Studia Scientiarum Mathematicarum Hungarica

VOLUME 30
NUMBERS 1–2
1995

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. It 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–1117 Budapest, Prielle Kornélia u. 19–35

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H–1364 Budapest

Tel.: (36)(1) 118–2875 Fax: (36)(1) 117–7166
e-mail: h3299mer @ ella.hu

or

merza @ math–inst.hu

Subscription information

For 1995 volumes 30–31 are scheduled for publication. The subscription price is $ 98.00 each, including normal postage, air delivery plus $ 20.00 per volume.

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O.Box 245
H–1519 Budapest

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 1–11

HOW TO COLOR SHIFT HYPERGRAPHS

N. ALON, I. KRIZ and J. NEŠETŘIL

Abstract

Let $g(k)$ denote the minimum integer m so that for every set S of m integers there is a k-coloring of the set of all integers so that every translate of S meets every color class. It is a well known consequence of the Local Lemma that $g(k)$ is finite for all k. Here we present a new proof for this fact, that yields a very efficient parallel algorithm for finding, for a given set S, a coloring as above. We also discuss the problem of finding colorings so that every translate of S has about the same number of points in each color. In addition, we prove that for large k

$$(1+o(1))k\log k \leqq g(k) \leqq (3+o(1))k\log k.$$

Introduction

Straus (cf. [8]) raised the following problem: Is there a function $g(k)(<\infty)$ such that for every set S of at least $g(k)$ integers there is a coloring of the integers by k colors so that every translate of S meets all the colors? This problem was solved by Erdős and Lovász [8], who proved that

$$g(k) \leqq O(k\log k). \tag{1}$$

The proof of [8] is probabilistic and uses the Lovász Local Lemma, a result that has been used for tackling many other combinatorial problems in numerous subsequent papers. As remarked in [12] there is no known proof for the finiteness of $f(k)$ that does not use the Local Lemma.

Our first result in the present short paper is such a proof, namely a solution of Straus' problem that does not apply the Local Lemma. Although our basic solution works only for sets S of cardinality at least $4k^2$ it has the advantage that it is more constructive than the original solution of [8], and yields very efficient deterministic and parallel algorithms for finding a coloring of the integers with the required properties. As is the case with

1991 *Mathematics Subject Classification.* Primary 05C65; Secondary 05C55.
Key words and phrases. Hypergraph coloring, probabilistic methods, sum-free sets.

Research supported in part by a United States Israel BSF Grant.

0081-6906/95/$ 4.00

many applications of the Local Lemma, the proof in [8] supplies neither a randomized nor a deterministic polynomial time algorithm for finding, given a set S of a sufficiently large cardinality, a k-coloring so that each translate of S meets every color. The recent technique of J. Beck [5] and its modification in [1] that supply efficient sequential and parallel deterministic algorithms for various applications of the Local Lemma do not seem to apply directly to the problem of Straus mentioned above, even when the cardinality of S is much larger than $\Theta(k \log k)$. Note that here the length of the input is the number of bits in the representation of S whereas it is not even clear if the output can be represented with a finite number of bits, as the output is a k-coloring of the infinite set of integers.

Our technique here yields a very efficient parallel algorithm that produces, for a given set S of at least $4k^2$ integers a k-coloring of the integers so that every translate of S meets every color class. In fact, the required coloring can be found in *constant* time in the standard model for parallel computation known as a CRCW PRAM with a polynomial number of parallel processors. (See, e.g., [11] for the exact definition of a CRCW PRAM; we assume that each processor is capable of adding, comparing, multiplying or dividing numbers of size as that of the members of S in constant time.) Since the required k-coloring is a coloring of an infinite set we agree that the k-coloring is produced successfully if it can be described by a polynomial number of bits that enable us, given any integer x, to compute the color of x efficiently; in our case this will be done by a constant number of modular additions and multiplications. The basic approach can be combined with the technique in [5] and yield efficient sequential coloring algorithms even if S has only $ck \log k$ elements for some (large) constant c. Still, we believe that the most interesting consequence of the argument is the new method for solving Straus' original problem.

Another result we prove here is the fact that the estimate in (1) is sharp.

In order to formulate our results and proofs in a more concise form we introduce two definitions. For a set of integers S, let $H = H(S)$ denote the infinite hypergraph whose set of vertices is the set Z of all integers and whose set of edges is the set of all translates of S, i.e., the set $\{x + S : x \in Z\}$. We call H the *shift hypergraph of* S. A k-coloring $c : Z \mapsto \{1, 2, \ldots, k\}$ is called *good (for H)*, if every edge of H meets every color class, i.e., if for every i, $1 \leqq i \leqq k$ and for every integer x there is an $s \in S$ so that $c(x + s) = i$.

In this notation, our two main results are the following.

THEOREM 1.1. *Let S be a set of at least $4k^2$ integers. Then there exists a good k-coloring c for the shift hypergraph $H(S)$. Such a coloring can be found in constant time, using a polynomial number of parallel processors on a CRCW PRAM. In addition, there exists a positive constant c such that for every set S of at least $ck \log k$ integers one can find a good k-coloring for the shift hypergraph $H(S)$ in (sequential) polynomial time.*

TEHOREM 1.2. *There exists an absolute positive constant a such that for*

every $k > 1$ there is a set $S = S_k$ of at least $ak \log k$ integers so that there is no good k-coloring for the shift hypergraph $H(S)$.

The proof of Theorem 1.1 is based on the ideas of [3] and is presented in the next section, together with some related extensions. In Section 3 we describe two proofs of Theorem 1.2; a probabilistic one and a constructive one. The final Section 4 contains some concluding remarks.

Finding a good coloring

PROOF OF THEOREM 1.1. Let S be a given set of $m = 4k^2$ distinct integers. Our objective is to find a good k-coloring c for the shift hypergraph $H(S)$. To do so, we first choose a prime p so that the members of S are pairwise distinct modulo p. Let $P = \{0, 1, \ldots, p-1\}$ be the set of all remainders modulo p and let us split P into k pairwise disjoint intervals of consecutive remainders $I_1, I_2, \ldots, I_k$, where $\lfloor p/k \rfloor \leq |I_i| \leq \lceil p/k \rceil$ for all $1 \leq i \leq k$.

For two integers a and b in P, let $c = c_{a,b}$ be the following k-coloring of the set of integers. For every integer y, $c(y)$ is the unique i so that $(ay + b) \pmod p \in I_i$. Define

$$Y = Y_{a,b} = \{(as + b) \pmod p : s \in S\}.$$

We claim that if a, b are chosen in such a way that the set Y intersects *every* (cyclic) interval of length $\lfloor p/k \rfloor$ in P then every translate of S intersects each color class of c. To see this, observe that if $x + S$ is a translate of S then the set

$$\{(ay + b) \pmod p : y \in x + S\}$$

is a cyclic translate of Y and hence it intersects every interval of length $\lfloor p/k \rfloor$ in P and in particular it intersects every I_i, implying the desired result. It thus suffices to choose a, b so that $Y_{a,b}$ has the above property. We next show that this can be done.

FACT. If a and b are chosen randomly and independently in P, according to a uniform distribution, then with positive probability the set $Y = Y_{a,b}$ intersects every cyclic interval of length at least $\lfloor p/k \rfloor$ in P.

PROOF. The argument essentially appears in [3], but since it is very short we repeat it here. Let $J_1, \ldots, J_{2k}$ be a fixed covering of P by $2k$ intervals of length $\lceil p/2k \rceil$ each. Observe that if Y intersects each J_i then it certainly satisfies the required property, since every cyclic interval of length $\lfloor p/k \rfloor$ must fully contain at least one interval J_i. Fix an i, $1 \leq i \leq 2k$ and put $m = |S| = 4k^2$. For each element $s \in S$ let X_s^i be the indicator random variable whose value is 1 if $(as + b) \pmod p \in J_i$ (and is 0 otherwise). Define $X^i = \sum_{s \in S} X_s^i$ and observe that $X^i = 0$ if and only if Y does not intersect J_i.

We estimate the probability of this event by computing the expectation and variance of X^i. By linearity of expectation

$$E(X^i) = \sum_{s \in S} E(X_s^i) = m\lceil p/2k \rceil / p.$$

The crucial (and simple) fact for computing the variance is the fact that the random variables X_s^i, $(s \in S)$ are pairwise independent. This follows from the fact that if s and t are two distinct members of S then when the pair (a, b) ranges over $P \times P$ so does the pair

$$((as+b) \pmod p, (at+b) \pmod p),$$

implying that X_s^i and X_t^i are independent. Hence,

$$Var(X^i) = \sum_{s \in S} Var(X_s^i) = m\frac{\lceil p/2k \rceil}{p}(1 - \frac{\lceil p/2k \rceil}{p}).$$

Therefore, by Chebyschev's Inequality,

$$Prob(X^i = 0) \leqq Prob(|X^i - E(X^i)| \geqq E(X^i)) \leqq \frac{Var(X^i)}{E(X^i)^2} < 2k/m.$$

Since there are $2k$ possible values of i the probability that $X^i = 0$ for some i is strictly smaller than $4k^2/m = 1$, completing the proof of the fact. □

Returning to the proof of Theorem 1.1 observe that the last fact implies that randomly chosen a and b supply a good k-coloring with positive probability. In particular, there is at least one such pair a, b and hence a good coloring exists.

For the algorithm, it is essential to choose a small prime p so that all members of S are distinct modulo p, that is, a prime which is smaller than some (fixed) polynomial in the length of the input. Fortunately, the existence of such a prime is simple. A prime p is not good if and only if it divides the product

$$\prod_{s,s' \in S, s > s'} (s - s').$$

If each number in S is at most 2^n (and at least 0, as we may assume since the problem is invariant under any additive shift of S), and S has m members, the last product is certainly at most 2^{nm^2}. Since it is well known that the product of all primes smaller than x is $e^{(1+o(1))x}$ this shows that there is a prime $p < nm^2$ that does not divide the product.

Therefore, for the algorithm, we simply check, in parallel, for every prime p up to nm^2 if it is good, i.e., if all the members of S are distinct modulo p. Once we find such a prime we check, in parallel, all pairs a, b and find

a pair for which the set $Y_{a,b}$ intersects every interval of length $\lfloor p/k \rfloor$ (we can afford checking all the intervals in parallel). One can check that all this can, indeed, be performed in constant time in parallel using polynomially many processors. (Recall that each processor can add, compare, multiply and divide numbers in the required range in constant time.) This completes the proof of the first part of the theorem.

For the second part of the proof of Theorem 1.1, namely the existence of the constant c, we only need to combine one simple ingredient of the proof above with the technique of [5]. Given a set S of $m \geq ck \log k$ positive integers, each at most 2^n, we can find efficiently a prime p so that $p < nm^2$ and all the members of S are distinct modulo p. Consider the hypergraph H whose set of vertices is the set Z_p of integers modulo p and whose set of edges is the set of all shifts of S modulo p, i.e., the set

$$\{ \{(x+s) \pmod p : s \in S\} : x \in Z_p\}.$$

H is clearly an m-uniform, m-regular hypergraph, and the technique in [5] can be applied to obtain in sequential polynomial time, a vertex k-coloring of H so that each edge meets every color class, provided $m \geq ck \log k$ for a sufficiently large c. Since this is very similar to the examples given in [5], we omit the detailed algorithm. The coloring can now be extended to a good integer k-coloring for $H(S)$ simply by letting the color of any integer y be the color of $y \pmod p$. This completes the proof of Theorem 1.1. □

Remarks

1. Combining the basic idea in the last proof with one of the results in [3] we can show that for a sufficiently large set S one can find a coloring with small discrepancy, in the following sense.

PROPOSITION 2.1. *For every set S of m integers, there is a k-coloring of the integers so that the number of points of each color in every translate of S deviates from m/k by at most*

$$O(m^{1/2}(\log m)^{3/2}).$$

Such a coloring can be found in constant time with polynomially many processors on a CRCW PRAM.

To prove this result we start, as before, by choosing a polynomially small prime p so that all the members of S are distinct modulo p. Next we use Theorem 2.3 in [3] which asserts that for every set T of cardinality m in Z_p there is an integer a so that the set $aT \pmod p$ is uniformly distributed in the sense that for every (cyclic) interval of length δp in Z_p the number of members of $aT \pmod p$ in the interval deviates from its expectation δm by

at most $O(m^{1/2}(\log m)^{3/2})$. Let $I_1, \ldots, I_k$ be a partition of Z_p into k almost equal intervals and define, for every integer y, the color of y as the unique i such that $ay(\mathrm{mod}\ p) \in I_i$, where a is chosen so that for $T = S(\mathrm{mod}\ p)$, $aT(\mathrm{mod}\ p)$ is uniformly distributed as described above. The same reasoning as in the last proof shows that this coloring satisfies the desired requirements. It is also obvious that it can be found in constant time with polynomially many processors in parallel, as before.

We note that one can obtain an even more uniform k-coloring for the shift hypergraph $H(S)$ by applying the local lemma, but we do not know to find such a coloring in constant time in parallel.

2. The basic argument in the proof of Theorem 1.1 can be modified and extended to the real case as we sketch next without discussing the algorithmic issue. Here, too, the Local Lemma yields a sharper result but our argument is more constructive.

PROPOSITION 2.2. *For every set S of at least $4k^2$ reals, there is a k-coloring c of the set of all real numbers R so that every (real) translate of S intersects each color-class.*

To prove this proposition first choose a real number t, so that all members of S are distinct modulo t. Let $I_1, \ldots, I_k$ be a set of k pairwise disjoint intervals of equal lengths that partition $[0, t)$, defined by $I_i = [(i-1)t/k, it/k)$. Next, let a be chosen randomly and uniformly in the interval $(0, M)$, where M is a large number, to be determined later, and let b be chosen randomly and uniformly in $(0, t)$. Define a coloring c of the real numbers as follows: For a real number y, the color $c(y)$ is the unique i so that $(ay+b)\ (\mathrm{mod}\ t) \in \in I_i$. One can imitate the proof of Theorem 1.1 and show that if S is a set of at least $4k^2$ reals, and M is sufficiently large, then the probability that c maps S into a set that intersects every cyclic interval of length at least t/k in $[0, t)$ is $\Omega(1)$. Such a coloring c will have the desired properties and the assertion of the proposition follows. We omit the details.

3. One can use a simple algebraic idea to extend the first part of Theorem 1.1 even further, to the case of an arbitrary torsion-free Abelian group (such as, e.g., any Euclidean space R^d). Here, too, the Local Lemma yields a sharper result but our argument is more constructive.

PROPOSITION 2.3. *Let G be a torsion-free Abelian group. Then for every subset S of G containing at least $4k^2$ elements there is a k-coloring of G so that every translate of S in G intersects each color class.*

PROOF. The set S spans a finitely generated and hence a free Abelian subgroup F of G. Select a homomorphism $\phi : F \mapsto Z$ such that ϕ restricted to S is injective (this is always possible). If $\chi : Z \mapsto \{1, \ldots, k\}$ is the required coloring of Z with respect to $\phi(S)$, then $\chi\phi$ is the required coloring of F with respect to S. To color G, color each F-coset separately by a translate of the coloring of F. □

Studying analogues of Straus' problem mentioned in the introduction for various actions of groups on sets seems to be an interesting program. It is worth mentioning that there is a whole area known as Geometric, or Euclidean Ramsey Theory (see, e.g., [9] for a few examples), which studies questions precisely opposite to Straus' problem.

$\Theta(k \log k)$ colors are necessary (and sufficient)

In this section we present two proofs of Theorem 1.2. Let $g(k)$ denote the smallest m so that for every set S of at least m integers there is a good k-coloring for the shift hypergraph $H(S)$. The proof in [8] gives that for large k:

$$g(k) \leqq (3 + o(1))k \log_e k.$$

The next proposition shows that this is sharp, up to a constant factor.

PROPOSITION 3.1. *If q is a prime, and $q > l^2 2^{2l-2}$ then*

$$g\left(\left\lceil \frac{2q}{l+1} \right\rceil\right) > \frac{q+1}{2}.$$

This implies that for large k:

$$g(k) \geqq (\frac{1}{8} + o(1))k \log_2 k.$$

The proof of the above proposition is by a construction that uses the properties of the quadratic residues and non-residues in the field Z_q. Recall that there are precisely $(q+1)/2$ quadratic residues modulo q. We need the following simple consequence of the well known theorem of Weil. For a derivation of this lemma from Weil's theorem, see [10] or [6].

LEMMA 3.2. *Let $q > l^2 2^{2l-2}$ be a prime and let $Z = \{z_1, \ldots, z_s\}$ be a set of $s \leqq l$ members of Z_q. Then there exists an element $y \in Z_q$ so that $z_i - y$ is a quadratic non-residue for all $1 \leqq i \leqq s$.*

PROOF OF PROPOSITION 3.1. Put $m = (q+1)/2$ and let S be the set of all m quadratic residues modulo q, considered here as usual integers. Suppose $k \geqq \frac{2q}{l+1}$ and let $c : Z \mapsto K = \{1, \ldots, k\}$ be a k-coloring of the integers. To complete the proof we show that there exists a translate of S that misses at least one color class. To this end, consider the colors of the integers in the set $Q = \{0, 1, \ldots 2q-2\}$ and let $j \in K$ be a color assigned to at most $|Q|/k < l+1$ members of this set. Let $y_1, \ldots, y_s$ $(s \leqq l)$ be all the members of Q satisfying $c(y_i) = j$. Let z_i be the elements of Z_q defined by $z_i \equiv y_i \pmod q$. By Lemma 3.2 there exists a $y \in Z_q$ so that $z_i - y$ is a quadratic non-residue for all $1 \leqq i \leqq s$. Consider, now, y as a usual integer. We claim that the

translate $y+S$ of S does not intersect color class number j. To see this, suppose it is false. Since $y+S \subset Q$ this means that there is an i, $1 \leqq i \leqq s$, and there is an $x \in S$ so that $y+x=y_i$. Reducing this equation modulo q we conclude that $x=(z_i-y)(\mathrm{mod}\ q)$. But this is impossible since z_i-y is a quadratic non-residue whereas x (like all the members of S) is a quadratic residue. Thus the claim holds, and the assertion of the proposition follows. □

Proposition 3.1 implies Theorem 1.2. We next present another, probabilistic proof of this theorem, that gives a slightly better lower bound for $g(k)$.

PROPOSITION 3.3. *For large k*

$$g(k) \geqq (1+o(1))k \log_e k.$$

The main part of the proof is the following somewhat technical lemma.

LEMMA 3.4. *For every fixed (small) $\epsilon > 0$ there exists a (small) $\delta > 0$ such that for every sufficiently large n the following holds; There exists a subset S of $N=\{1,2,\ldots,n\}$ of cardinality at least $(1-\frac{\epsilon}{10})\delta n$ so that for every set T of at most*

$$\frac{(1-\frac{\epsilon}{10})\log_e n}{(1+\frac{\epsilon}{10})\delta}$$

positive integers, each at most $(1+\frac{\epsilon}{10})n$, there is an integer $0 \leqq y \leqq \frac{\epsilon}{10}n$ so that $y+S$ does not intersect T.

PROOF. Let $\delta > 0$ satisfy

$$1-\delta > e^{-(1+\frac{\epsilon}{10})\delta}. \tag{2}$$

(Since

$$e^{-(1+\frac{\epsilon}{10})\delta} = 1-(1+\frac{\epsilon}{10})\delta + O(\delta^2)$$

any sufficiently small $\delta > 0$ satisfies (2)). Let n be sufficiently large (as a function of ϵ and δ), and let S be a random subset of $N=\{1,\ldots,n\}$ obtained by choosing each $i \in N$, randomly and independently, to be a member of S with probability δ. Denote the cardinality of S by m. By the standard estimates for binomial distributions (see, e.g., [4]), for a sufficiently large n, with high probability

$$m \geqq (1-\frac{\epsilon}{10})\delta n.$$

Fix a set T of at most

$$\frac{(1-\frac{\epsilon}{10})\log_e n}{(1+\frac{\epsilon}{10})\delta}$$

positive integers, each at most $(1+\frac{\epsilon}{10})n$, and fix an integer $y \leq \frac{\epsilon}{10}n$. The probability that $y+S$ does not intersect T is the probability that $t-y$ is not in S for all $t \in T$, which is, by (2), at least

$$(1-\delta)^{|T|} > e^{-(1+\frac{\epsilon}{10})\delta \frac{(1-\frac{\epsilon}{10})\log_e n}{(1+\frac{\epsilon}{10})\delta}} = \frac{1}{n^{1-\frac{\epsilon}{10}}}.$$

For the above fixed set T consider now all the possible shifts of S by an integer y satisfying $0 \leq y \leq \frac{\epsilon}{10}n$. For each such y we have the estimate above for the probability of the event E_y that $y+S$ does not intersect T. Moreover, if Y is a set of possible shifts and for every two distinct y and y' in Y, $T-y$ does not intersect $T-y'$ the events E_y $(y \in Y)$, are mutually independent. It is easy to see that there is such a set Y of cardinality at least $\frac{\epsilon}{10}n/|T|^2 = \Omega(n/(\log n)^2)$, where here the constant in the $\Omega(\cdot)$ notation depends on ϵ and δ but not on n. It follows that the probability that there is no shift y in the possible range so that $y+S$ does not intersect T is at most

$$(1-\frac{1}{n^{1-\frac{\epsilon}{10}}})^{\Omega(n/(\log n)^2)} = e^{-\Omega(n^{\frac{\epsilon}{10}}/(\log n)^2)}.$$

The total number of choices for a subset T as above is only

$$\sum_{i \leq \frac{(1-\frac{\epsilon}{10})\log_e n}{(1+\frac{\epsilon}{10})\delta}} \binom{(1+\frac{\epsilon}{10})n}{i} \leq e^{O((\log n)^2)}.$$

Therefore, the probability that there is a set T so that there is no shift $y+S$ of S that misses it is at most

$$e^{O((\log n)^2)} e^{-\Omega(n^{\frac{\epsilon}{10}}/(\log n)^2)},$$

which tends to 0 as n tends to infinity. This completes the proof of Lemma 3.4. □

PROOF OF PROPOSITION 3.3. Let ϵ be a fixed small positive constant. Our objective is to show that for all sufficiently large k, $g(k) \geq (1-\epsilon)k\log_e k$. Let δ, n, and S satisfy the assertion of Lemma 3.4. We assume, whenever it is needed, that ϵ is sufficiently small and that n is sufficiently large. Put $m = |S|$, then

$$n \geq m \geq (1-\frac{\epsilon}{10})\delta n.$$

Let k be an integer and let $c: Z \mapsto K = \{1, 2, \ldots, k\}$ be a good k-coloring for the hypergraph $H(S)$. Clearly $k \leq m$ $(\leq n)$. We claim that

$$k < \frac{(1+\frac{\epsilon}{10})^2\delta n}{(1-\frac{\epsilon}{10})\log_e n} \leq \frac{(1+\epsilon)m}{\log_e n} \leq \frac{(1+\epsilon)m}{\log_e k}, \tag{3}$$

and hence that

$$m \geqq (1-\epsilon)k \log_e k.$$

Since $\epsilon > 0$ is arbitrarily small (and for each such ϵ any sufficiently large n can be chosen) this, together with the obvious monotonicity of the function $g(k)$, imply the validity of Proposition 3.3. It thus remains to prove the claim. Let Q be the set of all positive integers which do not exceed $(1 + \frac{\epsilon}{10})n$. Fix a color $i \in K$ and let T be the set of all members of Q colored i. If

$$|T| \leqq \frac{(1-\frac{\epsilon}{10}) \log_e n}{(1+\frac{\epsilon}{10})\delta}$$

then, since S satisfies the assertion of Lemma 3.4, there is a translate $y + S$ of S contained in Q which misses T, contradicting the assumption that c is a good k-coloring for $H(S)$. Therefore, each of the k colors appears more than

$$\frac{(1-\frac{\epsilon}{10}) \log_e n}{(1+\frac{\epsilon}{10})\delta}$$

times in Q and hence

$$(1+\frac{\epsilon}{10})n \geqq |Q| > k\frac{(1-\frac{\epsilon}{10}) \log_e n}{(1+\frac{\epsilon}{10})\delta}.$$

This implies (3) and completes the proof. □

Concluding remarks

1. A *sum-free* set of integers is a set that contains no (not necessarily distinct) a, b and c so that $a + b = c$. An old result of Erdős [7] asserts that every set of n nonzero integers contains a sum-free subset of cardinality at least $n/3$. A very simple proof for this result (and some extensions of it) is given in [2], where the problem of obtaining a polynomial time deterministic algorithm for finding, for a given set S of n non-zero integers, a sum-free subset of at least $n/3$ of them, is raised. Our technique here supplies a very simple algorithm (which is also parallelizable), as follows. Given S, find a polynomially small prime $p = 3r + 2$ so that all the n members of S are distinct modulo p, and are all non-zero modulo p. Next check for every nonzero $a \in Z_p$ the number of members s of S so that $as \pmod p$ lies in the interval $r+1, \ldots, 2r+1$. An easy expectation argument shows that there is an a for which the number of these members is at least $\frac{r+1}{3r+2}n > n/3$, and it is easy to see that they form a sum-free subset of S.

2. As shown in Section 3, for large k

$$(1+o(1))k \log_e k \leqq g(k) \leqq (3+o(1))k \log_e k.$$

It would be interesting to find the correct constant in the expression for $g(k)$.

ACKNOWLEDGEMENT. We would like to thank Uri Zwick for helpful comments that improved the statement of Theorem 1.1 and the presentation of its proof.

REFERENCES

[1] ALON, N., A parallel algorithmic version of the local lemma, *Random Structures and Algorithms* **2** (1991), 367–378. *MR* **92i**:05220

[2] ALON, N. and KLEITMAN, D. J., Sum-free subsets, *A Tribute to Paul Erdős* (A. Baker, B. Bollobás and A. Hajnal eds.), Cambridge University Press, Cambridge, 1990, 13–26. *MR* **92f**:11020

[3] ALON, N. and PERES, Y., Uniform dilations, *Geom. Funct. Anal.* **2** (1992), 1–28. *MR* **93a**:11061

[4] ALON, N. and SPENCER, J. H., *The Probabilistic Method*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1992. *MR* **93h**:60002

[5] BECK, J., An algorithmic approach to the Lovász Local Lemma, I, *Random Structures and Algorithms* **2** (1991), 343–365. *MR* **92i**:05219

[6] BOLLOBÁS, B., *Random graphs*, Academic Press, London-New York, 1985, 318–319. *MR* **87f**:05152

[7] ERDŐS, P., Extremal problems in number theory, *Proc. Sympos. Pure Math.*, Vol. VIII, Amer. Math. Soc., Providence, R.I., 1965, 181–189. *MR* **30**#4740

[8] ERDŐS, P. and LOVÁSZ, L., Problems and results on 3-chromatic hypergraphs and some related questions, *Infinite and finite sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol.II, Colloq. Math. Soc. János Bolyai, Vol. 10, North-Holland, Amsterdam, 1975, 609–627. *MR* **52**#2938

[9] GRAHAM, R. L., ROTHSCHILD, B. L. and SPENCER, J. H., *Ramsey theory*, Wiley-Interscience Series in Discrete Mathematics, Wiley, New York, 1980. *MR* **82b**:05001

[10] GRAHAM, R. L. and SPENCER, J. H., A constructive solution to a tournament problem, *Canad. Math. Bull.* **14** (1971), 45–48. *MR* **45**#1798

[11] KARP, R. M. and RAMACHANDRAN, V., Parallel algorithms for shared memory machines, *Handbook of Theoretical Computer Science* (J. Van Leeuwen Ed.), Vol. A, Chapter **17**, Elsevier, Amsterdam, 1990, 869–941. (See **92d**:68001.)

[12] SPENCER, J., *Ten lectures on the probabilistic method*, CBMS-NSF Regional Conference Series in Applied Mathematics, 52, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1987, p. 60. *MR* **89g**:05002

(Received February 14, 1994)

DEPARTMENT OF MATHEMATICS
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCE
TEL AVIV UNIVERSITY
IL–69978 TEL AVIV
ISRAEL

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CHICAGO
CHICAGO, IL 60637
U.S.A.

DEPARTMENT OF MATHEMATICS
CHARLES UNIVERSITY
CZ–118 00 PRAHA 1
CZECH REPUBLIC

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 13–15

A NOTE ON THE PATH-DISCREPANCY OF TREES

I. BÁRÁNY and Gy. KÁROLYI

We will denote by $[n]$ the set of all positive integers $\{1, \dots, n\}$ and by $\mathcal{F}_n$ the set of all trees having vertex set $[n]$. For a two-colouring $f : [n] \mapsto \{-1, +1\}$ and $X \subset [n]$ we define

$$D(f, X) = \left| \sum_{x \in X} f(x) \right|.$$

Let $T_1, \dots, T_k \in \mathcal{F}_n$. The path-discrepancy of the trees $T_1, \dots, T_k$ with respect to f is

$$D_P(f; T_1, \dots, T_k) = \max_P D(f, P)$$

where the maximum is taken over all paths P that are a subgraph of some T_i. Finally, the path-discrepancy of the trees $T_1, \dots, T_k$ is

$$D_P(T_1, \dots, T_k) = \min_{f:[n] \mapsto \{-1,+1\}} D_P(f; T_1, \dots, T_k).$$

EXAMPLE. Let the tree T be just a path of length $n-1$ on the vertex set $[n]$. The ordering of the vertices along T is a permutation π of the set $[n]$. The subpaths of T correspond to intervals of π. Assume now that all trees $T_1, \dots, T_k$ are paths on vertex set $[n]$, and π_i is the permutation corresponding to T_i. Then the path-discrepancy of $T_1, \dots, T_k$ is exactly the discrepancy of the hypergraph consisting of the intervals of the permutations $\pi_1, \dots, \pi_k$. (For the definition see [1], [2].) Theorem 2.1 of G. Bohus [2] then states that $D_P(T_1, \dots, T_k) \ll k \log n$. The path-discrepancy of a single tree is clearly one and it is not difficult to see that $D_P(T_1, T_2) \leqq 2$, the details can be found in [2].

On the "Irregularities of distribution" workshop in Bielefeld Vera T. Sós asked how the function

$$D_P(n, k) = \max\{D_P(T_1, \dots, T_k) | T_i \in \mathcal{F}_n \text{ for } i = 1, \dots, k\}$$

behaves. Since the chromatic number of a tree is 2 we have $D_P(n, 1) = 1$.

1991 *Mathematics Subject Classification.* Primary 05C05.
Key words and phrases. Discrepancy, trees, paths.

0081–6906/95/$ 4.00

THEOREM.

$$\frac{\log n}{\log\log n} \ll D_P(n,2) \ll \log n, \text{ and}$$

$$D_P(n,k) \ll k\log^2 n \text{ for every } k \geqq 3.$$

PROOF. First we prove the lower bound. Let $d \geqq 2$ be an arbitrary integer. Define two trees $T_1, T_2 \in \mathcal{F}_{m(d)}$ where $m(d) = 1 + d + d^2 + \ldots + d^{d-1}$ in the following way. T_1 is a path with i connected to $i+1$ for every $1 \leqq i < m(d)$. To define T_2 connect vertex $i \leqq 1 + d + d^2 + \ldots + d^{d-2}$ to vertices $d(i-1)+2, d(i-1)+3, \ldots, di+1$. Thus T_2 is a complete d–ary tree with d levels. We note that this construction is related to an example of Hoffman (see [1], Example 1.8).

CLAIM. $D_P(T_1, T_2) \geqq d$.

PROOF. Suppose, by way of contradiction, that there is a two–colouring $f : [m(d)] \mapsto \{-1, +1\}$ with $D_P(f; T_1, T_2) \leqq d-1$. Assume that $i_1 = 1, i_2, \ldots, i_j$ have been defined so that $f(i_1) = \ldots = f(i_j)$ and

$$(*) \qquad i_{r+1} \in \{d(i_r - 1) + 2, d(i_r - 1) + 3, \ldots, di_r + 1\}$$

for every $1 \leqq r < j$. If $j < d$, then we can find

$$i_{j+1} \in \{d(i_j - 1) + 2, d(i_j - 1) + 3, \ldots, di_j + 1\}$$

satisfying $f(i_{j+1}) = f(i_j)$, since otherwise $d(i_j-1)+2, d(i_j-1)+3, \ldots, di_j+1$ is a monochromatic path in T_1 showing $D_P(f; T_1, T_2) \geqq d$. Therefore there exists a sequence of vertices $i_1, \ldots, i_d$ satisfying $(*)$ for every $r = 1, 2, \ldots, d-1$ and $f(i_1) = \ldots = f(i_d)$. This is a monochromatic path on d vertices in T_2, a contradiction. □

For large enough n we have $n > m(\frac{\log n}{\log\log n})$ proving the lower bound.

For the upper bound we will define, for every tree $T \in \mathcal{F}_n$, a permutation π_T of $[n]$ such that every path of T is the disjoint union of at most $O(\log n)$ intervals of π_T. Then the upper bounds will follow from Bohus's theorem mentioned above.

The definition of π_T goes as follows. Let $\pi_T(1) = 1$. Suppose that $s < n$ and we have already defined $\pi_T(1), \pi_T(2), \ldots, \pi_T(s)$. Now there is a maximal integer $r \leqq s$ with the property that there is a vertex adjacent to $\pi_T(r)$ which is not among the vertices $\pi_T(1), \ldots, \pi_T(s)$. Let these vertices be $v_1(s), \ldots, v_{i_s}(s)$. Thus any such vertex is adjacent to $\pi_T(r)$ and is distinct from $\pi_T(1), \ldots, \pi_T(s)$. Delete the edges $\pi_T(r)v_j(s)$ $(1 \leqq j \leqq i_s)$ of T and denote by $T_j(s)$ the component containing $v_j(s)$. This component is clearly a tree. Fix $j \in \{1, \ldots, i_s\}$ for which the number of vertices of $T_j(s)$ is maximal and define $\pi_T(s+1) = v_j(s)$. We will need this construction, in an induction argument, for a subtree T_0 of T where T_0 has $m < n$ vertices identified with

a subset $M \subset [n]$. In this case π_{T_0} will be a permutation of M which is constructed the same way as above using the natural ordering of M.

To finish the proof of the theorem it is enough to show that each path in T is the disjoint union of at most $O(\log n)$ intervals of π_T. A path in T will be called *monotone* if all of its vertices have different distances in T from vertex 1. Since every path in T is the disjoint union of at most two monotone paths, it will be enough to prove the following

LEMMA. *Let P be a monotone path in T. Then P is the disjoint union of at most $\lfloor \log(n+1) \rfloor$ intervals of π_T.*

PROOF. It is enough to prove the lemma for paths P starting at vertex 1. We use induction on n. The initial step is trivial. Suppose $n > 1$ and we have proved the statement for $1, \ldots, n-1$. Let v be the vertex of P adjacent to 1. Delete the edge $1v$ from T. Let T_0 be the component containing v. Then T_0 is a tree whose vertices form an interval of π_T. Let P_0 be the path in T_0 obtained from P by deleting edge $1v$. So by the induction hypothesis P_0 is the disjoint union of at most $\lfloor \log(m+1) \rfloor$ intervals of π_{T_0} where m is the number of vertices of T_0. These intervals are intervals of π_T as well. If $v = \pi_T(2)$, then the interval of π_T containing v can be extended to an interval containing 1, so P is the disjoint union of at most $\lfloor \log(m+1) \rfloor \leqq \lfloor \log(n+1) \rfloor$ intervals of π_T. On the other hand, if $v \neq \pi_T(2)$, then by the definition of π_T we have $m \leqq (n-1)/2$. Considering vertex 1 as an interval (of length 1) we see that P is the disjoint union of at most $\lfloor \log(m+1) \rfloor + 1 \leqq \lfloor \log(n+1) \rfloor$ intervals of π_T. □

ACKNOWLEDGEMENT. Both authors thank Géza Bohus for useful discussions and the Zentrum für interdisziplinäre Forschung in Bielefeld for the financial help supporting this research.

REFERENCES

[1] BECK, J. and SÓS, V. T., Irregularities of partitions, *Handbook of combinatorics*, ed. by R. L. Graham, M. Grötschel, L. Lovász, Springer, 1994

[2] BOHUS, G., On the discrepancy of 3 permutations, *Random Structures Algorithms* 1 (1990), 215–220. *MR* **92i**:05007

(Received February 14, 1994)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALGEBRA ÉS SZÁMELMÉLET TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 17–23

A NOTE ON PARADOXICAL METRIC SPACES

W. A. DEUBER, M. SIMONOVITS and V. T. SÓS

1. Introduction

Given a metric space, $(\mathcal{M}, d)$, we shall call a mapping $\varphi : \mathcal{M} \to \mathcal{M}$ **wobbling** if $d(x, \varphi(x))$ is bounded for $x \in \mathcal{M}$.

Such mapping were investigated by Laczkovich in his fundamental study on squaring the disk [6]. He considered sets which may be mapped into the regular grid by wobbling mappings.

The simple idea behind this concept is related also to physics and crystallography. Consider, for example, an amount of iron filings distributed in the plane to which an electrical field of finite energy is applied. The filings will move into an arranged position along the lines of the field. As long as the electrical field has small energy it is expected that no element is moved too far. Similarly, a faulty crystal can be imagined to be obtained from a regular crystal by moving certain elements by some small distance. Such mappings occur in many applications and may be treated in several ways [10].

In this note we outline some aspects of wobbling mappings in arbitrary metric spaces. The Banach-Tarski's theorem states that the unit ball in $\mathbb{R}^3$ may be decomposed into two parts which are piecewise congruent to the unit ball. We shall consider analogues of the Banach-Tarski's phenomenon in arbitrary metric spaces. We will characterize those metric spaces which may be decomposed into two parts, where both parts are equivalent to the whole metric space by a wobbling bijection.

2. Wobbling equivalences

DEFINITION 1. *Let $(\mathcal{M}, d)$ be a metric space and $X, Y \subseteq M$. An injective mapping $\varphi : X \to Y$ is called* k**–wobbling** *if*

$$\sup_{x \in X} d(\varphi(x), x) < k.$$

1991 *Mathematics Subject Classification.* Primary 54E50; Secondary 11K36.
Key words and phrases. Graphs, matchings, metric spaces, paradoxical sets.

We call X and Y **wobbling equivalent** *if for some k there is a k–wobbling bijection $\varphi : X \to Y$.*

Copying the proof of the Cantor-Bernstein theorem one gets

LEMMA 2.1. *Let X_1 and X_2 be subsets of a metric space $(\mathcal{M}, d)$ and φ_1 be a wobbling mapping of X_1 onto a subset $Y_2 \subseteqq X_2$ and φ_2 be a wobbling mapping of X_2 onto a subset $Y_1 \subseteqq X_1$. Then X_1 and X_2 are wobbling equivalent.*

DEFINITION 2. *For two sets $X, Y \subseteqq \mathcal{M}$ the bipartite k–distance graph $G_k(X, Y)$ is the bipartite graph with colour classes X and Y, where $x \in X$ is joined to $y \in Y$ iff $d(x, y) < k$.*

Clearly, X has a wobbling injection into Y if there exists a constant k such that $G_k(X, Y)$ contains a matching covering X. A metric space $(\mathcal{M}, d)$ is **discrete** if every bounded subset of $\mathcal{M}$ is finite. For discrete metric spaces the k–distance graphs $G_k(X, Y)$ are locally finite.

We denote by $N_k(Z)$ the k–neighbourhood of a set Z in M.

Applying the Rado–Hall theorem [10] for matchings in countable locally finite bipartite graphs to the k–distance graphs above gives the following

CLAIM 2.2. *Let $(\mathcal{M}, d)$ be a discrete countable metric space. Two subsets X, Y of $\mathcal{M}$ are wobbling–equivalent iff there exists a constant $k > 0$ such that*

(i) *For every finite subset X' of X $|N_k(X')) \cap Y| \geqq |X'|$.*

(ii) *For every finite subset Y' of Y $|N_k(Y') \cap X| \geqq |Y'|$.*

Sets which are equivalent to $\mathbb{Z}^d$ are called **uniformly spread** [7]. In the geometric setup one can make the transition from "counting" as in Claim 2.4 to "measuring volumes": Let $X \subseteqq \mathbb{R}$. To each $x \in X$ associate the unit cube with lower left corner in x:

$$C^d(x) = x + [0, 1)^d.$$

Heuristically one would say that if a set X is uniformly spread, then for every finite set $X' \subseteqq X$ the cardinality $|X'|$ may be approximated by the volume $\lambda_d(\bigcup_{x \in X} C^d(x))$ where λ_d is the d–dimensional volume.

For a subset $X \subseteqq \mathbb{R}^d$, $C \subseteqq \mathbb{R}^d$ $||X \cap C| - \lambda_d(C)|$ is called the discrepancy of X relative to C and denoted by $\Delta(X, C)$.

To prove his famous result on "squaring the disk" Laczkovich proved the following [6].

THEOREM 2.3 (Laczkovich). *A subset X of $\mathbb{R}^d$ is equivalent to $\mathbb{Z}^d$ if there exists a constant L such that for every measurable set $C \subset \mathbb{R}^d$ the following holds*

$$\Delta(X, Y) \leqq L\lambda_d(N_1(\partial C)),$$

where ∂C denotes the boundary of C.

For $d = 2$ there is a variant of this theorem [6].

THEOREM 2.4 (Laczkovich). *A subset X of $\mathbb{R}^2$ is equivalent to $\mathbb{Z}^2$ if there exists a constant L such that for every Jordan domain C of diameter at least 1 the following holds*

$$\Delta(X, C) \leqq L\lambda_1(\partial C).$$

As a corollary of the above theorem, one can easily show

COROLLARY 2.6 (see also P. Pleasants [9]). *Every Penrose tiling is equivalent to $\tau\mathbb{Z}^2$ for some $\tau \in \mathbb{R}$.*

PROOF. By De Bruijn's theorem [2] every Penrose tiling P is obtained as follows: There exists a 2-dimensional plane $E \subset \mathbb{R}^5$ and a constant l such that the orthogonal projection Π from $\mathbb{R}^5$ to E satisfies

$$\prod(N_l(E) \cap \mathbb{Z}^5) = P.$$

It is easy to verify that $N_l(E) \cap \mathbb{Z}^5$ satisfies the discrepancy condition of the theorem of Laczkovich. Then the projection — which is injective in this case — is a wobbling mapping as l is fixed. □

In a metric space much less is known in general about wobbling equivalence. Of course, there are general theorems guaranteeing the existence of injections such as the extensions of Hall' s theorem by Michael Holz, Klaus Peter Podewski, Karsten Steffens [4]. It could well be that an application of these theorems gives new insight in the context of wobbling equivalences.

PROBLEM 1. Characterize the sets which are wobbling equivalent to $\mathbb{Z}^2$.

The same problem could be of interest for $X \subset \mathbb{Q}^2$.

3. Paradoxical sets

DEFINITION 3. *Two sets A, B in $\mathbb{R}^3$ are called* **piecewise congruent** *if there exist decompositions $A = A_1 \dot\cup \ldots \dot\cup A_n$, $B = B_1 \dot\cup \ldots \dot\cup B_n$ such that each A_i is congruent to B_i.*

In their classical paper Banach and Tarski [1], see also Wagon [12] proved that the unit ball B in $\mathbb{R}^3$ is paradoxical in the following sense: B can be decomposed into two disjoint sets B_1, B_2 so that B_1, B_2 and B are pairwise piecewise congruent. Whenever one has an equivalence relation on the powerset of some set, one can define paradoxical sets. Here we define paradoxical sets only for the wobbling equivalence.

DEFINITION 4. *Let $(\mathcal{M}, d)$ be a metric space. $(\mathcal{M}, d)$ is* **paradoxical** *if there exists a decomposition $\mathcal{M} = M_1 \dot\cup M_2$ such that M_1, M_2 and $\mathcal{M}$ are pairwise wobbling equivalent.*

EXAMPLE 1. $\mathbb{R}^2$ is paradoxical. Take a checkerboard tiling of the plane. A translation moves the black tiles into the white ones. Moreover any single

square is equivalent to a domino. This shows that $\mathbb{R}^2$ and the set of black tiles are equivalent.

EXAMPLE 2. Let $\mathcal{M} = \{\log n | n \in \mathbb{N}\}$. Then $M_1 = \{\log(2n+1) \mid n \in \mathbb{N}\}$ and $M_2 = \{\log 2n | n \in \mathbb{N}\}$ show that $\mathcal{M}$ is paradoxical.

In order to characterize paradoxical sets (for wobbling equivalence) we introduce the following

DEFINITION 5. *Let $(\mathcal{M}, d)$ be a discrete metric space. $\mathcal{M}$ has* **exponential growth rate** *if*

$(*)$ *there exists a k (the doubling radius) such that for every finite set M' the k–neighbourhood $N_k(M')$ contains at least $2|M'|$ elements.*

REMARK. Obviously, the condition $(*)$ above is equivalent to that *for some fixed $q > 1$ there exists a k such that for every finite set M' the k–neighbourhood $N_k(M')$ has at least qM' elements.*

THEOREM 3.1. *Let $(\mathcal{M}, d)$ be a discrete countable metric space. Then the following are equivalent.*

(i) *M is paradoxical.*

(ii) *M has exponential growth rate.*

One should be aware that this theorem is not just a rewriting of definitions. To check exponential growth rate one has local tests: For every finite set one establishes the doubling radius. $\mathcal{M}$ is paradoxical if all these local doubling radii remain bounded. On the other hand paradoxity is a global property.

For the proof we need a variant of Hall's theorem.

DEFINITION 6. *Let $G = (A, B)$ be a bipartite graph. A set E of edges is an (l_1, l_2)-*__matching__ *if every vertex of A is contained in exactly l_1 edges of E and every vertex of B is contained in exactly l_2 edges of E.*

We need the following

GENERALIZED HALL-RADO THEOREM. *Let $G = (A, B)$ be a countable locally finite bipartite graph. G contains an (l_1, l_2)-matching iff the following two conditions are satisfied.*

(i) *For every finite subset A' of A there are at least $l_1|A'|$ neighbours in B.*

(ii) *For every finite subset B' of B there are at least $l_2|B'|$ neighbours in A.*

PROOF OF THE THEOREM. Let $\mathcal{M}$ be paradoxical. Then there exists a $k \in \mathbb{R}$ such that for every finite subset M' of $\mathcal{M}$ the k–neighbourhood $N_k(M')$ contains two disjoint sets of cardinality $|M'|$.

Indeed, let $\mathcal{M} = M_1 \dot\cup M_2$ be paradoxical decomposition with wobbling distance k. Then both $M_1 \cap N_k(M')$ and $M_2 \cap N_k(M')$ have at least $|M'|$ elements. Hence $\mathcal{M}$ has exponential growth rate.

To see the converse statement, observe that M is paradoxical iff for some k the k–distance graph $G_k(\mathcal{M},\mathcal{M})$ contains a (2,1) matching. To ensure a (2,1)-matching, we use the condition of the generalized Hall-Rado theorem for $G_k(\mathcal{M},\mathcal{M})$ with $(l_1,l_2)=(2,1)$. The exponential growth rate implies the Hall condition for $G_k(\mathcal{M},\mathcal{M})$ with $(2,2)$, and therefore with $(2,1)$ as well. □

4. Paradoxical graphs

Any graph G can be regarded as a metric space, where the distance $d(x,y)$ is the length of the shortest path between x and y in G.

PROBLEM 1. When is an infinite graph G paradoxical?

For trees this question can be answered easily. Let us call a path $P_k \subseteq G$ a **hanging chain** if all its inner vertices have degree 2 in G.

THEOREM 4.1. *A locally finite infinite tree T without endvertices is paradoxical iff the lengths of hanging chains in T is bounded.*

Here it should be remarked that when a tree T is decomposed into 2 subsets wobbling equivalent with each other and with the whole tree, these subsets are not trees.

COROLLARY 4.2. *An infinite tree is paradoxical if its minimum degree is at least* 3.

FIRST PROOF. Assume that T contains no hanging chain of K inner vertices (i.e. $K+1$ edges). For any $S \subseteq V(T)$ we define ∂S as the set of vertices joined to S but not in S. We may apply our characterization, Theorem 3.4 to T: the only thing to be proved is that if $S \subseteq V(T)$, then $|\partial S| \geqq c_K|S|$. Indeed, let F_n be the forest induced by the set $S \cup \partial S$. We shall count the vertices of degree 1 in F_n since all they belong to ∂S. Let n_i be the set of vertices of degree i, $n_{\geqq 3} = n_3 + n_4 + \ldots$. Then (for any tree or forest) $n_1 \geqq n_{\geqq 3} + 2$. Further, $n - n_2 > n/K$. Indeed, fix a vertex w of degree 1 and map each x of degree 2 to the y for which x is on a hanging chain yy^* and y is farther from w then y^*. We get each y at most K times. Thus $n_1 = n - n_2 - n_{\geqq 3} > n - n/K - n_1$ implying $|\partial S| > n/2K$. Hence T has exponential growth rate. □

One feels that in case of trees a directly constructed partition should also exist. One can easily provide the partition $V(T) = V_1 \dot\cup V_2$, e.g. if T is a tree of minimum degree 3.

PROBLEM 2. When is an infinite graph paradoxical? Is it true that if an infinite graph G is paradoxical, then there is an infinite spanning tree $T \subseteq G$ which is paradoxical?

5. Recursive sets

Often one would like to ensure some extra properties of the (wobbling) mapping or of the parts in a paradoxical partition under the condition that the original sets have additional properties. From the point of view of mathematical logic, those things are interesting for us which can be generated by a Turing Machine. This motivates the problems below.

PROBLEM 3. Let $X \subseteq \mathbb{Z}^2$ be recursive and wobbling equivalent to $\mathbb{Z}^2$. Is there a recursive wobbling bijection $X \to \mathbb{Z}^2$ which is recursive?

PROBLEM 4. Are there recursive paradoxical sets $\mathcal{M}$ in $\mathbb{Q}^2$ for which there is no recursive paradoxical decomposition $\mathcal{M} = M_1 \dot{\cup} M_2$?

REMARK. We do not think that there is a trivial positive answer. There are analogous situations with negative answers.

(a) There exists a recursive countable locally finite tree (i.e. the characteristic function of the edge set is recursive) which has no recursive infinite path [11].

(b) There exists a recursive k–regular bipartite graph $G(A, B)$ which has a 1-factor but has no recursive 1-factor. [8].

REFERENCES

[1] BANACH, ST. and TARSKI, A., Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fund. Math.* **6** (1924), 244–277. *Jb. Fortschritte Math.* **50**, 370

[2] BRUIJN, N. G. DE, Algebraic theory of Penrose's non-periodic tilings of the plane, *Nederl. Akad. Wetensch. Indag. Math.* **43** (1981), 39–52, 53–66. *MR* **82e**:05055

[3] HALL, P., On representatives of subsets, *J. London Math. Soc.* **10** (1935), 26–30. *Zbl* **10**, 345

[4] HOLZ, M., PODEWSKI, K.-P. and STEFFENS, K., *Injective choice functions*, Lecture Notes in Mathematics, 1238, Springer-Verlag, Berlin–New York, 1987. *MR* **88d**:04002

[5] KÖNIG, D., Über eine Schlussweise aus dem Endlichen ins Unendliche, *Acta Sci. Math. (Szeged)* **3** (1927), 121–130. *Jb. Fortschritte Math.* **53**, 170

[6] LACZKOVICH, M., Equidecomposability and discrepancy; a solution of Tarski's circle-squaring problem, *J. Reine Angew. Math.* **404** (1990), 77–117. *MR* **91b**:51034

[7] LACZKOVICH, M., Uniformly spread discrete sets in $\mathbb{R}^d$, *J. London Math. Soc.* (2) **46** (1992), 39–57. *MR* **93i**:11088

[8] MANASTER, A. B. and ROSENSTEIN, J. G., Effective matchmaking and k–chromatic graphs, *Proc. Amer. Math. Soc.* **39** (1973), 371–378. *MR* **49** #4838

[9] PLEASANTS, P., Personal communication, 1991.

[10] RADO, R., Note on the transfinite case of Hall's theorem on representatives, *J. London Math. Soc.* **42** (1967), 321–324. *MR* **35** #2758

[10] SOARDI, P. M., *Potential theory on infinite networks*, Lecture Notes in Mathematics, 1590, Springer-Verlag, Berlin-Heidelberg, 1994.

[11] SPECKER, E., Eine Verschärfung des Unvollständigkeitssatzes der Zahlentheorie, *Bull. Acad. Polon. Sci. Cl. III.* **5** (1957), 1041–1045. *MR* **19**–934

[12] WAGON, S., *The Banach-Tarski paradox*, Encyclopedia of Mathematics and its Applications, **24**, Cambridge University Press, Cambridge–New York, 1985. *MR* **87e**:04007

(Received February 14, 1994)

W. A. Deuber

FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
UNIVERSITÄTSTRASSE 25
D–33615 BIELEFELD
FEDERAL REPUBLIC OF GERMANY

M. Simonovits and V. T. Sós

MTA MATEMATIKAI KUTATÓINTÉZETE
PF. 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 25–46

ON THE BOOK SIZE OF GRAPHS WITH LARGE MINIMUM DEGREE

P. ERDŐS, R. FAUDREE and E. GYŐRI

The basic problem investigated in this paper was raised by the first two authors [4] in their paper on size Ramsey numbers. Among others, they dealt with the questions of which graphs must be contained in a graph whose complement does not contain a given tree as a subgraph. An even more special problem was raised in this paper, which is of special interest by its consequences and by the phenomenon described by it. To formulate this problem, we need the graph theoretical concept of "book" introduced by Shehan.

DEFINITION. The graph consisting of k triangles sharing a common edge is called a *book of k pages*. More precisely, the book B_k of k pages can be defined by the vertex set $V = \{a, b, x_1, x_2, \dots, x_k\}$ and the edge set $E = \{ab, ax_1, bx_1, ax_2, bx_2, \dots, ax_k, bx_k\}$.

Let G be a graph of n vertices with minimum degree d. Question: How thick a book must be contained in G? In other words, what is the maximum k such that a book of k pages is contained in any graph G of n vertices with minimum degree d. Naturally, if $d \leqq n/2$ then it is possible that G does not contain any triangle (think of the complete bipartite graph with $n/2$ vertices in each part) so the question is not interesting if $d \leqq n/2$.

The case $d \geqq \lfloor n/2 \rfloor + 1$ is interesting from several points of view. E.g., the results of Dirac [2] and Bondy [1] imply that every graph of n vertices and with minimum degree not less than $\lfloor n/2 \rfloor + 1$ is pancyclic (i.e., the graph contains a cycle of length k for any $3 \leqq k \leqq n$).

As to the books — as we will see — it is easy to prove that every graph of n vertices and with minimum degree not less than $\lfloor n/2 \rfloor + 1$ contains a book of $n/6$ pages:

THEOREM 1. *Every graph G of n vertices and with minimum degree not less than $\lfloor n/2 \rfloor + 1$ contains a book $B_{\lfloor n/6 \rfloor + 1}$ of $\lfloor n/6 \rfloor + 1$ pages as a subgraph.*

Theorem 1 is basically a direct consequence of the nice result of Edwards [3] that any graph with $\lfloor n^2/4 \rfloor + 1$ edges has a book with at least $n/6$ pages. However, we prove Theorem 1 for sake of self-containedness and since the proof of the more general statements starts in the very same way. (This is a weaker statement than Edwards' result, and the proof is simpler, as well.)

1991 *Mathematics Subject Classification.* Primary 05C35; Secondary 05C55.
Key words and phrases. Extremal graphs, triangles, degree conditions.

PROOF. The degree condition implies that G contains a triangle; furthermore, every edge of it is contained in a triangle. For an arbitrary triangle xyz, let us estimate the number of vertices of G not joined to any of the vertices x, y, z by means of principle of inclusion and exclusion. We obtain that

$$\begin{aligned} 0 \leqq |V(G) - (N(x) \cup N(y) \cup N(z))| = \\ = |V(G)| - |N(x)| - |N(y)| - |N(z)| + |N(x) \cap N(y)| + \\ + |N(x) \cap N(z)| + |N(y) \cap N(z)| - |N(x) \cap N(y) \cap N(z)| \\ \leqq n - 3(\lfloor n/2 \rfloor + 1) + |N(x) \cap N(y)| + |N(x) \cap N(z)| + |N(y) \cap N(z)|, \end{aligned}$$

which implies that one of the double intersections, say $N(x) \cap N(y)$ is of size at least $\lfloor n/2 \rfloor + 1 - n/3 > n/6$ what we wanted to prove since then x, y and $N(x) \cap N(y)$ define a book of $|N(x) \cap N(y)|$ pages. □

The surprisingly difficult question now is how sharp is Theorem 1. Is there a constant $\varepsilon > 0$ such that a book of $(1/6 + \varepsilon)n$ pages can be found in every graph G of n vertices and with minimum degree not less than $\lfloor n/2 \rfloor + 1$? We show that there exists such a constant; however, we do not calculate the value ε that can be deduced from the proof. Partly, since the arguments are very technical and it is very doubtful that the proof results in the best possible constant, and partly, since this phenomenon of a "jumping constant" is far more interesting, than the value of ε. The phenomenon is described in a stability theorem, more exactly, we prove its generalizations as well in a series of stability theorems. These theorems imply the interesting "jumps" of the function $b(c)$ defined in the interval $[0, 1]$ as the maximum b such that every graph G of (sufficiently large) n vertices and with minimum degree $cn + o(n)$ contains a book of bn pages.

We have seen in the proof above that G contains a book of $n/6$ pages. However, the same proof implies that if G does not contain a book of $(1/6 + \varepsilon)n$ pages then *every* edge of it is the common "spine" edge of a book of $n/6 + o(n)$ pages. Let us weaken the conditions of Theorem 1 a bit and suppose only that the degree of every vertex in G is at least $n/2 + o(n)$. This condition does not imply the existence of any triangle, although the original conditions imply that every edge is contained in a triangle. In order to be able to say something interesting, let us assume again that every edge is contained in a triangle and that any book in G has at most $n/6 + o(n)$ pages. Then these conditions essentially determine the structure of G in a unique way:

THEOREM 2. *Let G be a graph of n vertices such that the degree of every vertex is at least $n/2 + o(n)$ and every edge is contained in a triangle. If every book in G has at most $n/6 + o(n)$ pages, then deleting from and adding to G at most $o(n^2)$ edges, we can obtain the following graph G' of 15 (in some cases 12) classes of vertices:*

For the 15 classes V_{ij} $(1 \leqq i, j \leqq 4,\ i+j<8)$, *we have* $|V_{ij}| = n/12 + o(n)$ $(i=1,2,3; j=1,2,3)$, $|V_{4j}| = c_0 n/12 + o(n)$ $(j=1,2,3)$, $|V_{i4}| = (1-c_0)n/12 + o(n)$ $(i=1,2,3)$, *where* $0 \leqq c_0 \leqq 1$, *and the vertices* $x \in V_{i_1 j_1}$ *and* $y \in V_{i_2 j_2}$ *are joined to each other if and only if* $i_1 \neq i_2$, $j_1 \neq j_2$, *and* $i_1 + j_2, i_2 + j_1 < 8$.

This theorem can be generalized, and it is interesting that the structure of the graph is determined in an even more unique way in the general case. The description in the general case does not contain any free parameter as c_0 is in the special case above.

THEOREM 3. *Let* $k, l \geqq 3$ *be arbitrary integers such that* $\frac{k-1}{k}\frac{l-1}{l} > \frac{l-2}{l-1}$ *(i.e.,* $k > l^2 - 2l + 1$*). Let* G *be a graph of* n *vertices such that the degree of any vertex in* G *is at least* $\frac{k-1}{k}\frac{l-1}{l}n + o(n)$. *If any book in* G *has at most* $\frac{k-2}{k}\frac{l-2}{l}n + o(n)$ *pages, then deleting from and adding to* G *at most* $o(n^2)$ *edges, we can obtain the following graph* G' *with* kl *classes of vertices:*

For the kl *classes* V_{ij} $(1 \leqq i \leqq k,\quad 1 \leqq j \leqq l)$, $|V_{ij}| = \frac{n}{kl} + o(n)$ $(i = 1, \ldots, k; j = 1, \ldots, l)$, *and the vertices* $x \in V_{i_1 j_1}$ *and* $y \in V_{i_2 j_2}$ *are joined to each other if and only if* $i_1 \neq i_2$ *and* $j_1 \neq j_2$.

Now look at the consequences of these theorems concerning the behaviour of the function $b(c)$. It is obvious that the function b is monotone increasing and it is easy to see that it is continuous from the left-hand side, since deleting edges changes the sizes of the books sitting on the remaining edges as "spine" edges, and if an edge e is the spine of an exceptionally thick book then instead of this edge, delete other edges incident to the endvertices of e (e.g., the other edges of this book). What is more surprising and follows from Theorem 2 is

THEOREM 4. *There is an* $\varepsilon > 0$ *such that* $b(c) > 1/6 + \varepsilon$ *for any* $c > 1/2$.

PROOF. Suppose that all the degrees in G are greater than $n/2$ and all the books have at most $n/6 + o(n)$ pages. The minimum degree in G' is at most $n/2$ so G is not a subgraph of G'. Furthermore, since the order of magnitude of the number of vertices in G' with degree at most $n/2$ is n, the number of independent edges in $G - G'$ is of order of magnitude n, as well. It is easy to verify that if we add an edge e of G to G' then it is the spine of a book of $n/4 + o(n)$ pages in $G' \cup e$ and then G does not contain at least $n/12 + o(n)$ page edges of this book by the page number condition on G. But three books with independent spine edges have no common edges, (i.e., if we have books with independent spine edges then any edge is contained in at most two books), so we have to delete an edge set of order of magnitude n^2 from G' (and add some edges) to obtain the graph G with no books of size greater than $n/6 + o(n)$, a contradiction to Theorem 2. □

REMARK. The proof above is not precise enough, the use of orders of magnitude in case of a given graph is meaningless. However, these inaccuracies can be eliminated by means of the usual ε-technique. Actually, the proof

above would not be too complicated, but considering that this ε-technique would make the next proofs vast, we will follow this method and sacrifice the full preciseness for the sake of clearness.

Replacing Theorem 2 by Theorem 3, a similar argument shows

THEOREM 5. *There is a constant $\varepsilon(k,l)>0$ such that $b(c)>\frac{k-2}{k}\frac{l-2}{l}+$ $+\varepsilon(k,l)$ for any $c>\frac{k-1}{k}\frac{l-1}{l}$ if k and l satisfy the conditions of Theorem 3.*

Applying Theorem 5 and the construction of G', it can be seen easily that the function $b(c)$ is not continuous, it has some "jumpings" at the points $\frac{k-1}{k}\frac{l-1}{l}$ if k and l satisfy the conditions of Theorem 3.

Theorem 2 and the case $l=3$ of Theorem 3 can be proved in the same way not considering that the proof of some claims are much more involved in case of Theorem 2, i.e., if $k=4$. So, we will prove Theorem 2 and the case $l=3$ of Theorem 3 simultaneously. Then, we will prove Theorem 3 by induction on l.

PROOF OF THEOREM 2 AND THE CASE $l=3$ OF THEOREM 3. Let $k\geqq 4$ and let G be a graph of n vertices such that the degree of every vertex is at least $\frac{2(k-1)}{3k}n+o(n)$. Suppose that every edge is contained in a triangle (it automatically follows from the degree condition if $k>4$) and that every book in G has at most $\frac{k-2}{3k}n+o(n)$ pages. From now on, we refer to this second assumption as the page number condition.

Let xyz be a triangle in G and let us introduce the following notation:

$$\begin{aligned}
V_x &= N(x)-N(y)-N(z),\\
V_y &= N(y)-N(x)-N(z),\\
V_z &= N(z)-N(x)-N(y),\\
V_{xy} &= N(x)\cap N(y),\\
V_{xz} &= N(x)\cap N(z)-N(x)\cap N(y)\cap N(z),\\
V_{yz} &= N(y)\cap N(z)-N(x)\cap N(y)\cap N(z).
\end{aligned}$$

Like in the proof of Theorem 1, let us estimate the number of vertices not joined to any of the vertices x,y,z, by applying the principle of inclusion and exclusion, the degree condition, and the page number condition. We obtain that

$$\begin{aligned}
0\leqq |V(G)-(N(x)\cup N(y)\cup N(z))| &=\\
=|V(G)|-|N(x)|-|N(y)|-|N(z)|+|N(x)\cap N(y)|+&\\
+|N(x)\cap N(z)|+|N(y)\cap N(z)|-|N(x)\cap N(y)\cap N(z)|\leqq&\\
\leqq n-3\frac{2(k-1)}{3k}n+3\frac{k-2}{3k}n+o(n)=o(n),&
\end{aligned}$$

which implies that the estimates used are sharp apart from an error term $o(n)$, so, $d(x),d(y),d(z)=\frac{2(k-1)}{3k}n+o(n)$, $|N(x)\cap N(y)\cap N(z)|=o(n)$, $|V_{xy}|$,

$|V_{xz}|, |V_{yz}| = \frac{k-2}{3k}n + o(n)$. This implies that $|V_x|, |V_y|, |V_z| = \frac{2}{3k}n + o(n)$ and the six sets defined above cover the vertex set of G with exception of $o(n)$ vertices. Since xy was an arbitrary edge of G, it implies that $d(v) = \frac{2(k-1)}{3k}n + o(n)$ for every vertex v of G and every edge of G is the spine of a book of $\frac{k-2}{3k}n + o(n)$ pages. From now on, the degree condition and the page number condition can be used in this stronger form.

Now, we are going to determine the number of edges joining these six vertex classes and, as far as possible, the structure of the graph. Let $d(X, Y)$ and $d(X)$ denote the number of edges from X to Y and joining two elements of X, respectively.

Let $v \in V_{xy}$. Then applying the estimate above for the triangle xyv, we obtain that $|N(x) \cap N(y) \cap N(v)| = o(n)$, and $d(v, V_{xy}) = o(n)$, and so, $d(V_{xy}) = o(n^2)$. (Similarly, if $v \in V_{xz}$, then $d(v, V_{xz}) = o(n)$ and $d(V_{xz}) = o(n^2)$, and if $v \in V_{yz}$, then $d(v, V_{yz}) = o(n)$ and $d(V_{yz}) = o(n^2)$.)

