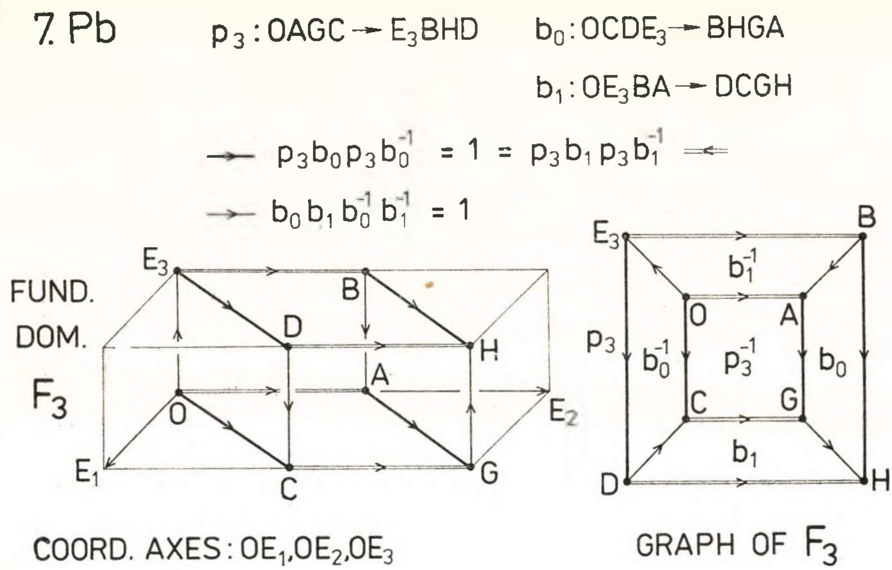


Fig. 9

Hence $p_2(I, e_2)=b^2$ and we have the first 3-generator presentation in Figure 7, the minimal domain $\mathcal{F}_1$ has unique combinatorial structure.

Let $b_3:=b$, $b_0:=bp_3^{-1}$, p_1 be a new system of generators. Then $p_3=b_0^{-1}b_3$ and we get the second minimal presentation according to the domain $\mathcal{F}_2$ in Figure 8.

Now let $b_0:=b$, $b_1:=bp_1$, p_3 be the third system of generators. Then $p_1=b_0^{-1}b_1$ and we get the minimal presentation according to $\mathcal{F}_3$ in Figure 9.

We get a new presentation of **Pb**, if we take $b_3:=b_0p_3$ in the previous one. This 3-generator presentation

$$\text{(4)}\qquad \mathbf{Pb}=(b_0,\, b_1,\, b_3 - 1 = b_0 b_1 b_0^{-1} b_1^{-1} = b_0^{-2} b_3^2 = b_0 b_3^{-1} b_1 b_0^{-1} b_3 b_1^{-1})$$

will not be a minimal one, although it can be written in another form, too.

Other collections of generators for minimal presentations are equivalent to these three ones by the affine normalizer N of **Pb**. We have

$$\text{(5)}\qquad N=(FL(2, Z)\otimes m)\circ(\mathbf{R}^2\oplus\mathbf{L}_1)$$

as a semi-direct product. In the usual representation

$$\text{(6)}\qquad N=\left\{\begin{pmatrix} n_1^1 & n_1^2 & 0\\ n_2^1 & n_2^2 & 0\\ 0 & 0 & \pm 1\end{pmatrix},\ (n^1\ \ n^2\ \ n^3)\right\}$$

where $n_i^j\in Z$; $\det(n_i^j)=\pm 1$; n_1^1 and n_2^2 odd; n_2^1 even; $n^1, n^2\in R$; $n^3\in Z/2$.

7. The space form $E^3/\mathbf{Bb}$

The B-centred monoclinic lattice $\mathbf{L}_G$ for the space group $G=\mathbf{Bb}$ can be derived from the primitive one of case **6**. We choose the base of $\mathbf{L}_G$ as follows:

$$\text{(1)}\qquad \begin{aligned}&\overrightarrow{OB_1} := \mathbf{b}_1 = \mathbf{e}_1, \quad \overrightarrow{OB_2} := \mathbf{b}_2 = \mathbf{e}_2,\\ &\overrightarrow{OB_3} := \mathbf{b}_3 = \frac{1}{2}(\mathbf{e}_1+\mathbf{e}_3).\end{aligned}$$

It may be surprising, but we have a 2-generator presentation for **Bb** described in Figure 10, which will be also a minimal one.

The generating glide reflections $\mathrm{b}(b, \mathbf{b})$ is given by

$$\text{(2)}\qquad \begin{pmatrix}\mathbf{b}_1 b\\ \mathbf{b}_2 b\\ \mathbf{b}_3 b\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & -1\end{pmatrix}\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{pmatrix}$$

and by the vector

$$\mathbf{b} := \overrightarrow{OO^{\mathrm{b}}}:$$

$$\text{(3)}\qquad \mathbf{b} = \begin{pmatrix}-1 & \frac{1}{2} & 2\end{pmatrix}\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{pmatrix},$$

hence $\mathrm{p}_2(I, \mathbf{b}_2)=b^2$ is a translation.

The second generating glide reflection, denoted, as usual, by $\mathrm{n}(n, \mathbf{n})$, is given by the linear transformation $n:=b$ in (2) and by the vector $\mathbf{n}:=\overrightarrow{OO^{\mathrm{n}}}$:

$$\text{(4)}\qquad \mathbf{n} = \begin{pmatrix}-1 & \frac{1}{2} & 3\end{pmatrix}\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{pmatrix},$$

hence $\mathrm{p}_1(I, \mathbf{b}_1)=\mathrm{n}^2\mathrm{b}^{-2}$ and $\mathrm{p}_3(I, \mathbf{b}_3)=\mathrm{b}^{-1}\mathrm{n}$ are translations. The combinatorial structure of $\mathcal{F}$ in Fig. 10 is unique again. The coordinates of the vertices in the base $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are

$$J\left(\frac{1}{4}; \frac{1}{2}; \frac{1}{8}\right), \quad B\left(0; 0; \frac{3}{8}\right), \quad H\left(\frac{1}{2}; 1; \frac{3}{8}\right),$$

$$A\left(0; -\frac{1}{2}; \frac{5}{8}\right), \quad C\left(0; \frac{1}{2}; \frac{5}{8}\right), \quad G\left(\frac{1}{2}; \frac{1}{2}; \frac{5}{8}\right),$$

$$I\left(\frac{1}{2}; \frac{3}{2}; \frac{5}{8}\right), \quad D\left(\frac{1}{4}; 0; \frac{7}{8}\right), \quad F\left(\frac{1}{4}; 1; \frac{7}{8}\right),$$

$$E\left(\frac{1}{4}; \frac{1}{2}; \frac{9}{8}\right).$$

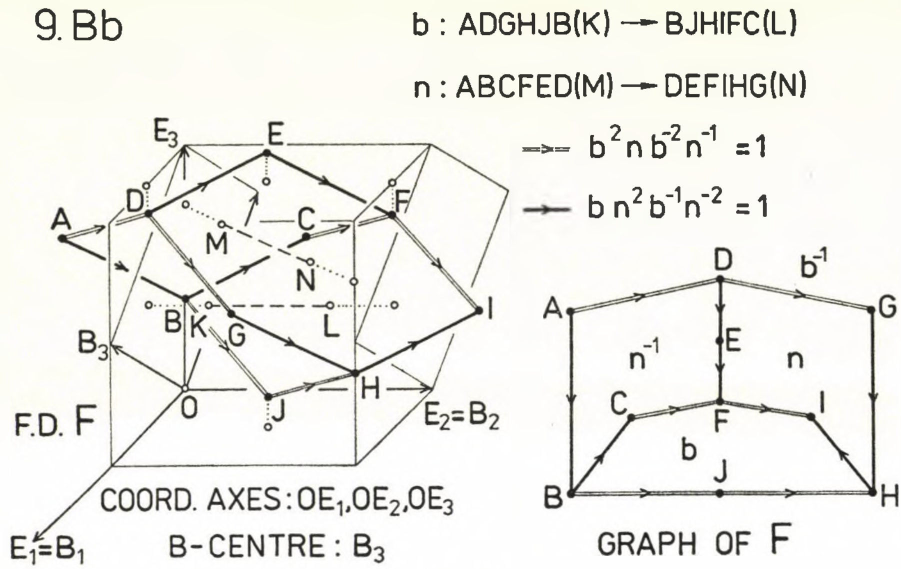


Fig. 10

For the construction of faces we have chosen the points

$$K\left(\frac{1}{4};\ \frac{1}{4};\ \frac{1}{2}\right),\quad L\left(\frac{1}{4};\ \frac{3}{4};\ \frac{1}{2}\right),\quad M\left(\frac{1}{8};\ \frac{1}{4};\ \frac{3}{4}\right),\quad N\left(\frac{3}{8};\ \frac{3}{4};\ \frac{3}{4}\right).$$

Another 2-generator presentation is

$$(5)\qquad \mathbf{Bb} = (\mathrm{b},\ \mathrm{p}_3 - \mathrm{b}^2\mathrm{p}_3\mathrm{b}^{-2}\mathrm{p}_3^{-1} = 1 = \mathrm{b}\mathrm{p}_3\mathrm{b}\mathrm{p}_3\mathrm{b}^{-1}\mathrm{p}_3^{-1}\mathrm{b}^{-1}\mathrm{p}_3^{-1})$$

whose length sum cannot be reduced, it is not minimal.

Other minimally presenting 2-generator collections are equivalent to the given ones by the affine normalizer N of **Bb**. In the base $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$, N has the following matrix-vector representation

$$(6)\qquad N = \left\{\begin{pmatrix} n_1^1 & n_1^2 & 0 \\ n_2^1 & n_2^2 & 0 \\ \frac{1}{2}(n_1^1 \pm 1) & \frac{1}{2} n_1^2 & \pm 1 \end{pmatrix},\ (n^1\ n^2\ n^3)\right\},$$

where $n_i^j \in Z$; $\det(n_i^j) = \pm 1$; n_1^1 and n_2^2 odd; n_1^2 even; $n^1, n^2 \in R$; $n^3 \in Z/2$ so that $\left(\frac{1}{2} n_2^1 - n^3\right) \in Z$.

8. The space form $E^3/\mathbf{P2_12_12_1}$

The primitive orthorhombic lattice $\mathbf{L}_G$ for the space group $G=\mathbf{P2_12_12_1}$ is spanned by the orthogonal base $\mathbf{e}_i$ in Figure 11,

(1) $$(\mathbf{e}_1;\ \mathbf{e}_2)=(\mathbf{e}_2;\ \mathbf{e}_3)=(\mathbf{e}_3;\ \mathbf{e}_1)=0.$$

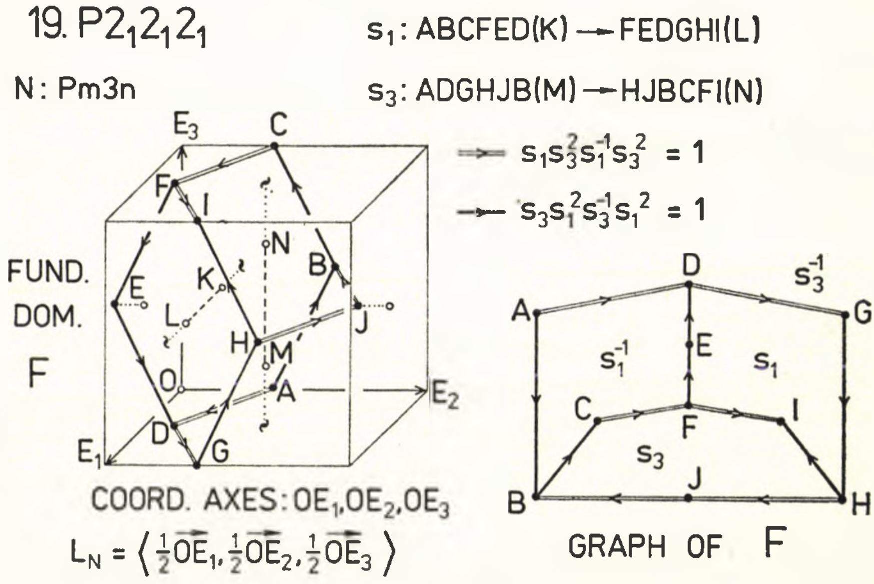


Fig. 11

We choose the 2_1 screw motion $\mathbf{s}_1(s_1, \mathbf{s}_1)$ defined by

$$\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\end{pmatrix},\quad \begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 1\end{pmatrix}$$

and the 2_1 screw motion $\mathbf{s}_3(s_3, \mathbf{s})$ defined by

$$\begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad \begin{pmatrix}1 & 1 & \frac{1}{2}\end{pmatrix}$$

as usual. So we get the minimal presentation described in Figure 11. The combinatorial structure of the fundamental domain $\mathscr{F}$ is uniquely determined. We have chosen the coordinates of the vertex E to be $\left(\frac{1}{2};\ -\frac{1}{8};\ \frac{1}{2}\right)$, then the other vertices of $\mathcal{F}$ are

$$D\left(\frac{1}{2};\frac{1}{8};0\right),\quad F\left(\frac{1}{2};\frac{1}{8};1\right),\quad A\left(0;\frac{3}{8};0\right),$$
$$C\left(0;\frac{3}{8};1\right),\quad G\left(1;\frac{3}{8};0\right),\quad I\left(1;\frac{3}{8};1\right),$$
$$B\left(0;\frac{5}{8};\frac{1}{2}\right),\quad H\left(1;\frac{5}{8};\frac{1}{2}\right),\quad J\left(\frac{1}{2};\frac{7}{8};\frac{1}{2}\right).$$

For the construction of faces for $\mathcal{F}$ we have taken the points:

$$K\left(\frac{1}{4};\frac{1}{4};\frac{1}{2}\right),\quad L\left(\frac{3}{4};\frac{1}{4};\frac{1}{2}\right),\quad M\left(\frac{1}{2};\frac{1}{2};\frac{1}{4}\right),\quad N\left(\frac{1}{2};\frac{1}{2};\frac{3}{4}\right).$$

The other 2-generator presentations are equivalent to this one by the affine normalizer N of $\mathbf{P2_12_12_1}$. We know that N is affinely equivalent to the space group **Pm3n,** whose lattice $\mathbf{L}_N$ is spanned by $\frac{1}{2}\mathbf{e}_1, \frac{1}{2}\mathbf{e}_2, \frac{1}{2}\mathbf{e}_3$. We do not discuss details here [11, 14].

9. The space form $E^3/\mathbf{Pca2_1}$

The primitive orthorhombic lattice $\mathbf{L}_G$ for the space group $G=\mathbf{Pca2_1}$, is spanned by the orthogonal base $\mathbf{e}_i$ as before (Fig. 12—15). We have got 4 minimal presentations by domains $\mathcal{F}_1-\mathcal{F}_4$ in Figures 12—15. We could follow the way as in the previous cases. We have to select 3 generators from 4 types of transformations. These are the translation $\mathrm{p}_2(I, \mathbf{e}_2)$, the glide reflection $\mathrm{a}(a, \mathbf{a})$ defined by the representation

$$\text{(1)}\qquad \begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad \left(\frac{1}{2}\ 1\ 0\right),$$

the glide reflection $\mathrm{c}(c, \mathbf{c})$ defined by

$$\text{(2)}\qquad \begin{pmatrix}-1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad \left(\frac{1}{2}\ 0\ \frac{1}{2}\right),$$

the 2_1 screw motion $\mathrm{s}(s, \mathbf{s})$ defined by

$$\text{(3)}\qquad \begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad \left(1\ 1\ \frac{1}{2}\right),$$

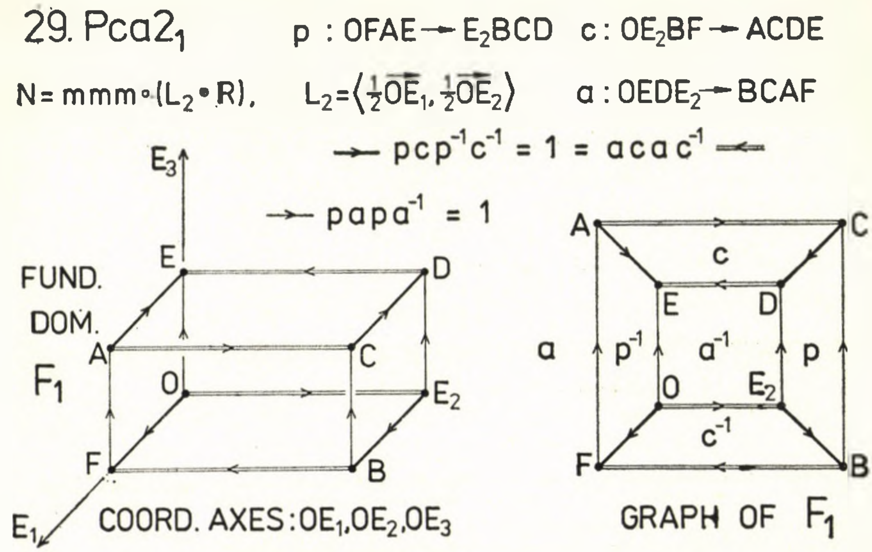

29. Pca2₁
p : OFAE → E₂BCD
c : OE₂BF → ACDE
N = mmm∘(L₂•R),
L₂ = ⟨½OE₁, ½OE₂⟩
a : OEDE₂ → BCAF
pcp⁻¹c⁻¹ = 1 = acac⁻¹
papa⁻¹ = 1
FUND. DOM. F₁
E₃
E
D
A
C
O
E₂
F
B
E₁
COORD. AXES: OE₁, OE₂, OE₃
A
C
c
E
D
a
p⁻¹
a⁻¹
p
O
E₂
c⁻¹
F
B
GRAPH OF F₁

Fig. 12

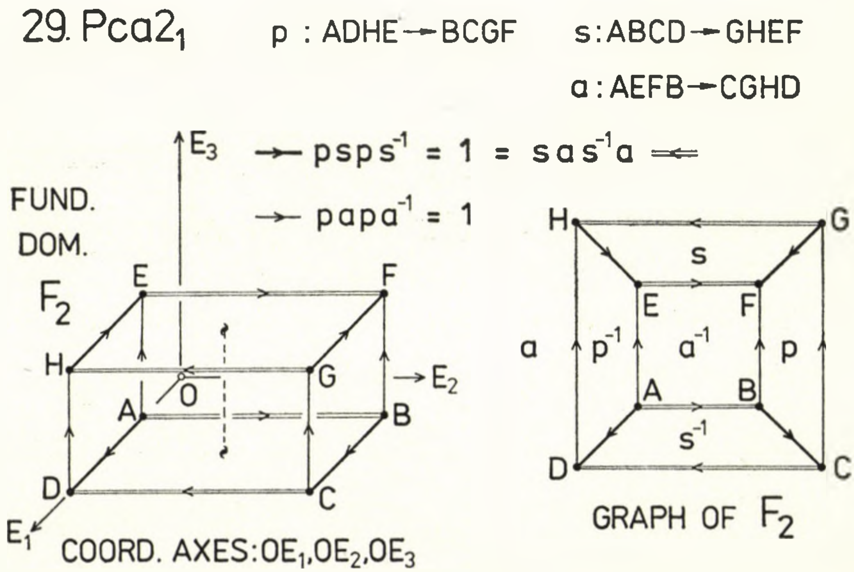

29. Pca2₁
p : ADHE → BCGF
s : ABCD → GHEF
a : AEFB → CGHD
E₃
psps⁻¹ = 1 = sas⁻¹a
papa⁻¹ = 1
FUND. DOM. F₂
E
F
H
G
E₂
A
O
B
D
C
E₁
COORD. AXES: OE₁, OE₂, OE₃
H
G
s
E
F
a
p⁻¹
a⁻¹
p
A
B
s⁻¹
D
C
GRAPH OF F₂

Fig. 13

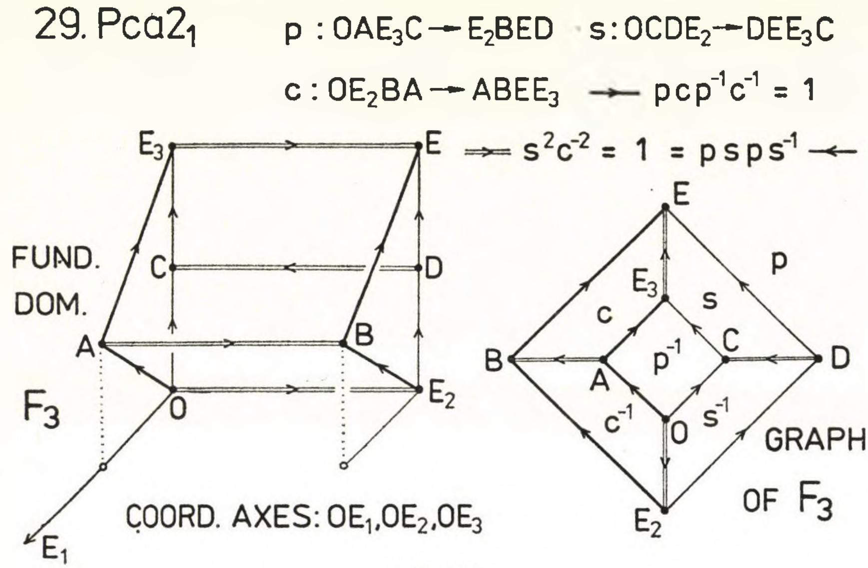

29. Pca2₁
p : OAE₃C → E₂BED
s : OCDE₂ → DEE₃C
c : OE₂BA → ABEE₃ → pcp⁻¹c⁻¹ = 1
→ s²c⁻² = 1 = psps⁻¹ ←
FUND. DOM.
F₃
E₃
E
C
D
A
B
O
E₂
E₁
COORD. AXES: OE₁, OE₂, OE₃
p
s
c
p⁻¹
c⁻¹
s⁻¹
GRAPH OF F₃

Fig. 14

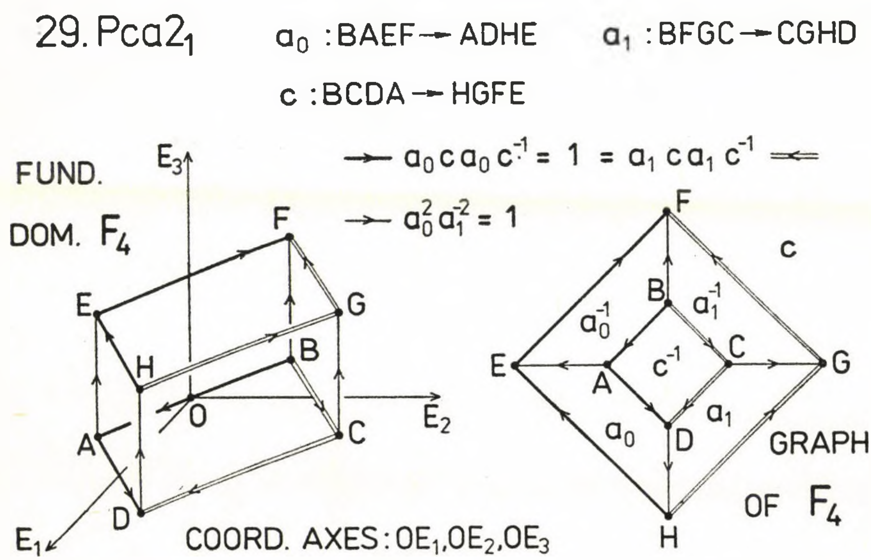

29. Pca2₁
a₀ : BAEF → ADHE
a₁ : BFGC → CGHD
c : BCDA → HGFE
→ a₀ca₀c⁻¹ = 1 = a₁ca₁c⁻¹ ←
→ a₀²a₁⁻² = 1
FUND. DOM. F₄
E₃
F
E
G
H
B
O
E₂
A
C
D
E₁
COORD. AXES: OE₁, OE₂, OE₃
c
a₀⁻¹
a₁⁻¹
c⁻¹
a₀
a₁
GRAPH OF F₄

Fig. 15

as it is described in Figure 12 and 13. Of course, we have also other 3-generator presentations whose length sum is bigger than 12, therefore they are not minimal. In Figure 12 we have given the normalizer N of $\mathbf{Pca2_1}$ as a semi-direct product, whose linear part is generated by three pairwise orthogonal plane reflections, the translational part is a direct sum, where

$$\mathbf{R} = \{c\mathbf{e}_3\colon c \in R \text{ (reals)}\}.$$

10. The space form $E^3/\mathbf{Pna2_1}$

The lattice $\mathbf{L}_G$ for the space group $G = \mathbf{Pna2_1}$ is as in 9. In Figure 16 we have given the unique minimal 2-generator presentation. The generating glide reflection $\mathrm{a}(a, \mathbf{a})$ is defined by

$$(1) \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \left(\frac{1}{2}\ \frac{1}{2}\ 0\right),$$

the generating 2_1 screw motion $\mathrm{s}(s, \mathbf{s})$ is defined by

$$(2) \qquad \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \left(1\ 1\ \frac{1}{2}\right)$$

in the base $\overrightarrow{OE_1}$, $\overrightarrow{OE_2}$, $\overrightarrow{OE_3}$, as usual.

The combinatorial structure of $\mathscr{F}$ allows us to choose $E\left(\frac{1}{2}; -\frac{1}{8}; \frac{1}{2}\right)$ for vertex, then the other vertices of $\mathscr{F}$ are

$$D\left(\frac{1}{2}; \frac{1}{8}; 0\right), \quad F\left(\frac{1}{2}; \frac{1}{8}; 1\right), \quad A\left(0; \frac{3}{8}; 0\right),$$

$$C\left(0; \frac{3}{8}; 1\right), \quad G\left(1; \frac{3}{8}; 0\right), \quad I\left(1; \frac{3}{8}; 1\right),$$

$$B\left(0; \frac{5}{8}; \frac{1}{2}\right), \quad H\left(1; \frac{5}{8}; \frac{1}{2}\right), \quad J\left(\frac{1}{2}; \frac{7}{8}; \frac{1}{2}\right).$$

For the construction of faces we have taken the points

$$K\left(\frac{1}{4}; \frac{1}{4}; \frac{1}{2}\right), \quad L\left(\frac{3}{4}; \frac{1}{4}; \frac{1}{2}\right), \quad M\left(\frac{1}{2}; \frac{1}{2}; \frac{1}{4}\right), \quad N\left(\frac{1}{2}; \frac{1}{2}; \frac{3}{4}\right).$$

The other minimal presentations are equivalent to this one by the affine normalizer N of **Pna2$_1$**. This normalizer has the form

$$N = \left\{\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix},\quad (n^1,\ n^2,\ n^3)\right\}, \tag{3}$$

where $n^1, n^2 \in Z/2$; $n^3 \in R$. Hence **Pca2$_1$** and **Pna2$_1$** have the same normalizer.

We mention the other 2-generator presentations which are not minimal ones. Defining the glide reflection $\mathrm{n} := \mathrm{as}$, we have two possibilities:

$$\mathbf{Pna2_1} = (\mathrm{n},\ \mathrm{s} - 1 = \mathrm{ns^2n^{-1}s^{-2}} = \mathrm{ns^{-1}n^2s^{-1}ns^{-2}}), \tag{4}$$

$$\mathbf{Pna2_1} = (\mathrm{a},\ \mathrm{n} - 1 = \mathrm{a^2na^2n^{-1}} = \mathrm{na^{-1}na^{-1}n^{-1}an^{-1}a}). \tag{5}$$

11. The space form $E^3/\mathbf{P3_1}$

The hexagonal lattice $\mathbf{L}_G$ for the space group $G=\mathbf{P3_1}$ is determined by the base $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ with Gram matrix

$$((\mathbf{e}_i,\ \mathbf{e}_j)) = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & c \end{pmatrix} \tag{1}$$

where $c>0$ is an arbitrary constant. In Figures 17—19 we have described the three minimal presentations according to the complicated polyhedra $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$, respectively.

The generating 3_1 screw motion $\mathrm{s}(s, \mathbf{s})$ for $\mathcal{F}_1$ will be defined by the rotational part s with the matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

and by the vector $\mathbf{s}\left(0; 0; \frac{1}{3}\right)$ in the base $\mathbf{e}_i$. Moreover, we choose the translation $\mathrm{p}(I, \mathbf{e}_1)$ as second generator. Thus $\mathrm{s}^3(I, \mathbf{e}_3)$ and $\mathrm{s^{-1}ps}(I, \mathbf{e}_2)$ are basic translations. Forming $\mathrm{s^{-2}ps^2}(I, -(\mathbf{e}_1+\mathbf{e}_2))$ we can state that the relations in Figure 17 give us a complete set of relations to define the space group **P3$_1$**. We can check that the length sum cannot be reduced and in fact, we get a minimal presentation which is also geometrically realizable.

First, we should form the two Poincaré cycles to construct the unique combinatorial structure of $\mathcal{F}_1$ indicated in Figure 17 in a similar manner as for **P4$_1$** elaborated in Section **3**.

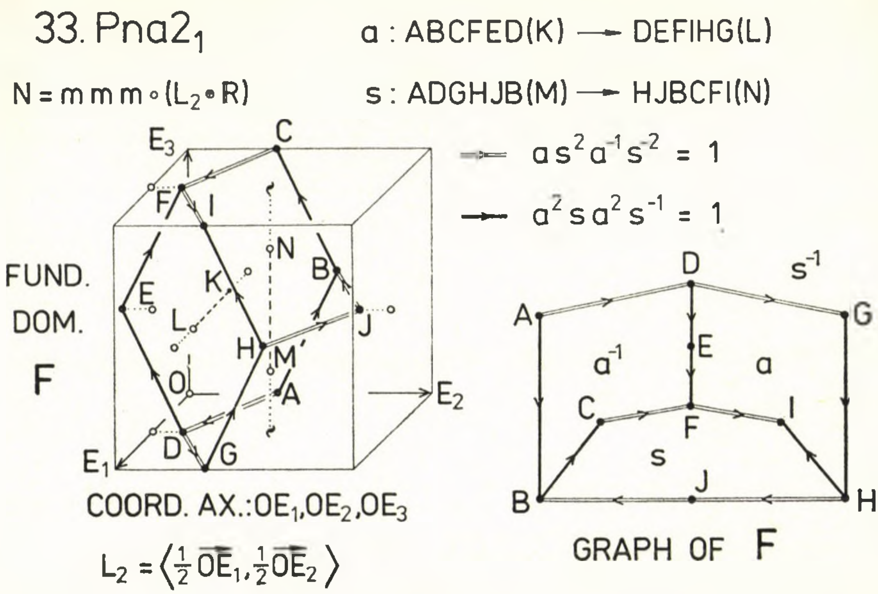

33. Pna2₁
N = mmm∘(L₂•R)
a : ABCFED(K) → DEFIHG(L)
s : ADGHJB(M) → HJBCFI(N)
as²a⁻¹s⁻² = 1
a²sa²s⁻¹ = 1
FUND.
DOM.
F
COORD. AX.:OE₁,OE₂,OE₃
L₂ = ⟨½ OE₁, ½ OE₂⟩
GRAPH OF F

Fig. 16

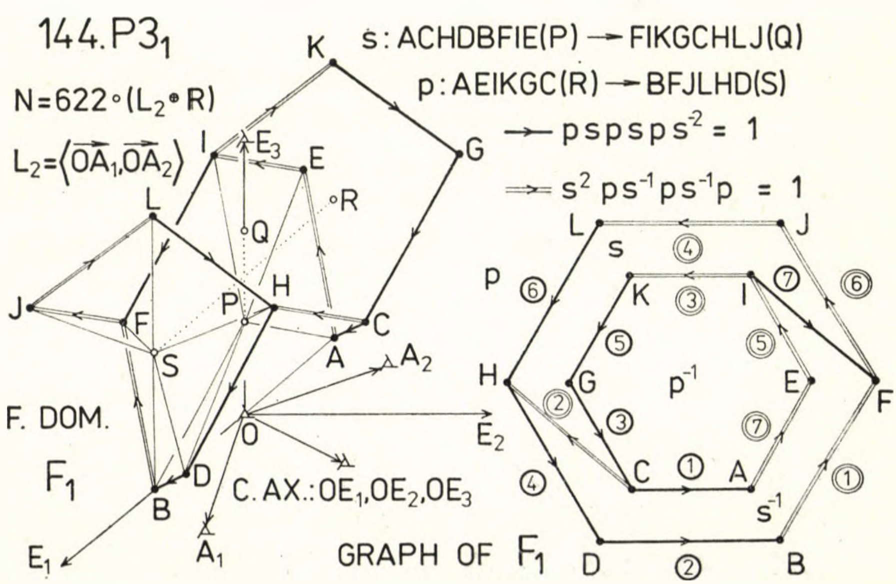

144. P3₁
N=622∘(L₂•R)
L₂=⟨OA₁,OA₂⟩
s : ACHDBFIE(P) → FIKGCHLJ(Q)
p : AEIKGC(R) → BFJLHD(S)
pspsps⁻² = 1
s²ps⁻¹ps⁻¹p = 1
F. DOM.
F₁
C. AX.:OE₁,OE₂,OE₃
GRAPH OF F₁

Fig. 17

144. $P3_1$

s_1: $OAGHDCE(K) \rightarrow GHDFE_3BA(L)$

s_2: $OECIJBA(M) \rightarrow CIJBE_3FD(N)$

$\Rightarrow s_1^3 s_2^{-3} = 1 = s_1 s_2^{-2} s_1 s_2 s_1^{-2} s_2 \Leftarrow$

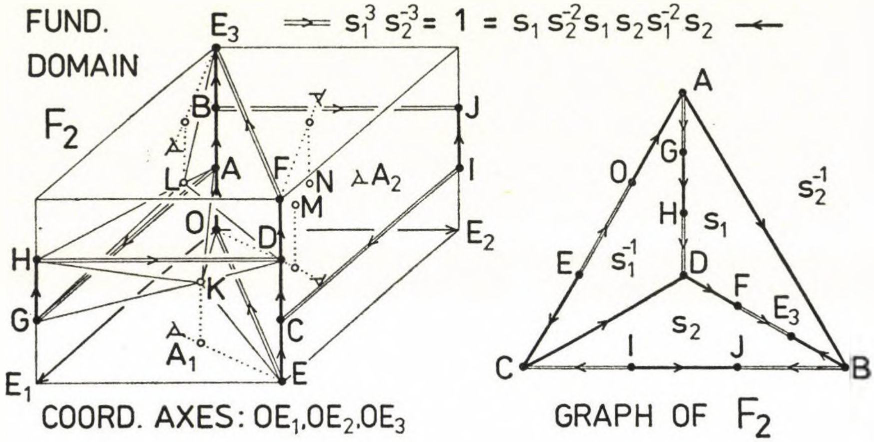


Fig. 18

144. $P3_1$

s_1: $ABJIE_1GOH(K) \rightarrow GFDCBJIE_1(L)$

s_2: $HOGFEA(M) \rightarrow ABCDFE(N)$

$\Rightarrow s_1^3 s_2^{-3} = 1 = s_2 s_1 s_2 s_1^{-2} s_2 s_1^{-2} \Leftarrow$

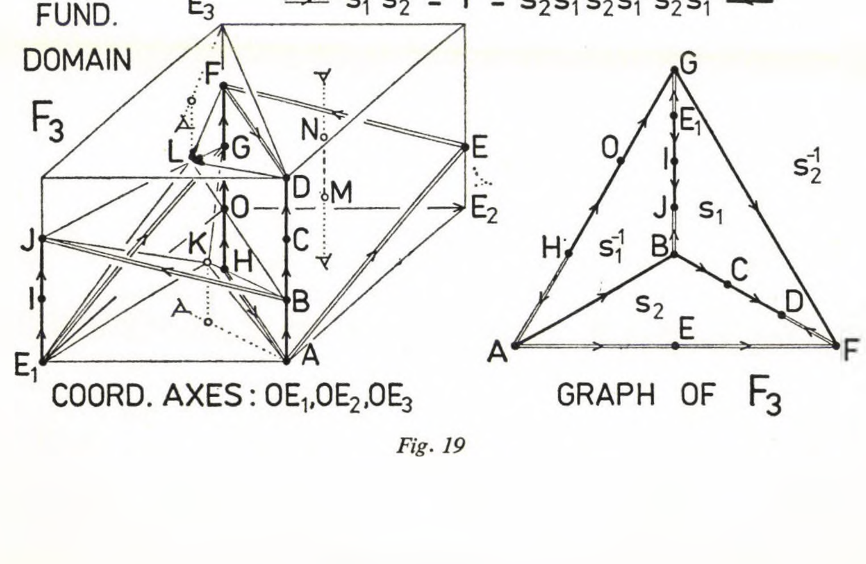


Fig. 19

After having solved this rather delicate combinatorial task we have a certain freedom in the metric realization as in the previous cases. We have chosen $A\left(-\frac{1}{2}; 0; 0\right)$ as starting point, then the other vertices are:

$$B\left(\frac{1}{2}; 0; 0\right), \quad C\left(0; \frac{1}{2}; \frac{1}{3}\right), \quad D\left(1; \frac{1}{2}; \frac{1}{3}\right),$$

$$E\left(-1; -\frac{1}{2}; \frac{1}{3}\right), \quad F\left(0; -\frac{1}{2}; \frac{1}{3}\right), \quad H\left(\frac{1}{2}; \frac{1}{2}; \frac{2}{3}\right),$$

$$G\left(-\frac{1}{2}; \frac{1}{2}; \frac{2}{3}\right), \quad I\left(-\frac{1}{2}; -\frac{1}{2}; \frac{2}{3}\right), \quad J\left(\frac{1}{2}; -\frac{1}{2}; \frac{2}{3}\right),$$

$$K\left(-\frac{1}{2}; 0; 1\right), \quad L\left(\frac{1}{2}; 0; 1\right).$$

For the usual construction of faces of $\mathcal{F}_1$ we were careful in taking the points $P\left(0; 0; \frac{1}{3}\right)$, $R\left(-\frac{1}{2}; 0; \frac{1}{2}\right)$ to ensure that the faces and the corresponding images do not intersect each other in their interiors.

In Figure 17 we have indicated the construction of $f_{s^{-1}}$ and f_p only. In the usual manner the affine normalizer N of $\mathbf{P3_1}$ is given in Figure 17. It shows that any 3 screw motion and any horizontal unit translation lead to a minimal presentation equivalent to that of $\mathcal{F}_1$.

In Figures 18—19 we have described the other minimal presentations belonging to the generating 3_1 motions $\mathbf{s}_1(s, \mathbf{s}_1)$, $\mathbf{s}_2(s, \mathbf{s}_2)$, where the common linear part has the matrix (2) and the translational parts are

$$\mathbf{s}_1\left(1; 0; \frac{1}{3}\right) \quad \text{and} \quad \mathbf{s}_2\left(1; 1; \frac{1}{3}\right), \tag{3}$$

respectively. The more symmetric presentation by $\mathcal{F}_2$ is uniquely determined by the relations there. Let the origin O be the starting vertex of $\mathcal{F}_2$. Then the other vertices are

$$E(1; 1; 0), \quad A\left(0: 0; \frac{1}{3}\right), \quad G\left(1; 0; \frac{1}{3}\right),$$

$$C\left(1; 1; \frac{1}{3}\right), \quad I\left(0; 1; \frac{1}{3}\right), \quad B\left(0; 0; \frac{2}{3}\right),$$

$$H\left(1; 0; \frac{2}{3}\right), \quad D\left(1; 1; \frac{2}{3}\right), \quad J\left(0; 1; \frac{2}{3}\right),$$

$$E_3(0; 0; 1), \quad F(1; 1; 1).$$

For the usual construction of faces we have chosen the points K and M to guarantee that the faces and the corresponding images do not intersect. We have taken a fixed parameter t with $0<t<1$ and so

$$K\left(\frac{2}{3}+t\frac{1}{3};\ \frac{1}{3}+t\frac{2}{3};\ \frac{1}{3}\right),\quad L=K^{s_1},$$

$$M\left(\frac{1}{3}-t\frac{1}{3};\ \frac{2}{3}-t\frac{2}{3};\ \frac{1}{3}\right),\quad N=M^{s_2}.$$

In Figure 18 the face $f_{s_1^{-1}}=OAGHDCE(K)$ and its s-image have been indicated only.

If we take $s_1 s_2^{-2}=s_1^{-2}s_2$ derived from $s_1^3 s_2^{-3}=1$ and we place it into the second relation of $\mathcal{F}_2$, then we get the presentation $\mathcal{F}_3$ in Figure 19, which is also minimal in the sense defined in the introduction. Here the vertices are:

$$H\left(0;\ 0;\ -\frac{1}{3}\right),\quad O(0;\ 0;\ 0),\quad E_1(1;\ 0;\ 0),$$

$$A(1;\ 1;\ 0),\quad G\left(0;\ 0;\ \frac{1}{3}\right),\quad I\left(1;\ 0;\ \frac{1}{3}\right),$$

$$B\left(1;\ 1;\ \frac{1}{3}\right),\quad E\left(0;\ 1;\ \frac{1}{3}\right),\quad F\left(0;\ 0;\ \frac{2}{3}\right),$$

$$J\left(1;\ 0;\ \frac{2}{3}\right),\quad C\left(1;\ 1;\ \frac{2}{3}\right),\quad D(1;\ 1;\ 1);$$

To construct the faces we choose $K\left(\frac{2}{3}+t\frac{1}{3};\frac{1}{3}+t\frac{2}{3};\ \frac{1}{3}\right)$, $L=K^{s_1}$, $M\left(\frac{1}{3};\ \frac{2}{3};\ \frac{1}{3}\right)$ and $N=M^{s_2}$.

Other "neighbouring" 3_1 screw motions lead to equivalent minimal presentations.

12. The space form $E^3/\mathbf{P6}_1$

The hexagonal lattice $\mathbf{L}_G$ for the space group $G=\mathbf{P6}_1$ is determined by the base $\mathbf{e}_i$ with Gram matrix **(11.1)**. The generating 6_1 screw motion is defined, as usual, by

$$\text{(1)}\qquad s_0(\mathit{s}_0,\mathbf{s}_0)\sim\begin{pmatrix}1&1&0\\-1&0&0\\0&0&1\end{pmatrix},\ \left(0\ 0\ \frac{1}{6}\right).$$

We may choose the translation $p_1(\mathit{1},\mathbf{e}_1)$ as second generator. We know, that the sum, in any order, of the translations p_1, $s_0^{-2}p_1s_0^2$, $s_0^{-1}p_1^{-1}s_0$ is the identity. Thus we get a presentation

$$\text{(2)}\qquad \mathbf{P6}_1=(s_0,p_1-1=p_1s_0^{-1}p_1^{-1}s_0^{-1}p_1s_0^2=p_1s_0p_1^{-1}s_0p_1s_0^{-2})$$

with length sum 14, but this will not be the shortest one.

Taking another 6_1 screw motion $s_1 := s_0 p_1$ as second generator we get a new presentation

(3) $$\mathbf{P6_1} = (s_0, s_1 - 1 = s_0^{-1} s_1^{-1} s_0^{-1} s_1 s_0 s_1 = s_0^{-2} s_1 s_0 s_1^{-2} s_0 s_1)$$

with length sum 14.

Now, we take the 2_1 screw motion $s_3 := s_0^3 p_1$ beside s_0, then

(4) $$\mathbf{P6_1} = (s_0, s_3 - 1 = s_0^6 s_3^{-2} = s_0 s_3 s_0 s_3^{-1} s_0 s_3^{-1}).$$

Defining the 3_1 screw motion $s_2 := s_0^2 p_1$, we pair this first with the 2_1 screw motion s_3. Then we get the presentation

(5) $$\mathbf{P6_1} = (s_2, s_3 - 1 = s_2^3 s_3^{-2} = s_2 s_3 s_2 s_3^{-1} s_2 s_3^{-1} s_2 s_3^{-1} s_2 s_3^{-1} s_2 s_3^{-1})$$

which is long enough.

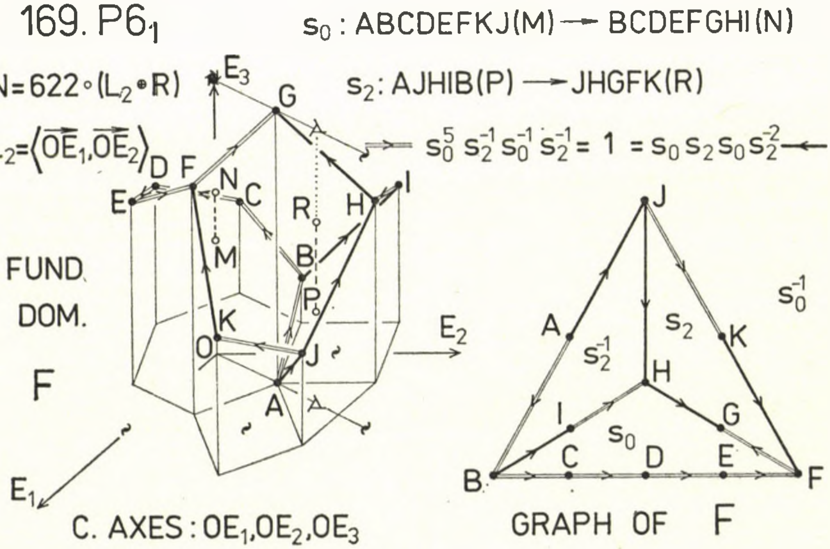


Fig. 20

Finally, we turn to the surprising minimal presentation generated by 6_1 and 3_1 screw motions above, but we have chosen another situation in Fig. 20 (the translation $p_2(I, \mathbf{e}_2)$ is taken instead of p_1). First we uniquely constructed the combinatorial structure of the domain $\mathcal{F}$ by means of the Poincaré cycles according to relations. For the metric constructions of $\mathcal{F}$ we took the starting vertex A, as in Figure 20, arbitrarily between the 6_1-axes MN and the 3_1-axes PR. Then the places of the other vertices have already been determined as follows: $B = A^{s_0}$, $C = A^{s_0 s_0}$, $D = A^{s_0 s_0 s_0}$, $E = A^{s_0 s_0 s_0 s_0}$, $F = A^{s_0 s_0 s_0 s_0 s_0}$, $G = A^{s_0 s_0 s_0 s_0 s_0 s_0} = A^{s_2 s_2 s_2}$, $H = A^{s_2 s_2}$, $J = A^{s_2}$, $I = J^{s_0} = A^{s_2 s_0}$, $K = B^{s_2} = A^{s_0 s_2}$. The points $M\left(0; 0; \frac{5}{12}\right)$, $N = M^{s_0}$, $P\left(\frac{1}{3}; \frac{2}{3}; \frac{1}{3}\right)$, $R = P^{s_2}$ determine the faces of $\mathcal{F}$ as usual. This presentation has a length sum 13, shorter than the earlier ones, moreover, this sum cannot be reduced.

The normalizer N of $\mathbf{P6_1}$ is also indicated in Figure 20 as a semi-direct product with our usual notations. The linear part **622** denotes the group generated by the 6-rotation s_0 in (1) and by the half-turn h defined by the matrix

$$h \sim \begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{6}$$

We have investigated all the minimal 2-generator presentations not equivalent under N.

I am indebted to Professor László Fejes Tóth whose beautiful book [8] "Regular Figures" inspired me to study the theory of discrete transformation groups.

REFERENCES

[1] AL-JUBOURI, N. K., On non-orientable hyperbolic 3-manifolds, *Quart. J. Math. Oxford Ser.* (2) **31** (1980), 9—18. *MR* **81b**: 57008.

[2] BEST, L. A., On torsion-free discrete subgroups of $PSL(2, C)$ with compact orbit space, *Canad. J. Math.* **23** (1971), 451—470. *MR* **44** #1767.

[3] BROWN, H., BÜLOW, R., NEUBÜSER, J., WONDRATSCHEK, H. and ZASSENHAUS, H., *Crystallographic groups of four-dimensional space,* Wiley Monographs in Crystallography, Wiley, New York, 1978. *MR* **58** #4109.

[4] BURAGO, YU. D., Appendix, In: WOLF, J. A., *Spaces of constant curvature,* Nauka, Moscow, 1982, 461—472 (in Russian).

[5] BURCKHARDT, J. J., *Die Bewegungsgruppen der Kristallographie,* 2. Aufl., Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, mineral.-geotechn. Reihe, Band II, Birkhäuser Verlag, Basel und Stuttgart, 1966. *MR* **34** #2708.

[6] COXETER, H. S. M. and MOSER, W. O. J., *Generators and relations for discrete groups,* Fourth edition, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Bd. 14, Springer-Verlag, Berlin—Heidelberg—New York, 1980. *MR* **81a**: 20001.

[7] DANZER, L. and SCHULTE, E., Reguläre Inzidenzkomplexe I, *Geom. Dedicata* **13** (1982), 295—308. *MR* **84h**: 51042.

[8] FEJES TÓTH, L., *Reguläre Figuren,* Akadémiai Kiadó, Budapest, 1965. *MR* **30** #3408.

[9] GUTSUL, I. S., Compact three-dimensional manifolds of constant negative curvature, *Trudy Mat. Inst. Steklov.* **152** (1980), 89—96 (in Russian). *MR* **83k**: 51035.

[10] HENRY, N. F. M. and LONSDALE, K., *Symmetry groups,* In: *International Tables for X-Ray Crystallography,* Vol. I, Kynoch Press, Birmingham, 1969.

[11] KOCH, E. and FISCHER, W., Automorphismengruppen von Raumgruppen und die Zuordnung von Punktlagen zu Konfigurationslagen, *Acta Cryst.* A **31** (1975), 88—95.

[12] KOCH, E., The implications of normalizers on group-subgroup relations between space groups, *Acta Cryst.* A **40** (1984), 593—600.

[13] MASKIT, B., On Poincaré's theorem for fundamental polygons, *Advances in Math.* **7** (1971), 219—230. *MR* **45** #7049.

[14] MOLNÁR, E., Konvexe Fundamentalpolyeder und einfache $D-V$-Zellen für 29 Raumgruppen, die Coxetersche Spiegelungsuntergruppen enthalten, *Beiträge Algebra Geom.* **14** (1983), 33—75. *MR* **84k**: 20023.

[15] MOLNÁR, E., Presentation of crystallographic groups by fundamental polyhedra, Manuscript postered at XIII Congress of IUC 9—18 August 1984, Hamburg, 20.2 Symmetry and its generalization, C 457.

[16] MOLNÁR, E., An infinite series of compact non-orientable 3-dimensional space forms of constant negative curvature, *Ann. Global Analysis and Geom.* **1** (1983), 37—49; **2** (1984), 253—254.

[17] MOLNÁR, E., Space forms and fundamental polyhedra, *Proc. Conf. Differential Geometry Applications,* Nové Město na Moravě, Czechoslovakia, 1983, Part 1, Differential Geometry, 91—103.

[18] MOLNÁR, E., Twice punctured compact euclidean and hyperbolic manifolds and their two-fold coverings, *Topics in differential geometry* (*Proc. Colloq.*, Debrecen (Hajdúszoboszló), Hungary, 1984), ed. by J. Szenthe and L. Tamássy, Colloq. Math. Soc. J. Bolyai, Vol. 46, North-Holland Publ. Co., Amsterdam—Oxford—New York, 1987 (to appear).
[19] SCHWARZENBERGER, R. L. E., *n-dimensional crystallography*, Research Notes in Mathematics, 41, Pitman, London, 1980. *MR* **82e**: 82018.
[20] SENECHAL, M., Morphisms of crystallographic groups: Kernel and images, *J. Mathematical Phys.* (to appear).
[21] THURSTON, W. P., Three-dimensional manifolds, Kleinian groups and hyperbolic geometry, *Bull. Amer. Math. Soc. (N. S.)* 6 (1982), 357—381. *MR* **83h**: 57019.
[22] WEBER, C. and SEIFERT, H., Die beiden Dodekaederräume, *Math. Z.* **37** (1933), 237—253. *Zbl.* **7**, 028.
[23] WOLF, J. A., *Spaces of constant curvature*, Second edition, Department of Mathematics University of California, Berkeley, Calif., 1972. *MR* **49** #7957.

(Received January 8, 1985)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RAKÓCZI ÚT 5
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 53—63

PROBLEMS AND RESULTS ON ADDITIVE PROPERTIES OF GENERAL SEQUENCES, III

P. ERDŐS, A. SÁRKÖZY and V. T. SÓS

To Professor L. Fejes Tóth on his seventieth birthday

1. Let $\mathcal{A}=\{a_1, a_2, \ldots\}$ $(a_1<a_2<\ldots)$ be an infinite sequence of positive integers, put

$$A(n) = \sum_{\substack{a\in\mathcal{A}\\ a\leqq n}} 1,$$

and for $n=0, 1, 2, \ldots$, let $R(\mathcal{A}, n)$ or briefly $R(n)$ denote the number of solutions of

$$a_x+a_y = n, \quad a_x\in\mathcal{A}, a_y\in\mathcal{A}.$$

In Part I of this paper [3], P. Erdős and A. Sárközy proved that if $F(n)$ is an arithmetic function satisfying $F(n)\to+\infty$, $F(n+1)\geqq F(n)$ for $n\geqq n_0$ and $F(n)=o\big(n(\log n)^{-2}\big)$, then

$$R(n)-F(n) = o\big((F(n))^{1/2}\big)$$

cannot hold. In Part II [4], they showed that this theorem is nearly best possible. (See [1], [2] and [5] for further related results and problems.) In this paper, we continue the study of the regularity properties of the function $R(n)$. In fact, here our goal is to show that under certain possibly simple assumptions on $\mathcal{A}$, $|R(n+1)-R(n)|$ cannot be bounded.

If $\mathcal{A}$ is "very thin" $\big(\mathcal{A}(n)=o(\sqrt[3]{n})\big)$ then $R(n)$ can be bounded and then also $|R(n+1)-R(n)|$ is bounded. On the other hand, if $\mathcal{A}$ is "very dense" (e.g. $\mathcal{A}=\{1, 2, \ldots, n, \ldots\}$) then clearly, $|R(n+1)-R(n)|$ can be bounded again. One may guess that if $\mathcal{A}$ is not "very thin" and not "very dense" then $|R(n+1)-R(n)|$ cannot be bounded. This is not so as the following theorem shows:

THEOREM 1. *Let* $S_1<S_2<\ldots$ *and* $t_1<t_2<\ldots$ *be positive integers satisfying*

$$S_k < t_k \leqq \frac{S_{k+1}}{2}-1 \qquad (\textit{for } k = 1, 2, \ldots), \tag{1}$$

and put

$$\mathcal{A} = \{a_1, a_2, \ldots\} = \bigcup_{k=1}^{+\infty} \{S_k, S_k+1, \ldots, t_k\}.$$

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1811.

1980 *Mathematics Subject Classification*. Primary 11B99.

Key words and phrases. Additive number theory, general sequences, gaps, thin and dense sequences.

Then we have

$$|R(n+1)-R(n)| \leqq 3$$

for all n.

This theorem shows that if $\mathscr{A}$ consists of a "few" blocks of consecutive integers then $|R(n+1)-R(n)|$ can be bounded (independently of the counting function $A(n)$). On the other hand, one may guess that if the number of these blocks up to n, i.e.,

$$B(\mathscr{A}, n) = \sum_{\substack{a \leqq n \\ a-1 \notin \mathscr{A}, a \in A}} 1$$

is "large" (in terms of n), then $|R(n+1)-R(n)|$ cannot be bounded. In fact, we will prove the following theorem:

THEOREM 2. *If* $\mathscr{A} \neq \emptyset$ *then*

$$S(N) \overset{\text{def}}{=} \sum_{n=1}^{N} (R(n+1)-R(n))^2 = o((B(\mathscr{A}, n))^2) \tag{2}$$

cannot hold.

The following corollaries are trivial consequences of Theorem 2.

COROLLARY 1. *If* $\mathscr{A} \neq \emptyset$, *then*

$$\max_{n \leqq N} |R(n+1)-R(n)| = o((B(\mathscr{A}, N)/N^{1/2})^{1/2}) \tag{3}$$

cannot hold.

In fact, Theorem 2 says that (3) is impossible in square mean.

COROLLARY 2. *If*

$$\lim_{N \to +\infty} \frac{B(\mathscr{A}, N)}{N^{1/2}} = +\infty$$

then $|R(n+1)-R(n)|$ *cannot be bounded.*

This is a consequence of Corollary 1.
Finally, we will show that Theorem 2 is nearly best possible:

THEOREM 3. *For all* $\varepsilon > 0$, *there exists an infinite sequence* $\mathscr{A}$ *such that*

(i) $$B(\mathscr{A}, n) \gg N^{1/2-\varepsilon}$$

and

(ii) $R(n)$ *is bounded (in fact,* $R(n) < 3+\varepsilon^{-1}$ *for large n) so that also* $|R(n+1)-R(n)|$ *is bounded.*

Furthermore, by using a construction of Erdős, we can show that there exists an infinite sequence $\mathscr{A}$ such that

$$\limsup_{N \to +\infty} \frac{B(\mathscr{A}, N)}{N^{1/2}} > 0$$

and $|R(n+1)-R(n)|$ is bounded.

There is some gap between the lower bounds and upper bounds given above. In fact, we conjecture that Corollary 2 can be sharpened in the following way: If

$$\limsup_{N\to+\infty} \frac{B(\mathcal{A}, N)}{N^{1/2}} = +\infty$$

or

$$\liminf_{N\to+\infty} \frac{B(\mathcal{A}\ N)}{N^{1/2}} > 0$$

(perhaps, it suffices to assume that

$$\liminf_{N\to+\infty} \frac{B(\mathcal{A}, N)\log N}{N^{1/2}} = +\infty),$$

then $|R(n+1)-R(n)|$ cannot be bounded. Unfortunately, we have not been able to prove this.

2. In this section, we prove Theorem 1.

For an arbitrary positive integer n, let us define the positive integer k by

$$t_{k-1} < n/2 \leqq t_k \tag{4}$$

(if $n/2 \leqq t_1$, then we put $k=1$). Then (1) and (4) yield that

$$2t_{k-1} < 2(n/2) = n \tag{5}$$

and

$$S_{k+1} \geqq 2(t_k+k) \geqq 2\left(\frac{n}{2}+1\right) = n+2. \tag{6}$$

Let m denote one of the numbers n, $n+1$. Then in view of (5) and (6)

$$a_x+a_y = m, \quad a_x \in \mathcal{A},\ a_y \in \mathcal{A} \tag{7}$$

implies that

$$t_{k-1} < \max(a_x, a_y) < S_{k+1}. \tag{8}$$

By the construction of the sequence $\mathcal{A}$, we have

$$\mathcal{A} \cap \{t_{k-1}, S_{k+1}\} = \{S_k, S_k+1, \ldots, t_k\}. \tag{9}$$

By (8) and (9),

$$S_k \leqq \max(a_x, a_y) \leqq t_k. \tag{10}$$

In view of (9) and (10), (a_x, a_y) is a solution of (7) if and only if it satisfies one and only one of the following equations:

$$a_x+a_y = m, \quad S_k \leqq a_x \leqq t_k, \quad a_y \leqq t_{k-1}, \quad a_y \in \mathcal{A}, \tag{11}$$

$$a_x+a_y = m, \quad a_x = t_{k-1}, \quad a_x \in \mathcal{A}, \quad S_k \leqq a_y \leqq t_k, \tag{12}$$

$$a_x+a_y = m, \quad S_k \leqq a_x \leqq t_k, \quad S_k \leqq a_y \leqq t_k. \tag{13}$$

Denoting the number of solutions of (11), (12) and (13) by $R_1(m)$, $R_2(m)$ and $R_3(m)$, respectively, clearly we have

$$R_2(m) = R_1(m)$$

and

$$R(m) = R_1(m)+R_2(m)+R_3(m) = 2R_1(m)+R_3(m). \tag{14}$$

If a_x, a_y is a solution of (11) or (13) with n in place of m, i.e.,

$$a_x+a_y = n, \quad S_k \leqq a_x \leqq t_k, \quad a_y \leqq t_{k-1}, \quad a_y \in \mathcal{A} \tag{15}$$

or

$$a_x+a_y = n, \quad S_k \leqq a_x \leqq t_k, \quad S_k \leqq a_y \leqq t_k, \tag{16}$$

then $a_u = a_x+1$, $a_v = a_y$ is a solution of

$$a_u+a_v = n+1, \quad S_k \leqq a_u \leqq t_k, \quad a_v \leqq t_{k-1}, \quad a_y \in \mathcal{A}, \tag{17}$$

or

$$a_u+a_v = n+1, \quad S_k \leqq a_u \leqq t_k, \quad S_k \leqq a_v \leqq t_k, \tag{18}$$

respectively, except at most the solution $a_x = t_k$, $a_y = n-t_k$ of (17) or (18). On the other hand, in this way we get all the solutions of (17) and (18), except at most the solution $a_u = S_k$, $a_v = n+1-S_k$. Thus we have

$$R_i(n)-1 \leqq R_i(n+1) \leqq R_i(n)+1 \quad \text{for} \quad i = 1, 3. \tag{19}$$

(14) and (19) yield that

$$|R(n+1)-R(n)| = \left|2\big(R_1(n+1)-R_1(n)\big)+\big(R_3(n+1)-R_3(n)\big)\right| \leqq$$

$$\leqq 2|R_1(n+1)-R_1(n)|+|R_3(n+1)-R_3(n)| \leqq 2\cdot 1+1 = 3$$

which completes the proof of Theorem 1.

3. Sections 3—6 will be devoted to the proof of Theorem 2. We start out from the indirect assumption that $\mathcal{A} \neq \emptyset$ is a sequence satisfying (2) in Theorem 2.

First we are going to show that there exist infinitely many integers N such that

$$\frac{B(\mathcal{A}, N+j)}{B(\mathcal{A}, N)} < \left(\frac{N+j}{N}\right)^2 \quad \text{for} \quad j = 1, 2, \dots. \tag{20}$$

In fact, if this inequality holds only for finitely many integers N, then there exists an integer N_0 such that

$$B(\mathcal{A}, N_0) \geqq 1$$

and for $N \geqq N_0$ there exists an integer $N' = N'(N)$ satisfying $N' > N$ and

$$\frac{B(\mathcal{A}, N')}{B(\mathcal{A}, N)} \geqq \left(\frac{N'}{N}\right)^2.$$

Then we get by induction that there exist integers $N_0<N_1<N_2<\ldots<N_j<\ldots$ such that

$$\frac{B(\mathscr{A}, N_{j+1})}{B(\mathscr{A}, N_j)} \geqq \left(\frac{N_{j+1}}{N_j}\right)^2 \qquad (\text{for } j = 1, 2, \ldots).$$

(In fact, N_{j+1} can be defined by $N_{j+1}=N'(N_j)$.) Hence

$$\frac{B(\mathscr{A}, N_{k+1})}{B(\mathscr{A}, N_0)} = \prod_{j=0}^{k} \frac{B(\mathscr{A}, N_{j+1})}{B(\mathscr{A}, N_j)} \geqq \prod_{j=0}^{k} \left(\frac{N_{j+1}}{X_j}\right)^2 = \left(\frac{N_{k+1}}{N_0}\right)^2$$

so that

(21) $$B(\mathscr{A}, N_{k+1}) \geqq \left(\frac{N_{k+1}}{N_0}\right)^2 B(\mathscr{A}, N_0) \geqq \frac{1}{N_0^2} N_{k+1}^2 > N_{k+1}^{3/2}$$

for large enough k. On the other hand, clearly we have

(22) $$B(\mathscr{A}, N_{k+1}) = \sum_{\substack{a \leqq N_{k+1} \\ a-1 \notin \mathscr{A},\, a \in \mathscr{A}}} 1 \leqq \sum_{a \leqq N_{k+1}} 1 = N_{k+1}.$$

(21) and (22) cannot hold simultaneously and this contradiction proves the existence of infinitely many integers N satisfying (20).

4. Throughout the remaining part of the proof of Theorem 2, we use the following notation:

N denotes a large integer satisfying (20). We write $e^{2\pi i\alpha}=e(\alpha)$, and we put $r=e^{-1/N}$, $z=re(\alpha)$ where α is a real variable (so that a function of form $p(z)$ is a function of the real variable α; $p(z)=p(re(\alpha))=P(\alpha)$). We write

$$f(z) = \sum_{j=1}^{+\infty} z^{a_j}.$$

(By $r<1$, this infinite series and all the other infinite series in the remaining part of the proof are absolutely convergent.)

We start out from the integral

$$\mathscr{J} = \int_0^1 |f(z)(1-z)|^2 \, d\alpha.$$

We will give lower and upper bounds for $\mathscr{J}$. The lower bound for $\mathscr{J}$ will be greater than the upper bound, and this contradiction will prove that the indirect assumption (2) cannot hold which will complete the proof of Theorem 2.

5. In this section, we give a lower bound for $\mathscr{J}$. We write

$$f(z)(1-z) = \sum_{n=1}^{+\infty} b_n z^n.$$

Then for $n-1\notin\mathcal{A}$, $n\in\mathcal{A}$ we have $b_n=1$, thus by the Parseval formula, we have

$$\mathcal{J}=\int_0^1 |f(z)(1-z)|^2\,d\alpha=\int_0^1 \Big|\sum_{n=1}^{+\infty} b_n z^n\Big|^2 d\alpha=\sum_{n=1}^{+\infty} b_n^2 r^{2n}\geqq \tag{23}$$
$$\geqq r^{2N}\sum_{\substack{n\leqq N\\ n-1\notin\mathcal{A},\,n\in\mathcal{A}}} b_n^2=e^{-2}\sum_{\substack{n\leqq N\\ n-1\notin\mathcal{A},\,n\in\mathcal{A}}} 1=e^{-2}B(\mathcal{A},N).$$

6. In this section, we give an upper bound for $\mathcal{J}$. By the Cauchy inequality and the Parseval formula, and in view of (2) and (20), for all $\varepsilon>0$ and for $N>N_0(\varepsilon)$ we have

$$\mathcal{J}=\int_0^1 |f(z)(1-z)^2|\,d\alpha=\int_0^1 |f^2(z)(1-z)|\,|1-z|\,d\alpha\leqq$$
$$\leqq\int_0^1 |f^2(z)(1-z)|\,(1+|z|)\,d\alpha\leqq 2\int_0^1 |f^2(z)(1-z)|\,d\alpha=$$
$$=2\int_0^1 \Big|\Big(\sum_{j=1}^{+\infty} z^{a_j}\Big)^2(1-z)\Big|\,d\alpha=2\int_0^1 \Big|\Big(\sum_{n=1}^{+\infty} R(n)z^n\Big)(1-z)\Big|\,d\alpha=$$
$$=2\int_0^1 \Big|\sum_{n=1}^{+\infty}(R(n)-R(n-1))z^n\Big|\,d\alpha\leqq 2\Big(\int_0^1 \Big|\sum_{n=1}^{+\infty}(R(n)-R(n-1))z^n\Big|^2 d\alpha\Big)^{1/2}=$$
$$=2\Big(\sum_{n=1}^{+\infty}(R(n)-R(n-1))^2 r^{2n}\Big)^{1/2}=2\left((1-r^2)\frac{1}{1-r^2}\sum_{n=1}^{+\infty}(R(n)-R(n-1))^2 r^{2n}\right)^{1/2}=$$
$$=2\Big((1-r^2)\sum_{n=1}^{+\infty} S(n-1)r^{2n}\Big)^{1/2}<2\Big((1-r^2)\sum_{n=1}^{+\infty} S(n)r^{2n}\Big)^{1/2}= \tag{24}$$
$$=2\Big((1-e^{-2/N})\Big(\sum_{n=1}^{N} S(n)r^{2n}+\sum_{n=N+1}^{+\infty} S(n)r^{2n}\Big)\Big)^{1/2}<$$
$$<2\left(\frac{2}{N}\Big(\sum_{n=1}^{N} S(N)+\sum_{n=N+1}^{+\infty} S(n)r^{2n}\Big)\right)^{1/2}<3\Big(S(N)+N^{-1}\sum_{n=N+1}^{+\infty} S(n)r^{2n}\Big)^{1/2}<$$
$$<3\Big(\varepsilon(B(\mathcal{A},N))^2+N^{-1}\sum_{n=N=1}^{+\infty}\varepsilon(B(\mathcal{A},n))^2 r^{2n}\Big)^{1/2}<$$
$$<3\left[\varepsilon(B(\mathcal{A},N))^2+\varepsilon N^{-1}\sum_{n=N+1}^{+\infty}\left(B(\mathcal{A},N)\left(\frac{n}{N}\right)^2\right)^2 r^{2n}\right]^{1/2}=$$
$$=3\varepsilon^{1/2}B(\mathcal{A},N)\Big(1+N^{-5}\sum_{n=N+1}^{+\infty} n^4 r^{2n}\Big)^{1/2}$$

since we have

$$1-e^{-x}=x-\frac{x^2}{2!}+\frac{x^3}{3!}-\ldots<x\quad\text{for}\quad 0<x<1.$$

For $0<x<1$ we have

$$(1-x)^{-5} = 1+\sum_{n=1}^{+\infty}\binom{n+4}{4}x^n > \frac{1}{24}\sum_{n=1}^{+\infty} n^4 x^n$$

thus we obtain from (24) that for large N,

$$\mathscr{J} < 3\varepsilon^{1/2}B(\mathscr{A}, N)\Big(1+N^{-5}\sum_{n=N+1}^{+\infty} n^4 r^{2n}\Big)^{1/2} <$$

$$< 3\varepsilon^{1/2}B(\mathscr{A}, N)\Big(1+N^{-5}\sum_{n=1}^{+\infty} n^4 r^{2n}\Big)^{1/2} <$$

$$< 3\varepsilon^{1/2}B(\mathscr{A}, N)\big(1+N^{-5}\cdot 24(1-r^2)^{-5}\big)^{1/2} = \tag{25}$$

$$= 3\varepsilon^{1/2}B(\mathscr{A}, N)\big(1+24N^{-5}1-e^{-2/N})^{-5}\big)^{1/2} <$$

$$< 3\varepsilon^{1/2}B(\mathscr{A}, N)\left(1+24N^{-5}\left(\frac{1}{N}\right)^{-5}\right)^{1/2} = 15\varepsilon^{1/2}B(\mathscr{A}, N)$$

since we have

$$1-e^{-x} = x-\frac{x^2}{2!}+\frac{x^3}{3!}-\ldots > x-\frac{x^2}{2!} = x\left(1-\frac{x}{2}\right) > \frac{x}{2} \quad \text{for} \quad 0<x<1.$$

7. In this section, we complete the proof of Theorem 2. By (23) and (25), for all ε and $N>N_0(\varepsilon)$ we have

$$e^{-2}B(\mathscr{A}, N) \leqq \mathscr{J} < 15\varepsilon^{1/2}B(\mathscr{A}, N)$$

hence

$$e^{-2}B(\mathscr{A}, N) < 15\varepsilon^{1/2}B(\mathscr{A}, N)$$

$$\frac{1}{15e^2} < \varepsilon^{1/2}.$$

But for sufficiently small ε (e.g., for $\varepsilon=3\cdot 10^{-5}$), this inequality cannot hold. Thus in fact, the indirect assumption (2) leads to a contradiction which completes the proof of Theorem 2.

8. Sections 8, 9 and 10 will be devoted to the proof of Theorem 3. The proof is based on the probabilistic method of Erdős and Rényi [1], [2]. The Halberstam—Roth book [5] contains an excellent exposition of this method thus we use the terminology and notation of this book. In this section, we give a survey of those notations, facts and results connected with this probabilistic method which will be needed in the proof of Theorem 3.

Let Ω denote the set of the strictly increasing sequences of positive integers.

LEMMA 1. *Let*

$$\alpha_1, \alpha_2, \alpha_3, \ldots \tag{26}$$

be real numbers satisfying

$$0 \leqq \alpha_n \leqq 1 \qquad (n = 1, 2, \ldots). \tag{27}$$

Then there exists a probability space (Ω, S, μ) *with the following two properties:*

(i) *For every natural number* n, *the event* $E^{(n)}=\{\mathcal{A}: \mathcal{A}\in\Omega, n\in\mathcal{A}\}$ *is measurable, and* $\mu(E^{(n)})=\alpha_n$.

(ii) *The events* $E^{(1)}, E^{(2)}, \dots$ *are independent.*

This is Theorem 13 in [5], p. 142.

We denote by $\varrho(\mathcal{A}, n)$ the characteristic function of the event $E^{(n)}$:

$$\varrho(\mathcal{A}, n) = \begin{cases} 1 & \text{if } n\in\mathcal{A} \\ 0 & \text{if } n\notin\mathcal{A} \end{cases}$$

so that

$$A(n) = \sum_{j=1}^{n} \varrho(\mathcal{A}, j).$$

Furthermore, we denote the number of solutions of

$$a_x + a_y = n, \quad a_x\in\mathcal{A}, \quad a_y\in\mathcal{A}, \quad a_x < a_y \tag{28}$$

by $r(n)=r(\mathcal{A}, n)$ so that

$$|R(\mathcal{A}, n) - 2r(\mathcal{A}, n)| \leqq 1 \tag{29}$$

(where $R(\mathcal{A}, n)$ is the number of solutions of (28) without the restriction $a_x<a_y$).

LEMMA 2. *If the sequence* (26) *satisfies* (27) *and*

$$\alpha_j = \alpha j^{-c} \quad \textit{for} \quad j \geqq j_0$$

where α, c *are constants such that* $0<\alpha$, $0<c<1$, *then, with probability* 1, *we have*

$$A(n) \sim \frac{\alpha}{1-c} n^{1-c}.$$

This lemma is a consequence of Lemmas 10 and 11 in [5], pp. 144—145.

The crucial point of the proof is the use of the following result of Erdős and Rényi [2]:

LEMMA 3. *If* $\varepsilon>0$ *and the sequence* (26) *is defined by*

$$\alpha_j = \frac{1}{2} j^{(2+\varepsilon)^{-1}-1} \quad \textit{for} \quad j = 1, 2, \dots, \tag{30}$$

then, with probability 1,

$$R(\mathcal{A}, n)\bigl(\leqq 2r(\mathcal{A}, n)+1\bigr) < 4(1+\varepsilon^{-1})+1 \quad \textit{for} \quad n > n_1(\varepsilon, \mathcal{A}).$$

See Theorem 2 and its proof in [5], pp. 111 and 151—152; see also (29).

We shall need also the Borel—Cantelli lemma:

LEMMA 4. *Let* (X, S, μ) *be a probability space and let* $F_1, F_2, \dots$ *be a sequence of measurable events. If*

$$\sum_{j=1}^{+\infty} \mu(F_j) < +\infty,$$

then, with probability 1, *at most a finite number of the events* F_j *can occur.*

See [5], p. 135.

9. For $\mathscr{A}\in\Omega$, we write

$$T(\mathscr{A},n)=\sum_{\substack{a\leq n\\ a-1\in\mathscr{A},\,a\in\mathscr{A}}}1$$

so that

$$B(\mathscr{A},n)+T(\mathscr{A},n)=\sum_{\substack{a\leq n\\ a-1\notin\mathscr{A},\,a\in\mathscr{A}}}1+\sum_{\substack{a\leq n\\ a-1\in\mathscr{A},\,a\in\mathscr{A}}}1=\sum_{\substack{a\leq n\\ a\in\mathscr{A}}}1=A(n). \tag{31}$$

LEMMA 5. *If the sequence* (26) *satisfies* (27) *and*

$$\sum_{j=1}^{+\infty}\alpha_j\alpha_{j+1}<+\infty, \tag{32}$$

then, with probability 1,

$$T(\mathscr{A},n)<4\log n\quad\textit{for}\quad n>n_2(\mathscr{A})$$

(where n_2 may depend on both the sequence (26) *and $\mathscr{A}$).*

PROOF. We have to give an upper bound for $\mu(\{\mathscr{A}\colon T(\mathscr{A},n)\geq 4\log n\})$. We write

$$\lambda_n=2\log n.$$

Then

$$T(\mathscr{A},n)=\sum_{\substack{a\leq n\\ a-1\in\mathscr{A},\,a\in\mathscr{A}}}1=\sum_{j=2}^{n}\varrho(\mathscr{A},j-1)\varrho(\mathscr{A},j)\geq 2\lambda_n$$

implies that either

$$T_1(\mathscr{A},n)\overset{\text{def}}{=}\sum_{1\leq i\leq n/2}\varrho(\mathscr{A},2i-1)\varrho(\mathscr{A},2i)\geq\lambda_n$$

or

$$T_2(\mathscr{A},n)\overset{\text{def}}{=}\sum_{1\leq i<n/2}\varrho(\mathscr{A},2i)\varrho(\mathscr{A},2i+1)\geq\lambda_n$$

holds so that

$$\begin{aligned}
&\mu(\{\mathscr{A}\colon T(\mathscr{A},n)\geq 2\lambda_n\})\leq\\
&\leq\mu(\{\mathscr{A}\colon T_1(\mathscr{A},n)\geq\lambda_n\})+\mu(\{\mathscr{A}\colon T_2(\mathscr{A},n)\geq\lambda_n\})=\\
&=\sum_{d\geq\lambda_n}\mu(\{\mathscr{A}\colon T_1(\mathscr{A},n)=d\})+\sum_{d\geq\lambda_n}\mu(\{\mathscr{A}\colon T_2(\mathscr{A},n)=d\})=\\
&=\sum_{d=\lambda_n}u_n(d)+\sum_{d\geq\lambda_n}v_n(d)
\end{aligned} \tag{33}$$

where

$$u_n(d)=\mu(\{\mathscr{A}\colon T_1(\mathscr{A},n)=d\})=\mu\Big(\Big\{\mathscr{A}\colon\sum_{1\leq i\leq n/2}\varrho(\mathscr{A},2i-1)\varrho(\mathscr{A},2i)=d\Big\}\Big)=$$

$$=\sum_{1\leq i_1<\ldots<i_d\leq n/2}\prod_{j=1}^{d}\alpha_{2i_j-1}\alpha_{2i_j}(1-\alpha_{2i_j-1}\alpha_{2i_j})^{-1}\prod_{1\leq i\leq n/2}(1-\alpha_{2i-1}\alpha_{2i})$$

and similarly,

$$v_n(d) = \mu\big(\{\mathcal{A}\colon T_2(\mathcal{A}, n) = d\}\big) = \mu\Big(\Big\{\mathcal{A}\colon \sum_{1 \le i < n/2} \varrho(\mathcal{A}, 2i)\varrho(\mathcal{A}, 2i+1) = d\Big\}\Big) =$$

$$= \sum_{1 \le i_1 < \dots < i_d < n/2} \prod_{j=1}^{d} \alpha_{2i_j}\alpha_{2i_j+1}(1-\alpha_{2i_j}\alpha_{2i_j+1})^{-1} \prod_{1 \le i < n/2} (1-\alpha_{2i}\alpha_{2i+1})$$

so that for any real number x,

$$U_n(x) \stackrel{\text{def}}{=} \sum_{0 \le d \le n/2} u_n(d)x^d = \sum_{1 \le i \le n/2} \big((1-\alpha_{2i-1}\alpha_{2i}) + \alpha_{2i-1}\alpha_{2i}x\big) \tag{34}$$

and

$$V_n(x) \stackrel{\text{def}}{=} \sum_{0 \le d < n/2} v_n(d)x^d = \prod_{1 \le i < n/2} \big((1-\alpha_{2i}\alpha_{2i+1}) + \alpha_{2i}\alpha_{2i+1}x\big). \tag{35}$$

By (32), (33), (34) and (35), we have

$$\mu\big(\{\mathcal{A}\colon T(\mathcal{A}, n) \ge 2\lambda_n\}\big) \le \sum_{d \ge \lambda_n} u_n(d) + \sum_{d \ge \lambda_n} v_n(d) \le$$

$$\le \sum_{d \ge \lambda_n} u_n(d)e^{d-\lambda_n} \sum_{d \ge \lambda_n} v_n(d)e^{d-\lambda_n} = e^{-\lambda_n}\Big(\sum_{d \ge \lambda_n} u_n(d)e^d + \sum_{d \ge \lambda_n} v_n(d)e^d\Big) \le$$

$$\le e^{-\lambda_n}\Big(\sum_{0 \le d \le n/2} u_n(d)e^d + \sum_{0 \le d < n/2} v_n(d)e^d\Big) = e^{-2\log n}\big(U_n(e) + V_n(e)\big) =$$

$$= n^{-2}\Big(\prod_{1 \le i \le n/2} (1+\alpha_{2i-1}\alpha_{2i}(e-1)) + \prod_{1 \le i < n/2} (1+\alpha_{2i}\alpha_{2i+1}(e-1))\Big) \le$$

$$\le n^{-2} \prod_{1 \le j < n} \big(1+\alpha_j\alpha_{j+1}(e-1)\big) < n^{-2} \prod_{1 \le j < n} \exp\big(\alpha_j\alpha_{j+1}(e-1)\big) =$$

$$= n^{-2}\exp\Big((e-1)\sum_{1 \le j < n} \alpha_j\alpha_{j+1}\Big) < n^{-2}\exp\Big(2\sum_{j=1}^{+\infty} \alpha_j\alpha_{j+1}\Big) < cn^{-2}$$

(where c depends on the sequence (26)) since

$$1+x < e^x \quad \text{for} \quad x > 0.$$

Thus we have

$$\sum_{n=1}^{+\infty} \mu\big(\{\mathcal{A}\colon T(\mathcal{A}, n) \ge 4\log n\}\big) < \sum_{n=1}^{+\infty} cn^{-2} < +\infty$$

so that by the Borel—Cantelli lemma (Lemma 4), with probability 1, at most a finite number of the events $T(\mathcal{A}, n) \ge 4\log n$ $(n=1, 2, \dots)$ can occur which completes the proof of the lemma.

10. In this section, we complete the proof of Theorem 3.

Let us define the sequence (26) by

$$\alpha_j = \frac{1}{2} j^{-(1/2)-\varepsilon}. \tag{36}$$

Then by Lemma 3 $\left(\text{with } \frac{4\varepsilon}{1-2\varepsilon} \text{ in place of } \varepsilon\right)$, with probability 1, $R(\mathscr{A}, n)$ is bounded. (In fact, for large n we have

$$R(\mathscr{A}, n) < 4\left(1+\left(\frac{4\varepsilon}{1-2\varepsilon}\right)^{-1}\right)+1 = 3+\varepsilon^{-1}.)$$

Furthermore, by Lemma 2, with probability 1, we have

$$A(n) \sim \frac{1}{2}\left(\frac{1}{2}-\varepsilon\right)^{-1} n^{(1/2)-\varepsilon}$$

so that, with probability 1,

(37) $$A(n) > \frac{1}{2}2n^{(1/2)-\varepsilon} = n^{(1/2)-\varepsilon}$$

for n large enough.

By Lemma 5 (note that, clearly, the sequence (36) satisfies (32)), with probability 1,

(38) $$T(\mathscr{A}, n) < 4 \log n$$

for n large enough.

In view of (31), (37) and (38) yield that, with probability 1,

$$B(\mathscr{A}, n) = A(n) - T(\mathscr{A}, n) > n^{(1/2)-\varepsilon} - 4 \log n > \frac{1}{2} n^{(1/2)-\varepsilon} \quad \text{for} \quad n > n_3(\varepsilon, \mathscr{A}).$$

Thus, with probability 1, both (i) and (ii) in Theorem 3 hold, so that there exists infinitely many sequences satisfying both (i) and (ii), which completes the proof of Theorem 3.

REFERENCES

[1] Erdős, P., Problems and results in additive number theory, *Colloque sur la Théorie des Nombres* (CBRM) (Bruxelles, 1955), Georges Thone, Liège; Masson et Cie, Paris, 1956, 127—137. *MR* **18**—18.

[2] Erdős, P. and Rényi, A., Additive properties of random sequences of positive integers, *Acta Arith.* **6** (1960), 83—110. *MR* **22** # 10970.

[3] Erdős, P. and Sárközy, A., Problems and results on additive properties of general sequences, I, *Pacific J. Math.* (to appear).

[4] Erdős, P. and Sárközy, A., Problems and results on additive properties of general sequences, II, *Acta Math. Acad. Sci. Hungar.*, (to appear).

[5] Halberstam, H. and Roth, K. F., *Sequences*, Second edition, Springer-Verlag, New York—Berlin, 1983. *MR* **83m**: 10094.

(Received January 20, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZET
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 65—73

TOTALLY SEPARABLE PACKING AND COVERING WITH CIRCLES

G. FEJES TÓTH

Dedicated to my father on his 70th birthday

A system of sets in the plane is said to be totally separable if any two sets of the system can be separated by a straight line which is disjoint from all sets of the system. In [2], where this notion was introduced, an upper bound was given for the density of a totally separable system of congruent convex domains: If for a convex domain K of area $a(K)$ $f(n)=f(n, K)$ denotes the minimal area of an n-gon containing K, then the upper density of a totally separable system of congruent copies of K is at most $a(K)/f(4)$. It was also pointed out in [2] that this bound cannot be exceeded if instead of congruent copies of K we consider a totally separable system of affine images of K the quotient of areas of any two of which is at least $\frac{f(4)-f(5)}{f(3)-f(4)}$. Moreover, the bound $a(K)/f(4)$ is best possible if K is centrally symmetric.

The definition above implies that a totally separable system of sets is automatically a packing. However, one can modify the definition so that it applies also for more general arrangements of sets. We shall say that the system $\mathcal{S}$ of sets in the plane is *totally separable* if each element S of $\mathcal{S}$ contains a subset S' such that the system $\mathcal{S}'$ of the sets S' is a totally separable packing and $\bigcup_{S\in\mathcal{S}} S=\bigcup_{S'\in\mathcal{S}'} S'$. In particular, this definition applies also for coverings, and using an argument analogous to the one given in [2] one easily proves a dual counterpart to the result of [2]:

Let $F(n)=F(n, K)$ denote the maximal area of an n-gon inscribed into K. Then the lower density d of a totally separable covering of the plane with affine images of K such that the quotient of the areas of any two domains of the covering is at least $\frac{F(5)-F(4)}{F(4)-F(3)}$, satisfies the inequality $d\geqq\frac{a(K)}{F(4)}$. The bound $\frac{a(K)}{F(4)}$ cannot be improved if K is centrally symmetric.

In the special case when K is a circle we have the following bounds: If $\mathcal{P}$ is a totally separable packing of circles such that the quotient of areas of any two circles from $\mathcal{P}$ is at least $\frac{4-5\tan\pi/5}{3\tan\pi/3-4}=0.307\ldots$ then the upper density of $\mathcal{P}$ is at most $\pi/4=0.785\ldots$. If $\mathcal{C}$ is a totally separable covering of the plane with circles such that

1980 *Mathematics Subject Classification.* Primary 52A45.
Key words and phrases. Totally separable packing, density bounds.

the quotient of the areas of any two circles from $\mathscr{C}$ is at least $\frac{5\sin 2\pi/5-4}{4-3\sin 2\pi/3}=$ $=0.538\ldots$ then the lower density of $\mathscr{C}$ is at least $\pi/2=1.570\ldots$. The bounds $\pi/4$ and $\pi/2$ are attained by the open incircles and the closed circumcircles of the squares in a square grid. For packing of congruent circles a stronger result is proved in [1].

What happens if we drop the restriction on the size of the circles? Of course, in order to guarantee the existence of the upper and lower densities, we have to consider arrangements of uniformly bounded sets. Let θ be the maximal density of a totally separable packing of circles of radius less than 1 and let Θ be the minimal density of a totally separable covering with circles of radius less than 1. The incircles of the faces of the Archimedean tiling (3, 6, 3, 6) form a totally separable packing with density $\frac{11\pi}{24\sqrt{3}}=0.831\ldots$. The circumcircles of the faces of the same tiling form a totally separable covering with density $\frac{5\pi}{6\sqrt{3}}=1.511\ldots$. Thus we have

$$\theta \geqq \frac{11\pi}{24\sqrt{3}}$$

and

$$\Theta \leqq \frac{5\pi}{6\sqrt{3}}.$$

It may be conjectured that we have equality in the last two inequalities. However it is not quite easy to give an upper bound for θ and a lower bound for Θ better than the trivial estimates $\theta \leqq 1 \leqq \Theta$. The following theorem contains such nontrivial bounds.

THEOREM. *We have*

$$\theta \leqq \min_{0<\alpha<\pi/4} \frac{4\pi+12\alpha-8\sin 2\alpha+\sin 4\alpha}{4(\pi-2a+2\tan\alpha)} < 0.979044 \tag{1}$$

and

$$\Theta \geqq \max_{0<\alpha<\pi/4} \frac{4\pi-12\alpha+8\sin 2\alpha-\sin 4\alpha}{4(\pi-\alpha+\sin\alpha\cos\alpha)} > 1.01158. \tag{2}$$

We emphasize the following

COROLLARY. *Let $\mathscr{L}$ be a set of countably many straight lines in the plane such that the diameter of each face of the tiling $\mathscr{T}$ generated by $\mathscr{L}$ is at most* 1. *Then the upper density of the incircles of the faces of $\mathscr{T}$ is at most* 0.98 *and the lower density of the circumcircles of the faces of $\mathscr{T}$ is at least* 1.01.

We start with the proof of inequality (1). Let $\mathscr{P}$ be a totally separable packing of circles of radii not exceeding 1. Let S be a square of side-length s and S^* a square of side-length $s+4$ concentric with and homothetic to S. We consider the subset $\mathscr{P}_S$ of those circles from $\mathscr{P}$ which intersect S. Obviously, these circles are packed into S^*. To each pair of circles from $\mathscr{P}_S$ we draw a straight line separating these circles and disjoint from all circles of $\mathcal{P}$. The resulting lines divide S^* into certain convex polygons; with each circle C from $\mathcal{P}_S$ we associate the polygon D_C from this subdivision which contains C. Then the polygons D_C, $C\in\mathcal{P}_S$, form a packing in S^*, thus we have

$$\sum_{C\in\mathcal{P}_S} a(D_C) \leqq (s+4)^2. \tag{3}$$

We shall denote the radius of a circle C by $r(C)$. Let C_1 and C_2 be two circles from $\mathcal{P}_S$ centred at c_1 and c_2, respectively such that $r(C_1)\leqq r(C_2)$. Let t_1 be a boundary-point of C_1 and t_2 a boundary-point of C_2 such that the line t_1t_2 is a common tangent of C_1 and C_2 and C_1 and C_2 are situated on the same side of this line. Further let s_1 and s_2 be boundary-points of C_1 and C_2, respectively such that the line s_1s_2 is tangent to C_1 and C_2, separates C_1 and C_2 and intersects the lines t_1t_2 and c_1c_2 in the points y and v, respectively, so that the order of points is y, s_1, v, s_2 (Fig. 1). Write

$$\alpha = \sphericalangle t_1c_1y.$$

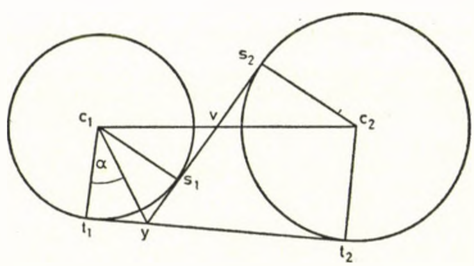


Fig. 1

Since $r(C_1)\leqq r(C_2)$, we have

$$\sphericalangle t_1c_1v = 2\alpha + \sphericalangle s_1c_1v \geqq \frac{\pi}{2},$$

and consequently

$$\sphericalangle s_1c_1v \geqq \frac{\pi}{2} - 2\alpha.$$

It follows that

$$\frac{r(C_1)}{vc_1} = \frac{r(C_2)}{vc_2} = \cos \sphericalangle s_1c_1v \leqq \cos\left(\frac{\pi}{2}-2\alpha\right) = \sin 2\alpha. \tag{4}$$

Hence we obtain the following

LEMMA. *If C_1 and C_2 are two circles from a totally separable packing, α is a real number from the interval $(0, \pi/4)$ and for $i=1,2$ the polygon D_{C_i} is contained in the circle of radius $r(C_i)/\cos\alpha$ concentric with C_i then the circles of radius $r(C_1)/\sin 2\alpha$ and $r(C_2)/\sin 2\alpha$ concentric with C_1 and C_2, respectively do not intersect.*

Further we observe that not only y but also its mirror image y' reflected in the line c_1c_2 belongs to D_{C_1}. Thus we have

$$\operatorname{conv}(\{y, y'\} \cup C_1) \subset D_{C_1}.$$

Write

$$h(\alpha) = \frac{\pi}{\pi - 2\alpha + 2\tan\alpha}.$$

Then we have for $\alpha \leqq \frac{\pi}{4}$ $a(\operatorname{conv}(\{y, y'\} \cup C_1) = a(C_1)/h(\alpha))$, and consequently

$$\frac{a(C_1)}{a(D_{C_1})} \leqq h(\alpha).$$

Observe that $h(\alpha)$ is a strictly decreasing function for $0 \leqq \alpha \leqq \frac{\pi}{4}$ taking the value 1 for $\alpha = 0$ and the value $\frac{\pi}{\pi+4}$ for $\alpha = \frac{\pi}{4}$.

For $0 < x < 1$ let $\mathscr{P}_S(x)$ consist of those circles from $\mathscr{P}_S$ for which $\frac{a(C)}{a(D_C)} \geqq x$. If $x \geqq \frac{2\pi}{\pi+4}$ and the circles C_1 and C_2 considered above are from $\mathscr{P}_S(x)$ then we have

$$x < \frac{a(C_1)}{a(D_{C_1})} \leqq h(\alpha). \tag{5}$$

We define a function $g(x)$ for $\frac{2\pi}{\pi+4} \leqq x \leqq 1$ by

$$g(x) = \sin^2 2\alpha,$$

where α is the unique root of the equation

$$h(\alpha) = x$$

in the interval $0 \leqq \alpha \leqq \pi/4$. Then (4) and (5) imply that the circles of radius $r(C_1)/\sqrt{g(x)}$ and $r(C_2)/\sqrt{g(x)}$ concentric with C_1 and C_2, respectively do not intersect. This means that the set of circles obtained by replacing each circle C from $\mathscr{P}_S(x)$ by a concentric circle of radius $r(C)/\sqrt{g(x)}$ is a packing. Since these circles are contained in a square of side-length $s+2+2/\sqrt{g(x)}$, we have for $\frac{2\pi}{\pi+4} \leqq$ $\leqq x \leqq 1$

$$\sum_{C \in \mathscr{P}_S(x)} a(C) \leqq \left(s+2+\frac{2}{\sqrt{g(x)}}\right)^2 g(x). \tag{6}$$

We choose a sequence of real numbers x_i, $i=0, \dots, n$, such that $0=x_0$, $\frac{2\pi}{\pi+4} \leqq x_1 < x_2 < \dots < x_n = 1$. For $i=1, \dots, n$ we write

$$v_i = \sum_{C \in \mathcal{P}_S(x_{i-1}) \setminus \mathcal{P}_S(x_i)} a(C)$$

and

$$w_i = \sum_{C \in \mathcal{P}_S(x_{i-1}) \setminus \mathcal{P}_S(x_i)} a(D_C).$$

Then we have by definition

$$v_i \leqq x_i w_i. \tag{7}$$

Furthermore (3) and (6) imply that

$$\sum_{i=k}^{n} v_i \leqq \left(s+2+2/\sqrt{g(x_{k-1})}\right)^2 g(x_{k-1}) \tag{8}$$

and

$$\sum_{i=1}^{n} w_i \leqq (s+4)^2. \tag{9}$$

It follows from (7) and (9) that

$$\sum_{C \in \mathcal{P}_S} a(C) = \sum_{i=1}^{n} v_i = v_1 + \sum_{i=2}^{n} v_i \leqq x_1 w_1 + \sum_{i=2}^{n} v_i \leqq$$

$$\leqq x_1 \left[(s+4)^2 - \sum_{i=2}^{n} w_i\right] + \sum_{i=2}^{n} v_i \leqq x_1 \left[(s+4)^2 - \sum_{i=2}^{n} \frac{v_i}{x_i}\right] + \sum_{i=2}^{n} v_i =$$

$$= x_1 \left[(s+4)^2 + \sum_{i=2}^{n} \left(\frac{1}{x_1} - \frac{1}{x_i}\right) v_i\right].$$

Using (8) we see that

$$\sum_{i=k}^{n} \left(\frac{1}{x_{k-1}} - \frac{1}{x_i}\right) v_i = \left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right) v_k + \sum_{i=k+1}^{n} \left(\frac{1}{x_{k-1}} - \frac{1}{x_i}\right) v_i \leqq$$

$$\leqq \left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right) \left[\left(s+2+\frac{2}{\sqrt{g(x_{k-1})}}\right)^2 g(x_{k-1}) - \sum_{i=k+1}^{n} v_i\right] + \sum_{i=k+1}^{n} \left(\frac{1}{x_{k-1}} - \frac{1}{x_i}\right) v_i =$$

$$= \left(s+2+\frac{2}{\sqrt{g(x_{k-1})}}\right)^2 g(x_{k-1}) \left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right) + \sum_{i=k+1}^{n} \left(\frac{1}{x_k} - \frac{1}{x_i}\right) v_i.$$

Combining this with the previous inequality we obtain by induction

$$\sum_{C \in \mathcal{P}_S} a(C) \leqq x_1 \left[(s+4)^2 + \sum_{i=2}^{n} \left(s+2+\frac{2}{\sqrt{g(x_{i-1})}}\right)^2 g(x_{i-1}) \frac{x_i - x_{i-1}}{x_i x_{i-1}}\right].$$

Provided that the numbers $x_1, \dots, x_n$ constitute a sufficiently fine subdivision of the interval $(x_1, 1)$ the sum $\sum_{i=2}^{n} \frac{g(x_i)}{x_{i-1} x_i} (x_i - x_{i-1})$ will get arbitrarily close to the integral $\int\limits_{x_1}^{1} \frac{g(x)}{x^2}\,dx$. Thus we can choose to any positive ε these numbers so that

$$\sum_{i=2}^{n} \frac{g(x_{i-1})}{x_i x_{i-1}}(x_i - x_{i-1}) < \int\limits_{x_1}^{1} \frac{g(x)}{x^2}\,dx + \varepsilon.$$

Further we choose s so that $(s+2+2/\sqrt{g(x_{n-1})})^2/s^2 < 1+\varepsilon$. Then we have

$$\sum_{C\in\mathcal{P}_S} a(C) \leqq (1+\varepsilon)^2 s^2 x_1\Bigl(1+\int\limits_{x_1}^{1}\frac{g(x)}{x^2}\,dx\Bigr),$$

or equivalently

$$\frac{\sum_{C\in\mathcal{P}_S} a(C)}{a(S)} \leqq (1+\varepsilon)^2 x_1\Bigl(1+\int\limits_{x_1}^{1}\frac{g(x)}{x^2}\,dx\Bigr).$$

This means that the upper density of $\mathcal{P}$, and thus θ as well, is at most $x_1\Bigl(1+\int\limits_{x_1}^{1}\frac{g(x)}{x^2}\,dx\Bigr)$. Letting α_1 be the number from the interval $[0, \pi/4]$ for which $h(\alpha_1)=x_1$ and evaluating the integral we obtain

$$\theta \leqq x_1\Bigl(1+\int\limits_{x_1}^{1}\frac{g(x)}{x^2}\,dx\Bigr) = h(\alpha_1)\Bigl[1+\frac{2}{\pi}\int\limits_{\alpha_1}^{0}(\sin^2 2\alpha - 4\sin^2\alpha)\,d\alpha\Bigr] =$$

$$= \frac{\pi}{\pi-2\alpha_1+2\tan\alpha_1}\Bigl[1+\frac{1}{\pi}\Bigl(3\alpha_1-2\sin 2\alpha_1+\frac{1}{4}\sin 4\alpha_1\Bigr)\Bigr] =$$

$$= \frac{4\pi+12\alpha_1-8\sin 2\alpha_1+\sin 4\alpha_1}{4(\pi-2\alpha_1+2\tan\alpha_1)}.$$

A numerical computation shows that $\min\limits_{0<\alpha<\pi/4}\frac{4\pi+12\alpha-8\sin 2\alpha+\sin 4\alpha}{4(\pi-2\alpha+2\tan\alpha)}$ is attained by some α from the interval $0.71<\alpha<0.72$ and it is less than 0.979044. Now the proof of inequality (1) is complete.

Let now $\mathcal{C}$ be a totally separable covering by circles of radii not exceeding 1. As above, let S be a square of side-length s, further let $\hat{S}$ be a square of side-length $s-4$ concentric with and homothetic to S. We consider the set $\mathcal{C}_S$ of those circles from $\mathcal{C}$ which are contained in S. By definition we can associate with each circle C from $\mathcal{C}$ a subset C' such that the sets C' form a totally separable tiling of the plane. We write $D_C = C'\cap\hat{S}$. Then it is clear that the set D_C, $C\in\mathcal{C}_S$, is either empty or a convex polygon contained in C and we have

$$\sum_{C\in\mathcal{C}_S} a(D_C) = (s-4)^2. \tag{10}$$

Let $\mathscr{C}_S(x)$ be the subset of $\mathscr{C}_S$ consisting of the circles C for which $\dfrac{a(C)}{a(D_C)} < x$. We write

$$H(\alpha) = \frac{\pi}{\pi - \alpha + \sin\alpha\cos\alpha}.$$

It is easy to check that $H(\alpha)$ is decreasing for $0<\alpha<\pi/4$, $H(0)=1$ and $H(\pi/4)= =\dfrac{4\pi}{3\pi+2}$. Let us suppose that $x \leqq \dfrac{4\pi}{3\pi+2}$ and let $\alpha=\alpha(x)$ be the root of the equation $x=H(\alpha)$ in the interval $0\leqq x\leqq \pi/4$. We observe that $H(\alpha)$ is nothing else but the quotient of the area of a circle and the area of a segment of angle $2(\pi-\alpha)$ of this circle. It follows that if $C\in\mathscr{C}_S(x)$ then the circle C of radius $r(C)\cos\alpha$ concentric with C is contained in D_C. This means that the system $\mathscr{S}$ of the circles C is a totally separable packing. Moreover, these circles can be separated by the same lines as the polygons D_C. Since $D_C \subset C$, we can apply the Lemma to the circles from $\mathscr{S}$ and see that the circles of radius $r(C)/2\sin\alpha$ concentric with C, $C\in\mathscr{C}_S(x)$, form a packing in the square of side-length $s-2+1/\sin\alpha$ concentric with and homothetic to S. Thus defining the function $G(x)$, $1<x<\dfrac{4\pi}{3\pi+2}$, implicitly by

$$G(x) = 4\sin^2\alpha, \quad x = H(\alpha), \quad 0<\alpha<\pi/4$$

we have

$$\sum_{C\in\mathscr{C}_S(x)} a(C) \leqq \left(s-2+2/\sqrt{G(x)}\right)^2 G(x). \tag{11}$$

We choose now a sequence of real numbers x_i, $i=0, \dots, n$ such that $1=x_0< <x_1<\dots<x_n\leqq \dfrac{4\pi}{3\pi+2}$. Further we write $x_{n+1}=\infty$ and $\mathscr{C}_S(\infty)=\mathscr{C}_S$. Defining the quantities V_i and W_i $(i=1, \dots, n+1)$ by

$$V_i = \sum_{C\in\mathscr{C}_S(x_i)\setminus\mathscr{C}_S(x_{i-1})} a(C)$$

and

$$W_i = \sum_{C\in\mathscr{C}_S(x_i)\setminus\mathscr{C}_S(x_{i-1})} a(D_C)$$

we have by definition

$$V_i \geqq x_{i-1}W_i \qquad (i=1, \dots, n+1),$$

by (10)

$$\sum_{i=1}^{n+1} W_i = (s-4)^2$$

and by (11)

$$\sum_{i=1}^{k} V_i \leqq \left(s-2+2/\sqrt{G(x_k)}\right)^2 \qquad (k=1, \dots, n).$$

Using these relations we obtain

$$\sum_{C\in\mathscr{C}_S} a(C) = \sum_{i=1}^{n+1} V_i = V_{n+1} + \sum_{i=1}^{n} V_i \geqq x_n W_{n+1} + \sum_{i=1}^{n} V_i =$$

$$= x_n\left[(s-4)^2 - \sum_{i=1}^{n} W_i\right] + \sum_{i=1}^{n} V_i \geqq x_n\left[(s-4)^2 - \sum_{i=1}^{n} \frac{V_i}{x_{i-1}}\right] + \sum_{i=1}^{n} V_i =$$

$$= x_n\left[(s-4)^2 - \sum_{i=1}^{n}\left(\frac{1}{x_{i-1}} - \frac{1}{x_n}\right)V_i\right]$$

and

$$\sum_{i=1}^{k}\left(\frac{1}{x_{i-1}} - \frac{1}{x_k}\right)V_i = \left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right)V_k + \sum_{i=1}^{k-1}\left(\frac{1}{x_{i-1}} - \frac{1}{x_k}\right)V_i \leqq$$

$$\leqq \left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right)\left[\left(s-2+\frac{2}{\sqrt{G(x_k)}}\right)^2 G(x_k) - \sum_{i=1}^{k-1} V_i\right] + \sum_{i=1}^{k-1}\left(\frac{1}{x_{i-1}} - \frac{1}{x_k}\right)V_i =$$

$$= \left(s-2+\frac{2}{\sqrt{G(x_k)}}\right)^2 G(x_k)\left(\frac{1}{x_{k-1}} - \frac{1}{x_k}\right) + \sum_{i=1}^{k-1}\left(\frac{1}{x_{i-1}} - \frac{1}{x_{k-1}}\right)V_i.$$

Hence we obtain by induction

$$\sum_{C\in\mathscr{C}_S} a(C) \geqq x_n\left[(s-4)^2 - \sum_{i=1}^{n}\left(s-2+\frac{2}{\sqrt{G(x_i)}}\right)^2 G(x_i)\frac{x_i - x_{i-1}}{x_i x_{i-1}}\right].$$

Given x_n and an arbitrary positive number ε, we can choose the numbers $x_1, \ldots, x_{n-1}$ and s so that the right-hand side of the last inequality is greater than $(1-\varepsilon)^2 s^2 x_n\left(1 - \int_1^{x_n} \frac{G(x)}{x^2}\,dx\right)$. It follows that

$$\Theta \geqq x_n\left(1 - \int_1^{x_n} \frac{G(x)}{x^2}\,dx\right).$$

Denoting by α_n the number from the interval $[0, \pi/4]$ for which $x_n = H(\alpha_n)$ and evaluating the integral we obtain

$$\Theta \geqq x_n\left(1 - \int_1^{x_n} \frac{G(x)}{x^2}\,dx\right) = H(\alpha_n)\left[1 - \frac{2}{\pi}\int_0^{\alpha_n} (1-\cos 2\alpha)^2\,d\alpha\right] =$$

$$= \frac{\pi}{\pi - \alpha_n + \sin\alpha_n\cos\alpha_n}\left[1 - \frac{1}{4\pi}(12\alpha_n - 8\sin 2\alpha_n + \sin 4\alpha_n)\right] =$$

$$= \frac{4\pi - 12\alpha_n + 8\sin 2\alpha_n - \sin 4\alpha_n}{4(\pi - \alpha_n - \sin\alpha_n\cos\alpha_n)}.$$

One easily checks that $\max\limits_{0<\alpha<\pi/4} \dfrac{4\pi-12\alpha+8\sin 2\alpha-\sin 4\alpha}{4(\pi-\alpha+\sin\alpha\cos\alpha)}$ is attained for $0.52<\alpha<0.53$ and it is greater than 1.01158. This completes the proof of the Theorem.

ACKNOWLEDGEMENT. This paper was written during a stay at the Geometrical Institute of the Technical University of Braunschweig. I thank the DFG (Deutsche Forschungsgemeinschaft) for the financial support making possible this visit.

REFERENCES

[1] BEZDEK, A., Locally separable circle packings, *Studia Sci. Math. Hungar.* **18** (1983), 371—375.
[2] FEJES TÓTH, G. and FEJES TÓTH, L., On totally separable domains, *Acta Math. Acad. Sci. Hungar.* **24** (1973), 229—232. *MR* **48** #1052.

(Received January 22, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 75—78

ON MULTIDIMENSIONAL UNIVERSAL FUNCTIONS

M. HORVÁTH

Dedicated to Professor L. Fejes Tóth on the occasion of his 70th birthday

J. Marcinkiewicz ([1], or [8, p. 118 Theorem 3.3]) proved the following

THEOREM A. *Let $0\neq\lambda_n\in\mathbf{R}$, $\lambda_n\to 0$. Then each $F\in C[a,b]$, except for a set of first category in $C[a,b]$, satisfies the following property:*

if $f\in S[a,b]$ (i.e. f is measurable and finite a.e. in $[a,b]$), there exists a subsequence $\{\lambda_{n_k}\}$ such that

$$\lim_{k\to\infty}\frac{F(x+\lambda_{n_k})-F(x)}{\lambda_{n_k}}=f(x)\quad \textit{a.e. on } [a,b].$$

Later, I. Joó [2, 3] generalized this theorem proving

THEOREM B. *Let $0\neq\lambda_n\in\mathbf{R}$, $\lambda_n\to 0$. Then for each $F\in L^1(a,b)$, except for a set of first category, the following statement is valid:*

for any $0\leqq p<1$ and $f\in L^p(a,b)$ there exists a subsequence $\{\lambda_{n_k}\}$ of $\{\lambda_n\}$ for which

$$\lim_{k\to\infty}\frac{F(x+\lambda_{n_k})-F(x)}{\lambda_{n_k}}=f(x)$$

in the sense of the L^p-metric.

He stated in [3] the problem: does Theorem B remain valid for $p\geqq 1$ or not?

The investigations of I. Joó were continued by A. Sövegjártó [4, 5], A. Bogmér [5, 6] and Z. Buczolich [7]. In [5] A. Sövegjártó and A. Bogmér gave negative answer for this problem and Z. Buczolich extended this result for the multidimensional case (when we consider a domain of $\mathbf{R}^N$ instead of (a,b)). The author of the present paper got an estimate (cf. (1)), which implies the negative answer for the question of I. Joó This result was obtained independently from that of [5], [7] and is based on a completely different method. Namely, we prove the following

THEOREM. *Let $\Omega\subset\mathbf{R}^N$ be a domain and $f\in L^1(\Omega)$. Denote for $\varepsilon>0$ $\Omega_{-\varepsilon}:=\{x\in\Omega\colon \operatorname{dist}(x,\partial\Omega)>\varepsilon\}$. Suppose that f is not constant in any direction, i.e.*

1980 *Mathematics Subject Classification.* Primary 28A20.

Key words and phrases. L^p-mean of difference quotients, density of difference quotients of a function.

there is no nonzero vector $e\in\mathbf{R}^N$ *for which* $f(x+\varepsilon e)=f(x)$ *holds a.e. in* $\Omega_{-\varepsilon}$ *for any* $\varepsilon>0$. *Then there are positive constants* $c_f>0$ *and* $\delta_f>0$ *such that*

$$\int_{\Omega_{-|\lambda|}} |f(x+\lambda)-f(x)|\,dx \geqq c_f|\lambda| \qquad (\lambda\in\mathbf{R}^N,\ 0<|\lambda|<\delta_f) \tag{1}$$

holds.

Proof. Consider first the case when Ω is bounded. Fix a sequence $(\varrho_n)\subset C_0^\infty(\mathbf{R}^N)$ satisfying the conditions

(a) $\varrho_n \geqq 0$,

(b) $\operatorname{supp}\varrho_n \subset S\left(0,\dfrac{1}{n}\right)$,

(c) $\varrho_n(x)=\varrho_n(-x) \quad (x\in\mathbf{R}^N)$,

(d) $\int_{\mathbf{R}^N}\varrho_n = 1.$

Let $0<\delta'<\delta$ be arbitrary numbers. For small λ and large n (this depends only on δ and δ') we have

$$\int_{\Omega_{-|\lambda|}} |f(x+\lambda)-f(x)|\,dx \geqq \int_{\Omega_{-|\lambda|}}\left|\int_{\Omega_{-|\lambda|}}[f(y+\lambda)-f(y)]\varrho_n(x-y)\,dy\right|dx \geqq$$

$$\geqq \int_{\Omega_{-\delta}}\left|\int_{\Omega_{-\delta'}} f(z)[\varrho_n(x-z+\lambda)-\varrho_n(x-z)]\,dz\right|dx =$$

$$= |\lambda|\int_{\Omega_{-\delta}}\left|\int_{\Omega_{-\delta'}} f(z)\left\langle \varrho_n'(x-z+\vartheta\lambda),\frac{\lambda}{|\lambda|}\right\rangle dz\right|dx,$$

where $0\leqq\vartheta=\vartheta(x-z,\lambda,n)\leqq 1$.

If

$$c_f := \liminf_{|\lambda|\to 0}\int_{\Omega_{-\delta}}\left|\int_{\Omega_{-\delta'}} f(z)\left\langle \varrho_n'(x-z+\vartheta\lambda),\frac{\lambda}{|\lambda|}\right\rangle dz\right|dx > 0,$$

then (1) is satisfied. Indirectly suppose that $c_f=0$, i.e. there exists a sequence $|\lambda_k|\to 0$ (n is fixed here!) such that

$$\int_{\Omega_{-\delta}}\left|\int_{\Omega_{-\delta'}} f(z)\left\langle \varrho_n'(x-z+\vartheta\lambda_k),\frac{\lambda_k}{|\lambda_k|}\right\rangle dz\right|dx \to 0. \tag{2}$$

Taking a subsequence, if it is necessary, we can suppose that $\dfrac{\lambda_k}{|\lambda_k|}\to e_0^{(n)}$ $(k\to\infty)$. Then (2) and the boundedness of Ω yield

$$\int_{\Omega_{-\delta}}\left|\int_{\Omega_{-\delta'}} f(z)\langle \varrho_n'(x-z),e_0^{(n)}\rangle\,dz\right|dx = 0, \quad \text{i.e.}$$

$$0 = \int_{\Omega_{-\delta'}} f(z)\langle\varrho_n'(x-z),e_0^{(n)}\rangle\,dz = D_{e_0^{(n)}}\Big(\int_{\Omega_{-\delta'}} f(z)\varrho_n(x-z)\,dz\Big) \tag{3}$$

for $x \in \Omega_{-\delta}$, i.e. $x \mapsto \int\limits_{\Omega_{-\delta'}} f(z)\varrho_n(x-z)\,dz$ is constant in the direction of $e_0^{(n)}$. Extracting a subsequence we can suppose that

$$e_0^{(n)} \to e_0$$

and then

$$\tag{4} \begin{aligned} &\int\limits_{\Omega_{-\delta-\varepsilon}} \Big| \int\limits_{\Omega_{-\delta'}} [f(z+\varepsilon e_0)-f(z+\varepsilon e_0^{(n)})]\,\varrho_n(x-z)\,dz\Big|\,dx \leqq \\ &\leqq \int\limits_{\Omega_{-\delta'}} |f(z+\varepsilon e_0)-f(z+\varepsilon e_0^{(n)})|\,dz \to 0 \qquad (n\to\infty,\ \varepsilon>0 \text{ fixed}). \end{aligned}$$

Thus

$$\tag{5} \begin{aligned} &\int\limits_{\Omega_{-\delta-\varepsilon}} \Big| \int\limits_{\Omega_{-\delta'}} [f(z+\varepsilon e_0)-f(z)]\,\varrho_n(x-z)\,dz\Big|\,dx \leqq \\ &\leqq \int\limits_{\Omega_{-\delta-\varepsilon}} \Big| \int\limits_{\Omega_{-\delta'}} [f(z+\varepsilon e_0)-f(z+\varepsilon e_0^{(n)})]\,\varrho_n(x-z)\,dz\Big|\,dx + \\ &+ \int\limits_{\Omega_{-\delta-\varepsilon}} \Big| \int\limits_{\Omega_{-\delta'}} [f(z+\varepsilon e_0^{(n)})-f(z)]\,\varrho_n(x-z)\,dz\Big|\,dx \to 0 \qquad (n\to\infty), \end{aligned}$$

the second term of the right-hand side being 0 by (3). It is well-known that

$$\int\limits_{\Omega_{-\delta'}} [f(z+\varepsilon e_0)-f(z)]\,\varrho_n(x-z)\,dz \to f(x+\varepsilon e_0)-f(x) \quad (n\to\infty) \quad \text{in} \quad L^1(\Omega_{-\delta-\varepsilon}).$$

Hence (5) implies

$$\int\limits_{\Omega_{-\delta-\varepsilon}} |f(z+\varepsilon e_0)-f(z)|\,dz = 0$$

for each $\varepsilon>0$, i.e. f is constant in $\Omega_{-\delta}$ in the direction of e_0.

Consider for a fixed $\delta>0$ the set of the directions in which f is constant in $\Omega_{-\delta}$. This set is not empty by the above arguments, forms a compact set in the unit ball of $\mathbf{R}^N$ and decreases with $\delta\to 0$, hence there is a "common" direction in which f must be constant in Ω; this contradiction proves the theorem in the case of a bounded domain.

Suppose finally, that Ω is not bounded. If (1) holds with

$$\Omega(R) := \Omega \cap \{x\in\mathbf{R}^N\colon\ \|x\| \leqq R\}$$

instead of Ω, then a fortiori it holds with Ω. If it is not so then there is a direction in which $f|_{\Omega(R)}$ is constant. Repeating the above idea we can find a "common" direction $\bar{e}$ for $R\to+\infty$ and f is constant in this direction in the whole Ω, contradiction.

The proof of the Theorem is complete.

REMARK 1. The Theorem remains valid if we replace (1) by

$$\tag{1'} \|f(x+\lambda)-f(x)\|_{L^p(\Omega)} \geqq c_{f,p}|\lambda|$$

$$(1 \leqq p < \infty,\ f\in L^p(\Omega),\ 0 \leqq |\lambda| \leqq \delta_{f,p}).$$

Indeed, we can suppose that Ω is bounded and then

$$\|f(x+\lambda)-f(x)\|_{L^p(\Omega)} \geqq c_{p,\Omega} \|f(x+\lambda)-f(x)\|_{L^1(\Omega)},$$

hence the Theorem can be applied.

REMARK 2. The assertion (1′) is not true for $p=\infty$ as the step functions show.

REFERENCES

[1] MARCINKIEWICZ, J., Sur les nombres dérivés, *Fund. Math.* **24** (1935), 305—308. *Zbl.* **11**, 107.

[2] JOÓ, I., Note to a theorem of Talaljan on universal series and to a problem of Nikišin, *Fourier analysis and approximation theory* (Proc. Colloq., Budapest, 1976), Vol. I, Colloq. Math. Soc. János Bolyai, 19, North-Holland, Amsterdam, 1978, 451—458. *MR* **81k**: 42025.

[3] JOÓ, I., On the divergence of eigenfunction expansions, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* (to appear).

[4] SÖVEGJÁRTÓ, A., Egy megjegyzés univerzális elemekről (A note on universal elements), *Mat. Lapok* **28** (1977—1980), 327—328. *MR* **82f**: 47044.

[5] SÖVEGJÁRTÓ, A. and BOGMÉR, A., On universal functions, *Acta Math. Hungar.* **49** (1987), 237—239.

[6] BOGMÉR, A., Lineáris operátorsorozatok univerzális elemeiről, *Mat. Lapok* **31** (1978—1983), 195—196. *MR* **85h**: 41049.

[7] BUCZOLICH, Z., Universal series and universal functions, *Acta Sci. Math. (Szeged)* (to appear).

[8] BRUCKNER, A. M., *Differentiation of real functions*, Lecture notes in mathematics, 659, Springer, Berlin—New York, 1978. *MR* **80h**: 26002.

(Received February 5, 1985)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 79—82

KUGELPACKUNGEN MIT VORGEGEBENEN NACHBARNZAHLEN

G. FEJES TÓTH und H. HARBORTH

Herrn Professor Dr. László Fejes Tóth zu seinem 70. Geburtstag

Eine Packung von kongruenten Kugeln im dreidimensionalen Raum heißt *n-Nachbarnpackung*, wenn jede Kugel genau n Nachbarn hat. Welches ist die kleinste Anzahl s_n von kongruenten Kugeln, so daß eine n-Nachbarnpackung existiert? Da mindestens $n+1$ Kugeln notwendig sind, folgt $s_n=n+1$ für $n=1, 2, 3$ aus den eindeutigen Beispielen, bei denen die Mittelpunkte paarweise den gleichen Abstand haben. Es läßt sich noch leicht zeigen, daß für $n=4$ mindestens 6 Kugeln notwendig sind, und daß 6 Kugeln nur eine 4-Nachbarnpackung bilden, wenn die Mittelpunkte in den Eckpunkten eines regulären Oktaeders liegen. Es gilt also $s_4=6$. In dieser Arbeit wollen wir beweisen:

SATZ. *Eine 5-Nachbarnpackung von kongruenten Kugeln besteht aus mindestens* 12 *Kugeln.*

Legt man die Mittelpunkte der Kugeln in die Eckpunkte eines regulären Ikosaeders, so ergibt sich damit dann $s_5=12$.

Für $n=6$ ist uns nur eine 6-Nachbarnpackung mit 240 Kugeln bekannt, die uns Gerd Wegner kürzlich mitgeteilt hat. In einem Dodekaeder sind dabei die Kanten durch 6 Kugeln ersetzt, deren Mittelpunkte ein reguläres Oktaeder bilden, und an den Eckpunkten des Dodekaeders werden diese Oktaeder-Kanten durch 3 Kugeln verbunden. In der Figur sind die Mittelpunkte der Kugeln und für die Nachbarschaften deren Verbindungen angedeutet. Für $n\geqq 10$ ist die Nichtexistenz von s_n bekannt (siehe [1], [4]). Auch für $n=9$ wird vermutet, daß s_n nicht existiert [2]. Für $n=7$ und $n=8$ liegen keine Ergebnisse vor. Die entsprechende Fragestellung für Kreise in der Ebene wurde in [3] behandelt.

Zum Beweis des Satzes betrachten wir eine Packung von Einheitskugeln $K_1, K_2, \dots$ mit den Mittelpunkten $M_1, M_2, \dots$. Zunächst beweisen wir:

HILFSSATZ. *Haben in einer Packung von Einheitskugeln zwei Kugeln K_1 und K_2 die Kugeln K_3, K_4 und K_5 als gemeinsame Nachbarn, so können höchstens zwei davon Nachbarn von i paarweise benachbarten Kugeln K_{5+i} sein* $(1\leqq i\leqq 4)$.

Diese Arbeit entstand während eines Gastaufenthaltes des erstgenannten Autors an der Technischen Universität Braunschweig. Wir danken der Deutschen Forschungsgemeinschaft für die Finanzierung.

1980 *Mathematics Subject Classification.* Primary 52A45.
Key words and phrases. n-neighbour packings of balls.

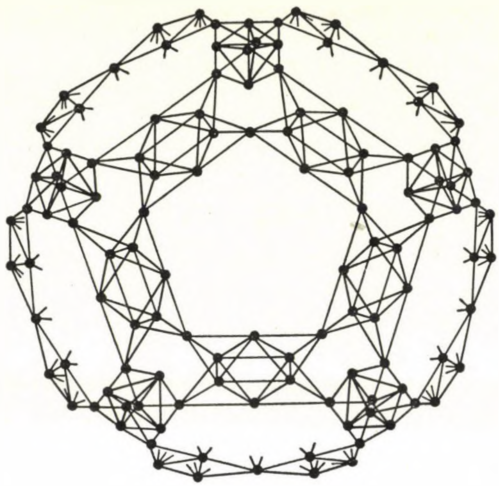


Fig. 1

BEWEIS DES HILFSSATZES. Die beiden Kugeln vom Radius 2 um die Mittelpunkte M_1 und M_2 schneiden sich in einem Kreis C, auf dem M_3, M_4 und M_5 mit Abständen mindestens 2 liegen. Die Halbierungspunkte H_3, H_4 und H_5 der Kreisbögen von M_3 zu M_4 ohne M_5, von M_4 zu M_5 ohne M_3 und von M_5 zu M_3 ohne M_4 bilden ein Dreieck, dessen Seiten ebenfalls nicht kürzer als 2 sind. Die orthogonalen Projektionen M'_{5+i} der Mittelpunkte M_{5+i} von zu K_3, K_4 oder K_5 benachbarten Kugeln K_{5+i} auf die durch C bestimmte Ebene E liegen in E nicht innerhalb von C und auf C nur, wenn schon M_{5+i} selbst auf C liegt, oder wenn $M_1M_2=2$ und M'_{5+i} in einen der Punkte M_3, M_4 oder M_5 fällt. Da die Kugeln K_{5+i} paarweise benachbart sind, haben dann die Projektionen M''_{5+i} von M'_{5+i} auf C paarweise Abstände kleiner als 2, mit den einzigen Ausnahmen, daß die beiden ursprünglichen Mittelpunkte selbst schon auf C liegen oder im Falle $M_1M_2=2$ senkrecht zu E über zwei der Kugelmittelpunkte M_3, M_4 oder M_5 liegen.

Nur von einer der Kugeln K_{5+i} können K_3, K_4 und K_5 nicht gleichzeitig berührt werden, da schon K_1 und K_2 die maximal 2 Kugeln sind, die jede 3 Kugeln (K_3, K_4 und K_5) berühren können.

Sind nur 2 Kugeln zu K_3, K_4 und K_5 benachbart, etwa K_6 zu K_3 und zu K_4 und K_7 zu K_5, so gilt $M''_6=H_3$, und M''_7 liegt auf dem Bogen von H_4 über M_5 zu H_5. Es ergibt sich der Widerspruch $M''_6M''_7>2$ (die obigen Ausnahmen können nicht vorkommen).

Wenn von 3 Kugeln, etwa K_6, K_7 und K_8, je eine mit K_3, K_4 und K_5 benachbart ist, dann bilden M''_6, M''_7 und M''_8 ein Dreieck, dessen Eckpunkte auf die 3 durch H_3, H_4 und H_5 bestimmten Kreisbögen von C verteilt sind. Daher können nicht alle Seiten des Dreiecks $M''_6M''_7M''_8$ kleiner als die kleinste Seite des Dreiecks $H_3H_4H_5$, also kleiner als 2 sein, was wiederum einen Widerspruch ergibt, da K_6, K_7 und K_8 in den beiden obigen Ausnahmefällen nicht paarweise benachbart sein können.

Beweis des Satzes. Eine 5-Nachbarnpackung mit s Einheitskugeln hat $5s/2$ Berührpunkte, so daß nur gerade Anzahlen s von Kugeln möglich sind. Da es keine 6 Punkte mit paarweise gleichen Abständen gibt, kann $s=6$ nicht sein. Für $s=8$ seien K_1 und K_2 nicht benachbarte Kugeln. Diese haben wenigstens 4 gemeinsame Nachbarn, von denen mindestens 3 mit einer weiteren Kugel benachbart sein müssen, was ein Widerspruch zum Hilfssatz ist.

Um auch noch $s=10$ auszuschließen, betrachten wir 2 nicht benachbarte Kugeln K_1 und K_2 und deren (I) 5, (II) 4, (III) 3 und (IV) 2 eventuelle gemeinsame Nachbarn.

(I) Von den 5 gemeinsamen Nachbarn zu K_1 und K_2 berührt jede der restlichen 3 Kugeln mindestens 3 im Widerspruch zum Hilfssatz.

(II) Die gemeinsamen Nachbarn von K_1 und K_2 seien K_3, K_4, K_5 und K_6. Ist K_1 noch zu K_7 und K_2 noch zu K_8 benachbart, so können K_9 und K_{10} wegen des Hilfssatzes jeweils zu höchstens 2 der Kugeln K_3, K_4, K_5 oder K_6 benachbart sein, so daß K_9 und K_{10} sowohl voneinander als auch von K_7 und von K_8 Nachbarn sein müssen. Nun können nach dem Hilfssatz die 4 fehlenden Nachbarschaften von K_9 und K_{10} nur zu 2 der Kugeln K_3, K_4, K_5 oder K_6, etwa zu K_3 und K_4, bestehen. Das wiederspricht jedoch auch dem Hilfssatz, da dann K_1, K_9 und K_{10} jeweils K_3, K_4 und K_7 berühren.

(III) Gemeinsame Nachbarn von K_1 und K_2 seien K_3, K_4 und K_5. Außerdem seien K_6 und K_7 von K_1, sowie K_8 und K_9 von K_2 die weiteren Nachbarn. Wäre K_{10} zu 2 der Kugeln K_3, K_4, K_5 benachbart, so hätten K_{10} zusammen mit K_1 oder K_2 dann 4 gemeinsame Nachbarn, was bei (II) schon betrachtet wurde. Also muß K_{10} etwa K_3 und dann K_6, K_7, K_8 und K_9 berühren. Jede der Kugeln K_4, K_5, K_6, K_7, K_8 und K_9 kann nicht zu beiden der Kugeln K_4 und K_5, K_6 und K_7 und K_8 und K_9 benachbart sein, denn berührt als ein Beispiel etwa K_4 sowohl K_6 als auch K_7, so haben im Widerspruch zum Hilfssatz alle 3 zu K_1 und K_{10} gemeinsamen Nachbarn K_3, K_6 und K_7 einen Berührpunkt mit dem benachbarten Kugelpaar K_2, K_4. Ohne Einschränkung können die 2 noch fehlenden Nachbarn von K_3 unter K_6, K_7, K_8 und K_9 angenommen werden. Daher müssen K_4 und K_5, K_4 und etwa K_6, sowie dann K_5 und K_7 Nachbarn sein. Dann bestehen aber im Widerspruch zum Hilfssatz Nachbarschaften zwischen den 3 paarweise benachbarten Kugeln K_2, K_4 und K_5 und den gemeinsamen Nachbarn K_3, K_6 und K_7 von K_1 und K_{10}.

(IV) Nun haben K_1 die Nachbarn K_3, K_4, K_5, K_6 und K_7 und K_2 die Nachbarn K_3, K_4, K_8, K_9 und K_{10}. Die Tripel K_5, K_6, K_7 und K_8, K_9, K_{10} sind paarweise untereinander benachbart, da sonst zum Beispiel K_5 mit K_2 mindestens 3 gemeinsame Nachbarn hat, was schon zu Widersprüchen führte. Weder K_3 noch K_4 kann 2 Kugeln eines der Tripel berühren, denn hat etwa K_3 die Nachbarn K_5 und K_6, so sind K_3 und K_7 nicht benachbart (sonst bilden die 5 Kugeln K_1, K_3, K_5, K_6 und K_7 eine 4-Nachbarnpackung), und K_3 und K_7 haben die 3 gemeinsamen Nachbarn K_1, K_5 und K_6, was schon bei (III) zum Widerspruch führte. Nun muß K_3 mit K_4 und etwa mit K_5 und K_8 benachbart sein. Auch K_4 muß K_5 und K_8 berühren, da sonst K_4 und K_5 oder K_4 und K_8 nicht benachbarte Kugeln mit 3 gemeinsamen Nachbarn sind. Damit folgen dann die paarweise Nachbarschaften von K_6, K_7, K_9 und K_{10}. Jetzt haben aber im Widerspruch zum Hilfssatz etwa die 3 gemeinsamen Nachbarn

K_1, K_2 und K_5 von K_3 und K_4 Nachbarschaften etwa mit den 3 paarweise benachbarten Kugeln K_6, K_7 und K_9.

Damit ist der Satz bewiesen.

Abschließend sei noch die Vermutung erwähnt, daß die Eckpunkte des regulären Ikosaeders wohl die einzige Möglichkeit für die Mittelpunkte der Kugeln einer 5-Nachbarnpackung mit 12 Kugeln bilden, also daß das Ikosaeder eindeutig bestimmt ist, wie auch für $n=3$ das Tetraeder und für $n=4$ das Oktaeder.

LITERATURVERZEICHNIS

[1] Fejes Tóth, G., Ten-neighbour packing of equal balls, *Period. Math. Hungar.* **12** (1981), 125—127. *MR* **82e**: 52013.

[2] Fejes Tóth, L. and Sachs, H., Research problem 17, *Period. Math. Hungar.* **7** (1976), 87—89.

[3] Österreicher, F. and Rohm, W., Die minimale 3-Nachbarnpackung kongruenter Kreise, *Math. Semesterber.* **30** (1983), 49—60. *MR* **84g**: 52024.

[4] Sachs, H., No more than nine unit balls can touch a closed hemisphere, *Studia Sci. Math. Hungar.* **21** (1986), 203—206.

(Eingegangen am 27. März 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

INSTITUT FÜR GEOMETRIE
TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG
POCKELSSTRASSE 14
D—3300 BRAUNSCHWEIG
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 83—94

GOTHIC CLASSES OF GROUPOID-LATTICES IN THE THEORY OF RADICALS

N. V. LOI and O. STEINFELD

To Professor L. Fejes Tóth on his 70th birthday

It is known (see e.g. [10]) that the absorbent of an element of a groupoid-lattice is a common generalization of the notions of the two-sided ideal of a ring (semigroup) and the normal subgroup of a group. The purpose of this paper is to define Gothic sets in a groupoid-lattice and Gothic classes of certain varieties, and to study these notions in the context of the general radical theory. For instance, the monotony and hereditary properties of a Gothic set can be characterized by a simple property of every absorbent of any element of a groupoid-lattice. Furthermore, in the cases of rings, groups and modules the Gothic semisimple classes with hereditary upper radical classes will be determined.

§ 1. Basic concepts and examples

A set with a binary operation $\cdot$ is said to be a *groupoid.* A partially ordered groupoid $(\mathbf{V};\ \cdot,\ \leqq)$ is called a *groupoid-lattice* if it is a complete lattice with respect to the partial ordering $\leqq$ and it has the properties

(i) $A^2 \leqq A$ for all $A \in \mathbf{V}$;

(ii) If 0 and E are the least and the greatest elements of the complete lattice $\mathbf{V}$, then $0 \cdot E = E \cdot 0 = 0$.

In this paper $\mathbf{V}$ always denotes a groupoid-lattice. The following notion has a leading part throughout this paper.

An element B of $\mathbf{V}$ is called an *absorbent* of the element A of $\mathbf{V}$ if

(iii) $B \leqq A$,

(iv) $AB \vee BA \leqq B$.

Groupoid-lattices were investigated in several papers of the second named author (e.g. [8], [9], [10]). Let us recall three important examples for groupoid-lattices.

1) The set $\mathbf{V}$ of all subrings of an associative ring A is a groupoid-lattice with respect to the inclusion and multiplication. From the definitions, it is clear that the ideals of A are just the absorbents of the element A of $\mathbf{V}$ (cf. [10] example 1). (Later we shall say that $\mathbf{V} = \mathbf{V}_A$ is the groupoid-lattice of A.)

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1813.

1980 *Mathematics Subject Classification.* Primary 06F99; Secondary 16A21, 16A63, 20F99.

Key words and phrases. Groupoid-lattice, Gothic set and class, radical and semisimple class of rings and groups, torsion theory of modules.

2) Let R be any associative ring and A be any left R-module (briefly module). The set $\mathbf{V}$ of all submodules of A is a groupoid-lattice with respect to the inclusion and the zero-multiplication, that is, if B and C are elements of $\mathbf{V}$ then $B \cdot C = 0$. Thus $\mathbf{V}$ is a groupoid-lattice, and each submodule of A is an absorbent of $A(\in \mathbf{V})$.

3) Now let $\mathbf{V}$ be the set of all subgroups of a group G. It is well-known that $\mathbf{V}$ is a complete lattice with respect to the inclusion $\subseteq$. Evidently, G is the greatest element and the unity element e of G is the least element of $\mathbf{V}$. If the lattice operations in $\mathbf{V}$ are denoted by $\wedge$ and $\vee$, then $H \wedge K$ is the intersection of the subgroups H and K. Furthermore, $H \vee K$ is the subgroup of G generated by H and K.

Let $[H, K]$ denote the subgroup of G generated by all commutators $[h, k] = h^{-1}k^{-1}hk$ $(h \in H, k \in K)$. Then $\mathbf{V}$ is a commutative groupoid with respect to the binary operation $[H, K]$, and thus $\mathbf{V}$ is a groupoid-lattice (cf. [10] example 3). By [8] a subgroup H of a group G is a normal subgroup of G if and only if H is an absorbent of the element G of $\mathbf{V}$.

In the sequel, the sign $\triangleleft$ will indicate the ideal of a ring or normal subgroup of a group. Radicals in the sense of Kurosh and Amitsur will be considered. Let us recall that a class $\mathbf{R}$ of rings is a *radical class*, if $\mathbf{R}$ is homomorphically closed, closed under extensions (i.e. if B and A/B are in $\mathbf{R}$, then also A is in $\mathbf{R}$), and $\mathbf{R}(A) = \sum (I \mid I \triangleleft A$ and $I \in \mathbf{R}) \in \mathbf{R}$ is satisfied for all rings A. One can define radical classes for groups and modules analogously. Let us mention that radical classes of modules are usually called *torsion classes*. We shall deal also with the *semisimple class* of a radical class $\mathbf{R}$ defined as

$$\mathcal{S}\mathbf{R} = \{A \mid \mathbf{R}(A) = 0\}$$

in the case of rings or modules (for groups $\mathbf{R}(G) = \{e\}$ is required). A class $\mathbf{S}$ of rings or modules (or groups) is a semisimple class if and only if $\mathbf{S}$ is *hereditary* (i.e. if $I \triangleleft A \in \mathbf{S}$, then $I \in \mathbf{S}$), $\mathbf{S}$ is closed under extensions, and under subdirect sums (products, respectively). The upper radical of a semisimple class $\mathbf{S}$ is determined by the operator $\mathcal{U}$ as

$$\mathcal{U}\mathbf{S} = \{A \mid \text{if } B \text{ is a homomorphic image of } A \text{ and } B \in \mathbf{S}, \text{ then } |B| = 1\}.$$

For the theory of radicals we refer to [16] and for torsion theory to [12].

§ 2. Gothic sets

Let $\mathbf{V}$ be a groupoid-lattice and $\mathbf{A}$ a subset of $\mathbf{V}$. An element V of $\mathbf{A}$ will be called briefly an $\mathbf{A}$-*element*. A subset $\mathbf{A}$ of $\mathbf{V}$ is called a *Gothic set*, if the following condition is satisfied:

(A) For every element V of $\mathbf{V}$ there exists a uniquely determined $\mathbf{A}$-absorbent $\mathbf{A}(V)$ of V, such that if A is an $\mathbf{A}$-absorbent of V, then $A \leqq \mathbf{A}(V)$. (If one considers the usual representation of the complete lattice $\mathbf{V}$ as a graph, then condition (A) considered in this representation may inspire the denomination Gothic.)

The following two results are immediate consequences.

COROLLARY 1. *An element V of $\mathbf{V}$ belongs to the Gothic set $\mathbf{A}$ if and only if $\mathbf{A}(V) = V$.*

COROLLARY 2. *For any element V of* $\mathbf{V}$ *it holds* $\mathbf{A}(V)=\mathbf{A}(\mathbf{A}(V))$.

The Gothic set $\mathbf{A}$ is called *hereditary* if $V\in\mathbf{A}$ implies that every absorbent of V is also in $\mathbf{A}$. We say that the Gothic set $\mathbf{A}$ has the *monotony property* if for every absorbent W of V we have $\mathbf{A}(W)\leqq\mathbf{A}(V)$.

THEOREM 3. *A Gothic set* $\mathbf{A}$ *in a groupoid-lattice* $\mathbf{V}$ *has the monotony and the hereditary properties if and only if every absorbent W of any element V of* $\mathbf{V}$ *satisfies*

$$\mathbf{A}(W) = \mathbf{A}(V)\wedge W.$$

PROOF. First assume that a Gothic set $\mathbf{A}$ has the monotony and the hereditary properties. Let W be an absorbent of an element $V\in\mathbf{V}$. By the monotonity of $\mathbf{A}$ we have $\mathbf{A}(W)\leqq\mathbf{A}(V)$. Furthermore, since $\mathbf{A}(W)$ is an absorbent of W it holds $\mathbf{A}(W)\leqq W$. Hence $\mathbf{A}(W)\leqq\mathbf{A}(V)\wedge W$. On the other hand, it is evident that $\mathbf{A}(V)\wedge W$ is an absorbent of $\mathbf{A}(V)$. This fact, Corollaries 1, 2 and hereditariness of $\mathbf{A}$ imply that $\mathbf{A}(V)\wedge W$ belongs to the Gothic set $\mathbf{A}$. Since $\mathbf{A}(V)\wedge W$ is an absorbent of W, too, by $\mathbf{A}(V)\wedge W\in\mathbf{A}$ and by (A) we get $\mathbf{A}(V)\wedge W\leqq\mathbf{A}(W)$. So we conclude $\mathbf{A}(W)=\mathbf{A}(V)\wedge W$.

Conversely, assume that every absorbent W of any element $V\in\mathbf{V}$ satisfies $\mathbf{A}(W)=\mathbf{A}(V)\wedge W$. Let B be an absorbent of an element A of $\mathbf{V}$. Then by assumption we have $\mathbf{A}(B)=\mathbf{A}(A)\wedge B$, whence $\mathbf{A}(B)\leqq\mathbf{A}(A)$, that is, the Gothic set $\mathbf{A}$ has the monotony property. Now we consider an arbitrary $\mathbf{A}$-element V of $\mathbf{V}$ and an absorbent W of V. In view of Corollary 1 we have $\mathbf{A}(V)=V$. So we conclude by assumption:

$$W= V\wedge W= \mathbf{A}(V)\wedge W= \mathbf{A}(W).$$

By Corollary 2 this fact means that W is an $\mathbf{A}$-absorbent of V. Thus the Gothic set $\mathbf{A}$ is hereditary.

In order to give a "converse" of Theorem 3 we need a new property of a Gothic set $\mathbf{A}$. We say that a Gothic set $\mathbf{A}$ is *partially left residuated* if the following condition is satisfied:

(R_l) Let E denote the greatest element of a groupoid-lattice $\mathbf{V}$. Assume that for every absorbent W of E and for $\mathbf{A}(W)$ there exists a uniquely determined absorbent $\mathbf{A}(W) : W$ of E such that for an absorbent X of E it holds $X\leqq\mathbf{A}(W) : W$ if and only if $XW\leqq\mathbf{A}(W)$.

THEOREM 4. *Assume that a Gothic set* $\mathbf{A}$ *in a groupoid-lattice* $\mathbf{V}$ *has the monotony, the hereditary and the partially left residuated properties. Then for every absorbent W of the greatest element E of* $\mathbf{V}$ *it holds*

$$\mathbf{A}(E) = \mathbf{A}(\mathbf{A}(W)\colon W).$$

PROOF. By condition (R_l) there exists the absorbent $\mathbf{A}(W) : W$ of E for every absorbent W of E. This fact and the monotony property imply the inequality

$$\mathbf{A}(\mathbf{A}(W)\colon W) \leqq \mathbf{A}(E).$$

Since $\mathbf{A}(E)$ and W are absorbents of E, we have $\mathbf{A}(E)W\leqq\mathbf{A}(E)\wedge W$. In view of Theorem 3 this inequality and the monotony and the hereditary properties imply

$$\mathbf{A}(E).W\leqq \mathbf{A}(E)\wedge W= A(W).$$

From this relation and condition (R_l) it follows

$$\mathbf{A}(E) \leqq \mathbf{A}(W){:}W.$$

Since $\mathbf{A}(E)$ is an absorbent of E, this implies that $\mathbf{A}(E)$ is an absorbent of $\mathbf{A}(W){:}\, W$, too. This fact, Corollary 2 and the monotony property imply

$$\mathbf{A}(E) = \mathbf{A}(\mathbf{A}(E)) \leqq \mathbf{A}(\mathbf{A}(W){:}\, W).$$

So we have proved $\mathbf{A}(E)=\mathbf{A}(\mathbf{A}(W){:}\, W)$.

A *partially right residuated* Gothic set $\mathbf{A}$ is defined by means of a dual condition (R_r). Theorem 4 and its dual imply

COROLLARY 5. *Assume that a Gothic set* $\mathbf{A}$ *in a groupoid-lattice* $\mathbf{V}$ *has the monotony, the hereditary and the* (R_l), (R_r) *properties. Then for every absorbent* W *of the greatest element* E *of* V *it holds*

$$\mathbf{A}(E) = \mathbf{A}(\mathbf{A}(W){:}W) = \mathbf{A}(\mathbf{A}(W){:}{:}W),$$

where $\mathbf{A}(W){:}{:}W$ *denotes the right dual of* $\mathbf{A}(W){:}\, W$ *in conditions* (R_r).

In the following we shall deal with rings, groups and modules. Let $\mathscr{C}$ denote any variety of these algebras. An abstract subclass $\mathscr{L}$ of $\mathscr{C}$ is called a *Gothic class* if for every $A\in\mathscr{C}$ the set $\mathbf{V}_A\cap\mathscr{L}$ is Gothic, where $\mathbf{V}_A$ is the groupoid-lattice of A. It may, of course, well happen that $\mathbf{V}_A$ and $\mathbf{V}_B$ $(A, B\in\mathscr{C})$ are isomorphic groupoid-lattices, but $\mathbf{V}_A\cap\mathscr{L}$ and $\mathbf{V}_B\cap\mathscr{L}$ do not correspond to each other, as A and B are not necessarily isomorphic.

Let us remark that if $\mathscr{C}$ is the variety of associative rings (or algebras) and $\mathscr{L}$ is a subvariety of $\mathscr{C}$, then $\mathscr{L}$ is a Gothic class if and only if it is an S-closed variety in the sense of Gardner [3], which means that $I, J\lhd A$ and $I, J\in\mathscr{L}$ imply $I+J\in\mathscr{L}$.

For groups, a Gothic class is exactly a preradical in the sense of Plotkin [6].

§ 3. Rings

Let $\mathscr{L}$ be any Gothic class of rings and A be a ring, then $\mathscr{L}\cap\mathbf{V}_A$ is a Gothic set. Denote $\mathscr{L}(A)$ the maximal $(\mathscr{L}\cap\mathbf{V}_A)$-absorbent of A for the Gothic set $\mathscr{L}\cap\mathbf{V}_A$. Notice that for any $B, C\in\mathbf{V}_A$, B is an absorbent of C if and only if B is an ideal of C. Then obviously $\mathscr{L}(A)$ has the form

$$\mathscr{L}(A) = \Sigma\{I \mid I\lhd A \text{ and } I\in\mathscr{L}\cap\mathbf{V}_A\}.$$

The Gothic class $\mathscr{L}$ has the *monotony property* if for any ideal B of a ring A it holds $\mathscr{L}(B)\subseteqq\mathscr{L}(A)$, and $\mathscr{L}$ has the *hereditary property* if $A\in\mathscr{L}$, then every ideal of A is also in $\mathscr{L}$. Theorem 3 implies

THEOREM 6. *The Gothic class* $\mathscr{L}$ *has the monotony and the hereditary properties if and only if for any ideal* I *of any ring* A *it holds*

$$\mathscr{L}(I) = \mathscr{L}(A)\cap I.$$

a) *Representation of radicals*

Let **R** be a radical class of rings and let A be an arbitrary associative ring. In view of Theorem 3.2 in [16] any radical class is Gothic. Furthermore, Corollary 5.3 of [16] implies that **R** has the monotony property. From this fact and from Theorem 6 it follows

COROLLARY 7 (cf. Anderson—Divinsky—Sulińsky [1] and [16] Theorem 13.1). *The radical class* **R** *has the hereditary property if and only if it holds:*

$$\mathbf{R}(I) = \mathbf{R}(A) \cap I$$

for every ideal I of any ring A.

Let A denote again an associative ring. If B and C are ideals of A, then quotients $B{:}C$ and $B{::}C$ are also ideals of A. This fact means that the greatest element A of the groupoid-lattice $\mathbf{V}_A$ has the properties (R_l) and (R_r). Thus Corollary 5 implies

COROLLARY 8. *Let A be a ring. If $\mathscr{L}$ is a Gothic class, which has the monotony and the hereditary properties, then for any ideal I of A, it holds*

$$\mathscr{L}(A) = \mathscr{L}(\mathscr{L}(I){:}I) = \mathscr{L}(\mathscr{L}(I){::}I).$$

Consider again a radical class **R** and an associative ring A, Corollary 8 has the following special case:

COROLLARY 9. *If the radical class* **R** *is hereditary, then for any ideal I of a ring A it holds*

$$\mathbf{R}(A) = \mathbf{R}(\mathbf{R}(I){:}I) = \mathbf{R}(\mathbf{R}(I){::}I).$$

This result is similar to [7] Theorem 4 and is a kind of generalization of [2] (see also [14] pp. 65—66).

REMARK. Similarly to Example 1, the set of all subsemigroups with 0 of a semigroup S with 0 is also a groupoid-lattice. Let **R** denote an *ideal-radical class* of semigroups with 0, i.e. a class which is closed under Rees factor homomorphisms, ideal extensions, and is such that if A_i $(i \in I)$ are **R**-ideals of a semigroup with 0, so is their union. Corollary 7 has an analogue for the ideal-radical class **R** (cf. R. S. Grigor [4]). Naturally, Corollary 9 has an analogue for this radical class, too.

b) *Semisimple classes, which are Gothic classes*

In the following we will show that there exist semisimple classes which are Gothic, but not always radical classes. Furthermore, the Gothic semisimple classes with hereditary upper radical classes will be determined. The ring A is called *hereditarily idempotent* if $I^2 = I$ for all ideals I of A.

PROPOSITION 10. *Let* **S** *be any semisimple class of rings. If* **S** *is Gothic, then either* **S** *consists of hereditarily idempotent rings, or* **S** *contains every nilpotent ring.*

PROOF. First, suppose that **S** contains a non-zero nilpotent ring. Since **S** is hereditary, it follows that **S** contains a non-trivial zero-ring. Now we will show that **S** contains all nilpotent rings. Since **S** is closed under extensions, it suffices to show that **S** contains every zero-ring. Indeed, let A be any zero-ring in **S**. By the hereditariness of S there is a cyclic abelian group I, such that the zero-ring I^0 is contained in S. Since I^0 is a homomorphic image of the zero-ring Z^0 over the integers, it follows $Z^0 \notin \mathscr{U}\mathrm{S}$. It is well-known that for any radical class **R** either $Z^0 \in \mathbf{R}$ or $Z^0 \in \mathscr{S}\mathbf{R}$. Thus for $\mathbf{R} = \mathscr{U}\mathrm{S}$ we get $Z^0 \in \mathscr{S}\mathbf{R} = \mathrm{S}$, and hence $Z^0 \oplus Z^0 \in \mathrm{S}$. Let us denote

$$T_1 = \{(Z, 0) | Z \in Z^0\}; \quad T_2 = \{(0, Z) | Z \in Z^0\},$$

and

$$\Delta_n = \{(nz, nz) | z \in Z^0\}.$$

Then Δ_n is an ideal of $Z^0 \oplus Z^0 \in \mathrm{S}$. Since $T_i \cap \Delta_n = \{0\}$, $i = 1, 2$, we have

$$\frac{T_1 + \Delta_n}{\Delta_n} \cong T_1 \cong Z^0 \simeq T_2 \cong \frac{T_2 + \Delta_n}{\Delta_n}.$$

On the other hand,

$$\frac{Z^0 \oplus Z^0}{\triangle_n} = \frac{T_1 + \triangle_n}{\triangle_n} + \frac{T_2 + \triangle_n}{\triangle_n}.$$

This fact implies that $\dfrac{Z^0 \oplus Z^0}{\Delta_n}$ is a sum of its ideals, which are isomorphic to $Z^0 \in \mathrm{S}$. Since S is a Gothic class, it follows $\dfrac{Z^0 \oplus Z^0}{\Delta_n} \in \mathrm{S}$. Obviously, the ring Δ_1/Δ_n as an ideal of $\dfrac{Z^0 \oplus Z^0}{\Delta_n}$ is also in S. However, we know that $\Delta_1/\Delta_n \cong Z^0/nZ^0$, whence $Z^0/nZ^0 \in \mathrm{S}$ holds for any arbitrary natural number n. Using again the fact that S is a Gothic class, and the fact that every zero-ring is the sum of subrings which are cyclic as abelian groups, we can show that S contains every zero-ring.

Conversely, assume that S does not contain a zero-ring. We have to show that S consists of only hereditarily idempotent rings. Indirectly, suppose that S contains a non-hereditarily idempotent ring. Then by hereditariness of S there exists a ring $A \in \mathrm{S}$, such that $A^2 \neq A$. For $a \in A \setminus A^2$, let B be the subring of A generated by A^2 and a. Let us denote $K = B/A^2$. Then B is an ideal of A and K is a zero-ring. Consider the ring $D = K \oplus B$ and

$$C = \{(na + A^2) \oplus (na + c) | c \in A^2, n \in Z\}.$$

Let $\varphi: C \to B$ be a mapping such that $\varphi\big((na + A^2) \oplus (na + c)\big) = na + c$. Then $\varphi\big((na + A^2) \oplus (na + c)\big) = 0$ if and only if $na + c = 0$, and this implies $na \in A^2$, that is, $(na + A^2) \oplus (na + c) = 0$. This fact means that φ is a monomorphism. It is clear that φ is surjective, too, hence C is isomorphic to B. We can see easily that C is an ideal of D, moreover $B + C = D$. Since B is an ideal of A, the hereditariness of S implies $B \in \mathrm{S}$. Since S is Gothic, we have $D \in \mathrm{S}$, and consequently $K \in \mathrm{S}$, a contradiction. Thus S consists only of hereditarily idempotent rings.

LEMMA 11. *Let A be a ring. If I, I_i, $(i \in \Lambda)$ are ideals of A, such that $I \cap I_i = \{0\}$ $i \in \Lambda$, and $A = \sum_{i \in \Lambda} I_i$, then $I^2 = 0$, that is, I is a zero-ring.*

PROOF. Straightforward.

The ring A is called *strongly idempotent* if every subring of A is idempotent. By [13] we have that if A is strongly idempotent, then A is a subdirect sum of finite fields.

LEMMA 12. *If $I \lhd A$, $I \neq 0$ is a strongly idempotent ring, then there are ideals J, K of A such that $J \neq 0$, $J \subseteq I$ and $A = J \oplus K$.*

PROOF. Since I is strongly idempotent, it follows by [13] Theorem 3.4 (see also [16] Theorem 33.3) that every subring $[X]$ of I is a finite field for $X \neq 0$. By [16] Lemma 33.4 I is commutative. Thus I contains a non-zero idempotent element e. Let J be the ideal of I generated by e. Then J is a ring with identity e. Let $\bar{J}$ denote the ideal of A generated by J. Since I is strongly idempotent, $\bar{J}$ is idempotent. Hence the Andrunakievich lemma yields

$$J \subseteq \bar{J} = \bar{J}^3 \subseteq J,$$

that is, $J = \bar{J}$. Since J has an identity, J is a direct summand of A.

LEMMA 13. *Let* $\mathbf{S}$ *be a semisimple class of rings. If* $\mathbf{S}$ *consists only of idempotent rings, then every ring in* $\mathbf{S}$ *is strongly idempotent.*

PROOF. Let $A \in \mathbf{S}$. We have to show that each subring [a] of A is idempotent. Consider the direct product $\prod_1^\infty A$ of infinite number copies of A. Then $\bigoplus_1^\infty A \lhd \prod_1^\infty A$ (where $\bigoplus_1^\infty A$ is the direct sum of infinite number copies of A), and $\bigoplus_1^\infty A$, $\prod_1^\infty A \in \mathbf{S}$. Let B be a subring of $\prod_1^\infty A$ generated by the diagonal element $\bar{a} = (\dots, a, \dots)$ of $\prod_1^\infty A$ and by $\bigoplus_1^\infty A$. Since $\bigoplus_1^\infty A \subseteq B$, B is a subdirect sum of copies of A. Hence $B \in \mathbf{S}$. Since $\mathbf{S}$ contains only idempotent rings, and since homomorphic images of idempotent rings are also idempotent, thus $B/\bigoplus_1^\infty A$ is also an idempotent ring. Hence $[a] \cong B/\bigoplus_1^\infty A$ is idempotent.

PROPOSITION 14. *If* $\mathbf{S}$ *is a semisimple class, which consists only of idempotent rings, then* $\mathbf{S}$ *is a Gothic class, and* $\mathcal{U}\mathbf{S}$ *is a hereditary radical class.*

PROOF. Indirectly, suppose $\mathbf{S}$ is not a Gothic class. In this case, without loss of generality we can assume that there is a ring $A \notin \mathbf{S}$, such that

$$A = \sum_{i \in \Lambda} \{I_i \mid I_i \lhd A \text{ and } I_i \in \mathbf{S}\}.$$

Since the class of all hereditarily idempotent rings is a radical class, by the assumption A is a hereditarily idempotent ring. Furthermore since $A \in \mathbf{S}$, we have $\mathbf{R}(A) \neq 0$ for $\mathbf{R} = \mathcal{U}\mathbf{S}$. Thus $\mathbf{R}(A)$ is a hereditarily idempotent ring. This fact and Lemma 11 imply that there is an index $i \in \Lambda$, such that $\mathbf{R}(A) \cap I_i \neq \{0\}$. Since $\mathbf{S}$ consists of idempotent rings, and $I_i \in \mathbf{S}$, it follows by Lemma 13 that I_i and also $I_i \cap \mathbf{R}(A)$ are strongly idempotent. Hence by Lemma 12 there are ideals $J \neq 0$ and K of

A such that $A=J\oplus K$, and $J\subseteq \mathbf{R}(A)\cap I_i$. On the other hand, by $\mathbf{R}(A)\subseteq\mathscr{U}\mathbf{S}$, it follows $J\in\mathscr{U}\mathbf{S}$, and since $J\lhd I_i\cap\mathbf{R}(A)$, we get $J\in\mathscr{U}\mathbf{S}\cap\mathbf{S}=\{0\}$, a contradiction. Thus $\mathbf{R}(A)=0$, and hence $A\in\mathbf{S}$.

Now we have to show that $\mathscr{U}\mathbf{S}$ is hereditary. This is equivalent to showing that $\mathbf{S}$ is closed under essential extensions. Now let $A\in\mathbf{S}$ and A be an essential ideal of a ring B. Then either $\mathbf{R}(B)=\{0\}$ or $\mathbf{R}(B)\cap A\neq\{0\}$. If $\mathbf{R}(B)\neq\{0\}$, that is, $\mathbf{R}(B)\cap A\neq\{0\}$, then using Lemmas 12 and 13, as above, we obtain a contradiction. Thus $\mathbf{R}(B)=\{0\}$, that is, $B\in\mathbf{S}$. The proof is complete.

PROPOSITION 15. *Let* $\mathbf{S}$ *be a semisimple class, which contains every nilpotent ring. If* $\mathscr{U}\mathbf{S}$ *is hereditary, then* $\mathbf{S}$ *is a Gothic class.*

PROOF. It is enough to show that if $A=\sum_{i\in\Lambda}\{I_i|I_i\lhd A \text{ and } I_i\in\mathbf{S}\}$, then $A\in\mathbf{S}$. Since $\mathbf{S}$ contains every nilpotent ring and hence every zero-ring, Lemma 11 implies that there exists an ideal I_i of A such that $I_i\cap\mathbf{R}(A)\neq\{0\}$, whenever $\mathbf{R}(A)\neq\{0\}$. Since $\mathscr{U}\mathbf{S}=\mathbf{R}$ is hereditary, it follows $I_i\cap\mathbf{R}(A)\in\mathscr{U}\mathbf{S}$. By hereditariness of $\mathbf{S}$ we get

$$I_i\cap\mathbf{R}(A)\in\mathscr{U}\mathbf{S}\cap\mathbf{S}=\{0\},$$

a contradiction. Thus $\mathbf{R}(A)=\{0\}$ and therefore $A\in\mathbf{S}$. This implies that $\mathbf{S}$ is a Gothic class.