Again, let $v \in V_{xy}$. The page number condition for the edge xv implies that $d(v, V_x) + d(v, V_{xy}) + d(v, V_{xz}) = \frac{k-2}{3k}n + o(n)$, and the same condition for the edge yv implies that $d(v, V_y) + d(v, V_{xy}) + d(v, V_{yz}) = \frac{k-2}{3k}n + o(n)$. But applying $d(v, V_{xy}) = o(n)$ and $d(v) = \frac{2(k-1)}{3k}n + o(n)$, it follows

$$d(v, V_x) + d(v, V_{xz}) = \frac{k-2}{3k}n + o(n),$$
$$d(v, V_y) + d(v, V_{yz}) = \frac{k-2}{3k}n + o(n),$$

and

$$d(v, V_z) = \frac{2}{3k}n + o(n).$$

From now on, we refer to this statement as the sum condition (for the appropriate edge). Thus, V_{xy} and V_z are joined to each other almost completely,

$$d(V_{xy}, V_z) = \frac{2(k-2)}{9k^2}n^2 + o(n^2).$$

Similarly,

$$d(V_{xz}, V_y) = \frac{2(k-2)}{9k^2}n^2 + o(n^2),$$

and

$$d(V_{yz}, V_x) = \frac{2(k-2)}{9k^2}n^2 + o(n^2).$$

Furthermore, if $v \in V_{xz}$ then

$$d(v, V_x) + d(v, V_{xy}) = \frac{k-2}{3k}n + o(n),$$
$$d(v, V_z) + d(v, V_{yz}) = \frac{k-2}{3k}n + o(n),$$
$$d(v, V_y) = \frac{2}{3k}n + o(n),$$

and if $v \in V_{yz}$, then

$$d(v, V_y) + d(v, V_{xy}) = \frac{k-2}{3k}n + o(n),$$
$$d(v, V_z) + d(v, V_{xz}) = \frac{k-2}{3k}n + o(n),$$
$$d(v, V_x) = \frac{2}{3k}n + o(n).$$

Consider the vertices $v \in V_x$. The page number condition for the edge xv says that $d(v, V_x) + d(v, V_{xy}) + d(v, V_{xz}) = \frac{k-2}{3k}n + o(n)$. On the other hand, $d(V_{yz}, V_x) = \frac{2(k-2)}{9k^2}n^2 + o(n^2)$ implies that $d(v, V_{yz}) = \frac{k-2}{3k}n + o(n)$ for a typical vertex $v \in V_x$ (i.e., for all vertices $v \in V_x$ except $o(n)$ ones). By the degree condition for a typical vertex $v \in V_x$, $d(v, V_y) + d(v, V_z) = \frac{2}{3k}n + o(n)$ and so $d(V_x, V_y) + d(V_x, V_z) = \frac{4}{9k^2}n^2 + o(n^2)$. Similarly, $d(V_y, V_x) + d(V_y, V_z) = \frac{4}{9k^2}n^2 + o(n^2)$ and $d(V_z, V_y) + d(V_z, V_x) = \frac{4}{9k^2}n^2 + o(n^2)$. So, $d(V_x, V_y) = \frac{2}{9k^2}n^2 + o(n^2)$, $d(V_x, V_z) = \frac{2}{9k^2}n^2 + o(n^2)$, and $d(V_y, V_z) = \frac{2}{9k^2}n^2 + o(n^2)$. It can be shown, as well, that if w is a typical vertex in V_y, then $d(w, V_{xz}) = \frac{k-2}{3k}n + o(n)$, and if w is a typical vertex in V_z, then $d(w, V_{xy}) = \frac{k-2}{3k}n + o(n)$. These results can be shown for every triangle or if just an edge is needed then for every edge, and we refer to them, as the structure condition.

Before continuing the proof of the Theorem, we prove

LEMMA 1. *The graph G does not contain K_4 as a subgraph.*

PROOF OF LEMMA 1. Suppose that the vertices x, y, z, v induce a K_4 in G and that the vertices x, y, z have the properties what we have seen above. We have $v \in V_{xy}$, so, as we have seen, $d(v, V_{xy}) = o(n)$, $d(v, V_x) + d(v, V_{xz}) = \frac{k-2}{3k}n + o(n)$, $d(v, V_y) + d(v, V_{yz}) = \frac{k-2}{3k}n + o(n)$. On the other hand, the page number condition for the edge vz and $d(v, V_z) = \frac{2}{3k}n + o(n)$ imply that $d(v, V_{xz}) + d(v, V_{yz}) = \frac{k-4}{3k}n + o(n)$. However, we saw that $d(v, V_x) + d(v, V_y) = \frac{k}{3k}n + o(n)$, a contradiction to the equalities $|V_x| = \frac{2}{3k}n + o(n)$ and $|V_y| = \frac{2}{3k}n + o(n)$ if $k > 4$.

Unfortunately, the proof in the case $k=4$ is much more complicated, it is much more difficult to get the desired contradiction. Then, all the six sets defined above have $n/6+o(n)$ elements, and as we have seen in this proof, $d(v,V_{xz}), d(v,V_{yz}) = o(n)$ and $d(v,V_x), d(v,V_y) = n/6+o(n)$. On the other hand, the same estimates can be shown for the triangles vxy, vxz, vyz, which imply that $d(V_x), d(V_y), d(V_z) = o(n^2)$ and $d(V_x,V_{xy}), d(V_x,V_{xz}), d(V_y,V_{xy})$, $d(V_y,V_{yz}), d(V_z,V_{xz}), d(V_z,V_{yz}), d(V_{xz},V_{xy}), d(V_{yz},V_{xy}), d(V_{xz},V_{yz}) = n^2/72+$ $o(n^2)$.

Now, we show Lemma 1 by means of three claims.

CLAIM 1. *With exception of $o(n^3)$ triangles, every triangle in G contains a vertex in V_{xy} and a vertex in V_z, or a vertex in V_{xz} and a vertex in V_y, or a vertex in V_{yz} and a vertex in V_x.*

PROOF. We have

$$d(V_{xy},V_z)+d(V_{xz},V_y)+d(V_{yz},V_x) = n^2/12+o(n^2).$$

The page number condition says that every edge is the spine of a book of $n/6+o(n)$ pages, so the number of triangles containing at least one of these edges is $n^3/72+o(n^3)$, since the number of triangles counted twice is $o(n^3)$ because $d(V_x), d(V_y), d(V_z), d(V_{xy}), d(V_{xz}), d(V_{yz}) = o(n^2)$. But this is the total number of triangles, as well, since the number of edges in G is $n^2/4+$ $+o(n^2)$ by the degree condition, and each edge is contained in $n/6+o(n)$ triangles, so counting every triangle three times, we count $n^3/24+o(n^3)$ triangles. □

CLAIM 2. *Let $w \in V_{xy}$ be an arbitrary vertex. Then $N(w) \cap V_{xz}$ and $V_{yz} -$ $- N(w)$, $N(w) \cap V_{yz}$ and $V_{xz} - N(w)$ are joined to each other completely with exception of $o(n^2)$ edges. Considering the symmetry of the vertices x, y, z, v, another 11 similar statements hold, as well.*

PROOF. Defining the appropriate sets for the triangle wxy, the statements follow from the fact that V_{wx} and V_y, V_{wy} and V_x are joined to each other completely with exception of $o(n^2)$ edges. □

CLAIM 3. *With exception of $o(n)$ vertices, either $d(w,V_x), d(w,V_{xz}) =$ $= n/12+o(n)$ or $d(w,V_y), d(w,V_{yz}) = n/12+o(n)$ for every vertex $w \in V_{xy}$. Similarly with exception of $o(n)$ vertices, $n/12+o(n)$ edges emanate from every vertex of G to at least two of the six defined sets.*

PROOF. We have $d(w,V_{xz}) = n/6 - d(w,V_x) + o(n)$ and $d(w,V_{yz}) = n/6 -$ $- d(w,V_y) + o(n)$ by the sum condition: it is sufficient to prove one of the equalities. Let us count the triangles containing a typical vertex w in two different ways.

The degree condition and the page number condition imply that each of the $n/2+o(n)$ edges incident to w is contained in $n/6+o(n)$ triangles; so,

the number of triangles containing w is $n^2/24 + o(n^2)$ since every triangle was counted twice in this way.

On the other hand, $d(V_z) = o(n^2)$ implies that the number of triangles containing any of the $n/6 + o(n)$ edges leading from w to V_z is $n^2/36 + o(n^2)$, and the number of triangles containing two of the remaining edges incident to w is $d(w, V_x)(n/6 - d(w, V_y)) + d(w, V_y)(n/6 - d(w, V_x)) + o(n^2)$ for a *typical* vertex w by Claim 1. And this is equal to $n^2/72 + o(n^2)$, only if either $d(w, V_x) = n/12 + o(n)$ or $d(w, V_y) = n/12 + o(n)$. □

Claims 2 and 3 imply that, for a typical vertex $w \in V_{xy}$, $N(w) \cap V_x$ and $N(w) \cap V_y$ define a partition of V_x and V_y, respectively, such that one of these two partitions is halving (apart from a possible error of $o(n)$) and, the other one is arbitrary, including that it may have an empty class, as well. One partition class in V_x is completely joined to one partition class in V_y and, the other partition class in V_x is completely joined to the other partition class in V_y, with exception of $o(n^2)$ edges. However, $d(V_x, V_y) = n^2/72 + o(n^2)$ implies that these two bundles contain all the edges from V_x to V_y with exception of $o(n^2)$ edges. It implies that every vertex $w \in V_{xy}$ (with exception of $o(n)$ vertices) defines almost the same partition in V_x and V_y. Thus, each (typical) vertex w is joined to exactly one of the partition classes in V_x and so, the graph of the edges from V_x to V_{xy} — whose structure is defined by a partition in the same way — partitions V_x in the very same way.

Now, we show that the edges joining V_x to V_y and the edges joining V_x to V_z define the very same partition of V_x if none of the partitions defined by the edges joining V_x, V_y and V_z has empty partition class or that of $o(n)$ elements. (Similar statements hold for V_x, V_{xy}, V_{xz}, for V_y, V_{xy}, V_{yz} and for V_z, V_{xz}, V_{yz}.)

If, say, a class P of the partition of the set V_x defined by the edges joining V_x and V_y contains more than $o(n)$ elements of both classes of the partition of the set V_x defined by the edges joining V_x and V_z then, (almost) every vertex of V_z can be reached from P by an edge and P is completely joined to cn vertices in V_y. There is no essentially empty partition class, so the number of edges leading from these vertices in V_y to V_z is of order of magnitude n^2 and we get $c_0 n^3$ triangles forbidden by Claim 1, a contradiction.

Now, suppose that there is no essentially empty partition class in G and so — as we have seen above — the partitions divide the six defined sets in two in the very same way. By Claim 1, there are only $o(n^3)$ triangles in $V_x \cup V_y \cup V_z$; thus, the partition classes in these sets can be colored with red and blue so that one class is red, the other one is blue in each partition, and exactly the classes of different colors in different sets are joined to each other almost completely. Again by Claim 1, a partition class in V_{xy} is joined almost completely to classes in V_x and V_y that are of the same color and similar statements hold for V_{xz} and V_{yz}. Thus, the coloring of the partition

classes can be extended to the classes in V_{xy}, V_{xz}, V_{yz} so that the classes of different colors are joined to each other in $V_{xy} \cup V_{xz} \cup V_{yz}$ and $V_x \cup V_y \cup V_z$, as well. And again by Claim 1, it yields a good coloring of the classes in V_{xy}, V_{xz}, V_{yz}, as well, exactly the classes of different colors in different sets are joined to each other almost completely. However, the edges from the red vertices in V_{xy} to the blue vertices in V_x are not contained in triangles such that the other two edges are from the almost complete bundles joining classes of different colors. The $o(n^2)$ edges not in these bundles result in very few triangles contradicting the page number condition.

Finally, suppose that the edges from V_x to V_y define a partition of V_x such that one of the classes has $o(n)$ elements. By Claim 3, the edges from V_x to V_y define a partition V_y', V_y'' of V_y such that $|V_y'|, |V_y''| = n/12 + o(n)$, V_y' is joined to V_x almost completely, and the number of edges from V_y' to V_x is $o(n^2)$. Claim 3 implies that the edges from V_y to V_{yz} define a partition V_{yz}', V_{yz}'' of V_{yz} such that $|V_{yz}'|, |V_{yz}''| = n/12 + o(n)$, and it essentially defines the partition V_y', V_y'' in V_y. Furthermore V_y' is joined to V_{yz}' and V_y'' is joined to V_{yz}'' almost completely, and the number of edges joining V_y' to V_{yz}'' and V_y'' to V_{yz}' is $o(n)$. Similarly, the edges from V_y to V_{xy} define a partition V_{xy}', V_{xy}'' of V_{xy} such that $|V_{xy}'|, |V_{xy}''| = n/12 + o(n)$, and it essentially defines the partition V_y', V_y'' in V_y. Furthermore V_y' is joined to V_{xy}' and V_y'' is joined to V_{xy}'' almost completely, and the number of edges joining V_y' to V_{xy}'' and V_y'' to V_{xy}' is $o(n)$. Both endvertices of a typical $V_y'V_{yz}'$-edge is joined to V_x almost completely. Thus, the page number condition implies that the number of edges from V_{yz}' to V_{xz} is $o(n^2)$; so since $d(V_{yz}, V_{xz}) = n^2/72 + o(n^2)$, V_{yz}'' is joined to V_{xz} almost completely. Similarly, the number of edges joining V_{yz}'' to V_{xz} is $o(n^2)$ and V_{yz}' is joined to V_{xz} almost completely. The number of $V_{yz}'V_{xy}'$-edges and $V_{yz}''V_{xy}''$-edges is $o(n^2)$, since such an edge is typically contained in $n/12 + o(n)$ triangles whose third vertex is in V_y, and Claim 1 says that the number of triangles of this type is $o(n^3)$. On the other hand, the number of $V_{yz}''V_{xy}'$-edges is $o(n^2)$, since such an edge is typically contained in $n/6 + o(n)$ triangles whose third vertex is in V_{xz}, and Claim 1 says that the number of triangles of this type is $o(n^3)$. However, these statements contradict the equality $d(V_{yz}, V_{xy}) = n^2/72 + o(n^2)$, and the proof of Lemma 1 is complete. □

Let us return to the proof of the Theorems.

From now on, let $d(V_x) = a, d(V_y) = b$ and $d(V_z) = c$. Turán's theorem implies that $a, b, c \leq \frac{n^2}{9k^2} + o(n^2)$, since G does not contain K_4 as a subgraph.

Then by the page number condition,

$$d(V_x, V_{xy}) + d(V_x, V_{xz}) + 2a = \frac{2(k-2)}{9k^2}n^2 + o(n^2),$$

$$d(V_x, V_{xy}) + d(V_{xy}, V_{xz}) = \frac{(k-2)^2}{9k^2}n^2 + o(n^2),$$

and

$$d(V_x, V_{xz}) + d(V_{xy}, V_{xz}) = \frac{(k-2)^2}{9k^2}n^2 + o(n^2).$$

It follows that

$$d(V_x, V_{xy}), d(V_x, V_{xz}) = \frac{k-2}{9k^2}n^2 - a + o(n^2),$$

and

$$d(V_{xy}, V_{xz}) = \frac{(k-3)(k-2)}{9k^2}n^2 + a + o(n^2).$$

Similarly,

$$d(V_y, V_{xy}), d(V_y, V_{yz}) = \frac{k-2}{9k^2}n^2 - b + o(n^2),$$

$$d(V_{xy}, V_{yz}) = \frac{(k-3)(k-2)}{9k^2}n^2 + b + o(n^2),$$

$$d(V_z, V_{xz}), d(V_z, V_{yz}) = \frac{k-2}{9k^2}n^2 - c + o(n^2),$$

and

$$d(V_{xz}, V_{yz}) = \frac{(k-3)(k-2)}{9k^2}n^2 + c + o(n^2).$$

Notice that a statement similar to Claim 1 can be shown about the triangles in G. Let $w \in V(G)$ be an arbitrary vertex. By Lemma 1, $N(w)$ does not contain any triangle. On the other hand, the degree condition and the page number condition imply that

$$d(N(w)) = \frac{(k-1)(k-2)}{9k^2}n^2 + o(n^2),$$

$$d(N(w), V(G) - N(w)) = \frac{2(k-1)}{9k}n^2 + o(n^2),$$

and so,

$$d(V(G) - N(w)) = \frac{2(k-1)}{9k^2}n^2 + o(n^2).$$

Note that the number of $N(w)(V(G) - N(w))$-edges in a triangle is 0 or 2, so by the page number condition, the number of triangles containing $N(w)(V(G) - N(w))$-edges is $\frac{(k-1)(k-2)}{9k^2}n^3 + o(n^3)$. However, by the degree condition and the page number condition, this is the total number of triangles in G (apart from an error of $o(n^3)$). Thus, the number of triangles in $V(G) - N(w)$ is $o(n^3)$. From now on, we refer to these results as to the triangle condition.

CLAIM 4. *We may assume that $o(n)$ is the number of vertices $v \in V_{xy}$ such that $d(v, V_x) = \frac{2}{3k}n + o(n)$. Similarly, we may assume that $o(n)$ is the number of vertices $v \in V_{xy}$ such that $d(v, V_y) = \frac{2}{3k}n + o(n)$, and analogous statements hold for vertices $v \in V_{xz}$ and $v \in V_{yz}$.*

PROOF. Suppose that n is the order of magnitude of the number of vertices $v \in V_{xy}$ such that $d(v, V_x) = \frac{2}{3k}n + o(n)$, and so $d(v, V_{xz}) = \frac{k-4}{3k}n + o(n)$ by the sum condition. Since G does not contain K_4, the existence of *one* v of this kind implies that $a = d(V_x) = o(n^2)$. For almost all neighbours $w \in V_x$ of such a *typical* v, the number of triangles containing vw is $d(v, V_{yz}) + d(w, V_z) + o(n)$ by the triangle condition for x and z. Hence (apart from the $o(n)$ possible exceptions), the edge numbers $d(w, V_z)$ for different vertices w differ from each other by $o(n)$, so, $d(w, V_z) = \frac{n}{3k} + o(n)$ for almost all vertices $w \in V_x$ because $d(V_x, V_z) = \frac{2}{9k^2}n^2 + o(n^2)$. Thus, $d(v, V_{yz}) = \frac{k-3}{3k}n + o(n)$ for typical such vertices v, and $d(v, V_y) = \frac{n}{3k} + o(n)$ because of the sum condition.

Let us fix such a typical vertex v for a while. Since $V_y - N(v)$ and V_x are joined to each other almost completely by the structure condition applied on the edge vx, the number of edges from $V_y \cap N(v)$ to V_x is $o(n^2)$ because $d(V_x, V_y) = \frac{2}{9k^2}n^2 + o(n^2)$. On the other hand, the sum condition implies that $V_y \cap N(v)$ and V_z are joined to each other almost completely and the number of edges from $V_y - N(v)$ to V_z is $o(n^2)$.

The structure condition on vy implies that $V_{xz} - N(v)$ and $V_{yz} \cap N(v)$ are joined to each other almost completely. Since G does not contain K_4 and so, the number of edges from $V_{xz} \cap N(v)$ to V_x is $o(n^2)$. Then, the sum condition implies that $V_{xz} \cap N(v)$ and V_{xy} are joined to each other almost completely. It follows that the number of edges from $V_{xz} \cap N(v)$ to V_z is $o(n^2)$, because if it is not the case, then such a typical edge is contained in $\frac{k-1}{3k}n + o(n)$ triangles with the third vertex in V_y and V_{xy}, a contradiction to the page number condition. On the other hand, the structure condition on the triangle vxy implies that $d(V_{xz} - N(v), V_z) = \frac{2}{9k^2}n^2 + o(n^2)$, and also by applying the results just proved above, it follows that $d(V_{xz}, V_z) = \frac{2}{9k^2}n^2 + o(n^2)$. Thirdly, the degree condition, the page number condition and the structure found so far imply that the degrees of almost all vertices in the graph of the edges joining V_x and V_z are $\frac{n}{3k} + o(n)$. This statement, the triangle condition and the fact that the total number of triangles of type $V_xV_yV_z$ is $o(n^3)$ imply that the number of triangles "sitting" on these $\frac{2}{9k^2}n^2 + o(n^2)$ edges is $\frac{n}{3k}(d(V_x, V_{xy}) + d(V_z, V_{yz})) + o(n^3) = \frac{n}{3k}(\frac{k-2}{9k^2}n^2 - c + \frac{k-2}{9k^2}n^2) + o(n^3)$, which is equal to $\frac{2(k-2)}{27k^3}n^3 + o(n^3)$ by the page number condition. Thus, $c = o(n^2)$, and now, the structure condition on the triangle xyz implies that $d(V_{xz}, V_z) = \frac{k-2}{9k^2}n^2 + o(n^2)$, a contradiction to the inequality obtained above, unless $k = 4$. (The case $k = 4$ is difficult again!)

From now on, we may assume that $k = 4$. In this case using the notation

$V'_y = V_y \cap N(v)$, $V''_y = V_y - N(v)$, $V'_{yz} = V_{yz} \cap N(v)$, $V''_{yz} = V_{yz} - N(v)$, we have seen that

$$|V'_y|, |V''_y|, |V'_{yz}|, |V''_{yz}| = n/12 + o(n),$$
$$d(V_x), d(V_z) = o(n^2), \quad d(V_y) = b \leqq n^2/144 + o(n^2),$$
$$d(V'_y, V_z) = n^2/72 + o(n^2), \quad d(V'_y, V_x) = o(n^2),$$
$$(V''_y, V_x) = n^2/72 + o(n^2), \quad d(V''_y, V_z) = o(n^2),$$
$$d(w, V_z) = n/12 + o(n) \text{ for almost all } w \in V_x,$$
$$d(w, V_x) = n/12 + o(n) \text{ for almost all } w \in V_z,$$

and

$$d(V_{xz}, V'_{yz}) = n^2/72 + o(n^2).$$

The last equality and $d(V_{xz}, V_{yz}) = n^2/72 + o(n^2)$ imply that $d(V_{xz}, V''_{yz}) = o(n^2)$. Applying the sum condition in $N(y)$ on the vertices of V_{yz}, it follows that $d(V_z, V''_{yz}) = n^2/72 + o(n^2)$ and $d(V_z, V'_{yz}) = o(n^2)$. Since $d(w, V_x) = n/12 + o(n)$ for almost all $w \in V_z$, applying the triangle condition and the page number condition for the $V''_{yz}V_z$-edges, we obtain that $d(u, V_{xy}) = n/12 + o(n)$ for almost all $u \in V''_{yz}$. Let us fix a typical vertex $u \in V''_{yz}$ for a moment. Then, a typical edge from u to V_{xy} is contained in $n/6 + o(n)$ triangles with third vertices in V_z; so, $d(V_{xy} \cap N(u), V_x) = o(n^2)$ because of the page number condition, and $d(V_{xy} - N(u), V_x) = n^2/72 + o(n^2)$ because $d(V_x, V_{xy}) = n^2/72 + o(n^2)$. The same argument shows that almost all vertices $u \in V''_{yz}$ have essentially the same neighbours in V_{xy}, and so for the sets $V'_{xy} = V_{xy} \cap N(u)$ and $V''_{xy} = V_{xy} - N(u)$, we have

$$|V'_{xy}|, |V''_{xy}| = n/12 + o(n),$$
$$d(V''_{xy}, V_x) = n^2/72 + o(n^2), \quad d(V'_{xy}, V_x) = o(n^2),$$
$$d(V'_{xy}, V''_{yz}) = n^2/144 + o(n^2), \quad d(V''_{xy}, V''_{yz}) = o(n^2).$$

Applying the sum condition to the vertices in V_{xy}, it follows that $d(V'_{xy}, V_{xz}) = n^2/72 + o(n^2)$, and $d(V''_{xy}, V_{xz}) = o(n^2)$.

Note that $d(V'_{yz}, V''_y) = o(n^2)$, since if not, then a typical $V'_{yz}V''_y$-edge is contained in $n/3 + o(n)$ triangles (with third vertices in V_x or V_{xz}), a contradiction to the page number condition. Similarly, $d(V'_{xy}, V'_y) = o(n^2)$.

Note that $d(V''_{yz}, V'_y) = o(n^2)$, since if not, then a typical $V''_{yz}V'_y$-edge is contained in $n/6 + o(n)$ triangles (with third vertices in V_z), a contradiction to the triangle condition if we have many edges of this kind. Similarly, $d(V'_{xy}, V'_y) = o(n^2)$.

Furthermore, $d(V'_{yz}, V'_y) = o(n^2)$ because of the K_4-freeness. This contradicts $d(V_{yz}, V_y) = n^2/72 - b + o(n^2)$, unless $b = n^2/144 + o(n^2)$, and $d(V''_{yz}, V''_y) =$

$n^2/144 + o(n^2)$. But then, it follows that $d(V'_{yz}, V_{xy}) = n^2/72 + o(n^2)$ because $d(V_{yz}, V_{xy}) = n^2/72 + b + o(n^2)$. Also, $d(V''_{xy}, V'_y) = n^2/144 + o(n^2)$, and $d(V'_{xy}, V''_y) = o(n^2)$, because $d(V_{xy}, V_y) = n^2/72 - b + o(n^2)$ and the degree condition.

Using $d(V'_{xy}, V_{xz}) = n^2/72 + o(n^2)$, $d(V''_{xy}, V_{xz}) = o(n^2)$, $d(V_{xz}, V'_{yz}) = n^2/72 + o(n^2)$, $d(V_{xz}, V''_{yz}) = o(n^2)$ and the sum condition in $N(x)$ and $N(z)$ for the vertices in V_{xz}, we obtain that $d(t, V_z), d(t, V_x) = n/12 + o(n)$ for almost all vertices $t \in V_{xz}$. The number of $V_{xz}V_xV_z$-triangles is $o(n^3)$ (triangle condition), so, it is possible only if we can define some sets $V'_{xz}, V''_{xz} = V_{xz} - V'_{xz}, V'_x, V''_x = V_x - V'_x, V'_z, V''_z = V_z - V'_z$ such that

$$|V'_{xz}|, |V''_{xz}|, |V'_x|, |V''_x|, |V'_z|, |V''_z| = n/12 + o(n),$$
$$d(V'_{xz}, V''_x), d(V'_{xz}, V''_z), d(V''_{xz}, V'_x), d(V''_{xz}, V'_z), d(V'_x, V''_z), d(V''_x, V'_z) = n^2/144 + o(n^2),$$

and

$$d(V'_{xz}, V'_x), d(V'_{xz}, V'_z), d(V''_{xz}, V''_x), d(V''_{xz}, V''_z), d(V'_x, V'_z), d(V''_x, V''_z) = o(n^2).$$

But then taking

$$V_{11} = V'_{xy}, V_{12} = V'_x, V_{13} = V'_y, V_{14} = V''_x, V_{21} = V''_{xy}, V_{22} = V'_{xz},$$
$$V_{23} = V''_{yz}, V_{24} = V''_{xz}, V_{31} = V''_y, V_{32} = V'_z, V_{33} = V'_{yz}, V_{34} = V''_z,$$

we obtain the structure described in Theorem 2 with $c_0 = 0$.

We obtained either a contradiction or the desired structure in all cases, so the proof of the Claim is complete. □

CLAIM 5. *The number of vertices $v \in V_x$ with $d(v, V_z) = \frac{2}{3k}n + o(n)$ or $d(v, V_y) = \frac{2}{3k}n + o(n)$ is $o(n)$. Similar statements hold for the vertices $v \in V_y$ and $v \in V_z$.*

PROOF. Because of symmetry reasons, it is sufficient to prove that the number of vertices $v \in V_x$ with $d(v, V_z) = \frac{2}{3k}n + o(n)$ is $o(n)$. Suppose that the number of such vertices v is of order of magnitude n. Applying the sum condition in $N(x)$ for v and that $d(v, V_z) = \frac{2}{3k}n + o(n)$, we have $d(v, V_y) = o(n)$ for every such a v. The triangle condition implies that an edge vw for a typical vertex v is contained in $d(w, V_{yz}) + d(v, V_{xy}) + o(n)$ triangles, which is equal to $\frac{k-2}{3k}n + o(n)$ by the page number condition. It implies that the fluctuation of $d(w, V_{yz})$ is $o(n)$ for the typical vertices $w \in V_z$, and since their average is $\frac{k-2}{6k}n - 3ck/2n + o(n)$, $d(w, V_{yz}) = \frac{k-2}{6k}n - 3ck/2n + o(n)$ for almost all vertices $w \in V_z$ and $d(v, V_{xy}) = \frac{k-2}{6k}n + 3ck/2n + o(n)$ for the considered vertices $v \in V_x$. Let $u \in V_{xy} \cap N(v)$ be a typical vertex. The page number condition on uv and the triangle condition imply that $d(u, V_{yz}) = \frac{k-4}{3k}n + o(n)$,

and so, the sum condition in $N(y)$ for u implies that $d(u,V_y)=\frac{2}{3k}n+o(n)$, a contradiction to Claim 4. □

CLAIM 6. *For a vertex $v\in V_{xy}$, $d(v,V_x)=o(n)$ holds if and only if $d(v,V_y)=o(n)$. The similar statements are true for the vertices in V_{xz} and V_{yz}.*

PROOF. Because of symmetry reasons, it is sufficient to prove only one direction. Suppose that $d(v,V_x)=o(n)$. The structure condition on yv implies that $V_y\cap N(v)$ and $V_x-N(v)$ are joined to each other almost completely. Applying Claim 5, it follows $d(v,V_y)=o(n)$. □

Claims 4 and 6 say that for almost every vertex $v\in V_{xy}$ (with exception of at most $o(n)$ ones), either $d(v,V_{xz}), d(v,V_{yz})=\frac{k-2}{3k}n+o(n)$, and $d(v,V_x), d(v,V_y)=o(n)$ — let V^*_{xy} denote the set of these vertices $v\in V_{xy}$ — or $d(v,V_{xz}), d(v,V_{yz})\neq\frac{k-2}{3k}n+o(n)$, $d(v,V_x), d(v,V_y)\neq o(n)$, and $d(v,V_x)$, $d(v,V_y)=\frac{2}{3k}n+o(n)$ — let V'_{xy} denote the set of these vertices $v\in V_{xy}$. Similar statements and definitions apply for the vertices in V_{xz} and V_{yz}.

CLAIM 7. *With exception of $o(n)$ vertices, either $d(v,V_x)=\frac{n}{3k}+o(n)$ or $d(v,V_y)=\frac{n}{3k}+o(n)$ for every vertex $v\in V'_{xy}$. Similar statements hold for the vertices in V'_{xz} and V'_{yz}.*

PROOF. We prove the statement by contradiction. The structure condition on vx and vy, resp., says that $V_x\cap N(v)$ is joined to $V_y-N(v)$ and $V_y\cap N(v)$ is joined to $V_x-N(v)$ almost completely. Furthermore $V_{xz}\cap N(v)$ is joined to $V_{yz}-N(v)$ and $V_{yz}\cap N(v)$ is joined to $V_{xz}-N(v)$ almost completely, respectively. The first half of the statement implies that if either $d(v,V_x)>\frac{n}{3k}+o(n)$ and $d(v,V_y)<\frac{n}{3k}+o(n)$, or $d(v,V_x)<\frac{n}{3k}+o(n)$ and $d(v,V_y)>\frac{n}{3k}+o(n)$, then $d(V_x,V_y)>\frac{2}{9k^2}n^2+o(n^2)$, a contradiction.

The structure condition on vx and vy, resp., also says that the number of edges joining $V_x\cap N(v)$ to $V_y\cap N(v)$ and $V_{xz}\cap N(v)$ to $V_{yz}\cap N(v)$ is $o(n^2)$.

We distinguish two cases according to the value of $d(v,V_x)$ and $d(v,V_y)$.

Case 1. $p=d(v,V_x)>\frac{n}{3k}+o(n)$, $\quad q=d(v,V_y)>\frac{n}{3k}+o(n)$.

First, suppose that $d(V_{xz},V_x\cap N(v))=o(n^2)$. A typical edge wu connecting V_{xz} to $V_x-N(v)$ is contained in at least $p+o(n)$ triangles such that the third vertex is in V_y; so, the page number condition implies that $d(w,V_{yz})\leqq\frac{k-2}{3k}n-p+o(n)$. Now, the sum condition in $N(z)$ for w implies that $d(w,V_z)\geqq p+o(n)>\frac{n}{3k}+o(n)$. On the other hand, $d(w,V_x)\leqq\frac{2}{3k}n-q+$ $+o(n)<\frac{n}{3k}+o(n)$ because of the starting assumption, a contradiction that $d(V_x,V_z)$ is too large, just as in case of v.

Secondly, suppose that the order of magnitude of $d(V_{xz},V_x\cap N(v))$ is n^2. Let us take a typical edge wu connecting V_{xz} to $V_x\cap N(v)$. As we have seen above, then $d(u,V_y)=\frac{2}{3k}n-p+o(n)$, and so $d(u,V_x)=p+o(n)$, because of the sum condition. Applying the structure condition for the triangle

wxz, we get that there is no edge connecting $V_x \cap N(w)$ to $V_z \cap N(w)$, i.e., $d(w, V_z) = p + o(n)$. Thus, for any vertex $s \in V_x$, especially for any vertex $s \in V_x - N(v)$, $d(s, V_z) \geqq \frac{2}{3k}n - p + o(n)$, and for any vertex $s \in V_x \cap N(v)$, $p + o(n)$ edges lead from s to V_z because of the sum condition and since there is no edge from s to $V_y \cap N(v)$. But then $d(V_x, V_z) \geqq (\frac{2}{3k}n - q)(\frac{2}{3k}n - p) + + pq + o(n^2) > \frac{2}{9k^2}n^2 + o(n^2)$, a contradiction.

Case 2. $p = d(v, V_x) < \frac{n}{3k} + o(n), \quad q = d(v, V_y) < \frac{n}{3k} + o(n)$.

Let us take the neighbours of the vertices $w \in V'_{xz}$ in V_x and V_z. The structure condition on wx and wz, says that $N(w) \cap V_x$ is joined to $V_z - N(w)$ and $N(w) \cap V_z$ is joined to $V_x - N(w)$ almost completely. For a typical vertex $w \in V'_{xz}$, the triangle condition says that the number of edges from $N(w) \cap V_x$ to $N(w) \cap V_z$ is $o(n^2)$. However, it follows that — with exception of $o(n)$ vertices — the sets $N(w) \cap V_x$ for the vertices $w \in V'_{xz}$ are either equal or disjoint to each other with an error of $o(n)$ elements, (i.e., we can define pairwise disjoint sets $V^i_{xz} \subseteqq V'_{xz}, V^i_x \subseteqq V'_x, V^i_z \subseteqq V'_z \quad (i = 1, \ldots, j)$ such that V^i_{xz} is joined to V^i_x and V^i_{xz} is joined to V^i_z almost completely, $d(V^i_{xz}, V_x - V^i_x), d(V^i_{xz}, V_z - V^i_z) = o(n^2)$, the sets V^i_{xz} cover V'_{xz} except $o(n)$ elements, and $j = O(1)$ because of the sizes of the pairwise disjoint sets V^i_x). We distinguish two cases.

Case 2.1. $j > 1$.

Let us take a typical edge wu from V^1_{xz} to V^1_x and apply the structure condition on it. We obtain that $N(w) \cap V_{yz}$ is joined to $V_{xz} - V^1_{xz}$ almost completely, and so, $N(w) \cap V_{yz} \subseteqq N(t)$ for a typical vertex $t \in V_{xz} - V^1_{xz}$. On the other hand, taking a typical edge ts and using that $j > 1$, we obtain that $N(t) \cap V_{yz} \subseteqq N(w)$, i.e., two typical vertices in V^1_{xz} are joined to the same elements of V_{yz}. Furthermore $j > 1$ implies that almost all elements of V'_{xz} are joined to the same elements of V_{yz}, i.e., to the elements of V^*_{yz} by definition. By definition, $V_{xz} - N(v) \subseteqq V'_{xz}$ and $V_{yz} \cap N(v)$ is joined to $V_{xz} - N(v)$ almost completely for a typical vertex v; so, $V_{yz} \cap N(v) \subseteqq V^*_{yz}$. It follows that $V_{yz} \cap N(v) \supseteqq V^*_{yz}$, because $V_{yz} \cap N(v) = V^*_{yz}$ (definition) with an error $o(n)$ and $V_{xz} \cap N(v) = V^*_{xz}$ holds, as well. But then $d(V_{xz}, V_{yz}) \geqq \geqq \frac{(k-2)^2}{9k^2}n^2 - pq + o(n^2) > \frac{(k-2)^2-1}{9k^2}n^2 + o(n^2)$, a contradiction to $d(V_{xz}, V_{yz}) = = \frac{(k-3)(k-2)}{9k^2}n^2 + c + o(n^2) \leqq \frac{(k-3)(k-2)+1}{9k^2}n^2 + o(n^2)$.

Similar argument can be applied if Case 2.1 holds for V_{yz}.

Case 2.2. $j = 1$ and we may assume that $V^1_{xz} = V'_{xz}$ and the analogouos equality holds for V_{yz}, as well.

Then, there are sets $V'_x \subseteqq V_x, V'_y \subseteqq V_y$ such that V'_x is joined to V'_{xz} and V'_y is joined to V'_{yz} almost completely. Furthermore the number of additional edges from V_x to V_{xz} and from V_y to V_{yz} is $o(n^2)$. Note that $d(V_x \cap N(v), V_{xz} \cap \cap N(v)) = o(n^2)$, because of the triangle condition (v is typical!). On the other hand, $d(V_x \cap N(v), V_{xz} - N(v)) = o(n^2)$ because, if not, then a typical edge

from $V_x \cap N(v)$ to $V_{xz} - N(v)$ is contained in at least $\frac{2}{3k}n - p + o(n)$ triangles with their third vertex in V_z and in at least $\frac{k-2}{3k}n - p + o(n)$ triangles with their third vertices in V_{yz}, a contradiction to the page number condition. Thus, $V_x' \subseteq V_x - N(v)$ and similarly $V_y' \subseteq V_y - N(v)$.

For a typical vertex $v \in V_{xy}$, $V_{xz}^* \subseteq V_{xz} \cap N(v)$ and $V_{xz} - N(v) \subseteq V_{xz}'$ (with a possible error $o(n)$). Let us take a typical edge wu from $V_{xz} - N(v)$ to V_x'. It is contained in $p + o(n)$ triangles with their third vertices in $V_y \cap N(v)$ and in $\frac{k-2}{3k}n - a + o(n)$ triangles with their third vertices in $V_{yz} \cap N(v)$; so, only $o(n)$ edges lead from a typical vertex $u \in V_x'$ to $V_y - N(v)$. Hence, $d(V_x', V_y - N(v)) = o(n^2)$ and a similar argument shows that $d(V_y', V_x - N(v)) = o(n^2)$. Suppose that $d(V_x - N(v) - V_x', V_y - N(v) - V_y') \neq o(n^2)$. Then a typical edge wu from $V_x - N(v) - V_x'$ to $V_y - N(v) - V_y'$ is contained in $o(n)$ triangles with their third vertices either in V_{xz} or in V_{yz} and in at most $\frac{2}{3k} - q + o(n)$ triangles with their third vertices in V_z, because $d(w, V_y) \geq q + o(n)$ and the sum condition. However, a typical edge is contained in $o(n)$ triangles with their third vertices in some other sets because of the triangle condition, so, we obtained a contradiction to the page number condition. Thus, $d(V_x - N(v), V_y - N(v)) = o(n^2)$ and $d(V_x \cap N(v), V_y \cap N(v)) = o(n^2)$ because of the triangle condition. Hence, $d(V_x, V_y) = p(\frac{2}{3k}n - q) + q(\frac{2}{3k}n - p) + o(n^2) < \frac{2}{9k^2}n^2 + o(n^2)$, a contradiction. The proof of Claim 7 is complete. □

Now, we return to the proof of the Theorems again.

So, there is a vertex $v \in V_{xy}'$ such that either $d(v, V_x) = \frac{n}{3k} + o(n)$ or $d(v, V_y) = \frac{n}{3k} + o(n)$, and, as we have seen, $X_1^z = V_x \cap N(v)$ is joined to $Y_2^z = V_y - N(v)$ and $Y_1^z = V_y \cap N(v)$ is joined to $X_2^z = V_x - N(v)$ almost completely. But $d(V_x, V_y) = \frac{2}{9k^2}n^2 + o(n^2)$ implies that the number of additional edges from V_x to V_y is $o(n^2)$. It follows that, for any other vertex $w \in V_{xy}'$, either $N(w) \cap V_x = X_1^z$ and $N(w) \cap V_y = Y_1^z$ or $N(w) \cap V_x = X_2^z$ and $N(w) \cap V_y = Y_2^z$ hold with a possible error of $o(n)$ vertices. Let Z_1 and Z_2 denote the vertices $w \in V_{xy}'$ of first and second type, respectively. Similarly, we can define the sets $X_1, X_2, Y_1^x, Y_2^x, Z_1^x, Z_2^x, Y_1, Y_2, X_1^y, X_2^y, Z_1^y, Z_2^y$, as well.

Case 1. The order of one of the 12 subsets defined in the sets V_x, V_y, V_z is not $\frac{n}{3k} + o(n)$, say, $|Z_1^y| = p < \frac{n}{3k} + o(n)$ and $|Z_2^y| = \frac{2}{3k}n - p + o(n) > \frac{n}{3k} + o(n)$.

Then, by Claim 7, $|X_1^y|, |X_2^y| = \frac{n}{3k} + o(n)$, and the sum condition implies that $|Z_1^x|, |Z_2^x|, |Y_1^x|, |Y_2^x| = \frac{n}{3k} + o(n)$ and that $d(u, V_y) = \frac{2}{3k}n - p + o(n)$ for almost all vertices $u \in X_2^y$ and $d(u, V_y) = p + o(n)$ for almost all vertices $u \in X_1^y$. But then, $|Y_2^z| = p + o(n)$, $|Y_1^z| = \frac{2}{3k}n - p + o(n)$, and $X_1^z = X_1^y$, $X_2^z = X_2^y$ (last two with a possible error $o(n)$).

Now, $d(V_{xz}, V_x) = \frac{k-2}{9k^2}n^2 - a + o(n^2) = \frac{n}{3k}(|Y_1| + |Y_2|) + o(n^2)$, and so, $|Y_1| + |Y_2| = \frac{k-2}{3k}n - \frac{3ka}{n} + o(n)$ and $|V_{xz}^*| = \frac{3ka}{n} + o(n)$. Similar argument shows that $|V_{xy}^*| = \frac{3ka}{n} + o(n)$ and that $|V_{yz}^*| = \frac{3kb}{n} + o(n) = \frac{3kc}{n} + o(n)$, which implies

$b - c = o(n^2)$, as well. All these imply that $|V^*_{xz}|, |V^*_{xy}|, |V^*_{yz}| \leq \frac{n}{3k} + o(n)$.

Now, we show that $|Y_1| = o(n)$. ($|Z_2| = o(n)$ can be shown similarly.) Suppose not and take a typical $Y_1 Z_1^y$-edge uw. The sum condition implies $d(u, V_{yz}) = \frac{k-2}{3k} n - p + o(n)$ and the triangle condition implies $d(w, V_{yz}) \leqq \leqq p + o(n)$. So if, say, Z_1^x, $|Z_1^x \cap Z_1^y| \neq o(n)$, then $|X_1| \leqq p + o(n)$, and a typical vertex $t \in Z_1^x - Z_1^y$ is not joined to the vertices in $V_z - Z_2^x$. Thus, $d(t, V_{xz}), |Y_2| \geqq \frac{k-3}{3k} n - p + o(n) \neq o(n)$, because of the page number condition. But then, as we have seen in Case 2.1 of the proof of Claim 7, there is no edge from V'_{xz} to V'_{yz}, contradicting that $\frac{k-2}{3k} n - p + o(n)$ edges connect a typical vertex of Y_1 to V_{yz}.

Now, we show that $d(Y_2, V'_{yz}) = o(n^2)$. ($d(Z_1, V'_{yz}) = o(n^2)$ can be shown similarly.) Suppose not. Then $d(u, V_z) = \frac{2}{3k} n - p + o(n)$ and $d(w, V_z) = \frac{n}{3k} + + o(n)$ for a typical edge uw from Y_2 to V'_{yz}; so, this edge is contained in $\frac{n}{3k} - p + o(n)$ triangles with their third vertices in V_z, a contradiction to the triangle condition.

These facts imply that $d(u, V_{yz}) = |V^*_{yz}| + o(n)$ for a typical vertex $u \in Y_2$. On the other hand, $d(u, V_{yz}) = \frac{k-4}{3k} n + p + o(n)$ because of the sum condition, which is a contradiction unless $k = 4$ and $|V^*_{yz}| = p + o(n)$.

Now, let us take typical vertices $u \in Z_1^y$ and $v \in N(u) \cap Y_1^z$. The triangle condition implies that the third vertex of almost all triangles containing uv is in V_x, V_{xy} or V_{xz}. The number of the triangles of these types is $n/12 + + o(n)$, $|V'_{xy}|$ and $o(n)$, respectively. But, then $|V'_{xy}|, |V^*_{xy}| = n/12 + o(n)$. Similarly, $|V'_{xz}|, |V^*_{xz}| = n/12 + o(n)$ and by the way, $a = n^2/144 = o(n^2)$. Thus $d(X_1^y), d(X_2^y) = o(n^2)$ because of the triangle condition; so, it follows that X_1^y is joined to X_2^y almost completely. Also $d(V_{xy}, V_{xz}) = 3n^2/144 + o(n^2)$, and the structure proved so far imply that $d(V'_{xz}, V'_{xy}) = o(n^2)$.

Note that $d(Z_1^y, Y_2^z) = o(n^2)$, since if not true, then a typical edge of this type is contained in $o(n)$ triangles with their third vertices in V_z, V_y or V_{yz}. The structure proved so far says that $o(n)$ triangles have their third vertex in the other sets, a contradiction to the page number condition. However, it follows that one of the sets Z_1^x and Z_2^x meets Z_1^y in $o(n)$ elements, say, $|Z_1^x \cap Z_1^y| = o(n)$, and then $|Y_1^x \cap Y_2^z| = o(n)$ holds, as well. Because of $p = = 12c/n + o(n)$, $d(V_z), d(V_y) = pn/12 + o(n^2)$. But, the triangle condition says that the number of edges in one of the defined sets is $o(n^2)$, so, it follows that Z_1^x is joined to Z_1^y and Y_1^x is joined to Y_2^z almost completely. Finally, the degree condition for the vertices in Z_1^x implies that $|X_1| = n/12 - p + o(n)$, and so $|X_2| = n/12 + o(n)$. The structure of G is described; it is exactly the structure defined in Theorem 2 with $c_0 = p$.

Case 2. Each of the 12 subsets defined in the sets V_x, V_y, V_z has $\frac{n}{3k} + o(n)$ elements.

Like in Case 1, we can determine the orders of the sets V^*_{xy}, V^*_{xz} and

V^*_{yz}. Now, we get stronger symmetry than $b-c=o(n^2)$, we obtain that $|V^*_{xy}|, |V^*_{xz}|, |V^*_{yz}| = \frac{3ka}{n} + o(n)$.

Case 2.1. $k > 4$.

Suppose that $|Y_1|, |Y_2| \neq o(n)$. Then, as we have seen in Case 2.1 of the proof of Claim 7, there is no edge from V'_{xz} to V'_{yz}, contradicting that $\frac{k-3}{3k}n + o(n)$ edges connect a typical vertex of V'_{xz} to V_{yz}. Thus, without loss of generality, we may assume that $|Y_2| = o(n)$, and similarly $|X_2|, |Z_2| = o(n)$. The number of edges from X_1 to Y_1 is not $o(n)$, a typical edge of this type is contained in $o(n)$ triangles with their third vertices in V_z because of the triangle condition, so $|Z_1^x \cap Z_1^y| = o(n)$. Similarly, $|X_1^z \cap X_1^y|, |Y_1^x \cap Y_1^z| = o(n)$, (i.e., the sets V_x, V_y, V_z are cut into two parts by the two partitions essentially in the same way).

A typical edge uv from Z_1^y to X_1^z is contained in $\frac{k-1}{3k}n - \frac{3ka}{n} + o(n)$ triangles by the triangle condition and the structure proved so far. Thus, $|V^*_{xy}|, |V^*_{xz}|, |V^*_{yz}| = \frac{3ka}{n} + o(n) = \frac{n}{3k} + o(n)$ by the page number condition. Applying the degree condition for the vertices in V_x, V_y, V_z and that, say, X_1^y does not contain edges by the triangle condition, we obtain that X_1^y is joined to X_1^z, Y_1^x is joined to Y_1^z and Z_1^y is joined to Z_1^x almost completely.

Take a typical edge uv from V'_{xz} to V'_{yz}. The number of triangles containing uv with their third vertices in V_x, V_y or V^*_{xy} and in V'_{xy} is $\frac{3}{3k}n + o(n)$ and $\frac{k-5}{3k}n + o(n)$, respectively. Because of the sum condition, it is a contradiction, unless $N(u) \cap V'_{xy}$ and $N(v) \cap V'_{xy}$ are as disjoint as possible, i.e., their union covers V'_{xy} with a possible error of $o(n)$ vertices. Repeating it for the other neighbours of u in V'_{yz}, we obtain that $V'_{xy} - N(u)$ is joined to $V'_{yz} \cap N(u)$ almost completely and that $d(V'_{xy} - N(u), V'_{yz} - N(u)) = o(n^2)$. Now, repeating it for the vertices in $V'_{yz} - N(u)$ and for their neighbours in V'_{xy}, we obtain that there is a similar subset of V_{xz} with $\frac{n}{3k} + o(n)$ elements such that the number of edges connecting this set to $V'_{xy} - N(u)$ or $V'_{yz} - N(u)$ is $o(n^2)$. Continuing, we can define pairwise disjoint subsets $V^1_{xy}, \ldots, V^{k-3}_{xy}, V^1_{xz}, \ldots, V^{k-3}_{xz}$ and $V^1_{yz}, \ldots, V^{k-3}_{yz}$ of $\frac{n}{3k} + o(n)$ elements of the sets V'_{xy}, V'_{xz} and V'_{yz}, respectively, such that two subsets with different subscripts are joined to each other almost completely if their superscripts are different, and, as well are joined to each other with $o(n^2)$ edges if their superscripts are equal. It is easy to see that we obtained the graph described in Theorem 3.

Case 2.2. $k = 4$.

First, we show that $a = n^2/144 + o(n^2)$. Suppose not, i.e., that $a < n^2/144 + o(n^2)$. (Inequality in other direction cannot hold by Turán's theorem, since V_x is triangle free by Lemma 1.) Then $|V^*_{xy}|, |V^*_{xz}|, |V^*_{yz}| = 12a/n + o(n) < n/12 + o(n)$. For a typical vertex $v \in V'_{xy}$, $d(v, V_{xz}) = n/12 + o(n)$

because of the sum condition, and so, $d(v, V'_{xz}) > o(n)$ and $d(V'_{xy}, V'_{xz}) > > o(n^2)$. The number of triangles containing a typical edge uv from V'_{xy} to V'_{xz} with their third vertices in V_z, V_y and V_{yz} is $n/12 + o(n)$, $n/12 + o(n)$ and $|V^*_{yz}| = 12a/n + o(n)$, respectively, and so $|V^*_{xy}|, |V^*_{xz}|, |V^*_{yz}| = o(n)$ and $a = o(n^2)$ by the page number condition.

Now, let us take a typical edge uv from V'_{xy} to V'_{xz}, which leads, say, from Z_1 to Y_1. The triangle condition implies that the number of triangles containing uv with their third vertices in V_x is $o(n)$, so, $|X_1^y \cap X_1^z| = o(n)$. Thus, V_x is divided into two parts in the same way by the two partition, and the analogous statements hold for V_y and V_z, as well. On the other hand, we already have $n/6 + o(n)$ triangles containing a typical V_xV_y-edge with their third vertices in V_{xz} or V_{yz}, so the number of triangles containing a typical V_xV_y-edge with their third vertices in V_z is $o(n)$. Without loss of generality, we may assume that $X_1^y = X_1^z, X_2^z = X_2^y, Y_1^x = Y_1^z, Y_2^z = Y_2^x, Z_1^y = Z_1^x$, and $Z_2^x = Z_2^y$ with a possible error of $o(n)$ vertices. Also two sets of different capitals are joined to each other almost completely if their subscripts are different and are joined to each other with $o(n^2)$ edges if their subscripts are equal. Furthermore, the triangle condition implies that two sets from among $X_1, X_2, Y_1, Y_2, Z_1, Z_2$ are joined to each other with $o(n^2)$ edges if their subscripts are equal. But then, the number of triangles containing a typical $X_1X_1^y$-edge or $X_2X_2^y$-edge is $o(n)$, a contradiction.

Hence, $a = n^2/144 + o(n^2)$ and $|V^*_{xy}|, |V^*_{xz}|, |V^*_{yz}| = n/12 + o(n)$. For a typical vertex $u \in V'_{xy}$, $d(u, V'_{xz}) = o(n)$; so, $d(V'_{xy}, V'_{xz}) = o(n^2)$ and similarly, $d(V'_{xy}, V'_{yz})$, $d(V'_{xz}, V'_{yz}) = o(n^2)$.

For a typical $X_1^zY_2^z$-edge and a typical $Y_1^zX_2^z$-edge, the number of triangles containing one of these edges with their third vertices in V_{xz} or V_{yz} is only $|X_1| + |X_2| + |Y_1| + |Y_2| + o(n) = n/6 + o(n)$, and the number of triangles containing one of these edges with their third vertices in V_x, V_y or V_{xy} is $o(n)$. Thus, the number of triangles containing one of these edges with their third vertices in V_z is $n/6 + o(n)$.

But then, the six subsets of $n/12 + o(n)$ elements in $V_x \cup V_y \cup V_z$ can be divided into two families of three subsets such that any two subsets in the same family is joined to each other almost completely. Thus e.g., the number of triangles containing a typical $Y_1^zX_2^z$-edge with their third vertices in V_z is $n/12 + o(n)$, and with their third vertices in V_{xz} or V_{yz} is $|X_i| + |Y_j| + o(n)$ for some i and j. So, the orders of the sets X_1 and X_2 are equal to the orders of the sets Y_1 and Y_2 in some pairing with a possible error $o(n)$. But then, it is easily seen that the structure of G is the structure described in Theorem 2 with $c_0 = 12|X_1|/n$ allowing $c_0 = 0$ and $c_0 = 1$, as well.

The proof of Theorem 2 and the case $l = 3$ of Theorem 3 is complete.

□

PROOF OF CASE $l > 3$ OF THEOREM 3. The case $l = 3$ was proved in Lemma 1 above. We prove the general case by induction on l. Let G be a

graph satisfying the conditions of Theorem 3.

The condition $\frac{k-1}{k}\frac{l-1}{l} > \frac{l-2}{l-1}$ (i.e., $k > l^2-2l+1$) implies that the degree of every vertex is greater than $\frac{l-2}{l-1}n$, so every vertex and every edge is contained in a clique of l vertices.

Take an arbitrary clique $x_1x_2\ldots x_l$. Let G_1 denote the subgraph of G induced by $N(x_1)$ and let $d_2,\ldots,d_l$ denote the degrees of the vertices $x_2,\ldots,x_l$ in G_1, respectively. Every book in G has at most $\frac{k-2}{k}\frac{l-2}{l}n+o(n)$ and so, $d_i \leqq \frac{k-2}{k}\frac{l-2}{l}n+o(n) \quad (i=2,\ldots,l)$. Now, let us estimate the total number D of the triangles containing any edge x_ix_j $(2 \leqq i < j \leqq l)$.

The number of triangles containing x_ix_j with their third vertices not in $N(x_1)$ is at least $d(x_i)-d_i+d(x_j)-d_j-(n-d(x_1))$, so, the total number D_1 of triangles of this type is at least

$$(l-2)\sum_{i=2}^{l} d(x_i) - (l-2)\sum_{i=2}^{l} d_i - \binom{l-1}{2}(n-d(x_1)).$$

On the other hand, if $v \in N(x_1)$ is the neighbour of $r(v)$ of the vertices $x_2,\ldots,x_l$, then it is contained in $\binom{r(v)}{2}$ counted triangles. By adding up $\binom{r(v)}{2}$ for all vertices $v \in N(x_1)$, we get the number $D-D_1$ of the triangles containing x_ix_j with their third vertices in $N(x_1)$. However, the sum of the $r(v)$'s is $\sum_{i=2}^{l} d_i$, which is at most $\frac{(k-2)(l-2)(l-1)}{kl}n+o(n)$. Deleting the edge x_1v from G for a vertex $v \in N(x_1)$, the two estimates above (together) decrease by $\binom{l-1}{2}+\binom{r(v)}{2}-r(l-2)>0$. And if to increase d_i, we add an edge $x_iv \quad (i>1)$ to G for a vertex $v \in N(x_1)-N(x_i)$, then the two estimates above (together) decrease by $(l-2)-(r-1)$, which is positive if $r(v) \leqq l-2$. On the other hand, if $d(x_i)=\frac{k-1}{k}\frac{l-1}{l}n+o(n) \quad (i=1,\ldots,l)$ and $d_i=\frac{k-2}{k}\frac{l-2}{l}n+o(n) \quad (i=2,\ldots,l)$, then we have the weakest lower bound for D if the $r(v)$'s are as equal as possible, i.e., $r(v)=l-2$ for $\frac{(k-l+1)(l-1)}{kl}n+o(n)$ and $r(v)=l-3$ for $\frac{(l-2)(l-1)}{kl}n+o(n)$ vertices $v \in N(x_1)$. Then, we obtain that

$$\begin{aligned} D \geqq{}& (l-2)(l-1)\frac{(k-1)(l-1)}{kl}n-(l-2)(l-1)\frac{(k-2)(l-2)}{kl}n \\ &-\binom{l-1}{2}\left(n-\frac{(k-1)(l-1)}{kl}n\right)+\frac{(k-l+1)(l-1)}{kl}n\binom{l-2}{2} \\ &+\frac{(l-2)(l-1)}{kl}n\binom{l-3}{2}+o(n)=\binom{l-1}{2}\frac{(k-2)(l-2)}{kl}n+o(n). \end{aligned}$$

However, we assumed that every edge is contained in $\frac{(k-2)(l-2)}{kl}n+o(n)$ triangles, so, this estimate and all the estimates used are sharp apart from a

possible error $o(n)$. Thus, $d(x_i) = \frac{(k-1)(l-1)}{kl}n + o(n) \qquad (i = 1, 2, \ldots, l)$, and since every vertex is contained in a clique of l vertices, the degree of every vertex in G is $\frac{(k-1)(l-1)}{kl}n + o(n)$. Also, it was "sharp" when we assumed that the $r(v)$'s are as equal as possible. (Actually, we use only that the number of vertices v with $r(v) = l-1$ is $o(n)$.) Then, we have $d_2 = \frac{(k-2)(l-2)}{kl}n + o(n)$, since it is of order of magnitude n of course, and if not, then we could find constant times n nonneighbours of x_2 such that joining x_2 to these vertices, the estimate decreases by constant times n, but the estimate holds for the resulting degrees, as well, i.e., some edge $x_i x_j$ is contained in too many triangles.

The degree of every vertex in G_1 is $\frac{(k-2)(l-2)}{kl}n + o(n)$. On the other hand, the estimate for D_1 implies that every edge in G_1 is contained in at least $\frac{k+l-5}{kl}n + o(n)$ triangles such that the third vertex is not in $N(x_1)$. Thus, every book in G_1 has at most $\frac{(k-2)(l-2)}{kl}n - \frac{k+l-5}{kl}n + o(n) = \frac{(k-3)(l-3)}{kl}n + o(n)$ pages. But then, we can apply the induction hypothesis for G_1, i.e., G_1 is essentially the appropriate $(k-1) \times (l-1)$-partite graph. It holds for the neighbourhood of any vertex, since x_1 was chosen arbitrarily.

From here, it is routine (but not too short) to prove that G has the desired structure. □

REMARK. The condition $\frac{k-1}{k}\frac{l-1}{l} > \frac{l-2}{l-1}$ (i.e., $k > l^2 - 2l + 1$) in Theorem 3 cannot be eliminated, since then the proof above does not work, the conditions do not imply that the graph contains a clique of l vertices, and counterexamples can be constructed on the base of it. Consider the case $k = l^2 - 2l + 1$. Then the degree condition says that the degree of every vertex is at least $\frac{l-2}{l-1}n + o(n)$, allowing that G is the $(l-1)$-partite Turán graph in which case every edge is contained in $\frac{l-3}{l-1}n + o(n) < \frac{k-2}{k}\frac{l-2}{l}n + o(n)$ triangles. For other small k's, the counterexamples are not so nice, but can be constructed similarly.

REFERENCES

[1] BONDY, J.A., Pancyclic graphs, I, *J. Combinatorial Theory Ser. B* **11** (1971), 80–84. *MR* **44**#2642

[2] DIRAC, G.A., Some theorems on abstract graphs, *Proc. London Math. Soc.* (3) **2** (1952), 69–81. *MR* **13**-856

[3] EDWARDS, C.S., A lower bound for the largest number of triangles with a common edge (unpublished manuscript).

[4] ERDŐS, P. and FAUDREE, R., Size Ramsey functions, *Sets, graphs and numbers* (Proc. Colloq. dedicated to the 60th birthday of A. Hajnal and V. T. Sós, Budapest, 1991), Colloq. Math. Soc. J. Bolyai, Vol. 60, North-Holland, Amsterdam, 1992, 219–238.

(Received February 14, 1994)

P. Erdős and E. Győri

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

R. Faudree

DEPARTMENT OF MATHEMATICS
MEMPHIS STATE UNIVERSITY
MEMPHIS, TN 38152
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 47–57

DISCREPANCY OF TREES

P. ERDŐS, Z. FÜREDI, M. LOEBL and V. T. SÓS

Abstract

We consider the question how large monochromatic part of a tree is present in any coloring of edges of a complete graph by two colors. It is proved that there exists a constant $c > 0$ such that for any given tree T_n on n vertices with maximum degree Δ the following holds. An arbitrary coloring of the edges of K_n with 2 colors contains a copy of T_n such that at least $(n-1)/2 + c(n-1-\Delta)$ edges of T_n get the same color.

1. Introduction, results

Discrepancy theory has originated from number theory. In the last two decades this subject has developed into an elaborate theory related also to geometry, probability theory, ergodic theory, computer science, combinatorics. The combinatorial setting of these problems proved to be a succesful approach. See the book of Beck and Chen [2], the chapter from the Handbook of Combinatorics [3], or [8].

One of the basic problems in combinatorial discrepancy theory is the following: Let $S = \{x_1, x_2, \dots, x_t\}$ be a finite set and $\mathcal{H} = \{A_1, \dots, A_m\}$ be a family of subsets of S. The goal is

(*) to find a partition $S_1 \cup S_2 = S$, $S_1 \cap S_2 = \emptyset$ which splits each of the set in the family $\mathcal{H}$ as equally as possible.

A partition of S can be given by a function $\varphi : S \to \{1, 2\}$. The discrepancy of $\mathcal{H}$ is defined by

$$\mathcal{D}(\mathcal{H}) := \min_{\varphi} \max_{A \in \mathcal{H}} \left| |\varphi^{-1}(1) \cap A| - \frac{|A|}{2} \right|.$$

This measures, in supremum norm, how well the set S can be partitioned in the sense of (*).

For a given $(S, \mathcal{H})$ we want to determine or estimate $\mathcal{D}(\mathcal{H})$. A large number of classical theorems in number theory, in geometry, in combinatorics can be formulated in this language. Here we consider the special case when

1991 *Mathematics Subject Classification.* Primary 05C05, 05D10, 05C55.

Key words and phrases. Discrepancy, Ramsey theory, extremal graphs, edge coloring, trees.

0081-6906/95/$ 4.00

the underlying set S is the edge set, $E(K_n)$, of a complete graph and the family $\mathcal{H}$ is given by isomorphic copies of a given graph.

Let L be an arbitrary fixed graph. Our goal is to two-color the edges of K_n so that in each subgraph L^* isomorphic to L the edge-set $E(L^*)$ is two-colored as equally as possible. Let the two-coloring be given by $\varphi: E(K_n) \to \{1,2\}$. The *discrepancy* of L is defined by

$$\mathcal{D}_n(L,\varphi) := \max_{\substack{L^* \subseteq K_n \\ L^* \sim L}} \left| |\varphi^{-1}(1) \cap E(L^*)| - \frac{|E(L^*)|}{2} \right|,$$

$$\mathcal{D}_n(L) := \min_{\varphi} \mathcal{D}_n(L,\varphi).$$

While Ramsey theory asks how large n should be so that any two-coloring of edges of K_n contains a monochromatic copy of a given graph L, discrepancy measures how large part of a graph L is present in any two-coloring. The case $L = K_t$ was investigated by Erdős and Spencer [5].

In this paper we consider the case when L is a tree T_n on n vertices. Put $\mathcal{D}(T_n) = \mathcal{D}_n(T_n)$.

Let S_n and P_n denote the star and the path on n vertices, respectively. It is obvious, that

$$\mathcal{D}(S_n) = \begin{cases} 0 & \text{for } n = 4k+1\,, \\ 1/2 & \text{for } n = 2k\,, \\ 1 & \text{for } n = 4k+3\,. \end{cases} \tag{1}$$

It is also easy to see that

$$\mathcal{D}(P_n) = \frac{1}{6}n + O(1)\,.$$

This follows from a theorem of Gerencsér and Gyárfás [6] stating

$$R(P_k) = \lfloor (3k+1)/2 \rfloor\,, \tag{2}$$

where $R(L)$ denotes the *Ramsey number* of the graph L.

In general, $R(L_1, L_2)$ denotes the minimum integer n such that the following holds: for each coloring of the edge-set of $E(K_n)$ with the colors $\{1,2\}$ one can find either a copy of L_1 of color 1 or a copy of L_2 consisting of edges of color 2; finally $R(L) := R(L,L)$.

Which are the basic relevant properties of T_n determining whether $\mathcal{D}(T_n)$ is small or large?

Let $\Delta(L)$ denote the maximal degree in L. A set $C \subseteq V(L)$ is called a *vertex cover* if each edge $e \in E(L)$ has at least one endpoint in C. Let $\tau(L_n)$ denote the minimum size of a vertex cover.

Here we prove that the order of magnitude of $\mathcal{D}(T_n)$ depends on $\Delta(T_n)$ and $\tau(T_n)$.

THEOREM 1.1. *Suppose that* $\Delta(T_n) \geqq 0.8n$. *Then* $\mathcal{D}(T_n) \geqq (n-1-\Delta)/6$.

For even n considering a two-coloring of $E(K_n)$ such that every color induces an $n/2$-regular graph, one sees that $\mathcal{D}(T_n) \leqq n-1-\Delta$.

THEOREM 1.2. *Suppose* $n > m_0$ *and* $\Delta(T_n) < 0.8n$. *Then* $\mathcal{D}(T_n) > > n10^{-3}$.

Here the value of m_0 comes from Corollary 2.8.

The next theorem describes a class of trees having discrepancies as large as possible, $n/2 - o(n)$ (if $\max(\Delta(T_n), \tau(T_n)) = o(n)$).

THEOREM 1.3. *If* $\Delta(T_n), \tau(T_n) \leqq k \leqq n/8$, *then* $\mathcal{D}(T_n) \geqq (n/2) - 4k$.