THEOREM 16. *Let* $\mathbf{S}$ *be a semisimple class, whose upper radical is hereditary. Then* $\mathbf{S}$ *is a Gothic class if and only if either* $\mathbf{S}$ *consists only of idempotent rings or* $\mathbf{S}$ *contains all nilpotent rings.*

PROOF. The assertion follows immediately from Propositions 10, 14 and 15.

THEOREM 17. *If* $\mathbf{S}$ *is a semisimple Gothic class, then* $\mathbf{S}$ *has the hereditary and the monotony properties.*

PROOF. By [16] Theorem 8.1 every semisimple class $\mathbf{S}$ is hereditary, that is, $\mathbf{S}$ has the hereditary property. We have only to show that $\mathbf{S}$ has the monotony property. By Proposition 10 either $\mathbf{S}$ consists of idempotent rings, or $\mathbf{S}$ contains all nilpotent rings. Let $\mathbf{S}(I)$ denote the greatest ideal of I, which is in $\mathbf{S}$, and $\overline{\mathbf{S}(I)}$ the ideal of A generated by $\mathbf{S}(I)$. If $\mathbf{S}$ consists of idempotent rings, then using the Andrunakievich Lemma we get

$$\mathbf{S}(I)\subseteq\overline{\mathbf{S}(I)}=\overline{\mathbf{S}(I)}^3\subseteq\mathbf{S}(I)$$

yielding $\mathbf{S}(I)=\overline{\mathbf{S}(I)}^3$. Since $\overline{\mathbf{S}(I)}\lhd A$, also $\mathbf{S}(I)=\overline{\mathbf{S}(I)}^3$ is an ideal of A.

If $\mathbf{S}$ contains all nilpotent rings, then by Andrunakievich Lemma $\overline{\mathbf{S}(I)}/\mathbf{S}(I)$ is nilpotent and hence it is in $\mathbf{S}$. Since $\mathbf{S}$ is closed under extensions we conclude $\overline{\mathbf{S}(I)}\in\mathbf{S}$, which implies $\mathbf{S}(I)=\overline{\mathbf{S}(I)}$. Thus $\mathbf{S}(I)$ is an ideal of A. Hence $\mathbf{S}(I)\subseteq\mathbf{S}(A)$ holds in both cases.

REMARK 1. Let $\mathbf{M}$ be any hereditary class of rings, which consists of idempotent rings. Then by [5] the lower radical of $\mathbf{M}$ is hereditary, and its semisimple class $\mathscr{S}\mathbf{M}$ contains all nilpotent rings. By Theorem 16, this semisimple class is Gothic. Since $\mathbf{S}$ does not contain every ring, by [13] (see also [16] Lemma 33.6) it follows that S is not a radical. Thus there exist semisimple classes, which are Gothic, but not radical.

REMARK 2. Since the intersection of Gothic classes is also Gothic therefore the intersection of any radical class and any semisimple Gothic class is Gothic. In this way we can construct Gothic classes, which are neither radical nor semisimple.

REMARK 3. In [3], Gardner considered S-closed varieties of rings (in our terminology: varieties of rings which are Gothic classes) and proved the following result: A non-trivial S-closed variety $\mathscr{V}$ of associative rings can contain no rings with torsion-free additive groups and consequently $\mathscr{V}$ is determined by subvarieties: $\mathscr{V}_p = \{A \in \mathscr{V} \mid \text{the additive group of } A \text{ is } p\text{-primary}\}$.

§ 4. Groups

Let $\mathscr{L}$ be any Gothic class of groups and G be a group. As in the case of rings let $\mathscr{L}(G)$ denote the maximal $(\mathscr{L} \cap \mathbf{V}_G)$-absorbent of G for the Gothic set $\mathscr{L} \cap \mathbf{V}_G$. Notice that for $A, B \in \mathbf{V}_G$, B is an absorbent of A if and only if B is a normal subgroup of A. Thus $\mathscr{L}(G)$ is the maximal normal subgroup of G, which contains every $(\mathscr{L} \cap \mathbf{V}_G)$-normal subgroup of G. This fact shows that $\mathscr{L}(G)$ is a characteristic subgroup of G. Radical classes are examples for classes of groups, which are Gothic. It is known (cf. [15]) that every radical class of groups has the monotony property. As in the case of rings the analogous results (see Theorem 6, Corollaries 7, 8 and 9) can be proved for groups.

Now we will study such semisimple classes of groups, which are Gothic. Let us notice that the proofs in the case of groups are sometimes similar to that of rings.

We shall call a group G *idempotent* if $G = [G, G]$. As in the case of rings, it can be proved that the class of all idempotent groups forms a radical class.

PROPOSITION 18. *In the variety of groups there does not exist a semisimple class* $(\neq e)$, *which consists only of idempotent groups.*

PROOF. Indirectly, let S be any non-trivial semisimple class, which consists of idempotent groups, and let $G \in \mathrm{S}$. Then the discrete direct product $\bigoplus_1^\infty G$ of copies of G is a normal subgroup of the complete direct product $\prod_1^\infty G$ of copies of G. For $e \neq g \in G$ consider the group K of $\prod_1^\infty G$ generated by the diagonal element $\bar{g} = (\ldots, g, \ldots)$ and $\bigoplus_1^\infty G$. As in the case of rings we can show that K is a subdirect product of copies of G, and hence $K \in \mathrm{S}$. Since S consists only of idempotent groups, it follows that every homomorphic image of any group of S is idempotent, hence so is $K / \bigoplus_1^\infty G$. But $K / \bigoplus_1^\infty G$ is isomorphic to the subgroup of G generated by g, and it is commutative, contradicting the assumption that S consists only of idempotent groups.

PROPOSITION 19. *Let* **S** *be any semisimple class of groups. If* **S** *is Gothic, then* **S** *contains all solvable groups.*

PROOF. Similarly to the case of rings in Proposition 10 we can show that if **S** is Gothic, then either **S** consists of idempotent groups or **S** contains all abelian groups, and hence all solvable ones. By Proposition 18 only the later case is possible. Thus if **S** is a Gothic class, then **S** must contain all solvable groups.

LEMMA 20. *Let* I_i, $i \in \Lambda$ *be normal subgroups of a group* G, *such that the subgroup* $\langle I_i, i \in \Lambda \rangle$ *of* G *generated by* I_i, $i \in \Lambda$, *coincides with* G. *If* K *is any normal subgroup of* G, *such that* $K \cap I_i = \{e\}$ *for all* $i \in \Lambda$, *then* K *is an abelian group.*

PROOF. Since $[I, K] \subseteq K \cap I$ for all normal subgroups K, I, the assertion is evident.

THEOREM 21. *Let* **S** *be any semisimple class of groups, whose upper radical class is hereditary. Then* **S** *is a Gothic class if and only if* **S** *contains all solvable groups.*

PROOF. By Proposition 19 the proof is similar to the case of Proposition 15 and Theorem 16.

THEOREM 22. *If the semisimple class* **S** *is a Gothic class, then* **S** *has the monotony property, too.*

PROOF. Since for any groups G the subgroup $\mathbf{S}(G)$ is a characteristic subgroup of G, the assertion is trivial.

REMARKS. As in the case of rings we can give examples for Gothic semisimple classes of groups, which are not radical. Now we give a concrete example for such a semisimple class. Here we would like to thank to Dr. P. P. Pálfy, who has given this example:

Let **M** be a class of groups, such that for any $G \in \mathbf{M}$ there does not exist a normal subgroup of G, which is finite, non-abelian, and simple. We can show that **M** is a semisimple class, and **M** is a Gothic class, too. The proof is based on Theorem 21. Here we omit the details.

The construction of Gothic classes, which are not radical and not semisimple, are similar to the case of rings, by using the fact that the intersection of any radical class and semisimple Gothic class is also Gothic.

§5. Modules

In the theory of modules it is usual to call a radical class a torsion class, and a semisimple class a torsionfree class. In the sequel we shall use this terminology.

For rings and groups we have shown that there are semisimple classes, which are Gothic, but not radical classes. In particular, for the modules this fact is not true, that is, if any torsionfree class is Gothic, then it must be also a torsion class. We will show this by the following stronger assertion.

PROPOSITION 23. *Let* **A** *be any class of modules, which is a Gothic class, and satisfies the following condition:*

(*) *If* $M \in \mathbf{A}$ *and* N *is a direct summand of* M, *then* $N \in \mathbf{A}$.

The following conditions are equivalent:

1) $\mathbf{A}(M/\mathbf{A}(M))=0$ *holds for each module* M,
2) $\mathbf{A}$ *is closed under extensions,*
3) $\mathbf{A}$ *is a torsion class.*

PROOF. The implications 3)$\Rightarrow$2)$\Rightarrow$1) are straightforward.

1)$\Rightarrow$3). Let M be any module in $\mathbf{A}$ and N be a submodule of $\mathbf{A}$. Denote

$$I=\{(i,i)\mid i\in N\}.$$

Then I is a submodule of the direct sum $M_1\oplus M_2$ of two copies of M, and I is contained in the diagonal submodule of $M_1\oplus M_2$. Since

$$\frac{M_1\oplus M_2}{I}=\frac{M_1+I}{I}+\frac{M_2+I}{I},$$

it follows $(M_1\oplus M_2)/I$ can be generated by the submodules $(M_1+I)/I$ and $(M_2+I)/I$. Since $M_1\cap I=0=M_2\cap I$, we have

$$\frac{M_1+I}{I}\cong M\in\mathbf{A}\ni M\cong\frac{M_2+I}{I}.$$

As $\mathbf{A}$ is a Gothic class, it follows $(M_1\oplus M_2)/I\in\mathbf{A}$. Now let K denote the diagonal submodule of $M_1\oplus M_2$, then K/I is a submodule of $(M_1\oplus M_2)/I$. Since

$$\frac{K}{I}\cap\frac{(M_1+I)}{I}=\{0\}$$

and

$$\frac{K}{I}+\frac{(M_1+I)}{I}=\frac{M_1\oplus M_2}{I},$$

it follows that K/I is a direct summand of $(M_1\oplus M_2)/I$. Since $(M_1\oplus M_2)/I\in\mathbf{A}$, we conclude by assumption (*) $K/I\in\mathbf{A}$. Since each homomorphic image of any module M is isomorphic to K/I for some I, it follows that every homomorphic image of M is also in $\mathbf{A}$. Hence $\mathbf{A}$ is closed under homomorphisms and $\mathbf{A}$ is a torsion class.

THEOREM 24. *A torsion-free class* $\mathbf{S}$ *of modules is a Gothic class if and only if* $\mathbf{S}$ *is a torsion class.*

PROOF. Since every torsion-free class is closed under extensions, and satisfies the condition (*) in Proposition 23, it follows by Proposition 23 that $\mathbf{S}$ is a torsion class. The converse is trivial.

ACKNOWLEDGEMENT. The authors are particularly indebted and grateful to Professor R. Wiegandt for his precious advices in the preparation of our paper.

REFERENCES

[1] ANDERSON, T., DIVINSKY, N. and SULIŃSKI, A., Hereditary radicals in associative and alternative rings, *Canad. J. Math.* **17** (1965), 594—603. *MR* **31** #215.

[2] ANDRUNAKIEVIČ, V. A., On a characteristic of a hyper-nilpotent radical, *Uspehi Mat. Nauk* **16** (1961), 127—130 (in Russian). *MR* **23** # A2442.

[3] GARDNER, B. J., Ring varieties closed under ideal sums, *Comment. Math. Univ. Carolinae* **18** (1977), 569—578. *MR* **58** # 5453.

[4] GRIGOR, R. S., On the theory of semigroup radicals, I, *Mat. Issled.* **6** (1971), no. 4, 37—55 (in Russian). *MR* **45** # 2055.

[5] HOFFMAN, A. E. and LEAVITT, W. G., Properties inherited by the lower radical, *Portugal. Math.* **27** (1968), 63—66. *MR* **41** # 6888.

[6] PLOTKIN, B. I., The functorials, radicals and co-radicals in groups, *Ural Gos. Univ. Mat. Zap.* **7** (1969/70), no. 3, 150—182 (in Russian). *MR* **44** # 2832.

[7] STEINFELD, O., Verbandstheoretische Betrachtung gewisser idealtheoretischer Fragen, *Acta Sci. Math. (Szeged)* **22** (1961), 136—149. *MR* **23** # A3187.

[8] STEINFELD, O., Über Gruppoid-Verbände, I, *Acta Sci. Math. (Szeged)* **31** (1970), 203—218. *MR* **43** # 3188.

[9] STEINFELD, O., Über Gruppoid-Verbände, II. Zerlegungssätze, *Period. Math. Hungar.* **4** (1973), 169—181. *MR* **49** # 7198.

[10] STEINFELD, O., On groupoid-lattices, *Contributions to general algebra* (Proc. Klagenfurt Conf., Klagenfurt, May 25—28, 1978), Heyn, Klagenfurt, 1979, 357—372. *MR* **80h**: 06018.

[11] STEINFELD, O., *Quasi-ideals in rings and semigroups,* Disquisitiones Mathematicae Hungaricae, 10, Akadémiai Kiadó, Budapest, 1978. *MR* **80e**: 16001.

[12] STENSTRÖM, Bo, *Rings and modules of quotients,* Lecture notes in mathematics, Vol. 237, Springer-Verlag, Berlin—Heidelberg—New York, 1971. *MR* **48** # 4010.

[13] STEWART, P. N., Semi-simple radical classes, *Pacific J. Math.* **32** (1970), 249—254. *MR* **41** # 254.

[14] SZÁSZ, F. A., *Radicals of rings,* Akadémiai Kiadó, Budapest, 1981; John Wiley & Sons, Ltd., Chichester, 1981. *MR* **84a**: 16012.

[15] TRAN VAN HAO (Chang Wang Hao), On semisimple classes of groups, *Sibirsk. Mat. Ž.* **3** (1962), 943—949 (in Russian).

[16] WIEGANDT, R., *Radical and semisimple classes of rings,* Queen's Papers in Pure and Applied Mathematics, no. 37, Queen's University, Kingston, Ont., 1974. *MR* **50** # 2227.

(Received April 25, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 95—101

BOUNDS ON THE NUMBER OF SMALL DISTANCES IN A FINITE PLANAR SET

K. VESZTERGOMBI

To Professor László Fejes Tóth on his 70th birthday

Abstract

Let m_j denote the number of times the j^{th} smallest distance occurs among a set S of m points in $\mathbf{R}^2$. We show that $m_j \leqq 3jm$ and $m_2 \leqq 5m$ and $m_1+m_2 \leqq 6m$. We give a construction with $m_2 \sim \frac{24}{7} m$.

Introduction

Let S be a set of m points in $\mathbf{R}^2$. We denote by t_1 the smallest distance between two points, by t_j the j^{th} smallest distance and by $m_1, \ldots, m_j, \ldots$ the number of occurrences of the distance $t_1, \ldots, t_j, \ldots$ resp. In [1] Harborth gave an exact upper bound for m_1. In the recent paper we give an upper bound on m_j, namely $m_j \leqq 3jm$, and a sharper bound on m_2; namely $m_2 \leqq 5m$. We give a construction in which $m_2 \sim \frac{24}{7} m$. In the last part we prove $m_1+m_2 \leqq 6m$, for which the lattice of regular triangles is sharp.

1

Let us assign a graph G_j to the case of the j^{th} distance: the vertices of the graph correspond to the points of the set S, and two vertices are connected by an edge if the distance of the corresponding points equals t_j. Let $d_j(v)$ denote the degree of $v \in S$ in G_j. We get the upper bound for m_j, by giving bounds for the degrees in G_j.

LEMMA 1. *For all $v \in S$, $d_j(v) \leqq 6j$.*

PROOF. Suppose a vertex v has at least $6j+1$ neighbours in G_j. Then if we take the circle of radius t_j around v, then there must exist an arc of angle $\leqq \frac{\pi}{3}$ on the circle which contains at least $j+2$ points (including the endpoints of the arc). Then this means that there are at least j different distances smaller than t_j, between the points on the arc, which contradicts the assumption that t_j is the j^{th}

1980 *Mathematics Subject Classification*. Primary 51D20, 52A37; Secondary 05B30.
Key words and phrases. Point set, distances, second smallest distances between points.

smallest distance. Furthermore, v can have $6j$ neighbours in G_j if the neighbours form a regular $6j$-gon on the circle of radius t_j around v.

From this lemma we get the upper bound:

PROPOSITION.

$$m_j \leqq 3jm.$$

In the case $j=1$ if we take the lattice of regular triangles we get the upper bound asymptotically in the sense that we do not count the deficiency of the boundary points of the set.

2

In this part we prove a sharper upper bound on m_2. By Lemma 1 we know that the degree of any point in G_2 is at most 12. But one can notice that the neighbours of a point of degree 12 must have essentially smaller degree. We call the sum of the degrees of the endpoints of an edge the *degree of the edge*. We prove for the degrees of the edges in G_2 the following upper bound:

LEMMA 2. *The degree of any edge of* G_2 *is at most* 20.

PROOF. Let uv be any edge of G_2; suppose that $d_2(v) \geqq d_2(u)$. If $d_2(v) \leqq 10$ then $d_2(u)+d_2(v) \leqq 20$, trivially. So there are two cases to consider:

(a) Case $d_2(v)=12$.

We want to show that the other endpoint u has degree at most 8. By Lemma 1 the neighbours of v in G_2 form a regular 12-gon. Let us denote them by u_i for $1 \leqq i \leqq 12$, $(u_1=u)$ and the circle around v of radius t_2 by C_1. The smallest distance t_1 must be equal to the side of this regular 12-gon, so $t_1=d(u_i, u_{i+1})$, obviously. Now let us consider u_1. Then u_1 already has some neighbours of distance t_2; namely u_3, u_{11} and evidently v. We draw the circle C_2 around u_1 of radius t_2. Let us denote by r_i the vertices of the regular 12-gon on C_2 (some of them coincide with earlier mentioned points; see Fig. 1). Then on the arc vu_3 $\left(\text{of angle } \frac{\pi}{3}\right)$ there cannot be

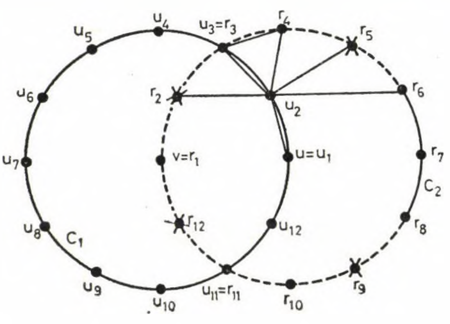


Fig. 1

any other point than r_2 because $d(v, r_2)=t_1$ and $d(v, u_3)=t_2$. But r_2 cannot belong to the set S because $t_1<d(r_2, u_2)<t_2$, obviously. Similarly on the arc vu_{11} $\left(\text{of angle } \frac{\pi}{3}\right)$ there cannot be any point of the set S. There cannot be any point except r_4 on the arc u_3r_4 $\left(\text{of angle } \frac{\pi}{3}\right)$ because $d(u_3, r_4)=t_1$, and again on the arc r_4r_5 $\left(\text{of angle } \frac{\pi}{6}\right)$ because it would give distance between t_1 and t_2. But r_5 is forbidden because $t_1<d(u_2, r_5)<t_2$. There cannot be any point s, $s\neq r_6$ on the arc r_5r_6 because $t_1<d(u_2, s)<d(u_2, r_6)=t_2$. This argument goes through symmetrically for the arc vr_8 $\left(\text{of angle } \frac{5\pi}{6}\right)$. So u_1 has at most 8 neighbours of distance t_2:

$$v, u_3 = r_3, r_4, r_6, r_7, r_8, r_{10}, r_{11} = u_{11}.$$

(b) Case $d_2(v)=11$.

(i) First we deal with the case when all the neighbours u_i $(1\leqq i\leqq 11)$ of v in G_2 form a regular 11-gon. Let us call C_1 the circle of radius t_2 around v. Obviously, $t_1=d(u_i, u_{i+1})$. Let us consider the neighbours of u_1 of distance t_2. We denote by C_2 the circle of radius t_2 around u_1. Let us denote by r_i $(1\leqq i\leqq 11)$ the vertices of the regular 11-gon on C_2 where $r_1=v$ (as on Fig. 2). On the arc vr_2 $\left(\text{of angle } \frac{2\pi}{11}\right)$

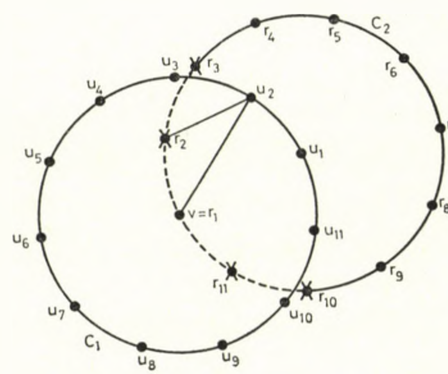


Fig. 2

there cannot be any point $s\in S$ at distance t_2 from u_1 because $d(v, s)<t_1$ would hold. r_2 is forbidden because

$$t_1 = d(v, r_2) < d(r_2, u_2) < d(v, u_2) = t_2,$$

because for the corresponding angles of the triangle we have the same inequalities

$$\frac{2\pi}{11} < \frac{5\pi}{22} < \frac{13\pi}{22}.$$

The next possibility would be the crossing point of the circles C_1, C_2 but it is an inner point of the arc u_2u_3 so it would give a smaller distance than t_1. Symmetrically the arc vr_{10} $\left(\text{of angle } 2\cdot\frac{2\pi}{11}\right)$ is forbidden. So what remained from C_2 is an arc of angle $<\frac{14\pi}{11}$. On this arc the set S can have less than 9 points, since there cannot be a smaller angle than $\frac{2\pi}{11}$, which belongs to t_1.

(ii) Secondly, we deal with the case when the u_i's do not form a regular 11-gon. Now there exist u_i, u_{i+1} (the i's are understood mod 11) for which $d(u_i, u_{i+1})>t_1$, but then it has to be at least t_2. On the other hand for all u_i, u_{i+2} pairs $d(u_i, u_{i+2})\geqq t_2$ must hold which means that the angle belonging to t_1 must be at least $\frac{\pi}{6}$. From this it follows that the only other possibility for having 11 neighbours of v in G_2 is that the u_i's form a regular 12-gon with one point deleted. Now those u_i's which are "far" from the missing point can have $\leqq 8$ neighbours by the same argument as in Case (a). So this kind of edges vu_i have degree $\leqq 20$.

We still have to check the u_i's which are "close" to the missing point, as indicated on Fig. 3.a, b. In the case when u_2 is the missing neighbour of v (see Fig. 3a) a similar argument as for $d_2(v)=12$ shows that we have to consider on the arc vr_4 $\left(\text{of angle } \frac{\pi}{2}\right)$ the point r_2. But $t_1<d(r_2, u_5)<t_2$ which means that $d_2(u_1)\leqq 9$ because r_9 and r_{12} are naturally forbidden.

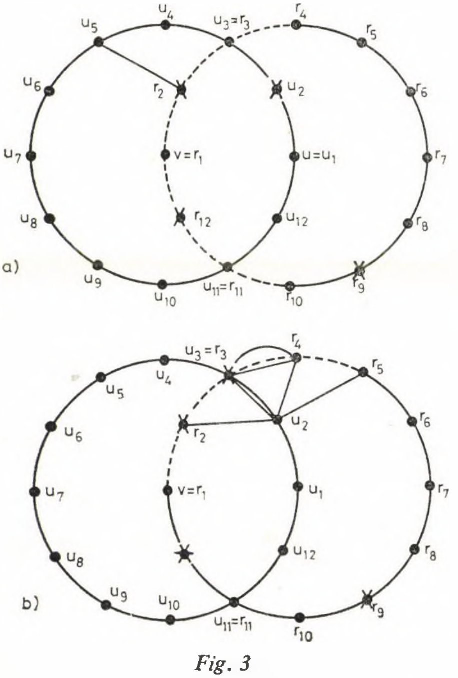


Fig. 3

In the case when u_3 is the missing neighbour of v (see Fig. 3b) then again the first possible neighbour of u_1 would be r_2, but $t_1 < d(r_2, u_2) < t_2$, so r_2 is forbidden. Since the triangle $u_2r_4u_3$ is regular, on the arc u_3r_4 we cannot have any point s because $d(u_2, s) < t_1$ would hold. So this shows $d_2(u_1) \leqq 8$. From this argument one can see that if any other u_i is missing the existing points forbid enough points to have the inequality for the degree of the edges vu_i.

THEOREM. $m_2 \leqq 5m$.

PROOF. From Lemma 2 we know that for any edge e of G_2 $d(e) \leqq 20$. Now if we sum for all edges e_i, $1 \leqq i \leqq m_2$ of G_2 we know the following:

$$d(e_1) + \ldots + d(e_{m_2}) \leqq 20m_2.$$

On the other hand we know:

$$\sum_{i=1}^{m_2} d(e_i) = \sum_{j=1}^{m} [d(v_j)]^2.$$

By the inequality for the algebraic and the quadratic mean for the degrees of the points:

$$\frac{2m_2}{m} = \frac{d(v_1) + \ldots + d(v_m)}{m} \leqq \sqrt{\frac{[d(v_1)]^2 + \ldots + [d(v_m)]^2}{m}} \leqq \sqrt{\frac{20m_2}{m}}.$$

From these we get the desired inequality.

3

We give a construction which gives more second smallest distances than the regular triangular lattice, but we do not know what is the best construction because what we get is far from the bound proved above. We take the regular hexagonal lattice. On every edge of the hexagons we take two more points so that in every hexagon these new points form a regular 12-gon (see Fig. 4). The set S consists of the centers c_i of these 12-gons and the vertices v_i of the 12-gons. One can easily check that the edge of the 12-gon is the smallest distance in the set. The second

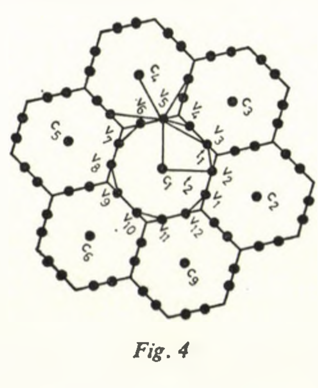


Fig. 4

smallest distance t_2 equals the radius of the circumscribed circle of the 12-gons. For the centers c_i we have $d_2(c_i)=12$, and for the other type of points v_i, $d_2(v_i)=6$, except for those points on the boundary of the figure. So in this construction:

$$m_2 = 3\frac{3}{7}m + o(m).$$

4

Next we give an upper bound for the number of edges of $G_1 \cup G_2$. The following theorem was suggested by P. Erdős:

THEOREM. $m_1 + m_2 \leqq 6m$.

This theorem follows immediately from the next lemma:

LEMMA 3. *For every vertex v, $d_1(v)+d_2(v) \leqq 12$.*

PROOF. Suppose $d_1(v)+d_2(v) \geqq 13$, then there exist two neighbours s_1, s_2 of v in $G_1 \cup G_2$ such that $s_1 v s_2 \sphericalangle < 30°$. Therefore $d(s_1, v) = d(s_2, v) = t_1$ is impossible, from this it follows that $t_2 > \sqrt{3}t_1$. By Lemma 1 we know that $d_1(v) \leqq 6$. If $d_1(v) \geqq 3$ one can easily check that $t_2 \leqq \sqrt{3}t_1$. So this contradicts the above inequality. So we may suppose $d_1(v) \leqq 2$. Then we have to consider the following two cases:

Case (a). $d_2(v) = 12$.

We saw in Lemma 1 that in this case the neighbours u_i of v in G_2 form a regular 12-gon. Obviously, $d(u_i, u_{i+1}) = t_1$. One can easily check that the radius of the circumscribed circle of the triangle vu_1u_2 equals t_1. Say s is the center of this circle (see Fig. 5). So any other point of the circle C in the triangle vu_1u_2 around v with radius t_1 is either from u_1 or from u_2 closer than t_1. So s is the only possibility in the triangle vu_1u_2 for having distance t_1 from v. On the other hand one can easily check

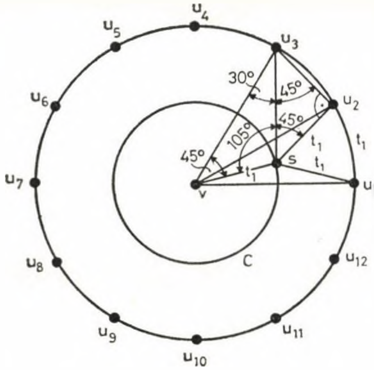


Fig. 5

that $d(u_3, s)=\sqrt{2}t_1$. So $t_1<d(u_3, s)<t_2$, which means that s is forbidden. So if $d_2(v)=12$ then $d_1(v)=0$.

Case (b). $d_2(v)=11$.

(i) We have seen in Case (a) if $d_2(v)=12$, then $d_1(v)=0$. If the 11 neighbours of v in G_2 form a regular 12-gon with one point missing there cannot be point on the circle of radius t_1 around v except inside an arc of angle 60°, but this allows only one point, so $d_1(v)\leqq 1$.

(ii) If the neighbours u_i of v in G_2 do not form a regular 12-gon with one point missing, then the u_i's form a regular 11-gon, by Lemma 2 (b). Let us take the circle C_1 around v of radius t_1. If r is a point on C_1 in the triangle vu_1u_2 then $d(u_1, r)<t_2$ and $d(u_2, r)<t_2$ (see Fig. 6). Furthermore $d(u_1, r)=d(u_2, r)=t_1$ is possible only for the regular 12-gon. So this means that in this case $d_1(v)=0$. So we finished the proof of our statement.

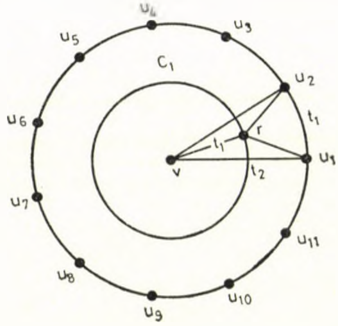


Fig. 6

REMARK. If we take the regular triangular lattice we get for all points v in the set S $d_1(v)+d_2(v)=12$, except for those points on the boundary of the figure.

REFERENCE

[1] HARBORTH, H. (unpublished).

(Received May 20, 1985)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 103—105

ÜBER DAS VERHALTEN EINER KRUMMEN FLÄCHE IN DER NÄHE EINES PARABOLISCHEN PUNKTES

J. STROMMER

L. Fejes Tóth zum 70. Geburtstag

P. Stäckel [1] hat die Frage nach dem Verhalten einer krummen Fläche in der Umgebung eines parabolischen Punktes für *analytische* Flächen beantwortet, indem er (unter Heranziehung des Weierstraßschen Vorbereitungssatzes für reelle Potenzreihen von zwei Veränderlichen) bewiesen hat, daß nur vier Fälle möglich sind, und zwar, je nachdem die Fläche ganz auf einer Seite der Tangentenebene liegt und mit dieser nur einen Punkt bzw. eine ganze Kurve gemein hat oder die Tangentenebene durchsetzt und die Schnittkurve aus zwei sich berührenden Zweigen besteht bzw. eine Spitze (erster oder zweiter Art) hat.

Man hat nun stillschweigend angenommen, daß auch bei beliebigen stetig gekrümmten Flächen dasselbe gilt (vgl. z. B. [2], S. 48). Wir wollen zunächst durch Beispiele zeigen, daß es selbst dann eine unberechtigte Verallgemeinerung ist, wenn die Fläche beliebig oft differenzierbar ist.

Zu dem Zwecke fasse man die Schiebungsfläche

$$z = f(x) - Cy^2 \quad (C \neq 0)$$

ins Auge. Sei $f(x)$ eine im Intervalle $-\infty < x < \infty$ stetige Funktion, welche für alle Werte des Arguments stetige Ableitungen aller Ordnungen besitzt und für $x=0$ gleichzeitig mit ihren Ableitungen erster und zweiter Ordnung verschwindet. Dann ist der Anfangspunkt ein parabolischer Punkt der Fläche, in dem die Tangentenebene die xy-Ebene ist.

1. Sei beispielweise

$$f(x) = \left(\operatorname{arc\,tg}\frac{1}{x} - \operatorname{arc\,ctg} x\right) e^{-(1/x^2)}, \quad x \neq 0; \quad f(0) = 0.$$

Dann hat die Fläche mit der berührenden Ebene des Anfangspunktes eine Kurve gemeinsam, die im Anfangspunkt einen Endpunkt bzw. Verzweigungspunkt hat, je nachdem C positiv oder negativ ist.

2. Sei nun

$$f(x) = e^{-(1/x^2)} \sin\frac{1}{x}, \quad x \neq 0; \quad f(0) = 0.$$

1980 *Mathematics Subject Classification*. Primary 53AXX; Secondary 53A05.
Key words and phrases. Parabolic points.

Dann durchsetzt die Fläche die Tangentenebene des Anfangspunktes in einer Kurve. Diese besteht aus unendlich vielen sich gegenseitig ausschließenden Ovalen, deren einzige Häufungsstelle der Anfangspunkt ist.

3. Sei

$$f(x) = e^{-(1/x^2)}\left(\sin\frac{1}{x}-1\right), \quad x \neq 0; \quad f(0) = 0.$$

Es sind hier zwei Fälle zu unterscheiden, je nachdem C positiv oder negativ ist.

a) Ist C positiv, so liegt die Fläche ganz auf der einen Seite der Tangentenebene und hat mit dieser außer dem Anfangspunkt unendlich viele isolierte Punkte gemeinsam. Der Anfangspunkt selbst ist Häufungspunkt derselben.

b) Ist C negativ, so durchsetzt die Fläche die Tangentenebene in einer Kurve. Diese besteht aus zwei Zweigen, die sich im Anfangspunkte berühren und außerdem in unendlich vielen Punkten schneiden. Der Berührungspunkt ist Häufungspunkt der Schnittpunkte.

4. Sei

$$f(x) = e^{-(1/x^2)}\sin^4\frac{1}{x}, \quad x \neq 0; \quad f(0) = 0.$$

Wenn dann $C>0$ ist, so besteht die Schnittlinie der Fläche mit der Tangentenebene gleichfalls aus zwei Zweigen, die sich außer dem Anfangspunkt in unendlich vielen Punkten berühren. Der Anfangspunkt ist Häufungspunkt der Selbstberührungspunkten.

5. Sei

$$f(x) = \left(\operatorname{arc\,ctg}\sin\frac{1}{x} - \operatorname{arc\,tg}\frac{1}{\sin 1/x}\right) e^{-\frac{x^2+\sin^2 1/x}{x^2\sin^2 1/x}}\sin^3\frac{1}{x}, \quad x \neq 0, \frac{1}{k\pi},$$

$$= 0, \quad x = 0, \frac{1}{k\pi}; \quad k = \pm 1, \pm 2, \ldots.$$

Dann gibt es zwei Fälle, je nachdem C positiv oder negativ ist.

a) Ist C positiv, so liegt die Fläche ganz auf einer Seite der Tangentenebene und hat mit dieser unendlich viele getrennt liegende geradlinige Strecke gemeinsam, deren einzige Häufungsstelle der Anfangspunkt ist.

b) Ist C negativ, so besteht die Schnittkurve der Fläche mit der berührenden Ebene des Anfangspunktes aus zwei sich in diesem Punkte berührenden Zweigen, die unendlich viele getrennte geradlinige Kurvenstücke gemeinsam haben. Der Anfangspunkt ist die einzige Häufungsstelle dieser Stücke.

Es lassen sich leicht verwickeltere Beispiele ähnlicher Art bilden, wovon hier abgesehen werden kann, da schon an Hand der eben betrachteten Beispiele liegt es nahe, daß man keinen alle möglichen Fälle umfassenden Satz über das Verhalten einer nicht-analytischen Fläche in der Umgebung eines parabolischen Punktes aufstellen kann. Man kann lediglich so viel aussagen, daß *der Tangentialschnitt von*

Geraden senkrecht zur Schmiegtangenten in der Nähe des Berührungspunktes höchstens in zwei Punkten getroffen wird.

Der Beweis verläuft so. Durch die Wahl des betreffenden Punktes P als Anfangspunkt und geeigneter Koordinatenachsen x, y, z, kann man leicht bewerkstelligen, daß die Fläche in der Umgebung von P durch eine Funktion

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

dargestellt wird, die für $x=0$ und $y=0$ gleichzeitig mit ihren Ableitungen erster und zweiter Ordnung, von z_{yy} abgesehen, verschwindet. Dann wird die Schmiegtangente in P die x-Achse.

Aus der Voraussetzung, daß der Tangentialschnitt in einer noch so kleinen Umgebung des Anfangspunktes mehr als zwei Punkte besitzt, die auf einer der y-Achse parallelen Geraden liegen, folgt durch zweimalige Anwendung des Rolleschen Satzes, daß der Schnitt parallel der yz-Ebene, dessen Spur in der berührenden Ebene die betreffende Gerade ist, mindestens einen in dieser Umgebung gelegenen Punkt enthält, wofür z_{yy} verschwindet. Damit wird man aber zu einem Widerspruch geführt, denn, der Stetigkeit von z_{yy} wegen, es eine Umgebung von P gibt, in welcher z_{yy} den Wert 0 niemals annehmen darf. Hiermit ist der Beweis erbracht.

LITERATURVERZEICHNIS

[1] STÄCKEL, P., Neue Beiträge zur Flächentheorie, *S.—B. Heidelberger Akad. Wiss. Math-Nat. Kl.* **1916**, 1. Abhandlung.

[2] STRUBECKER, K., *Differentialgeometrie III. Theorie der Flächenkrümmung,* Sammlung Göschen Bd. 1180/1180a, Walter de Gruyter and Co., Berlin, 1959. *MR* **21** #278.

(Eingegangen am 31. Mai 1985)

BUDAPESTI MŰSZAKI EGYETEM
GEOMETRIA TANSZÉKE
MŰEGYETEM RKP. 3—9
H—1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 107—113

ОБ ОДНОЙ ХАРАКТЕРИСТИКЕ РЕШЕТОК И НЕОДНОРОДНОЙ ПРОБЛЕМЕ МИНКОВСКОГО

Н. П. ДОЛБИЛИН

Посвящается Ласло Фейеш Тоту по случаю его 70-летия

1° $\varkappa$-характеристика решетки. Зафиксируем в евклидовом пространстве $\mathbf{R}^n$ ортонормированный базис. Введем обозначения:

Λ — n-мерная решетка; O — начало координат;

$\det\Lambda$ — определитель решетки Λ (объем её фундаментальной области);

$$\mathbf{x}=(x_1, \ldots, x_n)\in\mathbf{R}^n;$$

$\mathbf{x}\equiv\mathbf{x}' \pmod{\Lambda}$ означает, что $\mathbf{x}-\mathbf{x}'\in\Lambda$;

$$\Pi(\mathbf{x}) = \prod_{i=1}^{n}|x_i|;\quad \Pi(\Lambda;\mathbf{x}) = \inf_{\mathbf{x}'\equiv\mathbf{x} \pmod{\Lambda}} \Pi(\mathbf{x}');$$

$$\Pi(\Lambda) = \sup_{\mathbf{x}\in\mathbf{R}^n} \Pi(\Lambda;\mathbf{x}).$$

Под пустым координатным параллелепипедом P понимаем открытый n-мерный параллелепипед, удовлетворяющий трем условиям:

1) центр параллелепипеда P лежит в начале O;

2) его грани параллельны координатным плоскостям;

3) в параллелепипеде P кроме точки O других точек решетки Λ нет: $P\cap\Lambda=\{0\}$.

Для решетки Λ введем положительный функционал $\varkappa(\Lambda)$ посредством соотношения:

$$2^n\varkappa^n(\Lambda)\det\Lambda = \sup_P v(P),$$

где точная верхняя грань берется по пустым координатным параллелепипедам. Отметим, что согласно теореме Минковского о выпуклом теле, для любой решетки Λ имеем $\varkappa(\Lambda)\leqq 1$.

Функционал $\varkappa(\Lambda)$ инвариантен относительно линейных преобразований, сохраняющих систему координатных осей, например, относительно гомотетий, «гиперболических поворотов».

Характеристика $\varkappa(\Lambda)$ была введена в работе [1], где она, обозначенная там через $\delta(\Lambda)$, играла существенную роль при доказательстве теоремы о том, что в каждом классе гиперболически эквивалентных решеток существует

1980 *Mathematics Subject Classification*. Primary 10E15, 10E20, 11H46.

Key words and phrases. Lattices in R^n, inhomogeneous Minkowski problem, Čebotarev's theorem, empty parallelotopes, minimum of product of non-homogeneous linear forms.

решетка с «экономным» покрытием (имеется в виду решетчатое покрытие пространства $\mathbf{R}^n$ равными шарами). Этот факт является n-мерным продолжением направления, заданного методом Ремака—Дайсона—Скубенко доказательства при $n \leqq 5$ гипотезы Минковского (формулировку см. в п. 2°) о произведении неоднородных линейных форм. Дело в том, что важным компонентом этого метода является получение точной верхней оценки для радиуса покрытия, соответствующего наиболее экономной решетке в каждом классе гиперболически эквивалентных решеток.

С гипотезой Минковского связан и другой вопрос — вопрос об индексе «разделенного» параллелепипеда, то есть об отношении его объема к объему основного параллелепипеда решетки Λ. Б. Н. Делоне [2] показал, что при $n=2$ для любой решетки и любой её трансляции найдется разделенный параллелепипед индекса 1, а при $n \geqq 3$ в некоторых случаях минимальный индекс разделенного параллелепипеда больше, чем 1. В работе [1] было показано, что при любом n минимальный для данной решетки индекс разделенного параллелепипеда ограничен константой, зависящей только от n. Игравшая и здесь существенную роль $\varkappa$-характеристика, как мы видим, связана с кругом вопросов, примыкающих к проблеме Минковского. Один из аспектов упомянутой связи будет рассмотрен в п. 3° настоящей статьи.

Отметим, что $\varkappa$-характеристика в некотором смысле двойственна другой характеристике — арифметическому минимуму произведения однородных линейных форм. Действительно, последняя характеристика на языке решеток определяется посредством соотношения

$$2^n \mu(\Lambda) \det \Lambda = \inf_{P} v(\bar{P})$$

где $\bar{P}$ — замкнутый, не пустой (то есть содержаший помимо O другие точки решетки Λ) координатный параллелепипед.

2° Гипотеза Минковского. Гипотеза Минковского в наших обозначениях выглядит следующим образом:

Для любой n-мерной решетки Λ имеет место неравенство

$$\Pi(\Lambda) \leqq \frac{\det \Lambda}{2^n}.$$

Гипотеза подтверждена только для $n \leqq 5$ [3, 4, 5, а также 9]. Для произвольного n имеется классическая оценка Чеботарева [8]:

$$\Pi(\Lambda) \leqq \frac{\det \Lambda}{2^{n/2}}.$$

При больших n оценка Чеботарева последовательно улучшалась во многих работах и наиболее существенно в недавних работах Скубенко Б. Ф. и его учеников.

Обратим внимание на результат Коула [6]: для любого $\mathbf{x}$ найдется такой $\mathbf{x}' \equiv \mathbf{x} \pmod{\Lambda}$, что

$$x_i' > 0, \quad i = 1, \ldots, n-1, \quad \text{и} \quad \Pi(\mathbf{x}') < \frac{\det \Lambda}{2}.$$

Этот результат, наряду с более ранним результатом Чока [7], послужил основанием для следующего обобщения гипотезы Минковского.

Для любого $\mathbf{x}\in\mathbf{R}^n$ и $0\leqq k\leqq n$ найдется такой $\mathbf{x}'\equiv\mathbf{x} \pmod{\Lambda}$, для которого

$$x_i'>0,\quad i=1,\ldots,n-k \quad\text{и}\quad \Pi(\mathbf{x}')\leqq\frac{\det\Lambda}{2^k}.$$

При $k=n$ гипотеза Коула совпадает с гипотезой Минковского. При $k\geqq 3$ гипотеза остается открытой. Представляется правдоподобным усиление в терминах $\varkappa$-характеристики в её оценочной части:

$$\Pi(\mathbf{x}')\leqq\frac{\varkappa^k(\Lambda)}{2^k}\det\Lambda.$$

3° Обобшение теоремы Чеботарева. В том же направлении, в котором гипотеза Коула обобщает гипотезу Минковского, С. С. Рышков предложил автору обобщить оценку Чеботарева. Следующая теорема в основном является реализацией этого предложения.

Теорема 1. *Для любого* $\mathbf{x}_0\in\mathbf{R}^n$, $0\leqq k\leqq n$ *и* $\varepsilon>0$ *найдется* $\mathbf{x}\equiv\mathbf{x}_0 \pmod{\Lambda}$, *для которого* $x_i>0$, $i=1,\ldots,n-k$ *и*

$$\Pi(\mathbf{x})<\frac{\varkappa^k(\Lambda)}{2^{k/2}}\det\Lambda+\varepsilon.$$

Доказательство. Для простоты записи будем полагать $\det\Lambda=1$, что не влияет на общность задачи. Заметим, что результат теоремы при $k=n$ можно сформулировать в следующем виде

$$\Pi(\Lambda)\leqq\frac{\varkappa^n(\Lambda)}{2^{n/2}}\det\Lambda,$$

и он является непосредственным усилением теоремы Чеботарева в терминах $\varkappa$-характеристики ($\varkappa(\Lambda)\leqq 1$ по теореме Минковского о выпуклом теле).

Пусть k — целое, неотрицательное число, $0\leqq k\leqq n$, $\mathbf{x}_0\in\mathbf{R}^n$, $\varepsilon>0$. Покажем, что найдется такая точка $\mathbf{x}\equiv\mathbf{x}_0 \pmod{\Lambda}$, что

(1) $$x_1>0,\ \ldots,\ x_{n-k}>0,$$

(2) $$\Pi(\mathbf{x})\leqq\frac{\varkappa^k(\Lambda)}{2^{k/2}}+\varepsilon.$$

Поскольку среди точек $\Lambda+\mathbf{x}_0$ всегда найдется точка, удовлетворяющая неравенствам (1), то мы можем предполагать, что уже сама точка $\mathbf{x}_0$ удовлетворяет (1).

Обозначим через t такое положительное число, что

$$(\varkappa^k 2^{-k/2}+\varepsilon)t^k=1.$$

Отсюда

(3) $$0<\varkappa t<\sqrt{2}.$$

Пусть для $\mathbf{x}_0$ неравенство (2) не выполняется

$$\Pi(\mathbf{x}_0) > \varkappa^k 2^{-k/2} + \varepsilon = t^{-k}.$$

Параллелепипед, определяемый неравенствами

$$(4)\qquad \begin{aligned} |x_i| &< \varkappa x_{0i}, \quad i = 1, \ldots, n-k, \\ |\varkappa_j| &< \varkappa t |x_{0j}|, \quad j = n-k+1, \ldots, n, \end{aligned}$$

имеет объем

$$v = 2^n \varkappa^n t^k \Pi(\mathbf{x}_0) > 2^n \varkappa^n.$$

По определению $\varkappa$ всегда найдется точка $\mathbf{a} \in \Lambda$, $\mathbf{a} \neq \mathbf{0}$, координаты которой удовлетворяют неравенствам:

$$(5а)\qquad |a_i| < \varkappa x_{0i}, \quad i = 1, \ldots, n-k,$$

$$(5б)\qquad |a_j| < \varkappa t |x_{0j}|, \quad j = n-k+1, \ldots, n.$$

Будем предполагать, что при некотором $J \in \{1, \ldots, n\}$ выполняется либо неравенство

$$(6а)\qquad \frac{1}{2} \varkappa x_{0J} \leqq |a_y| < \varkappa x_{0J},$$

если $1 \leqq J \leqq n-k$, либо неравенство

$$(6б)\qquad \frac{1}{2} \varkappa t |x_{0J}| \leqq |a_J| < \varkappa t |x_{0J}|,$$

если $n-k+1 \leqq J \leqq n$.

Действительно, если для выбранной точки $\mathbf{a}$ это не так, то есть, если ни одна координата не удовлетворяет ни (6а), ни (6б), то найдется натуральное $m > 1$, такое, что с одной стороны, точка $2^m \mathbf{a}$ принадлежит параллелепипеду (4), с другой, — одна из его координат удовлетворяет соответственно либо (6а), либо (6б). И в качестве точки $\mathbf{a}$ можно взять точку $2^m \mathbf{a}$.

Таким образом, предположение о выполнении для $\mathbf{a}$ одного из двух неравенств (6а) или (6б) не ограничивает общность рассуждений.

Обозначим $\mathbf{x}^{\pm} = \mathbf{x}_0 \pm \mathbf{a}$.

В силу (5а) и неравенства $0 < \varkappa \leqq 1$,

$$(7)\qquad x_i^{\pm} = x_{0i} \pm a_i > 0, \quad i = 1, \ldots, n-k,$$

то есть первые $n-k$ координат каждой из точек $\mathbf{x}^+$ и $\mathbf{x}^-$ удовлетворяют неравенствам (1).

Далее

$$(8)\qquad \frac{\prod\limits_{i=1}^{n} x_i^+ \prod\limits_{i=1}^{n} x_i^-}{\prod\limits_{i=1}^{n} x_{0i}^2} = \prod_{i=1}^{n} \left(1 - \frac{a_i^2}{x_{0i}^2}\right).$$

При любом $i \in \{1, \ldots, n-k\}$

$$\left|\frac{a_i}{x_{0i}}\right| \leqq \varkappa \leqq 1,$$

откуда

(9)
$$0 < 1-\varkappa^2 \leqq 1-\frac{a_i^2}{x_{0i}^2} \leqq 1,$$

и в случае (6a)

(10)
$$0 < 1-\varkappa^2 \leqq 1-\frac{a_j^2}{x_{0j}^2} \leqq 1-\left(\frac{\varkappa}{2}\right)^2.$$

При любом $j \in \{n-k+1, \ldots, n\}$

(11)
$$\left|\frac{a_j}{x_{0j}}\right| < \varkappa t,$$

откуда

(12)
$$-1 < 1-(\varkappa t)^2 \leqq 1-\frac{a_j^2}{x_{0j}^2} \leqq 1,$$

и в случае (6б)

(13)
$$-1 < 1-(\varkappa t)^2 < 1-\frac{a_j^2}{x_{0j}^2} \leqq 1-\left(\frac{\varkappa t}{2}\right)^2.$$

Из неравенств (9)—(13), а также из (7) следует

$$\Pi(\mathbf{x}^+)\Pi(\mathbf{x}^-) = \prod_{i=1}^{n}\left|1-\frac{a_i^2}{x_{0i}^2}\right| \prod_{i=1}^{n} x_{0i}^2 \leqq s^2 \prod_{i=1}^{n} x_{0i}^2$$

где

$$s^2 = \max\left\{1-\varkappa^2, |1-(\varkappa t)^2|, 1-\left(\frac{\varkappa t}{2}\right)^2\right\}.$$

Для данной решетки Λ значение $\varkappa$ фиксировано,

$$0 < (\varkappa t)^2 < 2, \qquad 0 < \varkappa \leqq 1,$$

s^2 зависит только от $\varkappa$ и t. Поэтому

$$s^2 < 1-\delta(\varkappa, t)$$

где $\delta(\varkappa, t)$ — некоторая положительная константа.

Обозначим через $\mathbf{x}_1$ ту из двух точек $\mathbf{x}^+$ и $\mathbf{x}^-$, для которой

$$\Pi(\mathbf{x}_1) = \min\{\Pi(\mathbf{x}^+), \Pi(\mathbf{x}^-)\}.$$

Ясно, что $\mathbf{x}_1 \equiv \mathbf{x}_0 \pmod{\Lambda}$ и

$$\Pi(\mathbf{x}_1) < |s|\,\Pi(\mathbf{x}_0).$$

Если $\Pi(\mathbf{x}_1) > t^{-n}$, то точно таким же образом можно получить точку

$$\mathbf{x}_2 \equiv \mathbf{x}_0 \pmod{\Lambda},$$

для которой $x_{21}>0, \ldots, x_{2,n-k}>0$ и

$$\Pi(\mathbf{x}_2) < |s|\,\Pi(\mathbf{x}_1).$$

Ясно, что для данных $\mathbf{x}_0$ и t через конечное число шагов получим точку $\mathbf{x} \equiv \mathbf{x}_0 \pmod{\Lambda}$ для которой

$$x_1 > 0, \ldots, x_{n-k} > 0, \quad \Pi(\mathbf{x}) \leqq t^{-k} = \frac{\varkappa^k(\Lambda)}{2^{k/2}} + \varepsilon.$$

Теорема доказана.

4° Случай $\varkappa(\Lambda)=1$

Теорема 2. *Пусть* $\varkappa(\Lambda)=1$, *тогда*

$$\Pi(\Lambda) \leqq \frac{\det \Lambda}{2^n}.$$

Доказательство. Без ограничения общности будем предполагать, что $\det \Lambda = 1$.

Если значение $\varkappa(\Lambda)=1$ достигается на некотором пустом координатном параллелепипеде P, то параллелепипед $\frac{1}{2}\bar{P} = \{\mathbf{x} \in \mathbf{R}^n\colon 2\mathbf{x} \in \bar{P}\}$ содержит фундаментальную область и для любой точки $\mathbf{x} \in \frac{1}{2}\bar{P}$ выполняется

$$\Pi(\mathbf{x}) \leqq \frac{v\left(\frac{1}{2}\bar{P}\right)}{2^n} = \frac{1}{2^n}.$$

Менее тривиален случай, когда значение $\varkappa(\Lambda)=1$ не достигается ни на одном пустом координатном параллелепипеде. Тогда существует последовательность таких параллелепипедов $\{P_i\}$, что $\lim v(P_i) = 2^n$.

Пусть $\varepsilon>0$. Покажем для произвольного $\mathbf{x} \in \mathbf{R}^n$ найдется $\mathbf{x}' \equiv \mathbf{x} \pmod{\Lambda}$ такой, что $\Pi(\mathbf{x}') \leqq \left(\frac{1+\varepsilon}{2}\right)^n$.

Выберем число γ так, чтобы $0<\gamma<1$, $\delta(\gamma)<\gamma$ и $\gamma+\delta(\gamma)<1+\varepsilon$, где $\delta(\gamma) = \sqrt[n]{1-\gamma^n}$.

Из $\varkappa(\Lambda)=1$ следует, что найдется пустой координатный параллелепипед P' такой, что $v(P)=\gamma^n$ и его замыкание $\bar{P}'$ не содержит эквивалентных относительно Λ точек.

Пусть $Q = \frac{\delta(\gamma)}{\gamma}\bar{P}'$. Так как $\frac{\delta(\gamma)}{\gamma}<1$, то Q также не содержит эквивалентных точек.

Рассмотрим $U_{\mathbf{x}} = \bar{P}' \cup (Q+\mathbf{x})$, где $Q+\mathbf{x}$ означает трансляцию параллелепипеда Q на вектор $\mathbf{x}$.

Если $\bar{P}' \cap (Q+\mathbf{x}) \neq \emptyset$, то

$$\Pi(\mathbf{x}) < \left(\frac{\gamma+\delta}{2}\right)^n < \left(\frac{1+\varepsilon}{2}\right)^n.$$

Если же $\bar{P}'\cap(Q+\mathbf{x})=\emptyset$, то $v(U_{\mathbf{x}})=1$. Тогда $U_{\mathbf{x}}$ содержит пару эквивалентных точек $\mathbf{y}\equiv\mathbf{y}' \pmod{\Lambda}$, причем $\mathbf{y}'\in\bar{P}'$ и $\mathbf{y}'\in Q+\mathbf{x}$. Отсюда $P'\cap(Q+\mathbf{x}')\neq\emptyset$, где $\mathbf{x}'=\mathbf{x}+\mathbf{y}-\mathbf{y}'$. Поэтому

$$\Pi(\mathbf{x}')<\left(\frac{1+\varepsilon}{2}\right)^n$$

где $\mathbf{x}'\equiv\mathbf{x} \pmod{\Lambda}$.

Теорема 2 доказана.

Из теорем 1 и 2 вытекает, что гипотеза Минковского справедлива для решеток с $\varkappa(\Lambda)\leqq\frac{1}{\sqrt{2}}$ и $\varkappa(\Lambda)=1$.

ЛИТЕРАТУРА

[1] Долбилин, Н. П., О радиусе покрытия гиперболически эквивалентных решеток, *Dokl. Akad. Nauk SSSR* **214** (1974), № 5, 1002—1004. *MR* **49** # 225.

[2] Делоне, Б. Н., Алгоритм разделенных параллелограммов решетки, *Izvestiya Akad. Nauk SSSR Ser. Mat.* **11** (1947), 505—538. *MR* **9**—334.

[3] Remak, R., Verallgemeinerung eines Minkowskischen Satzes 1, 2, *Math. Z.* **17** (1923), 1—34; **18** (1924), 173—200.

[4] Dyson, F. J., On the product of four non-homogeneous linear forms, *Ann. of Math.* (2) **49** (1948), 82—109. *MR* **10**—19.

[5] Скубенко, Б. Ф., Доказательство гипотезы Минковского о произведении n линейных неоднородных форм с n переменными при $n\leqq 5$, *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **33** (1973), 6—36. *MR* **51** # 3061.

[6] Cole, A. J., On the product of n linear forms, *Quart. J. Math., Oxford Ser.* (2) **3** (1952), 56—62. *MR* **13**—726.

[7] Chalk, J. H. H., On the positive values of linear forms, *Quart. J. Math., Oxford Ser.* **18** (1947), 215—227. *MR* **9**—413. II. *Ibid.* **19** (1948), 67—80. *MR* **10**—18.

[8] Чеботарев, Н. Г., Заметки по алгебре и теории чисел, *Уч. зап. Казанск. Ун-та* **94** (1934), 3—16; см. также: Доказательство теоремы Минковского о неоднородных линейных формах, *Собрание сочинений*, Т.2, Izd. Akad. Nauk, Moscow—Leningrad, 1949, 361—364. *MR* **11**—572, 872.

[9] Касселс, Дж. (Cassels, J. W. S.), *Введение в геометрию чисел* (An introduction to the geometry of numbers), Мир, Москва, 1965. *MR* **31** # 5841.

(Поступила 14-ого июня 1985 г.)

STEKLOV INSTITUTE OF THE
ACADEMY OF SCIENCES
VAVILOVA, 42
SU—117 966 MOSCOW
SOVIET UNION

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 115—121

СТРОЕНИЕ КОНЕЧНЫХ ГРАНЕЙ ПОЛИЭДРА $\mu_n(m)$ ПРИ $n \leqq 4$

С. С. РЫШКОВ и М. X. УМАРОВ

Посвящается Ласло Фейеш Тоту по случаю его 70-летия

1. В геометрии положительных квадратичных форм (ПКФ) существенную роль играют два взаимных полиэдра: полиэдр Вороного $\Pi(n)$ и полиэдр $\mu_n(m)$ (см. [1, 2, 3, 4]). В то время как полиэдр $\Pi(n)$ изучен очень подробно до $n=5$, полиэдр $\mu_n(m)$ только начинает изучаться. Напомним, что полиэдр $\mu_n(m)$ это такая поверхность в пространстве коэффициентов квадратичных форм от n переменных, точки которой отвечают ПКФ с арифметическим минимумом равным m. Напомним также, что вершины полиэдра $\mu_n(m)$ эти совершенные формы, т. е. только среди вершин полиэдра $\mu_n(m)$ при любом n содержатся ПКФ, определяющие решетки, которые дают локально плотнейшие упаковки n мерных шаров.

Цель этой статьи — дать рисунки всех конечных граней полиэдра $\mu_n(m)$ при $n \leqq 3$ и всех не более чем четырехмерных граней полиэдра $\mu_4(m)$.

Отметим, что результаты для $n=2$ и $n=3$ были известны первому автору ещё во времени написания работы [3]. А вопросы создания рисунков при $n=4$ обсуждались З. Д. Ломакиной и Майклом Коном (Австралия) во время пребывания последнего на стажировке в Москве. Сейчас исследование было проведено на совершенно новой основе обоими авторами настоящей статьи совместно.

2. Мы рассматриваем квадратичную форму вида $f=f(\mathbf{x})=f(x_1, x_2, \ldots, x_n)= =\sum_{i<j} a_{ij}x_ix_j$, где $a_{ij}=a_{ji}$. Таким образом, пространство E^N коэффициентов квадратичной формы имеет размерность $N=n(n+1)/2$. В пространстве E^N действует «группа эквивалентности», т. е. группа аффинных преобразований E^N, порождённых целочисленными унимодулярными подстановками переменных $x_1, x_2, \ldots, x_n$.

ПКФ заполняет в пространстве E^N выпуклый открытый конус $\mathbf{K}$, заданный неравенствами Сильвестра. Все полиэдры $\mu_n(m)$ принадлежат конусу $\mathbf{K}$ и гомотетичны друг другу с центром гомотетии в начале координат (в вершине конуса $\mathbf{K}$). Поэтому в дальнейших рассуждениях ограничимся полиэдром $\mu_n(1)$.

Из результатов работы [5] следует, что каждая конечная грань полиэдра $\mu_n(m)$ есть грань некоторой максимальной конечной его грани. Таких максимальных конечных граней у полиэдра $\mu_n(1)$ при каждом $n \geqq 2$ с точностью

1980 *Mathematics Subject Classification*. Primary 11H55.

Key words and phrases. Positive definite quadratic forms, Voronoi polyhedron, finite faces of the polyhedron $\mu_n(m)$.

до эквивалентности лишь конечное число, все они имеют размерность $N-n$. При $n \leqq 4$ такой класс эквивалентности лишь один. Одна из максимальных конечных граней этого класса эквивалентности задается равенствами $a_{11}= =a_{22}=a_{33}=a_{44}=1$ и неравенствами $f(m_1, m_2, \ldots, m_n) \geqq 1$, где числа m_i берутся из табл. 4 [3]; (здесь нужная её часть приведена в виде табл. 1) для каждого n берутся строки, длиной не превосходящие n, в более коротких строках недостающие значения m_i полагаются равными нулю.

Таблица 1

$\pm m_i$ I	$\pm m_i$ II	$\pm m_i$ III	$\pm m_i$ IV
1	1		
1	1	1	
1	1	1	1

3. Случаи $n=2$ и $n=3$. При $n=2$ пространство E^N трехмерно, а поверхность $\mu_2(1)$ двумерна.

Проекция полиэдра $\mu_2(1)$ из вершины конуса **K** на плоскость перпендикулярную оси конуса **K** изображена на рис. 1 (жирные линии). На том же рисунке тонкими штриховыми линиями изображена аналогичная проекция полиэдра Вороного $\Pi(2)$, которому полиэдр $\mu_2(1)$ дуален (см. [6], стр. 188).

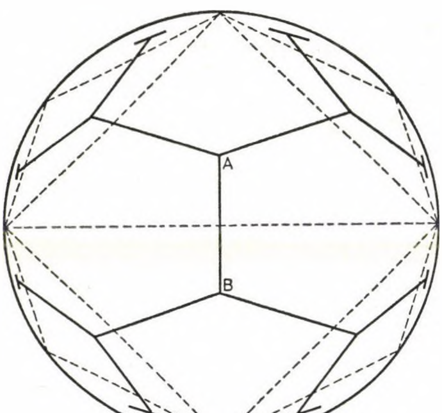


Рис. 1

Максимальные конечные грани здесь одномерны. Одной из них является ребро AB: u^2+v^2-uv, u^2+v^2+uv.

Максимальная конечная грань полиэдра $\mu_3(1)$, которую мы обозначим через Φ^3, ограничена на плоскости $a_{11}=a_{22}=a_{33}=1$, как это следует из вышесказанного, неравенствами:

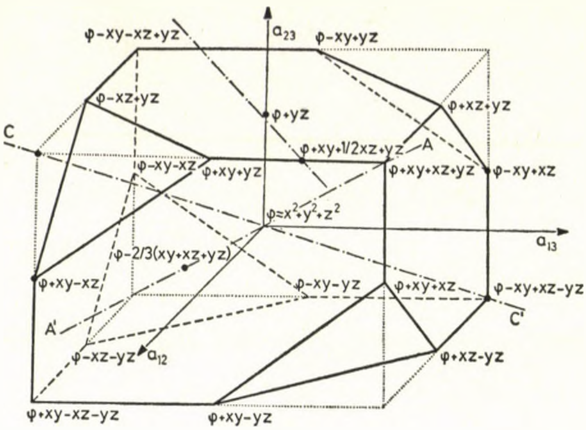


Рис. 2

$$f(1,\ 1, 0) \geqq 1, \qquad f(1, 0,\ 1) \geqq 1, \qquad f(0,\ 1,\ 1) \geqq 1,$$

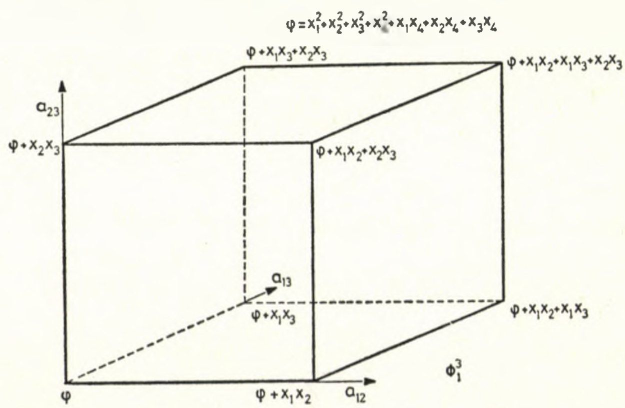
$$f(1,\ -1, 0) \geqq 1, \quad f(1, 0,\ -1) \geqq 1, \quad f(0,\ 1,\ -1) \geqq 1,$$

$$f(1,\ 1,\ 1) \geqq 1, \quad f(-1, 1,\ 1) \geqq 1, \quad f(1,\ -1,\ 1) \geqq 1, \quad f(1, 1,\ -1) \geqq 1.$$

Десятигранник Φ^3 в координатной системе a_{12}, a_{13}, a_{23} — это куб (заданный первыми шестью неравенствами) с отрезанными через одну вершинами (последние четыре неравенства).

Многогранник изображен на рис. 2. Тот же рисунок можно было получить другим способом, а именно, выписав из работы [7] все ПКФ соответствующие вершинам многогранника Φ^3 и построив их на чертеже.

4. Случай $n=4$. Здесь размерность максимальной конечной грани, которую мы обозначим через Φ^6, равна 6. Из дуальности полиэдра $\Pi(4)$ и $\mu_4(1)$



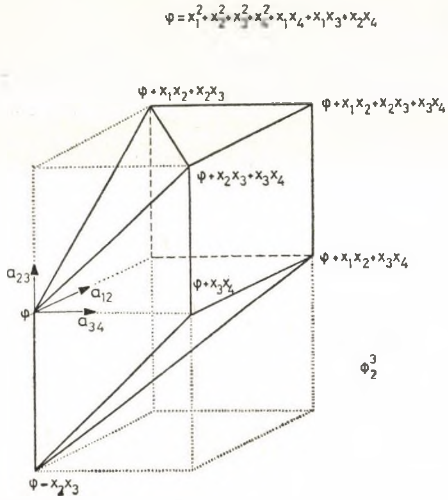
Φ_1^3: $f(1,\ 0,\ 0,\ -1) = 1$; $\quad f(0, 1,\ 0,\ -1) = 1$; $\quad f(0, 0,\ 1,\ -1) = 1$;

Рис. 3

следует, что она имеет 17 попарно неэквивалентных граней. Таких граней в размерности 0, 1, 2, 3, 4, 5 соответственно 2, 2, 2, 4, 4, 3 и мы их обозначаем соответственно через

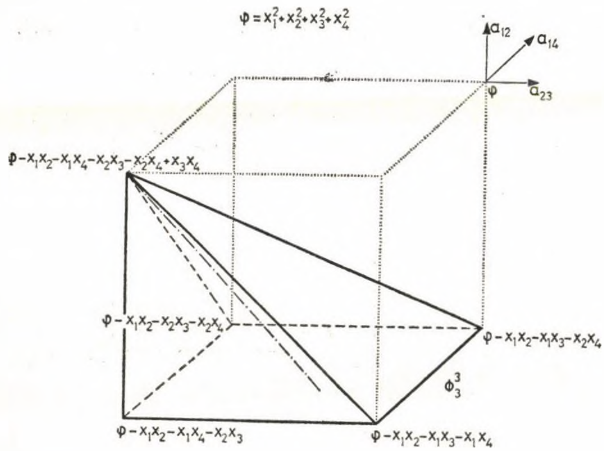
(1) $\Phi_1^0, \Phi_2^0, \Phi_1^1, \Phi_2^1, \Phi_1^2, \Phi_2^2, \Phi_1^3, \Phi_2^3, \Phi_3^3, \Phi_4^3, \Phi_1^4, \Phi_3^4, \Phi_4^4, \Phi_5^4, \Phi_1^5, \Phi_4^5, \Phi_5^5, \Phi^6$

(включив в список восемнадцатую грань саму Φ^6). Здесь верхний индекс — размерность, а нижний индекс выбран так, чтобы Φ_k^r была дуальна грани F_k^{9-r} полиэдра $\Pi(4)$ (см. [5]).



Φ_2^3: $f(1,\ 0,\ -1,\ 0) = 1$; $f(1, 0,\ 0,\ -1) = 1$; $f(0, 1,\ 0,\ -1) = 1$;

Рис. 4

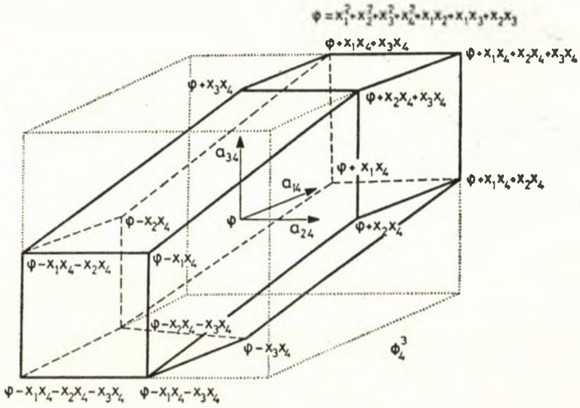


Φ_3^3: $f(1,\ 1,\ 1,\ 1) = 1$; $f(1, 1,\ 0,\ 1) = 1$; $f(1, 1,\ 1,\ 0) = 1$;

Рис. 5

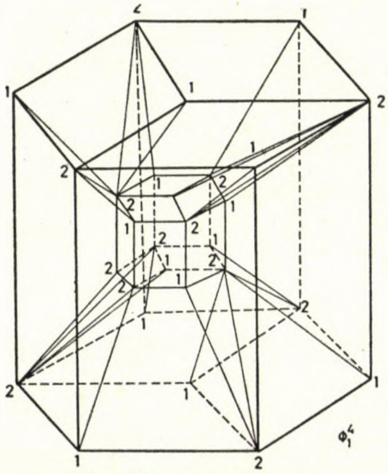
Так же из дуальности указанных полиэдров следует, что во всех гранях многогранника Φ^6 (кроме вершин эквивалентных ПКФ φ_1) схемы схождения граней более высоких размерностей такие же как у шестимерного симплекса.

Пользуясь теми же методами, что при изображении грани Φ^3, мы даем здесь чертежи всех трехмерных грани полиэдра $\mu_4(1)$ и эскизы центральных проекций четырехмерных граней. Рисунки для пятимерных и шестимерных граней трудно обозримы, и, следовательно, мало информативно, поэтому мы их не приводим.



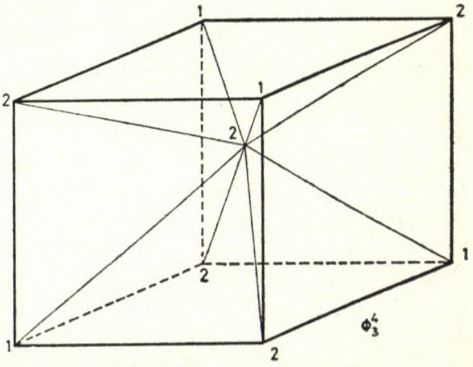
Φ_4^3: $f(1,\ -1,\ 0,\ 0) = 1$; $f(1, 0, -1,\ 0) = 1$; $f(0, 1, -1,\ 0) = 1$;

Рис. 6



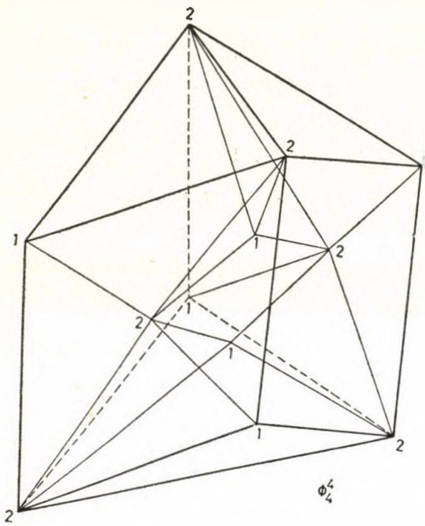
Φ_1^4: $f(1,\ 1,\ 0,\ 0) = 1$;
$f(0,\ 1,\ 1,\ 0) = 1$;

Рис. 7



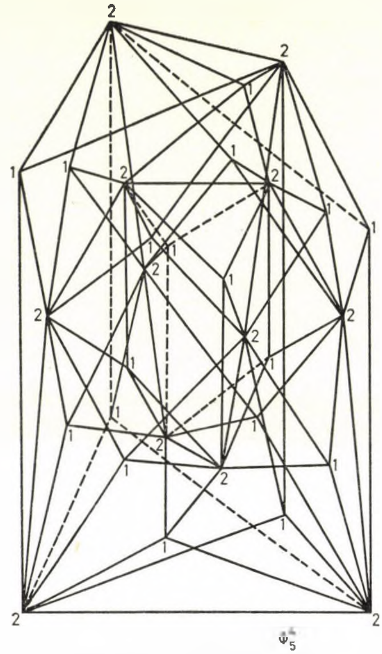
Φ_3^4: $f(1,\ 1,\ 1,\ 0) = 1$;
$f(0,\ 1,\ 1,\ 1) = 1$;

Рис. 8



Φ_4^4: $f(1,\ 1,\ 1,\ 1) = 1;$
$f(0,\ 1,\ 1,\ 1) = 1;$

Рис. 9



Φ_5^4: $f(1,\ 1,\ 0,\ 0) = 1;$
$f(0,\ 0,\ 1,\ 1) = 1;$

Рис. 10

На рисунках изображены трехмерные и четырехмерные грани, которые определяются несущими плоскостями, заданными системами уравнений $a_{11}=$ $=a_{22}=a_{33}=a_{44}=1$ и, соответственно.

Подчеркнем, что грани Φ_1^0 и Φ_2^0 это вершины, т. е. совершенные формы соответственно эквивалентные ПКФ

$$\varphi_0 = x_1^2+x_2^2+x_3^2+x_4^2+x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4, \quad \varphi_1 = \varphi_0-x_1x_2.$$

На рис. 3—6 ПКФ, отвечающие вершинам именно при указанных уравнениях выписаны полностью. На рис. 7—10 указан только тип вершины: 1 или 2, соответственно.

Кроме рисунков мы даем таблицу 2, показывающую для каждой грани (1) сколько граней какого вида в ней содержится. В таблице 2 занумерованы числами от 1 до 18, что соответствует номерам граней в приведенной последовательности (1). Столбцы занумерованы числами 3—18, которые имеют тот же смысл. На пересечении i-ой строки и j-го столбца указано сколько граней эквивалентны i-ой грани содержится в j-ой грани.

Авторы рады посвятить эту работу академику Ласло Фейеш Тоту к его семидесятилетию.

Таблица 2

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	0	1	2	4	4	2	8	16	4	6	18	60	12	24	200
2	1	2	2	2	4	4	3	6	12	5	6	12	36	10	18	104
3			2	4	12	12	6	24	64	16	24	72	300	60	120	1200
4			1	0	0	2	2	0	8	4	6	12	48	15	24	216
5					0	4	4	0	24	12	18	36	192	60	96	1080
6					6	4	1	12	36	6	9	36	204	30	72	960
7									2	1	0	0	12	0	6	80
8									8	0	3	18	84	15	24	504
9									4	6	6	0	60	30	48	576
10									2	0	0	0	12	0	3	64
11													12	0	3	96
12													0	0	6	48
13													12	10	8	176
14													2	0	0	0
15																12
16																8
17																16

ЛИТЕРАТУРА

[1] Делоне, Б. Н., Геометрия положительных квадратичных форм, *Uspehi Mat. Nauk* **3** (1937), 16—62; **4** (1938), 102—164.

[2] Рышков, С. С. и Барановский, Е. П., Классические методы теории решетчатых упаковок, *Uspehi Mat. Nauk* **34** (1979), no. 4, 3—63, 256. *MR* **81a:** 10045

[3] Рышков, С. С., Полиэдр $\mu(m)$ и некоторые экстремальные задачи геометрии чисел, *Dokl. Akad. Nauk SSSR* **194** (1970), 514—517. *MR* **43**#2613

[4] Делоне, Б. Н. и Рышков, С. С., Экстремальные задачи теории положительных квадратичных форм, *Trudy Mat. Inst. Steklov.* **112** (1971), 203—223, 387. *MR* **49**#4939

[5] Рышков, С. С., К проблеме отыскания соверщенных квадратичных форм от многих переменных, *Trudy Mat. Inst. Steklov.* **142** (1976), 215—239. *MR* **58**#27807

[6] Рышков, С. С., Максимальные конечные группы целочисленных $n\times n$-матриц и полные группы целочисленных автоморфизмов квадратичных форм (типы Бравэ), *Trudy Mat. Inst. Steklov.* **128** (1972), 183—211. *MR* **49**#8939

[7] Рышков, С. С., Кон, М. и Ломакина, З. Д., Вершины симметризованной области Минковского при $n\leqq 5$, *Trudy Mat. Inst. Steklov.* **152** (1980), 195—203. *MR* **82i**: 10040.

(Поступило 14-ого июня 1985 г.)

МАТЕМАТИЧЕСКИЙ ИНСТИТУТ
ИМ. В. А. СТЕКЛОВА
АКАДЕМИИ НАУК СССР
УЛ. ВАВИЛОВА 42
SU—117 966 МОСКВА
SOVIET UNION

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 123—137

ISOPERIMETRIC DIVISION INTO A FINITE NUMBER OF CELLS IN THE PLANE

MICHAEL N. BLEICHER

To Professor L. Fejes Tóth on his 70th birthday

I. History and overview

In its simplest form the Isoperimetric Theorem can be stated as follows: Among all plane closed curves of unit length the circle bounds the most area.

This fact has been known since antiquity and was applied by Dido, Princess of Tyre, when, legend has it, she founded Carthage in 814 B. C. She was told, according to the legend, that she could have as much land as she could surround with an ox-hide, whereupon she cut it into thin strips and laid it out in a large circular path. We do not know how she used the coastline as part of the boundary.

The earliest known mathematical work on the subject is attributed to Zenodorus (c. 150 B. C.) who published a book on isoperimetric problems. We know only fourteen of his theorems as reported by Theon and Pappus [9, 13, 23, 29, 36]. There were some difficulties in these early proofs [9, 31] especially regarding the existence of a solution.

Already in Zenodorus' work the simple theorem above had been generalized to results where the choice of optimizing curves had been reduced to special categories, e.g. triangles, polygons of n or fewer sides, etc. For theorems of this type see [6, 19, 22, 30, 31, 37, 38]. As time went by other generalizations of this theorem were considered.

These generalizations were still of the type which tried to bound one region, but constraints were put on the boundary, for example a part could come from a given fixed curve, the flexible part was made of straight line segments of preassigned length, etc., or the boundary was required to pass through certain points.

Other types of generalization try to give some measure of how much more efficient the circle is in terms of other geometric parameters, e.g. the ratio or difference of the inradius and circumradius of the figure. Of course, all these variations have been generalized to higher dimensions, non-Euclidean Geometry, and generally to surfaces [1, 3, 4, 5, 8, 11, 12, 14, 21, 24, 25, 26, 27, 32—36].

L. Fejes Tóth has considered the problem of packing a number of convex cells of fixed perimeter into a fixed region in such a way as to maximize the total area covered by the aggregate of cells, or, dually, a number of cells of fixed area are placed in a given region in such a way as to minimize the total perimeter [16, § 26].

Very little seems to be known about the shape of the cells if one wishes to con-

---

1980 *Mathematics Subject Classification*. Primary 52A45, 52A40.
Key words and phrases. Problem of Dido, isoperimetry, packing and covering.

struct in the plane a finite number of cells, each cell with its own preassigned area, with the minimal total boundary length.

This is true even though this knowledge might have interesting consequences in such diverse fields as physics, geology and biology [10, 15, 39, 40, 43 esp. § 2.3.2].

In fact an analogous problem of dividing part of space into cells of equal volume in the most efficient way was considered from the earliest times. See, for example, the introduction to the subject by Pappus on the habits of bees in constructing beehives. However, Coolidge makes a case for spiders being better geometers than bees [13]. Coolidge's position was perhaps strengthened by Fejes Tóth's work [20, cf. 7] showing a more efficient construction of a hive, more complicated than a beehive. However, there is considerable doubt if optimization is the motivation of the bees [2, Chap. 8].

It can be shown that no division into cells of equal area can be as efficient as the cells in a hexagonal tiling of the entire plane, but approaches the efficiency of a hexagonal tiling asymptotically as the number of cells gets large [17, 18]. The proof, however, says very little about the actual shape of the cells.

It is known from an unpublished result of A. Heppes that if one has a froth in three dimensions, i.e. a collection of cells in three dimensions where there are cells surrounded by several layers of other cells, then the configuration which minimizes the total surface area must contain a non-convex cell.

The only work known to the author which investigates the shape of the cells obtained when one wishes to divide a given region into a given number of equal area cells using the least total arc-length is the monograph by Tomonaga [41]. Upon reading this monograph, it seemed that several of these results could be improved by using some "well-known" theorems on isoperimetry. To the author's great surprise he was unable to find a reference to these results. After several years of querying colleagues and searching the literature, turned up, at best, vague references to these results, it becomes clear that even if these results are to be found in the literature, a more comprehensive statement and proof in a more accessible place are desirable.

Section II of this paper states some known preliminary results which will be used.

Section III contains the proof of the main result, namely

THEOREM. *If a network A of arcs of total arc length L is situated such that the complement of A has n finite components $n \geqq 1$ of given area ratio $a_1:a_2:\ldots:a_n$ in such a way that the total area is maximized, then:*

(1) *The arcs are all segments of circles or lines.*
(2) *Three arcs meet at each node.*
(3) *The angles at each node are $2\pi/3$.*
(4) *The number of nodes is $2(n-1)$ and the number of arcs is $3(n-1)$.*
(5) *The sum of the oriented curvature of the three arcs meeting at each node is zero.*
(6) *The network is connected.*

By definition, the oriented curvature is 0 for a line segment, $\frac{+1}{R}$ for a circle of radius R such that the inside of the circle lies to the right-hand side leaving the node and is $\frac{-1}{R}$ if the inside is to the left.

Section IV contains examples, applications and conjectures, and Section V, the bibliography.

II. Known results

The first theorem we use is a general version of the Blaschke Selection Theorem [16, 22, 28], namely: *Every infinite collection of uniformly bounded compact sets contains a convergent infinite sequence which converges to a non-empty compact subset.*

Also, convex sets converge to a convex set, Lebesgue measure is continuous, and the perimeter (hypersurface area), of the limit is at most the limit of the perimeter (hypersurface area) with equality if none of the perimeter coalesces in the limit. With these theorems it follows by routine arguments that there are solutions to the specific problems considered in this paper. Next, we shall need the classical isoperimetric theorem, stated in section 1.

We shall also use the theorem which G. Pólya calls the stick and string [31, p. 183]:

Of all curves of length L joining the end points of a segment, the unique curve which bounds, with the segment, the most area is a circular arc.

III. Proof of the main result

If $n=1$ the network is a circle by the classical isoperimetric theorem, which clearly satisfies the conditions of the theorem.

We note that any pair of the n bounded regions can have at most $n-1$ distinct components to their intersection, since there are exactly n bounded regions in the complement of the network.

PROOF of (1). Let $\overset{\frown}{AB}$ be an arc which is a connected component of the boundary between two regions, R_1 and R_2 and which has no points on the boundary of other regions. If the arc $\overset{\frown}{AB}$ is not circular then there is a point P, strictly between A and B where no subarc of $\overset{\frown}{AB}$ containing P is a circular arc. Since P is on the boundary of no regions other than R_1 and R_2, there is a circular neighborhood N centered at P contained completely in $R_1 \cup R_2$. Let N' be a circular neighborhood of P with one third the radius of N. Let A' be on $\overset{\frown}{AP} \cap N'$ and B' be on $\overset{\frown}{PB} \cap N'$ with $A' \neq P$ and $B' \neq P$, and such that $\overset{\frown}{A'P} \subseteq N'$ and $\overset{\frown}{PB'} \subseteq N'$. The line segment $\overline{A'B'}$ is within $N' \subseteq R_1 \cup R_2$. The subarc $\overset{\frown}{A'B'}$ of $\overset{\frown}{AB}$ is not circular. Let a be the signed area bounded by the arc $\overset{\frown}{A'B'}$ and the segment $\overline{A'B'}$ with the area in R_1 positive and the area in R_2 negative. Without loss of generality $a \geqq 0$. And since the arc and segment all lie within N', $a \leqq \frac{\text{area}\,(N)}{9}$. Replace $\widehat{A'B'}$ by the shortest arc which bounds, together with $\overline{A'B'}$, an area a on the side of R_1. By the upper bound on a this arc is inside N hence inside $R_1 \cup R_2$. Thus this change effects only the area of R_1 and R_2. Thus all the regions have the same area as before, but the replacement arc is shorter (by the stick and string theorem). Thus the total length of the system is less than L, and a dilation of ratio $\lambda > 1$ is possible enlarging the configuration to bring the length back to L. This dilation increases the areas of all the regions by the factor λ^2 so that all ratios are unchanged, which violates the maximality of the area. Thus, in a maximal system the arcs are circular. This proves (1).

We next prove (3), which immediately implies (2).

PROOF of (3). Since a node always has at least three arcs, it suffices to show that the angle between two arcs cannot be less than $2\pi/3$. Suppose at a node A two circular arcs AB and AC meet at an angle of less than $2\pi/3$.
The proof is in four steps.

Step (i). If the circular arcs are line segments, then by moving the node slightly we can obtain a saving of length in the arcs used. Given any sufficiently small $\varepsilon > 0$ we can do this so that the length saved is ε. In this process we change the areas of the regions involved by an amount of order ε^2.

Step (ii). The errors introduced by using line segment instead of circular arc will not change the facts of Step (i), although the implied constant may change in the area changes of order ε^2.

Step (iii). Using length less than $\frac{\varepsilon}{2}$ the area changes can be corrected to preserve the given area ratios.

Step (iv). A dilation of ratio greater than 1 can be applied to bring the total length back to L. This dilation increases the total area while preserving the ratios; since if the dilation is of ratio $\lambda > 1$ and the original areas are A_i, $i = 1, 2, \ldots, n$, then the new areas are $\lambda^2 A_i$, which have the same ratio.

Proof of Step (i). Consider the point N with three or more rays emanating from it, two of which form an angle $\alpha < 2\pi/3$. Pick points A and B on the sides of the angle with $\overline{NA} = \overline{NB}$ (see Figure 1).

Choose P on the bisector of $\sphericalangle ANB$ so that $\sphericalangle APB = 2\pi/3$. Let E, F be the intersections of the lines through P perpendicular to $\overline{AP}$ and $\overline{BP}$, respectively with $\overline{NA}$ and $\overline{NB}$, respectively. We must show that

$$(*) \qquad |\overline{NA}| + |\overline{NB}| > |\overline{AP}| + |\overline{BP}| + |\overline{NP}|.$$

Since $\overline{AE}$ is the hypotenuse of $\triangle APE$,

$$(1) \qquad |\overline{AE}| > |\overline{AP}|.$$

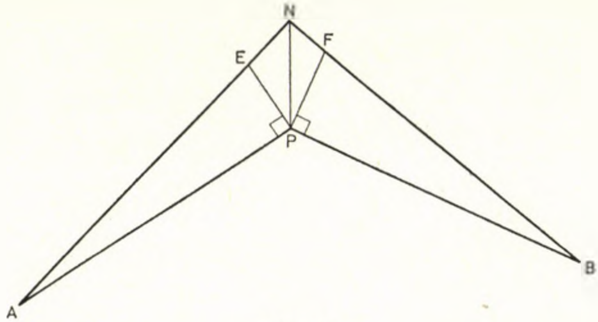


Fig. 1

Similarly

$$|\overline{BF}| > |\overline{BP}|. \tag{2}$$

Now $\sphericalangle ENP < \pi/3$ and $\sphericalangle EPN = \pi/6$, so $\sphericalangle PEN = \beta > \pi/2$. Also $\dfrac{|\overline{EN}|}{\sin \pi/6} = \dfrac{|\overline{PN}|}{\sin \beta}$, thus $|\overline{EN}| = |\overline{PN}| \dfrac{\sin \pi/6}{\sin \beta} > \dfrac{|\overline{PN}|}{2}$. Since $|\overline{EN}| = |\overline{FN}|$,

$$|\overline{EN}| + |\overline{FN}| > |\overline{PN}|. \tag{3}$$

Combining (1), (2), and (3), we obtain

$$|\overline{AE}| + |\overline{BF}| + |\overline{EN}| + |\overline{FN}| > |\overline{AP}| + |\overline{BP}| + |\overline{PN}|$$

which yields ($*$).

Thus there is a net saving in length which is directly proportional to the length, $|NP|$. If we choose the proper value for $|NP|$, the length saving will be ε and the area change will be of order ε^2, since the area is proportional to $|NP|^2$.

Proof of Step (ii). Suppose we have arcs of circles of radius R_1 and R_2 emanating from the node as sides of the angle α considered in Step (i). On each side of the angle there are two possibilities: The arc can go inside or outside the angle formed by the tangent rays, figures 2 and 3, respectively.

We take the case where the arc is inside the angle first. Choose A and B on the tangent line to the circles emanating from N, as before. Let C be the center of the circle from which the arc tangent to $\overline{AN}$ is taken. Let D be the point of the circle on the radius which when extended, passes through A. Let θ be the central angle of the circular arc DN, $\sphericalangle DCN$. Thus, for small values of θ, $\dfrac{1}{2}|\overline{AN}| < R\theta < |\overline{AN}|$.

Let $\varkappa$ be a positive constant and $\varepsilon > 0$. Choose θ to satisfy $\dfrac{\varkappa\varepsilon}{2} < \theta < \varkappa\varepsilon$. Since $\dfrac{\tan \theta}{\theta} \to 1$ as $\theta \to 0$ we know that $\dfrac{|\overline{AN}|}{|DN|} \to 1$ as $\varepsilon \to 0$. Thus if we change from

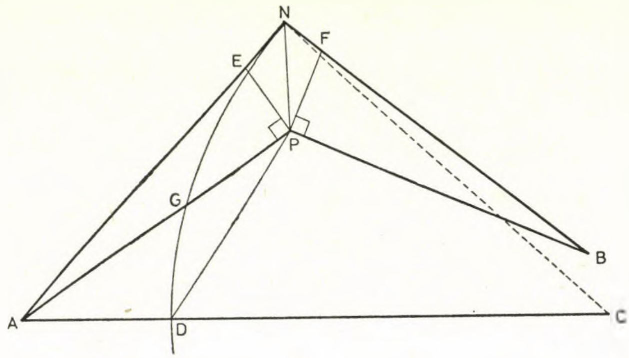


Fig. 2

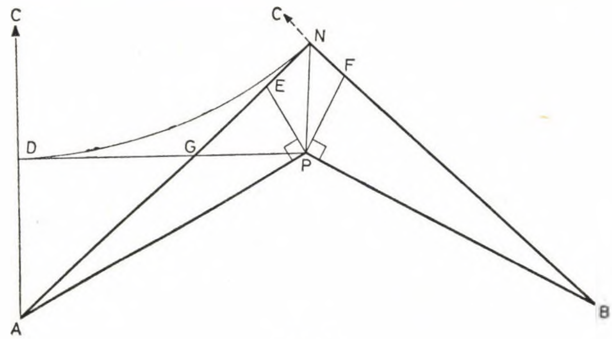


Fig 3

step (i) to this configuration we replace $\overline{PA}$ with $\overline{PD}$ and $\overline{AN}$ with $\overline{DN}$. Since $R(\sec\theta-1)=AD$ is of order θ^2 the change in replacing $\overline{PA}$ by the shorter $\overline{PD}$ is of order ε^2. Also $R(\tan\theta-\theta)=|\overline{AN}|-|ND|$ is of order θ^2, hence ε^2. Thus the total length saved is still of order ε. The change of areas is still of order ε^2.

In a similar way, we see in figure 3 that $|\overline{AD}|$ is of order ε^2 and thus replacing the straight segment by a circular arc will still allow a saving of ε in length while changing the areas by an amount of order ε^2.

Proof of Step (iii). We now correct any area changes while using at most half of the length saved. Suppose the area changes are less than $K\varepsilon^2$. Let L_1 and L_2 be the lengths of the arcs beginning at N and bounding the sides of the angle under consideration. See Figure 4. Let M_1 be the midpoint of $\overparen{NN_1}$ where N_1 is the node at the other end of the arc. Choose a point Q_1 on the perpendicular bisector of $\overparen{NN_1}$ outside the given angle and such that the area of $\triangle N_1M_1Q_1$ is equal to the area of $\triangle NPD$, namely $K_1\varepsilon^2$, $K_1\leqq K$ where $\overparen{N_1Q_1}$ is a circular arc of radius R_1. Thus $|\overline{M_1Q_1}|$ is of order ε^2. Consequently, $|N_1Q_1|-|N_1M_1|$ is of order ε^2. Hence replacing

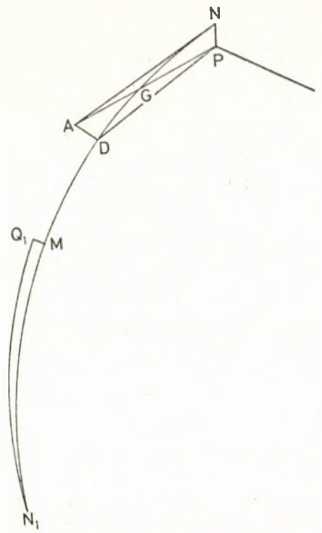


Fig. 4

$\widehat{M_1N_1}$ by $\overline{M_1Q_1}+\widehat{Q_1N_1}$ changes the length by an amount of order ε^2, while correcting the area changes of the regions bounded by the $\widehat{NN_1}$. A similar correction can be made on the arc $\widehat{NN_2}$. Thus for small $\varepsilon>0$, the length used to correct the area changes will be less than $\frac{\varepsilon}{2}$ while the area ratios of the regions will be left unchanged. Hence some length has been saved.

Proof of Step (iv). Obvious.

Proof of (2). Since the angles are at least $2\pi/3$ there can be at most 3 arcs. But if there are only two arcs the point must be on a circular arc separating two regions and hence is not a node.

Proof of (4). This follows from Euler's Formula

$$V-E+F=2$$

in conjunction with $F=n+1$ and $3V=2E$. The last equation comes from the fact that the network is trivalent.

Proof of (5). The first step is to derive formulae for the changes in area and arc length when a circular arc on a given chord is replaced by a circular arc of slightly larger or slightly smaller radius on the same chord. We use three parameters in this process, namely: the radius of the original arc, R, half the central angle of the original arc, θ, and the angle φ which is the difference between θ and the corresponding angle for the changed radius, see figure 5. Let $\widehat{AB}$ be an arc of a circle

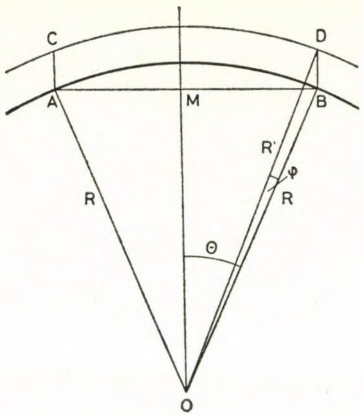


Fig. 5

with center O and radius R with central angle 2θ. Let M be the midpoint of $\widehat{AB}$ and let $\overline{AC}$ and $\overline{BD}$ be parallel to $\overline{OM}$ with C and D on the arc of radius R'. (In this figure we have shown $R'>R$.) Applying the law of sines to $\triangle BOD$ we obtain $R'=R\dfrac{\sin\theta}{\sin(\theta-\varphi)}$. Let δL denote the change in length from the arc of radius R to the arc of radius R' on a chord of length $|\overline{AB}|=|\overline{CD}|$. Thus

$$\delta L = 2R\left[\frac{(\theta-\varphi)\sin\theta}{\sin(\theta-\varphi)}-\theta\right]$$

where φ is negative for $R'<R$. Expanding $\dfrac{\sin\theta}{\sin(\theta-\varphi)}$ in powers of φ, we obtain,

$$\delta L = 2R\varphi(\theta\cot\theta-1)+O(\varphi^2). \tag{1}$$

We note that $\theta\cot\theta<1$, for $0<\theta<\dfrac{\pi}{2}$.

We next calculate δA. The area of a circular segment with central angle 2α and radius r is given by

$$a = r^2\left(\alpha-\frac{\sin 2\alpha}{2}\right). \tag{$*$}$$

Thus

$$\delta A = (R')^2\left(\theta-\varphi-\frac{\sin 2(\theta-\varphi)}{2}\right)-R^2\left(\theta-\frac{\sin 2\theta}{2}\right).$$

This can be rewritten

$$\delta A = R^2\left[\left[\left(\frac{\sin\theta}{\sin(\theta-\varphi)}\right)\right]^2\left[(\theta-\varphi)-\frac{\sin 2(\theta-\varphi)}{2}\right]-\left[\theta-\frac{\sin 2\theta}{2}\right]\right].$$

Expanding $\left[\dfrac{\sin\theta}{\sin(\theta-\varphi)}\right]^2$ and $\sin 2(\theta-\varphi)$ as a power series in φ we will only need to be careful with first order terms since φ will ultimately be chosen very small in comparison with θ, which will also be small. Hence,

$$(**)\qquad \left[\frac{\sin\theta}{\sin(\theta-\varphi)}\right]^2 = 1+2\varphi\cot\theta+O(\varphi^2)$$

and

$$(***)\qquad \sin 2(\theta-\varphi) = \sin 2\theta-2\varphi\cos 2\theta+O(\varphi^2).$$

Substituting in the above formula we obtain

$$\delta A = -\varphi R^2[1-\cos 2\theta-(2\theta-\sin 2\theta)\cot\theta]+O(\varphi^2),$$

but the expression in brackets can be simplified to obtain

$$(2)\qquad \delta A = -2\varphi R^2[1-\theta\cot\theta]+O(\varphi^2).$$

Let N be a node at which the total oriented curvature is not zero. Let the curvature of the three circular arcs be $\varkappa_1$, $\varkappa_2$, $\varkappa_3$ where we may suppose $\varkappa_1+\varkappa_2+\varkappa_3>0$. We may further suppose $\varkappa_1>0$. We wish to replace a segment of each circular arc leaving N by a segment of a different arc having the same chord. We may freely choose θ and φ_i for $i=1, 2, 3$.

There are really six cases to consider since we have assumed $\varkappa_1>0$. These are

(i) $\varkappa_2=\varkappa_3=0$

(ii) $\varkappa_3=0 \quad \varkappa_2>0$

(iii) $\varkappa_3=0 \quad \varkappa_2<0$

(iv) $\varkappa_2>0 \quad \varkappa_3>0$

(v) $\varkappa_2>0 \quad \varkappa_3<0$

(vi) $\varkappa_2<0 \quad \varkappa_3<0$.

In this paper cases (iii) and (v) will be written in detail. These illustrate all the techniques needed in the other cases. Also Case (iv) is impossible by Theorem 8 of Tomonaga [39], who observed that in this case replacing part of each arc by a line segment is more efficient. The same observation holds for Case (ii).

Case (iii). See Figure 6. Choose P_i on the arc with curvature $\varkappa_i$. Further choose P_i, $i=1, 2$, such that the arc $\overset{\frown}{P_iN}$ has central angle 2θ, where θ is chosen small enough that:

1) The circle centered at N containing P_1 and P_2 meets no arcs of the network except the three emanating from N,

2) $-\lambda=\theta\cot\theta-1<0$, and

3) $\lambda<10^{-1}|\overline{NP_3}|(\varkappa_1+\varkappa_2+\varkappa_3)$.

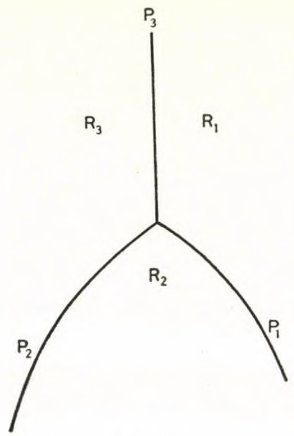


Fig. 6

Let $\varphi_i = \operatorname{sgn}(\varkappa_i)\varkappa_i^2$ $i=1, 2$, where we shall determine the positive value of ε later. Let A_i denote the area of the regions R_i indicated in Figure 6. Replace the arc $\overset{\frown}{NP_i}$ $i=1, 2$, by the arc obtained through the process indicated in Figure 5 for the given choices of θ and φ_i. Thus, using (1) and (2), we obtain

$$
\begin{aligned}
\delta L_1 &= 2\varepsilon\varkappa_1(\theta\cot\theta - 1)\\
\delta L_2 &= 2\varepsilon\varkappa_2(\theta\cot\theta - 1)\\
\delta A_1 &= +2\varepsilon[\theta\cot\theta - 1] + O(\varepsilon^2)\\
\delta A_2 &= -2\varepsilon[\theta\cot\theta - 1] + O(\varepsilon^2),
\end{aligned} \tag{3}
$$

where δL_i is the change in length of the arc $\overset{\frown}{P_iN}$ and δA_i is the change of area from the new arc from P_i to N. Hence on the boundary $\overset{\frown}{NP_1}$ an area of $|\delta A_1|$ is taken from R_1 and added to R_2 and on the boundary $\overset{\frown}{NP_2}$ an area of $|\delta A_2|$ is added to R_3 and taken from R_2. It follows that the change in area A_2 is of order ε^2. We may now rechoose P_1 so that θ is changed to θ_1 on the arc P_1N in such a way that $|\delta A_1| = |\delta A_2|$ and $\theta_1 = (1+\alpha\varepsilon)\theta$ where α is $O(1)$. Clearly δL_1 is now $2\varepsilon K_1(\theta\cot\theta - 1) + O(\varepsilon^2)$ so that $\delta L_1 + \delta L_2 = 2\varepsilon(\varkappa_1+\varkappa_2)(\theta\cot\theta - 1) + O(\varepsilon^2)$, where $\varkappa_1+\varkappa_2 = \varkappa_1+\varkappa_2+\varkappa_3 > 0$. Thus we have left the area of R_2 unchanged but the area of R_1 is too large by an amount of order ε while R_3 is too small by an equal amount, while causing a net saving in length of $\delta L_1 + \delta L_2$, which is of order ε. We can now correct the area perturbations by replacing the line segment $\overline{NP_3}$ by two line segments in the following way. Choose a point Q on the perpendicular bisector of $\overline{P_3N}$ on the side which initially goes through R_1 at that distance for which the area of the triangle $\triangle P_3NQ$ has the desired area which is $O(\varepsilon)$ (see Figure 7). The sum $|\overline{QN}| + |\overline{QP_3}|$ is equal to $|\overline{NP_3}| + O(\varepsilon^2)$, since

$$l < \sqrt{l^2 + \beta^2\varepsilon^2} < l + \frac{\beta^2\varepsilon^2}{2l}.$$

Thus the additional length required is of order ε^2 while the length saved by replacing the circular arcs is of order ε. Hence if ε is small enough there is a net saving of length, which shows the initial configuration was not optimal.

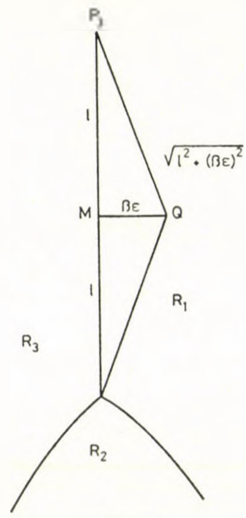


Fig. 7

Case (v). See Figure 8.

We choose P_1, P_2, and P_3 as in Case (iii). Again let $\varphi_i = \operatorname{sgn}(\varkappa_i)\varepsilon\varkappa_i^2$. Thus, as before, we obtain

$$\delta L_i = 2\varepsilon K_i(\theta \cot\theta - 1) \qquad i = 1, 2, 3$$

$$\delta A_i = +2\varepsilon[\theta\cot\theta - 1] + O(\varepsilon^2) \qquad i = 1, 2 \tag{4}$$

$$\delta A_3 = -2\varepsilon[\theta\cot\theta - 1] + O(\varepsilon^2).$$

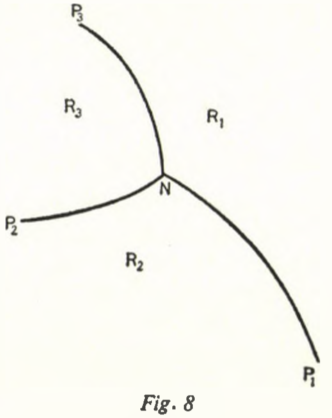


Fig. 8

By replacing θ by θ_i $i=1, 2, 3$ in the expressions for δL_i, δA_i in (4) we can get all the area changes to be equal while $\theta_i=(1+\alpha_i\varepsilon)\theta$, where $\alpha_i=O(1)$. For ε positive and small enough the total length used is reduced by this process since,

$$\delta L = \delta L_1+\delta L_2+\delta L_3 = 2\varepsilon(\varkappa_1+\varkappa_2+\varkappa_3)(\theta\cot\theta-1)+O(\varepsilon^2)$$

and $(\varkappa_1+\varkappa_2+\varkappa_3)(\theta\cot\theta-1)<0$. Thus the original configuration was not the best.

Proof of (6). If the network is not connected, slide two of its components until they touch. This point of contact can be considered as a node with more than three edges meeting. Hence the configuration, which has the same areas and total length as the original, is not best possible by Part 2.

This completes the proof of the theorem.

IV. Examples and conjectures

The isoperimetric configuration for constructing n cells of equal area in the plane is given in Figure 9 for $n=1, 2, 3$. The isoperimetric ratio for the configuration, $I(n)=\frac{(L/n)^2}{a}$, where L is the total length of all arcs and a is the area of each cell is also given.

$$I(1) = 4\pi \qquad I(2) = \frac{8\pi+3\sqrt{3}}{3} \qquad I(3) = 2\left(\pi+\frac{2}{\sqrt{3}}\right)$$

$$= 12.566 \qquad = 10.109 \qquad = 8.59$$

$$n = 1 \qquad n = 2 \qquad n = 3$$

We also know by standard limiting argument that $\lim_{n\to\infty} I(n)=2\sqrt{3}=3.464$, since the hexagonal tiling is the most efficient and the average length per cell is half the perimeter of the hexagon.

If we let $IH(n)$ denote the square of the average perimeter per cell, i.e. the square of the total length L divided by n, for a tiling by hexagonal tiles of area 1, then of course $IH(n)>I(n)$ but $\lim_{n\to\infty} IH(n)=\lim_{n\to\infty} I(n)$.

In his papers [14, 15], L. Fejes Tóth considered isoperimetric divisions in the plane by straight line segments. It follows from these works that the arrangement of hexagons is the most efficient for the shape they cover.

The conjecture that both these functions are monotone is still open. The following table gives the first few values of the functions.

The ratio in the last column of Table 1 is not monotone. It would be interesting to know if the lack of monotonicity is a one-time initial occurrence or if it happens again.

Finally it should be noted that the methods of the above proof also yield the results for joining n points by the shortest network [39]. The proof in this case is simpler since one does not need to worry about balancing area, consequently all arcs are line segments and everything has zero curvature.

Table 1

n	$I(n)$	$IH(n)$	$IH(n)/I(n)$
1	$4\pi = 12.5\ldots$	$8\sqrt{3} = 13.8\ldots$	$2\dfrac{\sqrt{3}}{\pi} = 1.102\ldots$
2	$\dfrac{8\pi+3\sqrt{3}}{3} = 10.1\ldots$	$\dfrac{121}{6\sqrt{3}} = 11.6\ldots$	$\dfrac{121}{16\sqrt{3}\pi+18} = 1.151\ldots$
3	$2\left(\pi+\dfrac{2}{\sqrt{3}}\right) = 8.59\ldots$	$\dfrac{50}{3\sqrt{3}} = 9.62\ldots$	$\dfrac{25}{3(\sqrt{3}\pi+2)} = 1.119\ldots$

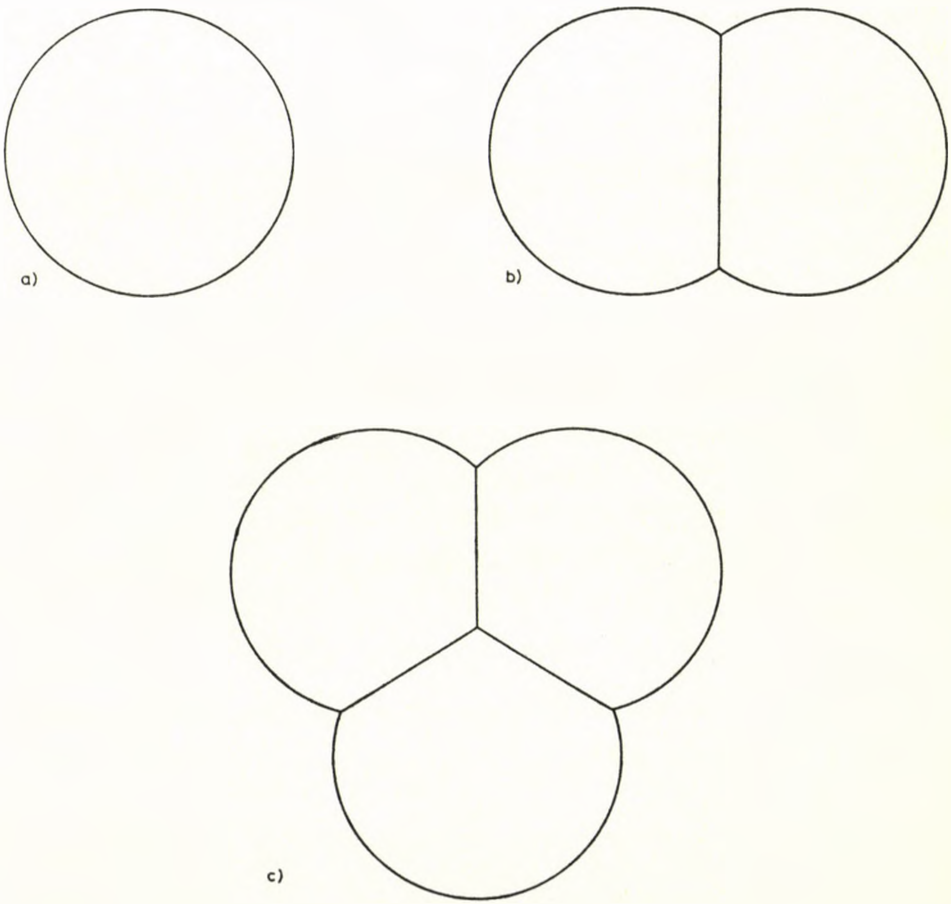


Fig. 9

REFERENCES

[1] Aleksandrov, A. D. and Strel'cov, V. V., The isoperimetric problem and estimates of the length of a curve on a surface, *Trudy Mat. Inst. Steklov.* **76** (1965), 67—80. *MR* **34** #1971.

[2] Ball, W. W. R., *Mathematical recreations and essays*, 9^{th} ed., MacMillan, London, 1920.

[3] Bandle, C., A generalization of the method of interior parallels, and isoperimetric inequalities for membranes with partially free boundaries, *J. Math. Anal. Appl.* **39** (1972), 166—176. *MR* **46** #5854.

[4] Bandle, C., A geometrical isoperimetric inequality and applications to problems of mathematical physics, *Comment. Math. Helv.* **49** (1974), 496—511. *MR* **50** #10565.

[5] Bandle, C., *Isoperimetric inequalities and applications*, Monographs and studies in mathematics, Vol. 7, Pitman Advanced Publishing Program, Boston, 1980. *MR* **81e**: 35095.

[6] Blaschke, W., *Kreis und Kugel*, Veit & Co., Leipzig, 1916.

[7] Bleicher, M. N. and Fejes Tóth, L., Two-dimensional honeycombs, *Amer. Math. Monthly* **72** (1965), 969—973.

[8] Bonnesen, T., *Les problèmes des isoperimètres des isépiphanes*, Collection de monographies sur la théorie des fonctions, Gauthier-Villars, Paris, 1929.

[9] Boyer, C. B., *A history of mathematics*, Wiley, New York, 1968. *MR* **38** #3105.

[10] Boys, C. V., *Soap bubbles: Their colours and the forces which mould them*, New and enlarged edition, London, 1931.

[11] Burago, Yu. D., Isoperimetric inequalities in the theory of surfaces of bounded external curvature, Seminars in Mathematics, V. A. Steklov Mathematical Institute, Leningrad, Vol. 10, Consultants Bureau, New York—London, 1970. *MR* **43** #2645.

[12] Burago, Yu. D. and Zalgaller, V. A., The isoperimetric problem in the case of limitation of the width of a domain on a surface, *Trudy Mat. Inst. Steklov.* **76** (1965), 81—87. *MR* **34** #1972.

[13] Coolidge, J. L., *A history of geometrical methods*, Oxford University Press, New York, 1940. *MR* 2—113, 419.

[14] Danelič, I. A., Estimate of the area of a surface of bounded absolute mean integral curvature in terms of its absolute integral mean curvature and the sum of the lengths of the boundary curves, *Sibirsk. Mat. Ž.* **7** (1966), 1199—1203. *MR* **35** #886.

[15] Dormer, K. J., *Fundamental tissue geometry for biologists*, Cambridge University Press, Cambridge, 1980.

[16] Eggleston, H. G., *Convexity*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 47, Cambridge University Press, New York, 1958. *MR* **23** #A2123.

[17] Fejes Tóth, L., On shortest nets with meshes of equal area, *Acta Math. Acad. Sci. Hungar.* **11** (1960), 363—370. *MR* **23** #A3510.

[18] Fejes Tóth, L., Isoperimetric problems concerning tessellations, *Acta Math. Acad. Sci. Hungar.* **14** (1963), 343—351. *MR* **28** #529.

[19] Fejes Tóth, L., *Regular figures*, International series of monographs in pure and applied math., Vol. 48, Macmillan, 1964. *MR* **29** #2705.

[20] Fejes Tóth, L., What the bees know and what they do not know, *Bull. Amer. Math. Soc.* **70** (1964), 468—481. *MR* **29** #524.

[21] Fiala, F., Le problème des isopérimètres sur les surfaces ouvertes à courbure positive, *Comment. Math. Helv.* **13** (1941), 293—346. *MR* 3—301.

[22] Hadwiger, H., *Vorlesungen über Inhalt, Oberfläche, und Isoperimetrie*, Springer-Verlag, Berlin, 1957. *MR* **21** #1561.

[23] Heath, T. L., *A history of Greek mathematics*, 2 Vols, Oxford, Clarendon, 1921.

[24] Huber, A., On the isoperimetric inequality on surfaces of variable Gaussian curvature, *Ann. of Math.* (2) **60** (1954), 237—247. *MR* **16**—508.

[25] Kohler-Jobin, M.-Th., Démonstration de l'inégalité isopérimétrique $P\lambda^2 \geqq \pi j_0^4/2$, conjecturée par Pólya et Szegő, *C. R. Acad. Sci. Paris Sér. A—B* **281** (1975), A119—A121. *MR* **52** #6551.

[26] Kohler-Jobin, M.-Th., Une propriété de monotonie isopérimétrique qui contient plusieurs théorèmes classiques, *C. R. Acad. Sci. Paris Sér. A—B* **284** (1977), A917—A920. *MR* **55** #7056.

[27] Luttinger, J. M., Generalized isoperimetric inequalities, *Proc. Nat. Acad. Sci. U.S.A.* **70** (1973), 1005—1006. *MR* **47** #9093.

[28] MacBeath, A. M., Compactness Theorems, *Seminar on Convex Sets V*, Institute for Advanced Study, Princeton, 19xx.
[29] Newman, J., *The world of mathematics*, 4 vols, Simon & Schuster, New York, 1956. *MR* **18**—453.
[30] Osserman, R., The isoperimetric inequality, *Bull. Amer. Math. Soc.* **84** (1978), 1182—1238. *MR* **58** # 18161.
[31] Pólya, G., *Induction and analogy in mathematics. Mathematics and plausible reasoning*, I, Princeton University Press, Princeton, 1954. *MR* **16**—556.
[32] Pólya, G. and Szegő, G., *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies, no. 27, Princeton University Press, Princeton, 1951. *MR* **13**—270.
[33] Pólya, G. and Schiffer, M., Convexity of functionals by transplantation, *J. Analyse Math.* **3** (1954), 245—346. *MR* **16**—591.
[34] Schmidt, E., Der Brunn—Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie, I, *Math. Nachr.* **1** (1948), 81—157. *MR* **10**—471.
[35] Schmidt, E., Der Brunn—Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie, II, *Math. Nachr.* **2** (1949), 171—244. *MR* **11**—534.
[36] Schmidt, W., Zur Geschichte der Isoperimetrie im Altertum, *Bibliotheca Mathematica* (3) **2** (1901), 5—8. *Jb. Fortschritte Math.* **32** (1901), 50.
[37] Schoenberg, I. J., *Mathematical time exposures*, Mathematical Association of America, Washington, D. C., 1982. *MR* **85b**: 00001.
[38] Steiner, J., *Gesammelte Werke*, Vol. 2, G. Reimer, Berlin, 1882.
[39] Stevens, P. S., *Patterns in nature*, Atlantic Little, 1974.
[40] Thompson, D. A. W., *On growth and form*, New edition, Cambridge University Press, Cambridge, 1942. *MR* **3**—291.
[41] Tomonaga, Yasuro, *Geometry of length and area*, 1, Dept. of Mathematics Utsunomiya University, Utsunomiya, 1974. *MR* **55** # 3977.
[42] Verblunsky, S., On the shortest path through a number of points, *Proc. Amer. Math. Soc.* **2** (1951), 904—913. *MR* **13**—577.
[43] Weaire, D. and Rivier, N., Soap cells and statistics — Random patterns in two dimensions, *Contemporary Phys.* **25** (1984), 59—99.

(Received May 20, 1986)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
MADISON, WI 53706
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 139—148

КОМБИНАТОРНАЯ ГЕОМЕТРИЯ И ДИСКРЕТНАЯ ГЕОМЕТРИЯ — ИХ МЕСТО В НАУКЕ И В СОВРЕМЕННОМ МИРЕ

И. М. ЯГЛОМ

Дорогому коллеге Ласло Фейешу Тоту к его 70-летию

Мне уже приходилось писать о месте, занимаемом сегодня комбинаторной и дискретной геометрией [1]; однако, возможно, уместно снова коснуться этого круга вопросов.

1° Европейская математика (и даже шире — все европейская культура, тесно с математикой связанная, ср. [2] или обсуждение параллелей в развитии (европейских) математики и искусства в заключительной части книги [3]) знала небольшое число переломных моментов, характеризующихся резкой сменой научной идеологии, сменой парадигм в терминах книги [4]: возникновение классической античной культуры в VI—V вв. до н. э.; гибель античной цивилизации в IV—V вв. и замена её принципально отличным от нее средневековым обществом; крах средневековой идеологии в эпоху итальянского Возрождения, непосредственно смыкающайея с промышленным переворотом (первой великой научно-технической революцией) XVII—XVIII вв.; научно-техническая революция наших дней, характеризующаяся систематическим использованием механизмов (компьютеров) для автоматизации умственной деятельности человека, подобно тому, как научный переворот прошлого ознаменовался заменой физического труда человека и животных гораздо более производительным машинным трудом. В истории математики переломным можно также считать XIX в., во многом подготовивший почву для современных исканий; среди революционных достижений этого столетия надо упомянуть создание математической логики и науки об основаниях математики, совершенно по новому поставившие вопрос о самом статусе математической науки и вновь приблизившие нас к установкам древнегреческих мыслителей, почти полностью зачеркнутых наукой XVII—XVIII вв. (ср., например, [5]), а также создание общих концепций *множества* (и возниковение теоретико-множественного подхода к математике) и *группы*. При этом каждая «общенаучная революция» сказывалась на резком изменении сопутствующего ей математического мышления и, в частности, на (единственно нас здесь интересующем) изменении положения геометрии в системе математических наук.

Математика как определенное научное направление, как дедуктивная структура, полностью отличная от (возникших примерно в то же время) естественных (физика) и гуманитарных (история, филология) наук, сложилась в древней Греции в VI—V вв. до н. э. (ионийская и пифагорейская школы) и

1980 *Mathematics Subject Classification.* Primary 05B40, 05B45, 52A45. Secondary 05CXX.
Key words and phrases. Combinatorial geometry, discrete geometry, history of geometry.

позже (афинская школа Платона и Аристотеля, ещё позже — александрийцы). При этом «эталонным» образцом математической системы являлась для древних греков прежде всего *геометрия*, занимавшая в их науке (и даже шире — в культуре) совершенно особое место: так, именно геометрия послужила базой создания первой четко структурированной аксиоматической схемы (Евклид). Напротив, крушение античной науки явилось, одновременно, периодом заката геометрии: в европейской средневековой математике (Леонардо Фибоначчи, «косисты»), как, впрочем, и в принявшей от древних греков эстафету научного знания арабской (точнее, арабскоязычной) культуре, геометрия ведущего места никак не занимала и основные достижения здесь были связаны отнюдь не с ней. Характерен также чисто алгебраический характер первых выдающихся достижений европейской математики эпохи Возрождения, коими явилось решение уравнений 3-й и 4-й степеней (Тарталья, Кардано, Ферарри, Бомбелли, XVI в.). Достижения алгебраистов XVI в., крупнейшим из которых был Виета, подготовили почву (и создали адекватный математический язык, а также необходимую символику) для великой научной революции XVII в., начало которой было положено, в определенном смысле, «закрытием» геометрии, сведением её к алгебре (Декарт, Ферма) и разработкой языка и символики, близких к современным (Декарт), а завершением явилось возникновение дифференциального и интегрального исчисления (Ньютон, Лейбниц). При этом характерно, что относящиеся к той же эпохе выдающиеся достижения в области геометрии (Дезарг, Паскаль) не были своевременно оценены и даже просто замечены (в силу чего основополагающий трактат Б. Паскаля по проективной геометрии не был никогда напечатан и впоследствии пропал).

Напротив, XIX в. явился периодом расцвета геометрии, что, разумеется, никак не было случайным: столь важная для века XX, как понимаем мы сегодня, попытка вновь серьезно продумать методологические и общефилософские позиции древних греков и сушность их подходе к математике (относящейся, по выражению Платона, к «миру умопостигаемому», в отличие от «мира видимого», коим занимаются естественные науки) вынудила снова обратиться к играющей столь важную роль в античной науке и философии геометрии — и первым революционным открытием XIX столетия явилось открытие неевклидовой геометрии (Гаусс, Я. Бойаи, Лобачевский), необходимое для более глубокого осмысления сущности аксиоматизированной (или математической) структуры. И если знаменитый Н. Бурбаки [6] датирует конец столь, относительно, близкого к нам хронологически и столь далекого идейно «золотого века» геометрии «Эрлангенской программой» Клейна (1872), то мне кажется более справедливым обозначить этот конец (и конец XIX столетия) появлением «„Начал" Евклида XX столетия» — «Оснований геометрии» Гильберта (1899), завершивших начатую греками работу по построению строго дедуктивной системы евклидовой геометрии и окончательно прояснивших смысл самого понятия «математическая структура», пришедщего к нам из Древней Греции, но древнегреческими мыслителями все же до конца не осознанного.

2° Если ведущей математической дисциплиной эпохи древней Греции следует считать *геометрию*, а с XVII и вплоть до первой половины XX в. эту роль твердо занимал *математический анализ*, то 2-ю половину XX в. вполне можно считать веком алгебры, о чем лучше всего свидетельствуют Éléments de Mathématique уже названного многоголового француза Н. Бурбаки трактующие всю математику как систематический анализ всевозможных «математических структур», которые наши предки, пожалуй, предпочли бы называть «алгебраическими структурами». [От древней Греции (или от уважения к древней Греции) пришло к нам многовековое употребление во многих языках слова «геометр» в смысле «математик»; научной революции XVII и следующих веков обязаны мы (в нашей стране ещё вполне жизненным) довольно нелепым термином «высшая математика», обозначающим дифференциальное и интегральное исчисление с приложениями; сегодня же вполне можно ожидать, что наши потомки будут употреблять в смысле «математик» термин «алгебраист».] Эта «алгебраизация математики» затронула как математический анализ, свидетельством чего может служить, например, вполне «бурбакистский» по своим установкам 9-томный учебник [7] (охватывающий, впрочем, в соответствии с традицией французских университетских учебников, не один лишь «математический анализ» в строгом смысле этого термина, но чуть ли не всю (университетскую) математику), так — и это нам здесь особенно интересно — и геометрию.

В самом деле, дифференциальная геометрия считавшаяся в 1-й половине нашего столетия ведущей ветвью геометрической науки, возникла в начале прошлого века как «приложения анализа к геометрии», где ударение, пожалуй, делалось все же на последнем слове; вершинным пунктом, достигнутым этим новым направлением следует считать общие концепции великого геометра (и великого физика) Римана, сформулированные на уровне чисто описательном, фактически — почти без всяких формул [8]. Однако дальнейшая расшифровка конструкций Римана вскрыла скрывающееся за ними формализованное исчисление [9] — и уже на рубеже прошлого и настоящего веков дифференциальная геометрия во многом приобрела черты громоздкой алгебраической системы с упором на формальные манипуляции с многоиндексными символами преславутого «жонглирования индексами» (ср., например, [10]); ныне же все это направление приобрело чисто «бурбакистский» вид учения о «связностях на расслоениях». Но и являвшаяся центром всех геометрических устремлений XIX в. проективная геометрия тоже ныне смогла сохранить свой научный авторитет лишь ценой полной трансформации, при которой она рассматривается сегодня скорее как глава линейной алгебры чем как раздел геометрии (см., например, основополагающий учебник [11]). С другой стороны даже и элементарная геометрия никак не смогла уберечься от «алгебраизирования» — для примера здесь достаточно назвать темпераментно написанный и не случайно не содержащий ни одного чертежа учебник [12], автор которого настойчиво проводит мысль о том, что элементарная геометрия — это есть раздел линейной алгебры, а никакой другой «элементарной геометрии» ныне нет и быть не должно. Но даже и принятая в книге [12] векторная аксиоматика элементарной (евклидовой) геометрии не является пределом её «алгебраизации» — так, пользующаяся большой и заслуженной известностью книга [13] начинает аксиоматику геометрии (в частности, и евклидовой) с фразы: «рассмотрим группу $\mathfrak{G}$ и систему $\mathfrak{S}$ её инволютивных образующих», а позднее называет элементы $\mathfrak{S}$ «точками» и «прямыми» рассматриваемого пространства, а в [14] обсуждение аксиоматики геометрии начинается с предложения: «назовем отрезок AB произведением точек A и B; это произведение коммутативно, ассоциативно, идемпотентно и дистрибутивно относительно сложения точек (или фигур), понимаемого как их объединение».

3° Присущая нашим дням чрезмерная «алгебраизация» математики, в последние десятилетия и годы поддержаная её «компьютеризацией» (или «алгоритмизацией»), о чем мы ещё скажем ниже, вполне может вызывать и некоторую тревогу. Хорошо известно, что мышление человека представляет собой достаточно сложный процесс, в котором тесно соединены (и могут быть разъеденены лишь условно) как интуитивная, так и логическая (дискурсивная) компоненты (ср. [15]), или как «геометрическая» («физическая») струя, несколько условно сопоставляемая обычно деятельности правого полушария головного мозга, так и «алгебраическая» («алгоритмическая») струя, условно свяваемая с левым полушарием мозга (см., например, [16]; достаточно схематичное обсуждение математических аспектов этого положения дела дано в [17] — ср. также [18]). При этом для человеческого мышления в целом безусловно необходимы как интуиция, так и логика; как и алгебра, так и геометрия — и игнорирование одной или другой стороны многостороннего процесса мышления кажется нам совершенно недопустимым.

Разумеется, разным областям человеческой культуры, как и разным эпохам или разным людям (на чем мы ещё остановимся ниже) может быть присуще доминирование той или иной стороны мышления; однако параллельное существование, скажем, принципиально алогичного искусства (а алогичным является всякое искусство высокого уровня, о чем см. хотя бы [15]) и строго формализованной математики (ср. учебник [19]) напоминает нам о важности сосуществования (и взаимодополняемости) в человеческой культуре интуитивной и дискурсивной составляющих (причем и в искусстве и в науке на самом деле присутствуют обе эти компоненты — различие заключается лишь в их «удельном весе»). Аналогично этому совсем не случайным кажется нам частое одновременное появление в истории науки и культуры «взаимно дополняющих» пар мыслителей, демонстрирующих принципиально разные подходы к процессу познания мира. Так у истоков европейской научной культуры стояли трезвый рационалист Галилей (логика) и страстный мистик Кеплер (интуиция), а неевклидову (гиперболическую) геометрию одновременно открыли яркий геометр с физическими и астрономическими интересами Н. И. Лобачевский и острый логик Я. Бойаи. Но наиболее поучительно с позиций настоящей статьи противопоставление столь грандиозных фигур из истории мировой культуры как глубокий мистик Платон с ярко выраженным художественным мышлением и картинным восприятием Вселенной и строгий логик-рационалист Аристотель, стоящий у истоков всей (в том числе — и современной) логики, или как физик (и тоже мистик) Ньютон и логик и алгоритмист, рационалист Лейбниц (см. по этому поводу [20]).

Необходимость для человеческого мышления как Платона, так и во всем, кажется, противоположного ему Аристотеля, как Ньютона, так и Лейбница хорошо иллюстрируется всей историей культуры. В Афинах IV в. до н. э. Платон определенно доминировал над Аристотелем — и не случайно первый (и, видимо, мало совершенный) корпус сочинений Аристотеля был составлен через ряд столетий после смерти их автора. Однако уже в период распространения в мире эллинистической культуры «маятник истории», пожалуй, качнулся от Платона к Аристотелю — и не случайно уже в этот период термином «Философ» без указания имени в литературе всегда обозначался именно Аристотель (но не Платон!), подобно тому как слово «Поэт» (такие без имени) всегда обозначало в древней Греции Гомера. Доминирование Аристотеля над Платоном имело ещё более безусловный характер в культуре европейского средневековья, когда Платон был практически почти забыт; напротив, эпоха Возрождения началась с бурного увлечения Платоном, рукописи которого были вновь завезены в Европу после крушения Византии, с создания «Платоновских академии», являвшихся основными очагами гуманистической культуры. Создание математической логики (XIX в.) снова повысило научный престиж Аристотеля; ныне же в европейской культуре наблюдается новый взрыв интереса к Платону — и не случайно через книгу [6] проходит непрекрытое восхищение Платоном и явственно ощущающееся скептическое отношение к Аристотелю (последним, кстати сказать, на мой взгляд вовсе не заслуженное). Аналогично этому, если «высшую математику» (calculus) мы безусловно получили из рук Лейбница, в то время как относящиеся к этой области знания терминология и символика Ньютона известны сегодня, пожалуй, одним лишь специалистам по истории науки, то не только прискорбный спор о приоритете в открытии дифференциального и интегрального исчисления в свое время бесспорно выиграл Ньютон (с чем связана и смерть Лейбница в бедности и почти в безвестности), но и, — что, разумеется, является куда более важным, — вся наша научная идеология XVII, XVIII и XIX веков шла «путем Ньютона», все эти столетия рассматривавшегося как «ученый № 1». Однако уже в XIX столетии оказались в определенной степени реализованными столь обогнавшие свое время мечты Лейбница о «геометрическом исчислении» и о «логическом исчислении»; XX же век частично осуществил и центральную идею Лейбница (не мало раздражавшую, кстати, видимо Ньютона, которому сама подобная постановка вопроса казалась концщунственно недопустимой!) о механизированном устройстве, осуществляющем логические процедуры и помогающем человеку в выполнении умственных операций — компютере (этого слова, разумеется, у Лейбница не было, — но сама идея была осознан им с завидной точностью и полнотой!). Наряду с этим 2-я половина XX в. принесла и полноценное обоснование лейбницевского «исчисления дифференциалов» — в форме нестандартного анализа [21], точно соответствующей, как мне кажется, достаточно отчетливым, хоть и в ту пору и вовсе не формализованым представлениям Лейбница, а также тоже идущее от Лейбница внимание к «конечным» разделам математики и, в частности, к комбинаторике. А современная «компьютерная» эпоха человеческой культуры вполне может быть расценена как решительная (хоть, надо думать, временая) победа научной идеологии Лейбница над ньютоновским духом в науке.

4° Не повторяя здесь сказанное в [1] относительно кажущегося неожиданым расцвета в наши дни *дискретной геометрии* (ср. [22], [23]), развиваемой, в первую очередь, сильной англо-саксонской школой (покойный Г. Давенпорт, К. А. Роджерс, Г. С. М. Коксетер и др.), венгерской школой Л. Фейеша Тота, пожалуй, несколько более ограниченной по тематике и объему работы московской школой недавно скончавшегося Б. Н. Делоне, и *комбинаторной геометрии* ([24], [25]), во многом дискретной геометрии родственной (но только имеющей своим «исходным пунктом» не геометрическую теорию чисел как дискретная геометрия, а «оптимизационную математику» наших дней), укажем, что в какой-то мере эти два направления математической науки, столь ярко геометрические по своим задачам и методам, по роли здесь «геометрического видения» и по обилию (и необходимости) выразительных и красивых чертежей, в значительной степени сосредоточили «правостороннюю компоненту мышления» современной математики (в терминах книги [16]), — чем можно объяснить и нынешнее процветание этих новых «математических наук», ещё раз напомнившее нам о важности всех возможных философских и общенаучных подходов к познанию действительности. В [1] мы говорили также о тесных связях дискретной и комбинаторной геометрии с некоторыми специфическими чертами современной математики (порожденными «веком компьютеров» или, напротив, эту эру вызвавшими — вряд ли уместно вдаваться здесь в старинный спор о том, что было раньше: яйцо или курица); о подчеркиваемых также и чисто лингвистически связях дискретной геометрии с современной *дискретной (конечной) математикой* и комбинаторной математики с переживающей ныне невиданный расцвет *комбинаторикой;* о тесной связи этих направлений науки с вышедшими на авансцену математической мысли *теорией выпуклых тел, теорией многогранников* или, скажем, *теорией графов.* Здесь же мы остановимся лишь на (в [1] обойденых) прямых связях дискретной и комбинаторной геометрии с современной компьютерной эрой.

Наиболее близко лежащим моментом здесь можно счесть естественность использования компьютеров для решения тех или иных задач дискретной и комбинаторной геометрии — в иных разделах геометрии указать напрашивающиеся на «машинное решение» задачи, видимо, гораздо труднее. В качестве простейшего — и не особенно глубокого — примера укажу тут на идущую от Л. Фейеша Тота [26] задачу нахождения *ньютоновых чисел* (плоских, выпуклых) фигур: под ньютоновым число фигуры F понимается наибольшее число $\mathcal{N}(F)$ равных F фигур $F_1, F_2, \ldots, F_{\mathcal{N}}$, которые можно «приложить» к F так, чтобы F и F_i имели общие граничные точки, но ни одной общей внутренней точки и также любые две фигуры F_i и F_j (где $i, j = 1, 2, \ldots, \mathcal{N}$; $i \neq j$) общих внутренних точек не имели. Ньютоновские числа всех правильных n-угольников (где $n > 3$ и $\neq 5$) были перечислены Бёрёцки [27], а случай $n = 5$ с помощью весьма громоздкого и мало изящного перечисления многих случаев расположения фигур рассмотрел Линхарт [28]; с помощью же компьютера (ЭВМ) эта задача решается куда проще ([29]).

Но, разумеется, гораздо больше оснований для соотнесения дискретной и комбинаторной геометрии с современной компьютеризацией науки и жизни дают прямые использования основной для дискретной геометрии задачи о плотнейших упаковках шаров (или двойственной задачи о редчайшем покрытии шарами) в реализуемых на компьютерах задачах численного анализа и обратное применение компьютерных вычислений для решения этих столь важных сегодня для приложений задач. Здесь слова «столь важных сегодня» апеллируют к идущему от Шеннона [30] использованию теорем об упаковках шаров к *теории кодирования,* огромная научная актуальность которой также тесно связана с «компьютеризацией общества» — и если когда то и создателями и «потребителями» дискретной геометрии являлись, в первую очередь, специалисты по теории чисел, то ныне и в той и в другой области чаще всего выступают специалисты по теории информации и теории кодирования и инженеры-связисты (ср. обзор [31], принадлежащий перу выдающегося авторитета в области теории кодирования и примыкающей сегодня к теории кодирования и также в компьютерный век весьма важной для приложений криптографии). Наконец, общеизвестна близость к компьютерной тематике комбинаторных задач теории графов или теории многогранников (ср. [32]), многие из которых вполне уместно отнести к комбинаторной геометрии.

5° Близость дискретной и комбинаторной геометрии к ведущим направлениям современной мысли создали также своеобразное положение, при котором эти научные направления можно считать интегрированными современной культурой. На первое место здесь, пожалуй, можно поставить тесную связь дискретной геометрии с соображениями *симметрии* (ср., например, книги [22] и [35]), играющими столь важную роль в науке и культуре сегодняшнего дня (ср. [33]—[39] или предисловие [40] к русскому переводу книги [33]). Расцвет дискретной и комбинаторной геометрии не случайно совпал также с резким повышением в современной жизни роли *дизайна*, а в современном искусстве — новых направлений, дизайном порожденных или с дизайном тесно связанных; при этом близость к дискретной геометрии творчества ряда популярных ныне художников (назовем здесь хотя бы французского художника В. Вазерелли, возможно, не случайно вышедшего из Венгрии, лидирующей в области дискретной геометрии) сразу бросается в глаза. В связи со сказанным в п. 4° уместно отметить также родство анализируемых геометрических дисциплин с современной орнаменталистикой (ср. глубокую книгу [41]) и, особенно, с *машинной* (компьютерной) *графикой*. При этом характерно возникновение в наши дни специализированного международного журнала *Leonardo*, широко затрагивающего вопросы дизайна и, в частности, машинной графики; даже внешний облик этого журнала (или специализированных выпусков [42] статей из него) сразу же обнаруживает глубинную связь журнала Leonardo с вопросами дискретной (и комбинаторной) геометрии. (Не случайным является также и оформление образцами машинной графики весьма современной по своим установкам и пользующейся ныне столь большой известностью книги [43]; в развивающей геометрическую струю этой книги монографию [44] включены ряд результатов и задач, прямо примыкающих к дискретной и комбинаторной геометрии.) Наконец, весьма характерным для нашего времени является «феномен М. К. Эшера», взрывоподнобный рост интереса к творчеству этого «математического графика», нашедший отражение как в обилии посвященной Эшеру (книжной и журнальной) литературы (см., например, [39] или [45], [46]), так и в недавнем (1985 г.) международном междисциплинарном конгрессе в Риме, специально посвященном М. К. Эшеру (см.[47])— а меж тем геометрические истоки творчества Эшера безусловно порождены дискретной геометрией. Характерным отражением сегодняшнего интереса к дискретной и комбинаторной геометрии является и неожиданная книга [48] такого яркого представителя современной культуры как недавно скончавшийся американский архитектор и литератор Р. Б. Фуллер (1895—1984), не случайно посвященная уже упоминавшемуся выше Г. С. М. Коксетеру и, в известной мере, вдохновленная геометрическими трудами Коксетера: обрашая стандартную постановку вопроса о «геометрическом мышлении» Фуллер ставит в своей книге вопрос о «геометрии мышления» (книга имеет подзаголовок: Explorations in the Geometry of Thinking) и пытается найти в геометрических конструкциях «коксетеровского типа» истоки умственной деятельности человека.

В заключении хочется указать, что все вышесказанное подчеркивает актуальность относящейся к дискретной и комбинаторной геометрии проблематики для современного преподавания. Это обстоятельство хорошо ощущают наиболее чуткие к современным веяниям математики и педагоги — здесь можно указать на большое место заимствованных из дискретной геометрии мотивов в лучших из современных учебников геометрии [49], [50], принадлежащих перу видных ученых (ср. также The Coxeter Festschrift [51]), в книгах и статьях такого видного пропагандиста и популяризаторе математики как М. Гарднер (у которого можно найти специальные эссе о М. К. Эшере и Г. С. М. Коксетере — см. также посвященную Гарднеру книгу [52]) или в одновремененной Коксетером книге [53] (я здесь ограничиваюсь лишь весьма немногочисленными примерами такого рода). Сложнее обстоит дело с комбинаторной геометрией (о чем см., например, [1]) — и настоящую статью мне хочется кончить обращением к венгерской геометрической школе Л. Фейеша Тота с призывом продолжить книги [22] и [35] иными сочинениями сходного плана, шире затрагивающими вопросы комбинаторной геометрии и, возможно, обращенными к более обширному кругу читателей.

ЛИТЕРАТУРА

[1] YAGLOM, I. M., Elementary geometry, then and now, *The geometric vein*, ed. by C. Davis, B. Grünbaum, F. A. Sherk, Springer-Verlag, New York—Berlin, 1981, 253—269. *MR* **84h**: 01059; for the volume: *MR* **83e**: 51003.

[2] KLINE, M., *Mathematics in western culture*, Oxford University Press, New York, 1964. *MR* **31** #3296.

[3] YAGLOM, I. M., *Mathematical structures and mathematical modelling*, Gordon and Breach, New York, 1985.

[4] KUHN, T. S., *The structure of scientific revolutions*, University of Chicago Press, Chicago, 1970.

[5] KLINE, M., *Mathematics. The loss of certainty*, Oxford University Press, New York, 1980. *MR* **82e**: 03013.

[6] BOURBAKI, N., *Éléments d'histoire des mathématiques*, Histoire de la pensée, IV, Hermann, Paris, 1960. *MR* **22** #4620.

[7] DIEUDONNÉ, J., *Éléments d'analyse*, Gauthier-Villars, Paris.
Tome 1: Fondements de l'analyse moderne, Cahiers scientifiques, XXVIII, 1963, 1968. *MR* **28** #5149; **38** #4246.
Tome 2: Chapitres XII à XV, Cahiers scientifiques, XXXI, 1968. *MR* **38** #4247.
Tome 3: Chapitres XVI et XVII, Cahiers scientifiques, XXXIII, 1970, *MR* **42** #5266.
Tome 4: Chapitres XVIII à XX, Cahiers scientifiques, XXXIV, 1971, 1977. *MR* **50** #14507; **57** #7632.
Tome 5: Chapitre XXI, Cahiers scientifiques, XXXVIII, 1975. *MR* **57** #7633.
Tome 6: Chapitre XXII, Cahiers scientifiques, XXXIX, 1975. *MR* **58** #29825a.
Tome 7: Chapitre XXIII, Première partie, Cahiers scientifiques, XL, 1978. *MR* **58** #13103a.
Tome 8: Chapitre XXIII, Deuxième partie, Cahiers scientifiques, XLI, 1978. *MR* **58** #13103b.
Tome 9: Chapitre XXIV, Cahiers scientifiques, XLII, 1982.

[8] RIEMANN, B., Über die Hypothesen, welche der Geometrie zu Grunde liegen, *Gesammelte mathematische Werke und Wissenschaftlicher Nachlass*, Dover Publications, Inc., New York, 1953, 272—287 and 391—404. *MR* **14**—610.

[9] SCHOUTEN, J. A., *Der Ricci-Kalkül*, Springer-Verlag, Berlin, 1924. English translation: *Ricci calculus*. An introduction to tensor analysis and its geometrical applications, 2nd

edition, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen usw., Bd. 10, Springer-Verlag, Berlin—Göttingen—Heidelberg, 1954. *MR* **16**—521.

[10] SCHOUTEN, J. A. and STRUIK, D. J., *Einführung in die neueren Methoden der Differentialgeometrie*, Band I, II, Springer-Verlag, Berlin, 1934, 1938.

[11] BAER, R., *Linear algebra and projective geometry*, Academic Press, Inc., New York, 1952. *MR* **14**—675.

[12] DIEUDONNÉ, J., *Algèbre linéaire et géométrie élémentaire*, Enseignement des sciences, VIII, Hermann, Paris, 1964. *MR* 30 # 2015.

[13] BACHMANN, F., *Aufbau der Geometrie aus dem Spiegelungsbegriff*, Zweite ergänzte Auflage, Die Grundlehren der mathematischen Wissenschaften, Band 96, Springer-Verlag, Berlin—New York, 1973. *MR* **49** # 11368.

[14] PRENOWITZ, W. and JANTOSCIAK, J., *Join geometries*. A theory of convex sets and linear geometry, Undergraduate texts in mathematics, Springer-Verlag, New York—Heidelberg, 1979. *MR* **80c**: 52001.

[15] FEINBERG, E. L., *Art in the science dominated world*, Gordon and Breach, New York, 1986.

[16] SPRINGER, S. P. and DEUTSCH, G., *Left brain, right brain*, Freeman, San Francisco, 1981.

[17] YAGLOM, I. M., Algebra und Geometrie als alternative Sprachen mathematischen Denkens, *Mitteilungen aus dem Math. Seminar Giessen*, Heft 166 (Coxeter Festschrift), Teil IV, 1984, 197—204.

[18] WEYL, H., Topologie und abstrakte Algebra als zwei Wege mathematischen Verständnisses, *Gesammelte Abhandlungen*, Band 3, Springer-Verlag, Berlin, 1968, 348—358. *MR* **37** # 6157.

[19] LANDAU, E., *Grundlagen der Analysis; Einführung in die Differentialrechnung und Integralrechnung*, Springer, Berlin, 1930, 1934.

[20] ЯГЛОМ, И. М., Почему высшую математику открыли одновременно Ньютон и Лейбниц? *Число и мысль*, вып. 6, Знание, Москва, 1983, 99—125.

[21] ROBINSON, A., *Non-standard analysis*, North-Holland, Amsterdam—New York, 1974.

[22] FEJES TÓTH, L., *Lagerungen in der Ebene, auf der Kugel und im Raum*, Zweite verbesserte und erweiterte Auflage, Die Grundlehren der mathematischen Wissenschaften, Band 65, Springer-Verlag, Berlin—New York, 1972. *MR* **50** # 5603.

[23] ROGERS, C. A., *Packing and covering*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 54, Cambridge University Press, New York, 1964. *MR* 30 # 2405.

[24] HADWIGER, H. and DEBRUNNER, H., *Combinatorial geometry in the plane*, With a new chapter and other additional material supplied by V. Klee, Holt, Rinehart and Winston, New York, 1964. *MR* **29** # 1577.

[25] BOLTJANSKI, W. G. und GOHBERG, I. Z., *Sätze und Probleme der kombinatorischen Geometrie*, Kleine Ergänzungsreihe zu den Hochschulbüchern für Mathematik, XXV; Mathematische Schülerbücherei, No. 61, VEB Deutscher Verlag der Wissenschaften, Berlin 1972. *MR* **50** # 3108.

[26] FEJES TÓTH, L., Remarks on a theorem of R. M. Robinson, *Studia Sci. Math. Hungar.* **4** (1969), 441—445. *MR* **40** # 7951.

[27] BÖRÖCZKY, K., Über die Newtonsche Zahl regulärer Vielecke, *Period. Math. Hungar.* **1** (1971), 113—119. *MR* **44** # 4644.

[28] LINHART, J., Die Newtonsche Zahl von regelmässigen Fünfecken, *Period. Math. Hungar.* **4** (1973), 315—328. *MR* **50** # 3111.

[29] ПАНКОВ, П. С., *Доказательные вычисления на электронных вычислительных машинах*, ИЛИМ, Фрунзе, 1978. *RŽ Mat.* 1978: 9B694; *MR* **82a**: 65003.

[30] SHANNON, C. E., Communication in the presence of noise, *Proc. I.R.E.* **37** (1949), 10—21. *MR* **10**—21.

[31] SLOANE, N. J. A., Sphere packing, *Scientific American* **250** (1984), no. 1.

[32] GRÜNBAUM, B., *Convex polytopes*, Pure and applied mathematics, Vol. 16, Interscience Publishers [John Wiley and Sons, Inc.], New York, 1967. *MR* **37** # 2085.

[33] WEYL, H., *Symmetry*, Princeton University Press, Princeton, N. J., 1952. *MR* **14**—16.

[34] THOMPSON, D'ARCY W., *On growth and form*, New edition, Cambridge University Press, Cambridge, 1942. *MR* 3—291.

[35] FEJES TÓTH, L., *Regular figures*, A Pergamon Press book, The Macmillan Co., New York, 1964. *MR* **29** # 2705.

[36] *Patterns of symmetry*, ed. by M. Senechal and G. Fleck, University of Massachusetts Press, Amherst, Mass., 1977. *MR* **58** # 5.

[37] Gardner, M., *The ambidextrous universe*, Basic Books, New York, 1964.
[38] Holden, A., *Shapes, space, and symmetry*, Columbia University Press, New York, 1971.
[39] MacGillavry, C. H., *Symmetry aspects of M. C. Escher's periodic drawings*, Holkema, Bohn, Scheltema and Utrecht, 1976.
[40] Яглом, И. М., *Герман Вейль*, Знание, Москва, 1967.
[41] Grünbaum, B. and Shephard, G. K., *Tilings and patterns*, Freeman, San Francisco, 1987.
[42] *Kinetic art: Theory and practice*, ed. by Frank J. Malina, Dover, New York, 1974; *Visual art, mathematics and computers*, ed. by Frank J. Malina, Pergamon, New York, 1979.
[43] Mandelbrot, B., *The fractal geometry of nature*, W. H. Freeman and Co., San Francisco, Calif., 1982. *MR* **84h**: 00021.
[44] Falconer, K. J., *The geometry of fractal sets*, Cambridge University Press, Cambridge, 1985.
[45] Bool, F. H., Ernst, B., Kist, J. R., Locher, J. L. and Wierda, F., *Escher*, General editor: J. L. Locher, Abrams, New York, 1982.
[46] Ernst, B., *The magic mirror of M. C. Escher*, Ballantine Books, New York, 1976.
[47] *M. C. Escher: Art and science*, ed. by H.S.M. Coxeter et al., North-Holland, Amsterdam, 1986.
[48] Fuller, R. B., *Synergetics*, Macmillan Co., New York, 1975.
[49] Coxeter, H. S. M., *Introduction to geometry*, Wiley, New York—London, 1961. *MR* **23** #A1251; Second edition, 1969. *MR* **49** #11369.
[50] Berger, M., *Géométrie*, CEDIC, Paris; Nathan Information, Paris, 1977. Vol. 1. Actions de groupes, espaces affines et projectifs; Vol. 2. Espaces euclidiens, triangles, cercles et sphères; Vol. 3. Convexes et polytopes, polyèdres réguliers, aires et volumes; Vol. 4. Formes quadratiques, coniques et quadriques; Vol. 5. La sphère pour elle-même, géométrie hyperbolique, l'espace des sphères. *MR* **81k**: 51001a—51001e.
[51] *The geometric vein*, The Coxeter Festschrift, ed. by C. Davis, B. Grünbaum and F. A. Sherk, Springer-Verlag, New York—Berlin, 1981. *MR* **83e**: 51003.
[52] *The mathematical Gardner*, ed. by David A. Klarner, Wadsworth International, Belmont, Calif; Prindle, Weber and Schmidt, Boston, Mass., 1981. *MR* **82b**: 00003.
[53] Ball, W. W. R. and Coxeter, H. S. M., *Mathematical recreations and essays*, Twelfth edition, University of Toronto Press, Toronto, Ont., 1974. *MR* **50** #4229.

(Поступило 1-ого июня 1986 г.)

1—Й ГОНЧАРНЫЙ ПЕР., 7, КВ. 17
SU—109 172 МОСКВА Ж-172
SOVIET UNION

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 149—167

STOPPING RULES FOR STOCHASTIC APPROXIMATION METHODS

GÁBOR HANÁK

Introduction

Let $(X, \| \ \|)$ be a normed space, x_n be random variables with elements in X and $\Theta_n \in X$ for $n=1, 2, \dots$. The sequence (x_n) is assumed to be an approximation of the sequence (Θ_n) in the sense that $(x_n-\Theta_n)$ tends to zero almost surely (or in probability). Our aim is to find a stopping rule such that the stopped sequence of the rv's (or a function of them) should be near to the stopped sequence of Θ_n's.

We are going to deal with two methods, first the method of the independent copies and second the sequential method. The essence of the former one consists of taking independent copies of the sequence of rv's, i.e. taking $x_n^{(i)}$ $i=1, 2, \dots, k$ where $x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)}$ are iid rv's for each $n=1, 2, \dots$; and stopping the sequence when the copies are near to each other. The essence of the other method is to stop the sequence when the k last member of it are near to each other. The method of the independent copies was suggested by R. Zieliński ([Z1]). This paper is an extension and continuation of his results. First we deal with the method of the independent copies and give certain extensions of Zieliński's original results. Then we state appropriate results for the sequential method.

The advantage of the method of the independent copies opposite to the other is the short time between the beginning of the observation and the stopping. Furthermore some problems proved to be insoluble using stopping times with respect to the σ-algebra generated by the sequence itself ([BR]) but solvable using this method ([Z2]). However, the method has the drawback of the necessity of the great number of observations.

Now let us formalize our problem. Let $k \in \mathrm{N}$ and denote $x_n^{(i)}$ $i=1, 2, \dots, k$ the ith of the k independent realizations of x_n, i.e. $x_n^{(i)}$ $i=1, 2, \dots, k$ are iid rv's with the distribution of x_n $(n=1, 2, \dots)$. Denote furthermore

(I.1) $$p_n(\delta) = \mathsf{P}(\|x_n^{(1)} - x_n^{(2)}\| < \delta) \quad \delta > 0, \quad n = 1, 2, \dots$$

(I.2) $$p_n = \mathsf{P}(x_n^{(1)} = x_n^{(2)}) = \lim_{\delta\to 0} p_n(\delta), \quad n = 1, 2, \dots$$

(I.3) $$p = \sup_n p_n$$

(I.4) $$q_n(\delta) = \sup_{x\in X} \mathsf{P}(\|x_n - x\| < \delta) \quad \delta > 0, \quad n = 1, 2, \dots$$

(I.5) $$q_n = \lim_{\delta\to 0} q_n(\delta) \quad n = 1, 2, \dots$$

(I.6) $$q = \sup_n q_n.$$

1980 *Mathematics Subject Classification.* Primary 60G40; Secondary 62L20.
Key words and phrases. Stopping times, stochastic approximation, weak dependence.

Assume that there exists an $r_k: X^k \to \mathbf{R}_+$ measurable mapping for each $k \geqq 2$ with the following property

$$\exists c_1, c_2 > 0:$$

$$c_1 \max_{1 \leqq i, j \leqq k} \|a^{(i)} - a^{(j)}\| \leqq r_k(a^{(1)}, \ldots, a^{(k)}) \leqq c_2 \max_{1 \leqq i, j \leqq k} \|a^{(i)} - a^{(j)}\| \tag{I.7}$$

$$\text{for every } (a^{(1)}, \ldots, a^{(k)}) \in X^k.$$

This r_k might be imagined, for example, as the diameter of the convex hull of k points; it will measure whether the nth copies are near to each other.

In what follows we will try to approximate the parameter sequence by a function of our observations, i.e. we assume we have a $g_k: X^k \to X$ measurable mapping for each $k \geqq 2$ with the following property

$$\exists c > 0:$$

$$\|g_k(a^{(1)}, \ldots, a^{(k)}) - a\| \leqq c \max_{1 \leqq i \leqq k} \|a^{(i)} - a\| \tag{I.8}$$

$$\text{for every } (a^{(1)}, \ldots, a^{(k)}, a) \in X^{k+1},$$

and we get $\hat{\Theta}_n = g_k(x_n^{(1)}, \ldots, x_n^{(k)})$ $(n = 1, 2, \ldots)$. For example we can choose

$$g_k(a^{(1)}, \ldots, a^{(k)}) = \frac{1}{k} \sum_{i=1}^{k} a^{(i)} \quad \text{(the sample mean)}$$

or

$$g_k(a^{(1)}, \ldots, a^{(k)}) = a^{(1)} \quad \text{(projection)}.$$

An attempt to make an effort for giving a solution of the problem stated above was made by R. Zieliński in [Z1]. He suggested the method of the independent copies and that it should be stopped whenever

$$r_k(x_n^{(1)}, \ldots, x_n^{(k)}) < \gamma$$

for the first time. More precisely he took

$$t = \inf \{n \geqq 1: \ r_k(x_n^{(1)}, \ldots, x_n^{(k)}) < \gamma\}, \quad \gamma > 0, \tag{I.9}$$

which is clearly a stopping time with respect to the sequence of σ-algebras $(\mathcal{F}_n)_1^\infty$ where

$$\mathcal{F}_n = \sigma\{x_1^{(1)}, \ldots, x_1^{(k)}, \ldots, x_n^{(1)}, \ldots, x_n^{(k)}\} \qquad n = 1, 2, \ldots.$$

Naturally, $t = t(k, \gamma)$.

He proved the following

Lemma. *If for each positive* η *the inequality* $\sum_{n=1}^{\infty} \mathsf{P}(\|x_n - \Theta_n\| \geqq \eta) < \infty$ *holds then* t *is an a.s. finite stopping time (a stopping rule), i.e.* $\mathsf{P}(t < \infty) = 1$;

Theorem A. *If for each positive* η *the inequality* $\sum_{n=1}^{\infty} \mathsf{P}(\|x_n - \Theta_n\| \geqq \eta) < \infty$ *holds and* $q = 0$ *then*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \forall k \geqq 2 \quad \exists \gamma > 0: \ \mathsf{P}(\|\hat{\Theta}_t - \Theta_t\| > \delta) \leqq \varepsilon;$$

THEOREM B. *If for each positive η there exists a positive β such that the inequality* $\sum_{m=n}^{\infty} \mathsf{P}(\|x_m-\Theta_m\|>\eta)=O(n^{-\beta})$ *holds and $q<1$ then*

$$\forall\varepsilon>0 \quad \forall\delta>0 \quad \exists k \geqq 2 \quad \exists\gamma>0\colon \mathsf{P}(\|\hat{\Theta}_t-\Theta_t\|>\delta) \leqq \varepsilon.$$

Note that in the results above the Θ_n's are assumed to be constant.

Our results appear in Chapter 1 (independent copies) and Chapter 2 (sequential method).

Theorems mentioned above can be interpreted as the statement that t is an ε-optimal stopping rule when we are looking for a stopping time that minimizes $\mathsf{P}(\|\hat{\Theta}_w-\Theta_w\|>\delta)$ (w is a stopping time. It is trivial that in the case of $\lim_n (x_n-\Theta_n)=0$ a.s., the stopping time which has the only value of $+\infty$ minimizes this set of probabilities since $\mathsf{P}(\|\hat{\Theta}_\infty-\Theta_\infty\|>\delta)=0$). We can also comprehend our goal that for given ε and δ we look for a stopping rule (and t is just like this) which yields a confidence interval of length 2δ and of level at least $1-\varepsilon$.