Color red a complete subgraph of size $n-(k/2)$ and blue the rest of the edges of K_n. Then the largest monochromatic part of a tree with $\tau(T) = k$ does not have more than $n-(k/2)$ edges. Hence $\mathcal{D}(T_n) \leqq n/2 - k/2$.

2. Conjectures, problems, lemmata

The proofs of the theorems above are closely related to extremal and Ramsey problems on trees. Here a new type of extremal problems arose, where the lower bound on the number of edges (in Turán type problems) is replaced by a lower bound on the number of vertices with high degrees.

CONJECTURE 2.1 ($n/2$–$n/2$–$n/2$ conjecture). *Let G be a graph with n vertices and let at least $n/2$ of them have degree at least $n/2$. Then G contains any tree on at most $n/2$ vertices.*

M. Ajtai, J. Komlós and E. Szemerédi [1] proved the following approximate version.

THEOREM 2.2 (Ajtai, Komlós and Szemerédi [1]). *For every $\eta > 0$ there is a threshold $n_0 = n_0(\eta)$ such that the following statement holds for all $n \geqq \geqq n_0$: if G is a graph on n vertices, and at least $(1+\eta)\frac{n}{2}$ vertices have degrees at least $(1+\eta)\frac{n}{2}$, then G contains, as subgraphs, all trees with at most $\frac{n}{2}$ edges.*

J. Komlós and V. T. Sós extended Conjecture 2.1 for trees of any size.

CONJECTURE 2.3. *If G is a graph on n vertices and more than $n/2$ vertices have degrees greater than or equal to k, then G contains, as subgraphs, all trees with k edges.*

J. Komlós announced proving an approximate version of Conjecture 2.3, too.

THEOREM 2.4 (Komlós [7]). *For every $\eta > 0$ there is a threshold $n_0 = n_0(\eta)$ such that the following statement holds for all $n \geq n_0$: if G is a graph on n vertices and at least $(1+\eta)\frac{n}{2}$ vertices have degrees at least $(1+\eta)k$ then G contains all trees with at most k edges.*

A weaker form of Theorem 2.4 which we will need follows analogously to the proof of Theorem 2.2 [1].

THEOREM 2.5. *For every $\eta > 0$ there is a threshold $n_0 = n_0(\eta)$ such that the following statement holds for all $\varepsilon \geq 0$ and $n \geq n_0$: if G is a graph on n vertices and at least $(1+\eta)\frac{n}{2}$ vertices have degrees at least $(1-\varepsilon+\eta)\frac{n}{2}$ then G contains all trees with at most $(1-3\varepsilon)\frac{n}{2}$ edges.*

SKETCH OF PROOF OF THEOREM 2.5. The proof goes in the same way as the proof of Theorem 2.2 in [1], with only one change: a combinatorial Lemma 6 of [1] is replaced by a lemma below proving a weaker property (from weaker assumptions) than the original Lemma 6.

LEMMA 2.6. *Let H be a graph on N vertices, and let U be the set of vertices of degree greater than $(1-\eta)\frac{N}{2}$. If $|U| \geq \frac{N}{2}+1$ then there are two vertices x, $y \in U$ and a (partial) matching M in H such that*

x and y are adjacent,

M covers at least $(1-3\eta)\frac{N}{2}-1$ neighbors of both x and y.

PROOF OF LEMMA 2.6. First observe that at least two vertices of U are joined by an edge of H. We will use the Gallai–Edmonds decomposition (GED). Let A be the set of vertices of H omitted by at least one maximum matching of H, let B the set of vertices of $H-A$ which have neighbors in A and let C be the set of remaining vertices of H. GED Theorem asserts that the connectivity components of $H-A$ are hypomatchable (a graph G is called hypomatchable if $G-v$ has a perfect matching for each vertex v of G), the connectivity components of $H-C$ have a perfect matching and any maximum matching of H covers completely B from A.

If a component of $H-B$ has two adjacent vertices of U then Lemma 2.6 follows. Hence U forms an independent set in each component of $H-B$. Let α denote the size of a maximum independent set. However, $\alpha(C) < \frac{|C|}{2}$ for any hypomatchable C with more than one vertex and $\alpha(C) \leq \frac{|C|}{2}$ for any C with a perfect matching. Since $|U| > \frac{|V(H)|}{2}$, there is a hypomatchable component C of $H-B$ consisting of exactly one vertex which moreover belongs to U. Hence $|B| \geq (1-\eta)\frac{n}{2}$ and by GED Theorem H has a matching which covers at least $n-\eta n$ vertices.

Hence the Lemma 2.6 is proved and Theorem 2.5 then follows analogously as Theorem 2.2 in [1]. □

Using Theorem 2.5, it is not difficult to prove a Ramsey type result, which will provide a basic tool in further considerations.

THEOREM 2.7. *For every $\varepsilon > 0$ there is a threshold $m_0 = m_0(\varepsilon)$ such that the following statement holds for all $n \geqq m_0$: if G is a graph on n vertices and T is a tree on at most $(1-\varepsilon)\frac{n}{2}$ vertices, then G or complement of G contains T.*

PROOF. Let $m_0 = \left(1 - \frac{2}{9}\varepsilon\right)^{-1} n_0\left(\frac{\varepsilon}{9}\right)$, where $n_0(\cdot)$ comes from Theorem 2.5. Let $\eta = \frac{\varepsilon}{9}$ and $\varepsilon' = \frac{\varepsilon}{3}$. If G satisfies the assumptions of Theorem 2.5 for ε' and η then Theorem 2.7 follows, otherwise complement of G has at least $(1-\eta)\frac{n}{2}$ vertices of degree at least $n - (1-\varepsilon'+\eta)\frac{n}{2} = \frac{n}{2}(1+\varepsilon'-\eta)$. Denote by S the set of these vertices. Let G' be a graph obtained from $\overline{G}$ by deleting $2\eta n$ vertices from $V(\overline{G}) - S$. $|V(G')| = n' = (1-2\eta)n$. Now, at least $|S| \geqq (1-\eta)\frac{n}{2} \geqq (1+\eta)\frac{n'}{2}$ vertices of G' have degree at least $(1+\varepsilon'-5\eta)\frac{n}{2} \geqq$ $\geqq (1-\varepsilon'+\eta)\frac{n'}{2}$. Since $n' \geqq n_0(\eta)$ we may apply Theorem 2.5 to G' and get that G' and hence also $\overline{G}$ has all trees on $(1-\varepsilon)\frac{n}{2}$ vertices. □

We will use only the following weaker version.

COROLLARY 2.8. *For $n > m_0$ the following holds. Every tree on at most $(\frac{1}{2} - 10^{-3})n$ vertices is contained in either G_n or in $\overline{G}_n$.*

Theorem 2.7 states that $R(T_k) \leqq 2k + o(k)$ as $k \to \infty$. Here we formulate the

CONJECTURE 2.9. *Let T_a and T_b be trees on a and b vertices, respectively, and let G be a graph on $a+b-2$ vertices. Then either G contains T_a or $\overline{G}$ contains T_b. Especially, $R(T_k) \leqq 2k-2$.*

We think that even more is true.

CONJECTURE 2.10. *There is a $c > 0$ such that $R(T_k) < (2-c)k + c\Delta$.*

We conclude this section by an easy observation.

LEMMA 2.11. *Let M_a be a star-forest on $a \geqq 2$ vertices and consider an arbitrary two-coloring of the complete graph, $E(K_n) = E(G_1) \cup E(G_2)$. If G_1 does not contain a monochromatic copy of M_a then there is a subset $A \subseteq$ $\subseteq V(K_n)$ such that every vertex in A has more than $n-a$ G_2-neighbors in A. Consequently, $R(M_a, T_b) \leqq a+b-2$.*

PROOF. If M consists of only one star, then the statement is trivial with $A = V(K_n)$. Otherwise, one can use induction on the number of stars in M.

If the degree of each vertex of the subgraph of G_2 induced on A is at least $n-a+1$, then G_2 has every tree on $n-a+2$ vertices. □

3. The case of large maximum degree

In this section we prove Theorem 1.1. Consider an arbitrary two coloring, φ, of the edge-set of the complete graph using the colors $\{1,2\}$. Let T be

an n vertex tree with $\Delta(T) \geqq 0.8n$. Suppose, on the contrary, that $\mathcal{D}_\varphi(T) = =: x < (n-1-\Delta)/6$. Then $x < (n-1)/30$. Let S_1 be a monochromatic star of K_n of maximum number of vertices. Denote its vertex set by A_1, let $A_2 := V(K_n) - A_1$, and $|A_1| - 1 = (n-1)/2 + m$. Here $m \leqq x$. We may suppose that the edges of S are colored by the color 1.

Let M be the forest having $(n-1-\Delta)$ edges obtained from T by deleting the edges adjacent to a vertex of maximum degree. M has a subforest consisting of vertex disjoint stars and containing at least half of its edges. Let M_1, M_2 be star-forests contained in M of sizes $|E(M_1)| = x - m + 1$ and $|E(M_2)| = 3x - m + 1$. As the vertex of maximum degree of T is adjacent to at least $0.6(n-1)$ vertices of degree 1, T contains a vertex-disjoint copy of a star T_i and the star-forest M_i such that their total number of edges is $(n-1)/2 + x + 1$. (This is, indeed, a special case of Lemma 4.1.)

There is no monochromatic copy of M_1 in A_2 in color 1, otherwise together with S_1 it would form a too large monochromatic part of a copy of T. Hence Lemma 2.11 implies that there exists an $A_2' \subseteq A_2$ such that every degree in color 2 in A_2' is at least $|A_2| - 2(x - m + 1) + 1$. As the maximum degree in color 2 is at most $|A_1| - 1$ we obtain that every vertex of A_2' is joined to at most $2x + 1$ vertices from A_1 using edges of color 2.

We also obtain that there is a star S_2 of at least $|A_2| - 2(x - m) + 1$ edges of color 2 contained in A_2. Thus, repeating the previous argument, A_1 does not contain a copy of M_2 of color 2. Hence Lemma 2.11 implies that there exists an $A_1' \subseteq A_1$ such that every vertex in A_1' has degree in color 1 at least $|A_1| - 2(3x - m + 1) + 1$. We obtain that every vertex of A_1' is joined to at most $(6x - 2m)$ vertices of A_2' using edges of color 1.

Altogether, considering the complete bipartite graph with parts A_1' and A_2' we get that

$$2x + 1 + (6x - 2m) \geqq \min(|A_1'|, |A_2'|) \geqq (n-1)/2 + 1 + m - (6x - 2m + 1) + 1 .$$

This yields $x \geqq (n-1)/28$, a contradiction. □

4. How to cut a tree

In this section we collect some technical lemmata about tree decompositions we are going to use in the next section for the proof of our main result, Theorem 1.2. As we are providing an asymptotic only, for simplicity, from now on in this and the next sections, we suppose that n is even.

LEMMA 4.1. *Let T be a tree on n vertices and let $\Delta(T) < 0.8n$. Then there is a subtree T' on $n/2$ vertices and a subgraph M of T such that the following properties hold.*

(1) *M is star-forest of at least $(n-1)/16$ edges;*

(2) *M is vertex-disjoint to T'.*

PROOF. If there is a cut edge, e, of T such that the deletion of e results two trees on $n/2$-$n/2$ vertices, then we are done. Otherwise, T has a (unique) vertex, v with the following property: considering the edges $vv_1, vv_2, \ldots, vv_t$ adjacent to v and the subtrees $T_1, \ldots, T_t$, obtained after deleting all of these edges ($v \notin T_i$, $v_i \in T_i$), $s_i = |V(T_i)|$, we get that $s_1 \leqq s_2 \leqq \ldots \leqq s_t < n/2$, (and, of course, $\sum s_i = n-1$). We have that $t \geqq 3$. Let j be defined by

$$1 + s_1 + \ldots + s_{j-1} < n/2 \leqq 1 + s_1 + \ldots + s_j.$$

Here $j < t$ (because $s_j < n/2$). Then T' can be any subtree of $v + T_1 + \ldots + T_j$ on $n/2$ vertices. Define M' as the forest $T_{j+1} + \ldots + T_t$. We claim that M' has at least $(n-1)/8$ edges. Indeed, if $s_j = 1$, then T' is a star and M' has at least $n - 1 - \Delta$ edges. Otherwise, for $s_j \geqq 2$ we have that

$$|E(M')| \geqq \sum_{i>j} (s_i - 1) \geqq \sum_{i>j} (s_j/2).$$

Here $\sum_{j>i} s_i > (n-1)/4$, because otherwise $s_j > (n-1)/4$ follows, and this again implies $(n-1)/4 < s_j \leqq s_{j+1}$. Finally, every forest contains a star-forest consisting of at least half of its edges, so there is an $M \subseteq M'$ of size at least $(n-1)/16$. □

Considering the decomposition, $v + T_1 + \cdots + T_{j-1}$, $v + T_j$, $v + T_{j+1} + + \cdots + T_t$ in the proof of Lemma 4.1 we obtain the following statement.

LEMMA 4.2. *The edge set of an arbitrary tree T can be partitioned into at most 3 trees each of sizes at most $|V(T)|/2$.*

Let W be the set of all neighbors of leaves of T. For each $w \in W$ choose a neighboring vertex of degree 1, we get the set W', $|W| = |W'|$. Applying Lemma 4.2 for the tree $T - W'$ one gets the following

COROLLARY 4.3. *$T - W'$ has a subtree T' on $\frac{n}{2}$ vertices, which contains at least $\frac{1}{3}|W|$ vertices of W.*

Let P be the set of pendant edges. Deleting $\deg(x) - 2$ edges from each vertex x of degree at least 3 one gets a subforest which is a path-forest, i.e., we obtain the following

LEMMA 4.4. *T_n has a subforest T' of at least $n - |P|$ edges consisting of vertex-disjoint paths, edges and isolated vertices.*

5. Proof of Theorem 1.2

Let $n > m_0$, and let T be a tree on n vertices satisfying $\Delta(T) < 0.8n$. Let φ be a two-coloring of the edges of K_n, and suppose, on the contrary, that $\mathcal{D}_\varphi(T) < n/258 =: l$.

CLAIM 5.1. *The vertices of K_n may be partitioned by $V(K_n) = A_1 \dot\cup A_2$, $|A_1| = |A_2| = \frac{n}{2}$ so that both graphs*

$$G_i = \{e \subset A_i : e \text{ has color } i\}, \quad i = 1, 2$$

contain all trees on $\frac{n}{2} - 8l$ vertices. Moreover, there are sets $B_i \subseteq A_i$ such that the minimum degree of the restriction of G_i to B_i is at least $n - 8l$.

PROOF. Let T_1' be a subtree on $\frac{n}{2} - l$ vertices provided by Lemma 4.1. Let T_2' be a subtree of T_1' of $n/2 - 3l$ edges. Also let M_1 and M_2 be star-forest contained in T vertex-disjoint to T_1' and T_2', respectively, of sizes $|E(M_1)| = 2l$, $|E(M_2)| = 4l$.

We use Corollary 2.8 to find a monochromatic copy of T_1', say color 1. Let A_1' be formed by the vertices of this copy of T_1'. There is no copy of M_1 of color 1 vertex disjoint to A_1', otherwise we obtain $\mathcal{D}_\varphi(T) \geq l$. By Lemma 2.11 we have that $V(K_n) - A_1'$ contains a copy of T_2' of color 2. Then define the sets A_1, A_2 such that $V(T_i') \subseteq A_i$, $|A_i| = n/2$, $A_1 \cup A_2 = V(K_n)$. The set A_i does not contain a copy of M_{3-i} of color $3 - i$. Hence Lemma 2.11 yields that A_i contains a set B_i satisfying the requirements and B_i contains all trees of color i of sizes at most $n/2 - 8l$. □

To finish the proof of Theorem 1.2 we distinguish three cases.

1. $|W| \geq 54l$, where W is the set of all neighbors of leaves of T. By Corollary 4.3 there is a subtree T' on $\frac{n}{2}$ vertices and a matching M such that each edge of M intersects T' in one vertex and $|M| \geq 18l$. In each A_i, $i = 1, 2$ take a copy of T_i' with at least $\frac{n}{2} - 8l$ edges of color i. Between $(M \cap T_i') \cap A_i$, $i = 1, 2$ there must be a monochromatic matching of at least $9l$ edges. This together with the corresponding copy of T' has at least $\frac{n}{2} + l$ edges of the same color. This finishes Case 1.

2. $|P| < n/4 - (3/2)l$, where P is the set of pendant edges. By Lemma 4.1, T contains a path-forest, T', of at least $n - |P|$ edges. We apply (2) that K_n contains a monochromatic path, H, of at least $(2/3)(n-1)$ edges. We can cover at least 2/3 of the edges of T' by H and conclude that T has a monochromatic part of at least $n/2 + l$ edges. This finishes Case 2.

3. If neither Case 1 nor Case 2 take place then let T' be a subtree of $(n/2) - |W| - 8l$ edges on $(n/2) - 8l$ vertices obtained from T by deleting edges in the following 3 steps. Let P' be a set of $(n/4) - (3/2)l$ pendant edges, delete these from T. Second, delete $|W| - 1$ edges such that the rest of the tree consists of $|W|$ components each component having exactly one vertex from W. Finally, trim leaves off these components to get the desired size such that we never cut off a vertex of W.

Without loss of generality we may assume that A_1 has a set N of $\frac{n}{4}$ vertices such that each of them is incident with at least $\frac{n}{4}$ edges of color 1 going to A_2. Fix a copy of T' in B_1 such that the vertices of W all come from $B_1 \cap N$. The edges of P' can be added to T' from the color 1 edges

between A_1 and A_2. We found a subgraph of T with at least $\frac{3n}{4} - 63.5l$ edges of color 1. This finishes Case 3, thus Theorem 1.2 is proved.

6. Proof of Theorem 1.3

Let T be a tree on n vertices and consider an arbitrary two-coloring of the edges of K_n using colors red and blue. We claim that K_n contains a subforest of T of at least $n - 4k$ edges of the same color, consisting of vertex-disjoint stars.

Let T^* be a maximum star-forest of T. T^* has at least $n - \tau$ edges. Let T' be a maximum monochromatic subgraph of T^*. If T' has at least $n - 4k$ edges we are done. In the rest of the proof we assume that T' has less than $n - 3k$ vertices.

Let us assume that the color of T' is red. Let x be a vertex of T' of degree 1. There are less than k red edges going from x to vertices out of T' in T, otherwise T' may be improved by replacing the edge incident with x by the red star of k edges rooted in x, whose leaves do not belong to T'. This new system of red stars contains a subgraph of T^* which is bigger than T'. Similar argument shows that red stars of $K_n - V(T')$ have at most $(k-1)$ edges. Let $W = V(K_n) - V(T')$ and let $W = W_1 \cup W_2$ be a partition of W such that $|W_1| = 3k$. We will construct a big blue subgraph of T^*. Its stars will be rooted in W_1 and leaves will be in $V(K_n) - W_1$. Let T'' denote the current part of this blue subgraph which we have already constructed. We enlarge T'' as follows. If there are at least $2k$ vertices of $M = W_2 \cup \{j;\ j$ is a vertex of T' of degree $1\}$ uncovered by T'' then observe that at least one vertex of $W_1 - T''$ is incident by blue edges with at least k vertices of M. Thus we enlarge T'' by adding this star to it. If less than $2k$ vertices of M are uncovered by T'' then we stop. In the end T'' has at least $n - 3k$ vertices out of W_1, hence it has at least $n - 4k$ edges. Hence Theorem 1.3 is proved.

7. Further problems and generalizations

Above the special case was considered when $E(K_n)$ was two-colored, and we investigated how large monochromatic portion of a given tree T_n must be contained in it. Here we give a list of some possible generalizations.

(1) Instead of K_n we can consider other sequence of underlying graphs, e.g., the complete bipartite $K_{n,n}$, t-partite graphs $K_{n,n,\dots,n}$;

(2) Instead of copies of a T_n some other family of graphs, even with different sizes can be investigated;

(3) Two coloring can be replaced by r-coloring;

(4) Instead of the measuring the disrepancy in supremum norm it is interesting to consider the average, e.g., the l_2 norm;

(5) Instead of considering the maximum distance from the evenly colored subgraphs (when the goal was to approach a $(1/2, 1/2)$ coloring) to consider for a given $\alpha \in (0,1)$ the discrepancy from an $(\alpha, 1-\alpha)$ coloring. Some applications lead these kind of questions, eventually α depends on n, $\alpha = \alpha(n)$;

Finally we mention two further problems.

1. A general method in discrepancy theory is to obtain an estimation from the discrepancy of the random coloring. One of the first problems is to decide when the random coloring yields the optimal or nearly optimal solutions. It is easy to see that when $|E(L)| = \omega(n) n \log n$ with $\omega(n) \to \infty$, for $n \to \infty$, then already the random coloring φ gives $\mathcal{D}_n(L, \varphi) = o(|E(L)|)$.

2. In our case (the case of spanning trees) the bounds on the discrepancy are in terms of the maximum degree, Δ, and the covering number, τ. It would be interesting to see what other graph parameters or structural properties of the sample graphs (and the underlying graphs) influence the discrepancy. For example, if the tree T_n has two vertices of degree $n/2$ (it is called a *broom*), then $\mathcal{D}(T_n) = n/4 + O(1)$. (This was also proved by Bondy [4].)

ACKNOWLEDGEMENT. The authors would like to thank János Komlós, Imre Ruzsa and Mikkel Thorup for helpful discussions. This research of the second author was partially supported by an NSF Grant. The research of the fourth author was partially supported by the Hungarian National Foundation for Scientific Research Grant No. 1909.

REFERENCES

[1] AJTAI, M., KOMLÓS, J. and SZEMERÉDI, E., On a conjecture of Loebl, *Proc. 7th International Conference on Graph Theory*, Kalamazoo, Michigan, 1993.

[2] BECK, J. and CHEN, W. L., *Irregularities of distribution*, Cambridge Tracts in Mathematics, 89, Cambridge Univ. Press, Cambridge–New York, 1987. *MR* **88m**: 11061

[3] BECK, J. and SÓS, V. T., Discrepancy theory, *Handbook of combinatorics*, Springer, Berlin–New York, 1994.

[4] BONDY, A., Personal communication.

[5] ERDŐS, P. and SPENCER, J., Imbalances in k-colorations, *Networks* 1 (1971/72), 379–385. *MR* **45** #8573

[6] GERENCSÉR, L. and GYÁRFÁS, A., On Ramsey-type problems, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **10** (1967), 167–170. *MR* **39** #1351

[7] KOMLÓS, J., Personal communication, 1993.

[8] SÓS, V. T., Irregularities of partitions: Ramsey theory, uniform distribution, *Surveys in combinatorics* (Southampton, 1983), London Math. Soc. Lecture Note Ser., 82, Cambridge Univ. Press, Cambridge–New York, 1983, 201–246. *MR* **85h**:05012

(Received February 14, 1994)

P. Erdős and V. T. Sós

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Z. Füredi

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

and

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, IL 61801–2917
U.S.A

M. Loebl

KATEDRA APLIKOVANÉ MATEMATIKY
MATEMATICKO–FYZIKÁLNI FAKULTA
UNIVERZITA KARLOVA
MALOSTRANSKÉ NÁM. 25
CZ–118 00 PRAHA 1
CZECH REPUBLIC

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 59–94

GEOMETRIC DISCREPANCY THEOREMS IN HIGHER DIMENSIONS

GY. KÁROLYI

1. Introduction and a brief survey

The classical problem

Let $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_N\} \subset [0,1)^K$ be a finite sequence of not necessarily distinct points. To investigate the uniformity of this sequence one considers the differences

$$D(\mathcal{P}, A) = |Z(\mathcal{P}, A) - N\mu(A)|$$

between the number of the points contained in A and the expected number of the points contained in A, for certain measurable subsets A of the unit cube, where $Z(\mathcal{P}, A)$ is the number of the points $\mathbf{p}_i$ that are contained in A. Usually the sets A are specified as the aligned boxes contained in the unit cube $[0,1)^K$. Accordingly, the discrepancy of the sequence $\mathcal{P}$ is defined as

$$D(\mathcal{P}) = \sup_A D(\mathcal{P}, A),$$

where the supremum is extended to the family of aligned boxes (i.e. direct products of intervals) $A \subseteqq [0,1)^K$. The less the discrepancy of the sequence $\mathcal{P}$ is the more uniform its distribution is. In case $K = 1$ clearly there exist very evenly distributed N-element sequences $\mathcal{P}$ for every natural number N: if $\mathcal{P} = \{0, \frac{1}{N}, \cdots, \frac{N-1}{N}\}$, then $D(\mathcal{P}, A) \leqq 2$ for every interval $A \subseteqq [0,1)$.

In higher dimensions the situation is different. In his basic paper published in 1954 Roth proved the following result.

THEOREM 1.1 (Roth [22]). *For an arbitrary integer $N \geqq 2$ and distribution $\mathcal{P} \subset [0,1)^K$ of N points*

$$D(\mathcal{P}) \gg_K (\log N)^{\frac{K-1}{2}} .$$

1991 *Mathematics Subject Classification.* Primary 11K38; Secondary 05C65, 52A27, 52B12.

Key words and phrases. Geometric discrepancy, irregularities of distribution, uniform distribution in high dimensions, discrepancy of matrices.

(We use the so-called Vinogradov symbol: $f \ll_I g$ means that there exists a constant c depending on the index set I such that $f < cg$.)

Pointsets with relatively small discrepancy was constructed at first by Halton, in every dimension. These kinds of sequences are very useful because of their application in numerical integration (for the background see e.g. [18], [14], [21]).

THEOREM 1.2 (Halton [13]). *For an arbitrary integer $N \geq 2$ there exists a distribution $\mathcal{P} \subset [0,1)^K$ of N points satisfying*

$$D(\mathcal{P}) \ll_K (\log N)^{K-1} .$$

One of the most important open problems of the subject is to eliminate the gap between the two bounds. It is suspected that the upper bound is exact. The following results seem to support this conjecture.

THEOREM 1.3 (Schmidt [26]). *For an arbitrary integer $N \geq 2$ and distribution $\mathcal{P}$ of N points in the unit square $[0,1)^2$*

$$D(\mathcal{P}) \gg \log N .$$

Halász [12] gave an alternate proof of Theorem 1.3 by modifying Roth's basic idea.

Theorem 1.1 was slightly improved recently by Beck in case $K = 3$. For higher dimensions there are not known any stronger lower estimate.

THEOREM 1.4 (Beck [5]). *For an arbitrary integer $N \geq 3$ and distribution $\mathcal{P} \subset [0,1)^3$ of N points*

$$D(\mathcal{P}) \gg_\varepsilon (\log N)(\log\log N)^{\frac{1}{8}-\varepsilon} ,$$

where ε is an arbitrarily small positive number.

Let us say a few words on the history of the subject. The first problems in this topic arised in number theory concerning infinite sequences. One can measure the uniformity of the distribution of an infinite sequence $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots\} \subset [0,1)^k$ by the sequence of its discrepancies

$$D_N(\mathcal{Q}) = D(\{\mathbf{q}_1, \dots, \mathbf{q}_N\}) .$$

The sequence $\mathcal{Q}$ is uniformly distributed in the unit cube if $D_N(\mathcal{Q}) = o(N)$. Well-known examples are the so-called van der Corput-sequence [11], or the $\{N\alpha\}$-sequence, if α is irrational. J.G. van der Corput [11] raised the question in 1935, if a sequence $\mathcal{Q}$ may be so uniformly distributed in the interval $[0,1)$ that the sequence of discrepancies $D_N(\mathcal{Q})$ is bounded. Note that the best possible upper bound for the sequences mentioned above is $D_N(\mathcal{Q}) \ll \ll \log N$. The negative answer was given by van Aardenne-Ehrenfest [1] in 1945. She also proved a stronger theorem in [2]. An essential improvement

of this result was carried out by Roth (Theorem 1.1). He observed that the problem can be translated to investigation of finite subsets of $[0,1)^2$. More precisely he showed the equivalence of the following two statements:

(1) for an arbitrary infinite sequence $\mathcal{Q} \subset [0,1)^k$ and natural number N there exists an integer $1 \leqq n \leqq N$ with $D_n(\mathcal{Q}) \gg f(N)$,

and

(2) for an arbitrary integer $N \geqq 2$ and a distribution $\mathcal{P} \subset [0,1)^K$ of N points $D(\mathcal{P}) \gg f(N)$ holds, where $K = k+1$.

In fact, Schmidt stated and proved Theorem 1.3 in the form (1). Furthermore, he proved ([25]), that for any infinite sequence $\mathcal{Q} \subset [0,1)$ the set of real numbers $0 < \alpha \leqq 1$ for which the sequence

$$D_N(\mathcal{Q}, \alpha) = D(\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}, [0, \alpha))$$

is bounded, is countable.

Geometric discrepancy

The uniformity and irregularity of sequences may be studied with respect to various kinds of geometric objects instead of aligned boxes. As the multidimensional analogue of intervals, it seems natural to consider aligned cubes or balls. We have to note that balls raise some technical difficulties. Indeed, the unit cube intersects an aligned box in an aligned box, but its analogue for cubes or balls is clearly not true.

Halász in 1985 proved (see [7]) that Theorem 1.3 remains valid if the family of aligned rectangles is replaced with the family of aligned squares. This observation is true in a much stronger sense. Ruzsa in 1991 discovered the following connection between two notions of discrepancy.

THEOREM 1.5 (Ruzsa [23]). *Denote by $\mathcal{N}$ and $\mathcal{T}$ the families of the aligned squares and rectangles contained in the unit square, respectively. For an arbitrary finite sequence of points $\mathcal{P} \subset [0,1)^2$ one has*

$$\sup_{A \in \mathcal{N}} D(\mathcal{P}, A) \gg \sup_{A \in \mathcal{T}} D(\mathcal{P}, A) \ .$$

(The reversed inequality is obvious.)

Theorem 1.1 has the following stronger form, too.

THEOREM 1.6 (Beck [7, Theorem 19A]). *Let $\mathcal{P} \subset [0,1)^K$ be a distribution of N points. There exists an aligned cube $A \subseteqq [0,1)^K$ satisfying*

$$D(\mathcal{P}, A) \gg_K (\log N)^{\frac{K-1}{2}} \ .$$

The analogue of Theorem 1.5 in higher dimensions is still open.

The first investigations concerning balls was made by Schmidt [24]. The following lower bound is essentially the best possible.

THEOREM 1.7 (Beck [3]). *Let $\mathcal{P} \subset [0,1)^K$ be a distribution of N points. There exists a ball $A \subseteqq [0,1)^K$ satisfying*

$$D(\mathcal{P}, A) \gg_{K,\varepsilon} N^{\frac{1}{2}-\frac{1}{2K}-\varepsilon} ,$$

where ε is an arbitrarily small positive number.

This theorem shows that with respect to balls sequences always have big discrepancy, and the order of magnitude is a power of N. The situation is the same if we consider cubes in arbitrary position instead of balls. One can find even bigger irregularities if the discrepancy of sequences is defined relative to the family of all convex sets.

THEOREM 1.8 (Schmidt [28]). *Let $\mathcal{P} \subset [0,1)^K$ be a distribution of N points. There exists a convex set $A \subseteqq [0,1)^K$ satisfying*

$$D(\mathcal{P}, A) \gg_K N^{1-\frac{2}{K+1}} .$$

This lower bound is essentially the best possible (see Stute [31] and Beck [6]).

It is worth mentioning that Theorem 1.7 is part of a more general phenomenon.

THEOREM 1.9 (Beck [3]). *Let $A \subseteqq [0,1)^K$ be a convex body, and denote by $\mathcal{A}$ the family of convex bodies obtained from A by a similarity transformation of ratio less than 1. Then for an arbitrary distribution $\mathcal{P} \subset [0,1)^K$ of N points*

$$\sup_{B \in \mathcal{A}} D(\mathcal{P}, B \cap [0,1)^K) \gg_A N^{\frac{1}{2}-\frac{1}{2K}} ,$$

and this lower bound is essentially the best possible.

The situation is basically different if rotation is not allowed. Then the discrepancy heavily depends on the body A itself. Indeed, let $A \subset \mathbb{R}^2$ be a convex region, and denote by A_l a convex l-gon of greatest area inscribed into A.

THEOREM 1.10 (Beck [4]). *Let $A \subseteqq [0,1)^2$ be a convex region and denote by $\mathcal{A}$ the family of convex regions obtained from A by reduction and translation ("homothetic copies of A"). Then for an arbitrary distribution $\mathcal{P} \subset [0,1)^2$ of N points*

$$\sup_{B \in \mathcal{A}} D(\mathcal{P}, B \cap [0,1)^2) \gg \sqrt{\xi_N(A)} (\log N)^{-\frac{1}{4}} .$$

On the other hand, for an arbitrary positive number ε and integer N there exists a distribution $\mathcal{P} \subset [0,1)^2$ of N points satisfying

$$\sup_{B \in \mathcal{A}} D(\mathcal{P}, B \cap [0,1)^2) \ll_{\varepsilon,A} \xi_N(A) (\log N)^{4.5+\varepsilon} ,$$

where the number $\xi_N(A)$ *is the smallest integer* $l \geqq 3$ *for which* $\mu(A \setminus A_l) \leqq \leqq l^2/N$ *holds.*

In this paper we will study irregularities of point distributions relative to convex polytopes having facets parallel to given hyperplanes. To see the connections with related fields of combinatorics, number theory and geometry, we refer to the surveys of Beck and T. Sós [9], [30].

2. New results

Let there be given a set of hyperplanes $\mathcal{A} = \{H_1, \dots, H_l\}$ in the K-dimensional Euclidean space $\mathbb{R}^K$. Denote by $\mathrm{POL}(\mathcal{A}) = \mathrm{POL}(H_1, \dots, H_l)$ the family of convex K-polytopes having facets parallel to the given hyperplanes. We will suppose that $\mathrm{POL}(\mathcal{A})$ is not empty. Define $\mathrm{POL}_0(\mathcal{A}) = \{A \cap \cap [0,1]^K \mid A \in \mathrm{POL}(\mathcal{A})\}$, and let ε be an arbitrarily small positive number. We will prove the following generalizations of some results of Beck and Chen [8, Theorem 3] and Beck [4, Theorem 4D] in higher dimensions.

THEOREM A. *There exists an infinite sequence of points* $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots\}$ *in* $\mathbb{R}^K$ *such that for any convex polytope* $A \in \mathrm{POL}(\mathcal{A})$

$$|D(\mathcal{Q}, A)| = |Z(\mathcal{Q}, A) - \mu(A)| \ll_{\mathcal{A},\varepsilon} (\log(d(A) + 2))^{3K-1+\varepsilon}.$$

COROLLARY. *For every integer* $N \geqq 2$ *there exists an* N*-element subset* $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_N\}$ *of the unit cube* $[0,1]^K$ *such that for any* $A \in \mathrm{POL}(\mathcal{A})$, $A \subseteq [0,1]^K$

$$|D(\mathcal{P}, A)| = |Z(\mathcal{P}, A) - N\mu(A)| \ll_{\mathcal{A},\varepsilon} (\log N)^{3K-1+\varepsilon}.$$

The combinatorial investigations of Section 4 allows us to prove the following stronger version of this result.

THEOREM B. *For every integer* $N \geqq 2$ *there exists an* N*-element subset* $\mathcal{P}$ *of the unit cube* $[0,1]^K$ *such that for any* $A \in \mathrm{POL}_0(\mathcal{A})$

$$|D(\mathcal{P}, A)| \ll_{\mathcal{A},\varepsilon} (\log N)^{\max\{\frac{3}{2}K+1+\varepsilon, 2K-1\}}.$$

The proof of these theorems are more or less parallel to that of the referred theorems of Beck and Chen. Sections 3 and 4 contain the new technical machinery we need to prove Theorems A and B. The proofs are presented in Section 5.

On the other hand, it follows from Theorem 1.1 that for every integer N and a distribution $\mathcal{P}$ of N points in $[0,1]^K$ there exists a convex polytope $A \in \mathrm{POL}(\mathcal{A})$ — namely a parallelepiped — such that

$$|D(\mathcal{P}, A)| \gg_{\mathcal{A}} (\log N)^{\frac{K-1}{2}}.$$

Theorem C gives a stronger form of Roth's theorem. Let $P \subset \mathbb{R}^K$ be a convex polytope ($K \geq 2$). Denote by $\mathrm{POL}(P)$ the set of convex polytopes contained in the unit cube $[0,1]^K$ having facets parallel to those of P. For example, if P is the unit cube itself, then $\mathrm{POL}(P)$ is the set of aligned boxes contained in $[0,1]^K$; if P is a simplex, then $\mathrm{POL}(P)$ is the set of simplices homothetic to P, contained in $[0,1]^K$.

THEOREM C. *For every positive integer N and a distribution $\mathcal{P}$ of N points in $[0,1]^K$ there exists a convex polytope $A \in \mathrm{POL}(P)$ such that*

$$|D(\mathcal{P}, A)| \gg_P (\log N)^{\frac{K-1}{2}}.$$

REMARK. For $K = 2$ Beck and Chen [8, Theorem 2] proved this theorem with lower bound $\log N$. It is reasonable to suspect the lower bound $(\log N)^{K-1}$ in Theorem C (and the same upper bound in Theorem B). Our proof in Section 6 will combine the analytic method developed by Roth [22] for the case $P = [0,1]^K$ with the idea of Beck and Chen for the planar case.

Let $\mathcal{R} = \{\mathbf{x}_1, \mathbf{x}_2, \dots\} \subset [0,1]^K$ be an arbitrary infinite sequence of points, and let A be a measurable subset of the unit cube $[0,1]^K$. Then

$$D_N(\mathcal{R}, A) = \left| \sum_{i=1}^{N} \chi_A(\mathbf{x}_i) - N\mu(A) \right|$$

is the discrepancy of the pointset consisting of the first N elements of $\mathcal{R}$ with respect to the set A. According to Theorem C, there exists a convex polytope $A \in \mathrm{POL}(P)$, for which

$$D_N(\mathcal{R}, A) \gg_P (\log N)^{\frac{K-1}{2}}.$$

We can prove a theorem that is similar to the result of Schmidt [25] what we have already mentioned.

THEOREM D. *Let $\mathcal{R} = \{\mathbf{x}_1, \mathbf{x}_2, \dots\} \subset [0,1]^K$ be an arbitrary infinite sequence of points. Then the set of the real numbers $\alpha \in [0,1]$ for which there exists a convex polytope $A \in \mathrm{POL}_0(\mathcal{A})$ with $\mu(A) = \alpha$ such that the sequence $D_N(\mathcal{R}, A)$ is bounded, is countable.*

We indicate the proof of this theorem in Section 6.

3. Geometric considerations

In the first part of this section we formulate a geometric lemma on approximation of simplices by convex polytopes of a certain type. First of all we have to discuss what kind of polytopes we will work with.

Define the basic simplex of $\mathbb{R}^K$ as follows:

$$\Delta^K = \left\{ \mathbf{x} \in \mathbb{R}^K \mid x_i \geqq 0 \, (i = 1, \ldots, K), \sum_{i=1}^{K} x_i \leqq 1 \right\}.$$

We will denote the vertices of Δ^K by $\mathbf{e}_1, \ldots, \mathbf{e}_{K+1}$, where $\mathbf{e}_{K+1}$ is the origin and $e_{ij} = \delta_{ij}$ for $1 \leqq i, j \leqq K$.

Let $d_i \in \mathbb{N}$, $\sum_{j=1}^{i} d_j = s_i$ $(s_0 = 0)$, $s_k = K$ and $\alpha_i > 0$. Define the polytope

$$E(d_1, \ldots, d_k; \alpha_1, \ldots, \alpha_k) =$$
$$= \{\mathbf{x} \in \mathbb{R}^K \mid x_j \geqq 0 \ (1 \leqq j \leqq K), \ x_{s_{i-1}+1} + \cdots + x_{s_i} \leqq \alpha_i \, (i = 1, \ldots, k)\}.$$

Consider the family of all the polytopes of the form

$$E(d_1, \ldots, d_k; 2^{n_1}, \ldots, 2^{n_k}) + \mathbf{v}_{n_1, \ldots, n_k}(\tau),$$

where $k \leqq K$, $d_i \in \mathbb{N}$ with $\sum_{i=1}^{k} d_i = K$, $n_i \in \mathbb{Z}$ and the translation vectors $\mathbf{v}_{n_1, \ldots, n_k}(\tau)$ run through all the elements

$$(l_{11}2^{n_1}, \ldots, l_{i1}2^{n_i}, \ldots, l_{id_i}2^{n_i}, \ldots, l_{kd_k}2^{n_k})$$

of $\mathbb{R}^K$, where the l_{ij}'s are integers. We will call them the *basic special polytopes* of $\mathbb{R}^K$ and denote this family of polytopes by $\mathrm{BSP}(\mathbb{R}^K)$. This is a subset of the family of *basic good polytopes* of $\mathbb{R}^K$

$$\mathrm{BGP}(\mathbb{R}^K) = \{E(d_1, \ldots, d_k; \alpha_1, \ldots, \alpha_k) + \mathbf{v}\},$$

where $k \leqq d_i$, $d_i \in \mathbb{N}$ with $\sum_{i=1}^{k} d_i = K$, $\alpha_i > 0$ for $1 \leqq i \leqq k$ and $\mathbf{v} \in \mathbb{R}^K$.

Each simplex Δ of $\mathbb{R}^K$ determines $(K+1)!$ linear transformations $A_i^{(\Delta)}$ of determinant ± 1 and vectors $\mathbf{v}_i^{(\Delta)} \in \mathbb{R}^K$ such that

$$\Delta = \lambda A_i^{(\Delta)}(\Delta^K) + \mathbf{v}_i^{(\Delta)},$$

where $\lambda = \mu(\Delta)^{1/K} \mu(\Delta^K)^{-1/K}$. With the help of these linear transformations we are able to define the polytopes we will consider.

DEFINITION 3.1. The family of good polytopes belonging to Δ is

$$\mathrm{GP}(\Delta) = \{A_i^{(\Delta)}(E) \mid E \in \mathrm{BGP}(\mathbb{R}^K), 1 \leqq i \leqq (K+1)!\};$$

and the family of special polytopes belonging to Δ is

$$\mathrm{SP}(\Delta) = \{A_i^{(\Delta)}(E) \mid E \in \mathrm{BSP}(\mathbb{R}^K), 1 \leqq i \leqq (K+1)!\}.$$

We also introduce the following notations:

$$\Phi(\Delta) = \{A_i^{(\Delta)} \mid 1 \leqq i \leqq (K+1)!\}$$

and

$$\Psi(\Delta) = \{A_i^{(\Delta)} + \mathbf{v}_i^{(\Delta)} \mid 1 \leqq i \leqq (K+1)!\}.$$

For technical reasons fix two sequences of constants (c_K^*) and (c_K^{**}) with

$$\frac{3}{4} = c_1^* < c_2^{**} < c_2^* < c_3^{**} < \ldots < 1.$$

The crucial result of this section is the following lemma.

LEMMA 3.2. *There exist polynomials p_K^* of degree at most $2K-1$ and positive constants D_K^* such that for any simplex Δ of $\mathbb{R}^K$ and positive integer $D > D_K^*$ there exist pairwise disjoint polytopes $S_1, \ldots, S_m \in \mathrm{SP}(\Delta)$ with $m < p_K^*(D)$ satisfying*

1) $\bigcup_{i=1}^m S_i \subseteqq \Delta$ *and*
2) $\mu(\Delta) - \sum_{i=1}^m \mu(S_i) < (c_K^*)^D \mu(\Delta)$.

NOTE. Throughout this paper we use the following convention. We say that two polytopes are disjoint, if their interiors are disjoint, i.e. we allow common boundary points. We use the symbol $\bigcup^*$ in this sense. Similarly, we define the characteristic function of the set S by

$$\chi_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathrm{int} S, \\ 0 & \text{if } \mathbf{x} \in \mathrm{ext} S, \end{cases}$$

but we do not fix its value on the boundary points, we allow 0 or 1 values depending on the situation.

We will prove Lemma 3.2 by induction on K. The case $K = 1$ is essentially covered by Lemma 8.16 of Beck and Chen [7]. For the sake of completeness we present the proof here with a slight modification.

LEMMA 3.3. *Let a, b be real numbers, $a < b$. For every positive integer D there exist pairwise disjoint intervals $I_1, \ldots, I_D$ of the form $I_i = [k_i 2^{n_i}, (k_i+1)2^{n_i}]$ with $k_i, n_i \in \mathbb{Z}$ such that*

1) $\bigcup_{i=1}^D I_i \subseteqq [a, b]$ *and*
2) $(b-a) - \sum_{i=1}^D 2^{n_i} < (3/4)^D (b-a)$.

PROOF. First let I_1 be the longest interval of the desired form contained in $[a, b]$, then the length of I_1, $2^{n_1} > \frac{1}{4}(b-a)$. For $i > 1$ we can define I_i inductively so as to satisfy the following properties:

1) I_i is the longest interval of the desired form, contained in $[a, b] \setminus \setminus \bigcup_{j=1}^{i-1} I_j$,

2) $\bigcup_{j=1}^{i} I_j$ is an interval and

3) $I_i \subseteq K_{i-1}$ if we write $[a,b] \setminus \bigcup_{j=1}^{i-1} I_j = K_{i-1} \cup J_{i-1}$, where J_{i-1} and K_{i-1} are disjoint intervals, J_{i-1} is not longer than K_{i-1}, and $J_{i-1} = \emptyset$ if $[a,b] \setminus \bigcup_{j=1}^{i-1} I_j$ is an interval.

It is easy to show with induction that $\mu(I_i) > \frac{1}{2}\mu(K_{i-1})$. Indeed, it implies that $\mu(I_i) > \frac{1}{2}\mu(K_i)$, furthermore I_i and K_i have a common endpoint of the form $k2^{n_i}$, $k \in \mathbb{Z}$, and hence we can find I_{i+1} with $\mu(I_{i+1}) > \frac{1}{2}\mu(K_i)$.

Thus for every $i > 1$ we have

$$\mu(I_i) > \frac{1}{2}\mu(K_{i-1}) \geqq \frac{1}{4}\left(b - a - \sum_{j=1}^{i-1} \mu(I_j)\right)$$

and the result follows immediately.

Before turning to the induction step, notice that it is enough to prove the assertion for simplices $\Delta \in \mathrm{BGP}(\mathbb{R}^K)$, i.e. for simplices homothetic to Δ^K. Indeed, for an arbitrary Δ we only have to apply a suitable linear transformation $\phi \in \Phi(\Delta)$. Let $K \geqq 2$ and suppose that Lemma 3.2 has already been proved for $1, \dots, K-1$. First we prove a similar assertion for good polytopes belonging to Δ, except of simplices.

LEMMA 3.4. *There exists a polynomial p_K^{**} of degree at most $2K-2$ and a positive constant D_K^{**} such that for any $P \in \mathrm{GP}(\Delta)$ which is not a simplex and arbitrary positive integer $D > D_K^{**}$ there exist pairwise disjoint polytopes $S_1, \dots, S_m \in \mathrm{SP}(\Delta)$ with $m < p_K^{**}(D)$ satisfying*

1) $\bigcup_{i=1}^{m} S_i \subseteq P$ *and*

2) $\mu(P) - \sum_{i=1}^{m} \mu(S_i) < (c_K^{**})^D \mu(P)$.

PROOF. P is of the form $E(d_1, \dots, d_k; \alpha_1, \dots, \alpha_k) + \mathbf{w}$, where $k \geqq 2$ and $\sum_{i=1}^{k} d_i = K$, therefore P is the direct product of simplices Δ_i of dimension $d_i < K$. More precisely, if we identify the d_i-dimensional coordinate-plane determined by the coordinates $x_{s_{i-1}+1}, \dots, x_{s_i}$ with $\mathbb{R}^{d_i}$, then $\Delta_i = \alpha_i \Delta^{d_i} + \mathbf{w}_i$, where $\mathbf{w}_i = (w_{s_{i-1}+1}, \dots, w_{s_i})$. Let $D > \max\{D_1^*, \dots, D_{K-1}^*\}$. We can apply the induction hypothesis to each Δ_i, then we obtain polytopes $S_{i1}, \dots, S_{im_i} \in \mathrm{SP}(\Delta^{d_i})$ with $m_i < p_{d_i}^*(D)$ satisfying

1) $\bigcup_{j=1}^{m_i} S_{ij} \subseteq \Delta_i$ and

2) $\mu(\Delta_i) - \sum_{j=1}^{m_i} \mu(S_{ij}) < (c_{d_i}^*)^D \mu(\Delta_i)$.

Consider the polytopes of the form $S_{1j_1} \times \cdots \times S_{kj_k}$. They are special polytopes belonging to Δ^K. To see this write $S_{ij} = \phi_{ij}(E_{ij})$, where $\phi_{ij} \in \Phi(\Delta_i)$ and $E_{ij} \in \mathrm{BSP}(\mathbb{R}^{d_i})$. Then clearly $E_{1j_1} \times \cdots \times E_{kj_k} \in \mathrm{BSP}(\mathbb{R}^K)$ and the polytope $S_{1j_1} \times \cdots \times S_{kj_k}$ is its image at the linear transformation $\phi \in \Phi(\Delta^K)$ composed from the ϕ_{ij_i}'s on the natural way. The number of these

polytopes is

$$\prod_{i=1}^{k} m_i < \prod_{i=1}^{k} p^*_{d_i}(D) \leqq p^{**}_K(D) = \sum_{\substack{k\geqq 2\\ d_1+\cdots+d_k=K}} \prod_{i=1}^{k} p^*_{d_i}(D)\ ,$$

where

$$\deg p^{**}_K \leqq \max_{\substack{k\geqq 2\\ d_1+\cdots+d_k=K}} \sum_{i=1}^{k} \deg p^*_{d_i} \leqq \max_{\substack{k\geqq 2\\ d_1+\cdots+d_k=K}} \sum_{i=1}^{k} (2d_i-1) = 2K-2.$$

The polytopes are disjoint and are contained in P. Finally

$$\mu(P) - \sum_{j_1=1}^{m_1} \cdots \sum_{j_k=1}^{m_k} \mu(S_{1j_1} \times \cdots \times S_{kj_k}) =$$

$$= \mu\left(\prod_{i=1}^{k} \Delta_i \setminus \prod_{i=1}^{k} \bigcup_{j=1}^{m_i}{}^{*} S_{ij}\right) < \sum_{i=1}^{k} \left(\prod_{l\neq i} \mu(\Delta_l)\right)\left(\mu(\Delta_i) - \sum_{j=1}^{m_i} \mu(S_{ij})\right) <$$

$$< \sum_{i=1}^{k} \left(\prod_{l\neq i} \mu(\Delta_l)\right) (c^*_{d_i})^D \mu(\Delta_i) = \left(\sum_{i=1}^{k} (c^*_{d_i})^D\right) \mu(P) \leqq (c^{**}_K)^D \mu(P),$$

if D is large enough.

We note that the polytopes $S_1, \ldots, S_m$ in Lemma 3.4 are not simplices.

The induction step can be derived the following way. On the one hand we will prove the analogue of the previous lemma for a certain type of simplices: a simplex $\Gamma \in \mathrm{BGP}(\mathbb{R}^K)$ is called *nice* if Γ is obtained from a basic special simplex of $\mathbb{R}^K$ cutting it with a hyperplane parallel to one of its facets. On the other hand, we will see that Δ can be decomposed to the disjoint union of nice simplices and good polytopes belonging to Δ which are not simplices. The existence of this kind of decomposition will follow from the next lemma, that we may regard as the "heart" of the induction step. A special case of this lemma will also be helpful in putting our first aim into the matter.

LEMMA 3.5. *Let* $\mathbf{p} \in \Delta^K$, $p_1 = \alpha$ $(0 < \alpha < 1)$ *and suppose that* $\mathbf{p}$ *is incident to an* h*-face of* Δ^K $(1 \leqq h \leqq K)$. *Then there exist polytopes* $G_1, \ldots, G_k \in \mathrm{GP}(\Delta^K)$, $k \leqq (h-1)\binom{K}{2} + K + 1$, *such that*

1) $\Delta^K = \bigcup_{i=1}^{k}{}^{*} G_i$,
2) G_1 *is the simplex* $\Delta^K \cap \{\mathbf{x} \in \mathbb{R}^K \mid x_1 \geqq \alpha\}$,
3) $\mathbf{p}$ *is a vertex of the simplex* G_k,
4) $G_2, \ldots, G_{k-1}$ *are not simplices.*

PROOF. We proceed by induction on h. Notice that if $\psi \in \Psi(\Delta^K)$, then ψ permutes the elements of $\mathrm{GP}(\Delta^K)$, bringing simplices to simplices. Therefore applying a suitable element of $\Psi(\Delta^K)$, if necessary, we may assume that $\beta = \sum_{i=1}^{K} p_i < 1$.

First we deal with the initial step $h = 1$, then $p_2 = \ldots = p_K = 0$.

For $1 \leqq i \leqq K+1$ define the polytope D_i by

$$D_i = \left\{ \mathbf{x} \in \mathbb{R}^K \mid x_j \geqq 0 \ (j \neq i),\ \sum_{j=1}^{i} x_j \geqq \alpha,\ \sum_{j=1}^{K} x_j \leqq 1,\ \sum_{j=1}^{i-1} x_j \leqq \alpha \right\}.$$

(Note that if $i = K+1$, then the second condition has no meaning and the third one is superfluous, so they may be omitted.)

It is easy to see, that $\Delta^K = \bigcup_{i=1}^{K+1} {}^* D_i$. Indeed, $\mathbb{R}^K = \bigcup_{i=1}^{K+1} {}^* R_i$, where

$$R_i = \left\{ \mathbf{x} \in \mathbb{R}^K \mid \sum_{j=1}^{i} x_j \geqq \alpha,\ \sum_{j=1}^{i-1} x_j < \alpha \right\},$$

and $D_i = R_i \cap \Delta^K$. On the other hand,

$$D_i = \phi_i(E(i-1, K-i+1; \alpha, 1-\alpha)) + \mathbf{e}_i,$$

where ϕ_i is the linear transformation defined by

$$\phi_i \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_i \\ x_{i+1} \\ \vdots \\ x_K \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ -x_1 - x_2 - \cdots - x_K \\ x_{i+1} \\ \vdots \\ x_K \end{pmatrix}.$$

It is clear that $\phi_i(\Delta^K) + \mathbf{e}_i = \Delta^K$, so $\phi_i + \mathbf{e}_i \in \Psi(\Delta^K)$, $\phi_i \in \Phi(\Delta^K)$. Therefore $D_i \in \mathrm{GP}(\Delta^K)$ and it is easy to check that the polytopes $G_i = D_i$ $(i = 1, 2, \ldots, k = K+1)$ satisfy the desired conditions.

Suppose now that $1 < h \leqq K$ and we have proved the lemma for $1, \ldots, h-1$. We may suppose that $\mathbf{p}$ is not contained in any $(h-1)$-face of Δ^K.

CLAIM 3.6. *If* $2 \leqq i \leqq K$, *then* $D_i^* = D_i \cap \left\{ \mathbf{x} \in \mathbb{R}^K \mid \sum_{j=1}^{K} x_i \geqq \beta \right\}$ *is the disjoint union of* $K - i + 1$ *good polytopes belonging to* Δ^K, *neither of which is a simplex.*

PROOF. $D_i^* = \phi_i(E_i^*) + \mathbf{e}_i$, where

$$E_i^* = E(i-1, K-i+1; \alpha, 1-\alpha) \cap \{\mathbf{x} \in \mathbb{R}^K \mid x_i \leqq 1-\beta\}.$$

It is enough to prove the assertion for E_i^* instead of D_i^*.

Consider the linear map $\omega_i : \mathbb{R}^K \longrightarrow \mathbb{R}^{K-i+1}$

$$\omega_i \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix} = \begin{pmatrix} x_i \\ \vdots \\ x_K \end{pmatrix}.$$

Then

$$\omega_i(E(i-1, K-i+1; \alpha, 1-\alpha)) =$$

$$= \left\{ \mathbf{x} \in \mathbb{R}^{K-i+1} \mid x_j \geqq 0 \ (1 \leqq j \leqq K-i+1), \ \sum_{j=1}^{K-i+1} x_j \leqq 1-\alpha \right\}$$

and

$$\omega_i(E_i^*) = \Big\{ \mathbf{x} \in \mathbb{R}^{K-i+1} \mid x_j \geqq 0 \ (1 \leqq j \leqq K-i+1), \ \sum_{j=1}^{K-i+1} x_j \leqq 1-\alpha, \ x_1 \leqq 1-\beta \Big\}.$$

Therefore $(1-\alpha)^{-1}\omega_i(E(i-1, K-i+1; \alpha, 1-\alpha)) = \Delta^{K-i+1}$ and we can apply the first part of the proof of the Lemma (in dimension $K-i+1$ instead of K) for the point $\frac{1-\beta}{1-\alpha}\mathbf{e}_1$. We obtain that $\omega_i(E_i^*) = \bigcup_{j=1}^{K-i+1} {}^*G_i^{(j)}$ with suitable polytopes $G_i^{(j)} \in \mathrm{GP}(\Delta^{K-i+1})$. More precisely, $G_i^{(j)}$ is of the form

$$G_i^{(j)} = \zeta_i^{(j)}(E(j, K-i+1-j; 1-\beta, \beta-\alpha)) + (1-\alpha)\mathbf{e}_{j+1},$$

where $\zeta_i^{(j)} \in \Phi(\Delta^{K-i+1})$.

If we define $\phi_i^{(j)} \in \Phi(\Delta^K)$ by

$$\phi_i^{(j)} \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_i \\ \vdots \\ x_K \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ y_1 \\ \vdots \\ y_{K-i+1} \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_{K-i+1} \end{pmatrix} = \zeta_i^{(j)} \begin{pmatrix} x_i \\ \vdots \\ x_K \end{pmatrix},$$

then $E_i^* = \bigcup_{j=1}^{K-i+1} {}^* E_i^{(j)}$, where

$$E_i^{(j)} = \phi_i^{(j)}(E(i-1, j, K-i+1-j; \alpha, 1-\beta, \beta-\alpha)) + (1-\alpha)\mathbf{e}_{j+1}$$

is a good polytope belonging to Δ^K, and is not a simplex, since $1 \leqq i-1 \leqq$ $\leqq K-1$. This completes the proof of the Claim.

For our further purpose denote $\phi_i(E_i^{(j)}) + \mathbf{e}_i$ by $D_i^{(j)}$.

Enlarge the simplex $\Delta^K \cap \left\{\mathbf{x} \in \mathbb{R}^K \mid \sum_{j=1}^K x_j \leqq \beta\right\}$ by the ratio β^{-1}. The image of the point $\mathbf{p}$ is contained in an $(h-1)$-face of Δ^K – the intersection of the original h-face of Δ^K containing $\mathbf{p}$ and the hyperplane $\left\{\mathbf{x} \in \mathbb{R}^K \mid \sum_{j=1}^K x_j = 1\right\}$. Thus we can apply the induction hypothesis to get polytopes $G_1', \ldots, G_l' \in \mathrm{GP}(\Delta^K)$, $l \leqq (h-2)\binom{K}{2} + K + 1$, with the following properties:

1) $\Delta^K = \bigcup_{i=1}^{l} {}^* G_i'$,
2) G_1' is the simplex $\Delta^K \cap \{\mathbf{x} \in \mathbb{R}^K \mid x_1 \geqq \alpha\beta^{-1}\}$,
3) $\beta^{-1}\mathbf{p}$ is a vertex of the simplex G_l',
4) $G_2', \ldots, G_{l-1}'$ are not simplices.

Now it is easy to check that with $k = l + \binom{K}{2}$ the polytopes $G_1 = D_1, G_2, \ldots, G_{k-1}, G_k = \beta G_l'$ satisfy the desired condition, if we set

$$\{G_2, \ldots, G_{k-1}\} = \{\beta G_i' \mid 2 \leqq i \leqq l-1\} \cup \bigcup_{i=2}^{K} \{D_i^{(j)} \mid 1 \leqq j \leqq K-i+1\}.$$

Next we formulate the analogue of Lemma 3.4 for nice simplices.

LEMMA 3.7. *There exists a polynomial p_K^{***} of degree at most $2K-1$ and a positive constant D_K^{***} such that for any nice simplex Γ and positive integer $D > D_K^{***}$ there exist pairwise disjoint polytopes $S_1, \ldots, S_m \in \mathrm{SP}(\Delta)$ with $m < p_K^{***}(D)$ satisfying*

1) $\bigcup_{i=1}^m S_i \subseteq \Gamma$ *and*
2) $\mu(\Gamma) - \sum_{i=1}^m \mu(S_i) < (c_K^*)^D \mu(\Gamma)$.

PROOF. We may assume that $\Gamma = \alpha\Delta^K$ for some $\alpha > 0$. There exists an integer n satisfying $\frac{\alpha}{2} < 2^n \leqq \alpha$. Then $2^n\Delta^K$ is the largest basic special simplex of $\mathbb{R}^K$ contained in Γ. Applying the previous lemma to $\Delta = \alpha^{-1}\Gamma$ and $\mathbf{p} = \frac{2^n}{\alpha}\mathbf{e}_1$ it follows that Γ is the disjoint union of $2^n\Delta$, the nice simplex $\Gamma' = \Gamma \cap \{\mathbf{x} \in \mathbb{R}^K \mid x_1 \geqq 2^n\}$ with $\mu(\Gamma') < 2^{-K}\mu(\Gamma)$, and $K-1$ good polytopes

belonging to Γ. If we translate Γ' with $-2^n\mathbf{e}_1$, then we can repeat this argument. Iterating the process D times we find that Γ is the disjoint union of D special simplices belonging to Γ, a nice simplex of volume $< 2^{-KD}\mu(\Gamma)$ and $D(K-1)$ good polytopes P_i belonging to Γ that are not simplices. Let $D > D_K^{**}$, and apply Lemma 3.4 to these good polytopes, respectively. Summing up, we obtain pairwise disjoint special polytopes belonging to Γ: $S_1, \dots, S_m$ with $m < D + (K-1)Dp_K^{**}(D) = p_K^{***}(D)$ satisfying

$$\mu(\Gamma) - \sum_{i=1}^{m} \mu(S_i) < 2^{-KD}\mu(\Gamma) + \sum_{i=1}^{(K-1)D} (c_K^{**})^D \mu(P_i) \leqq$$

$$\leqq \left((2^{-K})^D + (c_K^{**})^D\right)\mu(\Gamma) < (c_K^*)^D \mu(\Gamma)$$

if D is large enough.

LEMMA 3.8. *Δ is the disjoint union of a special simplex belonging to Δ, at most $K+1$ nice simplices and at most $((K-1)\binom{K}{2} + K + 1)(K+1)$ good polytopes belonging to Δ that are not simplices.*

PROOF. Let Δ_1 be the largest basic special simplex of $\mathbb{R}^K$ contained in Δ. Let $\mathbf{q}$ be a vertex of Δ_1 that is not a vertex of Δ. There exists a transformation $\psi \in \Psi(\Delta)$ so that

$$\psi^{-1}(\Delta_1) \subset \{\mathbf{x} \in \mathbb{R}^K \mid x_1 \geqq p_1\},$$

where $\mathbf{p} = \psi^{-1}(\mathbf{q}) \in \Delta^K$ with $p_1 > 0$. Applying Lemma 3.5 and the transformation ψ we obtain that Δ is the disjoint union of a nice simplex, at most $(K-1)\binom{K}{2} + K + 1$ good polytopes belonging to Δ that are not simplices, and a simplex $\Delta' \supseteq \Delta_1$ a facet of which contains a facet of Δ_1 incident to $\mathbf{q}$. Repeating this argument to Δ' we end with a simplex Δ'' containing Δ_1, and at least two facets of Δ_1 are contained in suitable facets of Δ''. We can iterate this step at most $K+1$ times, yielding a desired decomposition.

Now the induction step of Lemma 3.2 can be proved as follows. Let

$$\Delta = \bigcup_{i=1}^{K+2}{}^* \Delta_i \cup^* \bigcup_{i=1}^{t}{}^* P_i ,$$

where the simplices $\Delta_1, \dots, \Delta_{K+2}$ are nice ones, $\Delta_1 \in \mathrm{SP}(\Delta)$ and $P_i \in \mathrm{GP}(\Delta)$ for $1 \leqq i \leqq t \leqq ((K-1)\binom{K}{2} + K + 1)(K+1)$. Let $D > D_K^{***}$ $(\geqq D_K^{**})$ and

$$p_K^*(D) = 1 + (K+1)p_K^{***} + ((K-1)\binom{K}{2} + K + 1)(K+1)p_K^{**} ,$$

a polynomial of degree at most $2K-1$. Apply Lemmas 3.7 and 3.4 to the simplices $\Delta_2, \dots, \Delta_{K+2}$ and to the polytopes $P_1, \dots, P_t$, respectively. We

obtain special polytopes belonging to Δ: $S_1 = \Delta_1, \dots, S_m$ with $m < p_K^*(D)$ satisfying

$$\mu(\Delta) - \sum_{i=1}^{m} \mu(S_i) < \sum_{i=2}^{K+2} (c_K^*)^D \mu(\Delta_i) + \sum_{i=1}^{t} (c_K^{**})^D \mu(P_i) <$$

$$< (c_K^*)^D \left(\sum_{i=2}^{K+2} \mu(\Delta_i) + \sum_{i=1}^{t} \mu(P_i) \right) < (c_K^*)^D \mu(\Delta).$$

This completes the proof of Lemma 3.2 with $D_K^* = D_K^{***}$.

With the help of Lemma 3.2 we can approximate simplices by special polytopes from inside. Our next aim is to find an analogous statement on approximation of simplices by special polytopes from outside, too.

Let Δ be an arbitrary simplex. For every interval $I \subset \mathbb{R}$ there exist special intervals I_1, I_2 satisfying $I \subset I_1 \cup^* I_2$ and $\mu(I_1 \cup I_2) < 2\mu(I)$. Thus there exists a simplex $\Delta_1 \supsetneqq \Delta$ obtained from Δ by an enlargement of ratio < 2 that is the disjoint union of $K+1$ special polytopes T_i belonging to Δ. As in Lemma 3.8, we can use Lemma 3.5 to prove that $\Delta = \Delta_1 \setminus \bigcup_{i=1}^{t}{}^* P_i$, where $P_i \in \mathrm{GP}(\Delta)$ and $t \leqq ((K-1)\binom{K}{2} + K + 2)(K+1)$. Applying Lemmas 3.2 and 3.4 for the polytopes P_i – note that if $\Delta' \in \mathrm{GP}(\Delta)$, then $\mathrm{SP}(\Delta') = \mathrm{SP}(\Delta)$ – we can summarize our results in the following form.

LEMMA 3.9. *There exist polynomials p_K of degree at most $2K-1$ and positive constants $c_K < 1$ and D_K such that for any simplex Δ of $\mathbb{R}^K$ and positive integer $D > D_K$ there exist polytopes $S_1, \dots, S_m, T_1, \dots, T_n \in \mathrm{SP}(\Delta)$ with $m, n < p_K(D)$ satisfying*

1) $\bigcup_{i=1}^{m}{}^* S_i \subseteqq \Delta \subseteqq \bigcup_{i=1}^{K+1}{}^* T_i \setminus \bigcup_{i=K+2}^{n}{}^* T_i$,
2) $\mu(\Delta) - \sum_{i=1}^{m} \mu(S_i) < c_K^D \mu(\Delta)$,
3) $\mu(T_1) > 1, \dots, \mu(T_{K+1}) > 1$,
4) $\sum_{i=1}^{K+1} \mu(T_i) - \sum_{i=K+2}^{n} \mu(T_i) - \mu(\Delta) < c_K^D 2^K (\mu(\Delta) + 1)$ *and*
5) *S_i, T_j are contained in the neighbourhood of Δ of radius $d(\Delta) + 2$.*

In the remaining part of this section we wish to formulate a consequence of Lemma 3.9 to polytopes contained in $\mathrm{POL}(\mathcal{A})$. First we decompose them into simplices contained in $\mathrm{POL}(\mathcal{A})$. More precisely, it may happen that the normal vectors of the hyperplanes $H_1, \dots, H_l$ are not in general position, i.e. there exist K of them that are linearly dependent over $\mathbb{R}$. Therefore choose and fix a hyperplane H that is in general position with respect to $\mathcal{A}$: the normal vector of H is linearly independent of any $K-1$ of the normal vectors of the hyperplanes $H_1, \dots, H_l$.

LEMMA 3.10 (Károlyi–Lovász [17]). *For an arbitrary convex polytope* $A \in \mathrm{POL}(\mathcal{A})$ *there exist (not necessarily distinct) simplices* $\Delta_1, \dots, \Delta_t \in \mathrm{POL}(H_1, \dots, H_l, H)$ *contained in the neighbourhood of* A *of radius* $\ll_{\mathcal{A},H} d(A)$, *with* $t < c(K,l)$, *and* ± 1 *signs* $\varepsilon_1, \dots, \varepsilon_t$ *such that* $\chi_A = \sum_{i=1}^{t} \varepsilon_i \chi_{\Delta_i}$.

REMARK. As Günter M. Ziegler has showed it to me, Varchenko had proved a similar result earlier. Though his result as stated in [32] is less general than Lemma 3.10, it gives an explicit decomposition formula when the bounding hyperplanes of the polytope are in general position. A careful analysis of his proof yields also a stronger version of our result. See [16] for a short proof and a detailed discussion.

Define the set of special polytopes

$$\mathrm{SPEC}(\mathcal{A}, H) = \bigcup_{\Delta \in \mathrm{POL}(H_1,\dots,H_l,H)} \mathrm{SP}(\Delta) \ .$$

To prove Theorems A and B we will need the next result.

LEMMA 3.11. *For an arbitrary convex polytope* $A \in \mathrm{POL}(\mathcal{A})$ *there exist polytopes* $P_1, \dots, P_m, Q_1, \dots, Q_n \in \mathrm{SPEC}(\mathcal{A}, H)$ *and* ± 1 *signs* $\gamma_1, \dots, \gamma_m, \delta_1, \dots, \delta_n$ *with the following properties:*

1) $\sum_{i=1}^{m} \gamma_i \chi_{P_i} \leqq \chi_A \leqq \sum_{i=1}^{n} \delta_i \chi_{Q_i}$,
2) $\sum_{i=1}^{n} \delta_i \mu(Q_i) - \sum_{i=1}^{m} \gamma_i \mu(P_i) \ll_{K,l} (\log(d(A)+2))^{2K-1}$,
3) P_i, Q_j *are contained in the neighbourhood of* A *of radius* $\ll_{\mathcal{A},H} d(A) + 1$,
4) $m, n \ll_{K,l} (\log(d(A)+2))^{2K-1}$,
5) $\mu(P_i), \mu(Q_j) > 1$ *and*
6) *each point of* $\mathbb{R}^K$ *is covered by* $\ll_{K,l} 1$ *of the special polytopes* P_i, Q_j.

PROOF. Let $\chi_A = \sum_{i=1}^{t} \varepsilon_i \chi_{\Delta_i}$, where $\Delta_i \in \mathrm{POL}(H_1, \dots, H_l, H)$, $\varepsilon_i = \pm 1$ and $t < c(K,l)$. Then $d(\Delta_i) \ll_{H_1,\dots,H_l,H} d(A)$. Unfortunately, there is no analogous estimate for the corresponding volumes, that is the reason why we have to operate with diameters. Since $\log \mu(A) \ll_K \log d(A)$ we can apply Lemma 3.9 with $D \gg\ll \log(d(A)+2)$ to the simplices Δ_i, respectively, to obtain polytopes $S_{i1}, \dots, S_{im_i}, T_{i1}, \dots, T_{in_i} \in \mathrm{SP}(\Delta_i) \subseteqq \mathrm{SPEC}(\mathcal{A}, H)$ and ± 1 signs $\gamma_{i1}, \dots, \gamma_{im_i}, \delta_{i1}, \dots, \delta_{in_i}$ satisfying

1) $\sum_{j=1}^{m_i} \gamma_{ij} \chi_{S_{ij}} \leqq \chi_{\Delta_i} \leqq \sum_{j=1}^{n_i} \delta_{ij} \chi_{T_{ij}}$,
2) $\sum_{j=1}^{n_i} \delta_{ij} \mu(T_{ij}) - \sum_{j=1}^{m_i} \gamma_{ij} \mu(S_{ij}) \leqq 1$,
3) $m_i, n_i \ll_K (\log(d(A)+2))^{2K-1}$,
4) $\gamma_{ij} = +1$, and $\delta_{ij} = -1$ if $\mu(T_{ij}) \leqq 1$,
5) S_{ij}, T_{ij} are contained in the neighbourhood of A of radius $\ll_{\mathcal{A},H} d(A) + 1$ and
6) each point of $\mathbb{R}^K$ is covered by $\ll_{K,l} 1$ of the special polytopes S_{ij}, T_{ij}.

Therefore we have

$$\sum_{\varepsilon_i=+1}\sum_{j=1}^{m_i}\gamma_{ij}\chi_{S_{ij}}-\sum_{\varepsilon_i=-1}\sum_{j=1}^{n_i}\delta_{ij}\chi_{T_{ij}}\leqq\chi_A\leqq\sum_{\varepsilon_i=+1}\sum_{j=1}^{n_i}\delta_{ij}\chi_{T_{ij}}-\sum_{\varepsilon_i=-1}\sum_{j=1}^{m_i}\gamma_{ij}\chi_{S_{ij}}$$

and

$$\left(\sum_{\varepsilon_i=+1}\sum_{j=1}^{n_i}\delta_{ij}\mu(T_{ij})-\sum_{\varepsilon=-1}\sum_{j=1}^{m_i}\gamma_{ij}\mu(S_{ij})\right)-$$

$$-\left(\sum_{\varepsilon=+1}\sum_{j=1}^{m_i}\gamma_{ij}\mu(S_{ij})-\sum_{\varepsilon=-1}\sum_{j=1}^{n_i}\delta_{ij}\mu(T_{ij})\right)<c(K,l).$$

Thus the polytopes

$$\{P_1,\dots,P_m\}=\bigcup_{\varepsilon_i=+1}\{S_{i1},\dots,S_{im_i}\}\cup\bigcup_{\varepsilon_i=-1}\{T_{i1},\dots,T_{in_i}\}$$

with the corresponding ±1 signs $\gamma_i=\gamma_{ij}$ or $-\delta_{ij}$ and the polytopes

$$\{Q_1,\dots,Q_n\}=\bigcup_{\varepsilon_i=+1}\{T_{i1},\dots,T_{in_i}\}\cup\bigcup_{\varepsilon_i=-1}\{S_{i1},\dots,S_{im_i}\}$$

with the corresponding ±1 signs $\delta_i=\delta_{ij}$ or $-\gamma_{ij}$ clearly satisfies conditions 1)–4) and 6) of the lemma. If we omit the polytopes of volume $\leqq 1$, then conditions 3), 4) and 6) clearly remain satisfied and condition 5) holds also obviously. Moreover, both sides of the first inequality change in advantageous direction, so the first condition remains satisfied, too. Finally, we have changed the left-hand side of the second inequality $\ll_{K,l}(\log(d(A)+2))^{2K-1}$ times by at most 1, so condition 2) is also fulfilled.

4. Two-colourings of vector-systems

In this section we work out the combinatorial tools we will need in the next section. The proof of Theorem A will depend on the following lemma.

LEMMA 4.1 (Beck and Chen [7, Lemma 8.6]). *Suppose that* $X=\{x_1,\dots,x_p\}$ *is a finite set. For* $i=1,2,\dots$, *let* $\mathcal{Y}^{(i)}=\{Y_1^{(i)},Y_2^{(i)},\dots\}$ *be a partition of* X*:*

$$X=\bigcup_{j\geqq1}^{\bullet}Y_j^{(i)}.$$

Let us associate a real number $\alpha_k \in [0,1]$ with each point $x_k \in X$. Then for every $\eta > 0$ there exist integers $a_k \in \{0,1\}$ such that

$$\left| \sum_{x_k \in Y_j^{(i)}} (a_k - \alpha_k) \right| \ll_\eta i^{1+\eta}$$

for all $Y_j^{(i)}$ satisfying $i \geqq 1$ and $j \geqq 1$.

Given a set-system $\mathcal{Y}$ on the finite underlying set $X = \{1, \dots, p\}$, we can define its discrepancy as

$$\operatorname{disc}(\mathcal{Y}) = \min_{f: X \to \{-1,1\}} \max_{Y \in \mathcal{Y}} \left| \sum_{i \in Y} f(i) \right| .$$

If we consider $\{0,1\}$-colourings instead of $\{-1,1\}$-colourings we can normalize the discrepancy as follows:

$$\operatorname{disc}(\mathcal{Y}) = 2 \min_{f: X \to \{0,1\}} \max_{Y \in \mathcal{Y}} \left| \sum_{i \in Y} (f(i) - \tfrac{1}{2}) \right| .$$

In this approach it is easy to compare discrepancy with linear discrepancy,

$$\operatorname{lindisc}(\mathcal{Y}) = \max_{\alpha_1, \dots, \alpha_p \in [0,1]} \min_{f: X \to \{0,1\}} \max_{Y \in \mathcal{Y}} \left| \sum_{i \in Y} (f(i) - \alpha_i) \right| .$$

Then obviously $2\operatorname{lindisc}(\mathcal{Y}) \geqq \operatorname{disc}(\mathcal{Y})$. On the other hand, linear discrepancy can be estimated from above by hereditary discrepancy.

THEOREM 4.2 (Lovász–Spencer–Vesztergombi [19]).

$$\operatorname{lindisc}(\mathcal{Y}) \leqq \operatorname{herdisc}(\mathcal{Y}) = \max_{A \subseteq X} \operatorname{disc}(\mathcal{Y}|A),$$

where $\mathcal{Y}|A = \{Y \cap A \mid Y \in \mathcal{Y}\}$ is a set system on the underlying set A.

In order to prove Theorem B, in this section we extend a result of Beck ([4, Lemma 6.2]) to "weighted" set-systems, where the sets may contain their elements with positive or negative integer multiplicities. Therefore we introduce the notion of vector-system. Since an arbitrary subset of the underlying set X may be identified with a 0–1 vector of length p, the following definition seems to be natural. On a vector-system on the underlying set $X = \{1, \dots, p\}$ we mean a finite set $\mathcal{Y}$ of real vectors of length p. The ith coordinate of the vector $Y \in \mathcal{Y}$ we denote by $Y(i)$. We may regard every

set-system to a vector-system this way, and we can adopt the notion of discrepancy. In this context the discrepancy of the vector-system $\mathcal{Y}$ is

$$\operatorname{disc}(\mathcal{Y}) = \min_{f:X\to\{-1,1\}} \max_{Y\in\mathcal{Y}} \left| \sum_{i=1}^{p} Y(i)f(i) \right| .$$

If we regard the elements of the vector-system $\mathcal{Y}$ to the row vectors of a matrix, then this notion of discrepancy is more or less the same as the discrepancy of matrices introduced by Lovász, Spencer and Vesztergombi [19]. If the matrix $\mathbf{A}$ has p columns, then its discrepancy is

$$\operatorname{disc}(\mathbf{A}) = \min_{\mathbf{x}\in\{0,1\}^p} \|\mathbf{A}(\mathbf{x}-\mathbf{c})\|_\infty$$

where $\mathbf{c}$ is the vector $(\frac{1}{2},\dots,\frac{1}{2})$. The difference is only a multiplicative factor 2 arising from the difference between $\{0,1\}$- and $\{-1,1\}$-colourings. We have to note that the analogue of Theorem 4.2 for matrices was proved in [19] (the difference is a multiplicative factor on the right-hand side of the inequality), and therefore it is valid for vector-systems, too, if we define the linear and hereditary discrepancy of the vector-system $\mathcal{Y}$ by

$$\operatorname{lindisc}(\mathcal{Y}) = \max_{\alpha_1,\dots,\alpha_p\in[0,1]} \min_{f:X\to\{0,1\}} \max_{Y\in\mathcal{Y}} \left| \sum_{i=1}^{p} Y(i)(f(i)-\alpha_i) \right|$$

and

$$\operatorname{herdisc}(\mathcal{Y}) = \max_{A\subseteq X} \operatorname{disc}(\mathcal{Y}|A) ,$$

respectively, where the restriction of $\mathcal{Y}$ to $A=\{i_1,\dots,i_k\}\subseteq X$ is

$$\mathcal{Y}|A = \{(Y(i_1),\dots,Y(i_k)) \mid Y\in\mathcal{Y}\}.$$

Before stating our result, we have to introduce some more terminology. Let $\mathcal{Z}$ be a set-system, and denote by $\mathcal{Z}(k)$ the set of those vectors Y that can be written as the signed sum of at most k elements of $\mathcal{Z}$, i.e. there exist sets(=vectors) $Z_1,\dots,Z_t\in\mathcal{Z}$ $(t\leqq k)$ and signs $\varepsilon_1,\dots,\varepsilon_t\in\{-1,1\}$ such that $Y=\sum_{i=1}^{t}\varepsilon_i Z_i$. The vector $Y_1\in\mathcal{Z}(k)$ is said to be contained in the vector $Y_2\in\mathcal{Z}(k)$ $(Y_1\subseteqq Y_2)$, if $Y_1=\sum_{i\in I}\varepsilon_i Z_i$ and $Y_2=\sum_{i\in J}\varepsilon_i Z_i$, where $I\subseteqq J$, $|J|\leqq k$. For $\mathcal{Y}\subseteqq\mathcal{Z}(k)$ define the vector-system

$$\mathcal{Y}\downarrow = \{H\in\mathcal{Z}(k) \mid \exists Y\in\mathcal{Y},\ H\subseteqq Y\} .$$

THEOREM 4.3. *Let there be given a vector-system $\mathcal{Y}$ and a set-system $\mathcal{Z}$ on the underlying set $X=\{1,\dots,p\}$ such that $\mathcal{Y}\subseteqq\mathcal{Z}(k)$. With notations $d=\deg(\mathcal{Z})$, $q=|\mathcal{Y}|$ and $y=\max_{Y\in\mathcal{Y}\downarrow}\|Y\|_\infty$ we have*

$$\operatorname{disc}(\mathcal{Y}) \ll \sqrt{ykd\log(d+2)\log(q+2)}\log(p+2).$$

If we restrict the vector-systems $\mathcal{Y}$ and $\mathcal{Z}$ to an arbitrary subset A of X, then the condition of Theorem 4.3 is hereditary, and the quantities d, q and y do not increase. Therefore we can use Theorem 4.2 (completed with our previous note) to see that

$$\operatorname{lindisc}(\mathcal{Y}) \ll \sqrt{ykd\log(d+2)\log(q+2)}\log(p+2)\ .$$

Thus we obtain

COROLLARY 4.4. *Let $\mathcal{Y}$ and $\mathcal{Z}$ as in Theorem 4.3. Then, for any real number $0 \leqq \alpha \leqq 1$ there exists a function $f : X \longrightarrow \{1-\alpha, -\alpha\}$ satisfying*

$$\left|\sum_{i=1}^{p} Y(i)f(i)\right| \ll \sqrt{ykd\log(d+2)\log(q+2)}\log(p+2)$$

for every $Y \in \mathcal{Y}$.

The remaining part of this section is devoted to the proof of Theorem 4.3. We will need the following Chernoff[10]-type inequality to handle certain sums of binomial coefficients. Although this inequality is known in more general forms (e.g. Hoeffding [15, Theorem 2] or McDiarmid [20, Lemma 1.2]), let us present here a simple proof.

LEMMA 4.5. *Let $X_1, \ldots X_p$ be independent random variables with common distribution*

$$\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = \frac{1}{2}\ .$$

Let $\gamma > 0$, and let $\varepsilon_1, \ldots, \varepsilon_p$ be arbitrary real numbers. Then

$$\mathbf{P}\left(\left|\sum_{i=1}^{p} \varepsilon_i X_i\right| \geqq \gamma\right) \leqq 2\exp\left(\frac{-\gamma^2}{2\sum_{i=1}^{p}\varepsilon_i^2}\right).$$

PROOF. Let $\sum_{i=1}^{p} \varepsilon_i X_i = S$, then obviously $\mathbf{P}(|S| \geqq \gamma) = 2\mathbf{P}(S \geqq \gamma)$. For an arbitrary real parameter $\alpha > 0$ we have

$$\mathbf{P}(S \geqq \gamma) = \mathbf{P}(e^{\alpha S} \geqq e^{\alpha\gamma}) \leqq e^{-\alpha\gamma}\mathbf{E}(e^{\alpha S})\ ,$$

where $\mathbf{E}$ denotes expectation. As the random variables $X_1, \ldots, X_p$ are independent, we can write

$$\mathbf{E}(e^{\alpha S}) = \mathbf{E}(\prod_{i=1}^{p} e^{\alpha\varepsilon_i X_i}) = \prod_{i=1}^{p}\mathbf{E}(e^{\alpha\varepsilon_i X_i}) = \prod_{i=1}^{p}(\tfrac{1}{2}(e^{\alpha\varepsilon_i} + e^{-\alpha\varepsilon_i})).$$

Using inequality $\frac{1}{2}(e^y + e^{-y}) \leqq e^{\frac{1}{2}y^2}$, that may be checked immediately comparing Taylor-series, we obtain

$$\mathbf{P}(S \geqq \gamma) \leqq e^{-\alpha\gamma}\prod_{i=1}^{p} e^{\frac{1}{2}\alpha^2\varepsilon_i^2} = \exp(\tfrac{1}{2}\alpha^2\textstyle\sum_{i=1}^{p}\varepsilon_i^2 - \alpha\gamma)\ .$$

We can optimize the estimate choosing $\alpha = \gamma / \sum_{i=1}^{p} \varepsilon_i^2$, and the lemma follows immediately.

To prove Theorem 4.3 let $\mathcal{Z} = \mathcal{Z}^* \cup \mathcal{Z}^{**}$, where

$$\mathcal{Z}^* = \{Z \in \mathcal{Z} \mid |Z| < 100d \log d\}$$

and

$$\mathcal{Z}^{**} = \{Z \in \mathcal{Z} \mid |Z| \geqq 100d \log d\}.$$

Every element of $\mathcal{Y} = \{Y_1, \ldots, Y_q\}$ can be written in the form $Y_i = \sum_{j=1}^{t} \varepsilon_{ij} Z_{ij}$, where $t \leqq k$ and $\varepsilon_{ij} \in \{-1, 1\}$, $Z_{ij} \in \mathcal{Z}$. According to this decomposition, write $Y_i = Y_i^* + Y_i^{**}$, where

$$Y_i^* = \sum_{Z_{ij} \in \mathcal{Z}^*} \varepsilon_{ij} Z_{ij} \qquad \text{and} \qquad Y_i^{**} = \sum_{Z_{ij} \in \mathcal{Z}^{**}} \varepsilon_{ij} Z_{ij} \ .$$

Clearly $Y_i^*, Y_i^{**} \in \mathcal{Z}(k)$.

Our aim is to find a "partial" two-colouring $g : X \longrightarrow \{-1, 0, 1\}$ that have relatively small discrepancy on the vectors Y_i^*, more precisely

$$\left| \sum_{j=1}^{p} Y_i^*(j) g(j) \right| \ll \sqrt{ykd \log(d+2) \log(q+2)}$$

for every vector $Y_i \in \mathcal{Y}$;

that colours perfectly the vectors Y_i^{**}, i.e. for any $Y_i \in \mathcal{Y}$

$$\sum_{j=1}^{p} Y_i^{**}(j) g(j) = 0 \ ;$$

and that colours a positive percent (10%) of the coordinates really, which means

$$|\{i \in X \mid g(i) \neq 0\}| \geqq \frac{p}{10} \ .$$

If we can prove the existence of such a function g, then we may restrict the vector-systems $\mathcal{Y}$ and $\mathcal{Z}$ to the set $\{i \in X \mid g(i) = 0\}$ of noncoloured points. Now we are in the same position as in the beginning of the proof, and hence we can colour at least 10% of the remaining coordinates the same way. Repeating this procedure at most $\ll \log n$ times, we can colour each point of the underlying set X with a colour -1 or 1, and the theorem follows.

As the first step, consider the set $\mathcal{F}$ of two-colourings $f : X \longrightarrow \{-1, 1\}$, then $|\mathcal{F}| = 2^p$.

PROPOSITION 4.6. *Let*

$$\mathcal{F}' = \left\{ f \in \mathcal{F} \;\middle|\; \left|\sum_{j=1}^{p} Y_i^*(j) f(j)\right| < \kappa\sqrt{ykd\log(d+2)\log(q+2)},\ \forall\ Y_i \in \mathcal{Y} \right\} ,$$

where κ is a sufficiently large absolute constant, then $|\mathcal{F}'| > 2^{p-1}$.

PROOF. Colour the points of X independently of each other by the colours -1 and 1 with probability $\frac{1}{2} - \frac{1}{2}$. Let $f : X \longrightarrow \{-1, 1\}$ denote a random two-colouring of X. For a fixed vector $Y_i \in \mathcal{Y}$ we can apply Lemma 4.5 to obtain estimate

$$\mathbf{P}\left(\left|\sum_{j=1}^{p} Y_i^*(j) f(j)\right| \geqq \kappa\sqrt{ykd\log(d+2)\log(q+2)}\right) \leqq$$
$$\leqq 2\exp\left(\frac{-\kappa^2 ykd\log(d+2)\log(q+2)}{2\sum_{j=1}^{p}(Y_i^*(j))^2}\right) .$$

Since $Y_i^* \in \mathcal{Z}(k)$, by the definition of y we have $|Y_i^*(j)| \leqq y$ for every $1 \leqq i \leqq q$, $1 \leqq j \leqq p$. By the definition of Y_i^* we get

$$\sum_{j=1}^{p} |Y_i^*(j)| \leqq \sum_{Z_{ij} \in \mathcal{Z}^*} |Z_{ij}| < 100kd\log(d+2) .$$

We can summarize these observations in

$$\mathbf{P}\left(\left|\sum_{j=1}^{p} Y_i^*(j) f(j)\right| \geqq \kappa\sqrt{ykd\log(d+2)\log(q+2)}\right) \leqq$$
$$\leqq 2\exp\left(\frac{-\kappa^2\log(q+2)}{200}\right) < 2(q+2)^{-\kappa^2/200} .$$

Therefore we have

$$|\mathcal{F}'| \geqq 2^p(1 - 2q(q+2)^{-\kappa^2/200}) > 2^{p-1} ,$$

if κ is large enough.

Now we turn to the coloration of vectors Y_i^{**}. It is enough to desire relation

$$\sum_{j=1}^{p} Z(j) g(j) = 0$$

to hold for every set $Z \in \mathcal{Z}^{**}$, here we have $Z(j) \in \{0,1\}$. Let $\mathcal{Z}^{**} = \{Z_1, \dots, Z_M\}$. Associate with each two-coloration $f \in \mathcal{F}$ the vector

$$\mathbf{v}(f) = (v_1(f), \dots, v_M(f)) \in \mathbb{Z}^M ,$$

where

$$v_i(F) = \sum_{j=1}^{p} Z_i(j) f(j) = \sum_{j \in Z_i} f(j) .$$

PROPOSITION 4.7. $|\{\mathbf{v}(f) \mid f \in \mathcal{F}\}| < 2^{p/5}$.

PROOF. Count the number of values the ith coordinates of the vectors $\mathbf{v}(f)$ may have. Clearly

$$|v_i(f)| \leqq |Z_i| \qquad \text{and} \qquad v_i(f) \equiv |Z_i| \pmod 2 ,$$

so this number satisfy

$$|\{v_i(f) \mid f \in \mathcal{F}\}| \leqq |Z_i| + 1 \leqq 2|Z_i| ,$$

because $\emptyset \notin \mathcal{Z}^{**}$.

It is well-known that $\ln \beta \leqq \beta - 1$ for $\beta > 0$. Thus for arbitrary $\alpha > 1$ inequalities $\ln \alpha \leqq \frac{\alpha}{e}$, $\alpha \leqq e^{\alpha/e}$ hold, and so do they for $0 \leqq \alpha \leqq 1$. Therefore we can estimate as

$$|\{\mathbf{v}(f) \mid f \in \mathcal{F}\}| \leqq \prod_{i=1}^{M} 2|Z_i| = t^M \prod_{i=1}^{M} \left(\frac{2|Z_i|}{t}\right) \leqq t^M \exp\left(\sum_{i=1}^{M} \frac{2|Z_i|}{et}\right) ,$$

where the value of the positive parameter t we will fix later. As

$$\sum_{i=1}^{M} |Z_i| \leqq \sum_{Z \in \mathcal{Z}} |Z| \leqq pd ,$$

we can obtain

$$M \leqq \frac{1}{100d \log(d+2)} \sum_{i=1}^{M} |Z_i| \leqq \frac{p}{100 \log(d+2)} .$$

Finally choose $t = 10d$, then

$$|\{\mathbf{v}(f) \mid f \in \mathcal{F}\}| \leqq (10d)^{p/100 \log(d+2)} e^{2p/10e} =$$
$$= \exp\left\{p\left(\frac{\ln 100 + \ln d}{100 \log d} + \frac{1}{5e}\right)\right\} < e^{p/10} < 2^{p/5} .$$

Comparing Propositions 4.6 and 4.7 one can see the existence of a subset $\mathcal{F}'' \subseteq \mathcal{F}'$, that satisfy $|\mathcal{F}''| \geq 2^{4p/5-1}$ and $\mathbf{v}(f_1) = \mathbf{v}(f_2)$ for any $f_1, f_2 \in \mathcal{F}''$. Fixing an element f_1 of $\mathcal{F}''$, the family

$$\mathcal{G} = \{\tfrac{1}{2}(f - f_1) \mid f \in \mathcal{F}''\}$$

is consisting of partial two-colourings $g : X \longrightarrow \{-1, 0, 1\}$ that satisfy our first two conditions. Let $\mathcal{G}'$ be the family of those colourings $g : X \longrightarrow \{-1, 0, 1\}$ for which

$$|\{i \in X \mid g(i) = 0\}| \geq \tfrac{9p}{10} .$$

It is enough to show that $|\mathcal{G}'| < 2^{4p/5-1}$. Since

$$|\mathcal{G}'| = \sum_{i=0}^{[p/10]} \binom{p}{i} 2^i < 2^{p/10} \sum_{i=0}^{[p/10]} \binom{p}{i} ,$$

we can apply Lemma 4.5 again, in the special case $\varepsilon_1, \ldots, \varepsilon_p = 1$, with $\gamma = \frac{4p}{10}$:

$$|\mathcal{G}'| < 2^{p/10} 2^p e^{-8p/25} < 2^{p/10+1-2p \log e/25} < 2^{4p/5-1} .$$

This completes the proof of Theorem 4.3.

5. Proofs of Theorems A and B

PROOF OF THEOREM A. For an arbitrary positive integer t let us define the cube $C_t = [-M, M]^K$, where $M = 2^t$, and the finite set of points

$$\mathcal{P}_t = \left\{ \left(\frac{a_1}{M^{K-1}}, \frac{a_2}{M^{K-1}}, \ldots, \frac{a_K}{M^{K-1}} \right) + \mathbf{v}_t \;\middle|\; -M^K \leq a_i \leq M^K,\ a_i \in \mathbb{Z} \right\} ,$$

where the vector $\mathbf{v}_t \in [0, 1]^K$ is to be fixed later. Let $\alpha = M^{-K(K-1)}$, and for an arbitrary function $f_t : \mathcal{P}_t \longrightarrow \{-\alpha, 1 - \alpha\}$ define

$$\mathcal{P}_t(f_t) = \{x \in \mathcal{P}_t \mid f_t(x) = 1 - \alpha\}.$$

LEMMA 5.1. *For any convex polytope $B \subseteq C_t$,*

$$D(\mathcal{P}_t(f_t), B) = |Z(\mathcal{P}_t(f_t), B) - \mu(B)| < \Big| \sum_{x \in B \cap \mathcal{P}_t} f_t(x) \Big| + c_1(K).$$

PROOF.

$$D(\mathcal{P}_t(f_t), B) \leq \left| \sum_{x \in B \cap \mathcal{P}_t(f_t)} 1 - \alpha \sum_{x \in B \cap \mathcal{P}_t} 1 \right| + \alpha \left| \sum_{x \in B \cap \mathcal{P}_t} 1 - \alpha^{-1} \mu(B) \right| =$$

$$= \left| \sum_{x \in B \cap \mathcal{P}_t} f_t(x) \right| + M^{-K(K-1)} \left| \sum_{x \in \widetilde{B} \cap \widetilde{\mathcal{P}}_t} 1 - \mu(\widetilde{B}) \right| ,$$

where $\widetilde{B} = M^{K-1}B$ and $\widetilde{\mathcal{P}}_t = M^{K-1}\mathcal{P}_t$.

By a standard averaging argument,

$$\left| \sum_{x \in \widetilde{B} \cap \widetilde{\mathcal{P}}_t} 1 - \mu(\widetilde{B}) \right| < \sqrt{K}\mu(\partial \widetilde{B}) + c_2(K),$$

where $c_2(K)$ is the volume of a ball of radius $\sqrt{K}$. Finally we have

$$\mu(\partial \widetilde{B}) = (M^{K-1})^{K-1}\mu(\partial B) \leqq M^{(K-1)(K-1)}\mu(\partial C_t) = c_3(K)M^{K(K-1)},$$

and the assertion follows.

We will construct the set $\mathcal{Q}$ in terms of the sets $\mathcal{P}_t(f_t)$ with suitable functions $f_t : \mathcal{P}_t \longrightarrow \{-\alpha, 1-\alpha\}$. For the sake of simplicity, from now on we will assume that $\mathcal{A}$ contains the coordinate hyperplanes $H_i = \{\mathbf{x} \in \mathbb{R}^K \mid x_i = 0\}$ for $1 \leqq i \leqq K$. Note that it implies immediately that $\mathrm{POL}(\mathcal{A})$ is not empty.

LEMMA 5.2. *There exists a function $f_t : \mathcal{P}_t \longrightarrow \{-\alpha, 1-\alpha\}$ such that for every convex polytope $B \in \mathrm{SPEC}(\mathcal{A}, H)$ satisfying $\mu(B) > 1$, we have*

$$\left| \sum_{x \in B \cap \mathcal{P}_t} f_t(x) \right| \ll_{\mathcal{A},H,\varepsilon} (\log d(B) + 1)^{K+\varepsilon}.$$

PROOF. Define an equivalence relation on the set of special polytopes $\mathrm{SPEC}(\mathcal{A}, H)$ as follows. The polytopes $P_1, P_2 \in \mathrm{SPEC}(\mathcal{A}, H)$ are equivalent if and only if there exists a simplex $\Delta \in \mathrm{POL}(H_1, \ldots, H_l, H)$, a linear transformation $\phi \in \Phi(\Delta)$, a polytope $E = E(d_1, \ldots, d_k; 2^{n_1}, \ldots, 2^{n_k}) \in \mathrm{BSP}(\mathbb{R}^K)$ and vectors $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^K$ of the form

$$\mathbf{w}_\nu = (l_{11}^\nu 2^{n_1}, \ldots, l_{i1}^\nu 2^{n_i}, \ldots, l_{id_i}^\nu 2^{n_i}, \ldots, l_{kd_k}^\nu 2^{n_k}) \quad (\nu = 1, 2)$$

where $l_{ij}^\nu \in \mathbb{Z}$, so that $P_\nu = \phi(E + \mathbf{w}_\nu)$ for $\nu \in \{1, 2\}$.

Denote by $\mathrm{SET}(\mathcal{A}, H)$ the set of the classes of this equivalence relation. For $\mathcal{C} \in \mathrm{SET}(\mathcal{A}, H)$ we may define $\mu(\mathcal{C})$ and $d(\mathcal{C})$ as the common volume and diameter of the elements of $\mathcal{C}$, respectively. We may define a linear ordering $\preceq$ on the set

$$\{\mathcal{C} \in \mathrm{SET}(\mathcal{A}, H) \mid \mu(\mathcal{C}) > 1\}$$

by $\mathcal{C}_1 \prec \mathcal{C}_2$ if $d(\mathcal{C}_1) < d(\mathcal{C}_2)$, with the convention that the ordering is defined arbitrarily in the case of equal diameters.

Observe that for $B = E(d_1, \ldots, d_k; 2^{n_1}, \ldots, 2^{n_k}) \in \mathrm{BSP}(\mathbb{R}^K)$, the conditions $\mu(B) > 1$ and $d(B) < y$ for some $y > 1$ imply

$$y^{-K+1} < 2^{n_i} < y \ , \ -(K-1)\log y < n_i < \log y.$$

Therefore, with $H_0 = \{\mathbf{x} \in \mathbb{R}^K \mid x_1 + \cdots + x_K = 1\}$, we have

$$|\{\mathcal{C} \in \mathrm{SET}(H_1, \ldots, H_K, H_0) \mid \mu(\mathcal{C}) > 1,\ d(\mathcal{C}) < y\}| \ll_K (\log y)^K,$$

and finally

$$|\{\mathcal{C} \in \mathrm{SET}(\mathcal{A}, H) \mid \mu(\mathcal{C}) > 1,\ d(\mathcal{C}) < y\}| \ll_{\mathcal{A},H} (\log y + 1)^K.$$

Choose the vector $\mathbf{v}_t$ so that no point of $\mathcal{P}_t$ lie on the boundary of any special polytope $P \in \mathrm{SPEC}(\mathcal{A}, H)$ (every vector is allowed except of a set of measure 0). Each family $\mathcal{C} \in \mathrm{SET}(\mathcal{A}, H)$ defines a partition of $\mathcal{P}_t$:

$$\mathcal{P}_t = ((\mathbb{R}^K \setminus \cup_{B \in \mathcal{C}} B) \cap \mathcal{P}_t) \cup^* \bigcup_{B \in \mathcal{C}}^* (B \cap \mathcal{P}_t).$$

Let $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \ldots$ be the partitions of $\mathcal{P}_t$ defined by the families in

$$\{\mathcal{C} \in \mathrm{SET}(\mathcal{A}, H) \mid \mu(\mathcal{C}) > 1\}$$

ordered in $\preceq$. If we apply Lemma 4.1 to the set $X = \mathcal{P}_t$ and its partitions $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \ldots$ with $\alpha_1 = \alpha_2 = \cdots = \alpha$ and $\varrho = K^{-1}\varepsilon$, then the result follows immediately.

Observe that there exists a positive constant $c_4(K)$ such that $\mu(B) > 1$ implies $d(B) > 1 + c_4(K)$ for every convex body in $\mathbb{R}^K$. Therefore, using Lemma 5.1 we have

$$D(\mathcal{P}_t(f_t), B \cap C_t) \ll_{\mathcal{A},H,\varepsilon} (\log d(B) + 1)^{K+\varepsilon}$$

for every convex polytope $B \in \mathrm{SPEC}(\mathcal{A}, H)$ with $\mu(B) > 1$.

The proof of the following simple lemma can be found e.g. in Beck and Chen [7] or [8].

LEMMA 5.3. *Suppose that $A, P_1, \ldots, P_m, Q_1, \ldots, Q_n$ are measurable subsets of $\mathbb{R}^K$ and $\gamma_1, \ldots, \gamma_m, \delta_1, \ldots, \delta_n$ are ± 1 signs such that*

$$\sum_{i=1}^{m} \gamma_i \chi_{P_i} \leqq \chi_A \leqq \sum_{j=1}^{n} \delta_j \chi_{Q_j},$$

and

$$\sum_{j=1}^{n} \delta_j \mu(Q_j) - \sum_{i=1}^{m} \gamma_i \mu(P_i) \leqq D_1.$$

Suppose further that $\mathcal{P}$ is a discrete subset of $\mathbb{R}^K$ such that

$$\max\{|D(\mathcal{P}, P_i)|, |D(\mathcal{P}, Q_j)|\} \leqq D_2.$$

Then

$$|D(\mathcal{P}, A)| \leqq D_1 + D_2 \max\{m, n\}.$$

Let $A \in \mathrm{POL}(\mathcal{A})$, $A \subseteqq C_t$. First apply Lemma 3.11 to obtain the approximating polytopes $P_1, \ldots, P_m, Q_1, \ldots, Q_n \in \mathrm{SPEC}(\mathcal{A}, H)$, then apply Lemma 5.3 replacing P_i and Q_j by $P_i \cap C_t$ and $Q_j \cap C_t$, respectively. Then we gain the estimate

$$D(\mathcal{P}_t(f_t), A) \ll_{\mathcal{A},H,\varepsilon} (\log(d(A)+2))^{3K-1+\varepsilon}.$$

Now we are in a position to construct the infinite set $\mathcal{Q}$. First we write

$$\mathbb{R}^K = C_1 \cup^* \bigcup_{n=1}^{\infty}{}^*(C_{2^n} \setminus C_{2^{n-1}})$$

and observe that $C(n) = C_{2^n} \setminus C_{2^{n-1}}$ is the disjoint union of $2K$ aligned boxes $B_{n1}, \ldots, B_{n2K}$. Define

$$\mathcal{Q} = \mathcal{P}_1(f_1) \cup^* \bigcup_{n=1}^{\infty}{}^*(\mathcal{P}_{2^n}(f_{2^n}) \cap C(n)).$$

Consider an arbitrary convex polytope $A \in \mathrm{POL}(\mathcal{A})$. Then $A_n = A \cap C(n)$ is the disjoint union of the (occasionally empty) convex polytopes $A \cap B_{ni}$, $1 \leqq i \leqq 2K$. As the coordinate hyperplanes $H_1, \ldots, H_K \in \mathcal{A}$, these polytopes are elements of $\mathrm{POL}(\mathcal{A})$. Thus we have

$$D(\mathcal{Q}, A) = |Z(\mathcal{Q}, A) - \mu(A)| \leqq$$

$$\leqq D(\mathcal{P}_1(f_1), A \cap C_1) + \sum_{n=1}^{\infty}\sum_{i=1}^{2K} D(\mathcal{P}_{2^n}(f_{2^n}), A \cap B_{ni}) \ll_{\mathcal{A},H,\varepsilon}$$

$$\ll_{\mathcal{A},H,\varepsilon} \sum_{A_n \neq \emptyset} (\min\{\log(d(A)+2)), n\})^{3K-1+\varepsilon} \ll (\log(d(A)+2))^{3K-1+\varepsilon},$$

as it was to be proved.

Proof of Theorem B. With a refinement of our argument we can decrease the exponent $3K - 1 + \varepsilon$ in the finite version. We note that this modified argument does not work for the infinite version. Indeed, in the case when a small polytope is located far from the origin, we could give an upper bound only in the function of its distance from the origin.

Let $M = N^{1/K}$ and consider the finite set of points

$$X = \left\{ \left(\frac{a_1}{M^{K-1}}, \frac{a_2}{M^{K-1}}, \ldots, \frac{a_K}{M^{K-1}} \right) + \mathbf{v} \;\middle|\; 0 \leqq a_i < M^K,\ a_i \in \mathbb{Z} \right\}$$

contained in the cube $C = [0, M)^K$, where choosing the vector $\mathbf{v} \in [0,1)^K$ (or more precisely $\mathbf{v} \in [0, \{M\})^K$, if M is not an integer) we have the same

requirements than in the previous theorem. Let $\alpha = M^{-K(K-1)}$ again, and define

$$\mathcal{P}' = \mathcal{P}'(f) = \{x \in X \mid f(x) = 1 - \alpha\}$$

for an arbitrary function $f : X \longrightarrow \{-\alpha, 1-\alpha\}$. The following lemma is the analogue of Lemma 5.1.

LEMMA 5.4. *Let* $Y = \sum_{i=1}^{k} \varepsilon_i Z_i$, *where* $\varepsilon_i \in \{-1, 1\}$ *and* $Z_i \subseteqq C$ *is a convex polytope* $(1 \leqq k)$. *Then*

$$D(\mathcal{P}', Y) = \left|\sum_{i=1}^{k} \varepsilon_i (Z(\mathcal{P}', Z_i) - \mu(Z_i))\right| < \left|\sum_{i=1}^{k} \varepsilon_i \sum_{x \in Z_i \cap X} f(x)\right| + k c_1(K).$$

PROOF.

$$\begin{aligned} D(\mathcal{P}', Y) &\leqq \left|\sum_{i=1}^{k} \varepsilon_i Z(\mathcal{P}', Z_i) - \alpha \sum_{i=1}^{k} \varepsilon_i Z(X, Z_i)\right| + \\ &+ \alpha \left|\sum_{i=1}^{k} \varepsilon_i (Z(X, Z_i) - \alpha^{-1} \mu(Z_i)\right| \leqq \\ &\leqq \left|\sum_{i=1}^{k} \varepsilon_i \sum_{x \in Z_i \cap X} f(x)\right| + \sum_{i=1}^{k} \alpha |Z(X, Z_i) - \alpha^{-1} \mu(Z_i)| \,, \end{aligned}$$

and $\alpha |Z(X, Z_i) - \alpha^{-1} \mu(Z_i)| < c_1(K)$ can be proved in the same way as in Lemma 5.1.

Instead of using Lemma 5.2 we can follow the next method. Introduce the set-system

$$\mathcal{Z} = \{B \cap X \mid B \in \mathrm{SPEC}(\mathcal{A}, H),\ B \subseteqq [-c_5 M, c_5 M]^K,\ \mu(B) \geqq 1\}$$

where $c_5 = c_5(\mathcal{A}, H)$ is a sufficiently large positive constant. Then

$$d = \deg \mathcal{Z} \ll_{\mathcal{A}, H} (\log M + 1)^K \,,$$

since our conditions imply that the lengths of the edges of every polytope B are between $c_6(\mathcal{A}, H) M^{-K+1}$ and $c_7(\mathcal{A}, H) M$.

For a significant reduction of the size of the vector-system $\mathcal{Y}$ we are to introduce it is worth considering the next fact. Let

$$\mathcal{A}_0 = \{A \in \mathrm{POL}(\mathcal{A}) \mid A \subseteqq C\} \,.$$

LEMMA 5.5. *There exists a family of convex polytopes* $\mathcal{A}_1 \subseteq \mathcal{A}_0$, $|\mathcal{A}_1| \ll_{K,l} M^{2Kl}$ *such that for every convex polytope* $A \in \mathcal{A}_0$ *one can find polytopes* $A_1, A_2 \in \mathcal{A}_1$ *satisfying* $A_1 \subseteq A \subseteq A_2$ *and* $\mu(A_2 \setminus A_1) \leqq 1$.

PROOF. For every given hyperplane $G \in \mathcal{A}$ put parallels to G with equal distances $1/(c_8(K)M^{K-1})$. The number of the hyperplanes of a fixed direction obtained this way and cutting the cube C is at most $(\sqrt{K}+1)c_8(K)M^K$. Since having each hyperplane two sides, they determine at most

$$\left(\left(\sqrt{K}+1\right)c_8(K)M^K\right)^{2l} \ll_{K,l} M^{2Kl}$$

different convex polytopes that are contained in $\mathrm{POL}(\mathcal{A})$. Cutting these polytopes with the cube C we obtain a desired family of convex polytopes $\mathcal{A}_1$. Indeed, if $A \in \mathcal{A}_0$, then each facet of A lies between two neigbouring hyperplanes parallel to it. Therefore there exist convex polytopes $A_1, A_2 \in \in \mathcal{A}_1$, $A_1 \subseteq A \subseteq A_2$ such that

$$\mu(A_2 \setminus A_1) < \mu(\partial A_2)\frac{1}{c_8(K)M^{K-1}} .$$

Since $A_2 \subseteq C$ are convex polytopes, $\mu(\partial A_2) \leqq \mu(\partial C)$, and thus

$$\mu(A_2 \setminus A_1) < \mu(\partial C)\frac{1}{c_8(K)M^{K-1}} = \frac{2K}{c_8(K)} \leqq 1 ,$$

if $c_8(K) \geqq 2K$.

Assume for technical reasons, that $C \in \mathcal{A}_1$ (we will use this fact at the end of the proof). We can define the vector-system $\mathcal{Y}$ the following way. Let us associate to an arbitrary convex polytope $A \in \mathcal{A}_1$ the signed sums $A' = \sum_{i=1}^m \gamma_i P_i$ and $A'' = \sum_{i=1}^n \delta_i Q_i$ according to Lemma 3.11, and let

$$A' \cap X = \sum \gamma_i (P_i \cap X), \quad A'' \cap X = \sum_{i=1}^n \delta_i (Q_i \cap X) .$$

Create the vector-system $\mathcal{Y} = \{A' \cap X,\ A'' \cap X \mid A \in \mathcal{A}_1\}$. Taking into consideration assertion 3) of Lemma 3.11, $\mathcal{Y} \subseteq \mathcal{Z}(k)$, where $k \ll_{K,l} (\log M + 1)^{2K-1}$, if the constant c_5 used in the definition of $\mathcal{Z}$ is large enough. On the basis of assertion 6) of Lemma 3.11 we have the estimate

$$y = \max_{Y \in \mathcal{Y}} \|Y\|_\infty \ll_{K,l} 1 .$$

Therefore we can apply Corollary 4.4 to show the existence of a function $f : X \longrightarrow \{-\alpha, 1-\alpha\}$ satisfying

$$\left|\sum_{i=1}^p Y(i)f(i)\right| \ll_{\mathcal{A},H,\varepsilon} (\log M + 1)^{\frac{1}{2}K+1+\varepsilon}$$

for every $Y \in \mathcal{Y}$, since in the present situation $p = |X| = \lceil M \rceil^{K^2}$ and $q = |\mathcal{Y}| \leqq 2|\mathcal{A}_1| \ll_{K,l} M^{2Kl}$.

Comparing this result to Lemma 5.4 we get the estimate

$$D(\mathcal{P}', A') < \left| \sum_{i=1}^{m} \gamma_i \sum_{x \in P_i \cap X} f(x) \right| + mc_1(K) =$$

$$= \left| \sum_{i=1}^{p} (A' \cap X)(i) f(i) \right| + mc_1(K) \ll_{\mathcal{A},H,\varepsilon}$$

$$\ll_{\mathcal{A},H,\varepsilon} \max\{(\log M + 1)^{\frac{3}{2}K+1+\varepsilon},\ (\log M + 1)^{2K-1}\}$$

for every convex polytope $A \in \mathcal{A}_1$, and we have the same upper bound for $D(\mathcal{P}', A'')$, too. Thus we obtain

$$Z(\mathcal{P}', A) - \mu(A) \leqq \sum_{i=1}^{n} \delta_i Z(\mathcal{P}', Q_i) - \sum_{i=1}^{m} \gamma_i \mu(P_i) \leqq$$

$$\leqq D(\mathcal{P}', A'') + \left(\sum_{i=1}^{n} \delta_i \mu(Q_i) - \sum_{i=1}^{m} \gamma_i \mu(P_i)\right) \ll_{\mathcal{A},H,\varepsilon}$$

$$\ll_{\mathcal{A},H,\varepsilon} (\log M + 1)^{\max\{\frac{3}{2}K+1+\varepsilon, 2K-1\}}$$

for every convex polytope $A \in \mathcal{A}_1$. Estimating from below in the same way,

$$Z(\mathcal{P}', A) - \mu(A) \geqq \sum_{i=1}^{m} \gamma_i Z(\mathcal{P}', P_i) - \sum_{i=1}^{n} \delta_i \mu(Q_i) \geqq$$

$$\geqq -D(\mathcal{P}', A') - \left(\sum_{i=1}^{n} \delta_i \mu(Q_i) - \sum_{i=1}^{m} \gamma_i \mu(P_i)\right)$$

shows that

$$D(\mathcal{P}', A) = |Z(\mathcal{P}', A) - \mu(A)| \ll_{\mathcal{A},H,\varepsilon} (\log M + 1)^{\max\{\frac{3}{2}K+1+\varepsilon, 2K-1\}} .$$

Finally, because of Lemmas 5.5 and 5.3, it follows

$$D(\mathcal{P}', A) \ll_{\mathcal{A},H,\varepsilon} (\log M + 1)^{\max\{\frac{3}{2}K+1+\varepsilon, 2K-1\}} \ll_{\mathcal{A},H,\varepsilon}$$

$$\ll_{\mathcal{A},H,\varepsilon} (\log N)^{\max\{\frac{3}{2}K+1+\varepsilon, 2K-1\}}$$

for every polytope $A \in \mathcal{A}_0$, if $N > 2$. Since $C \in \mathcal{A}_1$, the same estimation holds for $||\mathcal{P}'| - N| = D(\mathcal{P}', C)$, too. Therefore we may assume that $\mathcal{P}'$ consists of exactly N points. To finish the proof we only have to note that $\mathcal{P}'$ can be transformed into a pointset $\mathcal{P}$ satisfying Theorem B with the help of the reduction of ratio $N^{1/K}$, centered at the origin. If the implicit constant in the theorem is greater than 4, then the assertion clearly holds for $N = 2$, too.

6. Proofs of Theorems C and D

PROOF OF THEOREM C. First choose and fix a polytope $A^* \in \mathrm{POL}(P)$ having a vertex $\mathbf{v}^*$ incident to exactly K facets of A^*. To see that such an A^* exists, consider an arbitrary polytope $A_0 \in \mathrm{POL}(P)$ contained in the interior of $[0,1]^K$. Let $\mathbf{v}_0$ be an arbitrary vertex of A_0, and let $F_1, \dots, F_l$ be the facets of A_0 that contain $\mathbf{v}_0$. If $l = K$, then we are done. Otherwise pushing the facets $F_{K+1}, \dots, F_l$ outwards a bit, we can obtain a desired polytope A^*. Indeed, $\mathbf{v}^* = \mathbf{v}_0$ will be a vertex of A^* contained in exactly K facets $F_1', \dots, F_K'$ of A^*, where F_i' is the facet of A^* that contains the original facet F_i of A_0.

Let $F_1, \dots, F_K$ denote the facets of A^* that contain the vertex $\mathbf{v}^*$, the remaining facets we denote by $F_{K+1}, \dots, F_l$. A^* is the intersection of halfspaces H_i supported by F_i $(i = 1, 2, \dots, l)$, defined by the inequalities, say, $\mathbf{a}_i\mathbf{x} \leqq b_i$. Let $A_{\mathbf{v}^*}$ be the convex hull of the vertices of A^* not incident to the facets $F_1, \dots, F_K$, then $A_{\mathbf{v}^*} \subset \operatorname{int} \bigcap_{i=1}^K H_i$. Therefore there exist positive numbers $\beta_1, \dots, \beta_K$ such that $A_{\mathbf{v}^*} \subset \operatorname{int} \bigcap_{i=1}^K H_i'$, where H_i' is the halfspace defined by the inequality $\mathbf{a}_i\mathbf{x} \leqq b_i - \beta_i$. For every $\mathbf{y} \in [0,1]^K$ we can define a convex polytope $A(\mathbf{y}) = \bigcap_{i=1}^l H_i(\mathbf{y})$, where $H_i(\mathbf{y})$ has defining inequality $\mathbf{a}_i\mathbf{x} \leqq b_i - \beta_i y_i$ if $i \in \{1, 2, \dots, K\}$ and $H_i(\mathbf{y}) = H_i$ otherwise. We will prove the existence of a point $\mathbf{y} \in [0,1]^K$ for which

$$|D(\mathcal{P}, A(\mathbf{y}))| \gg_{A^*} (\log N)^{\frac{K-1}{2}}.$$

We will construct an auxiliary function $F(\mathbf{y}) = F(\mathcal{P}, \mathbf{y})$ satisfying

$$\int_{[0,1]^K} F(\mathbf{x}) D(\mathbf{x}) \mathrm{d}\mathbf{x} \gg_{A^*} (\log N)^{K-1}$$

and

$$\int_{[0,1]^K} F^2(\mathbf{x}) \mathrm{d}\mathbf{x} \ll_K (\log N)^{K-1},$$

where $D(\mathbf{y}) = D(\mathcal{P}, A(\mathbf{y}))$ is the discrepancy function. Then applying Schwarz's inequality we obtain

$$\int_{[0,1]^K} D^2(\mathbf{x}) \mathrm{d}\mathbf{x} \gg_{A^*} (\log N)^{K-1}$$

and Theorem C follows.

We will define the auxiliary function $F(\mathbf{x})$ with the help of the Rademacher functions.

Any $x \in [0,1)$ can be written uniquely in the form

$$x = \sum_{j=0}^{\infty} \beta_j(x) 2^{-j-1},$$

where $\beta_j(x) \in \{0,1\}$ and the sequence $\beta_0, \beta_1, \ldots$ does not end with $1, 1, \ldots$.

For $r = 0, 1, \ldots$ define the r^{th} Rademacher function by

$$R_r(x) = (-1)^{\beta_r(x)}.$$

For a K-tuple $\mathbf{r} = (r_1, \ldots, r_K)$ of non-negative integers set

$$R_{\mathbf{r}}(\mathbf{x}) = R_{r_1}(x_1) \cdots R_{r_K}(x_K)$$

for every $\mathbf{x} = (x_1, \ldots, x_K) \in [0,1]^K$.

By an $\mathbf{r}$-box of the unit cube we mean a set of the form $I_1 \times \cdots \times I_K$, where I_i is an interval of the form

$$I_i = [m_i 2^{-r_i}, (m_i + 1)2^{-r_i})$$

with an integer $m_i \in [0, 2^{r_i})$.

By an $\mathbf{r}$-function we mean a real function $f(\mathbf{x})$ defined on $[0,1)^K$ satisfying $f = R_{\mathbf{r}}$ or $f = -R_{\mathbf{r}}$ on every $\mathbf{r}$-box.

Let n be a positive integer satisfying $2N \leqq 2^n < 4N$. For every K-tuple $\mathbf{r}$ with $|\mathbf{r}| = r_1 + \cdots + r_K = n$ let $f_{\mathbf{r}}$ be an arbitrary $\mathbf{r}$-function and let $F(\mathbf{x}) = \sum_{|\mathbf{r}|=n} f_{\mathbf{r}}(\mathbf{x})$. We recall a result of Schmidt [29] stating

$$\int_{[0,1]^K} F^2(\mathbf{x}) \mathrm{d}\mathbf{x} \ll_K (n+1)^{K-1}.$$

For a proof we refer to Lemma 2.4 in the monograph of Beck and Chen [7].

Therefore to finish the proof of Theorem C it is enough to specify the functions $f_{\mathbf{r}}$ so that the function F satisfy the first condition, too. For this aim choose an $\mathbf{r}$-function $f_{\mathbf{r}}$ for every $|\mathbf{r}| = n$ satisfying

$$\int_B f_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) \mathrm{d}\mathbf{x} \geqq 0$$

for every $\mathbf{r}$-box B, this can be done because of the definition of the $\mathbf{r}$-functions.

For a box $B = [c_1, d_1) \times \cdots \times [c_K, d_K)$ $(0 \leqq c_i < d_i \leqq 1)$ let $\widetilde{B}$ be the parallelepiped $\widetilde{B} = \bigcap_{i=1}^{K} (H_i^* \cap H_i^{**})$, where the halfspaces H_i^* and H_i^{**} are defined by inequalities $\mathbf{a}_i \mathbf{x} \leqq b_i - c_i \beta_i$ and $\mathbf{a}_i \mathbf{x} > b_i - d_i \beta_i$, respectively. Denote $\widetilde{[0,1)^K}$ by E.

LEMMA 6.1. *If B is an $\mathbf{r}$-box with $\widetilde{B} \cap \mathcal{P} = \emptyset$, then*

$$\int_B R_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) \mathrm{d}\mathbf{x} = (-1)^{K+1} N \mu(E) 2^{-2|\mathbf{r}|-2K}.$$

PROOF. Let $B = \prod_{i=1}^{K} [m_i 2^{-r_i}, (m_i+1)2^{-r_i})$ and introduce

$$B' = \prod_{i=1}^{K} \left[m_i 2^{-r_i}, \left(m_i + \frac{1}{2}\right) 2^{-r_i} \right).$$

Then

$$\int_B R_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} =$$

$$= \int_{B'} \sum_{\varepsilon_1=0}^{1} \cdots \sum_{\varepsilon_K=0}^{1} (-1)^{\varepsilon_1+\cdots+\varepsilon_K} D(\mathcal{P}, A(x_1 + \varepsilon_1 2^{-r_1-1}, \ldots, x_k + \varepsilon_K 2^{-r_K-1})) d\mathbf{x}.$$

Using an elementary sieving argument we obtain

$$\int_B R_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} = \int_{B'} (-1)^K D(\mathcal{P}, \widetilde{B(\mathbf{x})}) d\mathbf{x},$$

where

$$B(\mathbf{x}) = \prod_{i=1}^{K} [x_i, x_i + 2^{-r_i-1}).$$

Note that $B(\mathbf{x}) \subset B$ for every $\mathbf{x} \in B'$, hence $\widetilde{B(\mathbf{x})} \cap \mathcal{P} = \emptyset$. Therefore we have

$$\int_B R_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} = (-1)^K \int_{B'} -N\mu(\widetilde{B(\mathbf{x})}) d\mathbf{x} =$$

$$= (-1)^{K+1} \mu(B') N\mu(B(\mathbf{x}))\mu(E) = (-1)^{K+1} N\mu(E) 2^{-2|\mathbf{r}|-2K},$$

as it was to be proved.

Let $|\mathbf{r}| = n$. E is the disjoint union of parallelepipeds $\widetilde{B}$, where B is an $\mathbf{r}$-box. Since $|\mathcal{P}| \leqq 2^{n-1}$, there exist at least 2^{n-1} $\mathbf{r}$-boxes B for which $\widetilde{B}$ does not contain a point of $\mathcal{P}$. Therefore Lemma 6.1 yields

$$\int_{[0,1]^K} f_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} \geqq \sum_{B \cap \mathcal{P} = \emptyset} \left| \int_B R_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} \right| \geqq$$

$$\geqq N\mu(E) 2^{-n-2K-1}.$$

Finally

$$\int_{[0,1]^K} F(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} = \sum_{|\mathbf{r}|=n} \int_{[0,1]^K} f_{\mathbf{r}}(\mathbf{x}) D(\mathbf{x}) d\mathbf{x} \geqq$$

$$\geqq \binom{n+K-1}{K-1} N\mu(E) 2^{-n-2K-1} \gg_{A^*} (\log N)^{K-1}$$

shows the desired inequality.

PROOF OF THEOREM D. We may assume, that $\mathcal{A}$ contain the coordinate hyperplanes $H_i = \{\mathbf{x} \in \mathbb{R}^K \mid x_i = 0\}$. Then we can simply write

$$\mathrm{POL}_0(\mathcal{A}) = \{A \in \mathrm{POL}(\mathcal{A}) \mid A \subseteqq [0,1]^K\}.$$

We will represent the elements of $\mathrm{POL}_0(\mathcal{A})$ by the points of the $2l$-dimensional unit cube. Suppose that the hyperplane $H_i \in \mathcal{A}$ has defining equation $\mathbf{a}_i\mathbf{x} = b_i$, and let $\mathbf{a}_i\mathbf{x} = u_i$ and $\mathbf{a}_i\mathbf{x} = v_i$ $(u_i < v_i)$ denote the equations of the supporting hyperplanes to the cube $[0,1]^K$, parallel to H_i. Define the halfspaces

$$H_i(s) = \{\mathbf{x} \in \mathbb{R}^K \mid \mathbf{a}_i\mathbf{x} \geqq u_i + s(v_i - u_i)\}$$

and

$$H_{i+l}(s) = \{\mathbf{x} \in \mathbb{R}^K \mid \mathbf{a}_i\mathbf{x} \leqq u_i + s(v_i - u_i)\}$$

for every real number s. Then an arbitrary element of $\mathrm{POL}_0(\mathcal{A})$ is of the form $H_1(s_1) \cap \ldots \cap H_{2l}(s_{2l})$ with a suitable vector $\mathbf{s} = (s_1, \ldots, s_{2l}) \in [0,1]^{2l}$, and conversely, for every $\mathbf{s} \in [0,1]^{2l}$ the polytope $A(\mathbf{s}) = \bigcap_{i=1}^{2l} H_i(s_i)$ is an element of $\mathrm{POL}_0(\mathcal{A})$, that may be degenerated.

Call the sequence $s_1, s_2, \ldots$ of real numbers monotonic of type $\prec$, if $s_1 \prec s_2 \prec \ldots$. The sequence of vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots \in [0,1]^{2l}$ is called monotonic of type $(\prec_1, \ldots, \prec_{2l})$ (where $\prec_i \in \{<, =, >\}$), if the sequence $s_{1i}, s_{2i}, \ldots$ is monotonic of type $\prec_i$ for every $1 \leqq i \leqq 2l$.

LEMMA 6.2. *Suppose that the sequence of vectors* $\mathbf{s}_1, \mathbf{s}_2, \ldots \in [0,1]^{2l}$ *is monotonic. Then the intersection of any* $2l+1$ *of the sets*

$$A(\mathbf{s}_1) \triangle A(\mathbf{s}_2), A(\mathbf{s}_2) \triangle A(\mathbf{s}_3), \ldots$$

is empty.

($A \triangle B$ denotes the symmetric difference $(A \setminus B) \cup (B \setminus A)$ of the sets A and B.)

PROOF. Let $\mathbf{s}, \mathbf{t} \in [0,1]^{2l}$. If the point $\mathbf{x} \in [0,1]^K$ is contained in the halfspaces $H_i(s_i)$ and $H_i(t_i)$ for the same indices $1 \leqq i \leqq 2l$, then $\mathbf{x}$ is contained in either both of the sets $A(\mathbf{s})$ and $A(\mathbf{t})$ or in neither of them. Thus, if $\mathbf{x} \in A(\mathbf{s}) \triangle A(\mathbf{t})$, then there exists $1 \leqq i \leqq 2l$ such that $\mathbf{x} \in H_i(s_i) \triangle H_i(t_i)$. Therefore it is enough to prove that the point $\mathbf{x}$ is contained in at most one of the sets $H_i(s_{1i}) \triangle H_i(s_{2i}), H_i(s_{2i}) \triangle H_i(s_{3i}), \ldots$ for any fixed i. It is clear since the sequence $s_{1i}, s_{2i}, \ldots$ is monotonic, and it implies $H_i(s_{1i}) \subseteqq H_i(s_{2i}) \subseteqq \subseteqq \ldots$ or $H_i(s_{1i}) \supseteqq H_i(s_{2i}) \supseteqq \ldots$.

The function $f : [0,1]^{2l} \longrightarrow [0,1]$, $f(\mathbf{s}) = \mu(A(\mathbf{s}))$ is clearly a continuous function. Therefore the missing part of the proof is contained in Schmidt [27], Theorem 1 (see the note following the proof of Lemma 1). We omit the very complicated proof.

ACKNOWLEDGEMENTS. I am indebted to the Zentrum für interdisziplinäre Forschung in Bielefeld for the financial support to this research.

REFERENCES

[1] AARDENNE-EHRENFEST, T. VAN, Proof of the impossibility of a just distribution of an infinite sequence of points over an interval, *Nederl. Akad. Wetensch., Proc.* **48** (1945), 266–271 = *Indagationes Math.* **7** (1945), 71–76. *MR* **7**–376

[2] AARDENNE-EHRENFEST, T. VAN, On the impossibility of a just distribution, *Nederl. Akad. Wetensch., Proc.* **52** (1949), 734–739 = *Indagationes Math.* **11** (1949), 264–269. *MR* **11**–336

[3] BECK, J., Irregularities of distribution. I, *Acta Math.* **159** (1987), 1–49. *MR* **89c**:11117

[4] BECK, J., Irregularities of distribution. II, *Proc. London Math. Soc.* (3) **56** (1988), 1–50. *MR* **89c**:11118

[5] BECK, J., A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution, *Compositio Math.* **72** (1989), 269–339. *MR* **91f**:11054

[6] BECK, J., On the discrepancy of convex plane sets, *Monatsh. Math.* **105** (1988), 91–106. *MR* **89f**:11109

[7] BECK, J. and CHEN, W. W. L., *Irregularities of distribution*, Cambridge Tracts in Mathematics, 89, Cambridge Univ. Press, Cambridge – New York, 1987. *MR* **88m**:11061

[8] BECK, J. and CHEN, W. W. L., Irregularities of point distribution relative to convex polygons, *Irregularities of partitions* (Fertőd, 1986), Algorithms Combin.: Study Res. Texts, 8, Springer, Berlin – New York, 1989, 1–22. *MR* **90e**:11116

[9] BECK, J. and SÓS, V. T., Discrepancy theory, *Handbook of combinatorics*, Springer, Berlin – New York, 1994.

[10] CHERNOFF, H., A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23** (1952), 493–507. *MR* **15**–241

[11] CORPUT, J.G. VAN DER, Verteilungsfunktionen I–II, *Nederl. Akad. Wetensch., Proc.* **38** (1935), 813–821 and 1058–1066. *Zbl* **12**, 347 and *Zbl* **13**, 57

[12] HALÁSZ, G., On Roth's method in the theory of irregularities of point distributions, *Recent progress in analytic number theory*, Vol. 2 (Durham, 1979), Academic Press, London – New York, 1981, 79–94. *MR* **83e**:10072

[13] HALTON, J. H., On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals, *Numer. Math.* **2** (1960), 84–90. *MR* **22**#12688

[14] HLAWKA, E., Funktionen von beschränkter Variation in der Theorie der Gleichverteilung, *Ann. Math. Pura Appl.* (4) **54** (1961), 325–333. *MR* **25**#3029

[15] HOEFFDING, W., Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30. *MR* **26**#1908

[16] KÁROLYI, GY., On a decomposition formula of Varchenko, concerning configurations of affine hyperplanes in real space (manuscript).

[17] KÁROLYI, GY. and LOVÁSZ, L., Decomposition of convex polytopes into simplices (manuscript).