We consider the ε-optimality problem in this way and find weaker conditions than Zieliński's ones in order to guarantee that t is ε-optimal. We also deal with the interpretation of the ε-optimality when we want to minimize $\mathsf{E}\|\hat{\Theta}_w-\Theta_w\|$ and prove that certain conditions imply the ε-optimality of t.

Similar considerations are performed in the sequential method case.

0. Preliminaries

The goal of this chapter is to prepare ourselves, to make clear the essence of and the relations among different conditions. We will use the $*$-assumption:

($*$) $(X, \|\ \|)$ has finite dimension.

Introduce the following notations with a (y_n) nonnegative sequence of real rv's (and imagine y_n to be $\|x_n-\Theta_n\|$):

$$Q_n(y)=\sup_{m>n}\mathsf{P}(y_m>y) \qquad n=0,1,\ldots;$$

$$Q(x)=\int_x^\infty Q_0(y)\,dy \qquad x\geqq 0;$$

$$R_n(y)=\mathsf{P}(\sup_{m>n} y_m>y) \qquad n=0,1,\ldots;$$

$$R(x)=\int_x^\infty R_0(y)\,dy \qquad x\geqq 0;$$

$$S_n(y)=\sum_{m>n}\mathsf{P}(y_m>y) \qquad n=0,1,\ldots;$$

$$S(x)=\int_x^\infty S_0(y)\,dy \qquad x\geqq 0.$$

Obviously, $Q_n(y) \leqq R_n(y) \leqq S_n(y)$ $(n=0, 1, \dots$ and $y \geqq 0)$; $R_0(0) \leqq 1$; $Q(x) \leqq R(x) \leqq$ $\leqq S(x)$ $(x \geqq 0)$; all the functions are monotone decreasing in the argument at fixed index and in the index at fixed argument.

Now consider the following assumptions (underlined which are used by Zieliński)

(A) $$\lim_n y_n = 0 \quad \text{in probability};$$

(B) $$\forall y > 0\colon \lim_n Q_n(y) = 0;$$

(C) $$\lim_n y_n = 0 \quad \text{a.s.};$$

(D) $$\forall y > 0\colon \lim_n R_n(y) = 0;$$

$(\underline{\mathrm{E}})$ $$\forall y > 0\colon\ S_0(y) < \infty;$$

(F) $$\forall y > 0 \quad \exists \beta > 0\colon R_n(y) = O(n^{-\beta});$$

$(\underline{\mathrm{G}})$ $$\forall y > 0 \quad \exists \beta > 0\colon S_n(y) = O(n^{-\beta});$$

(H) $$(y_n) \quad \text{is uniformly integrable};$$

(I) $$Q(0) < \infty;$$

(J) $$\mathsf{E}(\sup_n y_n) < \infty;$$

(K) $$R(0) < \infty;$$

(L) $$\sum_{n=1}^{\infty} \mathsf{E} y_n < \infty;$$

(M) $$S(0) < \infty.$$

PROPOSITION 1.

$$\begin{array}{ccccccccccccc} \mathrm{M}) & \Leftrightarrow & (\mathrm{L}) & \Rightarrow & (\mathrm{K}) & \Leftrightarrow & (\mathrm{J}) & \Rightarrow & (\mathrm{I}) & \Rightarrow & (\mathrm{H}) & & \\ & & \Downarrow & & & & & & & & & & \\ (\underline{\mathrm{G}}) & \Rightarrow & (\underline{\mathrm{E}}) & \Rightarrow & (\mathrm{D}) & \Leftrightarrow & (\mathrm{C}) & \Rightarrow & (\mathrm{B}) & \Leftrightarrow & (\mathrm{A}) & & \\ & \searrow & & \nearrow & & & & & & & & & \\ & & (\mathrm{F}) & & & & & & & & & & \end{array}$$ □

Take into consideration the definitions (I.1)—(I.6) and let

(a) $$p_n < 1 \qquad n = 1, 2, \dots;$$

(b) $$\lim_m \sum_{n=1}^{m} p_n^{L/R_m(\delta)} = 0 \qquad \forall L > 0, \quad \forall \delta > 0;$$

(c) $$\lim_m m p^{L/R_m(\delta)} = 0 \qquad \forall L > 0, \quad \forall \delta > 0;$$

(d) $$p < 1;$$

$(\underline{\mathrm{e}})$ $$q < 1;$$

(f) $$p = 0;$$

$(\underline{\mathrm{g}})$ $$q = 0.$$

PROPOSITION 2.

$$
\begin{tikzcd}
(g) \arrow[r, Rightarrow, shift right] \arrow[dr, Rightarrow] & (f) \arrow[l, Rightarrow, shift right, "(*)"'] \arrow[d, Rightarrow, "(*)"] \arrow[dr, Rightarrow] & & (c) \arrow[dr, Rightarrow] & (b) \arrow[d, Rightarrow] \\
& (e) \arrow[r, Rightarrow, shift right] & (d) \arrow[l, Rightarrow, shift right, "(*)"'] \arrow[ur, Rightarrow, "(F)"] \arrow[rr, Rightarrow] & & (a)
\end{tikzcd}
$$

(Marked inductions are true only when the appropriate condition holds.) □

Before proving these propositions we state and prove some auxiliary ones.

PROPOSITION 3. *Let* ξ *be a nonnegative rv and* a *be an arbitrary nonnegative real number. Then*

$$\int_{\xi>a} \xi \, d\mathsf{P} = \int_a^{\infty} \mathsf{P}(\xi > s)\, ds + a\mathsf{P}(\xi > a).$$

In particular, $\mathsf{E}\xi = \int_0^{\infty} \mathsf{P}(\xi > a)\, da.$ □

PROOF.

$$\int_{\Omega} \xi \, d\mathsf{P} = \int_{\Omega} I(\xi > a)\xi \, d\mathsf{P} = \int_{\Omega} I(\xi > a) \cdot \int_0^{\xi} 1 \, ds \, d\mathsf{P} =$$

$$= \int_a^{\infty} \int_{\Omega} I(\xi > s)\, d\mathsf{P}\, ds + \int_0^{a} \int_{\Omega} I(\xi > a)\, d\mathsf{P}\, ds =$$

$$= \int_a^{\infty} \mathsf{P}(\xi > s)\, ds + a\mathsf{P}(\xi > a). \quad \square\square\square$$

PROPOSITION 4. *Let* f *and* g *be nonnegative integrable real functions on* $[0, \infty)$ *and*

$$\int_0^{\infty} f(s)\, ds < \infty, \quad \int_0^{\infty} g(s)\, ds = \infty.$$

Then $\liminf_{s\to\infty} \dfrac{f(s)}{g(s)} = 0.$ □

PROOF. Supposing the contrary evidently leads to a contradiction with the assumptions. □ □ □

Now let x and z be iid rv's with elements in X and

$$\bar{p} = \mathsf{P}(x = z);$$

$$\bar{q}(\delta) = \sup_{a\in X} \mathsf{P}(\|x-a\| < \delta) \qquad \delta > 0;$$

$$\bar{q} = \lim_{\delta\to 0} \bar{q}(\delta).$$

PROPOSITION 5.

$$\bar{p} > 0 \Leftrightarrow \exists\,(a_n)_{n=1}^{\infty} \subset X\colon\ \mathsf{P}(x = a_n) = \bar{p}_n > 0$$

and

$$\forall\, a \in X (a \neq a_n \forall\, n \in \mathbf{N})\colon\ \mathsf{P}(x = a) = 0.$$

In this case $\bar{p} = \sum_{n=1}^{\infty} \bar{p}_n^2$. □

PROPOSITION 6. $\bar{p}=0 \Leftrightarrow x$ *has continuous distribution.* □

PROPOSITION 7. $\bar{q}=0 \Leftrightarrow x$ *has uniformly continuous distribution.* □

PROPOSITION 8. $\bar{p}=1 \Leftrightarrow x$ *is a.s. constant.* □

PROPOSITION 9. $\bar{p} \leqq \bar{q}$. □

PROPOSITION 10. *If* $(*)$ *then* $\bar{p} \geqq \bar{q}^2$. □

PROOF OF PROPOSITION 5. $\bar{p} = \mathsf{P}(x=z) = \int\limits_X \int\limits_X I(u=v)\, dT(u)\, dT(v)$ where I stands for the indicator function of a set in X and T is the distribution of x. The inner integral differs from zero if and only if at a definit point $a \in X$ $T(a) > 0$. It is well-known that there can be only countable number of points in X with this property. Let $(a_n)_{n=1}^{\infty}$ be all the points satisfying $T(a_n) \overset{\text{def}}{=} \bar{p}_n > 0$. Then $\bar{p} = \sum_{n=1}^{\infty} \int\limits_{v=a_n} \bar{p}_n\, dT(v) = \sum_{n=1}^{\infty} \bar{p}_n^2$. □ □ □

PROOF OF PROPOSITION 6. An obvious consequence of Proposition 5. □ □ □

PROOF OF PROPOSITION 7. An equivalent formulation of the definition of uniform continuity. □ □ □

PROOF OF PROPOSITION 8. If x is a.s. constant $\bar{p}$ is equal to 1 trivially. On the other hand if $\bar{p}=1$ then by Proposition 5 there exists an $a \in X$ such that $\mathsf{P}(x=a) > 0$. But $\mathsf{P}(x=a) = \mathsf{P}(x=a=z) = \mathsf{P}^2(x=a)$, thus $\mathsf{P}(x=a)=1$. □ □ □

PROOF OF PROPOSITION 9. If $\bar{p}=0$ the inequality automatically holds. Now let $\bar{p}>0$, and $(a_n)_{n=1}^{\infty} \subset X$ be the sequence mentioned in Proposition 5. Obviously $\sum_{n=1}^{\infty} \bar{p}_n \leqq 1$ thus there exists an $N \in \mathbf{N}$ such that $\bar{p}_N = \sup_n \bar{p}_n$. By the definition of $\bar{q}$ we have $\bar{p}_N \leqq \bar{q}$. Multiplying the two inequalities we get $\bar{q} \geqq \bar{p}_N \sum_{n=1}^{\infty} \bar{p}_n \geqq \sum_{n=1}^{\infty} \bar{p}_n^2 = \bar{p}$ (the last equality holds by Proposition 5). □ □ □

PROOF OF PROPOSITION 10. If $\bar{q}=0$ the inequality holds automatically. Now let $\bar{q}>0$. By the definition of $\bar{q}$ for every natural number n there exists a sequence $(a_m(n))_{m=1}^{\infty} \subset X$ such that

$$\mathsf{P}\left(\|x - a_m(n)\| < \frac{1}{2n}\right) \geqq \bar{q} - \frac{1}{m}$$

$(a_m(n))_{m=1}^{\infty}$ is a bounded sequence since, by $\bar{q}>0$, there may be only finite number of disjoint subsets $\left\{\|x-a_m(n)\|<\frac{1}{2n}\right\}$. Since X has finite dimension we can choose a subsequence $(a_{m_k}(n))_{k=1}^{\infty}\subset X$ and a point $a_n\in X$ such that $\lim\limits_{k\to\infty} a_{m_k}(n)=a_n$. Let k be large enough that $\|a_n-a_{m_k}(n)\|<\frac{1}{2n}$. Then $\mathsf{P}\left(\|x-a_n\|<\frac{1}{n}\right)\geqq$ $\geqq\mathsf{P}\left(\|x-a_{m_k}(n)\|<\frac{1}{n}-\|a_n-a_{m_k}(n)\|\right)\geqq\mathsf{P}\left(\|x-a_{m_k}(n)\|<\frac{1}{2n}\right)\geqq\bar{q}-\frac{1}{m_k}$. Thus $\mathsf{P}\left(\|x-a_n\|<\frac{1}{n}\right)\geqq\bar{q}$.

By a similar way of arguments $(a_n)_{n=1}^{\infty}\subset X$ is a bounded sequence so we can choose an $(a_{n_i})_{i=1}^{\infty}\subset X$ subsequence of it and an $a\in X$ point such that $\lim\limits_{i\to\infty} a_{n_i}=a$ and we can prove $\mathsf{P}(\|x-a\|<\delta)\geqq\bar{q}$ for every $\delta>0$. Thus $\mathsf{P}(x=a)\geqq\bar{q}$ and we have

$$\bar{p}=\mathsf{P}(x=z)\geqq\mathsf{P}(x=z=a)=\mathsf{P}(x=a)\,\mathsf{P}(z=a)\geqq\bar{q}^2. \qquad \Box\Box\Box$$

Corollary. *If* $(*)$ *holds then*

(1) $\bar{p}=0\Leftrightarrow\bar{q}=0\Leftrightarrow x$ *has continuous distribution*;
(2) $\bar{p}=1\Leftrightarrow\bar{q}=1\Leftrightarrow x$ *is a.s. constant*;
(3) $\bar{q}^2\leqq\bar{p}\leqq\bar{q}$. $\Box\Box\Box$

Now we turn to the proof of the main propositions of this section.

Proof of Proposition 1.

(M)$\Leftrightarrow$(L) Proposition 3
(M)$\Rightarrow$(K) Trivial
(K)$\Leftrightarrow$(J) Proposition 3
(K)$\Rightarrow$(I) Trivial
(I)$\Rightarrow$(H) Since $\sup\limits_n \int\limits_{y_n>a} y_n\,d\mathsf{P}$ is monotone in a we can use the result of Proposition 4 if we divide, using Proposition 3, the expression into two parts:

$$\lim_{a\to\infty}\sup_n \int\limits_{y_n>a} y_n\,d\mathsf{P}=\lim_{a\to\infty}\sup_n\Big[\int\limits_a^{\infty}\mathsf{P}(y_n>u)\,du+a\mathsf{P}(y_n>a)\Big]\leqq$$

$$\leqq\liminf_{a\to\infty}\Big[\int\limits_a^{\infty}\sup_n\mathsf{P}(y_n>u)\,du+a\sup_n\mathsf{P}(y_n>a)\Big]=0.$$

(G)$\Rightarrow$(F) Trivial
(G)$\Rightarrow$(E) Trivial
(F)$\Rightarrow$(D) Trivial
(L)$\Rightarrow$(E) Trivial
(E)$\Rightarrow$(D) Trivial
(D)$\Leftrightarrow$(C) Well-known
(D)$\Rightarrow$(B) Trivial
(B)$\Leftrightarrow$(A) Trivial. $\Box\Box\Box$

PROOF OF PROPOSITION 2.

$(\underline{g}) \Rightarrow (f)$	Proposition 9
$(f) \overset{(*)}{\Rightarrow} (\underline{g})$	Proposition 10
$(\underline{g}) \Rightarrow (\underline{e})$	Trivial
$(f) \overset{(*)}{\Rightarrow} (\underline{e})$	Consequence of the two above
$(f) \Rightarrow (d)$	Trivial
$(\underline{e}) \Rightarrow (d)$	Proposition 9
$(d) \overset{(*)}{\Rightarrow} (\underline{e})$	Proposition 10
$(d) \overset{(F)}{\Rightarrow} (c)$	Follows from the different speed of the power and the exponential function.
$(d) \Rightarrow (a)$	Trivial
$(c) \Rightarrow (a)$	Trivial
$(b) \Rightarrow (a)$	Trivial. □ □ □

NOTE 1. If (A) holds, i.e. (y_n) converges to zero in probability then for every $\varepsilon > 0$ and for every nonnegative monotone decreasing real function f with the property $\lim_{x \to \infty} f(x) = 0$ we can choose a subsequence (y'_n) of (y_n) (depending on ε and f) such that

$$S'_n(\varepsilon) = \sum_{m>n} \mathsf{P}(y'_m > \varepsilon) = o(f(n))$$

so that $(\underline{G}')$ holds.

NOTE 2. If (A) holds, i.e. (y_n) converges to zero in probability then (H) is equivalent to the L_1-convergence of the sequence.

NOTE 3. If (y_n) converges to zero in L_1 then for every nonnegative monotone decreasing real function f with the property $\lim_{x \to \infty} f(x) = 0$ we can choose a subsequence (y'_n) of (y_n) (depending on f) such that

$$\sum_{m>n} \mathsf{E} y'_m = o(f(n))$$

so that (L') holds.

NOTE 4. (H)$\not\Rightarrow$(I). Let the probability space be the [0, 1] with the Borel sets and the Lebesgue measure and let

$$y_n(x) = \begin{cases} \dfrac{n}{\log n}, & \text{if } \ x \leqq \dfrac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then (y_n) tends to zero a.s. and in L_1, too, and it is a uniformly integrable sequence of rv's. But (I) does not hold.

The method of the independent copies

Now we turn to the examination of the stopping time t defined in the Introduction ((I. 9)). The first two lemmas state that t is a.s. finite and has finite expectation, respectively. Then the flow of our argument branches as we indicated at the end of the Introduction. First we state some theorems related to the probability ε-optimality then the appropriate ones related to the expectation ε-optimality.

LEMMA 1.1. *If* $(x_n - \Theta_n)$ *tends to zero in probability* ((A)) *then* t *is a.s. finite, more precisely* $\forall k \geqq 2\ \forall \gamma > 0: \mathsf{P}(t < \infty) = 1$. □

LEMMA 1.2. *If* (E) *holds then* t *has finite expectation, more precisely* $\forall k \geqq 2$ $\forall \gamma > 0: \mathsf{E}t < \infty$. □

PROOF OF LEMMA 1.1.

$$\mathsf{P}(t=\infty) = \mathsf{P}\Big(\bigcap_{n=1}^{\infty}\{r_k(x_n^{(1)}, \ldots, x_n^{(k)}) \geqq \gamma\}\Big) \leqq$$
$$\leqq \inf_n \mathsf{P}\big(r_k(x_n^{(1)}, \ldots, x_n^{(k)}) \geqq \gamma\big) \leqq$$
$$\leqq \inf_n \mathsf{P}\big(\max_{1\leqq i,j\leqq k} \|x_n^{(i)} - x_n^{(j)}\| \geqq \gamma/c_1\big) \leqq$$
$$\leqq k^2 \inf_n \mathsf{P}(\|x_n^{(1)} - x_n^{(2)}\|) \geqq \gamma/c_1) \leqq$$
$$\leqq 2k^2 \inf_n \mathsf{P}(\|x_n - \Theta_n\| \geqq \gamma/2c_1) = 0. \quad \square\square\square$$

PROOF OF LEMMA 1.2.

$$\mathsf{E}t = 1 + \sum_{n=1}^{\infty} \mathsf{P}(t > n) \leqq 1 + \sum_{n=1}^{\infty} \mathsf{P}\big(r_k(x_n^{(1)}, \ldots, x_n^{(k)}) \geqq \gamma\big) \leqq$$

(as we have seen just before)

$$\leqq 1 + 2k^2 \sum_{n=1}^{\infty} \mathsf{P}(\|x_n - \Theta_n\| \geqq \gamma/2c_1). \quad \square\square\square$$

Now we are going to formulate three theorems. All of them set up confidence intervals of length 2δ and of level at least $1-\varepsilon$.

THEOREM 1.1.A. *If* $(x_n - \Theta_n)$ *tends to zero a.s.* ((C)) *and each member of the sequence of rv's has continuous distribution* ((f)) *then* t *is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \forall k \geqq 2 \quad \exists \gamma > 0: \ \mathsf{P}(\|\hat{\Theta}_t - \Theta_t\| > \delta) \leqq \varepsilon. \quad \square$$

If we give up the requirement of the continuous distributions we cannot guarantee that t is ε-optimal for small k's:

THEOREM 1.1.B. *If* $(x_n - \Theta_n)$ *tends to zero a.s.* ((C)) *and* (b) *holds then* t *is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists k \geqq 2 \quad \exists \gamma > 0: \ \mathsf{P}(\|\hat{\Theta}_t - \Theta_t\| > \delta) \leqq \varepsilon. \quad \square$$

THEOREM 1.1.C. *If the functions g_k are projections and furthermore if $(x_n - \Theta_n)$ tends to zero ((C)) and none of the rv's is a.s. constant ((a)) then t is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists k \geqq 2 \quad \exists \gamma > 0\colon \mathsf{P}(\|x_t - \Theta_t\| > \delta) \leqq \varepsilon. \quad \square$$

The next three theorems concern the ε-optimality with respect to the expectation. They have some additional conditions compared with the appropriate theorems stated above.

THEOREM 1.2.A. *Assume the conditions of Theorem 1.1.A* ((C) *and* (f)) *and furthermore* (J). *Then t is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \forall k \geqq 2 \quad \exists \gamma > 0\colon E\|\hat{\Theta}_t - \Theta_t\| \leqq \varepsilon. \quad \square$$

THEOREM 1.2.B. *Assume*

(i) (F),

(ii) $\forall y > 0 \quad \exists \alpha > 0\colon n^{-\alpha} = O(R_n(y))$.

(iii) (J),

(iv) $R(x) = o(x^{-\alpha - 2(\alpha/\beta) + 1})$,

(v) (d).

Then t is ε-optimal in the sense

$$\forall \varepsilon > 0 \quad \exists k \geqq 2 \quad \exists \gamma > 0\colon \mathsf{E}\|\hat{\Theta}_t - \Theta_t\| \leqq \varepsilon. \quad \square$$

THEOREM 1.2.C. *If the functions g_k are projections and furthermore if $(x_n - \Theta_n)$ tends to zero a.s.* ((C)), (J) *holds and none of the rv's is a.s. constant* ((a)) *then t is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \exists k \geqq 2 \quad \exists \gamma > 0\colon \mathsf{E}\|x_t - \Theta_t\| \leqq \varepsilon. \quad \square$$

PROOF OF THEOREM 1.1.A. Let $\varepsilon > 0$, $\delta > 0$ and $k \geqq 2$ be arbitrary and let

$$N = \inf_n \{n \geqq 1\colon R_n(\delta/c) \leqq \varepsilon/2k\}.$$

We have to prove $P_1 + P_2 \leqq \varepsilon$ where

$$P_1 = \mathsf{P}(\|\hat{\Theta}_t - \Theta_t\| > \delta,\ t \leqq N)$$

and

$$P_2 = \mathsf{P}(\|\hat{\Theta}_t - \Theta_t\| > \delta,\ t > N).$$

$$P_1 \leqq \mathsf{P}(t \leqq N) = \mathsf{P}\Big(\bigcup_{n=1}^{N} \{r_k(x_n^{(1)}, \ldots, x_n^{(k)}) < \gamma\}\Big) \leqq$$

$$\leqq \sum_{n=1}^{N} \mathsf{P}\Big(\max_{1 \leqq i, j \leqq k} \|x_n^{(i)} - x_n^{(j)}\| < \gamma/c_2\Big) \leqq$$

$$\leqq \sum_{n=1}^{N} \mathsf{P}(\|x_n^{(1)} - x_n^{(2)}\| < \gamma/c_2, \ldots, \|x_n^{(2\cdot[k/2]-1)} - x_n^{(2\cdot[k/2])}\| < \gamma/c_2) = \sum_{n=1}^{N} p_n^{[k/2]}(\gamma/c_2).$$

Since each x_n has continuous distribution and N and k are fixed we can choose $\gamma>0$ small enough such that $P_1 \leqq \varepsilon/2$.

$$P_2 \leqq \mathsf{P}\big(\sup_{n>N} \|\hat{\Theta}_n - \Theta_n\| > \delta\big) \leqq$$

(by the definition of $\hat{\Theta}_n$)

$$\mathsf{P}\big(\sup_{n>N} \max_{1\leqq i\leqq k} \|x_n^{(i)} - \Theta_n\| > \delta/c\big) =$$

$$= \mathsf{P}\big(\bigcup_{i=1}^{k} \{\sup_{n>N} \|x_n^{(i)} - \Theta_n\| > \delta/c\}\big) \leqq$$

$$\leqq k\mathsf{P}\big(\sup_{n>N} \|x_n - \Theta_n\| > \delta/c\big) = kR_N(\delta/c) \leqq \varepsilon/2. \quad \square\square\square$$

Now we have proved the theorem and the next two inequalities, too:

$$P_1 \leqq \sum_{n=1}^{N} p_n^{[k/2]}(\gamma/c_2), \tag{1.1}$$

$$P_2 \leqq kR_N(\delta/c). \tag{1.2}$$

Proof of Theorem 1.1.B. Let $\varepsilon>0$ and $\delta>0$ be arbitrary and let

$$N = \inf\Big\{n \geqq 1:\ R_n(\delta/c) \leqq \varepsilon/8,\quad \sum_{i=1}^{n} p_i^{(\varepsilon/16)\cdot(1/R_n(\delta/c))} \leqq \varepsilon/4\Big\},$$

$$k = \left[\frac{\varepsilon}{4R_N(\delta/c)}\right] + 1.$$

Choose $\gamma>0$ small enough such that

$$\sum_{n=1}^{N} p_n^{[k/2]}(\gamma/c_2) \leqq \sum_{n=1}^{N} p_n^{[k/2]} + \varepsilon/4.$$

Then according to (1.1) and the definition of N and k we have

$$P_1 \leqq \sum_{n=1}^{N} p_n^{[k/2]} + \varepsilon/4 \leqq \varepsilon/4 + \varepsilon/4 = \varepsilon/2,$$

$$P_2 \leqq \left(\left[\frac{\varepsilon}{4R_N(\delta/c)}\right] + 1\right) R_N(\delta/c) \leqq \varepsilon/4 + \varepsilon/8 \leqq \varepsilon/2. \quad \square\square\square$$

Proof of Theorem 1.1.C. It is the same as the proof of Theorem 1.1.A, the only difference is that the multiplying factor at the estimation of P_2 does not appear so N can be defined as follows

$$N = \inf\{n \geqq 1:\ R_n(\delta/c) \leqq \varepsilon/2\}. \quad \square\square\square$$

PROOF OF THEOREM 1.2.A. Let $\varepsilon>0$ and $k\geqq 2$ be arbitrary.

$$\mathsf{E}\|\hat\Theta_t-\Theta_t\| = \int\limits_{\|\hat\Theta_t-\Theta_t\|\leqq\varepsilon/3} \|\hat\Theta_t-\Theta_t\|\,d\mathsf{P} + \int\limits_{\|\hat\Theta_t-\Theta_t\|>\varepsilon/3} \|\hat\Theta_t-\Theta_t\|\,d\mathsf{P} \leqq$$

$$\leqq \int\limits_{\|\hat\Theta_t-\Theta_t\|\leqq\varepsilon/3} \varepsilon/3\,d\mathsf{P} + \int\limits_{\|\hat\Theta_t-\Theta_t\|>\varepsilon/3} \|\hat\Theta_t-\Theta_t\|\,d\mathsf{P} = \varepsilon/3 + \int\limits_{\|\hat\Theta_t-\Theta_t\|>\varepsilon/3} (\|\hat\Theta_t-\Theta_t\|-\varepsilon/3)\,d\mathsf{P} =$$

$$= \varepsilon/3 + \int\limits_{\|\hat\Theta_t-\Theta_t\|>\varepsilon/3} \int\limits_{\varepsilon/3}^{\|\hat\Theta_t-\Theta_t\|} 1\,dy\,d\mathsf{P} = \varepsilon/3 + \int\limits_{\varepsilon/3}^{\infty} \int\limits_{\|\hat\Theta_t-\Theta_t\|>y} 1\,d\mathsf{P}\,dy =$$

$$= \varepsilon/3 + \int\limits_{\varepsilon/3}^{\infty} \mathsf{P}(\|\hat\Theta_t-\Theta_t\|>y)\,dy =$$

$$= \varepsilon/3 + \int\limits_{\varepsilon/3}^{L} \mathsf{P}(\|\hat\Theta_t-\Theta_t\|>y)\,dy + \int\limits_{L}^{\infty} \mathsf{P}(\|\hat\Theta_t-\Theta_t\|>y)\,dy \leqq$$

$$\leqq \varepsilon/3 + L\mathsf{P}(\|\hat\Theta_t-\Theta_t\|>\varepsilon/3) + \int\limits_{L}^{\infty} \mathsf{P}(\|\hat\Theta_t-\Theta_t\|>y)\,dy.$$

Using (1.2) with $N=0$ we have

$$\mathsf{E}\|\hat\Theta_t-\Theta_t\| \leqq \varepsilon/3 + L\mathsf{P}(\|\hat\Theta_t-\Theta_t\|>\varepsilon/3) + k\int\limits_{L}^{\infty} R_0(y/c)\,dy$$

thus

$$\mathsf{E}\|\hat\Theta_t-\Theta_t\| \leqq \varepsilon/3 + L\mathsf{P}(\|\hat\Theta_t-\Theta_t\|>\varepsilon/3) + ckR(L/c).$$

By assumption (J) L can be chosen such that the third part does not exceed $\varepsilon/3$. After L is fixed, by Theorem 1.1.A, we can choose a $\gamma>0$ such that the second part can also not exceed $\varepsilon/3$ which completes the proof. □ □ □

PROOF OF THEOREM 1.2.B. It is the same as the proof of Theorem 1.2.A, the only, however considerable, difficulty arises at choosing the suitable k. The problem is similar to the extension of the proof of Theorem 1.1.A to the proof of Theorem 1.1.B but a little bit more complicated. The additional assumptions in this theorem with respect to Theorem 1.2.A serve for guaranteeing the correct choice of k. The problem is enlighted by the following argument based on the last estimation of $\mathsf{E}\|\hat\Theta_t-\Theta_t\|$: from the third part L is determined by k; then on the base of this L and Theorem 1.2.A k and γ are determined. So let $\eta>0$ be a provisionally nonfixed number and define, depending on η, the three numbers:

$$N = \inf\Big\{n\geqq 1:\ R_n(\varepsilon/3c)\leqq\eta/8,\quad \sum_{i=1}^{n} p_i^{(\eta/16)\cdot(1/R_n(\varepsilon/3c))}\leqq\eta/4\Big\},$$

$$k = \left[\frac{\eta}{4R_N(\varepsilon/3c)}\right]+1,$$

$$L = \inf\{x>0:\ ckR(x/c)\leqq\varepsilon/3\}.$$

In order to complete the proof we only have to find an $\eta>0$ for which

$$\eta \leqq \frac{1}{L}\varepsilon/3$$

where L depends on η, of course. Since R is a monotone real function we can take its generalized inverse which has the form

$$R^{-1}(y) = \inf\{x>0:\ R(x) \leqq y\}.$$

With this terminology we have to find an $\eta>0$ such that

$$R^{-1}\left(\frac{\varepsilon}{3c\left(\left[\frac{\eta}{4R_N(\varepsilon/3c)}\right]+1\right)}\right) \leqq \frac{1}{\eta}\varepsilon/3c. \tag{1.3}$$

First we give an estimation of N. By (i) and (ii) we have a c_α and a c_β (depending on $\varepsilon/3c$) such that

$$c_\alpha n^{-\alpha} \leqq R_n(\varepsilon/3c) \leqq c_\beta n^{-\beta}$$

so if

$$c_\beta n^{-\beta} \leqq \eta/8 \tag{1.4}$$

and

$$n p^{\eta n^\beta/16c_\beta} \leqq \eta/4 \tag{1.5}$$

then $n \geqq N$. Now let

$$u = \frac{\eta}{16c_\beta} n^\beta \log\frac{1}{p}$$

and $b=2/\beta$. One can easily verify that if $u \geqq e^b$ then $e^{-u} \leqq u^{-b}$. It follows that

$$n \geqq e^{2/\beta^2}\left(\frac{16c_\beta}{\log\frac{1}{p}}\right)^{1/\beta}\frac{1}{\eta^{1/\beta}}$$

and

$$n \geqq 4\left(\frac{16c_\beta}{\log\frac{1}{p}}\right)^{2/\beta}\frac{1}{\eta^{1+2/\beta}}$$

imply (1.5). We may assume that c_β is large enough, so we can integrate these last two inequalities:

$$n \geqq 4e^{2/\beta^2}\left(\frac{16c_\beta}{\log\frac{1}{p}}\right)^{2/\beta}\frac{1}{\eta^{1+2/\beta}} = A\frac{1}{\eta^{1+2/\beta}} \tag{1.6}$$

and say: (1.6) implies (1.5). But if η is small enough (1.6) implies (1.4), too. So we have got the inequality for N:

$$N \leqq A \frac{1}{\eta^{1+2/\beta}} \quad \text{for some} \quad A > 0. \tag{1.7}$$

By (ii)

$$R_N(\varepsilon/3c) \geqq c_\alpha N^{-\alpha} \geqq c_\alpha A^{-\alpha} \eta^{\alpha+2\alpha/\beta}$$

so taking into consideration (iv) we can complete the proof. □□□

PROOF OF THEOREM 1.2.C. It is the same as the proof of Theorem 1.2.A, the only difference is that at the last inequality k does not appear as a multiplying factor in the third part so we can use directly Theorem 1.1.C. □□□

2. The sequential method

Instead of k independent copies of the original sequence we consider now the sequence itself. Let $k \geqq 2$ and $\gamma > 0$ be arbitrary and let

$$u = \inf\{n \geqq k\colon\ r_k(x_{n-k+1}, \ldots, x_n) < \gamma\}.$$

Now u is a stopping time with respect to the sequence of σ-algebras $(\mathcal{G}_n)$ where $\mathcal{G}_n = \sigma\{x_1, \ldots, x_n\}$ $(n = 1, 2, \ldots)$. In this case we have to assume that there exists a $\Theta \in X$ such that $\lim_n \Theta_n = \Theta$ in order to guarantee u to be a.s. finite. Furthermore, since we do not know anything about the common distribution of the members of the sequence, we need some weak dependence conditions on the sequence to assure that the results mentioned in Chapter 1 remain similar.

Now let $n \in \mathbf{N}$ be arbitrary and introduce the following notations:

$$p_n'(\delta) = \mathsf{P}(\|x_{n-1} - x_n\| < \delta), \quad \delta > 0;$$

$$p_n' = \mathsf{P}(x_{n-1} = x_n);$$

$$p' = \sup_n p_n'.$$

We define some conditions on p_n' parallel to (a)—(f).

$$\text{(a}'\text{)} \qquad p_n' < 1 \quad n = 1, 2, \ldots;$$

$$\text{(b}'\text{)} \qquad \lim_n \sum_{i=1}^{n} p_i'^{\,s/R_n(\delta)} = 0 \qquad \forall s > 0, \quad \forall \delta > 0;$$

$$\text{(c}'\text{)} \qquad \lim_n n p'^{\,s/R_n(\delta)} = 0 \qquad \forall s > 0, \quad \forall \delta > 0;$$

$$\text{(d}'\text{)} \qquad p' < 1;$$

$$\text{(f}'\text{)} \qquad p' = 0.$$

Parallel to Chapter 1 we first state two lemmas about the stopping time u (to be a.s. finite and has finite expectation, respectively) then we deal with probability ε-optimality and next expectation ε-optimality.

LEMMA 2.1. *If* $(x_n-\Theta_n)$ *tends to zero in probability* ((A)) *then* u *is a.s. finite (i.e. it is a stopping rule), more precisely*

$$\forall k \geqq 2 \quad \forall \gamma > 0: \quad \mathsf{P}(u < \infty) = 1. \quad \square$$

LEMMA 2.2. *If* ($\underline{\mathrm{E}}$) *holds then* u *has finite expectation, more precisely*

$$\forall k \geqq 2 \quad \forall \gamma > 0: \quad \mathsf{E}u < \infty. \quad \square$$

PROOF OF LEMMA 2.1. Let $k \geqq 2$ and $\gamma > 0$ be arbitrary and let $N \in \mathbf{N}$ such that

$$\sup_{n,m \geqq N-k+1} \|\Theta_n - \Theta_m\| \leqq \gamma/2c_1.$$

Then

$$\begin{aligned}
\mathsf{P}(u=\infty) &= \mathsf{P}\Big(\bigcap_{n=k}^{\infty} \{r_k(x_{n-k+1}, \ldots, x_n) \geqq \gamma\}\Big) \leqq \\
&\leqq \inf_{n \geqq N} \mathsf{P}\big(r_k(x_{n-k+1}, \ldots, x_n) \geqq \gamma\big) \leqq \\
&\leqq \inf_{n \geqq N} \mathsf{P}\big(\max_{n-k+1 \leqq i,j \leqq n} \|x_i - x_j\| \geqq \gamma/c_1\big) \leqq \\
&\leqq \inf_{n \geqq N} \sum_{i,j=n-k+1}^{n} \mathsf{P}(\|x_i - x_j\| \geqq \gamma/c_1) \leqq \\
&\leqq \inf_{n \geqq N} \sum_{i,j=n-k+1}^{n} \mathsf{P}(\|x_i - \Theta_i\| + \|x_j - \Theta_j\| \geqq \gamma/2c_1) \leqq \\
&\leqq 2k \inf_{n \geqq N} \sum_{i=n-k+1}^{n} \mathsf{P}(\|x_i - \Theta_i\| \geqq \gamma/4c_1) = 0. \quad \square\,\square\,\square
\end{aligned}$$

PROOF OF LEMMA 2.2. Taking the preceding proof into account, for some constant C, we have

$$\begin{aligned}
\mathsf{E}u &= k + \sum_{n=k}^{\infty} \mathsf{P}(u > n) \leqq \\
&\leqq C + \sum_{n=N}^{\infty} \mathsf{P}\big(r_k(x_{n-k+1}, \ldots, x_n) \geqq \gamma\big) \leqq \\
&\leqq C + 2k \sum_{n=N}^{\infty} \sum_{i=n-k+1}^{n} \mathsf{P}(\|x_i - \Theta_i\| \geqq \gamma/4c_1) \leqq \\
&\leqq C + 2k^2 \sum_{n=N-k+1}^{\infty} \mathsf{P}(\|x_n - \Theta_n\| \geqq \gamma/4c_1) < \infty. \quad \square\,\square\,\square
\end{aligned}$$

Now we turn on to formulate the theorems concerning the probability ε-optimality.

THEOREM 2.1.A. *If* $(x_n-\Theta_n)$ *tends to zero a.s.* ((C)) *and* (f′) *holds then* u *is ε-optimal in the sense*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \forall k \geqq 2 \quad \exists \gamma > 0: \mathsf{P}(\|\hat{\Theta}_u - \Theta_u\| > \delta) \leqq \varepsilon. \quad \square$$

In what follows we assume (except for Theorem 2.2.A) that our original sequence (x_n) is m-dependent, i.e. for each $n\in\mathbf{N}$ the σ-algebras $\sigma\{x_1, \ldots, x_n\}$ and $\sigma\{x_{n+m+1}, x_{n+m+2}, \ldots\}$ are independent (m is a nonnegative integer).

THEOREM 2.1.B. *If* $(x_n-\Theta_n)$ *is m-dependent and tends to zero a.s.* ((C)) *and* (c′) *holds then* u *is ε-optimal in the sense*

$$\forall\varepsilon>0 \quad \forall\delta>0 \quad \exists k\geqq 2 \quad \exists\gamma>0\colon \mathsf{P}(\|\hat{\Theta}_u-\Theta_u\|>\delta)\leqq\varepsilon. \quad \Box$$

THEOREM 2.1.C. *If the functions* g_k *are projections and furthermore if* $(x_n-\Theta_n)$ *is m-dependent and tends to zero* ((C)) *and none of the rv's and its successors are a.s. identical* ((a')) *then* u *is ε-optimal in the sense*

$$\forall\varepsilon>0 \quad \forall\delta>0 \quad \exists k\geqq 2 \quad \exists\gamma>0\colon \mathsf{P}(\|x_u-\Theta_u\|>\delta)\leqq\varepsilon. \quad \Box$$

The next three theorems concern the expectation ε-optimality. They are similar to Theorem 1.2.A—Theorem 1.2.C.

THEOREM 2.2.A. *Assume the conditions of Theorem 2.1.A* ((C) *and* (f′)) *and furthermore* (J). *Then* u *is ε-optimal in the sense*

$$\forall\varepsilon>0 \quad \forall k\geqq 2 \quad \exists\gamma>0\colon \mathsf{E}\|\hat{\Theta}_u-\Theta_u\|\leqq\varepsilon. \quad \Box$$

THEOREM 2.2.B. *Assume the conditions* (i)—(iv) *of Theorem 1.2.B and* (d′). *Assume furthermore that the sequence is m-dependent. Then* u *is ε-optimal in the sense*

$$\forall\varepsilon>0 \quad \exists k\geqq 2 \quad \exists\gamma>0\colon \mathsf{E}\|\hat{\Theta}_u-\Theta_u\|\leqq\varepsilon. \quad \Box$$

THEOREM 2.2.C. *If the functions* g_k *are projections, conditions of Theorem 2.1.C and* (J) *hold then* u *is ε-optimal in the sense*

$$\forall\varepsilon>0 \quad \exists k\geqq 2 \quad \exists\gamma>0\colon \mathsf{E}\|x_u-\Theta_u\|\leqq\varepsilon. \quad \Box$$

PROOF OF THEOREM 2.1.A. The proof is very similar to the proof of Theorem 1.1.A. Let $\varepsilon>0$, $\delta>0$ and $k\geqq 2$ be arbitrary and let

$$N=\inf\{n\geqq k\colon R_{n-k+1}(\delta/2c)\leqq\varepsilon/2k, \sup_{i,j\geqq n-k+1}\|\Theta_i-\Theta_j\|\leqq\delta/2c\}.$$

We have to prove $P_1'+P_2'\leqq\varepsilon$ where

$$P_1'=\mathsf{P}(\|\hat{\Theta}_u-\Theta_u\|>\delta,\ u\leqq N)$$

and

$$P_2'=\mathsf{P}(\|\hat{\Theta}_u-\Theta_u\|>\delta,\ u>N).$$

$$P_1'\leqq\mathsf{P}(u\leqq N)=\mathsf{P}\Big(\bigcup_{n=k}^{N}\{r_k(x_{n-k+1},\ldots,x_n)<\gamma\}\Big)\leqq$$

$$\leqq\sum_{n=k}^{N}\mathsf{P}\Big(\max_{n-k+1\leqq i,j\leqq n}\|x_i-x_j\|<\gamma/c_2\Big)\leqq\sum_{n=k}^{N}\prod_{i=0}^{[(k-2)/(m+2)]}p'_{n-k+1+i\cdot(m+2)+1}(\gamma/c_2)$$

where $m=k$ if the m-dependence does not hold; so in this case

$$P_1\leqq\sum_{n=k}^{N}p'_{n-k+2}(\gamma/c_2).$$

Since N and k are fixed now and (f′) holds we can choose $\gamma>0$ small enough such that $P_1'\leqq\varepsilon/2$.

$$P_2'\leqq \mathrm{P}\big(\sup_{n>N}\|\hat{\Theta}_n-\Theta_n\|>\delta\big)\leqq$$

$$\leqq \mathrm{P}\big(\sup_{n>N}\max_{n-k+1\leqq i\leqq n}\|x_i-\Theta_n\|>\delta/c\big)\leqq$$

$$\leqq \mathrm{P}\big(\sup_{n>N}\max_{n-k+1\leqq i\leqq n}\|x_i-\Theta_i\|>\delta/2c\big)=$$

$$=P\big(\bigcup_{i=1}^{k}\{\sup_{n>N}\|x_{n-i+1}-\Theta_{n-i+1}\|>\delta/2c)\}\leqq$$

$$\leqq \sum_{i=1}^{k}\mathrm{P}\big(\sup_{n>N}\|x_{n-i+1}-\Theta_{n-i+1}\|>\delta/2c\big)\leqq$$

$$\leqq kR_{N-k+1}(\delta/2c)\leqq\varepsilon/2.\quad\square\square\square$$

So we have proved the theorem and the next two inequalities, too:

(2.1) $$P_1'\leqq\sum_{n=k}^{N}\prod_{i=0}^{[(k-2)/(m+2)]}p'_{n-k+1+i(m+2)+1}(\gamma/c_2),$$

(2.2) $$P_2'\leqq k\cdot R_{N-k+1}(\delta/2c).$$

Proof of Theorem 2.1.B. By assumption (c′) we can finitely define

$$M=\inf\Big\{n\geqq 1:\ np'^{\left[\frac{\left[\frac{\varepsilon}{2R_n(\delta/2c)}\right]-2}{m+2}\right]+1}\leqq\varepsilon/4\Big\}.$$

Let $k=\left[\dfrac{\varepsilon}{2R_M(\delta/2c)}\right]$ and $N=M+k-1$. This way $M=N-k+1$. Since N and k now are fixed we can choose $\gamma>0$ such that

$$\sum_{n=k}^{N}\prod_{i=0}^{[(k-2)/(m+2)]}p'_{n-k+1+i(m+2)+1}(\gamma/c_2)\leqq\sum_{n=k}^{N}\prod_{i=0}^{[(k-2)/(m+2)]}p'_{n-k+1+i(m+2)+1}+\varepsilon/4\leqq$$

$$\leqq(N-k+1)p'^{[(k-2)/(m+2)]+1}+\varepsilon/4.$$

So we have

$$P_1'\leqq Mp'^{\left[\frac{\left[\frac{\varepsilon}{2R_M(\delta/2c)}\right]-2}{m+2}\right]+1}+\varepsilon/4\leqq\varepsilon/4+\varepsilon/4=\varepsilon/2$$

and

$$P_2'\leqq\left[\frac{\varepsilon}{2R_M(\delta/2c)}\right]R_M(\delta/2c)\leqq\varepsilon/2.\quad\square\square\square$$

Proof of Theorem 2.1.C. The proof is based on inequalities (2.1) and (2.2) but in this case in the last inequality k does not appear as a multiplying factor and has the form

$$P_2'\leqq R_N(\delta).$$

So if we fix N such that $R_N(\delta)\leqq\varepsilon/2$ then we can choose k and γ in the required manner. □ □ □

Proof of Theorem 2.2.A. The flow of the proof is the same as of Theorem 1.2.A. We can get the last but one inequality by putting $N=k-1$ and then we have

$$\mathsf{E}\|\hat{\Theta}_u-\Theta_u\|\leqq\varepsilon/3+L\mathsf{P}(\|\hat{\Theta}_u-\Theta_u\|>\varepsilon/3)+2ckR(L/2c)$$

where L is as large as $\sup\limits_{n,m>0}\|\Theta_n-\Theta_m\|\leqq L/c$. Then at the estimation of the second part we can refer to Theorem 2.1.A. □ □ □

Proof of Theorem 2.2.B. It is essentially the same as the proof of Theorem 1.2.B. □ □ □

Proof of Theorem 2.2.C. It is essentially the same as the proof of Theorem 1.2.C (refer to Theorem 2.2.A instead of Theorem 1.2.A). □ □ □

Note 5. We mentioned introducing this chapter that we need some weak dependence conditions in order to prove some theorems. But we only used the m-dependence. It was for the sake of simplicity: much weaker conditions make possible the same argument. There is only one critical point at the estimation of P_1' where we must take into account weak dependence.

For example consider the strong mixing condition. Let (x_n) be a sequence of rv's, $\mathscr{F}_i=\sigma\{x_1,\dots,x_i\}$ and $\mathscr{G}_i=\sigma\{x_i,x_{i+1},\dots\}$ $i=1,2,\dots$. The sequence is said to be strong mixing if there is a sequence (ϱ_n) of real numbers which monotone decreasingly tends to zero and

$$|\mathsf{P}(AB)-P(A)\mathsf{P}(B)|\leqq\varrho_n\quad\forall A\in\mathscr{F}_k\quad\forall B\in\mathscr{G}_{k+n}\quad(n,k\geqq 1).$$

(Many of weak dependence conditions imply this strong mixing one. See e.g. [DP] and [S].)

Now suppose m is as large that $\varrho_m<\varepsilon/8$. When we are at that point of the proof where N and k are fixed we can partition the indexes from $n-k+1$ to n into $\left[\frac{k-2}{m+2}\right]+1$ classes, each of them consists of $n-k+1+i(m+2)$ and the successor. As we have seen we can disregard the argument of p_n'. Using (d′) and our mixing condition successively we get the inequality

$$P_1'\leqq P+\varrho_m\frac{1-P}{1-p'}+\varepsilon/8$$

where

$$P=p'^{[(k-2)/(m+2)]+1}.$$

This makes possible to prove the theorems (exactly the corresponding theorems of Theorem 2.1.B and Theorem 2.2.B).

REFERENCES

[BR] Blum, J. R. and Rosenblatt, J., On fixed precision estimation in time series, *Ann. Math. Statist.* **40** (1969), 1021—1032. *MR* **39** # 5009.

[CRS] Chow, Y. S., Robbins, H. and Siegmund, D., *Great Expectations: The Theory of Optimal Stopping,* Hughton Mifflin, Boston, 1971. *MR* **48** # 10007.

[DP] Dehling, H. and Philipp, W., Almost sure invariance principle for weakly dependent vector-valued random variables *Ann. Probab.* **10** (1982), 689—701. *MR* **83m**: 60011.

[D1] Dupač, V., A dynamic stochastic approximation method, *Ann. Math. Statist.* **36** (1965), 1695—1702. *MR* **33** # 1939.

[D2] Dupač, V., Stochastic approximations in the presence of trend, *Czechoslovak Math. J.* **16** (1966), 454—462. *MR* **33** # 8061.

[S] Serfling, R. J., Contributions to central limit theory for dependent variables, *Ann. Math. Statist.* **39** (1968), 1158—1175. *MR* **37** # 3637.

[Z1] Zieliński, R., A stopping rule for sequential estimation processes, Inst. Math. Polish Acad. Sci., Warszawa, Preprint 109 (1977).

[Z2] Zieliński, R., A class of stopping rules for fixed precision sequential estimates, Inst. Math. Polish Acad. Sci. Warszawa, Preprint 147 (1978).

[Z3] Zieliński, R., Fixed precision estimate of mean of Gaussian sequence with unknown covariance structure, *Lecture Notes in Statist.* **2** (1980), 360—364. *MR* **81m**: 62161.

(Received May 7, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 169—174

AN APPROXIMATION THEOREM FOR THE POISSON PROCESSES DEFINED ON AN ABSTRACT SPACE

CZESŁAW GÓŹDŹ and MIECZYSŁAW POLAK

Summary

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and let $\mathcal{X}$ be an arbitrary set, $\mathcal{F}$ a ring of subsets of $\mathcal{X}$, and μ an additive set function on $(\mathcal{X}, \mathcal{F})$. Let $X(E, \omega)$, $E\in\mathcal{F}$, $\omega\in\Omega$ be a random element defined on $\mathcal{F}\times\Omega$ with its values in a measurable Abelian group $(G, \mathcal{G})$.

The aim of this paper is to give sufficient conditions such that the random element $X(E, \omega)$ has a Poisson distribution with a parameter $\lambda(E)$ where λ is an additive set function on $(\mathcal{X}, \mathcal{F})$. In particular the theorems presented extend the main result obtained by Győrfi [2].

1. Let $(G, \mathcal{G})$ be a measurable Abelian group, that is, an Abelian group on which a σ-field $\mathcal{G}$ has been selected in such a way that the map $(x, y)\to x+y$, from $G\times G$ to G, is measurable for the σ-fields $\mathcal{G}\times\mathcal{G}$ and $\mathcal{G}$. Let ν be a finite signed measure on $\mathcal{G}$. For any finite signed measure ν on $\mathcal{G}$ let $\|\nu\|$ be the norm defined by

$$\|\nu\| = |\nu|(G), \tag{1}$$

where $|\nu|(G)$ is the total variation of ν.

Let $\mathfrak{M}$ be the system of finite signed measure on $\mathcal{G}$. It is not difficult to show that $\mathfrak{M}$, with the norm defined as in (1), is a Banach space. It may also be showed that

$$\sup_{A\subset G} |\nu(A)| \leqq \|\nu\| \leqq 2 \sup_{A\subset G} |\nu(A)| \tag{2}$$

(see I. I. Gichman, A. W. Skorochod [1] for details).

Let $\mathcal{X}$ be an arbitrary abstract set, $\mathcal{F}$ a ring of subsets of $\mathcal{X}$, and μ an additive set function of $\mathcal{F}$. Further, we assume that for each $\varepsilon>0$ and $E\in\mathcal{F}$, there exists a disjoint decomposition $E_1, E_2, \ldots, E_n$ of E with $E_i\in\mathcal{F}$ and $\mu(E_i)<\varepsilon$, $i=1, 2, \ldots, n$. Let $(\Omega, \mathcal{A}, \mathrm{P})$ denote a probability space, and $X(E, \omega)$ be a random element taking its values in a measurable Abelian group $(G, \mathcal{G})$ that is, for each $E\in\mathcal{F}$, $X(E, \cdot)$ is a random variable on $(\Omega, \mathcal{A}, \mathrm{P})$, and for each $\omega\in\Omega$, $X(\cdot, \omega)$ is an additive function on $\mathcal{F}$, i.e.,

$$X(E_1+E_2, \cdot) = X(E_1, \cdot)+X(E_2, \cdot) \tag{3}$$

provided $E_1\cap E_2=\emptyset$,

1980 *Mathematics Subject Classification.* Primary 60F05; Secondary 60J75.
Key words and phrases. Abelian group, random element, Poisson distribution.

2. The main results. Let $X(E,\omega)$, $E\in\mathscr{F}$, $\omega\in\Omega$ be a random element as above. Suppose that for each $E_1, E_2\in\mathscr{F}$ such that $E_1\cap E_2=\emptyset$,

$$\mathsf{P}[X(E_1)\neq\Theta, X(E_2)\neq\Theta]\leqq C\mu(E_1)\mu(E_2), \tag{4}$$

where $C>0$ is a constant independent of E_1 and E_2, and Θ is the null-element of the group G.

Let us define

$$0<\tau(E)=\mathsf{P}[X(E)\neq\Theta].$$

The idea of the undermentioned considerations is due to T. Leżański [4].

LEMMA. *If $X(E,\omega)$ is a random element satisfying the conditions* (3) *and* (4) *then for every sequence of disjoint sets $E_1, E_2, \ldots, E_n$ there exists a constant $C>0$ independent on $E_1, E_2, \ldots, E_n$ such that*

$$\Big|\sum_{j=1}^{n}\tau(E_j)-\tau\Big(\sum_{j=1}^{n}E_j\Big)\Big|\leqq C\mu^2\Big(\sum_{j=1}^{n}E_j\Big). \tag{5}$$

PROOF. The proof of this Lemma will be settled by induction argument. Let $E_1, E_2\in\mathscr{F}$ and $E_1\cap E_2=\emptyset$ then using (4), we have

$$|\tau(E_1)+\tau(E_2)-\tau(E_1+E_2)|\leqq C\mu(E_1)\mu(E_2)\leqq C\mu^2(E_1+E_2).$$

From (5) and (4), we obtain

$$\Big|\sum_{j=1}^{n}\tau(E_j)-\tau\Big(\sum_{j=1}^{n}E_j\Big)\Big|\leqq\Big|\sum_{j=1}^{n-1}\tau(E_j)-\tau\Big(\sum_{j=1}^{n-1}E_j\Big)\Big|+\Big|\tau\Big(\sum_{j=1}^{n-1}E_j\Big)+\tau(E_n)-\tau\Big(\sum_{j=1}^{n}E_j\Big)\Big|\leqq$$

$$\leqq C\mu^2\Big(\sum_{j=1}^{n-1}E_j\Big)+C\mu\Big(\sum_{j=1}^{n-1}E_j\Big)\mu(E_n)\leqq C\mu^2\Big(\sum_{j=1}^{n}E_j\Big).$$

The last inequality ends the proof of the Lemma.

Now let E be an arbitrary set from $\mathscr{F}$, and let $\{E_j^k, j=1,2,\ldots,n_k;\ k\geqq 1\}$ be a sequence of disjoint decompositions of the set E.

We define

$$\lambda_1(E)=\lim_{\max\limits_{1\leqq j\leqq n_k}\mu(E_j^k)\to 0}\sum_{j=1}^{n_k}\tau(E_j^k). \tag{6}$$

THEOREM 1. *If the conditions* (3) *and* (4) *are satisfied, then λ_1 given by* (6), *is a well defined additive set function on $\mathscr{F}$.*

PROOF. Let $\{E_j, j=1,2,\ldots,n\}$ be a disjoint decomposition of E and let $\{E_j^k, k=1,2,\ldots,n_j\}$ be a disjoint decomposition of $E_j, j=1,2,\ldots,n$, respectively. Then

$$\Big|\sum_{j=1}^{n}\sum_{k=1}^{n_j}\tau(E_j^k)-\sum_{j=1}^{n}\tau(E_j)\Big|\leqq\sum_{j=1}^{n}\Big|\sum_{k=1}^{n_j}\tau(E_j^k)-\tau\Big(\sum_{k=1}^{n_j}E_j^k\Big)\Big|.$$

By our Lemma, we have

$$\left|\sum_{j=1}^{n}\sum_{k=1}^{n_j}\tau(E_j^k)-\tau\left(\sum_{j=1}^{n}E_j\right)\right| \leqq C\max_{1\leqq j\leqq n_k}\mu(E_j)\mu(E).$$

Now we will show that the function λ_1 is an additive set function on $\mathscr{F}$.

Let E and F be disjoint sets of $\mathscr{F}$ and let $\{E_j^k, F_j^k, j=1,2,\ldots,n_k;\ k\geqq 1\}$ be sequences of disjoint decompositions of the sets E and F, respectively. Using the Lemma with $n=2$, we get

$$\sum_{j=1}^{n_k}[\tau(E_j^k)+\tau(F_j^k)-C\mu(E_j^k)\mu(F_j^k)] \leqq \sum_{j=1}^{n_k}\tau(E_j^k+F_j^k) \leqq$$

$$\leqq \sum_{j=1}^{n_k}[\tau(E_j^k)+\tau(F_j^k)+C\mu(E_j^k)\mu(F_j^k)].$$

The decompositions of the sets E and F can be chosen in such a way that $\max_{1\leqq j\leqq n_k}\mu(E_j^k)\to 0$ and $\max_{1\leqq j\leqq n_k}\mu(F_j^k)\to 0$ as $k\to\infty$, thus the Theorem 1 is proved.

REMARK 1. If for every sequence of disjoint sets $E_1, E_2\in\mathscr{F}$ the random elements $X(E_1)$, $X(E_2)$ are independent and if, in addition, for each $E\in\mathscr{F}$, $\tau(E)=1-\exp\{-\mu(E)\}$, then the condition (4) is satisfied and $\lambda_1(E)$, defined by (6), is equal to $\mu(E)$.

Now we assume that $\{\Theta\}\in\mathscr{G}$, i.e., the set that contains only the null-element of the group G is element of the σ-field $\mathscr{G}$.

Let $\mathscr{L}(X(E))$ denote the distribution of the random element of $X(E,\cdot)$. For each $E\in\mathscr{F}$, let us put

$$\mathscr{L}(X(E))=I+\tau(E)[M_E-I], \tag{7}$$

where I is the probability measure whose mass is entirely concentrated at the point $\Theta\in G$, and M_E is a probability measure such that

$$M_E(\Theta)=0. \tag{8}$$

Let $\{E_j^k\}_1^{n_k}$, $k\geqq 1$ be a sequence of disjoint decompositions of the set E, and let us put

$$\lambda_2(\{E_j^k\}_1^{n_k}) = \sum_{j=1}^{n_k}\tau(E_j^k), \tag{9}$$

$$M_1(\{E_j^k\}_1^{n_k}) = [1/\lambda_2(\{E_j^k\}_1^{n_k})]\left\{\sum_{j=1}^{n_k}\{\tau(E_j^k)M_{E_j^k}\}\right. . \tag{10}$$

Define

$$G_E = \lim_{\max_{1\leqq j\leqq n_k}\mu(E_j^k)\to 0} M_1(\{E_j^k\}_1^{n_k}). \tag{11}$$

THEOREM 2. *If the conditions* (3) *and* (4) *are satisfied, then* G_E *given in* (11) *is a well defined probability measure on* $\mathscr{G}$.

PROOF. The proof is essentially the same as the proof of Theorem 1. In order to prove this Theorem it suffices to show that there exists a constant $C>0$ such that for every pair of disjoint sets $E_1, E_2 \in \mathscr{F}$

$$\|\tau(E_1+E_2)M_{E_1+E_2}-\tau(E_1)M_{E_1}-\tau(E_2)M_{E_2}\| \leqq C\mu(E_1)\mu(E_2). \tag{12}$$

Let us put

$$m = \tau(E_1+E_2)M_{E_1+E_2}-\tau(E_1)M_{E_1}-\tau(E_2)M_{E_2}.$$

Taking into account (7) and (8), we may write

$$m(A) = \begin{cases} \mathsf{P}[X(E_1+E_2)\in A]-\mathsf{P}[X(E_1)\in A]-\mathsf{P}[X(E_2)\in A] & \text{if } A \not\ni \Theta, \\ 0 & \text{if } A = \Theta. \end{cases}$$

For each $A \not\ni \Theta$, we define

$$\nu(A) = \mathsf{P}[X(E_1+E_2)\in A]-\mathsf{P}[X(E_1)\in A]-\mathsf{P}[X(E_2)\in A].$$

It is easy to see that

$$\nu(A) \geqq \mathsf{P}[X(E_1)\in A, X(E_2) = \Theta]+\mathsf{P}[X(E_1) = \Theta, X(E_2)\in A]+$$
$$-\mathsf{P}[X(E_1)\in A]-\mathsf{P}[X(E_2)\in A] \geqq -2C\mu(E_1)\mu(E_2).$$

For every $A \not\ni \Theta$, we put

$$\varkappa(A) = \mathsf{P}[X(E_1+E_2)\in A]-\mathsf{P}[X(E_1)\in A, X(E_2) = \Theta]-\mathsf{P}[X(E_1) = \Theta, X(E_2)\in A].$$

It may be seen that $\varkappa(A)\geqq 0$ and also $\nu(A)\leqq\varkappa(A)$. Hence

$$|\nu(A)| \leqq \max[2C\mu(E_1)\mu(E_2), \varkappa(A)].$$

Therefore

$$\sup_{\Theta\notin A\subset G} |\nu(A)| \leqq \max[2C\mu(E_1)\mu(E_2), \varkappa(G\setminus\Theta)].$$

It is obvious that if (5), then

$$\varkappa(G\setminus\Theta) = \mathsf{P}[X(E_1+E_2) \neq \Theta]-\mathsf{P}[X(E_1) \neq \Theta]+\mathsf{P}[X(E_1) \neq \Theta, X(E_2) \neq \Theta]-$$
$$-\mathsf{P}[X(E_2) \neq \Theta]+\mathsf{P}[X(E_1) \neq \Theta, X(E_2) \neq \Theta] \leqq 4C\mu(E_1)\mu(E_2),$$

which proves that

$$\sup_{\Theta\notin A\subset G} |\nu(A)| \leqq 4C\mu(E_1)\mu(E_2). \tag{13}$$

Taking into account (2) and (13) we obtain (12).

It may be showed that G_E is a probability measure on $(G, \mathscr{G})$.

REMARK 2. Let δ be a real valued function such that $\lim_{x\to 0}\delta(x)=0$. Furthermore, assume that for every $E\in\mathscr{F}$ there exists a constant $C>0$ such that $\tau(E)\geqq \geqq C\mu(E)$. If there exists a set $B\in\mathscr{G}$ such that for every $E\in\mathscr{F}$

$$\mathsf{P}[X(E)\in B] \leqq \mu(E)\,\delta(\mu(E)),$$

then $G_E(B)=0$.

Let $X(E, \omega)$, $E \in \mathscr{F}$, $\omega \in \Omega$ be a random element with values in a measurable group $(G, \mathscr{G})$ and satisfying the condition (3).

Assume that

1° there exists $C > 0$ such that for every $E \in \mathscr{F}$

$$\mathsf{P}[X(E) \neq \Theta] \leqq C\mu(E),$$

2° for any sequence of disjoint sets $E_1, E_2, \ldots, E_n \in \mathscr{F}$ the random variables $X(E_1), X(E_2), \ldots, X(E_n)$ are independent.

THEOREM 3. *If $X(E, \omega)$, $E \in \mathscr{F}$, $\omega \in \Omega$ is a random element satisfying the conditions* 1° *and* 2°, *then*

$$\left\|\mathscr{L}(X(E)) - \exp \lambda_2(\{E_j\}_1^n)[M_1(\{E_j\}_1^n) - I]\right\| \leqq 2C \sum_{j=1}^{n} \mu^2(E_j),$$

where $\lambda_2(\{E_j\}_1^n)$, $M_1(\{E_j\}_1^n)$ are defined by (9) *and* (10), *respectively.*

PROOF. The method used in the proof of this Theorem is essentially the same as the method given by Le Cam [3].

At first we see that the conditions 1° and 2° imply the condition (4).

Let us put $F_{E_j} = \exp \tau(E_j)(M_{E_j} - I)$, $j = 1, 2, \ldots, n$, and let

$$R_{E_1} = \prod_{j=1}^{n}{}^{*} \mathscr{L}(X(E_j)), \qquad R_{E_n} = \prod_{j=1}^{n-1}{}^{*} F_{E_j}.$$

For $1 < k < n$, we define

$$R_{E_k} = \Big(\prod_{1 \leqq j \leqq k-1}{}^{*} F_{E_j}\Big) * \Big(\prod_{j \geqq k+1}^{n}{}^{*} \mathscr{L}(X(E_j))\Big).$$

Then for $k = 1, 2, \ldots, n$, we have

$$R_{E_k} * F_{E_k} = R_{E_{k+1}} * \mathscr{L}(X(E_{k+1})).$$

Now, it is easy to see that

$$\prod_{j=1}^{n}{}^{*} \mathscr{L}(X(E_j)) - \prod_{j=1}^{n}{}^{*} F_{E_j} = \sum_{j=1}^{n} R_{E_j} * (\mathscr{L}(X(E_j)) - F_{E_j}).$$

Since R_{E_j} is a probability measure, this implies

$$\Big\|\prod_{j=1}^{n}{}^{*} \mathscr{L}(X(E_j)) - \prod_{j=1}^{n}{}^{*} F_{E_j}\Big\| \leqq \sum_{j=1}^{n} \|\mathscr{L}(X(E_j)) - F_{E_j}\|.$$

The difference $F_{E_j} - \mathscr{L}(X(E_j))$ can be written

$$F_{E_j} - \mathscr{L}(X(E_j)) = [e^{-\tau(E_j)} - 1 + \tau(E_j)]I + \tau(E_j)(e^{-\tau(E_j)} - 1)M_{E_j} +$$

$$+ e^{-\tau(E_j)} \sum_{k=2}^{\infty} \tau^k(E_j)(M_{E_j})^{*k}/k!.$$

Hence

$$\left\|F_{E_j} - \mathscr{L}(X(E_j))\right\| \leqq 2\tau(E_j)(1 - e^{-\tau(E_j)}) \leqq 2C\mu^2(E_j).$$

This proves the desired result.

THEOREM 4. *If $X(E, \omega)$, $E \in \mathscr{F}$, $\omega \in \Omega$ is a random element satisfying the conditions* 1° *and* 2°, *then*

$$\mathscr{L}(X(E)) = \exp \lambda_1(E)(G_E - I),$$

where $\lambda_1(E)$, G_E are defined in (6) *and* (11), *respectively.*

PROOF. At first we see that from Theorem 3, we have

$$\left\|\mathscr{L}(X(E)) - \exp \lambda_1(E)(G_E - I)\right\| \leqq 2C \sum_{j=1}^{n} \mu^2(E_j) +$$

$$\left\|\exp \lambda_2(\{E_j\}_1^n)\left[M_1(\{E_j\}_1^n) - I\right] - \exp \lambda_1(E)(G_E - I)\right\|.$$

Taking into account the Theorems 1 and 2 it is easy to see that the right-hand side of the last inequality tends to 0. This proves the desired result.

3. Special case. In this section, we consider the case where $(G, \mathscr{G})$ is a measurable Abelian subgroup of integer numbers. Let $X(E, \omega)$, $E \in \mathscr{F}$, $\omega \in \Omega$ be a random element taking its values in the set of nonnegative integer numbers and satisfying the condition (3).

THEOREM 5. *If $X(E, \omega)$, $E \in \mathscr{F}$, $\omega \in \Omega$ is a random element satisfying the conditions* 1° *and* 2°, *and if $G_E(1) = 1$, then*

$$\mathsf{P}[X(E) = k] = \exp[-\lambda_1(E)]\lambda_1^k(E)/k!, \qquad k = 0, 1, \ldots.$$

PROOF. The proof of this Theorem follows from Theorem 4 and from the fact that if $G_E(1) = 1$, then $\exp \lambda_1(E)(G_E - I)$ has the Poisson distribution with the parameter $\lambda_1(E)$.

The obtained result extends the main Theorem given by Győrfi [2].

ACKNOWLEDGEMENT. The authors wish to thank the referee for his helpful comments.

REFERENCES

[1] GIHMAN, I. I. and SKOROHOD, A. V., The theory of stochastic processes II, Springer-Verlag, New York—Heidelberg, 1975. *MR* **51** # 11656.

[2] GYŐRFI, L., Poisson processes defined on an abstract space, *Studia Sci. Math. Hungar.* **7** (1972), 243—248. *MR* **48** # 7365.

[3] LE CAM, L., An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.* **10** (1960), 1181—1197. *MR* **25** # 5567.

[4] LEŻAŃSKI, T., Sur l'intégration directe des équations d'évolution, *Studia Math.* **34** (1970), 149—163. *MR* **56** # 6042.

(Received May 8, 1984)

INSTYTUT MATEMATYKI
UNIWERSYTET MARII CURIE-SKŁODOWSKIEJ
PL. MARII CURIE-SKŁODOWSKIEJ 1
PL—20-031 LUBLIN
POLAND

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 175—188

LARGE PRIME FACTORS OF INTEGERS IN AN ARITHMETIC PROGRESSION

A. BALOG, J. FRIEDLANDER[1] and J. PINTZ

1. Introduction

Among a number of interesting results that have been proved by the Chebyshev—Hooley method is the theorem of Ramachandra [13] that, for arbitrary $\alpha>0$ there exists $\delta(\alpha)>0$ such that the interval $(X, X+X^\alpha]$ contains, for all $X>X_0$, an integer having a prime factor exceeding $X^{\alpha+\delta}$. The problem of getting good admissible values for δ has since received considerable attention beginning with Jutila [10] and most recently from Balog—Harman—Pintz [1].

The analogous problem for an arithmetic progression, which seems not to have been discussed in the literature, forms the topic of this paper. Our first result is

THEOREM 1. *Let $\alpha>1/2$ and $(a,q)=1$. There exists $\delta(\alpha)>0$ such that, for all sufficiently large q, there is some $m\leqq X=q^{1+\alpha}$ with $m\equiv a(\bmod q)$, and m having a prime factor exceeding $q^{\alpha+\delta}$.*

There are a variety of ways to attack this problem. The (rather complicated) values for δ obtained here are stated explicitly in Theorem 3, in the final section of the paper.

The absence of a non-trivial result for arbitrary $\alpha>0$ is due to the absence of an analogous result for short incomplete Kloosterman sums. Assumption of conjecture R^* of Hooley [7] gives Theorem 1 for any $\alpha>0$. In the case of short intervals the Kloosterman sums are replaced by van der Corput sums for which non-trivial estimates are known [10].

One conjectures that for any $\alpha>0$ we may take $\delta(\alpha)=1$. Indeed, for α sufficiently large, Linnik's theorem on the least prime in an arithmetic progression gives $\delta(\alpha)=1$, while Theorem 3 gives only $\lim\limits_{\alpha\to\infty}\delta(\alpha)=\frac{13}{16}$. On the other hand the methods given here provide a positive proportion of the integers (in the progression) with a large prime factor whereas Linnik's method gives far fewer.

It is well-known that the assumption of the Generalized Riemann Hypothesis gives primes $\ll q^{2+\varepsilon}$ and thus implies $\delta(\alpha)=1$ for $\alpha>1$. Moreover, as a consequence of the method of § 5 we are still able to derive a rather strong result from the following well-known weaker conjecture.

---

[1] Partially supported by NSERC Grant A5123.

1980 *Mathematics Subject Classification.* Primary 10H20, 10H30; Secondary 10H32, 10G05.
Key words and phrases. Primes, sieves, arithmetic progressions.

HYPOTHESIS. *For any* $\varepsilon>0$ *there exist* $\omega(\varepsilon)>0$, $q_0(\varepsilon)>0$ *such that whenever* $L>q^\varepsilon$ *and* $q>q_0$ *we have for all non-principal characters* $\chi \pmod q$

$$\Big|\sum_{l\leq L}\chi(l)\Big| < Lq^{-\omega}.$$

THEOREM 2. *Under the above hypothesis we have* $\delta(\alpha)=1-\varepsilon$ *for* $\alpha>2$, *that is, if* $X\geqq q^{3+\varepsilon}$ *and* $q>q_0(\varepsilon)$ *then for all a with* $(a,q)=1$ *there is an* $m\leqq X$, $m\equiv a \pmod q$ *such that the greatest prime factor* p *of* m *satisfies* $p>Xq^{-\varepsilon}$.

The paper is organized as follows. Sections 2 and 3 are preliminary, the former introducing the general idea of Chebyshev's method and reducing the problem to a sum over large prime powers, the latter reducing it to a sum over large primes. Sections 4, 5, 6 form the heart of the work giving three different treatments for the essential sum, each of which is stronger than the others for some range of α. In Section 7 we summarize the previous efforts in the statement of Theorem 3 and make some concluding remarks about possible improvements for δ.

2. Preliminary reduction of the proof

Throughout the paper ε denotes an arbitrary positive real number, not necessarily the same at each occurrence.

Letting $A(t,d)=\sum\limits_{\substack{j\leq t\\ j\equiv a(q)\\ j\equiv 0(d)}} 1$ we have

$$\sum_{\substack{j\leq X\\ j\equiv a(q)}}\log j = \sum_{\substack{j\leq X\\ j\equiv a(q)}}\sum_{d|j}\Lambda(d) = \sum_d \Lambda(d)A(X,d). \tag{1}$$

The left-hand side is

$$Y\log X+O(Y),$$

where throughout, $Y=X/q=q^\alpha$. The right-hand side is

$$\frac{X}{q}\sum_{\substack{(d,q)=1\\ d\leq Y}}\frac{\Lambda(d)}{d}+O(Y)+\sum_{d>Y}\Lambda(d)A(X,d).$$

Since $\sum\limits_{d|q}\frac{\Lambda(d)}{d}\ll\log\log q$, we have

$$\sum_{d>Y}\Lambda(d)A(X,d) = Y\log q+O(Y\log\log q). \tag{2}$$

For any $d>Y$ we have $A(X,d)\leqq 2$ and so the contribution to the sum in (2) from prime powers $d=p^k$ with $p\leqq Y$ is majorized by $\pi(Y)\log X$ and can be absorbed in the error term. For the remaining powers, where $p>Y$ we have $k\leqq 1+\alpha^{-1}$.

To complete the proof it suffices to show that for some positive ε and δ with $p=q^{\alpha+\delta}$, we have

$$\sum_{\substack{k,\,p\\ Y<p\leqq P}} A(X,p^k)\log p \leqq (1-\varepsilon)Y\log q. \tag{3}$$

In view of the trivial bound $A(X, p^k) \leqq A(X, p)$, it suffices to have

$$\sum_{Y<p\leqq P} A(X,p)\log p \leqq \left(\frac{\alpha}{\alpha+1}-\varepsilon\right) Y \log q. \tag{4}$$

This, however, gives a weaker value of δ than necessary. We instead show that, for $\alpha > 1/2$,

$$\sum_{p>Y} A(X, p^2) \ll_{\alpha} q^{1/2} \tag{5}$$

whence, since, $A(X, p^k) \leqq A(X, p^2)$ for $k > 2$ and since $k \leqq 1+\alpha^{-1}$, it is sufficient to prove

$$\sum_{Y<p\leqq P} A(X,p)\log p \leqq (1-\varepsilon) Y \log q, \tag{6}$$

for an appropriate δ. The methods for obtaining (6) occupy the major part of this work.

3. Multiples of higher powers

In this section we discuss methods for estimating the sum in (5). In the short interval problem the analogue of (5) is reduced to the estimation of exponential sums which can be treated by van der Corput's method for any $\alpha > 0$. Here the problem is more difficult but for the case $\alpha > 1/2$ an almost trivial method suffices.

Writing a typical integer contributing to $A(X, p^2)$ in the form kp^2, we have

$$\sum_{p>Y} A(X, p^2) = \sum_{\substack{k \leqq X/Y^2 \\ (k,q)=1}} S_k$$

where we fix, for each k, some $\bar{k}$ satisfying $k\bar{k} \equiv 1 \pmod q$, and let S_k be the number of integers j satisfying the conditions:

$$0 < jqk + ak\bar{k} \leqq X \tag{7}$$

and

$$jq + a\bar{k} = p_0^2 \quad \text{for some prime} \quad p_0 > Y. \tag{8}$$

Since $p_0 \leqq X^{1/2}$ and since the number of solutions in m of the congruence $km^2 \equiv a \pmod q$ is $\ll d(q)$, we have

$$\sum_{p>Y} A(X, p^2) \ll XY^{-2} d(q)(1+X^{1/2}q^{-1})$$

which gives (5) for $\alpha > 1/2$.

Remark. There are less trivial ways of attacking the sum (5); for the most part these also fail for small α.

The replacement of (8) by the condition that $jq + a\bar{k}$ be a quadratic residue modulo every "small" prime allows, if $\alpha > 1/2$, an application of the large sieve identity much as in [6, pp. 66—68]. The reduction of (5) to exponential sums (as done for (6) in § 4) leads to incomplete sums of the form

$$\sum_{\substack{v_1 \leqq m \leqq v_2 \\ (m,q)=1}} e\left(\frac{a\overline{m}^2}{q}\right) \tag{9}$$

which can be treated (see [4]) by Weil's method but this, too, fails for $\alpha \leqq 1/2$. Recent progress has been made by Heath-Brown [4] and from his treatment it follows that one has (5) for $\alpha > 4/9$.

4. Kloosterman sums and the sieve

In this section we prove:

PROPOSITION. *Let $\varepsilon > 0$ and $\alpha > 1/2$ be given. For all sufficiently large q and any ϱ, σ satisfying $\alpha \leqq \varrho < \sigma \leqq \alpha + 1 - \varepsilon$, and $R = q^\varrho$, $S = q^\sigma$, we have*

$$\sum_{R < p \leqq S} A(X, p) \leqq \frac{4+\varepsilon}{2\alpha - 1}(\sigma - \varrho)Y. \tag{10}$$

From (10) we deduce by considering a sufficiently fine partition of the interval $[\alpha, \alpha+\delta]$ that

$$\sum_{Y < p \leqq P} A(X, p) \log p \leqq \frac{4+\varepsilon'}{2\alpha - 1} Y \log q \int_{\alpha}^{\alpha+\delta} u\, du \tag{11}$$

which gives (6), and Theorem 1, with any δ such that $(\alpha+\delta)^2 - \alpha^2 < \alpha - 1/2$.

PROOF OF (10). Let $\{\lambda_d | d < D\}$ denote the upper bound Rosser or Selberg weights as usually used for an arithmetic progression mod q, having support on (a subset of) those square-free integers prime to q, all of whose prime factors are $< z$. We assume $z \leqq D \leqq Y$ so that

$$\sum_{R < p \leqq S} A(X, p) \leqq \sum_{R < m \leqq S} A(X, m) \sum_{d | m} \lambda_d = \sum_{d < D} \lambda_d \sum_{Rd^{-1} < r \leqq Sd^{-1}} \sum_{\substack{k \leqq X/dr \\ k \equiv a\bar{d}\bar{r}(q)}} 1 \tag{12}$$

where, again, $\bar{s}$ means $\bar{s}s \equiv 1\,(q)$. Throughout, r will be restricted to integers prime to q. The inner sum over k may be written as

$$\frac{X}{drq} + \psi\left(\frac{-a\bar{d}\bar{r}}{q}\right) - \psi\left(\frac{X}{drq} - \frac{-a\bar{d}\bar{r}}{q}\right),$$

where $\psi(t) = t - [t] - 1/2$. The main term $\dfrac{X}{drq}$ gives rise in (12) to a contribution

$$Y \sum_{d < D} \frac{\lambda_d}{d}\left(\frac{\varphi(q)}{q} \log(S/R) + O\left(\frac{d\tau(q)}{Y}\right)\right) =$$

$$= (\sigma - \varrho) Y q^{-1} \varphi(q) \log q \sum_d \frac{\lambda_d}{d} + O(Dq^\varepsilon) \leqq (2+\varepsilon)(\sigma - \varrho) Y \log q / \log D,$$

for all sufficiently large q, if we take say $D=z^2=q^{\alpha-\theta}$ for some positive θ. It thus suffices to show that, for $\theta>1/2$ and for each of $x=0$ and $x=X$, we have

(13) $$\sum_{d<D}\lambda_d \sum_{Rd^{-1}<r\leq Sd^{-1}} \psi\left(\frac{x}{drq}-\frac{a\bar{d}\bar{r}}{q}\right)=o(Y).$$

In fact we shall majorize the inner sum for each fixed d. The latter may be written as a sum of $\ll\log X$ sums

(14) $$\sum_{Td^{-1}<r\leq T'd^{-1}} \psi\left(\frac{x}{drq}-\frac{a\bar{d}\bar{r}}{q}\right)$$

where $T<T'\leq 2T$. For any $0<\Delta<1/2$, we have

$$|\psi(t)-A(t)|\leq B(t)$$

for two functions $A(t)=\sum_{h\neq 0} a_h e(ht)$, $e(t)=e^{2\pi it}$

$$B(t)=\Delta+\sum_{h\neq 0} b_h e(ht),$$

where $|a_h|\ll c_h$, $|b_h|\ll c_h$, $c_h=\min\left(\frac{1}{|h|},\frac{1}{\Delta^2|h|^3}\right)$. This may be seen by taking the functions in Lemma 2 of [3] and averaging them as in Vinogradov [15, p. 33]. Thus (14) is bounded by

(15) $$\Delta Td^{-1}+\sum_{h\neq 0} c_h\left|\sum_{Td^{-1}<r<T'd^{-1}} e\left(\frac{hx}{drq}-\frac{ha\bar{d}\bar{r}}{q}\right)\right|.$$

To the inner sum over r we apply the estimate

(16) $$\sum_{v_1<r\leq v_2} e\left(b\frac{\bar{r}}{q}\right)\ll q^{1/2+\varepsilon}(b,q)^{1/2}+\frac{(v_2-v_1)}{q}(b,q)$$

derived by Hooley [5, Lemma 1] from the Weil estimate for the Kloosterman sum. By partial summation

$$\sum_{Td^{-1}<r\leq T'd^{-1}} e\left(\frac{hx}{drq}-\frac{ha\bar{d}\bar{r}}{q}\right)\ll\left(q^{1/2+\varepsilon}(h,q)^{1/2}+\frac{T}{dq}(h,q)\right)\left(1+\frac{|h|x}{Tq}\right),$$

whence, after some computation, we see that (15) is majorized by

$$\Delta Td^{-1}+q^{1/2+\varepsilon}+X(\Delta T)^{-1}q^{-1/2+\varepsilon}+X(d\Delta)^{-1}q^{-2+\varepsilon}+Td^{-1}q^{-1+\varepsilon}.$$

Finally, the choices $D=q^{\alpha-1/2-\gamma}$ (for positive $\gamma<\alpha-1/2$) and $\Delta=T^{-1}q^{\alpha-1/2\gamma}$ ensure that the left-hand side of (13) is bounded by $q^{\alpha-\gamma/3}$. This completes the proof of (10).

5. Character sums and the Vaughan identity

If $\alpha>1$ then we can improve the result of § 4 by using a character sum approach instead of Kloosterman sums. We start with a definition.

(17) Definition. Let γ be a non-negative constant with the property that for any $\varepsilon>0$ there is an $\omega=\omega(\varepsilon)>0$ such that

$$\sum_{l\leqq L}\chi(l)\ll Lq^{-\omega}$$

for all non-principal characters $\chi \pmod q$ and all $L\geqq q^{\gamma+\varepsilon}$.

In this section we prove

(18) Proposition. *Let $\varepsilon_0>0$ and $\alpha>1$ be given. For all sufficiently large q and*

$$Y\leqq R<2R\leqq S\leqq q^{-\varepsilon_0}\min(Yq^{1/2}, Y^{4/3})$$

we have

$$\sum_{R<p\leqq S} A(X,p)\log p\leqq(1+\varepsilon_0)Y\log\frac{S}{R}.$$

(19) Proposition. *Let $\varepsilon_0>0$ and $\alpha>2$ be given. For all sufficiently large q and*

$$Y\leqq R<2R\leqq S\leqq Xq^{-\gamma-\varepsilon_0}$$

we have

$$\sum_{R<p\leqq S} A(X,p)\log p\leqq(1+\varepsilon_0)Y\log\frac{S}{R}.$$

Our knowledge about γ is

Lemma 1 (Burgess, [2]). *For any q we have $\gamma=3/8$ and for cube-free q we have $\gamma=1/4$.*

However, it is expected that

(20) Hypothesis. *For any q we have $\gamma=0$.*

This is just the hypothesis stated in the introduction. It is easy to see that for $\alpha>2$ (19) is stronger than (18) and it is immediate from (19) that the above hypothesis implies

$$\sum_{Y<p\leqq Xq^{-\varepsilon}} A(X,p)\log p\leqq(1-\varepsilon')Y\log q,$$

which gives Theorem 2.

The other effects of (18) and (19) on $\delta(\alpha)$ will be discussed later. We note that the bounds given in (18) and (19) are best possible; in fact, the same considerations give the asymptotic formula.

Proof of (18) and (19). After a splitting up argument we are interested in getting a good upper bound for

$$\mathcal{T}=\mathcal{T}(H)=\sum_{H<p\leqq\eta H} A(X,p)\log p \tag{21}$$

where $\eta=1+\varepsilon$ and $Y\leqq H<X$. For technical reasons sometimes we use weighted sums instead of $\mathcal{T}$ itself.

$$\mathcal{T}\leqq(1+\varepsilon)\mathcal{T}^*=(1+\varepsilon)\sum_{H<n\leqq\eta H}\Lambda(n)\sum_{\substack{l=1\\ nl\equiv a(q)}}^{\infty}w(l) \tag{22}$$

where either

$$w(l)=w_1(l)=e^{-(l/H_1)^h},\quad H_1=\frac{\eta X}{H},\quad h=\log^2 q$$

or

$$w(l)=w_2(l)=\begin{cases}1 & \text{if } l\leqq\dfrac{X}{H},\\ 0 & \text{otherwise.}\end{cases}$$

From the orthogonality of characters we have

$$\mathcal{T}^*=\frac{1}{\varphi(q)}\sum_{\chi}\bar{\chi}(a)\mathcal{T}_\chi \tag{23}$$

where

$$\mathcal{T}_\chi=\sum_{H<n\leqq\eta H}\sum_{l=1}^{\infty}\Lambda(n)\chi(n)w(l)\chi(l).$$

Computing the contribution of the principal character we get from (22) and (23) that for $H<Xq^{-\varepsilon}$

$$\mathcal{T}\leqq(1+5\varepsilon)Y\sum_{H<n\leqq\eta H}\frac{\Lambda(n)}{n}+O\left(\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|\mathcal{T}_\chi|\right).$$

We obtain (18) and (19) from this by proving that under the given conditions on α and H we have

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|\mathcal{T}_\chi|\ll Yq^{-\varepsilon}. \tag{24}$$

We transform the sum over primes into bilinear forms.

Lemma 2 (Vaughan's identity [14]). *For an arbitrary function $F(n)$ and real numbers $1\leqq X$, $1\leqq U\leqq X^{1/2}$ we have*

$$\sum_{n\leqq X}\Lambda(n)F(n)=\sum_{n\leqq U}\Lambda(n)F(n)+\sum_{\substack{mn\leqq X\\ n\leqq U}}\sum\mu(n)\log mF(mn)-$$

$$-\sum_{\substack{mn\leqq X\\ m\leqq U^2}}\sum\Big(\sum_{\substack{m=dt\\ d,t\leqq U}}\mu(d)\Lambda(t)\Big)F(mn)+\sum_{\substack{mn\leqq X\\ m,n>U}}\sum\Lambda(m)\Big(\sum_{\substack{n=dt\\ t>U}}\mu(t)\Big)F(mn).$$

Take

$$B_\chi=B_\chi(M,N)=\sum_{H<mn\leqq\eta H}\sum\sum_{l=1}^{\infty}a_m\chi(m)b_n\chi(n)w(l)\chi(l)$$

and

$$B^*_\chi=B^*_\chi(N)=\sum_{H<mn\leqq\eta H}\sum\sum_{l=1}^{\infty}\chi(m)b_n\chi(n)w(l)\chi(l)$$

where a_m, b_n are complex coefficients satisfying $|a_m|\leqq 1$, $|b_n|\leqq 1$; $a_m\neq 0$ implies $M<m\leqq 2M$; $b_n\neq 0$ implies $N<n\leqq 2N$ and $H/4\leqq MN\leqq 2H$ in B_χ, $N\leqq 2H$ in B^*_χ. From Vaughan's identity we have

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|\mathcal{T}_\chi| \ll \log^2 q \max_{N\leqq U^2}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B^*_\chi|+q^\varepsilon \max_{U<N\leqq H^{1/2}}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi|. \tag{25}$$

Thus we are interested in estimating the average over characters of general (B_χ) and special (B^*_χ) bilinear forms. We can detect the condition $H<mn\leqq\eta H$ by using the Perron integral formula

$$\begin{aligned} B_\chi = \frac{1}{2\pi i}\int_{-T}^{T}\sum\sum\sum_{l=1}^{\infty}\frac{a_m m^{-it}\chi(m)}{m^{1/2}}\frac{b_n n^{-it}\chi(n)}{n^{1/2}}w(l)\chi(l)\frac{(\eta H)^{1/2+it}-H^{1/2+it}}{1/2+it}dt+ \\ +O\left(\frac{Hq^\varepsilon}{T}\left|\sum_{l=1}^{\infty}w(l)\chi(l)\right|\right). \end{aligned} \tag{26}$$

With the choice $T=Xq$ the error term can be neglected. For a fixed $|t|<T$ we denote

$$a^*_m = a^*_m(t) = \frac{a_m m^{-it}}{(2m/M)^{1/2}}, \qquad b^*_n = b^*_n(t) = \frac{b_n n^{-it}}{(2n/N)^{1/2}}$$

and

$$B_{\chi,t} = \sum_m\sum_n\sum_{l=1}^{\infty}a^*_m\chi(m)b^*_n\chi(n)w(l)\chi(l).$$

We have

$$\begin{aligned} \frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi| \ll \int_{-T}^{T}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_{\chi,t}|\frac{dt}{1+|t|}+O(1) \ll \\ \ll 1+\log q\max_{|t|\leqq Xq}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_{\chi,t}| \end{aligned} \tag{27}$$

and the coefficients of $B_{\chi,t}$ have the same properties as those of B_χ. To prove (18) we use the weight $w_1(l)$. We need two further lemmas.

LEMMA 3. Reflection argument (Jutila, [11]). *If* $1\leqq K\leqq 2q$, $h=\log^2 q$, $K^*=\frac{q\log^6 q}{K}$ *and* $\chi\neq\chi_0$ *then*

$$\left|\sum_{l=1}^{\infty}\chi(l)e^{-(l/K)^h}\right| \ll 1+d(q)\int_{-h^2}^{h^2}\left|\sum_{l\leqq K^*}\frac{\chi(l)l^{i\tau}}{(l/K)^{1/2}}\right|d\tau.$$

LEMMA 4. Mean-square theorem (Montgomery, [12]). *For any complex numbers* a_m *we have*

$$\frac{1}{\varphi(q)}\sum_\chi\left|\sum_{n\leqq N}a_n\chi(n)\right|^2 \leqq \left(1+\frac{N}{q}\right)\sum_{n\leqq N}|a_n|^2.$$

We use the reflection argument for the factor $\sum w_1(l)\chi(l)$. Taking $H_2 = \dfrac{q\log^6 q}{H_1}$ and for a fixed $|\tau| \leqq h^2$, $L \leqq H_2$,

$$c_l = c_l(\tau, L) = \begin{cases} \dfrac{l^{i\tau}}{(l/L)^{1/2}} & \text{if} \quad L < l \leqq \min(2L, H_2) \\ 0 & \text{otherwise} \end{cases}$$

and

$$C_{\chi,t,\tau} = \sum\sum\sum a_m^* \chi(m) b_n^* \chi(n) c_l \chi(l)$$

we have

(28) $$\frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} |B_{\chi,t}| \ll q^\varepsilon \max_{L \leqq H^2} \max_{|\tau| \leqq h^2} \left(\frac{H_1}{L}\right)^{1/2} \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} |C_{\chi,t,\tau}|.$$

Finally we use the Cauchy—Schwartz inequality and the mean-square theorem to get

$$\frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} |C_{\chi,t,\tau}| \leqq \left(\frac{1}{\varphi(q)} \sum_{\chi} \left|\sum a_m^* \chi(m)\right|^2\right)^{1/2} \left(\frac{1}{\varphi(q)} \sum_{\chi} \left|\sum\sum b_n^* c_l \chi(nl)\right|^2\right)^{1/2} \ll$$

$$\ll q^\varepsilon \left(\left(1+\frac{M}{q}\right) M \left(1+\frac{NL}{q}\right) NL\right)^{1/2}.$$

Combining this with (28) and (27) we get

$$\frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} |B_\chi| \ll Y q^{-2\varepsilon}$$

provided (i) $\alpha \geqq 1 + 10\varepsilon$

(29) (ii) $H \leqq Y q^{1/2 - 5\varepsilon}$

(iii) $\dfrac{H}{Y} q^{8\varepsilon} \leqq N \leqq \dfrac{Y^2}{H} q^{-10\varepsilon}$.

In estimating B_χ^* we use $\mathscr{L}(1/2+it, \chi)$ instead of $\sum a_m \chi(m) m^{-1/2-it}$. Instead of (27) and (28) we have

$$\frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} |B_\chi^*| \ll q^\varepsilon \max_{T' \leqq Xq} \max_{L \leqq H_2} \max_{|\tau| \leqq h^2} \frac{\left(\dfrac{H_1 H}{NL}\right)^{1/2}}{T'+1} \times$$

$$\times \int_{T'}^{2T'+1} \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \left|\mathscr{L}(1/2+it) \sum b_n^* \chi(n) \sum c_l \chi(l)\right| dt + O(1).$$

We need the following lemma.

Lemma 5. Fourth-moment estimate (Montgomery, [12]). *For* $T \geqq 2$ *we have*

$$\sum_{\chi} \int_{-T}^{T} |\mathscr{L}(1/2+it, \chi)|^4 \, dt \ll qT \log^6 qT.$$

Using this with the mean-square theorem we arrive at

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi^*|\ll q^\varepsilon \max_{T'\leqq Xq}\max_{L\leqq H_2}\max_{|\tau|\leqq h^2}\left(\frac{X}{NL}\right)^{1/2}\times$$

$$\times\left(\frac{1}{T'+1}\int_{T'}^{2T'+1}\frac{1}{\varphi(q)}\sum_\chi|\mathcal{L}(1/2+it)\chi|^4\,dt\right)^{1/4}\times$$

$$\times\left(\frac{1}{T'+1}\int_{T'}^{2T'+1}\frac{1}{\varphi(q)}\sum_\chi\Big|\sum b_n^*\chi(n)\Big|^2\,dt\right)^{1/2}\times \tag{30}$$

$$\times\left(\frac{1}{T'+1}\int_{T'}^{2T'+1}\frac{1}{\varphi(q)}\sum_\chi\Big|\sum c_l\chi(l)\Big|^4\right)^{1/4}+O(1)\ll$$

$$\ll q^2\max_{L\leqq H_2}\left(\frac{X}{NL}\left(1+\frac{N}{q}\right)N\left(1+\frac{L^2}{q}\right)^{1/2}L\right)^{1/2}\ll q^{2\varepsilon}X^{1/2}\left(1+\frac{N}{q}\right)^{1/2}$$

provided $L^2\leqq q$ which follows from $H\leqq Yq^{1/2-5\varepsilon}$. These bounds give

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi^*|\ll Yq^{-2\varepsilon}$$

(31) provided (i) $\alpha\geqq 1+8\varepsilon$

(ii) $H\leqq Yq^{1/2-5\varepsilon}$

(iii) $N\leqq Yq^{-8\varepsilon}$.

We get the required bound (24) by substituting (29) and (31) into (25) provided we have a good choice of U satisfying $U^2\leqq Yq^{-8\varepsilon}$, $\frac{H}{Y}q^{8\varepsilon}\leqq U$, and $H^{1/2}\leqq\frac{Y^2}{H}q^{8\varepsilon}$. Under the conditions of (18), $U=Y^{1/2}q^{-4\varepsilon}$ is a good choice.

Next we turn to the proof of (19). We use the simpler weights $w_2(l)$ and we can assume $H\geqq Yq^{1/2-5\varepsilon}$; otherwise (18) gives the required result. We follow the above arguments except that we use (17) instead of the reflection, that is we have (cf. (28)) with $\omega=\omega(\varepsilon_0)$

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_{\chi,t}|\ll q^{-\omega}\frac{X}{H}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}\Big|\sum\sum a_m^*\chi(m)b_n^*\chi(n)\Big|+O(1)\ll$$

$$\ll q^{-\omega+\varepsilon}\frac{X}{H}\left(\left(1+\frac{M}{q}\right)M\left(1+\frac{N}{q}\right)N\right)^{1/2}.$$

From (27) we get (if $3\varepsilon<\omega$)

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi|\ll Yq^{-2\varepsilon}$$

(32) provided (i) $H\leqq Xq^{-\gamma-\varepsilon_0}$

(ii) $\min(M,N)\geqq q$.

For the special bilinear forms we have

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi^*|\ll O(1)+$$

$$+\log q\max_{T'\leqq X_q}\frac{(H/N)^{1/2}}{T'+1}\int_{T'}^{2T'+1}\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}\left|\mathscr{L}(1/2+it,\chi)\sum b_n^*\chi(n)\sum_{l\leqq H_1}\chi(l)\right|dt.$$

The same argument as (30) gives

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi^*|\ll q^{\varepsilon/2}\left(\frac{H}{N}\left(1+\frac{N}{q}\right)N\left(1+\frac{H_1^2}{q}\right)^{1/2}H_1\right)^{1/2}\ll q^{3\varepsilon}\left(X\left(1+\frac{N}{q}\right)\right)^{1/2}$$

provided $H_1^2\ll q^{1+10\varepsilon}$ which follows from $H\geqq Yq^{1/2-5\varepsilon}$. These bounds give

$$\frac{1}{\varphi(q)}\sum_{\chi\neq\chi_0}|B_\chi^*|\ll Yq^{-2\varepsilon}$$

(33) provided (i) $\alpha\geqq 1+10\varepsilon$

(ii) $Yq^{1/2-5\varepsilon}\leqq H\leqq Yq^{1-\varepsilon}$

(iii) $N\leqq Yq^{-10\varepsilon}$.

We get (24) by substituting (32) and (33) into (25) because under the given conditions $U=Y^{1/2}q^{-5\varepsilon}$ is a good choice of the parameter, if $\alpha\geqq 2+10\varepsilon$.

6. Sieve estimates and $\mathscr{L}$-functions

In this section we combine the $\mathscr{L}$-function techniques of § 5 with the Rosser—Iwaniec sieve to get upper bounds for $\sum A(X,p)$. Our main results are the following.

(34) PROPOSITION. *Let* $\varepsilon_0>0$ *and* $1<\alpha\leqq 3/2$ *be given. For all sufficiently large* q *and*

$$q^{-\varepsilon_0}Y^{4/3}\leqq R<2R\leqq S\leqq q^{-\varepsilon_0}Yq^{1/2}$$

we have

$$\sum_{R<p\leqq S}A(X,p)\log p\leqq(1+10\varepsilon_0)Y\int_R^S\frac{\log y}{y\log\frac{Y^2}{y}}\,dy.$$

(35) PROPOSITION. *Let* $\varepsilon_0>0$ *and* $\alpha>1$ *be given. For all sufficiently large* q *and*

$$Y_q^{1/2-\varepsilon_0}\leqq R<2R\leqq S<Xq^{-\varepsilon_0}$$

we have

$$\sum_{R<p\leqq S}A(X,p)\log p\leqq(1+10\varepsilon_0)\frac{\log^2 S-\log^2 R}{\log Y}Y.$$

Proof of (34) and (35). Let $\{\lambda_d | d<D\}$ denote the upper bound Rosser—Iwaniec weights of the sieve $S(\mathcal{A}, \mathcal{P}_q, z)$ with distribution level D. ($\mathcal{P}_q$ is the set of primes not dividing q.) We have

$$\mathcal{T} = \mathcal{T}(H) = \sum_{H<p\leqq\eta H} A(X,p)\log p \leqq$$

$$\leqq (1+\varepsilon)\log H \sum_{l=1}^{\infty} w(l) \sum_{\substack{H<p\leqq\eta H \\ pl\equiv a(q)}} 1 \leqq$$

$$\leqq (1+\varepsilon)\log H \sum_{l=1}^{\infty} w(l) \sum_{\substack{H<n\leqq\eta H \\ nl\equiv a(q)}} \sum_{d|n} \lambda_d =$$

$$= (1+\varepsilon)\log H \sum_{d<D} \lambda_d \sum_{H<md\leqq\eta H} \sum_{\substack{l=1 \\ mdl\equiv a(q)}}^{\infty} w(l)$$

where η and $w(l)$ are defined in § 5. We use again the orthogonality of characters. The principal character gives the main term of the sieve and the required results follow from

$$\frac{1}{\varphi(q)} \sum_{\chi\neq\chi_0} |\mathcal{T}_\chi| \ll Yq^{-\varepsilon}$$

where

$$\mathcal{T}_\chi = \sum_{d<D} \sum_{H<dk\leqq\eta H} \sum_{l=1}^{\infty} \lambda_d \chi(d)\chi(k)w(l)\chi(l)$$

and $D=Y^4H^{-2}q^{-20\varepsilon}$ for (34), $D=Yq^{-8\varepsilon}$ for (35). $\mathcal{T}_\chi$ has the same shape (after a splitting up argument) as B_χ^* thus choosing $w(l)=w_2(l)$ we get (35) from (33). To prove (34) we choose $w(l)=w_1(l)$ and now we use the bilinear form [8] of the weights λ_d. We have to investigate

$$\sum_{M<m\leqq 2M} \sum_{N<n\leqq 2N} \sum_{H<kmn\leqq\eta H} \sum_{l=1}^{\infty} a_m\chi(m)b_n\chi(n)\chi(k)w(l)\chi(l)$$

where $|a_m|\leqq 1$, $|b_n|\leqq 1$ and $M, N\leqq \frac{Y^2}{H}q^{-10\varepsilon}$. When $MN\leqq Yq^{-8\varepsilon}$ then collecting mn (31) gives the required bound. Otherwise at least one of M and N, say N satisfies $Y^{1/2}q^{-4\varepsilon}\leqq N\leqq Y^2H^{-1}q^{-10\varepsilon}$. Now collecting mk (29) gives the required bound provided $\frac{H}{Y}q^{8\varepsilon}\leqq Y^{1/2}q^{-4\varepsilon}$ which follows from $H\leqq Yq^{1/2-12\varepsilon}\leqq Y^{3/2}q^{-12\varepsilon}$ as $\alpha>1$. This completes the proof.

7. Conclusion

Combining the propositions (18), (19), (34) and (35) we get

(36) PROPOSITION. *Let* $\varepsilon_0>0$ *and* $\alpha>1>\delta>0$ *be given. For all sufficiently large* q *we have*

$$\sum_{q^\alpha<p\le q^{\alpha+\delta}} A(X,p)\log p \le (1+\varepsilon_0)Y\log q\int_\alpha^{\alpha+\delta} F(\alpha,\beta)\,d\beta$$

where $F(\alpha,\beta)$ *is defined by*

$$F(\alpha,\beta)=\left\{\begin{array}{ll|c|c} & 1<\alpha\le 3/2 & 3/2<\alpha\le 2 & 2<\alpha \\ \hline 1 & \alpha\le\beta<4\alpha/3 & \alpha\le\beta<\alpha+\frac{1}{2} & \alpha\le\beta<\alpha+\frac{5}{8} \\ \frac{\beta}{2\alpha-\beta} \quad \text{if} & 4\alpha/3\le\beta<\alpha+\frac{1}{2} & \dots & \dots \\ \frac{2\beta}{\alpha} & \alpha+\frac{1}{2}\le\beta<\alpha+1 & \alpha+\frac{1}{2}\le\beta<\alpha+1 & \alpha+\frac{5}{8}\le\beta<\alpha+1. \end{array}\right.$$

If q is cube-free then 5/8 can be replaced by 3/4 in the last column. Thus Theorem 1 is true for all $\alpha>1$, $\delta(\alpha)<\delta^*(\alpha)$ where $\delta^*(\alpha)$ is defined by

$$\int_\alpha^{\alpha+\delta^*} F(\alpha,\beta)\,d\beta=1.$$

From Propositions (10) and (36) a simple computation yields

THEOREM 3. *The conclusion of Theorem 1 is valid for any* $\alpha>1/2$ *and any* $\delta\le\delta^*$ *where* $\delta^*=\delta^*(\alpha)$ *is defined by*

$$\delta^*(\alpha)=\begin{cases}\sqrt{\alpha^2+\alpha-\frac{1}{2}}-\alpha & \text{if } \frac{1}{2}<\alpha\le 1\\ \sqrt{\frac{1}{3}\alpha^2+\frac{5}{2}\alpha+\frac{1}{4}-2\alpha^2\log\frac{4\alpha}{6\alpha-3}}-\alpha & \text{if } 1<\alpha\le\frac{3}{2}\\ \sqrt{\alpha^2+\frac{3}{2}\alpha+\frac{1}{4}}-\alpha & \text{if } \frac{3}{2}<\alpha\le 2\\ \sqrt{\alpha^2+\frac{13}{8}\alpha+\frac{25}{64}}-\alpha & \text{if } 2<\alpha.\end{cases}$$

Finally we note that there are some possible ways to improve the value $\delta^*(\alpha)$ for $\alpha>1$. In using the linear sieve we can compute the contribution of P_2's as well (see Balog—Harman—Pintz, [1]) and in bounding the bilinear forms we can use the Halász—Montgomery—Huxley inequality (see Iwaniec [9]).

REFERENCES

[1] Balog, A., Harman, G. and Pintz, J., Numbers with a large prime factor IV, *J. London Math. Soc.* **28** (1983), 218—226.

[2] Burgess, D. A., On character sums and $\mathscr{L}$-series I.—II, *Proc. London Math. Soc.* **12** (1962), 193—206. *MR* **24**# A2570; **13** (1963), 524—536. *MR* **26**# 6133.

[3] Friedlander, J. and Iwaniec, H., Quadratic polynomials and quadratic forms, *Acta Math.* **141** (1978), 1—15. *MR* **57**# 16232.

[4] Heath-Brown, D. R., The least square-free number in an arithmetic progression, *J. Reine Angew. Math.* **332** (1982), 204—220. *MR* **83i:** 10057.

[5] Hooley, C., On the Brun—Titchmarsh theorem, *J. Reine Angew. Math.* **255** (1972), 60—79. *MR* **46**# 3463.

[6] Hooley, C., *Applications of Sieve Methods to the Theory of Numbers*, Cambridge Tracts in Mathematics, 70, Cambridge, 1976. *MR* **53**# 7976.

[7] Hooley, C., On the greatest prime factor of a cubic polynomial, *J. Reine Angew. Math.* **303/304** (1978), 21—50. *MR* **80b:** 10061.

[8] Iwaniec, H., A new form of the error term in the linear sieve, *Acta Arith.* **37** (1980), 307—320. *MR* **82d:** 10069.

[9] Iwaniec, H., On the Brun—Titchmarsh theorem, *J. Math. Soc. Japan* **34** (1982), 95—123. *MR* **83a:** 10082.

[10] Jutila, M., On numbers with a large prime factor I—II, *J. Indian Math. Soc.* **37** (1973), 43—53. *MR* **50**# 12936; **38** (1974), 125—130. *MR* **53**# 2864.

[11] Jutila, M., Zero density estimates for $\mathscr{L}$functions, *Acta Arith.* **32** (1977), 55—62. *MR* **55**# 2800.

[12] Montgomery, H. L., *Topics in Multiplicative Number Theory*, Lecture Notes in Mathematics, Vol. 227, Springer, Berlin, 1971. *MR* **49**# 2616.

[13] Ramachandra, K., A note on numbers with a large prime factor II, *J. Indian Math. Soc.* **34** (1970), 39—48. *MR* **45**# 8616.

[14] Vaughan, R. C., Sommes trigonométriques sur les nombres premiers, *C. R. Acad. Sci. Paris, Ser. A.* **285** (1977), 981—983. *MR* **58**# 16555.

[15] Vinogradov, I. M., *The Method of Trigonometrical Sums in the Theory of Numbers*, Interscience, London, 1954. *MR* **15**—941.

(Received July 2, 1984)

A. Balog and J. Pintz:

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

J. Friedlander:

SCARBOROUGH COLLEGE
UNIVERSITY OF TORONTO
SCARBOROUGH, ONTARIO
M1C 1A4
CANADA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 189—195

0-MODULAR SEMILATTICES

C. JAYARAM

Summary

J. Varlet [5] has introduced the concept of 0-modular lattice and proved that a lattice L with 0 (0 is the least element in L) is 0-modular if and only if there exists in L, no non-modular five element sublattice including the element 0. The dual of 0-modularity is 1-modularity (1 is the greatest element). In this paper, we define 0-modular meet semilattices and prove that a bounded meet semilattice L is 0-modular if and only if the lattice $F(L)$ of all filters of L is 1-modular. Using this we establish some necessary and sufficient conditions for a bounded meet semilattice to be a finite Boolean algebra.

1. Preliminaries

Throughout this paper we shall be concerned with meet semilattices. We use the symbols $\vee$ (join) and $\wedge$ (meet) to indicate the least upper bound and the greatest lower bound of a subset when they exist. The least (greatest) element is denoted by 0 (1), when it exists.

A semilattice L with 0 is 0-distributive if, for any $a, b, c \in L$, $a \wedge b = 0 = a \wedge c$ implies $a \wedge d = 0$ for some $d \geqq b, c$. A semilattice L with 0 is called weakly 0-distributive if for any a, b, c in $L, a \wedge b = 0 = a \wedge c$ implies $a \wedge (b \vee c) = 0$ whenever $b \vee c$ exists in L. A semilattice L with 1 is 1-distributive if for any $a, b, c \in L$ $a \vee b$ and $a \vee c$ exist and equals 1, then $a \vee (b \wedge c)$ exists and equals 1. A semilattice L with 0 is semicomplemented if for any $a \in L$ ($a \neq 1$, if 1 exists), there is $b \in L$, $b \neq 0$ such that $a \wedge b = 0$. Dually a semilattice L with 1 is dual semicomplemented if for any $a \in L$ ($a \neq 0$, if 0 exists), there is $b \in L$, $b \neq 1$ such that 1 is the only upper bound of a and b. A semilattice L with 0 and 1 is complemented if for any a in L, there is b in L such that $a \wedge b = 0$ and 1 is the only upper bound of a and b. A semilattice L with 0 is weakly complemented if for any $a < b$, there is $c \in L$ such that $a \wedge c = 0$ and $b \wedge c \neq 0$. A lattice L with 0 (1) is 0-modular (1-modular) if for any $a, b, c \in L$, $a \leqq c$ ($a \geqq c$) and $b \wedge c = 0$ ($b \vee c = 1$) imply $(a \vee b) \wedge c = a$ $\big((a \wedge b) \vee c = a\big)$.

An element a of a semilattice L is meet prime if $b \wedge c \leqq a$ implies $b \leqq a$ or $c \leqq a$. A non-zero element x of a semilattice L with 0 is an atom if for any $y \in L$, $0 \leqq y \leqq x$ implies $0 = y$ or $y = x$. Dually, an element x ($x \neq 1$) of a semilattice L with 1 is a dual atom if $x \leqq y \leqq 1$ implies $x = y$ or $y = 1$.

A filter of a semilattice L is a non-empty subset F of L such that $x \wedge y \in F$ if and only if $x \in F$ and $y \in F$. The principal filter generated by an element $a \in L$, that is, the set $\{x \in L | x \geqq a\}$, is denoted by $[a)$. When ordered by inclusion, the set $F(L)$ of all filters of an up-directed semilattice L is a lattice called the filter lattice

1980 *Mathematics Subject Classification.* Primary 06-XX; Secondary 06A12.
Key words and phrases. 0-distributive semilattice, 0-modular semilattice, finite Boolean algebras.

of L. The lattice operations in $F(L)$ are denoted by $\vee$ and $\cap$, respectively. A filter F of L is prime if and only if it is a meet prime element of $(F(L), \cap)$. A proper filter F is maximal if the only filter strictly containing F is L. In (6) it is proved that a bounded semilattice is 0-distributive if and only if any maximal filter is meet prime.

Throughout this paper, L will denote a meet semilattice. All filters of L are assumed to be proper. Set inclusion and intersection are denoted by $\subseteq$ and $\cap$ respectively.

2. 0-modular semilattices

DEFINITION. A semilattice L with 0 is called 0-modular if $a \leqq c$ and $b \wedge c = 0$ $(a, b, c \in L)$ imply that there exists d in L such that $b \leqq d$ and $a = c \wedge d$.

Obviously, every modular semilattice is 0-modular. Now we give an example of a 0-modular semilattice which is non-modular.

EXAMPLE. Let $L_1 = \{0, a, b, c, d, e, f, g, a_0, a_1, a_2, \ldots, a_n, \ldots\}$ be a semilattice as shown in Fig 1. The arrow indicates an infinite chain. Let us observe that L_1 is not a lattice as e and g have no least upper bound. It can be easily seen that L_1 is a bounded 0-modular semilattice. It can also be observed that L_1 is neither 0-distributive nor modular.

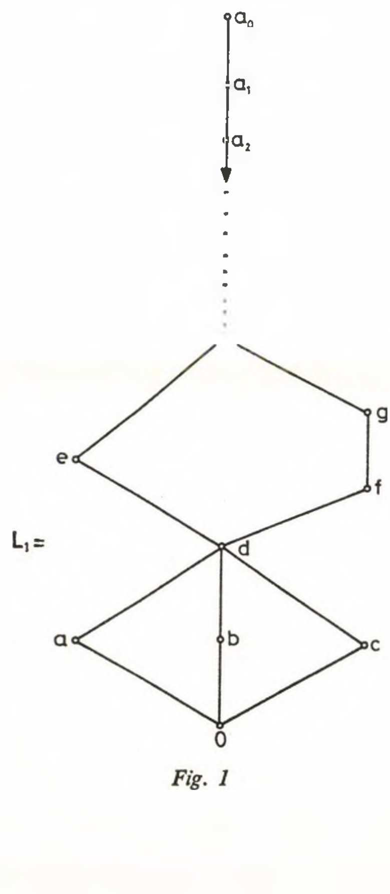


Fig. 1

It should be mentioned that our definition of 0-modularity coincides with that of Varlet [5] in a lattice.

Now we characterize 0-modularity as follows:

THEOREM 1. *A bounded semilattice L is 0-modular if and only if the lattice $F(L)$ of all filters of L is 1-modular.*

PROOF. Suppose L is 0-modular. Let F_1, F_2, F_3 be any three filters of L such that $F_1 \vee F_3 = [0)$ and $F_1 \subseteq F_2$. Clearly $F_1 \vee (F_3 \cap F_2) \subseteq F_2$. Let $x \in F_2$. Since $F_1 \vee F_3 = [0)$, it follows that $0 = y \wedge z$ for some $y \in F_1$ and $z \in F_3$. As $x \wedge y < y$ and $y \wedge z = 0$, by 0-modularity $x \wedge y = y \wedge z_1$ for some $z_1 \geqq z$. Observe that $z_1 \geqq x \wedge y \in F_2$ and so $z_1 \in F_3 \cap F_2$. Now $x \geqq x \wedge y = y \wedge z_1 \in F_1 \vee (F_3 \cap F_2)$ and therefore $F_1 \vee (F_3 \cap F_2) = F_2$. Thus $F(L)$ is a 1-modular lattice.

Conversely, suppose $a \leqq c$ and $b \wedge c = 0$ for some $a, b, c \in L$. Then $[b,) \vee [c,) = [0)$ and $[c) \subseteq [a)$, so that by hypothesis $[c) \vee ([b) \cap [a)) = [a)$. Consequently $a = c \wedge d$ for some $d \geqq b$. Thus L is 0-modular. This completes the proof of the theorem.

We now prove some lemmas to characterize finite Boolean algebras.

LEMMA 1. (a) *Every bounded semicomplemented and weakly 0-distributive semilattice is 1-distributive.* (b) *Every bounded dual semicomplemented and 1-distributive semilattice is weakly 0-distributive.*

PROOF. (a) Suppose $x \vee y$, $x \vee z$ $(x, y, z \in L)$ exist and are equal to 1. Let $x \leqq v$ and $y \wedge z \leqq v$ for some $v \in L$. Assume that $v \neq 1$. Since L is semicomplemented, we have $v \wedge a = 0$ for some $a \neq 0$, $a \in L$. Clearly $x \wedge a = 0 = (y \wedge z) \wedge a$. As L is weakly 0-distributive, it follows that $(a \wedge y) \wedge (x \vee z) = 0$ and so $a \wedge y = 0$. Again by weakly 0-distributivity, $0 = a \wedge (x \vee y) = a \wedge 1 = a$, a contradiction. Therefore 1 is the only upper bound of x and $y \wedge z$. Hence L is 1-distributive.

(b) Let $x \wedge y = 0 = x \wedge z$ and $y \vee z$ exists for $x, y, z \in L$. We have to show that $x \wedge (y \vee z) = 0$. Assume that $x \wedge (y \vee z) \neq 0$. Since L is dual semicomplemented, it follows that $(x \wedge (y \vee z)) \vee w$ exists and is equal to 1 for some $w \neq 1$, $w \in L$. Clearly $(w \vee x)$ and $(y \vee z) \vee w$ exist and are equal to 1. We claim that for any $a \in L$, if $a \geqq w$ and $a \geqq y$, then $a \vee z$ exists and is equal to 1. Suppose $a \geqq w, y$. If $m \geqq a, z$, then $m \geqq w, y \vee z$ and so $m = 1$. Therefore $a \vee z = 1$. Similarly, $a \vee x = 1$ and hence, by 1-distributivity, $a \vee (x \wedge z)$ exists and is equal to 1. As $x \wedge z = 0$, it follows that $a = 1$. Thus any upper bound of w and y is equal to 1. This shows that 1 is the only upper bound of w and y. Again by 1-distributivity; $w \vee (x \wedge y)$ exists and is equal to 1. Since $x \wedge y = 0$, it follows that $w = 1$, a contradiction. Therefore $x \wedge (y \vee z) = 0$ and hence L is weakly 0-distributive. This completes the proof of the lemma.

LEMMA 2. *Every complemented 0-modular semilattice L which is also weakly 0-distributive is a Boolean algebra.*

PROOF. By Corollary 1 of (2) (Corollary 1 (2): every weakly 0-distributive complemented semilattice which is also weakly complemented is a Boolean algebra) it is enough if we prove that L is weakly complemented. Suppose $a < b$ for some $a, b \in L$. Let c be a complement of a in L. Obviously, $a \wedge c = 0$. If $b \wedge c = 0$, then by 0-modularity $a = b \wedge d$ for some $d \geqq c$. Since $d \geqq a, c$ it follows that $d = 1$ and so $a = b$ which is a contradiction. Therefore $b \wedge c \neq 0$ and hence L is weakly complemented.

LEMMA 3. *Let L be a bounded* 0*-modular semilattice. If F and J are any two filters of L such that $F \vee J = [0)$ and $F \cap J = [1)$, then both F and J are principal filters of L.*

PROOF. Suppose $F \vee J = [0)$ and $F \cap J = [1)$ for some F and $J \in F(L)$. As $F \vee J = [0)$, we have $0 = x \wedge y$ for some $x \in F$ and $y \in J$. Now we show that $F = [x)$ and $J = [y)$. Clearly $[x) \subseteq F$. Let $z \in F$. Since $z \wedge x \leqq x$ and $x \wedge y = 0$, by 0-modularity, $z \wedge x = x \wedge w$ for some $w \geqq y$. Again as $w \geqq z \wedge x \in F$, it follows that $w \in F \cap \cap J = [1)$ and so $w = 1$. Consequently $z \wedge x = x$ and hence $z \in [x)$. Therefore $F = [x)$. Similarly $J = [y)$. Thus F and J are principal filters of L.

LEMMA 4. (3): *In a bounded semicomplemented semilattice every meet prime element is a dual atom.*

LEMMA 5. *Let L be a bounded semicomplemented semilattice. If* 0 *is the meet of a finite number of meet prime elements in L, then L is dual semicomplemented and* 0*-distributive.*

PROOF. We first show that L is dual semi-complemented. Let x be a non-zero element of L. Then, by hypothesis, there is a meet prime element p in L such that $x \not\leqq p$. Since L is semicomplemented, by Lemma 4, p is a dual atom and so 1 is the only upper bound of x and p. Therefore L is dual semicomplemented.

Next we claim that L is 0-distributive. Suppose $a \wedge b = 0 = a \wedge c$ for some $a, b, c \in \in L$. Let us assume that $0 = \bigwedge_{i=1}^{n} p_i$ (p_i's are meet prime elements in L). Observe that for each i, $p_i \geqq a \wedge b$ and $p_i \geqq a \wedge c$, so that for each i, $p_i \in [a)$ or $p_i \in [b) \cap [c)$. Therefore each $p_i \in [a) \vee ([b) \cap [c))$ and hence $[a) \vee ([b) \cap [c)) = [0)$. Consequently, $0 = a \wedge d$ for some $d \geqq b, c$. Hence L is 0-distributive.

LEMMA 6. *Let L be a bounded* 0*-modular semilattice. If $b \in L$ is a dual atom and $a \wedge b = 0$ for some $a \neq 0$, $a \in L$, then a is an atom.*

PROOF. Suppose $0 < c \leqq a$ for some $c \in L$. As $c \leqq a$ and $a \wedge b = 0$, by 0-modularity, $c = a \wedge d$ for some $d \geqq b$. Since $0 < c$, it follows that $b < d$ and so $d = 1$ as b is a dual atom. Consequently, $c = a$ and hence a is an atom.

LEMMA 7. *Let L be a bounded semicomplemented and* 0*-modular semilattice. If* 0 *is the meet of a finite number of meet prime elements in L, then* 1 *is the join of a finite number of atoms in L.*

PROOF. Let L be a bounded semicomplemented 0-modular semilattice. Assume that $0 = \bigwedge_{i=1}^{n} p_i$, where p_i's are meet prime elements in L. Observe that by Lemma 4, each p_i is a dual atom. Since each $p_i \neq 1$ and L is semicomplemented, there exist $q_i \neq 0$, $q_i \in L$ such that $p_i \wedge q_i = 0$ for $i = 1, 2, \ldots, n$. Also by Lemma 6, each q_i is an atom. Now we show that $\bigvee_{i=1}^{n} q_i$ exists and is equal to 1. Let $c \geqq q_i$ for $i = = 1, 2, \ldots, n$. Then for each $i \in \{1, 2, \ldots, n\}$, $c \vee p_i$ exists and equals 1. As L is bounded semicomplemented and 0 is the meet of a finite number of meet primes, by Lemma 5, L is 0-distributive and so again by Lemma 1, L is 1-distributive. Therefore $c =$

$=c\vee(\bigwedge_{i=1}^{n} p_i)$ is equal to 1. This shows that $1=\bigvee_{i=1}^{n} q_i$. Hence 1 is the join of a finite number of atoms.

The following theorem gives a necessary and sufficient condition for a bounded semilattice to be a finite Boolean algebra.

THEOREM 2. *A bounded semilattice L is a finite Boolean algebra if and only if L satisfies the following three conditions:*

(a) *L is* 1*-distributive,*
(b) *L is* 0*-modular,*
(c) *$F(L)$ is semicomplemented.*

PROOF. Obviously, every finite Boolean algebra satisfies the conditions (a), (b) and (c) of Theorem 2. Conversely, assume that L satisfies (a), (b) and (c). We first show that every filter of L is principal. Let F be a filter of L. Put $D(F)=\{x\in L|x\vee y$ exists and is equal to 1 for all $y\in F\}$. As L is 1-distributive, it follows that $D(F)$ is a filter of L. Observe that $[0)$ is the greatest element of $F(L)$ and $[1)$ is the least element of $F(L)$. We claim that $F\vee D(F)=[0)$. If $F\vee D(F)\neq[0)$, then by condition (c), there is a filter $J\in F(L)$ such that $(F\vee D(F))\cap J=[1)$ and $J\neq[1)$. Clearly $F\cap J=[1)$ and $D(F)\cap J=[1)$. Since $F\cap J=[1)$, it follows that $J\leqq D(F)$ and so $J=[1)$, a contradiction. Therefore $F\vee D(F)=[0)$. Also $F\cap D(F)=[1)$ and so by condition (b) and Lemma 3, both F and $D(F)$ are principal filters. Thus every filter of L is principal.

Next we claim that L is complemented. Let $x\in L$ and $(D(x)=\{t\in L'|x\vee t$ exists and is equal to $1\})$. Then by above observation, $[x)\vee D(x)=[0)$, $[x)\cap D(x)=[1)$ and $D(x)$ is a principal filter of L; say $D(x)=[y)$ for some $y\in L$. Clearly y is a complement of x in L. Therefore L is complemented.

Now as L is 0-modular, 1-distributive and complemented, by Lemmas 1 and 2, L is a Boolean algebra. Thus L is a Boolean algebra in which every filter is principal. Consequently, L is a finite Boolean algebra. This completes the proof of the theorem.

REMARK 1. We now give examples to show that the conditions (a), (b) and (c) of Theorem 2 are independent.

EXAMPLE A. Let S be the five element non-modular lattice. It can be easily seen that S is non 0-modular. It can also be observed that S is 1-distributive and $F(S)$ (the lattice of all filters of S) is semicomplemented.

EXAMPLE B. Let S be a finite chain with more than two elements. Obviously, $F(S)$ is non semicomplemented. It can be easily verified that S is 0-modular and 1-distributive.

EXAMPLE C. Let S be the five element modular non-distributive lattice. Clearly, S is 0-modular, and $F(S)$ is semicomplemented. Also it is easily seen that S is not 1-distributive.

THEOREM 3. *A bounded semilattice L is a finite Boolean algebra if and only if L satisfies the following three conditions:*

(i) *L is semicomplemented,*
(ii) *L is* 0*-modular,*

(iii) *0 is the meet of a finite number of meet primes.*

Proof. Clearly, every finite Boolean algebra satisfies the conditions (i), (ii) and (iii) of Theorem 3. Conversely, let us assume that L satisfies the conditions (i), (ii) and (iii) of Theorem 3. We first show that every maximal filter of L is a principal filter. Let M be a maximal filter of L. Since L satisfies the conditions (i), (ii) and (iii), by Lemma 7, 1 is the join of a finite number of atoms in L, say $1=\bigvee_{i=1}^{n} q_i$ (q_i's are atoms in L). As L is 0-distributive (by Lemma 5), M is a prime filter (see (6)) and so $q_i \in M$ for some $i \in \{1, 2, \ldots, n\}$. Consequently, $M=[q_i)$. Thus M is a principal filter.

Next we claim that L is complemented. Let $x \in L$. Put $T(x)=\{y \in L \mid x \wedge y=0\}$ and $D(x)=\{y \in L \mid 1$ is the only upper bound of x and $y\}$. If $T(x) \cap D(x) \neq \emptyset$ then we are through. So assume that $T(x) \cap D(x)=\emptyset$. Let $\mathcal{J}=\{J \mid J$ is a filter of L such that $D(x) \subseteq J$ and $J \cap T(x)=\emptyset\}$. Observe that by Lemmas 1 and 5, L is 1-distributive and so $D(x)$ is filter of L. Therefore $\mathcal{J} \neq \emptyset$. Now by Zorn's lemma, $\mathcal{J}$ has a maximal element, say F. Now one can easily see that $x \in F$ and F is a maximal filter of L (for proof see (4), Theorem 1, (ii)$\Rightarrow$(i)). As F is a maximal filter, we have $F=[a)$ for some atom $a \in L$. As $a \neq 0$, by (iii) we have $a \nleq b$ for some dual atom $b \in L$. Obviously, $a \wedge b=0$ and 1 is the only upper bound of a and b and so 1 is the only upper bound of x and b as $x \in F=[a)$. Therefore $b \in D(x) \subseteq F=[a)$ which is a contradiction. So $T(x) \cap D(x) \neq \emptyset$. This shows that L is complemented.

Again since L is complemented, by Lemmas 2 and 5, L a Boolean algebra. Again by dual of Theorem 1 of (1), every filter is a principal filter and hence L is a finite Boolean algebra. This completes the proof of the theorem.

The following examples show that the conditions assumed in Theorem 3 are independent.

Example D. Let S be the five element non-modular lattice. It can be easily seen that S satisfies the conditions (i) and (iii) of Theorem 3. It can also be observed that S is non 0-modular.

Example E. Let S be the five element modular non-distributive lattice. Obviously, S is semicomplemented and S is 0-modular. Also it is easily seen that the condition (iii) of Theorem 3 is not satisfied in S.

Example F. Let $S=\{0, a, b, c, 1\}$ with $0<a<c<1$ and $0<b<c$ be a lattice. Clearly, S satisfies the conditions (ii) and (iii) of Theorem 3. Also S is non semicomplemented.

I would like to thank Professor Iqbalunnisa for her valuable suggestions in the preparation of this paper.

REFERENCES

[1] CHANDRAN, V. R. and LAKSER, H., Distributive lattices whose prime ideals are principal, *Acta Sci. Math. (Szeged)* 39 (1977), 245. *MR* **57**#5847.
[2] JAYARAM, C., Complemented semilattices, *Math. Sem. Notes Kobe Univ.* **8** (1980), 259—267. *MR* **82a**: 06004.
[3] JAYARAM, C., Semilattices and finite Boolean algebras, *Algebra Universalis* **16** (1983), 390—394.
[4] RAMANA MURTHY, P. V., Ideal topology on a distributive lattice, *J. Austral. Math. Soc.* **18** (1974), 503—508. *MR* **51**#10191.
[5] VARLET, J., A generalization of the notion of pseudocomplementedness, *Bull. Soc. Roy. Sci. Liège* **37** (1968), 149—158. *MR* **37**#3971.
[6] VARLET, J., Distributive semilattices and Boolean lattices, *Bull. Soc. Roy. Sci. Liège* **41** (1972), 5—10. *MR* **46**#7106.

(Received July 5, 1984)

RAMANUJAN INSTITUTE FOR
ADVANCED STUDY IN MATHEMATICS
UNIVERSITY OF MADRAS
MADRAS—600 005
INDIA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 197—202

ITERATED DIFFERENCE SETS

IMRE Z. RUZSA[1]

1. Introduction

For a set A of natural numbers let

$$\Delta A = \{a-b:\ a\in A,\ b\in A,\ a > b\}$$

its difference set. Stewart and Tijdeman [4] proved that starting with any set of positive upper density and iterating this operation eventually we obtain the set of multiples of some number q, and this happens at most in

$$2\,{}^{2}\!\log \bar{d}(A)^{-1}$$

steps ($\bar{d}$, $\underline{d}$ denote upper, resp. lower asymptotic density).

If specially $A=\{a:a\equiv 0 \text{ or } 1 \pmod m\}$, then $\Delta^k A=\mathbf{N}$ if and only if $2^k \geqq m-1$. Our aim is to prove Tijdeman's conjecture that this example is extremal.

We introduce also the "density difference set"

$$\Delta_1 A = \{u:\ \bar{d}(A\cap(A+u)) > 0\}.$$

It will be convenient to consider Δ_1 together with Δ. We introduce the notation

$$Xm = \{xm:\ x\in X\},$$

to distinguish it from

$$mX = X+\ldots+X \quad (m \text{ summands}).$$

Let $q(A)$ be the greatest common divisor of the elements of ΔA and $q_1(A)$ the g.c.d. of the elements of $\Delta_1 A$. $q(A)$ can also be interpreted as the maximal number q such that A is contained in a single residue class modulo q. $q_1(A)$ is, however, not the maximal q for which A is almost contained in a residue class. E. g. if A consists of the even numbers in $(2^{2k}, 2^{2k+1})$ and of the odd numbers in $(2^{2k+1}, 2^{2k+2})$ for all k, then $\Delta_1 A$ is the set of even numbers, thus $q_1(A)=2$, but A is fairly equally divided between the modulo 2 residue classes.

THEOREM. a) *If* $(1+2^{k-1})\bar{d}(A)q(A)>1$, *then* $\Delta^k A=\mathbf{N}q(A)$.
b) *If* $(1+2^{k-1})\bar{d}(A)q_1(A)>1$, *then* $\Delta_1^k A=\mathbf{N}q_1(A)$.

---

[1] Paper finished when author was visiting the University of Ulm with a fellowship of the Alexander von Humboldt foundation.

1980 *Mathematical Subject Classification.* Primary 10L02; Secondary 10L05, 10J99.
Key words and phrases. Difference sets, addition of sets, additive number theory.

Thus, in terms of $\overline{d}(A)$, the best possible bound for the number of required steps is

$$2+[{}^{2}\log(\overline{d}(A)^{-1}-1)] \tag{1.1}$$

for $\overline{d}(A)\leqq 1/2$, and obviously 1 for $\overline{d}>1/2$.

2. Sums and differences

The density behaviour of sums is, due to the work of Kneser [2] (see also Halberstam—Roth [1]), rather well understood. Namely we know that $\underline{d}(A+B)\geqq \geqq \underline{d}(A)+\underline{d}(B)$ with certain well described exceptions. More exactly we have

LEMMA 1. *Let $A_1, \dots, A_k$ be sets of natural numbers, $B=A_1+\dots+A_k$. Either*

$$\underline{d}(B) \geqq \sum \underline{d}(A_i), \tag{2.1}$$

or there is a modulus m such that A_i is contained in some α_i residue classes modulo m, $B=B'\setminus C$ where B' is the union of certain β residue classes modulo m and C is finite, and

$$\beta \geqq \sum \alpha_i-(k-1). \tag{2.2}$$

For our purpose we deduce

LEMMA 2. *For every set A of natural numbers we have*

$$\underline{d}(kA) \geqq \min\left(q(A)^{-1},\ \frac{k+1}{2}\underline{d}(A)\right). \tag{2.3}$$

PROOF. Apply Lemma 1 with $A_i=A$. If (2.1) holds, we are ready. If not, and the modulus is $m\leqq q(A)$, then

$$\underline{d}(B) = \beta/m \geqq 1/m \geqq 1/q(A).$$

If $m>q(A)$, then A is not in a single residue class, thus $\alpha_i=\alpha\geqq 2$, hence by (2.2)

$$\underline{d}(B) = \beta/q \geqq \frac{k\alpha-(k-1)}{q} \geqq \frac{(k+1)\alpha}{q} \geqq \frac{k+1}{2}\underline{d}(A).$$

No similar result is available for differences; one would at once lead to our Theorem. To prove such an analogue seems to be a hard task; certainly Kneser's method does not work.

There is, however, some connection between sums and iterated differences. Namely let

$$DA = A-A = \Delta A\cup(-\Delta A)\cup\{0\}$$

be the *complete difference set* of A. It is easy to see by induction that

$$D^kA = D(2^{k-1}A) = 2^{k-1}D(A), \tag{2.4}$$

thus any information concerning sums can be utilized.

We can indeed easily establish the stability of $D^k(A)$.

LEMMA 3. *If* $(2^{k-1}+1)\underline{d}(A)q(A)>1$, *then* $D^kA=\mathbf{Z}q(A)$.

PROOF. Evidently, $D^kA\subset \mathbf{Z}q$ $(q=q(A))$. On the other hand, by Lemma 2

$$\underline{d}(2^{k-1}A) \geqq \min\left(q^{-1}, \frac{2^{k-1}+1}{2}\underline{d}(A)\right) > 1/(2q).$$

Hence for any $a\equiv 0 \pmod q$, the sets $2^{k-1}A$ and $2^{k-1}A+a$, both contained in $\mathbf{Z}q$, cannot be disjoint. This just means $a\in D(2^{k-1}A)$, i.e.

$$\mathbf{Z}q \subset D(2^{k-1}A) = D^kA$$

according to (2.4).

In the sequel we shall find connections between D, Δ and Δ_1 that finally lead to a proof of the Theorem.

3. Complete and density difference sets

Δ^kA generally does not contain all the positive elements of D^kA; an analogue of (2.4) does not hold. E.g. if $1\in A$ and $5\in A$, then $8=(5-1)-(1-5)\in D^2A$, but there is no guarantee that $8\in \Delta^2A$. This is, however, a rather exceptional phenomenon and does not hold for "typical" differences, as we shall see. The concept of a "typical" difference will be specified by the density difference set.

LEMMA 4. *If* $x, y\in \Delta_1A$, *then* $x, x+y\in \Delta_1^2A$, *and if* $x>y$, *then also* $x-y\in \Delta_1^2A$.

A reformulation of Lemma 4 is

$$\Delta\Delta_1 = \Delta_1^2,\quad 2\Delta_1\subset \Delta_1^2,\quad \Delta_1\subset \Delta_1^2. \tag{3.1}$$

PROOF. $x\in \Delta_1A$ means that there is a sequence S of positive upper density such that

$$S\subset A,\quad S+x\subset A.$$

Evidently, $\Delta_1S\subset \Delta_1A$; we shall see that also

$$(\Delta_1S)\pm x\subset \Delta_1A$$

(considering the positive elements only). Let $u\in \Delta_1S$ arbitrary. We have

$$A\cap(A+u+x) \supset (S+x)\cap(S+u+x) = \big(S\cap(S+u)\big)+x,$$

$$A\cap(A+u-x) \supset S\cap(S+x+u-x) = S\cap(S+u).$$

Since $\bar{d}(S\cap(S+u))>0$ by assumption, these sets have a positive upper density, which just means $u+x\in \Delta_1A$ and $u-x\in \Delta_1A$. Now

$$\Delta_1A\cap(\Delta_1A+x) \supset \Delta_1S+x$$

shows $x\in \Delta_1^2A$.

By the assumption $y\in \Delta_1A$ we get a sequence T of positive upper density such that $T\subset A$, $T+y\subset A$, and then a similar argument yields

$$\Delta_1T\subset \Delta_1A,\quad \Delta_1T\pm y\subset \Delta_1A.$$

By a theorem of Stewart and Tijdeman [3, Th. 1] the set

$$Z = \Delta_1 S \cap \Delta_1 T$$

has positive upper (even lower) density. Hence so do the sets

$$\Delta_1 A \cap (\Delta_1 A + x + y) \supset (Z+x) \cap (Z-y+x+y) = Z+x,$$

$$\Delta_1 A \cap (\Delta_1 A + x - y) \supset (Z+x) \cap (Z+y+x-y) = Z+x,$$

which shows that $x+y, x-y \in \Delta_1^2 A$, qu.e.d.

Lemma 5. *Writing*

$$Q_k = \Delta_1^k A \cup (-\Delta_1^k A) \cup \{0\}$$

we have $Q_k = D^{k-1} Q_1$.

Proof. $Q_{k+1} \subset DQ_k$ is obvious. On the other hand, every nonzero element of DQ_k is of one of the forms x, $-x$, $x+y$, $x-y$, $-x-y$ with $x, y \in \Delta_1^k A$. By Lemma 4 these numbers always belong either to $\Delta_1^{k+1} A$ or $-\Delta_1^{k+1} A$, and this shows $DQ_k \subset \subset Q_{k+1}$. The statement of the lemma follows by induction.

Remark. Though not all positive elements of $D^k A$ are in $\Delta^k A$, it is easy to see that they are contained in $\Delta^{k+1} A$ and this yields a bound 1 higher then (1.1). To remove this 1 we need these more refined considerations.

4. Iterated density-difference sets

Now we prove the part b) of the Theorem. We shall work with the

$$Q_k = \Delta_1^k A \cup (-\Delta_1^k A) \cup \{0\}$$

of Lemma 5 rather than with $\Delta_1^k A$ itself. Obviously, $\Delta_1^k A = \mathbf{N}q$ is equivalent to $Q_k = \mathbf{Z}q$.

By a theorem of Stewart and Tijdeman [3, Th. 5] there is a sequence B_0 such that

$$\Delta B_0 \subset \Delta_1 A, \quad \underline{d}(B_0) \geqq \bar{d}(A).$$

The first inclusion is equivalent to $DB \subset Q_1$. Now choose an arbitrary $u \in \Delta_1 A \setminus \Delta B_0$ if there is any. By $u \notin \Delta B_0$ the sets B_0 and $B_0 + u$ are disjoint, hence for

$$B_1 = B_0 \cup (B_0 + u)$$

we have $\underline{d}(B_1) = 2\underline{d}(B_0)$. Moreover

$$DB_1 = DB_0 + \{0, u, -u\} \subset Q_2$$

by Lemma 5. Now we repeat the whole process with (B_1, Q_2) instead of (B_0, Q_1). Sooner or later this process stops; then we have a B_j such that

$$\underline{d}(B_j) = 2^j \underline{d}(B_0), \quad DB_j \subset Q_{j+1},$$

but the next step is impossible because

$$DB_j = Q_{j+1}. \tag{4.1}$$

By (4.1) and Lemma 5 we obtain

$$Q_{j+r} = D^r B_j \quad (r \geqq 1).$$

By (4.1) we have $q(B_j)=q_1(Q_j)=q_1(A)$ (the second equality follows from Lemma 4), thus by Lemma 3 we get

$$Q_{j+r} = D^r B_j = \mathbf{Z} q_1(A)$$

if

(4.2) $$(2^{r-1}+1)2^j \underline{d}(B_1) q_1(A) > 1.$$

We have to check that if

(4.3) $$(2^{k-1}+1)\bar{d}(A) q_1(A) > 1,$$

then (4.2) holds with $r=k-j$, and also that $r\geqq 1$, i.e. $j<k$. Now if $j\geqq k$, then

$$\underline{d}(B_j) = 2^j \underline{d}(B_0) \geqq 2^k \bar{d}(A) > 1/q_1(A),$$

a contradiction. Hence (4.2) follows from (4.3) and the inequalities

$$\underline{d}(B_0) \geqq \bar{d}(A), \quad (2^{k-j-1}+1)2^j = 2^{k-1}+2^j \geqq 2^{k-1}+1.$$

5. Iterated difference sets

Now we prove the part a) of the Theorem. The proof will be quite similar to that of the part b) but somewhat simpler. It is based on the following observation:

Lemma 6. *For every set A of integers either $\Delta A=\Delta_1 A$, or $\bar{d}(\Delta A)\geqq 2\bar{d}(A)$.*

Proof. Let $u\in\Delta A\setminus\Delta_1 A$, if there is any. Then $u=a-b$ with some $a, b\in A$, the sets $A-a$ and $A-b$ are both contained in ΔA and their intersection has density 0 by $a-b\notin\Delta_1 A$, hence $\bar{d}(\Delta A)\geqq 2\bar{d}(A)$ immediately follows.

Now consider the sets $\Delta^k A$. Let j be the first index for which

(5.1) $$\Delta^{j+1}A = \Delta_1 \Delta^j A;$$

by Lemma 6 we conclude

$$\bar{d}(\Delta^j A) \geqq 2^j \bar{d}(A)$$

($j=0$ is possible, in that case we mean $\Delta^0 A=A$).

With $B=\Delta^j A$, (5.1) means $\Delta_1 B=\Delta B$, whence by Lemma 5 we have $\Delta^r B=\Delta_1^r B$ for all $r\geqq 1$. Thus, by the part a) of the Theorem that we have already proved,

$$\Delta^{j+r}A = \Delta_1^r B = \mathbf{N} q_1(B) = \mathbf{N} q(A)$$

if

(5.2) $$(2^{r-1}+1)\bar{d}(B) q(A) > 1$$

(we used the fact that $q_1(B)=q(B)=q(A)$). Now it can be checked like at the end of the proof of the part a) in the previous section that (5.2) holds with $r=k-j$ if

$$(2^{k-1}+1)\bar{d}(A) q(A) > 1.$$

REFERENCES

[1] Halberstam, H. and Roth, K. F., *Sequences*, Clarendon Press, Oxford, 1966. *MR* **35**# 1565.
[2] Kneser, M., Abschätzungen der asymptotischen Dichten von Summenmengen, *Math. Z.* **58** (1953), 459—484. *MR* **15**—104.
[3] Stewart, C. L. and Tijdeman, R., On infinite difference sets, *Canadian J. Math.* **31** (1979), 897—910.
[4] Stewart, C. L. and Tijdeman, R., On density difference sets of sets of integers, in: *Studies in pure mathematics to the memory of P. Turán* (ed. P. Erdős), Birkhäuser Verlag & Akadémiai Kiadó, Budapest, 1983, 701—710.

(Received June 5, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 203—206

A NOTE ON A THEOREM OF P. ERDŐS

K. KOVÁCS

1. In 1946 P. Erdős proved the following theorem [1]: If f is a real-valued additive function, which is monotonically increasing, then $f(n)=c\log n$. In the present work we shall consider the same problem when f is a complex-valued function and $|f|$ is monotonic.

For complex-valued completely additive functions we have a similar result. We shall prove the following theorem.

THEOREM. *If f is a complex-valued completely additive function and $|f|$ is monotonic, then $f(n)=c\log n$.*

PROOF. a) If $|f|$ is monotonically decreasing, then $f\equiv 0$. In order to show this we have only to observe that if $f(a)\neq 0$ then $|f(a^k)|=k|f(a)|>|f(a)|$ contradicts the monotonicity of the function $|f|$.

b) If $|f|$ is monotonically increasing and $f\not\equiv 0$ then, similarly to a) we have $|f(n)|>0$ for $n>n_0$.

Let $a<b$ be arbitrary natural numbers with $a>1$. If

$$a^s<b^t<a^{s+1}, \tag{1.1}$$

then $s\log a<t\log b<(s+1)\log a$, therefore

$$\frac{\log a}{\log b}<\frac{t}{s}<\frac{\log a}{\log b}\left(1+\frac{1}{s}\right).$$

If $t\to\infty$, then

$$\lim_{t\to\infty}\frac{t}{s}=\frac{\log a}{\log b}. \tag{1.2}$$

Since f is completely additive and

$$|f(a^s)|\leqq|f(b^t)|\leqq|f(a^{s+1})|,$$

we have

$$s|f(a)|\leqq t|f(b)|\leqq(s+1)|f(a)|.$$

Hence

$$\lim_{t\to\infty}\frac{t}{s}=\frac{|f(a)|}{|f(b)|}. \tag{1.3}$$

1980 *Mathematics Subject Classification.* Primary 10A20.
Key words and phrases. Additive function, the characterization of $\log N$, $|f|$ is monotonic.

Using (1.2) and (1.3) we obtain

$$\frac{|f(a)|}{|f(b)|} = \frac{\log a}{\log b}$$

for arbitrary natural numbers a, b, different from 1.

So we have $|f(n)|=c_1 \log n$ with a suitable $c_1>0$. Thus for arbitrary natural numbers a and b we have

$$|f(a)+f(b)| = |f(ab)| = |f(a)|+|f(b)|.$$

This means that the values of f can only lie on the same half-line of the complex plane. Hence $f(n)=c_2\,|f(n)|=c \log n$.

REMARK 1. It is possible to get the same result using the ideas of V. T. Sós' [2] proof of Erdős' theorem.

2. For additive functions we can prove the following weaker result.

PROPOSITION. *Let f be a complex-valued additive function.*

a) *If $|f|$ is monotonically decreasing, then $f\equiv 0$.*

b) *If $|f|$ is monotonically increasing, then there exists a line on the complex plane, such that the projection of $f(n)$ onto this line is $c \log n$.*

We give only a sketch of the proof.

a) We can prove more: If f is convergent, then $f\equiv 0$.

(i) If $\lim\limits_{n\to\infty} f(n)=0$, then for any $a\in \mathbf{N}$ $\lim\limits_{n\to\infty} f(an)=0$, too. This implies $f(a)=0$ for all $a\in\mathbf{N}$.

(ii) If $\lim f(n)=c\neq 0$, then for arbitrary $\varepsilon>0$ and for primes $p,q>c(\varepsilon)$ $(p\neq q)$ we have

$$c-\varepsilon < |f(p)|,\ |f(q)|,\ |f(pq)| < c+\varepsilon,$$

$$\operatorname{arc} f(p)-\operatorname{arc} f(q) = \frac{\pm 2\pi}{3}+O(\varepsilon) \pmod{2\pi}. \tag{2.1}$$

For a fixed prime $q>c(\varepsilon)$ we have infinitely many p_i with the properties above. So there will be a set p_{i_m} of the primes p_i, such that

$$\operatorname{arc} f(p_{i_{m+1}})-\operatorname{arc} f(p_{i_m}) = O(\varepsilon),$$

which contradicts to (2.1).

So we proved, that if $|f|$ is monotonically decreasing or if $|f|$ is bounded and monotonically increasing, then $f\equiv 0$, consequently it is enough to examine the case $|f|\to\infty$ in b).

b) LEMMA 1. *Let a denote an arbitrary fixed natural number. If $t_i\to\infty$ and $(a,t_i)=1$, then for any $\varepsilon>0$ there exists an $i_0=i_0(\varepsilon,a)$ such that*

$$\varphi_a(t_i) \leqq \frac{\pi}{2}+\varepsilon \quad \textit{for} \quad i>i_0,$$

where $\varphi_a(t_i)$ denotes the angle between $f(a)$ and $f(t_i)$ $(\varphi_a(t_i) = |\operatorname{arc} f(a)-\operatorname{arc} f(t)|)$.

PROOF. In the opposite case the angle opposite to $\overline{f(t_i)}$ will be an obtuse angle in the triangle $\overline{f(a)}$, $\overrightarrow{f(t_i)}$, $\overline{f(at_i)}$ for i large enough in view of $|f(t_i)|\to\infty$. This implies $|f(t_i)|>|f(at_i)|$ which contradicts to the monotonically increasing property of $|f|$.

There exists a line l on the complex plane, that for a suitable sequence p_j of primes $(j=1, \dots, \infty)$ the angle $\omega_l(p_j)$ between $f(p_j)$ and l converges to zero if $i\to\infty$. We shall prove that the projection f^* of f onto this line is monotonic. We note that for any $n\in\mathbf{N}$ we have for arbitrary $\varepsilon>0$

$$\omega_l(a) \leqq \varphi_a(p_j)+\omega_l(p_j) < \frac{\pi}{2}+2\varepsilon \quad \text{for} \quad j > j_0(\varepsilon, a).$$

Hence $\omega_l(n)\leqq\pi/2$ for any $n\in\mathbf{N}$.

LEMMA 2. *If* $n_1<n_2$ *then* $f^*(n_1)\leqq f^*(n_2)$.

PROOF. Let us write the cosinus-theorem for the following figure.

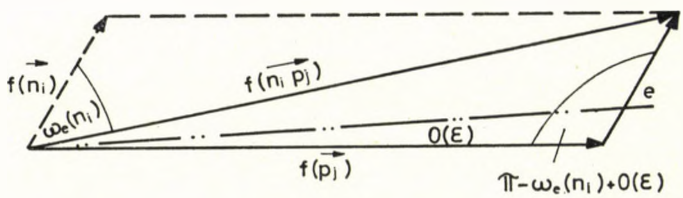


Fig. 1

$$|f^2(n_i p_j)| = |f^2(n_i)|+|f^2(p_j)|-2|f(n_i)|\,|f(p_j)| \cos[\pi-\omega_l(n_i)+O(\varepsilon)] =$$
$$= |f^2(n_i)|+|f^2(p_j)|+2|f(n_i)|\,|f(p_j)| \cos[\omega_l(n_i)+O(\varepsilon)]$$

for $i=1,2$, where $\omega_l(p_j)=O(\varepsilon)$ and $(p_j, n_i)=1$. In view of $|f(n_2p_j)|\geqq f|(n_1p_j)|$ we have

$$0 \leqq |f^2(n_2p_j)|-|f^2(n_1p_j)| =$$
$$= |f^2(n_2)|-|f^2(n_1)|+2|f(p_j)|\left\{|f(n_2)| \cos\left(\omega_l(n_2)+O(\varepsilon)\right)-|f(n_1)| \cos\left(\omega_l(n_1)+O(\varepsilon)\right)\right\}.$$

Using that ε can be arbitrary small and $|f(p_j)|\to\infty$ we obtain $|f(n_2)|\cos\omega_l(n_2)\geqq$ $\geqq|f(n_1)|\cos\omega_l(n_1)$, consequently

$$f^*(n_2) \geqq f^*(n_1).$$

So by Lemma 2 f^* is a real-valued additive function, which is monotonically increasing. Then using the theorem of Erdős we have $f^*(n)=c\log n$.

REMARK 2. We can prove, too, that $\omega_l(a)\neq\pi/2$ for $a\geqq a_0$. If there exists a number d such that $\omega_l(d)=\pi/2$ and d is large enough, then there are p_1 and p_2 primes such that $p_1<p_2<d$ and $\omega_l(p_1)<\pi/2$ and $\omega_l(p_2)<\pi/2$. Let $n_1:=p_1p_2$, $n_2:=p_1d$. Then using Lemma 2 we obtain

$$f^*(p_1p_2) \leqq f^*(p_1d),$$

but we have clearly

$$f^*(p_1) < f^*(p_1 p_2) \quad \text{and} \quad f^*(p_1 d) = f^*(p_1),$$

which is a contradiction.

REFERENCES

[1] Erdős, P., On the distribution function of additive functions, *Ann. of Math.* (2) **47** (1946), 1—20 *MR* 7—416.
[2] Sós, V. T., 28. feladat (Problem 28), *Mat. Lapok* **3** (1952), 91—94 (in Hungarian).

(Received July 27, 1984)

BERZSENYI DÁNIEL GIMNÁZIUM
HUBA U. 7.
H—1134 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 207—212

THE CHARACTERIZATION OF COMPLEX-VALUED ADDITIVE FUNCTIONS II

K. KOVÁCS

We examined the characterization of real-valued additive functions in [1]. We can generalize some results for complex-valued functions. Let f denote a complex-valued additive function and let $r_1=0$ and $r_2, \dots, r_k$ be fixed positive integers and let us define

$$H_n = \{f(n+r_i),\ i\in\{1, \dots, k\}\colon |f(n+r_i)| \text{ is maximal}\}.$$

Let $g(n)$ be an arbitrary but fixed element of H_n. It is possible that H_n has more elements and so there are many functions with the given property. Let us examine one of these functions, the function g, which is uniquely defined already.

In [2] we have proved two theorems. We can generalize these theorems in two directions as follows.

THEOREM 1. *If $|g|$ is monotonically decreasing on a set having upper density one, then $|g|\equiv c$ with a constant $c\in\mathbf{R}$. To any function $h(n)$ there exists a set $A=\{a_n\}_1^\infty$ such that $a_{n+1}-a_n>h(n)$ and if $|g|$ is monotonically decreasing on A, then $|g|\equiv c\in\mathbf{R}$.*

THEOREM 2. *If $\lim_{n\to\infty} g(n)=c$ on a set having upper density one, then $g\equiv c$ with a constant $c\in\mathbf{C}$. To any function $h(n)$ there exists a set $A=\{a_n\}_1^\infty$ such that $a_{n+1}-a_n>h(n)$ and if $\lim_{n\to\infty} g(a_n)=c$, then $g\equiv c$ with a constant $c\in\mathbf{C}$.*

PROOF OF THEOREM 1. a) Let B be a set having upper density one. If

$$\prod_{i\in J\subset\{1,\dots,n\}} t_i\in B$$

for any $n\in\mathbf{N}$ to all subset J and $(t_u, t_v)=1$ if $u\neq v$, then $|f(t_i)|>0$ only for finitely many t_i. Namely if $|f(t_i)|>0$ for infinitely many t_i, then to any $\varepsilon>0$ there exists an angular domain with the angle ε and the midpoint origo on the complex plane, which contains infinitely many $f(t_i)$ $(i=i_1, \dots, i_n, \dots;\ n\to\infty)$. If $\varepsilon<\pi/2$ and $n\to\infty$, then $\left|f\left(\prod_{j=1}^n t_{i_j}\right)\right|_{j\in\mathbf{N}}$ is strictly monotonically increasing, which is a contradiction.

If there exist x_i with $|g(x_i)|=c_i$ for $i=1, 2$, so we have $|f(x_i^*)|=c_i$ $(x_i^*\in\{x_i+r_1, \dots, x_i+r_k\})$. We show that $|g(x)|\geqq c_i$ $(x\in B)$ infinitely many times in

1980 *Mathematical Subject Classification.* Primary 10A20.

Key words and phrases. The characterization of constant functions, monotonic or convergent "maximum" functions on sets having upper density one and on rare sets.

this case, so using the monotonicity of $|g|$ on B it follows $|g|\equiv c=c_i$. Since B is a set having upper density one, there exists a set $(m_i)_{i\in \mathbf{N}}\subset B$ such that $\prod_{j\in J\subset\{1,\dots,n\}} m_j\in B$ for any $n\in\mathbf{N}$ to all subset J and

$$(m_u, m_v) = 1 \quad \text{for} \quad u\neq v \quad \text{and} \quad x_i m_j\in B, \quad \text{too, with} \quad (x_i, m_j) = 1.$$

Thus $|g(x_i m_j)|\geqq|f(x_i m_j)|=c_i$, if j is large enough. □

b) Let $A=\bigcup_{n=1}^{\infty} A_n$ where

$$A_n := \Big\{nq_{ni},\ i = 1, \dots, \infty;\ \prod_{i\in I\subset\{1,\dots,s\}} q_{ni}\Big\}$$

for all $s\in\mathbf{N}$ and to all subset I with $(n, q_{ni})=1$ and $(q_{nu}, q_{nv})=1$ for $u\neq v$.

Using that $\prod_{i\in I} q_{ni}\in A$ to all subset I, $f(q_{ni})=0$ if i is large enough. If there exists an x_0 such that $|g(x_0)|=c\neq 0$, then there exists an $x_1\in\mathbf{N}$ with $|f(x_1)|=c$ and similarly to a) we have

$$|g(x_1 q_{x_1 i})| \geqq |f(x_1 q_{x_1 i})| = c,$$

if i is large enough. Using the monotonicity of $|g|$ on the set $(x_1 q_{x_1 i})_{i\in\mathbf{N}}\subset A$, it follows $|g|\equiv c$.

The suitable rarity can be guaranteed with the suitable choice of q_{ni}. □

Proof of Theorem 2. a) Let B be a set having upper density one.

(i) In the proof we use the following fact from the proof of Theorem 1: To any $x\in\mathbf{N}$ there exists a set $(m_i)_{i\in\mathbf{N}}$ such that $xm_i\in B$ and $\prod_{i\in I\subset\{1,\dots,s\}} m_i\in B$ for any $s\in\mathbf{N}$ to all subset I and

$$(m_u, m_v) = 1 \quad \text{for} \quad u\neq v,\ (x, m_i) = 1 \quad \text{for} \quad \text{all} \quad i\in\mathbf{N}.$$

We prove, that for any $\varepsilon>0$ $|f(m_i)|<\varepsilon$, if i is large enough. Namely if $|f(m_i)|>\varepsilon$ infinitely many times, then to any $\varepsilon^*>0$ we can choose an angular domain with the angle ε^* and the midpoint origo, which contains infinitely many $f(m_i)$ ($i=i_1, \dots, i_n, \dots;\ n\to\infty$). So $\big|f\big(\prod_{j=1} m_{i_j}\big)\big|$ can be arbitrary large, which contradicts the condition $\lim_{n\to\infty} g(b_n)=c$, where $b_n\in B$.