[18] KOKSMA, J. F., Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1 [A general theorem from the theory of uniform distribution modulo 1], *Mathematica, Zutphen*, B, **11** (1942/43), 7–11 (in Dutch). *MR* **7**–370

[19] LOVÁSZ, L., SPENCER, J. H. and VESZTERGOMBI, K., Discrepancy of set-systems and matrices, *European J. Combin.* **7** (1986), 151–160. *MR* **88b**:05036

[20] MCDIARMID, C. J. H., On the method of bounded differences, *Surveys in combinatorics, 1989* (Norwich, 1989), London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge, 1989, 148–188. *MR* **91e**:05077

[21] NIEDERREITER, H., Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84** (1978), 957–1041. *MR* **80d**:65016

[22] ROTH, K. F., On irregularities of distribution, *Mathematika* **1** (1954), 73–79. *MR* **16**–575

[23] RUZSA, I. Z., The discrepancy of rectangles and squares, Österreichisch-Ungarisch-Slowakisches Kollokvium über Zahlentheorie, *Grazer Math. Ber.* **318** (1993), 135–140.

[24] SCHMIDT, W. M., Irregularities of distribution. IV, *Invent. Math.* **7** (1969), 55–82. *MR* **39**#6838

[25] SCHMIDT, W. M., Irregularities of distribution. VI, *Compositio Math.* **24** (1972), 63–74. *MR* **47**#152

[26] SCHMIDT, W. M., Irregularities of distribution. VII, *Acta Arith.* **21** (1972), 45–50. *MR* **47**#8474

[27] SCHMIDT, W. M., Irregularities of distribution. VIII, *Trans. Amer. Math. Soc.* **198** (1974), 1–22. *MR* **50**#12952

[28] SCHMIDT, W. M., Irregularities of distribution. IX, *Acta Arith.* **27** (1975), 385–396. *MR* **51**#12768

[29] SCHMIDT, W. M., Irregularities of distribution. X, *Number theory and algebra*, Academic Press, New York, 1977, 311–329. *MR* **58**#10803

[30] SÓS, V. T., Irregularities of partitions: Ramsey theory, uniform distribution, *Surveys in combinatorics* (Southampton, 1983), London Math. Soc. Lecture Note Ser., 82, Cambridge Univ. Press, Cambridge – New York, 1983, 201–246. *MR* **85h**:05012

[31] STUTE, W., Convergence rates for the isotrope discrepancy, *Ann. Probability* **5** (1977), 707–723. *MR* **56**#13336

[32] VARCHENKO, A. N., Combinatorics and topology of the arrangement of affine hyperplanes in the real space, *Funktsional Anal. i. Prilozhen* **21** (1987), no. 1, 11–22 (in Russian). *MR* **89c**:32029; Combinatorics and topology of the disposition of affine hyperplanes in real space, *Functional Anal. Appl.* **21** (1987), 9–19.

(Received February 14, 1994)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALGEBRA TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 95–103

TRIFFERENCE

J. KÖRNER and G. SIMONYI

Abstract

To distinguish n objects, we can label them by n binary sequences of length $\lceil \log_2 n \rceil$ each. Shorter sequences would not do. How about *tristinguishing* n objects? In this problem we use ternary sequences for labeling and require that any three of these be different in one and the same coordinate. This is the simplest unsolved case of a problem known as perfect hashing. We give a non-existence bound for a similar problem on binary sequences. We also deal with related problems of edge–colorings in graphs. It is shown that the minimum number of tricolorings needed to give every triangle of K_n all the three colors in at least one coloring is at most $\lceil \log_2 n \rceil$.

Introduction

To distinguish n objects, we can address them by n binary sequences of length $\lceil \log n \rceil$ each. Shorter sequences would not do. (Notice that here and in the sequel all log's and exp's are binary.) From this trivial observation a surprisingly short way takes us to a hard unsolved combinatorial problem that emerges in several important models of computer science. More importantly, we will try to show that this problem represents a stumbling block whose "removal" might lead to a spectacular extension of the information–theoretic approach to extremal set theory from the case of graphs to hypergraphs.

Subsets or bipartitions of an n–set can be represented by binary sequences of length n. Likewise, k-partitions of an n–set can be represented by sequences over a k–ary alphabet. It was shown in the papers [6] and [7] that many problems in combinatorics regarding subsets or partitions of a set can be reformulated within a common information–theoretic framework in which the key notion is for sequences to be *really different* in a particular way pertinent to the problem. Let us consider a graph G having as vertex

1991 *Mathematics Subject Classification.* Primary 05A18; Secondary 05D05.

Key words and phrases. Three-partitions, perfect hashing, edge-colorings of graphs, anti-Ramsey colorings.

The manuscript was prepared while the first author was visiting the IASI–CNR, Rome, Italy. The second author's research has partially been supported by the Hungarian National Foundation for Scientific Research Grant Nos. 1906 and 4264.

in some coordinate they differ in two elements of the alphabet which are the two endpoints of an edge of G leads us to the problem of Shannon capacity of the graph, [14]. In this problem, we ask how long each of n sequences over a k–ary alphabet must be if they have to be "really different" in the previous sense. This minimum length is easily shown to be asymptotically of the order $c \log n$ where the constant c is a characteristic of the graph. In fact, if we denote the above minimum length of the sequences by $l(G, n)$, then we can define the (logarithmic) Shannon capacity of the graph G as the always existing limit

$$\lim_{n\to\infty} \frac{\log n}{l(G,n)}.$$

In [2] Cohen, Körner and Simonyi extend this definition to families of graphs. Rather than requiring the occurrence of an edge of a fixed graph between sequences, they require the existence of an edge of any graph from a fixed family $\mathbf{G}$. They call the corresponding notion of capacity the Shannon capacity of the family of graphs $\mathbf{G}$. In [5] L. Gargano, J. Körner and U. Vaccaro introduce a further extension of this definition. Instead of simple graphs they consider directed graphs. Then, in case of a single digraph, the notion of really different requires, between any pairs of sequences the existence of arcs of the graph with opposite orientations. The corresponding analogon of Shannon capacity is called Sperner capacity.

Our paper [10] gives a simple example of the relevance of this kind of notions to extremal set theory. Subsequently, Gargano, Körner and Vaccaro have shown ([6], [7]) that the concept of Sperner capacity of a family of graphs offers a formally information–theoretic framework to treat and solve many interesting and even some long–standing open problems in extremal combinatorics in an asymptotic sense. These include various generalizations of Sperner's classical theorem on the maximum number of subsets of an n–set without one containing the other and the solution of Rényi's 1970 problem on the maximum number of pairwise qualitatively 2–independent k–partitions of an n–set, [13]. Beyond the above papers the interested reader is advised to consult [11] and [1] where the problem of computing the Shannon capacity and the Sperner capacity of a single graph are addressed.

If it is true that the Sperner capacity framework encompasses a great many combinatorial problems, we soon have to add that much more problems are left outside its scope. In fact, many problems in extremal combinatorics require more structure than what is offered by the framework of pairwise comparison of sequences. Whatever complicated notion of being really different we might come up with, it would not help; to formulate more problems in our language, we have to invoke some comparison of three or more sequences. Formally, this amounts to extend the investigation of capacities from graphs to hypergraphs, [9]. Of all such problems one is standing out. This is the problem of *trifference* discussed in the next section. We dare say that it is the conceptually simplest and most natural of them all. In one

way or the other, solving it would shed light on the rest.

In this paper we just want to present some results concerning trifferentiating objects in some restricted manner. Some of these related problems are defined on graphs. Instead of trifferentiating any triple of elements of an n-set we might want to trifferentiate just a particular subset of these. Problems of the latter kind bring us closer to the interesting topic of anti–Ramsey theorems in the sense of [15]. A typical problem in anti–Ramsey theorems is to ask how many colors are needed to color the edges of K_m so that any three edges that form a triangle obtain different colors. (In fact, the answer to this question is trivial, but problems of this kind soon get complicated, cf. [15].) Reversing this question, Vera T. Sós asked how many tricolorings of the edges of K_m are needed if the edges of every triangle have to get three different colors in at least one of them. Setting $n = \binom{m}{2}$ Vera Sós' question can be reformulated in our language as follows. What is the minimum length of ternary sequences we have to use for labeling in order to assign trifferent labels to any triple of edges forming a triangle in K_m? Similar questions can be asked about three edges forming other subgraphs.

Problems of tristinction are strongly connected to Rényi's still unsolved question of qualitatively 3–independent partitions of an n–set and some other problems in combinatorics which keep coming back under different disguises.

Perfect hashing, trifference and quasi-trifference

Perfect hashing is a purely combinatorial model for the hashing problem in computer science. Its history and importance can be best understood from the paper of A. Yao [16]. We shall adopt the terminology of Fredman and Komlós [4].

DEFINITION. A family of b–partitions of a set X is called a (b,k)–system of perfect hash functions if every k–element subset of X meets k different classes of at least one of the partitions in the family. We denote by $Y(b,k,n)$ the minimum number of partitions in any (b,k)–system for a set of n elements. For given b and k set

$$F(b,k) = \liminf_{n\to\infty} \frac{Y(b,k,n)}{\log n}.$$

(Notice that $F(b,k)$ is the reciprocal of the capacity of a particular uniform hypergraph in the sense of [9].) The exact value of $F(b,k)$ is unknown for $b \geqq k > 2$. The best available bounds are due to Fredman and Komlós [4] and Körner and Marton [8], cf. also [9]. In said papers rather sophisticated information–theoretic proof techniques are used to obtain lower bounds. In exchange, in [4] the upper bound is derived using plain random selection. In [8] and [3], independently, this upper bound was improved in the case

In [8] and [3], independently, this upper bound was improved in the case $b=k=3$, thus showing that random selection gives rather poor results for this problem. In this paper we concentrate on problems related to this particular case which we like to call the problem of trifference. We recall the corresponding bounds available in the literature.

Körner and Marton [8] have proved

$$\frac{1}{\log \frac{3}{2}} \leqq F(3,3) \leqq \frac{4}{\log \frac{9}{5}}. \tag{1}$$

Numerically, this means that

$$1.709 \leqq F(3,3) \leqq 4.717.$$

The upper bound is implicit also in [3].

Now we consider the following related problem.

DEFINITION. We call the binary sequences $\mathbf{x}=x_1,x_2,\ldots,x_t$, $\mathbf{y}=y_1,y_2,\ldots,y_t$ and $\mathbf{z}=z_1,z_2,\ldots,z_t$ quasi-trifferent if there exists a coordinate $1 \leqq i \leqq t-1$ for which the ordered pairs (x_i,x_{i+1}), (y_i,y_{i+1}), (z_i,z_{i+1}) are all different.

Let $Y_2(2,3,n)$ be the minimum number t for which there exist n binary sequences of length t such that every three of them are quasi-trifferent. Set

$$F_2(2,3)=\liminf_{n\to\infty} \frac{Y_2(2,3,n)}{\log n}.$$

(Note that any three pairwise different binary sequences are trifferent at *some* two coordinates, i.e., if there is no restriction on the (relative) location of these two coordinates.)

The result of this section is the following

THEOREM 1.

$$F_2(2,3) \geqq 2.$$

PROOF. Let D be a quasi-trifferent set of t-length binary sequences, i.e., any three sequences in D are quasi-trifferent. Let $B=\{0,1\}^{\lfloor t/2 \rfloor}$ and for each $y \in B$ let $A(y)$ denote the set of all t-length binary sequences the even numbered coordinates of which form y.

Now we use double counting for the pairs $(x,A(y))$ where $x \in D$, $y \in B$ and $x \in A(y)$. Since every $x \in D$ uniquely determines the corresponding $A(y)$, obviously $|\{(x,A(y)) : x \in A(y), x \in D\}| = |D|$. On the other hand, $|A(y) \cap D| \leqq 2$ for any fixed $A(y)$ since three binary sequences that coincide at every even-numbered coordinate could not be quasi-trifferent. This implies

$$|\{(x,A(y)) : x \in D, x \in A(y)\}| \leqq 2|\{A(y) : y \in \{0,1\}^{\lfloor t/2 \rfloor}\}| = 2 \cdot 2^{\lfloor t/2 \rfloor}.$$

Combining this inequality with the previous equality we get $|D| \leqq 2 \cdot 2^{\lfloor t/2 \rfloor}$. This implies $n \leqq 2 \cdot 2^{\lfloor Y_2(2,3,n)/2 \rfloor}$ and thus the theorem follows. □

Tedious calculations give an upper bound by random choice that we omit here because of its irrelevance.

Tricolored triangles

Let K_n denote the complete graph on n vertices. An edge–tricoloring of K_n is a partition of the edge set of K_n into three different classes. We refer to the members of these respective classes as red, blue and green edges, respectively. We call a triangle edge–tricolored (ET) if all its edges are colored differently. Let $t(n)$ denote the minimum number of edge–tricolorings needed to make every triangle ET in at least one of them. Write

$$T = \liminf_{n\to\infty} \frac{t(n)}{\log n}.$$

Determining T seems hard and our lower and upper bounds are far apart. The lower bound is trivial. The main interest of the next result is that the upper bound is obtained via an explicit construction for this does not seem to happen frequently with similar problems.

THEOREM 2.

$$\frac{1}{\log 3} \leqq T \leqq 1.$$

PROOF. The lower bound is trivial. In fact, fix any vertex and look at all the adjacent edges of which there are $n-1$. Clearly, any two of them must have differing colors in at least one tricoloring. This gives

$$\frac{\log(n-1)}{\log 3} \leqq t(n).$$

To prove the upper bound, assign to every node of K_n a different binary sequence of length $\lceil \log n \rceil$. We will define $\lceil \log n \rceil$ edge–tricolorings of K_n through these sequences. Let us look at the edge having the different vertices a and b as endpoints. Let $\mathbf{x} = x_1, x_2, \ldots, x_t$ and $\mathbf{y} = y_1, y_2, \ldots, y_t$ be the corresponding binary sequences. Define the i'th tricoloring of (a, b) as follows.

Let (a, b) be blue if $x_i = y_i$.

Let (a, b) be green if $x_i \neq y_i$ but $x_j = y_j$ for all $j < i$.

Let (a, b) be red else.

Let us say that the i'th coordinate *cuts the edge* (a, b) if the i'th coordinates of the sequences assigned to a resp. b are different, i. e., if $x_i \neq y_i$. We claim that the edges of every triangle get 3 different colors in a coloring in some coordinate $i > 1$. To prove this, notice that every edge of K_n is cut in some coordinate and that a coordinate cutting any edge of a triangle will cut exactly two of them. From these two observations it follows that in every triangle at least two different pairs of edges are cut in some coordinate.

Now fix a triangle and consider the smallest coordinate i for which all the edges of the triangle are cut in some coordinate with $j \leqq i$. This means by the foregoing that in this coordinate i there is an edge cut for the first

time and therefore never cut in a coordinate $j < i$. Notice, however, that the triangle cannot have more than one edge with these properties, for some pair of edges had to be cut before. Furthermore, there is an edge that is not cut in the i'th coordinate. This proves that our triangle has tricolored edges in this coordinate. Thus

$$t(n) \leqq \lceil \log n \rceil - 1. \qquad \square$$

It seems unlikely that the lower bound be tight. In this context, it is worth noticing what happens if we just want to bicolor every pair of adjacent edges. More precisely, let $u(n)$ be the minimum number of edge–tricolorings needed to make every pair of adjacent edges of K_n bicolored in at least one of the tricolorings. Write

$$U = \liminf_{n \to \infty} \frac{u(n)}{\log n}.$$

We have

PROPOSITION 3.

$$U = \frac{1}{\log 3}.$$

PROOF. The lower bound is true by the same argument as in Theorem 1. To prove the upper bound label every vertex of K_n by a different ternary sequence of length $\lceil \frac{\log n}{\log 3} \rceil$. Next label every edge by the modulo 3 sum of the ternary vectors assigned to its two endpoints. The i'th coordinates of all these vectors give rise to the i'th edge–coloring in the obvious way. It is immediate that this family of colorings satisfies our condition. $\square$

Other tricolored subgraphs

It follows from our previous observations that substantially more tricolorations of the edges of K_n are needed to tristinguish any triple of edges of the complete graph on n vertices than to tristinguish just the three edges of triangles. In fact, by definition, the number of tricolorings needed to tristinguish every triple of the edges is $Y(3,3,\binom{n}{2})$ which is about $2F(3,3)\log n \geqq$ $\geqq 3.4 \log n$ while we have seen that for the tristinction of the three edges of any triangle we need not more than about $\log n$ tricolorings of the edges. It is therefore interesting to understand what happens if we want to tristinguish the three edges of some other 3–edge subgraphs of K_n by the minimum number of 3–colorings of the edges of this complete graph. In particular, it is interesting to see whether there is a single type of subgraph which in itself is responsible for the total number of colorings needed to tristinguish all the edge triples of K_n, in the asymptotic sense.

Let $s(n)$ denote the minimum number of tricolorings of the edges of K_n needed to make every tristar edge–tricolored in at least one of them. (Here a tristar is a graph on 4 points with 3 edges all of which have a common endpoint. Just as for triangles, we say that a tristar is edge–tricolored if all its edges are colored differently.) Write

$$S = \liminf_{n\to\infty} \frac{s(n)}{\log n}.$$

We have

PROPOSITION 4.

$$S = F(3,3).$$

PROOF. The lower bound $F(3,3) \leqq S$ follows from the fact that if we want to tristinguish just the $n-1$ edges meeting in a single fixed vertex of K_n, then this is equivalent to the problem of trifference and thus

$$Y(3,3,n-1) \leqq s(n).$$

To prove the upper bound label every vertex of K_n by a trifferent ternary sequence of length $Y(3,3,n)$. Then label every edge by the modulo 3 sum of the ternary vectors assigned to its two endpoints. □

We will call trident a graph on 6 vertices with three vertex–disjoint edges. Let $r(n)$ denote the minimum number of tricolorings of the edges of K_n needed to tristinguish the three edges of any trident in at least one of them. Write further

$$R = \liminf_{n\to\infty} \frac{r(n)}{\log n}.$$

We claim that

PROPOSITION 5.

$$R = F(3,3).$$

PROOF. The lower bound $F(3,3) \leqq R$ is obvious. In fact, notice that a maximal matching of the graph consists of $\lfloor \frac{n}{2} \rfloor$ pairwise vertex–disjoint edges. If we only restrict ourselves to coloring these, we see that $Y(3,3,\lfloor \frac{n}{2} \rfloor) \leqq r(n)$.

To prove the upper bound, let us assign trifferent ternary sequences of minimum length to each of the vertices of K_n. Next assign to every edge one of the ternary sequences assigned to its two endpoints in a completely arbitrary manner. The sequences will define the tricolorings in the obvious way establishing

$$r(n) \leqq Y(3,3,n). \qquad \square$$

The last proposition is somewhat surprising for numberwise the configuration of three vertex–disjoint edges is dominant among all the configurations of three edges. The proof shows that for very general criteria for colorings of

k vertex–disjoint edges the minimum number of colorings is asymptotically the same as for criteria on vertex–colorings involving arbitrary k–tuples of vertices. The same remark applies for stars with k edges.

All our above constructions share the feature that the edge–colorings are constructed from vertex–colorings in a straightforward manner. We wonder whether this is due to our lack of imagination or something more relevant to the subject.

We have failed to give non–trivial bounds for the minimum number of tricolorings needed to tristinguish the edges of the two missing subgraphs.

ACKNOWLEDGEMENT. Warmfelt thanks are due to Vera T. Sós for useful discussions in a unique atmosphere. We thank Gábor Tardos for finding an error in an earlier version of our manuscript.

We are grateful to the Zentrum für interdisziplinäre Forschung of Bielefeld University for its generous hospitality.

REFERENCES

[1] CALDERBANK, R., FRANKL, P., GRAHAM, R. L., LI, W. and SHEPP, L., The cyclic triangle problem, *J. Alg. Comb.* (submitted).

[2] COHEN, G., KÖRNER, J. and SIMONYI, G., Zero-error capacities and very different sequences (preliminary version), *Sequences* (Naples/Positano, 1988), Springer, New York, 1990, 144–155. *MR* **90m**:00047

[3] ELIAS, P., Zero error capacity under list decoding, *IEEE Trans. Inform. Theory* **34** (1988), 1070–1074. *MR* **89k**:94038

[4] FREDMAN, M. L. and KOMLÓS, J., On the size of separating systems and families of perfect hash functions, *SIAM J. Algebraic Discrete Methods* **5**(1984), 61–68. *MR* **86a**:05009

[5] GARGANO, L., KÖRNER, J. and VACCARO, U., Sperner theorems on directed graphs and qualitative independence, *J. Combin. Theory Ser. A* **61**(1992), 173–192.

[6] GARGANO, L., KÖRNER, J. and VACCARO, U., Sperner capacities, *Graphs and Combinatorics* **9**(1993), 31–46. *MR* **94e**:05024

[7] GARGANO, L., KÖRNER, J. and VACCARO, U., Capacities: from information theory to extremal set theory, *J. Combin. Theory Ser. A* (to appear).

[8] KÖRNER, J. and MARTON, K., New bounds for perfect hashing via information theory, *European J. Combin.* **9**(1988), 523–530. *MR* **90a**:05011

[9] KÖRNER, J. and MARTON, K., On the capacity of uniform hypergraphs, *IEEE Trans. Inform. Theory* **36**(1990), 153–156. *MR* **91b**:05139

[10] KÖRNER, J. and SIMONYI, G., A Sperner-type theorem and qualitative independence, *J. Combin. Theory Ser. A* **59**(1992), 90–103. *MR* **92m**:05199

[11] LOVÁSZ, L., On the Shannon capacity of a graph, *IEEE Trans. Information Theory* **25**(1979), 1–7. *MR* **81g**:05095

[12] MACWILLIAMS, F. J. and SLOANE, N. J. A., *The theory of error-correcting codes*, Vol. I–II, North-Holland, Amsterdam, 1977 and New York, 1983. *MR* **57**#5408a, 5408b

[13] RÉNYI, A., *Foundations of probability*, Holden-Day, San Francisco, 1970; *Probability theory*, North-Holland Series in Applied Mathematics and Mechanics, Vol. 10, North-Holland, Amsterdam – London; American Elsevier Publ. Co., New York, 1970. *MR* **47**#4296

[14] SHANNON, C. E., The zero–error capacity of a noisy channel, *IRE Trans. Information Theory* **2**(1956), 8–19. *MR* **19**–623

[15] Simonovits, M. and Sós, V. T., On restricted colourings of K_n, *Combinatorica* **4** (1984), 101–110. *MR* **85m**:05044

[16] Yao, A. C., Should tables be sorted?, *J. Assoc. Comput. Mach.* **28**(1981), 615–628. *MR* **82f**:68099

(Received February 14, 1994)

DIPARTIMENTO DI SCIENZE DELL' INFORMAZIONE
UNIVERSITÀ "LA SAPIENZA"
VIA SALARIA 113
I–00198 ROMA
ITALY

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 105–109

DISCREPANCY ESTIMATES FOR SETS WITH SMALL BOUNDARY

M. LACZKOVICH

Let μ be a non-negative, finitely additive measure defined on the measurable subsets of the k-dimensional unit cube $I^k = [0,1)^k$. The quantity $D(\mu; H) = |\mu(H) - \lambda_k(H)|$ is called the *discrepancy* of μ with respect to the set $H \subset I^k$. The family of all cubes, intervals, and convex sets contained in I^k will be denoted by $\mathcal{Q}_k$, $\mathcal{I}_k$, and $\mathcal{C}_k$, respectively. We define

$$D_q(\mu) = \sup\{D(\mu; H) : H \in \mathcal{Q}_k\}, \quad D(\mu) = \sup\{D(\mu; H) : H \in \mathcal{I}_k\}, \quad \text{and}$$

$$D_c(\mu) = \sup\{D(\mu; H) : H \in \mathcal{C}_k\};$$

then we have $D_q(\mu) \leqq D(\mu) \leqq D_c(\mu)$ for every μ.

It was proved by E. Hlawka that $D_c(\mu) \leqq C_k D(\mu)^{1/k}$ holds for every μ, where C_k is a constant only depending on the dimension k (see [1], or [2], Theorem 1.6, p. 95).

A substantial generalization of this estimate was given by H. Niederreiter and J. M. Wills in [6]. Let $b : (0,\infty) \to (0,\infty)$ be an increasing function such that $\lim_{\varepsilon \to 0+} b(\varepsilon) = 0$, and let $\mathcal{M}_b$ denote the family of those subsets $H \subset I^k$ for which

$$\lambda_k(\{x \in I^k \setminus H : \operatorname{dist}(x, H) < \varepsilon\}) \leqq b(\varepsilon)$$

and

$$\lambda_k(\{x \in H : \operatorname{dist}(x, I^k \setminus H) < \varepsilon\}) \leqq b(\varepsilon)$$

hold for every $\varepsilon > 0$. Niederreiter and Wills prove that

$$D(\mu; H) \leqq 4b(2\sqrt{k} D(\mu)^{1/k})$$

holds for every $H \in \mathcal{M}_b$, supposing that $b(\varepsilon) \geqq \varepsilon$ for all $\varepsilon > 0$. They actually give a somewhat better, but more complicated estimate, which holds also without the condition $b(\varepsilon) \geqq \varepsilon$. If H is convex then $H \in \mathcal{M}_{C\varepsilon}$ and thus the special case when $b(\varepsilon) = C\varepsilon$ gives Hlawka's inequality.

For this special case W. M. Schmidt [8] proves the sharper estimate

$$D(\mu; H) \leqq C_k D_q(\mu)^{1/k};$$

that is, in this case $D(\mu)$ can be replaced by the smaller $D_q(\mu)$.

In this note we show that in the theorem by Niederreiter and Wills $D(\mu)$ can be replaced by $D_q(\mu)$ for every $b(\varepsilon)$.

1991 *Mathematics Subject Classification.* Primary 11K38.
Key words and phrases. Discrepancy.

0081-6906/95/$ 4.00

THEOREM. *For every increasing* $b : (0,\infty) \to (0,\infty)$ *and for every* $H \in \mathcal{M}_b$ *we have*

$$D(\mu; H) \leqq C_k b(\sqrt{k} D_q(\mu)^{1/k}),$$

where the constant C_k *only depends on* k.

PROOF. We shall use the following notation. If r is a positive integer, then we shall denote by $\mathcal{Q}_{k,r}$ the set of cubes

$$\prod_{i=1}^{k} \left[\frac{a_i - 1}{r}, \frac{a_i}{r}\right) \qquad (a_i = 1, \dots, r,\ i = 1, \dots, k).$$

For every $m \in \mathbf{N}$ we put

$$\mathcal{D}_k^{(m)} = \left\{ \prod_{i=1}^{k} [a_i \cdot 2^m, (a_i + 1)2^m) \ :\ a_i \in \mathbf{Z},\ i = 1, \dots, k \right\}.$$

Thus $\mathcal{D}_k^{(0)}$ is the set of unit cubes. The system of *dyadic cubes* is defined as

$$\mathcal{D}_k = \bigcup_{m=0}^{\infty} D_k^{(m)}.$$

For every $a \in \mathbf{R}$ and $A \subset \mathbf{R}^k$ we shall denote $aA = \{ax : x \in A\}$. The boundary and the closure of the set A will be denoted by ∂A and $\operatorname{cl} A$, respectively. The cardinality of a (finite) set E is denoted by $|E|$.

Let $b : (0,\infty) \to (0,\infty)$ be an increasing function and let $H \in \mathcal{M}_b$ be arbitrary. Since $D_q(\mu) \leqq 1$, there is a non-negative integer s such that

$$D_q(\mu)^{-1/k} \leqq 2^s < 2 D_q(\mu)^{-1/k}. \tag{1}$$

We put $r = 2^s$, $E = \{Q \in \mathcal{Q}_{k,r} : Q \cap \partial H \neq \emptyset\}$,

$$H_1 = \bigcup \{Q \in \mathcal{Q}_{k,r} : Q \subset H \text{ and } Q \cap \partial H = \emptyset\} = H \setminus \bigcup \{Q : Q \in E\},$$

and

$$H_2 = \bigcup \{Q \in \mathcal{Q}_{k,r} : Q \cap \operatorname{cl} H \neq \emptyset\} = H \cup \bigcup \{Q : Q \in E\}.$$

Then $H_1 \subset H \subset H_2$ and hence

$$\begin{aligned}(\mu(H_1) - \lambda_k(H_1)) + (\lambda_k(H_1) - \lambda_k(H)) \leqq \mu(H) - \lambda_k(H) \leqq \\ (\mu(H_2) - \lambda_k(H_2)) + (\lambda_k(H_2) - \lambda_k(H)),\end{aligned}$$

which implies

$$D(\mu; H) \leqq \max_{i=1,2} D(\mu; H_i) + \max_{i=1,2} |\lambda_k(H_i) - \lambda_k(H)|.$$

If $Q \in \mathcal{Q}_{k,r}$ then $|x-y| < \sqrt{k}/r$ for every $x, y \in Q$ and hence

$$H_2 \setminus H \subset \{x \in I^k \setminus H : \operatorname{dist}(x, H) < \sqrt{k}/r\},$$

and

$$H \setminus H_1 \subset \{x \in H : \operatorname{dist}(x, I^k \setminus H) < \sqrt{k}/r\}.$$

Therefore, by $H \in \mathcal{M}_b$ and by (1) we have

$$\max_{i=1,2} |\lambda_k(H_i) - \lambda_k(H)| \leqq b(\sqrt{k}/r) \leqq b(\sqrt{k} D_q(\mu)^{1/k}) \stackrel{\text{def}}{=} b_0,$$

and thus

$$D(\mu; H) \leqq \max_{i=1,2} D(\mu; H_i) + b_0. \tag{2}$$

Since $\lambda_k(\cup\{Q : Q \in E\}) = \lambda_k(H_2 \setminus H_1) \leqq 2b_0$, we have $|E| \leqq 2r^k b_0$. It is easy to see that $\partial H_i \cap \operatorname{int} I^k \subset \cup\{\partial Q : Q \in E\}$ $(i = 1, 2)$. Let λ_{k-1} denote the $k-1$-dimensional Hausdorff measure (surface area); then we have

$$\lambda_{k-1}(\partial H_i \cap \operatorname{int} I^k) \leqq \sum_{Q \in E} \lambda_{k-1}(\partial Q) = |E| 2kr^{1-k} \leqq 4krb_0 \qquad (i = 1, 2). \tag{3}$$

If $k = 1$ then this means that $\partial H_i \cap (0, 1)$ contains at most $4rb_0$ points and hence H_i is the union of at most $4rb_0$ non-overlapping intervals. Therefore

$$D(\mu; H_i) \leqq 4rb_0 D(\mu) = 4rb_0 D_q(\mu) \tag{4}$$

since, for $k = 1$, $D(\mu) = D_q(\mu)$. By (1) we have $rD_q(\mu) < 2$, and hence (2) and (4) give $D(\mu; H) \leqq 9b_0$. This proves the theorem for $k = 1$.

Now, in order to estimate $D(\mu; H_i)$ in the case $k \geqq 2$, we shall need some additional notation.

If $\mathcal{A}$ is a system of sets, then we shall denote by $S(\mathcal{A})$ the closure of $\mathcal{A}$ under the operations of disjoint union and proper difference ($A \setminus B$, where $B \subset A$) with the restriction that each element of $\mathcal{A}$ can be used only once. This system can be defined inductively as follows. We put $S_0(\mathcal{A}) = \mathcal{A} \cup \{\emptyset\}$. If $S_n(\mathcal{A})$ has been defined then we put a set A into $S_{n+1}(\mathcal{A})$ if and only if at least one of the following conditions is satisfied: (i) $A \in S_n(\mathcal{A})$; (ii) there are disjoint subsystems $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ and sets $A_1 \in S_n(\mathcal{A}_1)$, $A_2 \in S_n(\mathcal{A}_2)$ such that $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$; (iii) there are disjoint subsystems $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ and sets $A_1 \in S_n(\mathcal{A}_1)$, $A_2 \in S_n(\mathcal{A}_2)$ such that $A = A_1 \setminus A_2$ and $A_2 \subset A_1$. This defines $S_n(\mathcal{A})$ for every $n \in \mathbf{N}$. Finally, we take $S(\mathcal{A}) = \cup_{n=0}^{\infty} S_n(\mathcal{A})$.

If $A_1, \ldots, A_n$ are measurable subsets of I^k and $H \in S(\{A_1, \ldots, A_n\})$, then we have

$$D(\mu; H) \leqq \sum_{i=1}^{n} D(\mu; A_i).$$

This is an easy consequence of the definition of $S(\{A_1, \dots, A_k\})$, and the fact that $A \cap B = \emptyset$ implies

$$\mu(A \cup B) - \lambda_k(A \cup B) = (\mu(A) - \lambda_k(A)) + (\mu(B) - \lambda_k(B))$$

and that $B \subset A$ implies

$$\mu(A \setminus B) - \lambda_k(A \setminus B) = (\mu(A) - \lambda_k(A)) - (\mu(B) - \lambda_k(B)).$$

We denote $Q_0 = rI^k$, then $Q_0 \in \mathcal{D}_k^{(s)}$. Let $K_1 = H_1$ if $\lambda_k(H_1) \leqq 1/2$ and let $K_1 = I^k \setminus H_1$ otherwise. Then rK_1 is a finite union of unit cubes, $rK_1 \subset \subset Q_0$ and $\lambda_k(rK_1) \leqq \lambda_k(Q_0)/2$. By Lemma 3.2 of [3], this implies that there are dyadic cubes $Q_1, \dots, Q_n \subset Q_0$ such that $rK_1 \in S(\{Q_1, \dots, Q_n\})$ and for every $m \in \mathbf{N}$,

$$n_m \overset{\text{def}}{=} \left| \left\{ i : 1 \leqq i \leqq n,\ Q_i \in \mathcal{D}_k^{(m)} \right\} \right| \leqq C_k \frac{\lambda_{k-1}(\partial(rK_1) \cap \operatorname{int} Q_0)}{2^{m(k-1)}},$$

where C_k only depends on k. Since $k \geqq 2$, this, together with (3) imply

$$n = \sum_{m=0}^{\infty} n_m \leqq 2C_k \lambda_{k-1}(\partial(rK_1) \cap \operatorname{int} Q_0) \leqq 2C_k r^{k-1} \lambda_{k-1}(\partial K_1 \cap \operatorname{int} I^k) =$$

$$2C_k r^{k-1} \lambda_{k-1}(\partial H_1 \cap \operatorname{int} I^k) \leqq 8C_k k r^k b_0. \tag{5}$$

Putting $T_i = r^{-1}Q_i$ $(i = 1, \dots, n)$ we have $K_1 \in S(\{T_1, \dots, T_n\})$ and hence, by (1) and (5),

$$D(\mu; H_1) = D(\mu; K_1) \leqq \sum_{i=1}^{n} D(\mu; T_i) \leqq n D_q(\mu) \leqq 8C_k k r^k b_0 D_q(\mu) \leqq 8C_k k 2^k b_0.$$

The same estimate holds for $D(\mu; H_2)$ and hence, by (2), the theorem is proved.

REMARKS AND PROBLEMS. The theorem by Niederreiter and Wills has applications in numerical integration (see [5], p. 982). Recently it found an application also in the theory of equidecomposable sets. In [4] we proved that if $A, B \subset \mathbf{R}^k$ are bounded measurable sets with the same positive measure and if the box dimension of ∂A and ∂B is less than k, then A and B are equidecomposable using translations. The proof is based on a sufficient condition of equidecomposability using the discrepancy of some special sequences ([4], Theorem 1), and, in order to apply this condition for the sets A, B in question, we need the estimate given by Niederreiter and Wills. It is interesting to note that both this sufficient condition and our theorem above use the same combinatorial result (Lemma 3.2 of [3]). In connection with this result the following question arises.

Suppose that the set $H \subset \mathbf{R}^k$ is the union of finitely many unit cubes. Is it true that H is the union of at most $C_k \lambda_{k-1}(\partial H)$ non-overlapping cubes (with a constant C_k only depending on k)?

There is another important problem concerning our theorem above, namely the question whether or not it is really an improvement of the result by Niederreiter and Wills. The same question could have been asked about W. Schmidt's theorem [8] stating $D(\mu; H) \leqq C_k D_q(\mu)^{1/k}$ for $H \in \mathcal{M}_b$ with $b(\varepsilon) = C\varepsilon$. The point is that replacing $D(\mu)$ by $D_q(\mu)$ leads to an improvement only if $D_q(\mu)$ is smaller than $D(\mu)$. But is it really smaller? When I raised this question in the Research Group on Uniformity and Irregularity of Partitions, Imre Ruzsa proved that for $k = 2$ the answer is no: there is an absolute constant C such that $D(\mu) \leqq C D_q(\mu)$ for every μ defined on I^2 (see [7]). His proof does not work in higher dimensions, so the following problem remains open.

Does there exist a constant C_k depending only on k such that every measure μ defined on I^k satisfies $D(\mu) \leqq C_k D_q(\mu)$?

REFERENCES

[1] Hlawka, E., Funktionen von beschränkter Variation in der Theorie der Gleichverteilung, *Ann. Mat. Pura Appl.* (4) **54** (1961), 325–334. *MR* **25** #3029

[2] Kuipers, L. and Niederreiter, H., *Uniform distribution of sequences*, Pure and Applied Mathematics, John Wiley & Sons, New-York–London, 1974. *MR* **54** #7415

[3] Laczkovich, M., Uniformly spread discrete sets in $\mathbf{R}^d$, *J. London Math. Soc.* (2) **46** (1992), 39–57. *MR* **93i**:11088

[4] Laczkovich, M., Decomposition of sets with small boundary, *J. London Math. Soc.* (2) **46** (1992), 58–64. *MR* **93i**:11089

[5] Niederreiter, H., Quasi-Monte Carlo Methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84** (1978), 957–1041. *MR* **80d**:65016

[6] Niederreiter, H. and Wills, J.M., Diskrepanz und Distanz von Massen bezüglich konvexer und Jordanscher Mengen, *Math. Z.* **144** (1975), 125–134. *MR* **51** #12763 and **53** #7996

[7] Ruzsa, I. Z., The discrepancy of rectangles and squares, *Österreichisch-Ungarisch-Slowakisches Kolloquium über Zahlentheorie* (Maria Trost, 1992), Grazer Math. Ber., 318, Karl-Franzens-Univ. Graz, Graz, 1993, 135–140. *MR* **94j**:11070

[8] Schmidt, W., Irregularities of distribution IX, *Acta Arith.* **27** (1975), 385–396. *MR* **51** #12768

(Received February 14, 1994)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALIZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 111–122

LOW-DISCREPANCY SEQUENCES AND NONARCHIMEDEAN DIOPHANTINE APPROXIMATIONS

H. NIEDERREITER

1. Introduction

We consider discrepancy theory in the classical setting, namely for finite and infinite sequences of points in an s-dimensional unit cube $I^s = [0,1)^s$. Although it has been customary to speak of "finite sequences" in the theory of discrepancy, we prefer the term "point set" since the discrepancy of a finite sequence does not depend on the order of the terms. Here "point set" means the same as "multiset" in combinatorics, i.e., a set in which the multiplicity of elements is taken into account. Instead of "infinite sequence" we will just say "sequence".

Let P be the point set consisting of the N points $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1} \in I^s, s \geqq 1$. For a subinterval J of I^s the counting function $A(J;P)$ is defined as the number of integers n with $0 \leqq n \leqq N-1$ and $\mathbf{x}_n \in J$.

DEFINITION 1. The *(star) discrepancy* of the point set P is defined by

$$D_N^*(P) = \sup_J \left| \frac{A(J;P)}{N} - \mathrm{Vol}(J) \right|,$$

where the supremum is extended over all subintervals J of I^s of the form $J = \prod_{i=1}^s [0, t_i)$. For a sequence S of elements of I^s we define $D_N^*(S)$ to be the (star) discrepancy of the point set consisting of the first N terms of S.

Informally, a point set P is called a *low-discrepancy point set* if $D_N^*(P)$ is small, where N is the given number of points in P. A sequence S is called a *low-discrepancy sequence* if $D_N^*(S)$ is small for all $N \geqq 1$. In the s-dimensional case, "small" is usually interpreted to mean $O(N^{-1}(\log N)^s)$. Background material on low-discrepancy point sets and sequences can be found in Hua and Wang [3] and Niederreiter [8], [17].

The most promising current methods for the construction of low-discrepancy point sets and sequences are based on the theory of (t, m, s)-nets and (t, s)-sequences. These are point sets and sequences, respectively, which

1991 *Mathematics Subject Classification.* Primary 11J61, 11K38; Secondary 11J70.

Key words and phrases. Discrepancy, nets, diophantine approximations of formal Laurent series, continued fractions.

show a very regular distribution behavior with regard to special classes of subintervals of I^s. In the following, let $s \geqq 1$ be a given dimension and let $b \geqq 2$ be a fixed integer.

DEFINITION 2. An *elementary interval in base b* is a subinterval J of I^s of the form

$$J = \prod_{i=1}^{s} [a_i b^{-d_i}, (a_i + 1) b^{-d_i})$$

with integers a_i and d_i for $1 \leqq i \leqq s$.

DEFINITION 3. Let $0 \leqq t \leqq m$ be integers. A *(t, m, s)-net in base b* is a point set P of b^m points in I^s such that $A(J; P) = b^t$ for every elementary interval J in base b with $\mathrm{Vol}(J) = b^{t-m}$.

DEFINITION 4. Let $t \geqq 0$ be an integer. A sequence $\mathbf{x}_0, \mathbf{x}_1, \ldots$ of points in I^s is a *(t,s)-sequence in base b* if for all integers $k \geqq 0$ and $m > t$ the point set consisting of the $\mathbf{x}_n$ with $kb^m \leqq n < (k+1)b^m$ is a (t, m, s)-net in base b.

Definitions 3 and 4 were introduced by Sobol' [18] in the case $b = 2$; the general definitions are due to Niederreiter [10]. It is easily seen that any (t, m, s)-net in base b is also a (u, m, s)-net in base b for $t \leqq u \leqq m$ and that any (t, s)-sequence in base b is also a (u, s)-sequence in base b for $u \geqq t$. Therefore, smaller values of t mean stronger regularity properties. Upper bounds for the discrepancy of (t, m, s)-nets and (t, s)-sequences have been established in [10]. These bounds are completely explicit, but for the sake of simplicity we give them here in an abbreviated form. The implied constants in the Landau symbols depend only on b and s.

THEOREM 1. *If P is a (t, m, s)-net in base b with $s \geqq 2$, then*

$$D_N^*(P) \leqq B(s, b) b^t N^{-1} (\log N)^{s-1} + O(b^t N^{-1} (\log N)^{s-2}).$$

If either $s = 2$ or $b = 2, s = 3, 4$, then

$$B(s, b) = \left(\frac{b-1}{2 \log b} \right)^{s-1},$$

otherwise

$$B(s, b) = \frac{1}{(s-1)!} \left(\frac{\lfloor b/2 \rfloor}{\log b} \right)^{s-1}.$$

THEOREM 2. *If S is a (t, s)-sequence in base b with $s \geqq 2$, then*

$$D_N^*(S) \leqq C(s, b) b^t N^{-1} (\log N)^s + O(b^t N^{-1} (\log N)^{s-1}) \text{ for } N \geqq 2.$$

If either $s = 2$ or $b = 2, s = 3, 4$, then

$$C(s, b) = \frac{1}{s} \left(\frac{b-1}{2 \log b} \right)^{s},$$

otherwise

$$C(s,b)=\frac{1}{s!}\cdot\frac{b-1}{2\lfloor b/2\rfloor}\left(\frac{\lfloor b/2\rfloor}{\log b}\right)^s.$$

There remains, of course, a crucial problem, namely that of constructing (t,m,s)-nets and (t,s)-sequences explicitly. Some special constructions were given by Sobol' [18], Srinivasan [19], and Faure [2]. In Section 2 we describe a general principle for the construction of (t,m,s)-nets and (t,s)-sequences. The most powerful applications of this general construction principle are based on the machinery of formal Laurent series over finite fields and lead, in particular, to diophantine approximation problems with respect to the nonarchimedean degree valuation on the field of formal Laurent series, as will be discussed in Section 3.

2. A general construction principle

A general principle for the construction of (t,m,s)-nets and (t,s)-sequences was introduced in Niederreiter [10]. Although this construction principle works for arbitrary bases, we now consider only the special case where the base b is a prime power q. Let F_q be the finite field of order q and let $Z_q=\{0,1,\ldots,q-1\}$ be the least residue system modulo q. For the construction of nets we fix integers $m\geqq 1$ and $s\geqq 1$ and we choose:

(i) bijections $\psi_r: Z_q\to F_q$ for $0\leqq r\leqq m-1$;

(ii) bijections $\eta_{ij}: F_q\to Z_q$ for $1\leqq i\leqq s, 1\leqq j\leqq m$;

(iii) elements $c_{jr}^{(i)}\in F_q$ for $1\leqq i\leqq s, 1\leqq j\leqq m, 0\leqq r\leqq m-1$.

For $n=0,1,\ldots,q^m-1$ let

$$n=\sum_{r=0}^{m-1}a_r(n)q^r \quad \text{with all } a_r(n)\in Z_q$$

be the digit expansion of n in base q. Put

$$x_n^{(i)}=\sum_{j=1}^{m}y_{nj}^{(i)}q^{-j} \quad \text{for } 0\leqq n<q^m, 1\leqq i\leqq s,$$

with

$$y_{nj}^{(i)}=\eta_{ij}\left(\sum_{r=0}^{m-1}c_{jr}^{(i)}\psi_r(a_r(n))\right)\in Z_q \quad \text{for } 0\leqq n<q^m, 1\leqq i\leqq s,\ 1\leqq j\leqq m,$$

and define the point set

$$\mathbf{x}_n=\left(x_n^{(1)},\ldots,x_n^{(s)}\right)\in I^s \quad \text{for } 0\leqq n<q^m. \tag{1}$$

We collect the elements $c_{jr}^{(i)} \in F_q$ into the system C of vectors

$$\mathbf{c}_j^{(i)} = \left(c_{j0}^{(i)}, \ldots, c_{j,m-1}^{(i)}\right) \in F_q^m \quad \text{for } 1 \leqq i \leqq s, 1 \leqq j \leqq m.$$

DEFINITION 5. For the system $C = \{\mathbf{c}_j^{(i)} : 1 \leqq i \leqq s, 1 \leqq j \leqq m\}$ as above let $\varrho(C)$ be the largest integer d such that any system $\{\mathbf{c}_j^{(i)} : 1 \leqq j \leqq d_i, 1 \leqq i \leqq s\}$ with $0 \leqq d_i \leqq m$ for $1 \leqq i \leqq s$ and $\sum_{i=1}^{s} d_i = d$ is linearly independent over F_q (here the empty system is viewed as linearly independent).

Here we have followed Definition 4.27 in [17]. Note that this definition is slightly different from that in the original paper [10]. In fact, if $\varrho_1(C)$ is the quantity introduced in Definition 6.8 in [10] and $\varrho(C)$ is as in Definition 5 above, then $\varrho_1(C) = \varrho(C) + 1$. We always have $0 \leqq \varrho(C) \leqq m$. The following result was shown in [10, Theorem 6.10] (see also [17, Theorem 4.28]).

LEMMA 1. *The point set* (1) *is a* (t, m, s)*-net in base* q *with* $t = m - \varrho(C)$.

From Lemma 1 and the general discrepancy bound in Theorem 1 it follows that if P is the point set (1) and $s \geqq 2$, then

$$D_N^*(P) \leqq B(s,q) q^{-\varrho(C)} (\log N)^{s-1} + O\left(q^{-\varrho(C)} (\log N)^{s-2}\right),$$

where the implied constant depends only on q and s. According to [16, Theorem 1] we have the lower bound

$$D_N^*(P) \geqq \frac{q-1}{3q} q^{-\varrho(C)}.$$

These bounds show that $D_N^*(P)$ is small if and only if $\varrho(C)$ is large. A general study of how large one can make $\varrho(C)$ was carried out in [15].

The analogous method for the construction of (t, s)-sequences proceeds as follows. For a given $s \geqq 1$ we choose:
(i) bijections $\psi_r : Z_q \to F_q$ for $r \geqq 0$, with $\psi_r(0) = 0$ for all sufficiently large r;
(ii) bijections $\eta_{ij} : F_q \to Z_q$ for $1 \leqq i \leqq s$ and $j \geqq 1$;
(iii) elements $c_{jr}^{(i)} \in F_q$ for $1 \leqq i \leqq s, j \geqq 1, r \geqq 0$.

For $n = 0, 1, \ldots$ let

$$n = \sum_{r=0}^{\infty} a_r(n) q^r$$

be the digit expansion of n in base q, so that all $a_r(n) \in Z_q$ and $a_r(n) = 0$ for all sufficiently large r. Put

$$x_n^{(i)} = \sum_{j=1}^{\infty} y_{nj}^{(i)} q^{-j} \quad \text{for } n \geqq 0, 1 \leqq i \leqq s,$$

with

$$y_{nj}^{(i)} = \eta_{ij}\left(\sum_{r=0}^{\infty} c_{jr}^{(i)}\psi_r(a_r(n))\right) \in Z_q \quad \text{for } n \geqq 0, 1 \leqq i \leqq s, j \geqq 1,$$

and define the sequence

$$\mathbf{x}_n = \left(x_n^{(1)}, \ldots, x_n^{(s)}\right) \quad \text{for } n = 0, 1, \ldots. \tag{2}$$

We impose the condition that for each n and i we have $y_{nj}^{(i)} < q-1$ for infinitely many j, and in this way we guarantee that all points $\mathbf{x}_n$ are in I^s. The following result was established in [10, Theorem 6.23] (see also [17, Theorem 4.36]).

LEMMA 2. *Let $t \geqq 0$ be an integer. If for each integer $m > t$ the system $C^{(m)}$ given by*

$$\mathbf{c}_j^{(i)} = \left(c_{j0}^{(i)}, \ldots, c_{j,m-1}^{(i)}\right) \in F_q^m \quad \text{for } 1 \leqq i \leqq s, 1 \leqq j \leqq m,$$

satisfies $\varrho(C^{(m)}) \geqq m-t$, then the sequence (2) is a (t,s)-sequence in base q.

Constructions of (t,m,s)-nets and (t,s)-sequences based on the principles above have been carried out in Niederreiter [10], [11], [12]. Further applications of these principles will be given in the next section.

3. Constructions based on formal Laurent series

The method of formal Laurent series for the construction of (t,m,s)-nets and (t,s)-sequences was introduced by Niederreiter [11]. The idea is to use the coefficients of formal Laurent series over F_q for the construction of the elements $c_{jr}^{(i)} \in F_q$ that are needed in the general constructions in Section 2. Let $F_q((x^{-1}))$ be the field of formal Laurent series over F_q in the variable x^{-1}. Every nonzero $L \in F_q((x^{-1}))$ has the form

$$L = \sum_{k=w}^{\infty} t_k x^{-k},$$

where w is an integer, all $t_k \in F_q$, and $t_w \neq 0$. If we put $\nu(L) = -w$ and $\nu(0) = -\infty$, then ν is the degree valuation, which is a nonarchimedean discrete exponential valuation on $F_q((x^{-1}))$. Note that $F_q((x^{-1}))$ contains the field of rational functions over F_q.

The following construction of (t,m,s)-nets based on Laurent series expansions of rational functions over F_q was given in Niederreiter [16]. For

$s \geqq 2$ let $f \in F_q[x]$ with $\deg(f) = m \geqq 1$ and let $g_1, \ldots, g_s \in F_q[x]$. Consider the expansions

$$\frac{g_i(x)}{f(x)} = \sum_{k=w_i}^{\infty} u_k^{(i)} x^{-k} \in F_q((x^{-1})) \quad \text{for } 1 \leqq i \leqq s,$$

where $w_i \leqq 1$ for $1 \leqq i \leqq s$. Then define

$$c_{jr}^{(i)} = u_{r+j}^{(i)} \in F_q \quad \text{for } 1 \leqq i \leqq s, 1 \leqq j \leqq m, 0 \leqq r \leqq m-1.$$

With this choice of the elements $c_{jr}^{(i)}$ we then apply the general construction principle in Section 2. This yields the point set (1) consisting of q^m points in I^s. We denote this point set by $P(\mathbf{g}, f)$, where we write $\mathbf{g} = (g_1, \ldots, g_s) \in F_q[x]^s$. Put

$$\varrho(\mathbf{g}, f) = s - 1 + \min \sum_{i=1}^{s} \deg(h_i), \tag{3}$$

where the minimum is extended over all nonzero $(h_1, \ldots, h_s) \in F_q[x]^s$ with $\deg(h_i) < m$ for $1 \leqq i \leqq s$ and $\sum_{i=1}^{s} h_i g_i \equiv 0 \bmod f$, and where we use the convention $\deg(0) = -1$. Then, using Lemma 1, the following result was shown in [16].

THEOREM 3. *The point set $P(\mathbf{g}, f)$ is a (t, m, s)-net in base q with $t = m - \varrho(\mathbf{g}, f)$.*

From Theorems 1 and 3 we obtain a bound for the discrepancy of the point set $P(\mathbf{g}, f)$. In [16] it was also proved that if q is prime (so that F_q and Z_q can be identified), if every bijection η_{ij} is the identity map, and if $s \geqq 2$ and $f \in F_q[x]$ with $\deg(f) = m \geqq 1$ are fixed, then "on the average" we have $D_N^*(P(\mathbf{g}, f)) = O(N^{-1}(\log N)^s)$ with an implied constant depending only on s, where the average is taken over all $\mathbf{g} = (g_1, \ldots, g_s)$ with $\gcd(g_i, f) = 1$ and $\deg(g_i) < m$ for $1 \leqq i \leqq s$. For the special case $f(x) = x^m$ it was recently shown by Larcher [4] that by a suitable choice of $\mathbf{g}$ we can always obtain

$$D_N^*(P(\mathbf{g}, f)) = O(N^{-1}(\log N)^{s-1} \log\log(N+1))$$

with an implied constant depending only on s.

In the case $s = 2$ there is a connection between the quantity $\varrho(\mathbf{g}, f)$ in (3) and continued fractions for rational functions over F_q, where q is again an arbitrary prime power. Let $\mathbf{g} = (1, g)$ with $g \in F_q[x], \deg(g) < m$, and $\gcd(g, f) = 1$, and let

$$\frac{g}{f} = [0; A_1, A_2, \ldots, A_u]$$

be the continued fraction expansion of the rational function g/f, with partial quotients $A_d \in F_q[x]$ satisfying $\deg(A_d) \geqq 1$ for $1 \leqq d \leqq u$. Put

$$K\left(\frac{g}{f}\right) = \max_{1 \leqq d \leqq u} \deg(A_d). \tag{4}$$

Then it was shown in [16] that

$$\varrho(\mathbf{g}, f) = m + 1 - K\left(\frac{g}{f}\right). \tag{5}$$

The quantity $K(g/f)$ was studied in detail in [9]. The formula (5) makes it clear that there is a connection between the construction of (t, m, s)-nets in this section and diophantine approximations in $F_q((x^{-1}))$.

We now introduce an analogous construction for sequences. A formal Laurent series $L \in F_q((x^{-1}))$ is called *irrational* if it is not the expansion of a rational function over F_q. For $s \geqq 1$ we choose irrational $L_1, \ldots, L_s \in \in F_q((x^{-1}))$, say

$$L_i = \sum_{k=w_i}^{\infty} u_k^{(i)} x^{-k} \quad \text{for } 1 \leqq i \leqq s,$$

where $w_i \leqq 1$ for $1 \leqq i \leqq s$. Then define

$$c_{jr}^{(i)} = u_{r+j}^{(i)} \in F_q \quad \text{for } 1 \leqq i \leqq s, j \geqq 1, r \geqq 0. \tag{6}$$

With this choice of the $c_{jr}^{(i)}$ we then apply the general construction principle in Section 2, which yields the sequence (2). The assumption in Lemma 3 below guarantees that the condition imposed after (2) is satisfied; here η_{ij}^{-1} denotes the inverse map of the bijection η_{ij}.

LEMMA 3. *If the bijections η_{ij} are such that $\eta_{ij}^{-1}(q-1)$ is $\neq 0$ and independent of j for all sufficiently large j, then for each $n \geqq 0$ and $1 \leqq i \leqq s$ we have $y_{nj}^{(i)} < q-1$ for infinitely many j.*

PROOF. Suppose that for some n and i we had $y_{nj}^{(i)} = q-1$ for all sufficiently large j. Since for a suitable integer $R_n \geqq 0$ we have $\psi_r(a_r(n)) = 0$ for all $r > R_n$, it follows from (6) and the definition of the $y_{nj}^{(i)}$ that

$$\sum_{r=0}^{R_n} u_{r+j}^{(i)} \psi_r(a_r(n)) = \eta_{ij}^{-1}(q-1) \quad \text{for all sufficiently large } j.$$

By hypothesis, there exists a nonzero $m_i \in F_q$ such that $\eta_{ij}^{-1}(q-1) = m_i$ for all sufficiently large j. Thus, with a suitable integer $j_0 \geqq 1$ we have

$$\sum_{r=0}^{R_n} \psi_r(a_r(n)) u_{r+j}^{(i)} = m_i \quad \text{for all } j \geqq j_0.$$

Hence, if we put $v_k = u_{k+j_0}^{(i)}$ for $k \geq 0$, then the sequence $v_0, v_1, \ldots$ of elements of F_q satisfies a nontrivial linear recurrence relation and is thus ultimately periodic (see [6, Ch. 8]). Consequently, the sequence $u_1^{(i)}, u_2^{(i)}, \ldots$ is ultimately periodic, and so L_i is the expansion of a rational function over F_q, which is a contradiction. □

We assume henceforth that the condition in Lemma 3 is satisfied. Then the construction above yields the sequence (2) of points in I^s, and we denote this sequence by $S(L_1, \ldots, L_s)$. A fairly detailed analysis of such sequences can be carried out in the case $s = 1$, by again using continued fractions. For an irrational $L \in F_q((x^{-1}))$ let

$$L = [A_0; A_1, A_2, \ldots]$$

be its continued fraction expansion, with partial quotients $A_d \in F_q[x], d = 0, 1, \ldots$, satisfying $\deg(A_d) \geq 1$ for $d \geq 1$. In the usual notation, let

$$\frac{P_d}{Q_d} = [A_0; A_1, \ldots, A_d]$$

be the dth convergent to L. Then we have $\deg(Q_0) = 0$ and

$$\deg(Q_d) = \sum_{k=1}^{d} \deg(A_k) \quad \text{for } d \geq 1,$$

and furthermore

$$\nu\left(L - \frac{P_d}{Q_d}\right) = -\deg(Q_d) - \deg(Q_{d+1}) \quad \text{for } d \geq 0;$$

compare with [13]. We need the following result on best diophantine approximations to L.

LEMMA 4. *If $h \in F_q[x]$ with $0 \leq \deg(h) < \deg(Q_{d+1})$ for some $d \geq 0$, then*

$$\nu\left(L - \frac{b}{h}\right) \geq \nu\left(L - \frac{P_d}{Q_d}\right) \quad \textit{for all } b \in F_q[x].$$

PROOF. Suppose that for some $b \in F_q[x]$ we have

$$\nu\left(L - \frac{b}{h}\right) < \nu\left(L - \frac{P_d}{Q_d}\right).$$

Then

$$\begin{aligned} \nu(hL - b) &< \nu\left(L - \frac{P_d}{Q_d}\right) + \nu(h) \\ &= -\deg(Q_d) - \deg(Q_{d+1}) + \deg(h) < 0, \end{aligned} \tag{7}$$

and so it follows from [13, Lemma 3] that $h = \sum_{k=j}^{m} C_k Q_k$ with $C_k \in F_q[x]$ and $\deg(C_k) < \deg(A_{k+1})$ for $j \leqq k \leqq m$ and $C_j \neq 0$, and that

$$\nu(hL - b) = \deg(C_j) - \deg(Q_{j+1}) \geqq -\deg(Q_{j+1}).$$

On the other hand, from (7) we get $\nu(hL - b) < -\deg(Q_d)$, and so we must have $j \geqq d$. Since $\deg(h) < \deg(Q_{d+1})$, we cannot have $j \geqq d+1$, thus $j = d$ and $h = C_d Q_d$. This implies

$$\begin{aligned}\deg(C_d) - \deg(Q_{d+1}) &= \nu(hL - b) < -\deg(Q_d) - \deg(Q_{d+1}) + \deg(h) \\ &= \deg(C_d) - \deg(Q_{d+1}),\end{aligned}$$

which is a contradiction. □

In analogy with (4) we put

$$K(L) = \sup_{d \geqq 1} \deg(A_d),$$

where we may have $K(L) = \infty$.

THEOREM 4. *If $K(L) < \infty$, then the sequence $S(L)$ is a $(t, 1)$-sequence in base q with $t = K(L) - 1$.*

PROOF. Let $L = \sum_{k=w}^{\infty} u_k x^{-k}$ with $w \leqq 1$ and let $c_{jr} = u_{r+j}$ as in (6). By Lemma 2 it suffices to verify that for each integer $m > t$ the vectors

$$\mathbf{c}_j = (c_{j0}, \ldots, c_{j,m-1}) \in F_q^m, \quad 1 \leqq j \leqq m - t,$$

are linearly independent over F_q. Suppose that for some $m > t$ we had

$$\sum_{j=1}^{m-t} h_j \mathbf{c}_j = \mathbf{0} \in F_q^m,$$

where not all $h_j \in F_q$ are 0. Then

$$\sum_{j=1}^{m-t} h_j u_{r+j} = 0 \quad \text{for } 0 \leqq r \leqq m - 1. \tag{8}$$

With $h(x) = \sum_{j=1}^{m-t} h_j x^{j-1} \in F_q[x]$ we get

$$\begin{aligned}hL &= \left(\sum_{j=1}^{m-t} h_j x^{j-1}\right)\left(\sum_{k=w}^{\infty} u_k x^{-k}\right) = \sum_{j=1}^{m-t} h_j \sum_{k=w}^{\infty} u_k x^{-k+j-1} \\ &= \sum_{j=1}^{m-t} h_j \sum_{r=w-j}^{\infty} u_{r+j} x^{-r-1},\end{aligned}$$

and so by (8) the coefficient of x^{-r-1} in hL is 0 for $0 \leqq r \leqq m-1$. Thus, $\nu(hL-b) < -m$ for a suitable $b \in F_q[x]$. Since $\deg(h) \leqq m-t-1$, it follows that

$$\deg(h) + \nu(hL-b) < -t-1 = -K(L). \tag{9}$$

On the other hand, there exists a $d \geqq 0$ with $\deg(Q_d) \leqq \deg(h) < \deg(Q_{d+1})$, and so Lemma 4 yields

$$\deg(h) + \nu(hL-b) = 2\deg(h) + \nu\left(L - \frac{b}{h}\right) \geqq 2\deg(Q_d) + \nu\left(L - \frac{P_d}{Q_d}\right)$$
$$= \deg(Q_d) - \deg(Q_{d+1}) = -\deg(A_{d+1}) \geqq -K(L).$$

This is a contradiction to (9). □

EXAMPLE. Let $q=2$ and $L = \sum_{j=1}^{\infty} x^{1-2^j} \in F_2((x^{-1}))$. Then

$$L^2 = \sum_{j=1}^{\infty} x^{2-2^{j+1}} = x \sum_{j=2}^{\infty} x^{1-2^j} = xL + 1,$$

and so $L = x + L^{-1}$. This yields the periodic continued fraction expansion

$$L = [0; x, x, \ldots].$$

Thus, L is irrational with $K(L) = 1$. It follows from Theorem 4 that $S(L)$ is a $(0,1)$-sequence in base 2.

For an arbitrary irrational $L \in F_q((x^{-1}))$, Larcher and Niederreiter [5] have obtained a bound for the discrepancy of the sequence $S(L)$ in terms of the continued fraction expansion. If $q^{\deg(Q_{d-1})} < N \leqq q^{\deg(Q_d)}$ for some $d \geqq 1$, then

$$D_N^*(S(L)) \leqq cN^{-1} \sum_{k=1}^{d} q^{\deg(A_k)}$$

with an absolute constant c. Together with the probabilistic theory of continued fractions for formal Laurent series developed in Niederreiter [14], this yields probabilistic results on the order of magnitude of $D_N^*(S(L))$ for "almost all" $L \in F_q((x^{-1}))$, in the sense of an appropriate probability measure.

Now we consider sequences $S(L_1, \ldots, L_s)$ with any $s \geqq 1$. The following result of Larcher and Niederreiter [5] establishes a connection with simultaneous diophantine approximations in $F_q((x^{-1}))$. For any $L \in F_q((x^{-1}))$ let $\mathrm{Fr}(L)$ denote the *fractional part* of L, i.e., the part of L containing only negative exponents.

THEOREM 5. *If the irrationals $L_1, \ldots, L_s \in F_q((x^{-1}))$ are such that for some integer $t \geqq 0$ we have*

$$\nu\left(\mathrm{Fr}\left(\sum_{i=1}^{s} h_i L_i\right)\right) \geqq -s-t-\sum_{i=1}^{s} \deg(h_i)$$

for all nonzero $(h_1, \ldots, h_s) \in F_q[x]^s$, *then the sequence* $S(L_1, \ldots, L_s)$ *is a* (t, s)*-sequence in base* q.

An s-tuple $(L_1, \ldots, L_s)$ is said to be of *constant type* if the diophantine condition in Theorem 5 holds for some $t \geqq 0$. For $s = 1$ the irrationals L of constant type are exactly those with $K(L) < \infty$. For any $s \geqq 2$, Armitage [1] claimed the construction of an s-tuple of constant type. It was later shown by Taussat [20] that none of these s-tuples is of constant type. This leads to the first of the open problems that we pose in conclusion.

PROBLEM 1. Determine whether there exists an s-tuple $(L_1, \ldots, L_s)$ of constant type for $s \geqq 2$.

PROBLEM 2. Characterize the polynomials $f \in F_q[x]$ with $\deg(f) \geqq 1$ for which there exists a $g \in F_q[x]$ with $\gcd(g, f) = 1$ and $K(g/f) = 1$, where $K(g/f)$ is given by (4). Partial results can be found in Mesirov and Sweet [7] and Niederreiter [9].

PROBLEM 3. Given $s \geqq 2$ and $f \in F_q[x]$ with $\deg(f) \geqq 1$, develop methods for the explicit construction of s-tuples $\mathbf{g} \in F_q[x]^s$ with a large value of $\varrho(\mathbf{g}, f)$, where $\varrho(\mathbf{g}, f)$ is as in (3). Here, "large" means as close as possible to $\deg(f)$.