(ii) We prove that $g\equiv c$. Let $g(x_0)=c_0$. We show that there exists a set $(x_j)_{j\in\mathbf{N}}$ such that $\lim_{j\to\infty} g(x_j)=c_0$ (and $x_j\to\infty$).

Let us consider the congruence

$$x \equiv x_0 \big(\mathrm{mod} \prod_{i=1}^{k} (x_0+r_i)^2\big).$$

Infinitely many solutions of this congruence are in B. So it is possible to choose $\{v_j\}_1^{\infty}\subset\mathbf{N}$ such that we obtain $(x_j)_{j\in\mathbf{N}}$ with

$$x_j+r_i = (x_0+r_i)\left(1+v_j\frac{\prod_{v=1}^{k}(x_0+r_v)^2}{x_0+r_i}\right) = (x_0+r_i)V_{ji}$$

for $1\leqq i\leqq k$, where we choose v_j such that to any fixed i $(1\leqq i\leqq k)$ $(V_{ui}, V_{si})=1$

for $u\neq s$ and $\prod_{j\in J\subset\{1,\dots,n\}} V_{ji}\in B$ for any $n\in\mathbf{N}$ to all subset J. Let

$$v_j = m_j \prod_{z=1}^{j-1}\prod_{\alpha=1}^{k}\left(1+v_z\frac{\prod_{\nu=1}^{k}(x_0+r_\nu)^2}{x_0+r_\alpha}\right).$$

This choice ensures $(V_{ui}, V_{si})=1$ for $u\neq s$. We can also guarantee with a suitable choice of m_j that $\prod_{\mathscr{J}} V_{ji}\in B$ for all fixed i $(1\leqq i\leqq k)$. Let us assume that we have already constructed $m_1, m_2, \dots, m_N$. Let us denote the elements $\prod_{\mathscr{J}} V_{ji}$ with a_{si} $(1\leqq s\leqq 2^N-1,\ 1\leqq i\leqq k)$. The maximum of this a_{si} let be denoted by γ. We need to find an m_{N+1} such that

$$a_{si}V_{N+1,i} = a_{si}\left[1+m_{N+1}\prod_{z=1}^{N}\prod_{\alpha=1}^{k}\left(1+v_z\frac{\prod_{\beta=1}^{k}(x_0+r_\beta)^2}{x_0+r_\alpha}\right)\left(\frac{\prod_{\nu=1}^{k}(x_0+r_\nu)^2}{x_0+r_i}\right)\right]=$$
$$=a_{si}(1+m_{N+1}D_i)\in B \quad\text{for all}\quad s\in\{1,\dots,2^N-1\},\quad 1\leqq i\leqq k.$$

Let us consider the numbers of the form $a_{si}(1+mD_i)$ for $m=1,\dots,x$ for arbitrary choice of x. If all $m\in\{1,\dots,x\}$ numbers are "bad" for m_{N+1} then there exists an $s_0\in\{1,\dots,2^N-1\}$ and an $i_0\in\{1,\dots,k\}$ such that $a_{s_0i_0}(1+mD_{i_0})\notin B$ for at least $\dfrac{x}{(k+1)(2^N-1)}$ elements $m\in\{1,\dots,x\}$. Taking into account that all the elements $a_{s_0i_0}(1+mD_{i_0})$ are less then $\gamma(1+x\max_{1\leqq i\leqq k} D_i)=:xE$ this means that

$$\#\{b\in B;\ b\leqq xE\}\geqq\frac{x}{(k+1)(2^N-1)}.$$

Hence the upper density of B is less than 1, so this contradiction proves the possibility of the choice of m_{N+1}.

So using (i) $\lim_{j\to\infty} f(V_{ji})=0$ to any fixed i $(1\leqq i\leqq k)$, consequently

$$\lim_{j\to\infty} f(x_j+r_i)=\lim_{j\to\infty} f[(x_0+r_i)V_{ji}]=f(x_0+r_i)$$

for $1\leqq i\leqq k$, since $(x_0+r_i, V_{ji})=1$. Hence $\lim_{j\to\infty} g(y_j)=g(x_0)$. Therefore $g(x_0)= =\lim_{j\to\infty} g(y_j)=c$ for any $x_0\in\mathbf{N}$. □

b) Let us choose the sequences $\{w_{nj}\}_{j\in\mathbf{N}}$ and $\{\omega_{nij}\}_{j\in\mathbf{N}}$ for any fixed $n\in\mathbf{N}$ and fixed $i\in\{1,\dots,k\}$ as follows: Let

$$w_{n1}=m_{n1},\quad \omega_{ni1}=d_{ni1},\quad P_{n\beta\alpha}=1+w_{n\alpha}\frac{\prod_{\gamma=1}^{k}(n+r_\gamma)^2}{n+r_\beta}\quad(\beta\in\{1,\dots,k\}),$$

(1.1) $$Q_{nj}=\prod_{\mu=1}^{j}\prod_{\delta=1}^{k}\omega_{n\delta\mu}\quad\text{and}$$

$$w_{nj}=m_{nj}\Big(\prod_{\alpha=1}^{j-1}\prod_{\beta=1}^{k}P_{n\beta\alpha}\Big)Q_{nj}$$

with

$$m_{nj} \equiv 1 \pmod{Q_{nj}} \qquad \text{for} \quad j = 2, 3, \ldots, \infty,$$

(1.2)
$$\omega_{nij} = d_{nij}[(\prod_{\alpha=1}^{j-1} \prod_{\beta=1}^{k} P_{n\beta\alpha})Q_{n,j-1}+1]$$

with

$$d_{nij} \equiv 1 \left(\operatorname{mod} Q_{n,j-1}(\prod_{\alpha=1}^{j-1} \prod_{\beta=1}^{k} P_{n\beta\alpha})\right) \quad \text{for } j = 2, 3, \ldots, \infty.$$

So we have guaranteed that

a) $\omega_{nij} \equiv 1 \pmod{\omega_{nst}}$ for $j > t$, $1 \leqq s, i \leqq k$ using (1.2),

consequently

(1.3)
$$(\omega_{nij}, \omega_{nst}) = 1 \quad \text{for} \quad j \neq t;$$

b) $P_{n\beta\alpha}|w_{nu}$ for $u > \alpha$, $1 \leqq \beta \leqq k$ using (1.1), consequently

(1.4)
$$(P_{n\beta\alpha}, P_{n\gamma u}) = 1 \quad \text{for} \quad u \neq \alpha;$$

c) $\omega_{nit} \equiv 1 (\operatorname{mod} P_{n\beta\alpha})$ for $\alpha < t$, $1 \leqq i$, $\beta \leqq k$ using (1.2)

and

$$\omega_{nit}|w_{n\alpha} \quad \text{for} \quad t \leqq \alpha, \quad 1 \leqq i \leqq k \quad \text{using (1.1)}$$

consequently

(1.5)
$$(\omega_{nit}, P_{n\beta\alpha}) = 1 \quad \text{for} \quad t \leqq \alpha, \quad 1 \leqq i, \quad \beta \leqq k.$$

Let A be a sequence of monotonically increasing natural numbers such that for any fixed $n \in \mathbf{N}$ and for any fixed $i \in \{1, \ldots, k\}$

$$\prod_{j \in J \subset \{1,\ldots,m\}} \omega_{nij} \in A, \qquad \prod_{j \in J \subset \{1,\ldots,m\}} \omega_{nij} P_{nij} \in A \quad \text{and} \quad n + w_{nm} \prod_{\gamma=1}^{k} (n+r_\gamma)^2 \in A$$

for any $m \in \mathbf{N}$ with all subsets J. We make the construction of $\{w_{nj}\}_{j\in\mathbf{N}}$ and $\{\omega_{nij}\}_{j\in\mathbf{N}}$ for any fixed $n+j=2, 3, \ldots$ for all i $(1 \leqq i \leqq k)$. We can choose the factors m_{nj} and d_{nij} such that the rarity $a_{n+1} - a_n > h(n)$ could be guaranteed. First we construct w_{11} and ω_{1i1} such that P_{1i1}, ω_{1i1} and $P_{1i1}\omega_{1i1} \in A$. The rarity can be guaranteed with the suitable choice of d_{1i1} $(1 \leqq i \leqq k)$. The difference of the adjacent elements of A is greater than d_{1i1} for some i for these elements and we can guarantee $d_{111} > h(1), \ldots$ $\ldots, h(3k)$ already. For a fixed $n+j=z$ $(z>2)$ and for an arbitrary fixed $v=n-1$, $n-2, \ldots, 1$ we have

$$\delta_{v,z-v} = v + w_{v,z-v} \prod_{\gamma=1}^{k} (v+r_\gamma)^2 \in A$$

and

$$H_{vi} = \{(\prod_{t \in T \subset \{1,\ldots,z-v-1\}} \omega_{vit})\omega_{vi,z-v}, (\prod_{t \in T \subset \{1,\ldots,z-v-1\}} \omega_{vit} P_{vit})\omega_{vi,z-v} P_{vi,z-v}\} \subset A$$

for all subsequence T and for all i $(1 \leqq i \leqq k)$. With the suitable choice of $m_{v,z-v}$ we have $\delta_{v,z-v} \in A$ great enough. Then for any fixed i $(1 \leqq i \leqq k)$ the difference of the adjacent elements in the set H_{vi} is greater than $d_{vi,z-v}$, namely we have the new factor $\omega_{ni,z-v}$ in all these products. For the other differences we can achieve by suitable choices of d_{nij} and m_{nj} that

$$\min H_{v,s+1} - \max H_{v,s} > \frac{d_{v,s+1,z-v}}{2} \qquad (1 \leqq s \leqq k-1),$$

$$\delta_{v+1,z-v-1} - \max H_{v,k} > \frac{m_{v+1,z-v-1}}{2}$$

and

$$\min H_{v+1,1} - \delta_{v+1,z-v-1} > \frac{d_{v+1,1,z-v-1}}{2}$$

and so we can ensure the condition $a_{n+1} - a_n > h(n)$.

If $\prod_{t \in I \subset \{1,\dots,m\}} x_t \in A$ for all $m \in N$ and all I and $(x_u, x_v) = 1$ for $u \neq v$, then $\lim_{t\to\infty} f(x_t) = 0$ similarly to 2.a) (i). Using the definition of A, (1.3), (1.4) and (1.5) we can choose $x_t := \omega_{nit}$ and $x_t := \omega_{nit} P_{nit}$ $(t > 2)$ for arbitrary fixed $n \in \mathbf{N}$ and i $(1 \leqq i \leqq k)$. So we get

$$(1.6) \qquad \lim_{t\to\infty} f(\omega_{nit}) = 0 \quad \text{and} \quad \lim_{t\to\infty} f(\omega_{nit} P_{nit}) = 0,$$

consequently by (1.5)

$$(1.7) \qquad \lim_{t\to\infty} f(P_{nit}) = 0.$$

We prove that $g \equiv c$. Let $g(x_0) = c_0$. So with

$$y_j = x_0 + w_{x_0 j} \prod_{\gamma=1}^{k} (x_0 + r_\gamma)^2 \quad \text{we have} \quad f(y_j + r_i) = f(x_0 + r_i) + f(P_{x_0 ij})$$

for $1 \leqq i \leqq k$. Using (1.7) we have

$$\lim_{j\to\infty} f(y_j + r_i) = f(x_0 + r_i)$$

and so

$$\lim_{j\to\infty} g(y_j) = g(x_0) = c_0.$$

The sequence $\{y_j\}_{j\in\mathbf{N}} \subset A$, so

$$\lim_{j\to\infty} g(y_j) = c = c_0$$

is possible only.

REFERENCE

KOVÁCS, K., On the characterization of additive functions, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **25** (1982), 257—265.

(Received July 27, 1984)

BERZSENYI DÁNIEL GIMNÁZIUM
HUBA U. 7
H—1134 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 213—221

ON A NONLINEAR OPERATIONAL DIFFERENTIAL EQUATION SYSTEM

TAMÁS FÉNYES

Introduction

Let us consider the set L of complex (or real) valued functions in a real variable t which are locally integrable in the interval $0 \leqq t < \infty$. This set is a commutative ring with respect to the ordinary addition and multiplication defined by the Duhamel convolution.

By a theorem of Titchmarsh on convolution the ring L has no divisors of zero, so it can be extended to a quotient field (see [4]). This is the Mikusiński operator field denoted by M (see [3]). M contains the ring L and also the field of the complex numbers.

In this paper we shall deal with the following nonlinear algebraic differential equation system being defined in Mikusiński's operator field:

$$D(x_i) = \frac{a_i}{\sum\limits_{\substack{k=1\\k\neq i}}^{n} x_k} + (f_i - \gamma_i)x_i, \qquad i = 1, 2, \ldots, n \tag{1}$$

where $n > 1$ is an arbitrary integer, the operators $x_i \in M$ are the unknowns of system (1), D is the symbol of the well-known algebraic derivative (see [3]), the $a_i(t)$, and $f_i(t)$ are locally integrable functions in $0 \leqq t < \infty$, and the γ_i are given real numbers.

We assume that the conditions

$$a_i \neq 0 \qquad i = 1, 2, \ldots, n, \tag{2}$$

$$\sum_{i=1}^{n} a_i = 0 \tag{3}$$

hold. (2) means that none of the functions a_i is the zero element of L, (3) means that their sum equals to the zero element of L.

If there exist operators x_i satisfying the system (1), we say that (1) has an operational solution. If the x_i are locally integrable functions, i.e., $x_i \in L \subset M$ for every i, we say that the system (1) has a functional solution. It follows from (1) and (2) that $x_i \neq 0$ $(i = 1, 2, \ldots, n)$. In the sequel we give simple sufficient conditions guaranteeing the existence or non-existence of solutions for (1), moreover, we determine

---

1980 *Mathematics Subject Classification.* Primary 44A40; Secondary 44A35, 45E10.

Key words and phrases. Mikusiński's operator field, nonlinear differential equations, integral equations of convolution type.

the general operational solution of the system (1). The application of operational calculus is very useful in the discussion of (1) since, as we shall see, it reduces the non-linear problem to a linear one. This fact is the main motivation of the problem discussed in this paper.

We assume that the elements and notations of Mikusiński's operational calculus are familiar to the reader (see [3]).

1. On the system of differential equations (1)

Let us consider the system of differential equations

$$D(x_i) = \frac{a_i}{\prod\limits_{\substack{k=1\\k\neq i}}^{n} x_k} + (f_i-\gamma_i)x_i \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

which can be written in the form

$$D(x_i) \prod_{\substack{k=1\\k\neq i}}^{n} x_k = a_i + (f_i-\gamma_i) \prod_{k=1}^{n} x_k. \tag{1.2}$$

Obviously, in the ring L, (1.2) is equivalent with a system of non-linear integral equations of convolution type. For example, for $n=2$ we have

$$\int_0^t \tau x_1(\tau)x_2(t-\tau)\,d\tau + \int_0^t f_1(t-\tau)\int_0^\tau x_1(u)x_2(\tau-u)\,du\,d\tau - \gamma_1\cdot$$

$$\cdot\int_0^t x_1(\tau)x_2(t-\tau)\,d\tau + a_1(t) = 0,$$

$$\int_0^t \tau x_2(\tau)x_1(t-\tau)\,d\tau + \int_0^t f_2(t-\tau)\int_0^\tau x_1(u)x_2(\tau-u)\,du\,d\tau - \gamma_2\cdot$$

$$\cdot\int_0^t x_1(\tau)x_2(t-\tau)\,d\tau + a_2(t) = 0.$$

By adding the equations of system (1) and taking into account condition (3) we have

$$D\Big(\prod_{k=1}^{n} x_k\Big) = \sum_{i=1}^{n} (f_i-\gamma_i) \prod_{k=1}^{n} x_k. \tag{1.3}$$

By introducing

$$F = \prod_{k=1}^{n} x_k, \quad f = \sum_{i=1}^{n} f_i, \quad \gamma = \sum_{i=1}^{n} \gamma_i,$$

the following linear homogeneous algebraic differential equation is obtained:

$$D(F) - (f-\gamma)F = 0. \tag{1.4}$$

We have shown in [1] that if there exists a real number λ such that

$$\frac{f(t)-\lambda}{t} \in L, \tag{1.5}$$

then the general operational solution of (1.4) is of the form

$$F = cs^{\lambda}e^{-\gamma s}\exp\left\{\frac{f(t)-\lambda}{-t}\right\}, \tag{1.6}$$

and the latter is a function for $c \neq 0$ if and only if

$$\lambda < 0 \quad \text{and} \quad \gamma \geqq 0 \tag{1.7}$$

(see also Fényes [2]). In the sequel we assume that (1.5) holds. Here s denotes the differential operator and e^{-s} the shifting operator. We remark that the existence of the operator F does not imply the existence of an operational solution for (1.1). By substituting (1.6) in (1.1) we have

$$D(x_i) - \left(\frac{a_i}{c}s^{-\lambda}e^{\gamma s}\exp\left\{\frac{f(t)-\lambda}{t}\right\} - \gamma_i + f_i\right)x_i = 0, \qquad i = 1, 2, \ldots, n. \tag{1.8}$$

So the following Lemma holds.

LEMMA. *The system* (1.1) *has a solution in the operator field if and only if every equation of* (1.8) *has a nontrivial operational solution.*

Let x_{pi} denote a particular solution of the i-th equation of (1.8). Then the general solution is of the form

$$x_i = \beta_i x_{pi}, \tag{1.9}$$

where β_i is an arbitrary number. The set of the operators x_i $(i=1, 2, \ldots, n)$ is the general operational solution of (1.1) if and only if

$$\prod_{k=1}^{n} \beta_k = c \tag{1.10}$$

holds. This can be seen trivially and is plausible since the general operational solution of (1.1) contains exactly n parameters.

NON-EXISTENCE THEOREM. *Let* $\frac{f(t)-\lambda}{t} \in L$ *and* $\gamma > 0$, *moreover, let* (2), (3) *hold. If there exists an index* i *such that* $a_i(t)$ *does not vanish in the* γ *neighbourhood of the origin, then* (1.1) *has no operational solution.*

PROOF. Let i be an index for which $a_i(t)$ does not vanish in the γ neighbourhood of the origin. By the above Lemma it is enough to show that

$$D(x_i) - \left(\frac{a_i}{c}s^{-\lambda}e^{\gamma s}\exp\left\{\frac{f(t)-\lambda}{t}\right\} - \gamma_i + f_i\right)x_i = 0 \tag{1.11}$$

has no other solution than the trivial null solution. If $a_i(t)$ vanishes in a δ neighbourhood of the origin $(0 \leqq \delta < \gamma)$, we choose an arbitrary non-negative number r such that

$$\lambda - r < 0$$

and write (1.11) as

$$D(x_i) - \frac{h_1}{h_2} x_i = 0, \tag{1.12}$$

where

$$h_1 = \frac{a_i}{c} s^{-r} e^{\delta s} \exp\left\{\frac{f(t)-\lambda}{t}\right\} + (f_i - \gamma_i) s^{\lambda - r} e^{-(\gamma-\delta)s},$$

(1.13)

$$h_2 = s^{\lambda - r} e^{-(\gamma-\delta)s}.$$

Obviously, $h_1, h_2 \in L$, moreover, h_1 does not vanish in a neighbourhood of the origin and h_2 vanishes exactly in

$$0 \leqq t \leqq \gamma - \delta.$$

Let us assume that there exists an operator

$$x_i = \frac{u_1}{u_2}, \quad u_1, u_2 \in L, \quad u_1 \neq 0, \quad u_2 \neq 0$$

which satisfies (1.12). So we have

$$\frac{u_2 D(u_1) - u_1 D(u_2)}{u_2^2} = \frac{h_1 u_1}{h_2 u_2}$$

and

$$h_2 u_2 D(u_1) - h_2 u_1 D(u_2) = h_1 u_1 u_2. \tag{1.14}$$

Assuming that u_1 vanishes exactly in the α_1 neighbourhood of $t=0$ and u_2 vanishes exactly in the α_2 neighbourhood of $t=0$ where $\alpha_1, \alpha_2 \geqq 0$, by Titchmarsh's theorem (see [4]) we see that the function

$$h_1 u_1 u_2$$

vanishes exactly in $0 \leqq t \leqq \alpha_1 + \alpha_2$ and the function

$$h_2 u_2 D(u_1) - h_2 u_1 D(u_2)$$

vanishes at last in $0 \leqq t \leqq \alpha_1 + \alpha_2 + \gamma - \delta$. But this is impossible. Contradiction.

Taking into account the well-known shifting property of $e^{\gamma s}$, the theorem can be formalized also in another way.

If (1.1) *has an operational solution and the conditions* (2), (3), (1.5) *are satisfied, then for every* i *the operator* $a_i e^{\gamma s}$ *is a function.*

In order to find criteria for the existence of a solution of (1.1) we must assume that not only condition (1.5) but also the stronger condition

$$\frac{f_i(t) - \lambda_i}{t} \in L \qquad i = 1, 2, \ldots, n \tag{1.14}$$

holds for some real numbers λ_i. Here we have

$$\sum_{i=1}^{n} \lambda_i = \lambda.$$

By introducing the notation

$$q_i = \frac{a_i}{c} s^{-\lambda} e^{\gamma s} \exp\left\{\frac{f(t)-\lambda}{t}\right\} \tag{1.15}$$

we write (1.11) in the form

$$D(x_i)-(q_i+f_i-\gamma_i)x_i = 0, \qquad (i = 1, 2, \ldots, n). \tag{1.16}$$

If the operators q_i are functions and

$$\frac{q_i(t)}{t} \in L,$$

then from (1.4), (1.5), (1.6) it follows that the differential equations (1.16) have non-trivial solutions, and by the above Lemma we can see that the system (1.1) has operational solutions.

We distinguish the following cases.

I. $\lambda \geqq 0$ and $\gamma < 0$.

By the rules of operational calculus

$$q_i(t) \in L, \qquad i = 1, 2, \ldots, n.$$

Since the functions $q_i(t)$ vanish in a neighbourhood of the origin, we have

$$\frac{q_i(t)}{t} \in L.$$

II. $\lambda < 0$ and $\gamma < 0$.

We write

$$-\lambda = m-\varepsilon, \quad m = 1, 2, \ldots; \quad 0 \leqq \varepsilon < 1. \tag{1.17}$$

If the functions $a_i(t)$ are $m-1$ times continuously differentiable and the

$$\frac{d^{m-1} a_i(t)}{dt^{m-1}} \qquad i = 1, 2, \ldots, n$$

are absolutely continuous, further

$$\left.\frac{d^j a_i(t)}{dt^j}\right|_{t=0} = 0, \qquad i = 1, 2, \ldots, n;\ j = 0, 1, 2, \ldots, m-1,$$

then in (1.15) we have

$$s^{-\lambda}\{a_i(t)\} = s^{-\varepsilon}\{a_i^{(m)}(t)\}.$$

Consequently, the operators q_i are functions vanishing in a neighbourhood of the origin and so

$$\frac{q_i(t)}{t} \in L.$$

III. $\gamma \geqq 0$.

For $\gamma > 0$, the functions $a_i(t)$ must vanish in the γ neighbourhood of the origin since in the opposite case the Theorem would yield that (1.1) has no solution. We introduce

$$b_i = a_i e^{\lambda s} = \{a_i(t+\gamma)\}, \qquad i = 1, 2, \ldots, n.$$

First let us consider those indices i for which the functions b_i vanish in a neighbourhood of the origin. If $\lambda \geqq 0$ then, as in the case I, it can be easily seen that

$$q_i \in L, \quad \frac{q_i}{t} \in L$$

holds. For $\lambda < 0$ we obtain the same if we assume that the functions $b_i(t)$ are $m-1$ times continuously differentiable and

$$\frac{d^m b_i(t)}{dt^m} \in L.$$

Let us consider now those indices i for which the functions $b_i(t)$ do not vanish in a neighbourhood of the origin.

We need the following statement (see [2]).

(A) *If*

$$\frac{l_1(t)}{t} \in L, \quad l_2(t) \in L$$

then

$$\frac{\int_0^t l_1(\tau) l_2(t-\tau)\, d\tau}{t} \in L.$$

Let $\lambda > 1$. Since

$$s^{-\lambda} = \left\{\frac{t^{\lambda-1}}{\Gamma(\lambda)}\right\},$$

we have

$$t^{\lambda-2} \in L.$$

Obviously, $q_i \in L$ for every i. Applying the statement (A) we obtain

$$\frac{q_i(t)}{t} \in L.$$

Let $0 \leqq \lambda \leqq 1$. Assuming

$$\frac{b_i(t)}{t} \in L,$$

by the statement (A) we have

$$\frac{q_i(t)}{t} \in L.$$

Finally, let $\lambda < 0$. If the functions $b_i(t)$ are $m-1$ times continuously differentiable,

$$\left.\frac{d^j b_i(t)}{dt^j}\right|_{t=0} = 0, \qquad j = 0, 1, \ldots, m-1$$

and

$$\frac{d^m b_i(t)}{dt^m} \in L, \qquad \frac{1}{t}\frac{d^m b_i(t)}{dt^m} \in L,$$

then in (1.15) we have

$$a_i s^{-\lambda} e^{\gamma s} = b_i s^{m-\varepsilon} = s^{-\varepsilon}\{b_i^{(m)}(t)\},$$

so $q_i(t) \in L$ and an application of statement (A) gives that

$$\frac{q_i(t)}{t} \in L.$$

Taking into account (1.4), (1.5), (1.6), (1.7), (1.15), (1.16), the following Theorem holds:

EXISTENCE THEOREM. *Assume that* (2), (3) *hold, and let*

$$\frac{f_i(t) - \lambda_i}{t} \in L, \qquad i = 1, 2, \ldots, n.$$

If any of the following conditions holds, then the system of algebraic differential equations (1.1) *has an operational solution.*

(B$_1$) $\lambda \geqq 0$ *and* $\gamma < 0$.

(B$_2$) $\lambda < 0$ *and* $\gamma < 0$, *the functions* $a_i(t)$ *are* $m-1$ *times continuously differentiable for every* i,

$$a_i^{(j)}(0) = 0, \quad j = 0, 1, \ldots, m-1; \quad i = 1, 2, \ldots, n.$$

$$\frac{d^m a_i(t)}{dt^m} \in L,$$

where

$$-\lambda = m - \varepsilon, \quad m = 1, 2, \ldots; \quad 0 \leqq \varepsilon < 1.$$

(B$_3$) $\gamma \geqq 0$. *For* $\gamma > 0$ *the functions* $a_i(t)$ *vanish in the* γ *neighbourhood of the origin and either*

(D$_1$) $$\lambda > 1$$

or

(D$_2$) $$0 \leqq \lambda \leqq 1 \quad \textit{and} \quad \frac{a_i(t+\gamma)}{t} \in L, \quad i = 1, 2, \ldots, n,$$

or

(D_3) $\lambda<0$ *and the functions* $a_i(t+\gamma)$ *are* $m-1$ *times continuously differentiable, and*

$$\left.\frac{d^j a_i(t+\gamma)}{dt^j}\right|_{t=0} = 0, \quad j = 0, 1, 2, \dots, m-1; \qquad i = 1, 2, \dots, n.$$

$$a_i^{(m)}(t+\gamma)\in L, \qquad \frac{1}{t}\, a_i^{(m)}(t+\gamma)\in L.$$

The general operational solution of (1.1) *is of the form*

$$x_i = \beta_i s^{\lambda_i} e^{-\gamma_i s} \exp\left\{\frac{f_i(t)+q_i(t)-\lambda_i}{-t}\right\}, \qquad i = 1, 2, \dots, n \tag{1.18}$$

where

$$q_i = \frac{a_i}{c}\, s^{-\lambda} e^{\gamma s} \exp\left\{\frac{f(t)-\lambda}{t}\right\},$$

$$\sum_{i=1}^{n} f_i(t) = f(t), \quad \sum_{i=1}^{n} \gamma_i = \gamma, \quad \sum_{i=1}^{n} \lambda_i = \lambda.$$

The numbers $\beta_i \neq 0$, $c \neq 0$ *are such that*

$$\sum_{k=1}^{n} \beta_k = c.$$

If either (B_1) *or* (B_2) *holds, then* (1.1) *has no functional solution.*

If (B_3) *holds,* (1.18) *gives the general functional solution of* (1.1) *if and only if*

$$\lambda_i < 0 \quad \text{and} \quad \gamma_i \geqq 0 \quad i = 1, 2, \dots, n.$$

Remark 1. The differentiability conditions (D_3) seem to be very strong, however, we show that this is not so.

Example. Let us assume that

$$\gamma = 0, \quad f(t) = -2, \quad t \geqq 0,$$

and for some fixed index i

$$\gamma_i = 0,$$

$$a_i(t) = 1, \quad t \geqq 0$$

$$f_i(t) = -1, \quad t \geqq 0$$

hold. Then

$$\lambda_i = -1, \quad \lambda = -2,$$

and for this index i we have from (1.16)

$$D(x_i) - \left(\frac{s}{c} - \frac{1}{s}\right) x_i = 0. \tag{1.19}$$

Here $a_i(t)$ admits derivatives of all orders, but the condition

$$a(0) = a'(0) = 0$$

does not hold. (1.19) *has only the trivial solution.* Indeed, substituting

$$x_i = \frac{1}{s} y$$

we have

$$D(y) - \frac{s}{c} y = 0,$$

and it is well-known that the latter equation has only the trivial solution. On the other hand, condition (3) is very inconvenient. We want to deal with (1) by rejecting (3) in a subsequent paper.

REMARK 2. It can be seen that under the conditions given in the Existence Theorem every operational solution of (1.1) can be identified with a set of distributions of finite order.

REFERENCES

[1] FÉNYES, T., On an integral equation of the third kind, *Mat. Lapok* (in Hungarian) (to appear)
[2] FÉNYES, T., Anwendung der Mikusińskischen Operatorenrechnung zur Lösung von Integralgleichungen dritter Art vom Faltungstypus, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 9 (1965), 365—399. *MR* **32**#8071.
[3] MIKUSIŃSKI, J., *Operational calculus*, Pergamon Press — Państwowe Wydawnictwo, 1959. *MR* **21**#4333.
[4] TITCHMARSH, E. C., The zeros of certain integral functions, *Proc. London Math. Soc.* **25** (1926), 283—302.

(Received July 31, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 223—224

THE HALL—HIGMAN THEOREM FOR MONOMIALLY-CLOSED FINITE GROUPS

ROBERT W. van der WAALL

Dedicated to the memory of P. Hall (1904—1982)

Introduction

First of all, all groups in this paper are finite. Notations and conventions are standard and taken from Gorenstein's book [1] and from Huppert's book [3].

The theorem of P. Hall and G. Higman (Theorem B of [2]) has had a considerable influence on the study of simple groups and on certain problems concerning modular representations. Nowadays the reader is strongly referred to the recent book of B. Huppert and N. Blackburn [4], in which the whole chapter IX is devoted to this theorem, and where many applications are given.

The precise formulation of the Hall—Higman theorem is as follows.

THEOREM B of P. Hall and G. Higman [2]. *Let G be a p-solvable group of linear transformations in which $O_p(G)=1$ acting on a vector space V over a field F of characteristic p. Let x be an element of G of order p^n. Then the minimal polynomial of x on V is $(X-1)^r$, where either*

(i) *$r=p^n$, or*

(ii) *There exists an integer $n_0\leqq n$ such that $p^{n_0}-1$ is a power of a prime q and the S_q-subgroups of G are nonabelian. In this case, if n_0 is the least such integer, then*

$$p^{n-n_0}(p^{n_0}-1)\leqq r\leqq p^n.$$

A group T is *monomial*, if every complex irreducible representation of T is induced by a one-dimensional representation of a subgroup of T. The group S is *monomially-closed*, if S and all of its subgroups are monomial. (Hence all sections of S are monomial.) It is well-known, that not every monomial group is monomially-closed.

It is the purpose of this paper to show that the second case in the Hall—Higman theorem does not occur if G is monomially-closed. (Notice that any monomial group is p-solvable for all primes p; see [3], Theorem V. 18. 6.) The proof of our statement follows very closely the lines of the pages 359—363 of [1].

Statement of the theorem and its proof

THEOREM ("Hall—Higman" for monomially-closed groups). *Let the monomially-closed group G have the property that $O_p(G)=1$, p some prime number. Suppose that G acts faithfully as a group of linear transformations on the vector space V, of finite dimension over a field K of characteristic p. Let x be an element of order $p^n>1$ in G. Then the strong version of Theorem B of Hall and Higman holds, i.e. the degree of the minimal polynomial of x for the action on V is equal to p^n (thus the minimal polynomial itself is $(X-1)^{p^n}$).*

1980 *Mathematics Subject Classification.* Primary 20C15; Secondary 20D10.

Key words and phrases. Hall—Higman theorem, monomial representations of finite groups.

PROOF. Since any section of G is monomial by hypothesis, we argue by double induction on the order of G and on $\dim_K V$. One has to be careful here, as for the induction process only sections L of G for which $O_p(L)=1$ can be considered. Now look at the pages 359—363 of [1] in which the proof of a part of the ordinary Hall—Higman Theorem B has been given. By the end of page 359 we may assume that K is algebraically closed. Because of a corresponding reasoning as in the last two lines of page 360 and the first thirteen lines of page 361, and replacing the word "Theorem 1.1" by the words "the Theorem" everywhere on the pages 361, 362, 363 we conclude that the Theorem holds unless

1) either $G=CD$, where $C=O_{p'}(G)$, $D=\langle x\rangle\cong G/C$, D acts irreducibly on C/C', $\langle x^{p^{n-1}}\rangle$ acts trivially on C' but non-trivially on C/C', C is a special q-group for some prime number q, i.e. $C'=\Phi(C)=Z(C)$ and C is not abelian,

2) or $G=CD$, with the same conditions as under 1) but now with C elementary abelian.

Now, just as it is done on lines 14—34 of page 361, it follows that the Theorem is valid in either case unless perhaps G acts irreducibly on V. Thus we assume from now on that G acts irreducibly on V.

Suppose that case 2) applies. Then, by the last five lines of page 361 and the first five lines on page 362, it follows that the Theorem holds. Notice that the algebraically-closedness of K is used here and observe that the groups G occurring in 2) are indeed monomially-closed.

Thus from now on we assume that case 1) applies. Then we argue just as it is printed on the rest of the pages 362 and 363 after we have made the following alterations, written inside quotation marks.

a) Delete lines 9, 10, 11 from bottom on page 362 and replace these lines by "$(X-1)^{p^{n-a}}$.".

b) The integer r_0 has to be replaced everywhere where it occurs on the pages 362 and 363, by "p^{n-a}".

c) Lines 4, 5, 6 from top on page 363 have to be replaced by "$r\geqq p^{n-a}p^a=p^n$. So $r=p^n$. Hence the Theorem holds.".

Next we go through the lines 7—28 of page 363 verbatim. It follows then, that the Theorem holds, unless $G=CD$ as under 1), but with C extra special and with $Z(C)=\Phi(C)=C'=Z(G)$. However, it is well-known that G is not a monomial group in that exceptional situation. Therefore this exceptional case does not occur and so the proof of the Theorem is complete. □

REFERENCES

[1] GORENSTEIN, D., *Finite Groups*, Chelsea, New York, 1980. *MR* **81b**: 20002.
[2] HALL, P. and HIGMAN, G., The p-length of a p-soluble group and reduction theorems for Burnside's problem, *Proc. London Math. Soc.* **6** (1956), 1—42. *MR* **17**—344.
[3] HUPPERT, B., *Endliche Gruppen. I*, Springer-Verlag, Berlin—New York—Heidelberg, 1967. *MR* **37**#302.
[4] HUPPERT, B. and BLACKBURN, N., *Finite groups. II*, Springer-Verlag, Berlin—New York—Heidelberg, 1982.

(Received September 13, 1984)

MATHEMATISCH INSTITUUT
UNIVERSITEIT VAN AMSTERDAM
ROETERSSTRAAT 15
NL—1018 WB AMSTERDAM
THE NETHERLANDS

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 225—228

EQUATIONAL CLASSES OF PARTIAL ALGEBRAS

ISIDORE FLEISCHER

The embeddability of partial algebras, which satisfy (strongly) a set of identities, in total algebras which satisfy them — [S p. 45 Th. 30] and in improved form [F iii] — permits complementing the previous work on equational classes of partial algebras with a new HSP characterization of these "strong" equational classes.

A partial algebra is usually taken as a set on which the operations as specified by the type of a species of algebra need only be partially defined. A term (polynomial symbol [G]) built up from these type operations by formal composition may or may not be evaluable when specific elements from the algebra are substituted for its arguments. This leads to various possibilities for construing the identity of terms in such an algebra: "weak" validity = whenever both are evaluable they have the same value; "strong" (of [F], unqualified in [S]) = whenever either is evaluable then so is the other with the same value; and a still stronger, appropriately designated "total" (in [BNR]; called "strong" in [H]) = both everywhere defined with the same value. This last entails of course that the partial algebra is total for all the composites of operations used in building up either side of any identity; its equational theory is discussed in [H]; [B] features the embedding of such a partial in a total algebra of its corresponding HSP-class while [S] accomplishes this for strong equational classes; weak validity is in [K], [P], [H]. For an (apparently flawed) attempt at a comparative survey see [J].

Another option is to construe the algebra's "type" as comprising the full "clone" of terms — thus without specifying certain operations as generators, while retaining the original composition of terms as an "operation" internal to this "type". This leads to a more general notion of partial algebra: one in which each term is assigned a partial functional action independent of those of any constituent terms from which it might be composable (this makes it, in effect, an ordinary partial algebra with the terms as operations) — although one will still require term equalities to be respected insofar as both sides are evaluable (i.e. "weakly"): a term decomposable into subterms whose defined values would compose to determine its value at an argument tuple, should have the same value determined by any other determining decomposition. Taking then as morphisms the maps which preserve definability and value — i.e. the graph — of every term (thus, the usual notion for the terms as operations; in partial algebras with generating operations this agrees with the operational definition) one finds (also in agreement) their kernels to be "weak" congruences for the partial action by each of the terms; conversely, modulo such a con-

1980 *Mathematics Subject Classification.* Primary 08A55.
Key words and phrases. Partial algebras, equational classes.

gruence there is a well-defined term action, minimal for making the quotient map a morphism, and an injective morphism from the so-equipped quotient set to the original codomain. Since weak identities pass back to domains of injective morphisms i.e. are preserved by "weak subalgebras", the quotient inherits this type of partial algebra structure from the codomain. The quotient modulo an arbitrary congruence (for term action) should not be expected to satisfy the term equalities: that this occurs in the usual operational case is due to a term's being defined only if all its subterms are, which makes the equalities "strong" — and a quotient map preserves strong identities. (It might be noted that the term algebra derived from an operational quotient — by composing the quotient generating operations to act as the partial term functions — is a proper strengthening of the above quotient term algebra, in which the smaller quotients of terms act.)

The usual operational algebras are not only "weak partial term algebras" in the above sense but even satisfy the "strong" axiom laid down in [F] (which is the term analogue of [E]'s "incomplete algebra"): viz, if some decomposition into subterms, having defined values which would compose to determine the composed term's value at an argument tuple, then in any other decomposition which could determine the value at this argument tuple were the last subterm defined at the values of the next-to-last composites, the last subterm will indeed be defined there and take this value. This entails in particular that the partial algebra is composition closed, i.e. that a term's defined values include all those determined by composing values of subterms, and is thus more general than the usual operational form only in not having all the subterms of an evaluable term evaluable (at the appropriate arguments).

Each of the validities is inherited by a specific kind of subalgebra and (surjective) homomorphic image (all are preserved by products): Total identities by closed subalgebras (i.e. those closed for the action of the terms — operations suffice in the usual formulation — insofar as they are performable in the containing partial algebra) and by any homomorphic image (indeed a surjective homomorphism is a quotient for any totally defined term and quotients preserve strong identities); weak identities by weak subalgebras (i.e. domains of injective homomorphisms) and by images of strong homomorphisms[1] (i.e. quotient maps for whose kernel definability is saturated — this for the generating operations entails it for all terms in the usual partial algebras); strong identities by term images of term quotient maps and term "relative" subalgebras (i.e. the structures induced on subsets by letting the terms be defined just insofar as their values (for arguments in the subset) are already in the subset).

An operational identity may fail to pass to an operational relative subalgebra because the disappearance of an intermediate value has resulted in a term no longer being defined. (This difficulty does not arise for validity in the sense of Evans.) To circumvent this one can restrict to those subsets, call them "convex", which contain with the value of a composite term for arguments in the subset, all the intermediate operation values from the containing partial algebra. These subsets include the closed subalgebras as well as the subsets which (dually) contain with any operation value all its arguments [F′ p. 12], which we may dub "initial".

[1] Strong homomorphisms possibly defined on proper subsets have been called "partial" in [P] and "konformisms" in [K].

The strong HSP characterization may be obtained, in both its term and operational form, as a consequence of the embeddability of a partial in a total algebra of the equational class. The embeddability of a "strong partial term algebra" as a term relative subalgebra of a total algebra satisfying all its strong identities is from [F iii]; for an operational algebra this was already achieved in [S Th. 30][2] but the proof in [F] sends it on an initial subalgebra.

Thus: Every strong equational class is closed for (hence contains besides its total algebras every) convex (and even arbitrary term) relative subalgebra; conversely, in a strong equational class every operational algebra is a convex (indeed initial) relative subalgebra and every strong partial term algebra is a relative subalgebra, of a total one in the class.

Consequently a class of partial operational algebras is definable by strong equations just when it is closed under products, strong homomorphic images, convex relative subalgebras and its members are embeddable as initial relative subalgebras in total algebras of the class; for a class of term algebras, just when it is closed under products, term quotient maps, term relative subalgebras and its members admit embeddability as term relative subalgebras in total ones of the class.

One of the advantages of the operational over the term form is that the identities valid in a partial algebra are closed for the same transformations as are those in a total algebra; thus the same sets of identities will serve to define the class (the precautions taken in this regard in [S pp. 41—46] being superfluous). On the other hand, not every strong equational class has proper partial operational realizations for all its identities: for example, the identity $(x \vee y) \wedge y = y$ could not hold in a partial lattice nor $(xy)y^{-1} = x$ in a partial group. Indeed, by substituting a non-totally defined term for any variable appearing on only one side of an identity, one sees that any instance of such a "non-regular" identity entails that the operational strong algebra in which it holds is total.

REFERENCES

[B] Burmeister, P., An embedding theorem..., *Algebra Universalis* 3 (1973), 271—279. *MR* **50** # 2032.

[F] Fleischer, I., Extending a partial equivalence to a congruence and relative embeddings in universal algebras, *Fund. Math.* **106** (1980), 13—17. *MR* **81m**: 08007.

[F'] Fleischer, I., On extending congruences from partial algebras, *Fund. Math.* **88** (1975), 11—16. *MR* **52** # 3017.

[G] Grätzer, G., *Universal Algebra*, 2nd edn., Springer, New York, 1979. *MR* **80g**: 08001.

[H] Höft, H., Weak and strong equations in partial algebras, *Algebra Universalis* 3 (1973), 203—215. *MR* **50** # 2034.

[K] Kerkhoff, R., Gleichungsdefinierbare Klassen partieller Algebren, *Math. Ann.* **185** (1970), 112—133. *MR* **41** # 6754.

[J] John, R., Gültigkeitsbegriffe für Gleichungen in partiellen Algebren, *Math. Z.* **159** (1978), 25—35. *MR* **57** # 12348.

[P] Poythress, V. S., Partial morphisms on partial algebras, *Algebra Universalis* 3 (1973), 182—202. *MR* **48** # 8352.

[2] which should have been credited in [F]; it yields the main theorem of [B] on the basis of the complete HSP characterization claimed in [H]. It could also be deduced by embedding in the one-point completion, which is strong regular identity preserving and initial [F' top p. 14], once it is realized that only regular identities can hold non-vacuously in partial operational algebras (see below).

[S′] Słomiński, J., A theory of extensions of quasi-algebras to algebras. *Rozprawy Mat.* **40** (1964), 1—62. *MR* **31**#111.

[S] Słomiński, J., Peano-algebras and quasi-algebras, *Dissertationes Mathematicae, Rozprawy Mat.* **57** (1968), 1—60. *MR* **37**#5135.

[E] Evans, T., The word problem for abstract algebras, *J. London Math. Soc.* **26** (1951), 64—71. *MR* **12**—475.

[BNR] Bartol, W., Niwiński, D. and Rudak, L., Completion varieties (preprint).

(Received September 19, 1984)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WINDSOR
401 SUNSET AVENUE
WINDSOR
N9B 3P4
CANADA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 229—238

GOSSIPS BY CONFERENCE CALLS

ÁKOS SERESS

Introduction

There are n ladies each knowing a gossip not known by the others. They communicate by telephone, anybody can speak to anybody and whenever two ladies talk they inform the other one about all of the gossips they know at that time. We should like to give an "economical" sequence of calls such that everybody hears each gossip. Two possible definitions of "economical" are:

(i) the number of the calls is minimal,

(ii) everybody hears each gossip exactly once.

(i) was solved independently by B. Baker and R. Shostak [1], R. T. Bumby [3], A. Hajnal, E. C. Milner and E. Szemerédi [6], J. H. Spencer (unpublished) and R. Tijdeman [13], $2n-4$ calls are necessary and sufficient for $n \geqq 4$. Clearly (ii) cannot be done for all n's. The set F_2 of feasible n's was determined in [10]: $F_2=\{1, 2, 4, 8, 12, 16\} \cup \{n: n \geqq 20 \text{ and } 2|n\}$.

The main aim of this paper is to generalize problem (ii) for conference calls, i.e. when each conversation connects k ladies ($k \geqq 3$). We shall determine the feasible n's for all k with finitely many exceptions. Problem (i) for conference calls was solved by J.-C. Bermond [2], D. J. Kleitman and J. Shearer [8] and K. Lebensold [9], the necessary and sufficient number of calls is $\left\lceil \frac{n-k}{k-1} \right\rceil + \left\lceil \frac{n}{k} \right\rceil$ if $n \leqq k^2$ and $2\left\lceil \frac{n-k}{k-1} \right\rceil$ if $n \geqq k^2$. In the second paragraph we examine problem (ii) for directed graphs, i.e. when only one of the ladies tells her informations to the other one. Problem (i) was solved by M. C. Golumbic [5], F. Harary and A. J. Schwenk [7] in this case, $2n-2$ calls are necessary and sufficient. A complete answer of problem (ii) is also available.

To make the description easier we introduce the notion of the joined gossip (JG): when after some calls there are some gossips each of them of none of them known by each of the ladies, these gossips can be considered further as a single (joined) gossip.

The phrase "to organize the conversations" means to give a sequence of calls such that everybody hears each gossip exactly once.

1980 *Mathematics Subject Classification.* Primary 05C65; Secondary 05C99.
Key words and phrases. Gossip problem, algorithm.

§ 1

Denote by F_k the set of positive integers for which it is possible to organize the conversations by k-conference calls. Clearly $k^m \in F_k$ $(m=0, 1, 2, \ldots)$.

LEMMA 1. *Suppose* $n \in F_k$, $n>1$ *and let us consider a sequence of calls proving this property.*

(i) *Associating each lady to her first partners, we get disjoint k-tuples.*

(ii) *Associating each lady to her last partners, we get disjoint k-tuples, too.*

(iii) *If* $n \in F_k$, $n>1$ *then* $k|n$.

The — trivial — proof is well-known for $k=2$ ([10] or D. B. West [14]); the same proof is valid in the case $k \geqq 3$.

LEMMA 2. *If* $n \in F_k$ *then* $k-1|n-1$.

PROOF. Let us consider a sequence of calls proving $n \in F_k$ and a fixed gossip G. At the beginning G is known by one lady, at the end G is known by n ladies and G is told to 0 or $k-1$ ladies at each conversation. Hence $k-1|n-1$.

PROPOSITION 1. $n \in F_k$, $n>1 \Rightarrow n=k+xk(k-1)$ *(x is a non-negative integer).*

PROOF. The solution of the simultaneous congruence system $n \equiv 1 \bmod (k-1)$, $n \equiv 0 \bmod (k)$ is $n \equiv k \bmod (k^2-k)$.

THEOREM 1. *There exists a bound* $x(k)$ *for each* k *such that* $x \geqq x(k) \Rightarrow k+$ $+xk(k-1) \in F_k$. *Moreover* $x(k) \leqq 4k^2+2k$.

This theorem will be proved by a series of lemmas.

LEMMA 3. $s=1+p(k-1)$ *JG's* $(1 \leqq p \leqq k)$ *and* mk^2 *ladies are given. The ladies are divided into* s *groups,* $A_1, A_2, \ldots, A_s$, $|A_i|=a_i$, *and the members of* A_i *know the* i^{th} *JG. Suppose that the following conditions hold:*

a) $m \leqq a_i$,

b) $a_i \leqq mk$,

c) $k-1|a_i-a_j$ *and* $k-1|a_i-m$ *for all* $1 \leqq i,j \leqq s$.

Then it is possible to organize the conversations.

PROOF. By induction on m. If $m=1$ then $1 \leqq a_i \leqq k$ and $k-1|a_i-a_j$ for all i, j. Hence $a_i=1$ or k, consequently $\sum_{i=1}^{s} a_i=k^2$ implies that $a_i=1$ $pk-k$ times, and $a_i=k$ $k+1-p$ times. By Lemma 1 (i), it is possible to organize the conversations. Suppose we are done for m and let $m+1 \leqq a_1 \leqq a_2 \leqq \ldots \leqq a_s \leqq (m+1)k$, $\sum_{i=1}^{s} a_i=$ $=(m+1)k^2$. Let us choose one lady from A_i for $1 \leqq i \leqq pk-k$ and k ladies from A_i for $pk-k+1 \leqq i \leqq pk-p+1$. Each JG is represented in this new group and it is possible to organize the conversations. The conditions of the Lemma are fulfilled in the remaining part: if $a_{pk-k+1}=m+1$ (for the original a_{pk-k+1}) then $\sum_{i=1}^{s} a_i \leqq$ $\leqq (pk-k+1)(m+1)+(k-p)((m+1)k)=(m+1)(k^2-k+1)$, hence a) is true in the rest. If $a_{pk-k}=(m+1)k$ then $\sum_{i=1}^{s} a_i \geqq (pk-k+1)(m-1)+(k-p+2)((m+1)k)=$ $=(m+1)(k^2+k-1)$ hence b) is true. Clearly c) is also satisfied. By the inductional hypothesis it is possible to organize the conversations in the rest, too.

LEMMA 4. *If* $0 \leqq p \leqq k+1$ *then* $k^2+pk^2(k-1) \in F_k$.

PROOF. The case $p=0$ and $p=k+1$ is trivial. If $1 \leqq p \leqq k$ then divide the ladies into $s=1+p(k-1)$ groups, $A_1, A_2, \ldots, A_s$, such that $|A_i|=k^2$ for all i's and let us organize the conversations in each group. We get s JG's for which Lemma 3 can be applied.

LEMMA 5. *If* $1 \leqq p \leqq k-2$ *then* $k^4+pk^2(k-1) \in F_k$.

PROOF. Let us divide the ladies into $1+(p+2)(k-1)$ groups, $A_1, A_2, \ldots, A_{k-1}$, $B_1, B_2, \ldots, B_{k-1}$, $C_1, C_2, \ldots, C_{k-1}$, $D_1, D_2, \ldots, D_{(p-1)(k-1)+1}$, such that $|A_i|=k^3$, $|B_i|=|C_i|=|D_i|=k^2$, and let us organize the conversations in each group. After that we make $k-1$ new groups, $Y_1, Y_2, \ldots, Y_{k-1}$ of size k^3: in Y_j there are k^2 ladies from A_i for $1 \leqq i \leqq k-1$, k ladies from $B_{(j-1)(k-p)+1}, B_{(j-1)(k-p)+2}, \ldots, B_{(j-1)(k-p)+k-p}$ and one from each of the other groups. (The index of B_m must be understood mod $(k-1)$.) We choose a number $j(i)$ for all $1 \leqq i \leqq k-1$ such that there are k ladies from B_i in $Y_{j(i)}$, send $k-1$ ladies back to B_i and put $k-1$ ladies from C_i in $Y_{j(i)}$. Hence we obtain the groups Y_i'. In Y_i' there are k^2 ladies from $k-1$ groups, k ladies from $k-p$ groups and one lady from pk groups hence it is possible to organize the conversations in Y_i'. In the rest there are $k^2(pk+k-p)$ ladies altogether, in the remaining part of A_i there are $k^3-(k-1)k^2=k^2$ ladies, in B_i there are k^2- $-(k-p-1)k-p=pk+k-p$ ladies, in C_i $k^2-1\,k-(k-2)\,1=k^2-2k+2$ ladies and in D_i k^2-k+1 ladies. Since $pk+k-p \leqq k^2-2k+2$ the conditions of Lemma 3 are fulfilled in the rest and it is possible to organize the conversations (see Fig. 1).

Lemma 5 in the case $k=4$, $p=1$, $j(1)=j(2)=2$, $j(3)=1$

	Y_1'				Y_2'				Y_3'				*rest*
A_1	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	16
A_2	2222	2222	2222	2222	2222	2222	2222	2222	2222	2222	2222	2222	16
A_3	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333	16
B_1	4444				4				4444				7
B_2	5555				5				5555				7
B_3	6				6666				6666				7
C_1	7				7777				7				10
C_2	8				8888				8				10
C_3	9999				9				9				10
D_1	0				0				0				13

Fig. 1

LEMMA 6. $k^4+(k-1)^2k^2 \in F_k$.

PROOF. Let us divide the $k^4+(k-1)^2k^2$ ladies into $3k-2$ groups, there are k^3 ladies in $A_1, A_2, \ldots, A_{2k-3}$, k^2 ladies in $B_1, B_2, \ldots, B_{k+1}$ and let us organize the conversations in each group. After that we make $2k-3$ new groups, $Y_1, Y_2, \ldots$

$\dots, Y_{2k-3}$ of size k^3: in Y_j there are k^2 ladies from $A_{(j-1)(k-1)+1}, A_{(j-1)(k-1)+2}, \dots$ $\dots, A_{(j-1)(k-1)+k-1}$, k ladies from the other A_i's. If $1 \leqq j \leqq k-2$ then in Y_j there are k ladies from B_1 and one from each of the other B_i's, if $k-1 \leqq j \leqq 2k-4$ then in Y_j there are k ladies from B_2 and one from each of the other B_i's, in the case $j=2k-3$ in Y_j there are k ladies from B_3 and one from the other B_i's. Clearly it is possible to organize the conversations in each Y_j. In the rest there are $(k+1)k^2$ ladies altogether; the remaining part of A_i $(1 \leqq i \leqq 2k-3)$ is of size $k^3-(k-1)k^2-(k-2)k=2k$, in B_1, B_2 there are $k^2-(k-2)k-(k-1)\cdot 1=k+1$ ladies, in B_3 $k^2-k-(2k-4)\cdot 1=k^2-3k+4$, in $B_4, B_5, \dots, B_{k+1}$ k^2-2k+3 ladies. Since $k^2-3k+4 \geqq k+1$, Lemma 3 can be applied (see Fig. 2).

Lemma 6 in the case $k=5$

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	*rest*
A_1	25	25	5	25	5	25	5	10
A_2	25	5	25	25	5	25	5	10
A_3	25	5	25	5	25	25	5	10
A_4	25	5	25	5	25	5	25	10
A_5	5	25	25	5	25	5	25	10
A_6	5	25	5	25	25	5	25	10
A_7	5	25	5	25	5	25	25	10
B_1	5	5	5	1	1	1	1	6
B_2	1	1	1	5	5	5	1	6
B_3	1	1	1	1	1	1	5	14
B_4	1	1	1	1	1	1	1	18
B_5	1	1	1	1	1	1	1	18
B_6	1	1	1	1	1	1	1	18

Fig. 2

PROPOSITION 2. *$k^2+x(k-1)k^2 \in F_k$ for all $x \geqq 0$.*

PROOF. By induction on x. If $0 \leqq x \leqq 2k$ then we are done by the virtue of Lemmas 4, 5, 6. Suppose the statement is true when $x<x_0$, $x_0 \geqq 2k+1$. Let us write x_0 in the form $x_0=2+(2k-1)y+m$, $y \geqq 1$, $0 \leqq m<2k-1$. Because of the inductional hypothesis $k^2+y(k-1)k^2 \in F_k$ and $k^2+(y+1)(k-1)k^2 \in F_k$. Divide the $k^2+x_0(k-1)k^2$ ladies into $2k-1$ groups, $|A_1|=|A_2|=\dots=|A_m|=k^2+(y+1)(k-1)k^2$, $|A_{m+1}|=\dots=|A_{2k-1}|=k^2+y(k-1)k^2$ and let us organize the conversations in each group. Since $1+x_0(k-1) \leqq k^2(1+y(k-1))$ iff $m \leqq k-1+(k-1)^2y$ and $k^2(1+(y+1)(k-1)) \leqq k(1+x_0(k-1))$ iff $0 \leqq m+(k-1)(y-1)$ are true, Lemma 3 can be applied.

LEMMA 7. *Let $1 \leqq m \leqq k-2$. Suppose there are $(2m+1)k^2-mk$ ladies and $3k-2$ JG's. $k-1$ JG's are known by $m+1$ ladies, one JG by $mk+1$ ladies, $k-1$ JG's by $(m-1)k+2$ ladies, one JG by $m+k$ ladies and $k-2$ JG's by $(m+2)k-1$ ladies. Then it is possible to organize the conversations.*

PROOF. The organization of the conversations can be read from the table of Fig. 3. The groups A_i are as large as described above. First we choose one lady from each of $A_1, A_2, \dots, A_k$ and $A_{k+1}, A_{k+2}, \dots, A_{2k}$, these ladies tell their JG's to each other. After that we make new groups given by the columns. From those ladies whom we have chosen first we order two ladies to the first $k-m$ columns and one to others. Each information is represented in each column and clearly it is possible to organize the conversations anywhere.

$$
\begin{array}{ll|c|cccc}
|A_1| & = m+1 & 1 & & 1\ 1 \dots 1 & & \\
\vdots & & \vdots & & \vdots \quad \vdots & & \\
|A_{k-1}| & = m+1 & 1 & & 1\ 1 \dots 1 & & \\
|A_k| & = mk+1 & 1 & & k\ k \dots k & & \\
\hline
|A_{k+1}| & = (m-1)k+2 & 1 & & & k\ k \dots k & 1 \\
\vdots & & \vdots & & & \vdots \quad \vdots & \vdots \\
|A_{2k-1}| & = (m-1)k+2 & 1 & & & k\ k \dots k & 1 \\
|A_{2k}| & = m+k & 1 & & & 1\ 1 \dots 1 & k \\
\hline
|A_{2k+1}| & = (m+2)k-1 & & 1\ 1 \dots 1 & k\ k \dots k & 1\ 1 \dots 1 & k \\
\vdots & & & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \\
|A_{3k-2}| & = (m+2)k-1 & & \underbrace{1\ 1 \dots 1}_{k-m} & \underbrace{k\ k \dots k}_{m} & \underbrace{1\ 1 \dots 1}_{m-1} & \underbrace{k}_{1}
\end{array}
$$

Fig. 3

LEMMA 8. *Let* $1 \leqq m \leqq k-2$. *Then* $n=(m+1)k^2-mk+2(k^3-k^2) \in F_k$.

PROOF. Divide the n ladies into $3k-2$ groups, $|A_1|=\dots=|A_{k-m}|=k$, $|A_{k-m+1}|=\dots=|A_{3k-2}|=k^2$ and let us organize the conversations in each group. After that we divide the ladies into two groups given by the table of Fig. 4. In the first column it is possible to organize the conversations by Lemma 7. In the second column we make $k-m-2$ groups of size k^2 by the following method: we always choose one lady from each of the first $k-m$ rows and k ladies from the last m rows. We choose $k-m-2$ times k ladies and $k+m$ times one lady from the other rows, k ladies are chosen from those rows where the actual size of the remaining part of A_i is the largest. It can be easily seen that finally one lady remains in each of the first $k-m$ and the last m groups while the size of the rest of the other A_i's is between k and k^2. (The precise proof is similar to the proof of Lemma 3; the method works in the case $m=k-2$, when $k^2-(m+2)k+1=1$, too.) After the ladies from the first $k-m$ and last m rows tell their informations to each other we obtain $2k-1$ JG's and k^2 ladies, each JG is known by at least k and at most k^2 ladies, hence Lemma 3 can be applied.

$\lvert A_1\rvert$	$= k$	$m+1$	$k-m-1$
$\vdots$		$\vdots$	$\vdots$
$\lvert A_{k-m}\rvert$	$= k$	$m+1$	$k-m-1$
$\lvert A_{k-m+1}\rvert$	$= k^2$	$m+1$	k^2-m-1
$\vdots$		$\vdots$	$\vdots$
$\lvert A_{k-1}\rvert$	$= k^2$	$m+1$	k^2-m-1
$\lvert A_k\rvert$	$= k^2$	$mk+1$	k^2-mk-1
$\lvert A_{k+1}\rvert$	$= k^2$	$(m-1)\,k+2$	$k^2-mk+k-2$
$\vdots$		$\vdots$	$\vdots$
$\lvert A_{2k-1}\rvert$	$= k^2$	$(m-1)\,k+2$	$k^2-mk+k-2$
$\lvert A_{2k}\rvert$	$= k^2$	$m+k$	k^2-m-k
$\lvert A_{2k+1}\rvert$	$= k^2$	$(m+2)\,k-1$	$k^2-mk-2k+1$
$\vdots$		$\vdots$	$\vdots$
$\lvert A_{3k-2}\rvert$	$= k^2$	$(m+2)\,k-1$	$k^2-mk-2k+1$

Fig. 4

PROPOSITION 3. *Let* $1\leqq m\leqq k-2$. *If* $x\geqq 3k-2$ *then* $(m+1)k^2-mk+x(k^3-k^2)\in F_k$.

PROOF. Let $3k+2\leqq x\leqq 3k^2+3k-1$. Let us write x in the form $x=5+(3k-3)y+p$, $0\leqq p<3k-3$. Divide the $(m+1)k^2-mk+x(k^3-k^2)$ ladies into $3k-2$ groups, $|A_1|=(m+1)k^2-mk+2(k^3-k^2)$, $|A_2|=\ldots=|A_{p+1}|=k^2+(y+1)(k^3-k^2)$, $|A_{p+2}|=\ldots=|A_{3k-2}|=k^2+y(k^3-k^2)$ and let us organize the conversations in each group. (It is possible by Lemma 8 and Proposition 2.) We choose $(2m+1)k^2-mk$ ladies in distribution given by Lemma 7 and organize the conversations among them. In the remaining part the ladies' number is divisible by k^2. If $x\leqq 2k^2+(m+1)k+1$ then Lemma 3 can be applied. All conditions of Lemma 3 are fullfilled trivially except condition a) for the rest of A_1: $(m+1)k^2-mk+2(k^3-k^2)-(m+1)\geqq k^{-2}((m+1)k^2-mk+x(k^3-k^2))-((2m+1)k^2-mk))$ is equivalent with $x\leqq 2k^2+(m+1)k+1$. If $2k^2+(m+1)k+2\leqq x\leqq 3k^2+3k-1$ then we make k^2+2k-3 groups of size k^3: we always choose one lady from A_1. We select $k-1$ times k^2 ladies, $k-1$ times k ladies, $k-1$ times one lady from the other A_i's, k^2 ladies are chosen from those groups where the actual size of the rest is the largest and one from those groups where the actual size is the smallest. Finally

$$(m+1)k^2-mk+x(k^3-k^2)-\big((2m+1)k^2-mk\big)-(k+3)(k-1)k^3 =$$
$$= (x-k^2-3k)(k^3-k^2)-mk^2$$

ladies remained altogether, the rest of A_1 is of size

$$(m+1)k^2-mk+2(k^3-k^2)-(m+1)-(k^2+2k-3) =$$
$$= 2(k^3-k^2)+mk^2-(m+2)k-m+2.$$

Lemma 3 can be applied in the remaining part:

$$(x-k^2-3k)(k-1)-m \leqq 2(k^3-k^2)+mk^2-(m+2)k-m+2 \leqq$$
$$\leqq k\big((x-k^2-3k)(k-1)-m\big)$$

if $2k^2+(m+1)k+2\leqq x\leqq 3k^2+3k-1$. The size of the other A_i's can differ from the average size by at most k^2 hence the conditions of Lemma 3 hold for these groups, too. If $3k^2+3k\leqq x\leqq 9k^2-1$ then let us write x in the form $x=3k+5+(3k-3)y+p$, $0\leqq p<3k-3$. Divide the $(m+1)k^2-mk+x(k^3-k^2)$ ladies into $3k-2$ groups,

$$|A_1| = (m+1)k^2-mk+(3k+2)(k^3-k^2),$$
$$|A_2| = \ldots = |A_{p+1}| = k^2+(y+1)(k^3-k^2),$$
$$|A_{p+2}| = \ldots = |A_{3k-2}| = k^2+y(k^3-k^2)$$

and let us organize the conversations in each group. We choose $(2m+1)k^2-mk$ ladies in distribution given by Lemma 7. After that Lemma 3 can be applied since $(m+1)k^2-mk+(3k+2)(k^3-k^2)-(m+1)\leqq k\big(x(k^3-k^2)-mk^2\big)k^{-2}$ if $x\geqq 3k^2+3k$ and the other conditions of Lemma 3 are fulfilled trivially. Finally, let $x_0\geqq 9k^2$ and suppose $(m+1)k^2-mk+x(k^3-k^2)\in F_k$ for all $x<x_0$. Let us write x_0 in the

form $x_0=3+(3k-2)y+p$, $0\leq p<3k-2$, divide the ladies into $3k-2$ groups,

$$|A_1| = (m+1)k^2-mk+y(k^3-k^2)$$

$$|A_2| = \ldots = |A_{p+1}| = k^2+(y+1)(k^3-k^2),$$

$$|A_{p+2}| = \ldots = |A_{3k-2}| = k^2+y(k^3-k^2)$$

and let us organize the conversations in each group. Applying Lemmas 7 and 3 we obtain $(m+1)k^2-mk+x_0(k^3-k^2)\in F_k$.

LEMMA 9. *Suppose that there are $k+2(k-1)k^2$ ladies and $3k-2$ JG's. $2k-2$ JG's are known by k^2-k+1 ladies, 2 JG's by k ladies and $k-2$ JG's by $2k-1$ ladies. Then it is possible to organize the conversations.*

PROOF. The organization of the conversations can be read from the table of Fig. 5. First we choose one lady from each of $A_1, A_2, \ldots, A_k$ and $A_{k+1}, A_{k+2}, \ldots$ $\ldots, A_{2k}$, these ladies tell their JG's to each other. After that we make new groups given by the columns. From those whom we have chosen first we order two ladies to the first column and one to the others.

$\lvert A_1\rvert = k$	1			$1\ 1 \ldots 1$
$\lvert A_2\rvert = k^2-k+1$	1			$k\ k \ldots k$
$\vdots$	$\vdots$			$\vdots \quad \vdots$
$\lvert A_k\rvert = k^2-k+1$	1			$k\ k \ldots k$
$\lvert A_{k+1}\rvert = k$	1		$1\ 1 \ldots 1$	
$\lvert A_{k+2}\rvert = k^2-k+1$	1		$k\ k \ldots k$	
$\vdots$	$\vdots$		$\vdots \quad \vdots$	
$\lvert A_{2k}\rvert = k^2-k+1$	1		$k\ k \ldots k$	
$\lvert A_{2k+1}\rvert = 2k-1$		1	$1\ 1 \ldots 1$	$1\ 1 \ldots 1$
$\vdots$		$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$
$\lvert A_{3k-2}\rvert = 2k-1$		1	$1\ 1 \ldots 1$	$1\ 1 \ldots 1$
		$\underbrace{}_{1}$	$\underbrace{}_{k-1}$	$\underbrace{}_{k-1}$

Fig. 5

LEMMA 10. $k+4(k^3-k^2)\in F_k$.

PROOF. Divide the $k+4(k^3-k^2)$ ladies into $4k-3$ groups, $|A_1|=k$, $|A_2|=$ $=|A_3|=\ldots=|A_{4k-3}|=k^2$ and let us organize the conversations in each group. After that we divide the ladies into two groups given by the table of Fig. 6. We apply Lemma 9 in the second column. In the first column we organize the conversations among the ladies from $A_1, A_{k+1}, \ldots, A_{2k-1}$. Hence we get $3k-2$ JG's and $3k-2$ groups, $B_1, B_2, \ldots, B_{3k-2}$, $|B_1|=\ldots=|B_{k-1}|=k-1$, $|B_k|=\ldots=|B_{2k-3}|=$ $=k^2-2k+1$, $|B_{2k-2}|=|B_{2k-1}|=k^2-k$ and $|B_{2k}|=\ldots=|B_{3k-2}|=k^2-1$. After that we make $k-2$ groups of size k^2, we always choose one lady from $B_1, \ldots, B_{k-1}$ and k ladies from B_k. We select $k-3$ times k ladies and $k+1$ times one lady from the other groups, k ladies are chosen from those B_i's where the actual size of the group is the largest. Finally one lady remains in each of $B_1, B_2, \ldots, B_k$ while the size of the rest of the other B_i's is between k and k^2. After the ladies from $B_1, B_2, \ldots, B_k$ tell their JG's to each other Lemma 3 can be applied.

A_1	$k-1$	1
A_2	k^2-1	1
$\vdots$	$\vdots$	$\vdots$
A_k	k^2-1	1
A_{k+1}	$k-1$	k^2-k+1
$\vdots$	$\vdots$	$\vdots$
A_{2k-1}	$k-1$	k^2-k+1
A_{2k}	$k-1$	k^2-k+1
$\vdots$	$\vdots$	$\vdots$
A_{3k-2}	$k-1$	k^2-k+1
A_{3k-1}	k^2-k	k
A_{3k}	k^2-2k+1	$2k-1$
$\vdots$	$\vdots$	$\vdots$
A_{4k-3}	k^2-2k+1	$2k-1$

Fig. 6

PROPOSITION 4. *If* $x \geqq 4k+2$ *then* $k+x(k^3-k^2) \in F_k$.

The proof is the same as the proof of Proposition 3 and we omit it.

Theorem 1 is proved by Propositions 2, 3, 4.

As a closure of the paragraph we mention some open questions.

PROBLEM 1. *Can a linear upper bound for* $x(k)$ *be given? More precisely, is the following conjecture true?* $x \geqq 2k+1,\ x \neq 3k \Rightarrow k+xk(k-1) \in F_k$. (The conjecture is true in the cases $k=3, 4$.)

Problem 2 is originated from G. O. H. Katona.

PROBLEM 2. *For which values of* k_1, k_2 *does a bound* $n(k_1, k_2)$ *exist such that it is possible to organize the conversations for all* $n \geqq n(k_1, k_2)$ *using* k_1- *and* k_2-*conference calls?*

Theorem 2 which we state without proof summarizes our results.

THEOREM 2. a) *If* $n(k_1, k_2)$ *exists then* $(k_1, k_2)=1$ *and* $(k_1-1, k_2-1)=1$. b) $n(2, 2k+1)$ *exists, moreover* $n(2, 2k+1)=O(k)$.

PROBLEM 3. *The ladies communicate by telephone calls. The expense of a call is* x *and the expense of telling a gossip is* y. *What is the minimal cost of a sequence of calls such that everybody hears each gossip? Specially, what is the minimal number of calls when everybody hears each gossip exactly once?*

D. B. West [15] proved that $2.25n-6$ calls are enough if $4|n$. The same question can be asked for conference calls, too.

ADDED in proof. (1) In [12] it has been shown that $x(k) \leqq k^2+5k+2$. (2) For $n \in F_k$, denote by $f_k(n)$ the minimal number of k-conference calls s.t. everybody hears each gossip exactly once. In [11], it has been shown that

$$f_2(n) = 2.25n-6 \quad \text{if} \quad 4|n, \quad \text{and}$$

$$2.25n-4.5 \leqq f_2(n) \leqq 2.25n-3.5 \quad \text{if} \quad n \equiv 2 (\operatorname{mod} 4).$$

In [12], it has been proved that $f_k(n) < \frac{3k^2-k-1}{k^3-k^2} n-k$, for arbitrary k.

§ 2

In this paragraph we discuss problem (ii) of the introduction for directed graphs, i.e. when each conversation is between two ladies and only one of them tells the gossips known at that time to the other. If $A \to B$ is the first call in any sequence of calls then A hears her own gossip. Hence it is worth to examine a weaker version of problem (ii): does a sequence of calls exist such that everybody hears the gossips of the other ladies exactly once and her own gossip can be heard once more?

THEOREM 3. a) *A sequence of calls exists for all n's such that everybody hears each gossip exactly once and* $n-1$ *ladies hear their own gossips.*

b) *There is no sequence of calls such that at most* $n-2$ *ladies hear their own gossips.*

PROOF. a) Denote the ladies by $A_1, A_2, \ldots, A_n$.

$$A_1 \to A_n,\ A_2 \to A_n,\ \ldots,\ A_{n-1} \to A_n,\ A_n \to A_{n-1},\ A_n \to A_{n-2}, \ldots, A_n \to A_1$$

is a good sequence of calls.

b) Indirectly suppose that n is the least number such that at most $n-2$ ladies hear their own gossips. (Clearly $n>2$.) Let S be a sequence of calls showing this property and $A \to B$ the last conversation in S. A knows each information before this call, specially she knows B's gossip, B hears her own story. B knows none of the other gossips before the $A \to B$ call because she would hear these gossips at least twice. Hence all edges joining B and different from $A \to B$ are of the form $B \to C$. Let us rub out all edges of S joining B. We obtain a sequence T of calls among $n-1$ ladies. Executing T, 1) everybody hears each gossip; 2) nobody hears a gossip twice; 3) at most $n-3$ ladies hear their own information.

2) and 3) are true because nobody hears such gossip which she did not hear at S and B heard her own gossip at S. 1) is true because B did not take part in carrying informations out of her own gossip in S. T shows that n is not the least counter-example hence we obtain a contradiction.

REFERENCES

[1] BAKER, B. and SHOSTAK, R., Gossips and telephones, *Discrete Math.* **2** (1972), 191—193. *MR* **46**#68.

[2] BERMOND, J. C., Le problème des "ouvoirs", (hypergraph gossip problem), in J. C. Bermond, ed., *Problèmes Comb. et Théorie des Graphes*, C.N.R.S. Colloq. Paris, 1976.

[3] BUMBY, R. T., A problem with telephones, *SIAM J. Algebraic Discrete Methods* **2** (1981), 13—18. *MR* **82f**: 05083

[4] GERÉB, M., Problem appearing in: *Finite and infinite sets*, Colloq. Math. Soc. János Bolyai 37, North-Holland, Amsterdam, 1984.
[5] COLUMBIC, M. C., *The general gossip problem*, IBM Research Report, 1974.
[6] HAJNAL, A., MILNER, E. C. and SZEMERÉDI, E., A cure for the telephone disease, *Canad. Math. Bull.* **15** (1972), 447—450. *MR* **47**# 3184.
[7] HARARY, F. and SCHWENK, A. J., The communication problem on graphs and digraphs, *J. Franklin Inst.* **297** (1974), 491—495. *MR* **50**# 1980.
[8] KLEITMAN, D. J. and SHEARER, J., Further gossip problems, *Discrete Math.* **30** (1980), 151—156. *MR* **81d**: 05068.
[9] LEBENSOLD, K., Efficient communication by phone calls, *Studies in Appl. Math.* **52** (1973), 345—358. *MR* **49**# 4797.
[10] SERESS, Á., Gossiping old ladies, *Discrete Math.* **46** (1983), 75—81. *MR* **85b**: 05115
[11] SERESS, Á., Quick gossiping without duplicate transmissions, *Graphs and Combinatorics* (to appear).
[12] SERESS, Á., Quick gossiping by conference calls, *SIAM J. Algebraic Discrete Methods* (to appear).
[13] TIJDEMAN, R., On a telephone problem, *Nieuw Arch. Wisk.* **19** (1971), 188—192. *MR* **49**# 7151.
[14] WEST, D. B., A class of solutions to the gossip problem, Part I, *Discrete Math.* **39** (1982), 307—326.
[15] WEST, D. B., Gossiping without duplicate transmissions, *SIAM J. Algebraic Discrete Methods* **3** (1982), 418—419.

(Received September 20, 1983)

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 239—245

KRYLOV—STAYERMANN POLYNOMIALS ON THE JACOBI ROOTS

P. VÉRTESI

1. Introduction

For a given interpolatory matrix $X=\{x_{kn}\}\subset[-1, 1]$ with

$$(1.1)\qquad -1\leqq x_{nn}<x_{n-1,n}<\ldots<x_{2n}<x_{1n}\leqq 1,\quad n=1,2,\ldots,$$

let us consider the uniquely defined interpolatory procedure $K_n(f, X, x)$ discussed by Krylov and Stayermann [1] for a continuous $f(x)$ on $[-1, 1]$ (shortly $f\in C$) as follows

$$(1.2)\qquad \begin{cases} K_n(f, X, x_{kn})=f(x_{kn}), & k=1,2,\ldots,n,\\ K_n^{(i)}(f, X, x_{kn})=0, & k=1,2,\ldots,n,\quad i=1,2,3,\\ K_n \text{ is a polynomial of degree } \leqq 4n-1. \end{cases}$$

The question is, as usual, to obtain matrices X such that the relation

$$(1.3)\qquad \lim_{n\to\infty}\|K_n(f, X, x)-f(x)\|=0$$

should hold for any $f\in C$. (Generally, $\|f(x)\|_{[a,b]}=\|f\|_{[a,b]}=\max_{a\leqq x\leqq b}|f(x)|$; if $[a, b]=[-1, 1]$ we write $\|\cdot\|$.)

If the nodes are the roots of the Chebyshev polynomial $P_n^{(-1/2,-1/2)}(x)=T_n(x)=\cos n\theta$, $x=\cos\theta$, then (1.3) is true whenever $f\in C$ (see [1]).

Considering other Jacobi parameters, i.e. when the nodes $X=\{x_{kn}\}$ are $X^{(\alpha,\beta)}=\{x_{kn}^{(\alpha,\beta)}\}$, the roots of the n-th Jacobi polynomial $P_n^{(\alpha,\beta)}(x)$, $\alpha,\beta>-1$ (see [2, Ch. 4]), the next theorem can be proved.

THEOREM 1.1 (Laden [3], Knoop and Stockenberg [4]). a) *If* $-3/4\leqq\alpha,\beta<-1/4$ *then* (1.3) *holds for any* $f\in C$. *Moreover, if* $\alpha=\beta=-1/4$, *for a suitable* $f\in C$, (1.3) *is not true.*

b) *If* $-3/4\leqq\alpha,\beta\leqq-1/4$, *the* K_n *process is a positive operator.*

c) *For arbitrary* $\alpha,\beta>-1$ *and* $[a,b]\subset(-1,1)$,

$$\lim_{n\to\infty}\|K_n^{(\alpha,\beta)}(f,x)-f(x)\|_{[a,b]}=0\quad \textit{for any}\quad f\in C$$

($K_n^{(\alpha,\beta)}(f,x)$ *stands for* $K_n(f, X^{(\alpha,\beta)}, x)$).

REMARK 1.1. As it is well-known, for the Hermite—Fejér stepparabolas $H_n^{(\alpha,\beta)}(f,x)$ ($H_n^{(\alpha,\beta)}$ is the unique polynomial of degree $\leqq 2n-1$ with $H_n^{(\alpha,\beta)}(f,x_k)=f(x_k)$, $[H_n^{(\alpha,\beta)}(f,x_k)]'=0$, where $x_k:=x_{kn}^{(\alpha,\beta)}$) we have

1980 *Mathematics Subject Classification.* Primary 41A05; Secondary 41A25.
Key words and phrases. Interpolation, Jacobi polynomials.

a) If $-1<\alpha, \beta<0$ then for any $f\in C$

$$\lim_{n\to\infty} \|H_n^{(\alpha,\beta)}(f, x)-f(x)\| = 0. \tag{1.4}$$

Further, if $\max(\alpha, \beta)\geqq 0$, then for a certain $f\in C$, (1.4) does not hold.

b) The $H_n^{(\alpha,\beta)}$ process is a positive operator iff $-1<\alpha, \beta\leqq 0$.

c) For arbitrary $\alpha, \beta>-1$ and $[a, b]\subset(-1, 1)$,

$$\lim_{n\to\infty} \|H_n^{(\alpha,\beta)}(f, x)-f(x)\|_{[a,b]} = 0 \quad \text{for any} \quad f\in C$$

(see [2, 14.6]).

2. The $K_n^{(\alpha,\beta)}$ process for arbitrary $\alpha, \beta>-1$

2.1. In this part we are going to complete Theorem 1.1. Namely we prove as follows.

THEOREM 2.1. *Let* $-1<a<1$ *be fixed. Then*

$$\lim_{n\to\infty} \|K_n^{(\alpha,\beta)}(f, x)-f(x)\|_{[a,1]} = 0 \tag{2.1}$$

for arbitrary $f\in C$ *whenever* $-1<\alpha<-1/4$ *and* $\alpha-\beta<1/2$. *On the other hand, if* $\alpha\geqq-1/4$ *or* $\alpha-\beta>1/2$, *one can find a function* $f\in C$ *for which* (2.1) *does not hold.*

Using $P_n^{(\alpha,\beta)}(x)=(-1)^n P_n^{(\beta,\alpha)}(-x)$ ([2, (4.1.3)]), we can obtain the next

COROLLARY 2.2. *Let* $-1<a<1$ *be fixed. Then*

$$\lim_{n\to\infty} \|K_n^{(\alpha,\beta)}(f, x)-f(x)\|_{[-1,a]} = 0 \tag{2.2}$$

for arbitrary $f\in C$ *whenever* $-1<\beta<-1/4$ *and* $\beta-\alpha<1/2$. *On the other hand if* $\beta\geqq-1/4$ *or* $\beta-\alpha>1/2$, *one can find a function* $f\in C$ *for which* (2.2) *does not hold.*

By the above statements one can easily get the next

COROLLARY 2.3. *If* $-1<\alpha, \beta<-1/4$ *and* $0\leqq|\alpha-\beta|<1/2$, *then*

$$\lim_{n\to\infty} \|K_n^{(\alpha,\beta)}(f, x)-f(x)\| = 0 \quad \textit{for any} \quad f\in C. \tag{2.3}$$

(2.3) *does not hold for a certain* $f\in C$ *if* $\max(\alpha, \beta)\geqq-1/4$ *or* $|\alpha-\beta|>1/2$.

2.2. REMARK. If $-1<\alpha, \beta<-1/4$ but $|\alpha-\beta|=\frac{1}{2}$, we have not been able to prove or disprove (2.3). Our conjecture is that (2.3) holds true even in these cases.

3. Proofs

First we quote some formulas. Sometimes omitting the superfluous notations we have by [3, (6)]

$$K_n(f, x) = \sum_{k=1}^{n} f(x_k)u_k(x)l_k^4(x) := \sum_{k=1}^{n} f(x_k)\tau_k(x) \tag{3.1}$$

where l_k are the fundamental polynomials of Lagrange interpolation, i.e. if $\omega_n(x)= =c_n \prod_{k=1}^{n} (x-x_k)$,

$$l_k(x) = \frac{\omega_n(x)}{\omega_n'(x_k)(x-x_k)}, \quad k = 1, 2, \ldots, n, \tag{3.2}$$

and

$$\begin{aligned} u_k(x) = 1-2(x-x_k)\frac{\omega_n''(x_k)}{\omega_n'(x_k)} + \frac{(x-x_k)^2}{2}\left\{5\left[\frac{\omega_n''(x_k)}{\omega_n'(x_k)}\right]^2 - \frac{4}{3}\frac{\omega_n'''(x_k)}{\omega_n'(x_k)}\right\} + \\ + \frac{(x-x_k)^3}{6}\left\{-15\left[\frac{\omega_n''(x_k)}{\omega_n'(x_k)}\right]^3 + 10\frac{\omega_n''(x_k)\omega_n'''(x_k)}{[\omega_n'(x_k)]^2} - \frac{\omega_n^{(\mathrm{IV})}(x_k)}{\omega_n'(x_k)}\right\}. \end{aligned} \tag{3.3}$$

If $\omega_n(x)=P_n^{(\alpha,\beta)}(x)$, i.e. when $x_k=x_{kn}^{(\alpha,\beta)}$, by the differential equation for the Jacobi polynomials we can write (cf. [2, (4.2.4)] and [4, (2.1)])

$$\begin{aligned} u_k(x) = 1-2t_k(x-x_k)+\frac{11}{6}t_k^2(x-x_k)^2+ \\ +\frac{s_k(x)(x-x_k)^2}{6(1-x_k^2)^2}[4(M-\delta)(1-x_k^2)-8x_k(\gamma+\delta x_k)]- \\ -\left(t_k^3+\frac{x_k}{3}\frac{t_k^2}{1-x_k^2}+\frac{\delta+2}{6}\frac{t_k}{1-x_k^2}\right)(x-x_k)^3 := 1+\sum_{i=1}^{4} I_i, \end{aligned} \tag{3.4}$$

where

$$\begin{cases} \gamma = \alpha-\beta, \quad \delta = \alpha+\beta+2, \quad M = n(n+\delta-1), \\ t_k = \dfrac{\gamma+\delta x_k}{1+x_k^2}, \\ s_k(x) = 1-\dfrac{2\gamma+(2\delta-1)x_k}{1-x_k^2}(x-x_k). \end{cases} \tag{3.5}$$

To estimate the fundamental polynomials l_k we shall use

$$\begin{cases} P_n^{(\alpha,\beta)}(x) = (-1)^n P_n^{(\beta,\alpha)}(-x), \quad P_n(1)\sim n^\alpha, \\ |P_n(x)|\sim|\theta-\theta_j|\theta_j^{-\alpha-1/2}n^{1/2} \le c\dfrac{n^\alpha}{j^{\alpha+1/2}} \quad \text{if} \quad \theta\in[0,\pi-\varrho], \\ |P_n'(x_k)|\sim k^{-\alpha-3/2}n^{\alpha+2}, \quad \text{if} \quad \theta_k\in[0,\pi-\varrho], \\ \theta_k\sim\dfrac{k}{n}, \quad k=1,2,\ldots,n, \\ \theta_{k+1}-\theta_k\sim\dfrac{1}{n}, \quad k=0,1,\ldots,n, \quad \theta_0=0, \quad \theta_{n+1}=\pi. \\ |x-x_k|\sim\dfrac{(j+k)(|j-k|+1)}{n^2}, \quad k\ne j, \quad k=1,2,\ldots,n. \end{cases} \tag{3.6}$$

Here $P_n(x)=P_n^{(\alpha,\beta)}(x)$, $x=\cos\theta$, $x_k=x_{kn}^{(\alpha,\beta)}=\cos\theta_{kn}^{(\alpha,\beta)}=\cos\theta_k$, $x_j=\cos\theta_j$ is the nearest node to x (for a fixed n) and $0<\varrho<\pi$ is arbitrary fixed. The symbol

"$\sim$" ([2, 1.1]) does not depend on θ, k and n. (See [2, (4.1.3), (8.9.2)], [5] and [6, Lemma 3.2]). Here and later $c, c_1, c_2, \ldots$ are different positive constants.

3.1. To prove the direct part of Theorem 2.1, we use the next simple

LEMMA 3.1. *If for a fixed* $X\subset[-1, 1]$ *and for each fixed* η *with* $0<\eta\leqq\eta_0$ *we have both*

(i) $\displaystyle\sum_{|x-x_k|\leqq\eta}|\tau_k(x)|\leqq A,$

(ii) $\displaystyle\lim_{n\to\infty}\sum_{|x-x_k|>\eta}|\tau_k(x)|=0$

holding uniformly in $x\in[c, d]\subset[-1, 1]$, *then*

$$\lim_{n\to\infty}\|K_n(f, x)-f(x)\|_{[c,d]}=0 \quad \textit{for any} \quad f\in C.$$

Here $A>0$ *does not depend on* η.

Indeed, let $|f(x)-f(y)|\leqq\varepsilon$ if $|x-y|\leqq\eta$. Then, by $\sum\tau_k(x)\equiv 1$, (i) and (ii) we have in $[c, d]$

$$|K_n(f, x)-f(x)|\leqq\sum_{k=1}^{n}|f(x_k)-f(x)|\,|\tau_k(x)|=$$

$$=\sum_{|x-x_k|\leqq\eta}+\sum_{|x-x_k|>\eta}\leqq A\varepsilon+2\|f\|\varrho_n\leqq 2A\varepsilon \quad \text{when } n \text{ is big enough}$$

(because $\varrho_n\to 0$) as it was stated (cf. e.g. [3, Theorem A]).

3.2. Let us prove (2.1) if $-1<\alpha<-1/4$ and $\alpha-\beta<1/2$. We use the previous lemma. By (3.4), (3.5) and $(1-x_k^2)^{-1}\leqq cn^2$ we get

$$\begin{aligned}\sum_{|x-x_k|>\eta}|\tau_k(x)|&\leqq c\sum_{|x-x_k|>\eta}\left[\frac{1}{(1-x_k^2)^3}+\frac{n^2}{(1-x_k^2)^2}\right]\left[\frac{P_n(x)}{P_n'(x_k)}\right]^4\leqq\\ &\leqq c\sum_{|x-x_k|>\eta}\frac{n^2}{(1-x_k^2)^2}\left[\frac{P_n(x)}{P_n'(x_k)}\right]^4.\end{aligned}\tag{3.7}$$

By (3.6) and $x\in[a, 1]$ we get

$$\sum_{\ldots}=\sum_{\substack{|x-x_k|>\eta\\ x_k\geqq 0}}+\sum_{\substack{|x-x_k|>\eta\\ x_k<0}}\leqq c\frac{n^{4\alpha}}{j^{4\alpha+2}}\left(\sum_{k=1}^{c_2 n}\frac{n^6}{k^4}\frac{k^{4\alpha+6}}{n^{4\alpha+8}}+\sum_{k=1}^{c_3 n}\frac{n^6}{k^4}\frac{k^{4\beta+6}}{n^{4\beta+8}}\right)\leqq$$

$$\leqq c\left(\frac{n^{4\alpha+1}}{j^{4\alpha+2}}+\frac{n^{4\alpha-4\beta-2}}{j^{4\alpha+2}}\sum_{k=1}^{n}k^{4\beta+2}\right)=o(1)+c\frac{n^{4\alpha-4\beta-2}}{j^{4\alpha+2}}(1+n^{4\beta+3})=o(1)$$

whenever $-3/4<\alpha<-1/4$ and $\alpha-\beta<1/2$ from where we get (ii). The case $-1<\alpha\leqq-3/4$ can be treated similarly.

To get (i), we use that according to (3.4) and (3.5)

$$|u_k(x)|\leqq c\left[1+\frac{|x-x_k|}{1-x_k^2}+\left(\frac{x-x_k}{1-x_k^2}\right)^2+\frac{|x-x_k|^3}{(1-x_k^2)^3}+\frac{n^2|x-x_k|^3}{(1-x_k^2)^2}\right].$$

By this decomposition we write

$$\sum_{|x-x_k|\leq\eta} |u_k(x)|\, l_k^4(x) \leq c\left(\sum\nolimits_1+\sum\nolimits_2+\sum\nolimits_3+\sum\nolimits_4+\sum\nolimits_5\right).$$

Let us estimate Σ_4, say. If $a-\eta_0=-1+\varepsilon$ then $x_k\geq-1+\varepsilon$ whenever $|x-x_k|\leq\eta$ and $x\in[a, 1]$. That means, by (3.6)

$$\sum_{|x-x_k|\leq\eta} \frac{|x-x_k|^3}{(1-x_k^2)^3}\, l_k^4(x) = \sum_{\cdots} \frac{P_n^4(x)}{[P_n'(x_k)]^4|x-x_k|(1-x_k^2)^3} \leq$$

$$\leq c\sum_{\cdots} \frac{n^{4\alpha}}{j^{4\alpha+2}}\, \frac{k^{4\alpha+6}}{n^{4\alpha+8}}\, \frac{n^2}{(k+j)(|k-j|+1)}\, \frac{n^6}{n^6} \leq$$

$$\leq c\Big(\sum_{k<j/2} + \sum_{j/2\leq k\leq 2j} + \sum_{k\geq 2j}\Big) \frac{k^{4\alpha}}{j^{4\alpha+2}(k+j)(|k-j|+1)} \leq$$

$$\leq \frac{c}{j^{4\alpha+4}} \sum_{k=1}^{j} k^{4\alpha} + \frac{c}{j^3} \sum_{i=1}^{j}{}' \frac{1}{i} + \frac{c}{j^{4\alpha+2}} \sum_{k=j}^{n} k^{4\alpha-2} \leq c,$$

even if $\alpha\leq-1/4$.

Similar estimations hold for the other sums, too. That means we proved the direct part of the theorem.

3.3. Now we prove that (2.1) does not hold for a certain $f\in C$ if $\alpha\geq-1/4$ or when $\alpha-\beta>1/2$.

a) Let $\alpha\leq-1/4$ and $\alpha-\beta>1/2$, more exactly $\beta=-3/4-\varepsilon$, $0<\varepsilon<1/4$. If $|u_k(1)|\geq c_6\varepsilon(n-k)^2(1+x_k)^{-3}$ were true at least when $\varrho n\leq k\leq n-c_4$, $0<\varrho<1$, then

$$\sum_{k=1}^{n} |\tau_k(1)| \geq \sum_{k=\varrho n}^{n-c_4} |u_k(1)|\, l_k^4(1) \geq c_6\varepsilon \sum_{k=\varrho n}^{n-c_4} \frac{(n-k)^2}{(1+x_k)^3} \left[\frac{P_n(1)}{P_n'(x_k)}\right]^4 \sim$$

$$\sim n^{4\alpha} \sum_{i=c_4}^{(1-\varrho)n} \frac{n^6}{i^4}\, \frac{i^{4\beta+6}}{n^{4\beta+8}} \sim n^{4\alpha-4\beta-2} = n^{\delta}, \quad \delta>0.$$

So by the uniform boundedness principle we were ready. To prove the mentioned relation we remark that by (3.5)

$$2\gamma-2\delta+1 = -4\beta-3 = 4\varepsilon,$$

from where

$$|s_k(1)| \geq |1-3\varepsilon(1-x_k)(1-x_k^2)^{-1}| \geq 2\varepsilon(1+x_k)^{-1}$$

if ϱ is big enough. Then by (3.4)

$$|u_k(1)| \geq |I_3|-(1+|I_1|+|I_2|+|I_4|) \geq \frac{\varepsilon}{3(1+x_k)^3}\,[3n^2(1-x_k^2)-c_1]-$$

$$-c_2\left[1+\frac{1}{1+x_k}+\frac{1}{(1+x_k)^2}+\frac{1}{(1+x_k)^3}\right].$$

By

$$3n^2(1-x_k^2)-c_1 \geq c_3(n-k)^2 \quad \text{if} \quad n/2\leq k\leq n-c_4,$$

we have

$$|u_k(1)| \geqq c_3 \frac{\varepsilon}{3} \frac{(n-k)^2}{(1+x_k)^3} - \frac{c_5}{(1+x_k)^3} \geqq c_6 \varepsilon \frac{(n-k)^2}{(1+x_k)^3}$$

for $\varrho n \leqq k \leqq n - c_4$ (where ϱ and c_4 are suitably chosen) which was to be proved.

b) Now suppose $\alpha = -1/4 + \lambda$, $\lambda > 0$. As above, we shall see that $|u_k(1)| \geqq \geqq \lambda c_7 k^2$ if $c_8 \leqq k \leqq \mu n$, $0 < \mu < 1$. Then

$$\sum_{k=1}^{n} |\tau_k(1)| \geqq \sum_{k=c_8}^{\mu n} |u_k(1)|\, l_k^4(1) \geqq c_7 \lambda \sum_{k=c_8}^{\mu n} k^2 \left[\frac{P_n(1)}{P_n'(x_k)(1-x_k^2)}\right]^4 \sim$$

$$\sim n^{4\alpha} \sum_{k=c_8}^{\mu n} k^2 \frac{k^{4\alpha+6}}{n^{4\alpha+8}} \frac{n^8}{k^8} \sim \sum_{k=1}^{n} k^{4\alpha} \sim n^{4\alpha+1} \sim n^{4\lambda}$$

from where we get the statement by the uniform boundedness principle.

To estimate $u_k(1)$, let us remark that by (3.5) $2\gamma + 2\delta - 1 = 4\alpha + 3 = 4\lambda + 2$, i.e. $|s_k(1)| \geqq |1 - (3\lambda + 2)(1+x_k)^{-1}| \geqq 3\lambda(1+x_k)^{-1}$, if μ is small enough. So by (3.4)

$$|u_k(1)| \geqq |I_3| - (1 + |I_1| + |I_2| + |I_4|) \geqq \frac{\lambda}{2} \frac{1}{(1+x_k)^3} [3n^2(1-x_k^2) - c_9] -$$

$$- c_{10}\left[1 + \frac{1}{1+x_k} + \frac{1}{(1+x_k)^2} + \frac{1}{(1+x_k)^3}\right].$$

Here $\frac{1}{2} \leqq \frac{1}{1-x_k} < 1$ if $x_k > 0$, further $3n^2(1-x_k^2) - c_9 \geqq c_{11} k^2$ if $c_8 \leqq k \leqq n/2$. So we can write

$$|u_k(1)| \geqq \frac{\lambda c_{11}}{16} k^2 - 4c_{10} \geqq \lambda c_7 k^2 \quad \text{if} \quad c_8 \leqq k \leqq \mu n$$

(where μ and c_8 are suitably chosen), which was to be proved.

c) If $\alpha = -1/4$ and $\alpha - \beta = \frac{1}{2} - \varepsilon$, $\varepsilon > 0$, we get that $s_k(1) = 2\varepsilon(1-x_k)/(1+x_k)$ from where $s_k(1) \geqq \varepsilon$ if $|x_k|$ is small enough, $|x_k| \leqq 2\lambda < 1$, say. Then by (3.4) and (3.5) $u_k(1) \geqq \frac{\varepsilon}{2} n^2$ if $|x_k| \leqq 2\lambda$, $n \geqq n_0$. Now we define a "hat function" $f \in C$ as follows

$$f(x) = \begin{cases} 0 & \text{if} \quad -1 \leqq x \leqq -2\lambda \quad \text{or} \quad 2\lambda \leqq x \leqq 1 \\ 1 & \text{if} \quad -\lambda \leqq x \leqq \lambda \\ \text{linear in} & [-2\lambda, -\lambda] \quad \text{and} \quad [\lambda, 2\lambda]. \end{cases}$$

Then obviously $f(1) = 0$, but

$$K_n(f, 1) = \sum_{|x_k| \leqq 2\lambda} f(x_k)\tau_k(1) > \sum_{|x_k| \leqq \lambda} \tau_k(1) =$$

$$= \sum_{|x_k| \leqq \lambda} u_k(1) l_k^4(1) \geqq \frac{\varepsilon}{3} n^2 \sum_{|x_k| \leqq \lambda} n^{4\alpha - 2} > \varepsilon c > 0, \quad n \geqq n_0,$$

i.e. $\{K_n(f, 1)\}$ does not tend to $f(1)$. (If $\alpha = \beta = -1/4$, see Laden [3, p. 602].)

d) If $\alpha=-1/4$ and $\beta=-3/4$, by $s_k(1)=0$ and $t_k \leqq 0$ if $x_k \leqq -1/2$, we get that all the terms in (3.4) are $\geqq 0$. I.e. $u_k(1) \geqq |t_k|^3(1-x_k)^3$, say (if $x_k \leqq -1/2$). If

$$f(x)=\begin{cases} 1 & \text{for} \quad -1 \leqq x \leqq -3/4, \\ 0 & \text{for} \quad -1/2 \leqq x \leqq 1, \\ \text{linear} & \text{for} \quad -3/4 \leqq x \leqq -1/2, \end{cases}$$

we have that $f(1)=0$. On the other hand,

$$K_n(f, 1) = \sum_{x_k \leqq -1/2} f(x_k)\tau_k(1) \geqq \sum_{x_k \leqq -3/4} f(x_k)\tau_k(1) =$$

$$= \sum_{x_k \leqq -3/4} \tau_k(1) > \sum_{x_k \leqq -3/4} |t_k|^3 l_k^4(1) \sim \sum_{i=1}^{cn} \frac{n^6}{k^6} \frac{n^{4\alpha} i^{4\beta+6}}{n^{4\beta+8}} \sim$$

$$\sim n^{4\alpha-4\beta-2} \sum_{i=1}^{n} i^{4\beta} = \sum_{i=1}^{n} i^{-3} > 1,$$

which completes the proof of our theorem.

REFERENCES

[1] Krylov, N. M. and Stayermann, E., Sur quelques formulas d'interpolation.... *Bull. Acad. de l'Oucraine* **1** (1923), 13—16.

[2] Szegő, G., *Orthogonal Polynomials,* AMS Coll. Publ., 23, AMS Providence, Rhode Island, 1975. *MR* **46** # 9631.

[3] Laden, H. N., An application of the classical orthogonal polynomials to the theory of interpolation, *Duke Math. J.* **8** (1941), 591—610. *MR* **3**—115.

[4] Knoop, H. B. and Stockenberg, B., On Hermite—Fejér type interpolation, *Bull. Austral. Math. Soc.* **28** (1983), 39—51.

[5] Natanson, G. I., Two-sided estimates for the Lebesgue function of Lagrange interpolation based on Jacobi nodes, *Izv. Vysš. Učebn. Zaved. Matematika* **11** (1967), 67—74 (Russian). *MR* **36** # 4210.

[6] Vértesi, P., On Lagrange interpolation, *Period. Math. Hungar.* **12** (1981), 103—112. *MR* **82e**: 41011.

(Received October 4, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
P.O. BOX 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 247—255

ON THE PROJECTION LATTICE OF GW^*-ALGEBRAS

JÁNOS KRISTÓF

In this paper a new type of C^*-algebras will be introduced. I name them GW^*-algebras (i.e generalized W^*-algebras) since their class contains the class of W^*-algebras and their properties are somewhat close to those of W^*-algebras.

Here the emphasis will be laid on the lattice theoretical aspects concerning GW^*-algebras. The mathematical motivation leading to the investigations below will be explained in Section I.

I. Preliminaries

If A is a *-algebra with unit $\mathbf{1}$ then the set $L(A)$ of projections (i.e. self-adjoint idempotent elements) of A can be equipped with a lattice theoretical structure determined by

— a partial ordering $\leqq$, defined by $h \leqq g$ iff $gh=h$ and

— an orthocomplementation $\perp$, i.e. a mapping $A \to A$, $h \mapsto h^\perp$ defined as $h^\perp := \mathbf{1}-h$.

Then $L(A)$, equipped with this structure is an orthocomplemented partially ordered set. In the sequel $L(A)$ will always be thought of as an orthocomplemented partially ordered set whose structure is that of described above.

It is known that $L(A)$ is not a lattice, in general. However, $L(A)$ is always finitely additive, i.e. if $(e_i)_{i\in I}$ is a finite system of pairwise orthogonal projections in A then the least upper bound $\bigvee\limits_{i\in I} e_i$ of the set $\{e_i | i\in I\}$ exists in $L(A)$ and equals $\sum\limits_{i\in I} e_i$. On the other hand, $L(A)$ is orthomodular in the sense that if $g, h\in L(A)$ and $g\leqq h$ then there is a unique element $e\in L(A)$ such that $e\perp g$ and $e\vee g=h$ (of course, $e=h-g$).

Before formulating the main problem leading to the examinations of the present study, we mention that given a bounded orthocomplemented lattice L, a mapping $p\colon L\to[0, 1]$ is called a *state* on L if p assigns 1 to the greatest element of L and it is finitely additive, i.e. for any finite system $(e_i)_{i\in I}$ of pairwise orthogonal elements of L we have $p(\bigvee\limits_{i\in I} e_i)=\sum\limits_{i\in I} p(e_i)$. If p is countably additive, i.e. the last requirement is satisfied for any countable orthogonal system $(e_i)_{i\in I}$, provided the least upper bound $\bigvee\limits_{i\in I} e_i$ exists in L, the state p is called σ-additive. If p is a state on L

1980 *Mathematics Subject Classification*. Primary 06B30, 46L10; Secondary 46L30.
Key words and phrases. Lattice, projection, GW^*-algebra.

and for every system $(e_i)_{i\in I}$ of pairwise orthogonal elements of L we have $p(\bigvee_{i\in I} e_i) = \sum_{i\in I} p(e_i)$, provided $\bigvee_{i\in I} e_i$ exists in L, then we say that the state p is completely additive. As for the basic necessary lattice theoretical notions see [1] and [2].

In general probability theory we are currently concerned with orthocomplemented lattices which are the abstract objects replacing the event spaces of classical probability theory (see [3], [4], [5]). In the general framework probability measures are treated as states on orthocomplemented lattices. In order to build up a meaningful and analytic general probability theory we need such σ-complete orthocomplemented lattices which have a great number of σ-additive states.

However, it is well-known that given a unital *-algebra A, the orthocomplemented partially ordered set $L(A)$ is neither a lattice, nor a σ-complete lattice, in general. So the question arises; which conditions on A provide that $L(A)$ be a σ-complete lattice such that $L(A)$ possesses a separating set of σ-additive states? As far as I know this problem is unsolved at present. A partial answer appeared in [6]; if A is a Rickart *-algebra then $L(A)$ is a lattice, moreover, if A is a Banach *-algebra whose underlying *-algebra is a Rickart *-algebra, then $L(A)$ is a σ-complete orthomodular lattice. However, in the latter case even the existence of a single σ-additive state seems to be questionable.

In this paper a sufficient condition will appear, assuring not only that $L(A)$ be a σ-complete orthomodular lattice but the existence of a separating set of σ-additive states on $L(A)$ as well.

II. GW^*-algebras

The vector space of the linear forms on the *-algebra A will be denoted by A^* and the weak $\sigma(A^*, A)$ topology relates to the canonical duality between A^* and A.

If A is a unital *-algebra (whose unit is always denoted by $\mathbf{1}$) and P is a set of positive linear forms on A then the set $\{f\in P \mid f(\mathbf{1})\leqq 1\}$ will be denoted by $P(1)$. Further, assuming that $P(1)$ is non void and $\sigma(A^*, A)$-bounded, $\|\cdot\|_P$ denotes the mapping from A into $\mathbf{R}_+$ defined by

$$\|x\|_P := \sup_{f\in P(1)} \sqrt{f(x^*x)}$$

for all $x\in A$. It is obvious that $\|\cdot\|_P$ is a seminorm on A; the dual seminorm is denoted by $\|\cdot\|_P'$.

If S is a subset of A^* then the linear subspace of A^* spanned by S and the convex hull of S is denoted by $\mathrm{sp}(S)$ and $\mathrm{co}(S)$, respectively, while the $\sigma(A^*, A)$-closed linear subspace of A^* spanned by S and the $\sigma(A^*, A)$-closed convex hull of S is denoted by $\widetilde{\mathrm{sp}}(S)$ and $\widetilde{\mathrm{co}}(S)$, respectively. If the elements of S are $\|\cdot\|_P$-continuous forms (where P is a set of positive linear forms on A such that $P(1)$ is non void and $\sigma(A^*, A)$-bounded) then the $\|\cdot\|_P'$-closed linear subspace of A^* spanned by S and the $\|\cdot\|_P'$-closed convex hull of S is denoted by $\overline{\mathrm{sp}}(S)$ and $\overline{\mathrm{co}}(S)$, respectively, provided no confusion arises as for P.

If f is a linear form on the *-algebra A then for every $x\in A$ we define the linear forms $x\cdot f$ and $f\cdot x$ on A as the mappings $y\mapsto f(xy)$ and $y\mapsto f(yx)$, respectively. If $f\in A^*$ and $x, y\in A$ then $x\cdot f\cdot y$ stands for $(x\cdot f)\cdot y$.

DEFINITION. The pair (A, P) is called a *weak GW^*-algebra* if A is a unital $*$-algebra and P is a separating set of positive linear forms on A satisfying

(I) $P(1)$ is non void and $\sigma(A^*, A)$-bounded.
(II)$_w$ $\mathbf{R}_+P\subset P$ and $x^*\cdot P\cdot x\subset\widetilde{\mathrm{co}}(P)$ for all $x\in A$.
(III) $x\cdot P\subset\overline{\mathrm{sp}}(P)$ for all $x\in A$.
(IV) A is sequentially complete with respect to the uniform structure defined by the $\sigma(A, \mathrm{sp}(P))$-topology.

The pair (A, P) is called a *GW^*-algebra* if it is a weak GW^*-algebra and instead of (II$_w$) satisfies the more restrictive condition

(II) $\mathbf{R}_+P\subset P$ and $x^*\cdot P\cdot x\subset\overline{\overline{\mathrm{co}}}(P)$ for all $x\in A$.

The GW^*-algebra (A, P) is called *complete* if it satisfies

(IV)$_s$ A is quasi complete with respect to the uniform structure defined by the $\sigma(A, \mathrm{sp}(P))$-topology.

If (A, P) is a weak GW^*-algebra then A is a C^*-algebra whose C^*-norm equals $\|\cdot\|_P$. This is shown in [7], however, here a new proof will be presented (see Proposition 1).

If (A, P) is a weak GW^*-algebra then

— the $\sigma(A, \mathrm{sp}(P))$ and $\sigma(A, \overline{\mathrm{sp}}(P))$ topologies coincide in every C^*-norm bounded subset of A,

— the multiplication in A is left and right continuous on C^*-norm bounded sets with respect to the $\sigma(A, \mathrm{sp}(P))$ topology,

— the involution of A is continuous in the $\sigma(A, \mathrm{sp}(P))$ topology.

This statement is verified in [7] Lemma 1 and Lemma 2.

There are two important examples for GW^*-algebras.

EXAMPLE 1. Let A be a von Neumann algebra in the Hilbert space H and for all $z\in H$ define $f_z\colon A\to\mathbf{C}$ as $T\mapsto(Tz|z)$ for every $T\in A$. If $P:=\{f_z|z\in H\}$ then (A, P) is a complete GW^*-algebra. Here the $\sigma(A, \mathrm{sp}(P))$ and $\sigma(A, \overline{\mathrm{sp}}(P))$ topologies coincide with the weak and ultraweak operator topologies in A, respectively. (See [7].)

EXAMPLE 2. Let B be a σ-algebra of subsets of the set X and let A denote the $*$-algebra of complex bounded Borel functions on X. Let P be the set of integrals arising from positive σ-additive finite measures defined on B. Then (A, P) is a commutative GW^*-algebra. Here the $\sigma(A, \mathrm{sp}(P))$ and $\sigma(A, \overline{\mathrm{sp}}(P))$ topologies are equal. (See [7].)

The first example shows that every von Neumann algebra can be regarded as a complete GW^*-algebra while the second example yields a great number of commutative GW^*-algebras that are not $*$-isomorphic to any von Neumann algebra.
The following lemma makes possible to give a new proof of the completeness of a weak GW^*-algebra (A, P) with respect to the uniform structure defined by the norm $\|\cdot\|_P$.

LEMMA 1. *Let (A, P) be a weak GW^*-algebra, r a positive real number and*

$(x_i)_{i\in I}$ a generalized sequence in A such that $\|x_i\|_P\leqq r$ for all $i\in I$. If $x\in A$ and $x_i\to x$ in the $\sigma(A,\operatorname{sp}(P))$ topology then $\|x\|_P\leqq 2r$.

PROOF. Obviously, it is enough to prove for the special case $r=1$. Let f be a fixed element of $P(1)$ and $i\in I$. Then we have $f(x_i^*x_i)\leqq\|x_i\|_P^2\leqq 1$ and

$$|f((x-x_i)^*x_i)|\leqq\sqrt{f((x-x_i)^*(x-x_i))}\sqrt{f(x_i^*x_i)}\leqq\|x-x_i\|_P\|x_i\|_P,$$

i.e. $|f((x-x_i)^*x_i)|\leqq\|x\|_P+1$. From this we obtain

$$f(x^*x)=f(x^*(x-x_i))+f((x-x_i)^*x_i)+f(x_i^*x_i)\leqq|(x^*\cdot f)(x-x_i)|+\|x\|_P+2.$$

Since $x^*\cdot f\in\overline{\operatorname{sp}}(P)$, by (III), and $(x_i)_{i\in I}$ is a $\|\cdot\|_P$-bounded sequence in A, we conclude that $(x^*\cdot f)(x_i)\to(x^*\cdot f)(x)$ ([7] Lemma 1) thus $f(x^*x)\leqq\|x\|_P+2$. This means that $\|x\|_P\leqq\sqrt{\|x\|_P+2}$, i.e. $(\|x\|_P-2)(\|x\|_P+1)\leqq 0$, thus $\|x\|_P\leqq 2$.

PROPOSITION 1. *If (A, P) is a weak GW^*-algebra then A is a C^*-algebra whose C^*-norm equals $\|\cdot\|_P$.*

PROOF. The proof of the nontrivial fact that $\|\cdot\|_P$ is a C^*-seminorm on A can be taken from [7] Theorem 1. We have only to show the completeness of A with respect to the uniform structure defined by the norm $\|\cdot\|_P$. Let B_r denote the $\|\cdot\|_P$-closed ball in A with radius r, for $r>0$. Further, let V_r be the closure of B_r in the $\sigma(A,\operatorname{sp}(P))$ topology, for $r>0$. Then Lemma 1 implies that $B_{r/2}\subset V_{r/2}\subset B_r$ $(r>0)$ thus $(V_r)_{r>0}$ is a base at 0 for the $\|\cdot\|_P$-topology consisting of $\sigma(A,\operatorname{sp}(P))$-closed sets. Since the $\|\cdot\|_P$ topology is finer than $\sigma(A,\operatorname{sp}(P))$, (IV) results in the sequentially completeness of A with respect to the uniform structure defined by $\|\cdot\|_P$ (see [8] Ch. I, § 1, n° 5, Prop. 8).

III. The projection lattice of a weak GW^*-algebra

In this section we prove that the orthocomplemented partially ordered set of projections of a weak GW^*-algebra is a σ-complete orthomodular lattice admitting a separating set of σ-additive states. Moreover, we show that the projection lattice of a complete GW^*-algebra is a complete orthomodular lattice admitting a separating set of completely additive states.

THEOREM 1. *Assume that (A, P) is a weak GW^*-algebra.*

(i) *If $(e_k)_{k\in\mathbf{N}}$ is a sequence of pairwise orthogonal projections in A then the series $\sum_{k\in\mathbf{N}}e_k$ converges in the $\sigma(A,\operatorname{sp}(P))$ topology and its sum (denoted by $(\sigma)\sum_{k\in\mathbf{N}}e_k$) equals the least upper bound $\bigvee_{k\in\mathbf{N}}e_k$ of the set $\{e_k|k\in\mathbf{N}\}$ in $L(A)$.*

(ii) *If $(e_k)_{k\in\mathbf{N}}$ is a sequence of pairwise commutating projections in A then the product $\prod_{k\in\mathbf{N}}e_k$ converges in the $\sigma(A,\operatorname{sp}(P))$ topology and its product (denoted by $(\sigma)\prod_{k\in\mathbf{N}}e_k$) equals the greatest lower bound $\bigwedge_{k\in\mathbf{N}}e_k$ of the set $\{e_k|k\in\mathbf{N}\}$ in $L(A)$.*

(iii) *If $g,h\in L(A)$ then the sequences*

$$((gh)^k)_{k\in\mathbf{N}},\quad ((hg)^k)_{k\in\mathbf{N}},\quad ((hgh)^k)_{k\in\mathbf{N}},\quad ((ghg)^k)_{k\in\mathbf{N}}$$

converge in the $\sigma(A, \operatorname{sp}(P))$ *topology and their common limit equals* $g \wedge h$ *in the partially ordered set* $L(A)$.

(iv) $L(A)$ *is a* σ*-complete orthomodular lattice and the set* $\mathbf{P}_\sigma := \{f|_{L(A)} | f \in \widetilde{\overline{\mathrm{co}}}(P), f(1)=1\}$ *is a separating set of* σ*-additive states on* $L(A)$ *with the property that to every* $e \in L(A)$, $e \neq 0$ *there is a state* p *in* $\mathbf{P}_\sigma$ *such that* $p(e)=1$.

PROOF. (i) Let $(e_k)_{k\in\mathbf{N}}$ be a sequence of pairwise orthogonal projections in A and put $s_n := \sum_{k=0}^{n} e_k$ for all $n \in \mathbf{N}$. Then we have $0 \leqq s_n \leqq s_{n+1} \leqq 1$ in the *-algebra A thus $(f(s_n))_{n\in\mathbf{N}}$ is an increasing bounded sequence of positive real numbers for every $f \in P$. By virtue of (IV) we conclude the existence of an element $e \in A$ such that $s_n \to e$ in the $\sigma(A, \operatorname{sp}(P))$ topology showing that the series $\sum_{k\in\mathbf{N}} e_k$ is convergent in the $\sigma(A, \operatorname{sp}(P))$ topology (and, by the definition, its sum equals e). Since the involution of A is $\sigma(A, \operatorname{sp}(P))$-continuous and $s_n^* = s_n$ $(n \in \mathbf{N})$, the element e is self-adjoint. Further, for $n, m \in \mathbf{N}$, $m \geqq n$ we have $e_n = s_m e_n$, thus fixed a number $n \in \mathbf{N}$ we obtain $s_m e_n \to e e_n$ in the $\sigma(A, \operatorname{sp}(P))$ topology, since $\|s_m\|_P \leqq 1$ $(m \in \mathbf{N})$. This means that $e_n = e e_n$ $(n \in \mathbf{N})$, consequently, $s_m = e s_m$ for all $m \in \mathbf{N}$. Applying again the fact that the sequence $(s_m)_{m\in\mathbf{N}}$ is $\|\cdot\|_P$-bounded and $s_m \to e$ in the $\sigma(A, \operatorname{sp}(P))$ topology, we derive that e is idempotent in A, i.e. $e \in L(A)$. On the other hand, we already know that $e_n = e e_n$ $(n \in \mathbf{N})$, i.e. $e_n \leqq e$ for all $n \in \mathbf{N}$ in $L(A)$. If $g \in L(A)$ is an upper bound in $L(A)$ of the set $\{e_k | k \in \mathbf{N}\}$ then $e_k = g e_k$ $(k \in \mathbf{N})$ thus $s_m = g s_m$ for all $m \in \mathbf{N}$, consequently, $g s_m \to g e$ in the $\sigma(A, \operatorname{sp}(P))$ topology now gives that $e = ge$, i.e. $e \leqq g$ in $L(A)$, proving that e equals the least upper bound of the set $\{e_k | k \in \mathbf{N}\}$ in the orthocomplemented partially ordered set $L(A)$.

(ii) Let $(e_k)_{k\in\mathbf{N}}$ be a sequence of pairwise commutating projections in A and put $p_n := \prod_{k=0}^{n} e_k$ for all $n \in \mathbf{N}$. Then we have $0 \leqq p_{n+1} \leqq p_n \leqq 1$ in the *-algebra A thus $(f(p_n))_{n\in\mathbf{N}}$ is a decreasing sequence of positive real numbers for every $f \in P$. Then (IV) implies the existence of an element $e \in A$ with the property that $p_n \to e$ in the $\sigma(A, \operatorname{sp}(P))$ topology, proving that the product $\prod_{k\in\mathbf{N}} e_k$ is convergent in the $\sigma(A, \operatorname{sp}(P))$ topology (and, by the definition, its product equals e). Since the involution of A is $\sigma(A, \operatorname{sp}(P))$-continuous and $p_n^* = p_n$ $(n \in \mathbf{N})$, we obtain that $e^* = e$. For $n, m \in \mathbf{N}$, $m \geqq n$ we obviously have $p_m = e_n p_m$, thus given a number $n \in \mathbf{N}$ we deduce that $e = e_n e$, since $\|p_m\|_P \leqq 1$ and $p_m \to e$ in the $\sigma(A, \operatorname{sp}(P))$ topology. Consequently, $e = p_m e$ for all $m \in \mathbf{N}$, thus applying again the $\|\cdot\|_P$-boundedness of the sequence $(p_m)_{m\in\mathbf{N}}$ we obtain $e = e^2$, i.e. $e \in L(A)$. On the other hand, we already know that $e = e_n e$ $(n \in \mathbf{N})$, i.e. $e \leqq e_n$ in $L(A)$ for every $n \in \mathbf{N}$. If $g \in L(A)$ is a lower bound in $L(A)$ of the set $\{e_k | k \in \mathbf{N}\}$ then $g = e_n g$ $(n \in \mathbf{N})$, thus $g = p_m g$ for all $m \in \mathbf{N}$. Now $p_m g \to eg$ in the $\sigma(A, \operatorname{sp}(P))$ topology yields $g = eg$, i.e. $g \leqq e$ in $L(A)$, proving that e equals the greatest lower bound of the set $\{e_k | k \in \mathbf{N}\}$ in $L(A)$.

(iii) Let g and h be fixed projections in A and put

$$a_k := (gh)^k, \quad b_k := (hg)^k, \quad c_k := (hgh)^k, \quad d_k := (ghg)^k$$

for every $k \in \mathbf{N}$. Of course, all these sequences are $\|\cdot\|_P$-bounded. It is easy to see that

$$(1)\qquad a_{k+1} = d_k h, \quad b_{k+1} = hd_k, \quad c_{k+1} = hd_k h$$

for all $k \in \mathbf{N}$, thus if we prove that the sequence $(d_k)_{k\in\mathbf{N}}$ converges in the $\sigma(A, \mathrm{sp}(P))$ topology then the equalities in (1) combined with Lemma 2 in [7] provide that the other three sequences also converge in the same topology. In order to prove that $(d_k)_{k\in\mathbf{N}}$ converges in the $\sigma(A, \mathrm{sp}(P))$ topology, first we mention that $d_{2j}=d_j^* d_j$ and $d_{2j+1}=b_{j+1}^* b_{j+1}$ $(j\in\mathbf{N})$ thus d_k is a positive element in the *-algebra A for all $k\in\mathbf{N}$. On the other hand, we have

$$d_{2j}-d_{2j+1} = \big((1-h)\,g(ghg)^j\big)^*\big((1-h)\,g(ghg)^j\big)$$

for $j\in\mathbf{N}$ and

$$d_{2j-1}-d_{2j} = \big((1-g)\,h(ghg)^{j-1}\big)^*\big((1-g)\,h(ghg)^{j-1}\big)$$

for $j\in\mathbf{N}$, $j\neq 0$. This means that $0\leqq d_{k+1}\leqq d_k$ $(k\in\mathbf{N})$ in the *-algebra A, thus $(f(d_k))_{k\in\mathbf{N}}$ in a decreasing sequence of positive real numbers for every $f\in P$. Then (IV) yields the existence of an element d in A such that $d_k\to d$ in the $\sigma(A, \mathrm{sp}(P))$ topology. Hence we may put $a:=\lim\limits_k a_k$, $b:=\lim\limits_k b_k$, $c:=\lim\limits_k c_k$ and $d:=\lim\limits_k d_k$, where the limit operations relate to the $\sigma(A, \mathrm{sp}(P))$ topology. With regard to (1) we have

$$(2)\qquad a = dh, \quad b = hd, \quad c = hdh.$$

It remained to show that $a=b=c=d$ and this element is just the greatest lower bound in $L(A)$ of the set $\{g, h\}$. Since $d_k^*=d_k$ and $c_k^*=c_k$ $(k\in\mathbf{N})$ and the involution of A is $\sigma(A, \mathrm{sp}(P))$-continuous, the elements d and c are self-adjoint. If $k, n\in\mathbf{N}$ then $d_k d_n=d_{k+n}$ and $c_k c_n=c_{k+n}$, thus applying twice the $\|\cdot\|_P$-boundedness of the sequences $(d_k)_{k\in\mathbf{N}}$ and $(c_k)_{k\in\mathbf{N}}$, respectively, and the definitions of d and c, we obtain that $d^2=d$ and $c^2=c$, i.e. $d, c\in L(A)$. If $k, m, n\in\mathbf{N}$, then $d_k c_m d_n=d_{k+m+n+1}$ and $c_k d_m c_n=c_{k+m+n+1}$, hence applying successively three times the $\|\cdot\|_P$-boundedness of the sequences $(d_k)_{k\in\mathbf{N}}$ and $(c_k)_{k\in\mathbf{N}}$, respectively, and the definitions of d and c, we deduce the equalities $dcd=d$ and $cdc=c$. These equalities result in the following ones: $(d-cd)^*(d-cd)=0$ and $(c-dc)^*(c-dc)=0$. Since the involution of A is proper, it follows that $d=cd$ and $c=dc$. Taking the adjoint of the second equality we arrive at $d=c$. Finally, we prove that $d=g\wedge h$ in $L(A)$. Since $d_k=gd_k$ and $c_k=hc_k$ $(k\in\mathbf{N},\ k\neq 0)$, taking the limits in the $\sigma(A, \mathrm{sp}(P))$ topology we obtain $d=gd$ and $d=c=hc=hd$, i.e. $d\leqq g$ and $d\leqq h$ in $L(A)$. If $e\in L(A)$ and $g\geqq e$, $h\geqq e$ in $L(A)$ then we have $e=eg=he=eh$, hence $e=d_k e$ $(k\in\mathbf{N})$. Taking the $\sigma(A, \mathrm{sp}(P))$-limit of the last equality; $e=de$, i.e. $d\geqq e$ in $L(A)$. This shows that $d=g\wedge h$ in $L(A)$. Finally, by (2), we have $a=dh=d$ and $b=hd=d$, thus the proof is completed.

(iv) With regard to (i) and (iii) and the orthomodularity of the orthocomplemented partially ordered set $L(A)$, we deduce that $L(A)$ is a σ-complete orthomodular lattice (see [2]). The elements of the set $\mathbf{P}_\sigma:=\{f|_{L(A)}\,|\,f\in\widetilde{\overline{\mathrm{co}}}(P),\ f(1)=1\}$ are states on the lattice $L(A)$, evidently. If $(e_k)_{k\in\mathbf{N}}$ is a sequence of pairwise orthogonal projections in A then, by (i), we have $\bigvee\limits_{k\in\mathbf{N}} e_k=(\sigma)\sum\limits_{k\in\mathbf{N}} e_k$. If $f\in\widetilde{\overline{\mathrm{co}}}(P)$ then f is a $\|\cdot\|_P$-

continuous linear form on A (see [7], Theorem 1), thus $f(\bigvee_{k\in\mathbf{N}} e_k)=\sum_{k\in\mathbf{N}} f(e_k)$, proving that $f|_{L(A)}$ is a σ-additive state on $L(A)$ for all $f\in\overline{\mathrm{co}}(P)$, $f(\mathbf{1})=1$. If $g, h\in L(A)$ and $g\neq h$ then there is an element f in P such that $f(g)\neq f(h)$. Of course, $f\neq 0$, i.e. $f(\mathbf{1})\neq 0$ and the restriction of the linear form $f':=f/f(\mathbf{1})$ to $L(A)$ belongs to $\mathbf{P}_\sigma$ and $f'(g)\neq f'(h)$. This means that even the set

$$\mathbf{P}:=\{f|_{L(A)}|f\in P, f(\mathbf{1})=1\}$$

separates the points of $L(A)$. If $e\in L(A)$, $e\neq 0$ then there is an element f in P with the property $f(e)\neq 0$. Then the restriction to $L(A)$ of the linear form $f':=$ $=(e^*\cdot f\cdot e)/f(e)$ is in $\mathbf{P}_\sigma$ and the state $p:=f'|_{L(A)}$ satisfies the equality $p(e)=1$.

THEOREM 2. *Assume that* (A, P) *is a complete* GW^**-algebra.*

(i) *Let* I *be a right directed preordered set and* $(e_i)_{i\in I}$ *a family of projections in* A *with the property that* $i, i'\in I$, $i\leqq i'$ *implies* $e_i\leqq e_{i'}$ *in the lattice* $L(A)$. *Then* $\lim_i e_i$ *exists in the* $\sigma(A, \mathrm{sp}(P))$ *topology and equals the least upper bound* $\bigvee_{i\in I} e_i$ *of the set* $\{e_i|i\in I\}$ *in* $L(A)$.

(ii) *If* $(e_i)_{i\in N}$ *is an arbitrary system of pairwise orthogonal projections in* A *then the generalized series* $\sum_{i\in I} e_i$ *converges in the* $\sigma(A, \mathrm{sp}(P))$ *topology and its sum (denoted by* $(\sigma)\sum_{i\in I} e_i$*) equals the least upper bound in* $L(A)$ *of the set* $\{e_i|i\in I\}$.

(iii) *The projection lattice of* A *is a complete orthomodular lattice and the set* $\mathbf{P}_c:=\{f|_{L(A)}|f\in\overline{\mathrm{co}}(P),\ f(\mathbf{1})=1\}$ *is a separating set of completely additive states on* $L(A)$ *with the property that to all* $e\in L(A)$, $e\neq 0$ *there is a state* p *in* $\mathbf{P}_c$ *such that* $p(e)=1$.

PROOF. (i) If $f\in P$ then $(f(e_i))_{i\in I}$ is a right directed family in the compact interval $[0, f(\mathbf{1})]$ thus $\lim_i f(e_i)$ exists in $\mathbf{R}$. Consequently, $(e_i)_{i\in I}$ is a generalized Cauchy sequence with respect to the uniform structure defined by the $\sigma(A, \mathrm{sp}(P))$ topology. Since the set $\{e_i|i\in I\}$ is $\sigma(A, \mathrm{sp}(P))$-bounded, the quasi completeness of A in the uniform structure defined by the $\sigma(A, \mathrm{sp}(P))$ topology now gives that $e:=\lim_i e_i$ exists in the $\sigma(A, \mathrm{sp}(P))$ topology. The $\sigma(A, \mathrm{sp}(P))$-continuity of the involution of A implies that e is self-adjoint. If $i_0\in I$ is fixed then $e_{i_0}=e_i e_{i_0}$ for $i\in I$, $i_0\leqq i$, thus applying the $\|\cdot\|_P$-boundedly left continuity of the multiplication of A in the $\sigma(A, \mathrm{sp}(P))$ topology, we obtain $e_{i_0}=ee_{i_0}$ for all $i_0\in I$. Repeating the former reasoning we find that $e=e^2$, i.e. $e\in L(A)$. On the other hand, $e_i\leqq e$ for all $i\in I$. Suppose that $g\in L(A)$ and $e_i\leqq g$ $(i\in I)$. Then $e_i=ge_i$ $(i\in I)$ and taking the limit in the $\sigma(A, \mathrm{sp}(P))$ topology we deduce $e=ge$, i.e. $e\leqq g$ showing that $e=\bigvee_{i\in I} e_i$ in the lattice $L(A)$.

(ii) It is an immediate consequence of (i).

(iii) The completeness of the orthomodular lattice $L(A)$ follows from (i). Let $(e_i)_{i\in I}$ be a family of pairwise orthogonal projections in A. According to (ii), the series $\sum_{i\in I} e_i$ is summable in the $\sigma(A, \mathrm{sp}(P))$ topology and its sum equals $\bigvee_{i\in I} e_i$. The system of finite sums of the family $(e_i)_{i\in I}$ consists of projections in the C^*-algebra A, thus it is C^*-norm bounded. Since the $\sigma(A, \mathrm{sp}(P))$ and $\sigma(A, \overline{\mathrm{sp}}(P))$ topologies coincide in every C^*-norm bounded subset of A, given a linear form $f\in\overline{\mathrm{co}}(P)$, $f(\mathbf{1})=1$, we obtain $f(\bigvee_{i\in I} e_i)=f((\sigma)\sum_{i\in I} e_i)=\sum_{i\in I} f(e_i)$, i.e. $f|_{L(A)}$ is a completely additive state on $L(A)$.

We have seen in the proof of (iv) of Theorem 1 that the set $\mathbf{P}$ separates the points of $L(A)$ and, of course, $\mathbf{P}\subset\mathbf{P}_c$. Since (A, P) is a GW^*-algebra (i.e. it satisfies (II) instead of (II_w)), the linear form $f':=(e^*\cdot f\cdot e)/f(e)$ introduced in the proof of the last assertion in Theorem 1 now belongs to $\overline{\mathrm{co}}(P)$, thus our present assertion is proved.

IV. On the countable summability of partial isometries in GW^*-algebras

Here we remind of certain elementary notions from the theory of *-algebras; the details can be found e.g. in [6].

Let A be a unital *-algebra. The element w of A is called a *partial isometry* if both w^*w and ww^* are projections in A. If the involution of A is proper then $w\in A$ is a partial isometry if and only if $w^*ww^*=w^*$, or equivalently, $ww^*w=w$. A family $(w_i)_{i\in I}$ of partial isometries in A is called *orthogonal* if for all $i,j\in I$, $i\neq j$ we have $w_i^*w_iw_j^*w_j=0=w_iw_i^*w_jw_j^*$. We say that *the partial isometries are (countably) summable* in A if for every (countable) orthogonal family $(w_i)_{i\in I}$ of partial isometries in A there is a partial isometry w in A with the property that $w^*w=\bigvee_{i\in I} w_i^*w_i$ and $ww^*=\bigvee_{i\in I} w_iw_i^*$ in the partially ordered set $L(A)$ and $ww_i^*w_i=w_i=w_iw_i^*w$ $(i\in I)$.

If $g, h\in L(A)$ then the projections g and h are called *equivalent* if there is a partial isometry w in A such that $g=w^*w$ and $h=ww^*$. We say that *the equivalence of projections is (countably) additive,* if given two (countable) families $(g_i)_{i\in I}$ and $(h_i)_{i\in I}$ of pairwise orthogonal projections in A, the equivalence of g_i and h_i for all $i\in I$ implies that the projections $\bigvee_{i\in I} g_i$ and $\bigvee_{i\in I} h_i$ are equivalent, provided the least upper bounds exist in the partially ordered set $L(A)$.

It can be shown without difficulty that the (countable) summability of the partial isometries implies the (countable) additivity of the equivalence of projections. However, it is not known, whether the equivalence of projections is countably additive in a C^*-algebra whose underlying *-algebra is a Rickart *-algebra (see [6]).

For the case of weak GW^*-algebras we have an affirmative answer as follows.

Proposition 2. *If (A, P) is a weak GW^*-algebra then the partial isometries are countably summable in A and, consequently, the equivalence of projections is countably additive.*

Proof. If w is an arbitrary partial isometry in A and f a positive linear form on A then

$$|f(w)|^2=|f(w^*)|^2=|f(w^*(ww^*))|^2\leqq f((ww^*)^*ww^*)f(w^*w)= \tag{3}$$
$$=f(ww^*)f(w^*w).$$

Let there be given an orthogonal sequence $(w_k)_{k\in\mathbf{N}}$ of partial isometries in A. Combining Theorem 1 with the inequality (3) we easily deduce

$$\sum_{k\in\mathbf{N}} |f(w_k)| \leq \sum_{k\in\mathbf{N}} f(w_k w_k^*)^{1/2} f(w_k^* w_k)^{1/2} \leq$$

$$\leq \Big(\sum_{k\in\mathbf{N}} f(w_k w_k^*)\Big)^{1/2} \Big(\sum_{k\in\mathbf{N}} f(w_k^* w_k)\Big)^{1/2} = f\Big(\bigvee_{k\in\mathbf{N}} w_k w_k^*\Big)^{1/2} f\Big(\bigvee_{k\in\mathbf{N}} w_k^* w_k\Big)^{1/2}$$

for all $f\in P$. From this it follows, by (IV), that the series $\sum_{k\in\mathbf{N}} w_k$ converges in the $\sigma(A, \mathrm{sp}(P))$ topology; let w denote its sum.

Since $(w_k)_{k\in\mathbf{N}}$ is an orthogonal sequence of partial isometries in A, we have $w_k^* w_k w_n^* w_n = 0$ $(k, n\in\mathbf{N},\ k\neq n)$ thus it follows that $0 = w_k(w_k^* w_k w_n^* w_n) w_n^* = (w_k w_k^* w_k)(w_n^* w_n w_n^*) = w_k w_n^*$, and similarly, $w_k^* w_n = 0$ $(k, n\in\mathbf{N},\ k\neq n)$. Fixed a number $n\in\mathbf{N}$, we obtain $(\sigma)\sum_{k\in\mathbf{N}} w_k w_n^* = w_n w_n^*$ and $(\sigma)\sum_{k\in\mathbf{N}} w_k^* w_n = w_n^* w_n$ for all $n\in\mathbf{N}$. On the other hand, in a pre-C^*-algebra the norm of a partial isometry is less than or equal to 1, hence the sequences $(w_k)_{k\in\mathbf{N}}$ and $(w_k^*)_{k\in\mathbf{N}}$ are $\|\cdot\|_P$-bounded. Then the left $\|\cdot\|_P$-boundedly $\sigma(A, \mathrm{sp}(P))$-continuity of the multiplication of A combined with the results in Theorem 1 yield $(\sigma)\sum_{k\in\mathbf{N}} w_k w_n^* = w w_n^*$ and $(\sigma)\sum_{k\in\mathbf{N}} w_k^* w_n = w^* w_n$ $(n\in\mathbf{N})$, thus $w w_n^* = w_n w_n^*$ and $w^* w_n = w_n^* w_n$ $(n\in\mathbf{N})$, i.e. $w w_n^* w_n = w_n w_n^* w_n = w_n w_n^* w = w_n$ $(n\in\mathbf{N})$ and also $\bigvee_{n\in\mathbf{N}} w_n w_n^* = (\sigma)\sum_{n\in\mathbf{N}} w_n w_n^* = w w^*$ and $\bigvee_{n\in\mathbf{N}} w_n^* w_n = (\sigma)\sum_{n\in\mathbf{N}} w_n^* w_n = w^* w$, showing that w is the partial isometry in A whose existence is claimed.

REFERENCES

[1] *Trends in Lattice Theory*, ed. J. C. Abbot, van Nostrand, New York, 1970. *MR* **42**#5863.

[2] Maeda, F. and Maeda, S., *The Theory of Symmetric Lattices*, Springer-Verlag, New York—Berlin, 1970. *MR* **44**#123.

[3] Varadarajan, V. S., *Geometry of Quantum Theory*, vol. II., van Nostrand, New York, 1970. *MR* **57**#11400.

[4] Matolcsi, T., *A Concept of Mathematical Physics, vol. II.*, Akadémiai Kiadó, Budapest (to appear).

[5] Mackey, G. W., *The Mathematical Foundations of Quantum Mechanics*, W. A. Benjamin, Inc., New York, 1963. *MR* **27**#5501.

[6] Berberian, S. K., *Baer *-rings*, Springer-Verlag, Berlin, 1972. *MR* **55**#2983.

[7] Kristóf, J., C^*-norms defined by positive linear forms, *Acta Sci. Math. (Szeged)* (to appear).

[8] Bourbaki, N., Éléments de mathématique, XV et XVIII. Première partie: Les structures fondamentales de l'analyse. Livre V. Espaces vectoriels topologiques, Chap. I—II. et III—V. Actualités Sci. Ind., No. 1189 et 1229, Hermann et Cie, Paris 1953 et 1955. *MR* **11**—880 et **17**—1109.

[9] Bratteli, O. and Robinson, D. W., *Operator Algebras and Quantum Statistical Mechanics I.*, Springer-Verlag, New York—Heidelberg, 1979. *MR* **81a**:46070.

(Received November 12, 1984)

ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 257—263

NON-ADAPTIVE HYPERGEOMETRIC GROUP TESTING

F. K. HWANG and V. T. SÓS

Abstract

In a hypergeometric group testing problem we have a set of n items known to contain exactly d defectives. The problem is to identify all the defectives through group testing with a minimal number of tests where a test consists of a specified subset of the items and has the outcome *pure* if none of the items in the subset is defective and the outcome contaminated otherwise. A testing procedure is called *non-adaptive* if all tests have to be specified simultaneously. We translate this group testing problem into an extremal problem for set-systems and give estimates for the size of the extremal systems.

1. Introduction

In a *hypergeometric group testing* (HGT) problem we have a set of n items known to contain d defectives and $n-d$ good items. Any subset of the n items can be pooled for a test with two possible outcomes: the subset is *pure* if it contains no defectives and the subset is *contaminated* otherwise. The objective is to identify all the defectives using a minimal number of tests. In this paper "minimal" is defined as to minimize the maximum number of tests required (the worst-case number of tests).

A group testing algorithm is called *sequential* if the tests are given sequentially, that is, the decision of which subset to test currently may depend on the outcomes of tests already performed. A group testing algorithm is called *non-adaptive* if all tests have to be specified simultaneously, thus banning any possibility of using feedback information from tests. Since any non-adaptive algorithm can also be used sequentially, it is clear that the sequential algorithms should be expected to perform better than non-adaptive algorithms in general. The line between sequential algorithms and non-adaptive algorithms has not been clearly drawn historically; hence the group testing literature consists almost exclusively of sequential algorithms owing to their better performance as far as the number of tests is concerned. However, with parallel processing a possibility in many potential applications of group testing, one should no longer ignore the potential advantage of time saving in non-adaptive group testing. The purpose of this paper is to call attention to this fact and to provide some exploratory results in this direction.

1980 *Mathematics Subject Classification.* Primary 62K99; Secondary 05A05.
Key words and phrases. Group testing, extremal set-systems, d-complete design.

2. The group-testing problem and some extremal problems for set-systems

Suppose S is the set of n items which contains d defective elements. To design a nonadaptive HGT for finding the defective items we need a system $\mathcal{T}=\{T_1, \ldots, T_t\}$ of tests which satisfies the following:

for any $I=\{i|T_i$ is contaminated$\}$ there is exactly one d-tuple $D=\{u_1, \ldots, u_d\}\in S$ so that

$$D\cap T_i\neq\emptyset \quad \text{for} \quad i\in I$$

and

$$D\cap T_i=\emptyset \quad \text{for} \quad i\notin I.$$

For given n and d let $\varphi(n, d)$ be the minimum number of t so that there is a system $\mathcal{T}=\{T_1, \ldots, T_t\}$ which satisfies this condition. Our main result yields

$$c_1(d)\log n < \varphi(n; d) < c_2(d)\log n \tag{1}$$

where $c_1, c_2>0$ depend only on d. The lower bound follows from the simple Proposition 1, the upper bound follows from Theorem 3.

First we reformulate the problem in a dual form. This yields to different extremal problems for set-systems.

Now let $S=\{u_1, \ldots, u_n\}$ be a set of n elements, $\mathcal{T}=\{T_1, \ldots, T_t\}$ be a family of subsets of S. Define the *dual family*

$$\mathcal{C} = \{C_1, \ldots, C_n\}$$

by

$$C_i = \{j\,|u_i\in T_j\}, \qquad i = 1, \ldots, n.$$

Observe that for given defective elements $u_{i_1}, \ldots, u_{i_d}$ a test T_j yields "contaminated" iff $j\in \bigcup_{\nu=1}^{d} C_{i_\nu}$. So

$$I = \{j\,|T_j \text{ is "contaminated"}\} = \bigcup_{\nu=1}^{d} C_{i_\nu}.$$

This leads to a reformulation of the problem.

Definition 1. Let $\mathcal{C}=\{C_1, \ldots, C_n\}$ be a family of subsets of S. $\mathcal{C}$ is a *d-Sidon family,* if all the d-term unions are distinct:

$$\bigcup_{k=1}^{d} C_{i_k} \neq \bigcup_{k=1}^{d} C_{j_k} \tag{2}$$

if $\{i_1, \ldots, i_d\}\neq\{j_1, \ldots, j_d\}$.

Proposition 1. *$\mathcal{T}=\{T_1, \ldots, T_t\}$ is a system of tests of a parallel HGT, iff the dual system $\mathcal{C}=\{C_1, \ldots, C_n\}$ is a d-Sidon system.*

Proof. Suppose

$$\bigcup_{k=1}^{d} C_{i_k} = \bigcup_{k=1}^{d} C_{j_k}$$

but $\{i_1, \ldots, i_d\}\neq\{j_1, \ldots, j_d\}$. Then the test outcomes are identical whether $D=\{u_{i_1}, \ldots, u_{i_d}\}$ or $D=\{u_{j_1}, \ldots, u_{j_d}\}$. Thus $\mathcal{T}$ fails to identify the defectives.

Now suppose that $\mathscr{C}$ is a d-Sidon family. Then the set

$$\{j\colon\ T_j \text{ has a "contaminated" outcome}\}$$

corresponds to a unique d-element set $\{u_{i_1}, \ldots, u_{i_d}\}$ which is the set of defectives.

COROLLARY 1. *For any parallel HGT-system* $\mathscr{T}=\{T_1, \ldots, T_t\}$ *on n-elements with d defectives*

$$\binom{n}{d} \leqq 2^t. \tag{3}$$

This gives the lower bound in (1).

So we arrive at a dual form of the original problem which is an extremal problem for set-systems.

PROBLEM 1. Let t, d be given positive integers. Let $f(t, d)$ denote the maximum cardinality of a d-Sidon system on a t element set. Determine $f(t, d)$.

Recently Busch et al. [1] studied a stronger version of parallel HGT-systems, the d-complete designs.

DEFINITION. The family $\mathscr{T}=\{T_1, \ldots, T_t\}$ is called a d-complete design if for any d subscripts $\{i_1, \ldots, i_d\}$

$$\cup\Big\{T_j | j \notin \bigcup_{\nu=1}^{d} C_{i_\nu}\Big\} = S \setminus \{u_{i_1}, \ldots, u_{i_d}\}. \tag{4}$$

REMARK. By Proposition 1 a d-complete design can be used as a non-adaptive HGT when the number of defectives is d. To find the defectives is very simple:

$$D = S - \cup\{T_j | T_j \text{ is pure}\}.$$

Proposition 2 gives a characterization of d-complete designs.

PROPOSITION 2. $\mathscr{T}=\{T_1, \ldots, T_t\}$ *is a d-complete design iff for any d subscripts* $\{i_1, \ldots, i_d\}$

$$C_j \nsubseteq \bigcup_{k=1}^{d} C_{i_k} \quad \text{if} \quad j \notin \{i_1, \ldots, i_d\}. \tag{5}$$

PROOF. Suppose $C_1 \subseteq \bigcup_{k=2}^{d-1} C_k$. Then

$$\Big\{T_j | j \notin \bigcup_{k=2}^{d+1} C_k\Big\} = \cup\Big\{T_j | j \notin \bigcup_{k=1}^{d+1} C_k\Big\} \subseteq \{u_{d+2}, \ldots, u_n\} \neq S - \{u_2, \ldots, u_{d+1}\}.$$

Hence T is not a d-complete design.

Now suppose (5) holds. Then for any choice of distinct $i_0, i_1, \ldots, i_d$ there always exists a T_j such that

$$u_{i_0} \in T_j$$

but

$$u_{i_k} \notin T_j \quad \text{for} \quad k = 1, \ldots, d.$$

Thus

$$u_{i_0} \in \cup \{T_j | j \notin \bigcup_{k=1}^{d} C_{i_k}\}.$$

Consequently,

$$\{T_j | j \notin \bigcup_{k=1}^{d} C_{i_k}\} = S - \{u_{i_1}, \ldots, u_{i_d}\}.$$

The original HGT problem leads to the question what is the minimum number of sets in a d-design on n elements.

By Proposition 2 this yields to the following dual problem:

PROBLEM 2. Let t, d be given integers. Denote by $g(t, d)$ the maximum cardinality of a system $\mathscr{C}=\{C_1, \ldots, C_n\}$ on t elements which satisfies (5). Determine $g(t, d)$.

A system $\mathscr{C}=\{C_1, \ldots, C_n\}$ satisfies (5) if

$$(6) \qquad |C_i \cap C_j| < \frac{1}{d}|C_i| \qquad \forall i \neq j.$$

So we arrive to

PROBLEM 3. Let t, d be given integers. Denote by $h(t, d)$ the maximum cardinality of a system $\mathscr{C}=\{C_1, \ldots, C_n\}$ on t elements which satisfies (6). Determine $h(t, d)$.

Now we have three extremal problems for sets systems. Since the restriction (6) is stronger than (5) and this is stronger than (2), we have

$$(7) \qquad h(t, d) \leqq g(t, d) \leqq f(t, d).$$

In Theorem 3 we prove

$$(8) \qquad h(t, d) > c^t$$

where $c>1$ depends only on d. This proves in (1) the upper bound for $\varphi(n; d)$:

THEOREM 3.

$$h(t, d) > \frac{1}{2}\left(1+\frac{1}{(4d)^2}\right)^t.$$

PROOF. We use a sort of greedy algorithm to prove the theorem.

Let S be a t-element set, $r=\left[\frac{t}{(4d)^2}\right]$. Put $k=4dr$, $m=\frac{1}{d}\,k=4r$ and $[S]^k= =\{A | A\subset S,\ |A|=k\}$.

Choose $A_1 \in [S]^k$ arbitrarily. Delete all k-sets of S which intersect A_1 in at least m elements. I.e., let

$$\mathscr{B}_1 = \{B | B \in [S]^k,\ |B \cap A_1| \geqq m\}.$$

We define the sets A_i and the families B_i inductively. Suppose we have already

$A_1, \ldots, A_v, \mathscr{B}_1, \ldots, \mathscr{B}_v$. Then choose

$$A_{v+1} \in [S]^k \setminus \bigcup_{i=1}^{v} \mathscr{B}_i$$

arbitrarily. Define

$$\mathscr{B}_{v+1} = \{B \mid B \in [S]^k \setminus \bigcup_{i=1}^{v} \mathscr{B}_i,\ |B \cap A_{v+1}| \geqq m\}.$$

We proceed as long as we can. Suppose $A_1, \ldots, A_M$ have been chosen this way. Since

$$|\mathscr{B}_v| \leqq \sum_{i=m}^{k} \binom{k}{i}\binom{t-k}{k-i},$$

we certainly can continue unless

$$M \geqq \binom{t}{k}\left(\sum_{i=m}^{k}\binom{k}{i}\binom{t-k}{k-i}\right)^{-1}.$$

Put $b_i = \binom{k}{i}\binom{t-k}{k-i}$. Obviously

(9) $$\binom{t}{k} > b_r$$

and for $3r \leqq i < k = 4dr$

$$\frac{b_i}{b_{i-1}} = \frac{(k-i)^2}{i(t-2k+i)} \leqq \frac{1}{3}\frac{(4d-3)^2}{(4d)^2-8d+3} < \frac{1}{3}.$$

Hence

$$\binom{k}{m}\binom{t-k}{k-m} < b_r \prod_{i=3r}^{4r} \frac{b_{i+1}}{b_i} < b_r 3^{-r}$$

and

(10) $$\sum_{i=m}^{k}\binom{k}{i}\binom{t-k}{k-m} < 2b_r 3^{-r} < 2b_r\left(1+\frac{1}{(4d)^2}\right)^t.$$

By (9) and (10)

$$\binom{t}{k}\sum_{i=m}^{k}\binom{k}{m}\binom{t-k}{k-m} > \frac{1}{2}\left(1+\frac{1}{(4d)^2}\right)^t.$$

This means the above algorithm leads to a family $\mathscr{A} = \{A_1, \ldots, A_M\}$, which satisfies (6) and

$$M \geqq \frac{1}{2}\left(1+\frac{1}{(4d)^2}\right)^t.$$

By Corollary 1 and Theorem 3 we have

COROLLARY 2.

$$(\log(1+d^{-1}))^{-1}\log n < \varphi(n, d) < 2(\log 1+(4d)^{-2}))^{-1}\log n.$$

REMARK. A more careful computation would give a better constant instead of 4. We do not see how to diminish the gap between d^{-1} and d^{-2}.

Some open problems

1. We considered three different problems for set-systems. We have the trivial inequalities (7). What can be said on

$$\frac{g(t,d)}{h(t,d)} \quad \text{resp.} \quad \frac{f(t,d)}{g(t,d)}\,?$$

2. Is it possible to give an explicit construction for a "large" system which satisfies (2) resp. (5) or (6)?

Historical remarks

Hwang [8] gave a sequential HGT algorithm with $t=d\log_2\frac{n}{d}$ tests. Since sequential HGT algorithms also obey the inequality (with a different argument) as given in the Corollary of Theorem 3, Hwang's algorithm is asymptotically optimal for fixed d and large n.

For $d=2$, the best sequential HGT algorithm so far was given by Chang, Hwang and Lin [2]: For $t\geqq 4$,

$$n=\begin{cases}\lceil 43\cdot 2^{(t/2)-5}\rceil-1 & \text{for } t \text{ even},\\ \lceil 31\cdot 2^{(t-1/2)-4}\rceil-1 & \text{for } t \text{ odd}.\end{cases}$$

Freidlina [6] considered the non-adaptive HGT problem but did not require that all defectives be identified. He gave a construction for a non-adaptive algorithm in which each test consists of a set of random items with the probability of each item being included being q. He showed that for $q=1-2^{-1/d}$ and $t\geqq(1+\varepsilon)\log_2\binom{n}{d}$, then the probability of such an algorithm identifying all defectives is at least $1-\lambda$ where ε and λ are arbitrary numbers in $(0, 1)$.

Shapiro [10] and Fine [5] studied a different version of HGT called the "counterfeit coin" problem in which each test reveals the exact number of defectives contained in it and the number of defectives is unknown at the outset (of course it can be found out in one test). Söderberg and Shapiro [11] gave a non-adaptive algorithm with $t=O(n/\log n)$. Erdős and Rényi [4], and Lindström [9] proved that $t=n/\log_4 n)$ in an asymptotically optimal parallel algorithm for the counterfeit coin problem and gave a construction for such an algorithm.

A criticism of the HGT model is that in real applications one usually can determine only an upper bound but not the exact number of defectives. Now we know that a non-adaptive HGT algorithm or a d-complete design can also be applied to the case where only an upper bound d is known for the number of defectives. This is so because (2) implies

$$\bigcup_{\nu=1}^{k} C_{i_\nu} \neq \bigcup_{\nu=1}^{l} C_{j_\nu} \quad \text{if} \quad 1\leqq k\leqq l\leqq d$$

and

$$\{i_1, \ldots, i_k\} \neq \{j_1, \ldots, j_l\}.$$

ADDED in proof. Theorem 3 was proved independently by P. Erdős, P. Frankl and Z. Füredi [13].

REFERENCES

[1] BUSH, K. A., FEDERER, W. T., PESOTAN, H. and RAGHAVARAO, D., New combinatorial designs and their application to group testing, 1980 (preprint).

[2] CHANG, G. J., HWANG, F. K. and LIN, S., Group testing with two defectives, *Discrete Appl. Math.* **4** (1982), 97—102. *MR* **84e**: 05010.

[3] ERDŐS, P., FRANKL, P. and FÜREDI, Z., Families of finite sets in which no set is covered by the union of two others, *J. Combin. Theory Ser. A* **33** (1982), 158—166. *MR* **84e**: 05002.

[4] ERDŐS, P. and RÉNYI A., On two problems of information theory, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** (1963), 229—243. *MR* **29** #3268.

[5] FINE, N. J., Solution El 399, *Amer. Math. Monthly* **67** (1968), 697.

[6] FREIDLINA, V. L., A certain problem on the design of screening experiments, *Teor. Verojatnost. i Primenen.* **20** (1975), 100—114 (in Russian). *MR* **51** #9388.

[7] HALL, M., *Combinatorial Theory*, Blaisdell Publishing Co., Waltham, 1967. *MR* **37** #80.

[8] HWANG, F. K., A method for detecting all defective members in a population by group testing, *J. Amer. Statist. Assoc.* **67** (1972), 605—608.

[9] LINDSTRÖM, B., On a combinatorial detection problem, I., *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **9** (1964), 195—207. *MR* **29** # 5750.

[10] SHAPIRO, H. S., Problem El 399, *Amer. Math. Monthly* **67** (1960), 82.

[11] SÖDERBERG, S. and SHAPIRO, H. S., A combinatory detection problem, *Amer. Math. Monthly* **10** (1963), 1066—1070.

[12] SPERNER, E., Ein Satz über Untermengen einer endlichen Menge, *Math. Zeitschr.* **27** (1928), 544—548.

[13] ERDŐS, P., FRANKL, P. and FÜREDI, Z., Families of finite sets in which no set is covered by the union of *r* others, *Israel J. Math.* **51** (1985), no. 1-2.

(Received March 25, 1982; in revised form February 25, 1986)

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NJ 07974
U.S.A.

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 265—273

ON 3-SKEINS IN A 3-CONNECTED GRAPH

A. K. KELMANS

1. Introduction

H. Whitney has introduced and investigated [1], [2] the concept of cyclic isomorphism of graphs. A k-skein isomorphism of graphs [3] (see also 2.6 here) is a certain generalization of the concept of cyclic isomorphism, and the natural question arises when a k-skein isomorphism of graphs is induced by an isomorphism of the graphs.

2. Basic notions and notations

All the notions and notation used but not defined here may be found in [4].

2.1. Let G be an undirected graph without loops or multiple edges, VG be the set of vertices and EG be the set of edges of G. A *k-star* is a graph with k edges incident to a common vertex. K_n is a complete graph with n vertices.

2.2. A graph G is *k-connected* if G has at least $k+1$ vertices and upon removing any $k-1$ vertices from G the resulting graph is connected.

2.3. A chain T will sometimes be denoted by xTy or $[xTy]$ to identify x and y as the end-elements (vertices or edges) of T. A subchain of T with end-elements p and q is denoted by pTq. Let (xTy) and $(xTy]$ chains without x, y and without x, respectively. A vertex of (xTy) is called an *inner vertex* of T.

2.4. A *thread* of G is a maximal chain of G (by inclusion) whose inner vertices are of degree 2 in G. Two graphs are called *homeomorphic* if they may be turned into isomorphic graphs by changing the length of their threads.

2.5. The union of $k \geqq 1$ chains of G is called a *k-skein* [3] if they have common end-vertices (the *terminals* of the chain) and no pair of them has an inner vertex in common. In other words a k-skein is a homeomorph of the graph with two vertices and k parallel edges. A k-skein Z will sometimes be denoted by xZy to identify x and y as the terminals of Z. Put $W_k(G)=\{A\subseteqq EG : A \text{ is the edge-set of a } k\text{-skein in } G\}$.

2.6. A one-to-one map e: $EG\to EF$ is called a *k-skein isomorphism* of G onto F if $A\in W_k(G) \Leftrightarrow e(A)\in W_k(F)$. Since a 2-skein is a cycle, a 2-skein isomorphism is called a *cycle isomorphism.*

1980 *Mathematics Subject Classification.* Primary 05C40.
Key words and phrases. Isomorphisms of graphs.

2.7. A *cut* of a connected graph G is a set K of edges of G such that $G-K$ is disconnected but $G-K'$ is connected for any $K'\subset K$. A *k-cut* is a cut with k edges. Given a partition (X, Y) of VG put $E(X, Y)=\{(x, y)\in EG: x\in X, y\in Y\}$. If K is a cut of a connected graph G then $G-K$ has exactly two components G_1 and G_2. It means that K induces the partition (VG_1, VG_2) and $K=E(VG_1, VG_2)$.