REFERENCES

[1] ARMITAGE, J. V., An analogue of a problem of Littlewood, *Mathematika* **16** (1969), 101–105. *MR* **42**#1768a; Corrigendum and addendum, *ibid.* **17** (1970), 173–178. *MR* **42**#1768b

[2] FAURE, H., Discrépance de suites associées à un système de numération (en dimension s), *Acta Arith.* **41** (1982), 337–351. *MR* **84m**:10050

[3] HUA, L. K. and WANG, Y., *Applications of number theory to numerical analysis*, Springer, Berlin, 1981. *MR* **83g**:10034

[4] LARCHER, G., Nets obtained from rational functions over finite fields, *Acta Arith.* **63** (1993), 1–13.

[5] LARCHER, G. and NIEDERREITER, H., Kronecker-type sequences and nonarchimedean diophantine approximations, *Acta Arith.* **63** (1993), 379–396.

[6] LIDL, R. and NIEDERREITER, H., *Finite fields*, Encyclopedia of Mathematics and its Applications, 20, Addison-Wesley, Reading, Mass., 1983. *MR* **86c**:11106

[7] MESIROV, J. P. and SWEET, M.M., Continued fraction expansions of rational expressions with irreducible denominators in characteristic 2, *J. Number Theory* **27** (1987), 144–148. *MR* **89a**:11016

[8] NIEDERREITER, H., Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84** (1978), 957–1041. *MR* **80d**:65016

[9] NIEDERREITER, H., Rational functions with partial quotients of small degree in their continued fraction expansion, *Monatsh. Math.* **103** (1987), 269–288. *MR* **88h**:12002

[10] NIEDERREITER, H., Point sets and sequences with small discrepancy, *Monatsh. Math.* **104** (1987), 273–337. *MR* **89c**:11120

[11] NIEDERREITER, H., Low-discrepancy and low-dispersion sequences, *J. Number Theory* **30** (1988), 51–70. *MR* **89k**:11064

[12] NIEDERREITER, H., Quasi-Monte Carlo methods for multidimensional numerical integration, *Numerical Integration III* (Oberwolfach, 1987), Internat. Series of Numer. Math., Vol. **85**, Birkhäuser, Basel, 1988, 157–171. *MR* **91f**:65008

[13] NIEDERREITER, H., Sequences with almost perfect linear complexity profile, *Advances in Cryptology – EUROCRYPT '87* (Amsterdam, 1987), Lecture Notes in Computer Science, Vol. **304**, Springer, Berlin, 1988, 37–51. *Zbl* **651**.94003

[14] NIEDERREITER, H., The probabilistic theory of linear complexity, *Advances in Cryptology – EUROCRYPT '88* (Davos, 1988), Lecture Notes in Computer Science, Vol. **330**, Springer, Berlin, 1988, 191–209. *MR* **90d**:11138

[15] NIEDERREITER, H., A combinatorial problem for vector spaces over finite fields, *Discrete Math.* **96** (1991), 221-228. *MR* **92j**:11150

[16] NIEDERREITER, H., Low-discrepancy point sets obtained by digital constructions over finite fields, *Czechoslovak Math. J.* **42** (1992), 143–166. *MR* **93c**:11055

[17] NIEDERREITER, H., *Random number generation and quasi-Monte Carlo methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, 63, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. *MR* **93h**:65008

[18] SOBOL', I. M., Distribution of points in a cube and the approximate evaluation of integrals, *Ž. Vyčisl. Mat. i Mat. Fiz.* **7** (1967), 784–802 (in Russian). *MR* **36**#2321

[19] SRINIVASAN, S., On two-dimensional Hammersley's sequences, *J. Number Theory* **10** (1978), 421–429. *MR* **80f**:10065

[20] TAUSSAT, Y., Approximation diophantienne dans un corps de séries formelles, Thèse, Univ. de Bordeaux, 1986.

(Received February 14, 1994)

INSTITUT FÜR INFORMATIONSVERARBEITUNG
ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN
SONNENFELSGASSE 19
A–1010 WIEN
AUSTRIA

e-mail: nied@qiinfo.oeaw.ac.at

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 123–125

FEW MULTIPLES OF MANY PRIMES

I. Z. RUZSA

Erdős (1978) considers (among others) the following question. Let $Q = \{p_1 < p_2 < \ldots < p_n\}$ be a set of primes and I an interval of length N. Let $m(Q, I)$ be the number of those integers in I that are divisible by at least one p_j and put

$$m = m(Q, N) = \min m(Q, I),$$

where the minimum is taken over all intervals of length N. Estimate $m(Q, N)$.

If N is small, there may be no multiple in I. If $N \geqq p_n$, each prime must have at least one multiple, but they may all coincide. If $N \geqq 2p_n$ is assumed, then each prime has at least two multiples and under this assumption Erdős and Selfridge proved $m \geqq 2\sqrt{n+1}$. They also constructed examples where this is exact, and even $N > (3-\varepsilon)p_n$. The problem for $N > 3p_n$ is left open. We give an upper estimate of m for $N \sim \varrho p_n$ which is perhaps not very far from the actual size, though I cannot find any better lower estimate than that of Erdős and Selfridge given for $\varrho = 2$.

THEOREM. *Let $\varrho \geqq 3$ and write $k = [\varrho]$. There is a constant C depending only on ϱ such that for every $n > n_0(\varrho)$ there is a set $Q = \{p_1 < \ldots < p_n\}$ of primes satisfying*

$$m(Q, \varrho p_n) < C(n \log n)^{1-1/k}.$$

I cannot show the existence of infinitely many such sets Q.

PROOF. Define the numbers α and β by

$$\beta = 1/\varrho, \quad \alpha = \frac{1}{2}\left(\frac{1}{\varrho} + \frac{1}{k+1}\right).$$

These numbers satisfy

$$\frac{1}{k+1} < \alpha < \beta.$$

1991 *Mathematics Subject Classification.* Primary 11B75; Secondary 11A41, 11K99.

Key words and phrases. Prime numbers, combinatorial number theory, random construction.

Supported by Hungarian National Foundation for Scientific Research Grant No. 1901 and Zentrum für interdisziplinäre Forschung, Bielefeld, Germany.

Let $N = [Kn \log n]$, where the constant K will be specified later. Our primes will lie in the interval $(\alpha N, \beta N)$. The total number of primes in this interval is

$$L = \pi(\beta N) - \pi(\alpha N) \sim (\beta - \alpha)\frac{N}{\log N} \sim K(\beta - \alpha)n.$$

Let A be a random subset of the integers in $[1, N]$, where each integer is selected into A with probability $cN^{-1/k}$; later we specify c in terms of ϱ. The expectation of the cardinality of A is obviously $cN^{1-1/k}$.

We call a prime $p \in (\alpha N, \beta N)$ *useful*, if there is an integer a such that all integers

$$l \equiv a (\mathrm{mod}\, p), \quad l \in [1, N]$$

are in A, and *useless* otherwise. Let R be the set of useful primes; this is also a random set (it is uniquely determined by A).

Fix a prime $p \in (\alpha N, \beta N)$. We estimate the probability that it is useful. We consider only residues $a \in ((1 - k\alpha)N, \alpha N)$. The number of these a's is

$$\geqq ((k+1)\alpha - 1)N - 2 \geqq \gamma N$$

for large N with, say, $\gamma = ((k+1)\alpha - 1)/2 > 0$.

For such a number a the numbers $l \equiv a (\mathrm{mod}\, p)$ in $[1, N]$ are $a, a + p, \ldots, a + (k-1)p$. Indeed, $a < \alpha N < p$, so $a - p < 1$, while

$$a + kp > (1 - k\alpha)N + k\alpha N > N.$$

For each number $a + jp$ the probability that it is in A is $cN^{-1/k}$, thus the probability that all are in A is c^k/N. Hence

$$\mathbf{P}(a + jp \notin A \text{ for some } j) = 1 - c^k/N,$$

$$\mathbf{P}(p \text{ is useless}) \leqq \prod_a \mathbf{P}(a + jp \notin A \text{ for some } j)$$

$$\leqq \left(1 - c^k/N\right)^{\gamma N} \leqq e^{-\gamma c^k} < 1/2$$

if $c = (1/\gamma)^{1/k}$. We conclude that every prime is useful at least with probability $1/2$.

The expectation of $|R|$ is

$$\mathbf{E}|R| = \sum \mathbf{P}(p \text{ is useful}) \geqq L/2.$$

Since $|R|$ never exceeds L, we have

$$\mathbf{E}|R| \leqq L/4 + L\mathbf{P}(|R| > L/4),$$

and comparing these inequalities we obtain

$$\mathbf{P}(|R| > L/4) \geqq 1/4.$$

Since

$$\mathbf{E}|A| = cN^{1-1/k},$$

by Markov's inequality we have

$$\mathbf{P}\left(|A| > 4cN^{1-1/k}\right) < 1/4.$$

Consequently there must be a choice of A such that $|A| < 4cN^{1-1/k}$ and $|R| > L/4$. Since $L \sim K(\beta-\alpha)n$, with $K = 5/(\beta-\alpha)$ for large n we have $|R| > n$. Select a subset $Q \subset R$ with $|Q| = n$; this will be our set of primes.

For each $p \in Q$ let a_p be the number making p useful. Take an integer s that satisfies all the congruences

$$s \equiv -a_p (\mathrm{mod}\, p), \quad p \in Q.$$

In the interval $[s+1, s+N]$ all multiples of p lie in the set $s + A$, that is, their total number is

$$\leqq |A| \ll N^{1-1/k} \ll (n \log n)^{1-1/k}$$

as claimed. Since all the primes were $\leqq \beta N = N/\varrho$, the condition $N \geqq \varrho p_n$ is also satisfied.

REMARK. The same argument can be used to find an interval of length N, in which the total number of multiples of all primes $\alpha N \leqq p \leqq N$ is small. This problem was also proposed by Erdős. The difference is that we need all primes, not just a positive portion. This can be achieved if

$$\mathbf{P}(p \text{ is useless}) \leqq N^{-2}$$

for all $p \geqq \alpha N$. If $\alpha > 1/k$, k integer, then similar arguments show that this holds if the probability of selecting a number into A is $c(\log N)^{1/k} N^{-1/k}$ with a suitable c. In this way the minimum of the total number of multiples can be shown to be $O\left((\log N)^{1/k} N^{1-1/k}\right)$.

ACKNOWLEDGEMENTS. I profited much from discussing this problem with P. Erdős. The idea of applying a random construction to this problem was inspired by a remark of Erdős and Spencer (unpublished) about a related problem of mine concerning the maximal number of arithmetical progressions with different increments in a set of n numbers.

REFERENCE

[1] ERDŐS, P., Problems and results in combinatorial analysis and combinatorial number theory, *Proceedings of the Ninth Southeastern Conference on Combinatorics, Graph Theory, and Computing* (Florida Atlantic Univ., Boca Raton, Fla., 1978), *Congressus Numer.* **21**, Utilitas Math., Winnipeg, Man., 1978, 29–40. *MR* **80h:**05001

(Received February 14, 1994)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 127–148

SETS OF SUMS AND COMMUTATIVE GRAPHS

I. Z. RUZSA

1. Introduction

For two sets A, B (in any structure with an operation called addition) by their *sum* we mean the set

$$A+B=\{a+b: a\in A, b\in B\}.$$

We use $A-B$ similarly. For repeated addition we write

$$Ak = A+\cdots+A \quad (k \text{ times}),$$

in contrast to

$$kA=\{ka: k\in A\}.$$

We shall be interested in the "impact" of a fixed set B, that is, we want to know how "bigger" $A+B$ must be than A. There are different ways to measure size: cardinality of finite sets, density of infinite sets of integers, measure of sets of reals, and all give rise to similar problems. Here we focus on the finite case. We shall consider finite sets of integers, of lattice points, and of elements of an arbitrary commutative group.

For a fixed finite set B, we define its *impact function* by

$$\xi(n)=\xi_B(n)=\min\{|A+B|: |A|=n\}. \tag{1.1}$$

We want to understand the behaviour of this function. It turns out that the asymptotic behaviour of ξ depends only on a few simple properties of B. For instance, if $B\subset Z^d$ is a set of lattice points in d dimensions (which is not contained in any affine hyperplane), then we have

$$\xi(n)^{1/d}-n^{1/d}\to\frac{\mu(B^*)}{i}, \tag{1.2}$$

1991 *Mathematics Subject Classification.* Primary 11P99, 11B75; Secondary 11B13, 11P21, 05C90.

Key words and phrases. Sumsets, commutative graphs, additive impact.

Supported by Hungarian National Foundation for Scientific Research Grant No. 1901 and Zentrum für interdisziplinäre Forschung, Bielefeld, Germany.

where B^* is the convex hull of B, μ denotes volume (Lebesgue measure) and i is the index of the subgroup generated by $B - B$ in Z^d.

The basic tool in proving (1.2) and related results will be a graph-theoretic method developed by Plünnecke [13] and explained in detail in Section 6.

The paper consists of two parts. The first is a survey of related results, and the second is devoted to the proof of an effective version of (1.2).

We use this occasion to introduce a few basic conventions.

We shall work in a d-dimensional space R^d for a fixed d. Volume (Lebesgue measure) will be denoted by μ.

By a *lattice* we mean a discrete additive subgroup of R^d; *sublattice* means an additive subgroup of a lattice. For the most important lattice Z^d (points with integer coordinates) we preserve the notation L.

X^* denotes the convex hull of a set X.

$c_1, c_2, \ldots$ are constants that may depend on the dimension d.

PART I

2. Bases, essential components, density

From the related concepts of cardinality, density, measure, definitely cardinality is the simplest, but historically the first to get investigated was density.

In 1932, to establish his famous quasi-Goldbach theorem, Schnirelmann introduced a new concept of density. For a set A of integers we write

$$A(x) = |A \cap [1, x]|,$$

its *counting function*. (A may contain negative numbers, but they are ignored in the definition of $A(x)$.) The *Schnirelmann density* of A is the number

$$\sigma(A) = \inf_{n \in N} \frac{A(n)}{n}.$$

Observe that if $1 \notin A$, then automatically $\sigma(A) = 0$. In comparison with the *asymptotic density*

$$d(A) = \lim_{x \to \infty} \frac{A(x)}{x},$$

or the corresponding lower (upper) density in which the lim inf, lim sup is taken, there are fundamental differences. Asymptotic density is translation invariant and remains the same if finitely many numbers are added or

deleted. The Schnirelmann density does not have these nice properties. To compensate for this, Schnirelmann proved the inequality

$$\sigma(A+B) \geqq \sigma(A) + \sigma(B) - \sigma(A)\sigma(B) \tag{2.1}$$

for sets, at least one of which contains 0. Using (2.1), he proved that every set A with $\sigma(A) > 0$ is a *basis*, that is, $Ak \supset N$ for some k. He then applied this to the set

$$A = \{0, 1\} \cup \{p + q : p, q \text{ primes}\},$$

to prove the quasi-Goldbach theorem, which means that every sufficiently large number is the sum of k primes with a certain constant k.

A direct analog of (2.1) for the asymptotic density fails, as the example

$$A = B = \{\text{even numbers}\}$$

shows; here $d(A+B) = d(A) = d(B) = 1/2$. There are more complicated results that describe the connection of $d(A+B)$ to $d(A)$ and $d(B)$; for these and other improvements and variations of (2.1) we refer the reader to the book of Halberstam and Roth [6].

(2.1) has the consequence that if $0 \in B$ and $\sigma(A) > 0$, then

$$\sigma(A+B) > \sigma(A) \tag{2.2}$$

for every set A such that $0 < \sigma(A) < 1$. Sets B that have property (2.2) are called *essential components*. In 1935 Erdős proved that every basis satisfies (2.2) as well; he gave the estimate

$$\sigma(A+B) > \sigma(A) + \frac{\sigma(A)(1-\sigma(A))}{2k} \tag{2.3}$$

whenever B is a basis of order k, that is, $Bk \supset N$, and $0 \in B$.

We can define the (Schnirelmann) *impact function* of a set B for $0 \leqq x \leqq 1$ by

$$\xi(x) = \xi_B(x) = \inf\{\sigma(A+B) : \sigma(A) \geqq x\}. \tag{2.4}$$

Observe the analogy to (1.1); only the way of measurement of size is different. Every set with $\xi(x) > x$ for all $0 < x < 1$ is an essential component, and the converse is also true (though not completely obvious). (2.1) and (2.3) estimate the impact function of sets of positive Schnirelmann density and bases, respectively. The name itself is much younger than these results, it is probably due to Plünnecke [12] (Wirkungsfunktion).

Essential components that are not bases were first constructed by Linnik in 1942; see Ruzsa [18] for the thinnest possible ones.

A substantial improvement of (2.3) was given by Plünnecke [13]. He proved that for bases of order k we have

$$\sigma(A+B) \geqq \sigma(A)^{1-1/k}. \tag{2.5}$$

It is immediately seen that for $\sigma(A) \to 0$ (2.5) is of a much larger order of magnitude, but it is also better for any $\sigma(A)$.

Plünnecke's method of proving (2.5) will be one of our main tools and will be explained in detail in Section 6.

Various improvements of (2.5) were given for specific sets. Plünnecke [11] investigated squares and in general, sets of the form

$$B = \{[f(n)] : n \in N\}$$

where f is a smooth function. I considered squares and primes in [16], and prime-powers in [17]. Asymptotic density is also considered in these papers; there is often a striking difference between Schnirelmann and asymptotic density. For example, if P is the set of primes (and 0 and 1), then we have

$$\xi(x) \asymp \frac{1}{\log 2/x}$$

but

$$\xi'(x) \asymp \frac{1}{\log\log 3/x}$$

where ξ' denotes the impact function defined for asymptotic density (in (2.4) we replace ξ by d).

3. Results on finite sets

In comparison to the density, little attention was paid to the corresponding finite questions. We have the obvious inequality

$$|A+B| \geqq |A| + |B| - 1, \tag{3.1}$$

in which equality holds if and only if A and B are arithmetic progressions with a common difference. Freiman devoted a book [1,2; see also 4] to the important special case when $A = B$. His deep main theorem essentially characterizes those sets A that satisfy

$$|A+A| \leqq c|A|.$$

He also improves (3.1) for $A = B \subset R^d$. Assuming that A is *proper d-dimensional*, that is, it is not contained in an affine hyperplane, he shows

$$|A+A| \geqq (d+1)|A| - \frac{d(d+1)}{2}, \tag{3.2}$$

which is best possible in this generality.

Freiman, Heppes and Uhrin [5] consider another important subcase, namely $B = -A$. They prove the analog of (3.2):

$$|A - A| \geqq (d+1)|A| - \frac{d(d+1)}{2}. \tag{3.3}$$

However, unlike (3.2), (3.3) is probably not optimal for large $|A|$.

I proved [23] the following common generalization of (3.2) and (3.3). Let $A, B \subset R^d$ and assume that $A + B$ is proper d-dimensional (that is, A and B are not contained in parallel affine hyperplanes). Assume that $|A| = m \leqq |B| = n$. Then we have

$$|A + B| \geqq n + dm - \frac{d(d+1)}{2}. \tag{3.4}$$

Here the constant term can be improved to d if $n - m \geqq d$ (which is also exact).

The papers Freiman-Pigaev [3], Ruzsa [14, 15, 19, 20, 21] treat related questions in which not the structure of the sets is connected with the cardinality of sumsets, but different cardinalities to each other. For instance, if $|A| = n$, then with the notation $s = |A + A|/n$ we have

$$n\sqrt{s} \leqq |A - A| \leqq ns^2.$$

A common feature of (3.1–4) is that only the size of the sets is used to estimate the size of the sum. Now we plan to explore the dependence of the impact function on the structure, not just the size of the set B. We shall see that its behaviour at infinity depends essentially on one simple parameter only, the volume of the convex hull.

The main results of this paper sound as follows.

THEOREM 3.1. *Let $B \subset Z^d$. Assume that B is not contained in an affine hyperplane and define v by*

$$v = \mu(B^*)/i,$$

where i is the index of the sublattice generated by $B - B$ in L. We have

$$\xi(n)^{1/d} - n^{1/d} \to v^{1/d}. \tag{3.5}$$

An equivalent formulation of (3.5) is

$$\xi(n) = n + dv^{1/d}n^{1-1/d} + o(n^{1-1/d}). \tag{3.6}$$

We have the following effective improvements of (3.5–6).

THEOREM 3.2. *With the notations of the previous theorem, if $d \geqq 2$ and $n \geqq v$, we have*

$$\xi(n) \leqq n + dv^{1/d}n^{1-1/d} + c_1 v^{2/d} n^{1-2/d}, \tag{3.7}$$

$$\xi(n)^{1/d} - n^{1/d} \leqq v^{1/d} + c_2 v^{2/d} n^{-1/d}. \tag{3.8}$$

(c_1, c_2 depend on d.)

REMARK. For $d = 1$ we have the obvious inequality $\xi(n) \leqq n + v$.

THEOREM 3.3. *With the notations of the previous theorems and $m = |B|$, for large n we have*

$$\xi(n) \geqq n + dv^{1/d}n^{1-1/d} - c_3 v^{\frac{d+3}{2d}} m^{-1/2} n^{1-\frac{3}{2d}} \tag{3.9}$$

$$\xi(n)^{1/d} - n^{1/d} \geqq v^{1/d} - c_4 v^{\frac{d+3}{2d}} m^{-1/2} n^{-1/(2d)}. \tag{3.10}$$

Probably the real error terms are much smaller than my estimates.

It is possible to treat sets of points that are not necessarily lattice points, or even sets in an abstract commutative group. Let G be a commutative group; we use additive notation. Take a finite $B \subset G$, and assume for simplicity that $0 \in B$ (this simplifies the description; otherwise we just need to replace B by $B - b$ for any fixed $b \in B$, which does not change the impact). Let G' be the subgroup of G generated by B. By the structure theorem of finitely generated commutative groups, we have

$$G' \simeq H \times Z^d \tag{3.11}$$

for some integer d, where H is a finite group. Define the *dimension* of B as the integer d in (3.11). Let $\phi : G' \to Z^d$ be the homomorphism induced by the representation (3.11). Define the *volume* of B as

$$v = |H|\, \mu(\phi(B)^*). \tag{3.12}$$

It is easy to see that this volume does not depend on the particular choice of the homomorphism ϕ, which is typically not unique (and this also follows from the theorem below, since a function cannot have two different limits at ∞).

THEOREM 3.4. *Let G be an infinite commutative group, $B \subset G$ finite, and define d and v by* (3.11) *and* (3.12). *For $d \geqq 1$ we have*

$$\xi(n)^{1/d} - n^{1/d} \to v^{1/d}. \tag{3.13}$$

This theorem will be proved elsewhere.

If $d = 0$, then by taking unions of cosets of G' as A, we see that $\xi(n) = n$ for infinitely many n.

4. Results in one and two dimensions

If $d = 1$, then $\xi(n) - n$ is integer, so the only way it can be convergent is that $\xi(n) = n + v$ for $n > n_0(B)$. We describe the situation more exactly.

THEOREM 4.1. *Let B be a set of integers, $|B| = m$, and let v be the smallest number for which*

$$B \subset \{a, a+q, \dots, a+vq\}$$

for suitable a and q. For any set A with $|A| = n$ we have

$$|A+B| \geqq \min\left(n+v, \left(\sqrt{n} + \sqrt{\frac{m-1}{2}}\right)^2\right). \tag{4.1}$$

In particular, for $n \geqq v^2/(m-1)$ we have

$$|A+B| \geqq n+v.$$

This result will be proved elsewhere. Observe that the definition of v is consistent with the multidimensional case.

In two dimensions, the behaviour of ξ is already more varied. We use $\mathbf{e}_1, \mathbf{e}_2$ to denote the unit vectors.

THEOREM 4.2. *The impact function of the set $B = \{0, \mathbf{e}_1, \mathbf{e}_2\}$ satisfies*

$$\xi(n) - \sqrt{n} > \sqrt{v} \tag{4.2}$$

for all n.

The proof of this result will be published elsewhere.

THEOREM 4.3. *The impact function of the set*

$$B = \{0, \mathbf{e}_1, \mathbf{e}_2, -(\mathbf{e}_1 + \mathbf{e}_2)\} \tag{4.3}$$

satisfies

$$\xi(n) - \sqrt{n} < \sqrt{v} \tag{4.4}$$

for infinitely many n.

PROOF. In general, take a convex lattice polygon U with h lattice points on its boundary (including the vertices) and l lattice points in its interior. A familiar result from the geometry of numbers tells us that

$$\mu(U) = l + \frac{h}{2} - 1.$$

This implies that

$$|U \cap L| = h + l = \mu(U) + \frac{h}{2} + 1. \tag{4.5}$$

Put $B = U \cap L$; then $B^* = U$ and

$$v = \mu(B^*) = l + \frac{h}{2} - 1.$$

Now consider the sets $A_k = kU \cap L$. One easily sees that the number of points on the boundary of kU is kh, and then (4.5) implies

$$|A_k| = \mu(kU) + \frac{kh}{2} + 1 = k^2 v + \frac{kh}{2} + 1.$$

Since $A_k + B \subset A_{k+1}$, for $n = |A_k|$ we have

$$\xi(n) \leqq |A_{k+1}|.$$

A routine calculation shows that

$$\sqrt{|A_{k+1}|} - \sqrt{|A_k|} < \sqrt{v}$$

holds if (and only if)

$$h^2 < 16v. \tag{4.6}$$

The set B defined by (4.3) and its convex hull $U = B^*$ give $v = 3/2$, $h = 3$, thus (4.6) is satisfied.

I cannot decide what happens if (4.6) fails, except for a few simple sets like the one in Theorem 4.2. If (4.6) holds, we learned that $\xi(n) - \sqrt{n} < \sqrt{v}$ for infinitely many n. I cannot decide whether there is a set such that $\xi(n) - \sqrt{n} < \sqrt{v}$ for all n.

5. Results on measure

Take now (Borel) measurable sets in R^d. A famous inequality of Brunn, Minkowski, Lusternik (which also will play an important role in our proof) asserts that

$$\mu(A + B)^{1/d} \geqq \mu(A)^{1/d} + \mu(B)^{1/d}, \tag{5.1}$$

with equality only if A, B are homothetic convex sets. If we introduce the impact function by

$$\xi(x) = \inf\{\mu(A + B) : \mu(A) = x\},$$

(this is the third different concept for which we use the same name and letter), then this can be reformulated as

$$\xi(x)^{1/d} \geqq x^{1/d} + \mu(B)^{1/d}. \tag{5.2}$$

For this concept I can prove the following analog of Theorem 3.1.

THEOREM 5.2. *Let $B \subset R^d$, and assume that $\mu(B) > 0$. With $v = \mu(B^*)$ we have*

$$\xi(x)^{1/d} - x^{1/d} \to v^{1/d}. \tag{5.3}$$

A proof of this result will be published elsewhere.

Comparing this with the Brunn-Minkowski inequality, this tells that for large x, B behaves as if it filled its whole compact hull. I also have an analog of the estimate in Theorem 3.3. The estimate of Theorem 3.2 can be replaced by the obvious

$$\xi(x)^{1/d} - x^{1/d} \leqq v^{1/d},$$

which can be shown by taking sets homothetic to B^* in the place of A.

In [22] I prove the following measure analog of Theorem 4.1.

THEOREM 5.2. *Let A, B be bounded Borel sets of reals. Write $\mu(A) = a$, $\mu(B) = b$, $\mu(B^*) = v$. We have*

$$\mu(A+B) \geqq \min\left(a+v, \left(\sqrt{a} + \sqrt{b/2}\right)^2\right). \tag{5.4}$$

In particular, if

$$a \geqq \frac{L^2}{2b}, \tag{5.5}$$

then

$$\mu(A+B) \geqq a + v. \tag{5.6}$$

Observe that here v has a simple interpretation: it is equal to $\max B - \min B$, the diameter of B. I can also show that in 2 dimensions, $\sqrt{\xi(x)} - \sqrt{x}$ may not reach $\sqrt{v}$.

6. Plünnecke's method and an outline

In [13], Plünnecke developed a graph-theoretic method to study the Schnirelmann density of sumsets $A+B$, where A has a positive Schnirelmann density and B is a basis. In [19] I simplified his proof and applied his method to addition of finite sets. Let us quote his main result and some consequences.

We consider directed graphs $G = (V, E)$, where V is the set of vertices and E is that of the edges. If there is an edge from x to y, then we also write $x \to y$. A graph is *semicommutative*, if for every collection $(x; y; z_1, z_2, \ldots, z_k)$ of distinct vertices such that $x \to y$ and $y \to z_i$ there are distinct vertices $y_1, \ldots, y_k$ such that $x \to y_i$ and $y_i \to z_i$. G is *commutative*, if both G and the graph $\tilde{G}$ obtained by reverting every edge of G are semicommutative.

Our graphs will be of a special kind we call *bridging*. By a $(k+1)$-*bridging* graph we mean a graph with a fixed partition of the set of vertices

$$V = S_0 \cup S_1 \cup \ldots \cup S_k$$

into $k+1$ disjoint sets such that every edge goes from some S_{i-1} into S_i.

For $X, Y \subset V$, we define the *image* of X in Y as

$$\operatorname{im}(X,Y) = \{y \in Y : \text{ there is a directed path from some } x \in X \text{ to } y\}.$$

The corresponding *magnification ratio* is defined by

$$D(X,Y) = \min\left\{\frac{|\operatorname{im}(Z,Y)|}{|Z|} : Z \subset X, Z \neq \emptyset\right\}.$$

For a bridging graph we write

$$D_i(G) = D(S_0, S_i).$$

Now Plünnecke's result can be stated as follows.

THEOREM 6.1 (Plünnecke [13]). *In a commutative bridging graph $D_i^{1/i}$ is decreasing.*

Write $\alpha = |S_1|/|S_0|$. Since obviously $D_1 \leqq \alpha$, a consequence of Theorem 6.1 is $D_k \leqq \alpha^k$. This also can be formulated as follows.

STATEMENT 6.2. *There is a nonempty $X \subset S_0$ such that*

$$|\operatorname{im}(X, S_k)| \leqq \alpha^k |X|.$$

These results will be applied to the *addition graph*. Let A, B be subsets of a commutative group. We take the sets $S_0 = A$, $S_i = A + Bi$ $(i = 1, \ldots, k)$ (in different copies of the group for disjointness), and $x \to y$ for $x \in S_{i-1}$, $y \in S_i$ if $y - x \in B$. This graph is easily seen to be commutative, moreover this corresponds to the commutativity of the addition, which also explains this term. An application of Statement 6.2 to this graph yields the following result.

STATEMENT 6.3. *Let k be an integer, A, B sets and write $|A| = n$, $|A + B| = \alpha n$. There is an $X \subset A$, $X \neq \emptyset$ such that*

$$|X + Bk| \leqq \alpha^k |X|. \tag{6.1}$$

My applications of this method always use only Statement 6.3. The reader is invited to find other applications.

Now we outline the proof of Theorem 3.1. To get the upper estimate, we consider sets A of lattice points lying in λB^*. Then $A + B$ is contained in $(\lambda + 1)B^*$, so we expect to have

$$|A| \approx \mu(\lambda B^*) = \lambda^d v, \quad |A + B| \approx \mu((\lambda + 1)B^*) = (\lambda + 1)^d v.$$

The main difficulty is here that the error terms in the estimates for the number of lattice points in a domain are bigger than our second main term.

To get the lower estimate, we show that for large k, Bk contains a large portion of the lattice points in kB^*, except those near to the boundary. Then we show that its discrete impact is similar to the measure impact of kB^*, for which we can apply the Brunn-Minkowski inequality. Finally, Statement 6.3 is used to connect the impact of Bk with the impact of B itself.

PART II. ESTIMATES FOR THE IMPACT

7. Reduction to the standard case

DEFINITION 7.1. We say that a set B of lattice points is *standard*, if $0 \in B$ and B generates L.

We show that the description of impact functions of arbitrary sets of lattice points can be reduced to the standard case.

LEMMA 7.2. *Let $B \subset L$ be a set of lattice points, not contained in an affine hyperplane. Let L_1 be the lattice generated by $B - B$ and i the index of L_1 in L. There is a standard set B' such that*

$$\mu(B') = \mu(B)/i; \tag{7.1}$$

and the impact functions of B and B' are identical.

PROOF. We can achieve $0 \in B$ by a translation, which does not change the difference set $B - B$ or the impact function. So assume that $0 \in B$. In this case B also generates the lattice L_1. There exists an isomorphism $\phi : L \to L_1$ between L and L_1. It is a linear function and has a unique linear extension to R^d, for which we use the same letter. The determinant of the matrix of ϕ is the same as the index of L_1, hence the set

$$B' = \phi^{-1}(B)$$

satisfies (7.1), and it is obviously a standard set.

Let $\xi = \xi_B$, $\xi' = \xi_{B'}$ be the impact functions of B and B'. We show that $\xi \geqq \xi'$. To this end take an $A \subset L$ such that

$$|A| = n, \quad |A + B| = \xi(n).$$

Assume that A intersects k cosets of L_1, and write

$$A = \bigcup_{j=1}^{k} (A_j + \mathbf{t}_j),$$

where $A_j \subset L_1$ and each $\mathbf{t}_j$ is in a different coset. The sets $A_j + B + \mathbf{t}_j$ are disjoint, hence

$$|A + B| = \sum |A_j + B|.$$

Put $A'_j = \phi^{-1}(A_j)$; we have

$$|A'_j| = |A_j|, \quad |A'_j + B'| = |A_j + B|$$

by the isomorphy. Now define A' as

$$A' = \bigcup A'_j + \mathbf{z}_j,$$

where $\mathbf{z}_j$ is selected so that all sets $A'_j + \mathbf{z}_j$ are disjoint. We have

$$|A'| = \sum |A'_j| = n$$

and

$$|A' + B'| \leqq \sum |A'_j + B'| = \sum |A_j + B| = \xi(n),$$

thus $\xi'(n) \leqq \xi(n)$ as claimed.

Finally we show $\xi(n) \leqq \xi'(n)$. Take $A' \subset L$ so that $|A'| = n$, $|A' + B'| = \xi'(n)$. The set $A = \phi(A')$ satisfies $|A| = n$, $|A + B| = |A' + B'| = \xi'(n)$, so indeed $\xi(n) \leqq \xi'(n)$.

8. Fundamental sets

Definition 8.1. A measurable set $Q \subset R^d$ is a *fundamental set*, if for every $\mathbf{x} \in R^d$ there is a unique $\mathbf{y} \in Q$ such that $\mathbf{x} - \mathbf{y} \in L$.

This means that $L + Q$ fills the space exactly. A typical fundamental set is the unit cube. Any fundamental set obviously satisfies $\mu(Q) = 1$.

Lemma 8.2. *Every measurable set X such that $L + X = R^d$ contains a fundamental set.*

This is well-known (and easy).

Lemma 8.3. *If $\mathbf{u}_1, \dots, \mathbf{u}_d \in L$ are linearly independent, then the parallelotope*

$$P = \left\{ \sum \lambda_i \mathbf{u}_i : 0 \leqq \lambda_i < 1 \right\} \tag{8.1}$$

contains a fundamental set.

Proof. Take an arbitrary $\mathbf{x} \in R^d$. We can express it uniquely as

$$\mathbf{x} = \sum \alpha_i \mathbf{u}_i.$$

Then

$$\mathbf{y} = \sum [\alpha_i]\mathbf{u}_i \in L, \quad \mathbf{x} - \mathbf{y} = \sum \{\alpha_i\}\mathbf{u}_i \in P,$$

hence the previous lemma can be applied.

LEMMA 8.4. *For any set $A \subset L$ and fundamental set Q we have*

$$|A| = \mu(A + Q). \tag{8.2}$$

PROOF. This follows immediately from the disjointness of the sets $\mathbf{a} + A$, $\mathbf{a} \in A$.

LEMMA 8.5. *Let $U \subset R^d$, $Q \subset R^d$ and define $E \subset L$ by*

$$E = (U - Q) \cap L. \tag{8.3}$$

If $Q + L = R^d$ (in particular, if Q is a fundamental set), then we have

$$U \subset E + Q.$$

PROOF. Take an $\mathbf{x} \in U$ and express it as

$$\mathbf{x} = \mathbf{q} + \mathbf{y}, \quad \mathbf{q} \in Q, \mathbf{y} \in L.$$

We have $\mathbf{y} = \mathbf{x} - \mathbf{q} \in U - Q$, so $y \in E$.

LEMMA 8.6. *If $Q + L = R^d$ (in particular, if Q is a fundamental set), then for every $U \subset R^d$ the number of lattice points in $U - Q$ or $U + Q$ is at least $\mu(U)$.*

PROOF. Take a fundamental set $Q_1 \subset Q$ and define E by

$$E = (U - Q_1) \cap L;$$

our aim is to estimate its cardinality. We apply Lemma 8.5 for Q_1 in the place of Q to get $U \subset E + Q_1$, and then Lemma 8.4 yields

$$|E| = \mu(E + Q) \leqq \mu(U).$$

To get the result for $U + Q$, observe that if Q is a fundamental set, then so is $-Q$.

We now apply these lemmas to deduce a lower estimate for the number of lattice points in domains that are homothetic to lattice polytopes.

STATEMENT 8.7. *Let $V \subset R^d$ be a convex lattice polytope with volume $\mu(V) = v > 0$. For any $\lambda > d$ and $\mathbf{x} \in R^d$ the number of lattice points in $\lambda V + \mathbf{x}$ is at least*

$$(\lambda - d)^d v. \tag{8.4}$$

PROOF. We may assume that one of the vertices of V is 0. Take d other vertices $\mathbf{u}_1, \dots, \mathbf{u}_d$ that span R^d; these are lattice points. Take a fundamental set Q that is contained in the parallelotope (8.1). Since $\mathbf{u}_i \in V$, we have $Q \subset P \subset dV$. Now aply the previous lemma to the set $U = (\lambda - d)V + \mathbf{x}$. We obtain that the number of lattice points in

$$U + Q = (\lambda - d)V + Q + \mathbf{x} \subset (\lambda - d)V + dV + \mathbf{x} = \lambda V + \mathbf{x}$$

is at least

$$\mu(U) = (\lambda - d)^d v.$$

REMARK. A standard estimate for the number of lattice points is volume–surface. We need an estimate that depends only on the volume.

The example of a simplex shows that for $\lambda < d$ we cannot guarantee the existence of lattice points in $\lambda V + \mathbf{x}$. The same example shows that even for large λ, the d in (8.4) cannot be replaced by any number $< d/2$. On the other hand, (8.4) can probably be improved if v is assumed to be large.

9. Upper estimate for the impact

Here we prove Theorem 3.2.

LEMMA 9.1. $\xi(n+1) \geqq \xi(n) + 1$ *for all* n.

PROOF. Take a set A with $|A| = n+1$, $|A+B| = \xi(n+1)$. Take an $\mathbf{x} \in R^d$ such that all the scalar products $(\mathbf{a}, \mathbf{x})$, $a \in A$ and $(\mathbf{b}, \mathbf{x})$, $b \in B$ are different. Let $\mathbf{a}^* \in A, \mathbf{b}^* \in B$ be those points for which they are maximal. Then

$$(\mathbf{a} + \mathbf{b}, \mathbf{x}) \leqq (\mathbf{a}^* + \mathbf{b}^*, \mathbf{x})$$

for all $\mathbf{a} \in A, \mathbf{b} \in B$, and equality holds only if $\mathbf{a} = \mathbf{a}^*, \mathbf{b} = \mathbf{b}^*$. This shows that $\mathbf{a}^* + \mathbf{b}^*$ does not have any other representation in the form $\mathbf{a} + \mathbf{b}$. Hence for $A_1 = A \setminus \{\mathbf{a}^*\}$ we have $\mathbf{a}^* + \mathbf{b}^* \notin A_1 + B$, consequently

$$\xi(n+1) = |A + B| \geqq 1 + |A_1 + B| \geqq 1 + \xi(n).$$

PROOF OF THEOREM 3.2. By Lemma 7.2 it is sufficient to prove the theorem for standard sets, so assume that B is standard. Define λ so that

$$v(\lambda - d)^d = n,$$

and consider the sets

$$A_{\mathbf{x}} = (\lambda B^* + \mathbf{x}) \cap L.$$

We have $|A_{\mathbf{x}}| \geqq n$ by Statement 8.7, hence

$$|A_{\mathbf{x}} + B| \geqq \xi(|A_{\mathbf{x}}|) \geqq \xi(n) + |A_{\mathbf{x}}| - n \tag{9.1}$$

by the previous lemma. Let Q denote the unit cube. For any set $U \subset R^d$ we have

$$\int_Q |(U+\mathbf{x}) \cap L| \, d\mu(\mathbf{x}) = \mu(U),$$

thus by integrating (9.1) and taking into account that $(\lambda+1)B^* \supset \lambda B^* + B$ we obtain

$$\mu((\lambda+1)B^*) = (\lambda+1)^d v \geqq \xi(n) - n + \lambda^d v,$$

that is,

$$\xi(n) - n \leqq v\left((\lambda+1)^d - \lambda^d\right). \tag{9.2}$$

By a mean value theorem we have

$$(\lambda+1)^d - \lambda^d \leqq d(\lambda+1)^{d-1}. \tag{9.3}$$

Using the definition of λ and a mean value theorem again, we find

$$\begin{aligned}(\lambda+1)^{d-1} &= \left(\left(\frac{n}{v}\right)^{1/d} + d + 1\right)^{d-1} \\ &\leqq \left(\frac{n}{v}\right)^{\frac{d-1}{d}} + (d-1)\left(\left(\frac{n}{v}\right)^{1/d} + d + 1\right)^{d-2} \\ &\leqq \left(\frac{n}{v}\right)^{\frac{d-1}{d}} + c_3\left(\frac{n}{v}\right)^{\frac{d-2}{d}}\end{aligned} \tag{9.4}$$

for $n \geqq v$, with a constant c_3 depending on d.

By combining (9.2–4) we obtain (3.7), (3.8) follows from (3.7) and the inequality

$$(1+x)^{1/d} \leqq 1 + \frac{x}{d}, \tag{9.7}$$

valid for all $x \geqq 0$ and $d \geqq 1$.

10. Set addition in finite groups

This and the next sections contain some preparation to the proof of Theorem 3.3.

Let G be a finite commutative group, with additive notation.

LEMMA 10.1. *Let $A, B \subset G$. We have either*

$$|A+B| \geqq |A| + |B|, \tag{10.1}$$

or

$$A+B=G, \tag{10.2}$$

or there is a subgroup H of G such that $A+B=A+B+H$ and

$$|A+B|=|A+H|+|B+H|-|H|. \tag{10.3}$$

This is a very special case of a theorem of Kneser [7]. It is essentially contained in the papers [8], [9] of Mann (see also [10]), but it is not formulated there exactly in this way.

LEMMA 10.2. *Let $A, B \subset G$. Assume that $0 \in B$ and B generates G. We have*

$$|A+B| \geqq \min\left(|A|+\frac{|B|}{2}, |G|\right). \tag{10.4}$$

PROOF. If (10.1) or (10.2) holds, we are ready. Assume that (10.3) holds. Since B generates G, $B \not\subset H$, thus $B+H$ consists of at least two cosets of H. Consequently

$$|B+H| \geqq 2|H|, \quad |B+H|-|H| \geqq \frac{|B+H|}{2} \geqq \frac{|B|}{2},$$

and then (10.3) yields (10.4).

LEMMA 10.4. *Let $B \subset G$. Assume that $0 \in B$ and B generates G. For every positive integer k we have*

$$|Bk| \geqq \min\left(\frac{k+1}{2}|B|, |G|\right). \tag{10.5}$$

PROOF. The case $k=1$ is obvious, and (10.4) supplies the inductive step for a transition from k to $k+1$.

LEMMA 10.5. *Let $B \subset G$. Assume that $0 \in B$ and B generates G. For*

$$r=\left[\frac{2|G|}{|B|}\right]$$

we have $Br=G$.

PROOF. An immediate consequence of (10.5).

11. Estimating Bk

Throughout this section we assume that $B \subset L=Z^d$, $0 \in B$, B generates L, $|B|=m$ and $\mu(B^*)=v$.

LEMMA 11.1. *There is a fundamental set Q and a positive integer p such that*

$$Bp + Q \supset B^* d + Q + \mathbf{t} \tag{11.1}$$

for some $\mathbf{t} \in L$. Moreover, p can be bounded by

$$p < c_5 v/m. \tag{10.2}$$

REMARK. Without (11.2), (11.1) would be almost obvious.

PROOF. Select $\mathbf{b}_1, \dots, \mathbf{b}_d \in B$ so that the volume of the simplex

$$R = \left\{ \sum \lambda_i \mathbf{b}_i : \lambda_i \geqq 0, \sum \lambda_i \leqq 1 \right\}$$

is maximal, say $\mu(R) = w$. Since $R \subset B^*$, we have $w \leqq v$.

We claim that the parallelotope

$$P = \left\{ \sum \lambda_i \mathbf{b}_i : |\lambda_i| \leqq 1 \right\}$$

of volume $2^d d! w$ contains B. Indeed, every $\mathbf{b} \in B$ has a unique representation

$$\mathbf{b} = \sum \lambda_i \mathbf{b}_i.$$

If we had $|\lambda_i| > 1$ for some i, then replacing $\mathbf{b}_i$ by $\mathbf{b}$ we would get a larger simplex.

Consider the lattice L_1 generated by $3\mathbf{b}_1, \dots, 3\mathbf{b}_d$ and the factor group $G = L/L_1$, with the natural homomorphism $\psi : L \to G$. Observe that

$$|G| = [L : L_1] = 3^d d! w \leqq 3^d d! v = c_6 v.$$

Take $B' = \psi(B) \subset G$. We claim that

$$|B'| = |B| = m.$$

Indeed, take $\mathbf{b}, \mathbf{b}' \in B$ and represent them as

$$\mathbf{b} = \sum \lambda_i \mathbf{b}_i, \quad \mathbf{b}' = \sum \lambda_i' \mathbf{b}_i'.$$

The representation of $\mathbf{b} - \mathbf{b}'$ is

$$\mathbf{b} - \mathbf{b}' = \sum (\lambda_i - \lambda_i') \mathbf{b}_i', \quad |\lambda_i - \lambda_i'| \leqq 2,$$

thus $\mathbf{b} - \mathbf{b}' \notin L_1$.

Since B generates L, B' generates G. By Lemma 10.4 we have $B'r = G$ with

$$r = \left\lceil \frac{2|G|}{|B|} \right\rceil < c_7 v/m.$$

Take a fundamental set Q contained in the parallelotope

$$P_1 = \left\{ \sum \lambda_i \mathbf{b}_i : 0 \leqq \lambda_i \leqq 1 \right\}$$

(Lemma 8.3). Applying Lemma 8.5 for $U = dB^* + Q$ we obtain

$$dB^* + Q \subset E + Q,$$

where

$$E = (dB^* + Q - Q) \cap L.$$

Take an $\mathbf{e} \in E$ and express it as

$$\mathbf{e} = \sum \alpha_i \mathbf{b}_i. \tag{11.3}$$

From $B^* \subset P$ and $Q \subset P_1$ we deduce that

$$-1 < \alpha_i < d + 1. \tag{11.4}$$

Also, since $\phi(\mathbf{e}) \in G = B'r$, there are $\mathbf{x}_1, \dots, \mathbf{x}_r \in B$ such that

$$\mathbf{e} = \mathbf{x}_1 + \cdots + \mathbf{x}_r + s_1\mathbf{b}_1 + \cdots + s_d\mathbf{b}_d \tag{11.5}$$

with integers s_i (all multiples of 3, but we do not need this extra information).

Expressing each $\mathbf{x}_j$ as a linear combination of the $\mathbf{b}_i$'s (with coefficients that are at most 1 in absolute value by $B \subset P$) and then comparing the coefficients of $\mathbf{b}_i$ in (11.4) and (11.5) we see that

$$|s_i| \leqq |\alpha_i| + r \leqq r + d + 1 = r'.$$

Hence with

$$\mathbf{t} = r'(\mathbf{b}_1 + \cdots + \mathbf{b}_d)$$

we have

$$\mathbf{e} + \mathbf{t} = \mathbf{x}_1 + \cdots + \mathbf{x}_r + (s_1 + r')\mathbf{b}_1 + \cdots + (s_d + r')\mathbf{b}_d,$$

a combination with nonnegative integral coefficients. Here the sum of the coefficients is

$$\begin{aligned} r + (s_1 + r') + \cdots + (s_d + r') &= r'(d+1) + s_1 + \cdots + s_d \\ &\leqq r'(2d+1) = (r + d + 1)(2d + 1) \\ &= p \leqq c_8 v/m. \end{aligned}$$

Since $0 \in B$, we can add further terms to have exactly p summands.

In the last inequality we implicitly used the fact that v/m cannot be very small. To see this, decompose B^* into empty lattice simplices, each of which

contains exactly $d+1$ lattice points and has volume $\geqq 1/d!$. This shows that $v/m \geqq 1/(d+1)!$.

We proved that

$$E + \mathbf{t} \subset Bp,$$

hence

$$Bp + Q \supset E + Q + \mathbf{t} \supset dB^* + Q$$

as asserted.

LEMMA 11.2. *With the same p and $\mathbf{t}$ as in Lemma 11.1, we have*

$$kB^* + Q + \mathbf{t} \subset B(k+p) + Q \tag{11.6}$$

for every positive integer k.

PROOF. If $k \leqq d$, (11.6) follows immediately from (11.2). Assume $k > d$. We have

$$kB^* = B(k-d) + dB^* \tag{11.7}$$

by a special case of the Shapley-Folkman theorem. (For a direct proof of (11.7), take an arbitrary $\mathbf{b} \in B^*$, express it as a linear combination of at most $d+1$ vertices of B, multiply the coefficients by k and then separate the integer and fractional parts.) From (11.7) we obtain

$$kB^* + Q + \mathbf{t} \subset B(k-d) + dB^* + Q + t$$
$$B(k+p-d) + Q \subset B(k+p) + Q.$$

REMARK. Considering a set of the form

$$B = \{0, \mathbf{e}_1, \ldots, \mathbf{e}_d, 2\mathbf{e}_d, 3\mathbf{e}_d, \ldots, (m-d-1)\mathbf{e}_d, n\mathbf{e}_d\}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_d$ are the unit vectors and comparing the volumes for $k = p + + d + 1$ it can be shown that in our estimate of p at most the constant c_5 can be improved.

12. Lower estimate for the impact

We prove Theorem 3.3.

LEMMA 12.1. *For $x > 0$ and $y \leqq 1$ we have*

$$(1+x)^y \geqq 1 + y(x - x^2/2). \tag{12.1}$$

PROOF. We have

$$\log(1+x) \geqq x - \frac{x^2}{2}$$

for $x \geqq 0$. Indeed, there is equality at $x = 0$ and the derivative of the differences is

$$\frac{1}{1+x} - 1 + x = \frac{x^2}{1+x} \geqq 0.$$

Hence

$$(1+x)^y = \exp(y \log(1+x)) \geqq \exp(y(x - x^2/2)) \geqq 1 + y(x - x^2/2).$$

Take our fixed set B and a set A with $|A| = n$, $|A+B| = \xi(n)$. We want to show that $A+B$ is large. Define the number α by

$$|A+B| = \alpha n. \tag{12.2}$$

By Statement 6.3 for every k there is a nonempty $A' \subset A$ (possibly depending on k) such that

$$|A' + Bk| \leqq \alpha^k |A'|. \tag{12.3}$$

Take the number p and the fundamental set Q provided by Lemma 11.2. We have

$$Bk + Q \supset (k-p)B^* + Q + \mathbf{t}$$

for $k > p$, therefore

$$\begin{aligned} |A' + Bk| &= \mu(A' + Bk + Q) \geqq \mu(A' + (k-p)B^* + Q) \\ &\geqq \left(\mu(A'+Q)^{1/d} + \mu((k-p)B^*)^{1/d}\right)^d = \left(|A'|^{1/d} + (k-p)v^{1/d}\right)^d. \end{aligned}$$

Here we applied the Brunn-Minkowski theorem (5.1) for the sets $A'+Q$ and $(k-p)B^*$.

Combining this with (12.3) and taking d'th roots we obtain

$$\alpha^{k/d} |A'|^{1/d} \geqq |A'|^{1/d} + (k-p)v^{1/d}.$$

Hence

$$\alpha^{k/d} \geqq 1 + \frac{(k-p)v^{1/d}}{|A'|^{1/d}} \geqq 1 + \frac{(k-p)v^{1/d}}{n^{1/d}}.$$

We take k'th roots and estimate the right side by the previous lemma. With the notation $w = (v/n)^{1/d}$ this yields

$$\alpha^{1/d} \geqq 1 + w - w\left(\frac{p}{k} + \frac{kw}{2}\right). \tag{12.4}$$

We want to minimize the parenthetic term to optimize the result. Assume that $w < 1/p$. Then with the choice

$$k = \left[\sqrt{p/w} + 1\right] \leqq 2\sqrt{p/w}$$

we obtain

$$\frac{p}{k} + \frac{kw}{2} \leqq 2\sqrt{pw}. \tag{12.5}$$

This choice of k is admissible if $k > p$, that is, $w < 1/p$, which is equivalent to $n > vp^d$; for large n this is satisfied.

Substituting (12.5) into (12.4) we find

$$\alpha^{1/d} \geqq 1 + w - 2\sqrt{p}w^{3/2},$$

which in turn yields (recall that $w = (v/n)^{1/d}$)

$$\xi(n)^{1/d} - n^{1/d} = n^{1/d}(\alpha^{1/d} - 1) \geqq v - 2\sqrt{p}n^{\frac{-1}{2d}}v^{\frac{3}{2d}}.$$

Now (3.10) follows by the estimate $p < c_5 v/m$, and (3.9) follows from (3.10) and (9.7).

This concludes the proof of Theorem 3.3, and hence also that of Theorem 3.1.

REFERENCES

[1] FREIMAN, G. A., *Foundations of a structural theory of set addition*, Kazan Gos. Ped. Inst., Kazan, 1966 (in Russian). *MR* **50**#12943

[2] FREIMAN, G. A., *Foundations of a structural theory of set addition*, Translation of Mathematical Monographs, Vol. 37, Amer. Math. Soc., Providence, R. I., 1973. *MR* **50**#12944

[3] FREIMAN, G. A. and PIGAEV, V. P., The relation between the invariants R and T, *Number-theoretic studies in the Markov spectrum and in the structural theory of set addition*, Kalinin. Gos. Univ., Moscow, 1973, 172–174 (in Russian). *MR* **55**#7957

[4] FREIMAN, G. A, What is the structure of K if $K + K$ is small?, *Number theory* (New York, 1984–1985), Lecture Notes in Math., 1240, Springer–Verlag, New York–Berlin, 1987, 109–134. *MR* **88k**:11007

[5] FREIMAN, G., HEPPES, A. and UHRIN, B., A lower estimation for the cardinality of finite difference sets in R^n Problems of computer science (Budapest, 1987), *Tanulmányok – MTA Számítástech. Automat. Kutató Int.* Budapest, No. 202 (1987), 63–73. *MR* **89e**:05041; *Number Theory*, Vol. I. (Budapest, 1987), Coll. Math. Soc. J. Bolyai, **51**, North-Holland, Amsterdam, 1990, 125–139. *MR* **91g**:11008

[6] HALBERSTAM, H. and ROTH, K. F., *Sequences*, Vol. 1, Clarendon Press, London, 1966. *MR* **35**#1565

[7] KNESER, M., Summenmengen in lokalkompakten Abelschen Gruppen, *Math. Z.* **66** (1956), 88–110. *MR* **18**-403

[8] MANN, H. B., On products of sets of group elements, *Canad. J. Math.* **4** (1952), 64–66. *MR* **13**-720

[9] MANN, H. B., An addition theorem for sets of elements of Abelian groups, *Proc. Amer. Math. Soc.* **4** (1953), 423. *MR* **14**-1058

[10] MANN, H. B., *Addition theorems: The addition theorems of group theory and number theory*, Interscience Publishers [John Wiley& Sons], New York – London – Sydney, 1965. *MR* **31**#5854

[11] PLÜNNECKE, H., Über die Dichte der Summe zweier Mengen, deren eine die Dichte null hat, *J. Reine Angew. Math.* **205**(1960), 1–20. *MR* **23**#A2407

[12] PLÜNNECKE, H., Eigenschaften und Abschätzungen von Wirkungsfunktionen, Gesellshaft für Mathematik und Datenverarbeitung, Bonn, 1969. *MR* **40**#5569

[13] PLÜNNECKE, H., Eine zahlentheoretische Anwendung der Graphentheorie, *J. Reine Angew. Math.* **243**(1970), 171–183. *MR* **42**#1794

[14] RUZSA, I. Z., On the cardinality of $A+A$ and $A-A$, *Combinatorics* (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. J. Bolyai, **18**, North-Holland, Amsterdam – New York, 1978, 933–938. *MR* **80c**:05016

[15] RUZSA, I. Z., Sets of sums and differences, *Séminaire de Théorie des Nombres, Paris* 1982/83, *Progr. Math.*, **51**, Birkhäuser, Boston, Mass., 1984, 267–273. *MR* **86f**:11006

[16] RUZSA, I. Z., On an additive property of squares and primes, *Acta Arith.* **49**(1988), 281–289. *MR* **89d**:11008

[17] RUZSA, I. Z., An additive problem for powers of primes, *J. Number Theory* **33**(1989), 71–82. *MR* **90g**:11139

[18] RUZSA, I. Z., Essential components, *Proc. London Math. Soc.* (3) **54**(1987), 38–56. *MR* **88d**:11095

[19] RUZSA, I. Z., An application of graph theory to additive number theory, *Scientia Ser. A* **3**(1989), 97–109.

[20] RUZSA, I. Z., Diameter of sets and measure of sumsets, *Monatsh. Math.* **112**(1991), 323–328. *MR* **92k**:28011

[21] RUZSA, I. Z., On the number of sums and differences, *Acta Math. Hungar.* **59**(1992), 439–447. *MR* **93h**:11025

[22] RUZSA, I. Z., Sums of finite sets, *Number Theory* (New York, 1989) (to appear).

[23] RUZSA, I. Z., Sum of sets in several dimensions, *Combinatorica* (to appear).

(Received February 14, 1994)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 149–154

A PROBLEM IN COVERING PROGRESSIONS

J. SPENCER and P. ERDŐS

A natural question in what might be called extremal number theory is to ask for the largest $T \subseteq \{1, \dots, n\}$ that does not contain any sets A in some natural family $\mathcal{F}$. For example, when $\mathcal{F}$ is the set of arithmetic progressions of a fixed size k the celebrated theorem of Szemerédi gives that $|T| = o(n)$, though more precise results have been difficult. When $\mathcal{F}$ is the family of sets $\{x, 2x\}$ such a set T is called doublefree and the problem of finding its maximal size has appeared in a number of competitions. Fortunately the problem splits. For each odd $u \leq n$ let $C_u = \{u2^i : u2^i \leq n\}$. One now wants T so that each $T \cap C_u$ has maximal size. The solution is precisely to take T to be the set of all $u2^i$ where i is even. Its natural to expand the problem and let $\mathcal{F}$ be the family of all sets $\{x, 2x, 3x\}$. Again the problem splits, for each u relatively prime to $2, 3$ let $C_u = \{u2^i3^j : u2^i3^j \leq n\}$. Now the problem is not so clean. There is an asymptotic (in n) solution, at least in theory. For every fixed integer q there are asymptotically $\frac{n}{3q(q+1)}$ values $u \in [\frac{n}{q+1}, \frac{n}{q}]$ that are relatively prime to $2, 3$. For each of these C_u looks the same (when viewed appropriately as a lattice) and so $|T \cap C_u| = g(q)$ where $g(q)$ can be found (for each q) by a finite, though possibly involved, computation. Then asymptotically $|T| \sim n \sum_q \frac{g(q)}{3q(q+1)}$ so that $|T| \sim cn$ where c can be computed (theoretically) to any degree of accuracy by taking q appropriately large. Here we have examined the natural to extension to letting $\mathcal{F}$ be the family of sets $\{x, 2x, \dots, sx\}$. Instead of T it is more natural to consider its complement S. Thus we arrive at the following formulation:

DEFINITION. $f_s(n)$ be the minimal size of a set $S \subseteq \{1, \dots, n\}$ with the property that

$$S \cap \{t, 2t, \dots, st\} \neq \emptyset \quad \text{for} \quad 1 \leq t \leq \frac{n}{s}.$$

From our previous comments we know that for all s there is a positive constant c_s so that

$$f_s(n) \sim c_s n$$

as $n \to \infty$. Our concern in this paper will be the asymptotics of c_s.

1990 *Mathematics Subject Classification.* Primary 11B05.
Key words and phrases. Density, progressions.

THEOREM.

$$c_s = \Theta\left(\frac{1}{s \ln s}\right).$$

For convenience set $A_t = \{t, 2t, \ldots, ts\}$.

Let $M(t)$ denote the smallest prime p dividing t, with $M(1) = +\infty$. Observe that the density of integers with $M(t) > p$ is

$$\prod_{p_1 \leqq p} (1 - p_1^{-1}) \sim \frac{1}{\ln p},$$

where p_1 above runs over primes only. All t with $M(t) > s$ give disjoint sets A_t and hence

$$c_s > s^{-1} \prod_{p_1 \leqq s} (1 - p_1^{-1}) \sim \frac{1}{s \ln s}.$$

For the upper bound we shall find a set S that overlaps A_t for all *square-free* $t \leqq \frac{n}{s}$. The adjustment to overlapping A_t for all t is left to the end.

List the primes $2, 3, 5, \ldots$ in order for as long as their product is less than $s^{1/3}$. Let A denote the first prime not on that list. Basic bounds give $A \sim \ln s^{1/3} \sim \frac{1}{3} \ln s$. Define $SMALLP$ to be the product of all primes $p < A$ so that $SMALLP < s^{1/3}$. Now place the primes starting with A in consecutive blocks, each maximal with product less than $s^{1/3}$. Terminate this process when the next block would start with a prime $p > s^{-1}$. Let B be the final prime in the final block. Let F be the set of first primes of the blocks. For $p \in F$ let $BLOCK(p)$ denote the block of primes of which p is the first, let $p^{(L)}$ denote the largest prime of $BLOCK(p)$, let p^* denote the product of all primes in $BLOCK(p)$ and let $r = r(p)$ denote the size of $BLOCK(p)$. Rough bounds give that $p^{(L)} < 3p$ for all blocks. (Indeed, except for the first "few" blocks $p^{(L)} \sim p$.) As all $p' \in BLOCK(p)$ have $p' = p^{1+o(1)}$ we almost have $r \sim \frac{\ln s}{3 \ln p}$. The rounddown for r with p large makes this not quite accurate (the right-hand side being a small integer) but since we stop at $p \sim s^{-1}$ we can certainly say

$$r = r(p) > \frac{1}{4} \frac{\ln s}{\ln p}$$

for all $p \in F$ with room to spare.

With $p \in F$ note that (letting p' range over $BLOCK(p)$)

$$\frac{1}{p} = \frac{1}{r(p)} \sum_{p'} \frac{1}{p} \leqq \frac{1}{r(p)} \sum_{p'} \frac{3}{p'} \leqq 12 \sum p' \frac{\ln p}{p' \ln s} \leqq 12 \sum p' \frac{\ln p'}{p' \ln s}.$$

As $\ln p' \sim \ln p$ for all $p' \in BLOCK(p)$ we also have

$$\frac{1}{p \ln p} \leqq 13 \sum p' \frac{\ln p'}{p' \ln p' \ln s} \leqq \frac{13}{\ln s} \sum_{p'} \frac{1}{p'}.$$

We shall denote these as *smoothing ineqalities*, they shall prove useful in the analyses. Note that if we apply a smoothing inequality to each term of a sum over all $p \in F$ it becomes a sum over all primes p' with $A \leqq p' \leqq B$.

For any P the set of positive integers t with $M(t) > P$ has density $\prod_{p \leqq P}\left(1-\frac{1}{p}\right)$, p ranging over primes. This product is asymptotic in P to $\Theta(\frac{1}{\ln P})$. In application below we will have $A \leqq P \leqq B$ and count t up to $\frac{n}{ps}$ or $\frac{n}{pqs}$ but always up to some dn where d does not depend on n, though it will depend on s. In that case the number of such $t \leqq dn$ with $M(t) > P$ is at most $k_1 \frac{dn}{\ln P}$, with k_1 an absolute constant. We call this the *sieve inequality* in later analyses.

The set S will consist of four parts, *BASIC*, *CLOSEP*, *ONEP*, *NOP*, each designed to intersect different A_t. Together they will intersect all A_t with t squarefree, $t \leqq \frac{n}{s}$. Each will have size $O(\frac{n}{s \ln s})$. The analysis of each is done separately.

• *BASIC*. For each pair of distinct $p, q \in F$ with $p < q$ take all x of the form

$$x = (SMALLP)p^*q^*u$$

with

$$u \leqq \frac{n}{spq} \quad \text{and} \quad M(u) \geqq q.$$

Let $t \leqq \frac{n}{s}$ be squarefree, suppose t has at least two prime factors in $[A, B]$ and, further, letting $p_1 < q_1$ be the first two such primes, suppose p_1, q_1 lie in different blocks. We claim that then there is an $x \in BASIC$ with $x \in A_t$. Indeed: set $\alpha = \operatorname{lcm}(t, SMALLP)$. Let $p, q \in F$ be such that $p_1 \in BLOCK(p)$, $q_1 \in BLOCK(q)$. Write $t = \alpha p_1 q_1 u$ so that $M(u) > q_1 \geqq q$ and

$$u \leqq \frac{t}{p_1 q_1} \leqq \frac{n/s}{pq}.$$

We set

$$x = (SMALLP)p^*q^*u.$$

Then t divides x and

$$\frac{x}{t} = \frac{SMALLP}{\alpha}\frac{p^*}{p_1}\frac{q^*}{q_1} < (SMALLP)p^*q^* < s^{1/3}s^{1/3}s^{1/3} = s,$$

as desired.

The critical analysis is of the size of *BASIC*. For a given $p, q \in F$ with $p < q$ the sieve inequality gives at most $k_1 \frac{n}{spq \ln q}$ values u so that (employing

the smoothing inequalities)

$$|BASIC| \leqq k_1 \sum_{\substack{p,q\in F\\ p<q}} \frac{n}{spq\ln q}$$

$$\leqq k_2 \sum_{A\leqq p'<q'\leqq B} \frac{n}{sp'q'\ln q'} \frac{\ln s}{\ln p'} \frac{\ln s}{\ln q'}$$

$$< \frac{k_3 n}{s\ln^2 s} \sum_{A<q'\leqq B} \frac{1}{q'\ln q'} \sum_{A\leqq p<q} \frac{\ln p'}{p'}$$

$$< \frac{k_4 n}{s\ln^2 s} \sum_{A\leqq q'<B} \frac{1}{q'\ln q'} \ln q'$$

$$< \frac{k_5 n}{s\ln^2 s} \ln\ln B < \frac{k_6 n}{s\ln s}$$

as desired.