2.8. Partitions (X_1, X_2) and (Y_1, Y_2) of a set V are said to be *transversal* if $X_i\cap Y_j\neq\emptyset$ for any $i, j\in\{1, 2\}$, and *parallel* otherwise. Cuts $K=E(X_1, X_2)$ and $R=E(Y_1, Y_2)$ in G are *transversal (parallel)* if the corresponding partitions (X_1, X_2) and (Y_1, Y_2) of VG are transversal (parallel).

2.9. The set $S_x(G)$ of edges of G incident to a vertex x is called the *vertex star* of x. Put

$\mathscr{S}(G) = \{S_x(G)\colon x\in VG\}$,
$\mathscr{S}_k(G) = \{A \subseteq EG\colon A \text{ is a } k\text{-star}\}$,
$\mathscr{H}_k(G) = \{A \subseteq EG\colon A \text{ is a } k\text{-cut of } G\}$ and
$\mathscr{R}_k(G) = \mathscr{H}_k(G)\cup\mathscr{S}_k(G)$.

2.10. Given distinct $A, B, C\in\mathscr{R}_k(G)$ we say that B is *between* A and C and write $(A_BC)_k$ or $(C_BA)_k$ if (1) there is a k-skein Z in G such that $A, B, C\subseteq EZ$, and (2) for any k-skein T in G, $A, C\subseteq ET \Rightarrow B\subseteq ET$.

3. Preliminaries

R. Halin and H. A. Jung have proved [3] the following

THEOREM. *A k-skein isomorphism of two $(k+1)$-connected graphs, $k\geqq 2$, is induced by an isomorphism of the graphs.*

It is easy to see that the same assertion for 2-skeins in 2-connected graphs is wrong [2]. Obviously, every permutation of the edge-set of K_4 is a 3-skein automorphism of K_4 but not every such map is induced by an automorphism of K_4. In [3] we read: "... we were not able to find more complicated examples in the case $k=3$, nor could we find for $k\geqq 4$ an example of a k-skein isomorphism of k-vertex connected graphs which was not induced by an isomorphism."

Here it will be proved (see Theorem 4.4) that no other example of this kind exists for 3-connected graphs. We will see that a 3-connected graph G is uniquely defined up to isomorphism by $W_3(G)$.

The basic idea is to show that the set $\mathscr{S}(G)$ of the vertex-stars of a 3-connected graph G can be obtained from the collection $W_3(G)$ of the edge-sets of its 3-skeins. The same idea has been used in [5] to give a simple prove of the Whitney theorem on cyclic isomorphism of 3-connected graphs.

4. The main assertions

THEOREM 4.1 (see 5.5). *The following conditions are equivalent:*

(a) *G is a homeomorph of a* 3*-connected graph,*

(b) *G has no isolated vertex, has at least two* 3*-skeins, and four edges of G belong to a common* 3*-skein in G iff the threads of G containing these four edges do not form a homeomorph of a* 4*-star.*

COROLLARY. *A cubic graph is* 3*-connected if and only if any four edges in the graph belong to a common* 3*-skein.*

THEOREM 4.2 (see 6.6). *Suppose that G is* 3*-connected,* $|VG|\geqq 5$ *(that is* $G\neq K_4$*). Then A is a* 3*-cut of G if and only if for any* 3*-skein Z of G* $|A\cap EZ|\neq 1$.

THEOREM 4.3 (see 7.4). *Suppose that G is k-connected. Then B is a k-cut of G distinct from a k-star if and only if B is a k-cut and there exist two edge-sets A and C in G such that each of A and C is a k-star or a k-cut and B is between A and C (see Definition* 2.10*).*

THEOREM 4.4 (see 8.3). *Suppose that G is* 3*-connected, F has no isolated vertex,* $|VF|\geqq 5$, *and* $e\colon EG\to EF$ *is a one-to-one map such that* $A\in W_3(G)\Leftrightarrow e(A)\in W_3(F)$. *Then an isomorphism of G onto F exists and the only one which induces e.*

5. 3-skeins presentation of the vertex-stars with $\geqq 4$ edges in a 3-connected graph

LEMMA 5.1. *Suppose that G is* 3*-connected, A and B are disjoint cycles in G and* $u, v\in EB$. *Then A, u and v belong to a common* 3*-skein in G.*

PROOF. The lemma follows immediately from the fact that by the Menger theorem in G there are three disjoint chains between A and B. ∎

LEMMA 5.2. *Suppose that G is* 3*-connected, cycles A and B in G have only one vertex x in common,* $u, v\in EB$ *and v is not incident to x. Then in G there exists a* 3*-skein containing A, u and v.*

PROOF Let $u=(s,t)$, $v=(p,q)$ and the chains xB^uu and xB^vv in B do not contain v and u, respectively, and the chain tB^1q in B does not contain x, u and v. Let $e=(x,y)\in E(xB^vv)$ be the end-edge of the chain xB^vv. By the lemma's hypothesis $x\neq p$. Therefore $e\neq v$ and $V=V[yB^vq]$ has at least 2 vertices. In $A\cup B$ there are two disjoint chains from V to $A'=A\cup[xB^ut]$, namely $[yex]$ and $[qB^1t]$ (may be $q=t$, so that $qB^1t=\{q=t\}$) (see Fig. 1). Since G is 3-connected, by the Menger theorem and the Ford—Fulkerson theorem [5] in G there exist three chains $a_iC_ib_i$, $i=1,2,3$ with the following properties: (1) $C_i\cap(V\cup A')=\{a_i,b_i\}$, $a_i\in A'$ and $b_i\in V$, (2) $a_1C_1b_1=xey$, (3) $b_2=q$, $b_3\in V[yB^vq)=V-q$, $t\in\{a_2,a_3\}$, $x\notin\{a_2,a_3\}$, and (4) the chains C_i, $i=1,2,3$, are pairwise inner disjoint.

Suppose that $t=a_2$ (Fig. 1). If $a_3\in(xB^us]$, then the cycle $b_3C_3a_3B^ua_2C_2b_2B^vb_3$ contains u and v and has no vertex in common with A. By Lemma 5.1 A, u and v belong to a common 3-skein in G. If $a_3\in A-x$, then $A\cup(a_3C_3b_3B^vb_2C_2a_2B^ux)$ is a 3-skein containing A, u and v.

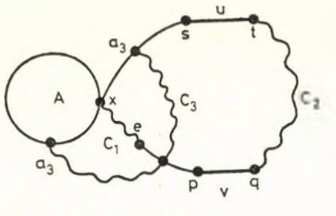


Fig. 1

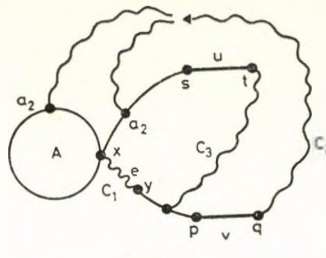


Fig. 2

Suppose instead that $t=a_2$ (Fig. 2). If $a_2\in(xB^us]$, then the cycle

$$b_3C_3a_3B^ua_2C_2b_2B^vb_3$$

contains u and v and has no vertex in common with A. By Lemma 5.1 there exists a 3-skein in G containing A, u and v. If $a_2\in A-x$, then $A\cup(a_2C_2b_2B^vb_3C_3a_3B^ux)$ is a 3-skein containing A, u and v. ▮

5.3. Obviously a 4-star does not belong to a 3-skein.

LEMMA. *In a 3-connected graph any four edges which do not form a 4-star belong to a common 3-skein.*

PROOF. Obviously, any 3 edges in a 2-connected graph belong to a common 3-skein. Suppose that a set $E=\{e_1, e_2, e_3, e_4\}$ of edges of a 3-connected graph G does not form a 4-star. If some 3 edges from E lie on a common cycle, then obviously E belongs to a 3-skein in G.

Suppose that e_1, e_2 and e_3 belong to distinct threads T_1, T_2 and T_3, respectively of a 3-skein $pZq=T_1\cup T_2\cup T_3$. Since G is 2-connected, there is a chain xT_4y in G such that $e_4\in ET_4$ and $Z\cap T_4=\{x, y\}$. If x and y belong to no common thread in Z or if for some $i\in\{1, 2, 3\}$ $x, y\in T_i$ and $e_i\notin xT_iy$, then $Z\cup T_4$ obviously has a 3-skein containing E. Therefore we suppose that $x, y\in T_1$ and $e_1\in xT_1y$. If $\{x, y\}\neq\{p, q\}$, then by Lemma 5.1 or by Lemma 5.2 and the fact that E is not a 4-star G has a 3-skein containing E. Now suppose that $\{x, y\}=\{p, q\}$, so that $T_1\cup T_2\cup T_3\cup T_4=F$ is a 4-skein. Without loss of generality we may assume that $ET_4\neq\{e_4\}$. Since G is 3-connected, there is a chain aPb in G such that $F\cap P=\{a, b\}$, $a\in(pT_4q)$ ($a\in(pT_4e_4)$, say) and $b\in F-T_4$ ($b\in(pT_1q)$, say). If $b\in(e_1T_1q)$ then $F\cup P$ obviously has a 3-skein containing E. Now $b\in(pT_1e_1)$. Since E is not a 4-star, by Lemma 5.2 G has a 3-skein containing E. ▮

LEMMA 5.4. *Suppose that G is not a homeomorph of a 3-connected graph and not a 4-star. Then there are four edges in G which do not form a 4-star and do not belong to a common 3-skein.*

PROOF. Without loss of generality we may assume that G is a homeomorph of 2-connected graph. Since G is not hom 3-connected, there are subgraphs A and B such that $G=A\cup B$, $EA\cap EB=\emptyset$, $VA\cap VB=\{x, y\}$, and A (B) has two edges a_1 and a_2 (b_1 and b_2) incident to x (y). Then the edges a_1, a_2, b_1 and b_2 do not form a 4-star and obviously cannot belong to a common 3-skein in G. ▮

From Lemmas 5.3 and 5.4 there follows

THEOREM 5.5. *The following conditions are equivalent:*
(a) *G is a homeomorph of a 3-connected graph,*
(b) *G has no isolated vertex, has at least two 3-skeins, and four edges of G belong to a common 3-skein iff the threads of G containing these four edges do not form a homeomorph of a 4-star.*

From Lemma 5.3 there follows directly

LEMMA 5.6. *Suppose that G is 3-connected. Then A is a vertex-star of G with $\geqq 4$ edges iff A is a maximal* (by inclusion) *set of edges of G such that any four edges of A do not belong to a common 3-skein in G.*

5.7. Obviously, B is a 3-star but not a 3-cut of G iff B belongs to a vertex-star with $\geqq 4$ edges.

6. 3-skeins presentation of 3-cuts in 3-connected graphs

LEMMA 6.1. *Suppose that G is 3-connected, A is a subgraph of G with two components one of which is a chain with one edge ae_1b and the other is the chain $x_1e_2x_2e_3x_3$ with two edges. Then there exists a 3-skein Z in G such that $|EA\cap EZ|=1$.*

PROOF. Since G is 3-connected, there are three inner disjoint chains aT_1x_1, aT_2x_2 and bT_3x_3 connecting $\{a, b\}$ with $\{x_1, x_2, x_3\}$ (Fig. 3). Since G is 3-connected, there is a chain x_1Py in G such that $P\cap(T_1\cup T_2\cup T_3)=\{x_1, y\}$. If $y\in T_1\cup T_3$ then the 3-skein $P\cup T_1\cup T_2\cup e_2$ has only one edge e_2 in common with A. Therefore suppose that $y\in T_3$. Since G is 3-connected, there is a chain pCq in G such that $q\in\{b, x_3\}$ and $C\cap(T_1\cup T_2\cup T_3\cup P)=\{p, q\}$. Since the cycles $x_1T_1ae_1bT_3yPx_1$ and $x_1T_1aT_2x_2e_3x_3T_3yPx_1$ have only one edge in common with A (e_1 and e_3, respectively), the subgraph $T_1\cup T_2\cup T_3\cup P\cup C\cup A$ has a 3-skein Z such that $|EZ\cap EA|=1$. ∎

LEMMA 6.2. *Suppose that G is 3-connected, E is a set of three pairwise non-adjacent edges in G and $G-E$ is connected. Then there is a 3-skein Z in G such that $|E\cap EZ|=1$.*

PROOF. Let $E=\{e_1, e_2, e_3\}$ and $e_i=(x_i, y_i)$, $i=1, 2, 3$. Since $G-E$ is connected, there is a chain $x_1T_1y_1$ in G such that $E\cap ET_1=\emptyset$, Since G is 3-connected,

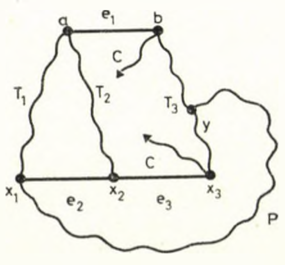


Fig. 3

there are three pairwise disjoint chains $(x_2 P_i z_i)$, $i=1, 2, 3$, in G from x_2 to T_1. If some two chains of P_i do not contain e_2 and e_3 then we are done. Otherwise let $e_2 \in P_2$ and $e_3 \in P_3$ (Fig. 3). Put $B=E \cup T_1 \cup P_1 \cup P_2 \cup P_3$. Since G is 3-connected, there is a chain $y_2 C p$ such that $B \cap C=\{y_2, p\}$. Since the cycles $x_1 e_1 y_1 T_1 x_1$, $x_2 e_2 y_2 P_2 z_3 T_1 z_1 P_1 x_2$ and $x_3 e_3 y_3 P_3 z_3 T_1 z_1 P_1 x_2 P_3 x_3$ have only one edge in common with E (e_1, e_2 and e_3, respectively) the subgraph $B \cup C$ has a 3-skein Z such that $|E \cap EZ|=1$. ∎

LEMMA 6.3. *Suppose that G is 3-connected, S is a 3-star in G and $G-ES$ is connected. Then there exists a 3-skein Z in G such that $|ES \cap EZ|=1$.*

PROOF. Let x be the vertex of degree 3 in S and $e=(x, y) \in ES$. Since $G-ES$ is connected, there is an edge $u=(x, z) \notin ES$ in G. Since G is 3-connected, in $G-x$ there are two inner disjoint chains $zT_1 y$ and $zT_2 y$. Then the 3-skein $(zuxey) \cup T_1 \cup T_2$ has only one edge e in S. ∎

LEMMA 6.4. *Suppose that G is 3-connected, $|VG| \geqq 5$ (that is $G \neq K_4$) and A is a triangle in G. Then there exists a 3-skein Z in G such that $|EA \cap EZ|=1$.*

PROOF. Put $VA=\{x_1, x_2, x_3\}$ and let $x \in VG-VA$. Since G is 3-connected, there are three disjoint chains $(xT_i x_i]$, $i=1, 2, 3$, such that $A \cap T_i=x_i$. Since $G \neq K_4$ and G is 3-connected, there is a chain pCq in G such that $C \cap (T_1 \cup T_2 \cup T_3)=$ $=\{p, q\}$. It is easy to see that $A \cup T_1 \cup T_2 \cup T_3 \cup C$ has a 3-skein Z such that $|EZ \cap EA|=1$. ∎

LEMMA 6.5. *Suppose that G is 3-connected, $|VG| \geqq 5$ (that is $G \neq K_4$) and P is a chain with 3 edges in G. Then there exists a 3-skein Z in G such that $|EP \cap EZ|=1$.*

PROOF. Let $P=x_1 e_1 x_2 e_2 x_3 e_3 x_4$. Since $|VG| \geqq 5$, there is $x \in G-P$. Since G is 3-connected, there are three disjoint chains $(xT_i t_i]$, $i=1, 2, 3$, such that $P \cap T_i=\{t_i\}$. We have two alternatives (up to notations): (1) $t_1=x_1$, $t_2=x_2$ and $t_3=x_3$, and (2) $t_1=x_1$, $t_2=x_2$ and $t_3=x_4$.

Case (1). Since G is 3-connected, there are two disjoint chains $(x_4 C_1 y_1]$ and $(x_4 C_2 y_2]$ such that $C_i \cap (T_1 \cup T_2 \cup T_3)=y_i$. Then $P \cup T_1 \cup T_2 \cup T_3 \cup C_1 \cup C_2$ obviously has a 3-skein Z containing only one edge from P (Fig. 4).

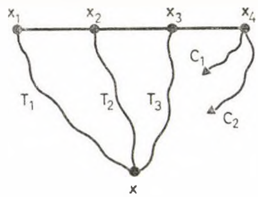


Fig. 4

Case (2). Since G is 3-connected, there is a chain $x_1 Cy$ in G such that $C \cap (T_1 \cup T_2 \cup T_3 \cup P)=y$. If $y \in T_1 \cup T_2 \cup T_3$ then $A=P \cup T_1 \cup T_2 \cup T_3 \cup C$ has a 3-skein containing only one edge from P. Therefore suppose $y=x_3$. Since G is 3-connected, there is a chain pRq in G such that $R \cap A=\{p, q\}$ and $p \in [x_4 T_3 x)$. Then $A \cup R$ contains a required 3-skein (Fig. 5). ∎

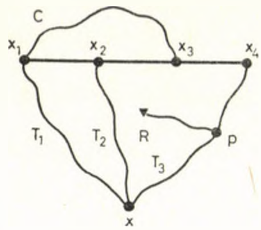


Fig. 5

THEOREM 6.6. *Suppose that G is 3-connected, $|VG| \geqq 5$ (that is $G \neq K_4$). Then E is a 3-cut of G if and only if $|E \cap EZ| \neq 1$ for any 3-skein Z of G.*

PROOF. (p1) Suppose that $G-E$ is disconnected. Since each edge of a 3-skein Z belongs to a cycle of Z, we have $|E \cap EZ| \neq 1$.

(p2) If $G-E$ is disconnected and $|E|=3$ then the 3-connectivity of G implies that the edges of E are either pairwise non-adjacent or have a vertex in common (that is they form a 3-star). Therefore from Lemmas 6.1—6.5 it follows that if $|E \cap EZ| \neq 1$ for any 3-skein Z then $G-E$ is disconnected. ∎

7. 3-skeins recognition of the star 3-cuts and the matching 3-cuts in a 3-connected graph

LEMMA 7.1. *Suppose that a minimal cut of G has an odd number of edges. Then any two minimal cuts of G are parallel (see Definition 2.8).*

PROOF. (p1) If (X_1, X_2) and (Y_1, Y_2) are transversal partitions of VG then

$$\sum \{|E(X_i \cap Y_j, \quad VG-(X_i \cap Y_j))|, \ i, j \in \{1,2\}\} = 2|E(X_1, X_2)| + 2|E(Y_1, Y_2)|.$$

Now suppose that $E(X_1, X_2)$ and $E(Y_1, Y_2)$ are minimal cuts of G with k edges. Then from the equality it follows that $E(X_i \cap Y_j, VG-(X_i \cap X_j))$ is a minimal cut for any $i, j \in \{1, 2\}$.

$$\text{(p2)} \ k = |E(X_1, X_2)| = |E(X_1 \cap Y_1, VG-(X_1 \cap Y_1))| +$$

$$|E(X_1 \cap Y_2), VG-(X_1 \cap Y_2))| - 2|E(X_1, X_2)| = 2k - 2|E(X_1, X_2)|.$$

Therefore if $E(X_1, X_2)$ and $E(Y_1, Y_2)$ are transversal and minimal then k is even. ∎

THEOREM 7.2. *Suppose that G is k-connected and k is odd. Suppose also that $A, B, C \in \mathscr{R}_k(G)$ (see Def. 2.9) and there exists a k-skein xZy in G containing A, B and C. Then one and only one of the relations $(A_B C)_k$, $(B_A C)_k$ and $(A_C B)_k$ is true.*

PROOF. (p1) Since A is a cut of G, we have $A = E(A_x, A_y)$. Since $A \subseteq EZ$, we obtain $x \in A_x$, $y \in A_y$. Also $B = E(B_x, B_y)$ and $C = E(C_x, C_y)$ and $x \in B_x \cap C_x$ and $y \in B_y \cap C_y$.

(p2) By Lemma 7.1 A, B and C are pairwise parallel. Therefore we may assume without loss of generality that $A_x \subset B_x$. Since C is parallel to A and B, and since

$x \in A_x \cap B_x \cap C_x$, we have the following 3 possibilities:

(1) $A_x \subset B_x \subset C_x$, (2) $A_x \subset C_x \subset B_x$ and (3) $C_x \subset A_x \subset B_x$.

(p3) Obviously, for any k-cut $K=E(X, Y)$ and a 3-skein aTb

$$(*) \qquad K \subseteqq ET \Leftrightarrow (a \in X, b \in Y) \quad \text{or} \quad (b \in Y, a \in X).$$

Suppose that $A_x \subset B_x \subset C_x$. If a 3-skein aTb contains A and C then by $(*)$ $a \in A_x \cap \cap C_x = A_x$ and $b \in A_y \cap C_y = C_y$. Therefore $B \subseteqq ET$. It means that $(A_B C)_k$. Let $u \in B_x - A_x$ and $v \in C_x - B_x$. Since G is k-connected there are k-skeins xPv and yQu. Obviously, $A, B \subseteqq EP$ and $C \nsubseteqq EP$ that is not $(A_C B)_k$, and $B, C \subseteqq EQ$ and $A \nsubseteqq EQ$ that is not $(B_A C)_k$.

In the same way we have in case (2) $(A_C B)_k$ and in case (3) $(C_A B)_k$. ▮

THEOREM 7.3. *Suppose that G is k-connected. Then $B \in \mathscr{H}_k(G) - \mathscr{S}_k(G)$ (that is B is a k-cut but not a k-star) if and only if $B \in \mathscr{H}_k(G)$ and there exist distinct $A, C \in \mathscr{H}_k(G) \cup \mathscr{S}_k(G) - \{B\}$ such that $(A_B C)_k$.*

PROOF. (p1) Suppose that B is a k-cut of G but not a k-star. Choose two vertices a and c in distinct components of $G-B$. Since G is k-connected, there is a k-skein aZc in G and $B \subseteqq EZ$. Put $A=S_a \cap Z$ and $C=S_c \cap Z$. Then, obviously, $(A_B C)_k$.

(p2) Suppose that $A, B, C \in \mathscr{H}_k(G) \cup \mathscr{S}_k(G)$ and $(A_B C)_k$. Then there is a k-skein xPy in G such that $A, B, C \subseteqq EP$. Suppose the contrary: $B=E(B_x, B_y)$ is a k-star that is $B_x=\{x\}$, say. Since $A, B, C \subseteqq EP$ we have A or $C \in \mathscr{H}_k(G) - \mathscr{S}_k(G)$ (A, say). Then $A=E(A_x, A_y)$, $x \in A_x$, $y \in A_y$.

Suppose $C \in \mathscr{S}_k(G)$ that is $C=S_y$. Since $A \neq B$ there is $z \in A_x - x$. Suppose $C \in \mathscr{H}_k(G) - \mathscr{S}_k(G)$. Then $C=E(C_x, C_y)$, $x \in C_x$, $y \in C_y$. Since G is k-connected, each cut of G with k edges is a matching. Therefore there is $z \in A_x \cap C_x - x$.

Since G is k-connected, there is a k-skein yQz, and $A, C \subseteqq EQ$ but $B \nsubseteqq EQ$. This contradicts to $(A_B C)_k$. ▮

REMARK. For odd k Theorem 7.3 follows in fact from Theorem 7.2.

8. A 3-skein isomorphism of 3-connected graphs implies an isomorphism of these graphs

8.1. Thus 5.6, 5.7, 6.6 and 7.3 give a procedure for obtaining the set $\mathscr{S}(G)$ of vertex stars from the collection $W_3(G)$ of the edge-set of 3-skeins for a 3-connected graph $G \neq K_4$. Therefore we have

THEOREM. *Suppose that G and F are 3-connected graphs, $|VG| \geqq 5$ (that is $G \neq K_4$) and e: $EG \to EF$ is a one-to-one map such that $A \in W_3(G)$ if and only if $e(A) \in W_3(F)$. Then an isomorphism v: $VG \to VF$ of G onto F exists and the only one which induces e.*

The uniqueness of an isomorphism which induces e follows from the fact that G has no isolated edge.

Below (see 8.3) we give a strengthening of the theorem showing that it is sufficient to require 3-connectedness only for one of G and F.

LEMMA 8.2. *In a 3-connected graph G for any two edges e_1 and e_2 there exists a 3-skein Z such that $e_1 \in EZ$ and $e_2 \notin EZ$.*

PROOF. Let $e_i=(x_i, y_i)$, $i=1, 2$. If in $G-e_2$ there is a 3-skein x_1Ty_1 then $T \cup e_1$ obviously contains a required 3-skein. Otherwise e_1 and e_2 belong to a 3-cut $R=\{e_1, e_2, e_3\}$, $e_3=(x_3, y_3)$. We may assume that x_1, x_2 and x_3 belong to a common component in $G-R$. Since G is 3-connected, R is a matching and in $G-e_2$ there is a 3-skein x_1Px_3 containing e_1 and e_3. ∎

COROLLARY. *Suppose that G is a homeomorph of a 3-connected graph. Then A is the edge-set of a thread in G if and only if $A \cap EZ = A$ or $\emptyset$ for any 3-skein Z in G.*

THEOREM 8.3. *Suppose that G is 3-connected, F has no isolated vertex $|VF| \geqq 5$, and e: $EG \to EF$ is a one-to-one map such that $A \in W_3(G)$ if and only if $e(A) \in W_3(F)$. Then an isomorphism v: $VG \to VF$ of G onto F exists and the only one which induces e.*

PROOF. Since e is a 3-skein isomorphism of G onto F and since G is 3-connected and therefore has at least six 3-skeins, by Theorems 5.5, 6.6, 7.3 F is a homeomorph of a 3-connected graph. Since each thread in G has exactly one edge, by Lemma 8.2, each thread of F has exactly one edge. Therefore F is 3-connected. Now the theorem follows from Theorem 8.1. ∎

The results were reported to the All-union seminar on graphs and hypergraphs (Odessa, September, 1980) and briefly stated in [7].

REFERENCES

[1] WHITNEY, H., Congruent graphs and the connectivity of graphs, *Amer. J. Math.* **54** (1932), 150—168.
[2] WHITNEY, H., 2-isomorphic graphs, *Amer. J. Math.* **55** (1933), 243—254.
[3] HALIN, R. and JUNG, H. A., Note on isomorphism of graphs, *J. London Math. Soc.* **42** (1967), 254—256. *MR* **34** #7402.
[4] HARARY, F., *Graph theory*, Addison-Wesley, Reading, Mass., 1969. *MR* **41** #1566.
[5] KELMANS, A. K., The concept of a vertex in a matroid, the non-separating cycles of a graph and a new criterion for graph planarity, *Algebraic methods in graph theory*, 345—388, *Coll. Math. Soc. János Bolyai, 25*, North-Holland, Amsterdam—New York, 1981. *MR* **84i:** 05040.
[6] FORD, L. R. and FULKERSON, D. R., *Flows in networks*, Princeton Univ. Press, Princeton, N. J., 1962. *MR* **28** #2917.
[7] KELMANS, A. K., 3-skeins in 3-connected graphs, I. All-union conference on statistical and discrete analysis of non-digital information, Moscow—Alma Ata, 1981.

(Received July 8, 1981)

INSTITUTE OF CONTROL SCIENCES
PROFSOYUZNAYA 65
SU—117 342 MOSCOW
SOVIET UNION

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 275—285

WEAK CONVERGENCE OF RANDOM SUMS TO INFINITELY DIVISIBLE DISTRIBUTIONS

K. S. KUBACKI

1. Let (X_{nk}) $(k\geq 1, n\geq 1)$ be a doubly infinite array (DIA) of random variables (r.vs) defined on a common probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Assume that (X_{nk}) is adapted to an array $(\mathcal{F}_{nk})$ $(k\geq 0, n\geq 1)$ of row-wise increasing sub-σ-fields of $\mathcal{F}$, i.e. X_{nk} is $\mathcal{F}_{nk}$-measurable and $\mathcal{F}_{n,k-1}\subset\mathcal{F}_{nk}\cdot\mathcal{F}_{n0}$ need not be the trivial σ-field $\{\emptyset, \Omega\}$. We put

$$S_{nk}=\sum_{j=1}^{k} X_{nj}, \qquad k\geq 1, \quad n\geq 1.$$

Recently, several papers have appeared which are devoted to the study of the limit distribution of S_{nN_n} as $n\to\infty$, where $\{N_n, n\geq 1\}$ is a sequence of positive integer-valued r.vs defined on the same probability space $(\Omega, \mathcal{F}, \mathsf{P})$. There have been two basic problems on the field. One is when a limit theorem is already given and the question is about the optimal conditions ensuring that the same theorem, or a mixtured version of it, remain true with random indices (see, e.g. [11], [8], [2], and their references). The other problem is to prove directly random-sums limit theorems, and to determine the class of possible limit distributions. Of course, the conditions ensuring the random limit theorem ought to be reduced to the classical ones when $N_n=n$. However, there are (X_{nk}) such that S_{nN_n} satisfies the random limit theorem whereas S_{nn} does not weakly converge (see, e.g. [10], Example 1).

The latter problem has been considered by many authors. The case when $N_n, X_{n1}, X_{n2}, \ldots$ are independent r.vs has been investigated in [11], [8] and [10], while the case when N_n are stopping times has been investigated in [9], [1] and [4] (see also references in these papers). A general case, when N_n need not be (for every n) independent of X_{nk} or a stopping time, has been considered in [5] under the assumption that X_{nk} have finite conditional variances.

The aim of the present paper is to consider the general case without any assumptions about the existence of moments of the r.vs X_{nk}.

In the sequel we use the notation

$$\mathsf{E}_{k-1}Z_{nk}=\mathsf{E}(Z_{nk}|\mathcal{F}_{n,k-1}), \qquad \mathsf{P}_{k-1}(A_{nk})=\mathsf{P}(A_{nk}|\mathcal{F}_{n,k-1}).$$

Throughout, $I(A)$ denotes the indicator function of the set A, and the various kinds of convergence, in L_p norm, in probability, and weak (in distribution) are denoted by $\xrightarrow{L_p}$, $\xrightarrow{\mathsf{P}}$ and $\xrightarrow{\mathcal{D}}$, respectively. All equalities and inequalities between r.vs

1980 *Mathematics Subject Classification.* Primary 60F05; Secondary 62E20, 60E15.

Key words and phrases. Weak convergence, randomly indexed partial sums, random limit theorems, stopping times.

are considered in the sense "with probability one", and all limits are taken as $n\to\infty$, unless stated otherwise.

We put

$$(1.1)\qquad A(t)=\exp\left[it\gamma+\int_{-\infty}^{+\infty}\left(\exp(itx)-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\,dK(x)\right],$$

where γ is a fixed real number, and K is a bounded and nondecreasing real function such that $K(-\infty)=0$. Then A is the characteristic function (ch.f.) of an infinitely divisible distribution in the Lévy—Khintchine representation (see, e.g. [7]).

The following routine lemma will be useful throughout this paper.

LEMMA 1 ([5]). *Let (X_{nk}) be a DIA of r.vs adapted to an array $(\mathcal{F}_{nk})$ of row-wise increasing sub-σ-fields, and let $\{N_n\}$ be a sequence of positive integer-valued r.vs such that*

$$(1.2)\qquad f_{nN_n}(t)=\prod_{j=1}^{N_n}\mathsf{E}_{j-1}\exp(itX_{nj})\xrightarrow{\mathsf{P}}A(t),$$

for some t, implies

$$(1.3)\qquad \mathsf{E}\left(\exp(itS^*_{nN_n})(f^*_{nN_n}(t))^{-1}\right)\to 1,$$

where

$$S^*_{nN_n}=\sum_{k=1}^{N_n}X^*_{nk},\qquad f^*_{nN_n}(t)=\prod_{k=1}^{N_n}\mathsf{E}_{k-1}\exp(itX^*_{nk})$$

and

$$X^*_{nk}=X_{nk}I(|f_{nk}(t)|\geqq|A(t)|/2).$$

Then, under (1.2), $\mathsf{E}\exp(itS_{nN_n})\to A(t)$.

REMARK 1. One can easily prove that (1.2) implies (1.3) for example in the following cases:

(a) N_n is for every n independent of $\{X_{nk},\ k\geqq 1\}$, and $\mathcal{F}_{n0}=\{\emptyset,\Omega\}$, $\mathcal{F}_{nk}=\sigma\{X_{n1},X_{n2},\dots,X_{nk}\}$;

(b) N_n is for every n a stopping time w.r.t. $\{\mathcal{F}_{nk},\ k\geqq 1\}$;

(c) $X_{nk}=X_k$ for every k and all n, and $N_n\xrightarrow{\mathsf{P}}\infty$;

(d) $\mathsf{P}(k_n\leqq N_n\leqq r_n)\to 1$ and $\max\limits_{k_n\leqq k\leqq r_n}\alpha(\mathcal{F}_{nk},\mathcal{A}_{nk})\to 0$,

where $\mathcal{A}_{nk}=\{\emptyset,\Omega,\{N_n=k\},\{N_n\neq k\}\}$, $k_n\leqq r_n$ are positive integers, and $\alpha(\mathcal{F},\mathcal{A})=\sup\{|\mathsf{P}(B\cap C)-\mathsf{P}(B)\mathsf{P}(C)|:\ B\in\mathcal{F},\ C\in\mathcal{A}\}$.

The proof of Remark 1 is based on the fact that for every n the sequence $Z^n_k=\exp(itS^*_{nk})(f^*_{nk}(t))^{-1}-1$, $k\geqq 1$, forms a uniformly bounded martingale w.r.t. $\{\mathcal{F}_{nk},\ k\geqq 1\}$ such that $\mathsf{E}Z^n_k=0$, and is not detailed here.

2. Let (τ_{nk}) $(k\geqq 1,\ n\geqq 1)$ be a DIA of nonnegative r.vs such that

$$\left|\mathsf{E}_{k-1}X_{nk}I(|X_{nk}|\leqq\tau_{nk})\right|<\infty\quad\text{for every } k \text{ and all } n.$$

We put

$$a_{nk}=\mathsf{E}_{k-1}X_{nk}I(|X_{nk}|\leqq\tau_{nk}),\qquad Y_{nk}=X_{nk}-a_{nk}.$$

What we consider to be the basic random limit theorem for partial sums of dependent r.vs, can now be formulated as follows.

THEOREM 1. *Let* (X_{nk}) *and* $\{N_n\}$ *be as in Lemma 1, and assume that*

$$\max_{1\leq k\leq N_n} \mathsf{P}_{k-1}(|Y_{nk}|\geqq\varepsilon)\xrightarrow{\mathsf{P}}0, \quad \textit{for all} \quad \varepsilon>0, \tag{2.1}$$

$$\sum_{k=1}^{N_n}\left(\mathsf{E}_{k-1}\frac{Y_{nk}}{1+Y_{nk}^2}\right)^2\xrightarrow{\mathsf{P}}0, \tag{2.2}$$

$$\sum_{k=1}^{N_n}\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}I(Y_{nk}\leqq y)\xrightarrow{\mathsf{P}}K(y) \tag{2.3}$$

for every point y *of continuity of* K *and for* $y=\pm\infty$, *and*

$$\sum_{k=1}^{N_n}\left\{a_{nk}+\mathsf{E}_{k-1}\frac{Y_{nk}}{1+Y_{nk}^2}\right\}\xrightarrow{\mathsf{P}}\gamma. \tag{2.4}$$

Then S_{nN_n} *weakly converges to an infinitely divisible distribution, whose ch.f.* A *is given by* (1.1).

Theorem 1 extends the main result of [5], Theorem 1, to partial sums of dependent r.vs such that their first moments need not exist.

In the special case when the DIA (X_{nk}) is "conditionally infinitesimal", i.e.

$$\max_{1\leq k\leq N_n} \mathsf{P}_{k-1}(|X_{nk}|\geqq\varepsilon)\xrightarrow{\mathsf{P}}0, \quad \text{for all } \varepsilon>0, \tag{2.5}$$

and $\tau_{nk}=\tau$, where τ is a positive constant, Theorem 1 yields

THEOREM 2. *Let* (X_{nk}) *and* $\{N_n\}$ *be as in Lemma 1, and assume that* (2.3), (2.4) *and* (2.5) *are satisfied. Then* S_{nN_n} *weakly converges to an infinitely divisible distribution, whose ch.f.* A *is given by* (1.1).

One can see that in the special case, when N_n is a sequence of constants and for every n the r.vs $X_{n1}, X_{n2}, \dots$ are independent, (2.3) and (2.4) with $\tau_{nk}=\tau$ are just the conditions that are necessary and sufficient in order that $S_{nN_n}\xrightarrow{\mathcal{D}}X$, where X has ch.f. A (see, e.g. [6], p. 309). In the case when for every n the r.vs $N_n, X_{n1}, X_{n2}, \dots$ are independent Theorem 2 reduces to Theorem 6.1 of [8]. Moreover, under the assumption that N_n is a sequence of stopping times Theorem 2 is the 1-dimensional analogy of Theorem 1 from [1]. Of course, in all these cases (1.2) implies (1.3).

As simple consequences of Theorem 1 we also get:

COROLLARY 1. *Let* (X_{nk}) *and* $\{N_n\}$ *be as in Lemma 1, and assume that* (2.1), (2.2) *and* (2.4) *are satisfied, and that*

$$\sum_{k=1}^{N_n}\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}\xrightarrow{\mathsf{P}}D, \tag{2.6}$$

where D is a positive number, and for every real y

$$(2.7)\qquad \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(Y_{nk} \leqq y) \xrightarrow{\mathsf{P}} \frac{D}{2} + \frac{D}{\pi} \operatorname{arc\,tg}(y).$$

Then S_{nN_n} weakly converges to the Cauchy distribution with ch.f. $A(t)=\exp(it\gamma - D|t|)$.

COROLLARY 2. *Let (X_{nk}) and $\{N_n\}$ be as in Lemma 1, and assume that* (2.1), (2.2), (2.4) *and* (2.6) *are satisfied with $\gamma=D$, and that*

$$(2.8)\qquad \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(|Y_{nk}-1| \geqq \varepsilon) \xrightarrow{\mathsf{P}} 0, \quad \textit{for all} \quad \varepsilon > 0.$$

Then S_{nN_n} weakly converges to the Poisson distribution with parameter $\lambda = 2D$.

COROLLARY 3. *Let (X_{nk}) and $\{N_n\}$ be as in Lemma 1, and assume that* (2.2), (2.4) *and* (2.6) *are satisfied, and that*

$$(2.9)\qquad \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(|Y_{nk}| \geqq \varepsilon) \xrightarrow{\mathsf{P}} 0, \quad \textit{for all} \quad \varepsilon > 0.$$

Then S_{nN_n} weakly converges to the normal distribution $\mathcal{N}(\gamma, D)$ with mean γ and variance D.

It is easy to verify that (2.9) is equivalent to

$$(2.10)\qquad \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|Y_{nk}| \geqq \varepsilon) \xrightarrow{\mathsf{P}} 0, \quad \text{for all } \varepsilon > 0.$$

Now we see that in the special case, when $\{N_n\}$ is a sequence of constants and $\tau_{nk}=0$, Corollary 3 reduces to Theorem 2.3 of [3].

From Corollary 3 one can deduce the following random version of the "normal convergence criterion" (cf. [6], p. 316).

COROLLARY 4. *Let (X_{nk}) and $\{N_n\}$ be as in Lemma 1, and assume that*

$$(2.11)\qquad \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|X_{nk}| \geqq \varepsilon) \xrightarrow{\mathsf{P}} 0, \quad \textit{for all} \quad \varepsilon > 0,$$

$$(2.12)\qquad \sum_{k=1}^{N_n} \mathsf{E}_{k-1} X_{nk} I(|X_{nk}| \leqq \tau) \xrightarrow{\mathsf{P}} \gamma,$$

$$(2.13)\qquad \sum_{k=1}^{N_n} \{\mathsf{E}_{k-1} X_{nk}^2 I(|X_{nk}| \leqq \tau) - \mathsf{E}_{k-1}^2 X_{nk} I(|X_{nk}| \leqq \tau)\} \xrightarrow{\mathsf{P}} D,$$

where $\tau > 0$ is finite and arbitrarily fixed. Then $S_{nN_n} \xrightarrow{\mathcal{D}} \mathcal{N}(\gamma, D)$.

In the special case, when $\{N_n\}$ is a sequence of stopping times, Corollary 4 reduces to Theorem 2.1 of [4].

3. In order to prove Theorem 1 we need the following auxiliary results.

LEMMA 2. (2.1) *holds if and only if*

$$d_{nN_n} = \max_{1 \le k \le N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} \xrightarrow{\mathsf{P}} 0. \tag{3.1}$$

Furthermore, (2.1) *implies that*

$$b_{nN_n} = \max_{1 \le k \le N_n} \left| \mathsf{E}_{k-1} \frac{Y_{nk}}{1+Y_{nk}^2} \right| \xrightarrow{\mathsf{P}} 0. \tag{3.2}$$

PROOF. Obvious.

PROPOSITION 1. *Let* (X_{nk}) *be a DIA of r.vs adapted to an array* $(\mathscr{F}_{nk})$ *of row-wise increasing sub-σ-fields, and let* $\{N_n\}$ *be a sequence of positive integer-valued r.vs satisfying* (2.1), (2.2), (2.3) *and* (2.4). *Then* (1.2) *holds.*

PROOF. For every t let us define

$$h_t(x) = \begin{cases} \left(\exp(itx) - 1 - \dfrac{itx}{1+x^2}\right) \dfrac{1+x^2}{x^2} & \text{for} \quad x \ne 0, \\ \dfrac{1}{2} t^2 & \text{for} \quad x = 0. \end{cases}$$

The function h_t is continuous and bounded (i.e. there exists a constant $L = L_t > 0$ such that $|h_t(x)| \le L_t$ for all x). Thus,

$$\begin{aligned} &\max_{1 \le k \le N_n} |\mathsf{E}_{k-1} \exp(itY_{nk}) - 1| = \\ &= \max_{1 \le k \le N_n} \left| it\mathsf{E}_{k-1} \frac{Y_{nk}}{1+Y_{nk}^2} + \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} \right| \le |t| b_{nN_n} + L_t d_{nN_n} \xrightarrow{\mathsf{P}} 0, \end{aligned} \tag{3.3}$$

as (3.1) and (3.2) hold.

Furthermore, from the Taylor series it is seen that

$$\sum_{k=1}^{N_n} |\log(1+z_{nk}) - z_{nk}| \le \sum_{k=1}^{N_n} |z_{nk}|^2.$$

This with $z_{nk} = \mathsf{E}_{k-1} \exp(itY_{nk}) - 1$, by (3.3), yields for fixed t

$$\begin{aligned} |A_n^t| &= \left| \sum_{k=1}^{N_n} \log \mathsf{E}_{k-1} \exp(itY_{nk}) - \sum_{k=1}^{N_n} (\mathsf{E}_{k-1} \exp(itY_{nk})) \right| \le \\ &\le \sum_{k=1}^{N_n} |\mathsf{E}_{k-1} \exp(itY_{nk}) - 1|^2 \le t^2 \sum_{k=1}^{N_n} \left(\mathsf{E}_{k-1} \frac{Y_{nk}}{1+Y_{nk}^2} \right)^2 + \\ &\quad + (2|t| L_t b_{nN_n} + L_t^2 d_{nN_n}) \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} \xrightarrow{\mathsf{P}} 0, \end{aligned}$$

as (2.1), (2.2) and (2.3) with $y = +\infty$ hold.

Since

$$\log f_{nN_n}(t) = it \sum_{k=1}^{N_n} \left\{ a_{nk} + \mathsf{E}_{k-1} \frac{Y_{nk}}{1+Y_{nk}^2} \right\} + \sum_{k=1}^{N_n} \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} + A_n^t,$$

then for the proof of (1.2) it is sufficient to show that

$$R_n^t = \sum_{k=1}^{N_n} \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} \xrightarrow{\mathsf{P}} \int_{-\infty}^{+\infty} h_t(x)\, dK(x). \tag{3.4}$$

For arbitrary positive ε_1, ε_2 and ε_3, choose an integer m sufficiently large and subdivision $x_0 < x_1 < \ldots < x_m$, all continuity points of K, so that

$$|B(m)| = \left| \sum_{j=1}^{m} h_t(x_{j-1})\big(K(x_j) - K(x_{j-1})\big) - \int_{-\infty}^{+\infty} h_t(x)\, dK(x) \right| < \varepsilon_1,$$

$$\max_{1 \le j \le m} |x_j - x_{j-1}| < \varepsilon_2, \qquad |h_t(x_j) - h_t(x_{j-1})| < \varepsilon_3,$$

and

$$K(x_0) + K(+\infty) - K(x_m) < \varepsilon_3. \tag{3.5}$$

(We recall that K is a bounded and nondecreasing function such that $K(-\infty)=0$, and that $x_0 \to -\infty$ and $x_m \to +\infty$ as $m \to \infty$.) Then, it follows that

$$|C_n(m)| = \left| \sum_{k=1}^{N_n} \sum_{j=1}^{m} \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} I(x_{j-1} < Y_{nk} \le x_j) - \right.$$

$$\left. - \sum_{k=1}^{N_n} \sum_{j=1}^{m} h_t(x_{j-1}) \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(x_{j-1} < Y_{nk} \le x_j) \right| \le \varepsilon_3 \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2},$$

$$|D_n(m)| = \left| R_n^t - \sum_{k=1}^{N_n} \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} I(x_0 < Y_{nk} \le x_m) \right| \le$$

$$\le L_t \left(\sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(Y_{nk} \le x_0) + \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} - \right.$$

$$\left. - \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(Y_{nk} \le x_m) \right),$$

where the last factor on the majorant side converges in probability to $K(x_0)+ +K(+\infty) - K(x_m)$, which is less than ε_3 by (3.5).

Thus, by the relations given above, we obtain

$$R_n^t = \sum_{k=1}^{N_n} \mathsf{E}_{k-1} h_t(Y_{nk}) \frac{Y_{nk}^2}{1+Y_{nk}^2} = \int_{-\infty}^{+\infty} h_t(x)\, dK(x) + B(m) +$$

$$+ \sum_{j=1}^{m} h_t(x_{j-1}) \left\{ \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} I(x_{j-1} < Y_{nk} \le x_j) - \big(K(x_j) - K(x_{j-1})\big) \right\} +$$

$$+ C_n(m) + D_n(m),$$

which, by (2.3), gives the desired result (3.4).

The proof of Theorem 1 is easily based on Proposition 1 and Lemma 1 and is not detailed here.

In order to prove Theorem 2 we need the following auxiliary results.

LEMMA 3. *Under* (2.1) *and* (2.3) *with* $y=+\infty$ (2.2) *is equivalent to*

$$\sum_{k=1}^{N_n} |\mathsf{E}_{k-1} \exp(itY_{nk})-1|^2 \xrightarrow{\mathsf{P}} 0. \tag{3.6}$$

PROOF. From the proof of Proposition 1 it follows that (2.2) implies (3.6). The reverse implication can be proved similarly.

LEMMA 4. *Suppose that* (2.5) *holds, and that* $\tau_{nk}=\tau$, *where* τ *is a positive constant. Then*

$$\max_{1 \le k \le N_n} |a_{nk}| \xrightarrow{\mathsf{P}} 0 \tag{3.7}$$

and there is a constant M depending only on τ *and* t *such that for every* k *and all* n

$$|\mathsf{E}_{k-1} \exp(itY_{nk})-1| \le M\mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2}. \tag{3.8}$$

Furthermore, in this case (2.1) *holds.*

PROOF. From (2.5) it follows that for every $0<\varepsilon<\tau$

$$\max_{1 \le k \le N_n} |a_{nk}| < \varepsilon+\tau \max_{1 \le k \le N_n} \mathsf{P}_{k-1}(|X_{nk}| \ge \varepsilon) \xrightarrow{\mathsf{P}} \varepsilon,$$

which states that (2.5) implies (3.7). Furthermore, one can see that (2.5) combined with (3.7) implies (2.1). Thus, for the proof of Lemma 4 we wish to show that (3.8) holds. To this end we note that by a routine technique of subsequences we can assume that

$$|a_{nk}| \le \frac{\tau}{2} \quad \text{for every } k \text{ and all } n. \tag{3.9}$$

Obviously, there is no loss of generality, because if (X_{nk}) does not satisfy (3.9), then we can set $\tilde{X}_{nk}=X_{nk} I\left(|a_{nk}| \le \frac{\tau}{2}\right)$. Then $(\tilde{X}_{nk})$ will form a DIA of r.vs adapted to $(\mathscr{F}_{nk})$ and, by (3.7),

$$\mathsf{P}\Big(\bigcup_{k=1}^{N_n} \{X_{nk} \ne \tilde{X}_{nk}\}\Big) \le \mathsf{P}\left(\max_{1 \le k \le N_n} |a_{nk}| > \frac{\tau}{2}\right) \to 0$$

and

$$\mathsf{P}\Big(\bigcup_{k=1}^{N_n} \{f_{nk}(t) \ne \prod_{j=1}^{k} \mathsf{E}_{j-1} \exp(it\tilde{X}_{nj})\}\Big) \le \mathsf{P}\left(\max_{1 \le k \le N_n} |a_{nk}| > \frac{\tau}{2}\right) \to 0.$$

Furthermore, (2.5) holds with X_{nk} replaced by $\tilde{X}_{nk}$, and

$$\mathsf{E}_{k-1}\tilde{X}_{nk}I(|\tilde{X}_{nk}|\leqq\tau)=a_{nk}I\left(|a_{nk}|\leqq\frac{\tau}{2}\right):=\tilde{a}_{nk},$$

$$\tilde{Y}_{nk}:=\tilde{X}_{nk}-\tilde{a}_{nk}=Y_{nk}I\left(|a_{nk}|\leqq\frac{\tau}{2}\right),$$

$$\mathsf{E}_{k-1}\frac{\tilde{Y}_{nk}}{1+\tilde{Y}_{nk}^2}=\mathsf{E}_{k-1}\frac{Y_{nk}}{1+Y_{nk}^2}I\left(|a_{nk}|\leqq\frac{\tau}{2}\right),$$

$$\mathsf{E}_{k-1}\frac{\tilde{Y}_{nk}^2}{1+\tilde{Y}_{nk}^2}=\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}I\left(|a_{nk}|\leqq\frac{\tau}{2}\right),$$

$$\mathsf{E}_{k-1}\exp(it\tilde{Y}_{nk})-1=\left(\mathsf{E}_{k-1}\exp(itY_{nk})-1\right)I\left(|a_{nk}|\leqq\frac{\tau}{2}\right),$$

from which we conclude that (2.1)—(2.4) and (3.6)—(3.8) hold with $\tilde{Y}_{nk}$ replacing Y_{nk}, enabling use to be made of property (3.9). Alternatively, we will prove Lemma 4 (and later on Theorem 2 and Corollary 4) as it stands and assume also that (3.9) holds.

We see that

$$|\mathsf{E}_{k-1}\exp(itY_{nk})-1|\leqq\left|\mathsf{E}_{k-1}\{\exp(itY_{nk})-1-itY_{nk}\}I(|Y_{nk}|\leqq\tau)\right|+$$

$$+\left|\mathsf{E}_{k-1}\{\exp(itY_{nk})-1\}I(|Y_{nk}|>\tau)\right|+\left|t\mathsf{E}_{k-1}Y_{nk}I(|Y_{nk}|\leqq\tau)\right|\leqq$$

$$\leqq\frac{1}{2}t^2\mathsf{E}_{k-1}Y_{nk}^2I(|Y_{nk}|\leqq\tau)+2\mathsf{P}_{k-1}(|Y_{nk}|>\tau)+$$

$$+\left|t\mathsf{E}_{k-1}Y_{nk}\{I(|Y_{nk}|\leqq\tau)-I(|X_{nk}|\leqq\tau)\}\right|+\left|t\mathsf{E}_{k-1}Y_{nk}I(|X_{nk}|\leqq\tau)\right|,$$

where the first and the second terms are less than

$$\frac{1}{2}t^2(1+\tau^2)\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}\quad\text{and}\quad 2\frac{1+\tau^2}{\tau^2}\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2},$$

respectively. Moreover, we have

$$\begin{aligned}&\left|\mathsf{E}_{k-1}Y_{nk}I(|X_{nk}|\leqq\tau)\right|=\left|a_{nk}-a_{nk}\mathsf{P}_{k-1}(|X_{nk}|\leqq\tau)\right|=\\ &=|a_{nk}|\mathsf{P}_{k-1}(|X_{nk}|>\tau)\leqq\frac{\tau}{2}\mathsf{P}_{k-1}\left(|Y_{nk}|\geqq\frac{\tau}{2}\right)\end{aligned}\tag{3.10}$$

and

$$\begin{aligned}&\left|\mathsf{E}_{k-1}Y_{nk}\{I(|Y_{nk}|\leqq\tau)-I(|X_{nk}|\leqq\tau)\}\right|\leqq\\ &\leqq\mathsf{E}_{k-1}|Y_{nk}|I\left(\frac{\tau}{2}\leqq|Y_{nk}|\leqq\frac{3\tau}{2}\right)\leqq\frac{3\tau}{2}\mathsf{P}_{k-1}\left(|Y_{nk}|\geqq\frac{\tau}{2}\right),\end{aligned}\tag{3.11}$$

as (3.9) holds. Hence,

$$|\mathsf{E}_{k-1}\exp(itY_{nk})-1| \leqq$$

$$\leqq \left\{\frac{1}{2}t^2(1+\tau^2)+2(1+\tau|t|)\frac{1+\left(\frac{\tau}{2}\right)^2}{\left(\frac{\tau}{2}\right)^2}\right\}\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}$$

which gives (3.8).

Proof of Theorem 2. We need to show that the assumptions of Theorem 2 imply (1.2). To this end it is sufficient to prove that (2.1) and (2.2) are satisfied.

From Lemmas 2 and 4 it follows that (2.1) holds, and that

$$\sum_{k=1}^{N_n}|\mathsf{E}_{k-1}\exp(itY_{nk})-1|^2 \leqq M^2 d_{nN_n}\sum_{k=1}^{N_n}\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}\xrightarrow{\mathsf{P}} 0,$$

i.e. (3.6) is satisfied. Hence, and by Lemma 3, we get (2.2).

Corollaries 1, 2 and 3 easily follow from Theorem 1 (see also [7], p. 93). In order to prove Corollary 4 we need the following auxiliary results.

Lemma 5. (2.11) *implies* (2.10).

Proof. By $\mathscr{F}_{n,k-1}$-measurability of a_{nk} we obtain for every $\varepsilon>0$

$$\sum_{k=1}^{N_n}\mathsf{P}_{k-1}(|Y_{nk}|\geqq\varepsilon)\leqq\sum_{k=1}^{N_n}\mathsf{P}_{k-1}\left(|X_{nk}|\geqq\frac{\varepsilon}{2}\right)+\sum_{k=1}^{N_n}I\left(|a_{nk}|\geqq\frac{\varepsilon}{2}\right).$$

But

$$\mathsf{P}\left(\sum_{k=1}^{N_n}I\left(|a_{nk}|\geqq\frac{\varepsilon}{2}\right)\geqq\varepsilon\right)=\mathsf{P}\left(\max_{1\leqq k\leqq N_n}|a_{nk}|\geqq\frac{\varepsilon}{2}\right)\to 0$$

as (3.7) holds. Hence, we get the desired implication.

Lemma 6. (2.6) *and* (2.11) *imply that*

$$\sum_{k=1}^{N_n}\mathsf{E}_{k-1}\frac{Y_{nk}}{1+Y_{nk}^2}\xrightarrow{\mathsf{P}} 0. \tag{3.12}$$

Proof. We note that for every $0<\varepsilon<\tau$

$$\left|\mathsf{E}_{k-1}Y_{nk}I(|Y_{nk}|\leqq\tau)-\mathsf{E}_{k-1}\frac{Y_{nk}}{1+Y_{nk}^2}\right|=$$

$$=\left|\mathsf{E}_{k-1}\frac{Y_{nk}^3 I(|Y_{nk}|\leqq\tau)}{1+Y_{nk}^2}+\mathsf{E}_{k-1}\frac{Y_{nk}I(|Y_{nk}|>\tau)}{1+Y_{nk}^2}\right|\leqq$$

$$\leqq\varepsilon\mathsf{E}_{k-1}\frac{Y_{nk}^2}{1+Y_{nk}^2}+(\tau+1)\mathsf{P}_{k-1}(|Y_{nk}|\geqq\varepsilon).$$

Hence, using (3.10) and (3.11), we get for every $0<\varepsilon<\frac{\tau}{2}$

$$\left|\sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}}{1+Y_{nk}^2}\right| \leqq$$

$$\leqq \varepsilon \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} + \left(1+\tau+\frac{\tau}{2}+\frac{3\tau}{2}\right) \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|Y_{nk}| \geqq \varepsilon),$$

which by (2.6) and (2.11) gives (3.12).

PROOF OF COROLLARY 4. Taking into account Corollary 3 and Lemmas 5 and 6 we only need to show that (2.6) is satisfied.

First we note that for every k and all n

$$\mathsf{E}_{k-1} Y_{nk}^2 I(|X_{nk}| \leqq \tau) = \mathrm{Var}_{k-1} X_{nk} I(|X_{nk}| \leqq \tau) - a_{nk}^2 \mathsf{P}_{k-1}(|X_{nk}| > \tau),$$

where $\mathrm{Var}_{k-1} Z_{nk} = \mathsf{E}_{k-1} Z_{nk}^2 - \mathsf{E}_{k-1}^2 Z_{nk}$. Since by (3.9) and (2.11)

$$\sum_{k=1}^{N_n} a_{nk}^2 \mathsf{P}_{k-1}(|X_{nk}| > \tau) \leqq \left(\frac{\tau}{2}\right)^2 \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|X_{nk}| \geqq \tau) \xrightarrow{\mathsf{P}} 0,$$

then

$$\sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|X_{nk}| \leqq \tau) \xrightarrow{\mathsf{P}} D, \tag{3.13}$$

as (2.13) holds. But (cf. 3.11))

$$\left|\mathsf{E}_{k-1} Y_{nk}^2 \{I(|Y_{nk}| \leqq \tau) - I(|X_{nk}| \leqq \tau)\}\right| \leqq \left(\frac{3\tau}{2}\right)^2 \mathsf{P}_{k-1}\left(|Y_{nk}| \geqq \frac{\tau}{2}\right).$$

Hence, and by (2.11), we get

$$\sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \tau) \xrightarrow{\mathsf{P}} D. \tag{3.14}$$

Moreover, by (2.11), we have for every $\varepsilon<\tau$

$$\left|\sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \tau) - \sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \varepsilon)\right| \leqq \tau^2 \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|Y_{nk}| \geqq \varepsilon) \xrightarrow{\mathsf{P}} 0,$$

and the same is true for every $\varepsilon>\tau$; suffices to interchange ε and τ on the majorant side. Thus, using (2.11) and (3.14), it follows that

$$\sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \varepsilon) \xrightarrow{\mathsf{P}} D, \quad \text{for all} \quad \varepsilon > 0. \tag{3.15}$$

Since for every $\varepsilon>0$

$$\frac{1}{1+\varepsilon^2} \sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \varepsilon) \leqq \sum_{k=1}^{N_n} \mathsf{E}_{k-1} \frac{Y_{nk}^2}{1+Y_{nk}^2} \leqq$$

$$\leqq \sum_{k=1}^{N_n} \mathsf{E}_{k-1} Y_{nk}^2 I(|Y_{nk}| \leqq \varepsilon) + \sum_{k=1}^{N_n} \mathsf{P}_{k-1}(|Y_{nk}| \geqq \varepsilon),$$

then (2.11) and (3.15) imply (2.6).

ACKNOWLEDGEMENTS. I am indebted to Professor Dominik Szynal for having introduced me to this subject and for his constant encouragement and useful advice. Also, I am grateful to Professor Zdzisław Rychlik for helpful discussion.

REFERENCES

[1] BEŚKA, M., KŁOPOTOWSKI, A. and SŁOMIŃSKI, L., Limit theorems for random sums of dependent d-dimensional random vectors, *Z. Wahrsch. Verw. Gebiete* **61** (1982), 43—57. *MR* **84c:** 60034.

[2] CSÖRGŐ, M. and RYCHLIK, Z., Asymptotic properties of randomly indexed sequences of random variables, *Canad. J. Statist.* **9** (1981), 101—107. *MR* **83b:** 60017.

[3] DVORETZKY, A., Asymptotic normality for sums of dependent random variables, *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** (1972), 513—535. *MR* **54** #3808.

[4] HELLAND, I. S., Central limit theorems for martingales with discrete or continuous time, *Scand. J. Statist.* **9** (1982), 79—94. *MR* **84a:** 60036.

[5] KUBACKI, K. S. and SZYNAL, D., Weak convergence of martingales with random indices to infinitely divisible laws, *Acta Math. Hung.* **42** (1983), 143—151. *MR* **85f:** 54026.

[6] LOÈVE, M., *Probability theory*, 2nd ed., Van Nostrand, Princeton, 1960. *MR* **23** #A670.

[7] LUKACS, E., *Characteristic functions*, Griffin, London, 1960. *MR* **23** #A1392.

[8] ROSIŃSKI, J., Limit theorems for randomly indexed sums of random vectors, *Colloq. Math.* **34** (1975), 91—107. *MR* **53** #9320.

[9] RYCHLIK, Z., A central limit theorem for martingales, *Litovsk. Mat. Sb.* **18** (1978), 139—145. *MR* **81j:** 60033.

[10] RYCHLIK, Z., Martingale random central limit theorems, *Acta Math. Acad. Sci. Hungar.* **34** (1979), 129—139. *MR* **80k:** 60061.

[11] SZÁSZ, D., Limit theorems for the distributions of the sums of a random number of random variables, *Ann. Math. Statist.* **43** (1972), 1902—1913. *MR* **50** #11399.

(Received May 24, 1983)

INSTYTUT ZASTOSOWAŃ MATEMATYKI
WYDZIALU TECHNIKI ROLNICZEJ AR
UL. AKADEMICKA 13
PL—20-934 LUBLIN
POLAND

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 287—297

ON APPROXIMATION BY BERNSTEIN-TYPE OPERATORS: BEST CONSTANTS

HEINZ H. GONSKA and JUTTA MEIER

1. Introduction and results

In recent years the problem of determining best constants in quantitative convergence theorems for Bernstein operators has attracted more and more interest, the most striking results being due to the work of several Dutch mathematicians (cf. our list of references). The most interesting problems in this field were those of giving the exact value of the two quantities

$$c := \sup_{n\in N} \sup_{\substack{f\in C[0,1]\\ f\not\equiv \text{constant}}} \frac{\|B_n f - f\|_\infty}{\omega_1(f, n^{-1/2})}$$

and

$$d := \sup_{n\in N} \sup_{\substack{f\in C^1[0,1]\\ f\neq \text{linear}}} \frac{n^{1/2}\|B_n f - f\|_\infty}{\omega_1(f', n^{-1/2})},$$

where for $n\in\mathbf{N}$

$$B_n\colon C[0,1]\ni f\mapsto \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} e_1^k(e_0-e_1)^{n-k}\in \Pi_n[0,1]$$

is the n-th Bernstein operator, where $\omega_1(f,\cdot)$ is the first order modulus of continuity of a function $f\in C[0,1]$, and where $\omega_1(f',\cdot)$ denotes that of the first derivative f' of a $C^1[0,1]$-function f. Moreover, e_i is defined by $e_i(x):=x^i$ for all $x\in[0,1]$ and $i=0,1$.

The above mentioned problems were solved by P. C. Sikkema [24], [25], who proved that c is equal to $\dfrac{4306+837\sqrt{6}}{5832}\approx 1.089887$, and by F. Schurer, F. W. Steutel [19], [20], [22], who showed that $d=0.25$.

Similar but somewhat weaker results, with respect to the constants c and d, can be obtained from general quantitative convergence theorems for positive linear operators, as shown by R. G. Mamedow [10], O. Shisha, B. Mond [23], R. A. DeVore [1], [2], B. Mond [13] and one of the authors [5], [6] among others.

For example the application of B. Mond's result to the sequence of Bernstein operators gives the inequality

$$\|B_n f - f\|_\infty \leqq \frac{5}{4}\omega_1(f, n^{-1/2}) \quad \text{for any} \quad f\in C[0,1];$$

1980 *Mathematics Subject Classification.* Primary 41A44, 41A36; Secondary 41A25, 41A10.

Key words and phrases. Bernstein-type operators, best constants, concave modulus of continuity.

moreover, the combination of Theorem 2.1 and Corollary 2.3 in [5] yields the estimate

$$\|B_n f-f\|_\infty \leqq \frac{5}{8} n^{-1/2}\omega_1(f', n^{-1/2}) \quad \text{for all} \quad f\in C^1[0, 1].$$

In view of the above discussion it is desirable to have quantitative assertions for general positive operators L of (at least) the Bernstein type (i.e. operators L, mapping $C[a, b]$ into $C[a, b]$ and satisfying the additional conditions $Le_i=e_i$ for $i=0, 1$) which at the same time give best constants. It is the aim of the present paper to prove such theorems in order to bridge the gap between special results for Bernstein operators and general theorems for approximation by positive linear mappings. This will be achieved by using the least concave majorants of $\omega_1(f, \cdot)$ and $\omega_1(f', \cdot)$ (denoted by $\omega_1^*(f, \cdot)$ and $\omega_1^*(f', \cdot)$, respectively) and a certain K-functional $\Omega(f, \ldots)$ which have already been employed for similar purposes by J. Peetre [16] and one of the authors [5].

The use of these tools leads to simple proofs and the values of the best constants are determined with the aid of functions playing a crucial role in approximation by positive linear operators in general. So it once more becomes apparent what an important role the least concave majorant of the modulus of continuity of a function plays in statements about the magnitude of the approximation.

2. Basic estimates

Suppose ω: $[0, \infty)\to[0, \infty)$ is a modulus of continuity, namely a function satisfying the properties (see e.g. G. G. Lorentz [9]) a) $\omega(h)\to 0$ for $h\to 0$, b) ω is increasing, c) ω is subadditive, i.e. $\omega(h_1+h_2)\leqq\omega(h_1)+\omega(h_2)$ for all $(h_1, h_2)\in[0, \infty)^2$.

The least concave majorant ω^* of ω is given by the equation (cf. J. Peetre [15], see also R. T. Rockafellar [17], p. 36)

$$\omega^*(t) = \sup\Big\{\sum_{i=1}^{n} \lambda_i\omega(t_i): n\in\mathbf{N}, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad \sum_{i=1}^{n} \lambda_i t_i = t, \quad \lambda_i \geqq 0\Big\}, \quad t\in[0, \infty).$$

Both functions are related by the inequalities (see J. Peetre [14])

$$\omega(t) \leqq \omega^*(t) \leqq 2\omega(t), \qquad t \geqq 0.$$

If f: $C[a, b]\to\mathbf{R}$ is a continuous function, the modulus of continuity of f given by

$$\omega_1(f, \cdot)\colon [0, \infty)\ni h \mapsto \omega_1(f, h) = \sup_{\substack{|\delta|\leqq h\\ x, x+h\in[a, b]}} \{|f(x)-f(x+\delta)|\}$$

is a modulus of continuity as defined above. Its least concave majorant $\omega_1^*(f, \cdot)$ will be very important in what follows.

We now sketch briefly how the functional Ω: $C[a, b]\times\mathbf{R}_+^3\to\mathbf{R}_+$ mentioned above is constructed. Let E denote a real vector space, U a subspace of E, and p and $\bar{p}$ seminorms on E and U, respectively. We define $\tilde{K}$: $\mathbf{R}_+^2\times E\to\mathbf{R}_+$ by

$$\tilde{K}(t_1, t_2, f; (E, p), (U, \bar{p})) := \inf\{p(f-g)+t_1 p(g)+t_2\bar{p}(g)\colon\ g\in U\},$$

and K: $\mathbf{R}_+ \times E \to \mathbf{R}_+$ by

$$K(t,f;(E,p),(U,\bar{p})) := \inf\{p(f-g)+t\bar{p}(g):\ g\in U\}.$$

We write for the sake of simplicity $\tilde{K}(t_1, t_2, f)$ and $K(t,f)$, respectively, if it is clear what (E,p) and $(U,\bar{p})$ are. It is readily verified that for fixed (t_1, t_2) the functional $\tilde{K}(t_1, t_2, \cdot)$ is a seminorm on E; thus it makes sense to use it as the seminorm $\bar{p}$ when defining $K(t, \cdot)$.

We now consider the spaces $C^i[a,b]$, $i\in\{1,2\}$, of i-times continuously differentiable functions defined on the finite interval $[a,b]$ with the seminorms $\|\cdot^{(i)}\|_\infty$, where $\cdot^{(i)}$ denotes i-fold differentiation. Then, for fixed t_1, t_2, the seminorm $\tilde{K}(t_1, t_2, f;\ (C^1[a,b], \|\cdot^{(1)}\|_\infty), (C^2[a,b], \|\cdot^{(2)}\|_\infty))$ may be used in the definition of

$$\Omega(f;t,t_1,t_2) := K(t,f;(C[a,b],\|\cdot\|_\infty),\ (C^1[a,b],\ \tilde{K}(t_1,t_2,\cdot))),\quad t\ge 0,\quad f\in C[a,b].$$

The following assertion is a partial improvement of an estimate contained in [6].

THEOREM 2.1. *Let Ω be defined as above. Then the following inequalities are true for any* $(f;t,t_1,t_2)\in C[a,b]\times\mathbf{R}_+^3$:

$$\Omega(f;\ t,t_1,t_2) \le \begin{cases} \dfrac{1}{2}\omega_1^*(f,2t), & \text{(i)}\\[2mm] t\left\{\min(1,\ t_1)\cdot\|f'\|_\infty+(1+t_1)\chi_{[0,1]}(t_1)\dfrac{1}{2}\omega_1^*\left(f',\dfrac{2t_2}{1+t_1}\right)\right\}, & \text{(ii)}\end{cases}$$

where $\chi_{[0,1]}$ denotes the characteristic function of $[0,1]$, *$\omega_1^*(f,\cdot)$ is the least concave majorant of $\omega_1(f,\cdot)$ and* (ii) *is valid for $f\in C^1[a,b]$ only.*

PROOF. (i) Putting $f_2=0$ ($f_2\in C^2[a,b]$), the definition of Ω gives immediately

$$\Omega(f;t,t_1,t_2)\le\inf\{\|f-f_1\|_\infty+t\|f_1'\|_\infty:\ f_1\in C^1[a,b]\} =$$
$$= K(t,f;(C[a,b],\|\cdot\|_\infty),\ (C^1[a,b],\|\cdot^{(1)}\|_\infty)),$$

the latter functional being equal to $\dfrac{1}{2}\omega_1^*(f,2t)$ (see J. Peetre [14], B. S. Mitjagin, E. M. Semenov [12]).

(ii) If f is continuously differentiable, then the definition of Ω, being a special K-functional, gives

$$\Omega(f;t,t_1,t_2)\le t\tilde{K}(t_1,\ t_2,f).$$

Several properties of functionals $\tilde{K}$ were investigated in [4]. In particular, it was shown that two functionals $\tilde{K}$ and K constructed with the aid of the same pairs (E,p) and $(U,\bar{p})$ are related by the inequality (see [4], Lemma 1.2, (ii))

$$\tilde{K}(t_1,t_2,f;\ (E,p),(U,\bar{p}))\le$$
$$\le\min(1,t_1)\cdot p(f)+(1+t_1)\chi_{[0,1]}(t_1)K\left(\frac{t_2}{1+t_1},f;\ (E,p),(U,\bar{p})\right),\quad t_1,t_2\ge 0.$$

Putting $(E, p)=(C^1[a, b], \|\cdot^{(1)}\|_\infty)$, $(U, \bar{p})=(C^2[a, b], \|\cdot^{(2)}\|_\infty)$ for the moment, we obtain

$$t\tilde{K}(t_1, t_2, f) \leqq t\left\{\min(1, t_1)\cdot\|f'\|_\infty+(1+t_1)\chi_{[0,1]}(t_1)K\left(\frac{t_2}{1+t_1}, f\right)\right\}.$$

As in part (i) of this proof, we have

$$K\left(\frac{t_2}{1+t_1}, f;\ (C^1[a, b], \|\cdot^{(1)}\|_\infty, (C^2[a, b], \|\cdot^{(2)}\|_\infty))\right) = \frac{1}{2}\,\omega_1^*\left(f', \frac{2t_2}{1+t_1}\right).$$

Inserting this in the above estimates we arrive at assertion (ii) in Theorem 2.1. ∎

COROLLARY 2.2 (H. Gonska [6]). *If Ω is defined as above and $(f; t, t_1, t_2)\in C[a, b]\times\times\mathbf{R}_+^3$, then we have*

$$\begin{array}{ll}\text{(i)} & \\ & \Omega(f;\ t, t_1, t_2) \leqq \begin{cases}\dfrac{1}{2}\,\omega_1^*(f, 2t), \\ t\, t_1\|f'\|_\infty+t\,\omega_1^*(f', 2t_2),\end{cases} \\ \text{(ii)} & \end{array}$$

(ii) *being valid for $f\in C^1[a, b]$ only.*

This is an immediate consequence of $\min(1, t_1)\leqq t_1$, $(1+t_1)\chi_{[0,1]}(t_1)\leqq 2$ and $\dfrac{1}{1+t_1}\leqq 1$.

The following theorem which will be needed below was proved in [5].

THEOREM 2.3 (H. Gonska [5]). *Let $L\colon C[a, b]\to C[a, b]$ be a positive linear operator satisfying $Le_0=e_0$. Then for any $f\in C[a, b]$ and all $x\in[a, b]$ the estimate*

$$|L(f, x)-f(x)| \leqq 2\Omega\left(f;\ \frac{L(|e_1-x|, x)}{2},\ \frac{|L(e_1-x, x)|}{L(|e_1-x|, x)},\ \frac{1}{2}\,\frac{L((e_1-x)^2, x)}{L(|e_1-x|, x)}\right)$$

holds. The estimate remains true if one or more of the three "differences" occurring on the right side are replaced by majorants such that the quotients remain finite.

Thus, by combining Theorems 2.1 and 2.3, it is easy to arrive at estimates for $|L(f, x)-f(x)|$ containing $\omega_1^*(f, \cdot)$ and $\omega_1^*(f', \cdot)$, respectively. The great advantage of this combination is due to the fact that in both cases we obtain optimal constants, as will be seen below.

3. Optimal constants in the approximation of arbitrary continuous functions

THEOREM 3.1. *Let $L\colon C[a, b]\to C[a, b]$ be a positive linear operator satisfying $Le_0=e_0$. Then we have for any $f\in C[a, b]$ and all $x\in[a, b]$:*

$$|L(f, x)-f(x)| \leqq \omega_1^*(f, L(|e_1-x|, x)).$$

If the function $f\in C[a, b]$ possesses a concave modulus of continuity $\omega_1(f, \cdot)$, we have

$$|L(f, x)-f(x)| \leqq \omega_1(f, L(|e_1-x|, x)).$$

The proof is an immediate consequence of Corollary 2.2 and Theorem 2.3.

In the case of Bernstein polynomials $(B_n)_{n\geq 1}$ the paper of F. Schurer, F. W. Steutel [19] gives more information on the majorant occurring on the right side in the above estimate.

EXAMPLE 3.2. For the n-th Bernstein operator B_n, for all $f\in C[0, 1]$ and every $x\in[0, 1]$ the following inequality holds:

$$|B_n(f, x)-f(x)| \leq \omega_1^*\left(f, \frac{2}{n}(n-r)\binom{n}{r}x^{r+1}(1-x)^{n-r}\right),$$

where $r:=[nx]$ denotes the largest integer not exceeding nx.

PROOF. F. Schurer and F. W. Steutel proved the equality

$$B_n(|e_1-x|, x) = \frac{2}{n}(n-r)\binom{n}{r}x^{r+1}(1-x)^{n-r}$$

where r is defined as above. ∎

The following considerations will show that the constant 1 occurring on the right side of the estimate of Theorem 3.1 cannot be replaced by a constant c, $c<1$.

THEOREM 3.3. *Let* L: $C[a, b]\to C[a, b]$ *be a positive linear operator satisfying* $Le_0=e_0$. *Then for every* $x\in[a, b]$ *there exists a function* $f_x\in C[a, b]$ *so that the equality*

$$|L(f_x, x)-f_x(x)| = \omega_1^*(f_x, L(|e_1-x|, x))$$

is fulfilled.

PROOF. We put $f_x(t):=|t-x|$ for $a\leq t\leq b$. Then we have for the modulus of continuity $\omega_1(f_x, \cdot)$ of f_x:

$$\omega_1(f_x, h) = \min\{h, \max\{b-x, x-a\}\} \quad \text{for each} \quad h\in[0, b-a].$$

Thus, $\omega_1(f_x, \cdot)$ is a concave function. Furthermore we have

$$|L(f_x, x)-f_x(x)| = |L(f_x, x)-0| = |L(|e_1-x|, x)| =$$
$$= \omega_1(f_x, L(|e_1-x|, x)) = \omega_1^*(f_x, L(|e_1-x|, x)).$$

The next to the last equality holds, since we have, because of the positivity of L and the condition $Le_0=e_0$:

$$L(|e_1-x|, x) \leq L(\max\{b-x, x-a\}, x) = \max\{b-x, x-a\}.$$

With this the considered equality results from the above representation of $\omega_1(f_x, \cdot)$. The last equality follows from the fact that $\omega_1(f_x, \cdot)$ is concave. This completes the proof. ∎

EXAMPLE 3.4. In the estimate of Example 3.2, namely

$$|B_n(f, x)-f(x)| \leq \omega_1^*\left(f, \frac{2}{n}(n-r)\binom{n}{r}x^{r+1}(1-x)^{n-r}\right),$$

where $r=[nx]$, the constant 1 appearing before $\omega_1^*(f, \ldots)$ cannot be replaced for any $x\in(0, 1)$ and any $n\in\mathbf{N}$ by a constant $c<1$.

Obviously, the points $x=0$ and $x=1$ play a special role, since in those cases zero occurs on both sides of the inequality in Example 3.2.

Whereas we have treated in the above optimal constants for pointwise estimates, we now turn to uniform statements. For this we once more refer to the report of F. Schurer and F. W. Steutel [19], where the authors prove the following

LEMMA 3.5 (F. Schurer, F. W. Steutel [19]). *For the functions $S_n(x)$, defined for every $x\in[0, 1]$ and all $n\in\mathbf{N}$ by*

$$S_n(x):= n^{-1/2}(n-r)\binom{n}{r}x^{r+1}(1-x)^{n-r} \qquad (r=[nx])$$

the following inequalities are valid:

$$\frac{1}{4} = \|S_1\|_\infty > \|S_3\|_\infty > \|S_5\|_\infty > \ldots,$$

$$\frac{4\sqrt{2}}{27} = \|S_2\|_\infty > \|S_4\|_\infty > \|S_6\|_\infty > \ldots.$$

Making use of Lemma 3.5 the following theorem results:

THEOREM 3.6. *If $(B_n)_{n\in\mathbf{N}}$ denotes the sequence of Bernstein operators and if for $f\in C[0, 1]$ the symbol $\omega_1^*(f, \cdot)$ denotes the least concave majorant of its first order modulus of continuity, then*

$$\sup_{n\in\mathbf{N}} \sup_{\substack{f\in C[0,1]\\ f\neq\text{constant}}} \frac{\|B_n f-f\|_\infty}{\omega_1^*\left(f, \frac{1}{2}n^{-1/2}\right)} = 1.$$

PROOF. As proved in Example 3.2 we have for $n\in\mathbf{N}, f\in C[0, 1]$ and $0\leqq x\leqq 1$, $(r=[nx])$:

$$|B_n(f, x)-f(x)| \leqq \omega_1^*\left(f, \frac{2}{n}(n-r)\binom{n}{r}x^{r+1}(1-x)^{n-r}\right) =$$

$$= \omega_1^*(f, 2n^{-1/2}S_n(x)) \leqq \omega_1^*(f, 2n^{-1/2}\sup_{n\in\mathbf{N}}\|S_n\|_\infty) = \omega_1^*\left(f, \frac{1}{2}n^{-1/2}\right),$$

the latter estimate arising from Lemma 3.5.

It remains to show that there is an $n\in\mathbf{N}$ and an $f_0\in C[0, 1]$ which give equality in the above inequality. The nice thing about this proof is that we can choose $n=1$ and $f_0=\left|e_1-\frac{1}{2}\right|$.

As it was discussed in the proof of Theorem 3.3, f_0 possesses the concave modulus of continuity $\omega_1(f_0, \cdot)$ given by $\omega_1(f_0, h)=\min\left\{h, \frac{1}{2}\right\}$ for $0\leqq h\leqq 1$. Thus,

$\omega_1^*(f_0, \cdot)=\omega_1(f_0, \cdot)$. In view of the above representation of $\omega_1(f_0, \cdot)$ for $n=1$ we have $\omega_1\left(f_0, \frac{1}{2}n^{-1/2}\right)=\frac{1}{2}$.

Since the graph of $B_1 f_0$ is the straight line connecting the points $(0, f_0(0))= =\left(0, \frac{1}{2}\right)$ and $(1, f_0(1))=\left(1, \frac{1}{2}\right)$ one also has $\|B_1 f_0-f_0\|_\infty=\frac{1}{2}$. This proves the above equality. ∎

The foregoing discussion also shows the validity of the following

COROLLARY 3.7. (i) *Suppose that the function* $f\in C[0, 1]$ *possesses a concave first order modulus of continuity* $\omega_1(f, \cdot)$. *Then we have*

$$\|B_n f-f\|_\infty \leqq \omega_1\left(f, \frac{1}{2}n^{-1/2}\right).$$

(ii) *For any* $n\in\mathbf{N}, f\in C[0, 1]$ *the estimate*

$$\|B_n f-f\|_\infty \leqq 2\omega_1\left(f, \frac{1}{2}n^{-1/2}\right)$$

holds.

Obviously, these two estimates cannot be obtained by applying P. C. Sikkema's result [23], [24] which would yield

$$\|B_n f-f\|_\infty \leqq 2\frac{4306+837\sqrt{6}}{5832}\cdot\omega_1\left(f, \frac{1}{2}n^{-1/2}\right) \approx 2.179\,774\,7\omega_1\left(f, \frac{1}{2}n^{-1/2}\right).$$

4. Optimal constants in the approximation of continuously differentiable functions

In this section we shall prove estimates analogous to the ones given in Section 3. As far as pointwise assertions are concerned we have

THEOREM 4.1. *Let* L: $C[a, b]\to C[a, b]$ *be a positive linear operator satisfying* $Le_i=e_i$ $(i=0, 1)$ *where e_i is the i-th monomial, and* $x\in[a, b]$. *Then for all* $f\in C^1[a, b]$ *the estimate*

$$|L(f, x)-f(x)| \leqq \frac{1}{2}L((e_1-x)^2, x)^{1/2}\omega_1^*(f', L((e_1-x)^2, x)^{1/2}).$$

holds.

If the function $f\in C^1[a, b]$ *possesses a concave modulus of continuity* $\omega_1(f', \cdot)$ *it simplifies to*

$$|L(f, x)-f(x)| \leqq \frac{1}{2}L((e_1-x)^2, x)^{1/2}\omega_1(f', L((e_1-x)^2, x)^{1/2}).$$

PROOF. We make use of the remark terminating the formulation of Theorem 2.3, replacing the quantity $L(|e_1-x|, x)$ by its majorant $L((e_1-x)^2, x)^{1/2}$. The possibility of doing so is a consequence of the Cauchy—Schwarz inequality. Moreover, because of $Le_i = e_i$ $(i=0, 1)$ the second parameter $|L(e_1-x, x)|/L(|e_1-x|, x)$ in Ω is equal to 0. Thus Theorem 2.1 yields the above assertion. ∎

It is well-known for Bernstein operators that $B_n((e_1-x)^2, x) = \frac{x(1-x)}{n}$. This is used in

EXAMPLE 4.2. For the n-th Bernstein operator B_n, all $f \in C^1[0, 1]$ and every $x \in [0, 1]$ we have

$$|B_n(f, x) - f(x)| \leq \frac{1}{2}\left(\frac{x(1-x)}{n}\right)^{1/2} \omega_1^* \left(f', \left(\frac{x(1-x)}{n}\right)^{1/2}\right).$$

We shall show in the sequel that the constant $\frac{1}{2}$ occurring on the right side in the estimate of Theorem 4.1 cannot be replaced by a constant $c < \frac{1}{2}$.

THEOREM 4.3. *Let $L\colon C[a, b] \to C[a, b]$ be a positive linear operator satisfying $Le_i = e_i$ for $i \in \{0, 1\}$. Then for the second monomial e_2 we have*

$$|L(e_2, x) - e_2(x)| = \frac{1}{2} L((e_1-x)^2, x)^{1/2} \omega_1^*(e_2', L((e_1-x)^2, x)^{1/2}).$$

PROOF. Since $e_2' = 2e_1$, the function $\omega_1(e_2', \cdot)$ given by $\omega_1(e_2', h) = 2h$ for $0 \leq h \leq b-a$ is concave. Thus $\omega_1^*(e_2', \cdot) = \omega_1(e_2', \cdot)$, and the right side in the inequality in Theorem 4.1 is equal to $L((e_1-x)^2, x)$.

The same quantity occurs on the left side in that inequality, which is easily verified using the fact that $Le_i = e_i$ for $i \in \{0, 1\}$. ∎

EXAMPLE 4.4. In the estimate of Example 4.2 concerning Bernstein operators, the constant $\frac{1}{2}$ cannot be replaced by a constant $c < \frac{1}{2}$, except for the points $x=0$ and $x=1$.

We now consider uniform assertions. The main result in this direction is

THEOREM 4.5. *For the sequence $(B_n)_{n \in \mathbf{N}}$ of Bernstein operators one obtains*

$$\sup_{n \in \mathbf{N}} \sup_{\substack{f \in C^1[0,1] \\ f \neq \text{linear}}} \frac{n^{1/2} \|B_n f - f\|_\infty}{\omega_1^*\left(f', \frac{1}{2} n^{-1/2}\right)} = \frac{1}{4}.$$

PROOF. This proof is again a consequence of the concluding remark in Theorem 2.3. From Example 4.2 we first arrive at $(0 \leq x \leq 1)$:

$$|B_n(f, x) - f(x)| \leq \frac{1}{2}\left(\frac{x(1-x)}{n}\right)^{1/2} \omega_1^* \left(f', \left(\frac{x(1-x)}{n}\right)^{1/2}\right),$$

which leads to

$$\|B_n f - f\|_\infty \leq \frac{1}{4} n^{-1/2} \omega_1^* \left(f', \frac{1}{2} n^{-1/2}\right),$$

where, of course, linear functions l (i.e. $l(t)=at+b$) may be excluded from the discussion. This gives the "$\leqq$"-statement. In order to prove that equality holds we may take the function $\tilde{f}\in C^1[0, 1]$ with $\tilde{f}(t)=\left(t-\frac{1}{2}\right)^2$. Because of $\tilde{f}'(t)=2t-1$, the modulus of continuity $\omega_1(\tilde{f}', h)=2h$ is a concave function which guarantees $\omega_1^*(\tilde{f}', h)=\omega_1(\tilde{f}', h)$. Thus $\omega_1^*\left(\tilde{f}', \frac{1}{2}n^{-1/2}\right)=n^{-1/2}$.

Arguing as in the proof of Theorem 3.6, we get $\|B_1\tilde{f}-\tilde{f}\|_\infty=1/4$ since the graph of $B_1\tilde{f}$ is the line connecting the points $(0, 1/4)$ and $(1, 1/4)$. Thus we have $1^{-1/2}\|B_1\tilde{f}\|_\infty/\omega_1^*\left(\tilde{f}, \frac{1}{2}\cdot 1^{-1/2}\right)=1/4$, which proves Theorem 4.5. ∎

Theorem 4.5 immediately implies the following

COROLLARY 4.6. (i) *If the function $f\in C^1[0, 1]$ has a concave first order modulus of continuity of its first derivative $\omega_1(f', \cdot)$ the uniform estimate*

$$\|B_n f-f\|_\infty \leqq \frac{1}{4} n^{-1/2}\omega_1\left(f', \frac{1}{2} n^{-1/2}\right), \qquad n\in\mathbf{N},$$

holds.

(ii) *For any $f\in C^1[0, 1]$ we have*

$$\|B_n f-f\|_\infty \leqq \frac{1}{2}n^{-1/2}\omega_1\left(f', \frac{1}{2}n^{-1/2}\right), \qquad n\in\mathbf{N}.$$

REMARK 4.7. The foregoing result should be compared to Theorem 5.1 in the paper [22] of F. Schurer and F. W. Steutel. While the estimate in Corollary 4.6 (i) cannot be obtained from their assertion, the one in 4.6 (ii) is also an obvious consequence of Schurer's and Steutel's theorem. However, they only treated the particular case of Bernstein polynomials, while our inequalities with optimal constants arose from general theorems for the approximation by positive linear operators of Bernstein-type.

5. Concluding remark

As proved in 1979 (see [4], p. 104),

$$\sup_{n\in\mathbf{N}} \sup_{\substack{f\in C[0,1]\\ f\neq\text{linear}}} \frac{\|B_n f-f\|_\infty}{\omega_2(f, n^{-1/2})} \leqq 3.25,$$

where $\omega_2(f, \cdot)$ denotes the second order modulus of continuity of f. As far as we know it is an open question whether the constant 3.25 is optimal in this sense. Since we do not think so we would be grateful for any comments on this subject.

REFERENCES

The references [3], [7], [8], [11], [18], [21] and [26] are not cited in our article but closely related to the subject.

[1] DeVore, R. A., Optimal convergence of positive linear operators, *Proceedings of the Conference on Constructive Theory of Functions*, Akadémiai Kiadó, Budapest, 1972, 101—120. *MR* **53** #1117.

[2] DeVore, R. A., *The Approximation of Continuous Functions by Positive Linear Operators*, Springer, Berlin—Heidelberg—New York, 1972. *MR* **54** #8100.

[3] Esseen, C. G., Über die asymptotisch beste Approximation stetiger Funktionen mit Hilfe von Bernstein-Polynomen, *Numer. Math.* **2** (1960), 206—213. *MR* **24** #A2784.

[4] Gonska, H. H., Quantitative Aussagen zur Approximation durch positive lineare Operatoren, Dissertation, Universität Duisburg, November, 1979.

[5] Gonska, H. H., A note on pointwise approximation by Hermite—Fejér type interpolation polynomials, *Functions, series, operators*, (Budapest, 1980), 525—537, Colloq. Math. Soc. János Bolyai, 35, North-Holland, Amsterdam—New York, 1983. *MR* **85h:** 41004.

[6] Gonska, H. H., On quasi-Hermite—Fejér interpolation: pointwise estimates. *Constructive Function Theory*, 81 (Varna, 1981), 328—335, Bulgar. Acad. Sci., Sofia, 1983. *MR* **84m:** 41002.

[7] van Iperen, H., A theorem about the determination of a certain best constant in the approximation by some linear operators, *Indag. Math.* **30** (1968), 303—311. *MR* **37** #3253.

[8] Lorentz, G. G., *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953. *MR* **15,** 217.

[9] Lorentz, G. G., *Approximation of Functions*, Holt, Rinehart and Winston, New York, N. Y., 1967. *MR* **35** #4642.

[10] Mamedov, R. G., On the order of approximation of functions by linear positive operators (in Russian), *Dokl. Akad. Nauk SSSR* **128** (1959), 674—676. *MR* **22** #900.

[11] van der Meer, P. J. C., On the degree of approximation by certain linear positive operators, *Indag. Math.* **40** (1978), 467—478. *MR* **80c:** 41013.

[12] Mitjagin, B. S. and Semenov, E. M., Absence of interpolation of linear operators in spaces of smooth functions, *Izv. Akad. Nauk SSSR* **41** (1977), 1289—1328 (in Russian). *MR* **58** #2234

[13] Mond, B., On the degree of approximation by linear positive operators, *J. Approximation Theory* **18** (1976), 304—306. *MR* **54** #10949.

[14] Peetre, J., Exact interpolation theorems for Lipschitz continuous functions, *Ricerche Mat.* **18** (1969), 239—259. *MR* **42** #841.

[15] Peetre, J., Concave majorants of positive functions, *Acta Math. Acad. Sci. Hungar.* **21** (1970), 327—333. *MR* **42** #7841.

[16] Peetre, J., On the connection between the theory of interpolation spaces and approximation theory, *Proceedings of the Conference on Constructive Theory of Functions*, 351—363, Akadémiai Kiadó, Budapest, 1972. *MR* **53** #8760.

[17] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, N. J., 1970. *MR* **43** #445.

[18] Schurer, F., Sikkema, P. C. and Steutel, F. W., On the degree of approximation with Bernstein polynomials, *Indag. Math.* **38** (1976), 231—239. *MR* **53** #6155.

[19] Schurer, F. and Steutel, F. W., On the degree of approximation of functions in $C^1[0, 1]$ by Bernstein polynomials, T. H.-Report 75-WSK-07 (Onderafdeling der Wiskunde, Technische Hogeschool Eindhoven, The Netherlands, 1975).

[20] Schurer, F. and Steutel, F. W., On an inequality of Lorentz in the theory of Bernstein polynomials, *Spline Functions* (*Proceedings of the International Symposium*, Karlsruhe, 1975), 332—338, Springer, Berlin—Heidelberg—New York, 1976. *MR* **58** #12094.

[21] Schurer, F. and Steutel, F. W., Note on the asymptotic degree of approximation of functions in $C^1[0, 1]$ by Bernstein polynomials, *Indag. Math.* **39** (1977), 128—130. *MR* **56** #6213.

[22] Schurer, F. and Steutel, F. W., The degree of local approximation of functions in $C_1[0, 1]$ by Bernstein polynomials, *J. Approximation Theory* **19** (1977), 69—82. *MR* **55** #10913.

[23] Shisha, O. and Mond, B., The degree of convergence of sequences of linear positive operators, *Proc. Nat. Acad. Sci. U.S.A.* **60** (1968), 1196—1200. *MR* **37** #5582.

[24] Sikkema, P. C., Über den Grad der Approximation mit Bernstein-Polynomen, *Numer. Math.* **1** (1959), 221—239. *MR* **22** # 1060.
[25] Sikkema, P. C., Der Wert einiger Konstanten in der Theorie der Approximation mit Bernstein-Polynomen, *Numer. Math.* **3** (1961), 107—116. *MR* **23** # A 459.
[26] Sikkema, P. C. and van der Meer, P. J. C., The exact degree of local approximation by linear positive operators involving the modulus of continuity of the p-th derivative, *Indag. Math.* **41** (1979), 63—76. *MR* **80h:** 41010.

(Received June 21, 1983)

H. H. Gonska and J. Meier

UNIVERSITÄT DUISBURG
GESAMTHOCHSCHULE
FACHBEREICH 11—MATHEMATIK
LOTHARSTRASSE 65
D—4100 DUISBURG 1
FEDERAL REPUBLIC OF GERMANY

H. H. Gonska

DEPARTMENT OF MATHEMATICS AND
COMPUTER SCIENCE
DREXEL UNIVERSITY
PHILADELPHIA, PA 19104
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 299—303

EXTENSIONS OF OPERATOR GROUPS

C. G. CHEHATA, L. NASHED and H. H. SHERIF

1. Introduction

In this paper we consider partial M-endomorphisms of M-groups and seek conditions under which these mappings could be continued to total ones. Conditions analogous to those obtained by B. H. Neumann and H. Neumann [5] are necessary and sufficient for embedding an M-group with a partial M-endomorphism $\mu\colon A\to B(A<_M G, B<_M G;\ <_M$ is read "M-subgroup of") into an M-group G^* with an M-endomorphism μ^* whose restriction to A is μ. For this end we use the generalized free product of M-groups.

Then necessary and sufficient conditions for the simultaneous extension of any well-ordered number of partial M-endomorphisms of G to total ones of some supergroup $G^* \underset{M}{\succ} G$ are given. All the results obtained hold also true for M-modules and N-rings where M is a ring and N a near-ring.

2. Single extensions

Let M be a set of operators and for each i in some index set I, let G_i be an M-group. If, in the free product $\overset{*}{\prod} G_i$, we define an operation $m[f]=[mf]$ for every m in M and every class $[f]$ of words equivalent to

$$f=g_{\alpha_1}\cdots g_{\alpha_i}\cdots g_{\alpha_n},\quad g_{\alpha_i}\in G_j,$$

then $\overset{*}{\prod} G_i$ becomes an M-group. If each G_i contains an M-subgroup H_i which is M-isomorphic to a fixed M-group H, then as in [4] we can define the free product G of the G_i with H amalgamated and with $m[f]=[mf]$, G is again an M-group. The following proposition characterizes G. As for its uniqueness see [3].

PROPOSITION 2.1. *For embeddings* $\gamma_i\colon H\to G_i$ *the following two conditions hold:*

(1) *There exist M-monomorphisms* $\sigma_i\colon G_i\to G$ *such that* $\sigma_i\gamma_i=\sigma_j\gamma_j$.

(2) *If A is an M-group with M-homomorphisms* $\varphi_i\colon G_i\to A$ *satisfying* $\varphi_i\gamma_i=\varphi_j\gamma_j$ *for each i,j in I, there exists a unique M-homomorphism* $\psi\colon G\to A$ *such that* $\sigma_i\psi=\varphi_i$ *for each i.*

THEOREM 2.2. *Let G be an M-group,* $A<_M G$, $B<_M G$ *and* $\mu\colon A\to B$ *a partial M-endomorphism of G. For μ to be totally extendable to an M-endomorphism μ^**

1980 *Mathematics Subject Classification.* Primary 20E06; Secondary 20K35.

Key words and phrases. Partial endomorphisms, operator groups, simultaneous extensions.

of an M-group $G^*>G$, it is necessary and sufficient that there exists an ascending sequence $L_1<L_2<\ldots<L_n<\ldots$ of normal M-subgroups of G such that

$$(L_{n+1}\cap A)\mu = L_n\cap B. \tag{2.1}$$

The proof of this theorem runs along the same lines as in [5]. As for necessity we have only to remark that if G^* with the required properties exists, $G=G_1$, $G_1\mu^*=H_1$ and $\{G_1, H_1\}=G_2$ then G_2 is an M-group for if $g=x_1x_2\ldots x_n$ is an arbitrary element in G_2 then for any m in M

$$mg = m(x_1x_2\ldots x_n) = (mx_1)(mx_2)\ldots(mx_n)$$

since each x_i is in G^*. From the fact that G_1 and H_1 are M-groups then mx_i is in G_1 or H_1 according as x_i is in G_1 or H_1 and hence mg is in G_2. If μ_1 is the restriction of μ^* to G_1 then the kernels of μ and μ_1 are M-subgroups and the kernels K_n^* of the powers μ^{*n} are M-groups for if k is in K_n^* then $(mk)\mu^{*n}=m(k\mu^{*n})=m\,1=1$ and mk is in K_n^*.

To establish the sufficiency of the conditions, we need the analogues of Lemmas 4.1—4.3 in [5]. To this end we first prove the following lemma.

LEMMA 2.3. *Let K be a subset of an M-group G and let*

$$K^G = \bigcap_\alpha (N_\alpha \lhd G\colon K < N_\alpha)$$

be the normal closure of K in G. Further let

$$K_M^G = \bigcap_\beta (I_\beta \lhd_M G\colon K < I_\beta)$$

be the smallest normal M-subgroup of G containing K. If K is an M-subgroup then $K^G=K_M^G$.

PROOF. Let K be an M-subgroup of G. Obviously, $K^G<K_M^G$. Also we have

$$K^G = \Big\{\prod_{\text{finite}} g_i^{-1}k_ig_i\colon\ g_i\in G,\ k_i\in K\Big\}.$$

If $x\in K^G$, $m\in M$ then $x=\prod_i g_i^{-1}k_ig_i$ and

$$mx = \prod_i (mg_i^{-1})(mk_i)(mg_i)$$

$$= \prod_i (mg_i)^{-1}(mk_i)(mg_i)\in K^G.$$

Thus K^G is an M-subgroup of G, and since it is normal and contains K then $K^G=I_\beta$ for some β and $K_M^G<K^G$.

Applying this lemma we obtain the analogue of Lemma 4.1 in [5]. Using the universal property of Proposition 2.1 we prove the following lemma which is the analogue of Lemma 4.2 in [5].

LEMMA 2.4. *Let P be the free product of two M-groups Q, R with an amalgamated M-subgroup S and U be a normal M-subgroup of G. Then*

$$U^P \cap Q = U, \tag{2.2}$$

$$(U \cap S)^R \cap S = U \cup S \tag{2.3}$$

are equivalent and imply

$$U^P \cap R = (U \cap S)^R.$$

PROOF. That (2.2) implies (2.3) is similar to the proof in [5]. Assume (2.3) and consider the M factor groups $Q_1 = Q/U$ and $R_1 = R/(U \cap S)^R$. One easily shows that the two isomorphic subgroups

$$S_1 = (S \cup U)/U < Q_1,$$

$$S_1' = S \cup (U \cap S)^R/(U \cap S)^R < R_1$$

are M-subgroups and M-isomorphic under the mapping φ defined by $\varphi(sU) = s(U \cap S)^R$.

The free product $P_1 = \{Q_1 * R_1;\ S_1 = S_1'\}$ is an M-group with M-homomorphisms:

$$\varphi_1\colon Q \to P_1, \quad Q\varphi_1 = Q_1$$

and

$$\varphi_2\colon R \to P_1, \quad R\varphi_2 = R_1.$$

Let $\gamma_1\colon S \to Q$ and $\gamma_2\colon S \to R$ be M-embeddings. Then $s\gamma_1\varphi_1 = sU \in S_1$ and $s\gamma_2\varphi_2 = s(U \cap S)^R \in S_1'$. Since in P_1, sU is identified with $s(U \cap S)^R$ then $\gamma_1\varphi_1 = \gamma_2\varphi_2$. Thus by Proposition 2.1 there exists a unique M-homomorphism $\theta\colon P \to P^1$ such that $\varphi_1 = \sigma_1\theta$ and $\varphi_2 = \sigma_2\theta$ where σ_1, σ_2 are the M-embeddings. At this step the proof of Lemma 2.4 may be completed as in [5].

The analogue of [5], Lemma 4.3 now follows and this enables us to construct $G^*{}_M > G$ with a total M-endomorphism μ^* whose restriction to A is μ.

3. Simultaneous extensions

For the simultaneous extension of two (or more) partial M-endomorphisms to total M-endomorphisms of one and the same M-supergroup, the necessary and sufficient conditions of Theorem 2.2 now involve the semigroup freely generated by the given partial M-endomorphisms. In case of two partial M-endomorphisms μ, ν of an M-group G the construction of the extension group $G^*{}_M > G$ follows, under similar conditions, the same lines as in [2] making the required alterations due to the fact that G is an operator group. In this case we extend μ and ν alternatively and indefinitely to obtain the sequence of extensions

$$G_1 <_M G_2 <_M \dots$$

and the extension group G^* will then be $G^* = \bigcup_{i=1}^{\infty} G_i$ with μ^*, ν^* defined in the usual manner that makes them M-mappings.

By a classical procedure of transfinite induction we obtain necessary and sufficient conditions for the simultaneous extension of any well-ordered number of partial M-endomorphisms as follows.

THEOREM 3.1. *Let G be an M-group and for every λ in some well-ordered set Λ, let $\mu(\lambda)$: $A_\lambda \to B_\lambda (A_\lambda <_M G,\ B_\lambda <_M G)$ be a partial M-endomorphism of G. Then there exists $G^*{}_M > G$ with M-endomorphisms $\mu^*(\lambda)$ extending $\mu(\lambda)$ for every λ in Λ if and only if for every σ in the semigroup Σ, freely generated by all $\mu(\lambda)$, there exists an M-normal subgroup L_σ of G such that*

$$L_\sigma < L_{\sigma\sigma'}, \tag{3.1}$$

$$L_{\mu(\lambda)} \cap A_\lambda \quad \text{is the kernel of} \quad \mu(\lambda), \tag{3.2}$$

$$(L_{\mu(\lambda)\sigma} \cap A_\lambda)\mu(\lambda) = L_\sigma \cap B_\lambda \tag{3.3}$$

for every σ, σ' in Σ and every λ in Λ.

The following corollaries give sufficient conditions for the existence of the prescribed embeddings in certain special cases.

COROLLARY 3.2. *It is sufficient for the existence of G^* that there exists for each $\lambda \in \Lambda$, a non-decreasing sequence*

$$L_1(\lambda) < L_2(\lambda) < \ldots < L_n(\lambda) < \ldots$$

of M-normal subgroups of G such that

$$L_1(\lambda) \cap A_\lambda \quad \text{is the kernel of } \mu(\lambda), \tag{3.4}$$

$$(L_{n+1}(\lambda) \cap A_\lambda)\mu(\lambda) = L_n(\lambda) \cap B_\lambda, \tag{3.5}$$

$$L_n(\lambda) \cap B_{\lambda'} = 1 \tag{3.6}$$

for all λ, $\lambda'(\neq \lambda)$ in Λ.

For if we set

$$L_{\mu^n(\lambda)\sigma} = L_{\mu^n(\lambda)} = L_n(\lambda)$$

for every λ in Λ, σ in Σ and $n = 1, 2, \ldots$ then conditions (3.1)—(3.3) will be satisfied.

COROLLARY 3.3. *If Ω is a subset of Λ such that for every ω in Ω, $\mu(\omega)$ is moreover a partial M-automorphism of G, then it is sufficient for the existence of G^* that the sequences of Corollary* 3.2 *exist for every λ in $\Lambda - \Omega$ such that conditions* (3.4)—(3.6) *are satisfied for every λ in $\Lambda - \Omega$ and $\lambda'(\neq \lambda)$ in Λ.*

For if, for every ω in Ω, we set

$$L_1(\omega) = L_2(\omega) = \ldots = L_n(\omega) = \ldots = 1$$

then conditions of Corollary 3.3 will be satisfied for every λ in Λ.

COROLLARY 3.4. *If for every λ in Λ, $\mu(\lambda)$ is a partial M-automorphism of G then $G^*{}_M > G$ with total M-automorphisms $\mu^*(\lambda)$ extending $\mu(\lambda)$ always exists.*

This follows immediately from Corollary 3.3 by setting $L_n(\lambda) = 1$ for all λ in Λ and $n = 1, 2, \ldots$. This result had been proved differently in [3], Theorem 4.1.

4. Abelian extensions

Finally we remark that if G is abelian, a partial M-endomorphism of G is always extendable to a total one. If in our construction we use the generalized direct product instead of the free product then the resulting extension group G^* will come out to be also abelian. Applying transfinite induction as in [2] we may prove the following result.

THEOREM 4.1. *If for every λ in a well-ordered set Λ, the abelian M-group G possesses a partial M-endomorphism $\mu(\lambda)$, then there exists an abelian M-supergroup $G^* > G$ with M-endomorphisms $\mu^*(\lambda)$ extending $\mu(\lambda)$ for every λ in Λ.*

REFERENCES

[1] CHEHATA, C. G., Simultaneous extension of partial endomorphisms of groups, *Proc. Glasgow Math. Assoc.* **2** (1954), 37—46. *MR* **16**—10.
[2] CHEHATA, C. G., Embedding theorems for abelian groups, *Canad. J. Math.* **15** (1963), 766—770. *MR* **27** #4852.
[3] CHEHATA, C. G. and SHERIF, H. H., Embedding theorems for operator groups, *Math. Sci.* **4** (1979), 99—104. *MR* **81c**: 20027.
[4] HALL, M., JR., *The theory of groups*, The Macmillan Co., New York, 1959. *MR* **21** #1996.
[5] NEUMANN, B. H. and NEUMANN, HANNA, Extending partial endomorphisms of groups, *Proc. London Math. Soc.* **2** (1952), 337—348. *MR* **14**—351.

(Received August 3, 1983)

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
ALEXANDRIA UNIVERSITY
MOHARREM BEY
ALEXANDRIA
EGYPT

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 305—313

WALKING IN FINITE DIRECTED GRAPHS

M. BOGNÁR

The centers of the old towns are mostly full with one-way streets and the traffic directions of the streets are always chosen so that each point of the town should be connected by one-way routes with each other one. Now what is a good wandering method for a driver who has never been before in this town and has no map from the town to reach a given point of the center starting from another one.

In the language of the theory of graphs where the terminology and notations are taken from [1] the problem may be formulated as follows.

Given a finite directed strongly connected graph $G=(X, U)$ with at least one arc, what is a good algorithm to obtain a path in G including all the arcs of G?

We have the following

THEOREM. *Consider the following algorithm.*

Let u_1 be an arbitrary arc of G. Suppose that n is a positive integer and for $1 \leqq i \leqq n$ the arc u_i has already been taken such that

$$\mu_n = (u_1, \dots, u_n)$$

is a path in G. Let x_n be the terminal endpoint of μ_n. Now select u_{n+1} such that among the arcs with initial endpoint x_n it should appear a minimal number of times in μ_n.

Now let m be the order of G and let k be the maximum of the outer demi-degrees of the vertices. Let

$$s(m, k) = k+k^2+\dots+k^m$$

and

$$t(m, k) = s(m, k)-k^m+k^{m-1}.$$

Then each arc of G appears at least once in any path in G formed by the given algorithm and being of the length $s(m, k)$. If in addition G has no loops then $s(m, k)$ can be replaced by $t(m, k)$. Moreover, these upper bounds $s(m, k)$ and $t(m, k)$ are strict.

PROOF. First for any arc u of G let $x_*(u)$ denote the initial and $x^*(u)$ the terminal endpoint of u.

Next for any path

$$\mu' = (u_1', \dots, u_r')$$

1980 *Mathematics Subject Classification.* Primary 05C20; Secondary 05C38.
Key words and phrases. Regular path, critical path.

in G and for any arc u of G let

(1) $$\alpha_{\mu'}(u) = \left|\{i\in\{1, \ldots, r\}/u'_i = u\}\right|.$$

Moreover for the path μ' and for any vertex x of G let

(2) $$\alpha_{\mu'}(x) = \min\{\alpha_{\mu'}(u)/x_*(u) = x\}$$

(3) $$\beta_{\mu'}(x) = \sum_{x_*(u)=x} \alpha_{\mu'}(u).$$

If in addition u' is an arc of G with $x^*(u')=x$ then we have obviously

(4) $$\alpha_{\mu'}(u') \leqq \beta_{\mu'}(x)+1.$$

Moreover one clearly has

(5) $$\sum_{x\in X} \beta_{\mu'}(x) = r.$$

In the sequel a path in G is said to be *regular* if it is formed by the given algorithm.

Let

$$\mu = (u_1, \ldots, u_n)$$

be a regular path in G. Let

$$k' = k+1.$$

1° Let u and u' be arcs of G with the same initial endpoint x. Then

$$|\alpha_\mu(u)-\alpha_\mu(u')| \leqq 1.$$

In fact for $i=1, \ldots, n$ let

$$\mu'_i = (u_1, \ldots, u_i).$$

μ'_i is obviously a regular path, too. Moreover we clearly have

$$|\alpha_{\mu'_1}(u)-\alpha_{\mu'_1}(u')| \leqq 1.$$

To prove the given inequality we proceed by induction.

Suppose that $1<i\leqq n$ and that

$$|\alpha_{\mu'_{i-1}}(u)-\alpha_{\mu'_{i-1}}(u')| \leqq 1$$

holds.

If $\alpha_{\mu'_{i-1}}(u)=\alpha_{\mu'_{i-1}}(u')$ then we have obviously

$$|\alpha_{\mu'_i}(u)-\alpha_{\mu'_i}(u')| \leqq 1.$$

Considering now the case $\alpha_{\mu'_{i-1}}(u)\neq\alpha_{\mu'_{i-1}}(u')$ there is no loss of generality by supposing

$$\alpha_{\mu'_{i-1}}(u') = \alpha_{\mu'_{i-1}}(u)+1$$

and thus

$$\alpha_{\mu'_{i-1}}(u') \neq \alpha_{\mu'_{i-1}}(x).$$

Hence by the regularity of μ_i' we have $u' \neq u_i$ and this implies

$$\alpha_{\mu_i'}(u') = \alpha_{\mu_{i-1}'}(u').$$

Now in the case $u=u_i$ we have $\alpha_{\mu_i'}(u)=\alpha_{\mu_{i-1}'}(u)+1$ which yields $\alpha_{\mu_i'}(u')= =\alpha_{\mu_i'}(u)$ and if $u\neq u_i$ then $\alpha_{\mu_i'}(u)=\alpha_{\mu_{i-1}'}(u)$ and thus $\alpha_{\mu_i'}(u')=\alpha_{\mu_i'}(u)+1$. In both cases one clearly has

$$|\alpha_{\mu_i'}(u)-\alpha_{\mu_i'}(u')| \leqq 1.$$

The inequality

$$|\alpha_\mu(u)-\alpha_\mu(u')| = |\alpha_{\mu_n'}(u)-\alpha_{\mu_n'}(u')| \leqq 1$$

has been proved.

2° Let x be a vertex of G and u an arc of G such that $x_*(u)=x$. Then for any positive integer q

$$\alpha_\mu(u) < q$$

implies

$$\beta_\mu(x) < k'q-1.$$

In fact for any arc u' of G with initial endpoint x and distinct from u we have by 1°

$$\alpha_\mu(u') \leqq q$$

and thus

$$\beta_\mu(x) < q+(k'-2)q = (k'-1)q \leqq k'q-1$$

indeed.

3° Let u and u' be arcs of G such that $x^*(u')=x_*(u)$ and let q be a positive integer. Then by 2° and (4)

$$\alpha_\mu(u) < q$$

implies

$$\alpha_\mu(u') < k'q.$$

4° It is immediate from 3° that for any path $\mu'=(u_1', \ldots, u_r')$ in G and for any positive integer q the relation

$$\alpha_\mu(u_r') < q$$

implies

$$\alpha_\mu(u_1') < k'^{r-1}q.$$

5° Lemma. *If the given regular path μ in G is of the length mk'^m then it includes all the arcs of G.*

Proof of the lemma. Suppose the existence of an arc u of G not appearing in μ, i.e. for which

$$\alpha_\mu(u) < 1.$$

Let x be the initial endpoint of u and let x' be an arbitrary vertex of G.

If $x=x'$ then we have by 2°

$$\beta_\mu(x') < k'-1 < k'^m.$$

Suppose now $x' \neq x$. Since G is strongly connected there is a path $\mu' = (u_1', \ldots, u_t')$ in G with initial endpoint x' and terminal endpoint x and of the length $t \leqq m-1$. Let $\mu'' = (u_1', \ldots, u_t', u)$. μ'' is a path in G of the length less or equal than m. Hence $\alpha_\mu(u) = 0 < 1$ implies by 4°

$$\alpha_\mu(u_1') < k'^{m-1} \cdot 1 = k'^{m-1}.$$

Consequently, since u_1' is an arc of G such that $x_*(u_1') = x'$ it follows by 2°

$$\beta_\mu(x') < k'k'^{m-1} - 1 < k'^m.$$

Hence for each vertex x' of G one has $\beta_\mu(x') < k'^m$ and thus

$$\sum_{x' \in X} \beta_\mu(x') < mk'^m.$$

Therefore by (5) we have

$$n < mk'^m.$$

The length n of path μ is less than mk'^m which proves the lemma.

6° The regular path $\mu = (u_1, \ldots, u_n)$ in G is said to be *critical* if it includes all the arcs of G and $\alpha_\mu(u_n) = 1$.

Now to prove the first two statements of the theorem by Lemma 5° we need only to show that if the given path μ is critical then the length of it is at most $s(m, k)$ ($t(m, k)$, respectively).

Suppose in the remainder of the proof that the regular path $\mu = (u_1, \ldots, u_n)$ is critical.

For $i = 1, \ldots, m$ let us define the vertex x^i and the positive integer $r(i)$ in a recursive way as follows.

Let $r(1) = n$ and let $x^1 = x_*(u_{r(1)})$. Suppose that $1 < i \leqq m$ and that the vertices $x^1, \ldots, x^{i-1}$ have already been defined. Let $u_{r(i)}$ be the last member in the sequence μ incident into the set $\{x^1, \ldots, x^{i-1}\}$ and $r(i)$ the index of it. Let $x^i = x_*(u_{r(i)})$.

Thus we have obviously

$$X = \{x^1, \ldots, x^m\} \tag{6}$$

and

$$r(m) < \ldots < r(1). \tag{7}$$

Now for $i = 1, \ldots, m$ let

$$\mu_i = (u_1, \ldots, u_{r(i)})$$

and if $r(i) > 1$ then let

$$\mu_i' = (u_1, \ldots, u_{r(i)-1}).$$

μ_i and μ_i' are regular paths in G and for $1 \leqq i < m$ μ_i' is still defined.

Observe that for $1 \leqq j < i \leqq m$ according to (7) we clearly have

$$\alpha_{\mu_i}(u_{r(j)}) < \alpha_\mu(u_{r(j)}). \tag{8}$$

Moreover for any arc u of G and for $i=1, \ldots, m$ one has

$$\alpha_{\mu_i}(u) \leqq \alpha_\mu(u). \tag{9}$$

If in addition $x_*(u)=x^i$ then we have obviously

$$\alpha_{\mu_i}(u) = \alpha_\mu(u) \tag{10}$$

and if the condition $u \neq u_{r(i)}$ is also satisfied then $r(i)>1$ and we get

$$\alpha_{\mu_i'}(u) = \alpha_\mu(u) \tag{11}$$

and

$$\alpha_{\mu_i'}(u_{r(i)}) < \alpha_\mu(u_{r(i)}). \tag{12}$$

7° We now show that for $i=1, \ldots, m$ and for any arc u of G with initial endpoint x^i we have

$$\alpha_\mu(u) \leqq k^{i-1}.$$

In fact let $i=1$. Since the path μ is critical it follows for $u=u_{r(1)}=u_n$

$$\alpha_\mu(u) = 1 \leqq k^{1-1}. \tag{13}$$

Now let u' be an arc of G with initial endpoint x^1 and distinct from $u_{r(1)}$. Then by (12) and (13) we have $\alpha_{\mu_1'}(u_{r(1)})=0$ and thus by 1° one obtains $\alpha_{\mu_1'}(u')\leqq 1$. Taking also (11) into account we get

$$\alpha_\mu(u') \leqq 1 = k^{1-1}$$

as required.

Suppose now that the assertion is true for $i=1, \ldots, q-1$ where $q \leqq m$. We first show that

$$\alpha_\mu(u_{r(q)}) \leqq k^{q-1}. \tag{14}$$

Let x^j be the terminal endpoint of $u_{r(q)}$. Then $j<q$ and thus by the induction hypothesis for each arc u with $x_*(u)=x^j$ we have

$$\alpha_\mu(u) \leqq k^{j-1} \leqq k^{q-2}, \tag{15}$$

consequently by (9) we get

$$\alpha_{\mu_q}(u) \leqq k^{q-2}. \tag{16}$$

However, by (8) and (15) one has

$$\alpha_{\mu_q}(u_{r(j)}) < \alpha_\mu(u_{r(j)}) \leqq k^{q-2}$$

and thus taking also (16) into account we get

$$\beta_{\mu_q}(x^j) < kk^{q-2} = k^{q-1}$$

which implies by (10) and (4)

$$\alpha_\mu(u_{r(q)}) = \alpha_{\mu_q}(u_{r(q)}) \leqq \beta_{\mu_q}(x^j)+1 \leqq k^{q-1}.$$

The inequality (14) has been verified.

Now let u be an arc of G with initial endpoint x^q and distinct from $u_{r(q)}$. Then by (12) and (14) we have $\alpha_{\mu'_q}(u_{r(q)})<k^{q-1}$, consequently 1° shows that $\alpha_{\mu'_q}(u)\leqq k^{q-1}$ and this implies by (11) the required

$$\alpha_\mu(u) \leqq k^{q-1}.$$

8° The preceding segment shows that for $i=1, \ldots, m$ we have

(17) $$\beta_\mu(x^i) \leqq kk^{i-1} = k^i$$

and thus by (5) and (6) we get

$$n = \sum_{i=1}^{m} \beta_\mu(x^i) \leqq k+k^2+\ldots+k^m = s(m, k).$$

Here the first statement of the theorem has been proved.

9° Suppose now that G has no loops. The order of G is in this case obviously at least 2. For proving the inequality $n\leqq t(m, k)$ by (17), (5) and (6) we only need to show that

$$\beta_\mu(x^m) \leqq k^{m-1}.$$

First observe that for an arbitrary path μ' in G and for each vertex x of G we have

(18) $$\beta_{\mu'}(x) \leqq \Big(\sum_{x^*(u)=x} \alpha_{\mu'}(u)\Big)+1.$$

Next, $x_*(u_{r(1)})=x^1\neq x^m$ shows that

(19) $$\beta_\mu(x^m) = \beta_{\mu'_1}(x^m).$$

Moreover we have obviously

$$\alpha_{\mu'_1}(u_{r(1)}) = 0.$$

Observe also that for $i=2, \ldots, m-1$ one clearly has

$$x^*(u_{r(i)}) \neq x^m.$$

These latter two relations show that for $i=1, \ldots, m-1$ there are at most $(k-1)$ arcs u in G satisfying the conditions

$$x_*(u) = x^i, \quad x^*(u) = x^m \quad \text{and} \quad \alpha_{\mu'_1}(u) \neq 0.$$

Accordingly, since G has no loops it follows by (19), (18) and 7° that

$$\beta_\mu(x^m) \leqq \Big(\sum_{x^*(u)=x^m} \alpha_{\mu'_1}(u)\Big)+1 \leqq (k-1)\cdot 1+\ldots+(k-1)\cdot k^{m-2}+1 = k^{m-1}$$

as required.

Also the second statement of the theorem has been proved.

10° We are going to prove that the upper bounds $s(m, k)$ and $t(m, k)$ are strict.

First for $k=1, 2, \ldots$ we construct an infinite directed graph G_k.

Let $x_1, x_2, \ldots$ be distinct vertices. Moreover let $u_1, u_2, \ldots$ and $u_1', u_2', \ldots$ be distinct arcs such that for $i=1, 2, \ldots$ and $r=1, \ldots, k$ the relations

$$x_*(u_{k(i-1)+r}) = x_*(u_i') = x_i,$$

$$x^*(u_{k(i-1)+r}) = x_1$$

and

$$x^*(u_i') = x_{i+1}$$

hold. G_k is the graph with the vertices $x_1, x_2, \ldots$ and with the arcs $u_1, u_2, \ldots$ and $u_1', u_2', \ldots$.

Now let

$$X(1, k) = \{x_1\},$$

$$U(1, k) = \{u_1, \ldots, u_k\}$$

and for $m=2, 3, \ldots$ let

$$X(m, k) = \{x_1, \ldots, x_m\},$$

$$U(m, k) = (\{u_1, \ldots, u_{mk}\} \setminus \{u_k, u_{2k}, \ldots, u_{(m-1)k}\}) \cup \{u_1', \ldots, u_{m-1}'\}$$

and

$$U'(m, k) = U(m, k) \setminus \{u_1, u_2, \ldots, u_{k-1}\}.$$

Thus for $k=1$ we have

$$U(m, k) = U'(m, k).$$

In such a way the subgraphs

$$G(m, k) = (X(m, k),\ U(m, k))$$

and

$$G'(m, k) = (X(m, k),\ U'(m, k))$$

of G_k (the latter one is only defined for $m \geqq 2$) are strongly connected directed graphs they have the order m and the maximum of the outer demi-degrees of their vertices is k. Moreover, $G'(m, k)$ has no loops.

For $m=3$ and $k=2$ $G(m, k)$ and $G'(m, k)$ are drawn below.

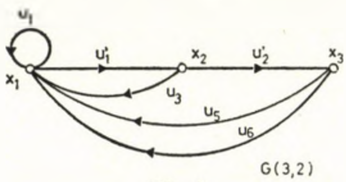


Fig. 1

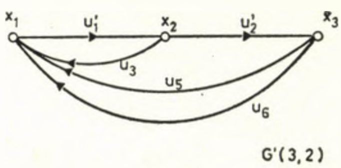


Fig. 2

We are going to construct regular paths $\mu(m, k)$ and $\mu'(m, k)$ in $G(m, k)$ ($G'(m, k)$, respectively) of the length $s(m, k)-1$ ($t(m, k)-1$, respectively) not including all the arcs of $G(m, k)$ (of $G'(m, k)$, respectively).

(a) If $k=1$ and $m \geqq 2$ then let

$$\mu(m, 1) = \mu'(m, 1) = (u_1', \ldots, u_{m-1}').$$

This path is regular, it is of the length $s(m, 1)-1=t(m, 1)-1$ and fails to include the arc u_m of $G(m, 1)=G'(m, 1)$.

(b) Let $k\geqq 2$. Let

$$\mu(1, k) = (u_1, \ldots, u_{k-1}).$$

$\mu(1, k)$ is a regular path in $G(1, k)$ of the length $s(1, k)-1$ not including the arc u_k of $G(1, k)$.

Now suppose that the regular path $\mu(m, k)$ in $G(m, k)$ of the length $s(m, k)-1$ has already been constructed, it is of the form

$$\mu(m, k) = (u_1, \ldots, u^*)$$

where $x^*(u^*)=x_m$ and it fails to include the arc u_{mk} of $G(m, k)$ and let

$$\mu(m+1, k) = (\underbrace{u_1, \ldots, u^*}_{\mu(m,k)}, \underbrace{u'_m}_{1}, u_{mk+1}, \underbrace{u_1, \ldots, u^*}_{\mu(m,k)},$$

$$\underbrace{u'_m}_{2}, \ldots, \underbrace{u'_m}_{k-1}, u_{mk+(k-1)}, \underbrace{u_1, \ldots, u^*}_{\mu(m,k)}, \underbrace{u'_m}_{k}).$$

Then $\mu(m+1, k)$ is clearly a regular path in $G(m+1, k)$, its first arc is u_1, its terminal endpoint is x_{m+1}, it fails to include the arc $u_{(m+1)k}$ of $G(m+1, k)$ and it is of the length

$$k(s(m, k)-1)+2(k-1)+1 = s(m+1, k)-1.$$

E.g. $\mu(3, 2)=(u_1, u'_1, u_3, u_1, u'_1, u'_2, u_5, u_1, u'_1, u_3, u_1, u'_1, u'_2)$.

(c) Let

$$\mu'(2, k) = (\underbrace{u'_1}_{1}, u_{k+1}, \underbrace{u'_1}_{2}, \ldots, \underbrace{u'_1}_{k-1}, u_{k+(k-1)}, \underbrace{u'_1}_{k}).$$

$\mu'(2, k)$ is a regular path in $G'(2, k)$, it fails to include the arc u_{2k} of $G'(2, k)$ and it is of the length $t(2, k)-1$.

Now suppose that $m\geqq 2$ and the regular path $\mu'(m, k)$ in $G'(m, k)$ of the length $t(m, k)-1$ has already been constructed, it is of the form

$$\mu'(m, k) = (u'_1, \ldots, u'_{m-1})$$

and it fails to include the arc u_{mk} of $G'(m, k)$ and let

$$\mu'(m+1, k) = (\underbrace{u'_1, \ldots, u'_{m-1}}_{\mu'(m,k)}, \underbrace{u'_m}_{1}, u_{mk+1},$$

$$\underbrace{u'_1, \ldots, u'_{m-1}}_{\mu'(m,k)}, \underbrace{u'_m}_{2}, \ldots, \underbrace{u'_m}_{k-1}, u_{mk+(k-1)},$$

$$\underbrace{u'_1, \ldots, u'_{m-1}}_{\mu'(m,k)}, \underbrace{u'_m}_{k}).$$

Then $\mu'(m+1, k)$ is clearly a regular path in $G'(m+1, k)$, it is of the form

$$\mu'(m+1, k) = (u'_1, \ldots, u'_m)$$

it fails to include the arc $u_{(m+1)k}$ of $G'(m+1, k)$ and it is of the length

$$k(t(m, k)-1)+2(k-1)+1 = t(m+1, k)-1.$$

E.g.

$$\mu'(3, 2) = (u_1', u_3, u_1', u_2', u_5, u_1', u_3, u_1', u_2').$$

(d) Observe finally that for $m=k=1$ the upper bound $s(1, 1)=1$ is strict indeed.

Hence for each couple (m, k) of positive integers the upper bound $s(m, k)$ is strict and for each couple (m, k) of positive integers where $m\geqq 2$ the upper bound $t(m, k)$ is strict as well.

The proof of the theorem is complete.

11° A finite directed strongly connected graph G of order m with at least one arc and being such that the maximum of the outer demi-degrees of its vertices is k is said to be critical if either $m=k=1$ or there exists a regular path in G of the length $s(m, k)-1$ (if G has no loops then of the length $t(m, k)-1$, respectively) not including all the arcs of G.

And now we can establish the problem of characterizing the set of critical graphs.

The problem can be easily solved.

For $k=1$ the only critical graphs are the $G(m, 1)-s$ and there is no other critical graph of this kind.

For $k\geqq 2$ we can replace in $G(m, k)$ the arc u_{mk} by an arc u_{mk}^* such that $x_*(u_{mk}^*)=x_m$ and $x^*(u_{mk}^*)$ is an arbitrary element of $X(m, k)$. Thus we have got all the critical graphs of this kind.

If in addition we require that the graph in question has no loops then beside the preceding modification each loop u_j of $G(m, k)$ must be either omitted or replaced by an arc u_j^* incident out of vertex x_1 and incident into vertex x_2. Thus we clearly obtain all the critical graphs without loops.

Remark. We do not know whether there exists a better simple algorithm solving the same original problem.

REFERENCE

[1] Berge, C., *Graphs and Hypergraphs*, North-Holland Publ. Co., Amsterdam—London, 1973. *MR* **50** #9640.

(Received September 15, 1983)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIAI TANSZÉK
RÁKÓCZI ÚT 5
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 315—320

DECOMPOSITIONS OF GRAPHS INTO COMPLETE SUBGRAPHS OF GIVEN ORDER

E. GYŐRI and ZS. TUZA

1. Introduction

The best known result of extremal graph theory is Turán's theorem. For $n=t(k-1)+r$, $0\leqq r\leqq k-2$, let $T_{k-1,n}$ denote the complete $(k-1)$-partite graph of n vertices such that r color classes contain $t+1$ vertices and $k-r-1$ color classes contain t vertices. Then

$$t_{k-1,n}:=|E(T_{k-1,n})|=\frac{n(n-t)-r(t+1)}{2}$$

and $T_{k-1,n}$ does not contain any complete subgraph K_k of k vertices. Turán [10] proved in 1941 that any other graph of n vertices with at least $t_{k-1,n}$ edges contains K_k as a subgraph. (The case $k=3$ was proved already in 1907 by Mantel [7].) Erdős, Goodman and Pósa [3] (for $k=3$) and Bollobás [1] (for $k\geqq 4$) extended Turán's theorem in some sense. Their results can be formulated in the following form.

THEOREM A. *The edge set of every graph of n vertices can be decomposed into at most $t_{k-1,n}$ edge disjoint K_k's and edges. If $k>3$ then $T_{k-1,n}$ is the only extremal graph. For $k=3$, the complete graphs K_4, K_5 and the graphs $K_{2,n}$ $(n=1, 2, \ldots)$ are the extremal graphs.*

Stronger extremality of Turán graphs $T_{2,n}$ was proved by Chung [2] and Győri and Kostochka [4]. For any graph G, let

$$p(G)=$$
$$=\min\left\{\sum_{i=1}^{m}|V(G_i)|:\ G_i\text{'s are edge disjoint complete subgraphs, }\bigcup_{i=1}^{m}E(G_i)=E(G)\right\}.$$

The following theorem was proved in [2] and [4].

THEOREM B. *For any graph G of n vertices,*

$$p(G)\leqq\left[\frac{n^2}{2}\right]=2t_{2,n}$$

and equality holds only for $T_{2,n}$.

In this paper, we prove the still stronger extremality of Turán graphs $T_{k,n}$ $(k\geqq 3)$. Actually, we improve Theorem A for $k\geqq 4$. For an arbitrary graph G, let

---

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1812.

1980 *Mathematics Subject Classification*. Primary 05C35, 05C70.
Key words and phrases. Turán's theorem, complete subgraphs, edge decomposition.

$p_k(G)=\min\{\sum_{i=1}^{m}|V(G_i)|:\ G_i$'s are K_k's and edges, $\bigcup_{i=1}^{m}E(G_i)=E(G)\}$. Naturally, if G is K_k-free then $p_k(G)=2|E(G)|$ and, in particular, $p_k(T_{k-1,n})=2t_{k-1,n}$. The following theorem states that $2t_{k-1,n}$ is the maximum of $p_k(G)$ for graphs G of n vertices.

THEOREM 1. *If $k\geqq 4$ and G is an arbitrary graph of n vertices then*

$$p_k(G)\leqq 2t_{k-1,n}$$

and equality holds only for $T_{k-1,n}$.

Covering the complete graph K_{6m-2} by edge disjoint triangles (K_3's) and edges, the edges used in the cover form a subgraph such that the degrees are odd. But this subgraph cannot be a 1-factor because the number of edges in the complement of a 1-factor is $(3m-2)(6m-2)$ which is not divisible by 3. Thus we have to use at least $3m$ edges and so it is easy to see that

$$p_3(K_{6m-2})\geqq|E(K_{6m-2})|+3m=2t_{2,6m-2}+1.$$

For $k=3$, we prove the following

THEOREM 2. *For any graph G of n vertices, $p_3(G)<\frac{9}{16}n^2$.*

From the proof of Theorem 2, it will be seen easily that the argument can be improved to obtain the slightly stronger inequality $p_3(G)\leqq\frac{9}{16}n^2-\frac{3}{16}n+\frac{7}{16}+o(1)$. However, we did not find it worth giving the numerical details of computation because we conjecture that this stronger estimate is still far from the best upper bound.

CONJECTURE. For every graph G of n vertices, $p_3(G)\leqq\frac{1}{2}n^2+O(n)$.

We note that the not necessarily edge disjoint (edge) coverings of graphs also have an extended literature. References of such results can be found in [11].

NOTATION. For a real number x, $\lfloor x\rfloor$ and $\lceil x\rceil$ denote the maximum integer not greater than x and the minimum integer not smaller than x, respectively.

Graph means simple graph without loops and multiple edges. As usual, $V(G)$, $E(G)$ and $\bar{G}$ denote the vertex set, the edge set and the complement of G, respectively.

For $V_0\subset V(G)$, $G(V_0)$ and $G-V_0$ denote the subgraphs induced by V_0 and by $V(G)-V_0$, respectively. If $V_0=\{v\}$ then we write $G-v$ instead of $G-\{v\}$. For $E_0\subset E(G)$, $G-E_0$ denotes the graph with vertex set $V(G)$ and with edge set $E(G)-E_0$.

For $v\in V(G)$, the degree of v is denoted by $d_G(v)$ or sometimes simply by $d(v)$ if it does not cause any misunderstanding. The neighbourhood $N(v)$ of v is the set of vertices connected to v and so $|N(v)|=d(v)$.

2. Decomposition into complete subgraphs of given order $\geqq 4$

For proving Theorem 1, the main result of our paper, we need the following deep theorem of Hajnal and Szemerédi [5].

THEOREM C. *Let G be a graph of n vertices such that the degree of every vertex in G is at least d. Then there exist vertex disjoint complete subgraphs $G_1, \ldots, G_{n-d}$ of G such that $\bigcup_{i=1}^{n-d} V(G_i)=V(G)$ and the numbers of the vertices of the subgraphs G_i differ from each other by at most one, i.e., $|V(G_i)|$ is $\left\lceil \frac{n}{n-d} \right\rceil$ or $\left\lfloor \frac{n}{n-d} \right\rfloor$ for $1\leqq i\leqq$ $\leqq n-d$.*

PROOF OF THEOREM 1. We prove the theorem by induction on n. If $n<k$ then the statement holds obviously because then $T_{k-1,n}=K_n$. Suppose that Theorem 1 holds for graphs of at most $n-1$ vertices. Let $n-1=t(k-1)+r$, $0\leqq r\leqq k-2$, $t\geqq 1$. Then

$$2t_{k-1,n}-2t_{k-1,n-1} = 2(n-t-1) = 2\big(t(k-2)+r\big).$$

Consider an arbitrary graph G of n vertices and let x be a vertex of G with minimum degree, $d_G(x)=d$. If d is relatively small then the induction step can be done easily as follows. Suppose that $d\leqq n-1-t$, then

$$p_k(G) \leqq 2d+p_k(G-x) \leqq 2\big(t(k-2)+r\big)+2t_{k-1,n-1} = 2t_{k-1,n}$$

with equality if and only if $d=t(k-2)+r=n-t-1$ and $G-x=T_{k-1,n-1}$. But in this case $|E(G)|=t_{k-1,n}$, therefore $p_k(G)<2t_{k-1,n}$ unless G is K_k-free. Thus, Turán's theorem implies $G=T_{k-1,n}$.

From now on, suppose $d\geqq t(k-2)+r+1=n-t$.

We will prove that $p_k(G)<2t_{k-1,n}$ in this case.

By the minimality of $d(x)$, the degree of every vertex in $G\big(N(x)\big)$ is at least $2d-n$. Applying Theorem C to $G\big(N(x)\big)$, we obtain that there exist vertex disjoint complete subgraphs $G_1, \ldots, G_{n-d}$ in $G\big(N(x)\big)$ such that $\bigcup_{i=1}^{n-d} V(G_i)=N(x)$ and $|V(G_i)|=\left\lceil \frac{d}{n-d} \right\rceil$ or $\left\lfloor \frac{d}{n-d} \right\rfloor$ $(i=1, 2, \ldots, n-d)$. Now

$$\frac{d}{n-d} \geqq \frac{t(k-2)+r+1}{t} > k-2,$$

so if $|V(G_i)|<k-1$ then $|V(G_i)|=k-2$ $(i=1, \ldots, n-d)$. Now we distinguish between two cases according to the orders of G_i's.

Case 1. There exists at least one G_i with $|V(G_i)|=k-2$.

We may suppose that $|V(G_i)|=k-1$ for $i=1, 2, \ldots, p$ and $|V(G_i)|=k-2$ for $i=p+1, p+2, \ldots, n-d$ where $p\leqq n-d-1$. Then $d=p(k-1)+(n-d-p)(k-2)$ which implies that

$$p = d(k-1)-n(k-2). \tag{1}$$

Then the edges incident to x can be covered by p edge disjoint K_k's $\big(G(V(G_i)\cup\{x\})$ for $i=1, 2, ..., p\big)$ and $(n-d-p)(k-2)$ edges. The sum of their orders is

$$kp+2(n-d-p)(k-2) = n(k-2)^2-d(k^2-3k) \leqq$$
$$\leqq (r+1)(4-k)+2tk-4t \leqq 2t(k-2)+2r$$

by the assumption $d \geqq n-t$.

It follows by the induction hypothesis that

$$p_k(G) \leqq 2t(k-2)+2r+p_k\big(G-x-\bigcup_{i=1}^{n-d} E(G_i)\big) \leqq$$
$$\leqq 2t(k-2)+2r+2t_{k-1,n-1} = 2t_{k-1,n}.$$

From here, it is easy to see that $p_k(G)$ is strictly smaller than $2t_{k-1,n}$: equality would hold only if $d=n-t$, $k=4$, $r=0$ and $G-x=T_{k-1,n-1}$. Then every vertex has degree $n-t-1$ in $G-x$, therefore each of them is adjacent to x, by the minimality of $d_G(x)$. Consequently, $n-t=d=n-1$, $t=1$, $n=4$, $G=K_4$ and so $p_4(G)=4<2t_{3,4}$.

Case 2. $|V(G_i)| \geqq k-1$ for $i=1, 2, ..., n-d$.

Now

$$d \geqq (k-1)(n-d)$$

by the equality $\sum_{i=1}^{n-d} |V(G_i)|=d$. Let

$$d = s(k-1)+q, \qquad 0 \leqq q \leqq k-2. \tag{2}$$

Then $s \geqq n-d \geqq 1$ by the previous inequality and $s \leqq t$ by $d \leqq n-1$. The minimum degree in $G\big(N(x)\big)$ is at least $d-(n-d) \geqq d-s$ and so, according to Theorem C, there exist vertex disjoint complete subgraphs $G_1', ..., G_s'$ of $\lfloor d/s \rfloor$ or $\lceil d/s \rceil$ vertices in $G\big(N(x)\big)$ such that $\bigcup_{i=1}^{s} V(G_i')=N(x)$. Then

$$|V(G_i')| \geqq \lfloor d/s \rfloor \geqq k-1 \qquad (i = 1, 2, ..., s)$$

by (2), i.e., there exist vertex disjoint complete subgraphs $G_1'', ..., G_s''$ of $k-1$ vertices in $G\big(N(x)\big)$. Thus the edges incident to x can be covered by s edge disjoint K_k's $\big(G\big(V(G_i'')\cup\{x\}\big)$, $i=1, ..., s\big)$ and q edges. The sum of the orders of those complete graphs is $ks+2q$. Let $G_0=G-x-\bigcup_{i=1}^{s} E(G_i'')$. Using the inequalities $k \geqq 4$, $s \leqq t$, $q \leqq k-2$, $r \geqq 0$ and that $q \leqq r$ if $s=t$, we obtain by the induction hypothesis that

$$p_k(G) \leqq ks+2q+p_k(G_0) \leqq (2k-4)t+2r+2t_{k-1,n-1} = 2t_{k-1,n} \tag{3}$$

with equality only if $k=4$, $G_0=T_{k-1,n-1}$ and either $s=t$, $q=r$ (and so $d=n-1$, $G=K_n$) or $s=t-1$, $r=0$, $q=k-2$ (and so $d=n-2$).

In the first case $G-x=K_{n-1}$ and it is easy to see that $G=K_{10}$ and there can be found five edge disjoint K_4's implying

$$p_4(K_{10}) \leqq 50 < 2t_{3,10} = 66.$$

In the second case, $d=n-2$, i.e. there exists a vertex $y\in V(G-x)$ such that $xy\notin E(G)$, $y\notin N(x)$. Then y is not covered by the subgraphs $G_1'', \dots, G_s''$ and by the edges incident to x. Thus $d_{G_0}(y)=d_G(y)$. Now $d_{G_0}(y)=n-2$ by the minimality of $d_G(x)$ and by $V(G_0)=n-1$. On the other hand, as $s\geqq 1$, there exists a $z\in V(G_1'')$ with $d_{G_0}(z)\leqq n-4\leqq d_{G_0}(y)-2$. Since the degrees of the vertices of $T_{3,n-1}$ differ from each other by at most one, the contradiction $G_0\neq T_{3,n-1}$ follows.

Thus we have proved that strict inequality holds somewhere in (3) which completes the proof of the theorem. ∎

3. Decomposition into triangles

In this section we prove Theorem 2. We derive it from the following stronger theorem. Define

$$c(n)=\begin{cases}0, & \text{if } n\equiv 1 \text{ or } 3 \pmod 6,\\ 4/3, & \text{if } n\equiv 5 \pmod 6,\\ n/6, & \text{if } n\equiv 0 \text{ or } 2 \pmod 6,\\ n/6+1/3, & \text{if } n\equiv 4 \pmod 6.\end{cases}$$

THEOREM 3. *Every graph of n vertices and $m>n^2/4$ edges contains at least* $\dfrac{4m^2-mn^2}{3n}\left(\dfrac{1}{n-2}-\dfrac{c(n)}{\binom{n}{3}}\right)$ *(pairwise) edge disjoint triangles.*

Moreover, we have

PROPOSITION 4. *Let $t(G)$ denote the number of triangles (subgraphs isomorphic to K_3) in an arbitrary graph G of n vertices. Then G contains at least* $\dfrac{t(G)}{n-2}-\dfrac{c(n)}{\binom{n}{3}}t(G)$ *edge disjoint triangles.*

PROOF OF PROPOSITION 4. It has been proved by Spencer [9] that in the complete graph K_n of n vertices there can be found a (partial) Steiner system $\mathbf{S}_n$ of $s_n=\frac{1}{3}\binom{n}{2}-$ $-c(n)$ edge disjoint triangles.

Let G be a graph of n vertices and denote by $\mathbf{T}$ the set of triangles of G (then $|\mathbf{T}|=t(G)$). For an arbitrary permutation π of $V(G)$, put $t_\pi=|\mathbf{T}\cap\pi(\mathbf{S}_n)|$ where $\pi(\mathbf{S}_n)$ is the image of $\mathbf{S}_n$ after applying π.

Observe that for each pair T, T' of triangles satisfying $T\in\mathbf{T}$ and $T'\subset K_n$ there are exactly $6(n-3)!$ permutations such that $T=\pi(T')$ and this number *does not depend* on the choice of T and T'. On the other hand, $\mathbf{S}_n$ contains s_n of the $\binom{n}{3}$ triangles of K_n, therefore the average value of t_π is $t(G)s_n/\binom{n}{3}$. Consequently, there exists a π for which $\pi(\mathbf{S}_n)$ contains at least $t(G)s_n/\binom{n}{3}$ edge disjoint triangles of G. ∎

PROOF OF THEOREM 3. If $m \leqq n^2/4$, the expression is not positive and we have nothing to prove. Otherwise, if $n^2/4 < m \leqq \binom{n}{2}$, a result of Moon and Moser [8] states that G contains at least $t(G) \geqq (4m^2 - mn^2)/3n$ triangles, therefore Proposition 4 implies the validity of our theorem. ∎

PROOF OF THEOREM 2. The statement is obvious for $n \leqq 5$, so assume $n > 5$. Moreover, $p_3(G) \leqq 2m = 2|E(G)|$, therefore we may suppose $n^2/4 < m \leqq \binom{n}{2}$.

If $E(G)$ is decomposed into K_3's and K_2's, the edges of K_3's are counted once in $p_3(G)$ while the other edges are counted twice. Therefore, if $t_0(G)$ denotes the maximal number of edge disjoint triangles in G, we have

$$p_3(G) = 2|E(G)| - 3t_0(G).$$

Put $q_n = (n^2 - 2n - 2)/(n^2 - 3n + 2)$, then $1 < q_n < 4$ for every $n \geqq 5$. Now Theorem 3 gives $t_0(G) \geqq m(4m - n^2)q_n/3n^2$, therefore putting $r = m/n^2$ we gain

$$p_3(G) \leqq 2m - m(4m - n^2)q_n/n^2 = n^2 r(2 + q_n - 4q_n r) \leqq$$

$$\leqq \frac{n^2}{4q_n}\left(\frac{2+q_n}{2}\right)^2 = \frac{n^2}{16}\left(q_n + 4 + \frac{4}{q_n}\right) < \frac{9}{16} n^2. \quad ∎$$

REFERENCES

[1] BOLLOBÁS, B., On complete subgraphs of different orders, *Math. Proc. Cambridge Philos. Soc.* **79** (1976), 19—24. *MR* **55** #12572.

[2] CHUNG, F. R. K., On the decomposition of graphs, *SIAM J. Algebraic and Discrete Methods* **2** (1981), 1—12.

[3] ERDŐS, P., GOODMAN, A. and PÓSA, L., The representation of graphs by set intersections, *Canad. J. Math.* **18** (1966), 106—112. *MR* **32** #4034.

[4] GYŐRI, E. and KOSTOCHKA, A. V., On a problem of G. O. H. Katona and T. Tarján, *Acta Math. Acad. Sci. Hungar.* **34** (1979), 321—327. *MR* **83h:** 05052.

[5] HAJNAL, A. and SZEMERÉDI, E., Proof of a conjecture of Erdős, *Combinatorial Theory and its Applications*, Coll. Math. Soc. J. Bolyai 4, North-Holland, Amsterdam, 1970, 601—623.

[6] KIRKMAN, T. P., On a problem in combinatorics, *Cambridge and Dublin Math. J.* **2** (1847), 191—204.

[7] MANTEL, W., Problem 28, solution by H. Gouwentak, W. Mantel, J. Teixera de Mattes, F. Schuh and W. A. Wythoff, *Wiskundige Opgaven* **10** (1907), 60—61.

[8] MOON, J. W. and MOSER, L., On a problem of Turán, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 283—286.

[9] SPENCER, J., Maximal consistent families of triplets, *J. Combinatorial Theory* **5** (1968), 1—8. *MR* **37** #2616.

[10] TURÁN, P., On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok* **48** (1941), 436—452. *MR* **8**—284.

[11] TUZA, ZS., Covering of graphs by complete bipartite subgraphs; complexity of 0—1 matrices, *Combinatorica* **4** (1984), 111—116. *MR* **85m:** 05045.

(Received November 16, 1983)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST

MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE
VICTOR HUGO U. 18—22
H—1132 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 321—332

LIMIT THEOREMS FOR ERDŐS—RÉNYI TYPE PROBLEMS

E. CSÁKI, A. FÖLDES and J. KOMLÓS

Introduction

A simple theorem of Rényi says that the length of the longest run in n tosses of a fair coin is about $\log n/\log 2$. This was generalized to excessive blocks by the Erdős—Rényi [4] laws of large numbers. A block $B_i=(X_i, X_{i+1}, \dots, X_{i+m-1})$ of the i.i.d random variables $X_1, X_2, \dots$ is called excessive if the average $\frac{1}{m}\sum_{j=i}^{i+m-1} X_j$ is greater than x, where x is some number greater than $\mathsf{E}X_1$. Erdős and Rényi have shown that the length m of the longest excessive block of $X_1, \dots, X_n$ is around $f(x)\log n$, where $f(x)=-1/\log \varrho(x)$ and $\varrho(x)$ is the so-called Chernoff function. (An important property of this $f(x)$ is that it determines the distribution of X_1 uniquely.)

This law of large numbers has been extended to a limit theorem by Komlós and Tusnády [8]. The complication in the proofs was caused by the fact that near-by blocks are strongly dependent. This dependence is measured by $p=p(x)$ a function determined by the distribution of X_1.

The method used applies in numerous other situations; the purpose of this paper is to rephrase this method into an explicit "cookbook recipe" and to illustrate it by three examples.

Another Erdős—Rényi type problem was solved by Csáki and Földes [1]: The block B_i is called an α-tube, if $|S_j-S_{i-1}|<\alpha$ for $j=i, i+1, \dots, i+m-1$, where $X_1, X_2, \dots, X_n$ are independent Bernoulli variables and $S_j=X_1+X_2+\dots+X_j$. Then the length of the longest α tube is around $c(\alpha)\log n$, where $c(\alpha)$ is determined by the distribution of X_1. Here we extend this theorem to a limit theorem.

Révész [9] proved that the longest monotone block $(X_i<X_{i+1}<\dots<X_{i+m-1})$ of uniform variables $X_1, \dots, X_n$ has a length around $\log n/\log\log n$. This will also be extended to a limit theorem.

As the simplest example, we will apply the Main Lemma to the original longest run problem (which was earlier extended to a law of large numbers by Erdős and Révész [5]).

The Main Lemma together with standard Borel—Cantelli type arguments yield also upper and lower classes for the random variables concerned. We state these results without proofs since no new idea is needed. For the proof of similar results we refer to Erdős and Révész [5], Guibas and Odlyzko [7], and Samarova [10].

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1808.

1980 *Mathematics Subject Classification.* Primary 60F05; Secondary 60F15.

Key words and phrases. Limit distribution, weakly dependent blocks, longest head run, monotone sequences, narrowest tube of a random walk.

§ 1. The Main Lemma

The general model

Let $X_1, X_2, \dots$ be a sequence of independent random variables, and let $\mathcal{F}_{n,m}$ denote the σ-algebra generated by the block of variables $X_n, X_{n+1}, \dots, X_{n+m-1}$. We will deal with a sequence $A_{n,m}$ of events, where $A_{n,m} \in \mathcal{F}_{n,m}$. We will often suppress m in the notation and write A_n for $A_{n,m}$.

Since A_n is determined by $X_n, \dots, X_{n+m-1}$, the events A_{n_1} and A_{n_2} are independent if $|n_1 - n_2| \geqq m$, but the dependence of "neighbouring events" makes the investigation of the sequence A_n interesting. This dependence is reflected by the constant p in the Main Lemma. Our main goal is to find good approximations to the probabilities $\mathsf{P}(\bigcup_{i=1}^{n} A_i) = 1 - \mathsf{P}(\bar{A}_1 \dots \bar{A}_n)$; in other words, find the limit distribution of the random variable τ_m = first n such that A_n occurs.

The Main Lemma. Although the lemma (implicit in Komlós—Tusnády [8]) is simple algebra, it is the main tool in establishing limit theorems for the above problem. We first state it in a limiting form, then a finite version.

Since in most applications the sequence of events $A_n = A_{n,m}$ is stationary for every fixed m (i.e. $\mathsf{P}(A_{i_1+d} A_{i_2+d} \dots A_{i_k+d})$ is independent of d), we first state the lemma for stationary events.

We adopt the convention that $\mathsf{P}(A_i A_{i+1} \dots A_j) = 1$ if $j < i$.

MAIN LEMMA (stationary case, limit form). *If, for any fixed m, A_n is stationary, and*

(i)
$$p = \lim_{k\to\infty} \lim_{m\to\infty} \mathsf{P}(\bar{A}_2 \dots \bar{A}_k | A_1) > 0$$

(ii)
$$\lim_{k\to\infty} \sup_{m} \sum_{k \leqq i \leqq 2m} \mathsf{P}(A_i | A_1) = 0$$

(iii)
$$\lim_{m\to\infty} (m\mathsf{P}(A_1)) = 0$$

then

$$\lim_{m\to\infty} \frac{\mathsf{P}(\bar{A}_2 \dots \bar{A}_n | A_1)}{\mathsf{P}(\bar{A}_2 \dots \bar{A}_n)} = p$$

uniformly in n. Consequently, if $n(m)$ satisfies $\lim_{m\to\infty} n(m)\mathsf{P}(A_1) = \lambda$, *then* (i), (ii), (iii) *imply*

$$\lim_{m\to\infty} \mathsf{P}(\bar{A}_1 \bar{A}_2 \dots \bar{A}_{n(m)}) = e^{-p\lambda}.$$

REMARK. (ii) and (iii) are technical conditions, the essential task is to determine the number p in (i). If we want to use this limit theorem for getting strong laws, we have to worry about its range; i.e. to see whether we can replace λ in the lemma with $\lambda(m)$, where $\lambda(m)$ diverges to infinity or converges to 0 at a certain rate. To this, we need the following finite form of the lemma (which, of course, will imply the original limit form). Another flexibility of this finite form is that we have the freedom of choosing k as an appropriate function of m (e.g. $k = m/2$); while in the limit form k was kept fixed as m approached infinity. ε will denote sufficiently small (in terms of p) positive numbers, and m is thought to be fixed.

MAIN LEMMA (stationary case, finite form). *Assume that A_n is stationary (m is fixed), and there is a number p, $0<p\leqq 1$, such that the following three conditions hold for some $k<m$, $\varepsilon>0$:*

(SI) $$|\mathsf{P}(\bar{A}_2 \dots \bar{A}_k|A_1)-p| < \varepsilon$$

(SII) $$\sum_{k\leqq i\leqq 2m} \mathsf{P}(A_i|A_1) < \varepsilon$$

(SIII) $$\mathsf{P}(A_1) < \varepsilon/m.$$

Then, for all $N>1$,

$$\left|\frac{\mathsf{P}(\bar{A}_2 \dots \bar{A}_N|A_1)}{\mathsf{P}(\bar{A}_2 \dots \bar{A}_N)}-p\right| < 7\varepsilon$$

and

$$e^{-(p+10\varepsilon)N\mathsf{P}(A_1)-2m\mathsf{P}(A_1)} < \mathsf{P}(\bar{A}_1 \dots \bar{A}_N) < e^{-(p-10\varepsilon)N\mathsf{P}(A_1)+2m\mathsf{P}(A_1)}.$$

The above lemma can easily be extended to non-stationary events, and we are going to prove the lemma in this last, most general form.

MAIN LEMMA (non-stationary case, finite form). *Assume that there is a number p, $0<p\leqq 1$, such that the following three conditions hold. For some $k<m$, $\varepsilon>0$, and all n, $m<n\leqq N$:*

(NI) $$|\mathsf{P}(\bar{A}_{n-1}\bar{A}_{n-2} \dots \bar{A}_{n-k}|A_n)-p| < \varepsilon$$

(NII) $$\sum_{n-2m<i<n-k} \mathsf{P}(A_i|A_1) < \varepsilon$$

(NIII) $$\max_{1\leqq i\leqq N} \mathsf{P}(A_i) < \varepsilon/m.$$

Then, for $m<n\leqq N$,

$$\left|\frac{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1})}-p\right| < 7\varepsilon$$

whence

$$e^{-(p+10\varepsilon)\lambda-2\sum_{i=1}^{m}\mathsf{P}(A_i)} < \mathsf{P}(\bar{A}_1 \dots \bar{A}_N) < e^{-(p-10\varepsilon)\lambda}$$

where

$$\lambda = \sum_{i=m+1}^{N} \mathsf{P}(A_i).$$

PROOF OF THE MAIN LEMMA. Let $n=lm+r$ $(0\leqq r<m)$ and denote

$$E_i = \bar{A}_1\bar{A}_2 \dots \bar{A}_{n-mi} \qquad (i = 0, 1, \dots, l).$$

Suppose first that $l \geqq 3$. Then by (NI)

$$(1.1)\qquad \frac{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1})} \leqq \frac{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-2m}\bar{A}_{n-k}\bar{A}_{n-k+1} \dots \bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1})} \leqq$$

$$\leqq \frac{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-2m})}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_n)} \mathsf{P}(\bar{A}_{n-k} \dots \bar{A}_{n-1}|A_n) < \frac{\mathsf{P}(E_2)}{\mathsf{P}(E_0)}(p+\varepsilon).$$

On the other hand

$$\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1}|A_n) \geqq \mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-2m}\bar{A}_{n-k}\bar{A}_{n-k+1} \dots \bar{A}_{n-1}|A_n) -$$

$$- \sum_{n-2m<j<n-k} \mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-2m}A_j\bar{A}_{n-k}\bar{A}_{n-k+1} \dots \bar{A}_{n-1}|A_n) \geqq$$

$$\geqq \mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-2m})\mathsf{P}(\bar{A}_{n-k} \dots \bar{A}_{n-1}|A_n) - \sum_{n-2m<j<n-k} \mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-3m}A_j|A_n).$$

Hence by (NI) and (NII)

$$(1.2)\qquad \frac{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1})} > \frac{\mathsf{P}(E_2)}{\mathsf{P}(\bar{A}_1 \dots \bar{A}_{n-1})}(p-\varepsilon) - \frac{\mathsf{P}(E_3)}{\mathsf{P}(E_0)}\varepsilon \geqq$$

$$\geqq (p-\varepsilon) - \frac{\mathsf{P}(E_3)}{\mathsf{P}(E_0)}\varepsilon.$$

Now we show that $\mathsf{P}(E_2)/\mathsf{P}(E_0)$ and $\mathsf{P}(E_3)/\mathsf{P}(E_0)$ are around 1. This part of the proof (quoted from Komlós—Tusnády [8]) can be considered as an early version of the combinatorial Lovász' Lemma (see Erdős—Lovász [3]). We start from the inequality

$$\mathsf{P}(E_i) = \mathsf{P}(E_{i+1}) - \mathsf{P}\Big(\bigcup_{n-m(i+1)<j\leqq n-mi} E_{i+1}A_j\Big) \geqq$$

$$\geqq \mathsf{P}(E_{i+1}) - \mathsf{P}(E_{i+2}) \sum_{n-m(i+1)<j\leqq n-mi} \mathsf{P}(A_j) \geqq \mathsf{P}(E_{i+1}) - \varepsilon\mathsf{P}(E_{i+2})$$

whence

$$(1.3)\qquad \mathsf{P}(E_i|E_{i+1}) \geqq 1 - \frac{\varepsilon}{\mathsf{P}(E_{i+1}|E_{i+2})}.$$

Since

$$\mathsf{P}(E_{l-1}|E_l) = \frac{\mathsf{P}(E_{l-1})}{\mathsf{P}(E_l)} \geqq \mathsf{P}(E_{l-1}) \geqq 1 - \sum_{j=1}^{2m} \mathsf{P}(A_j) > 1 - 2\varepsilon \geqq \frac{1}{2}$$

for $\varepsilon \leqq \frac{1}{4}$, from (1.3) we have by induction on i, the inequality

$$\mathsf{P}(E_i|E_{i+1}) \geqq \frac{1}{2} \quad \text{for all} \quad i,$$

whence

$$(1.4)\qquad \mathsf{P}(E_i|E_{i+1}) \geqq 1 - 2\varepsilon \quad \text{for all} \quad i.$$

(In what follows we will have various inequalities which hold if ε is small enough.) All of these hold if $\varepsilon<\varepsilon_0=\frac{1}{42}$. In particular

(1.5) $$\mathsf{P}(E_2)/\mathsf{P}(E_0) = \mathsf{P}(E_2|E_1)\mathsf{P}(E_1|E_0) \leqq \frac{1}{(1-2\varepsilon)^2} < 1+5\varepsilon$$

and similarly

(1.6) $$\mathsf{P}(E_3)/\mathsf{P}(E_0) \leqq \frac{1}{(1-2\varepsilon)^3} < 1+7\varepsilon$$

if $\varepsilon<\varepsilon_0$. Now from (1.1) and (1.5) we have

$$\frac{\mathsf{P}(E_2)}{\mathsf{P}(E_0)}(p+\varepsilon) < (1+5\varepsilon)(p+\varepsilon) < p+7\varepsilon$$

if $\varepsilon<\varepsilon_0$, and from (1.2) and (1.6)

$$p-\varepsilon-\frac{\mathsf{P}(E_3)}{\mathsf{P}(E_0)}\varepsilon > p-\varepsilon-(1+7\varepsilon)\varepsilon > p-3\varepsilon$$

if $\varepsilon<\varepsilon_0$, proving the first statement of our lemma, if $l\geqq 3$.