- *CLOSEP.* For each $p\in F$ take all x of the form

$$x = (SMALLP)p^* u$$

with

$$u \leqq \frac{n}{sp^2} \text{ and } M(u) > p.$$

Let $t \leqq \frac{n}{s}$ be squarefree, suppose t has at least two prime factors in $[A,B]$ but now, letting $p_1 < q_1$ denote the first two such primes, suppose p_1, q_1 lie in a common $BLOCK(p)$. We claim there is an $x\in CLOSEP$ with $x\in A_t$. As before, set $\alpha = \text{lcm}(t, SMALLP)$. Then $t = \alpha p_1 q_1 u$ with $M(u) > q_1 > p$ and $u \leqq \frac{t}{p_1 q_1} < \frac{n}{sp^2}$. Set

$$x = (SMALLP)p^* u.$$

Then t divides x and

$$\frac{x}{t} = \frac{SMALLP}{\alpha} \frac{p^*}{p_1 q_1} < s^{1/3} s^{1/3} < s,$$

completing the claim. Now we bound the size of *CLOSEP*. For a given $p\in$ $\in F$ the sieve inequality gives that there are less than $k_7 \frac{n}{sp^2\ln p}$ values u so, employing the smoothing inequality

$$|CLOSEP| \leqq k_7 \sum_{p\in F} \frac{n}{sp^2\ln p}$$

$$\leqq k_8 \sum_{A\leqq p'\leqq B} \frac{n}{s(p')^2\ln p'} \frac{\ln p'}{\ln s}$$

$$\leqq \frac{k_8 n}{s\ln s} \sum_{A\leqq p'} \frac{1}{(p')^2}$$

which is actually $o(\frac{n}{s \ln s})$.

- *ONEP*. For each $p \in F$ take all x of the form

$$x = (SMALLP)p^* u$$

with

$$M(u) > B \text{ and } u \leqq \frac{n}{sp}.$$

Let $t \leqq \frac{n}{s}$ be squarefree and suppose t has precisely one prime factor, call it p_1 in $[A, B]$. We claim there is an $x \in ONEP$ with $x \in A_t$. Let $p \in F$ be such that $p_1 \in BLOCK(p)$. Set $\alpha = \operatorname{lcm}(t, SMALLP)$. Then $t = \alpha p_1 u$ with $M(u) > p_1 \geqq p$ and $u \leqq \frac{t}{p_1} \leqq \frac{n}{sp}$. Set

$$x = (SMALLP)p^* u.$$

Then t divides x and

$$\frac{x}{t} = \frac{SMALLP}{\alpha}\frac{p^*}{p_1} < s^{1/3}s^{1/3} < s,$$

giving the claim. Now we bound the size of *ONEP*. For each $p \in F$ the sieve inequality gives that there are less than $k_9 \frac{n}{sp \ln B}$ different u so, employing the smoothing inequality and recalling $\ln B = \Theta(\ln s)$,

$$\begin{aligned} |ONEP| &\leqq k_{10} \sum_{p \in F} \frac{n}{sp \ln s} \\ &< \frac{k_{11} n}{s \ln s} \sum_{A \leqq p' \leqq B} \frac{1}{p'} \frac{\ln p'}{\ln s} \\ &< \frac{k_{11} n}{s \ln^2 s} \sum p' \leqq B \frac{\ln p'}{p'} \\ &< \frac{k_{12} n}{s \ln^2 s} \ln B < \frac{k_{12} n}{s \ln s}, \end{aligned}$$

as desired.

- *NOP*. All x of the form

$$x = (SMALLP)u$$

with

$$M(u) > B \text{ and } u \leqq \frac{n}{s}.$$

Let $t \leqq \frac{n}{s}$ be squarefree and suppose t has no prime factors in $[A, B]$. We claim there is an $x \in NOP$ with $x \in A_t$. Set $\alpha = \operatorname{lcm}(t, SMALLP)$. Then $t = \alpha u$ with $M(u) > B$ and $u \leqq t \leqq \frac{n}{s}$. Set

$$x = (SMALLP)u$$

Then t divides x and

$$\frac{x}{t} = \frac{SMALLP}{\alpha} \leqq s^{1/3} < s,$$

giving the claim. Now we bound the size of NOP. By the sieve inequality

$$|NOP| \leqq k_{13} \frac{n}{s} \frac{1}{\ln B} \leqq k_{14} \frac{n}{s \ln s},$$

as desired.

Hence the total size of S is less than $k_{15} \frac{n}{s \ln s}$.

To complete the proof we must find an S that overlaps A_t for all $t \leqq \leqq \frac{n}{s}$, not just the squarefree t. Let $m = m(n)$ slowly approach infinity, for definiteness take $m = \lfloor \ln n \rfloor$. For $1 \leqq i \leqq m$ let S_i be a set that overlaps all A_t for t squarefree, $1 \leqq t \leqq \frac{n}{si^2}$. The argument above still applies as ni^{-2} is approaching infinity for fixed s and so there is such a set of size at most $k_{15} \frac{n}{si^2 \ln s}$. The set $i^2 S_i$ has the same size and overlaps all A_t for $t \leqq \frac{n}{s}$ where i^2 is the maximal square dividing t. Let S^0 be the union of the $i^2 S_i$ for $1 \leqq i \leqq m$. Then, making critical use of the convergence of $\sum i^{-2}$, S^0 has size at most $k_{16} \frac{n}{s \ln s}$. Let S^1 be all $t \leqq \frac{n}{s}$ with a square factor j^2, $j > m$. As $m = m(n) \to \infty$, the size of S^1 is $o(n)$ (fixed s, $n \to \infty$). The set $S = S^0 \cup S^1$ then overlaps *all* A_t, $t \leqq \frac{n}{s}$ and has size less than $k_{17} \frac{n}{s \ln s}$, completing the proof.

(Received February 14, 1994)

COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NY 10012-1110
U.S.A.

e-mail: spencer@cs.nyu.edu

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 155–163

ON THE MINIMUM NUMBER OF EMPTY POLYGONS IN PLANAR POINT SETS[1]

P. VALTR[2]

Abstract

We describe a configuration (related to Horton's constructions) of n points in general position in the plane with less than $1.8n^2$ empty triangles, less than $2.42n^2$ empty quadrilaterals, less than $1.46n^2$ empty pentagons, and less than $n^2/3$ empty hexagons. It improves the constants shown by Bárány and Füredi.

1. Introduction

We say that a set $\mathcal{P}$ of points in the plane is in *general position* if no three points of $\mathcal{P}$ lie on a line. Erdős and Szekeres [4] proved that for any k there is an integer $n(k)$ such that any set of $n(k)$ points in general position in the plane contains k points which are vertices of a convex k-gon.

We call a subset A of k points in $\mathcal{P}$ *an empty k-gon* if the convex hull of A contains no point of $\mathcal{P}$ in its interior. Erdős [3] asked whether the following sharpening of the Erdős–Szekeres theorem is true. Is there an $N(k)$ such that any set of $N(k)$ points in general position in the plane contains an empty k-gon? He pointed out that $N(4) = 5$ and Harborth [5] proved $N(5) = 10$. On the other hand, Horton [6] showed that $N(k)$ does not exist for $k \geqq 7$. The question about the existence of $N(6)$ is still open.

Denote by $f_k(\mathcal{P})$ the number of empty k-gons in $\mathcal{P}$ and let

$$f_k(n) = \min\{f_k(\mathcal{P}) : |\mathcal{P}| = n \text{ and } \mathcal{P} \text{ is in general position}\}.$$

Katchalski and Meir [7] proved that there is a constant $K < 200$ such that for any $n \geqq 3$

$$\binom{n-1}{2} \leqq f_3(n) \leqq Kn^2.$$

1991 *Mathematics Subject Classification.* Primary 52A10; Secondary 52C10.
Key words and phrases. Convex polygons, empty polygons.

[1] The work on this paper was partially done when the author participated at the workshop Uniformity and Irregularity of Partitions, University Bielefeld, Bielefeld, Germany.

[2] Graduiertenkolleg "Algorithmische Diskrete Mathematik", Fachbereich Mathematik, Freie Universität Berlin, Arnimallee 2–6, W–1000 Berlin 33, Germany, supported by "Deutsche Forschungsgemeinschaft" Grant We 1265/2-1.

0081–6906/95/$ 4.00

Horton [6] constructed configurations giving $f_k(n) = 0$, for $k \geqq 7$. Bárány and Füredi [2] proved

$$n^2 - O(n \log n) \leqq f_3(n) \leqq 2n^2,$$
$$\frac{1}{4}n^2 - O(n) \leqq f_4(n) \leqq 3n^2,$$
$$\left\lfloor \frac{n}{10} \right\rfloor \leqq f_5(n) \leqq 2n^2,$$
$$f_6(n) \leqq \frac{1}{2}n^2.$$

They proved the upper bounds only when n is a power of 2. However, one can prove them with a bit more effort for any integer n. To show the upper bounds Bárány and Füredi used the construction of Horton [6] giving $f_k(n) = 0$, for $k \geqq 7$. Bárány [1] still improved the lower bound on $f_4(n)$ to

$$\frac{1}{2}n^2 - O(n) \leqq f_4(n).$$

In Section 2 we describe two simple random configurations where the expected number of empty triangles is $2n^2 + o(n^2)$.

In Section 3 we show a construction giving the following better upper bounds:

$$f_3(n) < 1.8n^2, \qquad f_4(n) < 2.42n^2,$$
$$f_5(n) < 1.46n^2, \qquad f_6(n) < \frac{1}{3}n^2.$$

Let us note that the construction in Section 3 is a simplified version of a complicated construction which gives still a bit better estimations (see also the remark at the end of the paper).

2. Random constructions

Bárány and Füredi [2] proved that the following random construction gives a similar upper bound of $f_3(n)$ as Horton's construction.

THEOREM 1. *Let $I_1, I_2, \ldots, I_n$ be parallel unit intervals in the plane, $I_i = \{[x, y] : x = i, 0 \leqq y \leqq 1\}$. For any i, choose a random point p_i from I_i. Then the expected number of empty triangles in the set $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ is at most $2n^2 + O(n \log n)$.*

In the following we show that another random construction gives a similar result:

THEOREM 2. *Let K be a bounded convex area in the plane. Let $\mathcal{P}$ be a set of n points placed randomly (according to a uniform distribution) and independently inside K. Then the expected number of empty triangles in $\mathcal{P}$ is at most $2n^2 - 2n$.*

PROOF. Without loss of generality, assume the area of K equals 1. Consider two points p_i, p_j from the set $\mathcal{P} := \{p_1, p_2, \ldots, p_n\}$, and denote the Euclidean distance between p_i and p_j by l. Define the axes so that $p_i = [0,0]$ and $p_j = [l,0]$. Let S_{ij} be the strip of width l between the $y-$axis and the line $x = l$. For all triangles $p_ip_jp_k$ with the longest side p_ip_j, the vertex p_k lies obviously inside S_{ij}. The expected number of points p_k from $\mathcal{P} \cap S_{ij}$ such that $p_ip_jp_k$ is an empty triangle can be easily estimated. For any real number y, define the line segment $I_y := \{[x,y] : 0 \leqq x \leqq l\}$, and let $|I_y \cap K|$ denote the length of the line segment $I_y \cap K$. If $|y| > \frac{2}{l}$ then $I_y \cap K = \emptyset$ (otherwise the area of K would exceed 1). For any $k, 1 \leqq k \leqq n, k \neq i, k \neq j$,

$$\begin{aligned}
&\text{Prob}(p_k \in S_{ij} \text{ and } p_ip_jp_k \text{ is an empty triangle}) = \\
&= \int_{-\infty}^{\infty} |I_y \cap K| \cdot \text{Prob}(p_ip_jp_k \text{ is empty} \mid p_k \in I_y)\, dy = \\
&= \int_{-\frac{2}{l}}^{\frac{2}{l}} |I_y \cap K| \left(1 - \frac{l|y|}{2}\right)^{n-3} dy \leqq \int_{-\frac{2}{l}}^{\frac{2}{l}} l \left(1 - \frac{l|y|}{2}\right)^{n-3} dy = \frac{4}{n-2}.
\end{aligned}$$

Hence, for any pair $\{i,j\}$, the expected number of empty triangles $p_ip_jp_k$, where p_ip_j is the longest side, is at most $\frac{4}{n-2}(n-2) = 4$, and the overall expected number of empty triangles is at most $4\binom{n}{2} = 2n^2 - 2n$. □

Note that the estimations of the number of empty triangles for the above three configurations (Horton's construction, the random constructions from Theorems 1 and 2) are the best possible in the sense that the (expected) number of empty triangles in each of them is at least $2n^2 - o(n^2)$. In Section 3 we show a configuration with a smaller number of empty triangles.

Let us mention that the method of the proof of Theorem 2 can be extended to the higher dimension for the counting of the number of empty simplices. In $\mathbb{R}^d, d \geqq 2$, $d+1$ points of a set $\mathcal{P} \subset \mathbb{R}^d$ form an *empty simplex*, if they are vertices of a simplex containing no other points of $\mathcal{P}$.

THEOREM 3. *The expected number of empty simplices in a set of n points chosen independently and randomly (according to a uniform distribution) from a full-dimensional bounded convex body in $\mathbb{R}^d, d \geqq 2$, is smaller than $2d\binom{n}{d} < \frac{2}{(d-1)!}n^d$.*

We omit the proof of Theorem 3 here, it will appear elsewhere.

3. Construction

We start with Horton's construction: For any positive integer n, we will define a point set $H(n)$ of n points. In $H(n)$ the set of the first coordinates is just $\{0, 1, \dots, n-1\}$. First we define by induction a set $H(n)$ when n is a power of 2. Let $H(1) = \{(0,0)\}$ and $H(2) = \{(0,0),(1,0)\}$. When $H(n)$ is defined, set

$$H(2n) = \{(2x, y) : (x, y) \in H(n)\} \cup \{(2x+1, y+d_n) : (x, y) \in H(n)\}$$

where the numbers d_n are fastly growing, say $d_n = 3^n - 1$. These sets $H(n)$ are just the sets defined by Horton [6]. Now let n be a positive integer, and let n' be the least power of 2 which is not smaller than n. Set

$$H(n) = \{(x, y) \in H(n') : x < n\}.$$

All y–coordinates of points of $H(n)$ are smaller than 3^n. The building blocks of our construction are sets $Q(n)$ which are obtained from $H(n)$ by replacing each point (x, y) by $(x, (12+n)^{-1}3^{-n}y)$. Obviously, all points of $Q(n)$ lie in the $(12+n)^{-1}$–neighborhood of the x–axis ($(12+n)^{-1}$ is no specific number; it is only a sufficiently small positive number). Now let $m = 4n$ be a positive integer divisible by 4 (for simplicity). We construct an m–point set S_m in the following way:

$$S_m = Q_1 \cup Q_2 \cup Q_3 \cup Q_4,$$

where

$$Q_1 = Q(n), \qquad Q_2 = Q(n) + (\tfrac{1}{4}, 1),$$
$$Q_3 = Q(n) + (0, 2), \qquad Q_4 = Q(n) + (\tfrac{1}{4}, 3).$$

$Q(n) + (a, b)$ denotes the set $Q(n)$ shifted by the vector (a, b). So the points of S_m lie in the $(12+n)^{-1}$–neighborhoods of points of the set $\overline{S_m} = N \cup (N + (\frac{1}{4}, 1)) \cup (N + (0, 2)) \cup (N + (\frac{1}{4}, 3))$, where $N = \{(0,0), (1,0), \dots, (n-1, 0)\}$. Note now that the number $(12+n)^{-1}$ is small enough in order that the set S_m is combinatorially equivalent to the set $\overline{S_m}$, except that the sets $Q_i, i = 1, 2, 3, 4$, do not lie on a line.

The shifts $(\frac{1}{4}, 1)$, $(0, 2)$, $(\frac{1}{4}, 3)$ in the definition of S_m were chosen to ensure that e.g. no triangle with one point in Q_4 and two points in Q_1 is empty. This, and some related properties are used in the proof of Lemma 4 below.

Define, for any $s \geqq 3$, the following two sets:

$$G_s(3) = \{g : g \text{ is an empty } s\text{-gon in } Q_1 \cup Q_2 \cup Q_3, g \cap Q_1 \neq \emptyset, g \cap Q_3 \neq \emptyset\},$$
$$G_s(4) = \{g : g \text{ is an empty } s\text{-gon in } Q_1 \cup Q_2 \cup Q_3 \cup Q_4, g \cap Q_1 \neq \emptyset, g \cap Q_4 \neq \emptyset\}.$$

LEMMA 4.

$$|G_3(3)| < 3n^2, \qquad |G_3(4)| \leqq \frac{8}{3}n^2,$$
$$|G_4(3)| < 3n^2, \qquad |G_4(4)| \leqq \frac{8}{3}n^2,$$
$$|G_5(3)| < n^2, \qquad |G_5(4)| \leqq \frac{4}{3}n^2,$$
$$|G_6(3)| = 0, \qquad |G_6(4)| \leqq \frac{1}{3}n^2.$$

PROOF. For $i = 1, 2, 3, 4$, denote the elements of Q_i by $q_{i,j}, j = 1, 2, \ldots, n$ in the order according to their x–coordinates. First we estimate the sizes of the sets $G_s(4)$. Each empty s–gon $g \in G_s(4)$ contains only one point of Q_1 and only one point of Q_4. For $i, j = 1, 2, \ldots, n$, we can easily count the number of empty polygons g such that $g \cap Q_1 = \{q_{1,i}\}$ and $g \cap Q_4 = \{q_{4,j}\}$.

If $i \equiv j \pmod 3$, then $g \subseteq \{q_{1,i}, q_{2,\frac{2i+j}{3}}, q_{3,\frac{i+2j}{3}}, q_{4,j}\}$ and g is one of the two empty triangles $q_{1,i}\, q_{2,\frac{2i+j}{3}}\, q_{4,j}$ and $q_{1,i}\, q_{3,\frac{i+2j}{3}}\, q_{4,j}$ or the empty quadrilateral $q_{1,i}\, q_{2,\frac{2i+j}{3}}\, q_{3,\frac{i+2j}{3}}\, q_{4,j}$.

If $i \equiv j - 1 \pmod 3$, then $g \subseteq \{q_{1,i}, q_{2,\lfloor\frac{2i+j}{3}\rfloor}, q_{3,\lceil\frac{i+2j}{3}\rceil}, q_{4,j}\}$ and g is again one of two certain triangles or a certain quadrilateral.

If $i \equiv j + 1 \pmod 3$, then

$$g \subseteq \{q_{1,i}, q_{2,\lfloor\frac{2i+j}{3}\rfloor}, q_{2,\lceil\frac{2i+j}{3}\rceil}, q_{3,\lfloor\frac{i+2j}{3}\rfloor}, q_{3,\lceil\frac{i+2j}{3}\rceil}, q_{4,j}\}$$

and g is one of four triangles, six quadrilaterals, and four pentagons, or a hexagon.

There are $\left\lfloor\frac{n^2}{3}\right\rfloor$ pairs $\{i, j\}$ such that $i \equiv j + 1 \pmod 3$, and there are $\left\lceil\frac{2n^2}{3}\right\rceil$ other pairs $\{i, j\}$. Therefore

$$|G_3(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 4 + \left\lceil\frac{2n^2}{3}\right\rceil \cdot 2 \leqq \frac{8}{3}n^2,$$
$$|G_4(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 6 + \left\lceil\frac{2n^2}{3}\right\rceil \cdot 1 \leqq \frac{8}{3}n^2,$$
$$|G_5(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 4 \leqq \frac{4}{3}n^2,$$
$$|G_6(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 1 \leqq \frac{1}{3}n^2.$$

Now we estimate the sizes of the sets $G_s(3)$. Each empty s–gon $g \in G_s(3)$ contains either one or two consecutive points from Q_1. In the second case

the points from g are from one of the $\frac{n(n-1)}{2}$ sets

$$\{q_{1,i},\ q_{1,i+1},\ q_{2,i+\Delta},\ q_{2,i+\Delta+1},\ q_{3,i+2\Delta+1}\},$$
$$1 \leqq i \leqq n-1, \qquad \left\lceil -\frac{i}{2} \right\rceil \leqq \Delta \leqq \left\lfloor \frac{n-i-1}{2} \right\rfloor .$$

Each of these sets contains one triangle $(q_{1,i}\ q_{1,i+1}\ q_{3,i+2\Delta+1})$ from $G_3(3)$, two quadrilaterals $(q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta}\ q_{3,i+2\Delta+1}$ and $q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta+1}\ q_{3,i+2\Delta+1})$ from $G_4(3)$, and one pentagon $(q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta}\ q_{2,i+\Delta+1}\ q_{3,i+2\Delta+1})$ from $G_5(3)$.

Consider now the empty s–gons $g \in G_s(3)$ containing only one point of the set Q_1. Most of these polygons are contained in one of the $\frac{n(n-1)}{2}$ sets

$$\{q_{1,i},\ q_{2,i+\Delta-1},\ q_{2,i+\Delta},\ q_{3,i+2\Delta-1},\ q_{3,i+2\Delta}\},$$
$$1 \leqq i \leqq n, \qquad \left\lceil \frac{2-i}{2} \right\rceil \leqq \Delta \leqq \left\lfloor \frac{n-i}{2} \right\rfloor . \qquad \square$$

Each of these sets contains 5 triangles from $G_3(3)$, 4 quadrilaterals from $G_4(3)$, and one pentagon from $G_5(3)$.

For odd $i > 1$, the points of g can be also from the set $\{q_{1,i},\ q_{2,\frac{i-1}{2}},\ q_{2,\frac{i+1}{2}},\ q_{3,1}\}$. There are $\left\lfloor \frac{n-1}{2} \right\rfloor$ such sets, each with two triangles from $G_3(3)$ and one quadrilateral from $G_4(3)$. For $i = 1$, we have to consider only the triangle $\{q_{1,1}\ q_{2,1}\ q_{3,1}\}$.

If $i \not\equiv n \pmod 2$, then the points of g can be still from the set $\{q_{1,i},\ q_{2,\frac{n+i-1}{2}},\ q_{2,\frac{n+i+1}{2}},\ q_{3,n}\}$. There are $\left\lfloor \frac{n}{2} \right\rfloor$ such sets, each with two triangles from $G_3(3)$ and one quadrilateral from $G_4(3)$.

The required bounds follow:

$$|G_3(3)| = \frac{n(n-1)}{2} + \frac{n(n-1)}{2} \cdot 5 + \left\lfloor \frac{n-1}{2} \right\rfloor \cdot 2 + 1 + \left\lfloor \frac{n}{2} \right\rfloor \cdot 2 < 3n^2,$$
$$|G_4(3)| = \frac{n(n-1)}{2} \cdot 2 + \frac{n(n-1)}{2} \cdot 4 + \left\lfloor \frac{n-1}{2} \right\rfloor + \left\lfloor \frac{n}{2} \right\rfloor < 3n^2,$$
$$|G_5(3)| = \frac{n(n-1)}{2} + \frac{n(n-1)}{2} < n^2,$$
$$|G_6(3)| = 0. \qquad \square$$

THEOREM 5.

$$f_3(S_m) < 1.8m^2, \qquad f_4(S_m) < 2.42m^2,$$
$$f_5(S_m) < 1.46m^2, \qquad f_6(S_m) < \frac{1}{3}m^2.$$

PROOF. Let $\mathcal{P}$ be a point set in the plane and consider two points $u_1, u_2 \in \mathcal{P}, u_1 = (x_1, y_1), u_2 = (x_2, y_2), x_1 < x_2$. We say that the line segment u_1u_2 is *open from below* if there is no point of $\mathcal{P}$ inside the strip $S = \{(x, y) : x_1 < x < x_2$ and (x, y) lies below the line $u_1u_2\}$. A subset X of $\mathcal{P}$ is called *open from below* if all the line segments connecting two points of X are open from below. Analogously, we define *open from above*.

For any positive integer r, denote by $h_r^-(\mathcal{P})$ and $h_r^+(\mathcal{P})$ the number of $r-$point subsets in $\mathcal{P}$ empty from below and above, respectively.

Bárány and Füredi [2] showed

$$h_2^-(H(n)) < 2n, \qquad h_2^+(H(n)) < 2n,$$

and

$$h_3^-(H(n)) < n, \qquad h_3^+(H(n)) < n.$$

They proved the above inequalities when n is a power of 2. However, one can prove them for any positive integer n.

The construction of $H(n)$ is done so that, for any $r > 3$,

$$h_r^-(H(n)) = h_r^+(H(n)) = 0.$$

Obviously, all the above relations are satisfied for the set $H(n)$ as well as for the sets $Q(n)$ and $Q_i, i = 1, 2, 3, 4$.

For any $s \geqq 3$, and any $r, 0 < r < s$, the number of empty $s-$gons G in $Q_1 \cup Q_2$ with $|G \cap Q_1| = r$ is equal to $h_r^+(Q_1) \cdot h_{s-r}^-(Q_2)$. This is carried out by the construction (more precisely, by the fact that the set Q_2 lies entirely above any line containing two points of Q_1 and similarly the set Q_1 lies entirely below any line containing two points of Q_2). Thus

$$f_s(Q_1 \cup Q_2) = f_s(Q_1) + f_s(Q_2) + \sum_{r=1}^{s-1} h_r^+(Q_1) h_{s-r}^-(Q_2).$$

Since $f_s(Q_2 \cup Q_3) = f_s(Q_1 \cup Q_2)$ and $f_s(Q_2) = f_s(Q_1)$ we obtain

$$f_s(Q_1 \cup Q_2 \cup Q_3) = f_s(Q_1 \cup Q_2) + f_s(Q_2 \cup Q_3) - f_s(Q_2) + g_s(3) =$$

$$= 2f_s(Q_1 \cup Q_2) - f_s(Q_1) + g_s(3) = 3f_s(Q_1) + 2\sum_{r=1}^{s-1} h_r^+(Q_1) h_{s-r}^-(Q_2) + g_s(3).$$

Similarly

$$f_s(Q_1 \cup Q_2 \cup Q_3 \cup Q_4) =$$
$$f_s(Q_1 \cup Q_2 \cup Q_3) + f_s(Q_2 \cup Q_3 \cup Q_4) - f_s(Q_2 \cup Q_3) + g_s(4) =$$
$$2(3f_s(Q_1) + 2\sum_{r=1}^{s-1} h_r^+(Q_1)h_{s-r}^-(Q_2) + g_s(3)) -$$
$$-(2f_s(Q_1) + \sum_{r=1}^{s-1} h_r^+(Q_1)h_{s-r}^-(Q_2)) + g_s(4) =$$
$$= 4f_s(Q_1) + 3\sum_{r=1}^{s-1} h_r^+(Q_1)h_{s-r}^-(Q_2) + 2g_s(3) + g_s(4).$$

Now the required bounds follow:

$$f_3(S_m) < 4 \cdot 2n^2 + 3 \cdot (n \cdot 2n + 2n \cdot n) + 2 \cdot 3n^2 + \frac{8}{3}n^2 = \frac{86}{3}n^2 = 1.791\ldots m^2,$$
$$f_4(S_m) < 4 \cdot 3n^2 + 3 \cdot (n \cdot n + 2n \cdot 2n + n \cdot n) + 2 \cdot 3n^2 + \frac{8}{3}n^2 = \frac{116}{3}n^2 =$$
$$= 2.416\ldots m^2,$$
$$f_5(S_m) < 4 \cdot 2n^2 + 3 \cdot (2n \cdot n + n \cdot 2n) + 2 \cdot n^2 + \frac{4}{3}n^2 = \frac{70}{3}n^2 = 1.458\ldots m^2,$$
$$f_6(S_m) < 4 \cdot \frac{1}{2}n^2 + 3 \cdot (n \cdot n) + 2 \cdot 0 + \frac{1}{3}n^2 = \frac{16}{3}n^2 = \frac{1}{3}m^2. \qquad \square$$

The proof that for any positive integer m (not necessarily divisible by 4) there is a set S_m satisfying Theorem 3 requires only more computation.

REMARK. The author [8] constructed, for any n, a set A_n of n points in general position in the plane with the following unrelated property: The ratio between the maximum and minimum distance is at most $\Theta(\sqrt{n})$, and the set A_n does not contain more than $\mathcal{O}(n^{1/3})$ vertices of a convex polygon. This is essentially the best possible result. Imre Bárány suggested that the sets A_n might be used to improve the best known upper bound of $f_3(n)$. Indeed, the set A_n contains less than $1.68n^2$ empty triangles, for any large n. However, the proof of this fact which we know is involved and so we considered the simpler construction which gives slightly worse results.

ACKNOWLEDGEMENTS. The author thanks Imre Bárány and Emmerich Welzl for fruitful discussions.

REFERENCES

[1] BÁRÁNY, I., personal communication.
[2] BÁRÁNY, I. and FÜREDI, Z., Empty simplices in Euclidean space, *Canad. Math. Bull.* **30** (1987), 436–445. *MR* **89g**:52004

[3] Erdős, P., On some problems of elementary and combinatorial geometry, *Ann. Mat. Pura. Appl.* (4) **103** (1975), 99–108. *MR* **54** #113

[4] Erdős, P. and Szekeres, G., A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470. *Zbl* **12**, 270

[5] Harborth, H., Konvexe Fünfecke in ebenen Punktmengen, *Elem. Math.* **33** (1978), 116–118. *MR* **80a**:52003

[6] Horton, J.D., Sets with no empty convex 7-gons, *Canad. Math. Bull.* **26** (1983), 482–484. *MR* **85f**:52007

[7] Katchalski, M. and Meir, A., On empty triangles determined by points in the plane, *Acta. Math. Hungar.* **51** (1988), 323–328. *MR* **89f**:52021

[8] Valtr, P., Convex independent sets and 7-holes in restricted planar point sets, *Discrete Comput. Geom.* **7** (1992), 135–152. *MR* **93e**:52037

(Received February 14, 1994)

KATEDRA APLIKOVANÉ MATEMATIKY
MATEMATICKO-FYZIKÁLNI FAKULTA
UNIVERZITA KARLOVA
CZ–118 00 PRAHA 1
MALOSTRANSKÉ NÁM. 25
CZECH REPUBLIC

Typeset by TypoTEX Ltd., Budapest
PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda, Budapest

CONTENTS

HU ISSN 0081-6906

Studia Scientiarum Mathematicarum Hungarica

VOLUME 30
NUMBERS 3–4
1995

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. It 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–1117 Budapest, Prielle Kornélia u. 19–35

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H–1364 Budapest

Tel.: (36)(1) 118–2875 Fax: (36)(1) 117–7166
e-mail: merza @ math–inst.hu

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O.Box 245
H–1519 Budapest

For 1995 volume 30 is scheduled for publication. The subscription price is $ 98.00, air delivery plus $ 20.00.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 165–182

STRONG STABILITY OF HILBERT SPACE CONTRACTION SEMIGROUPS

K. N. BOYADZHIEV and N. LEVAN

Abstract

This paper studies strong stability of strongly continuous semigroups of contraction operators on Hilbert space by: (i) a canonical decomposition type approach, and (ii) a Tauberian type approach. In the former we derive a decomposition of the Hilbert space. This then results in conditions for the semigroup to be strongly stable, as a consequence, we are able to show that a contraction semigroup with a strictly dissipative generator need not be strongly stable. In the latter, necessary and sufficient conditions for strong stability are obtained, in terms of the behavior of the resolvent of the generator on the imaginary axis. We then give a method for generating "new" strongly stable, weakly stable, or strictly contractive semigroups from a given one which has the same properties.

1. Introduction

A semigroup $T(t)$, $t \geqq 0$, over a Hilbert space H is said to be strongly stable if

$$\lim_{t\to\infty} \|T(t)x\| = 0, \qquad \forall x \in H.$$

This paper will study strong stability of Hilbert space contraction semigroups from two directions.

The first one is based on the canonical decomposition of contractions due to Langer, Nagy and Foiaş [1]. Here we make good use of the *strong stability operator* of a contraction semigroup. First, the canonical decomposition of this operator and that of its defect operator are combined to give a decomposition which results in conditions for the semigroup to be strongly stable (Theorem 1). It has been claimed that a contraction semigroup with strictly dissipative generator is strongly stable [2]. However, this is not the case in general, as will be shown using either our decomposition, or using a basic property of strict contractions (Lemma 2). Secondly, properties of strongly stable contraction semigroups (Proposition 2) as well as stability properties of strict contraction semigroups (Theorem 2) are obtained directly from the strong stability operator. Counterexamples will then be given.

1991 *Mathematics Subject Classification.* Primary 47D03; Secondary 47A45.

Key words and phrases. Hilbert space contraction semigroups, strong stability and strict contractions, Tauberian criterion for strong stability, generation of strongly stable, weakly stable or strict contraction semigroups from a given one of the same type.

0081-6906/95/$ 4.00

The second direction is a Tauberian type approach. We derive a Tauberian type criterion for strong stability using the Nagy–Foiaş Model Theory for Hilbert space contractions [1]. The key tool of our development is the *Characteristic Operator Function* (COF) — of the adjoint of the cogenerator of a contraction semigroup. Necessary and sufficient conditions for strong stability will be given in terms of the behavior of the resolvent of the generator, on the imaginary axis (Theorem 5).

We must note that very interesting results have been obtained recently by Arendt and Batty [3], Lyubich and Vu [4], and Batty and Vu [5], for strong stability of semigroups on Banach spaces. Their techniques are different from ours and the conditions are more restrictive.

A nice and interesting consequence of our study is that, by means of the Functional Calculus of Hilbert space contractions [1], we are able to generate strongly stable contraction semigroups from a given strongly stable one (Theorem 6). Thus our paper has three "cluster points", uniting several topics and results in one place for convenience.

The main results are given in Section 2. We close the paper with two remarks. The first one involves a necessary condition for strong stability, while the second one relates to the generation of strictly contractive, or weakly stable contraction semigroups from a given one which admits the same properties (Theorem 7).

Throughout the paper, by semigroups we always mean strongly continuous, i.e., of the class C_0, semigroups of bounded linear operators over separable Hilbert spaces. A contraction semigroup $T(t)$, $t \geqq 0$, with generator A will be simply written as $[T(t)]$ or, at times, as e^{At}.

2. Main results

2.1. Strong stability, strict contraction semigroups. Let $[T(t)]$ be a strongly continuous semigroup of contraction operators over a separable complex Hilbert space H, with inner product $[.,.]$ and norm $\|.\|$, and let A be its generator.

The Cayley Transform of A, denoted by T,

$$T = [A + I][A - I]^{-1},$$

is called *the cogenerator* of $[T(t)]$, and it is also a contraction operator. Moreover, a contraction semigroup and its cogenerator share a number of important properties. We recall here a few key ones and refer to [1, p. 140 and p. 143] for details.

PROPOSITION 1. (i) *For* $x\ (\neq 0) \in H$: $\lim_{t\to\infty} \|T(t)x\| = \lim_{n\to\infty} \|T^n x\|$.

(ii) *A contraction semigroup* $[T(t)]$ *is self-adjoint, normal, isometric or unitary, if and only if its cogenerator* T *is of the same type.*

(iii) *A subspace of H is invariant for $[T(t)]$ if and only if it is invariant for T.*

Let H_u denote the unitary subspace of $[T(t)]$, hence of T also, i.e., the maximal subspace which reduces $[T(t)]$ to a unitary semigroup. Then its orthogonal complement in H, denoted by H_{cnu}, is called the completely nonunitary (cnu) subspace. This means that the only subspace of H_{cnu} which reduces $[T(t)]$ to a unitary one is the trivial subspace. These two subspaces play a key role in the well-known Langer–Nagy–Foiaş canonical decomposition of Hilbert space contractions [1, p. 9].

It is clear that a strongly stable contraction semigroup is necessarily completely nonunitary. But a cnu contraction semigroup is, in general, only *weakly stable*:

$$\text{for } x \text{ and } y \in H: \lim_{t\to\infty} [T(t)x, y] = 0.$$

This is a consequence of a result on weak stability of contractions due to Foguel [6].

Now let us define the *isometric subspace* of $[T(t)]$, and of T also, to be the closed invariant subspace on which the semigroup acts isometrically, i.e., the subspace

$$(2.1)\quad H_i(T) = \{x \in H: \|T(t)x\| = \|x\|, t \geqq 0\} = \{x \in H: \|T^n x\| = \|x\|, n \geqq 0\}.$$

It is evident that if $[T(t)]$ is strongly stable then $H_i(T)$ is trivial. This, of course, implies that the semigroup is cnu. However, a cnu semigroup can be isometric, e.g., a right shift semigroup.

To proceed, we define the *strong stability operator* of $[T(t)]$ to be the non-negative contraction C given by

$$(2.2)\qquad C^2 = \operatorname*{strong\,lim}_{t\to\infty} T(t)^* T(t) = \operatorname*{strong\,lim}_{t\to\infty} T^{*n} T^n.$$

This definition makes sense since the contractions $T(t)^*T(t)$, $t \geqq 0$, are non-increasing. It follows that, for $x \in H$

$$\|Cx\|^2 = \lim_{t\to\infty} \|T(t)x\|^2.$$

Therefore,

$$(2.3)\qquad \ker C = \{x \in H: \lim_{t\to\infty} \|T(t)x\| = 0\}$$

is invariant for $[T(t)]$ and the semigroup is strongly stable on it. Thus we shall refer to $\ker C$ as the *strongly stable subspace* of $[T(t)]$ and denote it by $H_{ss}(T)$.

Next, let D be the defect operator of the non-negative contraction C, i.e.,

$$D = (I - C^2)^{1/2}.$$

Then D is again a non-negative contraction, moreover, it commutes with C [1, p. 7],

$$(2.4)\qquad CD = DC.$$

We have

LEMMA 1. *For a contraction semigroup* $[T(t)]$ *with cogenerator* T *and strong stability operator* C,

$$H_u(C) = \ker(I-C) = \ker D = H_i(T) \tag{2.5}$$

and

$$H_u(D) = \ker(I-D) = \ker C = H_{ss}(T). \tag{2.6}$$

PROOF. We have, by definition and since C is self-adjoint,

$$H_u(C) = \{x \in H : \|C^n x\| = \|x\|,\ n \geqq 0\}.$$

Then, since C is a non-negative contraction,

$$\|Cx\| = \|x\| \Leftrightarrow Cx = x \Leftrightarrow x \in \ker(I-C).$$

Thus, for $x \in \ker(I-C)$,

$$C^n x = x, \quad \forall n \geqq 0.$$

Therefore

$$\ker(I-C) \subseteqq H_u(C).$$

Next we have, again by the fact that C^n is a non-negative contraction,

$$x \in H_u(C) \Leftrightarrow C^n x = x, \quad \forall n \geqq 0.$$

Therefore

$$H_u(C) = \bigcap_{n \geqq 0} \ker(I-C^n) \subseteqq \ker(I-C).$$

Thus one half of (2.5) is proven, while the other half follows readily from the definitions of the operators C and D. Exactly the same argument applies to (2.6). This completes the proof of the Lemma.

Now, with respect to the non-negative contractions C and D, we have the following unique orthogonal decompositions for H:

$$H = H_u(C) \oplus H_{cnu}(C) = H_u(D) \oplus H_{cnu}(D). \tag{2.7}$$

Next, it follows easily from Lemma 1 that

$$H_u(C) \subseteqq H_{cnu}(D), \quad H_u(C) \perp H_u(D), \text{ and } H_u(D) \subseteqq H_{cnu}(C). \tag{2.8}$$

Combining (2.7) and (2.8) we obtain the unique orthogonal decomposition:

$$H = H_u(C) \oplus H_{cnu}(C) \cap H_{cnu}(D) \oplus H_u(D). \tag{2.9}$$

We are now ready to state the next theorem.

THEOREM 1. *Let $[T(t)]$ be a contraction semigroup over H. Then H admits the unique orthogonal decomposition:*

$$H = H_{\mathrm{i}}(T) \oplus L(T) \oplus H_{\mathrm{ss}}(T), \tag{2.10}$$

where any one of the subspaces on the right-hand side may be trivial, $H_{\mathrm{i}}(T)$ is invariant for $[T(t)]$ and $T(t)|H_{\mathrm{i}}(T)$ is isometric, $H_{\mathrm{ss}}(T)$ is invariant for $[T(t)]$ and the semigroup is strongly stable on it, while $L(T)$ is invariant for $[T(t)^*]$, and

$$\text{for } x\ (\neq 0) \in L(T) : 0 < \lim_{t\to\infty} \|T(t)x\| < \|x\|.$$

Moreover, $[T(t)]$ is strongly stable if and only if the subspaces $H_{\mathrm{i}}(T)$ and $L(T)$ are trivial.

PROOF. The proof is all but trivial. The decomposition (2.10) is actually the decomposition (2.9) in which we have set

$$L(T) = H_{\mathrm{cnu}}(C) \cap H_{\mathrm{cnu}}(D).$$

This subspace is invariant for $[T(t)^*]$ since it is the intersection of two invariant subspaces of the semigroup. Finally, for $x\ (\neq 0) \in L(T)$ we must have:

$$\|Cx\| < \|x\| \quad \text{and} \quad \|Dx\| < \|x\|.$$

Otherwise x would either belong to $H_{\mathrm{i}}(T)$, or to $H_{\mathrm{ss}}(T)$, which leads to a contradiction. The last statement of the theorem is self-evident from (2.10). This finishes the proof of the theorem.

We note that a decomposition similar to (2.10) was obtained in [7] for completely nonunitary contractions, without using the defect operator D. Also, $H_{\mathrm{u}}(T) = \{0\}$ is a sufficient condition for weak stability, while $H_{\mathrm{i}}(T) = \{0\}$ does not, in general, imply strong stability!

More can be obtained from the non-negative contraction C as is shown below.

It follows readily from (2.2) that, for $t \geqq 0$:

$$T(t)^* C^2 T(t) = C^2. \tag{2.11}$$

Therefore, for $x \in H$ and for $t \geqq 0$:

$$\|CT(t)x\| = \|Cx\|. \tag{2.12}$$

Following Nagy and Foiaş [1], we define $V(t)$, $t \geqq 0$, by

$$V(t)Cx = CT(t)x, \quad \text{for } x \in H \text{ and for } t \geqq 0.$$

Then it is evident from (2.12) that $V(t)$, $t \geqq 0$, is a well-defined isometric semigroup on Cl Range (C) — the closure of the range space of C. Thus if $C > 0$ then we have

PROPOSITION 2. *If a contraction semigroup* $[T(t)]$ *is "totally" unstable on* H, *i.e.,*

$$\text{for } x\ (\neq 0) \in H\colon \lim_{t\to\infty} \|T(t)x\| > 0,$$

then it is a quasi-affine transform [1, p. 70] *of an isometric semigroup.*

We note that the semigroup of Example 2 below satisfies the condition of this proposition. Also, if both $[T(t)]$ and $[T(t)^*]$ are totally unstable then, as has been shown in [1, p. 79], $[T(t)]$ is quasi-similar to a unitary semigroup.

To proceed we define

DEFINITION 1. (i) A closed densely defined operator A in H is *strictly dissipative* if

$$\operatorname{Re}[Ax, x] < 0, \text{ for } x\ (\neq 0) \in D(A) \text{ — the domain of } A.$$

(ii) A contraction semigroup $[T(t)]$ over H is *strictly contractive*, or simply "strict", if

$$\|T(t)x\| < \|x\|, \text{ for } t > 0 \text{ and } x\ (\neq 0) \in H.$$

It is evident that if the generator A of a contraction semigroup $[T(t)]$ is strictly dissipative then $[T(t)]$ is strict. Moreover, if $[T(t)]$ is strict then its isometric subspace $H_i(T)$ is trivial. Therefore a strict contraction semigroup is completely nonunitary. More is true as it is shown in the next Lemma.

LEMMA 2. *If* $[T(t)]$ *is a strict contraction semigroup over* H, *then so is* $[T(t)^*]$.

PROOF. Suppose that $[T(t)]$ is strict and for some $y \neq 0$:

$$\|T(t)^* y\| = \|y\|, \quad \text{for any } t > 0.$$

Then, for such a $t > 0$:

$$\|y\|^2 = \|T^*(t)y\|^2 = [T(t)T(t)^* y, y]$$
$$\leqq \|T(t)(T(t)^* y)\|\, \|y\| < \|T(t)^* y\|\, \|y\| = \|y\|\, \|y\| = \|y\|^2.$$

This is not possible, therefore $[T(t)^*]$ is strict as claimed. This finishes the proof.

Let $[T(t)]$ be a strict contraction semigroup and suppose that the implication "strict $\Rightarrow$ strong stability" holds. Then, since $[T(t)^*]$ is also strict by Lemma 2, both $[T(t)]$ and $[T(t)^*]$ are strongly stable. This certainly is not the case in general. In fact, if $[T(t)]$ and $[T(t)^*]$ are strongly stable then the semigroups belong to the class C_{00} of Nagy and Foiaş [1, p. 72].

Another way of seeing that "strict need not imply strong stability" is via the decomposition (2.10) of Theorem 1. First, if $[T(t)]$ is strict then, as

discussed above, its isometric subspace $H_i(T)(= H_u(C) = \ker(I - C)) = \{0\}$. Therefore, (2.10) becomes

$$H = L(T) \oplus H_{ss}(T),$$

where $L(T)$ is now $H_{cnu}(D)$ which, by (2.6), is equal to Cl Range (C). Thus it need not be trivial, even though $\ker(I - C)$ is trivial. Therefore $H_{ss}(T)$ need not be all of H.

The next theorem shows, yet, another stability property of the class of strict contraction semigroups.

THEOREM 2. *Let A be strictly dissipative and generate a contraction semigroup $[T(t)]$ over H. Then there is a norm $\|.\|_D$ and a Hilbert space H_D containing H such that the semigroup can be extended to a strongly stable semigroup $[T_D(t)]$ on H_D.*

PROOF. We associate with $[T(t)]$, as in the above, the stability operator C and the defect operator D. Then, since $[T(t)]$ is strict, C^2 is also strict, hence D is positive. Therefore we can define the inner product:

$$[x, y]_D = [Dx, Dy], \quad \text{for } x \text{ and } y \in H,$$

and the norm:

$$\|x\|_D = \|Dx\|, \quad \text{for } x \in H.$$

Let H_D be the completion of H in the norm $\|.\|_D$; then of course, H is dense in H_D.

Next we have from (2.11) and from the definition of D, for $x \in H$ and for $t \geqq 0$:

$$\|DT(t)x\|^2 = \|T(t)x\|^2 - \|Cx\|^2 = \|T(t)x\|_D^2.$$

Letting $t \to \infty$ we obtain:

$$\lim_{t\to\infty} \|T(t)x\|_D = 0. \tag{2.13}$$

If $[T_D(t)]$ is the extension by continuity of $[T(t)]$ from H to H_D, then it follows that the semigroup $[T_D(t)]$ is strongly stable. This finishes the proof of the theorem.

We note that if the generator A is strictly dissipative, we can define the norm

$$\|x\|_n^2 = -2\,\mathrm{Re}\,[Ax, x] > 0, \quad \text{for } x \in D(A).$$

Let $\mathbf{K}$ be the completion of $D(A)$ in this norm. We have, for $x \in D(A)$ and for $t \geqq 0$:

$$\frac{d}{dt}\|T(t)x\|^2 = 2\,\mathrm{Re}\,[AT(t)x, T(t)x] = -\|T(t)x\|_n^2.$$

Therefore it is easy to see that, for $x \in D(A)$:

$$\|Cx\|^2 - \|x\|^2 = -\int_0^\infty \|T(t)x\|_n^2 \, dt. \tag{2.14}$$

This shows that, for $x \in D(A)$, the function $T(t)x$ belongs to the space $L^2(R^+, \mathbf{K})$. However, it is clear from (2.14) that the semigroup $[T(t)]$ need not be strongly stable! This is another way of showing that the strict contraction semigroup $[T(t)]$ generated by a strictly dissipative generator A need not be strongly stable. Indeed it follows from (2.14) that $[T(t)]$ is strongly stable if and only if:

$$\text{for } x \in D(A): \quad \|x\|^2 = \int_0^\infty \|T(t)x\|_n^2 \, dt. \tag{2.15}$$

If the generator A is only dissipative then $\|.\|_n$ is a seminorm. In this case the space $\mathbf{K}$ is taken to be the completion of $D(A)$ modulo the "null" vectors. In either case, (2.15) results in a *representation* of a strongly stable contraction semigroup, namely *a strongly stable contraction semigroup is unitarily equivalent to a part (i.e., the restriction to an invariant subspace) of the left shift semigroup — over the space* $L^2(R^+, \mathbf{K})$. We refer to [1] and [8] for details. We note also that, from (2.14), if $[T(t)]$ extends to a C_0 semigroup on $\mathbf{K}$, then it is exponentially stable on $\mathbf{K}$ in the norm $\|.\|_n$, by a result of Datko [9].

We now give examples of strict contraction semigroups which are not strongly stable.

EXAMPLE 1. Let $[T(t)]$ be an isometric semigroup over a Hilbert space H with generator A. Let BB^* be a linear bounded positive operator on H, $BB^* > 0$. Then $A - BB^*$ again generates a contraction semigroup $[S(t)]$ (say). We have, for $x \in D(A)$:

$$\operatorname{Re}[(A - BB^*)x, x] = \operatorname{Re}[Ax, x] - [BB^*x, x] = 0 - \|B^*x\|^2 < 0.$$

Thus $A - BB^*$ is strictly dissipative, hence $[S(t)]$ is strict. Let C_S denote the strong stability operator of $[S(t)]$, then as in the above

$$\|C_S x\|^2 - \|x\|^2 = -2\int_0^\infty \|B^*S(t)x\|^2 \, dt, \quad \text{for } x \in H.$$

This shows that $[S(t)]$ need not be strongly stable. It is plain that $[S(t)]$ is strongly stable if and only if

$$\text{for } x \in H: \quad \|x\|^2 = 2\int_0^\infty \|B^*S(t)x\|^2 \, dt.$$

Otherwise, the subspace $H_{ss}(S)$ is characterized by

$$H_{ss}(S) = \left\{ x \in H : \|x\|^2 = 2 \int_0^\infty \|B^* S(t)x\|^2 \, dt \right\}.$$

Then, since $[S(t)]$ is strict, $H_i(S)$ is trivial. Therefore

$$L(S) = H_{ss}(S)^\perp.$$

One explicit case of this example is the next example.

EXAMPLE 2. Let $H = L^2(\mathbf{R}^+)$ and define A by

$$(Au)(x) = -u'(x),$$

and

$$D(A) = \{u \in C^1(\mathbf{R}^+) \cap L^\infty(\mathbf{R}^+),\ u(0) = 0 = u(\infty)\}.$$

Let $b(x)$ be such that

$$b \in L^1(\mathbf{R}^+) \cap L^\infty(\mathbf{R}^+),$$

$$b(x) > 0, \quad \text{for all } x \in \mathbf{R}^+,$$

and define BB^* by

$$(BB^* u)(x) = b(x)u(x).$$

Then

$$((A - BB^*)u)(x) = -u'(x) - b(x)u(x),$$

and

$$[(A - BB^*)u, u] = -\int_0^\infty u'(x)\overline{u(x)}\, dx - \int_0^\infty b(x)|u(x)|^2 \, dx.$$

From which it follows easily that, for all u $(\neq 0)$:

$$\operatorname{Re}[(A - BB^*)u, u] = -\int_0^\infty b(x)|u(x)|^2 \, dx < 0,$$

i.e., $A - BB^*$ is strictly dissipative.

Next, let $[S(t)]$ denote the contraction semigroup generated by $A - BB^*$, we find

$$(S(t)u)(x) = \begin{cases} 0, & \text{for } x \leqq t, \\ \exp\left(-\int_{x-t}^{x} b(\sigma)d\sigma \right) u(x-t), & \text{for } x > t. \end{cases}$$

Therefore

$$\|S(t)u\|_2^2 = \int_t^\infty \exp\Big(-2\int_{x-t}^{x} b(\sigma)d\sigma\Big)|u(x-t)|^2\,dx,$$

$$= \int_0^\infty \exp\Big(-2\int_{x-t}^{x} b(\sigma)\,d\sigma\Big)|u(x-t)|^2\,dx.$$

From which it follows that

(2.16)

$$\|S(t)u\|_2^2 = \int_0^\infty \exp\Big(-2\int_y^{y+t} b(\sigma)d\sigma\Big)|u(y)|^2 dy \geqq \int_0^\infty \exp\Big(-2\int_0^\infty b(\sigma)d\sigma\Big)|u(y)|^2 dy$$

$$\geqq \exp(-2\|b\|_1)\|u\|_2^2.$$

Therefore,

$$\|S(t)u\| \not\to 0 \quad \text{for all } u \neq 0, \text{ as } t \to \infty,$$

i.e., $[S(t)]$ is totally unstable: $H_{ss}(S)$ is trivial while $L(S)$ is all of $L^2(\mathbf{R}^+)$. We note that, it follows easily from the equality in (2.16) that $[S(t)]$ is indeed strict.

2.2. Strong stability of contraction semigroups: A Tauberian type result. We now turn to our second approach to strong stability. We investigate behavior of the resolvent of the generator A of a contraction semigroup $[T(t)]$:

$$(\omega I - A)^{-1} = \int_0^\infty e^{-\omega t}T(t)\,dt$$

near the imaginary axis $i\mathbf{R}$, which results in strong stability of the semigroup. Such behavior is clearly of Tauberian type.

We must note that Tauberian type results for exponential stability of contraction semigroup were obtained by Gearhart [10].

THEOREM 3 ([10], see also [11]). *Let e^{At} be a contraction semigroup with generator A in H. Let $\varrho(.)$ denote the resolvent set and $r(.)$ denote the spectral radius. The following statements are equivalent:*

(i) $i\mathbf{R} \subset \varrho(A)$ *and* $\|(itI - A)^{-1}\| \leqq K < \infty$, *for* $t \in \mathbf{R}$, *where K is a constant;*

(ii) $C = \{|z| = 1\} \subset \varrho(e^A)$;

(iii) $r(e^A) < 1$;

(iv) $\|e^{At}\| \leqq e^{-at}$, *for some $a > 0$ and all t sufficiently large.*

It is clear that the correspondence between (i) and (iv) is a Tauberian type criterion, in the sense that the behavior of the resolvent on $i\mathbf{R}$ determines the behavior of the semigroup at infinity. Gearhart's proof is based on the Nagy–Foiaş theory. Here, we shall show that, in the same way, this theory also helps to answer similar questions regarding strong stability.

Now let $[T(t)]$ be a contraction semigroup over H, with generator A and cogenerator T. We shall take $[T(t)]$ to be completely nonunitary, hence so is the contraction T. Therefore T does not admit any eigenvalues on the unit circle C. This implies that the generator A has no eigenvalues on $i\mathbf{R}$. To see this we only have to note that to each z in the spectrum $\sigma(T)$ of T, there corresponds $(z+1)(z-1)^{-1}$ in $\sigma(A)$, and vice versa, and the fact that the transformation $z \to (z+1)(z-1)^{-1}$ maps the unit disk $\mathbf{D}$ onto the left half-plane, and the unit circle C on the imaginary axis $i\mathbf{R}$.

To proceed, we need to recall the definition of the Characteristic Operator Function (COF) of the contraction operator T^* [1, p. 237]:

$$\Theta(z) = -T^* + z\sqrt{(I - T^*T)}(I - zT)^{-1}\sqrt{(I - TT^*)} \tag{2.17}$$

which is defined for all z in the open unit disk, and acts from Hilbert space $\mathrm{Cl}(D_{T^*}H)$ to $\mathrm{Cl}(D_T H)$, where D_T and D_{T^*} are the defect operators of T and T^*, respectively.

We have

LEMMA 3 ([1, pp. 238, 241]). *The COF $\Theta(z)$ satisfies the following properties:*

(i) *For $|z| < 1$ and for $x \in H$:*

$$\|D_{T^*}x\|^2 - \|\Theta(z)D_{T^*}x\|^2 = (1 - |z|^2)\|(I - zT)^{-1}D_{T^*}^2 x\|^2. \tag{2.18}$$

(ii) *$\Theta(z)$ is a bounded function with $\|\Theta(z)\| \leqq 1$ on $\mathrm{Cl}(D_{T^*}H)$. Its boundary values on the unit circle C exist almost everywhere in the strong operator topology, and*

$$\Theta(e^{is})x = \lim \Theta(z)x, \quad \text{for } x \in \mathrm{Cl}\,(D_{T^*}H),$$

when $|z| < 1$ and $z \to e^{is}$ is non-tangentially for a.e. $s \in (0, 2\pi)$.

The boundary values of $\Theta(z)$ determine when the cogenerator T is strongly stable, i.e., $T^n x \to 0$, when $n \to \infty$, for all x in H, hence when the semigroup $[T(t)]$ is strongly stable, by Proposition 1. This is due to the fact that from the functional model of T [1, Theorem 2.3*, p. 248] we obtain

THEOREM 4. *The cnu contraction T is strongly stable if and only if $\Theta(e^{is})$ is an isometric operator on $\mathrm{Cl}(D_{T^*}H)$ for a.e. $s \in (0, 2\pi)$.*

Now, let

$$w = (1 + z)(1 - z)^{-1}.$$

Then

$$\mathrm{Re}(w) = \frac{w+\overline{w}}{2} = \frac{1-|z|^2}{|1-z|^2}.$$

Therefore $|z| < 1$ corresponds to $\mathrm{Re}(w) > 0$ and C corresponds to $i\mathbf{R}$. We have

$$I - zT = (1-z)(A-wI)(A-I)^{-1},$$

and

$$\begin{aligned}(I-zT)^{-1} &= (1-z)^{-1}(A-I)(A-wI)^{-1} \\ &= (1-z)^{-1}[I+(w-1)(A-wI)^{-1}].\end{aligned}$$

Therefore (2.18) becomes

$$\begin{aligned}&\|D_{T^*}x\|^2 - \|\Theta(z)D_{T^*}x\|^2 \\ &= \|\mathrm{Re}(w)^{1/2}D_{T^*}^2 x + (w-1)(\mathrm{Re}(w))^{1/2}(A-wI)^{-1}(I-TT^*)x\|^2.\end{aligned} \tag{2.19}$$

This allows us to express the condition of Theorem 4 in terms of the behavior of $(A-wI)^{-1}$ near $i\mathbf{R}$.

We are now ready to prove our *Tauberian Criterion* for strong stability.

THEOREM 5. *Let $[T(t)]$ be a cnu contraction semigroup over H, with generator A and cogenerator T. The following statements are equivalent:*

(i) *$[T(t)]$ is strongly stable;*

(ii) *There exists a set M of $i\mathbf{R}$ such that $i\mathbf{R} \setminus M$ has measure zero, and*

$$\mathrm{Re}(w)^{1/2}(A-wI)^{-1}y \to 0, \quad \text{for all } y \in (I-TT^*)H, \tag{2.20}$$

when $\mathrm{Re}(w) > 0$ and w tends to a point in M non-tangentially (convergence may be considered only on horizontal lines, i.e., with $\mathrm{Im}(w)$ fixed, and $\mathrm{Re}(w) \to 0^+$).

PROOF. Suppose (2.20) holds. To the set M there corresponds the set

$$S = \{s \in (0, 2\pi) : (1+e^{is})(1-e^{is})^{-1} \in M\}$$

whose complement in $(0, 2\pi)$ has measure zero.

Now as $w \to w_0$ in M non-tangentially, the right-hand side of (2.19) goes to 0, while $z \to e^{is}$, $s \in S$, non-tangentially. It then follows from Lemma 3 (ii) that $\Theta(e^{is})$ is isometric on $D_{T^*}H$, hence on its closure also, for a.e. $s \in (0, 2\pi)$. Therefore T, and $[T(t)]$ also, is strongly stable by Theorem 4. This proves one half of the theorem.

Conversely, if T is strongly stable, then $\Theta(e^{is})$ is isometric on $D_{T^*}H$ for a.e. s by Theorem 4. Let S be a subset of $(0, 2\pi)$ such that $(0, 2\pi) \setminus S$ has measure 0, and

$$\lim \|\Theta(z)D_{T^*}x\| = \|\Theta(e^{is})D_{T^*}x\| = \|D_{T^*}x\|, \quad \text{for } x \in H,$$

when $s \in S$ and $z \to e^{is}$ non-tangentially — see Lemma 3 (ii). Let M be the set:

$$M = \{w = (1+e^{is})(1-e^{is})^{-1}, \text{ for } s \in S\}.$$

Then, for $\mathrm{Re}(w) > 0$ and $w \to w_0$ in M non-tangentially, $z \to e^{is}$, $(s \in S)$, non-tangentially, and as the left-hand side of (2.19) goes to 0 so does the right-hand side. Finally, as $\mathrm{Re}(w)^{1/2} D_T^2 \cdot x \to 0$, we must have

$$\mathrm{Re}(w)^{1/2}(A - wI)^{-1}(I - TT^*)x \to 0.$$

This completes the proof of the theorem.

We note that the operator $I - TT^*$ can be expressed as

$$I - TT^* = -2(A-I)^{-1} - 4(A-I)^{-1}(A^*-I)^{-1} - 2(A^*-I)^{-1},$$

where $-2(A-I)^{-1} - 4(A-I)^{-1}(A^*-I)^{-1}$ maps H into $D(A)$, and $-2(A^*-I)^{-1}$ maps H into $D(A^*)$. Thus

$$(I - TT^*)H \subseteq D(A) \cup D(A^*)$$

and we have the following sufficient condition for strong stability.

COROLLARY 1. *Let $u = \mathrm{Re}(w)$ and $v = \mathrm{Im}(w)$. If*

$$u^{1/2}(A - ivI - uI)^{-1}x \to 0, \quad \text{for } x \in D(A) \cup D(A^*),$$

as $u \to 0^+$ and for a.e. $v \in \mathbf{R}$, then e^{At} is strongly stable.

Another immediate result is Proposition 6.7 in Nagy–Foiaş [1, p. 85] which can be expressed in the spirit of this paper as follows.

COROLLARY 2. *If $\sigma(A) \cap i\mathbf{R}$ has measure zero then e^{At} is strongly stable.*

As indicated in the Introduction, for linear bounded C_0 semigroups on a Banach space, one has the following Tauberian type results for strong stability.

THEOREM ([3, 4, 5]). *Let e^{At} be a bounded linear C_0 semigroup on a Banach space. If $\sigma(A) \cap i\mathbf{R}$ is countable and $\sigma_{\mathrm{point}}(A^*) \cap i\mathbf{R} = \emptyset$ then e^{At} is strongly stable.*

The techniques in [3, 4, 5] are quite different from ours.

2.3. Generating strongly stable contraction semigroups from a given one. We now close the section by showing how to "generate" strongly stable contraction semigroups from a given one. This is given in Theorem 6 below.

THEOREM 6. *Let e^{At} be a strongly stable contraction semigroup over a Hilbert space H. Let $f(.)$ be a holomorphic function defined on the open left half-plane* **LHP**: $\{z \in \mathbf{C}: \operatorname{Re} z < 0\}$, *and $f(.)$ maps* **LHP** *into* **LHP**. *Then $f(A)$ generates the strongly stable semigroup $e^{f(A)t}$.*

To prove the theorem we first recall some preliminaries.

Let **D** denote the open unit disc. Then the transformation

$$\lambda = \frac{z+1}{z-1}$$

maps **LHP** onto **D**, and

$$z = \frac{\lambda+1}{\lambda-1}$$

maps **D** onto **LHP**, as

$$-\operatorname{Re}(z) = \frac{1-|\lambda|^2}{|\lambda-1|^2}.$$

Since $f(.)$: **LHP** $\to$ **LHP**, the function

$$\frac{f\left\{\frac{\lambda+1}{\lambda-1}\right\}+1}{f\left\{\frac{\lambda+1}{\lambda-1}\right\}-1} = \varphi_f(\lambda) \quad \text{(say)} \tag{2.21}$$

maps **D** into itself, so that φ_f belongs to H^∞ and $|\varphi_f(\lambda)| < 1$ for λ in **D**.

As before, let T be the cogenerator of e^{At}, and define $\varphi_f(T)$ as in [1]. Then, by von Neumann's inequality

$$\|\varphi_f(T)\| \leqq \|T\| \sup_{\lambda \in \mathbf{D}} |\varphi_f(\lambda)| \leqq \|T\|.$$

Therefore $\varphi_f(T)$ is again a contraction. This allows us to define $f(A)$ as

$$f(A) = [\varphi_f(T) + I][\varphi_f(T) - I]^{-1},$$

and it is maximal dissipative. Moreover, if T is cnu then so is $\varphi_f(T)$ [1, p. 113].

We now prove the next Lemma before giving the proof of Theorem 6.

LEMMA 4. *Let T be a completely nonunitary and strongly stable contraction. Then*

(i) *If $\varphi(.): \mathbf{D} \to \mathbf{D}$ is a holomorphic function with $\varphi(0) = 0$, then $\varphi(T)$ is strongly stable.*

(ii) *If*

$$\varphi_\alpha(\lambda) = \frac{\lambda - \alpha}{1 - \overline{\alpha}\lambda}, \quad |\alpha| < 1,$$

then $\varphi_\alpha(T)$ is also strongly stable.

PROOF. For (i) we have,

$$\varphi(\lambda) = \lambda\psi(\lambda),$$

where $\psi \in H^\infty(\mathbf{D})$, and $\|\psi\| \leqq 1$. Therefore, for x in H:

$$\|\varphi(T)^n x\| = \|\{T\psi(T)\}^n x\| = \|\psi(T)^n T^n x\| \leqq \|T^n x\| \to 0, \text{ as } n \to \infty.$$

Part (ii) follows from the fact that T is strongly stable if and only if the contraction $T_\alpha = \varphi_\alpha(T) = [T - \alpha I][I - \overline{\alpha}T]^{-1}$ is strongly stable. This is so because of the relationship between the COF's of T^* and T_α^* (see [1, p. 240, (1.7)], and Theorem 4). Hence the proof of the Lemma is completed.

We are now ready to give the proof of Theorem 6.

PROOF OF THEOREM 6. Let $f(.)$: **LHP** $\to$ **LHP** and consider $\varphi_f(\lambda)$ as defined in (2.21). We need to show that $\varphi_f(T)$ is strongly stable. If $\varphi_f(0) = 0$ then this is indeed the case, by Lemma 4(i). Suppose now that $\varphi_f(0) \neq 0$. Set $\alpha = \varphi_f(0)$ and define

$$\psi(\lambda) = \frac{\varphi_f(\lambda) - \alpha}{1 - \overline{\alpha}\varphi_f(\lambda)},$$

then ψ maps **D** into **D**, and $\psi(0) = 0$. Therefore $\psi(T)$ is strongly stable. But

$$\varphi_f(T) = [\psi(T) + \alpha][1 + \overline{\alpha}\psi(T)]^{-1},$$

therefore, by Lemma 4 (ii), $\varphi_f(T)$ is strongly stable. Hence so is $e^{f(A)t}$. This completes the proof of the theorem.