Consider now the case $l<3$. Then we have by (NI) and (NIII)

$$\frac{\mathsf{P}(\bar{A}_1\bar{A}_2\ldots\bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1})} \leqq \frac{\mathsf{P}(\bar{A}_{n-k}\ldots\bar{A}_{n-1}|A_n)}{1-\mathsf{P}(A_1\cup\ldots\cup A_{n-1})} < \frac{p+\varepsilon}{1-3\varepsilon} < p+7\varepsilon$$

if $\varepsilon<\varepsilon_0$.

On the other hand (adopting the convention that $A_j=\emptyset$ for $j\leqq 0$)

$$\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1}|A_n) \geqq \mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-2m}\bar{A}_{n-k}\ldots\bar{A}_{n-1}|A_n)-$$

$$-\sum_{n-2m<j<n-k}\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-2m}A_j\bar{A}_{n-k}\ldots\bar{A}_{n-1}|A_n) \geqq \mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-2m})(p-\varepsilon)-\varepsilon.$$

Hence

$$\frac{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1}|A_n)}{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1})} \geqq p-\varepsilon-\frac{\varepsilon}{1-3\varepsilon} > p-3\varepsilon$$

if $\varepsilon<\varepsilon_0$, which proves the first statement when $l<3$. To prove the second statement of our lemma start from the identity

$$\mathsf{P}(\bar{A}_1\ldots\bar{A}_N) = \mathsf{P}(\bar{A}_1\ldots\bar{A}_m)\prod_{n=m+1}^{N}(1-P(A_n|\bar{A}_1\ldots\bar{A}_{n-1})).$$

Then we have (using $1-x\leqq e^{-x}$)

$$\mathsf{P}(\bar{A}_1\ldots\bar{A}_m)\prod_{n=m+1}^{N}\left(1-\frac{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1}A_n)}{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1})}\right) =$$

$$= \mathsf{P}(\bar{A}_1\ldots\bar{A}_m)\prod_{n=m+1}^{N}\left(1-\frac{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1}|A_n)\mathsf{P}(A_n)}{\mathsf{P}(\bar{A}_1\ldots\bar{A}_{n-1})}\right) \leqq$$

$$\leqq e^{-\sum_{n=m+1}^{N}(p-3\varepsilon)\mathsf{P}(A_n)}.$$

To get the lower bound apply the inequality

$$1-x \geqq e^{-x-x^2}, \qquad 0 \leqq x \leqq \frac{1}{3}.$$

Using (NIII) we get for $\varepsilon<\varepsilon_0$;

$$\mathsf{P}(\bar{A}_1 \dots \bar{A}_N) \geqq$$

$$\geqq \Big(1-\mathsf{P}\big(\bigcup_{j=1}^{m} A_j\big)\Big)\exp\Big\{-\sum_{n=m+1}^{N}\{(p+7\varepsilon)\mathsf{P}(A_n)+(p+7\varepsilon)^2\mathsf{P}^2(A_N)\}\Big\} \geqq$$

$$\geqq \Big(1-\sum_{j=1}^{m}\mathsf{P}(A_j)\Big)\exp\Big\{-\sum_{n=m+1}^{N}\{(p+7\varepsilon)\mathsf{P}(A_n)+(p+7\varepsilon)^2\varepsilon\mathsf{P}(A_n)\}\Big\} \geqq$$

$$\geqq \exp\Big\{-\sum_{j=1}^{m}\mathsf{P}(A_j)-\Big(\sum_{j=1}^{m}\mathsf{P}(A_j)\Big)^2\Big\}\exp\Big\{-\sum_{n=m+1}^{N}\mathsf{P}(A_n)(p+10\varepsilon)\Big\} \geqq$$

$$\geqq \exp\Big\{-2\sum_{j=1}^{m}\mathsf{P}(A_j)-(p+10\varepsilon)\lambda\Big\}.$$

§ 2. Two simple examples

1. To illustrate how the lemmas work, we first choose the simplest possible example, that of the original longest run problem. Let X_n be a Bernoulli sequence, $\mathsf{P}(X_n=1)=\alpha$, $\mathsf{P}(X_n=-1)=1-\alpha$. If $A_n=A_{nm}$ stands for the event $X_n=X_{n+1}=$ $=\dots=X_{n+m-1}=1$, then

$$\mathsf{P}(A_n)=\alpha^m, \quad \mathsf{P}(\bar{A}_2\dots\bar{A}_k|A_1)=\mathsf{P}(X_{m+1}=-1)=1-\alpha, \qquad 2\leqq k\leqq m$$

thus $p=1-\alpha$. Conditions (ii) and (iii) follow from the formula

$$\mathsf{P}(A_i|A_1)=\begin{cases}\alpha^{i-1} & \text{for } 2\leqq i\leqq m\\ \alpha & \text{for } i>m.\end{cases}$$

Thus, the Main Lemma leads to the following limit distribution for τ_m, the initial index of the first run of length m:

$$\lim_{m\to\infty}\mathsf{P}(\tau_m>\lambda\alpha^{-m})=e^{-(1-\alpha)\lambda}.$$

Of course, this can also be obtained easily from the well-known explicit formula for $\mathsf{P}(\tau_m>n)$.

The application of the finite form of the lemma with $k=m$, $\varepsilon=2m\alpha^m$, and standard arguments (Borel—Cantelli lemma, Chebyshev's inequality) lead to the following strong laws of large numbers:

$$\text{(2.1)} \qquad \mathsf{P}\left(\tau_m<\varphi(m)\frac{\alpha^{-m}}{1-\alpha} \text{ for infinitely many } m\right)=1 \text{ or } 0$$

according to whether $\sum\varphi(m)$ is divergent or convergent, and

$$\mathsf{P}\left(\tau_m > \varphi(m)\,\frac{\alpha^{-m}}{1-\alpha} \quad \text{for infinitely many } m\right) = 1 \text{ or } 0 \tag{2.2}$$

according to whether $\sum e^{-\varphi(m)}$ is divergent or convergent.

Hence, for $m(n)$, the length of the longest run up to n, we have with probability 1,

$$\begin{aligned} m(n) &\geqq \lfloor f(n)-\varepsilon\rfloor \quad \text{for} \quad n > n_0(\omega) \\ m(n) &< \lceil f(n)+\varepsilon\rceil \quad \text{for infinitely many } n, \end{aligned} \tag{2.3}$$

where $f(n)=(\log n-\log\log\log n+\log(1-\alpha))/\log(1/\alpha)$ and $\lfloor . \rfloor$, $\lceil . \rceil$ stand for the floor and ceiling functions, respectively (rounding down, rounding up).

The above statements are known and can be found in the quoted papers of Erdős—Révész [5], Guibas—Odlyzko [7], Samarova [10].

2. Longest monotone block:

$$A_{nm} = \{X_n < X_{n+1} < \ldots < X_{n+m-1}\},$$

where X_i are i.i.d. uniform random variables. We have, for $2\leqq k\leqq m$,

$$\mathsf{P}(\bar{A}_2 \ldots \bar{A}_k | A_1) = \mathsf{P}(X_{m+1} < X_m | A_1) = \frac{m}{m+1}$$

thus $p=1$. (ii) and (iii) are easily checked.

The Main Lemma implies

$$\lim_{m\to\infty} \mathsf{P}(\tau_m > \lambda m!) = e^{-\lambda}.$$

The finite form $(k=m,\ \varepsilon=1/m)$ leads to the strong laws (2.1), (2.2) and (2.3) with $\dfrac{\alpha^{-m}}{1-\alpha}$ replaced by $m!$, and $f(n)$ by the inverse of the (continuous) $n=m!$.

§ 3. The tube for simple symmetric random walk

Now consider i.i.d random variables $X_1, X_2, \ldots$ with the distribution

$$\mathsf{P}(X_1 = +1) = \mathsf{P}(X_1 = -1) = 1/2,$$

and put $S_0=0$, $S_i=X_1+\ldots+X_i$ $(i\geqq 1)$. Let the events A_j be defined by

$$A_j = \left\{\max_{0\leqq i\leqq m-1} |S_{j+i}-S_{j-1}| < \alpha\right\} \qquad (j = 1, 2, \ldots)$$

where m and α are integers. The block $\{X_j, X_{j+1}, \ldots, X_{j+m-1}\}$ if A_j holds is called an α-tube of length m. From a formula (see Ellis [2]) it follows that

$$\mathsf{P}(A_1) = \frac{1}{\alpha}\sum_{k=1}^{2\alpha-1}\left(\cos\frac{k\pi}{2\alpha}\right)^m \sin\frac{k\pi}{2}\,\frac{1+\cos\dfrac{k\pi}{2\alpha}}{\sin\dfrac{k\pi}{2\alpha}}\left(\frac{1-(-1)^k}{2}\right) \tag{3.1}$$

and that (see Csáki and Földes [1]):

$$(3.2)\qquad K_1\left(\cos\frac{\pi}{\alpha+\beta}\right)^m \leqq \mathsf{P}(-\beta < S_k < \alpha,\ k = 1, \dots, m) \leqq K_2\left(\cos\frac{\pi}{\alpha+\beta}\right)^m$$

with positive constants K_1, K_2 depending only on $\alpha+\beta$ but not on m.

Denote further

$$\tau_m = \min\left(n\colon \max_{0\leqq i\leqq m-1} |S_{n+i}-S_{n-1}| < \alpha\right)$$

$$\nu_n = \max\left(m\colon \min_{1\leqq j\leqq n}\max_{0\leqq i\leqq m-1} |S_{j+i}-S_{j-1}| < \alpha\right),$$

i.e. ν_n is the length of the longest α-tube starting not later than n.

In Csáki and Földes [1] it was shown that with probability 1 for large enough N,

$$\min_{0\leqq j\leqq N-a_N}\max_{0\leqq j\leqq a_N} |S_{j+i}-S_j| = [\alpha(c)]$$

where $a_N=[c\log N]$ and $\alpha(c)$ is defined by $\cos(\pi/(2\alpha))=\exp(-1/c)$, if $\alpha(c)$ is not an integer. Moreover

$$\min_{0\leqq j\leqq N-a_N}\max_{0\leqq j\leqq a_N} |S_{j+1}-S_j|$$

takes the values $\alpha(c)-1$ and $\alpha(c)$ infinitely often and all other values only finitely often if $\alpha(c)$ is an integer.

To verify the conditions of the Main Lemma (stationary case, finite form) we prove our

LEMMA 3.1. *For* $1<k<m$

$$(3.3)\qquad \left|\mathsf{P}(\bar{A}_2\dots\bar{A}_k|A_1)-\frac{1}{\alpha}\right| < K_3(\varrho_1^k+k^2\varrho_2^{m-k})$$

$$(3.4)\qquad \sum_{k\leqq i\leqq 2m}\mathsf{P}(A_i|A_1) < K_4\left(\left(\cos\frac{\pi}{2\alpha}\right)^k + m\left(\cos\frac{\pi}{2\alpha}\right)^m\right)$$

$$(3.5)\qquad \mathsf{P}(A_1) < K_2\left(\cos\frac{\pi}{2\alpha}\right)^m$$

with some constants $K_2, K_3, K_4>0$, $0<\varrho_1<1$, $0<\varrho_2<1$.

PROOF. Let T denote the first return to zero of the random walk $S_1, S_2, \dots$. Then we can write

$$\mathsf{P}(\bar{A}_2\dots\bar{A}_k|A_1) = \sum_{0<2j\leqq k}\mathsf{P}(\bar{A}_2\dots\bar{A}_k, T=2j|A_1)+\mathsf{P}(\bar{A}_2\dots\bar{A}_k,\ T>k|A_1) =$$

$$= \sum_{0<2j\leqq k}\mathsf{P}(\bar{A}_{2j},\ T=2j\,|A_1)-\sum_{0<2j\leqq k}\mathsf{P}((A_2\cup\dots\cup A_k)\bar{A}_{2j},\ T=2j\,|A_1)+$$

$$+\mathsf{P}(\bar{A}_2\dots\bar{A}_k,\ T>k|A_1).$$

Hence

$$Q = |\mathsf{P}(\bar{A}_2 \dots \bar{A}_k | A_1) - \sum_{0<2j\leq k} \mathsf{P}(\bar{A}_{2j},\ T = 2j\,|A_1)| \leq$$

$$\leq \mathsf{P}(T > k|A_1) + \sum_{0<2j\leq k} \sum_{i=2}^{k} \mathsf{P}(A_i \bar{A}_{2j},\ T = 2j\,|A_1).$$

It is easy to see that

$$\mathsf{P}(A_1 A_i \bar{A}_{2j},\ T = 2j) \leq$$

$$\leq 2\mathsf{P}(-\alpha < S_l < \alpha,\ l = 1, \dots, k,\ -\alpha < S_l < \alpha - 1,\ l = k+1, \dots, m) \leq$$

$$\leq 2K_2^2 \left(\cos \frac{\pi}{2\alpha}\right)^k \left(\cos \frac{\pi}{2\alpha - 1}\right)^{m-k}$$

for $2 \leq i \leq k$, and since $\mathsf{P}(A_1) \geq K_1 \left(\cos \frac{\pi}{2\alpha}\right)^m$, we have

$$\sum_{0<2j\leq k} \sum_{i=2}^{k} \mathsf{P}(A_i \bar{A}_{2j},\ T = 2j\,|A_1) \leq k^2 \frac{K_2^2}{K_1} \left(\frac{\cos \frac{\pi}{2\alpha - 1}}{\cos \frac{\pi}{2\alpha}}\right)^{m-k}.$$

Obviously

$$\mathsf{P}(T > k|A_1) \leq \frac{K_2}{K_1} \left(\frac{\cos \frac{\pi}{\alpha}}{\cos \frac{\pi}{2\alpha}}\right)^k,$$

hence

$$Q \leq K_1^* (\varrho_1^k + k^2 \varrho_2^{m-k})$$

with some constants $K_1^* > 0$, $0 < \varrho_1 < 1$ and $0 < \varrho_2 < 1$.

Furthermore

$$\sum_{0<2j\leq k} \mathsf{P}(\bar{A}_{2j},\ T = 2j\,|A_1) = 1 - \mathsf{P}(T > k|A_1) - \sum_{0<2j\leq k} \mathsf{P}(A_{2j},\ T = 2j\,|A_1) =$$

$$= 1 - \mathsf{P}(T > k|A_1) - \sum_{0<2j\leq k} \frac{\mathsf{P}(A_1 A_{2j},\ T = 2j)}{\mathsf{P}(A_1)}.$$

It is easy to see that

$$\mathsf{P}(A_1 A_{2j},\ T = 2j) = \mathsf{P}(|S_l| < \alpha,\ l = 1, \dots, 2j,\ T = 2j)\mathsf{P}(A_1)$$

therefore

$$\sum_{0<2j\leq k} \mathsf{P}(\bar{A}_{2j},\ T = 2j\,|A_1) = 1 - \mathsf{P}(T > k|A_1) -$$

$$- \mathsf{P}(|S_l| < \alpha,\ l = 1, 2, \dots, T) + \mathsf{P}(|S_l| < \alpha,\ l = 1, 2, \dots, T,\ T > k).$$

The event $\{|S_l| < \alpha,\ l = 1, 2, \dots, T\}$ means that the random walk returns to 0 before reaching $+\alpha$ or $-\alpha$ and this has the probability $(\alpha - 1)/\alpha$. On the other hand,

$$P(|S_l| < \alpha,\ l = 1, 2, \ldots, T,\ T > k) \leqq$$

$$\leqq P(0 < S_l < \alpha,\ l = 1, \ldots, k) \leqq K_2 \left(\cos \frac{\pi}{\alpha}\right)^k.$$

Combining these estimates, we get (3.3).

To get (3.4) we can easily see that

$$\sum_{k \leqq i \leqq 2m} P(A_i|A_1) = \sum_{k \leqq i \leqq 2m} \frac{P(A_1 A_i)}{P(A_1)} \leqq$$

$$\leqq \sum_{k \leqq i \leqq m} \frac{K_1^2 \left(\cos \frac{\pi}{2\alpha}\right)^{i+m}}{K_2 \left(\cos \frac{\pi}{2\alpha}\right)^m} + K_1 m \left(\cos \frac{\pi}{2\alpha}\right)^m \leqq K^* \left(\left(\cos \frac{\pi}{2\alpha}\right)^k + m \left(\cos \frac{\pi}{2\alpha}\right)^m\right).$$

Finally, (3.5) follows from (3.2).

In order to obtain limit distributions and upper and lower classes for τ_m and ν_n, we note that from (3.1) one obtains

$$P(A_1) \sim \frac{\varrho^m}{\alpha \sin \frac{\pi}{2\alpha}} (1 + \varrho + (-1)^{m+\alpha-1}(1-\varrho)) =$$

$$= \begin{cases} \dfrac{2\varrho^m}{\alpha \sin \frac{\pi}{2\alpha}} & \text{if} \quad m+\alpha-1 \quad \text{is even} \\[2ex] \dfrac{2\varrho^{m+1}}{\alpha \sin \frac{\pi}{2\alpha}} & \text{if} \quad m+\alpha-1 \quad \text{is odd,} \end{cases}$$

where

$$\varrho = \cos \frac{\pi}{2\alpha}.$$

For fixed integer α, let m run through either even or odd integers so that $m+\alpha-1$ be even. Then from Lemma 3.1 by choosing $k = m/2$ and from the Main Lemma one can easily obtain that for large enough m,

$$\exp\left\{-\frac{2N\varrho^m}{\alpha^2 \sin \frac{\pi}{2\alpha}}(1+\tilde{\varrho}^m)\right\} \leqq P(\bar{A}_1 \bar{A}_2 \ldots \bar{A}_N) \leqq \exp\left\{-\frac{2N\varrho^m}{\alpha^2 \sin \frac{\pi}{2\alpha}}(1-\tilde{\varrho}^m)\right\}$$

with some $0 < \tilde{\varrho} < 1$.

By using standard methods (Borel—Cantelli lemma, etc.) the following results can be obtained.

COROLLARY.

(A) $$\lim_{m\to\infty} \mathsf{P}\left(2\varrho^m \tau_m < x\alpha^2 \sin\frac{\pi}{2\alpha}\right) = 1-e^{-x}, \qquad 0<x<\infty.$$

(B) *Let $n(m)$ be an increasing function of m, such that $n(m)\varrho^m$ is increasing. Then*

$$\mathsf{P}(\tau_m > n(m) \text{ i.o.}) = 0 \quad or \quad 1$$

according as

$$\sum_m{}' \exp\left\{-\frac{2n(m)\varrho^m}{\alpha^2 \sin\frac{\pi}{2\alpha}}\right\} < \infty \quad or \quad = \infty.$$

Furthermore

$$\mathsf{P}(\tau_m < n(m) \text{ i.o.}) = 0 \quad or \quad 1$$

according as

$$\sum_m{}' n(m)\varrho^m < \infty \quad or \quad = \infty.$$

Here $\sum_m{}'$ *means that m runs through even or odd integers so that $m+\alpha-1$ should be even.*

(C) $$\mathsf{P}\left(\nu_n - 2\left[\frac{c}{2}\log n\right] < s\right) = \exp\left[-\frac{2\varrho^{s-2\{c/2\log n\}}}{\alpha^2\sin\frac{\pi}{2\alpha}}\right] + o(1),$$

where $c = -1/(\log\varrho) = -1/\left(\log\cos\frac{\pi}{2\alpha}\right)$ *and s runs through even or odd integers according as $\alpha-1$ is even or odd. $\{x\}$ denotes the fractional part of x.*

(D) *Let $m(n)$ be an increasing function of n, taking on even or odd integers according as $\alpha-1$ is even or odd. Then*

$$\mathsf{P}(\nu_n > m(n) \text{ i.o.}) = 0 \quad or \quad 1$$

according as

$$\sum_n \varrho^{m(n)} < \infty \quad or \quad = \infty.$$

Furthermore for any $\varepsilon>0$ with probability 1,

$$\nu_n < \lceil f(n)+\varepsilon\rceil_{\alpha-1} \quad \text{i.o.}$$

$$\nu_n < \lfloor f(n)-\varepsilon\rfloor_{\alpha-1} \quad \text{f.o.},$$

where

$$f(n) = \frac{\log n - \log\log\log n + \log\left(2/(\alpha^2\sin(\pi/2\alpha))\right)}{\log(1/\varrho)}.$$

$\lceil\cdot\rceil_{\alpha-1}$ *means rounding up to the nearest even or odd integer according as $\alpha-1$ is even or odd and $\lfloor\cdot\rfloor_{\alpha-1}$ means rounding down to the nearest even or odd integer according as $\alpha-1$ is even or odd.*

REFERENCES

[1] Csáki, E. and Földes, A., The Narrowest Tube of a Recurrent Random Walk, *Z. Wahrsch. Verw. Gebiete* **66** (1984), 387—403. *MR* **86b:** 60121.
[2] Ellis, R. E., On the solution of equations in finite differences, *Cambridge Mathematical Journal* **4** (1844), 182.
[3] Erdős, P. and Lovász, L., Problems and results on 3-chromatic hypergraphs and some related questions, *Infinite and Finite Sets* (edited by A. Hajnal, R. Rado, V. T. Sós), 609—628, North-Holland, 1975. *MR* **52** #2938.
[4] Erdős, P. and Rényi, A., On a new law of large numbers, *J. Analyse Math.* **23** (1970), 103—111. *MR* **42** #6907.
[5] Erdős, P. and Révész, P., On the length of the longest head-run, *Topics in Information Theory, Colloq. Math. Soc. János Bolyai,* **16**, 219—228, North-Holland, Amsterdam, 1977. *MR* **57** #17788.
[6] Földes, A., The limit distribution of the length of the longest head run, *Period. Math. Hungar.* **10** (1979), 301—310. *MR* **81b:** 60022.
[7] Guibas, L. J. and Odlyzko, A. M., Long repetitive patterns in random sequences, *Z. Wahrsch. Verw. Gebiete* **53** (1980), 241—262. *MR* **81m:** 60047.
[8] Komlós, J. and Tusnády, G., On sequences of "pure heads", *Ann. Probability* **3** (1975), 608—617. *MR* **55** #11353.
[9] Révész, P., Three problems on the lengths of increasing runs, *Stochastic Process. Appl.* **15** (1983), 169—179. *MR* **84h:** 60064.
[10] Samarova, S. S., On the length of the longest head-run for the Markov chain with two states, *Teor. Verojatnost. i Primenen.* **26** (1981), 510—520 (in Russian). *MR* **83a:** 60104.

(Received March 7, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 333—340

ON THE CONTROL OF STRONGLY NONLINEAR SYSTEMS I

V. KOMORNIK

One of the most interesting problems of optimal control theory is to obtain necessary conditions for the optimality. We are interested in problems where the state equation is a partial differential equation. The first monograph in this field is due to J. L. Lions [8]. In this book well-posed linear problems are studied. Another, recent monograph [6] is consacrated to the control of ill-posed, nonlinear problems. Such problems often occur for example in chemical and biological investigations.

In the case of ill-posed problems the general purpose is to derive necessary conditions of the same type as for well-posed ones. However, the methods are generally very different. The method of Lions [6] is very effective if the nonlinear term of the state equation has polynomial growth but it is not directly applicable for more general (e.g. for exponential) nonlinearities. The aim of this paper is to develop the method of Lions to become applicable for far more general nonlinearities. Thus we will solve a problem raised in [6].

We show that after having changed (slightly) the usual cost function, practically no growth condition is needed to derive necessary conditions. In the polynomial case the two cost functions are very similar but in the general case they may be very different. The new cost function seems to be more adequate to the state equation and is not more complicated than the old one. Another advantage of our approach is that this make possible to avoid the L^p $(p\neq 2)$ regularity theory of the solutions of the state equation in many cases (one can compare in this direction the main result of this paper with Theorem 3.2.1 in [6] in dimension $n\leq 3$).

In this note elliptic problems are considered. Parabolic and hyperbolic systems will be investigated later, in [9] and [10].

Our notations are the same as in [5], [6] or [8]; in particular $W^{m,\gamma}(\Omega)$ denotes the Sobolev space of real functions $f\in L^\gamma(\Omega)$ having all partial derivatives (in distributional sense) of order $\leq m$ in $L^\gamma(\Omega)$; $W_0^{m,\gamma}(\Omega)$ is the closure of $C_c^\infty(\Omega)$.

Throughout this paper let Ω denote a bounded open set in $\mathbf{R}^n$ $(n\in\mathbf{N})$ with the boundary Γ of class C^∞ and let $f\colon \mathbf{R}\to\mathbf{R}$ be an arbitrary function of class C^2. Fix the numbers $1<\alpha<\infty$, $1<\beta<\infty$, $N>0$ arbitrarily and put for brevity $\gamma=\min\{\alpha,\beta\}$. Let $z_d\in L^\alpha(\Omega)$ be arbitrarily given and let U_{ad} be a non-void convex,

1980 *Mathematics Subject Classification.* Primary 93C20; Secondary 35J65.
Key words and phrases. Elliptic differential equation, distributed control.

closed subset in $L^\beta(\Omega)$. Furthermore we put

(1) $$J(v, z) = \frac{1}{\alpha}\|f(z)-z_d\|^\alpha_{L^\alpha(\Omega)} + \frac{N}{\beta}\|v\|^\beta_{L^\beta(\Omega)}.$$

A pair (v, z) is said to be *admissible* if

(2) $$v\in U_{ad},\quad z\in W_0^{1,\gamma}(\Omega),\quad f(z)\in L^\alpha(\Omega)\quad \text{and}\quad \Delta z+f(z)+v=0.$$

We shall assume that

(3) there exists at least one admissible pair.

This is satisfied for example if $U_{ad}=L^\beta(\Omega)$.

A pair (u, y) is said to be *optimal* if it is admissible and if

(4) $$J(u,y)=\inf\{J(v,z)\colon (v,z) \text{ is admissible}\}.$$

The first result of this paper is the following

THEOREM 1. *There exists at least one optimal pair.* ∎

Given an open subset ω of Ω we denote by $\widetilde{\mathscr{D}(\omega)}$ the set of extensions by 0 of the functions from $\mathscr{D}(\omega)=C_0^\infty(\omega)$ to Ω. Assume that

(5) $$\gamma > \frac{n}{2}$$

and

(6) there exists $v_0\in U_{ad}$ and a non-void open subset ω of Ω such that $v_0+\widetilde{\mathscr{D}(\omega)}\subset U_{ad}$.

Under these assumptions we prove the

THEOREM 2. *To any optimal pair (u, y) there exists a triplet (u, y, p) such that*

(7) $$u\in U_{ad}\quad y\in W^{2,\gamma}(\Omega)\cap W_0^{1,\gamma}(\Omega),\quad p\in W^{2,\alpha'}(\Omega)\cap W_0^{1,\alpha'}(\Omega),$$

(8) $$\Delta y+f(y)+u=0,$$

(9) $$\Delta p+f'(y)p+|f(y)-z_d|^{\alpha-1}\operatorname{sgn}(f(y)-z_d)f'(y)=0,$$

(10) $$\int_\Omega (p+N|u|^{\beta-1}\operatorname{sgn} u)(v-u)\,dx \geqq 0,\quad \forall v\in U_{ad}.$$

REMARKS. Theorem 2 solves a problem raised by J. L. Lions in his book [6] (Chapitre 3, Paragraph 16, Problem 25) for the case $f(x)=e^x$. In view of this theorem, to find the optimal pairs it is worthwhile to seek first the triplets (u, y, p) satisfying (7)—(10).

The author is grateful to J. L. Lions for the fruitful discussions.

Before turning to the proof of the theorems, for the reader's convenience we recall an important proposition on the weak solutions of the Dirichlet problem.

PROPOSITION. *Let us give two functions* $f, g \in L^{\mu}(\Omega)$ $(1<\mu<\infty)$ *and assume that*

$$\int_{\Omega} f \Delta\xi dx = \int_{\Omega} g\xi\, dx, \qquad \forall \xi \in W^{2,\mu'}(\Omega) \cap W_0^{1,\mu'}(\Omega). \tag{11}$$

Then $f \in W^{2,\mu}(\Omega) \cap W_0^{1,\mu}(\Omega)$ *and*

$$\|f\|_{W^{2,\mu}(\Omega)} \leqq C\|g\|_{L^{\mu}(\Omega)} \tag{12}$$

where C is a constant.

PROOF. It follows from the (deep) regularity results of S. Agmon, A. Douglis and L. Nirenberg [2] that there exists $f^* \in W^{2,\mu}(\Omega) \cap W_0^{1,\mu}(\Omega)$ such that (11) and (12) hold true for f^* instead of f. Then $f-f^* \in L^{\mu}(\Omega)$ and

$$\int_{\Omega} (f-f^*)\Delta\xi\, dx = 0, \qquad \forall \xi \in W^{2,\mu'}(\Omega) \cap W_0^{1,\mu'}(\Omega).$$

Furthermore, applying again the above regularity theorem, there exists $\xi \in W^{2,\mu'}(\Omega) \cap$ $\cap W_0^{1,\mu'}(\Omega)$ such that

$$\Delta\xi = |f-f^*|^{\mu-1}\operatorname{sgn}(f-f^*).$$

Therefore $\int_{\Omega} |f-f^*|^{\mu}\, dx = 0$ whence $f=f^*$ and the proposition is proved. ∎

PROOF OF THEOREM 1. Let (v_k, z_k) be a minimizing sequence of admissible pairs, i.e. such that

$$J(v_k, z_k) \to \inf\{J(v, z)\colon (v, z) \text{ is admissible}\}.$$

Then by (1), (2) and (12) the sequences

$$\|v_k\|_{L^{\beta}(\Omega)}, \quad \|f(z_k)\|_{L^{\alpha}(\Omega)}, \quad \|z_k\|_{W^{2,\gamma}(\Omega)}$$

are bounded. Passing, if it is needed, to subsequences, we may assume that

$$v_k \to u \quad \text{weakly in } L^{\beta}(\Omega), \tag{13}$$

$$z_k \to y \quad \text{weakly in } W^{2,\gamma}(\Omega). \tag{14}$$

It follows from the Rellich—Kondrasov theorem that the imbedding $W^{2,\gamma}(\Omega) \subset$ $\subset L^1(\Omega)$ is compact; therefore (14) implies that

$$z_k \to y \quad \text{strongly in } L^1(\Omega) \tag{15}$$

and then by the Riesz lemma we may also suppose that

$$z_k \to y \quad \text{almost everywhere in } \Omega. \tag{16}$$

Now $f(z_k)$ is bounded in $L^{\alpha}(\Omega)$ and $f(z_k) \to f(y)$ almost everywhere in Ω, therefore by Lemma 1.3 in [4], Chapitre 1

$$f(z_k) \to f(y) \quad \text{weakly in } L^{\alpha}(\Omega). \tag{17}$$

It follows from (13), (14) and (17) that (u, y) is admissible (we note that U_{ad} is weakly closed) and that

$$J(u, y) \leqq \underline{\lim}\, J(v_k, z_k),$$

i.e. (u, y) is an optimal pair. ∎

Proof of Theorem 2. The proof will be divided into several parts. Let (u, y) be an arbitrarily fixed optimal pair and put for each $\varepsilon > 0$

$$J_\varepsilon(v, z) = \frac{1}{\alpha}\|f(z)-z_d\|^\alpha_{L^\alpha(\Omega)} + \frac{N}{\beta}\|v\|^\beta_{L^\beta(\Omega)} +$$

$$+\frac{1}{\varepsilon\gamma}\|\Delta z+f(z)+v\|^\gamma_{L^\gamma(\Omega)} + \frac{1}{\gamma}\|z-y\|^\gamma_{L^\gamma(\Omega)} + \frac{1}{\beta}\|v-u\|^\beta_{L^\beta(\Omega)}.$$

A pair (v, z) is called *ε-admissible* if

(18) $$v \in U_{\mathrm{ad}}, \quad z \in W^{2,\gamma}(\Omega)\cap W_0^{1,\gamma}(\Omega) \quad \text{and} \quad f(z)\in L^\alpha(\Omega).$$

It follows from the proposition that every admissible pair is also ε-admissible.

A pair $(u_\varepsilon, y_\varepsilon)$ is called *ε-optimal* if it is ε-admissible and if

(19) $$J_\varepsilon(u_\varepsilon, y_\varepsilon) = \inf\{J_\varepsilon(v, z)\colon (v, z) \text{ is } \varepsilon\text{-admissible}\}.$$

Lemma 1. *For each $\varepsilon > 0$ there exists at least one ε-optimal pair $(u_\varepsilon, y_\varepsilon)$.*

Proof. One can repeat the argument used in Theorem 1. ∎

Let us fix for each $\varepsilon > 0$ and ε-optimal pair $(u_\varepsilon, y_\varepsilon)$.

Lemma 2. *The following relations holds true as ε tends to* 0:

(20) $$u_\varepsilon \to u \quad \textit{strongly in} \quad L^\beta(\Omega),$$

(21) $$y_\varepsilon \to y \quad \textit{strongly in} \quad W^{2,\gamma}(\Omega)\cap W_0^{1,\gamma}(\Omega),$$

(22) $$f(y_\varepsilon) \to f(y) \quad \textit{strongly in} \quad L^\alpha(\Omega).$$

Proof. It follows from the obvious relation

(23) $$J_\varepsilon(u_\varepsilon, y_\varepsilon) \leqq J_\varepsilon(u, y) = J(u, y)$$

that the sequences

$$\|u_\varepsilon\|_{L^\beta(\Omega)}, \quad \|f(y_\varepsilon)\|_{L^\alpha(\Omega)}, \quad \|y_\varepsilon\|_{W^{2,\gamma}(\Omega)}$$

are bounded and that

(24) $$\Delta y_\varepsilon + f(y_\varepsilon) + u_\varepsilon \to 0 \quad \text{strongly in } L^\gamma(\Omega).$$

Therefore every subsequence of $(u_\varepsilon, y_\varepsilon)$ has another subsequence such that

(25) $$u_\varepsilon \to \hat{u} \quad \text{weakly in} \quad L^\beta(\Omega),$$

(26) $$y_\varepsilon \to \hat{y} \quad \text{weakly in} \quad W^{2,\gamma}(\Omega)\cap W_0^{1,\gamma}(\Omega),$$

(27) $$y_\varepsilon \to \hat{y} \quad \text{almost everywhere in} \quad \Omega,$$

(28) $$f(y_\varepsilon) \to f(\hat{y}) \quad \text{weakly in} \quad L^\alpha(\Omega).$$

It follows from (23)—(26), (28) that $(\hat{u}, \hat{y})$ is admissible and

$$J(\hat{u}, \hat{y}) \leqq \underline{\lim}\, J(u_\varepsilon, y_\varepsilon) \leqq \overline{\lim}\, J_\varepsilon(u_\varepsilon, y_\varepsilon) \leqq J(u, y).$$

Being (u, y) optimal we conclude

$$J(\hat{u}, \hat{y}) = \lim J(u_\varepsilon, y_\varepsilon) = \lim J_\varepsilon(u_\varepsilon, y_\varepsilon) = J(u, y)$$

whence

$$u_\varepsilon \to u \quad \text{strongly in} \quad L^\beta(\Omega), \tag{29}$$

$$y_\varepsilon \to y \quad \text{strongly in} \quad L^\gamma(\Omega). \tag{30}$$

(25), (26), (29), (30) imply

$$\hat{u} = u, \quad \hat{y} = y. \tag{31}$$

Furthermore (28), (29) (or (25)), (31) and the relation $\lim J(u_\varepsilon, y_\varepsilon) = J(u, y)$ imply (22).

Finally, (24), (29) (which is identical with (20)) and (22) imply

$$\Delta y_\varepsilon \to -f(y) - u \quad \text{strongly in } L^\gamma(\Omega). \tag{32}$$

Using the proposition, (30) and (32) yield (21). ∎

Lemma 3. *The following relations hold true as* ε *tends to* 0:

$$y_\varepsilon \to y \quad \textit{strongly in} \quad C(\bar{\Omega}), \tag{33}$$

$$f(y_\varepsilon) \to f(y) \quad \textit{strongly in} \quad C(\bar{\Omega}), \tag{34}$$

$$f'(y_\varepsilon) \to f'(y) \quad \textit{strongly in} \quad C(\bar{\Omega}). \tag{35}$$

Proof. By the Sobolev imbedding theorem and by property (5) $W^{2,\gamma}(\Omega) \subset C(\bar{\Omega})$, therefore (21) implies (33). (34) and (35) hence follow because f and f' are locally uniformly continuous. ∎

Let us now set

$$p_\varepsilon = \frac{1}{\varepsilon} |\Delta y_\varepsilon + f(y_\varepsilon) + u_\varepsilon|^{\gamma-1} \operatorname{sgn}(\Delta y_\varepsilon + f(y_\varepsilon) + u_\varepsilon),$$

then $p_\varepsilon \in L^{\gamma'}(\Omega)$ for all $\varepsilon > 0$.

Lemma 4. *For any* $\xi \in W^{2,\alpha}(\Omega) \cap W_0^{1,\alpha}(\Omega)$ *we have*

$$\int_\Omega p_\varepsilon(\Delta\xi + f'(y_\varepsilon)\xi) + |f(y_\varepsilon) - z_d|^{\alpha-1} \operatorname{sgn}(f(y_\varepsilon) - z_d) f'(y_\varepsilon)\xi + \\ + |y_\varepsilon - y|^{\gamma-1} \operatorname{sgn}(y_\varepsilon - y)\xi \, dx = 0.$$

Proof. By the optimality of $(u_\varepsilon, y_\varepsilon)$ we have

$$\frac{d}{dt} J_\varepsilon(u_\varepsilon, y_\varepsilon + t\xi)|_{t=0} = 0$$

provided that this derivative exists. And this is true, because, using Lebesgue's dominated convergence theorem we obtain

$$\frac{J_\varepsilon(u_\varepsilon, y_\varepsilon+t\xi)-J_\varepsilon(u_\varepsilon, y_\varepsilon)}{t}=\frac{1}{\alpha}\int_\Omega \frac{|f(y_\varepsilon+t\xi)-z_d|^\alpha-|f(y_\varepsilon)-z_d|^\alpha}{t}\,dx+$$

$$+\frac{1}{\gamma\varepsilon}\int_\Omega \frac{|\Delta(y_\varepsilon+t\xi)+f(y_\varepsilon+t\xi)+u_\varepsilon|^\gamma-|\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon|^\gamma}{t}dx+$$

$$+\frac{1}{\gamma}\int_\Omega \frac{|y_\varepsilon+t\xi-y|^\gamma-|y_\varepsilon-y|^\gamma}{t}\,dx\to\int_\Omega |f(y_\varepsilon)-z_d|^{\alpha-1}\operatorname{sgn}(f(y_\varepsilon)-z_d)f'(y_\varepsilon)\xi+$$

$$+\frac{1}{\varepsilon}|\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon|^{\gamma-1}\operatorname{sgn}(\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon)(\Delta\xi+f'(y_\varepsilon)\xi)+$$
$$+|y_\varepsilon-y|^{\gamma-1}\operatorname{sgn}(y_\varepsilon-y)\xi\,dx$$

and the lemma is proved. ∎

LEMMA 5. *For any* $v\in U_{ad}$ *we have*

$$\int_\Omega (p_\varepsilon+N|u_\varepsilon|^{\beta-1}\operatorname{sgn}u_\varepsilon)(v-u_\varepsilon)+|u_\varepsilon-u|^{\beta-1}\operatorname{sgn}(u_\varepsilon-u)(v-u_\varepsilon)\,dx\geqq 0.$$

PROOF. By the optimality of $(u_\varepsilon, y_\varepsilon)$ we have

$$\frac{d}{dt}J_\varepsilon(u_\varepsilon+t(v-u_\varepsilon), y_\varepsilon)\big|_{t=0}\geqq 0$$

if this derivative exists. (We cannot say more in general because u_ε may be eventually a boundary point of U_{ad}.) Taking into account that the sequence u_ε is bounded in $L^\beta(\Omega)$, by Lebesgue's dominated convergence theorem we obtain

$$\frac{J_\varepsilon(u_\varepsilon+t(v-u_\varepsilon), y_\varepsilon)-J_\varepsilon(u_\varepsilon, y_\varepsilon)}{t}=\frac{N}{\beta}\int_\Omega \frac{|u_\varepsilon+t(v-u_\varepsilon)|^\beta-|u_\varepsilon|^\beta}{t}\,dx+$$

$$+\frac{1}{\gamma\varepsilon}\int_\Omega \frac{|\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon+t(v-u_\varepsilon)|^\gamma-|\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon|^\gamma}{t}\,dx+$$

$$+\frac{1}{\beta}\int_\Omega \frac{|u_\varepsilon-u+t(v-u_\varepsilon)|^\beta-|u_\varepsilon-u|^\beta}{t}\,dx\to\int_\Omega N|u_\varepsilon|^{\beta-1}(\operatorname{sgn}u_\varepsilon)(v-u_\varepsilon)+$$

$$+\frac{1}{\varepsilon}|\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon|^{\gamma-1}\operatorname{sgn}(\Delta y_\varepsilon+f(y_\varepsilon)+u_\varepsilon)(v-u_\varepsilon)+|u_\varepsilon-u|^{\beta-1}\operatorname{sgn}(u_\varepsilon-u)(v-u_\varepsilon)\,dx$$

and the lemma is proved. ∎

LEMMA 6. *The sequence* p_ε *is bounded in* $L^{\gamma'}(\Omega)$.

PROOF. Assume on the contrary that

$$\|p_\varepsilon\|_{L^{\gamma'}(\Omega)}\to\infty \tag{36}$$

for some subsequence. Being the sequence $\frac{p_\varepsilon}{\|p_\varepsilon\|_{L^{\gamma'}(\Omega)}}$ bounded in $L^{\gamma'}(\Omega)$, it is also bounded in $W^{2,\alpha'}(\Omega)$ by Lemmas 3, 4 and by Proposition. Applying (5) and the Rellich—Kondrasov theorem there exists therefore another subsequence such that

$$\frac{p_\varepsilon}{\|p_\varepsilon\|_{L^{\gamma'}(\Omega)}} \to q \quad \text{strongly in} \quad L^{\gamma'}(\Omega). \tag{37}$$

Passing to limit in Lemmas 4, 5 and using (20), (33)—(37) we obtain

$$\int_\Omega q(\Delta\xi + f'(y)\xi)\,dx = 0, \quad \forall \xi \in W^{2,\alpha}(\Omega) \cap W_0^{1,\alpha}(\Omega), \tag{38}$$

$$\int_\Omega q(v-u)\,dx \geqq 0, \quad \forall v \in U_{\mathrm{ad}}. \tag{39}$$

(6) and (39) imply $q\equiv 0$ in ω. Therefore, applying a unicity result of W. O. Amrein, A. M. Berthier and V. Georgescu [1], from (38) we conclude $q\equiv 0$ in Ω. But this contradicts to (37) whence $\|q\|_{L^{\gamma'}(\Omega)}=1$. ∎

Now we are ready to prove our theorem. Applying Lemmas 2, 3, 4, 6 and Proposition, there exists a subsequence p_ε such that

$$p_\varepsilon \to p \quad \text{weakly in} \quad W^{2,\alpha'}(\Omega) \cap W_0^{1,\alpha'}(\Omega). \tag{40}$$

Now (7) follows from (21) and (40) while (8) is obvious because (u, y) is admissible. (9) and (10) follow from Lemmas 4, 5 if we pass to limit and use the relations (20), (33)—(35) and (40).

The theorem is proved. ∎

Remarks. (i) It is not necessary to apply the strong unicity result of W. O. Amrein, A. M. Berthier and V. Georgescu. Indeed, being $f'(y)\in C(\bar{\Omega})$ and $q\in W^{2,\alpha'}(\Omega)\cap W_0^{1,\alpha'}(\Omega)$ (by the proposition) a classical unicity result is also sufficient for our purposes.

(ii) The special case $\mu=2$ of the proposition is much easier: instead of the results of S. Agmon, A. Douglis and L. Nirenberg it is then sufficient to use the regularity results only in the Hilbert space case. It is much more simple, see e.g. [3], [4]. However, this restricts our investigation to the case $\gamma=2$, i.e. (in view of condition (5)) to the case $n\leqq 3$.

(iii) The condition (6) (due to J. L. Lions) may be replaced by other conditions, for example by $U_{\mathrm{ad}}=\{v\in L^\beta(\Omega)\colon v\geqq 0$ almost everywhere in $\Omega\}$ (this is due to F. Murat) or by $0\in U_{\mathrm{ad}}$ and $\|z_{\mathrm{d}}\|_{L^\alpha(\Omega)}$ is sufficiently small (this type of conditions is due to P. Rivera); see [6], pp. 327—328. It is then natural to ask whether Theorem 2 remains valid without any further assumption on U_{ad}. In view of some recent results of M. Ramaswamy [7] this does not seem to be true.

REFERENCES

[1] AMREIN, W. O., BERTHIER, A. M. and GEORGESCU, V., L^p-inequalities for the Laplacian and unique continuation, *Ann. Inst. Fourier* **31** (1981), 153—168. *MR* **83g:** 35011.

[2] AGMON, S., DOUGLIS, A. and NIRENBERG, L., Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions, I, *Comm. Pure Appl. Math.* **12** (1959), 623—727, **17** (1964), 35—92. *MR* **23** #A2610, **28** #5252.

[3] BRÉZIS, H., *Analyse fonctionnelle. Théorie et applications*, Masson, Paris, 1983. *MR* **85a:** 46001.

[4] LIONS, J. L. and MAGENES, E., *Problèmes aux limites non homogènes et applications* I., Dunod, Paris, 1968. *MR* **40** #512.

[5] LIONS, J. L., *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod—Gauthier-Villars, Paris, 1969. *MR* **41** #4326.

[6] LIONS, J. L., Contrôle des systèmes distribués singuliers, Dunod, Paris, 1983.

[7] RAMASWAMY, M., Quelques problèmes non-linéaires: homogénéisation et comportement global des solutions d'une équation différentielle non-linéaire, Thèse présentée à l'Université Pierre et Marie Curie (Paris VI) pour obtenir le diplôme de docteur de 3ème cycle, 1983.

[8] LIONS, J. L., *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod, Paris, 1968. *MR* **39** #5920.

[9] KOMORNIK, V., On the control of strongly nonlinear problems II (to appear).

[10] KOMORNIK, V. and TIBA, D., On the control of strongly nonlinear hyperbolic systems (to appear).

[11] KOMORNIK, V. and TIBA, D., Contrôle de systèmes fortement non-linéaires, *C. R. Acad. Sci. Paris Sér. I Math.* **300** (1985), 393-396.

(Received March 22, 1984)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 341—345

OPTIMAL ALLOCATION OF TREATMENTS IN BLOCK DESIGNS

C. STĘPNIAK

Abstract

Possible allocations of treatments in the block designs with fixed or random block effects are compared by estimation of the general mean and treatment effects. It is shown that the orthogonal allocation is optimal.

1. General background

Consider models where the expectation of an observable random vector X depends linearly on several parameters, but only some of the parameters are of interest. The assumption can be written in

$$\mathrm{E}X = A\alpha + B\beta \tag{1.1}$$

and

$$\operatorname{Cov} X = V_\gamma, \tag{1.2}$$

where A and B are known matrices, α and β are unknown vectors of parameters, while V_γ is a symmetric non negative definite matrix depending on a parameter $\gamma \in \Gamma$. It is assumed that the parameter α is of interest while β and γ are nuisance parameters.

We shall say that a random vector X is subject to the linear model $L(A\alpha + B\beta, V_\gamma | \gamma \in \Gamma)$ if the conditions (1.1) and (1.2) hold. The usual block design with fixed or random block effects defines such a model. If $\Gamma = \{\gamma_0\}$, i.e. the case of known covariance, we may write $L(A\alpha + B\beta, V)$ instead of $L(A\alpha + B\beta, V_\gamma | \gamma = \gamma_0)$. Similarly, in the case when the all parameters of the expectation are of interest, we may write $L(A\alpha, V_\gamma | \gamma \in \Gamma)$ instead of $L(A\alpha + 0\beta, V_\gamma | \gamma \in \Gamma)$.

Suppose X and Y are subject, respectively, to the linear models

$$L(A_1\alpha + B_1\beta,\ V_\gamma | \gamma \in \Gamma) \quad \text{and} \quad L(A_2\alpha + B_2\beta,\ W_\gamma | \gamma \in \Gamma).$$

We shall write $L(A_1\alpha + B_1\beta, V_\gamma | \gamma \in \Gamma) \succ L(A_2\alpha + B_2\beta, W_\gamma | \gamma \in \Gamma)$ if for any function ψ of α and for any unbiased estimator $b'Y$ of ψ, provided a such exists, there is an unbiased estimator $a'X$ of ψ such that $\operatorname{Var}(a'X) \leqq \operatorname{Var}(b'Y)$ for all possible values of α, β and γ.

1980 *Mathematics Subject Classification.* Primary 62K10; Secondary 62K05.

Key words and phrases. Block design, fixed block effect, random block effect, allocation of treatments, comparison of allocations, optimal allocation.

The ordering $\succ$ will be used to the comparison of the allocations of treatments in the block designs with fixed or random treatment effects. For some results in this case see Kiefer [3], Stępniak [5] and Stępniak, Wang and Wu [8].

Throughout this paper the usual matrix notation will be used. Among others if A is a matrix then A' and A^- will denote, respectively, the transposition and a generalized inverse of A. For any two symmetric matrices A and B the symbol $A \geqq B$ means that $A-B$ is non negative definite (n.n.d.). Moreover, the symbol R^n stands for the n-dimensional Euclidean space represented by the column vectors.

2. The results

Consider an allocation of v treatments with replications $t_1, \ldots, t_v$ in k blocks of sizes $b_1, \ldots, b_k$, where $\sum t_i = \sum b_j = n$. It will be convenient to order the n observations in the way "block by block" and to identify the allocation with a matrix $D=(d_{ij})$ of $n \times v$ defined by

$$d_{ij} = \begin{cases} 1 & \text{if the } i\text{-th observation refers to the } j\text{-th treatment,} \\ 0 & \text{otherwise.} \end{cases}$$

The allocation D may be considered in the block design with fixed block effects, corresponding with the model

$$L([1_n \vdots D]\alpha + B\beta, \gamma I_n | \gamma > 0) \tag{2.1}$$

or in the block design with random block effects, corresponding with the model

$$L([1_n \vdots D]\alpha, V(\gamma_0, \gamma_1) | \gamma_0 > 0, \ \gamma_1 \geqq 0), \tag{2.2}$$

where 1_n is the column of n ones,

$$B = \operatorname{diag}(1_{b_1}, \ldots, 1_{b_k})$$

and

$$V(\gamma_0, \gamma_1) = \gamma_0 I_n + \gamma_1 \operatorname{diag}(1_{b_1} 1'_{b_1}, \ldots, 1_{b_k} 1'_{b_k}).$$

The parameters $\alpha = (\mu, \alpha_1, \ldots, \alpha_v)'$ and $\beta = (\beta_1, \ldots, \beta_k)'$ in the model (2.1) refer to the general mean, the treatment effects and the block effects, while in the model (2.2) the block effects are represented by the variance component γ_1.

We shall say that an allocation D_0 is at least as good as an allocation D in the block design with fixed block effects if

$$L([1_n : D_0]\alpha + B_0\beta, \gamma I_n | \gamma > 0) \succ L([1_n : D]\alpha + B\beta, \gamma I_n | \gamma > 0). \tag{2.3}$$

Similarly, D_0 is at least as good as D in the block design with random block effects if

$$L([1_n : D_0]\alpha, V_0(\gamma_0, \gamma_1) | \gamma_0 > 0, \gamma_1 \geqq 0) \succ L([1_n : D]\alpha, V(\gamma_0, \gamma_1) | \gamma_0 > 0, \gamma_1 \geqq 0). \tag{2.4}$$

Let $\mathscr{D}$ be a class of allocations. An allocation $D_0 \in \mathscr{D}$ is said to be optimal within $\mathscr{D}$ in the block design with fixed/random block effects if the condition (2.3)/(2.4) holds for all $D \in \mathscr{D}$.

Some another criterions of optimality as A-, D- and E-optimality were considered by several authors, among others by Kiefer [2]—[4]. It is known that the optimality defined above implied A-, D- and E-optimality.

To each allocation D corresponds the incidence matrix $N_D=D'B$. An allocation D is said to be orthogonal if $N_D=\frac{1}{n}tb'$, where $t=(t_1, \ldots, t_v)'$ and $b=(b_1, \ldots, b_k)'$.

Denote by $\mathcal{D}(t_1, \ldots, t_v)$ the class of possible allocations of treatments with replications $t_1, \ldots, t_v$ over all possible blocks with total capacity $\sum t_i$, i.e. for all $k \leqq \sum t_i$ and for all $b_1, \ldots, b_k$ such that $\sum b_i = \sum t_i$. Let D_0 and D be members of $\mathcal{D}(t_1, \ldots, t_v)$. A necessary and sufficient condition for (2.3), given by Kiefer ([3], p. 288), is

$$A_0'A_0 - A_0'B_0(B_0'B_0)^- B_0'A_0 - A'A + A'B(B'B)^- B'A \geqq 0,$$

where $A=[1_n \vdots D]$. Thus we get immediately

LEMMA 1. *For any $D_0, D \in \mathcal{D}(t_1, \ldots, t_v)$ the allocation D_0 is at least as good as the allocation D in the block design with fixed block effects if and only if*

$$N_D \operatorname{diag}\left(\frac{1}{b_1}, \ldots, \frac{1}{b_k}\right) N_D' - N_{D_0} \operatorname{diag}\left(\frac{1}{b_1^0}, \ldots, \frac{1}{b_k^0}\right) N_{D_0}'$$

is n.n.d.

Now consider the allocations D_0 and D in the block design with random effects.

LEMMA 2. *For any $D_0, D \in \mathcal{D}(t_1, \ldots, t_v)$ the allocation D_0 is at least as good as the allocation D in the block design with random block effects if and only if*

$$H_D(\varrho) \geqq H_{D_0}(\varrho) \quad \text{for all} \quad \varrho \geqq 0,$$

where $H_D(\varrho) = N_D \operatorname{diag}\left(\frac{1}{1+b_1\varrho}, \ldots, \frac{1}{1+b_k\varrho}\right) N_D'$.

PROOF. It follows from a result of Stępniak and Torgersen [7] (see also Stępniak [5], [6]) that a necessary and sufficient condition for (2.2) is

$$\begin{bmatrix} 1_n' \\ D_0' \end{bmatrix} V^{-1}(\gamma_0, \gamma_1)[1_n \vdots D_0] - \begin{bmatrix} 1_n' \\ D' \end{bmatrix} V^{-1}(\gamma_0, \gamma_1)[1_n \vdots D] \geqq 0 \tag{2.5}$$

for all $\gamma_0 > 0$ and $\gamma_1 \geqq 0$. As

$$V^{-1}(\gamma_0, \gamma_1) = \frac{1}{\gamma_0}\left[I_n - \varrho \operatorname{diag}\left(\frac{1}{1+b_1\varrho} 1_{b_1}1_{b_1}', \ldots, \frac{1}{1+b_k\varrho} 1_{b_k}1_{b_k}'\right)\right],$$

where $\varrho = \gamma_1/\gamma_0$, the condition (2.5) is reduced to

$$\left[\begin{array}{c|c} 1_v'H_D1_v & 1_v'H_D \\ \hline H_D1_v & H_D \end{array}\right] - \left[\begin{array}{c|c} 1_v'H_{D_0}1_v & 1_v'H_{D_0} \\ \hline H_{D_0}1_v & H_{D_0} \end{array}\right] \geqq 0. \tag{2.6}$$

Thus we only need to show that, for any two symmetric matrices A_1 and A_2 of dimension $v\times v$, the condition $A_1-A_2\geqq 0$ is equivalent to $B_1-B_2\geqq 0$, where

$$B_i = \left[\begin{array}{c:c} 1_v' A_i 1_k & 1_v' A_i \\ \hdashline A_i 1_v & A_i \end{array}\right], \quad i = 1, 2.$$

If $B_1-B_2\geqq 0$ then $A_1-A_2\geqq 0$ evidently. Conversely, if $A_1-A_2\geqq 0$ then, for any vector x of $v\times 1$ and for any scalar x_0 we get

$$[x_0, x'](B_1-B_2)\left[\begin{array}{c} x_0 \\ \hdashline x \end{array}\right] = (x_0 1_v' + x')(A_1-A_2)(x_0 1_v + x) \geqq 0.$$

This completes the proof. ▮

THEOREM 1. *Any orthogonal allocation D_0 of v treatments with replications $t_1, \ldots, t_v$ is optimal within the class $\mathscr{D}(t_1, \ldots, t_v)$ in the block design with fixed block effects.*

PROOF. It follows from the definition of the orthogonal allocation that

$$N_{D_0} \operatorname{diag}\left(\frac{1}{b_1^0}, \ldots, \frac{1}{b_k^0}\right) N_{D_0}' = \frac{1}{n} tt',$$

where $t=(t_1, \ldots, t_v)'$. On the other hand, by definition of $\mathscr{D}$ we get $t=N_D 1_k$ to each $D\in\mathscr{D}$. Thus, by Lemma 1, we only need to show the condition

$$n \operatorname{diag}\left(\frac{1}{b_1}, \ldots, \frac{1}{b_k}\right) - 1_k 1_k' \geqq 0 \tag{2.7}$$

for all $D\in\mathscr{D}$. The condition may be easily obtained from the following result of Farebrother [1]: For any matrix A of $k\times k$, for any vector x of $k\times 1$ and for any positive scalar c the matrix $cA-xx'$ is non negative definite if and only if $x'A^{-1}x\leqq c$. This completes the proof. ▮

Now consider the problem of optimal allocation of treatments in the block design with random block effects. Denote by $\mathscr{D}(t_1, \ldots, t_v; b_1, \ldots, b_k)$ the class of possible allocations of v treatments with replications $t_1, \ldots, t_v$ in k blocks of sizes $b_1, \ldots, b_k$. We shall restrict our attention to the case $b_1=\ldots=b_k=\frac{n}{k}$.

Let t_i/k be integer for $i=1, \ldots, v$. Then there exists an orthogonal allocation D_0 with the incidence matrix $N_{D_0}=(n_{ij})$, where $n_{ij}=t_i/k$.

THEOREM 2. *The orthogonal allocation D_0 is optimal within the class*

$$\mathscr{D}\left(t_1, \ldots, t_v; \frac{n}{k}, \ldots, \frac{n}{k}\right)$$

in the block design with random block effects.

PROOF. The Theorem follows immediately from Lemma 2 and from the result of Farebrother, mentioned above. ▮

REMARK. The assumption $b_1 = \ldots = b_k$ is substantial. To see it consider two allocations D_0 and D with the incidence matrices

$$N_{D_0} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \quad \text{and} \quad N_D = \begin{bmatrix} 2 & 2 \\ 1 & 7 \end{bmatrix},$$

respectively. Then

$$H_{D_0}(0) = 10 \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad H_D(0) = 2 \begin{bmatrix} 4 & 8 \\ 1 & 7 \end{bmatrix}$$

and the condition $H_D(\varrho) \geqq H_{D_0}(\varrho)$ is not satisfied. Thus, by Lemma 2, the allocation D_0 is not optimal because it is not at least as good as D.

REFERENCES

[1] FAREBROTHER, R. W., Further results on the mean squared error of ridge regression. *J. Roy. Statist. Soc. B* **38** (1976), 248—250. *MR* **58** #31610.

[2] KIEFER, J., On the nonrandomized optimality and randomized nonoptimality of symmetrical designs, *Ann. Math. Statist.* **29** (1958), 675—699. *MR* **20** #4910.

[3] KIEFER, J., Optimum experimental designs, *J. Roy. Statist. Soc. Ser. B* **21** (1959), 272—319. *MR* **22** #4101.

[4] KIEFER, J., General equivalence theory for optimum designs (approximate theory), *Ann. Statist.* **2** (1974), 849—879. *MR* **50** #8856.

[5] STĘPNIAK, C., Uniform theory of comparison of linear models, *Statistics and Probability* (Visegrád, 1982), 323—334, Reidel, Dordrecht—Boston, 1984. *MR* **85g:** 60005.

[6] STĘPNIAK, C., Optimal allocation of units in experimental designs with hierarchical and cross classification, *Ann. Inst. Statist. Math.* **35** (1983), 461—473. *MR* **85m:** 62166.

[7] STĘPNIAK, C. and TORGERSEN, E., Comparison of linear models with partially known covariances with respect to unbiased estimation, *Scand. J. Statist.* **8** (1981), 183—184. *MR* **82m:** 62160.

[8] STĘPNIAK, C., WANG, S. G. and WU, C. F. J., Comparison of linear experiments with known covariances, *Ann. Statist.* **12** (1984), 358—365. *MR* **85b:** 62064.

(Received April 11, 1984)

DEPARTMENT OF APPLIED MATHEMATICS
AGRICULTURAL UNIVERSITY OF LUBLIN
UL. AKADEMICKA 13
PL—20-934 LUBLIN
POLAND

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 347—350

EXPECTATION IN METRIC SPACES

R. PICK

We have been working for a while on the following problem: Let $t_1, t_2, \ldots$ be i.i.d.r.v.-s of exponential distribution with parameter λ, and set $T_i = t_1 + t_2 + \ldots + t_i$. Suppose that at time T_k a point P_k is settled on the unit sphere S, the points $P_1, P_2, \ldots$ are i.i.d.r.v.-s of uniform distribution on S, furthermore the t_i-s and P_i-s are independent. Each of the points P_i generates a process $v_{i,j}$; $Q_{i,j}$ similar to the process t_k, P_k but these processes have only finite number of elements V_i and the generated points $Q_{i,j}$ on S follow the cosine law with center P_i. We do not know the parameters of these processes, and we are able to observe only the points $Q_{i,j}$ at the time $T_{i,j} = T_i + v_{i,1} + v_{i,2} + \ldots + v_{i,j}$ of their appearence. The problem is to estimate the process $\{(P_k, T_k, U_k);\ k=1, 2, \ldots\}$ where U_k is the time of extinction of the process generated by P_k. To the solution of this problem we would like to apply the Kalman-filter technic, which works with conditional expectations.

What makes the problem rather sophisticated is the fact that the target of filtration, i.e. the system constituted from the simultaneously existing points P_i is not an Euclidean space. That is why we extend here the concept of mathematical expectation to the case of metric spaces. Such extensions were considered previously by Hans [3], Sverdrup—Thygeson [5] and Grenander [2]. Hans has introduced the concept of generalized random variable and extended the concept of mathematical expectation to Banach spaces. Sverdrup—Thygeson gave a definition similar to ours and proved an a.s. theorem on the empirical expectation. Grenander continued the investigation of Banach space valued random variables. We shall prove here continuity of expectation with respect to the Prohorov metric under a suitable moment constraint.

We would like to emphasize that our result does not give a direct solution to the problem described above, it only represents a step that appears necessary towards such a solution.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, (X, d) a metric space, and let σ-algebra $\mathcal{B}$ in X be generated by open subsets of X. Let y be a generalized random variable from $(\Omega, \mathcal{A}, \mu)$ to $(X, \mathcal{B})$ with probability distribution P on $(X, \mathcal{B})$.

Assumption 1. (Property of the space X.) Every closed and bounded subset of X is compact.

1980 *Mathematics Subject Classification.* Primary 60B05; Secondary 62F10.

Key words and phrases. Generalized expectation, Prohorov metric, Kalman filter, Poisson processes.

Assumption 2. (Property of the variable y.) There exists $x_0 \in X$ such that $\int\limits_X d^2(y, x_0)\, dP < \infty$.

DEFINITION. Let $V(x) = \int\limits_X d^2(y, x) dP(y)$, i.e. $V(x)$ is the expectation of the r.v. $d^2(y, x)$ for fixed $x \in X$. Let us define in the metric space X the expectation of y by

$$\mathsf{E}y = \{x \in X : V(x) = \inf_{t \in X} V(t)\}.$$

Thus the expectation is a subset of X. If X is an Euclidean space then $\mathsf{E}y$ is the expectation in the usual sense.

EXAMPLE. Let X be a two-dimensional Euclidean space with distance $d(x, y) = \max\limits_{i=1,2} \{|x_i - y_i|\}$ where x_i, y_i are coordinates of x and y, respectively. Let P be uniform distribution on the Cartesian product of the sets $\{(-3 < x_1 < -2) \cup (2 < x_1 < 3)\}$ and $\{-1 < x_2 < 1\}$. Then $\mathsf{E}y$ is the Cartesian product of the sets $\{x_1 = 0\}$ and $\{-1 < x_2 < 1\}$. This example shows that $\mathsf{E}y$ need not be a single point.

LEMMA 1. *Let Assumption 2 be satisfied, then the function $V(x) = Ed^2(y, x)$ is continuous.*

PROOF.

$$|V(x) - V(x_0)| \leqq \int\limits_X |d^2(y, x) - d^2(y, x_0)|\, dP \leqq$$

$$\leqq \int\limits_X ((2d(y, x_0) + d(x_0 x)))\, d(x_0, x) dP = 2d(x_0, x) \int\limits_X d(y, x_0) dP + d^2(x_0, x).$$

Using Assumption 2 we have that $V(x) \to V(x_0)$ if $x \to x_0$. From the above argument it can be seen that $\mathsf{E}d^2(y, x) < \infty$ for arbitrary $x \in X$. Thus Assumption 2 implies that $V(x) < \infty$ for arbitrary $x \in X$ and we can substitute x_0 in the above argument with an arbitrary x.

LEMMA 2. *Let Assumption 2 be satisfied, and $D_0^2 = \mathsf{E}d^2(y, x_0)$. Then*

(1) $\quad Ed^2(y, x) \geqq (d(x, x_0) - D_0)^2$ *for any* $x \in X$ *such that* $d(x, x_0) \geqq D_0$.

PROOF. The triangular inequality implies that

$$d(x, x_0) \leqq \mathsf{E}d(y, x) + \mathsf{E}d(y, x_0) \leqq \sqrt{\mathsf{E}d^2(y, x)} + \sqrt{\mathsf{E}d^2(y, x_0)}$$

which in turn implies (1).

REMARK 1. If $D_0^2 = \mathsf{E}d^2(y, x_0) < \infty$ for some $x_0 \in X$, then by Lemma 2, $d(x, x_0) > kD_0$ implies $\mathsf{E}d^2(y, x) > [(k-1)D_0]^2$. In applying this for $k = 2$ we get that $\mathsf{E}y$ is contained in the sphere of center x_0 and radius $2D_0$.

COROLLARY OF LEMMAS 1 AND 2. *Assumption 2 implies that $V(x) = \mathsf{E}d^2(y, x)$ is bounded and uniformly continuous in any compact subset of X. If Assumption 1 holds, too, then $V(x)$ has a minimum value, thus $\mathsf{E}y$ is not void.*

Let y and z random variables in (X, d) with distributions P and Q, respectively. Let us denote the Prohorov distance of P and Q by $L(\mathsf{P}, \mathsf{Q})$, and the ε-neighbourhood of a set H by H_ε. Let us remember that the Prohorov distance is defined by

$$L(\mathsf{P}, \mathsf{Q}) = \inf \{\delta\colon\ \mathsf{P}(F) < \mathsf{Q}(F_\delta)+\delta \text{ for all closed } F\subset X\}.$$

Further on we need the corollary of Theorem 11 of Strassen [2]:

Let (X, d) be complete separable metric space. Then

$$L(\mathsf{P}, \mathsf{Q}) = \inf \{\delta\colon \text{there is a probability measure } \mathsf{R} \text{ in } X\times X \text{ with marginals } \mathsf{P} \text{ and } \mathsf{Q} \text{ such that } \mathsf{R}\big(d(x, x') > \delta\big) < \delta\big\}.$$

LEMMA 3. *Let Assumption 1 be satisfied. Then (X, d) is complete separable metric space.*

PROOF. According to the Assumption 1 every closed ball is compact. Thus the unit ball is complete and separable. This implies, that the space X is complete and separable because necessarily every Cauchy sequence is part of an enough big ball.

LEMMA 4. *Let (X, d) be a metric space satisfying Assumption 1. Let y and z be random variables in (X, d) with distribution P and Q, respectively.*
If for a δ such that $0<\delta<1$ we have

(i) $L(\mathsf{P}, \mathsf{Q}) < \delta$

and there is a $K<\infty$ such that for some $x_0\in X$

(ii) $\mathsf{E}d^4(y, x_0) < K$

(iii) $\mathsf{E}d^4(z, x_0) < K$

then

$$|\mathsf{E}d^2(z, x)-\mathsf{E}d^2(y, x)| < A\delta,$$

if

$$d(x, x_0) < \sqrt[4]{K};\quad A = 8\sqrt{K}+4\sqrt[4]{K}.$$

PROOF. According to Lemma 2 we can use Corollary of Theorem 11 of Strassen's. Hence if $L(\mathsf{P}, \mathsf{Q})<\delta$ a joint distribution R of the random variable (y, z) can be given in the space $X\times X$ such that $\mathsf{R}\big(d(y, z)>\delta\big)<\delta$.

Then

$$|\mathsf{E}d^2(z, x)-\mathsf{E}d^2(y, x)| \leqq$$

$$\leqq \int\limits_{d(y,z)<\delta} |d^2(z, x)-d^2(y, x)|\,d\mathsf{R}+ \int\limits_{d(y,z)>\delta} d^2(y, x)\,d\mathsf{R}+ \int\limits_{d(y,z)\geqq\delta} d^2(z,x)\,d\mathsf{R} = J_1+J_2+J_3.$$

Here for the first term we have

$$J_1 < \delta\big(\mathsf{E}d(z, x)+\mathsf{E}d(y, x)\big).$$

We can estimate the next term in the following way

$$J_2^2 = \big[\mathsf{E}d^2(y, x)\,I\big(d(y, z) > \delta\big)\big]^2 \leqq \mathsf{E}d^4(y, x)\,\mathsf{R}\big(d(y, z) > \delta\big) \leqq 16K\delta$$

where I stands for the indicator variable.

Here the triangular inequality implies that

$$\mathsf{E}d^4(x, y) < 16K \quad \text{if} \quad d(x, x_0) < \sqrt[4]{K}$$

and the estimation of J_3 is similar. Summing up our estimations we get

$$|\mathsf{E}d^2(z, x) - \mathsf{E}d^2(y, x)| \leqq 4\delta \sqrt[4]{K} + 8\sqrt{\delta}\sqrt{K} < \delta(4\sqrt[4]{K} + 8\sqrt{K}).$$

THEOREM. *Let Assumption 1 be satisfied and y be an X-valued r.v. with distribution* P *such that* $\mathsf{E}d^4(y, x_0) < K$ *for some* $x_0 \in X$. *Then to any* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that for another r.v. z with distribution* Q *satisfying* $\mathsf{E}d^4(z, x_0) < K$, *the condition* $L(\mathsf{P}, \mathsf{Q}) < \delta$ *implies* $\mathsf{E}z \subset (\mathsf{E}y)_\varepsilon$.

PROOF. Let $\varepsilon > 0$ be fixed. Assumption 1 and Remark 1 imply that for the quantities $D_\varepsilon^2 = \inf_{x \notin (\mathsf{E}y)_\varepsilon} \mathsf{E}d^2(y, x)$ and

$$D^2 = \inf_{x \in X} \mathsf{E}d^2(y, x)$$

one has $D_\varepsilon^2 > D^2$.

Let A be the same constant as in Lemma 4 and let δ be defined by $\delta = (D_\varepsilon^2 - D^2)/2A$. Then for this δ and for any Q meeting the conditions of the theorem, Lemma 4 implies that for any $x \in \mathsf{E}y \subset \{x\colon d(x, x_0) < 4\sqrt[4]{K}\}$ we have $\mathsf{E}d^2(z, x) \leqq D^2 + \delta A < D_\varepsilon^2$ which results in $\mathsf{E}z \subset (\mathsf{E}y)_\varepsilon$.

ACKNOWLEDGEMENT. The author wish to thank Professors I. Csiszár and G. Tusnády for a number of very helpful and valuable discussions and suggestions.

REFERENCES

[1] MUKHERJEA, A. and POTHOVEN, K. L., *Real and functional analysis*, Plenum Press, New York—London, 1978. *MR* **58** #11294.

[2] GRENANDER, U., *Abstract inference*, Wiley, New York, 1981. *MR* **82g**: 62004.

[3] HANS, O., Generalized random variables, *Transactions of the first Prague conference on information theory, statistical decision functions, random processes*, Prague, 1957, 61—103. *MR* **20** #7331.

[4] STRASSEN, V., The existence of probability measures with given marginals, *Ann. Math. Statist.* **36** (1965), 423—429. *MR* **31** #1693.

[5] SVERDRUP-THYGESON, H., Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces, *Ann. Statist.* **9** (1981), 141—145. *MR* **82b**: 60031.

(Received May 4, 1984)

ARATÓ U. 11
H—1121 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 351—356

ON THE DECOMPOSITION OF A LATTICE-REGULAR MEASURE WITH RESPECT TO A MAPPING

P. GRASSI

A generalization of a measure decomposition theorem is presented. Measures defined on locally compact topological spaces are now considered in the more general lattice setting and a decomposition is obtained via a lattice continuous mapping from one lattice space into another. The noted topological application, as well as others, are immediate consequences.

1. Introduction

Our principal aim in this paper is to generalize an important theorem of Herz [7] to an abstract lattice-regular measure situation. The theorem of Herz has been utilized by Bauer [3] to investigate conditions under which a continuous map between locally compact spaces will be conservative. Here, we consider a map $T: X \to Y$ where X and Y are arbitrary sets, $\mathbf{L}_1$ and $\mathbf{L}_2$ are lattices of subsets of X and Y, respectively, and T is $\mathbf{L}_1-\mathbf{L}_2$ continuous (see below).

If $\nu \in M_R^\sigma(\mathbf{L}_2)$, the σ-smooth $\mathbf{L}_2$-regular measures on $\mathbf{A}(\mathbf{L}_2)$, the algebra generated by $\mathbf{L}_2$, we seek conditions for ν so that ν can be represented as the sum of an induced measure μT^{-1}, $\mu \in M_R^\sigma(\mathbf{L}_1)$, and a suitable $\nu' \in M_R^\sigma(\mathbf{L}_2)$. Specific topological choices for X and Y and the lattices $\mathbf{L}_1$ and $\mathbf{L}_2$ yield Herz's result. However, it will be clear that many other choices are possible. In particular, we show how these ideas yield quite readily a new proof of an important result of Hardy and Lacey [6].

We begin with a brief review of the lattice and measure theoretic terminology used throughout, and then proceed to a consideration of extending and restricting lattice-regular measures. The main treatment of the decomposition theorem then follows with several applications.

2. Notation and terminology

$\mathbf{L}$ will always denote a *lattice* of subsets of some abstract set X. Without loss of generality, we will assume that X and $\emptyset$ belong to $\mathbf{L}$. $\mathbf{L}$ is a *δ-lattice* if $\mathbf{L}$ is closed under countable intersections. $\mathbf{A}(\mathbf{L})$ and $\sigma(\mathbf{L})$ denote the algebra and σ-algebra respectively generated by $\mathbf{L}$, and $\delta(\mathbf{L})$ will represent the lattice obtained by taking

1980 *Mathematics Subject Classification.* Primary 28A60; Secondary 28A32.

Key words and phrases. Lattice, lattice regular measure, conservative map, ν-thick, $\mathbf{L}_1-\mathbf{L}_2$ continuous map, countably paracompact.

the collection of all countable intersections of elements in $\mathbf{L}$. $\mathbf{L}$ is said to be *countably compact* provided that whenever X is covered by countably many sets, of the form L_i', $L_i \in \mathbf{L}$, there exists a finite subcovering. If $\mathbf{L}_1 \subset \mathbf{L}_2$ where both $\mathbf{L}_1$ and $\mathbf{L}_2$ are lattices on the same set X, $\mathbf{L}_2$ *is* $\mathbf{L}_1$ *countably paracompact* provided that whenever $L_i \downarrow \emptyset$, $L_i \in \mathbf{L}_2$, there exist $\tilde{L}_i \in \mathbf{L}_1$ such that $L_i \subset \tilde{L}_i'$ for each i and $\tilde{L}_i' \downarrow \emptyset$.

$M_R(\mathbf{L})$ will denote the collection of all finitely additive, positive and finite valued, $\mathbf{L}$-regular measures on $\mathbf{A}(\mathbf{L})$, and $M_R^\sigma(\mathbf{L})$ will represent those members of $M_R(\mathbf{L})$ which are σ-smooth on $\mathbf{A}(\mathbf{L})$. We will often assume that a measure $\mu \in M_R^\sigma(\mathbf{L})$ has been extended to $\sigma(\mathbf{L})$ without any change of notation. It is also noted that if $\mu \in M_R(\mathbf{L})$ then μ is $\delta(\mathbf{L})$-regular on $\sigma(\mathbf{L})$. The outer and inner measures of $\mu \in M_R(\mathbf{L})$ will be written μ^* and μ_*, respectively.

Further notation consistent with the above may be found in Bachman and Sultan [2] and Grassi [4].

3. Regular measures on sublattices

We begin by considering conditions for which σ-smooth, lattice-regular measures on X may be restricted to a lattice determined by suitable subspaces of the lattice space while preserving the regularity. We will also attempt to extend the regularity from subspaces of X to the whole space.

LEMMA 3.1. *Suppose* $\mu \in M_R^\sigma(\mathbf{L})$ *and* $B \in \sigma(\mathbf{L})$. *Define* $\mu_0(B \cap E) = \mu(B \cap E)$ *where* $E \in \sigma(\mathbf{L})$. *Then* $\mu_0 \in M_R^\sigma(B \cap \mathbf{L})$.

PROOF. Clear. ∎

In order for the converse of Lemma 3.1 to be true we must take $B \in \mathbf{L}$ as we see in the following lemma.

LEMMA 3.2. *Suppose* $\mu_0 \in M_R^\sigma(B \cap \mathbf{L})$ *and* $B \in \mathbf{L}$. *Define* $\mu(E) = \mu_0(B \cap E)$ *where* $E \in \sigma(\mathbf{L})$. *Then* $\mu \in M_R^\sigma(\mathbf{L})$.

PROOF. If $A \in \mathbf{A}(\mathbf{L})$, there exists $L \in \mathbf{L}$ such that $\mu_0(B \cap A) - \mu_0(B \cap L) < \varepsilon$, $B \cap L \subset B \cap A$. But $\mu(A) - \mu(B \cap L) = \mu_0(B \cap A) - \mu_0(B \cap L) < \varepsilon$ and $B \cap L \subset A$. Since $B \cap L \in \mathbf{L}$ it follows that $\mu \in M_R^\sigma(\mathbf{L})$. ∎

DEFINITION. Suppose $B \subset X$. B is μ*-thick* if $\mu^*(B) = \mu(X)$ where $\mu \in M_R^\sigma(\mathbf{L})$.

The following measure is thus well-defined. (Cf. Halmos [5].)

LEMMA 3.3. *Suppose* $\mu \in M_R^\sigma(\mathbf{L})$ *and assume* B *is* μ*-thick. Define* $\mu_0(B \cap E) = \mu(E)$ *where* $E \in \sigma(\mathbf{L})$. *Then* $\mu_0 \in M_R^\sigma(B \cap \mathbf{L})$.

PROOF. Clear. ∎

We note that if B is μ-thick then $\mu_*(B') = 0$. Therefore $\mu^*(B \cap A) = \mu(A) = \mu_0(B \cap A)$, $A \in \sigma(\mathbf{L})$. We state this result formally.

LEMMA 3.4. *Suppose* $\mu \in M_R^\sigma(\mathbf{L})$ *and assume* B *is* μ*-thick. If* μ_0 *is as in Lemma 3.3, then* $\mu_0 = \mu^*$ *on* $\sigma(B \cap \mathbf{L})$.

We now show that for any subset W of X and for a given measure $\mu \in M_R^\sigma(\mathbf{L})$, we can construct a measure ν on a subspace of X on which W is ν-thick.

LEMMA 3.5. *Suppose* $\mu\in M_R^\sigma(\mathbf{L})$ *and* $\mathbf{L}$ *is a* δ*-lattice. Let* $W\subset X$. *Then there exists* $B\in\sigma(\mathbf{L})$ *and a measure* $\nu\in M_R^\sigma(B\cap\mathbf{L})$ *such that* W *is* ν*-thick.*

PROOF. Since μ is $\mathbf{L}$-regular on $\sigma(\mathbf{L})$,

$$\mu^*(W)=\inf_{W\subset L'}\mu(L'),\quad L\in\mathbf{L}.$$

Therefore,

$$\mu^*(W)=\lim_{n\to+\infty}\mu(L_n')=\mu\Big(\bigcap_{n=1}^{\infty}L_n'\Big)$$

where $W\subset L_n'$, $L_n'\downarrow$ and $L_n\in\mathbf{L}$. Let $B=\bigcap_{n=1}^{\infty}L_n'$. Define $\nu(B\cap E)=\mu(B\cap E)$, $E\in\sigma(\mathbf{L})$. By Lemma 3.1, $\nu\in M_R^\sigma(B\cap\mathbf{L})$. Also, since $W\subset B$, $\nu^*(W)=\mu^*(W)=$ $=\mu(B)=\nu(B)$. ∎

Assuming the conditions of Lemma 3.5 are satisfied, let us define a measure ν_0 on $\sigma(W\cap\mathbf{L})$ as follows:

$$\nu_0(W\cap E)=\nu(B\cap E),\qquad E\in\sigma(\mathbf{L}).$$

By Lemma 3.3, $\nu_0\in M_R^\sigma(W\cap\mathbf{L})$. Since $\mu^*=\nu^*$ on all subsets of B, $\mu^*=\nu^*$ on all subsets of W. But $\nu^*=\nu_0$ on $\sigma(W\cap\mathbf{L})$ by Lemma 3.4. Therefore, $\mu^*=\nu_0$ on $\sigma(W\cap\mathbf{L})$. Summarizing we have:

LEMMA 3.6. *Suppose* $\mu\in M_R^\sigma(\mathbf{L})$ *and assume* $\mathbf{L}$ *is a* δ*-lattice. If* $W\subset X$ *then* $\mu^*|_{W\cap\mathbf{L}}\in M_R^\sigma(W\cap\mathbf{L})$.

Our final result of this section investigates the relationship between a given outer measure on some lattice and the outer measure obtained by restricting the given measure to a sublattice.

LEMMA 3.7. *Suppose* $\mathbf{L}_1$ *and* $\mathbf{L}_2$ *are both lattices on* X *and assume* $\mathbf{L}_1\subset\mathbf{L}_2$. *Let* $\nu\in M_R^\sigma(\mathbf{L}_2)$ *and* $\nu|_{\mathbf{L}_1}=\mu$. *Then*

1. $\nu^*\leqq\mu^*$
2. $\nu^*=\mu^*$ *iff* $\mu^*=\nu$ *on* $\mathbf{L}_2'$.

PROOF. 1. Clear.

2. If $\nu^*=\mu^*$ then clearly $\mu^*=\nu$ on $\mathbf{L}_2'$. Suppose $\mu^*=\nu$ on $\mathbf{L}_2'$.

$$\nu^*(S)=\inf_{S\subset\bigcup_{i=1}^{\infty}A_i}\sum_{i=1}^{\infty}\nu(A_i)\qquad A_i\in\mathbf{A}(\mathbf{L}_2).$$

Since ν is $\mathbf{L}_2$-regular, for each $A_i\in\mathbf{A}(\mathbf{L}_2)$ there exists $L_i'\in\mathbf{L}_2'$ such that $\nu(A_i)\geqq$ $\geqq\nu(L_i')-\varepsilon/2^i$, $A_i\subset L_i'$. Thus,

$$\nu^*(S)+\varepsilon\geqq\sum_{i=1}^{\infty}\nu(A_i)\geqq\sum_{i=1}^{\infty}\nu(L_i')-\varepsilon$$

for some sequence of sets $\{A_i\}$,

$$S\subset\bigcup_{i=1}^{\infty}A_i\subset\bigcup_{i=1}^{\infty}L_i'.$$

Hence

$$\nu^*(S)+2\varepsilon \geqq \sum_{i=1}^{\infty} \nu(L_i') = \sum_{i=1}^{\infty} \mu^*(L_i') \geqq \mu^*(\bigcup_{i=1}^{\infty} L_i') \geqq \mu^*(\bigcup_{i=1}^{\infty} A_i) \geqq \mu^*(S).$$

Combining this result with that of part 1 we see that $\nu^*=\mu^*$. ∎

4. The decomposition theorem

In this section we prove a general decomposition theorem which states that if T: $X\to Y$ is $\mathbf{L}_1-\mathbf{L}_2$ continuous then under the stated conditions, a measure $\nu\in M_R^\sigma(L_2)$ can be expressed as the sum of the measures μT^{-1} and ν', where $\mu\in M_R^\sigma(\mathbf{L}_1)$ and μT^{-1}, $\nu'\in M_R^\sigma(\mathbf{L}_2)$. We will also show that ν' vanishes on a certain class of subsets of Y.

Definition. Let T: $X\to Y$ where $\mathbf{L}_1$ and $\mathbf{L}_2$ are lattices on X and Y, respectively. T is $\mathbf{L}_1-\mathbf{L}_2$ *continuous* if $T^{-1}(L_2)\in\mathbf{L}_1$ for each $L_2\in\mathbf{L}_2$.

We begin with some preliminary results.

Lemma 4.1. *Suppose* T: $X\to Y$. *Let* $\nu\in M_R^\sigma(L_2)$ *and let* $\{S_\alpha\}$ *be a class of subsets of* $\mathbf{L}_1$ *such that* $TS_\alpha\in\mathbf{L}_2$ *for all* α. *If* $\mathbf{S}$ *equals the* σ*-ring generated by* $\{TS_\alpha\}$ *and if* $a=\sup_{B\in\mathbf{S}}\nu(B)$, *then:*

1. $\mathbf{S}\subset\sigma(\mathbf{L}_2)$.
2. *There exist* $B_1,\dots,B_n,\dots$ *such that* $\lim_{n\to+\infty}\nu(B_n)=a$ *and* $\nu(B_0)=a$ *where* $B_0=\bigcup_{i=1}^{\infty} B_i$ *and each* $B_i\in\mathbf{S}$.
3. *If* $\mathbf{L}_2$ *is a* δ*-lattice define* $\nu_0(E)=\nu(E\cap B_0)$ *and* $\nu'(E)=\nu(E\cap B_0')$, $E\in\sigma(\mathbf{L}_2)$. *Then* ν_0, $\nu'\in M_R^\sigma(\mathbf{L}_2)$ *and* $\nu=\nu_0+\nu'$. *Furthermore,* $\nu'(TS_\alpha)=0$ *for all* α.

Proof 1. Follows immediately from the fact that

$$\{TS_\alpha\}\subset\mathbf{L}_2\subset\sigma(\mathbf{L}_2).$$

2. Clear.

3. Let $E\in\sigma(\mathbf{L}_2)$ and let $\varepsilon>0$ be given. Then $\nu(E)-\nu(L)<\varepsilon$ for some $L\subset E$, $L\in\mathbf{L}_2$. But $\nu_0(E)-\nu_0(L)=\nu((E\cap B_0)-L)\leqq\nu(E)-\nu(L)<\varepsilon$. Therefore $\nu_0\in M_R^\sigma(\mathbf{L}_2)$. Similarly, $\nu'\in M_R^\sigma(\mathbf{L}_2)$. Also,

$$\nu(E) = \nu(E\cap(B_0\cup B_0')) = \nu(E\cap B_0)+\nu(E\cap B_0') =$$
$$= \nu_0(E)+\nu'(E), \qquad E\in\sigma(\mathbf{L}_2).$$

Suppose $\nu'(TS_\alpha)=\nu(TS_\alpha\cap B_0')>0$. Let $A=B_0\cup TS_\alpha\in\mathbf{S}$. Then $A=B_0\cup(TS_\alpha\cap B_0')$ and hence $\nu(A)>\nu(B_0)$, a contradiction. ∎

Lemma 4.2. *Suppose* $\nu_n\in M_R^\sigma(\mathbf{L})$ *and* $\nu_n\uparrow\nu$. *Assume* $\nu_n(X)<c$ *for all* n *and for some constant* c. *Then* $\nu\in M_R^\sigma(\mathbf{L})$.

Proof. If $A\in\mathbf{A}(\mathbf{L})$ then clearly $\nu(A)$ is finite. Let $\varepsilon>0$ be given. Then $\nu_n(A)-\nu_n(L_n)<\varepsilon/2$, $L_n\in\mathbf{L}$, $L_n\subset A$ for all n. Let $B=\bigcup_{n-1}^{\infty} L_n$. We assume $L_n\uparrow B$. Then

$B \in \sigma(\mathbf{L})$ and $\nu_n(A) - \nu_n(B) < \varepsilon/2$, $B \subset A$. Therefore, $\lim_{n \to +\infty} (\nu_n(A) - \nu_n(B)) = \nu(A) - \nu(B) \leqq \varepsilon/2$. But $\nu(B) - \nu(L_N) < \varepsilon/2$ for N large since ν is σ-smooth on $\sigma(\mathbf{L})$. Thus $\nu(A) - \nu(L_N) < \varepsilon$ and $L_N \subset A$. ∎

LEMMA 4.3. *Suppose T: $X \to Y$. Assume T is $\mathbf{L}_1 - \mathbf{L}_2$ continuous and $\mathbf{L}_2$ is a δ-lattice. Let $\mu \in M_R^\sigma(\mathbf{L}_1)$ and $\mu T^{-1} \in M_R^\sigma(\mathbf{L}_2)$. Suppose $\{S_\alpha\}$ is as in Lemma 4.1. If there exists a set $S \in \{S_\alpha\}$ such that $\mu(X) - \mu(S) < \varepsilon$ then using the notation of Lemma 4.1, $(\mu T^{-1})_0 = \mu T^{-1}$ and $(\mu T^{-1})' = 0$.*

PROOF. Let B_0 and $\mathbf{S}$ be as in Lemma 4.1 and let $a = \mu T^{-1}(B_0)$. Suppose $S \in \{S_\alpha\}$. Then $a \geqq \mu T^{-1}(TS) \geqq \mu(S)$. Therefore, $\mu(X) - a \leqq \mu(X) - \mu(S_\varepsilon) < \varepsilon$ for some $S_\varepsilon \in \{S_\alpha\}$ and hence $\mu(X) = a$. Let $E \in \sigma(\mathbf{L}_2)$. $(\mu T^{-1})'(E) = \mu T^{-1}(E \cap B_0') \leqq \mu T^{-1}(B_0') = \mu T^{-1}(Y) - \mu T^{-1}(B_0) = \mu(X) - a = 0$. Therefore, $\mu T^{-1} = (\mu T^{-1})_0$. ∎

DEFINITION. If $\mu \in M_R^\sigma(\mathbf{L})$, μ is *supported* by $W \in \sigma(\mathbf{L})$ if $\mu(W) = \mu(X)$.

LEMMA 4.4. *Suppose T: $X \to Y$ is $\mathbf{L}_1 - \mathbf{L}_2$ continuous where $\mathbf{L}_2$ is a δ-lattice. Let $\{S_\alpha\}$ and $\mathbf{S}$ be as in Lemma 4.1 and let $\varrho \in M_R^\sigma(\mathbf{L}_2)$ where ϱ is supported by $W \in \mathbf{S}$. Define $R_n = \bigcup_{i=1}^{n} S_i$, $S_i \in \{S_\alpha\}$ and $R_\infty = \bigcup_{i=1}^{\infty} S_i$. Finally, assume that for each R_n and for each $\nu \in M_R^\sigma(TR_n \cap \mathbf{L}_2)$ there exists $\lambda \in M_R^\sigma(R_n \cap \mathbf{L}_1)$ such that $\nu = \lambda T^{-1}$. Then $\varrho = \mu T^{-1}$ for some $\mu \in M_R^\sigma(\mathbf{L}_1)$.*

PROOF. Clearly, $W \subset \bigcup_{i=1}^{\infty} TS_i$ for some sequence $\{S_i\}$, $S_i \in \{S_\alpha\}$ and hence ϱ is supported by TR_∞. Consider $\varrho_n = \varrho|_{TR_n \cap \mathbf{L}_2}$. From Lemma 3.1, $\varrho_n \in M_R^\sigma(TR_n \cap \mathbf{L}_2)$. Since $\varrho(TR_\infty \cap E) = \varrho(E)$, $\varrho_n \uparrow \varrho$. By assumption there exists $\mu_n \in M_R^\sigma(R_n \cap \mathbf{L}_1)$ such that $\varrho_n = \mu_n T^{-1}$. If $\hat{\mu}_n(\hat{E}) = \mu_n(R_n \cap \hat{E})$, $\hat{E} \in \sigma(\mathbf{L}_1)$, then from Lemma 3.2, $\hat{\mu}_n \in M_R^\sigma(\mathbf{L}_1)$. Let μ be such that $\hat{\mu}_n \uparrow \mu$. Now $\varrho(TR_\infty) \geqq \varrho_n(TR_n) = \mu_n T^{-1}(TR_n) \geqq \mu_n(R_n) = \hat{\mu}_n(X)$. From Lemma 4.2 it follows that $\mu \in M_R^\sigma(\mathbf{L}_1)$. Furthermore, if $E \in \sigma(\mathbf{L}_2)$, $\varrho(E) = \lim_{n \to \infty} \varrho_n(E) = \mu T^{-1}(E)$. ∎

The following result is due to Bachman and Sultan [1] and is needed in the sequel.

LEMMA 4.5. *Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are lattices on X and suppose $\mathbf{L}_1 \subset \mathbf{L}_2$. If $\mathbf{L}_2$ is $\mathbf{L}_1$-countably paracompact or if $\mathbf{L}_2$ is countably compact then any $\mu \in M_R^\sigma(\mathbf{L}_1)$ extends to $\nu \in M_R^\sigma(\mathbf{L}_2)$.*

The following decomposition theorem generalizes a theorem found in Herz [7].

THEOREM 4.1. *Let T: $X \to Y$ be $\mathbf{L}_1 - \mathbf{L}_2$ continuous where $\mathbf{L}_2$ is a δ-lattice and let $\{S_\alpha\}$ be a class of subsets of $\mathbf{L}_1$ which is closed under finite unions and such that $TS_\alpha \in \mathbf{L}_2$ for all α. If $S \cap \mathbf{L}_1$ is $S \cap T^{-1}(\mathbf{L}_2)$ countably paracompact or if $S \cap \mathbf{L}_1$ is countably compact for all $S \in \{S_\alpha\}$ then for any $\nu \in M_R^\sigma(\mathbf{L}_2)$, $\nu = \mu T^{-1} + \nu'$ where μT^{-1}, $\nu' \in M_R^\sigma(\mathbf{L}_2)$, $\mu \in M_R^\sigma(\mathbf{L}_1)$ and $\nu'(TS) = 0$ for all $S \in \{S_\alpha\}$.*

PROOF. Let $\nu \in M_R^\sigma(\mathbf{L}_2)$. From Lemma 4.1, $\nu = \nu_0 + \nu'$ where $\nu_0, \nu' \in M_R^\sigma(\mathbf{L}_2)$ and $\nu'(TS) = 0$ for all $S \in \{S_\alpha\}$. If B_0 is as in our previous discussion, ν_0 is supported by B_0. Consider T restricted to S and let $\tau \in M_R^\sigma(TS \cap \mathbf{L}_2)$. Define $\varrho T^{-1}(B) = \tau(B)$, $B \in \sigma(TS \cap \mathbf{L}_2)$. Clearly, $\varrho \in M_R^\sigma(S \cap T^{-1}(\mathbf{L}_2))$. From Lemma 4.5 it follows that ϱ

can be extended to $\lambda \in M_R^\sigma(S \cap \mathbf{L}_1)$. Therefore, $\lambda T^{-1} = \tau$ and thus from Lemma 4.4, $v_0 = \mu T^{-1}$ for some $\mu \in M_R^\sigma(\mathbf{L}_1)$. ∎

COROLLARY 4.1. *Suppose* $T\colon X \to Y$ *is* $\mathbf{L}_1 - \mathbf{L}_2$ *continuous and onto where* $\mathbf{L}_2$ *is a* δ*-lattice. Suppose* $X \in \{S_\alpha\}$ *where* $\{S_\alpha\}$ *is as in Theorem 4.1. If* $S \cap \mathbf{L}_1$ *is* $S \cap T^{-1}(\mathbf{L}_2)$ *countably paracompact or if* $S \cap \mathbf{L}_1$ *is countably compact for each* $S \in \{S_\alpha\}$ *then* $v = \mu T^{-1}$ *for any* $v \in M_R^\sigma(\mathbf{L}_2)$ *where* $\mu \in M_R^\sigma(\mathbf{L}_1)$.

PROOF. $v'(TX) = v'(Y) = 0$. The proof now follows immediately from Theorem 4.1. ∎

Let $\mathbf{F}_X$ and $\mathbf{F}_Y$ denote the closed sets of the T_2-topological spaces X and Y, respectively, and let $\mathbf{K}_X$ and $\mathbf{K}_Y$, respectively, denote the lattices of compact sets of X and Y in the topological sense. (We note that this particular setting is that given in Herz [7].)

COROLLARY 4.2 (Hardy and Lacey [6]). *Suppose* $T\colon X \to Y$ *where* X *and* Y *are compact Hausdorff spaces. If* T *is continuous and onto then* $v = \mu T^{-1}$ *for any* $v \in M_R^\sigma(\mathbf{F}_Y)$, *where* $\mu \in M_R^\sigma(\mathbf{F}_X)$.

PROOF. Let $\mathbf{L}_1 = \mathbf{F}_X = \mathbf{K}_X = \{S_\alpha\}$. $\mathbf{L}_2 = \mathbf{F}_Y = \mathbf{K}_Y$. The result follows immediately from Corollary 4.1. ∎

In conclusion we note that we may obtain a somewhat more general result than that of Corollary 4.2 since Y need not be compact. In this case simply let $\mathbf{L}_2 = \mathbf{K}_Y \subset \mathbf{F}_Y$.

REFERENCES

[1] BACHMAN, G. and SULTAN, A., On regular extensions of measures, *Pacific J. Math.* **86** (1980), 389—395. *MR* **82f:** 28016.
[2] BACHMAN, G. and SULTAN, A., Regular lattice measures: mappings and spaces, *Pacific J. Math.* **67** (1976), 291—321. *MR* **58** #22476.
[3] BAUER, H., Measures avec une image donnée, *Rev. Roumaine Math. Pures Appl.* **11** (1966), 747—752. *MR* **34** #2814.
[4] GRASSI, P., On subspaces of replete and measure replete spaces, *Canad. Math. Bull.* **27** (1984), 58—64. *MR* **85b** : 28006.
[5] HALMOS, P., *Measure Theory*, Van Nostrand, New York, 1950. *MR* **11**—504.
[6] HARDY, J. and LACEY, H., Extensions of regular Borel measures, *Pacific J. Math.* **24** (1968), 277—282. *MR* **36** #5291.
[7] HERZ, C. S., The spectral theory of bounded functions, *Trans. Amer. Math. Soc.* **91** (1960), 181—232. *MR* **24** #A1627.

(Received June 13, 1984)

DEPARTMENT OF MATHEMATICS
HOFSTRA UNIVERSITY
HEMPSTEAD, N. Y. 11550
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 357—359

RINGS IN WHICH ADDITIVE MAPPINGS ARE MULTIPLICATIVE

S. DHOMPONGSA and J. SANWONG

In [4] Sullivan raised the problem of determining all rings R that satisfy the property: every additive mapping from R into itself is multiplicative; we shall refer to this property as property (A), and note in passing that if a ring R satisfies the property: every multiplicative mapping from R into itself is additive, then R can be highly non-trivial (see [2]).

Our first aim in this paper is to show that near-rings with an identity and property (A) are isomorphic to Z_2. After doing this we consider the situation when an identity is not assumed.

Recall [3] that a (right) near-ring is a set R with two operations $+$ and $\cdot$ such that $(R, +)$ is a group, $(R, \cdot)$ is a semigroup and $\cdot$ is right distributive over $+$.

THEOREM 1. *If R is a near-ring with identity* 1 *and property* (A) *then* $R=\{0, 1\}$.

PROOF. The map $f(x)=xa$ is additive and so $xya=xaya$ for all $x, y, a\in R$. In particular, $y0=1y0=0y0=0$ (since $0z=0$ for all $z\in R$) and $a=a^2$ for all $a\in R$. Hence $-1=(-1)^2=1$ and so $x+x=0$ for all $x\in R$. Thus, $-x=x=$ $=x(-1)\cdot 1\cdot(-1)=x(-1)$ and so (see [3], Proposition 1.109(a)) the operation $+$ is commutative. We can now regard R as a vector space over Z_2 with scalars written on the right (we thank Dr. N. Punnim for this observation). Suppose $|R|>2$, fix $a\in R\setminus\{0, 1\}$ and choose a basis B for R such that $1, a\in B$ (see [1], Ch. 9). Now let $f(1)=a$, $f(a)=1$, $f(x)=0$ for all $x\in B\setminus\{1, a\}$ and extend f to a linear transformation from R into itself (again with scalars on the right). Then f is additive and property (A) implies $0=f(0)=f[(1+a)a]=1+a$, a contradiction.

We now consider near-rings without identity: unfortunately, to obtain any satisfactory result in this situation we must assume the near-ring is in fact a ring.

THEOREM 2. *If R is a ring without non-zero nilpotents but satisfies property* (A) *then* $|R|\leqq 2$.

PROOF. The map $k(x)=x+x$ implies $xy+xy=0$ for all $x, y\in R$ and so $x^2+x^2=0$ for all $x\in R$. Hence $(x+x)^2=0$ and so $x+x=0$ for all $x\in R$. Also, using the map $k(x)=ax$ for each fixed $a\in R$, we obtain $x^3=x^4$ for all $x\in R$.

Note that we can again regard R as a vector space over Z_2. Suppose there exist distinct $e_1, e_2\in R\setminus 0$ such that $e_1e_2\notin\text{span}\,\{e_1, e_2\}$ and choose a basis B for R

1980 *Mathematics Subject Classification.* Primary 20M; Secondary 20M25.

Key words and phrases. Additive mappings, multiplicative mappings, near-rings, commutative rings.

such that $e_1, e_2, e_1e_2 \in B$. Let $f(e_1)=f(e_2)=e_1$, $f(e_1e_2)=e_2$ and $f(x)=0$ for all $x\in B\setminus\{e_1, e_2, e_1e_2\}$. Since the linear extension of f to the whole of R is additive, we have $e_1^2=f(e_1)f(e_2)=f(e_1e_2)=e_2$. Since $0=(e_1^3+e_1^2)^2$, we have $e_1^3+e_1^2=0$. Hence $e_1e_2=e_2$, a contradiction. Hence we may suppose $e_1e_2\in\{0, e_1, e_2, e_1+e_2\}$ for distinct $e_1, e_2\in R\setminus 0$. We choose a basis C for R such that $e_1, e_2\in C$. If $e_1e_2=e_1$, let $g(e_1)=e_1, g(e_2)=0$ and extend g to a linear map on R. Then $0=e_1 0=g(e_1)g(e_2)= =g(e_1e_2)=g(e_1)=e_1$, a contradiction. If $e_1e_2=e_1+e_2$, let $h(e_1)=h(e_2)=e_1$ and extend h to a linear map on R. Then $0=e_1+e_1=h(e_1)+h(e_2)=h(e_1+e_2)=h(e_1e_2)= =h(e_1)h(e_2)=e_1^2$, a contradiction. So, we can now assume $e_1e_2=0$ for all distinct $e_1, e_2\in R\setminus 0$; in particular, for such e_1, e_2 we have $e_1(e_1+e_2)=0$ and so $e_1^2=e_1e_2=0$ which is impossible. Hence, $|R|\leqq 2$ as required.

If we now consider rings that possibly contain nilpotents and satisfy property (A), the most we can establish is the following.

THEOREM 3. *If R is a ring that satisfies property* (A) *and contains an element satisfying $a^3\neq 0$ then R is commutative.*

PROOF. By considering the additive maps $f(x)=ax$, $f(x)=x+x$, $f(x)=x+ax$ and $f(x)=x+xa$ for each fixed $a\in R$ we obtain $axy=axay$, $xy+xy=0$, and $xya=xay=axy$. Consequently,

$$x^3+y^3 = x^4+y^4 = (x+y)^4 = (x+y)^3 = x^3+y^2x+x^2y+y^3$$

which implies $x^2y=y^2x$. We choose, by Zorn's Lemma, a maximal ideal M of R so that $a^3\notin M$. Thus M has the property: $x^3y^3\in M\Rightarrow x^3\in M$ or $a^3y^3\in M$. For suppose $x^3y^3\in M$ and $x^3\notin M$. Then $M+Rx^3$ is an ideal containing $M\cup\{x^3\}$ (since $(x^3)^2=x^3$) and so $a^3\in M+Rx^3$. That is, $a^3=m+rx^3$ for some $m\in M$, $r\in R$ and then $a^3y^3=my^3+rx^3y^3\in M$. Now fix $b\in R$ and define φ by: $\varphi(x)=0$ if $x^3\in M$ and $f(x)=b^2$ if $x^3\notin M$. If $x^3, y^3\in M$, then $(x+y)^3=x^3+y^3\in M$ and $\varphi(x+y)= =0=\varphi(x)+\varphi(y)$; if $x^3\in M$, $y^3\notin M$, then $(x+y)^3\notin M$ and $\varphi(x+y)=b^2=\varphi(x)+ +\varphi(y)$; if $x^3, y^3\notin M$, then $a^3x^3y^3\notin M$ (otherwise, $a^3x^3y^3\in M\Rightarrow y^3\in M$ or $(a^3)^2x^3\in M\Rightarrow \Rightarrow a^3x^3\in M\Rightarrow x^3\in M$ or $a^3=(a^3)^2\in M$, a contradiction). But $x^3y^3(x+y)^3=0\in M$ where $a^3x^3y^3\notin M$. Hence $(x+y)^3\in M$ and so in this case we have $\varphi(x)+\varphi(y)= =b^2+b^2=0=\varphi(x+y)$. That is, φ is additive. Now since M is an ideal and $a^3\notin M$, we have $a, a^2\notin M$ and so, by property (A), $\varphi(a^2)=(\varphi(a))^2$. That is $b^2=(b^2)^2$ for all $b\in R$. Hence,

$$(x+y)^2 = (x+y)^4 = x^4+y^4 = x^2+y^2$$

for all $x, y\in R$ and we conclude that $xy=yx$ for all $x, y\in R$.

Acknowledgement. We are grateful to Dr. R. P. Sullivan for his helpful guidance during the preparation of this paper.

REFERENCES

[1] Jacobson, N., *Lectures in Abstract Algebra*, Vol. 2, Van Nostrand, New York, 1953. *MR* **14**—837.

[2] Martindale, W. S., *When are multiplicative mappings additive?*, *Proc. Amer. Math. Soc.* **21** (1969), 695—698. *MR* **39** #1483.

[3] Pilz, G., *Near-rings*, North-Holland, New York, 1977. *MR* **57** #9761.

[4] Sullivan, R. P., *Research problems 23* (ed. A. Hajnal), *Periodica Mathematica Hungarica* **5** (1977), 813—814.

(Received June 19, 1984)

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY
CHIANG MAI 50002
THAILAND

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 361—364

ON PSEUDO-QUADRATIC PROGRAMMING

D. G. KABE

1. Introduction

Dragomirescu [1] considers a class of pseudo-quadratic programs of the type

$$\text{Min}\, \varphi(c'x, x'Vx) \qquad (x \in X), \tag{1}$$

where c is a given n component (column) vector, V is a given $n \times n$ positive definite symmetric matrix, and X is the polyhedron

$$X = \{Ax \geqq b,\ x \geqq 0\}. \tag{2}$$

By assuming $\varphi = \varphi(u_1, u_2)$ to be convex on X, and φ to have continuous partial derivatives φ_{u_1}, φ_{u_2}, and $\varphi_{u_2} > 0$, he shows that (1) is equivalent to the parametric quadratic programming problem

$$\text{Min}\, (\lambda c'x + x'Vx), \quad (x \in X), \tag{3}$$

where λ is a real parameter.

Indeed, the problem (1) is equivalent to problem (3). However, Dragomirescu's [1] procedure for solving (3) is in error, and he solves his numerical illustrative example incorrectly. He considers the problem

$$\text{Min}\, \{\lambda(-4x_1 - 6x_2) + x_1^2 + 2x_1x_2 + 2x_2^2\}, \tag{4}$$

subject to

$$x_1 + x_2 \leqq 2, \quad 3x_1 + 2x_2 \geqq 3, \quad x_1, x_2 \geqq 0, \tag{5}$$

and obtains the solution $\lambda = 1/11$, $x_1 = 1$, $x_2 = 0$. The correct solution to the problem is $\lambda = 1/5$, $x_1 = 1$, $x_2 = 0$, which yields the minimum $\frac{1}{5}$. Dragomirescu [1] obtains the minimum to be $\frac{7}{11}$.

Our purpose in this paper is to give the correct solution to (4). We use the methodology of vanishing Jacobian theory, see e.g., Stewart [2], pp. 178—192, which is given in the next section, and Section 3 solves (4).

1980 *Mathematics Subject Classification.* Primary 90C.
Key words and phrases. Quadratic programming, vanishing Jacobian theory, optimal solution.

2. Vanishing Jacobian theory

In the simplest case of vanishing Jacobian theory, we have n functions, $\varphi_1, \varphi_2, \ldots, \varphi_n$ of n variables each, and it is required to find a critical point of φ_1, under the restrictions

$$\varphi_i(y_1, \ldots, y_n) = a_i, \qquad i = 2, \ldots, n, \tag{6}$$

where a_i are known constants. From (6) we observe that we have only one independent variable, say y_1, and hence the first order conditions are

$$\sum_{i=1}^{n} \frac{\partial \varphi_j}{\partial y_i} \frac{dy_i}{dy_1} = 0, \qquad j = 1, \ldots, n. \tag{7}$$

On eliminating $(n-1)$ quantities $dy_2/dy_1, \ldots, dy_n/dy_1$, from (7), we find the eliminant Jacobian to be

$$\begin{vmatrix} \dfrac{\partial \varphi_1}{\partial y_1} & \cdots & \dfrac{\partial \varphi_1}{\partial y_n} \\ \cdot & \cdots & \cdot \\ \dfrac{\partial \varphi_n}{\partial y_1} & & \dfrac{\partial \varphi_n}{\partial y_n} \end{vmatrix} = 0. \tag{8}$$

Thus a solution y_0 of (8) and (6) is a critical point. Now the second order conditions are obtained as follows. Note that

$$\frac{d^2 \varphi_1}{dy_1^2} = \lambda_1 + \sum_{j=2}^{n} \frac{\partial \varphi_1}{\partial y_j} \frac{d^2 y_j}{dy_1^2}, \tag{9}$$

where

$$\lambda_k = \sum_{i,j=1}^{n} \frac{\partial^2 \varphi_k}{\partial y_i \partial y_j} \frac{dy_i}{dy_1} \frac{dy_j}{dy_1}, \qquad k = 1, \ldots, n. \tag{10}$$

Similarly, from (6), we find that

$$\lambda_i + \sum_{j=2}^{n} \frac{\partial \varphi_i}{\partial y_j} \frac{d^2 y_j}{dy_1^2} = 0, \qquad i = 2, \ldots, n. \tag{11}$$

On eliminating $d^2 y_j/dy_1^2$, $j=2, \ldots, n$ from (9) and (11), we find $d^2\varphi_1/dy_1^2$ to be

$$\begin{vmatrix} \lambda_1 & \dfrac{\partial \varphi_1}{\partial y_2} & \cdots & \dfrac{\partial \varphi_1}{\partial y_n} \\ \cdot & \cdot & \cdots & \cdot \\ \lambda_n & \dfrac{\partial \varphi_n}{\partial y_2} & \cdots & \dfrac{\partial \varphi_n}{\partial y_n} \end{vmatrix} \Bigg/ \begin{vmatrix} \dfrac{\partial \varphi_2}{\partial y_2} & \cdots & \dfrac{\partial \varphi_2}{\partial y_n} \\ \cdot & \cdots & \cdot \\ \dfrac{\partial \varphi_n}{\partial y_2} & \cdots & \dfrac{\partial \varphi_n}{\partial y_n} \end{vmatrix}. \tag{12}$$

We note that the numerator of (12) is fixed, and once the sign of this numerator is determined, the sign of (12) is the sign of the denominator of (12). This denominator is the cofactor of $\partial\varphi_1/\partial y_1$ in (8).

The critical point y_0 establishes an optimal relationship between the functions $\varphi_1, \varphi_2, \dots, \varphi_n$, which may be explained as follows. Let $\varphi_1(y_0)=a_1$. Then consider the inverse optimization problem

$$\text{optimize } \varphi_i(y), \tag{13}$$

subject to

$$\varphi_j(y) = a_j, \qquad j = 1, \dots, n, \quad i \neq j, \tag{14}$$

and note that (13) will have either a maximum or minimum at y_0. The nature of this extremum is determined by the sign of $d^2\varphi_i/dy_1^2$. This sign is the sign of the cofactor of $\partial\varphi_i/\partial y_1$ in (8).

We now proceed to solve (4).

3. Numerical example

Dragomirescu [1] shows that a solution to (4) satisfies

$$x_1+x_2+x_3 = 2, \quad 3x_1+2x_2 = 3, \quad x \geqq 0. \tag{15}$$

An arbitrary solution to (15) is

$$x = [1, 0, 1]', \tag{16}$$

and a general solution to (15) is

$$x = [1-2t, 3t, 1-t]', \tag{17}$$

we determine t such that (17) is optimal to (4). Our three functions are

$$\begin{aligned} \varphi_1 &= \lambda(-4x_1-6x_2)+x_1^2+2x_1x_2+2x_2^2 \\ \varphi_2 &= x_1+x_2+x_3 = 2 \\ \varphi_3 &= 3x_1+2x_2 = 3. \end{aligned} \tag{18}$$

The Jacobian equation of (18) is

$$5\lambda-x_1-4x_2 = 0. \tag{19}$$

On substituting (17) in (19) we have that

$$5\lambda-1+10t = 0, \tag{20}$$

and this yields

$$\lambda = \frac{1-10t}{5}. \tag{21}$$

On substituting (21) and (17) in (4) we obtain

$$\text{Min } \{30t^2+8t+1/5\}, \quad t \geqq 0. \tag{22}$$

Now, we note (22) increases with t and hence $t=0$ yields the unique minimum, and hence from (21) $\lambda=1/5$ and from (15) $x_1=1$, $x_2=0$, $x_3=1$.

REFERENCES

[1] Dragomirescu, M., Pseudo-quadratic programming, *Studia Sci. Math. Hungar.* **7** (1972), 167—179. *MR* **47** #6316.
[2] Stewart, C. A., *Advanced Calculus*, Methuen, London, England, 1940. *MR* **1**—299.

(Received August 23, 1984)

SAINT MARY'S UNIVERSITY
HALIFAX, NOVA SCOTIA
B3H 3C3
CANADA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 365—377

ON THE RADEMACHER—MENSHOV THEOREM IN VON NEUMANN ALGEBRAS

S. GOLDSTEIN and A. ŁUCZAK

Introduction

A great deal of work has been done lately to generalize to the context of von Neumann algebras various theorems of probability theory. The aim of the paper is twofold. Firstly, we prove a non-commutative version of the Rademacher—Menshov theorem, a fundamental result in the theory of orthogonal random variables. We think that in operator algebras "orthogonality" is a more natural concept than "independence" (cf. [1]). Secondly, we try to demonstrate that there may be many ways of defining "almost sure" convergence for operators, especially when it is natural to formulate a theorem to be proved for some L^p-space over a von Neumann algebra rather than for the algebra itself. More precisely, let M be a von Neumann algebra with a faithful normal state ϱ. Following Kosaki [4], we consider, for each $\Theta\in[0, 1]$, the imbedding $\eta_\Theta\colon M\to M_*$ given by $\eta_\Theta(x)=\sigma^\varrho_{-i\Theta}(x)\varrho$ and define $L^p_\Theta(M, \varrho)$ as the p-th complex interpolation space for the pair of Banach spaces $(\eta_\Theta(M), M_*)$. This procedure gives a one-parameter family of L^p-spaces and the usual definition of almost uniform convergence of operators (which is a suitable replacement for the almost sure convergence in the non-commutative case) seems appropriate only for $\Theta=0$. In the sequel, we formulate all the results in two versions, called respectively "left" and "symmetric" which correspond to the cases $\Theta=0$ and $\Theta=\frac{1}{2}$. In the left version, we put no restriction on the algebra M and prove the Rademacher—Menshov theorem for orthogonal sequences of vectors taken from the closure in $L^2(M, \varrho)$ of the self-adjoint part M_{sa} of the algebra. In the symmetric version, we manage to prove the theorem for arbitrary elements of $L^2(M, \varrho)$, at the expense of requiring the pair (M, ϱ) to satisfy some additional condition which we call, by abuse of language, the "$\natural$-boundedness" of M. This condition is satisfied, if the state ϱ has bounded density with bounded inverse with respect to a n.s.f. trace on M.

1980 *Mathematics Subject Classification.* Primary 46L50; Secondary 46L10, 60B99.

Key words and phrases. Rademacher—Menshov theorem, positive cone of a von Neumann algebra, convergences in von Neumann algebras.

1. Preliminaries and notation

Throughout the paper, M is a σ-finite von Neumann algebra with a normal faithful state ϱ. We assume that M acts, in the standard way, in a Hilbert space H with a cyclic and separating vector ξ_0 such that $\varrho(x)=(x\xi_0, \xi_0)$ for x in M. Furthermore, J and Δ are the modular conjugation and the modular operator, respectively, both canonically associated with (M, H, ξ_0); $(\sigma_t\colon t\in R)$ is the modular automorphism group of M defined as $\sigma_t(x)=\Delta^{it}x\Delta^{-it}$ for x in M. M_0 stands for the set of entire elements of M for the group (σ_t), i.e. the set of these elements x of M for which the function $t\to\sigma_t(x)$ has an entire analytic continuation. In particular, M_0 is strongly dense in M.

$\mathcal{P}^{\#}$ and $\mathcal{P}^{\natural}$ are the cones $(M_+\xi_0)^-$ and $(\Delta^{1/4}M_+\xi_0)^-$, respectively. For ξ, η in H, we write $\xi\leqq\eta$ if $\eta-\xi\in\mathcal{P}^{\natural}$.

M_{sa} is the set of self-adjoint elements of M and we put $H_{sa}=(M_{sa}\xi_0)^-$.

To simplify the notation, the element $xJxJ$ is denoted by $\tilde{x}$. Let us note that, for x, y in M (or M'), $\widetilde{xy}=\tilde{x}\tilde{y}$.

For each vector ξ in H, we define two normal functionals ω_ξ and φ_ξ by:

$$\omega_\xi(x) = (x\xi, \xi), \qquad x\in M,$$
$$\varphi_\xi(x) = (\Delta^{1/4}x\xi_0, \xi), \quad x\in M.$$

The functionals φ_ξ were used by Kosaki in [5].

PROPOSITION 1.1. *For each x in M,*

$$\omega_{\Delta^{1/4}x\xi_0} = \varphi_{\tilde{x}\xi_0}.$$

PROOF. For an arbitrary y in M_0, we have

$$\varphi_{\tilde{x}\xi_0}(y) = (\Delta^{1/4}y\xi_0, xJxJ\xi_0) = (\sigma_{-i/4}(y)\xi_0, JxJx\xi_0) =$$
$$= (\sigma_{-i/4}(y)Jx^*J\xi_0, x\xi_0) = (\Delta^{1/4}y\Delta^{-1/4}Jx^*\xi_0, x\xi_0) =$$
$$= (y\Delta^{-1/4}JJ\Delta^{1/2}x\xi_0, \Delta^{1/4}x\xi_0) = (y\Delta^{1/4}x\xi_0, \Delta^{1/4}x\xi_0) = \omega_{\Delta^{1/4}x\xi_0}(y).$$

Since M_0 is dense in M, the assertion follows. ∎

The following form of the Cauchy—Schwarz inequality will be useful in the sequel:

LEMMA 1.2. *For arbitrary $\xi_1, \dots, \xi_n$ in H, we have*

$$\omega_{\xi_1+\dots+\xi_n} \leqq n(\omega_{\xi_1}+\dots+\omega_{\xi_n}).$$

The next proposition establishes the properties of the cone $\mathcal{P}^{\natural}$ we shall need throughout (see [9], Sec. 10.23, 10.25).

PROPOSITION 1.3. (i) $\mathcal{P}^{\natural}=\{\tilde{x}\xi_0\colon x\in M\}^-$,

(ii) *$\mathcal{P}^{\natural}$ is invariant with respect to J,*

(iii) *$\mathcal{P}^{\natural}$ is self-dual, i.e. $\mathcal{P}^{\natural}=\{\xi\in H\colon (\xi, \eta)\geqq 0\}$ for each $\eta\in\mathcal{P}$,*

(iv) *the mapping $\xi\in\mathcal{P}^{\natural}\to\omega_\xi\in M_*^+$ is bijective.*

From (iii) it follows, in particular, that φ_ξ belongs to M_*^+ for ξ in $\mathscr{P}^\natural$.

For x in M, ξ in H, φ in M_*, the symbols $\|x\|$, $\|\xi\|$, $\|\varphi\|$ denote the usual norms in M, H and M_*, respectively. These ones are used while proving theorems in the "left version". The "symmetric version" uses the following "norms".

DEFINITION 1.4. For x in M, ξ in $\mathscr{P}^\natural$ and φ in M_*^+, we define $\|x\|_2$, $\|\xi\|_\infty$ and $\|\varphi\|_\infty$ as follows:

$$\|x\|_2 = \|\Delta^{1/4}x\xi_0\|;$$

$$\|\xi\|_\infty = \inf\{c\colon \xi \leqq c\xi_0\};$$

$$\|\varphi\|_\infty = \inf\{c\colon \varphi \leqq c\varrho\} \quad \text{where} \quad \inf\emptyset = +\infty.$$

It is easily seen that $\|\cdot\|_2$ is a true norm given by the inner product $(x, y)_2 = (\Delta^{1/4}x\xi_0, \Delta^{1/4}y\xi_0)$. Moreover,

$$\|x\|_2^2 = (\Delta^{1/4}x\xi_0, \Delta^{1/4}x\xi_0) = \overline{(J\Delta^{1/2}x\xi_0, Jx\xi_0)} =$$

$$= \overline{(x^*\xi_0, Jx\xi_0)} = (xJx\xi_0, \xi_0) = (\bar{x}\xi_0, \xi_0).$$

The "norms" $\|\cdot\|_\infty$ in $\mathscr{P}^\natural$ and M_*^+ are positively homogeneous and subadditive.

PROPOSITION 1.5. (i) *For an arbitrary x in M_+,* $\|\Delta^{1/4}x\xi_0\|_\infty = \|x\|$.

(ii) *For an arbitrary ξ in $\mathscr{P}^\natural$,* $\|\varphi_\xi\|_\infty = \|\xi\|_\infty$.

PROOF. (i) Since $x \leqq \|x\|$, we have $\Delta^{1/4}x\xi_0 \leqq \|x\|\,\xi_0$ and, thus, $\|\Delta^{1/4}x\xi_0\|_\infty \leqq \leqq \|x\|$. Putting $c = \|\Delta^{1/4}x\xi_0\|_\infty$, we have $\Delta^{1/4}x\xi_0 \leqq c\xi_0$ and, consequently, for each y in M,

$$(\Delta^{1/4}x\xi_0, \Delta^{1/4}y^*y\xi_0) \leqq c(\xi_0, \Delta^{1/4}y^*y\xi_0) = c\|y\xi_0\|^2 = c\|JyJ\xi_0\|^2.$$

On the other hand,

$$(\Delta^{1/4}x\xi_0, \Delta^{1/4}y^*y\xi_0) = (x\xi_0, Jy^*JJyJ\xi_0) = (xJyJ\xi_0, JyJ\xi_0).$$

Taking into account the relation $JMJ = M'$ and the density of $M'\xi_0$ in H, we obtain that $\|x\| \leqq c$.

(ii) Assume that $\xi \leqq c\xi_0$. Then, for y in M_+,

$$\varphi_\xi(y) = (\Delta^{1/4}y\xi_0, \xi) \leqq (\Delta^{1/4}y\xi_0, c\xi_0) = c\varrho(y),$$

thus $\varphi_\xi \leqq c\varrho$.

Now, if $\varphi_\xi \leqq c\varrho$, then there exists h in M'_+ such that $\varphi_\xi(y) = (yh\xi_0, h\xi_0)$ for y in M, and $\|h\| \leqq \sqrt{c}$. Put $x = Jh^2J$, $\eta = \Delta^{1/4}x\xi_0$. Then

$$\varphi_\eta(y) = (\Delta^{1/4}y\xi_0, \Delta^{1/4}Jh^2J\xi_0) = (y\xi_0, \Delta^{1/2}Jh^2J\xi_0) =$$

$$= (y\xi_0, h^2\xi_0) = (yh\xi_0, h\xi_0) = \varphi_\xi(y).$$

Since $\Delta^{1/4}M\xi_0$ is dense in H ($\Delta^{1/4}M_+\xi_0$ is dense in $\mathscr{P}^\natural$ and $H = (\mathscr{P}^\natural - \mathscr{P}^\natural) + i(\mathscr{P}^\natural - \mathscr{P}^\natural)$), the equality $\varphi_\eta = \varphi_\xi$ yields $\xi = \eta$, that is, $\xi = \Delta^{1/4}x\xi_0$. It follows that $\xi \leqq c\xi_0$, hence $\varphi_\xi \leqq c\varrho$ if and only if $\xi \leqq c\xi_0$, which proves (ii). ∎

Now, we are going to introduce the notion of the $\natural$-boundedness of the algebra M.

An element x of M is said to be analytic in the vertical strip

$$\{\alpha\in\mathbf{C}:\ -\varepsilon_1 \leqq \operatorname{Re}\alpha \leqq \varepsilon_2\} \qquad (0 \leqq \varepsilon_1, \varepsilon_2 < \infty)$$

if there exists an M-valued function F, defined and ultraweakly continuous on the strip and analytic inside the strip, such that $F(it)=\sigma_t(x)$ for t in R. For each α from this strip, we put $\sigma_{-i\alpha}(x)=F(\alpha)$. For α in $\mathbf{C}$ and x in M, we write $x\in\mathscr{D}(\sigma_\alpha)$ if x is analytic in some vertical strip containing $i\alpha$.

Definition 1.6. Algebra M is said to be $\natural$-bounded if $M\subset\mathscr{D}(\sigma_{-i/4})$.

We shall now define three types of almost uniform convergence. The first one is well-known and was widely used, e.g. by Lance in [6] and [7], in the context of the algebra M itself, and by Gol'dštein in [3], in the context of the space H, as below. The next two refer to the "symmetric version" and seem to be appropriate, though a little bit sophisticated, counterparts of the first.

Definition 1.7. In all those definitions the first part is the same and reads: "For each $\varepsilon>0$, there is a projection e in M with $\varrho(e)>1-\varepsilon$ such that ...", so we shall not repeat it each time.

(i) $\xi_n\to\xi$ almost uniformly $(\xi_n, \xi\in H)$ if $e(\xi_n-\xi)=x_n\xi_0$ for some x_n in M and $\|x_n\|\to 0$;

(ii) $\xi_n\to\xi$ $\natural$-almost uniformly $(\xi_n, \xi\in H)$ if $e\in\mathscr{D}(\sigma_{-i/4})$ and

$$\|\sigma_{-i/4}(e)\omega_{\xi_n-\xi}\sigma_{i/4}(e)\|_\infty\to 0;$$

(iii) $x_n\to x$ $\natural$-almost uniformly $(x_n, x\in M)$ if $\|\tilde{e}(x_n-x)^{\sim}\xi_0\|_\infty\to 0$.

The following propositions make the above definitions clearer.

Proposition 1.8. *Let M be a $\natural$-bounded von Neumann algebra. Then $x_n\to x$ $\natural$-almost uniformly if and only if $\Delta^{1/4}x_n\xi_0\to\Delta^{1/4}x\xi_0$ $\natural$-almost uniformly.*

Proof. By definition and Propositions 1.1, 1.5 (ii), we have $\Delta^{1/4}x_n\xi_0\to\Delta^{1/4}x\xi_0$ $\natural$-almost uniformly if and only if

$$\|\sigma_{-i/4}(e)\omega_{\Delta^{1/4}(x_n-x)\xi_0}\sigma_{i/4}(e)\|_\infty = \|\omega_{\Delta^{1/4}e(x_n-x)\xi_0}\|_\infty =$$
$$= \|\varphi_{e(x_n-x)^{\sim}\xi_0}\|_\infty = \|\tilde{e}(x_n-x_n)^{\sim}\ \xi_0\|_\infty \to 0,$$

that is, if $x_n\to x$ $\natural$-almost uniformly. ∎

Proposition 1.9. *Let M be a $\natural$-bounded von Neumann algebra. Then $x_n\to x$ $\natural$-almost uniformly if, for each $\varepsilon>0$, there is a projection e in M with $\varrho(e)>1-\varepsilon$ such that $\|\sigma_{i/4}(e(x_n-x))\|\to 0$.*

Proof. By Proposition 1.5 (i), we have

$$\|e(x_n-x)Je(x_n-x)J\xi_0\|_\infty = \|\Delta^{1/4}\sigma_{i/4}(e(x_n-x))\sigma_{i/4}(e(x_n-x))^*\xi_0\|_\infty =$$
$$= \|\sigma_{i/4}(e(x_n-x))\|^2$$

and the assertion follows. ∎

2. Maximal theorems

The following immediate consequence of the maximal ergodic theorem from [3; Th. 1.2] is a basic tool in proving various theorems on almost uniform convergence:

THEOREM 2.1. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of positive operators from M and $\{\varepsilon_n\}_{n=1}^{\infty}$ — a sequence of positive numbers. Then there exists a projection e in M such that $\varrho(e)\geqq 1-2\sum_{n=1}^{\infty}\varepsilon_n^{-1}\varrho(x_n)$ and $\|ex_ne\|\leqq 2\varepsilon_n$ for $n=1, 2, \ldots$.*

Now, we shall prove a "symmetric" version of this theorem.

THEOREM 2.2. *Let $\{x_i\}_{i\in I}$ be a family of operators from M and $\{\varepsilon_i\}_{i\in I}$ — a family of positive numbers. Assume, moreover, that M is $\natural$-bounded. Then there exist a positive constant c and a projection e in M, such that $\|\bar{e}\bar{x}_i\xi_0\|_\infty\leqq c\varepsilon_i$ for each $i\in I$ and $\|e\|_2^2\geqq 1-\sum_{i\in I}^{\infty}\varepsilon_i^{-1}\|x_i\|_2^2$.*

The proof of the theorem will be an outcome of several lemmas.

LEMMA 2.3. $\{\bar{x}\xi_0\colon x\in M\}\subset\Delta^{1/4}\mathscr{P}^{*}$.

PROOF. The set $M\xi_0$ with the multiplication $(x\xi_0)(y\xi_0)=xy\xi_0$ and the involution $(x\xi_0)^{\#}=x^*\xi_0$ is a left Hilbert algebra. Let $\mathscr{T}$ be the Tomita algebra associated with $M\xi_0$ (see [9], pp. 298, 300).

Take an arbitrary x in M and choose x_n in M_0, such that
(i) $x_n\to x$ strongly;
(ii) $x_n^*\to x^*$ strongly;
(iii) $\|x_n\|\leqq\|x\|$ for $n=1, 2, \ldots$

(cf. [9], Cor. 2, p. 303).

From the properties of $\mathscr{T}$ we have $J\Delta^{1/4}x_n\xi_0\in\mathscr{T}$, $n=1, 2, \ldots$. Let $y_n\in M$ be such that $y_n\xi_0=J\Delta^{1/4}x_n\xi_0$. Then

$$\Delta^{1/4}y_n^*y_n\xi_0=(\Delta^{1/4}y_n^*\xi_0)(\Delta^{1/4}y_n\xi_0)=(x_n\xi_0)(Jx_n\xi_0)=\bar{x}_n\xi_0,$$

(see [9], Th. 10.20, p. 298). Conditions (i), (ii), (iii) yield $\bar{x}_n\xi_0\to\bar{x}\xi_0$ (in norm) and, consequently, $\Delta^{1/4}y_n^*y_n\xi_0\to\bar{x}\xi_0$. On the other hand,

$$\|y_n^*y_n\xi_0-y_m^*y_m\xi_0\|^2=\|\Delta^{-1/4}(\bar{x}_n-\bar{x}_m)\xi_0\|^2=$$
$$=\|\Delta^{-1/4}J(\bar{x}_n-\bar{x}_m)\xi_0\|^2=\|J\Delta^{1/4}(\bar{x}_n-\bar{x}_m)\xi_0\|^2=$$
$$=\|\Delta^{1/4}(\bar{x}_n-\bar{x}_m)\xi_0\|^2=\big(\Delta^{1/2}(\bar{x}_n-\bar{x}_m)\xi_0,(\bar{x}_n-\bar{x}_m)\xi_0\big)=$$
$$=\big(J(\bar{x}_n-\bar{x}_m)\xi_0,(\bar{x}_n^*-\bar{x}_m^*)\xi_0\big)\to 0\quad\text{as}\quad n,\ m\to\infty$$

since, by (i), (ii), (iii), $\bar{x}_n^*\xi_0\to\bar{x}^*\xi_0$.

Thus $y_n^*y_n\xi_0$ tends to some $\xi\in\mathscr{P}^*$, $\Delta^{1/4}$ being closed, so we infer that $\Delta^{1/4}\xi=\bar{x}\xi_0$, which ends the proof. ∎

LEMMA 2.4. *If the algebra M is $\natural$-bounded, then, for each projection e in M, there is a positive constant c such that $\tilde{e}\xi_0 \leqq c\xi_0$, i.e. $\|\tilde{e}\xi_0\|_\infty \leqq c$.*

PROOF. On account of Lemma 2.3, $\tilde{e}\xi_0 \in \mathscr{D}(\Delta^{-1/4})$ and, consequently,

$$\tilde{e}\xi_0 = \Delta^{1/4}\Delta^{-1/4}\tilde{e}\xi_0 = \Delta^{1/4}(\Delta^{-1/4}e\Delta^{1/2}e\Delta^{-1/4})\xi_0 = \Delta^{1/4}|\sigma_{-i/4}(e)|^2\xi_0.$$

By Proposition 1.5 (i), we have $\|\tilde{e}\xi_0\|_\infty = \|\sigma_{-i/4}(e)\|^2 < \infty$, which establishes the claim. ∎

LEMMA 2.5. *Let $\{x_i\}_{i\in I}$ be a family of operators from M and $\{\varepsilon_i\}_{i\in I}$ — a family of positive numbers. Let us further assume that there exists η in $\mathscr{P}^\natural$, $\eta \leqq \xi_0$, such that, for an arbitrary ξ in $\mathscr{P}^\natural$, $\xi \leqq \eta$,*

$$(*) \qquad (\tilde{x}_i\xi_0, \xi) \leqq \varepsilon_i(\xi_0, \xi) \quad \textit{for each} \quad i \in I.$$

Then there exists a projection e in M such that

$$\|e\|_2^2 \geqq (\varepsilon_0, \eta) \quad \textit{and} \quad \tilde{e}\tilde{x}_i\xi_0 \leqq \varepsilon_i\tilde{e}\xi_0 \quad \textit{for each} \quad i \in I.$$

PROOF. Let e be the support (in M) of the functional ω_η. Then $\tilde{e}\mathscr{P}^\natural$ is the closure of the face generated in $\mathscr{P}^\natural$ by η (see [2], Lemma 4.5). Hence $\|e\|_2^2 = (\tilde{e}\xi_0, \xi_0) \geqq (\tilde{e}\eta, \xi_0) = (\eta, \xi_0)$; moreover, condition $(*)$ holds for each element of $\tilde{e}\mathscr{P}^\natural$. Thus, for an arbitrary ξ in $\mathscr{P}^\natural$, $(\tilde{x}_i\xi_0, \tilde{e}\xi) \leqq \varepsilon_i(\xi_0, e\xi)$, which gives $\tilde{e}\tilde{x}_i\xi_0 \leqq \varepsilon_i\tilde{e}\xi_0$ for $i \in I$. ∎

LEMMA 2.6. *Let $\{x_i\}_{i\in I}$ be a family of operators from M and $\{\varepsilon_i\}_{i\in I}$ — a family of positive numbers. Then there exists an element η in $\mathscr{P}^\natural$, $\eta \leqq \xi_0$, such that*

(i) *for each* $\xi \in \mathscr{P}^\natural$, $\xi \leqq \eta$, $(\tilde{x}_i\xi_0, \xi) \leqq \varepsilon_i(\xi_0, \xi)$, $i \in I$;

(ii) $(\xi_0, \eta) \geqq 1 - \sum_{i\in I} \varepsilon_i^{-1}\|x_i\|_2^2$.

PROOF. We can assume that $\sum_{i\in I} \varepsilon_i^{-1}\|x_i\|_2^2 < \infty$. Consider the space H^I with the product topology where H is endowed with the weak topology. Put

$$\mathscr{L} = \{(\xi_i)_{i\in I} \in H^I \colon \xi_i \in \mathscr{P}^\natural \text{ for } i \in I, \ \sum_{i\in I} \xi_i \leqq \xi_0\}$$

where the sum $\sum_{i\in I}$ is convergent in the weak sense. $\mathscr{L}$ is compact in H^I because the set $\{\xi \in \mathscr{P}^\natural : \xi \leqq \xi_0\}$ is closed and bounded in norm in H. We have

$$\sum_{i\in I} \varepsilon_i^{-1}(\tilde{x}_i\xi_0, \xi_i) \leqq \sum_{i\in I} \varepsilon_i^{-1}(\tilde{x}_i\xi_0, \xi_0) = \sum_{i\in I} \varepsilon_i^{-1}\|x_i\|_2^2 < \infty$$

and, consequently, the function $g \colon \mathscr{L} \to R$ defined as

$$g((\xi_i)_{i\in I}) = \sum_{i\in I} \varepsilon_i^{-1}(\tilde{x}_i\xi_0, \xi_i) - \sum_{i\in I} (\xi_0, \xi_i)$$

is continuous on $\mathcal{L}$. Let $(\eta_i)_{i\in I}$ be a point in $\mathcal{L}$ in which g takes its greatest value. Then $g((\eta_i)_{i\in I})\geqq 0$. Put $\eta=\xi_0-\sum_{i\in I}\eta_i$. Then $\eta\in\mathcal{P}^\natural$ and $\eta\leqq\xi_0$.

We shall show that conditions (i) and (ii) hold. Take arbitrary $\xi\leqq\eta$ and $i_0\in I$. Let $\xi_i=\eta_i$ for $i\neq i_0$ and $\xi_{i_0}=\eta_{i_0}+\xi$. Then $(\xi_i)_{i\in I}$ belongs to $\mathcal{L}$ and, therefore, $g((\xi_i)_{i\in I})\leqq g((\eta_i)_{i\in I})$, which proves the inequality in (i). Moreover,

$$(\xi_0,\eta)=1-\sum_{i\in I}(\xi_0,\eta_i)\geqq 1-\sum_{i\in I}\varepsilon_i^{-1}(\tilde{x}_i\xi_0,\eta_i)\geqq$$

$$\geqq 1-\sum_{i\in I}\varepsilon_i^{-1}(\tilde{x}_i\xi_0,\xi_0)=1-\sum_{i\in I}\varepsilon_i^{-1}\|x_i\|_2^2,$$

which completes the proof. ∎

The proof of Theorem 2.2 is now an immediate consequence of Lemmas 2.4, 2.5 and 2.6.

3. Sufficient condition for almost uniform convergence

We shall now prove a theorem on almost uniform convergence in H and its symmetric version.

THEOREM 3.1. *Let $\{\xi_n\}$ be a sequence of elements from H_{sa} such that $\sum_{n=1}^{\infty}\|\xi_n\|^2<\infty$. Then $\xi_n\to 0$ almost uniformly.*

PROOF. We choose operators x_{nk} in M_{sa} such that $\xi_n=\sum_{k=1}^{\infty}x_{nk}\xi_0$ for $n=1,2,\dots$ (the series is convergent in norm in H) and $\|x_{nk}\xi_0\|\leqq 2^{-k+1}\|\xi_n\|$ for all k and n.

Let us take positive numbers δ_n such that $\delta_n\to 0$ and $\sum_{n=1}^{\infty}\delta_n^{-1}\|\xi_n\|^2<\infty$. Put $\varepsilon_{nk}=2^{-k+1}\delta_n$. Then

$$\sum_{n,k=1}^{\infty}\varepsilon_{nk}^{-1}\varrho(x_{nk}^2)=\sum_{n,k=1}^{\infty}\varepsilon_{nk}^{-1}\|x_{nk}\xi_0\|^2\leqq\sum_{n,k=1}^{\infty}4^{-k+1}2^{k-1}\delta_n^{-1}\|\xi_n\|^2=$$

$$=2\sum_{n=1}^{\infty}\delta_n^{-1}\|\xi_n\|^2<\infty.$$

Given $\varepsilon>0$, take N such that $\sum_{n=N}^{\infty}\sum_{k=1}^{\infty}\varepsilon_{nk}^{-1}\varrho(x_{nk}^2)<\varepsilon/2$. From Theorem 2.1 it follows that there is a projection e in M with $\varrho(e)\geqq 1-2\sum_{n=N}^{\infty}\sum_{k=1}^{\infty}\varepsilon_{nk}^{-1}\varrho(x_{nk}^2)>1-\varepsilon$ and $\|ex_{nk}^2e\|\leqq 2\varepsilon_{nk}$ for $n\geqq N$, $k=1,2,\dots$. Hence $\|ex_{nk}\|\leqq\sqrt{2}\,\varepsilon_{nk}^{1/2}$ for $n\geqq N$, $k=1,2,\dots$ and, for $n\geqq N$,

$$\sum_{k=1}^{\infty}\|ex_{nk}\|\leqq\sqrt{2}\sum_{k=1}^{\infty}\varepsilon_{nk}^{1/2}=\sqrt{2}\sum_{k=1}^{\infty}(2^{-k+1}\delta_n)^{1/2}=2/(\sqrt{2}-1)\delta_n^{1/2}.$$

Thus, for $n \geqq N$, the series $\sum_{k=1}^{\infty} ex_{nk}$ is convergent in norm to some operator x_n in M and, consequently, $\sum_{k=1}^{\infty} ex_{nk}\xi_0$ is convergent in norm in H to $x_n\xi_0$.

On the other hand, $e\xi_n = e\sum_{k=1}^{\infty} x_{nk}\xi_0$; therefore $e\xi_n = x_n\xi_0$. Moreover, $\|x_n\| \leqq \leqq \sum_{k=1}^{\infty} \|ex_{nk}\| \to 0$ as $n \to \infty$, which finishes the proof. ∎

THEOREM 3.2. *Let $\{\xi_n\}$ be a sequence of elements from H such that $\sum_{n=1}^{\infty} \|\xi_n\|^2 < \infty$. Furthermore, let us assume that M is $\natural$-bounded. Then $\xi_n \to 0$ $\natural$-almost uniformly.*

PROOF. Put $\eta_n = (\xi_n + J\xi_n)/2$, $\zeta_n = (\xi_n - J\xi_n)/2i$. Then $\sum_{n=1}^{\infty} \|\eta_n\|^2 < \infty$ and $\sum_{n=1}^{\infty} \|\zeta_n\|^2 < \infty$. By virtue of the equalities $J\eta_n = \eta_n$, $J\zeta_n = \zeta_n$, there exist decompositions $\eta_n = \eta_n^+ - \eta_n^-$, $\zeta_n = \zeta_n^+ - \zeta_n^-$; $\eta_n^+, \eta_n^-, \zeta_n^+, \zeta_n^- \in \mathcal{P}^\natural$, $\eta_n^+ \perp \eta_n^-$, $\zeta_n^+ \perp \zeta_n^-$ for $n = 1, 2, \ldots$. Hence

$$\sum_{n=1}^{\infty} \|\eta_n^+\|^2 + \sum_{n=1}^{\infty} \|\eta_n^-\|^2 + \sum_{n=1}^{\infty} \|\zeta_n^+\|^2 + \sum_{n=1}^{\infty} \|\zeta_n^-\|^2 < \infty.$$

On account of Lemma 1.2, the $\natural$-almost uniform convergence of the sequence $\eta_1^+, \eta_1^-, \zeta_1^+, \zeta_1^-, \eta_2^+, \eta_2^-, \zeta_2^+, \zeta_2^-, \ldots$ yields the $\natural$-almost uniform convergence of the sequence $\{\xi_n\}$. Therefore, we can assume, with no loss of generality, that $\xi_n \in \mathcal{P}^\natural$ for all $n \in N$.

Now, the proof follows the lines of that of Theorem 3.1. We take x_{nk} in M_+ such that $\xi_n = \sum_{k=1}^{\infty} \Delta^{1/4} x_{nk}\xi_0$ and $\|x_{nk}\|_2^2 \leqq 2^{-k+1}\|\xi_n\|$ for all k and n. δ_n and ε_{nk} are defined as before. Given $\varepsilon > 0$, there is an N such that $\sum_{n=N}^{\infty} \sum_{k=1}^{\infty} \varepsilon_{nk}^{-1} \|x_{nk}\|_2^2 < \varepsilon$.

From Theorem 2.2 it follows that there are a positive constant c and a projection e in M, such that

$$\varrho(e) \geqq \|e\|_2^2 \geqq 1 - \sum_{n=N}^{\infty} \sum_{k=1}^{\infty} \varepsilon_{nk}^{-1} \|x_{nk}\|_2^2 > 1 - \varepsilon$$

and

$$\|\omega_{\Delta^{1/4} e x_{nk}\xi_0}\|_\infty = \|\varphi_{\tilde{e}\tilde{x}_{nk}\xi_0}\|_\infty = \|\tilde{e}\tilde{x}_{nk}\xi_0\|_\infty \leqq c\varepsilon_{nk}$$

for $n \geqq N$, $k = 1, 2, \ldots$. For x in M_+, we have

$$(\sigma_{-i/4}(e)\omega_{\xi_n}\sigma_{i/4}(e))(x) = \omega_{\sigma_{-i/4}(e)\xi_n}(x) = \omega_{\sum_{k=1}^{\infty} \sigma_{-i/4}(e)\Delta^{1/4}x_{nk}\xi_0}(x) = \omega_{\sum_{k=1}^{\infty} \Delta^{1/4}ex_{nk}\xi_0}(x) \leqq$$

$$\leqq \Big(\sum_{k=1}^{\infty} [\omega_{\Delta^{1/4}ex_{nk}\xi_0}(x)]^{1/2}\Big)^2.$$

But, according to the inequality $\varphi(x) \leqq \|\varphi\|_\infty \varrho(x)$ for $\varphi \in (M_*)_+$, $x \in M_+$, and the preceding estimation, the last expression can be estimated as follows:

$$\Big(\sum_{k=1}^{\infty} [\omega_{\Delta^{1/4} e x_{nk} \xi_0}(x)]^{1/2}\Big)^2 \leqq \Big(\sum_{k=1}^{\infty} [\|\omega_{\Delta^{1/4} e x_{nk} \xi_0}\|_\infty \varrho(x)]^{1/2}\Big)^2 =$$

$$= \Big(\sum_{k=1}^{\infty} [\|\tilde{e}\tilde{x}_{nk}\xi_0\|_\infty \varrho(x)]^{1/2}\Big)^2 \leqq c\varrho(x)\Big(\sum_{k=1}^{\infty} \varepsilon_{nk}^{1/2}\Big)^2 = c\varrho(x)\frac{2}{(\sqrt{2}-1)^2}\delta_n.$$

Hence $\|\sigma_{-i/4}(e)\omega_{\xi_n}\sigma_{i/4}(e)\|_\infty \leqq [2/(\sqrt{2}-1)^2]c\delta_n \to 0$ as $n \to \infty$, which proves our assertion. ∎

4. The Menshov inequality

Now, we shall prove two versions of the Menshov inequality, slightly modified, so as to apply it to "nearly" orthogonal operators.

THEOREM 4.1. *Let* $x_1, \dots, x_{2^n}$ *belong to* M_{sa}. *Assume that* $|\varrho(x_j x_i)| \leqq c$ *for some non-negative constant c,* $1 \leqq i, j \leqq 2^n$ *and put* $s_k = \sum_{i=1}^{k} x_i$. *Then there exists an operator t in* M_+ *such that*

(i) $s_k^2 \leqq t$ *for* $k = 1, \dots, 2^n$;

(ii) $\varrho(t) \leqq (n+1)^2 2^{2n} c$.

PROOF. Let $x_{l,m} = \sum_{i=l+1}^{m} x_i$ for $0 \leqq l < m \leqq 2^n$. From the inequality $(z_1 + \dots + z_h)^2 \leqq h(z_1^2 + \dots + z_h^2)$ valid for any positive integer h and $z_1, \dots, z_h$ from M_{sa} we obtain that

$$s_k^2 \leqq (n+1)^2[x_{0,2^n}^2 + (x_{0,2^{n-1}}^2 + x_{2^{n-1},2^n}^2) + (x_{0,2^{n-2}}^2 + x_{2^{n-2},2\cdot 2^{n-2}}^2 + x^2_{2\cdot 2^{n-2}3\cdot 2^{n-2}} + x^2_{3\cdot 2,2^n}) + \dots$$

$$\dots + (x_{0,1}^2 + x_{1,2}^2 + \dots + x_{2^n-1,2^n})]$$

for $k = 1, \dots, 2^n$. Let t be the operator occurring on the right-hand side of the above inequality. Since

$$\varrho(x_{0,2^n}^2) \leqq \sum_{i,j=1}^{2^n} |\varrho(x_j x_i)| \leqq 2^{2n} c,$$

$$\varrho(x_{0,2^{n-1}}^2 + x_{2^{n-1},2^n}^2) \leqq \sum_{i,j=1}^{2^n} |\varrho(x_j x_i)| \leqq 2^{2n} c,$$

....

$$\varrho(x_{0,1}^2 + x_{1,2}^2 + \dots + x_{2^n-1,2^n}^2) \leqq \sum_{i,j=1}^{2^n} |\varrho(x_j x_i)| \leqq 2^{2n} c,$$

therefore $\varrho(t) \leqq (n+1)^2 2^{2n} c$, which completes the proof. ∎

THEOREM 4.2. *Let* $x_1, \ldots, x_{2^n}$ *belong to* M. *Assume that* $|(x_i, x_j)_2| \leqq c$ *for some non-negative constant* c, $1 \leqq i, j \leqq 2^n$, *and put* $s_k = \sum_{i=1}^{k} x_i$. *Then there exists a vector* η *in* $\mathscr{P}^\natural$ *such that*

(i) $$\tilde{s}_k \xi_0 \leqq \eta \quad \text{for} \quad k = 1, \ldots, 2^n;$$

(ii) $$(\eta, \xi_0) \leqq (n+1)^2 2^{2n} c.$$

PROOF. Let us define the operators $x_{l,m}$ as in Theorem 4.1. For x, y in M, we have $xJyJ\xi_0 + yJxJ\xi_0 \leqq \tilde{x}\xi_0 + \tilde{y}\xi_0$ because $\tilde{x}\xi_0 + \tilde{y}\xi_0 - (JxJy\xi_0 + JyJx\xi_0) = = J(x-y)J(x-y) \in \mathscr{P}^\natural$, and as a consequence, for an arbitrary positive integer h and $z_1, \ldots, z_h$ from M, the inequality $(z_1 + \ldots + z_n)^\sim \xi_0 \leqq h(\tilde{z}_1 + \ldots + \tilde{z}_h)\xi_0$ holds. From this inequality it follows that

$$\tilde{s}_k \xi_0 \leqq (n+1)[\tilde{x}_{0,2^n} + (\tilde{x}_{0,2^{n-1}} + \tilde{x}_{2^{n-1},2^n}) + (\tilde{x}_{0,2^{n-2}} + \tilde{x}_{2^{n-2},2^{n-1}} + \tilde{x}_{2^{n-1},3\cdot 2^{n-2}} + \tilde{x}_{3\cdot 2^{n-2},2^n}) + \\ + \ldots + (\tilde{x}_{0,1} + \tilde{x}_{1,2} + \ldots + \tilde{x}_{2^n-1,2^n})]\xi_0$$

for $k = 1, \ldots, 2^n$. Let η be the vector occurring on the right-hand side of the above inequality. We have the following estimations:

$$(\tilde{x}_{0,2^n}\xi_0, \xi_0) = ((\sum_{i=1}^{2^n} x_i)^\sim \xi_0, \xi_0) = \sum_{i,j=1}^{2^n} (x_i J x_j J \xi_0, \xi_0) = \sum_{i,j=1}^{2^n} (x_i, x_j)_2 \leqq 2^{2n} c,$$

....

$$((\tilde{x}_{0,1} + \tilde{x}_{1,2} + \ldots + \tilde{x}_{2^n-1,2^n})\xi_0, \xi_0) \leqq 2^{2n} c$$

and the assertion follows. ∎

5. The Rademacher—Menshov theorem

This section is devoted to the non-commutative Rademacher—Menshov theorem in the "left" and "symmetric" setups. We begin with a classical lemma which can be found, for instance, in [8], Th. 2.2.1, p. 16.

LEMMA 5.1. *Let* $\{\xi_n\}$ *be a sequence of pairwise orthogonal vectors from some Hilbert space, such that* $\sum_{n=1}^{\infty} (\log^2 n) \|\xi_n\|^2 < \infty$. *Then* $\sum_{n=1}^{\infty} \sum_{k=2^n+1}^{\infty} \|\xi_k\|^2 < \infty$.

THEOREM 5.2. *Let* $\{\xi_n\}$ *be a sequence of pairwise orthogonal vectors from* H_{sa}. *If* $\sum_{n=1}^{\infty} (\log^2 n) \|\xi_n\|^2 < \infty$, *then* $s_n = \sum_{i=1}^{\infty} \xi_i$ *converges almost uniformly.*

PROOF. Let us choose x_n in M_{sa} such that $\|x_n \xi_0 - \xi_n\| < 2^{-n}$ and define

$$r_n = \sum_{i=n+1}^{\infty} \xi_i, \quad q_n = \sum_{i=n+1}^{\infty} x_i \xi_0, \quad p_n = \sum_{i=1}^{n} x_i.$$

For sufficiently large k and i, j such that $2^k < i, j \leqq 2^{k+1}$, we have

$$|\varrho(x_j x_i)| \leqq |(x_i\xi_0, x_j\xi_0-\xi_j)|+|(x_i\xi_0-\xi_i, \xi_j)|+|(\xi_i, \xi_j)| \leqq$$
$$\leqq \|x_i\xi_0\|\,\|x_j\xi_0-\xi_j\|+\|x_i\xi_0-\xi_i\|\,\|\xi_j\| \leqq$$
$$\leqq \|x_i\xi_0-\xi_i\|+\|x_j\xi_0-\xi_j\| < 2^{-i}+2^{-j} < 2^{1-2^k} \quad \text{for} \quad i \neq j,$$
$$\varrho(x_i^2) \leqq 2^{-i}+\|\xi_i\|^2.$$

By virtue of Theorem 4.1, there exists a sequence $\{t_k\}$ of operators from M_+ such that

(i) $$(p_n-p_{2^k})^2 \leqq t_k \quad \text{for} \quad 2^k < n \leqq 2^{k+1}, \quad k = 1, 2, \ldots;$$

(ii) $$\sum_{k=1}^{\infty} \varrho(t_k) < \infty.$$

Assertion (ii) follows from the convergence of the series

$$\sum_{k=1}^{\infty}(k+1)^2\Big[2^{2k}2^{1-2^k}+\sum_{i=2^k+1}^{2^{k+1}} 2^{-i}\Big] \quad \text{and} \quad \sum_{k=1}^{\infty}(k+1)^2 \sum_{i=2^k+1}^{2^{k+1}} \|\xi_i\|^2.$$

We have $\sum_{n=1}^{\infty} \|r_n-q_n\|^2 < \infty$ and from Lemma 5.1 it follows that $\sum_{n=1}^{\infty} \|r_{2^n}\|^2 < \infty$ and, accordingly, $\sum_{n=1}^{\infty} \|q_{2^n}\|^2 < \infty$.

Now, we form a sequence $\{\zeta_n\}$ having the following terms:

$$r_1-q_1,\; r_2-q_2,\; t_1^{1/2}\xi_0,\; q_{2^1},\; r_3-q_3,\; r_4-q_4,\; t_2^{1/2}\xi_0,\; q_{2^2},$$
$$r_5-q_5,\; r_6-q_6,\; r_7-q_7,\; r_8-q_8,\; t_3^{1/2}\xi_0,\; q_{2^3}, \ldots.$$

From what has been said before it follows that $\sum_{n=1}^{\infty} \|\zeta_n\|^2 < \infty$; hence $\zeta_n \to 0$ almost uniformly. Given $\varepsilon > 0$, there is a projection e in M with $\varrho(e) > 1-\varepsilon$ and $e\zeta_n = y_n\xi_0$ for some y_n in M, $\|y_n\| \to 0$; in particular, $\|et_k^{1/2}\| \to 0$. For sufficiently large n, we can write

$$er_n = e(r_n-q_n)+eq_{2^k}+e(p_{2^k}-p_n)\xi_0$$

where $2^k < n \leqq 2^{k+1}$. Since all the three summands on the right-hand side belong to $M\xi_0$, therefore er_n belongs to $M\xi_0$ for sufficiently large n. Let us observe that

$$\|e(p_{2^k}-p_n)\|^2 = \|e(p_{2^k}-p_n)^2 e\| \leqq \|et_k e\| = \|et_k^{1/2}\|^2.$$

Thus the convergence $\zeta_n \to 0$ almost uniformly yields the convergence $r_n \to 0$ almost uniformly, which finishes the proof. ▌

THEOREM 5.3. *Let M be a $\natural$-bounded von Neumann algebra and $\{\xi_n\}$ a sequence of pairwise orthogonal vectors from H. If $\sum_{n=1}^{\infty} (\log^2 n)\|\xi_n\|^2 < \infty$, then $s_n = \sum_{i=1}^{\infty} \xi_i$ converges $\natural$-almost uniformly.*

PROOF. Let us choose x_n in M such that $\|\Delta^{1/4}x_n\xi_0-\xi_n\|<2^{-n}$ (which is possible because $H=(\mathscr{P}^\natural-\mathscr{P}^\natural)+i(\mathscr{P}^\natural-\mathscr{P}^\natural)$ and $\mathscr{P}^\natural=\Delta^{1/4}M_+\xi_0^-$) and define

$$r_n=\sum_{i=n+1}^{\infty}\xi_i,\qquad q_n=\sum_{i=n+1}^{\infty}\Delta^{1/4}x_n\xi_0,$$

$$p_n=\sum_{i=1}^{n}x_i.$$

Similarly as in the proof of Theorem 5.2, the estimation of $|(x_i,x_j)_2|$ and Theorem 4.2 allow us to form a sequence $\{\eta_k\}$ of vectors from $\mathscr{P}^\natural$ such that

(i) $\qquad (p_n-p_{2^k})^\sim\xi_0\leqq\eta_k$ for $2^k<n\leqq 2^{k+1}$, $k=1,2,\dots;$

(ii) $$\sum_{k=1}^{\infty}(\eta_k,\xi_0)<\infty.$$

Again, we build the sequence $\{\zeta_n\}$ as in Theorem 5.2, the only difference being that instead of $t_k^{1/2}\xi_0$ we take vectors Θ_k from $\mathscr{P}^\natural$ such that $\varphi_{\eta_k}=\omega_{\Theta_k}$. We have $\sum_{n=1}^{\infty}\|\zeta_n\|^2<\infty$ because $\|\Theta_k\|^2=\omega_{\Theta_k}(1)=\varphi_{\eta_k}(1)=(\xi_0,\eta_k)$ and, consequently, $\zeta_n\to 0$ $\natural$-almost uniformly. We can write

$$\sigma_{-i/4}(e)r_n=\sigma_{-i/4}(e)(r_n-q_n)+\sigma_{-i/4}(e)q_{2^k}(e)+\sigma_{-i/4}(e)\Delta^{1/4}(p_{2^k}-p_n)\xi_0$$

where e is the projection, chosen for given $\varepsilon>0$, as in the definition of $\natural$-almost uniform convergence of the sequence $\{\zeta_n\}$. Since the terms r_n-q_n and q_{2^k} occur in the sequence $\{\zeta_n\}$, therefore

$$\|\omega_{\sigma_{-i/4}(e)(r_n-q_n)}\|_\infty\to 0\quad\text{and}\quad\|\omega_{\sigma_{-i/4}(e)q_{2^k}}\|_\infty\to 0.$$

On account of Lemma 1.2, it is enough to show that

$$\|\omega_{\sigma_{-i/4}(e)\Delta^{1/4}(p_{2^k}-p_n)\xi_0}\|_\infty\to 0.$$

We have

$$\|\omega_{\sigma_{-i/4}(e)\Delta^{1/4}(p_{2^k}-p_n)\xi_0}\|_\infty=\|\omega_{\Delta^{1/4}e(p_{2^k}-p_n)\xi_0}\|_\infty=\|\varphi_{\tilde e(p_{2^k}-p_n)^\sim\xi_0}\|_\infty\leqq\|\varphi_{\tilde e\eta_k}\|_\infty.$$

Furthermore,

$$\varphi_{\tilde e\eta_k}(y)=(\Delta^{1/4}y\xi_0,\tilde e\eta_k)=(eJeJ\sigma_{-i/4}(y)\xi_0,\eta_k)=$$
$$=(e\Delta^{1/4}y\Delta^{-1/4}Je\xi_0,\eta_k)=(e\Delta^{1/4}y\Delta^{1/4}e\xi_0,\eta_k)=$$
$$=(\Delta^{1/4}\sigma_{i/4}(e)y\sigma_{-i/4}(e)\xi_0,\eta_k)=\big(\sigma_{-i/4}(e)\varphi_{\eta_k}\sigma_{i/4}(e)\big)(y)=\big(\sigma_{-i/4}(e)\omega_{\Theta_k}\sigma_{i/4}(e)\big)(y)$$

for each $y\in M$; hence

$$\|\varphi_{\tilde e\eta_k}\|_\infty=\|\sigma_{-i/4}(e)\omega_{\Theta_k}\sigma_{i/4}(e)\|_\infty\to 0\quad\text{as}\quad k\to\infty$$

because the Θ_k's occur in the sequence $\{\zeta_n\}$. We have thus obtained

$$\|\sigma_{-i/4}(e)\omega_{r_n}\sigma_{i/4}(e)\|_\infty\to 0\quad\text{as}\quad n\to 0,$$

which proves our theorem. ∎

REFERENCES

[1] Batty, C. J. K., The strong law of large numbers for states and traces of a W^*-algebra, *Z. Wahrsch. Verw. Gebiete* **48** (1979), 177—191. *MR* **80k:** 46073.

[2] Connes, A., Caractérisation des espaces vectoriels ordonnés sous-jacents aux algèbres de von Neumann, *Ann. Inst. Fourier (Grenoble)* **24** (1974), 121—155. *MR* **51** #13705.

[3] Gol'dštein, M. Š., Almost sure convergence theorems in von Neumann algebras (in Russian), *J. Operator Theory* **6** (1981), 233—311. *MR* **84g:** 46096.

[4] Kosaki, H., Applications of the complex interpolation method to a von Neumann algebra (Non-commutative L^p-spaces), preprint.

[5] Kosaki, H., A Radon—Nikodym theorem for natural cones associated with von Neumann algebras, *Proc. Amer. Math. Soc.* **84** (1982), 207—211. *MR* **83a:** 46076.

[6] Lance, E. C., Ergodic theorems for convex sets and operator algebras, *Invent. Math.* **37** (1976), 201—214. *MR* **55** #1089.

[7] Lance, E. C., Martingale convergence in von Neumann algebras, *Math. Proc. Cambridge Philos. Soc.* **84** (1978), 47—56. *MR* **80d:** 46115.

[8] Stout, W., *Almost Sure Convergence*, Academic Press, New York, San Francisco, London, 1974. *MR* **56** #13334.

[9] Stratila, S. and Zsidó, L., *Lectures on von Neumann Algebras*, Abacus Press, Tunbridge Wells, Kent, England, 1979. *MR* **81j:** 46089.

(Received October 9, 1984)

INSTITUTE OF MATHEMATICS
ŁÓDŹ UNIVERSITY
BANACHA 22
PL—90-238 ŁÓDŹ
POLAND

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 379—381

CATEGORICITY OF SIMPLICIAL COMPLICES

R. GUREVIČ

One can regard simplicial complices as models of the similarity type consisting of the relations $\{P_i\}_{i\geqq 1}$ with the intended meaning $P_i(x_0, \ldots, x_i)=$ "$x_0, \ldots, x_i$ are the vertices of an i-simplex". Let L be the corresponding first order language. Topologists often subdivide the simplicial complices which represent their polyhedra. Note that a simplicial complex is locally finite if and only if its space is locally compact (see any reference book on algebraic topology).

From [1] it is clear that there are many natural geometric properties which are not expressible in L even for finite connected simplicial complices. In many instances, however, the elementary equivalence of infinite simplicial complices implies their isomorphism.

PROPOSITION 1. *Let A be a countable locally finite simplicial complex and let each infinite connected component of A have* $\dim>1$. *Then there is a subdivision B of A such that for any locally finite C with* $\dim>1$ *for each infinite connected component of C and having at most countably many isolated points $C\equiv_L B$ implies that C is isomorphic to B.*

PROPOSITION 2. *Every connected locally finite simplicial complex A possesses a subdivision B such that for every connected C if $C\equiv_L B$ then C is isomorphic to B.*

Let A be a simplicial complex. A *tail* is an infinite sequence $\{x_i\}_{i\geqq 0}$ of vertices such that $P_1(x_i, y)$ iff $y=x_{i-1}$ or $y=x_{i+1}$ for every y and $i\geqq 1$. The corresponding notion for a polyhedron X is a proper embedding $[0, 1)\to X$ which is open on $(0, 1)$.

PROPOSITION 3. *Let A be a countable locally finite simplicial complex without tails. Then there is a subdivision B of A such that for any locally finite C without tails and with at most countably many isolated points if $C\equiv_L B$ then C is isomorphic to B.*

Consequently, each locally compact separable polyhedron possesses a "categorically" triangulable strong deformational retract.

REMARK. The condition on local finiteness cannot be dropped.

1980 *Mathematics Subject Classification.* Primary 03C35, 03C20; Secondary 57Q05.

Key words and phrases. Categoricity, locally compact polyhedrum, locally finite simplicial complex, subdivision of simplicial complex, ultrapower, Shelah isomorphic ultrapowers theorem.

Lemmata

A monomorphism $M \to N$ is called an *elementary embedding* if the image of M in N is an elementary submodel of N.

LEMMA 1. *Let $Q \subseteq R$ be an elementary embedding of simplicial complices and assume that Q is locally finite. Then either Q is a connected component of R or it is the union of some connected components.*

PROOF. Straightforward.

LEMMA 2. *Every simplicial complex A has a subdivision B such that every edge of A is divided, and for every n*
(1) *the number of edges of A divided into $\leqq n$ parts is finite,*
(2) *the number of* $\dim > 1$ *B-simplices with all vertices incident to $\leqq n$ simplices is finite.*

PROOF. Subdivide the edges of A to satisfy (1). Let $k > 1$; assume that the subdivision of the $k-1$-simplices is done. For every k-simplex s of A take its baricenter b_s and join it with the subdivided ∂s.

LEMMA 3. *If A is locally finite the subdivision in Lemma 2 can be chosen to satisfy the additional*
(3) *all bridges in B have different lengths (a bridge is a finite sequence $x_0, \dots, x_m (m > 1)$ of vertices maximal with respect to the property that $P_1(x_i, y)$ holds iff $y = x_{i-1}$ or $y = x_{i+1}$ for every vertex y and $1 \leqq i \leqq m-1$; m is the length of the bridge),*
(4) *let $u(x) = \text{card}\,\{y\colon P_1(x, y)\}$, let us call x a maximal vertex if for every y $P_1(x, y)$ implies $u(x) > u(y)$; then every* $\dim > 1$ *connected component of B contains exactly one vertex v such that v is maximal, v is incident to a* $\dim > 1$ *simplex and $u(v)$ is prime; and the primes corresponding to those v's are different for different* $\dim > 1$ *connected components of B.*

PROOF. Subdivide first the edges which are not contained in $\dim > 1$ simplices of A to satisfy (1) and (3). Notice that for $\dim(s) > 1$

$$u(b_s) = \sum \{\text{the number of new } B\text{-vertices on } A\text{-edges of } s\} +$$
$$+ 2^{\dim(s)+1} - 2 - \binom{\dim(s)+1}{2}.$$

Fix an enumeration of the $\dim > 1$ connected components of A; let C be the first not treated yet. Let $k > 1$ and $s = (x_0, \dots, x_k)$ be a maximal simplex in C, i.e. for no y $P_{k+1}(y, x_0, \dots, x_k)$ holds. In view of the above formula for $u(b_s)$ divide the edges of s into sufficiently many pieces to ensure the maximality of b_s in B and to make $u(b_s)$ a prime greater than the other primes produced so far. Fix an enumeration of the remaining edges of all $\dim > 1$ simplices of C and proceed as follows. Take the first edge not divided yet, (x, y), let $s_1, \dots$ be all the maximal C-simplices containing (x, y). Divide (x, y) into sufficiently many pieces to ensure the maximality of $\{b_{s_i}\}_i$ having chosen the number of pieces so that if all but (x, y) edges of s_i were already subdivided then $u(b_{s_i})$ will be composite. To ensure (1) and (2) it suffices to claim that the number of pieces into which (x, y) is divided must be greater than those numbers for the edges of C already divided and greater than the prime chosen for C.

Proofs of the propositions

The isomorphic ultrapowers theorem of Shelah [3] states that if M, N are elementary equivalent models then there exists a set S with an ultrafilter F on it such that the ultrapowers $\Pi_F M$ and $\Pi_F N$ are isomorphic. All three propositions are easy consequences of Shelah's theorem, their proofs almost coincide. Below I shall use also the fact that the model M is naturally ("diagonally") embedded into its ultrapower $\Pi_F M$.

Let A be a locally finite simplicial complex and let B be a subdivision as given in Lemma 3. Let $C \equiv_L B$ and $\Pi_F C$ be isomorphic to $\Pi_F B$. Assume that F is a non-principal ultrafilter. Then by Lemma 1 the image $\operatorname{im}(B)$ of B in $\Pi_F B$ is a union of some connected components of $\Pi_F B$ and by (1)—(3) $\operatorname{im} B$ contains all the locally finite components with $\dim > 1$, all finite components with $\dim = 1$, and all bridges of $\Pi_F B$. ($\operatorname{im}(B)$, of course, can be larger.) Propositions 1, 2, 3 now give that $\operatorname{im}(C)$ is just $\operatorname{im}(B)$ (except in a degenerate case in Proposition 2).

Proof of Proposition 1. In this case $\operatorname{im}(B) - \{\text{the isolated points}\} =$ the union of all locally finite $\dim > 1$ components plus all finite $\dim = 1$ components of $\Pi_F B$. Consequently, by Lemma 1 (and the condition of the proposition) we have $\operatorname{im}(C) - \{\text{the isolated points}\} =$ a union of some connected components of $\operatorname{im}(B)$. On the other hand, by (3), (4) we have $\operatorname{im}(C) \supset \operatorname{im}(B) - \{\text{the isolated points}\}$. (Note that every finite simplicial simplex with $\dim = 1$ contains bridges.) The number of isolated points $\operatorname{im}(B)$ and $\operatorname{im}(C)$ must clearly coincide.

Proof of Proposition 2. Consider first the case $\dim(A) > 1$. Then also $\dim(C) > 1$ and $\operatorname{im}(B)$ is the unique $\dim > 1$ locally finite component of $\Pi_F B$. Consequently, $\operatorname{im}(C) = \operatorname{im}(B)$. Now let $\dim(A) = 1$. If B contains a bridge then by (3) C must contain a bridge of appropriate length whence $\operatorname{im}(B) \cap \operatorname{im}(C) \neq \emptyset$ and by Lemma 1 then $\operatorname{im}(C) = \operatorname{im}(B)$. If B contains no bridge then A is a bouquet of tails, i.e. is a union of $n \geqq 1$ different tails $\{x_{ij}\}_{i \geqq 0}$, $1 \leqq j \leqq n$, such that the x_{0j}'s coincide. For C connected one can easily express "C is a bouquet of n tails" by a countable conjunction of L-sentences.

Proof of Proposition 3. In this case $\operatorname{im}(C) - \{\text{the isolated points}\}$ is the union of some components of $\operatorname{im}(B)$ and by (3), (4) we have $\operatorname{im}(C) - \{\text{the isolated points}\} = = \operatorname{im}(B) - \{\text{the isolated points}\}$, and of course B, C must have the same number of isolated points.

REFERENCES

[1] Gaifman, H., On local and non-local properties, *Proc. of the Herbrand Symp.* (Marseille 1981), 105—135, North-Holland, 1982. *MR* **85k:** 03020.
[2] Gurevič, R., A categoricity of simplicial complices, *Abstracts AMS*, 1984.
[3] Shelah, S., Every two elementary equivalent models have isomorphic ultrapowers, *Israel J. Math.* **10** (1971), 224—233. *MR* **45** #6608.
[4] Spanier, E., *Algebraic topology*, McGraw-Hill, 1966. *MR* **35** #1007.

(Received October 26, 1984)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
273 ALTGELD HALL
1409 WEST GREEN STREET
URBANA, IL 61801
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 383—393

A 3-PART SPERNER THEOREM

P. L. ERDŐS and G. O. H. KATONA

1. Introduction

Let X be a finite set of n elements and let $\mathcal{F} \subset 2^X$ be a family of different subsets of X such that every pair of members F_1, F_2 of $\mathcal{F}$ $(F_1 \neq F_2)$ satisfies $F_1 \not\subset F_2$. Sperner [6] proved that in this case

$$|F| \leq \binom{n}{\left\lfloor \frac{n}{2} \right\rfloor}. \tag{1}$$

If the inequality is realized by equality then either

$$\mathcal{F} = \left\{F \subset X \colon |F| = \left\lfloor \frac{n}{2} \right\rfloor\right\} \quad \text{or} \tag{2}$$

$$\mathcal{F} = \left\{F \subset X \colon |F| = \left\lceil \frac{n}{2} \right\rceil\right\} \quad \text{if} \quad n \equiv 1 \pmod 2.$$

Kleitman [5] and Katona [4] independently proved: If X is divided into two disjoint parts $(X = X_1 \cup X_2,\ X_1 \cap X_2 = \emptyset)$ and the family $\mathcal{F}$ contains no two different members F_1, F_2 such that:

$$F_1 \subset F_2 \quad \text{and} \quad F_2 \setminus F_1 \subset X_i \qquad (i = 1 \text{ or } 2) \tag{3}$$

then (1) holds.

By an analogous way, the more-part Sperner problem can be defined. Let X be a finite set of n elements and $X = X_1 \cup \ldots \cup X_M$ where $X_i \cap X_j = \emptyset$ $(i \neq j)$. The set $\mathcal{F} \subset 2^X$ of subsets of X is an M-part Sperner family, if no two members of $\mathcal{F}$ satisfy:

$$F_1 \subset F_2 \quad \text{and} \quad F_2 \setminus F_1 \subset X_i \quad \text{for some} \quad i \in \{1, \ldots, M\}. \tag{4}$$

As it is shown in [4], if $M \geqq 3$, then inequality (1) is not true for every M-part Sperner family $\mathcal{F}$. Füredi [2], Griggs, Odlyzko and Shearer [3] found good asymptotic results for the maximum size of M-part Sperner families. But the exact value is not known even for $M = 3$.

The aim of this paper is to determine this exact maximum size for the very modest case $M = 3$ and $|X_3| = 1$. Exactly, we prove

---

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1812.

1980 *Mathematics Subject Classification*. Primary 05A05; Secondary 05C65.
Key words and phrases. Sperner theorem.

THEOREM 1. *Let* $X=X_1\cup X_2\cup X_3$ *be a partition, where* $|X_1|=n_1\leqq|X_2|=n_2$; $|X_3|=1$; $n_1+n_2+1=n$. *Then*

$$\max\{|\mathcal{F}|:\ \mathcal{F}\text{ is a 3-part Sperner family}\}=$$

$$(5)\qquad =\begin{cases}\dbinom{n}{\left\lfloor\frac{n}{2}\right\rfloor} & \text{if } n_1\not\equiv n_2 \pmod 2,\\[2ex] 2\dbinom{n-1}{\frac{n-1}{2}} & \text{if } n_1\equiv n_2\equiv 1 \pmod 2,\\[2ex] 2\dbinom{n-1}{\frac{n-1}{2}}-\left(\dbinom{n_1}{\frac{n_1+2}{2}}-\dbinom{n_1}{\frac{n_1}{2}}\right)\left(\dbinom{n_2}{\frac{n_2+2}{2}}-\dbinom{n_2}{\frac{n_2}{2}}\right) & \text{if } n_1\equiv n_2\equiv 0 \pmod 2.\end{cases}$$

In the proof, the next theorem of Griggs, Odlyzko and Shearer [3] has a fundamental role. (A new proof of this theorem and a generalization of it to the extreme points of the polytop of the M-part Sperner families can be found in [1].)

THEOREM (GOS [3]). *There is an M-part Sperner family* $\mathcal{F}$ *such that*

$$|\mathcal{F}|=\max\{|\mathcal{F}'|:\ \mathcal{F}'\text{ is an } M\text{-part Sperner family}\}$$

and $F\in\mathcal{F}$ *implies that all sets* $G\subset X$ *satisfying*

$$|F\cap X_j|=|G\cap X_j|$$

for all j $(1\leqq j\leqq M)$ *belong to* $\mathcal{F}$.

2. Proof of the main theorem

Let $\mathcal{F}\subset 2^X$ be a family of subsets of the set $X=X_1\cup X_2\cup X_3$. The 3-dimensional matrix $P(\mathcal{F})=(p_{i_1,i_2,i_3}(\mathcal{F}))$ $i_j=0,\dots,n_j$ is called the *profile-matrix* of $\mathcal{F}$, where

$$(6)\qquad p_{i_1,i_2,i_3}(\mathcal{F}):=\left|\{F\in\mathcal{F}:\ \forall j\,|F\cap X_j|=i_j\}\right|.$$

According to the theorem of Griggs, Odlyzko and Shearer there is a maximum sized 3-part Sperner family $\mathcal{F}$ such that

$$p_{i_1,i_2,i_3}(\mathcal{F})=\begin{cases}\dbinom{n_1}{i_1}\dbinom{n_2}{i_2}\dbinom{n_3}{i_3},\\ \text{or}\quad 0.\end{cases}$$

Let $J=\{(i_1,i_2,i_3):\ p_{i_1,i_2,i_3}\neq 0\}$. Then according to the definition of the 3-part Sperner families the set J is a *partial transversal*, that is, if (i_1,i_2,i_3), $(i_1',i_2',i_3')\in\mathcal{J}$ and they are identical in at least two components; then $(i_1,i_2,i_3)=(i_1',i_2',i_3')$.

To the proof of Theorem 1 we need several lemmas. It is easy to see that the projection of a partial transversal (it is an $(n_1+1)\times(n_2+1)\times 2$ matrix) into its

$(n_1+1)\times(n_2+1)$ "face" has at most 2 elements in each row and each column. This justifies the following definition. $I\subset\{1, \dots, u\}\times\{1, \dots, v\}$ is a *partial 2-transversal* iff no column or row contains more than 2 elements of I. Let the values $a_1\geqq\dots\geqq a_u\geqq 1$, $b_1\geqq\dots\geqq b_v\geqq 1$ be fixed. We will consider the matrix $(a_i b_j)_{1\leqq i\leqq u, 1\leqq j\leqq v}$. The partial 2-transversal I will be called *optimal* iff

1) it maximizes

$$\sum_{i,j\in I} a_i b_j \tag{7}$$

among all partial 2-transversals;

2) It minimizes

$$\sum_{(i,j)\in I} (i+j) \tag{8}$$

among all partial 2-transversals satisfying 1) and

3) it maximizes

$$\sum_{(i,j)\in I} i\cdot j \tag{9}$$

among all partial 2-transversals satisfying 1) and 2).

In the proofs of the lemmas the following 3 transformations of partial 2-transversals will be used.

Transformation 1. If I contains at most one element in the i-th row and at most one in the j-th column, add (i,j) to I. This transformation increases (7).

Transformation 2. Move the element $(i,j)\in I$ into (i,k) if $k<j$ and the k-th column of I contains at most one element, or move $(i,j)\in I$ into (l,j) if $l<i$ and the l-th row contains at most one element. This transformation does not decrease (7) but decreases (8).

Transformation 3. Let $i<k$ and $j<l$. Suppose that $(i,l),(k,j)\in I$. The transformation changes I for $I'=(I-\{(i,l),(k,j)\})\cup\{(i,j),(k,l)\}$. It does not decrease (7) because

$$\sum_{(i,j)\in I'} a_i b_j - \sum_{(i,j)\in I} a_i b_j = a_i b_j + a_k b_l - a_i b_l - a_k b_j = (a_i-a_k)(b_j-b_l)\geqq 0.$$

It does not change (8), but it increases (9):

$$\sum_{(i,j)\in I'} i\cdot j - \sum_{(i,j)\in I} i\cdot j = ij+kl-il-kl = (i-k)(j-l)>0.$$

The following lemma is an easy consequence.

LEMMA 2.1. *Transformations 1—3 cannot be applied for an optimal partial 2-transversal.*

In what follows, we will study the structure of the optimal partial 2-transversals.

LEMMA 2.2. *An optimal partial 2-transversal has non-increasing number of elements in the rows (columns).*

PROOF. The number of elements of I in the i-th row (column) is denoted by ϱ_i ($\varkappa_i$) ($0 \leqq \varrho_i \leqq 2$, $0 \leqq \varkappa_i \leqq 2$). Suppose that $i<j$ and $\varrho_i > \varrho_j$. Consider an element $(i, k) \in I$. Transformation 2 with $(i, k) \to (j, k)$ could be applied contradicting Lemma 2.1. ∎

LEMMA 2.3. *Let $u=v$. An optimal partial 2-transversal satisfies* $\varrho_1 = \ldots = \varrho_{u-1} = \varkappa_1 = \ldots = \varkappa_{u-1} = 2$ *and either* $\varrho_u = \varkappa_u = 2$ *or* $\varrho_u = \varkappa_u = 1$.

PROOF. Suppose that I is optimal, consequently it satisfies the conditions of Lemma 2.2.

$\zeta_u = 0$ implies $|I| \leqq 2u-2$. Hence $\varkappa_u < 2$ follows. Transformation 1 could be applied with (u, u). This is a contradiction by Lemma 2.1. $\varrho_u \geqq 1$ is proved. $\varkappa_u \geqq 1$ can be seen in the same way.

Suppose that $\varrho_u = \varkappa_u = 1$. Let $\varrho_{u-1} = 1$. Either $(u, u) \notin I$ or $(u-1, u) \notin I$ holds, so Transformation 1 could be applied with one of them, contradicting the optimality of I. This proves $\varrho_{u-1} = 2$ and $\varkappa_{u-1} = 2$ can be proved analogously.

Suppose now that one of ϱ_u and $\varkappa_u$ equals 2. Then $|I| = 2u$, thus all ϱ's and $\varkappa$'s are equal to 2. ∎

LEMMA 2.4. *Let $u<v$. An optimal partial 2-transversal satisfies* $\varrho_1 = \ldots = \varrho_u = \varkappa_1 = \ldots = \varkappa_{u-1} = 2$, $\varkappa_{u+2} = \ldots = \varkappa_v = 0$ *and either* $\varkappa_u = \varkappa_{u+1} = 1$ *or* $\varkappa_u = 2$, $\varkappa_{u+1} = 0$.

PROOF. Suppose that I is optimal, consequently it satisfies the conditions of Lemma 2.2.

$\varrho_u \leqq 1$ implies $|I| \leqq 2u-1$. Hence $\varkappa_u \leqq 1$ and $\varkappa_{u+1} \leqq 1$ follow. One of (u, u) and $(u, u+1)$ is not in I, thus adding it to I by Transformation 1 it leads to a contradiction. $\varrho_1 = \ldots = \varrho_u = 2$ is proved.

If $\varkappa_u = 2$ then $|I| = \sum_{i=1}^{u} \varrho_i = 2u = \sum_{i=1}^{v} \varkappa_i$ implies $\varkappa_1 = \ldots = \varkappa_u = 2$, $\varkappa_{u+1} = \ldots = \varkappa_v = 0$.

Suppose now that $\varkappa_u = 1$. It implies $\varkappa_{u+1} = 1$. However, $\varkappa_{u-1} = 1$ leads to a contradiction. Indeed, $(i, u+1) \in I$ holds for some i. Transformation 2 can be used with $(i, u+1)$ and with either (i, u) or $(i, u-1)$, because both ones cannot belong to I. Hence we have $\varkappa_{u-1} = 2$. $\varkappa_{u+2} = \ldots = \varkappa_v = 0$ trivially follows. ∎

LEMMA 2.5. *Let $u \leqq v$, $1 \leqq i < u$ and $1 \leqq j < u-1$. Suppose that I is an optimal partial 2-transversal and its subset $A \subset I$ satisfies the following conditions:*

(10) *A contains at most one element in the ith row, at most one in the jth column and at most one in the $(j+1)$st column,*

and

(11) $I-A$ *has no element of the form* (i, k), $k<j$ *or* (l, j), $l<i$ *or* $(m, j+1)$, $m<i$.

Then either $(i, j) \in I$ or $(i, j+1) \in I$ holds. The roles of the rows and columns can be interchanged.

PROOF. We use an indirect way. Suppose that

$$(i,j)\notin I \quad \text{and} \quad (i,j+1)\notin I.$$

Lemmas 2.3, 2.4 and (10) imply the existence of an $(i,k)\in I-A$. $k\neq i, i+1$ by the assumption. On the other hand, (11) results in $j+1<k$. The same arguments show the existence of an $(s,j)\in I-A$ and a $(t,j+1)\in I-A$ where $s,t>i$ can be assumed. We distinguish 2 cases:

(i) $s=t$. The j-th column contains two elements of I: (s,j) and $(s,j+1)$. Therefore we have $(s,k)\notin I$. Transformation 3 can be applied for (i,k) and (s,j). This contradiction proves the statement in this case.

(ii) $s\neq t$. (s,k) and (t,k) cannot be both in I. Suppose e.g. that $(s,k)\notin I$. Transformation 3 can be applied for (i,k) and (s,j), again. This case is also settled. ∎

LEMMA 2.6. *Let* $u,v\geqq 2$. *If I is an optimal partial 2-transversal then* $(1,1), (1,2), (2,1)\in I$.

PROOF. Suppose that $u,v\geqq 4$ and apply Lemma 2.5 with $A=\emptyset$, $i=1$, $j=1$. Either $(1,1)\in I$ or $(1,2)\in I$ can be stated. We distinguish these two cases.

a) $(1,1)\in I$. Apply Lemma 2.5 with $A=\{(1,1)\}$, $i=1, j=2$. Two subcases are distinguished: aa) $(1,2)\in I$, ab) $(1,3)\in I$, $(1,2)\notin I$.

aa) $(1,2)\in I$. Lemma 2.5 can be applied, again, with $A=\{(1,1),(1,2)\}$ $i=2$, $j=1$. If we obtain $(2,1)\in I$ we are done. Suppose that $(2,2)\in I$, $(2,1)\notin I$. By Lemmas 2.3 and 2.4 there exists a $(k,1)\in I$, $k>2$. Transformation 3 can be applied with $(2,2)$ and $(k,1)$ because $(k,2)\notin I$. This contradiction proves the lemma in this case.

ab) $(1,3)\in I$. Apply Lemma 2.5 with $A=\{(1,1),(1,3)\}$, $i=2, j=1$. Two subcases will be distinguished:

aba) $(2,1)\in I$. We may continue: either $(2,2)\in I$ or $(2,3)\in I$. In the first case the change of $(2,2)$ and $(1,3)$ (Transformation 3) leads to the desired contradiction. In the latter case there is a $(k,2)\in I$ $(k>2)$ by Lemmas 2.3 and 2.4. $(k,2)$ and $(1,3)$ give the contradiction.

abb) $(2,2)\in I$, $(2,1)\notin I$. Transformation 3 with $(2,2)$ $(1,3)$ gives rise to a contradiction unless $(2,3)\in I$. In this latter case there is a $(k,1)\in I$ $(k>2)$. The application of Transformation 3 for $(k,1)$ and $(2,3)$ settles this case.

b) $(1,2)\in I$ but $(1,1)\notin I$. By Lemmas 2.3 and 2.4 there are $(k,1)$, $(l,1)\in I$ $(k,l>1, k\neq l)$. One of $(k,2)$ and $(l,2)$, say $(k,2)$, is missing from I, therefore Transformation 3 can be applied with $(k,1)$ and $(1,2)$. A contradiction.

The cases when $v\geqq u=2, 3$ can be proved similarly.

LEMMA 2.7. *Let* $u,v\geqq 3$. *Suppose that I is an optimal partial 2-transversal and* $(1,1), (1,2), (2,1)\in I$, $(2,2)\notin I$. *Then* $(2,3), (3,2), (3,3)\in I$.

PROOF. Suppose that $u,v\geqq 5$. Use Lemma 2.5 with $A=\{(1,1),(1,2),(2,1)\}$, $i=2, j=2$. $(2,2)$ is not in I by the assumption, thus we have $(2,3)\in I$. Apply Lemma 2.5 now with $A=\{(1,1),(1,2),(2,1),(2,3)\}$, $(3,2)$ and $(4,2)$. If $(3,2)\in I$ then Transformation 3 could be used for $(2,3)$ and $(3,2)$ except in the case $(3,3)\in I$. The statement is proved in this case.

If $(4, 2)\in I$ then $(4, 3)\in I$ can be supposed, again by Transformation 3. The third row contains at least two elements, therefore there exists a $(3, k)\in I$ satisfying $k>4$. $(4, k)\notin I$ is obvious, thus Transformation 3 is applicable with $(4, 2)$ and $(3, k)$. This contradiction proves the lemma for $u, v\geqq 5$.

The cases $v\geqq u=3, 4$ can be proved similarly. ∎

LEMMA 2.8. *If* $u=1\leqq v$ *then the optimal partial* 2*-transversal* I *consists of* $(1, 1)$ *and* $(1, 2)$. *If* $u=2\leqq v$ *then* $I=\{(1, 1), (1, 2), (2, 1), (2, 2)\}$.

The proof is trivial.

It is easy to prove by induction, using Lemmas 2.6—2.8, that the optimal partial 2-transversal consists of blocks

$$\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix} \qquad \begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix}$$

along the diagonal (i, i) and it might end with an

$$11$$

if exactly one row remains at the end. Not making any additional condition on the a's and b's nothing else can be said about the blocks. However, we want to use these results for binomial coefficients. They are ordered in natural order, therefore we have equal pairs among them. Under this condition the structure of the optimal partial 2-transversal can be described rather well.

First we investigate some further transformations.

Transformation 4.

$$\begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{matrix} \rightarrow \begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1. \end{matrix}$$

It is understood that this transformation is made somewhere along the diagonal (i, i) of the matrix $(a_i\cdot b_j)$. Denote by $c_1\geqq\ldots\geqq c_5$ and $d_1\geqq\ldots\geqq d_5$ the values a, resp. b corresponding the rows and columns, resp. The values of the subsums what these submatrices give from $\sum_{(i,j)\in I} a_i b_j$ are

$$c_1d_1+c_1d_2+c_2d_1+c_2d_3+c_3d_2+c_3d_3+c_4d_4+c_4d_5+c_5d_4+c_5d_5$$

and

$$c_1d_1+c_1d_2+c_2d_1+c_2d_2+c_3d_3+c_3d_4+c_4d_3+c_4d_5+c_5d_4+c_5d_5.$$

An easy calculation shows that the second sum is less than or equal to the first sum under the assumption $c_2=c_3$, $d_2=d_3$.

We say that the transformation is *non-increasing*. The *constant* and *non-decreasing* transformations are defined analogously. Easy calculations show the following lemmas.

LEMMA 2.9. *Transformation 4 is non-increasing if* $c_2=c_3$, $d_2=d_3$, *non-decreasing if* $c_3=c_4$, $d_3=d_4$ *and constant if* $c_2=c_3$, $d_3=d_4$ *or* $c_3=c_4$, $d_2=d_3$.

LEMMA 2.10. *Transformation 5 which is defined by*

$$\begin{array}{ccccccccccccccc}
c_1 & 1 & 1 & 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 & 0 & 0 \\
c_2 & 1 & 0 & 1 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 & 0 & 0 \\
c_3 & 0 & 1 & 1 & 0 & 0 & 0 & & 0 & 0 & 1 & 1 & 0 & 0 \\
c_4 & 0 & 0 & 0 & 1 & 1 & 0 & \to & 0 & 0 & 1 & 1 & 0 & 0 \\
c_5 & 0 & 0 & 0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 & 1 & 1 \\
c_6 & 0 & 0 & 0 & 0 & 1 & 1 & & 0 & 0 & 0 & 0 & 1 & 1, \\
 & d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & & & & & & &
\end{array}$$

is non-increasing if $c_2=c_3$, $c_4=c_5$, $d_2=d_3$, $d_4=d_5$, *non-decreasing if* $c_3=c_4$, $d_3=d_4$ *and constant if* $c_2=c_3$, $c_4=c_5$, $d_3=d_4$ *or* $c_3=c_4$, $d_2=d_3$, $d_4=d_5$.

LEMMA 2.11. *Transformation 6 which is defined by*

$$\begin{array}{ccccc}
c_1 & 1 & 1 & 0 & 0 \\
c_2 & 1 & 0 & 1 & 0 \\
c_3 & 0 & 1 & 1 & 0 \\
c_4 & 0 & 0 & 0 & 1
\end{array}
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}
\to
\begin{array}{cccc}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1
\end{array}
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$d_1\, d_2\, d_3\, d_4\, d_5$$

is non-decreasing if $c_3=c_4$, $d_2=d_3$, $(d_4=d_5)$ *or* $c_2=c_3$, $d_3=d_4$ *or* $c_3=c_4$, $d_3=d_4$.

LEMMA 2.12. *Transformation 7 which is defined by*

$$\begin{array}{cccccccccc}
c_1 & 1 & 1 & 0 & 0 & & 1 & 1 & 0 & 0 \\
c_2 & 1 & 0 & 1 & 0 & \to & 1 & 1 & 0 & 0 \\
c_3 & 0 & 1 & 1 & 0 & & 0 & 0 & 1 & 1 \\
 & d_1 & d_2 & d_3 & d_4 & & & & &
\end{array}$$

is constant if $c_2=c_3$ *and* $d_3=d_4$.

LEMMA 2.13. *Transformation 8 which is defined by*

$$\begin{array}{cccc}
c_1 & 1 & 1 & 0 \\
c_2 & 1 & 1 & 0 \\
c_3 & 0 & 0 & 1
\end{array}
\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}
\to
\begin{array}{ccc}
1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 1
\end{array}
\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

is non-decreasing if $c_2=c_3$ *and* $d_2=d_3$.

Using the above transformations we are now able to describe one of the optimal partial 2-transversals. We know that an optimal partial 2-transversal consists of 2×2 and 3×3 blocks. It is called *superoptimal* iff (i) it minimizes the number of 3×3 blocks and (ii) it minimizes the sum of the coordinates of the starting points of the 3×3 blocks among all optimal ones satisfying (i).

LEMMA 2.14. *Let* $u=n_1+1\leqq n_2+1=v$ *and suppose that* $a_1\geqq\ldots\geqq a_u$ *and* $b_1\geqq\ldots\geqq b_v$ *are the binomial coefficients* $\binom{n_1}{i}$ *and* $\binom{n_2}{i}$, *resp. Then the superoptimal partial* 2*-transversal consists of* 2×2 *blocks with two exceptions. If* $n_1\equiv n_2\equiv 0$ (mod_2) *then the first block is a* 3×3 *one. At the end of the diagonal a block of the form* 1 *or* 1 1 *can occur.*

PROOF. Suppose that I is a superoptimal partial 2-transversal. Three cases will be distinguished in the proof.

1) $n_1\not\equiv n_2$ (mod_2). Transformation 5 is constant by Lemma 2.10. It decreases the number of 3×3 blocks. This proves that I cannot contain two neighbouring 3×3 blocks.

However, Transformation 4 is also constant in this case (Lemma 2.9). Applying it backwards, it decreases the sum of the coordinates of starting points of the 3×3 blocks. Therefore I cannot contain a 3×3 block following a 2×2 one. Hence I can have at most one 3×3 block, at the beginning only.

We prove now that even this only one 3×3 block is excluded.

11) $n_1<n_2\equiv 0$ (mod_2). As n_1+1 is even, I ends with a block of the form 1 1. Transformation 4 is constant in this case, we may move the 3×3 block toward the end while I remains optimal. Finally we arrive to the configuration of the left-hand side of Transformation 6. This is non-decreasing, but it decreases the number of 3×3 blocks. I was not superoptimal. This contradiction proves the statement for this case.

12) $n_1<n_2\equiv 1$ (mod_2). n_1+1 is odd. We can repeat the argument of case 11), but Transformation 7 should be used in place of Transformation 6. I cannot contain any 3×3 block in this case.

2) $n_1\equiv n_2\equiv 1$ (mod 2). Consider a 3×3 block B_1 following a 2×2 block. Transformation 4 can be applied backwards unless the coordinate of its starting point of B_1 is odd. Thus we may suppose that this is the case. If B_1 is followed by a 3×3 block then Transformation 5 gives rise to a contradiction. Denote by B_2 the first 3×3 block occuring after B_1. It is easy to see that the coordinate of the starting point of B_2 is even. The converse of Transformation 4 leads to a contradiction. Therefore B_1 cannot be followed by a 3×3 block.

The first two blocks cannot be 3×3 ones because of Lemma 2.10. If the first block is a 3×3 one then the first other 3×3 block following it $(=B_1)$ has an even starting coordinate. This contradiction proves that there is at most one 3×3 block in I and its starting coordinate is odd.

Suppose that there is a 3×3 block. By Transformation 4 we can move this block until the end and I remains optimal. n_1+1 is even, thus the end looks like the left-hand side of Transformation 6. Lemma 2.11 gives the contradiction. I cannot contain any 3×3 block in this case.

3) $n_1 \equiv n_2 \equiv 0 \pmod{2}$. If the first block is a 2×2 then Transformation 4 can be used backwards for the first 3×3 block, with a contradiction. However, all blocks could not be 2×2 because otherwise Transformation 5 would lead to a contradiction, using it backwards. (If $n_1<5$, this argument does not work. For $n_1=2$ and 4 Transformation 8 can be used.) This proves that the first block has to be a 3×3 one. Disregarding the first block, the rest can be treated like case 2). In this case we obtained that the first block is a 3×3 one, all other blocks are 2×2. ∎

Proof of Theorem 1. First we prove that $\max\{|\mathcal{F}|: \mathcal{F}$ is a 3-part Sperner family$\}$ cannot exceed the values given in the theorem. This maximum equals (by Theorem GOS)

$$\max \sum \binom{n_1}{i}\binom{n_2}{j}\binom{1}{k} \tag{12}$$

where we sum over $(i, j, k)\in J$ and the maximum is taken over all partial transversals J in $\{0, \ldots, n_1\}\times\{0, \ldots, n_2\}\times\{0, 1\}$. It is easy to see that the set $I=\{(i,j): (i,j,k)\in J\}$ is a partial 2-transversal in $\{0, \ldots, n_1\}\times\{0, \ldots, n_2\}$. Therefore (12) can be upperbounded with

$$\max \sum_{(i,j)\in I} \binom{n_1}{i}\binom{n_2}{j} \tag{13}$$

where the max runs over all partial 2-transversals I. (13) can be determined by Lemma 2.14.

1) One of n_1 and n_2 is odd. Denoting by a_i and b_i the respective binomial coefficients, (13) can be expressed as

$$\sum (a_i+a_{i+1})(b_i+b_{i+1}) = \sum_{i=1,2,\ldots} a_i b_i + \sum_{i=1,3,\ldots} a_i b_{i+1} + \sum_{i=1,3,\ldots} a_{i+1} b_i. \tag{14}$$

If n_1 is odd then $a_i=a_{i+1}$ $(i=1, 3, \ldots)$ hence

$$\sum_{i=1,3,\ldots} a_i b_{i+1} = \sum_{i=1,3,\ldots} a_{i+1} b_{i+1} = \sum_{i=2,4,\ldots} a_i b_i$$

and

$$\sum_{i=1,3,\ldots} a_{i+1} b_i = \sum_{i=1,3,\ldots} a_i b_i$$

follow. Substituting these into (14) we obtain $2\sum_{i=1,2,\ldots} a_i b_i$. The case when n_2 is odd can be obtained in the same way.

Let n_1 be odd and n_2 be even. Then

$$\sum_{i=1,2,\ldots} a_i b_i = \binom{n_1}{\frac{n_1-1}{2}}\binom{n_2}{\frac{n_2}{2}} + \binom{n_1}{\frac{n_1+1}{2}}\binom{n_2}{\frac{n_2-2}{2}} + \binom{n_1}{\frac{n_1-3}{2}}\binom{n_2}{\frac{n_2+2}{2}} +$$

$$+\binom{n_1}{\frac{n_1+3}{2}}\binom{n_2}{\frac{n_2-4}{2}} + \ldots = \binom{n_1+n_2}{\frac{n_1+n_2-1}{2}}.$$

Multiplying it by 2 we obtain

$$\binom{n_1+n_2+1}{\frac{n_1+n_2+1}{2}} = \binom{n}{\frac{n}{2}}.$$

The case when n_2 is odd and n_1 is even can be treated analogously.

Finally, if $n_1 \equiv n_2 \equiv 1$ (mod$_2$) then

$$\sum_{i=1,3,\dots} a_i b_i = \binom{n_1}{\frac{n_1-1}{2}}\binom{n_2}{\frac{n_2+1}{2}} + \binom{n_1}{\frac{n_1+1}{2}}\binom{n_2}{\frac{n_2-1}{2}} + \dots = \binom{n_1+n_2}{\frac{n_1+n_2}{2}} = \binom{n-1}{\frac{n-1}{2}}$$

proves the upper bound in this case.

2) $n_1 \equiv n_2 \equiv 0$ (mod 2). Lemma 2.14 gives

$$a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_3 + a_3 b_2 + a_3 b_3 +$$
$$+ \sum_{i=4,6,\dots} (a_i + a_{i+1})(b_i + b_{i+1}) =$$
$$= a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_3 + a_3 b_2 + a_3 b_3 + 2 \sum_{i=4,5,6,\dots} a_i b_i =$$
$$= -a_1 b_1 + a_1 b_2 + a_2 b_1 - 2a_2 b_2 + a_2 b_3 + a_3 b_2 - a_3 b_3 + 2 \sum_{i=1,2,\dots} a_i b_i =$$
$$= -\binom{n_1}{\frac{n_1}{2}}\binom{n_2}{\frac{n_2}{2}} + \binom{n_1}{\frac{n_1}{2}}\binom{n_2}{\frac{n_2+2}{2}} + \binom{n_1}{\frac{n_1-2}{2}}\binom{n_2}{\frac{n_2}{2}} - 2\binom{n_1}{\frac{n_1-2}{2}}\binom{n_2}{\frac{n_2+2}{2}} +$$
$$+ \binom{n_1}{\frac{n_1-2}{2}}\binom{n_2}{\frac{n_2-2}{2}} + \binom{n_1}{\frac{n_1+2}{2}}\binom{n_2}{\frac{n_2+2}{2}} - \binom{n_1}{\frac{n_1+2}{2}}\binom{n_2}{\frac{n_2-2}{2}} + 2\binom{n_1+n_2}{\frac{n_1+n_2}{2}} =$$
$$= 2\binom{n-1}{\frac{n-1}{2}} - \left(\binom{n_1}{\frac{n_1+2}{2}} - \binom{n_1}{\frac{n_1}{2}}\right)\left(\binom{n_2}{\frac{n_2+2}{2}} - \binom{n_2}{\frac{n_2}{2}}\right).$$

We have proved that the right-hand side in the theorem is an upper bound. We need constructions proving the equality.

It is easy to check that the following families are 3-part Sperner families and their size is optimal:

$$\mathcal{F} = \left\{F\colon |F| = \frac{n_1+n_2-1}{2},\ |X_3 \cap F| = 0\right\} \cup$$
$$\cup \left\{F\colon |X_2 \cap F| - |X_1 \cap F| = \frac{n_2-n_1-1}{2},\ |X_3 \cap F| = 1\right\}$$

if n_1 is odd, n_2 is even,

$$\mathcal{F} = \left\{F\colon |F| = \frac{n_1+n_2-1}{2},\ |X_3\cap F| = 0\right\} \cup$$

$$\cup \left\{F\colon |X_2\cap F| - |X_1\cap F| = \frac{n_1-n_1+1}{2},\ |X_3\cap F| = 1\right\}$$

if n_1 is even, n_2 is odd,

$$\mathcal{F} = \left\{F\colon |F| = \frac{n_1+n_2}{2},\ |X_3\cap F| = 0\right\} \cup$$

$$\cup \left\{F\colon |X_2\cap F| - |X_1\cap F| = \frac{n_2-n_1}{2},\ |X_3\cap F| = 1\right\}$$

if both n_1 and n_2 are odd and finally

$$\mathcal{F} = \left\{F\colon |F| = \frac{n_1+n_2}{2},\ |X_1\cap F| \neq \frac{n_1}{2},\ \frac{n_1}{2}+1,\ |X_3\cap F| = 0\right\} \cup$$

$$\cup \left\{F\colon |X_1\cap F| = \frac{n_1}{2},\ |X_2\cap F| = \frac{n_2}{2}-1,\ |X_3\cap F| = 0\right\} \cup$$

$$\cup \left\{F\colon |X_1\cap F| = \frac{n_1}{2}+1,\ |X_2\cap F| = \frac{n_2}{2},\ |X_3\cap F| = 0\right\} \cup$$

$$\cup \left\{F\colon |X_2\cap F| - |X_1\cap F| = \frac{n_2-n_1}{2},\ |X_3\cap F| = 1\right\}$$

if both n_1 and n_2 are even. ∎

REFERENCES

[1] Erdős, P. L. and Katona, G. O. H., Convex hulls of more-part Sperner families, *Graphs and Combinatorics* **2** (1986), 123—134.

[2] Füredi, Z., A Ramsey—Sperner theorem, *Graphs and Combinatorics* **1** (1985), 51—56.

[3] Griggs, J. R., Odlyzko, A. M. and Shearer, J. B., *k*-color Sperner theorems, *J. Combinatorial Theory Ser. A* **42** (1986), 31—54.

[4] Katona, G. O. H., On a conjecture of Erdős and a stronger form of Sperner's theorem, *Studia Sci. Math. Hungar.* **1** (1966), 59—63. *MR* **34** #5690.

[5] Kleitman, D. J., On a lemma of Littlewood and Offord on the distribution of certain sums, *Math. Z.* **90** (1965), 251—259. *MR* **32** #2336.

[6] Sperner, E., Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27** (1928), 544—548.

(Received October 31, 1984)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 395—400

SOME REMARKS ON INFINITE SERIES

P. ERDŐS, I. JOÓ and L. A. SZÉKELY

Dedicated to Professor K. Tandori on the occasion of his 60th birthday

In the present paper we investigate the following problems. Suppose $a_n>0$ for $n\geqq 1$ and $\sum_{n=1}^{\infty} a_n=\infty$.

N° 1. Does there exist a sequence of natural numbers $N_0=0$, $N_i\nearrow\infty$, such that it decomposes the series monotone decreasingly:

$$\sum_{j=N_i+1}^{N_{i+1}} a_j \geqq \sum_{j=N_{i+1}+1}^{N_{i+2}} a_j \qquad (i=0,1,2,\ldots)? \tag{1}$$

In order to state the second problem we define the index $n_k(c)$ as the minimum m such that

$$kc \leqq \sum_{j=1}^{m} a_j. \tag{2}$$

Now the second problem is as follows.

N° 2. What is the relation between the behaviour of $\sum_1^{\infty} a_n^2$ and the typical behaviour of $\sum_{k=1}^{\infty} a_{n_k(c)}$ (c is variable)? As it turns out, the two problems are related. Problem N° 1 is motivated by the fact, that for every non-negative continuous function $f\colon [0,\infty)\to\mathbf{R}$ it is easy to define a sequence $x_i\nearrow\infty$ such that $\int_{x_n}^{x_{n+1}} f\geqq$ $\geqq \int_{x_{n+1}}^{x_{n+2}} f$ $(n=0,1,\ldots)$.

THEOREM 1. *Suppose* $a_n>0$, $a_n\geqq a_{n+1}$ *for every* $n\geqq 1$, $\sum_{n=1}^{\infty} a_n=\infty$. *Then for every* $c>0$

$$\sum_{n=1}^{\infty} a_n^2 \quad \text{and} \quad \sum_{k=1}^{\infty} a_{n_k(c)}$$

are equiconvergent.

1980 *Mathematics Subject Classification.* Primary 40A99; Secondary 40A30.
Key words and phrases. Decomposition of series.

PROOF.[1] We may suppose $a_n \searrow 0$, since in the opposite case the statement is trivial. Hence we have for $k > K(c)$

$$n_{k+1}(c) > n_k(c)$$

and

$$\sum_{i=n_k(c)+1}^{n_{k+1}(c)} a_i = c + o(1).$$

In view of monotonicity of (a_n) for $k > K(c)$

$$\Big(\sum_{i=n_k(c)+1}^{n_{k+1}(c)} a_i\Big) a_{n_k(c)} \geqq \sum_{i=n_k(c)+1}^{n_{k+1}(c)} a_i^2 \geqq \Big(\sum_{i=n_k(c)+1}^{n_{k+1}(c)} a_i\Big) a_{n_{k+1}(c)},$$

and the equiconvergence holds. ∎

Theorem 1 makes possible to give a partial solution for problem N° 1.

THEOREM 2. *Suppose* $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$.

(i) *If* (a_n) *has a majorant* $(b_n) \in l_2$ *with* $b_n \geqq b_{n+1}$ *for* $n \geqq 1$, *then* $\sum a_n$ *has the decomposition required in* (1).

(ii) *If* $a_n \geqq a_{n+1}$ *for* $n \geqq 1$, $(a_n) \notin l_2$, *then there exists a series* $\sum b_n$ *having no decomposition and* $1/3 < a_n/b_n < 3$.

PROOF. In the first step we prove the existence of the required decomposition (1) for (b_n). Let $N_0 = 0$. We define N_1 so large, that

$$K_1 := \sum_{j=1}^{N_1} b_j$$

obeys

(3) $$K_1/6 > \max_n b_n$$

(4) $$\sum_{k=1}^{\infty} b_{n_k(K_1/3)} < K_1/2.$$

The number N_1 exists, since $\sum_{k=1}^{\infty} b_{n_k(c)}$ is finite by Theorem 1 and monotone decreasing in c, and K_1 is as large as we want.

Suppose $N_0, N_1, \ldots, N_i, N_{i+1}$ are defined and

$$K_i := \sum_{j=N_i+1}^{N_{i+1}} b_j \geqq K_1/2.$$

Let N_{i+2} be the largest index for which

$$\sum_{j=N_i+1}^{N_{i+1}} b_j \geqq \sum_{j=N_{i+1}+1}^{N_{i+2}} b_j.$$

[1] The present simple proof is due to G. Petruska.

By (3) we have $N_{i+2}>N_{i+1}$. We prove $K_{i+1}:=\sum_{j=N_{i+1}+1}^{N_{i+2}} b_j \geqq K_1/2$, what means, N_i and K_i are defined for $i>0$ with $K_1\geqq K_2\geqq K_3\geqq \dots$.

Assume m is the least integer with $K_{m+1}<K_1/2$. First, $K_m\geqq K_1/2$ and by the choice of N_i's and by (3) $K_m-K_{m+1}<K_1/6$, hence $K_{m+1}\geqq K_1/3$. On the other hand

$$K_1-K_{m+1}\leqq \sum_{i=0}^{m+1} b_{N_{i+1}+1}\leqq \sum_{k=1}^{m+1} b_{n_k(K_{m+1})}\leqq \sum_{k=1}^{\infty} b_{n_k(K_1/3)}.$$

Using (4) we have $K_{m+1}\geqq K_1/2$, a contradiction.

In the second step set $M_0=0$, select M_1 so large that

$$K_1<\sum_{j=1}^{M_1} a_j$$

and let M_{i+2} be the largest integer with

$$\sum_{j=M_i+1}^{M_{i+1}} a_j \geqq \sum_{j=M_{i+1}+1}^{M_{i+2}} a_j .$$

Set

$$L_i:=\sum_{j=M_i+1}^{M_{i+1}} a_j .$$

We have to prove $M_{m+2}>M_{m+1}$ for $m>0$. Obviously, $M_i\geqq N_i$ and

$$L_1-L_{m+1}\leqq \sum_{i=0}^{m+1} a_{M_{i+1}+1}\leqq \sum_{i=0}^{m+1} b_{M_{i+1}+1}\leqq \sum_{i=0}^{m+1} b_{N_{i+1}+1}\leqq \sum_{k=1}^{\infty} b_{n_k(K_1/2)}<K_1/2,$$

what means $L_{m+1}>K_1/2$, i.e. $M_{m+2}>M_{m+1}$. In order to prove (ii) suppose without loss of generality $a_1<1$ and set $f(0)=0$,

$$f(n):=\left|\{k\colon 2^{-n}\leqq a_k<2^{-n+1}\}\right|$$

for $n\geqq 1$. It is well-known that

$$\sum_{n=1}^{\infty} a_n^2<\infty$$

if and only if $\sum_{n=1}^{\infty} f(n)4^{-n}<\infty$. If $f(n)>0$ we define a strictly monotone increasing sequence $\varepsilon_{n,j}$ $(j=1,2,\dots,f(n))$ obeying $0\leqq\varepsilon_{n,j}\leqq 4^{-n}$. For every natural number i there exists a unique m with

$$f(0)+f(1)+\dots+f(m-1)<i\leqq f(0)+f(1)+\dots+f(m).$$

We define

(5) $$b_i:=2^{-m}+\varepsilon_{m,\,i-\sum_{j=0}^{m-1} f(j)},$$

and prove that $\sum b_i$ satisfies the requirements of (ii). Obviously, $1/3<a_n/b_n<3$. The sequence (b_i) is monotone increasing in the intervals

$$\left(\sum_{j=0}^{m-1} f(j),\ \sum_{j=0}^{m} f(j)\right]$$

of indices, by (5).

Suppose there exists a decomposition required in (1) for $\sum b_i$ with indices $N_0=0<N_1<N_2<\dots$ and

$$K_i = \sum_{j=N_i+1}^{N_{i+1}} b_j.$$

We are going to prove $K_1=\infty$, a contradiction. If

$$\sum_{j=0}^{m-1} f(j) \leqq N_i < N_{i+1} < \sum_{j=0}^{m} f(j) \tag{6}$$

then $K_i-K_{i+1}\geqq 2^{-m}$, since $N_{i+2}-N_{i+1}<N_{i+1}-N_i$ by the strictly monotone increasingness of (b_i) in the above considered interval. Since $K_1\geqq K_2\geqq K_3\geqq\dots$ by (1), we have

$$\left|\{i\colon (6) \text{ holds for } i\}\right| \geqq \frac{f(m)}{K_1\cdot 2^m}-3.$$

Comparing our estimates we have

$$K_1 \geqq \sum_{i=0}^{\infty}(K_i-K_{i+1}) \geqq \sum_{(6)\text{ holds for } i}(K_i-K_{i+1}) \geqq \sum_m 2^{-m}\left(\frac{f(m)}{K_1\cdot 2^m}-3\right)=\infty. \quad \blacksquare$$

M. Szegedy noted, that with a bit more effort one can prove (ii) with $b_i=a_i(1+o(1))$. We have conjectured that $(a_n)\in l_2$ is sufficient for having a decomposition. Recently, the conjecture was proved by M. Szegedy and G. Tardos [1].

Now we investigate what happens if we drop the condition $a_n\geqq a_{n+1}$ from Theorem 1. It is clear, that dropping the condition a counterexample can be given for a fixed c, but we have

THEOREM 3. *Suppose* $a_n>0$, $\sum_{n=0}^{\infty} a_n=\infty$. *If*

$$\sum_{n=0}^{\infty} a_n^2<\infty, \quad \textit{then} \quad X:=\left\{c\colon \sum_{k=1}^{\infty} a_{n_k(c)}=\infty\right\}$$

is of measure zero, and if

$$\sum_{n=0}^{\infty} a_n^2=\infty, \quad \textit{then} \quad Y:=\left\{c\colon \sum_{k=1}^{\infty} a_{n_k(c)}<\infty\right\}$$

is meagre (i.e. of first category).

PROOF. In the first case we have for $0<a<b<\infty$

$$\sum_{k=1}^{\infty}\int_a^b a_{n_k(c)}\,dc<\infty,$$

what proves the first statement by Beppo Levi's theorem. Indeed, we have for $k>K(c)$

$$\int_a^b a_{n_k(c)}\,dc \leqq \frac{1}{k}\sum_{ka\leqq \sum_{i=1}^{j} a_i<kb} a_j^2$$

and

$$\sum_{k=1}^{\infty} \int_a^b a_{n_k(c)}\, dc \leqq \sum_{j=1}^{\infty} a_j^2 \sum_{\frac{1}{b} \sum_{i=1}^{j} a_i \leqq k < \frac{1}{a} \sum_{i=1}^{j} a_i} \frac{1}{k} = \sum_{j=1}^{\infty} a_j^2 \left(\log \frac{b}{a} + o(1)\right) < \infty.$$

In the second case we prove for $0<a<b<\infty$

$$\sum_{k=1}^{\infty} \int_a^b a_{n_k(c)}\, dc = \infty.$$

It is trivial, if $\inf_n a_n = \varepsilon > 0$. If not, the previous estimates will be repeated for $a<a'<$ $<b'<b$ in the inverse direction and

$$\sum_{k=1}^{\infty} \int_a^b a_{n_k(c)}\, dc \geqq \sum_{j=1}^{\infty} a_j^2 \left(\log \frac{b'}{a'} + o(1)\right) = \infty.$$

The function $c \to f(c) := \sum_{k=1}^{\infty} a_{n_k(c)}$ is lower semicontinuous from the left side since $\lim_{c \to c_0 -} f(c) \geqq f(c_0)$, so

$$H_i := \left\{c\colon \sum_{k=1}^{\infty} a_{n_k(c)} > i\right\}$$

contains a dense open set $G_i \subset (0, \infty)$. This way

$$\left\{c\colon \sum_{k=1}^{\infty} a_{n_k(c)} = \infty\right\} = \bigcap_i H_i \supset \bigcap_i G_i$$

and

$$\left\{c\colon \sum_{k=1}^{\infty} a_{n_k(c)} < \infty\right\}$$

is meagre. ∎

The size of an exceptional set in Theorem 3 is still an open question. A particular answer is given by the next construction.

THEOREM 4. *X can be residual, and Y can be of cardinality continuum.*

PROOF. We construct $\sum_{n=1}^{\infty} a_n^2 < \infty$ with a residual X. Suppose $\{\alpha_i\colon i \in \mathbf{N}\}$ is dense in $(0, \infty)$ and let β_i be $\beta_i = \alpha_{i - \binom{k}{2}}$ if $\binom{k}{2} < i \leqq \binom{k+1}{2}$. For every β_i set some segments $a_j\colon j \in I_i$, so, that

— I_i finite, $a_j\colon j \in I_i$ are disjoint,

— on the ray $(0, \infty)$ all $a_j\colon j \in I_i$ is on the right hand from all $a_j\colon j \in I_k$, where $k<i$,

— $\sum_{j \in I_i} a_j^2 < \frac{1}{2^i}$, $\sum_{j \in I_i} a_j \geqq 1$,

— all the segments a_j have in their interior a multiple of β_i.

We cover the rest of the ray with segments $a_j\colon j \in J$ such that $\sum_{j \in J} a_j^2 < \infty$.

If β_i is the n-th repetition of α_k, there is a neighbourhood V_k^n of α_k, such that $m_j\alpha_k\in a_j$ $(m_j\in\mathbf{N})$ implies $m_jV_k^n\subset a_j$ $(j\in I_i)$. Now clearly $\bigcap_n\big(\bigcup_k V_k^n\big)$ is residual and X contains it.

Now we construct a perfect set Y (i.e. of cardinality continuum) in the following way. Set $I_0^1=[100, 101]$, we are going to define closed intervals I_n^i $(i=1, \ldots, 2^n)$ for $n=1, 2, \ldots$ with the property: I_n^i contains the disjoint intervals I_{n+1}^{2i} and I_{n+1}^{2i-1}. We have a perfect set $\bigcap_n\big(\bigcup_i I_n^i\big)=Y$. In $\bigcup_i I_n^i$ we select 2^{n+1} numbers $x_1, \ldots, x_{2^{n+1}}$ independent over the field of rationals, two of which are in $\operatorname{int} I_n^i$ $(i=1, \ldots, 2^n)$. By Kronecker's Theorem for infinitely many α_j

$$|\alpha_j-k_{i,j}x_i|<0{,}001$$

for $i=1, 2, \ldots, 2^{n+1}$, $k_{i,j}$ integer. We are interested only in $\alpha_1, \ldots, \alpha_n$. We set an interval $J_m^{(n)}$ $(m=1, \ldots, n)$, $|J_m^{(n)}|=1/200$ close to α_j but right to it, $J_m^{(n)}$ not containing any multiple of $x_1, x_2, \ldots, x_{2^{n+1}}$, right from the previous $J_i^{(l)}$ $(l<n;\ 1\leq i\leq 2^l)$. Now we define I_{n+1}^i as short intervals centered at x_i, so that none of the $J_m^{(n)}$ $(m=1, \ldots, n)$ intersect any multiple of I_{n+1}^i. Finally we define the series $\sum_{n=1}^{\infty} a_n$. All the intervals $J_m^{(n)}$ $(n=1, 2, \ldots;\ m=1, 2, \ldots, n)$ occur as some $a_{s(n,m)}$ with

$$\sum_{i=1}^{s(n,m)} a_i=\text{the right endpoint of } J_m^{(n)}.$$

The "undefined gaps" in $\sum a_n$ we fill with small numbers tending quickly to zero.

It is easy to check, that $\sum a_n=\infty$, $\sum a_n^2=\infty$, since $a_n\not\to 0$. $c\in Y$ implies $\sum a_{n_k(c)}<\infty$, since the multiples of c avoid all the intervals $J_m^{(n)}$.

Remark. With a little care we can construct a series with the above properties with $a_n\to 0$.

Problem 1. Is there a topological property φ such that

$$\{c\colon \sum a_{n_k(c)}<\infty\}\in\varphi \quad \text{if and only if} \quad \sum a_n^2<\infty?$$

Problem 2. Is there a series $\sum a_n^2<\infty$ in Theorem 3 with Y of positive measure?

REFERENCE

[1] Szegedy, M. and Tardos, G., On infinite series, *Studia Sci. Math. Hungar.* (to appear).

(Received November 12, 1984)

P. Erdős

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST

I. Joó and L. A. Székely

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6—8
H—1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 401—406

A CHARACTERIZATION OF THE HELLY DIMENSION OF CONVEX BODIES

I. BÁRÁNY and J. KINCSES

1. Introduction

We prove that the Helly dimension, $\operatorname{him} K$ of a convex body $K \subset \mathbf{R}^d$ equals the positive dimension, $\operatorname{md}^+ S$ of an arbitrary dense subset S of the set of extreme points of K^*, the dual of K.

V. G. Boltjanskij proved a general theorem in [1] in which he determined the Helly dimension of certain systems of convex bodies. In a special case his theorem gives that $\operatorname{him} K = \operatorname{md}^+ \operatorname{ext} K^*$ under the assumption that for each extreme supporting hyperplane H of K, $H \cap K$ contains a regular point of K. The aim of this paper is to generalize this fact. As a corollary to our main result we answer a question of Boltjanskij and Soltan (see [2], p. 131). Namely, we prove that if H is a subset of the unit sphere such that it is not contained in a closed hemisphere, then $\operatorname{md}^+ H = \operatorname{md}^+ \bar{H}$ where $\bar{H}$ is the closure of H.

2. The main results

If $S \subseteqq \mathbf{R}^d$ then $\operatorname{conv} S$, $\operatorname{con} S$, $\operatorname{int} S$, $\bar{S}$ will denote the convex hull of S, the convex cone hull of S, the interior of S and the closure of S, resp. and $\operatorname{ext} S$ is the set of extreme points of S and S^* is the dual set of S that is

$$S^* = \{x \in \mathbf{R}^d,\ \langle x, s \rangle \leqq 1 \text{ for all } s \in S\}.$$

A convex body is a compact convex set with nonempty interior.

DEFINITION 2.1. The *Helly dimension* of a convex body K is the smallest integer k such that any finite family of positive homothetic images of K has nonempty intersection whenever any $k+1$ of them has nonempty intersection.

The Helly dimension of K will be denoted by $\operatorname{him} K$.

DEFINITION 2.2. Let $x_1, \ldots, x_n \in \mathbf{R}^d$. We say that $\{x_i\}_{i=1}^n$ is a minimal positive linearly dependent set if the vectors x_i are linearly dependent with positive coefficients but no proper subset of $\{x_i\}_{i=1}^n$ is linearly dependent. The positive dimension, $\operatorname{md}^+ S$ of $S \subseteqq \mathbf{R}^d$ is the largest integer k such that S contains a minimal positive linearly dependent set with $k+1$ elements.

1980 *Mathematics Subject Classification.* Primary 52A20; Secondary 52A35.

Key words and phrases. Convex sets, Helly type theorems, Helly dimension, duality, positive dimension.

The scalar product of $x, y \in \mathbf{R}^d$ is denoted by $\langle x, y\rangle$. If $a \in \mathbf{R}^d$ and α is a scalar, then (a, α) denotes the vector in $\mathbf{R}^{d+1}$ whose first d coordinates coincide with those of a and the last coordinate is α. We will use the following theorem of Ky Fan [3].

THEOREM 2.3. *The system of linear inequalities*

$$\langle a_\nu, x\rangle \leqq \alpha_\nu \qquad \nu \in I, \quad a_\nu \in \mathbf{R}^d, \quad \alpha_\nu \in \mathbf{R}$$

is solvable if and only if

$$(0, -1) \notin \overline{\operatorname{con}}\, \{(a_\nu, \alpha_\nu) \in \mathbf{R}^{d+1},\ \nu \in I\}.$$

Now we can state the main theorem.

THEOREM 2.4. *If* $K \subseteqq \mathbf{R}^d$ *is a convex body with* $0 \in \operatorname{int} K$ *and* S *a dense subset of* $\operatorname{ext} K^*$ *then*

$$\operatorname{him} K = \operatorname{md}^+ S.$$

PROOF. Let $\operatorname{him} K = k$ and $\operatorname{md}^+ S = l$.

First part. We show that $k \leqq l$. Suppose, on the contrary, that there exists a finite family of positive homothetic images of K

$$b_i + \lambda_i K \qquad b_i \in \mathbf{R}^d, \quad \lambda_i > 0, \quad i = 1, \dots, n,$$

such that

$$\bigcap_{j=1}^{l+1} (b_{i_j} + \lambda_{i_j} K) \neq \emptyset \quad \text{for all} \quad 1 \leqq i_1 < i_2 < \dots < i_{l+1} \leqq n$$

but

$$\bigcap_{i=1}^{n} (b_i + \lambda_i K) = \emptyset.$$

Since S is dense in $\operatorname{ext} K^*$ we have

$$K = S^* = \{x \in \mathbf{R}^d,\ \langle a, x\rangle \leqq 1 \text{ for each } a \in S\} =$$
$$= \bigcap_{a \in S} \{x \in \mathbf{R}^d, \langle a, x\rangle \leqq 1\}.$$

Applying this we get

$$\emptyset = \bigcap_{i=1}^{n} (b_i + \lambda_i K) = \bigcap_{i=1}^{n} \Big(b_i + \lambda_i \Big[\bigcap_{a \in S} \{x \in \mathbf{R}^d, \langle a, x\rangle \leqq 1\}\Big]\Big) =$$

$$= \bigcap_{i=1}^{n} \bigcap_{a \in S} \{b_i + \lambda_i x \in \mathbf{R}^d, \langle a, x\rangle \leqq 1\} =$$

$$= \bigcap_{i=1}^{n} \bigcap_{a \in S} \left\{y \in \mathbf{R}^d, \left\langle \frac{y - b_i}{\lambda_i},\ a\right\rangle \leqq 1\right\} =$$

$$= \bigcap_{i=1}^{n} \bigcap_{a \in S} \{y \in \mathbf{R}^d, \langle a, y\rangle \leqq \langle a, b_i\rangle + \lambda_i\} =$$

$$= \bigcap_{a \in S} \Big\{y \in \mathbf{R}^d, \langle a, y\rangle \leqq \alpha(a) = \min_{1 \leqq i \leqq n} (\langle a, b_i\rangle + \lambda_i)\Big\}.$$

This yields that the system of linear inequalities

$$\langle a, y\rangle \leqq \alpha(a), \quad a \in S$$

has no solution. Using Theorem 2.3 we have

$$(0, -1) \in \overline{\operatorname{con}}\{(a, \alpha(a)) \in \mathbf{R}^{d+1}, a \in S\} = C.$$

If $(0, -1)$ is on the boundary of C then applying the separation theorem for $(0, -1)$ and $\operatorname{int} C$ we get a point $(z, \xi) \in \mathbf{R}^{d+1}$, $(z, \xi) \neq (0, 0)$ such that

$$\langle (z, \xi), (0, -1)\rangle = 0 \quad \text{and} \quad \langle (z, \xi), (a, \alpha(a))\rangle \leqq 0 \quad \text{for all} \quad a \in S.$$

But then $\xi = 0$ and $\langle z, a\rangle \leqq 0$ for all $a \in S$, which implies that $K + z \subseteq K$ and this contradicts to the fact that K is compact. Thus $(0, -1) \in \operatorname{con}\{(a, \alpha(a)) \in \mathbf{R}^{d+1}, a \in S\}$ and so there exist a finite subset $\{a_j\}_{j=1}^m$ of S and positive numbers $\mu_j > 0$ $(j = 1, \ldots, m)$ such that

$$\sum_{j=1}^m \mu_j a_j = 0 \quad \text{and} \quad \sum_{j=1}^m \mu_j \alpha(a_j) = -1.$$

Obviously, we may assume that the set $\{a_j\}_{j=1}^m$ is a minimal positive linearly dependent set and so $m \leqq l + 1$. Now choose for every $1 \leqq j \leqq m$ an index $1 \leqq i_j \leqq n$ such that

$$\alpha(a_j) = \langle a_j, b_{i_j}\rangle + \lambda_{i_j} = \min_{1 \leqq i \leqq n} (\langle a_j, b_i\rangle + \lambda_i).$$

Using again Theorem 2.3 we get that the intersection of the halfspaces

$$\{y \colon \langle a_j, y\rangle \leqq \langle a_j, b_{i_j}\rangle + \lambda_{i_j}\}$$

is empty. But we have

$$b_{i_j} + \lambda_{i_j} K \subset \{y \colon \langle a_j, y\rangle \leqq \langle a_j, b_{i_j}\rangle + \lambda_{i_j}\}$$

and this implies that

$$\bigcap_{j=1}^m (b_{i_j} + \lambda_{i_j} K) = \emptyset$$

which is a contradiction since $m \leqq l + 1$. Thus we proved that $k \leqq l$.

Second part. We show that $l \leqq k$. To prove this we need the following lemma, which is probably folklore.

LEMMA. *Let $P \subseteq \mathbf{R}^d$ be a closed convex set with nonempty interior and $H_0 = \{x \colon \langle x - p_0, x_0\rangle = 0\}$ be an extremal supporting hyperplane of P at the point p_0 such that*

$$P \subseteq \{x \colon \langle x - p_0, x_0\rangle \leqq 0\}.$$

If $T \subset \{x \colon \langle x - p_0, x_0\rangle < 0\}$ is a compact set then there exists a positive homothetic image $b + \lambda P$ of P such that

$$T \subset \operatorname{int}(b + \lambda P) \subset \{x \colon \langle x - p_0, x_0\rangle < 0\}.$$

PROOF OF THE LEMMA. Without loss of generality we may assume that T is a Euclidean ball $B(a_0)$ with centre a_0 and $p_0=0$. We use induction on d. For $d=1$ the lemma is clear. Suppose that $d>1$ and let C_0 be the supporting cone of P at 0 that is

$$C_0=\overline{\bigcup_{\lambda\geqq 0}\lambda P}.$$

Since H_0 is extremal supporting hyperplane there is a halfline $l_0^+=\{\lambda x_1:\ \lambda\geqq 0\}$ such that $H_0\cap C_0\supseteqq l_0^+$. Now let

$$C_1=\overline{\bigcup_{\lambda\geqq 0}(C_0-\lambda x_1)}.$$

Then it is easy to show that C_1 is a closed convex cone and

$$(*)\qquad \begin{aligned}&\operatorname{int}(C_0-\lambda x_1)\supseteqq\operatorname{int}(C_0-\lambda' x_1)\quad\text{if}\quad\lambda\geqq\lambda'\\ &\operatorname{int}C_1=\bigcup_{\lambda\geqq 0}\operatorname{int}(C_0-\lambda x_1).\end{aligned}$$

Let H_1 be the hyperplane through a_0 and orthogonal to x_1. Then obviously $H_1\cap C_1$ is a closed convex cone and $H_1\cap H_0$ is an extremal supporting hyperplane of $H_1\cap C_1$ in H_1. By the induction hypothesis we obtain a translate $(H_1\cap C_1)+b_0$ of $H_1\cap C_1$ in H_1 such that

$$H_1\cap B(a_0)\subseteq\operatorname{rel}\operatorname{int}\big((H_1\cap C_1)+b_0\big)\subset H_1\cap\{x:\ \langle x_0,x\rangle<0\}.$$

But then, since C_1 contains the line $l_0=\{\lambda x_1:\ \lambda\in\mathbf{R}\}$, we have that the cylinder $(B(a_0)\cap H_1)+l_0$ is contained in $\operatorname{int}(C_1+b_0)=\operatorname{rel}\operatorname{int}\big((H_1\cap C_1)+b_0\big)+l_0$. So we get

$$B(a_0)\subseteqq(B(a_0)\cap H_1)+l_0\subseteqq\operatorname{int}(C_1+b_0)\subseteqq\{x:\ \langle x_0,x\rangle<0\}.$$

Using the relations $(*)$ and the compactness of $B(a_0)$ there exists a $\lambda_0>0$ such that

$$B(a_0)\subseteqq\operatorname{int}(C_0-\lambda_0x_1+b_0)\subseteqq\{x:\ \langle x_0,x\rangle<0\}.$$

Since $\operatorname{int}C_0=\bigcup_{\lambda>0}\lambda\operatorname{int}P$ then again by compactness of $B(a_0)$ there exists a $\lambda_1>0$ such that

$$B(a_0)\subseteqq\operatorname{int}(\lambda_1P-\lambda_0x_1+b_0)\subseteqq\{x:\ \langle x_0,x\rangle<0\}$$

and this proves the statement for d which completes the proof of the Lemma.

Returning to the proof of the second part of Theorem 2.4 let $\{a_i\}_{i=1}^{l+1}$ be a minimal positive linearly dependent subset of S. Then, by definition, there are positive constants μ_i $(1\leqq i\leqq l+1)$ such that

$$\sum_{i=1}^{l+1}\mu_ia_i=0\quad\text{and}\quad\sum_{i=1}^{l+1}\mu_i=1.$$

By Theorem 2.3 the system of linear inequalities

$$\langle a_i,x\rangle\leqq -1\qquad(i=1,\dots,l+1)$$

has no solution but for each $1 \leqq i \leqq l+1$ there is a point x_i:

$$x_i \in \bigcap_{\substack{j=1 \\ j \neq i}}^{l+1} \{x: \langle a_j, x \rangle < -1\}.$$

Let $T_i = \{x_j: j \neq i\}$ for $i=1, \ldots, l+1$. Since $\{x: \langle a_i, x \rangle = -1\}$ is an extremal supporting hyperplane of $K - 2\frac{a_i}{\|a_i\|^2}$ and $T_i \subseteqq \{x: \langle a_i, x \rangle < -1\}$ then, by the lemma, we get a $\lambda_i > 0$ and a $b_i \in \mathbf{R}^d$ such that

$$T_i \subset \operatorname{int}(b_i + \lambda_i K) \subseteqq \{x: \langle a_i, x \rangle < -1\}.$$

This means that

$$\bigcap_{i=1}^{l+1} (b_i + \lambda_i K) = \emptyset$$

and

$$\bigcap_{\substack{i=1 \\ i \neq j}}^{l+1} (b_i + \lambda_i K) \neq \emptyset$$

and this implies $l \leqq k$. This completes the proof of the theorem.

We need one more definition. The set $\{x_1, \ldots, x_n\} \subseteqq \mathbf{R}^d$ is a minimal linearly dependent set if the vectors x_i are linearly dependent but no proper subset of $\{x_1, \ldots, x_n\}$ is linearly dependent. For $S \subset \mathbf{R}^d$, $\operatorname{md} S$ denotes the largest integer k such that S contains a minimal linearly dependent set with $k+1$ elements. If S is centrally symmetric with respect to the origin then obviously $\operatorname{md}^+ S = \operatorname{md} S$. Thus we have

Corollary 2.5. *If the convex body $K \subseteqq \mathbf{R}^d$ is centrally symmetric with respect to the origin then*

$$\operatorname{him} K = \operatorname{md} \operatorname{ext} K^*.$$

Boltjanskij and Soltan asked the following question (see [2], p. 131): Assume $H \subseteqq \mathbf{R}^d$ is a subset of the unit sphere such that H is not contained in a closed hemisphere. Is it true then, that $\operatorname{md}^+ H = \operatorname{md}^+ \overline{H}$? This question is answered in the affirmative by

Theorem 2.6. *If $H \subseteqq \mathbf{R}^d$ is a subset of the unit sphere such that H is not contained in a closed hemisphere then*

$$\operatorname{md}^+ H = \operatorname{md}^+ \overline{H}.$$

Proof. If we take $K = \operatorname{conv} \overline{H}$ then obviously $\operatorname{ext} K = \overline{H}$ and $0 \in \operatorname{int} K$. Using Theorem 2.4. we get

$$\operatorname{md}^+ H = \operatorname{him} K^* = \operatorname{md}^+ \operatorname{ext} K = \operatorname{md}^+ \overline{H}.$$

Boltjanskij and Soltan in [2] prove several theorems of this type: The Helly number of a certain family of convex sets equals $\operatorname{md}^+ \overline{H}$ where H is a set determined by the family. By Corollary 2.5, $\mathrm{md}^+ \bar{H}$ can be replaced by $\mathrm{md}^+ H$ because $H \subset S^{d-1}$ is not contained in a closed hemisphere in these theorems.

We mention finally that a weaker version of Theorem 2.4 was originally proved by the second author. He used a theorem of A. Lima [5] and applied the result to classify the k-Helly dimensional convex bodies for $k \leqq 4$. Earlier this was known only for $k \leqq 2$.

REMARK. The question of this paper and the Boltjanskij's conjecture was solved by R. Živaljević [6], using different methods.

REFERENCES

[1] BOLTJANSKIJ, V. G., Helly's theorem for H-convex sets, *Dokl. Akad. Nauk. SSSR* **226** (1976), 249—252. *MR* **54** #3593.
[2] BOLTJANSKIJ, V. G. and SOLTAN, P. S., *Combinatorial geometry of various classes of convex sets*, Stiinca, Kishinev, 1978 (in Russian). *MR* **80g**: 52001.
[3] FAN, K., On infinite systems of linear inequalities, *J. Math. Anal. Appl.* **21** (1968), 475—478. *MR* **37** #3305.
[4] KINCSES, J., The classification of 3 and 4 dimensional convex bodies, *Geometriae Dedicata* (to appear).
[5] LIMA, Å., Intersection properties of balls and subspaces in Banach spaces, *Trans. Amer. Math. Soc.* **227** (1977), 1—63. *MR* **55** #3752.
[6] ŽIVALJEVIC, R., md H = md $\bar{H}$, *Publ. Inst. Math. N. S.* **26** (40) (1979), 307-311.

(Received January 14, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

JÓZSEF ATTILA TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
BOLYAI INTÉZET
ARADI VÉRTANÚK TERE 1
H—6720 SZEGED
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 407—444

PROLONGEMENT ANALYTIQUE A TRAVERS UN T-FILTRE

MARIE-CLAUDE SARMANT et ALAIN ESCASSUT

I. Introduction et principaux résultats

§1. Introduction

Soit $(K, |\cdot|)$ un corps muni d'une valeur absolue ultramétrique, complet et algébriquement clos.

Rappelons qu'une partie D de K est dite *infraconnexe* si pour tout $a \in D$ l'application φ_a définie dans D par $\varphi(x)=|x-a|$, a une image globale dans $\mathbf{R}_+$ dont l'adhérence est un intervalle de $\mathbf{R}_+$.

Soit D un fermé borné de K. Rappelons qu'on note $R(D)$ l'algèbre des fractions rationnelles $h(x) \in K(x)$ sans pôle dans D et on note $H(D)$ l'algèbre de Banach complétée de $R(D)$ pour la norme de la converg'nce uniforme $\|\cdot\|_D$ sur D [13], [3].

Soit D un infraconnexe fermé borné, soit $\mathfrak{F}$ un T-filtre à plage vide de D et soit $\mathfrak{J}(\mathfrak{F})$ l'idéal des éléments tel que $\lim_{\mathfrak{F}} f(x)=0$ [5].

Nous montrons d'abord que, si les diamètres des trous de D sont minorés par un nombre $\delta>0$, la topologie d'algèbre de Banach quotient de $H(D)/\mathfrak{J}(\mathfrak{F})$ est équivalente à celle de la valeur absolue définie sur $H(D)/\mathfrak{J}(\mathfrak{F})$ par le quotient de la semi-norme multiplicative définie par $\mathfrak{F}$ sur $H(D)$.

Ce résultat est généralement faux si on ne fait pas cette hypothèse sur les diamètres des trous, comme le montre un contre-exemple. On généralise ainsi un théorème obtenu en 1971 dans le cas particulier où chaque cercle centré à l'origine contenait un seul trou [2].

Alors l'équivalence entre les deux topologies évoquées ci-dessus permet d'établir que tout élément analytique défini sur la plage $\mathfrak{P}(\mathfrak{F})$ d'un T-filtre $\mathfrak{F}$ d'un infraconnexe D, dont les diamètres des trous sont minorés par $\delta>0$, se prolonge « à travers le T-filtre », (d'une infinité de façons) en élément de $H(D)$ et s'il existe une T-suite idempotente alors le résultat qui précède montre l'existence d'un prolongement méromorphe sur $\mathfrak{F}$, à pôles simples.

Par exemple, soit $f(x)$ une série de Taylor convergente dans $d(0, R)$, soit b_n une suite de K telle que $|b_n|>|b_{n+1}|$, $\lim_{n\to\infty} |b_n|=R$, $\prod_{n=1}^{\infty} \frac{R}{|b_n|}=0$ et pour tout $\varrho>0$ soit $\Lambda_\varrho=K\setminus \bigcup_{n=1}^{\infty} d^-(b_n, \varrho)$. Alors f se prolonge en éléments analytiques sur Λ_ϱ, méromorphe dans K admettant chaque b_n pour pôle simple.

Cette dernière conclusion s'applique particulièrement au cas d'un T-filtre dont les diamètres de trous appartiennent à un intervalle $[\varrho_1, \varrho_2] \subset]0, \operatorname{diam}(\mathfrak{F})[$ puisque dans ce cas on sait qu'il existe des T-suites idempotentes [17].

1980 *Mathematics Subject Classification.* Primary 32E27; Secondary 46F05.

Ceci apporte donc un élément de réponse à la question du prolongement d'une série de Taylor dont le disque de convergence est circonférencié évoqué dans [16].

D'autre part, quand cette équivalence entre les deux topologies ci-dessus est réalisée, on peut montrer que l'algèbre quotient $H(D)/\mathfrak{J}_0(\mathfrak{F})$ est isomorphe (algébriquement et topologiquement) à $H(\mathfrak{P}(\mathfrak{F}))$, en notant $\mathfrak{J}_0(\mathfrak{F})$ l'idéal des éléments nuls sur $\mathfrak{P}(\mathfrak{F})$. Cet isomorphisme n'existe plus si les deux topologies considérées ne sont plus équivalentes.

Enfin, en considérant une série de Taylor f convergeant dans le disque $d(0, R)$, nous construisons concrètement un infraconnexe D avec un T-filtre dont la plage est $d(0, R)$ et nous définissons un prolongement analytique de f dans D par son développement de Mittag-Leffler. On utilise pour cela notamment une matrice infinie de van der Monde en montrant qu'elle est inversible grâce aux propriétés du T-filtre considéré. Cette construction s'effectue à l'aide d'une série de Taylor convergente dans $d^-(0, R)$ satisfaisant certaines propriétés et on montre que de telles séries existent.

§ 2. Notations et définitions

L'énoncé des résultats utilise les définitions et notations habituelles qu'il faut rappeler.

Pour tout $a \in K$, $r \in \mathbf{R}_+$, on note

$$d(a, r) = \{x \mid |x-a| \leqq r\}$$

$$d^-(a, r) = \{x \mid |x-a| < r\}$$

$$C(a, r) = \{x \mid |x-a| = r\}$$

et si $R > r$, on note

$$\Gamma(a, r, R) = \{x \mid r < |x-a| < R\}.$$

Soit D un infraconnexe de K et soit $a \in K$ et soit $R \in \mathbf{R}_+^*$. Rappelons qu'on appelle *filtre croissant de centre a, de diamètre R*, le filtre de D qui admet pour base les ensembles $\Gamma(a, r, R) \cap D$ quand r parcourt $]0, R[$. On note $\mathfrak{P}(\mathfrak{F}) = \{x \in D \mid |x-a| \geqq R\}$.

On appelle *filtre décroissant, de diamètre* $R > 0$ un filtre admettant une base de la forme $D_m = (\Delta_m \setminus \bigcap_{j=1}^{\infty} \Delta_j) \cap D$ où la suite Δ_m est une suite décroissante de disques de K telle que $\lim_{m \to \infty} \operatorname{diam}(\Delta_m) = R$. Tout point de $\bigcap_{m=1}^{\infty} \Delta_m$ est appelé *centre* de $\mathfrak{F}$, et on note $\mathfrak{P}(\mathfrak{F}) = (\bigcap_{m=1}^{\infty} \Delta_m) \cap D$.

On appelle *filtre monotone* de D un filtre $\mathfrak{F}$ croissant ou décroissant et $\mathfrak{P}(\mathfrak{F})$ est appelé *plage* de $\mathfrak{F}$.

Un filtre croissant $\mathfrak{F}$ et un filtre décroissant $\mathfrak{G}$ sont dit *complémentaires* si $\mathfrak{P}(\mathfrak{F}) \cup \mathfrak{P}(\mathfrak{G}) = D$ [6].

Soit D un fermé borné de K, de diamètre R et soit $a \in K$. On note $\tilde{D} = d(a, R)$; alors $\tilde{D} \setminus D$ admet une partition par une famille de disques non circonférenciés maximaux appelés *trous* de D [3].

Alors $\|\cdot\|$ *et* $\bar{\psi}$ *sont deux normes de K-algèbres équivalentes.*

On remarque que le théorème 1 s'applique en particulier au cas où les diamètres des trous de D sont minorés par une constante $\varrho>0$. En revanche, si l'on ne dispose d'aucune hypothèse de minoration sur les diamètres des trous, la conclusion du théorème 1 est fausse dans le cas général comme le montre le théorème 2, sur un exemple représentatif.

THÉORÈME 2. *Soit D un infraconnexe fermé borné admettant un T-filtre croissant* $\mathfrak{F}$ *et dont l'ensemble des trous est une suite* $(T_n)_{n\in \mathbf{N}}$ *telle que* $\lim_{n\to\infty} \operatorname{diam}(T_n)=0$. *Soit* ψ *la semi-norme multiplicative définie par* $\psi(f)=\lim_{\mathfrak{F}} |f(x)|$. *Alors la norme d'algèbre de Banach quotient de* $H(D)/\mathfrak{J}(\mathfrak{F})$ *définit une topologie strictement plus fine que celle de la valeur absolue quotient de* ψ *par* $\mathfrak{J}(\mathfrak{F})$.

Soit $R>0$. On note $\mathfrak{M}(R)$ l'ensemble des fonctions f définies dans des ensembles de la forme $d^-(0, R)\setminus\{b_1, \ldots, b_n, \ldots\}$ où $(b_n)_{n\in\mathbf{N}}$ est une suite telle que $|b_n|<R$, $\lim_{n\to\infty} |b_n|=R$, et telles que pour tout $r<R$ il existe $P_r(x)\in K[x]$ tel que $P_r(x)f(x)\in H(d(0, r))$ [4]. On peut alors choisir P_r tel que la fonction $f_r=P_r f$ ne s'annule pas sur les b_n tels que $|b_n|\leqq r$. Chaque zéro b_n d'ordre q_n de P_r est appelé *pôle d'ordre* q_n *de* f. (Il est clair que cette définition ne dépend pas de r quand $r>|b_n|$.) Alors on sait [4] que $\mathfrak{M}(R)$ est le corps de fractions de l'anneau $\mathfrak{a}(R)$ des séries de Taylor convergentes dans $d^-(0, R)$.

Soit D un infraconnexe fermé admettant un T-filtre $\mathfrak{F}$ croissant (resp. décroissant) de centre a, de diamètre R. Nous dirons qu'un élément $f\in H(D)$ est *méromorphe sur* $\mathfrak{F}$ *et admet b pour pôle d'ordre q*, si $g(x)=f(a+x)$ appartient à $\mathfrak{M}(R)$ $\left(\text{resp. } g(x)=f\left(a+\frac{1}{x}\right) \text{ appartient à } \mathfrak{M}\left(\frac{1}{R}\right)\right)$, et admet $b-a$ $\left(\text{resp. } \frac{1}{b-a}\right)$ pour pôle d'ordre q.

Soit D un fermé de K, non borné à complémentaire borné, et soit $H_b(D)$ l'ensemble des éléments bornés de $H(D)$. Il résulte de l'étude générale des ensembles $H(D)$ [3] que si T est un trou de D, alors l'inverse D' de D par une inversion centrée en un point $a\in T$ a une adhérence $\bar{D}'=D\cup\{0\}$ telle que $H(\bar{D}')$ soit isométriquement isomorphe à $H_b(D)$. On pourra donc appliquer aux algèbres de Banach $H_b(D)$ les propriétés des algèbres $H(D)$.

Soit D un infraconnexe fermé borné (resp. fermé à complémentaire borné). Si $\mathfrak{F}$ est un T-filtre de D, à plage non vide, on notera $\mathfrak{J}_0(\mathfrak{F})$ l'idéal des $f\in H(D)$ (resp. $f\in H_b(D)$) tels que $f(x)=0$ pour tout $x\in\mathfrak{P}(\mathfrak{F})$. On sait [4], [5] que $\mathfrak{J}_0(\mathfrak{F})\subset\subset\mathfrak{J}(\mathfrak{F})$.

Rappelons d'autre part qu'une T-suite $(T_{m,i}; q_{m,i})_{1\leqq i\leqq k_m}$ est dite *idempotente* [2] si $q_{m,i}=0$ ou 1 $\forall i=1, \ldots, k_m$, $\forall m\in\mathbf{N}$.

Le théorème 3 qui va suivre montre que si un infraconnexe D admet un T-filtre $\mathfrak{F}$ avec une T-suite idempotente et bien percée, toutes les classes modulo l'idéal $\mathfrak{J}_0(\mathfrak{F})$ contiennent des éléments méromorphes sur $\mathfrak{F}$. La démonstration fait appel à un procédé de calcul très spécifique, ainsi qu'aux procédés du théorème 1.

THÉORÈME 3. *Soit D un infraconnexe fermé contenant* $K\setminus d^-(0, R)$, *admettant*

Considérons une famille $(T_{m,i}; q_{m,i})_{\substack{1\le i\le k_m \\ m\in\mathbf{N}}}$ où $T_{m,i}$ est un trou de D de centre $b_{m,i}$, de diamètre $\varrho_{m,i}$ et où $q_{m,i}\in\mathbf{N}$.

La famille $(T_{m,i}; q_{m,i})_{\substack{1\le i\le k_m \\ m\in\mathbf{N}}}$ est appelée T-*suite associée au filtre croissant* (resp. *décroissant) de centre a, de diamètre R* si les trous $(T_{m,i})_{1\le i\le k_m}$ sont inclus dans un cercle $C(a; d_m)$ avec $d_m<d_{m+1}$ (resp. $d_m>d_{m+1}$) et vérifient la condition $(\mathfrak{T})$

$$(\mathfrak{T})\quad \begin{aligned}&\lim_{m\to\infty}\Big(\sum_{j=1}^{m}|\log d_m-\log d_j|\sum_{i=1}^{k_j}q_{j,i}\Big)-\max_{1\le i\le k_m}\big[q_{m,i}(\log d_m-\log \varrho_{m,i})+\\ &\quad+\sum_{\substack{1\le j\le k_m\\ j\ne i}}q_{m,j}(\log d_m-\log|b_{m,j}-b_{m,i}|)\big]\Big)=+\infty \quad [17].\end{aligned}$$

De même, nous dirons que la famille $(T_{m,i}, q_{m,i})$ est une T-suite associée à un filtre décroissant dépourvu de centre, si les $(T_{m,i})_{1\le i\le k_m}$ sont inclus dans le cercle $C(b_{m+1,1}, d_m)$ où $d_m=|b_{m+1,1}-b_{m,1}|$ et $d_{m+1}<d_m$ et si la suite $D_m=d(b_{m+1,1}, d_m)\cap D$ a une intersection vide, et si la famille $(T_{m,i}; q_{m,i})$ vérifie la condition $(\mathfrak{T})$.

Nous appellerons *percement* d'une T-suite $(T_{m,i}; q_{m,i})_{\substack{1\le i\le k_m \\ m\in\mathbf{N}}}$ le nombre $\varrho=\inf_{\substack{1\le i\le k_m \\ m\in\mathbf{N}}}(\operatorname{diam}(T_{m,i}))$ et nous dirons qu'elle est *bien percée* si $\varrho>0$.

Un filtre monotone est appelé T-*filtre* s'il existe une T-suite associée à $\mathfrak{F}$.

REMARQUE. On sait que cette définition est bien équivalente à la définition donnée dans [5] et elle sera ici plus maniable. Cette équivalence a notamment été montrée dans [17], Lemme 1.

Si $\mathfrak{F}$ est un T-filtre d'un infraconnexe fermé borné D on notera $\mathfrak{J}(\mathfrak{F})$ l'idéal (premier fermé) [4] des $f\in H(D)$ tels que $\lim_{\mathfrak{F}} f(x)=0$.

Nous dirons qu'un T-filtre est *bien percé* s'il admet une T-suite associée bien percée.

Les ensembles D tels que le complété $H(D)$ de l'algèbre $R(D)$ pour la topologie de la convergence uniforme soit une K-algèbre ont été caractérisés dans [2], [3].

Outre les fermés bornés (pour lesquels $H(D)$ est algèbre de Banach) nous savons que les fermés à complémentaires bornés appartiennent à cette classe.

Tout filtre monotone $\mathfrak{F}$ d'un infraconnexe D tel que $H(D)$ soit une algèbre définit sur $H(D)$ une semi-norme multiplicative $\varphi_{\mathfrak{F}}$ continue pour la topologie de $H(D)$, par la relation [4], [9]

$$\varphi_{\mathfrak{F}}(f)=\lim_{\mathfrak{F}}|f(x)|.$$

§ 3. Les principaux résultats

THÉORÈME 1. *Soit D un infraconnexe fermé borné admettant un T-filtre $\mathfrak{F}$ à plage vide bien percé, et soit A l'algèbre quotient $H(D)/\mathfrak{J}(\mathfrak{F})$. Soit $\|\cdot\|$ la norme définie sur A par le quotient de $\|\cdot\|_D$ par $\mathfrak{J}(\mathfrak{F})$. Soit ψ la semi-norme multiplicative définie sur $H(D)$ par $\psi(f)=\lim_{\mathfrak{F}}|f(x)|$ et soit $\bar\psi$ la valeur absolue définie sur A par le quotient de ψ par $\mathfrak{J}(\mathfrak{F})$.*

un T-filtre croissant $\mathfrak{F}$ *de centre* 0, *de diamètre R, avec un T-suite idempotente bien percée* $(T_{m,i};\, 1)_{\substack{1\le i\le k_m\\ m\in \mathbf{N}}}$ *et soient* $\beta_{m,i}\in T_{m,i}$ $(1\leqq i\leqq k_m,\ m\in\mathbf{N})$.

Pour tout $f\in H(D)$ *il existe* $F\in H(D)\cap\mathfrak{M}(R)$ *tel que chaque pôle de F soit simple et soit l un des points* $\beta_{m,i}$ $(1\leqq i\leqq k_m,\ m\in\mathbf{N})$, *et tel que* $F-f\in\mathfrak{J}_0(\mathfrak{F})$.

Remarque. Le procédé de calcul employé dans la démonstration du théorème 3 s'applique à des pôles simples et sa complexité ne rend pas immédiate une généralisation aux pôles multiples. Toutefois il est naturel de conjecturer l'existence d'un élément F méromorphe à pôles multiples lorsque la T-suite n'est plus supposée idempotente.

Le théorème 1 permet d'établir l'existence du prolongement analytique à travers un T-filtre bien percé. Le théorème 3 permet de préciser qu'il existe un prolongement méromorphe à pôles simples lorsque le T-filtre admet un T-suite idempotente.

Théorème 4. *Soit D un infraconnexe fermé borné (resp. un infraconnexe fermé à complementaire borné) admettant un T-filtre* $\mathfrak{F}$ *de diamètre R, tel que* $\mathfrak{P}(\mathfrak{F})\neq\emptyset$ *avec une T-suite bien percée* $(T_{m,i};\, q_{m,i})_{\substack{1\le i\le k_m\\ m\in \mathbf{N}}}$.

Soit $a\in D\setminus\mathfrak{P}(\mathfrak{F})$ *et soit* $b\in K$. *Alors tout élément* $f\in H_b(\mathfrak{P}(\mathfrak{F}))$ *se prolonge en un élément* $h\in H_b(D)$ *tel que* $h(a)=b$ *et l'ensemble des prolongements de f en éléments de* $H_b(D)$ *est égal à* $h+\mathfrak{J}_0(\mathfrak{F})$.

Théorème 5. *Soit D un infraconnexe fermé dont le complémentaire est borné. On suppose que D admet un T-filtre décroissant* $\mathfrak{F}$, *de diamètre R, tel que* $\mathfrak{P}(\mathfrak{F})\neq\emptyset$ *avec une T-suite bien percée* $(T_{m,i};\, q_{m,i})_{\substack{1\le i\le k_m\\ m\in \mathbf{N}}}$. *Soit* $l\in K$. *Alors tout élément* $f\in H(\mathfrak{P}(\mathfrak{F}))$ *se prolonge en un élément* $h\in H_b(D)$ *tel que* $\lim\limits_{|x|\to\infty} h(x)=l$ *et l'ensemble des prolongements de f en éléments de* $H(D)$ *est* $h+\mathfrak{J}_0(\mathfrak{F})$.

En outre si la T-suite $(T_{m,i};\, q_{m,i})$ *est idempotente et si* $b_{m,i}\in T_{m,i}$ $(1\leqq i\leqq k_m;\ m\in\mathbf{N})$, *on peut choisir h méromorphe sur* $\mathfrak{F}$ *avec pour seuls pôles des points* $b_{m,i}$, *pôles simples.*

Soit D un infraconnexe fermé admettant un T-filtre $\mathfrak{F}$ de diamètre R. Quand les diamètres des trous de D appartiennent tous à un certain intervalle $[\varrho_1, \varrho_2]\subset$ $\subset]0, R[$, on sait grâce au théorème 1 de [2] que $\mathfrak{F}$ admet des T-suites idempotentes (et bien percées par hypothèses). On obtient donc les corollaires suivants:

Corollaire 1. *Soit D un infraconnexe fermé borné, ou fermé à complémentaire borné, admettant un T-filtre* $\mathfrak{F}$ *de diamètre R, et tel que les diamètres des trous de D appartiennent à un intervalle* $[\varrho_1, \varrho_2]\subset]0, R[$. *Alors tout élément de* $H(\mathfrak{P}(\mathfrak{F}))$ *se prolonge en un élément h de* $H_b(D)$ *méromorphe sur* $\mathfrak{F}$, *dont tous les pôles sur* $\mathfrak{F}$ *sont simples. En outre si* $l\in K$ *et si* $a\in\mathfrak{P}(\mathfrak{F})$, *on peut imposer à h de vérifier* $h(a)=l$ $(a\in\mathfrak{P}(\mathfrak{F}))$ *(ou* $\lim\limits_{\substack{|x|\to\infty\\ x\in D}} h(x)=l$ *si* $\mathfrak{F}$ *est décroissant).*

Corollaire 2. *Soit* $f(x)$ *une série de Taylor convergente dans* $d(0, R)$, *soit* b_n *une suite de K telle que* $|b_n|>|b_{n+1}|$, $\lim\limits_{n\to\infty}|b_n|=R$, $\prod\limits_{n=1}^{\infty}\dfrac{R}{|b_n|}=0$ *et pour tout* $\varrho>0$

soit $A_\varrho = K \setminus \bigcup_{n=1}^{\infty} d^-(b_n, \varrho)$. *Alors f se prolonge en éléments analytiques sur* A_ϱ, *méromorphe dans K admettant chaque* b_n *pour pôle simple.*

La conjecture qui suit le théorème 3, conduit à conjecturer aux théorèmes 3 et 4 que l'hypothèse de la T-suite idempotente est superflue pour la dernière conclusion sur l'existence de prolongement analytique méromorphe.

Le théorème 1 permet aussi de connaître les quotients $H(D)/\mathfrak{J}_0(\mathfrak{F})$.

THÉORÈME 6. *Soit D un infraconnexe fermé borné admettant un T-filtre à plage non vide* $\mathfrak{F}$, *avec une T-suite bien percée.*

a) *L'application* $f \to \hat{f}$ *qui à chaque* $f \in H(D)$ *associe sa restriction* $\hat{f}$ *à* $\mathfrak{P}(\mathfrak{F})$ *est une surjection de* $H(D)$ *sur* $H(\mathfrak{P}(\mathfrak{F}))$.

b) *L'algèbre* $H(D)/\mathfrak{J}_0(\mathfrak{F})$ *est isomorphe à* $H(\mathfrak{P}(\mathfrak{F}))$ *algébriquement et topologiquement.*

Rappelons que deux filtres monotones $\mathfrak{F}_1$ et $\mathfrak{F}_2$ d'un infraconnexe D sont dits *complémentaires* si $\mathfrak{P}(\mathfrak{F}_1) \cup \mathfrak{P}(\mathfrak{F}_2) = D$ [6].

COROLLAIRE 3. *Soit D un infraconnexe fermé borné admettant un T-filtre à plage non vide* $\mathfrak{F}$ *avec une T-suite bien percée et aucun T-filtre complémentaire à* $\mathfrak{F}$. *Alors l'algèbre* $H(D)/\mathfrak{J}(\mathfrak{F})$ *est isomorphe à* $H(\mathfrak{P}(\mathfrak{F}))$, *algébriquement et topologiquement.*

THÉORÈME 7. *Soit D un infraconnexe fermé borné admettant un T-filtre* $\mathfrak{F}$ *à plage non vide. On suppose que l'ensemble des trous de D est une suite* T_n *telle que* $\lim_{n\to\infty} (\mathrm{diam}\,(T_n)) = 0$. *Alors* $H(\mathfrak{P}(\mathfrak{F}))$ *n'est pas isomorphe à* $H(D)/\mathfrak{J}_0(\mathfrak{F})$.

NOTATION. D'un point de vue beaucoup plus concret, le théorème 8 permet de construire de tels prolongements analytiques. Pour chaque série de Taylor $h(x) = \sum_{n=0}^{\infty} \alpha_n x^n$ soit $\bar{h}$ la forme linéaire définie sur $K[x]$ par $\bar{h}\left(\sum_{k=0}^{q} u_k x^k\right) = \sum_{k=0}^{q} \alpha_k u_k$.

THÉORÈME 8. *Soit une série de Taylor* $f(x) = \sum_{n=0}^{\infty} a_n x^n$ *telle que* $\lim_{n\to\infty} a_n = 0$ *et soit* $\varphi(x) = \sum_{n=0}^{\infty} \lambda_n x^n$ *une série de Taylor telle que* $\left|\frac{\lambda_n}{\lambda_{n+1}}\right| < \left|\frac{\lambda_{n+1}}{\lambda_{n+2}}\right|$, $\lim_{n\to\infty}\left|\frac{\lambda_n}{\lambda_{n+1}}\right| = 1$, $\lim_{n\to\infty} |\lambda_n| = +\infty$ *et* $\lim_{n\to\infty} \lambda_n a_n = 0$. *Alors* φ *converge dans* $d^-(0, 1)$ *et l'ensemble de ses zéros est une suite de zéros simples* $(\beta_i)_{i\in\mathbf{N}^*}$ *tels que* $|\beta_i| < |\beta_{i+1}|$; *on notera* $b_i = \frac{1}{\beta_i}$.

Pour tout $i \in \mathbf{N}^*$, *soit* $\psi_i(x) = \frac{\varphi(x)}{1-(x/\beta_i)}$, *et soit* $\varphi_i(x) = \frac{\psi_i(x)}{\psi_i(\beta_i)}$ *et soit* $\|\cdot\|_\varphi$ *la norme définie sur* $K[x]$ *par* $\left\|\sum_{n=0}^{q} u_n x^n\right\|_\varphi = \max_{0\le n\le q} |\lambda_n u_n|$. *Alors f appartient au complété E de* $K[x]$ *pour la norme* $\|\cdot\|_\varphi$, *chaque* $\bar{\varphi}_i$ *est continu pour cette norme et se prolonge à E et l'on a* $\lim_{i\to\infty} \bar{\varphi}_i(f) = 0$.

Pour tout $\varrho>0$, *soit* $\Lambda_\varrho=K\setminus\bigcup_{i=1}^{\infty} d^-(b_i,\varrho)$. *Alors la série* $f(x)$ *se prolonge en un élément* $\tilde{f}$ *de* $H(\Lambda_\varrho)$ *égal à*

$$\sum_{i=1}^{\infty}\frac{\bar{\varphi}_i(f)}{1-(x/b_i)}.$$

En outre Λ_ϱ *admet un T-filtre décroissant* $\mathfrak{F}$ *de centre* 0, *de diamètre* 1, *et pour tout* $h\in\mathfrak{I}(\mathfrak{F})$, $\tilde{f}+h$ *est un autre prolongement de* f *en un élément de* $H(\Lambda_\varrho)$.

Remarque. Concrètement, on voit que pour chaque $i\in\mathbf{N}^*$, la série $\sum_{n=0}^{\infty}\lambda_{i,n}a_n$ converge vers une limite ε_i et que $\lim_{i\to\infty}\varepsilon_i=0$; alors le prolongement $\tilde{f}$ est égal à:

$$\sum_{i=1}^{\infty}\frac{\varepsilon_i}{1-(x/b_i)}.$$

Le théorème 9 qui suit assure que l'hypothèse du théorème 8 n'est pas vide. Sa démonstration montre comment construire une suite $r_n=|\lambda_n|$ permettant de prolonger la série f considérée, en posant $u_n=|a_n|$.

Théorème 9. *Soit une suite* $(u_n)_{n\in\mathbf{N}}$ *de* $\mathbf{R}_+$ *telle que* $\lim_{n\to\infty}u_n=0$. *Il existe des suites* $(r_n)_{n\in\mathbf{N}}$ *de* $\mathbf{R}_+\cap|K|$ *telles que*

a) $(r_n)/(r_{n+1})<(r_{n+1})/(r_{n+2})$,

b) $\lim_{n\to\infty}\dfrac{r_n}{r_{n+1}}=1$,

c) $\lim_{n\to\infty}r_n=+\infty$,

d) $\lim_{n\to\infty}r_nu_n=0$.

Corollaire 4. *Pour toute série de Taylor* $\sum_{n=0}^{\infty}a_nx^n$ *telle que* $\lim_{n\to\infty}a_n=0$, *il existe des séries de Taylor* $\sum_{n=0}^{\infty}\lambda_nx^n$ *satisfaisant l'hypothèse du théorème 8.*

II. Résultats préliminaires

Définition. Soit une T-suite $(T_{m,i};q_{m,i})_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ et pour tout $m\in\mathbf{N}$, soit d_m le diamètre du plus petit disque contenant tous les trous $(T_{m,i})_{1\le i\le k_m}$.

Nous appellerons *suite monotone de la* T-*suite*, la suite $(d_m)_{m\in\mathbf{N}}$.

Pour tout couple $(m,i)_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$, soit $\alpha_{m,i}\in T_{m,i}$ et soit $\varrho_{m,i}=\operatorname{diam}(T_{m,i})$ et soit

$$\gamma_m=\max_{1\le i\le k_m}\left(\left(\frac{d_m}{\varrho_{m,i}}\right)^{q_{m,i}}\prod_{\substack{1\le j\le k_m\\ j\ne i}}\left(\frac{d_m}{|\alpha_{m,j}-\alpha_{m,i}|}\right)^{q_{m,j}}\right).$$

Nous appellerons *affaissement de la T-suite* le nombre

$$\sup_{m\in\mathbf{N}}\left[\log\gamma_m-\left|\log\left(\prod_{j=1}^{m}\left(\frac{d_j}{d_m}\right)^{q_j}\right)\right|\right],\quad \text{où}\quad q_m=\sum_{i=1}^{k_m}q_{m,i}.$$

Lemme 1. *Soit D un infraconnexe fermé borné de diamètre R et soit $(T_{m,i};\, q_{m,i})_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ une T-suite de D de percement $\varrho>0$. Soit $A>0$.*

Alors D admet une T-suite $(T_{m,i};\, b_{m,i})_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ d'affaissement $\lambda\leqq A+2\log\left(\frac{R}{\varrho}\right)$.

Preuve. Soit μ l'affaissement de la T-suite $(T_{m,i};\, q_{m,i})$. Pour tout couple $(m,i)_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ soit $\alpha_{m,i}\in T_{m,i}$ et $\varrho_{m,i}=\operatorname{diam}(T_{m,i})$.

Soit $(d_m)_{m\in\mathbf{N}}$ la suite monotone de la T-suite et pour tout $m\in\mathbf{N}$ soit $q_m=\sum_{i=1}^{k_m}q_{m,i}$ et soit

$$\gamma_m=\sup_{1\le i\le k_m}\left(\frac{d_m}{\varrho_{m,i}}\right)^{q_{m,i}}\prod_{\substack{j\ne i\\ 1\le j\le k_m}}\left(\frac{d_m}{|\alpha_{m,i}-\alpha_{m,j}|}\right)^{q_{m,j}}$$

et on a donc

(1)
$$\lim_{m\to\infty}\left(\log(\gamma_m)-\left|\log\prod_{j=1}^{m}\left(\frac{d_j}{d_m}\right)^{q_j}\right|\right)=-\infty$$

et

(2)
$$\max\left(\log\gamma_m-\left|\log\prod_{j=1}^{m}\left(\frac{d_j}{d_m}\right)^{q_j}\right|\right)=\mu.$$

Pour tout couple $(m,i)_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ soit $t_{m,i}=\operatorname{Int}\left(\frac{A}{\mu}q_{m,i}\right)$ et soit $u_{m,i}=\frac{A}{\mu}q_{m,i}-t_{m,i}$.

Alors d'après le Lemme 2 de [17], on sait qu'il existe des entiers $(v_{m,i})_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ égaux à 0 ou 1 tels que

(3)
$$0\leqq\sum_{i=1}^{k_m}v_{m,i}-\sum_{i=1}^{k_m}u_{m,i}<1$$

et

(4)
$$\sup_{1\le i\le k_m}\left(\sum_{\substack{j\ne i\\ 1\le j\le k_m}}v_{m,j}\log\left(\frac{d_m}{|\alpha_{m,j}-\alpha_{m,i}|}\right)\right)\leqq$$
$$\leqq\sup_{1\le i\le k_m}\left(\sum_{\substack{j\ne i\\ 1\le j\le k_m}}u_{m,j}\log\left(\frac{d_m}{|\alpha_{m,j}-\alpha_{m,i}|}\right)\right)+\log\left(\frac{d_m}{\varrho}\right).$$

Posons $b_{m,i}=t_{m,i}+v_{m,i}$ $(1\leqq i\leqq k_m;\, m\in\mathbf{N})$ et pour tout $m\in\mathbf{N}$ soit $b_m=\sum_{i=1}^{k_m}b_{m,i}$ et

$$\gamma'_m=\sup_{1\le i\le k_m}\left[\left(\frac{d_m}{\varrho_{m,i}}\right)^{b_{m,i}}\prod_{\substack{1\le j\le k_m\\ j\ne i}}\left(\frac{d_m}{|\alpha_{m,j}-\alpha_{m,i}|}\right)^{b_{m,j}}\right].$$

On voit donc que

$$\log \gamma'_m = \max_{1 \leq i \leq k_m} \left((t_{m,i} + v_{m,i}) \log \left(\frac{d_m}{\varrho_{m,i}} \right) + \sum_{\substack{1 \leq j \leq k_m \\ j \neq i}} (t_{m,j} + v_{m,j}) \log \frac{d_m}{|\alpha_{m,j} - \alpha_{m,i}|} \right)$$

d'où grâce à (4), et au fait que

$$(t_{m,i} + v_{m,i}) \log \left(\frac{d_m}{\varrho_{m,i}} \right) \leqq (t_{m,i} + u_{m,i}) \log \left(\frac{d_m}{\varrho_{m,i}} \right) + \log \left(\frac{d_m}{\varrho} \right)$$

on obtient

$$\log \gamma'_m \leqq$$

$$\leqq \sup_{1 \leq i \leq k_m} \left[(t_{m,i} + u_{m,i}) \log \left(\frac{d_m}{\varrho_{m,i}} \right) + \sum_{\substack{1 \leq j \leq k_m \\ j \neq i}} (t_{m,j} + u_{m,j}) \log \left(\frac{d_m}{|\alpha_{m,j} - \alpha_{m,i}|} \right) + 2 \log \left(\frac{d_m}{\varrho} \right) \right]$$

et par suite

$$\log \gamma'_m \leqq \frac{A}{\mu} \log \gamma_m + 2 \log \frac{d_m}{\varrho}. \tag{5}$$

Alors il est clair que la suite $(T_{m,i}; b_{m,i})$ est une T-suite d'affaissement $\leqq A + 2 \log \dfrac{R}{\varrho}$.

En effet d'après (3) on a

$$b_m \geqq \sum_{i=1}^{k_m} t_{m,i} + \sum_{i=1}^{k_m} u_{m,i} = \frac{A}{\mu} q_m$$

et par suite

$$\sum_{j=1}^{m} b_j \left| \log \frac{d_m}{d_j} \right| \geqq \frac{A}{\mu} \sum_{j=1}^{m} q_j \left| \log \frac{d_m}{d_j} \right|,$$

d'où, d'après (5), on a:

$$\log \gamma'_m - \sum_{j=1}^{m} b_j \left| \log \left(\frac{d_j}{d_m} \right) \right| \leqq \frac{A}{\mu} \left(\log \gamma_m - \sum_{j=1}^{m} q_j \left| \log \left(\frac{d_j}{d_m} \right) \right| \right) + 2 \log \left(\frac{R}{\varrho} \right). \tag{6}$$

D'après (1), on voit donc que

$$\lim_{m \to \infty} \left(\log (\gamma_m) - \left| \log \left(\frac{d_j}{d_m} \right)^{b_j} \right| \right) = -\infty$$

donc la suite $(T_{m,i}; b_{m,i})$ est une T-suite.

Enfin, par hypothèse $\log \gamma_m + \sum_{j=1}^{m} q_j \left| \log \left(\dfrac{d_j}{d_m} \right) \right| \leqq \mu$, donc (6) montre que l'affaissement de cette T-suite est majoré par $A + 2 \log \left(\dfrac{R}{\varrho} \right)$.

Le Lemme 2 qui suit est immédiat:

LEMME 2. *Soient* $\alpha_1, \dots, \alpha_q \in d(0, 1)\setminus\{0\}$ *et soit* $Q(x) = \prod_{j=1}^{q}\left(1-\frac{x}{\alpha_j}\right) = \sum_{j=0}^{q} a_j x^j$. *Soit* $m = \min_{1\leq j\leq q} |\alpha_j|$. *Alors* $|a_j| \leqq \frac{1}{m^j}$.

Soit $\mathfrak{F}$ un filtre décroissant d'un infraconnexe D. On appelle *base canonique* de $\mathfrak{F}$ une base de $\mathfrak{F}$ de la forme $D_m = D\cap(\Delta_m\setminus(\bigcap_{j=1}^{\infty}\Delta_j))$ où la suite Δ_m est une suite décroissante de disques.

PROPOSITION 1. *Soit D un infraconnexe fermé borné, admettant un T-filtre $\mathfrak{F}$ croissant (resp. décroissant) de diamètre R, de centre a (resp. de base canonique D_m) avec une T-suite* $(T_{m,i}; q_{m,i})_{\substack{1\leq i\leq k_m\\ m\in\mathbf{N}}}$ *de percement $\varrho>0$. Soit $\varepsilon>0$ (resp. soit $Q(\varepsilon)\in\mathbf{N}$ tel que* $\operatorname{diam}(D_{Q(\varepsilon)}) < R(1+\varepsilon)$*).*

Il existe $\varphi_\varepsilon\in\mathfrak{J}_0(\mathfrak{F})$, méromorphe sur $\mathfrak{F}$, tel que

a) $|\varphi_\varepsilon(x)-1|\leqq\varepsilon$ *pour tout* $x\in D\cap d(a; R(1-\varepsilon))$ *(resp.* $|\varphi_\varepsilon(x)-1|\leqq\varepsilon$ *pour tout* $x\in D\setminus(D_{Q(\varepsilon)}\cup\mathfrak{P}(\mathfrak{F}))$*),*

b) $\|\varphi_\varepsilon\|_D \leqq \left(\varepsilon+\frac{1}{\varrho^2}\right)R^2$,

c) *les pôles de φ_ε sont des points $b_{m,i}\in T_{m,i}$ d'ordre $u_{m,i}\leqq q_{m,i}$* $1\leqq i\leqq k_m$, $m\in\mathbf{N}$.

PREUVE. Pour construire φ_ε, on va reprendre dans ses grandes lignes la démonstration du lemme I.6.A de [5] (prouvant que tout T-filtre admet des éléments strictement annulés). Toutefois on doit imposer ici des contraintes supplémentaires très fortes dues à a) et b), qui se répercutent sur toute la récurrence, ce qui interdit de se reporter simplement à la démonstration de ce lemme de [5].

Construisons donc $\varphi_\varepsilon\in\mathfrak{J}_0(\mathfrak{F})$, satisfaisant a), b), c). Il est immédiat de se ramener au cas où $R=1$ et on le supposera donc. Soit $a>0$ tel que $2\log\left(\frac{1}{\varrho}\right)+a\leqq$ $\leqq\log\left(\frac{1}{\varrho^2}+\varepsilon\right)$;

Soit $(T_{m,i}; q_{m,i})_{\substack{1\leq i\leq k_m\\ m\in\mathbf{N}}}$ une T-suite associée à $\mathfrak{F}$, de suite monotone $(d_m)_{m\in\mathbf{N}}$ $\left(\text{resp. } (d'_m)_{m\in\mathbf{N}} \text{ et soit } d_m=\frac{1}{d'_m}\right)$.

Fixons $\varepsilon>0$. Soit $L\in\mathbf{N}$ tel que $d_L\geqq 1-\varepsilon$ et soit $w(0)\in\mathbf{N}$ vérifiant

$$(\mathscr{U}_0)\qquad \left(\frac{d_L}{d_{L+1}}\right)^{w(0)}\leqq\varepsilon.$$

Soit $q=L+w(0)$.

On voit que la famille $(T_{m,i}; q_{m,i})_{\substack{1\leq i\leq k_m\\ m>q}}$ est encore une T-suite associée à $\mathfrak{F}$. Soit θ son affaissement. Alors d'après le lemme 1, on sait qu'il existe une T-suite de la forme $(T_{m,i}; u_{m,i})_{\substack{1\leq i\leq k_m\\ m>q}}$ d'affaissement $\lambda\leqq 2\log\left(\frac{1}{\varrho}\right)+a$. On notera $u_m=$ $=\sum_{i=1}^{k_m} u_{m,i}$ $(m>q)$.

Pour tout couple $(m, i)_{\substack{1\leqq i\leqq k_m\\ m\in \mathbf{N}}}$, soit $\alpha_{m,i}\in T_{m,i}$ et soit $\varrho_{m,i}=\operatorname{diam}(T_{m,i})$. Pour tout $m\geqq \mathbf{N}$, notons

$$\gamma_m = \sup_{1\leqq i\leqq k_m}\left(\frac{d_m}{\varrho_{m,i}}\right)^{u_{m,i}} \prod_{\substack{j\neq i\\ 1\leqq i\leqq k_m}}\left(\frac{d_m}{|\alpha_{m,i}-\alpha_{m,j}|}\right)^{u_{m,j}}$$

$$\left(\text{resp. } \gamma_m = \sup_{1\leqq i\leqq k_m}\left(\frac{d'_m}{\varrho_{m,i}}\right)^{u_{m,i}} \prod_{\substack{j\neq i\\ 1\leqq j\leqq k_m}}\left(\frac{d'_m}{|\alpha_{m,i}-\alpha_{m,j}|}\right)^{u_{m,j}}\right).$$

Alors on voit que

$$\lim_{m\to\infty}\gamma_m\prod_{j=N}^{m}\left(\frac{d_j}{d_m}\right)^{u_j}=0 \tag{1}$$

et puisque $\lambda\leqq 2\log\left(\frac{1}{\varrho}\right)+a$ on a

$$\log\left(\gamma_m\prod_{j=N}^{m}\left(\frac{d_j}{d_m}\right)^{u_j}\right)\leqq\lambda\leqq 2\log\left(\frac{1}{\varrho}\right)+a\leqq\log\left(\frac{1}{\varrho^2}+\varepsilon\right). \tag{2}$$

Nous allons construire par récurrence des suites d'entiers $s(n)$, $l(n)$, $w(n)$ $(n\in\mathbf{N})$ vérifiant

$$\left(\frac{d_{l(n)}}{d_{s(n)}}\right)^{w(n)}<\frac{\varepsilon}{n+1} \tag{$\mathcal{U}_n$}$$

$$\frac{1}{d_{s(n-1)}^{w(n-1)}\,d_{s(n)}^{w(n)}}\prod_{j=s(n-1)+1}^{m}\left(\frac{d_j}{d_m}\right)^{u_j}\gamma_m<\frac{\varepsilon}{n+1} \tag{$\mathcal{V}_n$}$$

pour tout $m\geqq s(n)$.

Supposons en effet déjà définies les trois suites jusqu'au rang n.

Alors on construit facilement les trois suites au rang $n+1$. On choisit d'abord $l(n+1)$ tel que

$$\gamma_m\prod_{j=l(n)}^{m}\left(\frac{d_j}{d_m}\right)^{u_j}<\frac{\varepsilon}{n+1}\quad\text{pour tout}\quad m\geqq l(n+1).$$

On peut choisir ensuite $w(n+1)$ tel que $\left(\frac{d_{l(n+1)}}{d_{l(n+1)+1}}\right)^{w(n+1)}<\frac{\varepsilon}{n+2}$. Alors d'après (1) on peut choisir $s(n+1)>\max\big(l(n+1), s(n)+1+w(n)\big)$

$$\gamma_m\frac{1}{d_{s(n)}^{w(n)}\,d_{s(n+1)}^{w(n+1)}}\prod_{j=s(n)+1}^{m}\left(\frac{d_j}{d_m}\right)^{u_j}<\frac{\varepsilon}{n+2}\quad\text{pour}\quad m\geqq s(n+1). \tag{$\mathcal{V}_n$}$$

Il reste donc à construire $s(1)$, $l(1)$, $w(1)$, $s(2)$, $l(2)$, $w(2)$ satisfaisant $(\mathcal{U}_2)$ et $(\mathcal{V}_2)$. Choisissons d'abord $l(1)>q$ tel que

$$\gamma_m\prod_{j=q+1}\left(\frac{d_j}{d_m}\right)^{u_j}<\frac{\varepsilon}{2}\quad\text{pour tout}\quad m\geqq l(1).$$

On peut alors choisir $w(1)$ tel que $\left(\frac{d_{l(1)}}{d_{l(1)+1}}\right)^{w(1)} < \frac{\varepsilon}{2}$. On choisit enfin $s(1) > l(1)$ vérifiant

$$(\mathscr{W}) \qquad \frac{1}{d_{s(1)}^{w(1)}} \prod_{j=q+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j} < \frac{\varepsilon}{2} \quad \text{pour tout } m \geqq s(1).$$

Il est alors immédiat de choisir $l(2)$ vérifiant $(\mathfrak{T}_1)$ puis $w(2)$ vérifiant $(\mathscr{U}_2)$ et enfin $s(2) > l(2)$ vérifiant $(\mathscr{V}_2)$. Les suites $s(n)$, $l(n)$, $w(n)$ sont ainsi définies pour tout n.

On va maintenant supposer $\mathfrak{F}$ croissant et on peut naturellement supposer $a=0$.

Pour tout $m \in \mathbf{N}$ soit $Q_m = \prod_{i=1}^{k_m} \left(1 - \frac{x}{\alpha_{m,i}}\right)^{u_{m,i}}$.

Soit $H_n(x) = \prod_{m=s(n)+1}^{s(n+1)} Q_m$ et soit $t(n) = \deg(H_n)$.

Développons $H_n(x) = \sum_{h=0}^{t(n)} a_{n,h} x^h$ avec $a_{n,0} = 1$ et notons

$$G_n(x) = \sum_{h=0}^{w(n)} a_{n,h} x^b \quad \text{pour} \quad n \in \mathbf{N}^*.$$

Pour tout $n \in \mathbf{N}$, soit $R_n(x) = \frac{G_n(x)}{H_n(x)}$, soit $P(x) = \prod_{m=L+1}^{q} \left(1 - \frac{x}{\alpha_{m,1}}\right)$ et soit $H_0(x) = P(x) \prod_{m=q+1}^{s(1)} Q_m(x)$. On peut développer $H_0(x)$ sous la forme $\sum_{h(0)}^{t(0)} a_{0,h} x^h$.

Soit $G_0(x) = \sum_{h=0}^{w(0)} a_{0,h} x^h$ et soit $R_0(x) = \frac{G_0(x)}{H_0(x)}$. Remarquons que d'après le Lemme 2 on voit que l'on a

$$(\mathscr{K}_n) \qquad \|G_n\|_D \leqq \frac{1}{d_{s(n)}^{w(n)}}, \quad \text{pour tout} \quad n \in \mathbf{N}^*.$$

Remarquons d'abord que

$$(\mathscr{Q}_0) \qquad |R_0(x) - 1| \leqq \varepsilon \quad \text{pour tout} \quad x \in D \quad \text{tel que} \quad |x| \leqq d_L$$

$$(\mathscr{Q}_n) \qquad |R_n(x) - 1| \leqq \frac{\varepsilon}{n+1} \quad \text{pour tout} \quad x \in D \cap d(0, d_{l(n)}).$$

En effet, on voit que

$$|R_0(x) - 1| = \left| \frac{\sum_{h=w(0)+1}^{t(0)} a_{0,h} x^h}{H_0(x)} \right|.$$

Or $|H_0(x)| \geqq 1$ pour tout $x \in D \cap (d(0, d_L))$ car $H_0(x)$ n'a aucun zéro dans $d(0, D_L)$.

En outre d'après le Lemme 2 on voit que

$$\Big|\sum_{h=w(0)+1}^{t(0)} a_{0,h}x^h\Big| \leqq \max_{w(0)+1\leqq h\leqq t(0)} \left(\frac{|x|^h}{d_{L+1}^h}\right) \leqq \max_{w(0)+1\leqq h\leqq t(0)} \left(\frac{d_L}{d_{L+1}}\right)^h = \left(\frac{d_L}{d_{L+1}}\right) w(0)+1$$

d'où finalement $|R_0(x)-1|\leqq\varepsilon$ d'après $(\mathscr{U}_0)$.

On montre de la même façon $(\mathscr{Q}_n)$. En effet $H_n(x)$ n'a aucun zéro dans $d^-(0, d_{s(n)+1})$; alors $|H_n(x)|=1$ pour $|x|\leqq d_{l(n)}$ et d'après le Lemme 2

$$|G_n(x)-H_n(x)| \leqq \left(\frac{d_{l(n)}}{d_{s(n)+1}}\right)^{w(n)} \leqq \left(\frac{d_{l(n)}}{d_{l(n)+1}}\right)^{w(n)} \frac{\varepsilon}{n+1} \quad \text{(d'après } (\mathscr{U}_n)).$$

Les relations $(\mathscr{Q}_n)$ étant établies nous allons montrer maintenant les relations

$$(\mathscr{R}_{n,h}) \qquad |\pi_n(x)| \leqq \frac{\varepsilon}{k+1} \quad \text{pour} \quad x\in D\setminus d(0, d_{s(k)}), \quad 1\leqq k\leqq n.$$

Pour cela supposons déjà établies les relations $(\mathscr{R}_{n,k})$ lorsque $n\leqq N$ et montrons les pour $n=N+1$.

Il est clair que les zéros de H_{N+1} et G_{N+1} n'appartiennent pas à $d(0, d_{s(N+1)})$ d'où

$$|R_{N+1}(x)| = 1 \quad \text{pour} \quad |x| \leqq d_{s(N+1)}.$$

Puisque la relation $(\mathscr{R}_{N,k})$ est supposée satisfaite on voit donc qu'elle entraîne directement $(\mathscr{R}_{N+1,k})$ pour $k\leqq N$. Il reste donc à établir $(\mathscr{R}_{N+1,N+1})$.

Supposons d'abord $x\notin d(0, d_{s(N+2)})$. On voit donc que

$$|H_{N+1}(x)| = \prod_{j=s(N+1)+1}^{s(N+2)} \left(\frac{|x|}{d_j}\right)^{u_j} \geqq \prod_{j=s(N+1)+1}^{s(N+2)} \left(\frac{d_{s(N+2)}}{d_j}\right)^{u_j}$$

et par suite d'après $(\mathscr{K}_{N+1})$ on obtient

$$|R_{N+1}(x)| \leqq \frac{1}{d_{s(N+1)}^{w(N+1)}} \prod_{j=s(N+1)+1}^{s(N+2)} \left(\frac{d_j}{d_{s(N+2)}}\right)^{u_j},$$

d'où $|R_{N+1}(x)|\leqq\dfrac{\varepsilon}{N+2}$ d'après (2), et donc a fortiori $|\pi_{N+1}(x)|\leqq\dfrac{\varepsilon}{N+2}$ d'après $(\mathscr{R}_{N,N})$.

Supposons maintenant $d_{s(N+1)}\leqq|x|<d_{s(N+2)+1}$. Par exemple, supposons $d_m\leqq$ $\leqq|x|<d_{m+1}$ pour $s(N+1)\leqq m<s(N+2)$, on voit que

$$\frac{1}{|H_{N+1}(x)|} \leqq \gamma_m \prod_{j=s(N+1)+1}^{m} \left(\frac{d_j}{|x|}\right)^{u_j},$$

$$|R_{N+1}(x)| \leqq \gamma_m \frac{1}{d_{s(N+1)}^{w(N+1)}} \prod_{j=s(N+1)+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j}.$$

Alors

$$|R_N(x)\,R_{N+1}(x)| \leqq \frac{\prod\limits_{j=s(N)+1}^{s(N+1)} \left(\dfrac{d_j}{d_m}\right)^{u_j}}{d_{s(N)}^{w(N)}}\,\gamma_m\,\frac{\prod\limits_{j=s(N+1)+1}^{m} \left(\dfrac{d_j}{d_m}\right)^{u_j}}{d_{s(N+1)}^{w(N+1)}} =$$

$$= \gamma_m \frac{1}{d_{s(N)}^{w(N)}\, d_{s(N+1)}^{w(N+1)}} \prod_{j=s(N)+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j}$$

et ceci est $\leqq \dfrac{\varepsilon}{N+2}$ d'après $(\mathscr{V}_{N+1})$ puisque $m > s(N+1)$. Alors puisque $|\pi_{N-1}(x)| < 1$, on voit que la relation $|\pi_{N+1}(x)| < \dfrac{\varepsilon}{N+2}$ est vraie pour tout $x \in D \setminus d^-(0, d_{s(N+1)})$ ce qui achève d'établir $(\mathscr{R}_{N+1,N+1})$.

Il reste donc seulement à établir $(\mathscr{R}_{1,1})$, pour amorcer la récurrence, c'est-à-dire

$$|R_0(x) R_1(x)| \leqq \frac{\varepsilon}{2} \quad \text{pour} \quad x \in D \setminus d(0, d_{s(1)})$$

ou encore

$$\left|\frac{G_0(x)}{P(x)}\right| \left|\frac{1}{\left(\prod\limits_{m=q+1}^{s(1)} Q_m(x)\right)}\right| \left|\frac{G_1(x)}{\prod\limits_{m=s(1)+1}^{s(2)} Q_m(x)}\right| \leqq \frac{\varepsilon}{2}$$

quand $|x| \geqq d_{s(1)}$, $x \in D$.

Comme G_0 a des coefficients égaux aux coefficients de H_0 dont l'indice va de 0 à $w(0)$, on voit que G_0 a lui aussi un zéro unique sur chaque cercle $C(0, d_m)$ pour $L+1 \leqq m \leqq q$ et aucun autre zéro dans K.

Alors il est immédiat de voir que

$$\left|\frac{G_0(x)}{P(x)}\right| \leqq \frac{1}{\varrho} \quad \text{pour tout} \quad x \in D \tag{3}$$

et plus particulièrement

$$\left|\frac{G_0(x)}{P(x)}\right| = 1 \quad \text{pour tout} \quad x \in D \setminus d(0, d_q) \tag{4}$$

d'où

$$|R_0(x) R_1(x)| = \frac{|G_1(x)|}{\left|\prod\limits_{m=q+1}^{s(2)} Q_m(x)\right|} \quad \text{pour} \quad |x| > d_q.$$

Si $x \notin d(0, d_{s(2)})$ on voit que

$$\frac{1}{\left|\prod\limits_{m=q+1}^{s(2)} Q_m(x)\right|} \leqq \prod_{m=q+1}^{s(2)} \left(\frac{d_j}{d_{s(2)}}\right)^{u_j}$$

d'où

$$\frac{G_1(x)}{\left|\prod_{m=q+1}^{s(2)} Q_m(x)\right|} \leqq \frac{1}{d_{s(1)}^{w(1)}} \prod_{j=q+1}^{s(2)} \left(\frac{d_j}{d_m}\right)^{u_j}.$$

Si $d_m \leqq |x| < d_{m+1}$ avec $s(1) \leqq m \leqq s(2)$ on voit que

$$\frac{1}{\left|\prod_{m=q+1}^{s(2)} Q_m(x)\right|} \leqq \left(\prod_{j=q+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j}\right) \gamma_m$$

d'où

$$\frac{|G_1(x)|}{\left|\prod_{m=q+1}^{s(2)} Q_m(x)\right|} \leqq \frac{\gamma_m}{d_{s(1)}^{w(1)}} \prod_{j=q+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j}$$

et cette dernière expression est majorée par $\frac{\varepsilon}{2}$ d'après $(\mathscr{W})$.

On a donc établi que $|R_0(x) R_1(x)| \leqq \frac{\varepsilon}{2}$ pour tout $x \in D \setminus d(0, d_{s(1)})$ ce qui donne $(\mathscr{R}_{1,1})$ et amorce la récurrence sur les relations $(\mathscr{R}_{n,k})$.

Notons maintenant $\pi_n(x) = \prod_{j=0}^{n} R_j(x)$.

Grâce aux relations $(\mathscr{Q}_n)$ et $(\mathscr{R}_{n,k})$ il est immédiat de voir que la suite π_n converge dans $H(D)$ vers un élément φ_ε vérifiant

(5) $$|\varphi_\varepsilon(x) - R_0(x)| \leqq \varepsilon \quad \text{pour} \quad x \in D \cap d(0, d_L)$$

et $(\mathscr{G}_n)$ $|\varphi_\varepsilon(x)| = |\pi_n(x)|$ quand $x \in D \cap d(0, d_{s(n)})$.

D'après $\mathscr{Q}_0$ et (5) on obtient la relation

(6) $$|\varphi_\varepsilon(x) - 1| \leqq \varepsilon \quad \text{quand} \quad x \in d(0, d_L).$$

Il reste donc seulement à montrer que $\|\varphi_\varepsilon\|_D \leqq \varepsilon + \frac{1}{\varrho^2}$.

D'après les relations $(\mathscr{G}_n)$ et $(\mathscr{R}_{n,k})$ vraies pour $n \geqq 1$, il est clair que $|\varphi_\varepsilon(x)| < 1$ quand $|x| > d_{s(1)}$ et il reste donc à montrer que

$$|\varphi_\varepsilon(x)| \leqq \varepsilon + \frac{1}{\varrho^2} \quad \text{quand} \quad x \in D \cap d(0, d_{s(1)})$$

c'est-à-dire, en tenant compte de (3) et de (6), quand $d_q \leqq |x| \leqq d_{s(1)}$. Alors on voit que $|R_n(x)| = 1$ pour tout $n \in \mathbf{N}^*$ et d'après (4) on voit que $|R_0(x)| = \frac{1}{\left|\prod_{m=q+1}^{s(1)} Q_m(x)\right|}$.

Supposons par exemple $d_m \leqq |x| < d_{m+1}$ avec $q+1 \leqq m \leqq s(1)$. Alors

$$|R_0(x)| \leqq \gamma_m \prod_{j=q+1}^{m} \left(\frac{d_j}{d_m}\right)^{u_j}$$

et d'après (2) on voit que $|R_0(x)|\leqq\frac{1}{\varrho^2}+\varepsilon$, ce qui achève de montrer que $\|\varphi_\varepsilon\|_D\leqq$ $\leqq\frac{1}{\varrho^2}+\varepsilon$. La proposition est donc établie quand $\mathfrak{F}$ est croissant.

Supposons maintenant $\mathfrak{F}$ décroissant (peut-être dépourvu de centre). Fixons $n\in\mathbf{N}$, soit $\omega_n\in T_{s(n+1),1}$, soit $\mathfrak{J}(x)=\frac{1}{x-\omega_n}$ et soit $D'=\mathfrak{J}(D)$. On voit que D' admet une suite finie de trous $(T'_{m,i})_{\substack{1\leqq i\leqq k_m\\ 1\leqq m<s(n+1)}}$ où $T'_{m,i}$ est inclus dans $C(0, d'_m)$. Alors la construction effectuée dans le cas croissant permet de définir $R_n(y)\in R(D')$ dont les pôles appartiennent aux trous $(T'_{m,i})_{\substack{1\leqq i\leqq k_m\\ s(n)\leqq m<s(n+1)}}$.

On peut donc noter $\bar{R}_n(x)=R_n(\mathfrak{J}(x))$ et $\pi_n(x)=\prod_{j=1}^{n}\bar{R}_j(x)$. Pour établir que la suite π_n converge vers un élément φ_ε de $\mathfrak{J}(\mathfrak{F})$, il suffit de reprendre les majorations obtenues (grâce au cas croissant) pour les R_j en remarquant que le changement d'origine de ω_n à ω_{n+1} ne change rien à ces majorations pour tout $j\leqq n$.

En notant $Q(\varepsilon)$ l'entier L défini ci-dessus, on voit que $|\varphi_\varepsilon(x)-1|\leqq\varepsilon$ quand $|x-\omega_n|>1+\varepsilon$, c'est-à-dire quand $x\notin D_{Q(\varepsilon)}\cup\mathfrak{P}(\mathfrak{F})$.

III. Démonstration des théorèmes 1 et 2

Théorème 1. *Soit D un infraconnexe fermé borné admettant un T-filtre $\mathfrak{F}$ bien percé, à plage vide, et soit A le corps quotient $\frac{H(D)}{\mathfrak{J}(\mathfrak{F})}$. Soit $\|\cdot\|$ la norme quotient de $\|\cdot\|_D$ par $\mathfrak{J}(\mathfrak{F})$.*

Soit ψ la semi-norme multiplicative définie sur $H(D)$ par $\psi(f)=\lim_{\mathfrak{F}}|f(x)|$ et soit $\bar{\psi}$ la valeur absolue définie sur A par le quotient de ψ par $\mathfrak{J}(\mathfrak{F})$.

Alors $\|\cdot\|$ et $\bar{\psi}$ sont deux normes de K-algèbres équivalentes.

Preuve. Soit θ la surjection canonique de $H(D)$ sur A et pour tout $f\in H(D)$ soit $\bar{f}=\theta(f)$.

L'inégalité $\bar{\psi}(u)\leqq\|u\|$ est classique et immédiate. En effet, il est d'abord clair que pour tout $f\in H(D)$ on a $\psi(f)\leqq\|f\|_D$, car $\psi(f)=\lim_{\mathfrak{F}}|f(x)|$, et $|f(x)|\leqq$ $\leqq\|f\|_D$ pour tout $x\in D$. Alors soit $\varepsilon>0$; il existe $f\in H(D)$ tel que $\bar{f}=u$ et $\|f\|_D\leqq$ $\leqq\|u\|+\varepsilon$. On a donc $\bar{\psi}(u)\leqq\|u\|+\varepsilon$. Ceci étant vrai pour tout $\varepsilon>0$, on a bien $\bar{\psi}(u)\leqq\|u\|$.

Réciproquement, montrons qu'il existe $M\in\mathbf{R}_+$ tel que $\|u\|\leqq M\bar{\psi}(u)$ pour tout $u\in A$.

Supposons $\mathfrak{F}$ croissant et soit $f\in H(D)$. Nous allons montrer que pour tout $\varepsilon>0$ il existe $\varphi\in\mathfrak{J}(\mathfrak{F})$ tel que $\|f-\varphi\|_D\leqq(\psi(f)+\varepsilon)\left(\frac{R}{\varrho}\right)^2$ ce qui montrera bien l'équivalence des deux normes $\|\cdot\|$ et $\bar{\psi}$.

Pour cela fixons $\varepsilon\in\left]0, \frac{\psi(f)}{\|f\|_D}\right[$ et soit $\eta\in]0,\varepsilon[$ tel que $|f(x)|\leqq\psi(f)+\varepsilon$ quand $x\in D\setminus d(0, R(1-\eta)^-)$.

D'après la proposition 1 il existe $\varphi_\eta \in \mathfrak{J}(\mathfrak{F})$ tel que $|\varphi_\eta(x)-1| \leqq \eta$ pour $|x| \leqq$ $\leqq R(1-\eta)$ et $\|\varphi_\eta\|_D \leqq 1+\frac{R^2}{\varrho^2}$. Alors $\|\bar{f}\| \leqq \|f(1-\varphi_\eta)\|_D$.

Or, on voit que $|f(x)(1-\varphi_\eta(x))| \leqq \eta |f(x)|$ quand $|x| \leqq R(1-\eta)$ et par suite $|f(x)(1-\varphi_\eta(x))| \leqq \varepsilon |f(x)| < \psi(f)$ d'après le choix de ε.

D'autre part, si $|x| \geqq R-\eta$, on sait que $|f(x)| \leqq \psi(f)+\varepsilon$ et comme $\|\varphi_\eta\|_D \leqq$ $\leqq 1+\left(\frac{R}{\varrho}\right)^2$. On voit que $|f(x)(1-\varphi_\eta(x))| \leqq (\psi(f)+\varepsilon)\left[1+\left(\frac{R}{\varrho}\right)^2\right]$. On a donc bien $\|\bar{f}\| \leqq (\bar{\psi}(\bar{f})+\varepsilon)\left[1+\left(\frac{R}{\varrho}\right)^2\right]$ et ceci est vrai pour tout $\varepsilon>0$, d'où $\|\bar{f}\| \leqq$ $\leqq \bar{\psi}(\bar{f})\left[1+\left(\frac{R}{\varrho}\right)^2\right]$.

Si $\mathfrak{F}$ est décroissant centré, on peut aisément se ramener au cas croissant par inversion. Si $\mathfrak{F}$ est décroissant dépourvu de centre, on doit considérer une base canonique D_m de $\mathfrak{F}$ et pour tout $\varepsilon>0$, on choisit d'abord un rang $L(\varepsilon)$ tel que $|f(x)| \leqq \bar{\psi}(f)+\varepsilon$ quand $x \in D_{L(\varepsilon)}$. Grâce à la proposition 1 il existe $\varphi_\varepsilon \in \mathfrak{J}(\mathfrak{F})$ tel que $\|\varphi_\varepsilon\|_D \leqq 1+\left(\frac{R}{\varrho}\right)^2$, d'où une conclusion analogue.

On va voir sur un exemple très représentatif, que le théorème 1 ne peut pas être généralisé de façon significative.

Théorème 2. *Soit D un infraconnexe fermé borné admettant un T-filtre $\mathfrak{F}$ à plage vide, et dont l'ensemble des trous est une suite $(T_n)_{n\in\mathbf{N}}$ telle que* $\lim_{n\to\infty} \operatorname{diam}(T_n)=0$. *Soit ψ la semi-norme multiplicative définie par $\psi(f)=\lim_{\mathfrak{F}} |f(x)|$. Alors la norme d'algèbre de Banach quotient de $\frac{H(D)}{\mathfrak{J}(\mathfrak{F})}$ définit une topologie strictement plus fine que celle de la valeur absolue quotient de ψ par $\mathfrak{J}(\mathfrak{F})$.*

Preuve. On peut naturellement supposer que $0 \in T_0$ et que $R=1$, et que la suite d_n des distances $d(T_0, T_n)$ vérifie $d_n \leqq d_{n+1}$ pour tout $n \in \mathbf{N}$.

Notons $\varrho_n = \operatorname{diam}(T_n)$ et $u_n = \frac{1}{x^n} \in R(D)$ et $\mathfrak{M} = \mathfrak{J}(\mathfrak{F})$.

Soit $\chi = \frac{H(D)}{\mathfrak{M}}$, soit θ la surjection canonique de $H(D)$ sur χ et pour tout $f \in H(D)$, soit $\bar{f} = \theta(f)$.

Supposons que les topologies définies par la norme quotient $\|\cdot\|$ de celle de $H(D)$ et par la valeur absolue $\bar{\psi}$ quotient de ψ soient équivalentes.

Il existe donc $M>1$ tel que

$$\|\bar{u}_n\| \leqq M\bar{\psi}(\bar{u}_n) \quad \text{pour tout} \quad n \in \mathbf{N}.$$

En fait, il est clair que $\psi(u_n)=1$ pour tout $n \in \mathbf{N}$ et on doit donc avoir $\|u_n\| \leqq M$ pour tout n. Cela signifie que pour tout $n \in \mathbf{N}$ et pour tout $\varepsilon>0$, il existe $\varphi \in \mathfrak{M}$ tel que $\|u_n+\varphi\|_D \leqq M+\varepsilon$.

On va chercher un entier n tel que, pour tout $\varphi \in \mathfrak{M}$, on ait $\|u_n+\varphi\|_D \geqq 2M$.

Il existe naturellement un rang q tel que tous les trous de D inclus dans $K \setminus d^-(0, d_q)$ aient un diamètre $\leqq \frac{1}{2M}$.

Soit $N \geqq q$ tel que $\frac{1}{d_q^N} \geqq 2M$. Quand $|x| < d_N$, on voit que

$$|u_N(x)| \geqq \frac{1}{d_N^N} \geqq 2M.$$

Supposons qu'il existe $\varphi \in \mathfrak{M}$ tel que $\|u_N + \varphi\|_D < 2M$. On a donc $|\varphi(x)| = |u_N(x)|$ pour $|x| \leqq d_N$, et par suite on a:

(1) $$v(u_N, \mu) = v(\varphi, \mu) = -N\mu \quad \text{pour} \quad \mu \geqq -\log d_N.$$

Puisque φ est strictement annulé par $\mathfrak{F}$ on sait qu'il existe $N' \geqq N$ tel que

$$v_d'(\varphi, -\log d_{N'}) > v_g'(\varphi, -\log d_{N'}).$$

Soit L le plus petit entier $\geqq N$ tel que

(2) $$v_d'(\varphi, -\log d_L) > v_g'(\varphi, -\log d_L).$$

On voit que la fonction $\mu \to v(\varphi, \mu)$ est concave dans l'intervalle $[-\log d_L, -\log d_N]$ et par conséquent, d'après (1) on obtient

$$v(\varphi, \mu) \leqq v(u_N, \mu) \quad \text{pour} \quad \mu \geqq -\log d_L$$

et en particulier

$$v(\varphi, -\log d_L) \leqq -N \log d_L.$$

Soit $\tau = C(0, d_L)$. Alors d'après les propriétés classiques des éléments analytiques [5] on sait que (2) implique

(3) $$\|\varphi\|_{\tau \cap D} \geqq \Big(\lim_{\substack{|x| \to d_L \\ |x| \neq d_L}} |\varphi(x)|\Big) \frac{d_L}{\gamma} \geqq \frac{1}{d_L^N} \frac{d_L}{\gamma}$$

où γ désigne la borne supérieure des diamètres des trous de D inclus dans le cercle τ. Alors d'après l'hypothèse faite sur N, on sait que $\frac{d_L}{\gamma} \geqq 2M$, et donc $\|\varphi\|_{\tau \cap D} \geqq 2M$. Alors on voit que

$$\|u_N\|_{\tau \cap D} = \|\varphi\|_{\tau \cap D} = \frac{1}{d_L^N} \quad \text{d'où} \quad \lim_{\substack{|x| \to d_L \\ |x| \neq d_L}} |\varphi(x)| < \frac{1}{d_L^N}$$

c'est-à-dire $v(\varphi, -\log d_L) > -N \log d_L$ ce qui contredit (3) et montre que $\|u_N\| \geqq 2M$. La non équivalence des deux normes est donc établie.

IV. Démonstration du théorème 3

LEMME 3. *Soit D un infraconnexe fermé contenant $K\setminus d^-(0,1)$, admettant un trou $T=d^-(0,r)$, et admettant un T-filtre croissant bien percé, $\mathfrak{F}$ de centre 0, de diamètre R avec une T-suite bien percée idempotente $(T_{m,i};1)_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ de percement ϱ et pour tout couple $(m,i)_{\substack{1\le i\le k_m\\ m\in\mathbf{N}}}$ soit $b_{m,i}\in T_{m,i}$. On suppose $|b_{m,i}|\geqq \geqq r>0$ $\forall m$, $\forall i$.*

Soit $\varepsilon\in]0,1[$ soit $\Phi\in\mathfrak{M}(1)\cap\mathfrak{I}(\mathfrak{F})$ tel que chaque pôle de Φ soit simple et soit l'un des points $(b_{m,i})_{\substack{1\le i\le k_m\\ M\in\mathbf{N}}}$ et tel que

$$|\Phi(x)-1|\leqq\varepsilon \quad \textit{pour} \quad |x|\leqq r \quad \textit{et que} \quad \Phi(0)=1.$$

Soit $f(x)=\sum_{k=0}^{\infty}\varepsilon_k/x^k$ une série convergente pour $|x|\geqq r$. Alors il existe $F\in H(D)\cap\mathfrak{M}(1)$ tel que chaque pôle de F soit simple et soit l'un des points $(b_{m,i})_{1\le i\le k_m}$, et tel que $f(x)=F(x)$ pour tout $x\in\mathfrak{P}(\mathfrak{F})$ et $\|F\|_D\leqq\|\Phi\|_D\|f\|_D$.

PREUVE. On peut naturellement réindexer les $(b_{m,i})$ sous la forme $n\to b_n$ $(n\in\mathbf{N})$ où $|b_n|\leqq|b_{n+1}|$, et l'on a donc $|b_0|\geqq r$ sur l'infraconnexe D, Φ admet une décomposition de Mittag—Lefflerienne

$$\sum_{n=0}^{\infty}\frac{\lambda_n}{x-b_n}, \quad \text{avec} \quad \lim_{n\to+\infty}|\lambda_n|=0. \tag{1}$$

Posons maintenant $\varphi_k(x)=(\Phi(x))/(x^k)$; on voit que $\varphi_k\in\mathfrak{M}(1)\cap\mathfrak{I}_0(\mathfrak{F})$ et la décomposition Mittag—Lefflérienne de φ_k sur D est de la forme:

$$(\alpha_{k,k}/x^k)+\dots+(\alpha_{k,1}/x)+\sum_{n=0}^{\infty}\lambda_n/(b_n^k(x-b_n)=(\Phi(x))/x^k). \tag{2}$$

Soit $\sum_{k=0}^{\infty}\Phi_k x^k$ le développement de Taylor de Φ (avec $\Phi_0=1$) dans le disque $d^-(0,r)$. Puisque $|\Phi(x)-1|\leqq\varepsilon$ $\forall x\in d(0,r)$ il est clair que $|\Phi_k|r^k\leqq\varepsilon$ $\forall k\geqq 1$ d'où

$$|\Phi_k|\leqq\frac{1}{r^k} \quad \forall k\in\mathbf{N}. \tag{3}$$

Alors, pour $x\in d^-(0,\omega)$, on voit que:

$$\sum_{k=0}^{\infty}\Phi_k x^k=\sum_{n=0}^{\infty}\lambda_n/(x-b_n)=-\sum_{n=0}^{\infty}\lambda_n/(b_n(1-(x/b_n)))$$

$$\sum_{k=0}^{\infty}\Phi_k x^k=-\sum_{k=0}^{\infty}\Big(\sum_{n=0}^{\infty}\lambda_n/(b_n^{k+1})\Big)x^k.$$

D'où, par identification dans $d^-(0,\omega)$:

$$\alpha_{k,i}=\Phi_{k-i}=-\sum_{n=0}^{\infty}\lambda_n/(b_n^{k+1-i}) \quad \text{pour} \quad 1\leqq i\leqq k. \tag{4}$$
$$k\in\mathbf{N}^*$$

Remarquons maintenant, en appliquant la décomposition Mittag—Lefflérienne à tous les φ_s $(1\leqq s\leqq k)$ comme en (2), que les $(1/x^s)$ vérifient, $\forall x\in D, \forall s=1,\dots,k$, le système:

$$(\alpha_{1,1})/x = (\Phi(x))/x - \sum_{n=0}^{\infty} (\lambda_n)/(b_n(x-b_n))$$

$$(5) \qquad (\alpha_{2,1})/x + (\alpha_{2,2})/(x^2) = (\Phi(x)/(x^2)) - \sum_{n=0}^{\infty} (\lambda_n)/(b_n^2(x-b_n))$$

$$\vdots$$

$$(\alpha_{k,1})/x + (\alpha_{k,2})/(x^2) + \ldots + (\alpha_{k,k})/(x^k) = (\Phi(x)/x^k) - \sum_{n=0}^{\infty} (\lambda_n)/(b_n^k(x-b_n)).$$

Posons:

$$\theta_i(x) = (\Phi(x))/(x^i) - \sum_{h=0}^{\infty} (\lambda_n)/(b_n^i(x-b_n))$$

et considérons les matrices:

$$A_k = \begin{pmatrix} \alpha_{1,1}, & 0, \ldots, & 0 \\ \alpha_{2,1}, & \alpha_{2,2}, \ldots, & 0 \\ \vdots & & \\ \alpha_{k,1}, & \ldots\ldots, & \alpha_{k,k} \end{pmatrix}$$

$$X_k(x) = \begin{pmatrix} 1/x \\ 1/(x^2) \\ \vdots \\ 1/(x^k) \end{pmatrix} \qquad \Theta_k(x) = \begin{pmatrix} \theta_1(x) \\ \theta_2(x) \\ \vdots \\ \theta_k(x) \end{pmatrix}.$$

Alors on peut écrire (5) sous forme matricielle:

$$\Theta_k = A_k X_k.$$

Remarquons, que pour tout $k \in \mathbf{N}$, $\alpha_{k,k} = \Phi(0) = 1$ et que d'après (3) pour $0 \leqq i \leqq k-1$, $k \in \mathbf{N}^*$, on a:

$$(6) \qquad |\alpha_{k,i}| = |\Phi_{k-i}| \leqq \frac{1}{r^{k-i}}.$$

Appelons $(a_{1,k}; a_{2,k}; \ldots; a_{k,k})$ la dernière ligne de la matrice $(A_k)^{-1}$. En inversant (5), nous avons:

$$1/(x^k) = \sum_{i=1}^{k} a_{i,k}\theta_i(x) \qquad \forall k \in \mathbf{N}^*, \ \forall x \in D$$

ou encore, en remplaçant chaque θ_i par sa valeur:

$$1/(x^k) = \Phi(x)\Big[\sum_{i=1}^{k} (a_{i,k})/(x^i)\Big] - \sum_{n=0}^{\infty} \lambda_n\big((a_{1,k}/b_n) + \ldots + (a_{k,k}/b_n^k)\big)/(x-b_n).$$

Posons:

$$(7) \qquad g_k(x) = \sum_{i=1}^{k} a_{i,k}/(x^i) \qquad \forall x \in D, \ \forall k \in \mathbf{N}^*$$

$$(8) \qquad l_k(x) = \sum_{n=0}^{\infty} \lambda_n g_k(b_n)/(x-b_n) \qquad \forall x \in D, \ \forall k \in \mathbf{N}^*.$$

Alors:

$$1/(x^k) = \Phi(x)g_k(x) - l_k(x). \tag{9}$$

Nous allons maintenant essayer de majorer $\|g_k\|_D$. Fixons k. Chaque $a_{i,k}$ est de la forme:

$$a_{i,k} = (-1)^{k-i}\delta_{k-i} \quad \text{si} \quad 1 \leqq i \leqq k, \quad \text{avec} \quad a_{k,k} = 1 \tag{10}$$

δ_{k-i} étant le cofacteur correspondant de la matrice A_k:

$$\delta_j = \begin{vmatrix}
\alpha_{1,1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \ddots & & & & \\
\alpha_{k-j-1,1} & \cdots & \alpha_{k-j-1,k-j-1} & 0 & \cdots & \cdots & 0 \\
\alpha_{k-j+1,1} & \cdots & \cdots & \alpha_{k-j+1,k-j} & \alpha_{k-j+1,k-j+1} & 0 \cdots & 0 \\
\vdots & & & & \ddots & & \\
\alpha_{k,1} & \cdots & \cdots & \cdots & \alpha_{k,k-j+1} & \cdots & \alpha_{k,k}
\end{vmatrix}.$$

En développant par rapport à la 1[ière] ligne et en tenant compte du fait que $\alpha_{ii} = 1 \ \forall i$, on obtient encore:

$$\delta_j = \begin{vmatrix}
\alpha_{2,2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & & & & & \\
\alpha_{k-j-1,2} & \cdots & \alpha_{k-j-1,k-j-1} & 0 & \cdots & \cdots & 0 \\
& & & \ddots & & & \\
\alpha_{k-j+1,2} & \cdots & \cdots & \alpha_{k-j+1,k-j} & \alpha_{k-j+1,k-j+1} & 0 \cdots & 0 \\
\vdots & & & \ddots & \ddots & & \\
\alpha_{k-1,2} & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha_{k-1,k-1} \\
& & & & & & \ddots \\
\alpha_{k,2} & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha_{k,k-1}
\end{vmatrix}$$

d'où par une récurrence immédiate, après $k-j-1$ opérations,

$$\delta_j = \begin{vmatrix} \alpha_{k-j+1,k-j} & \alpha_{k-j+1,k-j+1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \ddots \\ \alpha_{k-1,k-j} & \cdots & \cdots & \cdots & \alpha_{k-1,k-1} \\ \alpha_{k,k-j} & \cdots & \cdots & \cdots & \alpha_{k,k-1} \end{vmatrix}.$$

Et finalement d'après (6) on voit que

$$(11) \qquad \delta_j = \begin{vmatrix} \Phi_1 & 1 & 0 & 0 \cdots & 0 \\ \Phi_2 & \Phi_1 & 1 & 0 & 0 \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ \Phi_{j-1} & & & & \Phi_1 \end{vmatrix}$$

(pour $2 \leqq j \leqq k$, sachant que $\delta_1 = 1$).

Majorons maintenant ces δ_j:

$$\delta_1 = 1$$

$$\delta_2 = \Phi_1^2 - \Phi_2.$$

D'après (10) on voit que:

$$\delta_2 \leqq \sup(|\Phi_1^2|, |\Phi_2|) \leqq \sup[(1/r)^2, 1/(r^2)] \leqq 1/(r^2).$$

Nous allons montrer par récurrence la relation

$$(12) \qquad |\delta_s| \leqq 1/(r^s) \qquad \forall s \in \mathbf{N}^*.$$

Supposons la relation établie pour $s \leqq q-1$ et étudions δ_q: nous développons δ_q par rapport à sa première ligne:

$$\delta_q = \Phi_1 \delta_{q-1} - \begin{vmatrix} \Phi_2 & 1 & 0 & \cdots & 0 \\ \Phi_3 & \Phi_1 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 1 \\ \Phi_{q-1} & \Phi_{q-2} & \cdots & \cdots & \Phi_1 \end{vmatrix}.$$

De même, en développant les $q-1$ déterminants successifs par rapport à leur première ligne:

$$\delta_q = \sum_{s=1}^{q-1} (-1)^{s-1} \Phi_s \delta_{q-s}$$

et par suite $|\delta_q| \leqq \sup_{1 \leqq s \leqq q-1} |\Phi_s||\delta_{q-s}|$.

Supposons vraie jusqu'au rang $q-1$ la relation (12). Alors grâce à (6) on voit que

$$|\delta_q| \leqq \max_{1 \leqq s \leqq q-1} \big(1/(r^s)\big)\big(1/(r^{q-s})\big) = \frac{1}{r^q}. \tag{13}$$

Nous pouvons donc majorer les $a_{i,k}$, d'après (10) et (13):

$$|a_{i,k}| \leqq 1/(r^{k-i}) \quad \text{si} \quad 1 \leqq i \leqq k, \quad k \in \mathbf{N}^*.$$

Maintenant, nous pouvons majorer $g_k(x)$; grâce à (7) et (13)

$$\|g_k(x)\|_D = \Big\| \sum_{i=1}^{k} a_{i,k}/(x^i) \Big\|_D \leqq \max_{1 \leqq i \leqq k} \big(1/(r^{k-i})\big)\big(1/(r^i)\big) = 1/(r^k) \qquad \forall k \in \mathbf{N}^*. \tag{14}$$

Considérons alors une fonction $f \in H(D)$ de la forme:

$$f(x) = \sum_{k=0}^{\infty} \varepsilon_k/(x^k),$$

série convergente pour $|x| \geqq r$.

Alors, $\forall x \in D$, nous avons d'après (9)

$$f(x) = \sum_{k=0}^{\infty} \varepsilon_k [\Phi(x) g_k(x) - l_k(x)],$$

donc

$$f(x) = \Phi(x) \sum_{k=0}^{\infty} \varepsilon_k g_k(x) - \sum_{k=0}^{\infty} \varepsilon_k l_k(x).$$

Mais $\sum_{k=0}^{\infty} \varepsilon_k g_k(x)$ est une série de Laurent bornée et convergente sur D (rappelons que $f(x)$ est une série de Laurent convergente pour $|x| \geqq r$, donc que $\lim_{k \to +\infty} \varepsilon_k/(r^k) = 0$, c'est-à-dire d'après (14) que $\lim_{k \to +\infty} \varepsilon_k \|g_k(x)\|_D = 0$). Soit A sa borne supérieure:

$$\Big|\Phi(x) \sum_{k=0}^{\infty} \varepsilon_k g_k(x)\Big| < A |\Phi(x)| \qquad \forall x \in D.$$

Donc

$$\Phi(x) \sum_{k=0}^{\infty} \varepsilon_k g_k(x) \in \mathfrak{J}_0(\mathfrak{F}) \cap H(D).$$

Alors, pour tout $x \in \mathfrak{P}(\mathfrak{F})$ on a:

$$f(x) = -\sum_{k=0}^{\infty} \varepsilon_k l_k(x).$$

Posons:

$$F(x) = -\sum_{k=0}^{\infty} \varepsilon_k l_k(x). \tag{15}$$

Nous allons montrer que $\|F\|_D \leqq \|f\|_D \|\Phi\|_D$. En effet, d'après la définition (8) de l_k on voit que:

$$\|l_K\|_D \leqq \sup_{n\in\mathbf{N}} |\lambda_n|\, |g_k(b_n)|\, \|1/(x-b_n)\|_D$$

donc, d'après (14):

$$\|l_K\|_D \leqq 1/(r^k) \sup_{n\in\mathbf{N}} |\lambda_n|\, \|1/(x-b_n)\|_D.$$

Or, d'après la décomposition Mittag—Lefflérienne de Φ (1), on sait que

$$\|\Phi\|_D = \sup_{n\in\mathbf{N}} |\lambda_n|\, \|1/(x-b_n)\|_D$$

d'où

$$\|l_k\|_D \leqq \big(1/(r^k)\big) \|\Phi\|_D.$$

Mais, d'après la définition de F (15) on voit que:

$$\|F\|_D \leqq \sup_{k\in\mathbf{N}} |\varepsilon_k|/(r^k) \|\Phi\|_D,$$

d'où finalement, puisque $\|f\|_D = \sup_{k\in\mathbf{N}^*} |\varepsilon_k|/(r^k)$, on obtient bien:

$$\|F\|_D \leqq \|f\|_D \|\Phi\|_D.$$

THÉORÈME 3. *Soit D un infraconnexe fermé contenant $K \setminus d^-(0, R)$, admettant un T-filtre croissant $\mathfrak{F}$ de centre 0, de diamètre R avec une T-suite idempotente bien percée $(T_{m,i}; 1)_{\substack{1\leqq i\leqq k_m \\ m\in\mathbf{N}}}$ et soient $b_{m,i} \in T_{m,i}$ $(1 \leqq i \leqq k_m,\ m\in\mathbf{N})$.*

Pour tout $f\in H(D)$, il existe $F\in H(D)\cap\mathfrak{M}(R)$ tel que chaque pôle de F soit simple et soit l'un des points $b_{m,i}$ $(1\leqq i\leqq k_m;\ m\in\mathbf{N})$, et tel que $f(x)=F(x)$ pour tout $x\in\mathfrak{P}(\mathfrak{F})$.

PREUVE. Il est immédiat de se ramener au cas où $R=1$, ce que nous supposerons donc. D'autre part, on peut se ramener au cas où $\lim_{|x|\to\infty} f(x)=0$. En effet, on sait [2, 3] que f se décompose sous la forme $P(x)+f_0(x)$ où $P\in K[x]$ et $f_0\in H(D)$ et $\lim_{|x|\to\infty} f_0(x)=0$. On supposera donc $\lim_{|x|\to\infty} f(x)=0$.

Considérons la décomposition Mittag—Lefflérienne de f sur D. Soit $(\tau_n)_{n\in\mathbf{N}}$ la suite des trous de D qui sont associés à une partie singulière de f, et pour tout n soit $\tau_n' = K\setminus\tau_n$. Alors f s'écrit de façon unique sous la forme $\sum_{n=1}^{\infty} f_n$ où $f_n\in H(\tau_n')$ et $\lim_{|x|\to+\infty} f_n(x)=0$.

Pour tout $n\in\mathbf{N}$, soit $\tau_n = d^-(\alpha_n, r_n)$, soit $\varepsilon\in]0, 1[$ et soit ϱ le percement de la T-suite.

Alors grâce à la proposition 1, il existe donc $\Phi_n \in \mathfrak{J}_0(\mathfrak{F}) \cap \mathfrak{M}(1)$ tel que les pôles de Φ_n soient simples et soient des points $(\beta_{m,i})_{\substack{1 \le i \le k_m \\ m \in \mathbf{N}}}$, et que $|\Phi_n(x) - 1| \le \varepsilon$ pour $|x - a_n| \le r_n$ et $\Phi_n(\alpha_n) = 1$ et que

$$\|\Phi_n\|_D \le \left(\varepsilon + \frac{1}{\varrho^2}\right).$$

Soit $r > 0$ tel que $|b_{m,i}| \ge r \ \forall i, \ \forall m \in \mathbf{N}$. Alors grâce au lemme 3 il existe $F_n \in H(D) \cap \mathfrak{M}(1)$ tel que $f_n(x) = F_n(x)$ pour tout $x \in \mathfrak{P}(\mathfrak{J})$ et tel que F_n n'ait que des pôles simples et que chaque pôle soit l'un des points $b_{m,i}$ et que

$$\|F_n\|_D \le \|f_n\|_D \|\Phi_n\|_D \le \|f_n\|_D \left(\varepsilon + \frac{1}{\varrho^2}\right).$$

Par hypothèse la suite $\|f_n\|_D$ tend vers 0. La série $\sum_{n=1}^{\infty} F_n$ est donc convergente dans $H(D)$ et sa limite F vérifie $F(x) = f(x)$ pour tout $x \in \mathfrak{P}(\mathfrak{F})$, appartient à $\mathfrak{M}(1)$, n'a que des pôles simples, chacun étant l'un des points $b_{m,i}$. La démonstration est donc achevée.

V. Démonstration des théorèmes 4, 5, 6, 7

Rappelons qu'on appelle *filtre circulaire de centre a, de diamètre R*, le filtre $\mathfrak{C}$ qui admet pour système générateur la famille des couronnes $\Gamma(\alpha, r', r'')$ où $\alpha \in d(a, R)$ et $r' < R < r''$, et que $\mathfrak{C} \cap d(a, r)$ est un filtre croissant de $d(a, r)$.

Alors on sait [9] que chaque filtre circulaire $\mathfrak{C}$ de K définit sur $K(x)$ une valeur absolue $\psi_{\mathfrak{C}}$ par $\psi_{\mathfrak{C}}(h) = \lim_{\mathfrak{C}} |h(x)|$.

En outre, si $\mathfrak{C}$ est un filtre circulaire de K sécant à un infraconnexe D, alors $\mathfrak{C} \cap D$ est un filtre $\mathfrak{C}'$ qui définit une semi-norme multiplicative sur $H(D)$, continue pour la topologie de $H(D)$ [4, 9], $\psi_{\mathfrak{C}'}(f) = \lim_{\mathfrak{C}'} |f(x)|$.

PROPOSITION 2. *Soit $R > 0$, soit D un infraconnexe fermé borné contenant $d(0, R)$ et admettant un T-filtre $\mathfrak{F}$ décroissant bien percé de centre 0 et de diamètre R et soit* $A = \dfrac{H(D)}{\mathfrak{J}(\mathfrak{F})}$.

Alors la valeur absolue $\bar{\psi}_{\mathfrak{F}}$ quotient de la semi-norme multiplicative $\psi_{\mathfrak{F}}$ définie par $\psi_{\mathfrak{F}}(f) = \lim_{\mathfrak{F}} |f(x)|$ définit sur A une topologie équivalente à celle de la norme $\|\cdot\|$ d'algèbre de Banach quotient de $H(D)$.

PREUVE. On a toujours $\bar{\psi}_{\mathfrak{F}} \le \|u\|$ pour tout $u \in A$. Soit $B = d(0, R)$ et soit $D' = D \setminus B$.

Pour tout $f \in H(D)$ soit $\bar{f}$ son image dans A par la surjection canonique de $H(D)$ sur A.

D'après le théorème 1 il existe M tel que

(1) $\quad \|f\|_{D'} \le M\psi_{\mathfrak{F}}(f)$ pour tout $f \in H(D')$ et donc pour tout $f \in H(D)$.

Soit $\mathfrak{G}$ le filtre croissant de centre 0, de diamètre R, soit $\mathfrak{C}$ le filtre circulaire de centre 0 et de diamètre R et soit $\mathfrak{C}'=\mathfrak{C}\cap D$. Alors

(2) $$\psi_{\mathfrak{C}'}(f)=\psi_{\mathfrak{G}}(f)=\psi_{\mathfrak{F}}(f) \quad \text{pour tout} \quad f\in H(D) \quad [4,9].$$

Mais puisque $B\subset D$ on sait que $\psi_{\mathfrak{G}}(f)=\|f\|_B$, d'où d'après (1) et (2) on voit que $\|f\|_D=\max(\|f\|_{D'}, \|f\|_B)\leqq\max(\psi_{\mathfrak{G}}(f), M\psi_{\mathfrak{F}}(f))\leqq M\psi_{\mathfrak{F}}(f)$, ce qui achève la démonstration.

Comme corollaire il est immédiat de déduire par inversion la proposition 3.

PROPOSITION 3. *Soit* $R>0$, *soit* D *un infraconnexe contenant* $K\setminus d^-(0, R)$ *et admettant un T-filtre bien percé croissant de centre* 0, *de diamètre* R. *Soit* $A=(H_b(D))/\mathfrak{J}(\mathfrak{F})$. *Alors la valeur absolue quotient de la semi-norme multiplicative* $\psi_{\mathfrak{F}}$ *associée à* $\mathfrak{F}$ *définit sur* A *une topologie équivalente à celle de la norme d'algèbre de Banach quotient.*

PROPOSITION 4. *Soit* $R>0$, *soit* $D=K\setminus d^-(0, R)$ *et soit* D' *un infraconnexe inclu dans* $d^-(0, R)$, *admettant une T-suite croissante bien percée de centre* 0, *de diamètre* R. *Alors pour tout* $f\in H_b(D)$ *il existe* $\tilde{f}\subset H(D\cup D')$ *tel que* $f(x)=\tilde{f}(x)$ $\forall x\in D$.

PREUVE. Soit $\mathfrak{F}$ le T-filtre de D' de centre 0, de diamètre R. On peut naturellement se ramener au cas où D' admet un trou de centre 0 puisque $d^-(0, R)$ contient des trous de D'. Alors x est inversible dans $H(D'\cup D)$. Soit $A=\dfrac{H(D'\cup D)}{\mathfrak{J}(\mathfrak{F})}$, soit ω la surjection canonique de $H(D'\cup D)$ sur A, et soit $|\cdot|$ la valeur absolue de A, quotient de la semi-norme multiplicative $\psi_{\mathfrak{F}}$ définie sur $H(D')$ par $\mathfrak{F}$.

Soit $f\in H(D)$; on sait que $f(x)$ est de la forme $\sum_{n=0}^{\infty}\frac{a_n}{x^n}$. Soit $\bar{x}=\omega(x)$. Alors $\left|\frac{1}{\bar{x}}\right|=\frac{1}{R}$ de sorte que la série $\sum_{n=0}^{\infty}\frac{a_n}{\bar{x}^n}$ converge dans A puisque A est complète pour $|\cdot|$ d'après la proposition 3. 2

Soit $h\in H(D'\cup D)$ tel que $\omega(h)=\sum_{n=0}^{\infty}\frac{a_n}{\bar{x}^n}$, soit $\varepsilon>0$. Soit $N(\varepsilon)$ tel que $\frac{|a_n|}{R^n}\leqq\varepsilon$ pour $n\geqq N(\varepsilon)$ et soit $f_n=\sum_{i=0}^{n}\frac{a_i}{x^i}\in R(D'\cup D)$ pour $n\geqq N(\varepsilon)$.

On sait que $\left\|\frac{a_i}{x^i}\right\|_D=\frac{|a_i|}{R^i}$ et par suite

(1) $$\|f_n-f\|_D\leqq\varepsilon \quad \text{pour} \quad n\geqq N(\varepsilon).$$

D'autre part

(2) $$|\omega(f_n-h)|=\left|\sum_{i=n+1}^{\infty}\frac{a_i}{\bar{x}^i}\right|\leqq\sup_{i\geqq n+1}\left(\frac{|a_i|}{R^i}\right)\leqq\varepsilon.$$

Or par définition $|\omega(f_n-h)|=\lim_{\mathfrak{F}}|f_n(x)-h(x)|$.

Soit $\mathfrak{G}$ le filtre décroissant de D de centre 0, de diamètre R et soit $\psi_{\mathfrak{G}}$ la semi-norme multiplicative de $H(D)$ associée à $\mathfrak{G}$. Alors si $g \in H(D \cup D')$ on sait que

$$\psi_{\mathfrak{F}}(g) = \psi_{\mathfrak{G}}(g) \quad [4, 9]$$

d'où d'après (2),

$$\lim_{\mathfrak{G}} |f_n(x) - h(x)| \leqq \varepsilon.$$

Mais grâce à (1) on a donc $\lim_{\mathfrak{G}} |f(x) - h(x)| \leqq \varepsilon$. Ceci est vrai pour un $\varepsilon > 0$ quelconque, d'où

$$\lim_{\mathfrak{F}} (f(x) - h(x)) = 0.$$

Mais puisque D est sans T-filtre, la semi-norme multiplicative $\psi_{\mathfrak{G}}$ définie sur $H(D)$ par $\psi_{\mathfrak{G}}(u) = \lim_{\mathfrak{G}} |u(x)|$ est une valeur absolue, d'où la restriction de h à D est égale à f.

Preuve du Théorème 4. On peut naturellement se ramener au cas où $\mathfrak{F}$ est de centre 0. Soit $D' = \mathfrak{P}(\mathfrak{F})$; alors D' admet un trou $T = d^-(0, r)$. Soit $f \in H(D')$. La décomposition Mittag-Lefflerienne de f sur D' montre que f peut s'écrire $f_T + \tilde{f}$ où f_T est une série de Laurent convergente dans $K \setminus T$, telle que $\lim_{|x| \to \infty} f_T(x) = 0$ et $\tilde{f} \in H(D \cup T)$. Alors d'après la proposition 2, f se prolonge en un élément h_T de $H(D)$ et on voit donc que l'élément $h = h_T \tilde{f}$ appartient à $H(D)$, et que $h(x) = f(x)$ pour tout $x \in D'$.

Par ailleurs, il est clair que l'ensemble des autres prolongements de f en éléments de $H(D)$ est $h + \mathfrak{J}_0(\mathfrak{F})$. Il est immédiat d'en déduire un prolongement h_0 tel que $h_0(a) = b$. En effet $\mathfrak{J}_0(\mathfrak{F})$ contient trivialement des éléments φ non nuls en a et on voit que $h_0 = h + \dfrac{(b - h(a))}{\varphi(a)} \varphi$ convient.

Supposons maintenant que la T-suite $(T_{m,i}; q_{m,i})$ soit idempotente. Soit $D'' = D \cup (K \setminus d^-(0, r))$. Alors en appliquant la décomposition Mittag-Lefflerienne de g sur l'infraconnexe D on voit que h peut s'écrire sous la forme $g_1 + g_2$ où $g_1 \in H(D'')$ et $g_2 \in H(D \cup T)$, et $\lim_{|x| \to \infty} g_1(x) = 0$.

Alors grâce au théorème 3, il existe $G_1 \in \mathfrak{M}(R) \cap H(D'')$ tel que $G_1(x) = g_1(x)$ pour $x \in \mathfrak{P}(\mathfrak{F})$ et tel que chaque pôle de G_1 soit simple et soit l'un des points $\beta_{m,i}$.

Alors on voit que l'élément $G = G_1 + g_2$ est un prolongement de f, méromorphe sur $\mathfrak{F}$, avec seulement des pôles simples sur $\mathfrak{F}$ qui sont des points $\beta_{m,i}$. On en déduit encore un prolongement particulier G_0 tel que $G_0(a) = b$ en considérant $\varphi \in \mathfrak{M}(R) \cap \mathfrak{J}_0(\mathfrak{F})$ tel que chaque pôle de φ soit simple et soit l'un des points $\beta_{m,i}$ et vérifie $\varphi(a) \neq 0$. Un tel φ existe d'après la proposition 1 et on voit qu'on peut choisir $G_0 = G + \left(\dfrac{b - G(a)}{\varphi(a)}\right) \varphi$, ce qui achève la démonstration du théorème 4.

Preuve du Théorème 5. On peut naturellement se ramener au cas où $0 \in \mathfrak{P}(\mathfrak{F})$. Alors par l'inversion $x \to \dfrac{1}{x}$ on voit que $D \setminus \{0\}$ est transformé en un infraconnexe D' fermé, admettant un T-filtre décroissant et la conclusion découle du Théorème 4.

THÉORÈME 6. *Soit D un infraconnexe fermé borné admettant un T-filtre à plage non vide $\mathfrak{F}$, avec une T-suite bien percée.*

a) *L'application $f \to \tilde{f}$ qui à chaque $f \in H(D)$ associe sa restriction $\tilde{f}$ à $\mathfrak{P}(\mathfrak{F})$ est une surjection de $H(D)$ sur $H(\mathfrak{P}(\mathfrak{F}))$.*

b) *L'algèbre $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$ est isomorphe à $H(\mathfrak{P}(\mathfrak{F}))$ algébriquement et topologiquement.*

PREUVE. On voit que a) est une conséquence immédiate du théorème 4 si $\mathfrak{F}$ est croissant et du théorème 5 si $\mathfrak{F}$ est décroissant.

Montrons maintenant b). Soit θ la surjection canonique de $H(D)$ sur $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$ et soit ω la surjection de $H(D)$ sur $H(\mathfrak{P}(\mathfrak{F}))$ définie par a). On voit que $\operatorname{Ker}\omega = \operatorname{Ker}\theta$ et par suite d'après le théorème de factorisation, il existe un isomorphisme ψ tel que $\theta = \psi\circ\omega$ et $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$ est isomorphe à $H(\mathfrak{P}(\mathfrak{F}))$.

COROLLAIRE. *Soit D un infraconnexe fermé borné admettant un T-filtre à plage non vide $\mathfrak{F}$, avec une T-suite bien percée et aucun T-filtre complémentaire à $\mathfrak{F}$.*

L'algèbre $\dfrac{H(D)}{\mathfrak{J}(\mathfrak{F})}$ est isomorphe à $H(\mathfrak{P}(\mathfrak{F}))$ algébriquement et topologiquement.

THÉORÈME 7. *Soit D un infraconnexe fermé borné admettant un T-filtre $\mathfrak{F}$ à plage non vide. On suppose que l'ensemble des trous de D est une suite T_n telle que* $\lim\limits_{n\to\infty} (\operatorname{diam}(T_n)) = 0$.

Alors $H(\mathfrak{P}(\mathfrak{F}))$ n'est pas isomorphe à $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$.

PREUVE. Supposons par exemple $\mathfrak{F}$ décroissant, et supposons d'abord que $\mathfrak{P}(\mathfrak{F})$ soit un disque $d(a, r)$. Considérons la valeur absolue quotient $\bar{\psi}$ de la semi-norme multiplicative ψ définie sur $H(D)$ par $\mathfrak{F}$. On constate que $R(\mathfrak{P}(\mathfrak{F}))$ s'identifie à une sous-algèbre de $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$ car tout polynôme sans pôle dans $\mathfrak{P}(\mathfrak{F})$ a une image inversible par la surjection canonique θ de $H(D)$ sur $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$.

On voit donc que $\dfrac{H(D)}{\mathfrak{J}_0(\mathfrak{F})}$ (resp. $H(\mathfrak{P}(\mathfrak{F}))$) est le complété de $R(\mathfrak{P}(\mathfrak{F}))$ pour la topologie de la norme quotient (resp. pour la topologie définie par $\bar{\psi}$).

Les deux algèbres ne sont donc pas isomorphes d'après le théorème de Banach.

VI. Démonstration du théorème 8

Avant de démontrer le théorème 8 on doit d'abord établir les lemmes qui suivent.

On notera $\limsup\limits_{n\to\infty} u_n$ la limite supérieure d'une suite réelle.

Rappelons d'abord le lemme 4 établi dans [20] à l'aide de considérations classiques.

LEMME 4. *Soit* $f(x)=\sum_{n=0}^{\infty}\lambda_n x^n$ *une série de Taylor convergente dans* $d^-(0, R)$. *Alors l'ensemble des zéros de f est une suite de zéros simples* β_n *tels que* $|\beta_{n+1}|>|\beta_n|$ $(n\in\mathbf{N}^*)$ *si et seulement si* $\left|\frac{\lambda_n}{\lambda_{n+1}}\right|<\left|\frac{\lambda_{n+1}}{\lambda_{n+2}}\right|<R$ *pour tout* $n\in\mathbf{N}$.

Si ces conditions sont réalisées alors $|\beta_n|=\left|\frac{\lambda_{n-1}}{\lambda_n}\right|$ *pour tout* $n\in\mathbf{N}^*$.

LEMME 5. *Soit* $(\lambda_n)_{n\in\mathbf{N}}$ *une suite de K telle que* $\lim\limits_{n\to\infty}|\lambda_n|=+\infty$, $\left|\frac{\lambda_n}{\lambda_{n+1}}\right|<\left|\frac{\lambda_{n+1}}{\lambda_{n+2}}\right|$, $\lim\limits_{n\to\infty}\left|\frac{\lambda_n}{\lambda_{n+1}}\right|=1$. *Soit* $\varphi(x)=\sum_{n=0}^{\infty}\lambda_n x^n$. *Alors* φ *converge dans* $d^-(0, 1)$, *les zéros de* φ *sont simples et forment une suite* β_n *telle que* $|\beta_n|=\left|\frac{\lambda_n}{\lambda_{n+1}}\right|$ *et l'on a* $\lim\limits_{n\to\infty}|\lambda_n|\,|\beta_n|^n=$ $=+\infty$.

Pour tout $i\in\mathbf{N}^*$, *soit* $\psi_i(x)=\dfrac{\varphi(x)}{1-\dfrac{x}{\beta_i}}$ *et posons* $\psi_i(x)=\sum_{n=0}^{\infty}\alpha_{i,n}x^n$.

Alors la suite $n\to\alpha_{i,n}$ *vérifie encore* $\lim\limits_{n\to\infty}|\alpha_{i,n}|=+\infty$,

$$\left|\frac{\alpha_{i,n}}{\alpha_{i,n+1}}\right|<\left|\frac{\alpha_{i,n+1}}{\alpha_{i,n+2}}\right|,\quad \lim_{n\to\infty}\left|\frac{\alpha_{i,n}}{\alpha_{i,n+1}}\right|=1.$$

En outre on a $|\alpha_{i,n}|\leqq|\lambda_n|$ *pour tout* $n\in\mathbf{N}$ *et* $|\psi_i(\beta_i)|=|\lambda_i||\beta_i|^i$ *pour tout* $i\in\mathbf{N}^*$.

PREUVE. On sait grâce au lemme 4 que les zéros de φ sont simples et forment une suite β_n telle que $|\beta_n|=\left|\frac{\lambda_{n-1}}{\lambda_n}\right|$, $(n\in\mathbf{N}^*)$. Puisque $\lim\limits_{n\to\infty}|\lambda_n|=+\infty$, on sait que $\lim\limits_{\mu\to 0^+}v(\varphi,\mu)=-\infty$ et donc $\lim\limits_{n\to\infty}|\lambda_n||\beta_n|^n=+\infty$ car $v(\varphi, v(\beta_n))=v(\lambda_n)+$ $+nv(\beta_n)$ [1]. Les zéros de ψ_i sont ceux de φ à l'exception de β_i, et donc on a réciproquement:

$$\left|\frac{\alpha_{i,n}}{\alpha_{i,n+1}}\right|<\left|\frac{\alpha_{i,n+1}}{\alpha_{i,n+2}}\right|,\quad \lim_{n\to\infty}\left|\frac{\alpha_{i,n}}{\alpha_{i,n+1}}\right|=1.$$

D'autre part, on sait que pour un élément $\varphi\in\mathfrak{A}(1)$ l'hypothèse $(\lim\limits_{n\to+\infty}\lambda_n=+\infty)$ est équivalente à $\lim\limits_{\mu\to 0^+}v(\varphi,\mu)=-\infty$. Or il est clair que $\lim\limits_{\mu\to 0^+}v(\psi_i,\mu)=-\infty$, et l'on a donc encore $\lim\limits_{n\to+\infty}|\alpha_{i,n}|=+\infty$.

Enfin, on sait que

$$v(\varphi, v(\beta_n))=v(\lambda_n)+nv(\beta_n) \tag{1}$$

puisque β_n est le $n^{\text{ième}}$ zéro de φ dans l'ordre des valeurs absolues croissantes. De même β_n est le $n^{\text{ième}}$ zéro de ψ_i si $n<i$ et le $(n-1)^{\text{ième}}$ si $n>i$. On aura donc

$$v(\psi_i, v(\beta_n))=v(\alpha_{i,n})+nv(\beta_n)\quad\text{si}\quad n<i, \tag{2}$$

$$v(\psi_i, v(\beta_n))=v(\alpha_{i,n-1})+(n-1)v(\beta_n)\quad\text{si}\quad n>i. \tag{3}$$

Enfin, d'après la définition de ψ_i on voit que

$$v(\varphi, \mu) = v(\psi_i, \mu) \quad \text{si} \quad \mu > v(\beta_i), \tag{4}$$

et

$$v(\varphi, \mu) = v(\psi_i, \mu) + \mu - v(\beta_i) \quad \text{si} \quad \mu \leqq v(\beta_i). \tag{5}$$

Alors grâce à (1), (2), (4) on voit que pour $n<i$ on a $v(\alpha_{i,n})=v(\lambda_n)$.
Grâce à (1), (3), (5) on voit que pour $n>i$ on a

$$v\big(\psi_i, v(\beta_n)\big) = v(\alpha_{i,n-1}) + (n-1)v(\beta_n) = v\big(\varphi, v(\beta_n)\big) - v(\beta_n) + v(\beta_i) =$$
$$= v(\lambda_n) + nv(\beta_n) - v(\beta_n) + v(\beta_i)$$

d'où finalement

$$v(\alpha_{i,n-1}) = v(\lambda_n) + v(\beta_i),$$

et donc

$$|\alpha_{i,n-1}| \leqq |\lambda_n|.$$

On vérifie enfin que $|\psi_i(\beta_i)| = |\lambda_i||\beta_i|^i$. En effet, puisque ψ n'a pas de zéro sur $C(0, |\beta_i|)$ et puisque les zéros dans $d(0, |\beta_i|)$ sont $\beta_1, \ldots, \beta_{i-1}$, on a

$$v\big(\psi_i(\beta_i)\big) = v\big(\psi_i, v(\beta_i)\big) = v(\alpha_{i,i-1}) + (i-1)v(\beta_i).$$

Mais d'après ce qui précède, $v(\alpha_{i,i-1}) = v(\lambda_{i-1})$ et $v(\lambda_{i-1}) + (i-1)v(\beta_i) = v(\lambda_i) + iv(\beta_i)$ d'où

$$|\psi_i(\beta_i)| = |\lambda_i|\,|\beta_i|^i$$

ce qui achève la démonstration.

DÉFINITION. Nous appellerons *matrice infinie* une famille $A=(a_{ij})_{i\in\mathbf{N}^*, j\in\mathbf{N}^*}$ $(a_{ij}\in K)$ que nous noterons

$$\left|\begin{matrix} a_{11} \ldots a_{1j} \ldots \\ a_{21} \ldots a_{2j} \ldots \\ \ldots\ldots\ldots\ldots\ldots \\ a_{i1} \ldots a_{ij} \ldots \\ \ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots \end{matrix}\right.$$

De même nous appellerons *ligne infinie* U (resp. *colonne infinie* V) une suite $(u_n)_{n\in\mathbf{N}^*}$ $(u_n\in K)$ notée $(u_1, \ldots, u_n, \ldots)$ (resp. une suite $(v_n)_{n\in\mathbf{N}^*}$ $(v_n\in K)$ notée

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \\ \vdots \end{pmatrix};$$

On voit que le produit matriciel $UV = \sum_{n=1}^{\infty} u_n v_n$ existe si et seulement si $\lim_{n\to\infty} u_n v_n = 0$.

De même le produit matriciel AV existe et est égal à une colonne infinie

$$\begin{pmatrix} \sum_{j=1}^{\infty} a_{1j}v_j \\ \vdots \\ \sum_{j=1}^{\infty} a_{ij}v_j \end{pmatrix}$$

si et seulement si pour tout entier i, on a $\lim_{j\to\infty} a_{ij}v_j=0$.

Le produit matriciel UA existe et est égal à la ligne infinie

$$\Big(\sum_{i=1}^{\infty} u_i a_{i1}, \ldots, \sum_{i=1}^{\infty} u_i a_{ij}, \ldots\Big),$$

si et seulement si pour tout j fixé, on a $\lim_{i\to\infty} (u_i a_{ij})=0$.

Si $B=(b_{i,j})_{\substack{i\in\mathbf{N}^*\\ j\in\mathbf{N}^*}}$ est une autre matrice infinie le produit matriciel AB existe et est égal à la matrice infinie $\Big(\sum_{j=1}^{\infty} a_{ij}b_{jk}\Big)_{\substack{i\in\mathbf{N}^*\\ k\in\mathbf{N}^*}}$ si et seulement si pour tous i et k fixés on a $\lim_{j\to\infty} (a_{ij}b_{jk})=0$.

Définition. Nous dirons qu'une ligne infinie (resp. une colonne infinie) $(u_n)_{n\in\mathbf{N}^*}$ est *évanescente* si $\lim_{n\to\infty} u_n=0$.

Nous dirons qu'une matrice infinie $A=(a_{i,j})_{\substack{i\in\mathbf{N}^*\\ j\in\mathbf{N}^*}}$ est *évanescente en ligne* (resp. *en colonne*) si chacune de ses lignes infinies (resp. de ses colonnes infinies) est évanescente.

Nous dirons que A est *globalement évanescente* si pour tout $\varepsilon>0$, il existe $N(\varepsilon)$ tel que si $i+j\geqq N(\varepsilon)$ on ait $|a_{ij}|\leqq\varepsilon$.

Le lemme 6 est immédiat en analyse ultramétrique.

Lemme 6. *Soit* $A=(a_{ij})_{\substack{i\in\mathbf{N}^*\\ j\in\mathbf{N}^*}}$ *une matrice infinie globalement évanescente.*

Alors la famille (a_{ij}) *est sommable suivant le filtre des complémentaires des parties finies de* $\mathbf{N}^*\times\mathbf{N}^*$.

Corollaire. *Si une matrice infinie* $A=(a_{ij})_{i\in\mathbf{N}^*}$ *est globalement évanescente les séries* $\sum_{i=1}^{\infty}\Big(\sum_{j=1}^{\infty} a_{ij}\Big)$ *et* $\sum_{j=1}^{\infty}\Big(\sum_{i=1}^{\infty} a_{ij}\Big)$ *sont convergentes et ont même somme.*

On notera $\mathfrak{B}$ la matrice infinie de van der Monde où $(\beta_i)_{i\in\mathbf{N}^*}$ est une suite de K telle que $|\beta_i|<|\beta_{i+1}|$, $\lim_{i\to+\infty} |\beta_i|=1$

$$\begin{pmatrix} 1 \ldots 1 \ldots \\ \beta_1 \ldots \beta_i \ldots \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ \beta_1^n \ldots \beta_i^n \ldots \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \end{pmatrix}$$

LEMME 7. *Soit* (λ_n) *une suite de K telle que* $\left|\frac{\lambda_n}{\lambda_{n+1}}\right| < \left|\frac{\lambda_{n+1}}{\lambda_{n+2}}\right|$ $\forall n \in \mathbf{N}$ *et* $\lim_{n\to\infty} |\lambda_n| = +\infty$, $\limsup_{n\to\infty} |\lambda_n|^{1/n} = 1$. *Soit* $\varphi(x) = \sum_{n=0}^{\infty} \lambda_n x^n \in \mathfrak{A}(1)$ *et soit* $(\beta_i)_{i\in\mathbf{N}^*}$ *la suite des zéros de* φ *dans* $d^-(0, 1)$ *et soit* $\mathfrak{B}$ *la matrice infinie de van der Monde associée. On note* $\psi_i(x) = \frac{\varphi(x)}{1-\frac{x}{\beta_i}}$ *et* $\varphi_i(x) = \frac{\psi_i(x)}{\psi_i(\beta_i)}$ *et on pose* $\varphi_i(x) = \sum_{n=0}^{\infty} \lambda_{i,n} x^n$. *Alors la matrice infinie*

$$\mathfrak{B}' = \left|\begin{matrix} \lambda_{1,0} \dots \lambda_{1,n} \dots \\ \dots\dots\dots\dots \\ \dots\dots\dots\dots \\ \lambda_{i,0} \dots \lambda_{i,n} \dots \\ \dots\dots\dots\dots \\ \dots\dots\dots\dots \end{matrix}\right.$$

est évanescente en colonne et vérifie $\mathfrak{B}\mathfrak{B}' = \mathfrak{B}'\mathfrak{B} = I$. *En outre on a* $|\lambda_{i,n}| \leqq \frac{|\lambda_n|}{|\lambda_i|\,|\beta_i|^i}$ *et pour toute colonne évanescente* $A = (a_n)_{n\in\mathbf{N}}$ *on a* $\mathfrak{B}(\mathfrak{B}'A) = A$.

Pour tout $\varrho > 0$, soit $\Delta_\varrho = d^-(0, 1) \setminus \bigcup_{i=1}^{\infty} d^-(\beta_i, \varrho)$ et pour tout $i \in \mathbf{N}^*$ soit $\theta_i = \psi_i(\beta_i)$. D'après le lemme 4 on sait que $|\beta_i| < |\beta_{i+1}|$, que chaque β_i est un zéro simple et on en déduit que $v(\varphi(x)) \leqq v(\varphi, v(x)) + (v(x) - \log \varrho)$ d'où on voit que $\lim_{\substack{|x|\to 1^- \\ x\in\Delta_\varrho}} \varphi(x) = +\infty$ car on sait que $\lim_{\mu\to 0^+} v(\varphi, \mu) = -\infty$ du fait que $\lim_{n\to+\infty} |\lambda_n| = +\infty$. Alors d'après [18, 20] on sait que $\frac{1}{\varphi} \in H(\Delta_\varrho)$. Soit $u_n = \frac{x^n}{\varphi} \in H(\Delta_\varrho)$; il est clair que dans l'infraconnexe Δ_ϱ, u_n admet un développement de Mittag-Leffler sous la forme $\sum_{i=1}^{\infty} \frac{\alpha_{i,n}}{1-(x/\beta_i)}$. Calculons $\alpha_{i,n}$. Pour cela considérons $v_{n,i} = \left(1 - \frac{x}{\beta_i}\right) u_n$. Alors $v_{n,i}(\beta_i) = \alpha_{i,n}$. Or $\varphi = (\varphi_i)\left(1 - \frac{x}{\beta_i}\right)\theta_i$, d'où:

$$v_{n,i}(\beta_i) = \frac{\beta_i^n}{\varphi_i(\beta_i)\theta_i}$$

et comme $\varphi_i(\beta_i) = 1$ on a donc $\alpha_{i,n} = \frac{\beta_i^n}{\theta_i}$, d'où finalement

$$\frac{x^n}{\varphi(x)} = \sum_{i=1}^{\infty} \left(\frac{\beta_i^n}{\theta_i}\right) \frac{1}{(1-(x/\beta_i))}$$

et par suite $x^n = \sum_{i=1}^{\infty} \frac{(\beta_i)^n}{\theta_i} \frac{\varphi(x)}{1-(x/\beta_i)}$ est vrai pour tout $x \neq \beta_1, \dots, \beta_i, \dots$.

On a donc $x^n=\sum_{i=1}^{\infty}\beta_i^n\varphi_i(x)$. Soit $\varphi_i(x)=\sum_{n=0}^{\infty}\lambda_{i,n}x^n$ ($i\in\mathbf{N}^*$). Alors on voit que $\sum_{i=1}^{\infty}\beta_i^n\lambda_{i,n}=1$ et que $\sum_{i=1}^{\infty}\beta_i^n\lambda_{i,k}=0$ $\forall k\neq n$.

La matrice infinie

$$\mathfrak{B}'=\left|\begin{matrix}\lambda_{1,0}\ldots\lambda_{1,n}\ldots\\ \ldots\ldots\ldots\ldots\ldots\\ \ldots\ldots\ldots\ldots\ldots\\ \lambda_{i,0}\ldots\lambda_{i,n}\ldots\\ \ldots\ldots\ldots\ldots\ldots\\ \ldots\ldots\ldots\ldots\ldots\end{matrix}\right.$$

vérifie donc $\mathfrak{B}\mathfrak{B}'=I$.

On montre de même que $\mathfrak{B}'\mathfrak{B}=I$. En effet, $\varphi_i(\beta_k)=0\ \forall k\neq i$; on a $\sum_{n=0}^{\infty}\lambda_{i,n}\beta_k^n=0$ pour tout $i\neq k$, et $\varphi_i(\beta_i)=1$ d'où $\sum_{n=0}^{\infty}\lambda_{i,n}\beta_i^n=1$ ce qui montre bien que $\mathfrak{B}'\mathfrak{B}=I$.

D'autre part d'après le lemme 5 on sait que $|\psi_i(\beta_i)\alpha_{i,n}|\leqq|\lambda_n|$ et que $|\psi_i(\beta_i)|=$ $=|\lambda_i|\,|\beta_i|^i$. On a donc

$$|\lambda_{i,n}|\leqq\frac{|\lambda_n|}{|\lambda_i|\,|\beta_i|^i}. \tag{1}$$

Pour montrer que $(\mathfrak{B}\mathfrak{B}')A=\mathfrak{B}(\mathfrak{B}'A)$ il suffit d'établir que pour chaque ligne $U_k=(\beta_1^k,\ldots,\beta_i^k,\ldots)$ de $\mathfrak{B}$ on a bien $(U_k\mathfrak{B}')A=U_k(\mathfrak{B}'A)$.

Considérons la matrice infinie $(c_{i,n})_{\substack{i\in\mathbf{N}^*\\ n\in\mathbf{N}}}$ avec $c_{i,n}=\beta_i^k\lambda_{i,n}a_n$.

D'après le corollaire du lemme 6, il suffit de montrer qu'elle est globalement évanescente, c'est-à-dire que la famille (indexée par i et n) $(\beta_i^k\lambda_{i,n}a_n)_{\substack{i\in\mathbf{N}^*\\ n\in\mathbf{N}}}$ tend vers 0 quand $i+n$ tend vers $+\infty$.

Or d'après (1) on a $|\beta_i^k\lambda_{i,n}a_n|\leqq\dfrac{|a_n|\,|\lambda_n|}{|\beta_i|^{i-k}|\lambda_i|}$. Maintenant on sait que $\lim_{n\to\infty}\lambda_n a_n=0$ et que $\lim_{i\to+\infty}|\lambda_i|\,|\beta_i|^i=+\infty$. Soit $M=\sup_{n\in\mathbf{N}}|a_n\lambda_n|$ et soit $m=\inf_{i\in\mathbf{N}}|\beta_i|^{i-k}|\lambda_i|$. Fixons $\varepsilon>0$; il existe donc des entiers P et $Q\in\mathbf{N}$ tels que $|\lambda_n a_n|\leqq\varepsilon m$ pour $n\geqq P$ et $|\beta_i|^{i-k}|\lambda_i|\geqq\dfrac{M}{\varepsilon}$ pour $i\geqq Q$. Alors on voit que quand $i+n\geqq P+Q$ on a $\dfrac{|\lambda_n a_n|}{|\beta_i|^{i-k}|\lambda_i|}\leqq\varepsilon$ car ou bien $n\geqq P$ et alors

$$|\lambda_n a_n|\leqq\varepsilon m,\quad |\beta_i^{i-k}\lambda_i|\geqq m,$$

ou bien $i\geqq Q$ et alors

$$|\beta_i|^{i-k}|\lambda_i|\geqq\frac{M}{\varepsilon},\quad |\lambda_n a_n|\leqq M.$$

On a donc montré que la famille $\theta_{i,n}=\beta_i^k\lambda_{i,n}a_n$ tend vers 0 quand $i+n$ tend vers $+\infty$, ce qui achève la démonstration.

DÉMONSTRATION DU THÉORÈME 8. Pour chaque $i\in\mathbf{N}^*$, soit $\varphi_i(x)=\sum_{n=0}^{\infty}\lambda_{i,n}x^n$. D'après le lemme 5 il est clair que $|\lambda_{i,n}|\leq\frac{|\lambda_n|}{|\lambda_i|\,|\beta_i|^i}$ ce qui montre que $|\bar{\varphi}_i(P)|\leq$ $\leq\frac{\|P\|_\varphi}{|\lambda_i|\,|\beta_i|^i}$ et les $\bar{\varphi}_i$ sont donc continues. Par ailleurs il est clair que la série $\sum_{n=0}^{\infty}a_nx^n$ converge dans E pour la norme $\|\cdot\|_\varphi$ puisque $\lim_{n\to\infty}|\lambda_n a_n|=0$. Notons $\varepsilon_i=\bar{\varphi}_i(f)$ $(i\in\mathbf{N}^*)$.

Alors $|\varepsilon_i|\leq\sup_{n\in\mathbf{N}}|\lambda_{i,n}a_n|\leq\frac{\sup_{n\in\mathbf{N}}|\lambda_n a_n|}{|\lambda_i|\,|\beta_i|^i}$. Mais d'après le lemme 5 on sait que $\lim_{i\to\infty}|\lambda_i|\,|\beta_i|^i=+\infty$, ce qui montre que $\lim_{i\to\infty}\varepsilon_i=0$.

Alors d'après le lemme 7 la matrice infinie

$$\mathfrak{B}=\begin{pmatrix} 1 \ldots 1 \ldots \\ \beta_1 \ldots \beta_i \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ \beta_1^n \ldots \beta_i^n \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \end{pmatrix}$$

admet pour inverse

$$\mathfrak{B}'=\begin{pmatrix} \lambda_{1,0} \ldots \lambda_{1,n} \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ \lambda_{i,0} \ldots \lambda_{i,n} \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \end{pmatrix}$$

Mais on voit que la matrice colonne infinie évanescente C

$$C=\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \end{pmatrix}$$

est égale au produit convergeant $\mathfrak{B}'A$ où A est la matrice colonne

$$\begin{pmatrix} a_0 \\ \vdots \\ a_n \\ \vdots \end{pmatrix}$$

Puisque $\mathfrak{B}$ est bornée et que C est évanescente, on voit que le produit $\mathfrak{B}C$ converge et d'après le lemme 7 on a $\mathfrak{B}(\mathfrak{B}'A)=(\mathfrak{B}\mathfrak{B}')A$ donc $A=\mathfrak{B}C$.

Soit $\bar{f}(x)=\sum_{i=1}^{\infty}\frac{\varepsilon_i}{1-\frac{x}{b_i}}$. Il est immédiat de voir que $\bar{f}\in H(\Lambda_\varrho)$ et nous allons montrer que pour $|x|\leqq 1$ on a $\bar{f}(x)=\sum_{n=0}^{\infty}a_n x^n$.

En effet quand $x\in d(0,1)$ on peut écrire $\frac{1}{1-\frac{x}{b_i}}=\sum_{n=0}^{\infty}\left(\frac{x}{b_i}\right)^n=\sum_{n=0}^{\infty}(\beta_i x)^n$ et donc $\bar{f}(x)$: $\sum_{i=1}^{\infty}(\sum_{n=0}^{\infty}\varepsilon_i\beta_i^n x^n)=\sum_{n=0}^{\infty}x^n(\sum_{i=1}^{\infty}\varepsilon_i\beta_i^n)$ (d'après le corollaire du lemme 5). Alors chaque série $\sum_{i=1}^{\infty}\varepsilon_i\beta_i^n$ converge vers a_n, d'où $\bar{f}(x)=\sum_{n=0}^{\infty}a_n x^n$, ce qui achève la démonstration.

Preuve du Théorème 9. Du fait de la densité de $|K|$ dans $\mathbf{R}_+$, il suffit de montrer l'existence de suites $(r_n)_{n\in\mathbf{N}}$ dans $\mathbf{R}_+$, vérifiant a), b), c), d), pour en déduire l'existence de suites dans $\mathbf{R}_+\cap|K|$ vérifiant encore a), b), c), d).

D'autre part, nous allons montrer d'abord que si l'on prouve l'existence de suites r_n vérifiant a), c), d), alors il existe des suites r_n vérifiant a), b), c), d). En effet supposons que pour toute suite u_n de $\mathbf{R}_+$ telle que $\lim_{n\to\infty}u_n=0$, il existe des suites r_n de $\mathbf{R}_+$ vérifiant a), c), d). Il est immédiat de se ramener au cas où $\lim_{n\to\infty}\sqrt[n]{u_n}=1$. Soit $u_n'=\max\left(u_n,\frac{1}{n}\right)$; on voit que $\lim_{n\to\infty}u_n'=0$. En outre puisque $\lim_{n\to\infty}u_n=0$, on sait que $\limsup_{n\to\infty}\sqrt[n]{u_n}\leqq 1$ et puisque $\lim_{n\to\infty}\sqrt[n]{\frac{1}{n}}=1$, on voit que $\lim_{n\to\infty}\sqrt[n]{u_n'}=1$. Alors en supposant vérifiés a), c), d), il existe des suites r_n de $\mathbf{R}_+$ vérifiant a), c) et $\lim_{n\to\infty}r_n u_n'=0$ d'où, a fortiori, $\lim_{n\to\infty}r_n u_n=0$. Mais on a aussi $\lim_{n\to\infty}\sqrt[n]{r_n u_n'}\leqq 1$ puisque $\lim_{n\to\infty}r_n u_n'=0$. Or puisque $\lim_{n\to\infty}\sqrt[n]{u_n'}=1$, on voit que $\limsup_{n\to\infty}\sqrt[n]{r_n}\leqq 1$. Par ailleurs puisque $\lim_{n\to\infty}r_n=+\infty$, on voit que $\limsup_{n\to\infty}\sqrt[n]{r_n}\geqq 1$, donc $\limsup_{n\to\infty}\sqrt[n]{r_n}=1$. Enfin puisque la suite $\frac{r_n}{r_{n+1}}$ est monotone, elle converge vers une limite qui est aussi celle de $\sqrt[n]{r_n}$, et cette limite est donc 1.

Il reste donc à montrer l'existence de suite r_n vérifiant a), c), d).

On va procéder par des méthodes d'analyse élémentaires et notamment de convexité.

Soit $z_m=-\log u_m$ $(m\in\mathbf{N})$. On voit que $\lim_{m\to\infty}z_m=+\infty$ et que l'on peut trouver une suite θ_m vérifiant pour tout $m \in \mathbf{N}$

(1) $$\theta_m \leqq \frac{z_m}{2},$$

(2) $$0 < \theta_m < \theta_{m+1},$$

(3) $$\lim_{m\to\infty} \theta_m = +\infty,$$

(4) $$\lim_{m\to\infty} \frac{\theta_m}{m} = 0.$$

Nous allons montrer qu'on peut extraire une suite croissante d'entiers $q_n \in \mathbf{N}$ vérifiant

(5) $$\frac{\theta_{q_{n+1}}-\theta_{q_n}}{q_{n+1}-q_n} < \frac{\theta_{q_n}-\theta_{q_{n-1}}}{q_n-q_{n-1}}, \quad n\in\mathbf{N}^*,$$

(6) $$\frac{\theta_{q_{n+1}}-\theta_{q_n}}{q_{n+1}-q_n} < \frac{\theta_m-\theta_{q_n}}{m-q_n} \text{ pour } q_n < m < q_{n+1}, \quad n\in\mathbf{N}.$$

En effet supposons construite cette suite q_n jusqu'au rang N et définissons q_{N+1}. D'après (2) on voit que $\frac{\theta_{q_N}-\theta_{q_{N-1}}}{q_N-q_{N-1}} > 0$ et d'après (4) il est clair que (7)

$$\lim_{m\to\infty}\left(\frac{\theta_m-\theta_{q_N}}{m-q_N}\right) = 0.$$

Il existe donc un rang $q_{N+1} > q_N$ tel que

(5) $$\frac{\theta_{q_{N+1}}-\theta_{q_N}}{q_{N+1}-q_N} < \frac{\theta_{q_N}-\theta_{q_{N-1}}}{q_N-q_{N-1}}$$

et

(6) $$\frac{\theta_{q_{N+1}}-\theta_{q_N}}{q_{N+1}-q_N} < \frac{\theta_m-\theta_{q_N}}{m-q_N} \text{ pour } q_N < m < q_{N+1}.$$

La suite q_n est donc définie par récurrence de cette façon en choisissant $q_0=0$ et $q_1=1$.

Grâce à (5) et (6) il est immédiat de définir une suite $(\omega_m)_{m\in\mathbf{N}}$ telle que $\omega_{q_n}=\theta_{q_n}$ et

a′) $$0 < \omega_{m+1}-\omega_m < \omega_m-\omega_{m-1} \qquad \forall m\in\mathbf{N}^*$$

c′) $$\omega_m \leqq \theta_m \qquad \forall m\in\mathbf{N}$$

par exemple. Supposons déjà construit les ω_m pour $m \leqq q_n$ vérifiant a′) et c′) ainsi que

(7) $$\omega_{q_n}-\omega_{q_n-1} > \frac{\omega_{q_{n+1}}-\omega_{q_n}}{q_{n+1}-q_n},$$

et construisons les ω_m pour $q_n < m \leqq q_{n+1}$.

Soit h_n la fonction affine telle que $h_n(q_n)=\theta_{q_n}$ et $h_{n+1}(q_{n+1})=\theta_{q_{n+1}}$ et considérons un cercle Σ passant par les points (q_n, θ_{q_n}) et $(q_{n+1}, \theta_{q_{n+1}})$, et dont le centre $C=(a, b)$ est tel que $b<0$. Soit φ la fonction définie sur l'intervalle $I=[q_n, q_{n+1}]$ où $\varphi(x)$ est l'ordonnée du point de Σ d'abscisse x. On voit que si b tend vers $-\infty$, alors φ converge uniformément vers h sur I. Pour chaque b fixé posons $\omega_m=\varphi(m)$ pour $q_n<m\leqq q_{n+1}$. Alors d'après la convexité du cercle on voit que $\omega_{m+1}-\omega_m<\omega_m-\omega_{m-1}$ pour $q_n<m\leqq q_{n+1}$.

En outre quand $|y|$ est assez grand $(y<0)$ on voit que $\omega_m\leqq\theta_m$ et que

$$\omega_{q_n+1}-\omega_{q_n}<\omega_{q_n}-\omega_{q_n-1}$$

et enfin $\omega_{q_{n+1}}-\omega_{q_{n+1}-1}>\dfrac{\omega_{q_{n+2}}-\omega_{q_{n+1}}}{q_{n+2}-q_{n+1}}$ ce qui donne (7) au rang $n+1$ et permet de poursuivre la définition par récurrence.

La suite ω_m étant construite, vérifiant a′) et c′) il est clair qu'elle vérifie aussi b′) $\lim\limits_{m\to\infty}\omega_m=+\infty$ d'après a′) et du fait que $\lim\limits_{n\to\infty}\omega_{q_n}=+\infty$ d'après (3).

Alors en prenant $r_m=p^{\omega_m}$ on voit que a′), b′), c′) entraînent respectivement a), b), c) ce qui achève la démonstration du théorème.

RÉFÉRENCES

[1] AMICE, Y., *Les nombres p-adiques*, Presses Universitaires de France, Paris, 1975. *MR* **56** #5510.

[2] ESCASSUT, A., Les algèbres de Krasner, *C.R. Acad. Sci. Paris* **272** (1971), 598—601. *MR* **43** #5308.

[3] ESCASSUT, A., Algèbres d'éléments analytiques en analyse non archimédienne, *Indag. Math.* **36** (1974), 339—351. *MR* **51** #10671.

[4] ESCASSUT, A., Éléments analytiques et filtres percés sur un ensemble infraconnexe, *Ann. Mat. Pura Appl.* **110** (1976), 335—352. *MR* **54** #13132.

[5] ESCASSUT, A., T-filtres, ensembles analytiques et transformation de Fourier p-adique, *Ann. Inst. Fourier (Grenoble)* **25** (1975), 45—80. *MR* **52** #11111.

[6] ESCASSUT, A., Algèbres de Krasner intègres et noethériennes, *Indag. Math.* **38** (1976), 109—130. *MR* **53** #11381, 11382.

[7] ESCASSUT, A., Spectre maximal d'une algèbre de Krasner, *Colloq. Math.* **38** (1978), 339—357. *MR* **58** #22659.

[8] ESCASSUT, A., The ultrametric spectral theory, *Period. Math. Hungar.* **11** (1980), 7—60. *MR* **81i**: 46098.

[9] GARANDEL, G., Les semi-normes multiplicatives sur les algèbres d'éléments analytiques au sens de Krasner, *Indag. Math.* **37** (1975), 327—341. *MR* **52** #11112.

[10] KRASNER, M., Prolongement analytique dans les corps valués complets: préservation de l'analycité par des opérations rationnelles; quasi-connexité et éléments analytiques réguliers, *C. R. Acad. Sci. Paris* **244** (1957), 1599—1602. *MR* **19**—395.

[11] KRASNER, M., Prolongement analytique dans les corps valués complets: uniformité des fonctions analytiques; l'analycité des fonctions méromorphes, *C. R. Acad. Sci. Paris* **244** (1957), 1996—1999. *MR* **19**—395.

[12] KRASNER, M., Prolongement analytique dans les corps valués complets: éléments analytiques, préliminaires du théorème d'unicité, *C. R. Acad. Sci. Paris* **239** (1954), 468—470. *MR* **16**—799.

[13] KRASNER, M., Prolongement analytique dans les corps valués complets: préservation de l'analycité par la convergence uniforme et par la dérivation; théorème de Mittag-Leffler généralisé pour les éléments analytiques, *C. R. Acad. Sci. Paris* **244** (1957), 2570—2573. *MR* **19**—395.

[14] KRASNER, M., Prolongement analytique uniforme et multiforme dans les corps valués complets, *Les tendances géométriques en algèbre et théorie des nombres*, Clermont-Fer-

rand, 1964, 97—141. Centre National de la Recherche Scientifique, 1966. *MR* **34** #4246.

[15] Motzkin, E. et Robba, Ph., Ensembles d'analytique en analyse p-adique, *Séminaire de Théorie des Nombres*, Secrétariat mathématique, Paris, 1969. *MR* **41** #7155.

[16] Robba, Ph., Fonctions analytiques sur les corps valués ultramétriques complets, *Prolongement analytique et algèbres de Banach ultramétriques*, Astérisque **10** (1973), 109—220. *MR* **50** #10307.

[17] Sarmant, M.-C. et Escassut, A., T-suites idempotentes, *Bull. Sci. Math.* **106** (1982), 289—303. *MR* **84j**: 30086.

[18] Sarmant, M.-C., Décomposition en produit de facteurs de fonctions méromorphes, *C. R. Acad. Sci. Paris* **292** (1981), 127—130. *MR* **82b**: 12017.

[19] Sarmant, M.-C., Produits méromorphes, *Bull. Sci. Math.* **109** (1985), 155—178.

[20] Sarmant, M.-C., Fonctions analytiques et produits croulants, *Collectanea Math.* **36** (1985), 199—218.

(Reçu le 5 février 1985)

UNIVERSITÉ PIERRE ET MARIE CURIE
MATHÉMATIQUES
TOUR 45—46
4, PLACE JUSSIEU
F—75 230 PARIS Cedex 05

MATHÉMATIQUES ET INFORMATIQUE
UNIVERSITÉ DE BORDEAUX I
351, COURS DE LA LIBÉRATION
F—33405 TALENCE Cedex
FRANCE

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 445—446

A NOTE ON RAMSEY NUMBERS

H. LEFMANN

Let k, m be positive integers. The Ramsey number $r_k(m)$ denotes the least positive integer n such that $n \to (m)_k^2$; i.e., for every k-coloring of the edges of the complete graph K_n there is a monochromatic K_m subgraph.

Erdős [1] proved that $r_k(m) > cmk^{m/2}$ for a positive constant c and all positive integers k, m. In [2] Erdős and Szemerédi raised the question whether this lower bound for $r_k(m)$ can be improved. In this note we prove the following

THEOREM. *For all positive integers k, m with $m \geqq 3$:*

$$r_{2k}(m) > 2^{(m/2)k}.$$

The following observation occurred in [3]:

LEMMA. *If k, l, m are positive integers, then*

$$r_{k+l}(m) \geqq (r_k(m)-1)(r_l(m)-1)+1.$$

PROOF. Let k, l, m be positive integers and let $G_i = (V(G_i), E(G_i))$ for $i=1, 2, 3$ be complete graphs with vertex sets

$$V(G_1) = \{(0, j)\ j = 1, 2, \ldots, r_k(m)-1\},$$

$$V(G_2) = \{(i, 0)\ i = 1, 2, \ldots, r_l(m)-1\},$$

$$V(G_3) = \{(i, j)\ i = 1, 2, \ldots, r_l(m)-1 \quad \text{and} \quad j = 1, 2, \ldots, r_k(m)-1\}.$$

Let Δ_1: $E(G_1) \to \{1, \ldots, k\}$ and Δ_2: $E(G_2) \to \{k+1, \ldots, k+l\}$ be colorings such that there is no K_m subgraph of G_1 resp. G_2 with all edges colored the same by Δ_1 resp. Δ_2. Consider the coloring Δ_3: $E(G_3) \to \{1, 2, \ldots, k+l\}$ defined by

$$\Delta_3(\{(i_1, j_1), (i_2, j_2)\}) = \begin{cases} \Delta_1(\{(0, j_1), (0, j_2)\}), & \text{if} \quad i_1 = i_2, \\ \Delta_2(\{(i_1, 0), (i_2, 0)\}), & \text{if} \quad i_1 \neq i_2. \end{cases}$$

Clearly, G_3 contains no monochromatic K_m subgraph.

PROOF OF THEOREM. By [1] we have $r_2(m) \geqq 2^{m/2}+1$ for every integer $m \geqq 3$. The lemma yields $r_{2k}(m) \geqq (r_{2k-2}(m)-1)(r_2(m)-1)+1$ so by induction $r_{2k}(m) > 2^{(m/2)k}$ for all integers $k \geqq 1$, $m \geqq 3$.

1980 *Mathematics Subject Classification.* Primary 05A17.
Key words and phrases. Ramsey numbers.

REFERENCES

[1] Erdős, P., Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292—294. *MR* **8**—479.

[2] Erdős, P. and Szemerédi, E., On a Ramsey type theorem, *Per. Math. Hung.* **2** (1972), 295—299.

[3] Lefmann, H., Ramsey und Paris—Harrington Resultate in der Partitionstheorie, Diplomarbeit, Universität Bielefeld, 1982.

(Received March 2, 1985)

UNIVERSITÄT BIELEFELD
FAKULTÄT FÜR MATHEMATIK
POSTFACH 8640
D—4800 BIELEFELD 1
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 447—448

ON THE SEVEN POINTS PROBLEM OF P. ERDŐS

ILONA PALÁSTI

Dedicated to Professor P. Erdős on his 75th birthday

Abstract

Seven points in general position are given in the plane, i.e. no three on a line, no four on a circle, and they determine 6 distinct distances (being not in any order) so that the i-th distance occurs i times, $i=1, 2, \ldots, 6$.

1. Introduction. P. Erdős [1] asked the following question: can you find n points in the plane in general position, that is no three on a line, no four on a circle, which determine $n-1$ distinct distances so that the i-th distance occurs i times? The distances are not ordered by size or in any other way. For $n=4$ he gives the isosceles triangle and takes its centre.

Pomerance showed a construction for $n=5$. Two Hungarian students found an example for $n=6$ (but Erdős and everybody have forgotten both the examples and their names).

Here we present seven points for which the same is true.

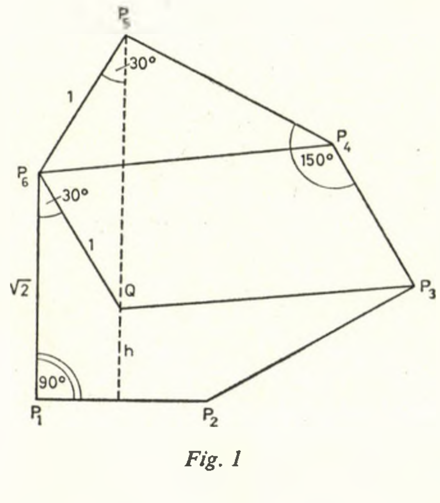


Fig. 1

1980 *Mathematics Subject Classification.* Primary 51L15; Secondary 51M15, 51A20.
Key words and phrases. Arrangements of points, Erdős' problem.
Research (partially) supported by the Hungarian National Foundation for Scientific Research Grant No. 1808.

2. The seven points construction. Consider the convex hexagon $P_1P_2\dots P_6$ with side length 1 and $\sqrt{2}$ ($P_{2i-1}P_{2i}=1$, $P_{2i}P_{2i+1}=\sqrt{2}$ for $i=1, 2, 3$) and angles 90° and 150° ($P_{2i-1}P_{2i}P_{2i+1}\sphericalangle=150°$, $P_{2i}P_{2i+1}P_{2i+2}\sphericalangle=90°$). See Figure 1. Then the vertex P_5 is on the perpendicular bisector h of P_1P_2. The angle P_6P_5h is 30°. Let Q be the point on h inside the hexagon such that $P_5Q=\sqrt{3}$. Then $P_6Q=1$ and $P_6Q\parallel P_4P_3$.

We claim that the point-set $\mathcal{P}_7=\{P_1, \dots, P_6, Q\}$ fulfils the Erdős' demands. Indeed, the distance $\sqrt{3+\sqrt{6}}=2.3\dots$ occurs 6 times (P_1P_3, P_1P_4, P_1P_5, P_2P_5, P_3P_5, P_3P_6), $\sqrt{3}=1.7\dots$ five times (P_2P_4, P_4P_6, P_6P_2 and P_5Q, P_3Q), 1 four times (P_1P_2, P_3P_4, P_5P_6 and P_6Q), $\sqrt{2}=1.4\dots$ three times (P_2P_3, P_4P_5, P_6P_1), $\sqrt{3-\sqrt{6}}=$ $=0.7\dots$ twice (P_1Q, P_2Q) and $\sqrt{5-\sqrt{6}}=1\cdot5\dots$ once (P_4Q). It is easy to check that no three points are on a line and no four on a circle.

3. Examples for $n\leqq 6$. Consider the regular triangle ABC with center 0. Let A' be defined by $OA'=AA'$, $BA'=BA$, $OAA'B$ is a convex quadrilateral. Let $A'OB\sphericalangle=120°$, $OA'=OB'$. Then the configurations $\mathcal{P}_6=\{O, A, B, C, A', B'\}$ and $\mathcal{P}_5=\{O, A, B, C, A'\}$ fulfil Erdős' constraints for $n=6$ and 5, resp. (see Figure 2). The arrangement $\mathcal{P}_5$ was found by C. Pomerance.

The author expresses her thanks to Z. Füredi for his valuable remarks.

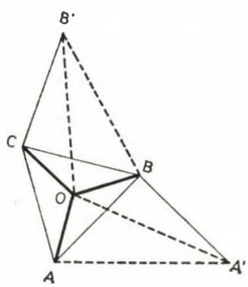


Fig. 2

REFERENCE

[1] Erdős, P., Combinatorial problems in geometry, *Math. Chronicle* **12** (1983), 34—35. *MR* **84i:** 52014.

(Received March 14, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 449—455

ON THE HANKEL-TRANSFORMATION OF SCHWARTZ DISTRIBUTIONS

I. FENYŐ

There are several definition known for the Hankel-transformation of certain classes of distributions. One of them was given by us [1], this definition is based on a known relation between the Fourier- and Hankel-transforms A different way was followed in order to give an applicable definition of the Hankel-transformation of generalized functions by A. H. Zemanian [2], [3], [4]. The aim of the present paper is to give a new definition of the Hankel-transformation of distributions based on the idea followed by Gelfand and Schilow used to define the Fourier-transforms of generalized functions [5], p. 153—155. The background of this is the Paley—Wiener theorem [6] which has an analogous concerning the Hankel-transformation. Such a theorem was proved by T. M. McRobert in 1931 [7], p. 53 and recently by Griffith [8]. The advantage of our present definition is its simplicity, we do not need to introduce special test-function spaces as it was made by Zemanian. Its disadvantage in respect to the definition of Zemanian is, that it works only for positive integer index n.

1. Let $D(a)$ be the subspace of the space of Schwartz testing functions which have their support on $(0, a)$ $(a>0)$. For a given nonnegative integer n we denote by $G_n(a)$ the space of functions ψ fulfilling the following conditions:

a) $\psi(s)$ is an entire function $(s\in\mathbf{C})$,

b) $|s^k\psi(s)|\leqq C_k e^{a|\mathrm{Im}\, s|}$ for $|s|>\lambda_\psi$ where λ_ψ is a positive constant depending only on ψ.

c$_n$) $$\psi(-s)=(-1)^n\psi(s);$$

d$_n$) $$\int_0^\infty s^{k-1}\psi(s)\,ds=0; \qquad k=n,\ n+2,\ n+4, \ldots$$

e$_n$) $$|\psi(s)|=O(|s|^n) \quad (s\to 0).$$

In what follows we consider the following expression (if it exists for every $s\in\mathbf{R}$)

$$H_n(f)(s)=\int_0^\infty tJ_n(st)f(t)\,dt \qquad (n=0,1,2,\ldots) \tag{1}$$

as the Hankel-transform of order n of the function $f\colon \mathbf{R}_+\to\mathbf{R}$. Here J_n denotes the Bessel function of first kind.

1980 *Mathematics Subject Classification.* Primary 46F12; Secondary 44A15.
Key words and phrases. Hankel-transformation, distributions, Paley—Wiener theorem.

THEOREM 1. *The Hankel-transformation H_n is an algebraic isomorphism between $D(a)$ and $G_n(a)$* $(a>0,\ n=0, 1, 2, \ldots)$.

PROOF. What we have to prove is that $\psi \in G_n(a)$ if and only if there exists a unique function $\varphi \in D(a)$ for which $\psi = H_n(\varphi)$ holds.

First let us consider a function $\varphi \in D(a)$. Its Hankel-transform (of order n) is

$$H_n(\varphi)(s) = \psi(s) = \int_0^a tJ_n(ts)\varphi(t)\,dt. \tag{2}$$

As J_n is an entire function condition a) is obviously valid.

In order to show that ψ fulfils also the condition b) let us consider the Bessel differential operator B_n defined for an arbitrary function $x \in \mathbf{C}^2$ as follows:

$$B_n(x)(t) = x'' + \frac{1}{t}x' - \frac{n^2}{t^2}x \qquad (t>0). \tag{3}$$

If we now substitute for x our test function φ, then $B_n(\varphi) \in D(a)$. By a well-known theorem (see e.g. [7], p. 61, formula (32))

$$H_n(B_n\varphi)(s) = -s^2 H_n(\varphi)(s). \tag{4}$$

But all testing functions of $D(a)$ fulfils the conditions (1) and (2) of the theorem of Griffith [8], therefore by the quoted theorem

$$|s^{5/2} H_n(\varphi)(s)| = |s^{1/2} H_n(B_n\varphi)(s)| \leqq C' e^{a|\mathrm{Im}\, s|} \tag{5}$$

for values great enough of $|s|$. This implies immediately b) for $k=0, 1$ and 2.

Applying the relation (4) to the second iterated B_n^2 of B_n, we get

$$H_n(B_n^2\varphi)(s) = s^4 H_n(\varphi)(s). \tag{6}$$

As $B_n^2\varphi \in D(a)$, we get

$$|s^{9/2} H_n(\varphi)(s)| = |s^{1/2} H_n(B_n^2\varphi)(s)| \leqq C'' e^{a|\mathrm{Im}\, s|},$$

from this follows b) for $k=3, 4$. By induction we see the validity in the same way for every nonnegative integer k.

The property c$_n$) is an immediately consequence of (2).

Let us now prove that also condition d$_n$) holds. By b) the Hankel-transform (of the order n) exists and by the well-known inversion theorem of Hankel-transforms (2) implies

$$\varphi(t) = \int_0^\infty sJ_n(ts)\psi(s)\,ds. \tag{7}$$

As $\varphi \in D(a)$, all derivatives of φ vanishes at $t=0$. We will now form the derivatives of φ differentiating under the sign of integration and put $t=0$. In order to show the legality of this step we have to prove that

$$\int_0^\infty s^{k+1} J_n^{(k)}(ts)\psi(s)\,ds \qquad (k=0, 1, 2, \ldots) \tag{8_k}$$

exists uniformly with respect to t. To do so let us consider first

$$\int_0^\infty s^2 J_n'(ts)\psi(s)\,ds. \tag{8'}$$

This is by a well-known relation (see e.g. [9] p. 360) equal to

$$\frac{1}{2}\int_0^\infty s^2[J_{n-1}(ts)-J_{n+1}(ts)]\psi(s)\,ds$$

which exists uniformly with respect to t by (2). This argument is valid if $n=1, 2, \dots$. For $n=0$ we use the relation $J_0'(t)=-J_1(t)$. In this way the uniform existence of (8′) with respect to t is shown. As

$$\int_0^\infty s^3 J_n''(ts)\psi(s)\,ds = \frac{1}{2}\int_0^\infty s^3[J_{n-1}'(ts)-J_{n-1}'(ts)]\psi(s)\,ds$$

it follows, that (8_2) exists uniformly with respect to t. In the same way by induction we see, that the proposition on (8_k) holds for every $k=0, 1, 2, \dots$. By this, (7) implies

$$\varphi^{(k)}(t) = \int_0^\infty s^{k+1} J_n^{(k)}(ts)\psi(s)\,ds \tag{9}$$

and therefore

$$\varphi^{(k)}(0) = J_n^{(k)}(0)\int_0^\infty s^{k+1}\psi(s)\,ds = 0 \qquad (k = 0, 1, 2, \dots).$$

Using now the well-known power series expansion of the Bessel functions, we see, that $J_n^{(k)}(0)\neq 0$ for $k=n, n+2, n+4, \dots$, this yields the property d_n).

The property e_n) is exactly the statement (5) in the quoted Theorem of Griffith.

Now we will show, that also the opposite is true, i.e. if $\psi\colon \mathbf{C}\to\mathbf{C}$ is a function satisfying the properties a)—e), then the function φ defined by (7) belongs to the class $D(a)$.

The Hankel-transform (7) exists obviously because a) and b) and the same is also valid for the integrals (8_k) $(k=0, 1, 2, \dots)$ and therefore by (9) all derivatives of φ exists and are given by the formula (9). But on the other side the conditions a)—e) imply the properties (1)—(5) in Griffiths Theorem [8] and therefore $\varphi(t)=0$ for $t>0$. By representation (9) and condition d_n) $\varphi^{(k)}(0)=0$ $(k=0, 1, 2, 3, \dots)$ and define $\varphi(t)=0$ for $t\leqq 0$, $\varphi\in D(a)$.

The unanimous correspondence between φ and ψ follows from the corresponding property of Hankel-transformation. This completes the proof.

2. Let us introduce in $D(a)$ the usual topology by the following system of norms:

$$\|\varphi\|_p = \sup_{\substack{t\in(0,a)\\ k=0,1,\dots,p}} |\varphi^{(k)}(t)| \tag{10}$$

and in $G_n(a)$ a topology generated by the following system of norms:

$$\|\psi\|_p = \sup_{\substack{s\in C \\ k=0,1,2,\dots}} |s^k\psi(s)|e^{-a|\mathrm{Im}\, s|} \qquad (p=0,1,2,\dots). \tag{11}$$

Looking at this topologies by (9) the mapping H_n is not only an algebraic but also a topologic isomorphism between $G_n(a)$ and $D(a)$. We denote the inductive limit (for fixed n) of the spaces $G_n(a)$ and $D(a)$ for $a\to\infty$ by G_n resp. D_+; the notation of the dual spaces should be as usually G_n' resp. D_n'. If we denote by Z the ultra-distribution testing function space, then we see immediately (by a) and b)), that

$$G_n \subset Z, \tag{12}$$

this inclusion is proper and the topology inducted by Z in G_n is just that was defined under (11).

It is important to remark, that the topology in G_n is independent of the properties c$_n$), d$_n$) and e$_n$).

3. Let $u\in D_+'$ be an arbitrary distribution. Our aim is now to give a definition of the Hankel-transformation $H_n(u)$ of the distribution u. In order to make it, we need the following Lemma:

LEMMA. *Let* $\psi\in G_n$, *then*

$$s^{2p+1}\psi(s)\in G_{n+1}, \qquad s^{2p}\psi(s)\in G_n \qquad (p=0,1,2,\dots).$$

PROOF. The conditions a) and b) are fulfilled for all members of all the spaces G_n $(n=0,1,2,\dots)$, therefore for $s^{2p+1}\psi$ the conditions c$_{n+1}$), d$_{n+1}$) and e$_{n+1}$); for $s^{2p}\psi$ the conditions c$_n$), d$_n$) and e$_n$) are satisfied.

We can now give a definition for the Hankel-transform of a distribution in the following way: Let $u\in D_+'$ be an arbitrary distribution with support on $\mathbf{R}_+$, then $H_n(u)$ will be defined as that element of the space G_{n+1}' for which

$$\langle H_n(u), s\psi\rangle = \langle u, tH_n(\psi)\rangle \tag{13}$$

holds for every $\psi\in G_n$. Here $\langle\cdot,\cdot\rangle$ means the scalar product of a functional with the corresponding testing function. The definition (13) makes sense, as $H_n(\psi)= =\Phi\in D_+$ and by the Lemma $s\psi\in G_{n+1}$. We see immediately, that $H_n(u)$ is a special ultradistribution.

It is easy to see that the definition (13) overgoes into the classical Hankel-transformation definition if u is a regular distribution generated by a function f for which $t^{1/2}f(t)\in L^2(\mathbf{R}_+)$. In this case the right-hand side of (13) will be

$$\int_0^\infty tf(t)\varphi(t)\,dt \tag{14}$$

and this is by the Parseval identity concerning the usual Hankel-transformation ([6] Theorem 1) equal to

$$\int_0^\infty sH_n(f)(s)H_n(\varphi)(s)\,ds. \tag{15}$$

We know [5] if φ runs over the functional space D_+, then the set of values of the integrals (14) defines f (a.e.) uniquely and therefore also (15) defines $H_n(f)$ which is the classical Hankel-transform of f.

4. We will now give a definition of the Hankel-transformation H_n of a linear and continuous functional v defined over G_{n+1}. Also this can be done using the Parseval identity. $H_n(v)$ will be defined as a distribution in D'_+ as follows:

$$\langle H_n(v), tH_n(\psi)\rangle = \langle v, s\psi\rangle \qquad (\psi \in G_n). \tag{16}$$

This definition makes sense by the fact that $H_n(\psi)\in D_+$ $(\psi\in G_n)$.

We can now prove the following statement:

THEOREM 2. *Let* $u\in D_+$ *be a distribution, then*

$$H_n(H_n(u)) = u \tag{17}$$

and for any $v\in G'_{n+1}$

$$H_n(H_n(v)) = v \tag{18}$$

holds.

PROOF. Using the notation $H_n(\psi)=\varphi$, we have $\psi = H_n(\varphi)$, so we get immediately by (16) and (13)

$$\langle H_n(H_n(u)), t\varphi\rangle = \langle H_n(H_n(u)), tH_n(\psi)\rangle = \langle H_n(u), s\psi\rangle = \langle u, t\varphi\rangle$$

for all $\varphi\in D_+$, this proves the statement.

The relation (18) can be proved in a similar way.

5. Our aim is now to show that the Hankel-transform defined under (13) have the same formal properties as the classical Hankel-transforms of functions.

THEOREM 3. *Let* $u\in D'_+$ *be an arbitrary distribution and* B_n *the Bessel differential operator as defined in* (3), *then*

$$H_n(B_n u) = -s^2 H_n(u). \tag{19}$$

PROOF. The adjoint operator of B_n, denoted by B_n^* is the following

$$B_n^* f = f'' - \left(\frac{f}{t}\right)' - \frac{n^2}{t^2} f \qquad (n = 0, 1, 2, \ldots) \tag{20}$$

then

$$\langle B_n u, \varphi\rangle = \langle u, B_n^*\varphi\rangle \quad (\varphi\in D_+) \tag{21}$$

holds. A simple calculation shows

$$\frac{1}{t} B_n^*(t\varphi) = B_n(\varphi)$$

and therefore by a well-known theorem ([7], p. 61, (32) and p. 62, (35))

$$H_n\left(\frac{1}{t} B_n^*(t\varphi)(s)\right) = H_n(B_n(\varphi))(s) = -s^2\psi(s), \tag{22}$$

where $\psi=H_n(\varphi)$. By the inversion theorem of Hankel-transformation

$$-tH_n(s^2\psi)(t)=B_n^*(t\varphi)(t). \tag{23}$$

Using the definition (13) and the relations (21), (23), we get

$$\langle H_n(B_n(u)), s\psi\rangle=\langle B_n(u), t\varphi\rangle=\langle u, B_n^*(t\varphi)\rangle=-\langle u, tH_n(s^2\psi)\rangle=$$
$$=-\langle sH_n(u, s^2\psi\rangle=-\langle s[s^2H_n(u)], \psi\rangle$$

for any $\psi\in G_{n+1}$. This proves the statement (19).

6. Also an other well-known formula of the theory of Hankel-transformation of usual functions can be carried over for our generalized Hankel-transformation.

THEOREM 4. *Let $u\in D'_+$ be a distribution and denote by Du its (distributional) derivative. Then*

$$H_n(Du)=-\frac{1+n}{2n}sH_{n-1}(u)-\frac{1-n}{2n}H_{n+1}(u)\qquad (n=1,2,3,\ldots).$$

PROOF. By (13) we can write

(24)

$$\langle H_n(Du), t\psi\rangle=\langle Du, tH_n(\psi)\rangle=-\left\langle u, \frac{d}{dt}tH_n(\psi)\right\rangle=-\langle u, H_n(\psi)\rangle-\left\langle u, t\frac{d}{dt}H_n(\psi)\right\rangle.$$

But

$$(25)\qquad \begin{aligned} H_n(\psi)+t\frac{d}{dt}H_n(\psi)&=\int_0^\infty [sJ_n(ts)+ts^2J_n'(ts)]\psi(s)\,ds=\\ &=\int_0^\infty s\frac{d}{ds}(sJ_n(ts))\psi(s)\,ds.\end{aligned}$$

Applying a well-known relation (see [7], p. 512, formula (5) and (7)) we can continue (25) in the following way:

$$H_n(\psi)+t\frac{d}{dt}H_n(\psi)=\int_0^\infty [s(1-n)J_n(ts)+ts^2J_{n-1}(ts)]ds=$$
$$=\int_0^\infty\left[\frac{1-n}{2n}ts^2\frac{2n}{ts}J_n(ts)+ts^2J_{n-1}(ts)\right]\psi(s)\,ds=$$
$$=\int_0^\infty\left[\frac{1-n}{2n}ts^2J_{n-1}(ts)+\frac{1-n}{2n}ts^2J_{n+1}(ts)+ts^2J_{n-1}(ts)\right]\psi(s)\,ds=$$
$$=\int_0^\infty\left[\frac{1+n}{2n}ts^2J_{n-1}(ts)+\frac{1-n}{2n}ts^2J_{n+1}(ts)\right]\psi(s)\,ds=$$
$$=\frac{1+n}{2n}tH_{n-1}(s\psi)+\frac{1-n}{2n}tH_{n+1}(s\psi).$$

Substituting this expression into (24), we get

$$\langle sH_n(Du), \psi\rangle = -\frac{1+n}{2n}\langle u, tH_{n-1}(s\psi)\rangle = -\frac{1-n}{2n}\langle u, tH_{n+1}(s\psi)\rangle =$$

$$= -\frac{1+n}{2n}\langle sH_{n-1}(u), s\psi\rangle - \frac{1-n}{2n}\langle sH_{n+1}(u), s\psi\rangle =$$

$$= -\frac{1+n}{2n}s\langle H_{n-1}(u), s\psi\rangle - \frac{1-n}{2n}s\langle H_{n+1}(u), s\psi\rangle$$

for every $\psi \in G_{n+1}$. This completes the proof.

REFERENCES

[1] Fenyő, I. S., Hankel-Transformation verallgemeinerter Funktionen, *Mathematica* **8** (1966), 235—242. *MR* **35** #3384.

[2] Zemanian, A. H., A distributional Hankel transformation, *SIAM J. Appl. Math.* **14** (1966), 561—576. *MR* **34** #1807.

[3] Zemanian, A. H., The Hankel transformation of certain distributions of rapid growth, *SIAM J. Appl. Math.* **14** (1966), 678—690. *MR* **35** #2093.

[4] Zemanian, A. H., Hankel transformation of arbitrary order, *Duke Math. J.* **34** (1967), 761—770. *MR* **36** #6883.

[5] Gelfand, I. M. and Schilow, G. E., *Verallgemeinerte Funktionen*, I. Berlin, 1960. *MR* **22** #5889.

[6] Paley, R. E. A. C. and Wiener, N., Fourier Transforms in the Complex Domain, *Amer. Math. Soc. Coll. Publ.* **19** (1934).

[7] Sneddon, I. N., *Fourier-Transforms*, McGraw-Hill Book Co., New York—Toronto—London, 1951. *MR* **13**—29.

[8] Griffith, J. L., Hankel transforms of functions zero outside a finite interval, *J. Proc. Roy. Soc. New South Wales* **89** (1955), 109—115. *MR* **17**—1066.

[9] Whittaker J. M. and Watson, G. N., *A Course of Modern Analysis*, Cambridge, 1946.

(Received March 25, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
REÁLTANODA U. 13—15.
H—1053 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 457—461

A GENERALIZATION OF URYSOHN'S METRIZATION THEOREM AND ITS SET-THEORETIC CONSEQUENCES

B. M. PÖTSCHER

Abstract

In this paper we show that, for $\mu>0$, the existence of an embedding of every ω_μ-metrizable space X with weight less or equal to ω_μ into an ω_μ-compact, ω_μ-metrizable space depends on the underlying model of set theory thus establishing another link between the theory of ω_μ-metrizable spaces and set theory. In contrast to this in the case $\mu=0$ the proof of Urysohn's metrization theorem yields an embedding of every separable, metrizable space into a compact, metrizable one, namely the Hilbert cube. As a by-product, generalizing Urysohn's metrization theorem we obtain a characterization of locally ω_μ-separable, ω_μ-metrizable spaces and in the case $\mu=0$ a characterization of those metrizable spaces which can be embedded into locally compact, metrizable ones.

1. Introduction[1]

Let us first recall some definitions concerning ω_μ-metric spaces. Let $(G, +, \leqq)$ be an ordered abelian group with cofinality ω_μ, i.e. there exists a strictly decreasing ω_μ-sequence $\{g_\alpha\colon \alpha<\omega_\mu\}$ in G converging to the element $0\in G$ and ω_μ is the smallest ordinal with this property. Clearly ω_μ has then to be regular. Let X be a set and $d\colon X^2\to G$ be a function such that (i) $d(x, y)\geqq 0$, (ii) $d(x, y)=0$ iff $x=y$, (iii) $d(x, y)=d(y, x)$ and (iv) $d(x, y)\leqq d(x, z)+d(z, y)$. Such a function d is called an ω_μ-metric and (X, d) an ω_μ-metric space.[2] A topology on (X, d) is defined in an obvious way. A topological space is called ω_μ-metrizable if its topology can be generated by some ω_μ-metric. The ω_0-metrizable spaces are exactly the metrizable ones. For further details on ω_μ-metrizable spaces see e.g. [4], [5], [9], [10], [11], [14]. A space is called ω_μ-additive if the intersection of fewer than ω_μ open sets is open again. Note that every ω_μ-metrizable space is ω_μ-additive and that every topological space is ω_0-additive. Furthermore every ω_μ-metrizable space is paracompact. A space is ω_μ-compact iff every open cover admits a subcover of cardinality less than ω_μ. For ω_μ-metrizable spaces this is equivalent to the property that every ω_μ-sequence has at least one cluster point. A space is called locally ω_μ-compact if every point has an ω_μ-compact neighborhood. Following [6] we denote by $w(X)$ and $d(X)$ the weight and density of X, respectively. Obviously, for ω_μ-metric spaces $w(X)\leqq\omega_\mu$

1980 *Mathematics Subject Classification.* Primary 54E35, 54A35, 54D35, 54E15.

Key words and phrases. ω_μ-metrizable spaces, Urysohn's theorem, ω_μ-compactness, consistency.

[1] I would like to thank H. C. Reichel for helpful comments.

[2] Note that the definition of an ω_μ-metric, e.g. in [7], differs slightly from ours. In [7] ω_μ is not required to be the smallest ordinal with the above properties. If d is an ω_ν-metric in the sense of [7] it is an ω_μ-metric in our sense where $\omega_\mu=\operatorname{cof}\omega_\nu$ and vice versa. (A typographical error occurred in line 2 of [7] where μ should be replaced by ω_μ.)

iff $d(X)\leqq\omega_\mu$. A space is called locally ω_μ-separable (has locally an ω_μ-basis, respectively) if every point $x\in X$ has a neighborhood $U(x)$ with $d(U(x))\leqq\omega_\mu$ ($w(U(x))\leqq\omega_\mu$, respectively) as a subspace. We use ω_μ as an ordinal as well as a cardinal; ω_μ^+ denotes its cardinal successor. $2^{\underset{\smile}{\omega_\mu}}$ denotes $\sup\{2^{\omega_\nu}\colon \nu<\mu\}$. If $\mathcal{U}$ is a family of subsets of X then $\operatorname{ord}(x,\mathcal{U})=\operatorname{card}\{U\in\mathcal{U}\colon x\in U\}$. By $\operatorname{st}(U,\mathcal{U})$ we denote the set $\bigcup\{V\in\mathcal{U}\colon V\cap U\neq\emptyset\}$ and $\operatorname{st}^n(U,\mathcal{U})=\operatorname{st}(\operatorname{st}^{n-1}(U,\mathcal{U}),\mathcal{U})$.

2. The main result

Before presenting the main result we give a generalization of Urysohn's metrization theorem and as a corollary obtain a characterization of metrizable spaces which can be embedded into locally compact, metrizable ones.

LEMMA 1. *Let X be a topological space. If there is an open cover $\mathcal{U}$ of X such that any $U\in\mathcal{U}$ meets at most ω_μ many $V\in\mathcal{U}$ and if for every $U\in\mathcal{U}$ $d(U)\leqq\omega_\mu$ ($w(U)\leqq\omega_\mu$, respectively) holds then X is the disjoint union of spaces X_α, $\alpha\in A$, with $d(X_\alpha)\leqq\omega_\mu$ ($w(X_\alpha)\leqq\omega_\mu$, respectively).*

PROOF. Define for $U\in\mathcal{U}$ the set $G(U)=\bigcup\{st^n(U,\mathcal{U})\colon n\in\mathbf{N}\}$. Then $G(U)$ is of course open and $d(G(U))\leqq\omega_\mu$. Obviously, $G(U)\neq G(V)$ implies $G(U)\cap G(V)=\emptyset$. The proof of the statement in parenthesis is exactly the same.

LEMMA 2. *Let X be a topological space admitting an open cover $\mathcal{U}$ such that $\operatorname{ord}(x,\mathcal{U})\leqq\omega_\mu$ for all $x\in X$ and such that every member $U\in\mathcal{U}$ fulfils $d(U)\leqq\omega_\mu$. Then every $U\in\mathcal{U}$ meets at most ω_μ many members $U'\in\mathcal{U}$.*

PROOF. Fix $U\in\mathcal{U}$ and let S be a dense set in U of cardinality $\leqq\omega_\mu$. Define a function f from the set $\mathcal{V}=\{U'\in\mathcal{U}\colon U'\cap U\neq\emptyset\}$ to S by $f(U')=s$ where $s\in S$ is chosen such that $s\in U'\cap U$. Then $\mathcal{V}=\bigcup\{f^{-1}(s)\colon s\in S\}$ and $\operatorname{card}\mathcal{V}\leqq\leqq\omega_\mu\cdot\sup\{\operatorname{card}f^{-1}(s)\colon s\in S\}\leqq[\omega_\mu\cdot\sup\operatorname{ord}(s,\mathcal{U})]\leqq\omega_\mu$.

THEOREM 1. *Let X be a T_1-space and ω_μ be regular. Then the following are equivalent:*

(i) *X is ω_μ-metrizable and locally ω_μ-separable.*

(ii) *X is regular, ω_μ-additive and can be decomposed into disjoint open (and hence also closed) pieces X_α, $\alpha\in A$, each of which fulfils $w(X_\alpha)\leqq\omega_\mu$.*

(iii) *X is a regular metacompact, ω_μ-additive Hausdorff space and has locally an ω_μ-basis.*

(iv) *X is regular, ω_μ-additive and there exists an open cover $\mathcal{U}$ of X with $\operatorname{ord}(x,\mathcal{U})\leqq\omega_\mu$ and every $U\in\mathcal{U}$ fulfils $w(U)\leqq\omega_\mu$.*

PROOF. That (iv) implies (ii) follows from Lemma 1 and 2. (ii) clearly implies (iv). Also (ii) implies (iii) and (i) since every ω_μ-additive, regular T_1-space X with $w(X)\leqq\omega_\mu$ is ω_μ-metrizable, see [11]. (i) implies (iii), since every ω_μ-metrizable space is paracompact, see e.g. [9], [10]. Moreover, (iii) implies (iv): suppose (iii) holds, and for every $x\in X$, take an open neighborhood $V(x)$ with $w(V(x))\leqq\omega_\mu$. The collection $\{V(x)\colon x\in X\}$ now admits a point finite open refinement $\mathcal{U}$ and every $U\in\mathcal{U}$ fulfils $w(U)\leqq\omega_\mu$ since $U\subset V(x)$ for some $x\in X$. Now $\mathcal{U}$ fulfils exactly the requirements of (iv).

REMARK. (ii), (iv) are equivalent also if "$w(\cdot)\leqq\omega_\mu$" is replaced by "$d(\cdot)\leqq\omega_\mu$". If (ii), (iii) or (iv) hold for singular ω_μ then X will be discrete. Hence (ii)$\leftrightarrow$(iii)$\leftrightarrow$(iv) again. Note that every discrete space is ω_μ-metrizable for every (regular) ω_μ and fulfils (i)—(iv) for every ω_μ. Furthermore note that a non-discrete space is ω_μ-metrizable at most for one ω_μ, see e.g. [10].

The next corollary characterizes locally separable, metrizable spaces as those which can be embedded into locally compact, metrizable ones.

COROLLARY 1. *Let X be a T_1-space and $\mu=0$. Then the following is equivalent to each of* (i)—(iv) *of Theorem 1:*
(v) *X can be densely embedded into a locally compact metrizable space.*

PROOF. (v) clearly implies (i). (ii) implies (v) since every component of X can be densely embedded into a compact metric space by Urysohn's theorem.

The analogous question for $\mu>0$, whether any X fulfilling one of the conditions of Theorem 1 can be densely embedded into a locally ω_μ-compact, ω_μ-metrizable space Y, is not so easy to answer. First of all, because of Theorem 1 (ii), we can assume without loss of generality that X fulfils $w(X)\leqq\omega_\mu$. Furthermore, if such a Y exists we can construct the one-point ω_μ-compactification Y^* (in a similar way as the usual one-point compactification) and obviously Y^* is ω_μ-metrizable since $d(X)=d(Y)\leqq\omega_\mu$. Hence without loss of generality we are looking for ω_μ-compact Y's. Now trying to mimic the proof of the Urysohn metrization theorem one ends up with an embedding of X into $(2^{\omega_\mu})_{\omega_\mu}$ (i.e. 2^{ω_μ} with the smallest ω_μ-additive topology containing the Tychonoff product-topology, see e.g. [1]) which is ω_μ-metrizable, see e.g. [10]. But for $\mu>0$ this space is ω_μ-compact if and only if ω_μ is a weakly compact cardinal, whose existence is independent of ZFC and is not known to be consistent with ZFC, see e.g. [1], [8]. But this does not definitely answer our question. Let us first concentrate on the case $\omega_\mu=\omega_1$: let X be an ω_1-metrizable space with $w(X)\leqq\omega_1$ and card $X>\omega_1$ (e.g. $X=(2^{\omega_1})_{\omega_1}$ which fulfils $w(X)\leqq\omega_1$ iff CH holds, see [11], p. 130) and assume that it is provable in ZFC+CH that X admits an embedding into an ω_1-compact, ω_1-metrizable Y. Now clearly card $Y\geqq$ $\geqq$card $X>\omega_1$. Since it is proved in [7] that the existence of an ω_1-compact, ω_1-metrizable space with cardinality bigger than ω_1 is equivalent to the existence of a Kurepa tree with no Aronszajn subtree, we would have that the existence of Kurepa trees is provable in ZFC+CH. But now it is shown in [2], [3], [12], that ZFC+CH+ + "there exists no Kurepa tree" is consistent provided that ZFC+ "there exists an inaccessible cardinal" is consistent. So we have arrived at the following:

THEOREM 2. *Provided that ZFC+"there exists an inaccessible cardinal" is consistent, the statement that every ω_1-metrizable space X with $w(X)\leqq\omega_1$ can be embedded into an ω_1-compact, ω_1-metrizable space Y is independent of ZFC+CH.*

We do not know whether it is consistent with ZFC or ZFC+CH to assume the existence of such an embedding. Since under CH the space $(2^{\omega_1})_{\omega_1}$ is such an X as already said above and since every other ω_1-metrizable X with $w(X)\leqq\omega_1$ can be embedded into $(2^{\omega_1})_{\omega_1}$ it is sufficient to consider the case $X=(2^{\omega_1})_{\omega_1}$ when looking for consistency; in the light of the proof of Theorem 3, (ii)$\to$(iii), of [7], p. 226, one would have to show that it is consistent to assume that $(2^{\omega_1})_{\omega_1}$ embeds into a Kurepa line associated to a (very normal) Kurepa tree with no Aronszajn subtree. In case $\mu>1$ and ω_μ is a successor cardinal an analogue of Theorem 2 holds if CH is replaced by GCH: Firstly observe that $X=(2^{\omega_\mu})_{\omega_\mu}$ fulfils $w(X)\leqq\omega_\mu$ as long as $2^{\underset{\smile}{\omega_\mu}}\leqq\omega_\mu$, see [11], which is implied by GCH. Secondly, similar as before, the existence of an ω_μ-compact, ω_μ-metrizable space Y with $\operatorname{card} Y>\omega_\mu$ is equivalent to the existence of an ω_μ-Kurepa tree[3] with no ω_μ-Aronszajn subtree as is shown in [7]. Thirdly, by a result of Silver [12], it is independent of ZFC+GCH (again provided ZFC+ "there exists an inaccessible cardinal" is consistent) that an ω_μ-Kurepa tree exists.

Now if ω_μ is a strong limit cardinal (every limit is strong limit under GCH) then an ω_μ-Kurepa tree obviously exists, see e.g. [8], p. 313, so in this case we cannot prove a result analogous to Theorem 2 in the same way. Without assuming GCH there may be limit cardinals ω_μ which are not strong limits and nothing is then known about the existence of ω_μ-Kurepa trees (in this case $2^{\underset{\smile}{\omega_\mu}}>\omega_\mu$ might occur and then we cannot longer use $(2^{\omega_\mu})_{\omega_\mu}$ as a candidate for X since then $w\big((2^{\omega_\mu})_{\omega_\mu}\big)>\omega_\mu$. In fact the existence of an ω_μ-metrizable space X with $w(X)\leqq\omega_\mu$, $\operatorname{card} X>\omega_\mu$ is equivalent[4] to the existence of an $(\omega_\mu, \omega_\mu^+)$-tree T with more than ω_μ cofinal branches, as is seen as follows: given such an X we can find a basis $\mathscr{B}=\bigcup\{\mathscr{B}_\alpha\colon \alpha<\omega_\mu\}$, $\operatorname{card}\mathscr{B}\leqq\omega_\mu$ such that $\mathscr{B}_\alpha$ is a partition of X into clopen sets and $\mathscr{B}_\alpha$ refines $\mathscr{B}_\beta$ when $\alpha>\beta$, see e.g. [10]. The levels of T are $\mathscr{B}_\alpha$, $\alpha<\omega_\mu$, the partial order is given by the reversed set-theoretic inclusion and every point $x\in X$ gives raise to one cofinal branch of T in an obvious way. Conversely, given T let $X=\{b\subset T\colon b \text{ is a cofinal branch}\}$ and $\mathscr{U}=\big\{U_t\colon U_t=\{b\colon t\in b\},\ t\in T\big\}$ as in [7]. Then it is easily seen using similar arguments as in [7] that $\mathscr{U}$ is a base for a topology on X such that $\operatorname{card}\mathscr{U}=\operatorname{card} T\leqq\omega_\mu$, $\mathscr{U}$ is a rank one base, i.e. every two members of $\mathscr{U}$ are disjoint or one is contained in the other, and such that X is an ω_μ-additive T_1-space. But then X is clearly regular and hence ω_μ-metrizable, see [11], p. 129).

Note that it is shown in [13], Theorems 5.5 and 6.1, that (for every regular ω_μ) the existence of an ω_μ-Kurepa tree with no ω_μ-Aronszajn subtree is consistent with ZFC and this shows that it is consistent with ZFC to assume the existence of an ω_μ-compact, ω_μ-metrizable space Y with $\operatorname{card} Y>\omega_\mu$, see [7]. However, as in the case $\mu=1$, we do not know whether it is consistent with ZFC to assume the existence of an embedding of X into Y or not. Furthermore note that GCH holds in the model constructed in Theorem 5.5 of [13], so in this model all limits are strong limits.

REFERENCES

[1] Comfort, W. W. and Negrepontis, S., *The theory of ultrafilters*, Springer-Verlag, Berlin—Heidelberg—New York, 1974. *MR* **53** #135.
[2] Devlin, K. J., $\aleph_1$-trees, *Ann. Math. Logic* **13** (1978), 267—330. *MR* **80c:** 03053.
[3] Devlin, K. J., Concerning the consistency of the Souslin hypothesis with the continuum hypothesis, *Ann. Math. Logic* **19** (1980), 115—125. *MR* **82a:** 03049.

[3] We use the notion of an ω_μ-Kurepa tree as defined in [13]. If ω_μ is a successor cardinal an ω_μ-Kurepa tree exists iff an ω_μ-Kurepa tree in the sense of [8] exists.

[4] Without loss of generality we can assume X to have no isolated points: decompose $X=X_1\cup X_0$ where X_1 consists of all $x\in X$ such that every neighbourhood contains more than ω_μ many points. Then clearly $\operatorname{card} X_0\leqq\omega_\mu$ and X_1 is dense in itself.

[4] HUŠEK, M. and REICHEL, H. C., Topological characterizations of linearly uniformizable spaces, *Topology Appl.* **15** (1983), 173—188. *MR* **84d:** 54058.
[5] JUHÁSZ, I., Untersuchungen über ω_μ-metrisierbare Räume, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **8** (1965), 129—145. *MR* **33** #3257.
[6] JUHÁSZ, I., *Cardinal functions in topology — ten years later*, Math. Centre Tracts 123, Mathematisch Centrum, Amsterdam, 1980. *MR* **82a:** 54002.
[7] JUHÁSZ, I. and WEISS, W., On a problem of Sikorski, *Fund. Math.* **100** (1978), 223—227. *MR* **80g:** 54006.
[8] LEVY, A., *Basic set theory*, Perspectives in Math. Logic, Springer-Verlag, Berlin—Heidelberg—New York, 1979. *MR* **80k:** 04001.
[9] NYIKOS, P. and REICHEL, H. C., On uniform spaces with linearly ordered bases II, *Fund. Math.* **93** (1976), 1—10. *MR* **55** #13390.
[10] PÖTSCHER, B. M., Some results on ω_μ-metric spaces, *Annales Univ. Sci. Budapest. Eötvös Sect. Math.* **25** (1982), 3—18. *MR* **84c:** 54046. (Corrections, *Annales Univ. Sci. Budapest. Eötvös Sect. Math.* **28** (1985), 283.)
[11] SIKORSKI, R., Remarks on some topological spaces of high power, *Fund. Math.* **37** (1950), 125—136. *MR* **12**—727.
[12] SILVER, J. H., The independence of Kurepa's conjecture and two-cardinal conjectures in model theory, *Axiomatic Set Theory*, Proc. Symp. in Pure Math., Am. Math. Soc., 1971, 383—390. *MR* **43** #3112.
[13] TODORČEVIĆ, S. B., Trees, subtrees and order types, *Ann. Math. Logic* **20** (1981), 233—268. *MR* **82m:** 03062.
[14] WANG SHU-TANG, Remarks on ω_μ-additive spaces, *Fund. Math.* **55** (1964), 101—112. *MR* **29** #4022.

(Received March 26, 1985)

INSTITUT FÜR ÖKONOMETRIE UND
OPERATIONS RESEARCH
ABTEILUNG FÜR ÖKONOMETRIE
TECHNISCHE UNIVERSITÄT WIEN
ARGENTINIERSTRASSE 8
A—1040 WIEN
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 463—469

ÜBER EINE SYMMETRISCHE SCHROTUNG MIT EINER CAYLEYFLÄCHE ALS GRUNDFLÄCHE

M. HUSTY

Herrn o. Prof. Dr. H. Vogler zum 50. Geburtstag gewidmet

1. J. Krames hat einer Reihe bemerkenswerter Arbeiten [3]—[9] Zwangläufe Σ/Σ' untersucht, die sich durch gleichsinnige Kongruenz von Gang- und Rastaxoid auszeichnen. Die einzelnen Lagen des Gangsystems Σ werden dabei durch Spiegelungen an einer stetigen Schar von Geraden $e(t)\in\Sigma'$ erhalten. Die Geraden $e(t)$ erfüllen eine Regelfläche Γ, die als *Grundfläche* der symmetrischen Schrotung bezeichnet wird (vgl. [3], S. 395). Wir untersuchen in dieser Arbeit jene symmetrischen Schrotungen, die eine *Cayleyfläche* zur Grundfläche haben.

Da zwei benachbarte Erzeugende e, e' durch eine Schraubung um das Gemeinlot ineinander übergehen, gilt dies auch für die zugehörigen Systemlagen Σ und Σ'. Aus dem Grenzübergang $e\to e'$ folgt nun, daß die momentane Schraubachse mit der Zentraltangente von Γ zusammenfällt, und das Rastaxoid Π' von der Zentraltangentenfläche gebildet wird. Das Gangaxoid Π ist eine dazu gleichsinnig kongruente Fläche und geht aus Π' durch Spiegelung an der gemeinsamen Erzeugenden e hervor.

Die Bahnkurven der Punkte von Σ' werden bei einer symmetrischen Schrotung nach [3] durch zentrische Verdoppelung der Fußpunktskurven der Grundfläche Γ erhalten.

2. In einem kartesischen Koordinatensystem $1:x:y:z=x_0:x_1:x_2:x_3$ kann eine Cayleysche Fläche Γ in der Normalform

$$3x-3zy+z^3=0 \tag{2.1}$$

dargestellt werden. Sie besitzt in dieser Aufstellung die Ferngerade der Grundrißebene als Torsallinie und die Fernebene ω $(x_0=0)$ als Torsalebene. In Plückerkoordinaten lautet (2.1)

$$p_1:p_2:p_3:p_4:p_5:p_6=\left(u:1:0:u:-u^2:\frac{1}{3}u^3\right) \tag{2.2}$$

und man erhält nach [2], S. 319 die zu Γ gehörige symmetrische Schrotung in der Form:

$$\begin{aligned} X&=\frac{1}{u^2+1}\left[(u^2-1)x+2uy-\frac{2}{3}u^3\right]\\ Y&=\frac{1}{u^2+1}\left[2ux+(1-u^2)y+\frac{2}{3}u^4\right]\\ Z&=2u-z \qquad u\in\mathbf{R}(-\infty,+\infty). \end{aligned} \tag{2.3}$$

1980 *Mathematics Subject Classification.* Primary 53A17.
Key words and phrases. Raumkinematik, symmetrische Schrotung.

Das Rastaxoid dieser Bewegung ist ein Zylinder über dem Grundriß der Striktionslinie von Γ mit der Gleichung:

$$9x^2 = 4y^3. \tag{2.4}$$

Dieser Grundriß ist eine *Neilsche Parabel* mit dem Brennpunkt $\left(0, -\frac{1}{3}\right)$. Die Grundrißbewegung kann daher als eine symmetrische Rollung von zwei bezüglich ihrer gemeinsamen Tangenten spiegelbildlichen Neilschen Parabeln gedeutet werden.[1] Abb. 1 zeigt die Grundrißbewegung, wobei neben dem Polkurvenpaar p, p_0 einige Punktbahnen dargestellt sind.

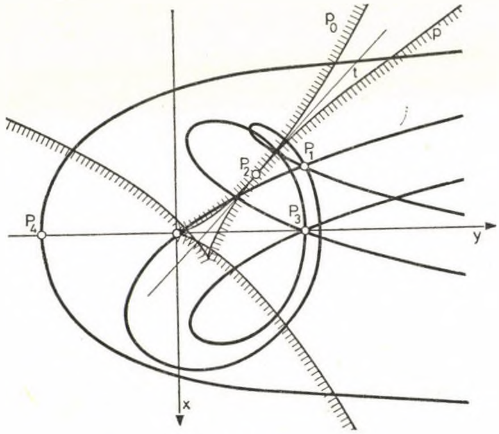


Abb. 1

Die Bahnkurven der Punkte bei diesem Zwanglauf sind zirkulare Quartiken, die im Fernpunkt der y-Achse die Fernebene berühren. Da die Neilsche Parabel eine Kurve dritter Klasse ist, ist der Grundriß des Ausgangspunktes dreifacher Punkt des Grundrisses der Bahnkurven. Damit besitzen die Raumkurven die erstprojizierende Gerade durch den Ausgangspunkt als Trisekante. Die Bahnkurven sind somit Raumkurven 4. Ordnung zweiter Art. Sie liegen auf orthogonalen hyperbolischen Paraboloiden mit den gemeinsamen Fernerzeugenden $x_0 = x_1 = 0$ und $x_0 = x_3 = 0$. Diese hyperbolischen Paraboloide besitzen für einen festen Ausgangspunkt (x_0, y_0, z_0) die Darstellung

$$(x - x_0)(z + z_0) + 2(y - y_0) = 0. \tag{2.5}$$

Wir erhalten damit eine Schiebschar von untereinander kongruenten hyperbolischen Paraboloiden, deren Scheitel S aus dem Ausgangspunkt durch Spiegelung an der Ebene $z = 0$ hervorgehen; die Spiegelungsebene ist die asymptotische Ebene der Mittelerzeugenden der Cayleyfläche.

---

[1] Diese Grundrißbewegung tritt auch bei J. Krames in [9] auf, wo als Grundfläche eine konoidale Regelfläche fünften Grades verwendet wird.

SATZ 1. *Das Gang- bzw. Rastaxoid einer symmetrischen Schrotung mit einer Cayleyfläche als Grundfläche sind kongruente Zylinder, deren Normalschnitte Neilsche Parabeln sind. Die Bahnkurven dieses Bewegungsvorganges sind zirkulare Quartiken zweiter Art, die auf orthogonalen hyperbolischen Paraboloiden liegen.*

3. Bei den Bahnkurven einer symmetrischen Schrotung kann es nach Krames ([3], S. 399) nur dann zu einer Verringerung der Bahnkurvenordnung $2n$ (mit n wird die Ordnung der Grundfläche bezeichnet) kommen, wenn die Grundfläche Fernerzeugenden besitzt oder der Punkt des bewegten Systems Σ den durch die isotropen Erzeugenden der Grundfläche gelegten Minimalebenen angehört. Da die Existenz der Ferntorsalerzeugenden der Cayleyfläche die Reduktion der Bahnkurvenordnung auf vier bewirkt, ist eine weitere Reduktion nur für Punkte in den isotropen Tangentialebenen durch die beiden Minimalerzeugenden der Grundfläche zu erwarten. Diese Minimalebenen schneiden sich in einer z-parallelen Geraden g durch den Brennpunkt der Neilschen Parabel (2.4). Die Punkte von g besitzen die Bahnkurven

$$X = \frac{2}{3} u$$
$$Y = \frac{2}{3} u^2 - \frac{1}{3} \tag{3.1}$$
$$Z = 2u - z_0 \qquad z_0 \in \mathbf{R}(-\infty, +\infty)$$

und erweisen sich damit als *Parabeln* in den untereinander parallelen kreuzrißprojizierenden Ebenen

$$z = 3x - z_0. \tag{3.2}$$

Zusammenfassend gilt somit

SATZ 2. *Bei der auf die Cayleyfläche gegründeten symmetrischen Schrotung gibt es eine einparametrige Schar ebener Bahnkurven. Diese Bahnkurven sind Parabeln und entstehen, wenn man Punkte auf der z-parallelen Geraden durch den Brennpunkt der Neilschen Parabel* (2.4) *dem Bewegungsvorgang unterwirft.*

4. Unterwirft man eine Gerade g

$$\mathfrak{x} = \begin{vmatrix} a_1 \\ a_2 \\ a_3 \end{vmatrix} + \lambda \begin{vmatrix} v_1 \\ v_2 \\ v_3 \end{vmatrix} \tag{4.1}$$

der Bewegung (2.3), so entsteht eine Regelfläche Φ mit der Parameterdarstellung

$$X = \frac{1}{u^2+1}\left[(u^2-1)\lambda v_1 + 2u\lambda v_2 - \frac{2}{3} u^3\right]$$
$$Y = \frac{1}{u^2+1}\left[2u\lambda v_1 - (u^2-1)\lambda v_2 + \frac{2}{3} u^4\right] \tag{4.2}$$
$$Z = -\lambda v_3 + 2u$$

wobei o. B. d. A. in (4.1) $a_1=a_2=a_3=0$ gesetzt wurde. Diese Regelfläche ist i. a. von vierter Ordnung wie aus ihrer Darstellung in Plückerkoordinaten

$$\begin{aligned}(4.3)\quad & p_1:p_2:p_3:p_4:p_5:p_6 = v_1(u^2-1)+2u:2v_1u-(u^2-1)v_2:-(u^2-1)v_3: \\ & :\frac{2}{3}u^4v_3-2u[2v_1u-(u^2-1)v_2]:2u[v_1(u^2-1)+2uv_2]-\frac{2}{3}v_3u^3:-\frac{2}{3}u^3(v_1u+v_2)\end{aligned}$$

sofort ersichtlich ist und besitzt i. a. einen Drehkegel als Richtkegel. Abb. 2 zeigt einen Ausschnitt der Fläche für $v_1=1$, $v_2=-\frac{1}{2}$, $v_3=\frac{1}{2}$ in einer axonometrischen Darstellung.

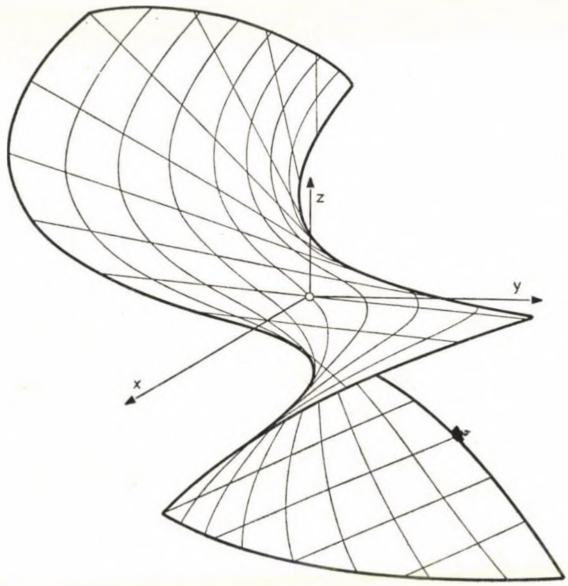


Abb. 2

Als Sonderfall ergibt sich eine Reduktion der Ordnung der entstehenden Regelfläche, wenn für die Ausgangsgerade g: $v_1=v_2=0$ gesetzt wird. Damit ergibt sich in (4.3)

$$(4.4)\qquad p_1:p_2:p_3:p_4:p_5:p_6 = 2u:1-u^2:0:2u(u^2-1):\ 4u^2:-\frac{2}{3}u^3$$

und die Fläche besitzt die algebraische Gleichung

$$(4.5)\qquad 12zy-12x+3z^2x-z^3=0.$$

Wie unmittelbar durch Homogenisieren von (4.5) einzusehen ist, besitzt diese Fläche die Ferngerade $x_0=x_3=0$ als Doppelgerade und (nach komplexer Erweiterung) die beiden konjugiert komplexen Torsalebenen

$$z = \pm 2i. \tag{4.6}$$

Sie ist daher eine Fläche *3. Ordnung vom Typ II* (vgl. [10], S. 176).

Ein weiterer Sonderfall ergibt sich, wenn die Ausgangsgerade parallel zur Grundrißebene $z=0$ ($v_3=0$) gewählt wird. Die Fläche besitzt dann die algebraische Gleichung:

$$24y\left(\frac{z^2}{4}-1+zv_2\right)-z^4 = 24x\left[z-v_2\left(\frac{z^2}{4}-1\right)\right]+2v_2z^3. \tag{4.7}$$

Als Schnitt mit der Fernebene $x_0=0$ ergibt sich die vierfache Gerade $x_3=x_0=0$, die über $\dot{p}_1\dot{p}_4+\dot{p}_2\dot{p}_5+\dot{p}_3\dot{p}_6=0$ (Bedingung für Torsallinien) als Torsallinie 2. Ordnung und einfache Leitgerade erkannt wird. Da sich in Ebenen $z=$konst. neben der Fernleitgeraden nur mehr jeweils eine Gerade ergibt stellt diese Ferngerade die Doppelkurve der Fläche dar. Die Fläche ist damit von XII. Art der Sturmschen Einteilung (vgl. [10], S. 269). Sie besitzt neben der Torsallinie 2. Ordnung noch zwei konjugiert komplexe Torsallinien in den Torsalebenen $z=\pm 2i$. Zusammenfassend gilt der

SATZ 3. *Unterwirft man eine Gerade der symmetrischen Schrotung* (2.3), *so entsteht eine Regelfläche 4. Ordnung. Ist die Ausgangsgerade parallel zur Grundrißebene, so entsteht eine konoidale Regelfläche 4. Ordnung der XII. Sturmschen Art; ist die Ausgangsgerade parallel zur y-Achse des Koordinatensystems, so entsteht eine Regelfläche 3. Ordnung vom Typ II.*

5. Weitere interessante Flächen ergeben sich, wenn Kreise der symmetrischen Schrotung (2.3) unterworfen werden. Wir wollen uns hier auf Kreise in Ebenen parallel zur Grundrißebene beschränken und werden dabei auf Flächen stoßen die in der erst kürzlich von O. Röschel in [15] ausführlich studierten Geometrie des galileischen Raumes $\mathbf{G}_3$ eine besondere Rolle spielen.

Wird ein allgemein liegender Kreis $((a+R\cos t,\ b+R\sin t,\ 0);\ t\in(+\infty, -\infty),\ R=\text{konst.}\in\mathbf{R})$ dem Zwanglauf unterworfen, so entsteht eine Fläche Ψ mit der algebraischen Gleichung:

$$\left(x-a-\frac{z}{3}\right)^2+\left(y-b-\frac{z^2}{3}\right)^2 = R^2. \tag{5.1}$$

Die Fläche ist somit eine Kreisfläche 4. Ordnung und schneidet die Fernebene $x_0=0$ in der doppelt zu zählenden Geraden $x_0=x_3=0$.

Bemerkung. Macht man die Ferngerade der Grundrißebene $x_0=x_3=0$ und die auf ihr liegenden absoluten Kreispunkte, sowie die Fernebene ω zum Absolutgebilde einer Cayley—Kleinschen Metrik, so erhält man den von O. Röschel [15] näher untersuchten galileischen Raum $\mathbf{G}_3$. In diesem Raum können die in (5.1) gefundenen Flächen als *Zykliden* gedeutet werden.

Abb. 3 zeigt eine isometrische Darstellung eines Teils der Fläche Ψ.

Diese Fläche kann neben der Erzeugung (2.3) auch als Kreisschiebfläche erzeugt werden. Die Schiebkurven sind dabei zur Bahn des Mittelpunktes bei der Bewegung (2.3) kongruente Kurven.

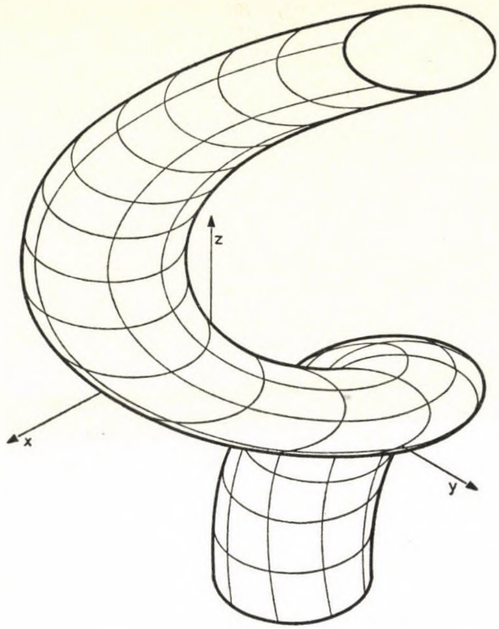


Abb. 3

Als Sonderfall ergibt sich in (5.1) mit $a=0$ und $b=-\frac{1}{3}$ die Fläche

$$\left(x-\frac{z}{3}\right)^2+\left(y+\frac{1}{3}-\frac{z^2}{3}\right)^2=R^2. \tag{5.2}$$

Da der Mittelpunkt des bewegten Kreises auf der Parabel (3.1) geführt wird, kann die Fläche durch Schiebung dieses Kreises längs einer Parabel oder nach obiger Bemerkung längs eines Kreises des $\mathbf{G}_3$ erzeugt werden. Sie kann daher in diesem speziell metrisierten Raum auch als *Torus vom Typ B* (vgl. [14]) aufgefaßt werden, wobei die in [14] angegebene Normalform durch die Transformation

$$x'=z, \qquad y'=y+\frac{1}{3}, \qquad z'=x-\frac{z}{3} \tag{5.3}$$

erreicht wird. Zusammenfassend gilt der

SATZ 4. *Wird ein der Ebene $z=0$ angehörender Kreis der symmetrischen Schrotung (2.3) unterworfen, so entsteht eine Kreisfläche 4. Ordnung, die auch als Zyklide des galileischen Raumes $\mathbf{G}_3$ aufgefaßt werden kann. Unter diesen Zykliden entstehen für $a=0$ und $b=-\frac{1}{3}$ in (5.1) Torusflächen vom Typ B des $\mathbf{G}_3$.*

LITERATURVERZEICHNIS

[1] BEREIS, R., Über die symmetrische Rollung, *Österreich. Ing.-Arch.* **7** (1953), 243—246. *MR* **15**—259.

[2] BOTTEMA, O. und ROTH, B., *Theoretical kinematics*, North-Holland Series, Amsterdam, 1979. *MR* **81c:** 70001.

[3] KRAMES, J., Über Fußpunktkurven von Regelflächen und eine besondere Klasse von Raumbewegungen (Über symmetrische Schrotungen I), *Monatsh. Math.* **45** (1937), 394—406.

[4] KRAMES, J., Zur Bricard'schen Bewegung, deren sämtliche Bahnkurven auf Kugeln liegen (Über symmetrische Schrotungen II), *Monatsh. Math.* **45** (1937), 407—417.

[5] KRAMES, J., Zur aufrechten Ellipsenbewegung des Raumes (Über symmetrische Schrotungen III), *Monatsh. Math.* **46** (1937), 48—50.

[6] KRAMES, J., Zur kubischen Kreisbewegung des Raumes (Über symmetrische Schrotungen IV), *Österreich. Acad. Wiss. Math.-Natur. Kl. Sitzungber. II* **146** (1937), 145—158.

[7] KRAMES, J., Zur Geometrie des Bennett'schen Mechanismus (Über symmetrische Schrotungen V), *Österreich. Acad. Wiss. Math.-Natur. Kl. Sitzungber. II* **146** (1937), 159—173.

[8] KRAMES, J., Die Borel—Bricard-Bewegung mit punktweise gekoppelten orthogonalen Hyperboloiden (Über symmetrische Schrotungen VI), *Monatsh. Math.* **146** (1937), 172—195.

[9] KRAMES, J., Über eine konoidale Regelfläche fünften Grades und die darauf gegründete symmetrische Schrotung, *Österreich. Acad. Wiss. Math.-Natur. Kl. Sitzungber. II* **190** (1981), 221—230. *MR* **83h:** 51044.

[10] MÜLLER, E. und KRAMES, J., *Konstruktive Behandlung der Regelflächen* (Vorlesungen über Darstellende Geometrie III), Leipzig—Wien, 1931.

[11] MÜLLER, H. R., *Sphärische Kinematik*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1962. *MR* **26** #3243.

[12] RÖSCHEL, O., Räumliche Zwangläufe mit einer zweiparametrigen Schar ebener Bahnkurven I, *Österreich. Acad. Wiss. Math.-Natur. Kl. Sitzungber. II* (im Druck).

[13] RÖSCHEL, O., Räumliche Zwangläufe mit einer zweiparametrigen Schar ebener Bahnkurven II, *Österreich. Acad. Wiss. Math.-Natur. Kl. Sitzungber. II* (im Druck).

[14] RÖSCHEL, O., Torusflächen im galileischen Raum G_3 (in Vorbereitung).

[15] RÖSCHEL, O., Die Geometrie des galileischen Raumes, Habilitationsschrift MU — Leoben, 1—20 (1984). Arbeitsbericht 20/1984 Inst. Math. Angew. Geom. MU — Leoben.

[16] SCHÖNFLIES, A., Über Bewegungen starrer Systeme im Fall cylindrischer Axenflächen, *Math. Ann.* **40** (1982), 317—331.

[17] TÖLKE, J., Ebene euklidische und sphärische symmetrische Rollungen, *Mech. Mach. Theory* **13** (1978), 187—198.

[18] WUNDERLICH, W., *Ebene Kinematik*, Bibliographisches Institut, Mannheim, 1970.

[19] WUNDERLICH, W., Kubische Zwangläufe, *Sitzungsber. Österr. Akad. Wiss.* (im Druck).

(Eingegangen am 10. Juni 1985)

INSTITUT FÜR MATHEMATIK UND ANGEWANDTE GEOMETRIE
MONTANUNIVERSITÄT LEOBEN
FRANZ-JOSEF-STRASSE 18
A—8700 LEOBEN
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 22 (1987), 471—484

ON A DISCRETE NONLINEAR OPERATIONAL DIFFERENTIAL EQUATION SYSTEM BASED ON THE DIRICHLET PRODUCT

T. FÉNYES

To the memory of K. Szilárd

Introduction

In the paper [1] we have discussed the system

$$D(x_i) = \frac{a_i}{\prod\limits_{\substack{k=1\\k\neq i}}^{n} x_k} + (f_i - \gamma_i)x_i, \qquad i = 1, 2, \ldots, n \tag{1}$$

of nonlinear algebraic differential equations defined in the Mikusiński operator field M. Here $n>1$ is an arbitrary integer, the M-operators $x_i \in M$ are the unknowns of the system (1), D is the symbol of the well-known algebraic derivative (see [3]), $a_i(t), f_i(t)$ are locally integrable functions given in $0 \leqq t < \infty$, moreover γ_i are given real numbers. The symbols are those used by Mikusiński (see [3]).

On the other hand, in the paper [2] we have given a discrete operational theory based on the so called Dirichlet product

$$\{a(n)\}\{b(n)\} = \left\{\sum_{\nu|n} a(\nu) b\left(\frac{n}{\nu}\right)\right\}, \qquad n = 1, 2, \ldots. \tag{2}$$

Here $a = \{a(n)\}$, $b = \{b(n)\}$ denote real-valued functions defined on the set of the natural numbers (see also [4]).

In this paper we shall give an operational treatment of the following nonlinear, algebraic differential equation system based on the Dirichlet product:

$$D(x_i) = \frac{a_i}{\prod\limits_{\substack{k=1\\k\neq i}}^{m} x_k} + f_i x_i, \qquad i = 1, 2, \ldots, m; \; m > 1. \tag{3}$$

Here a_i, f_i are given real-valued functions defined on the set of the natural numbers. The discrete operators x_i are the unknowns of (3). D denotes the "discrete" algebraic derivative (see [2], [4]).

We assume that the conditions

$$a_i \neq 0, \qquad i = 1, 2, \ldots, m \tag{4}$$

$$\sum_{i=1}^{m} a_i = 0 \tag{5}$$

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 6032/6319.

1980 *Mathematics Subject Classification.* Primary 44A40.
Key words and phrases. Mikusiński's operator calculus, operational differential equation system.

hold. (4) means that none of the functions a_i is identically zero, (5) means that their sum equals to the identically zero function.

We say that (3) has an operational solution if there exist M-operators x_i satisfying the system (3). If x_i is a function for every value of i, we say that (3) has a function solution. From (4) it follows that $x_i \neq 0$.

Papers [2], [4] contain number-theoretical applications of the operational calculus. In the discussion of (3) there appears the opposite case. We shall show that the existence problem of the operational solutions of (3) can be reduced — in general — to an interesting number-theoretical problem.

In Chapter 1 we briefly summarize the results of the paper [2], giving some generalizations of them, Chapter 2 contains the operational theory of the nonlinear problem defined earlier.

In what follows Z will denote the set of natural numbers.

§1. Discrete Mikusiński operators based on the Dirichlet product

Let $a = \{a(n)\}$ be an arbitrary real-valued function defined on Z. The symbol $a(n)$ denotes the value of this function for arbitrary fixed n.

Let E denote the set of the discrete functions. If we introduce in E the following two operations

(i) $$a+b := \{a(n)\}+\{b(n)\} = \{a(n)+b(n)\}, \quad \text{addition}$$

(ii) $$ab := \{a(n)\}\{b(n)\} = \left\{\sum_{\nu|n} a(\nu)\, b\left(\frac{n}{\nu}\right)\right\}, \quad \text{multiplication,}$$

then E becomes a commutative ring without divisor of zero and can be extended to a quotient field.

This is called the discrete Mikusiński operator field and is denoted by M. The elements of M are called M-operators.

The definition and properties of the "discrete" Dirac-function

We define the discrete Dirac-function by

$$\delta(N) = \{\delta(n, N)\},$$

where

$$\delta(n, N) = \begin{cases} 0, & \text{for} \quad n \neq N, \\ 1, & \text{for} \quad n = N, \end{cases} \qquad N \in Z.$$

For later purposes we enumerate some properties of the Dirac function.

PROPERTY 1. $\delta(N)\{a(n)\} = \{b(n)\}$,

(1.1) $$b(n) = \begin{cases} a\left(\dfrac{n}{N}\right), & \text{for} \quad N|n \\ 0, & \text{otherwise.} \end{cases}$$

(1.2) $$\delta(N_1)\delta(N_2) = \delta(N_1 N_2), \quad N_1, N_2 \in Z.$$

PROPERTY 2.

(1.3) $$x = \frac{\{a(n)\}}{\delta(N)} \in E, \qquad N \in Z$$

holds if and only if

(1.4) $$a(n) = 0$$

for those values of n for which N is not a divisor of n. If (1.4) holds, then

(1.5) $$x = \{a(nN)\}.$$

The field K of the real numbers can be embedded isomorphically into the operator field M. The common unit element of K, E, M is the function $\delta(1)$ and we write

$$\delta(1) = 1.$$

Moreover,

$$c\delta(1) = c, \quad c\{a(n)\} = \{ca(n)\}; \quad c \in K, \quad a \in E.$$

Every operator of the form

$$x = \frac{\{a(n)\}}{\{b(n)\}}$$

is a function if $b(1) \neq 0$.

The operator function $\delta(\alpha)$

For arbitrary rational number $\alpha = \frac{N_1}{N_2}$ we define

(1.6) $$\delta(\alpha) = \frac{\delta(N_1)}{\delta(N_2)}.$$

From this definition it follows that for $\alpha = N$ we have

$$\delta(\alpha) = \delta(N) = \{\delta(n, N)\}.$$

If

$$\frac{N_1}{N_2} = \frac{N_3}{N_4},$$

then

$$\delta\left(\frac{N_1}{N_2}\right) = \delta\left(\frac{N_3}{N_4}\right)$$

holds.

PROPERTY 3. Let α, β be arbitrary positive rational numbers, then

(1.7) $$\delta(\alpha)\delta(\beta) = \delta(\alpha\beta)$$

and it is easily seen that

(1.8) $$\delta\left(\frac{1}{\alpha}\right) = \frac{1}{\delta(\alpha)}$$

is also true.

The definition of the ring E^*

Let $E^* \subset M$ be the subset of M whose elements are of the form

$$x = \frac{\{a(n)\}}{\delta(N)} \qquad N \in Z, \ a \in E. \tag{1.9}$$

E^* is a ring and, by choosing $N=1$, we have

$$E \subset E^*.$$

PROPERTY 4. Obviously

$$x = \frac{\{a(n)\}}{\delta(\alpha)} \in E^*, \quad \alpha = \frac{N_1}{N_2}, \quad (N_1, N_2 \text{ are relatively primes}).$$

Moreover, $x \in E$ if and only if

$$a(n) = 0$$

for those values of n for which N_1 is not a divisor of n. If the condition is satisfied, we have

$$x(n) = \begin{cases} a\left(\dfrac{nN_1}{N_2}\right), & \text{for} \quad N_2 | n, \\ 0, & \text{otherwise.} \end{cases}$$

Definition of the convergence in the ring E

Let $\{a_k(n)\} \in E$, $(k=1, 2, \ldots)$ be an infinite sequence of functions. By definition

$$\lim_{k \to \infty} \{a_k(n)\} = \{a(n)\} \tag{1.10}$$

if for every fixed n

$$\lim_{k \to \infty} a_k(n) = a(n)$$

(see [5]). This convergence can be extended to infinite series of functions as usual. Let

$$f(z) = \sum_{k=0}^{\infty} \beta_k z^k, \qquad \beta_k \in K$$

be an arbitrary entire function of the complex variable z. Then

$$f(a) = \sum_{k=0}^{\infty} \beta_k \{a(n)\}^k, \qquad a \in E, \ a^0 = 1 \tag{1.11}$$

holds in the sense of the convergence defined above. We have

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!},$$

having the property

$$e^a e^b = e^{a+b}, \qquad a, b \in E,$$

moreover, if we write

$$e^a = \{e_a(n)\},$$

so

(1.12) $$e_a(1) = e^{a(1)}$$

holds.

The algebraic derivation and integration (see also [4]).

For the sake of easy reading we recapitulate some definitions and facts of the algebraic derivation and integration (see also [4]).

(1.13)
$$D(a) = \{-\log n \cdot a(n)\}, \quad a \in E,$$
$$D\left(\frac{a}{b}\right) = \frac{bD(a) - aD(b)}{b^2}, \quad a, b \in E, \quad \frac{a}{b} \in M.$$

PROPERTY 5.

(1.14)
$$D\left[\frac{a}{\delta(\alpha)}\right] = \frac{\left\{-\log\frac{n}{\alpha} \cdot a(n)\right\}}{\delta(\alpha)} \in E^*, \quad a \in E, \quad \alpha = \frac{N_1}{N_2},$$
$$D[\delta(\alpha)] = -\log \alpha \cdot \delta(\alpha).$$

PROPERTY 6.

(1.15) $$D(e^a) = D(a)e^a.$$

If for a given $x \in M$ there exists a $y \in M$ such that

$$D(y) = x,$$

we say that x is algebraic integrable and we write

$$y = \int x.$$

PROPERTY 7. If $x \in M$ and

$$D(x) = 0,$$

then x is an arbitrary number. (In paper [2] we have shown this fact under the stronger assumption $x \in E^*$.)

Let

$$x = \frac{a}{b}, \quad a, b \in E.$$

Since

$$D(x)=\frac{D(a)b-aD(b)}{b^2}=0,$$

we have

$$D(a)b-aD(b)=0.$$

From this equality we can deduce the relation

$$a(n)=kb(n), \qquad k\in K, \quad n=1,2,\dots .$$

So x is an arbitrary number.

Two algebraic integrals of an operator may differ only by an arbitrary number.

The algebraic differentiation and integration is a linear operation over the field of the real (complex) numbers.

The algebraic integration in E^*

PROPERTY 8. The operator

$$x=\frac{a}{\delta(\alpha)}, \quad a\in E, \quad \alpha=\frac{N_1}{N_2} \tag{1.16}$$

is algebraic integrable in the ring E^* if and only if

$$\text{either} \quad \alpha\neq N, \quad N\in Z,$$

$$\text{or} \quad \alpha=N \quad \text{and} \quad a(N)=0$$

holds true. Every algebraic integral of (1.16) belonging to E^* is given by

$$\int\frac{a}{\delta(\alpha)}=\frac{\left\{-\dfrac{a(n)}{\log\dfrac{n}{\alpha}}\right\}}{\delta(\alpha)}+c, \quad c\in K, \tag{1.17}$$

where in the case of $\alpha=N$ the symbol

$$\frac{a(N)}{\log\dfrac{N}{N}}$$

denotes an arbitrary real (complex) number.

For $\alpha=1$ we have that a is integrable if and only if $a(1)=0$, and

$$\int a=\left\{-\frac{a(n)}{\log n}\right\}+c.$$

The homogeneous first order differential equation

Let us consider the differential equation

$$D(x)-fx=0, \qquad f\in E. \tag{1.18}$$

In the paper [2] we have proved the following

THEOREM. *If* $\alpha=e^{-f(1)}$ *is irrational, then* (1.18) *has only the trivial zero solution in* E^*. *If* $\alpha=e^{-f(1)}$ *is rational, then the general solution of* (1.18) *in* E^* *is of the form*

$$x = c\delta(\alpha)\exp\left[\int(f-f(1))\right], \quad c\in K \tag{1.19}$$

being a function for $c\neq 0$ *if and only if* α *is natural.*

We shall show the converse statement. If the differential equation

$$D(x)-qx = 0, \qquad q\in M \tag{1.20}$$

has a nontrivial solution $x\in M$, then $q\in E$, and $e^{-q(1)}$ is rational.
Indeed, let

$$x = \frac{a}{b}\in M \tag{1.21}$$

a nontrivial solution of (1.20). We have

$$\frac{bD(a)-aD(b)}{b^2}-q\frac{a}{b} = 0$$

and

$$q = \frac{D(a)}{a}-\frac{D(b)}{b}.$$

We have only to show that the operator

$$r = \frac{D(a)}{a} = \frac{\{-\log n\cdot a(n)\}}{\{a(n)\}} \tag{1.22}$$

is a function and $e^{-r(1)}$ is integer. The values of the function $r=\{r(n)\}$ can be determined from the equation

$$-\log n\cdot a(n) = \sum_{\nu|n} a(\nu)r\left(\frac{n}{\nu}\right). \tag{1.23}$$

This is trivial for $a(1)\neq 0$. Let now

$$a(1) = a(2) = \ldots = a(N-1) = 0, \qquad a(N)\neq 0, \quad N>1.$$

By choosing $n=N$, we have

$$-\log N\cdot a(N) = \sum_{\nu|N} a(\nu)r\left(\frac{N}{\nu}\right) = a(N)r(1)$$

so

$$r(1) = -\log N$$

holds. If we have already determined the values

$$r(1), r(2), \ldots, r(k),$$

then $r(k+1)$ can be calculated from (1.23) by choosing $n=(k+1)N$:

$$-\log(k+1)N\cdot a[(k+1)N] = \sum_{\nu|(k+1)N} a(\nu) r\left[\frac{(k+1)N}{\nu}\right] = \tag{1.24}$$

$$= a(N)r(k+1)+\sum_{\substack{\nu|(k+1)N\\ \nu>N}} a(\nu) r\left[\frac{(k+1)N}{\nu}\right].$$

From Property 7 it follows that (1.20) cannot have two, or more linearly independent solutions in M (see [3]). So we can formulate the following Theorem, which is more general than the corresponding one of the paper [2].

THEOREM 1.1. *The differential equation*

$$D(x)-qx=0, \quad q\in M$$

has a nontrivial solution in the operator field M if and only if q is a function, and $\alpha=e^{-q(1)}$ is rational. The general solution is given by (1.19) *and it is a function for every $c\in K$ if and only if $\alpha=e^{-q(1)}$ is natural. Every solution belongs to the ring E^*.*

§ 2. On the nonlinear differential equation system (3)

Let us consider the differential equation system (3)

$$D(x_i) = \frac{a_i}{\prod\limits_{\substack{k=1\\ k\neq i}}^{m} x_k} + f_i x_i, \quad i=1,2,\ldots,m;\ m>1 \tag{2.1}$$

which can be written in the form

$$D(x_i)\prod_{\substack{k=1\\ k\neq i}}^{m} x_k = a_i + f_i \prod_{k=1}^{m} x_k, \quad i=1,2,\ldots,m. \tag{2.2}$$

By summing up the equations (2.2) for all i's and taking into account condition (5) we have

$$D\Big(\prod_{k=1}^{m} x_k\Big) = \sum_{i=1}^{m} f_i \prod_{k=1}^{m} x_k. \tag{2.3}$$

By introducing

$$F=\prod_{k=1}^{m} x_k, \qquad f=\sum_{i=1}^{m} f_i,$$

we shall obtain the following differential equation:

$$D(F)-fF=0. \tag{2.4}$$

So the following lemma holds.

LEMMA 2.1. *If* (2.1) *has a solution in* M *then* (2.4) *has also a nontrivial solution in* M.

The converse does not hold, since the existence of $F \in E^*$ $(F \neq 0)$ does not imply the existence of the operational solution of (2.1).

By Theorem of the preceding chapter and Lemma 2.1 we obtain the following simple

NON-EXISTENCE THEOREM 2.1. *If* $\alpha = e^{-f(1)}$ *is irrational, then the system* (2.1) *has no solution in* M.

In the sequel we assume that α is rational. By (1.19) we have

$$F = c\delta(\alpha) \exp\left[\int (\{f(n)\} - f(1))\right]. \tag{2.5}$$

Substituting (2.5) in (2.1) and taking into account (1.8), we have

$$D(x_i) - \left[\frac{a_i}{c}\delta\left(\frac{1}{\alpha}\right)\exp\left(-\int (f - f(1))\right) + f_i\right] x_i = 0, \qquad i = 1, 2, \ldots, m. \tag{2.6}$$

Consequently, we can state

LEMMA 2.2. *The system* (2.1) *has an operational solution if and only if every equation of* (2.6) *has a nontrivial operational solution.*

REMARK 1. Let x_{pi} denote a particular solution of the i-th equation of (2.6). The general solution is of the form

$$x_i = \beta_i x_{pi} \qquad \beta_i \in K.$$

It is easily seen that the set of operators x_i $(i = 1, 2, \ldots, m)$ is the general solution of (2.1) if and only if

$$\prod_{k=1}^{m} \beta_k = c. \tag{2.7}$$

We prove the following

NON-EXISTENCE THEOREM 2.2. *Let* $\alpha = e^{-f(1)} = \dfrac{N_1}{N_2}$, N_1, $N_2 \in Z$, $N_1 > 1$, *where* N_1, N_2 *are relatively primes. Moreover, let exist an index i such that* $a_i(n)$ *does not vanish for every value of n, for which* N_1 *is not a divisor of n. Then* (2.1) *has no solution in* M.

PROOF. Let us choose such an index i, and denote it by j. Then the application of Property 4 of the preceding chapter shows that

$$\frac{a_j}{c}\delta\left(\frac{1}{\alpha}\right) = \frac{a_j \delta(N_2)}{c\delta(N_1)} \notin E.$$

Consequently,

$$z_j = \frac{a_j}{c}\delta\left(\frac{1}{\alpha}\right)\exp\left[-\int (f - f(1))\right] \notin E. \tag{2.8}$$

(Indeed, if z_j were a function, then

$$z_j \exp\left[\int (f-f(1))\right] = \frac{a_j}{c}\delta\left(\frac{1}{\alpha}\right)$$

would also be a function.) Since $f_j \in E$, $a_j \neq 0$, so we have

$$\frac{a_j}{c}\delta\left(\frac{1}{\alpha}\right)\exp\left[-\int (f-f(1))\right]+f_j \notin E.$$

Applying the Theorem 1.1 we obtain that the differential equation

$$D(x_j)-\left(\frac{a_j}{c}\delta\left(\frac{1}{\alpha}\right)\exp\left[-\int (f-f(1))\right]+f_j\right)x_j = 0 \tag{2.9}$$

has only the trivial solution in M. The application of Lemma 2.2 concludes the proof.

Assuming that the functions $a_i(n)$ vanish for those n, for which N_1 is not a divisor of n, we have by Property 4 that

$$\frac{a_i}{c}\delta\left(\frac{1}{\alpha}\right) = \frac{1}{c}\{r_i(n)\}, \tag{2.10}$$

where

$$r_i(n) = \begin{cases} a_i\left(\dfrac{nN_1}{N_2}\right), & \text{for} \quad N_2|n, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

So

$$\frac{a_i}{c}\delta\left(\frac{1}{\alpha}\right)\exp\left[-\int (f-f(1))\right]+f_i = \frac{r_i}{c}\exp\left[-\int (f-f(1))\right]+f_i \in E$$

holds for every i.

By introducing the notation

$$p_i = r_i \exp\left[-\int (f-f(1))\right] \tag{2.12}$$

we have the following

EXISTENCE THEOREM 2.3. *Let* $\alpha = e^{-f(1)} = \frac{N_1}{N_2}$, $N_2 > 1$ *where* N_1, N_2 *are relatively primes. Moreover, for* $N_1 > 1$, *let the functions* $a_i(n)$ *vanish for the values of* n, *for which* N_1 *is not a divisor of* n. *The system* (2.1) *has a solution in* M *if and only if*

$$\alpha_i = e^{-f_i(1)} \qquad i = 1, 2, \ldots, m$$

are rational. The general solution is of the form

$$x_i = \beta_i \delta(\alpha_i)\exp\left[\int\left(\frac{p_i}{c}+f_i-f_i(1)\right)\right], \quad i = 1, 2, \ldots, m; \ \beta_i, \ c \in K, \tag{2.13}$$

where

$$\prod_{i=1}^{m} \beta_i = c. \tag{2.14}$$

(2.1) *has no solution in E.*

PROOF. Since $N_2 > 1$, we have by (2.11) that

$$r_i(1) = 0, \qquad i = 1, 2, \ldots, m.$$

Consequently,

$$p_i(1) = 0,$$

$$\frac{p_i(1)}{c} + f_i(1) = f_i(1).$$

Applying Theorem 1.1, Lemma 2.2 and (2.7) it is easy to see that the first half of the Existence Theorem holds.

(2.13) would give the general function solution of (2.1) if and only if the numbers

$$\alpha_i = e^{-f_i(1)}$$

were integer for every value of i. But this is impossible since

$$\alpha = e^{-f(1)} = \prod_{i=1}^{m} \alpha_i = \frac{N_1}{N_2} \quad (N_2 > 1)$$

is not an integer. Therefore there exists no solution in E. The theorem is proved. Now let us consider the case $N_2 = 1$.

DEFINITION. Let I be the set of those indices i, for which $a_i(N_1) \neq 0$. Let Q and $Q^* \subset Q$ be the sets of those numbers $c \in K$, for which

$$\gamma_i = \exp\left[-\left(\frac{a_i(N_1)}{c} + f_i(1)\right)\right], \quad i \in I \tag{2.15}$$

are rational and integer, respectively, for every $i \in I$.

EXISTENCE THEOREM 2.4. *Let* $\alpha = e^{-f(1)} = N_1$, *moreover, for* $N_1 > 1$, *let the functions* $a_i(n)$ *vanish for those values of* n, *for which* N_1 *is not a divisor of* n. *If the set* I *is empty, then* (2.1) *has a solution in* M *if and only if*

$$\alpha_i = e^{-f_i(1)} \qquad i = 1, 2, \ldots, m$$

are rational. The general solution is of the form (2.13) *being the general function solution of* (2.1) *if and only if*

$$\alpha_i = e^{-f_i(1)} \qquad i = 1, 2, \ldots, m$$

are integer. If the set I *is non-empty, then* (2.1) *has a solution in* M *if and only if the set* Q *is non-empty and*

$$\alpha_i = e^{-f_i(1)} \qquad i \notin I$$

are rational. The general solution is of the form

$$x_i = \beta_i \delta(\alpha_i) \exp\left[\int\left(\frac{p_i}{c} + f_i - f_i(1)\right)\right], \quad i \notin I, \quad c \in Q, \tag{2.16}$$

$$x_i = \beta_i \delta(\gamma_i) \exp\left[\int\left(\frac{p_i}{c} + f_i - \frac{a_i(N_1)}{c} - f_i(1)\right)\right], \quad i \in I, \quad c \in Q, \tag{2.17}$$

where

$$\prod_{k=1}^{m} \beta_k = c.$$

(2.1) *has a solution in E if and only if the subset* $Q^* \subset Q$ *is non-empty, and*

$$\alpha_i = e^{-f_i(1)} \qquad i \notin I$$

are integer. The general function solution is given by (2.16), (2.17) *by replacing* $c \in Q$ *by* $c \in Q^*$.

Proof. Since $N_2 = 1$, we have by (2.11) that

$$r_i(n) = a_i(nN_1) \qquad n = 1, 2, \ldots$$

holds. If I is empty then

$$r_i(1) = 0 \qquad i = 1, 2, \ldots, m.$$

Applying the idea of the proof of Existence Theorem 2.3, and taking into account that all the numbers

$$\alpha_i = e^{-f_i(1)}$$

may be integer, since

$$\alpha = e^{-f(1)} = \prod_{i=1}^{m} \alpha_i = N_1.$$

So, the case of the empty I is settled.

Let I be non-empty, then

$$r_i(1) = a_i(N_1) \neq 0, \qquad i \in I.$$

From (1.12), (1.17) follows that we can choose the value of

$$e^{\int [f - f(1)]}$$

for $n = 1$ to be 1. By (2.12) we have

$$p_i(1) = r_i(1) = a_i(N_1) \neq 0, \qquad i \in I. \tag{2.18}$$

We consider those differential equations in (2.6), for which $i \in I$. Obviously, they have nontrivial solutions only for those values of c, for which (2.15) are rational.

Applying (2.18) we obtain (2.17). For $i \notin I$, the differential equations (2.6) have nontrivial solutions in M if and only if

$$\alpha_i = e^{-f_i(1)} \qquad i \notin I$$

are rational. Using Theorem 2.3 we have that (2.16) holds. The operators (2.17) are functions only for those values of c, for which the numbers (2.15) are integer. Moreover, the operators (2.16) are functions if and only if

$$\alpha_i = e^{-f_i(1)} \qquad i \notin I$$

are integer. The application of Lemma 2.2 concludes the proof.

REMARK 2. Simple examples show the existence or non-existence of the sets Q, Q^*.

EXAMPLE 1. Let

$$m = 2, \quad f_1(1) = f_2(1) = 0, \quad a_1(N_1) \neq 0.$$

We have

$$\gamma_1 = e^{-a_1(N_1)/c}, \quad \gamma_2 = e^{a_1(N_1)/c}.$$

Every $c \in Q$ is of the form

$$c = \frac{a_1(N_1)}{\log \frac{R_1}{R_2}}, \quad R_1, R_2 \in Z, \quad R_1 \neq R_2.$$

The set Q^* is empty.

EXAMPLE 2. Let

$$m = 3, \quad a_1(N_1) = -1, \quad a_2(N_1) = -2, \quad a_3(N_1) = 3, \quad f_1(1) = f_2(1) = -1,$$
$$f_3(1) = 2.$$

We have

$$\gamma_1 = e^{1/c+1}, \quad \gamma_2 = e^{2/c+1}, \quad \gamma_3 = e^{-3/c-2}.$$

Let γ_1 be rational for some $c \in Q$. Consequently, γ_2 is not rational. Indeed,

$$\frac{\gamma_2}{\gamma_1} = e^{1/c}$$

and if γ_2 were rational, then $e^{1/c}$ would also be rational. But this is impossible since

$$e = \gamma_1 e^{-1/c}$$

is not rational. The set Q is empty.

EXAMPLE 3. Let

$$m = 3, \quad a_1(N_1) = 1, \quad a_2(N_1) = 0, \quad a_3(N_1) = -1,$$
$$f_1(1) = -\log 2, \quad f_2(1) = f_3(1) = 0.$$

We have

$$\gamma_1 = 2e^{-1/c}, \quad \gamma_3 = e^{1/c}.$$

It can easily be seen that the set Q^* contains exactly one element, namely

$$c = \frac{1}{\log 2}.$$

In the general case, the determination of the existence criteria of Q, Q^* and of the explicit forms of these sets is an interesting number-theoretical problem.

REFERENCES

[1] Fényes, T., On a nonlinear operational differential equation system, *Studia Sci. Math. Hungar.* (to appear).

[2] Fényes, T. and Szilárd, K., Über diskrete Mikusińskische Operatoren, die auf Grund der Dirichletschen Produktenformel erzeugt werden, *Studia Sci. Math. Hungar.* **11** (1976), 181—200. *MR* **81b:** 44014b.

[3] Mikusiński, J., *Operational calculus*, Pergamon Press, Państwowe Wydawnictwo Naukowe, Warsaw, 1959. *MR* **21** #4333.

[4] Gesztelyi, E., The application of the operational calculus in the theory of numbers, *Number Theory*, Colloq. Math. Soc. János Bolyai **2**, 51—104, North-Holland, Amsterdam, 1970. *MR* **42** #5922.

[5] Butzer, P. and Schulte, H., *Ein Operatorenkalkül zur Lösung gewöhnlicher und partieller Differenzengleichungssysteme von Funktionen diskreter Veränderlicher und seine Anwendungen*, Westdeutscher Verlag, Köln und Opladen, 1965. *MR* **33** #511.

(Received June 26, 1985)

MTA MATEMATIKAI KUTATÓ INTÉZETE
POSTAFIÓK 127
H—1364 BUDAPEST
HUNGARY

RECENTLY ACCEPTED PAPERS

SZÁZ, Á., Projective and inductive limits of preseminormed spaces by linear relations
LETTL, G., Durch Normen definierte Idealklassengruppen
ZAHAROV, V. K., Borel cover and Borel extension
KATSURA, M., Leaves of precodes assigned to finite Moore automata
CAKAR, Ö. and COLAK, R., Banach limits and related matrix transformations
KRISHNAN, V. S., A note on the category of s.m.f. spaces and related categories
HA LE ANH, On two problems of Katona concerning sums of vectors
AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., On a theorem of N. K. Bari
GENSEMER, S. H. and WEINERT, H.-J., On 0-semigroups and 0*-semigroups
SWARTZ, C. W., Pshenichnyi's necessary condition for nonsmooth programming
DEÁK, J., Bimerotopies I
FÉNYES, T., On an operational differential equation system
NGUYEN VAN DUNG, Some conditions for a self-injective ring to be quasi-Frobenius
FILEP, L. and MAURER, GY., Compatible fuzzy relations and groups
MÓRI, T. F., On the number of different patterns preceding a given one
JUHÁSZ, I., On a problem of van Douwen
SACHS, H., Vollständig zirkuläre Kurven n-ter Ordnung der isotropen Ebene

CONTENTS *(continued)*

CONTENTS

(continued inside)