It is of interest to note that all functions which preserve **LHP** are given by the following formula:

$$f(z) = az + ib + \int_{-\infty}^{\infty} \frac{1 - itz}{z - it} d\mu(t),$$

where $a \geqq 0$, $b \in \mathbf{R}$ are constants, and μ is a non-negative measure on **R** such that

$$\int_{\mathbf{R}} d\mu(t) < \infty.$$

See, for instance [12, p. 22]. This together with the results of Theorem 6 allow us to "generate" new strongly stable contraction semigroup from a given one.

3. Concluding remarks

We have seen in Corollary 2 that a sufficient condition for a contraction semigroup $[T(t)]$, with generator A, to be strongly stable is that the intersection of the spectrum $\sigma(A)$ and $i\mathbf{R}$ has measure 0. It is natural to ask whether this condition is also necessary?

Now returning to the necessary and sufficient conditions for strong stability of Theorem 5 we can show that, by adding a further condition, we come close to answering the above question.

First, we know that, for a contraction semigroup,

$$\|(wI - A)^{-1}\| < \frac{1}{\mathrm{Re}(w)}, \tag{3.1}$$

for $w \in \mathbf{C}$ and $\mathrm{Re}(w) > 0$. Moreover, [13, p. 566],

$$\|(wI - A)^{-1}\| \geqq \frac{1}{\mathrm{dist}\,\{w, \sigma(A)\}}. \tag{3.2}$$

Suppose now that for some $b \in \mathbf{R}$, $ib \in \sigma(A)$, and let $w = a + ib$, $a > 0$. Then $\mathrm{dist}\,\{w, \sigma(A)\} = \mathrm{Re}(w)$. It then follows from (3.1), (3.2), and since $\sigma(A)$ is contained in **LHP**,

$$\|(wI - A)^{-1}\| = \frac{1}{\mathrm{Re}(w)} \quad \text{on the ray} \quad \{\mathrm{Im}(w) = b,\ \mathrm{Re}(w) > 0\},$$

and,

$$\sqrt{\mathrm{Re}(w)}\|(wI - A)^{-1}\| = \frac{1}{\sqrt{\mathrm{Re}(w)}} \to \infty, \ \text{as } \mathrm{Re}(w) \to 0^+.$$

Therefore, we conclude that the condition:

$$\sqrt{\mathrm{Re}(w)}\|(wI - A)^{-1}\| \to 0, \ \text{as } \mathrm{Re}(w) \to 0^+, \ \text{for a.e. } b = \mathrm{Im}(w),$$

in the uniform operator topology, implies that

$$\mathrm{mes}\{\sigma(A) \cap i\mathbf{R}\} = 0. \tag{3.3}$$

This in turn implies strong stability by Corollary 2. It follows from this and from Theorem 5 that the condition:

$$\sqrt{\mathrm{Re}(w)}(wI - A)^{-1}x \to 0, \ \text{as } \mathrm{Re}(w) \to 0^+,$$

$$\text{for a.e. } b = \mathrm{Im}(w) \text{ and } x \in \mathrm{Range}\,(I - TT^*),$$

is "close" to implying that (3.3) is necessary for strong stability!

Finally, we note that functions which preserve the **LHP**, preserve also other interesting properties. These are given in the next theorem.

THEOREM 7. *Let $f(.)$ be a holomorphic function which preserves the* **LHP**. *Then $f(.)$ preserves the strict dissipativity property, as well as the weak stability property, of a contraction semigroup.*

As in the case of Theorem 6, we have the following Lemma which is the analogue of Lemma 4.

LEMMA 5. *Let T be a strict contraction on H: $\|Ty\| < \|y\|$ for all y $(\neq 0) \in H$. Then*

(i) *If $\varphi(.)$: $\mathbf{D} \to \mathbf{D}$ is a holomorphic function with $\varphi(0) = 0$, then $\varphi(T)$ is also a strict contraction, and*

(ii) *$T_\alpha = \varphi_\alpha(T) = [T - \alpha I][I - \overline{\alpha}T]^{-1}$ is also a strict contraction for all $|\alpha| < 1$.*

PROOF. (i) As in the proof of Lemma 4, we have, since $\varphi(z) = z\psi(z)$ and $\psi(.)$: $\mathbf{D} \to \mathbf{D}$,

$$\|\varphi(T)y\| = \|T\psi(T)y\| < \|\psi(T)y\| \leqq \|y\|.$$

(ii) This follows from the fact that [1, p. 14], for $y = (I - \overline{\alpha}T)^{-1}x$,

$$\|x\|^2 - \|T_\alpha x\|^2 = (1 - |\alpha|^2)(\|y\|^2 - \|Ty\|^2)$$

and we complete the proof of the Lemma.

PROOF OF THEOREM 7. Let $[T(t)]$ be a contraction semigroup on H with generator A and cogenerator T. Then, for $x \in D(A)$:

$$\|(A + I)x\|^2 - \|(A - I)x\|^2 = 4\mathrm{Re}[Ax, x].$$

Thus A is strictly dissipative if and only if T is strictly contractive, since for y $(\neq 0) \in H$:

$$\|Ty\| = \|(A + I)(A - I)^{-1}y\| < \|(A - I)(A - I)^{-1}y\| = \|y\|.$$

Then, if T is a strict contraction then so is $\varphi(T)$ for all holomorphic functions $\varphi(.)$: $\mathbf{D} \to \mathbf{D}$ by Lemma 5. Thus, as in the proof of Theorem 6, we apply this to the function $\varphi_f(.)$ and we are through.

Similarly, suppose that $[T(t)]$ is completely nonunitary then so is its cogenerator T and vice versa. Moreover, $[T(t)]$ is weakly stable if it is completely nonunitary. Thus we only need to show that if T is cnu, then so is $\varphi(T)$ for every holomorphic function $\varphi(.)$: $\mathbf{D} \to \mathbf{D}$. However, this is indeed the case as has been shown in [1, p. 113]. As before, applying this to the function $\varphi_f(.)$ and we are through. This finishes the proof of the theorem.

Acknowledgements

We are indebted to Professor J. Prüss (Paderborn, FDR) for providing us with Example 2, and to Professor G. Greiner (Tübingen, FDR) for valuable comments.

REFERENCES

[1] Sz.-Nagy, B. and Foiaş, C., *Harmonic analysis of operators on Hilbert space*, North-Holland Publ. Co., Amsterdam–London; American Elsevier Publ. Co., New York; Akadémiai Kiadó, Budapest, 1970. *MR* **43** #947

[2] D'yachenko, S. V., Asymptotic stability of semigroups with a strictly dissipative generating operator, *Mat. Zametki* **28** (1980), no. 1, 75–78, 168 (in Russian). *MR* **81m**:47018. See also *Math. Notes* (1) **28** (1980), 502–503.

[3] Arendt, W. and Batty, C. J. K., Tauberian theorems and stability of one-parameter semigroups, *Trans. Amer. Math. Soc.* **306** (1988), 837–852. *MR* **89g**:47053

[4] Lyubich, Yu. I. and Vu, Q. P., Asymptotic stability of linear differential equations in Banach spaces, *Studia Math.* **88** (1988), 37–42. *MR* **89e**:47062

[5] Batty, C. J. K. and Vu, Q. P., Stability of individual elements under one-parameter semigroups, *Trans. Amer. Math. Soc.* **322** (1990), 805–818. *MR* **91c**:47072

[6] Foguel, S. R., Powers of a contraction in Hilbert space, *Pacific J. Math.* **13** (1963), 551–562. *MR* **29** #473

[7] Levan, N., Canonical decompositions of completely nonunitary contractions, *J. Math. Anal. Appl.* **101** (1984), 514–526. *MR* **86c**:47020

[8] Fillmore, P. A., *Notes on operator theory*, Van Nostrand Reinhold Mathematical Studies, No. 30, Van Nostrand Reinhold, New York, 1970. *MR* **41** #2414

[9] Datko, R., Extending a theorem of A. M. Liapunov to Hilbert space, *J. Math. Anal. Appl.* **32** (1970), 610–616. *MR* **42** #3614

[10] Gearhart, L., Spectral theory for contraction semigroups on Hilbert space, *Trans. Amer. Math. Soc.* **236** (1978), 385–394. *MR* **57** #1191

[11] Prüss, J., On the spectrum of C_0-semigroups, *Trans. Amer. Math. Soc.* **284** (1984), 847–857. *MR* **85f**:47044

[12] Donoghue, Jr., W. F., *Monotone matrix-functions and analytic continuation*, Die Grundlehren der mathematischen Wissenschaften, Band 207, Springer-Verlag, New York, 1974. *MR* **58** #6279

[13] Dunford, N. and Schwartz, J., *Linear operators*. Part I, *General theory*, Pure and Applied Mathematics, Vol. 7, Interscience, New York, 1958. *MR* **22** #8302

(Received July 12, 1991)

DEPARTMENT OF MATHEMATICS
MEYER HALL
OHIO NORTHERN UNIVERSITY
ADA, OH 45810
U.S.A.

DEPARTMENT OF ELECTRICAL ENGINEERING
56-125B, ENGR IV
UNIVERSITY OF CALIFORNIA IN LOS ANGELES
LOS ANGELES, CA 90024-1594
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 183–187

COMMUTATIVITY OF SEMIPRIME RINGS WITH POWER CONSTRAINTS

H. A. S. ABUJABAL, H. E. BELL*, M. S. KHAN and M. A. KHAN

For thirty years, various authors have studied commutativity in rings satisfying polynomial identities of the form

$$(*) \qquad (xy)^n = x^n y^n, \qquad n > 1$$

(see [7] for a fairly inclusive list of references). In the last few years there have been studies of rings satisfying variable-n versions of $(*)$ or conditions of the form $[(xy)^n - x^n y^n, x] = 0$ or $[(xy)^n - x^n y^n, z] = 0$ [1, 2]. Most recently, H. E. Bell and A. A. Klein [8] have proved that a semiprime ring R must be commutative if for each $x \in R$ there exists an integer $n = n(x) > 1$ such that $(xy)^n - x^n y^n$ is central for each $y \in R$.

Naturally enough, some authors have explored conditions of the form

$$(**) \qquad (xy)^n = y^n x^n,$$

but apparently only for fixed n [4, 5, 6]. It is our purpose to study variable-n versions of $(**)$ and generalizations thereof.

Throughout the paper, $[x, y]$ will denote the commutator $xy - yx$, and Z or $Z(R)$ will denote the center of the ring R.

1. Rings with $[(xy)^n - y^n x^n, x] = 0$

THEOREM 1. *Let R be a ring with no nonzero nil ideals. If for each $x, y \in R$ there exists an integer $n = n(x, y) \geqq 1$ such that $[(xy)^n - y^n x^n, x] = 0 = [(yx)^n - x^n y^n, x]$, then R is commutative.*

PROOF. Let $x, y \in R$. Then there exists $n = n(x, y) \geqq 1$ such that

$$[(xy)^n - y^n x^n, x] = 0 \qquad \text{and} \qquad [(yx)^n - x^n y^n, x] = 0.$$

1991 *Mathematics Subject Classification.* Primary 16U80; Secondary 16U99.

Key words and phrases. Commutativity theorems, semiprime rings, power constraints, commutator constraints.

*Supported by the Natural Sciences and Engineering Research Council of Canada Grant No. A3961.

The first of these conditions may be written as

$$x((xy)^n - (yx)^n) = xy^n x^n - y^n x^{n+1}; \tag{1}$$

the second may be written as

$$((xy)^n - (yx)^n)x = x^{n+1}y^n - x^n y^n x. \tag{2}$$

Right-multiplying (1) by x and left-multiplying (2) by x gives

$$xy^n x^{n+1} - y^n x^{n+2} = x^{n+2}y^n - x^{n+1}y^n x,$$

that is, $x[y^n, x^{n+1}] = [y^n, x^{n+1}]x$. It follows that $[[y^n, x^{n+1}], x^{n+1}] = 0$, hence R is commutative by Theorem 1 of [9].

Making use of this result, we obtain an analogue of a result in [2].

THEOREM 2. *Let R be a semiprime ring, and suppose that for each $x \in R$, there exists an integer $n = n(x) > 1$ such that $[(xy)^n - y^n x^n, x] = 0 = [(yx)^n - x^n y^n, x]$ for all $y \in R$. Then R is commutative.*

This theorem is an immediate consequence of Theorem 1, once we establish the following lemma.

LEMMA 1. *Let R be semiprime, and suppose that for each $x \in R$ there exists an integer $n = n(x) > 1$ for which $[(xy)^n - y^n x^n, x] = 0$ for all $y \in R$. Then R has no nonzero nilpotent elements.*

PROOF. Let $a \in R$ with $a^2 = 0$. Then there exists $n = n(a) > 1$ such that $[(ay)^n - y^n a^n, a] = 0$ for all $y \in R$. It follows at once that $(ay)^n a = 0$, so that $(ay)^{n+1} = 0$ for all $y \in R$. Thus aR is a nil right ideal of bounded index, which must be trivial by a well-known result of Levitzki [11]. Hence $a = 0$.

2. Semiprime rings with $(xy)^n - y^n x^n \in Z$

Our final theorem, a companion theorem to Theorem 4 of [8], is

THEOREM 3. *Let R be a semiprime ring with the property that for each $x \in R$ there exists an integer $n = n(x) > 1$ such that $(xy)^n - y^n x^n \in Z(R)$ for all $y \in R$. Then R is commutative.*

Before beginning the general proof, we dispose of an important special case, which parallels the result in [1].

LEMMA 2. *Let R be a prime ring such that for each $x \in R$ there exists an integer $n = n(x) > 1$ for which $(xy)^n = y^n x^n$ for all $y \in R$. Then R is commutative.*

PROOF. By Lemma 1, R has no nonzero nilpotent elements; and since R is prime, it has no nonzero divisors of zero. In fact, R is an Ore domain, hence

embeddable in a division ring; therefore we shall be able to use multiplicative inverses in appropriate contexts.

Let $x, y \in R \setminus \{0\}$, and take $n, m > 1$ such that

$$(xz)^n = z^n x^n \quad \text{for all } z \in R, \tag{3}$$

$$(yz)^m = z^m y^m \quad \text{for all } z \in R. \tag{4}$$

It follows by induction that

$$\begin{gathered}(x^t z)^n = z^n x^{nt} \text{ and } (y^t z)^m = z^m y^{mt} \\ \text{for all } z \in R \text{ and all positive integers } t.\end{gathered} \tag{5}$$

Now right-multiplying (3) by x and left-multiplying by x^{-1} gives

$$(zx)^n = x^{-1} z^n x^{n+1} \quad \text{for all } z \in R.$$

Thus

$$(yx)^{nm} = ((yx)^n)^m = (x^{-1} y^n x^{n+1})^m = x^{-1}(y^n x^n)^m x,$$

and using (5), we get

$$(yx)^{nm} = x^{-1}(x^{nm} y^{nm})x = x^{nm-1} y^{nm} x. \tag{6}$$

On the other hand, (4) and (5) give

$$(yx)^{nm} = ((yx)^m)^n = (x^m y^m)^n = y^{nm} x^{nm}, \tag{7}$$

and comparing (6) and (7) yields $(x^{nm-1} y^{nm} - y^{nm} x^{nm-1})x = 0$. Since R is a domain, we now have $x^{nm-1} y^{nm} = y^{nm} x^{nm-1}$, and R is commutative by a well-known theorem of Herstein [10].

Proof of Theorem 3. Since R is a subdirect product of prime rings, we may assume that R is prime, hence (by Lemma 1) a domain. By Lemma 2, we need only consider the case of $Z(R) \neq \{0\}$, so we can localize at $Z(R) \setminus \setminus \{0\}$, thereby embedding R in a domain R^* with 1 which satisfies our original hypothesis. Our immediate goal is to show that R^* is a division ring.

Consider $x \in R \setminus \{0\}$. If for each $v \in xR$ there exists $m = m(v) > 1$ such that $(vy)^m = y^m v^m$ for all $y \in xR$, then xR is a nonzero commutative right ideal of R, which forces R to be commutative. Thus, we may assume that for each $x \in R \setminus \{0\}$, there exists $xw \in xR$ such that

$$(xwy)^{n(xw)} - y^{n(xw)}(xw)^{n(xw)} \in Z \setminus \{0\} \quad \text{for some } y \in xR.$$

In particular, there exists $u \in R$ for which $xu \in Z \setminus \{0\}$; and noting how R^* is constructed, we see that R^* is indeed a division ring.

Now let x be an arbitrary element of R^*. Since there exists $n = n(x) > 1$ such that $(xy)^n - y^n x^n \in Z(R^*)$ for all $y \in R^*$, we have

$$xyy^n x^n = y^n x^n xy \quad \text{for all} \quad y \in R^*. \tag{8}$$

Thus R^* satisfies a generalized polynomial identity, hence is finite-dimensional over $Z(R^*)$ by Theorem 13 of [3]. In view of Wedderburn's theorem on finite division rings, we assume that $Z(R^*)$ is infinite.

We now apply a typical Vandermonde argument. For $x \in R^*$, choose n such that (8) holds, and replace y by $y + \lambda$ for $\lambda \in Z(R^*)$, thereby obtaining

$$x(y+\lambda)^{n+1} x^n - (y+\lambda)^n x^{n+1} (y+\lambda) = 0. \tag{9}$$

In view of (8) and the fact that the y-free terms of (9) sum to 0, we get

$$\sum_{i=1}^{n} \lambda^i w_i(x, y) = 0, \tag{10}$$

where $w_i(x, y)$ denotes the "coefficient" of λ^i on the left side of (9). Doing this for n distinct nonzero λ, we get a system $Aw = 0$, where A is an $n \times n$ Vandermonde matrix with nonzero determinant and w is an $n \times 1$ matrix with i^{th} entry $w_i(x, y)$. Thus, $w_i(x, y) = 0$ for each i; and examining $w_n(x, y)$, we get $(n+1)xyx^n - x^{n+1}y - nyx^{n+1} = 0$ for all $y \in R^*$. But this condition can be written as $[x^{n+1}, y] = (n+1)[x, y]x^n$ for all $y \in R^*$; hence R^* is commutative by Theorem 1 of [8].

REFERENCES

[1] Abu-Khuzam, H., A commutativity theorem for semiprime rings, *Bull. Austral. Math. Soc.* **27** (1983), 221–224. *MR* **84i**:16037

[2] Abu-Khuzam, H. and Yaqub, A., Commutativity of certain semiprime rings, *Studia Sci. Math. Hungar.* **24** (1989), 33–36. *MR* **90b**:16041

[3] Amitsur, S. A., Generalized polynomial identities and pivotal monomials, *Trans. Amer. Math. Soc.* **114** (1965), 210–226. *MR* **30** #3117

[4] Ashraf, M., A study of certain commutativity condition for associative ring, Ph. D. Thesis, Aligarh Muslim University, Aligarh, India, 1986.

[5] Ashraf, M. and Quadri, M. A., On commutativity of rings with some polynomial constraints, *Bull. Austral. Math. Soc.* **41** (1990), 201–206. *MR* **91e**:16041

[6] Awtar, R., On the commutativity of nonassociative rings, *Publ. Math. Debrecen* **22** (1975), 177–188. *MR* **52** #5752

[7] Bell, H. E., The identity $(XY)^n = X^n Y^n$: does it buy commutativity?, *Math. Mag.* **55** (1982), 165–170. *MR* **83j**:16045

[8] Bell, H. E. and Klein, A. A., Two commutativity problems for rings, *Studia Sci. Math. Hungar.* **28** (1993), 159–162.

[9] Chuang, C. L. and Lin, C. H., On a conjecture by Herstein, *J. Algebra* **126** (1989), 119–138. *MR* **90i**:16028

[10] Herstein, I. N., A commutativity theorem, *J. Algebra* **38** (1976), 112–118. *MR* **53** #549

[11] KLEIN, A. A., A new proof of a result of Levitzki, *Proc. Amer. Math. Soc.* **81** (1981), 8. *MR* **81j**:16011

(Received July 12, 1991)

H. A. S. Abujabal and M. A. Khan

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
KING ABDULAZIZ UNIVERSITY
P. O. BOX 31464
JEDDAH 21497
SAUDI ARABIA

H. E. Bell

DEPARTMENT OF MATHEMATICS
BROCK UNIVERSITY
ST. CATHARINES, ONTARIO
L2S 3A1
CANADA

M. S. Khan

DEPARTMENT OF MATHEMATICS AND COMPUTING
COLLEGE OF SCIENCE
SULTAN QABOOS UNIVERSITY
P.O.BOX 32486
AL-KHOD
MUSCAT
SULTANATE OF OMAN

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 189–195

COMPACT ABELIAN LIE GROUP ACTION AND THE GROUP $N_*^G[F]$

N. J. DEV and S. S. KHARE

Abstract

The object of this paper is to construct suitable family $\bar{F}_j$ of G-slice types for every compact abelian Lie group so that the bordism theory $N_*^G[\bar{F}_j]$ vanishes.

§ 1. Introduction

In [4] Kosniowski constructed, for a finite abelian group, an equivariant bordism theory which vanishes. His results were later extended in [1] for a finite group (not necessarily abelian) and in [2] for a torus. Here we consider the action of a compact abelian Lie group G and construct a family $\bar{F}_j$ of G-slice types such that $N_*^G[\bar{F}_j]$ is zero.

§ 2. Preliminaries

By the structure theorem, a compact abelian Lie group G can be decomposed as $T^k \times \mathcal{G}$ where T^k is the k-torus and $\mathcal{G}$ is a finite abelian group. By a G-slice type we mean a pair $[H; V]$, H is a subgroup of G and V is an H-module containing no trivial H-submodule. For a G-manifold M, if we take G_x to be the isotropy subgroup at $x \in M$ then it is well-known that there exists a G_x invariant neighbourhood of x which is equivariantly diffeomorphic to a G_x-module $\overline{V}_x = V_x \oplus V_x'$, where V_x is the G_x-submodule in which no nonzero vector remains fixed by all of G_x and V_x' is one on which G_x acts trivially. The pair $[G_x; V_x]$ is designated as the slice type of x. A family F of G-slice types is a collection such that $[H; V] \in F$ implies that $[G_x; V_x] \in F$ for every $x \in G \times_H V$.

Denote an element of G by $(y_1, y_2, \ldots, y_k, g)$, where $y_i \in S^1$, $1 \leqq i \leqq k$ and $g \in \mathcal{G}$, and define homomorphisms $p_i : G \to G$, $1 \leqq i \leqq k$ by $p_i(y_1, \ldots, y_k, g) = (0, 0, \ldots, y_i, 0, \ldots, 0, e)$. Consider the elementary abelian 2-group $\mathbf{Z}_2^k$

1991 *Mathematics Subject Classification.* Primary 57R85.
Key words and phrases. Equivariant bordism, G-slice types.

0081–6906/95/$ 4.00

contained in $T^k \subset G$. If x_i is the generator of $p_i(\mathbf{Z}_2^k)$ then we see that $\{x_1, x_2, \ldots, x_k\}$ form a basis of $\mathbf{Z}_2^k$. Consider

$$F_j = \{[H;V] : p_i(H) = \text{finite or } S^1, \forall i \text{ and } \langle x_i \rangle \not\subset p_i(H) \text{ for at least } (k-j) \text{ values of } i\}.$$

Clearly, $F_0 \subset F_1 \subset \ldots \subset F_k$ are families of G-slice types. For $[K;U] \in F_j$, $0 \leqq j < k$, we consider $H = K \times \langle x_{i_1} \rangle$, i_1 being the first place for which $\langle x_i \rangle \not\subset p_i(K)$. The projection $p : H \to K$ enables us to extend the K-action to H-action on U. We denote the corresponding H-module by p^*U. Also the real line with the antipodal action of $\langle x_{i_1} \rangle$ becomes an H-module via the projection map $q : H \to \langle x_{i_1} \rangle$ and will be denoted by $V(K)$ or sometimes by $V(x_{i_1})$. As in [3], we define an extension map e on each of the families F_j, $0 \leqq j < k$ by

$$e[K;U] = [H; V(K) \oplus p^*U]$$

for $[K;U] \in F_j$. By writing $\bar{F}_j = F_j \cup e(F_j)$, we see that F_j is also family with $F_j \subset \bar{F}_j \subset F_{j+1}$.

Let $\mathcal{F} \subset \mathcal{F}'$ be a pair of families of G-slice type. A compact G-manifold M^n is said to be $(\mathcal{F}', \mathcal{F})$-free if the G-slice type $[G_x; V_x] \in \mathcal{F}'$, $\forall x \in M^n$ and $[G_x; V_x] \in \mathcal{F}$, $\forall x \in \partial M^n$. If $\mathcal{F} = \emptyset$, M^n is called $\mathcal{F}'$-free. Two $(\mathcal{F}', \mathcal{F})$-free compact G-manifolds M_1^n and M_2^n are said to be $(\mathcal{F}', \mathcal{F})$-bordant, if $\exists$ an $(n+1)$-dimensional $(\mathcal{F}', \mathcal{F})$-free compact G-manifold W^{n+1} and an n-dimensional $(\mathcal{F}, \mathcal{F})$-free G-manifold $V^n \subset W^{n+1}$ for which the disjoint union $M_1^n \cup M_2^n$ is equivariantly embedded into ∂W^{n+1} and $(M_1^n \cup M_2^n) \cup V^n = \partial W^{n+1}$ with $\partial M_1^n \cup \partial M_2^n = (M_1^n \cup M_2^n) \cap V^n = \partial V^n$. This is an equivalence relation in the set of all compact $(\mathcal{F}', \mathcal{F})$-free compact G-manifolds. The set of equivalence classes forms an abelian group $N_n^G[\mathcal{F}', \mathcal{F}]$, the operation being induced by disjoint union. We denote $N_n^G[\mathcal{F}', \emptyset]$ by $N_n^G[\mathcal{F}']$. $N_*^G[\mathcal{F}', \mathcal{F}] = \bigoplus_n N_n^G[\mathcal{F}', \mathcal{F}]$ is a graded N_*-module, N_* being the bordism ring.

For a given G-slice type $\rho = [H;U]$, a G-vector bundle E over a manifold M is said to be of type ρ if the set of points in E having slice type ρ is homeomorphic to M. A G-vector bundle E_1 over M_1 of type ρ is said to be bordic to another G-vector bundle E_2 over M_2 of type ρ if there exists a G-vector bundle F over W of type ρ such that $\partial W = M_1 \cup M_2$, $F|_{M_1} = E_1$ and $F|_{M_2} = E_2$. This bordism relation leads to the bundle bordism group $N_n^G[\rho]$, where n is the total dimension of the vector bundles in consideration.

Let $\mathcal{F} \subsetneq \mathcal{F}'$ be families of G-slice types with $\mathcal{F}' = \mathcal{F} \cup \{\rho\}$. We have a natural N_*-module homomorphism

$$\nu_\rho : N_*^G[\mathcal{F}', \mathcal{F}] \to N_*^G[\rho]$$

given by $\nu_\rho([M, \theta]) =$ the bordism class of normal bundle to the submanifold N in M, where N consists of all $x \in M$ with slice type $[G_x; V_x]$ at x

being equal to ρ. The homomorphism ν_ρ is an isomorphism, for the inverse homomorphism μ can be defined as

$$\mu\colon\ N_*^G[\rho] \to N_*^G[\mathcal{F}', \mathcal{F}]$$

given by $\mu[\xi] =$ the bordism class of the top space of the disc bundle $D(\xi)$. Also we have the following commutative diagram:

$$\begin{array}{ccc} N_*^G[\mathcal{F}', \mathcal{F}] & \xrightarrow{\partial} & N_{*-1}^G[\mathcal{F}] \\ & \approx\searrow \nu_\rho \qquad \delta\nearrow & \\ & N_*^G[\rho] & \end{array}$$

where $\delta[\xi] =$ the bordism class of the top space of the sphere bundle $S(\xi)$. Therefore the long exact sequence

$$\cdots \longrightarrow N_n^G[\mathcal{F}] \xrightarrow{i} N_n^G[\mathcal{F}'] \xrightarrow{j} N_n^G[\mathcal{F}', \mathcal{F}] \xrightarrow{\partial} N_{n-1}^G[\mathcal{F}] \longrightarrow \cdots$$

associated to the pair $(\mathcal{F}', \mathcal{F})$ gives the following exact sequence

$$\longrightarrow N_*^G[\mathcal{F}] \xrightarrow{i} N_*^G[\mathcal{F}'] \xrightarrow{\tilde{\nu}_\rho} N_*^G[\rho] \xrightarrow{\delta} N_{*-1}^G[\mathcal{F}] \longrightarrow \cdots,$$

where $\tilde{\nu}_\rho[M, \theta] =$ the bordism class of the normal bundle to the submanifold N of M consisting of points with slice type ρ.

The proof of the following is similar to that given for the Lemma 4.5.8 of [3].

LEMMA 2.1. *If ρ is a G-slice type of F_j, $0 \leqq j < k$, then*

$$N_*^G[e(\rho)] = N_{*-1}^G[\rho]. \qquad \square$$

This lemma together with long exact sequence provide an inductive method to calculate $N_n^G[\tilde{F}_j]$.

§ 3. Ordering the G-slice types

For $0 \leqq j < k$, the family $\tilde{F}_j$ is at first partitioned into subsets

$$\tilde{F}_j^n = \{[H; V] \in \tilde{F}_j : \dim V = n\}.$$

Each of these subsets $\tilde{F}_j^n$ is further partitioned into subsets

$$\tilde{F}_j^{n,r} = \{[H; V] \in \tilde{F}_j^n :\ \text{the maximal torus of } H \text{ is } r\text{-dimensional}\}$$

$0 \leqq r \leqq j$. For $[H;V] \in \bar{F}_j^{n_r}$, we look at $\lambda = |H/(S^1)^r|$ and order the G-slice types of $\bar{F}_j^{n_r}$ in increasing order of λ. A typical subgroup H with $[H;V] \in \bar{F}_j$ and with maximal torus of H being r-dimensional will look like

$$\{(e^{in_{11}\theta_1}, \ldots, e^{in_{1t_1}\theta_1}), \ldots, (e^{in_{r1}\theta_r}, \ldots, e^{in_{rt_r}\theta_r}) : 0 \leqq \theta_i \leqq 2\pi\} \times H,$$

where H is a finite subgroup and the greatest common divisor of $n_{i1}, \ldots, n_{it_i}$ is 1, $i = 1, \ldots, r$. Therefore G-slice types having the same value of λ are countable and are ordered arbitrarily. But once an ordering is chosen for them we stick to it throughout. We start ordering the members of $\bar{F}_j^{0_0}$ as

$$\rho_0, \quad \rho_1, \quad \rho_2, \cdots$$

by the ordinals less than ω. The members of $F_j^{0_1}$ are ordered as

$$\rho_\omega, \quad \rho_{\omega+1}, \quad \rho_{\omega+2}, \cdots$$

by the ordinals less than $\omega(2)$. We order members of $F_j^{0_j}$ as

$$\rho_{\omega(j)}, \quad \rho_{\omega(j)+1}, \cdots$$

by the ordinals less than $\omega(j+1)$. Similarly, we order G-slice types of $\bar{F}_j^1$ starting from $\rho_{\omega(j+1)}$ as below

$$\begin{aligned}
&\rho_{\omega(j+1)}, \quad \rho_{\omega(j+1)+1}, \cdots \\
&\rho_{\omega(j+2)}, \quad \rho_{\omega(j+2)+1}, \cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\rho_{\omega(2j+1)}, \quad \rho_{\omega(2j+1)+1}, \cdots < \rho_{\omega(2j+2)}.
\end{aligned}$$

In general, members of $\bar{F}_j^n$ are ordered as

$$\begin{aligned}
&\rho_{\omega(nj+n)}, \quad \rho_{\omega(nj+n)+1}, \cdots \\
\cdots\ &\rho_{\omega(nj+n+1)}, \quad \rho_{\omega(nj+n+1)+1}, \cdots \\
\cdots\ &\rho_{\omega(nj+n+j)}, \quad \rho_{\omega(nj+n+j)+1}, \cdots < \rho_{\omega((n+1)j+(n+1))}.
\end{aligned}$$

It is clear that for every countable ordinal α, the collection $\mathcal{F}_\alpha = \{\rho_t : 0 \leqq t \leqq \alpha\}$ is a family and for a limit ordinal α, both $\mathcal{F}_\alpha$ and $\bar{\mathcal{F}}_\alpha = \mathcal{F}_\alpha - \{\rho_\alpha\}$ are families.

Next we construct three mutually disjoint sets A_α, B_α, C_α, whose union is $\mathcal{F}_\alpha$ for every $\alpha \geqq 0$. We start with

$$A_0 = \{\rho_0\}, \quad B_0 = \emptyset \quad \text{and} \quad C_0 = \emptyset.$$

Now suppose A_β, B_β and C_β are defined for all $\beta < \alpha$ for some countable ordinal α with $\mathcal{F}_\beta$ as the disjoint union of A_β, B_β and C_β. Then

$$C_\alpha = \Big(\bigcup_{\beta<\alpha} C_\beta \Big) \cup \{\rho\} \quad \text{and} \quad B_\alpha = \Big(\bigcup_{\beta<\alpha} B_\beta \Big) \cup \{\rho_\alpha\}$$

if $\rho_\alpha = e(\rho)$ for some $\rho \in A_\gamma$, $\gamma < \alpha$ and

$$C_\alpha = \Big(\bigcup_{\beta<\alpha} C_\beta \Big) \quad \text{and} \quad B_\alpha = \Big(\bigcup_{\beta<\alpha} B_\beta \Big)$$

if $\rho_\alpha \neq e(\rho)$ for any $\rho \in A_\gamma$, $\gamma < \alpha$. In either case we take

$$A_\alpha = \mathcal{F}_\alpha - (B_\alpha \cup C_\alpha).$$

That the construction is meaningful will be clear from the following Lemma.

LEMMA 3.1. *Corresponding to any ρ_α there exists at most one G-slice type $\rho \in A_\gamma$, $\gamma < \alpha$ with $e(\rho) = \rho_\alpha$.*

PROOF. For $\rho_\alpha = [H; V]$ to be the image of a G-slice type $\rho = [K; U] \in \in A_\gamma$, $\gamma < \alpha$, we must have $H = K \times \langle x_{i_1} \rangle$ and $V = V(K) \oplus P^*U$, where x_{i_1} is chosen minimally and $p: H \to K$ is the projection. For the sake of definiteness, we assume that the $\mathbb{Z}_2$ copies which can be factored out of H in different coordinates of T^k be $\langle x_{i_1} \rangle, \ldots, \langle x_{i_m} \rangle$ and $i_1 < i_2 < \ldots < i_m$. Therefore ρ_α can be equal to $e(\rho)$ for at most m different values of ρ. Denote these possible G-slice types by ρ_s, $1 \leqq s \leqq m$ and write $\rho_s = [K_s, U_s]$, where the least value of i for which $\langle x_i \rangle \not\subset p_i(K_s)$ is i_s. Clearly

$$H = K_s \times \langle x_{i_s} \rangle \quad \text{and} \quad V = V(K_s) \oplus (p_s)^* U_s,$$

where $p_s: H \to K_s$ is the projection. Note that V contains one and only one copy of each $V(K_s)$.

Now consider ρ_1 and ρ_s, $2 \leqq s \leqq m$. Writing $L = K_1 \cap K_s$, we get $K_1 = = L \times \langle x_{i_s} \rangle$, $K_s = L \times \langle x_{i_1} \rangle$ and $H = L \times \langle x_{i_1} \rangle \times \langle x_{i_s} \rangle$. Also

$$V = V(K_1) \oplus V(K_s) \oplus W_s$$

for some H-module W_s. The action of $\langle x_{i_1} \rangle \times \langle x_{i_s} \rangle \subset H$ on W_s is clearly trivial. In fact H acts on W_s via its projection on L. But W_s considered as an L-module cannot contain any trivial representation of L, as if it does then it is bound to contain a trivial representation of H.

Considering W_s as an L-module we look at the G-slice type $[L; W_s]$. Since L cannot be factored out by any $\langle x_i \rangle$ for $i < i_1$, $[L; W_s]$ cannot be extension of any G-slice type and

$$e[L; W_s] = [L \times \langle x_{i_1} \rangle; V(L) \oplus q^* W_s] = [K_s; U_s]$$

$q: L \times \langle x_{i_1} \rangle \to L$ being the projection. This shows that $[K_s; U_s] \notin A_\gamma$ for any $\gamma < \alpha$. □

§ 4. Vanishing of the group $N_*^G[\tilde{F}_j]$

Lemma 3.1 establishes a one to one correspondence between C_α and B_α for every α. The proof of this lemma together with the fact that $\tilde{F}_j = F_j \cup \cup e(F_j)$ reveals that if a G-slice type $\rho \in A_\alpha$ then it must have an extension. This shows that for every n, there exists a countable ordinal number β, sufficiently large, such that A_β has G-slice types of dimension higher than n. Also B_β and C_β are countable. We rewrite the elements of C_β and B_β in increasing order of G-slice types as

$$C_\beta = \{\rho_0, \rho_2, \cdots, \rho_\omega, \rho_{\omega+2}, \cdots, \rho_{\omega(2)}, \rho_{\omega(2)+2}, \cdots\},$$
$$B_\beta = \{\rho_1, \rho_3, \cdots, \rho_{\omega+1}, \rho_{\omega+3}, \cdots, \rho_{\omega(2)+1}, \rho_{\omega(2)+3}, \cdots\}.$$

THEOREM 4.1. *For $0 \leqq j < k$, the group $N_n^G[\tilde{F}_j] = 0$.*

PROOF. First we note that $\tilde{F}_j = \tilde{\mathcal{F}}_\beta = B_\beta \cup C_\beta$ for some countable ordinal β, sufficiently large. Therefore $N_n^G[\tilde{F}_j] = N_n^G[\tilde{\mathcal{F}}_\beta] = N_n^G[B_\beta \cup C_\beta]$. Further $\mathcal{F}_i = \{\rho_\alpha \in \tilde{F}_j : 0 \leqq \alpha \leqq i\}$ is a family, $\forall i \geqq 0$. Also $e : C_\beta \to B_\beta$ is a bijective map. Obviously,

$$N_*^G[\mathcal{F}_0] \cong N_*^G[\rho_0].$$

The long exact sequence

$$\cdots \longrightarrow N_*^G[\mathcal{F}_0] \to N_*^G[\mathcal{F}_1] \to N_*^G[\rho_1] \to N_{*-1}^G[\mathcal{F}_0] \longrightarrow \cdots$$

together with Lemma 2.1 shows that $N_*^G[\mathcal{F}_1] = 0$. Now, suppose that $N_*^G[\mathcal{F}_{2i}] \cong N_*^G[\rho_{2i}]$ for some $i < \beta$. The long exact sequence

$$\cdots \longrightarrow N_*^G[\mathcal{F}_{2i}] \to N_*^G[\mathcal{F}_{2i+1}] \to N_*^G[\rho_{2i+1}] \to N_{*-1}^G[\mathcal{F}_{2i}] \longrightarrow \cdots$$

shows that $N_*^G[\mathcal{F}_{2i+1}] = 0$. Therefore the long exact sequence

$$\cdots \longrightarrow N_*^G[\mathcal{F}_{2i+1}] \to N_*^G[\mathcal{F}_{2i+2}] \to N_*^G[\rho_{2i+2}] \to N_{*-1}^G[\mathcal{F}_{2i+1}] \longrightarrow \cdots$$

immediately gives that

$$N_*^G[\mathcal{F}_{2(i+1)}] \cong N_*^G[\rho_{2(i+1)}].$$

Thus, by transfinite induction, we have

$$N_*^G[\mathcal{F}_{2j}] = N_*^G[\rho_{2j}] \quad \text{and} \quad N_*^G[\mathcal{F}_{2j+1}] = 0,$$

$\forall j < \beta$. By taking the direct limit, we have

$$N_*^G[\tilde{\mathcal{F}}_\beta] = 0 = N_*^G[\tilde{F}_j]. \qquad \square$$

REFERENCES

[1] Dev, N. J. and Khare, S. S., Finite group action and vanishing of $N_*^G[\bar{F}]$, *Pacific J. Math.* **122** (1986), 57–71. *MR* **87b**:57023

[2] Dev, N. J. and Khare, S. S., Action of k-torus and vanishing of the group $N_*^G[\bar{F}]$, *Nat. Acad. Sci. Letters* **12** (1989), no. 11.

[3] Kosniowski, C., *Actions of finite abelian groups,* Research Notes in Mathematics, vol. 18, Pitman, Boston, Mass.–London, 1978. *MR* **80e**:57047

[4] Kosniowski, C., Some equivariant bordism theories vanish, *Math. Ann.* **242** (1979), 59–68. *MR* **80e**:57041

(Received July 8, 1988)

DEPARTMENT OF MATHEMATICS
LADY KEANE COLLEGE
SHILLONG
INDIA

DEPARTMENT OF MATHEMATICS
NORTH-EASTERN HILL UNIVERSITY
PERMANENT CAMPUS, MAWLAI
IND–793022 SHILLONG
INDIA

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 197–203

ÜBER EIN KREISÜBERDECKUNGSPROBLEM AUF DER SPHÄRE

G. BLIND und R. BLIND

1. Einleitung

Auf der Einheitssphäre S^2 seien $n \geqq 3$ kongruente, abgeschlossene, sphärische Kreise gegeben. Die 2 klassischen Problem dazu lauten:

1. Die Kreise bilden speziell eine Packung. Wie groß ist die maximale Dichte $\overline{D}_n$ einer Kreispackung aus n kongruenten Kreisen?

2. Die Kreise bilden speziell eine Überdeckung. Wie groß ist die minimale Dichte $\underline{D}_n$ einer Kreisüberdeckung aus n kongruenten Kreisen?

Es sei $\omega_n := \frac{n}{n-2}\,\frac{\pi}{6}$. Dann gelten die Abschätzungen

$$(1) \qquad \overline{D}_n \leqq \frac{n}{2}\left(1 - \frac{1}{2\cos\omega_n}\right) \quad \text{und} \quad \underline{D}_n \geqq \frac{n}{2}\left(1 - \frac{1}{\sqrt{3}\tan\omega_n}\right).$$

Diese Abschätzungen sind für $n = 3,\ 4,\ 6,\ 12$ scharf; die Kreismittelpunkte sind dann die Ecken eines regulären Dreiecks, Tetraeders, Oktaeders bzw. Ikosaeders. Für $n \to \infty$ gehen die Schranken in (1) gegen die optimalen Dichten der entsprechenden ebenen Probleme (siehe [4, S. 114]). Für andere Werte von n ist $\overline{D}_n$ bzw. $\underline{D}_n$ in einzelnen Fällen bekannt, in manchen Fällen gibt es Vermutungen oder gute Abschätzungen, siehe etwa [5] und [6].

Ein Kreissystem zerlegt die S^2 in mehrfach, einfach und überhaupt nicht überdeckte Bereiche. Beim Problem 1 wird der von den Kreisen einfach überdeckte Teil der S^2 abgeschätzt unter der Voraussetzung, daß es keinen mehrfach überdeckten Teil gibt. In [4, S. 97] wird nun folgendes (ebene) Problem gestellt: Der wievielte Teil der Ebene läßt sich durch *beliebig* gelegene kongruente Kreise einfach überdecken? Geht man von der dichtesten Packung kongruenter Kreise aus und vergrößert die Kreise konzentrisch, bis jeder Kreis von den 6 benachbarten in den Ecken eines regulären 12-Ecks geschnitten wird, so überdeckt das entstehende Kreissystem $100(\sqrt{48}-6)\%$ der Ebene einfach. Es gilt: $\vartheta := \sqrt{48} - 6$ ist die maximale Dichte des von

1991 *Mathematics Subject Classification.* Primary 52A45.

Key words and phrases. Arrangement of congruent circles on the sphere, simply covered area, density.

0081–6906/95/$ 4.00

einem beliebigen System kongruenter Kreise einfach überdeckten Bereichs des E^2. Dies wurde unter starken Voraussetzungen an das Kreissystem in [4] bzw. in [1] bewiesen (siehe auch [7]), und schließlich ohne jede Voraussetzung in [2]. Analog läßt sich auf der Sphäre fragen:

3. Gegeben seien $n \geqq 3$ kongruente sphärische Kreise $K_1, \ldots, K_n$ mit Radius ρ $(0 < \rho < \pi)$. $E(K_1, \ldots, K_n)$ sei der davon einfach überdeckte Bereich der S^2. Wie groß ist

$$\vartheta_n := \max \frac{|E(K_1, \ldots, K_n)|}{4\pi},$$

wobei sich das Maximum auf *alle* Familien aus n kongruenten Kreisen bezieht? Wir vermuten, daß gilt:

$$\vartheta_n \leqq \frac{3}{\pi}(n-2)\left(\omega_n - \pi + 2\arccos\frac{1}{4\cos\frac{\omega_n}{2}}\right) =: S_n \quad (n \geqq 3). \tag{2}$$

Diese Abschätzung ist für $n = 3, 4, 6, 12$ scharf; die Kreismittelpunkte sind dann die Ecken eines regulären Dreiecks, Tetraeders, Oktaeders bzw. Ikosaeders, und die Kreisradien sind (analog zum ebenen Fall) so groß, daß jeder Kreis die k benachbarten in den Ecken eines regulären $2k$-Ecks schneidet.

S_n ist wachsend in n weil

$$S_n = \frac{1}{\omega_n - \frac{\pi}{6}}\left(\omega_n - \pi + 2\arccos\frac{1}{4\cos\frac{\omega_n}{2}}\right)$$

in ω_n fallend ist, wie man durch Ableiten sieht. Es ist $S_3 = 0,809\,839\ldots$ und $S_4 = 0,881\,423\ldots$. Für $n \to \infty$ geht S_n gegen die maximale Dichte des ebenen Problems.

Nachdem schon das ebene Problem recht unzugänglich ist, wird hier zunächst gezeigt:

SATZ 1. *Die obere Vermutung ist richtig für $n = 3$. Es ist also $\vartheta_3 = S_3$; beim zugehörigen Kreissystem liegen die Kreismittelpunkte äquidistant auf einem Großkreis von S^2 und der Kreisradius ist $\rho_0 := \arccos\frac{1}{\sqrt{7}}$.*

SATZ 2. *Für ein Kreissystem $\{K_1, \ldots, K_n\}$ aus $n \geqq 4$ kongruenten Kreisen mit Radius ρ gilt: Ist $\rho \geqq \rho_0 = \arccos\frac{1}{\sqrt{7}}$ oder liegen alle Kreismittelpunkte in einer abgeschlossenen Halbsphäre, so ist*

$$\frac{|E(K_1, \ldots, K_n)|}{4\pi} < S_n. \tag{3}$$

Satz 2 besagt, daß zum Beweis der oberen Vermutung nur Kreissysteme "in allgemeiner Lage" und mit Kreisradius $\rho < \rho_0$ berücksichtigt werden müssen. Dies ist für weitere Untersuchungen wesentlich, siehe [3].

2. Vorbereitungen

∂M bezeichne den Rand einer Menge M. Zu einem Kreissystem $\{K_1, \ldots, K_n\}$ sei $T(K_1, \ldots, K_n)$ der mindestens zweifach überdeckte Bereich.

HILFSSATZ 1. *Für ein Kreissystem* $\{K_1, \ldots, K_n\}$ *aus* $n \geqq 3$ *kongruenten Kreisen mit Radius* $\rho \geqq \frac{\pi}{2}$ *gilt*

$$\frac{|E(K_1, \ldots, K_n)|}{4\pi} < S_n.$$

BEWEIS. Offensichtlich ist

$$|K_1| + \ldots + |K_n| \leqq |K_1 \cup \ldots \cup K_n| + (n-1)|T(K_1, \ldots, K_n)|.$$

Nach Voraussetzung ist $|K_1| + \ldots + |K_n| \geqq n \cdot 2\pi$, und weil die Kreise höchstens die ganze S^2 überdecken, ist $|K_1 \cup \ldots \cup K_n| \leqq 4\pi$. Deshalb ist

$$|T(K_1, \ldots, K_n)| \geqq \frac{1}{n-1}(n \cdot 2\pi - 4\pi) \geqq \pi.$$

Aus

$$|E(K_1, \ldots, K_n)| + |T(K_1, \ldots, K_n)| = |K_1 \cup \ldots \cup K_n| \leqq 4\pi$$

folgt damit

$$\frac{|E(K_1, \ldots, K_n)|}{4\pi} \leqq 0{,}75 < S_3 \leqq S_n.$$

HILFSSATZ 2. *Seien* K_1, K_2 *kongruente Kreise mit Radius* $\rho < \frac{\pi}{2}$. *Dann ist mit den Bezeichnungen von Fig. 1*

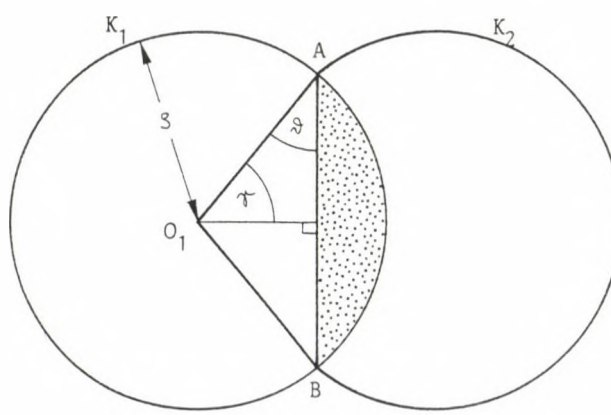


Fig. 1

$$|T(K_1, K_2)| = 2(\pi - 2\gamma \cos\rho - 2\vartheta). \tag{4}$$

BEWEIS. Der Kreisausschnitt mit Winkel 2γ hat den Flächeninhalt $2\gamma(1-\cos\rho)$, das gleichschenklige Dreieck O_1AB hat den Flächeninhalt $2\gamma+2\vartheta-\pi$. Daraus folgt (4).

HILFSSATZ 3. *Seien K_1, K_2, K_3 abgeschlossene, kongruente Kreise mit Radius $\rho<\frac{\pi}{2}$. Sei $K_1\cap K_2\cap K_3\neq\emptyset$. In einer Lage mit minimalem $|T(K_1,K_2,K_3)|$ enthält dann $K_1\cap K_2\cap K_3$ genau einen Punkt p, und die Kreise liegen symmetrisch bzgl. p. In dieser Lage ist*

$$|T(K_1,K_2,K_3)|=2\pi-12\cos\rho\arctan\frac{1}{\sqrt{3}\cos\rho}. \tag{5}$$

BEWEIS. Als erstes zeigen wir, daß die Behauptung über die Lage der Kreise richtig ist, wenn $K_1\cap K_2\cap K_3$ nur einen Punkt p enthält. Denn sonst wäre z.B. $|K_1\cap K_2|>|K_1\cap K_3|$. Dann ist auch $|\partial K_1\cap K_2|>|\partial K_1\cap K_3|$. Bei festem K_2 und K_3 drehe man nun K_1 um p so, daß $|K_1\cap K_2|$ verkleinert wird. Bei einer genügend kleinen Drehung wird $|K_1\cap K_2|+|K_1\cap K_3|$ kleiner, so daß auch $|T(K_1,K_2,K_3)|$ kleiner wird im Widerspruch dazu, daß es minimal ist.

Enthält $K_1\cap K_2\cap K_3$ nicht nur einen Punkt, so sind 2 Fälle möglich.

(i) $K_1\cap K_2\cap K_3$ ist ein Kreisbogenzweieck, d.h. z.B. $K_3\supset K_1\cap K_2$. Dann ist $|T(K_1,K_2,K_3)|$ höchstens dann minimal, wenn K_3 bzgl. des Großkreises durch die Ecken von $K_1\cap K_2$ symmetrisch liegt, und wenn eine solche Ecke p auf ∂K_3 liegt. Drehung von K_1 bzw. K_2 um p zeigt danach die Behauptung.

(ii) $K_1\cap K_2\cap K_3$ ist ein Kreisbogendreieck. Dann ist $|T(K_1,K_2,K_3)|$ höchstens dann minimal, wenn K_3 bzgl. des Großkreises G durch die Ecken von $K_1\cap K_2$ symmetrisch liegt. Dann sei $b_1:=\partial K_3\cap(K_1\cap K_2)$, $b_2:=\partial K_3\setminus(K_1\cup K_2)$ und B sei der zu G symmetrische Halbkreis von ∂K_3 durch b_1. Es ist $|\partial K_3\cap K_1|+|b_2|=|\partial K_3\setminus K_1|+|b_1|$, und aus $|\partial K_3\setminus K_1|>|\partial K_3\cap K_1|$ folgt $|b_2|>|b_1|$. Verschiebt man nun K_3 längs G so, daß $|K_1\cap K_2\cap K_3|$ abnimmt, so wird $|T(K_1,K_2,K_3)|$ kleiner: Im Fall $|b_2|>\frac{1}{2}|\partial K_3|$ ist dies klar, und sonst gilt $|B\setminus b_1|+|b_2|>\frac{1}{2}|\partial K_3|$. Bei minimalem $|T(K_1,K_2,K_3)|$ enthält also auch hier $K_1\cap K_2\cap K_3$ genau einen Punkt.

Das minimale $T(K_1,K_2,K_3)$ berechnet man nach (4).

3. Beweis von Satz 1

Es seien also K_1,K_2,K_3 kongruente Kreise mit Radius ρ. Wegen Hilfssatz 1 kann $\rho<\frac{\pi}{2}$ angenommen werden.

Es ist $|E(K_1,K_2,K_3)|=|K_1|+|K_2|+|K_3|-2|T(K_1,K_2,K_3)|-|K_1\cap K_2\cap K_3|$. Wegen Hilfssatz 3 ist also $|E(K_1,K_2,K_3)|$ maximal für minimales $|T(K_1,K_2,K_3)|$, und dann enthält $K_1\cap K_2\cap K_3$ höchstens einen Punkt. $|T(K_1,K_2,K_3)|$ ist also minimal, wenn die Mittelpunkte von

K_1, K_2, K_3 äquidistant auf einem Großkreis liegen. Variiert man nun ρ, so ist $|E(K_1, K_2, K_3)|$ maximal, wenn jeder Kreis die beiden anderen in den Ecken eines regulären 4-Ecks schneidet, was $\rho = \arccos \frac{1}{\sqrt{7}}$ ergibt. Jetzt ist $|T(K_1, K_2, K_3)| = 3|T(K_1, K_2)| = 6\left(\pi - \frac{\pi}{2} \frac{1}{\sqrt{7}} - 2 \arccos \frac{\sqrt{2}}{4}\right)$ nach (4). Daraus folgt $\vartheta_3 = S_3$.

4. Beweis von Satz 2

Es seien also $n \geqq 4$ kongruente Kreise $K_1, \ldots, K_n$ mit Radius ρ gegeben. Wegen Hilfssatz 1 kann $\rho < \frac{\pi}{2}$ angenommen werden. Es ist $S_4 = 0,881\,423 \ldots \leqq S_n$ $(n \geqq 4)$, und

$$|E(K_1, \ldots, K_n)| = 4\pi - |T(K_1, \ldots, K_n)| - |S^2 \setminus (K_1 \cup \ldots \cup K_n)|.$$

Wegen $T(K_1, \ldots, K_n) \supset T(K_1, \ldots, K_4)$ ist also (3) richtig, wenn

$$|T(K_1, \ldots, K_4)| + |S^2 \setminus (K_1 \cup \ldots \cup K_n)| > 1,4901.$$

Dies wird als erstes für $\rho \geqq \rho_0$ gezeigt. $|T(K_1, \ldots, K_4)|$ wächst in ρ bei fester Lage der Kreismittelpunkte, so daß $|T(K_1, \ldots, K_4)|$ für $\rho = \rho_0$ abgeschätzt wird. 2 Fälle werden unterschieden:

(i) $T(K_1, \ldots, K_4)$ besteht nur aus doppelt überdeckten Bereichen. Zu je 3 verschiedenen Kreisen $K_i, K_j, K_k \in \{K_1, \ldots, K_4\}$ ist dann

$$|T(K_i, K_j, K_k)|$$

minimal, wenn die Kreismittelpunkte äquidistant auf einem Großkreis liegen. Zusammen mit (4) und wegen $\rho = \rho_0$ gilt also

$$|T(K_i, K_j, K_k)| > 0,7741 \ldots .$$

Die Bezeichnung sei so gewählt, daß $|T(K_1, K_4)| \geqq |T(K_2, K_3)|$. Dann ist

$$\begin{aligned} |T(K_1, \ldots, K_4)| = \\ = |T(K_1, K_2)| + |T(K_1, K_3)| + |T(K_1, K_4)| + \\ + |T(K_2, K_3)| + |T(K_2, K_4)| + |T(K_3, K_4)| \geqq \\ \geqq |T(K_1, K_2, K_3)| + |T(K_2, K_3, K_4)| \geqq 1.5482 \ldots > 1.4901. \end{aligned}$$

(ii) O.B.d.A. ist $K_1 \cap K_2 \cap K_3 \neq \emptyset$. Dann ist

$$|T(K_1, \ldots, K_4)| \geqq |T(K_1, \ldots, K_3)| \geqq 1,7877 \ldots$$

nach (5) mit $\rho = \rho_0$.

Als zweites wird der Fall betrachtet, daß alle Kreismittelpunkte in einer abgeschlossenen Halbsphäre H liegen. Dann enthält $S^2 \setminus (K_1 \cup \ldots \cup K_n)$

einen Kreis K um den Mittelpunkt der zu H komplementären Halbsphäre. Es ist

$$|S^2 \setminus (K_1 \cup \ldots \cup K_n)| \geqq |K| \geqq 2\pi \left(1 - \cos\left(\frac{\pi}{2} - \rho\right)\right) = 2\pi(1 - \sin\rho),$$

was in ρ monoton fällt. Zur Abschätzung von $|T(K_1, \ldots, K_4)|$ unterscheidet man wieder 2 Fälle:

(i) $T(K_1, \ldots, K_4)$ besteht nur aus doppelt überdeckten Bereichen. Es wird gezeigt, daß dann $\frac{1}{2}|T(K_1, \ldots, K_4)|$ abgeschätzt werden kann durch den Flächeninhalt des von 2 Kreisen mit Mittelpunktabstand $\frac{\pi}{2}$ doppelt überdeckten Bereichs, d.h.

$$|T(K_1, \ldots, K_4)| \geqq 4\left(\pi - 2\cos\rho \arccos\frac{1}{\tan\rho} - 2\arcsin\frac{1}{\sqrt{2}\sin\rho}\right),$$

was in ρ monoton wächst für $\rho \geqq \frac{\pi}{4}$.

Zum Beweis kann o.B.d.A. angenommen werden, daß mindestens 3 Kreismittelpunkte O_1, O_2, O_3 auf dem Rand ∂H liegen. Die Behauptung ist klar, wenn auch $O_4 \in \partial H$ gilt, oder wenn O_1, O_2, O_3 schon in einer Hälfte von ∂H liegen. Ist O der Mittelpunkt von H, so kann also angenommen werden, daß die Dreiecke O_1O_2O, O_2O_3O und O_3O_1O eine Zerlegung von H bilden. Deshalb gilt z.B. $O_4 \in O_1O_2O$. Dann ist $|\overline{O_1O_4}| + |\overline{O_2O_4}| \leqq |\overline{O_1O}| + |\overline{O_2O}|$, so daß $|T(K_1, K_4)| + |T(K_2, K_4)|$ für alle Lagen von O_4 in O_1O_2O minimal ist für $O_4 = O$. Aus $|T(K_1, \ldots, K_4)| \geqq |T(K_1, K_4)| + |T(K_2, K_4)|$ folgt die Behauptung.

(ii) O.B.d.A. ist $K_1 \cap K_2 \cap K_3 \neq \emptyset$. Nach (5) ist dann

$$|T(K_1, \ldots, K_4)| \geqq 2\pi - 12\cos\rho \arctan\frac{1}{\sqrt{3}\cos\rho},$$

was in ρ monoton wächst.

In beiden Fällen wird $|T(K_1, \ldots, K_4)| + |S^2 \setminus (K_1 \cup \ldots \cup K_n)|$ durch die Summe zweier Funktionen von ρ abgeschätzt. Unter Ausnutzung der Monotonie der Summanden rechnet man in beiden Fällen nach, daß für $\rho \leqq \rho_0$ die Summe einen Wert $> 1,4901$ hat.

LITERATURVERZEICHNIS

[1] Balázs, J., Über ein Kreisüberdeckungsproblem, *Acta Math. Acad. Sci. Hungar.* **24** (1973), 377–382. *MR* **48** # 9561

[2] Blind, G. und Blind, R., Ein Kreisüberdeckungsproblem, *Studia Sci. Math. Hungar.* **21** (1986), 35–57. *MR* **88m**:52023

[3] Blind, G. und Blind, R., Ein Kreisüberdeckungsproblem auf der Sphäre, *Studia Sci. Math. Hungar.* **29** (1994), 107–164.

[4] Fejes Tóth, L., *Lagerungen in der Ebene, auf der Kugel und im Raum*, 2. Auflage, Die Grundlehren der mathematischen Wissenschaften, Band 65, Springer-Verlag, Berlin–New York, 1972. *MR* **50** # 5603

[5] GÁSPÁR, Zs. und TARNAI, T., Multisymmetric close packings of equal spheres on the spherical surface, *Acta Cryst. Sect. A* **43** (1987), 612–616. *MR* **89d**:52031

[6] GÁSPÁR, Zs. und TARNAI, T., Covering the sphere with 11 equal circles, *Elem. Math.* **41** (1986), 35–38. *MR* **88c**:52014

[7] MAKAI, E., Research problem 20, *Period. Math. Hungar.* **7** (1976), 319–320.

(Eingegangen am 15. November 1990)

MATHEMATISCHES INSTITUT B
UNIVERSITÄT STUTTGART
PFAFFENWALDRING 57
D–70550 STUTTGART
FEDERAL REPUBLIC OF GERMANY

WALDBURGSTRASSE 88
D–70563 STUTTGART
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 205–216

ON THE A POSTERIORI ERROR ESTIMATES FOR STIRLING'S METHOD

I. K. ARGYROS

Abstract

New a posteriori error estimates for Stirling's method are given under natural assumptions. It is shown that they are better from the viewpoint of the accuracy and the cost of the information used than the ones already in the literature under similar assumptions.

1. Introduction

A fixed point x^* of an operator F defined on a subset D_F of a Banach space E and taking values into itself satisfies the equation

$$x = F(x). \tag{1}$$

We want to construct a sequence $\{x_n\}_{n\geq 0} \subset D_F$ converging to x^* for a suitable starting value x_0. To achieve this construction we attach to the pair (F, x_0) an operator $P\colon D_P \subset E \to E$ and consider the iteration

$$x_{n+1} = P(x_n), \qquad n = 0, 1, 2, \ldots, \tag{2}$$

where P is given by

$$P(x) = x - (I - F'(F(x)))^{-1}(x - F(x)). \tag{3}$$

This particular choice of P in (2) defines the so-called Stirling's method [10], [11]. Stirling's method can be viewed as a combination of the method of successive substitutions and Newton's method.

We derive new a posteriori error estimates for Stirling's method. We also show that they are better from the viewpoint of the accuracy and the cost of the information used than the ones already in the literature under similar assumptions [10], [11]. Moreover, we provide an example where we show that for the same starting point x_0 Newton's method fails to converge but Stirling's method converges to the fixed point x^* of F. Finally, we provide a simple example of a two point boundary value problem where our results compare favourably with the ones obtained before under similar assumptions [10], [11]. We are interested in the case where $x_n \in D_P$, $n = 0, 1, 2, \ldots$.

1991 *Mathematics Subject Classification.* Primary 65J10, 65J15; Secondary 47D15.
Key words and phrases. Banach space, Stirling's method, fixed point.

0081–6906/95/$ 4.00

From now on for simplicity we will assume $D_F = E$.

Let α, γ, r_0 be real numbers such that $0 < \alpha < 1$, $\gamma > 0$, and $r_0 \geqq 0$. Let us denote by $C(\alpha, \gamma, r_0)$ the class of all pairs (F, x_0) satisfying the conditions:

(a_1) F is an operator defined on E and with values into itself and x_0 is a point of E.

(a_2) The Fréchet-derivative F' of F is uniformly bounded on E. That is, there exists α with $0 < \alpha < 1$ such that

$$\|F'(x)\| \leqq \alpha \quad \text{for all} \quad x \in E. \tag{4}$$

(a_3) The Fréchet-derivative F' of F is Lipschitz continuous on E. That is, there exists $\gamma > 0$ such that

$$\|F'(x) - F'(y)\| \leqq \gamma \|x - y\| \quad \text{for all} \quad x, y \in E. \tag{5}$$

(a_4) The following inequality is true:

$$\frac{\|x_0 - F(x_0)\|}{1 - \alpha} \leqq r_0 < \min(q_0^{-1}, q^{-1}) \equiv \bar{q} \tag{6}$$

where we denote

$$q_0 = \frac{\gamma(1 + 2\alpha)}{2(1 - \alpha)}$$

and

$$q = \frac{\gamma(3 - 2\alpha)}{2(1 - \alpha)}.$$

We now define a convergent iterative procedure for the class $C(\alpha, \gamma, r_0)$ by associating with each pair $(F, x_0) \in C(\alpha, \gamma, r_0)$ the iterative algorithm (P, x_0) with P given by (3). The iteration (2) becomes

$$x_{n+1} = x_n - (I - F'(F(x_n)))^{-1}(x_n - F(x_n)), \qquad n = 0, 1, 2, \ldots . \tag{7}$$

2. Error bounds for Stirling's method

For the proof of the main theorem, we will use the method of nondiscrete mathematical induction [6], [8], [9]. Let T be either the positive real axis or an interval of the form $T = \{r \in \mathbb{R};\ 0 < r \leqq r_0\}$.

DEFINITION 5. A function $w: T \to T$ is called a rate of convergence on T (see, e.g. [6], p. 65) if the series

$$\sigma(r) = \sum_{n=0}^{\infty} w^{(n)}(r) \tag{8}$$

is convergent for each $r \in T$, where the iterates $w^{(n)}$ of w are defined as follows:

$$w^{(0)}(r) = r, \quad w^{(n+1)}(r) = w(w^{(n)}(r)), \quad n = 0, 1, 2, \ldots .$$

The functions w and σ from Definition 5 obviously satisfy the following functional equation:

$$\sigma(w(r)) = \sigma(r) - r, \qquad r \in T. \tag{9}$$

Using Definition 5 and (8) it is easy to check that the function

$$w(r) = qr^2 \tag{10}$$

is a rate of convergence on $T = \{r \in \mathbb{R};\ 0 < r \leqq r_0\}$ and the corresponding series σ is given by

$$\sigma(r) = \frac{1}{q} \sum_{k=0}^{\infty} (qr)^{2^k} \quad \text{for all } 0 < r \leqq r_0. \tag{11}$$

It can easily be seen that σ converges if $0 < qr < 1$. Moreover, the iterates $w^{(n)}$ of w are given by

$$w^{(n)}(r) = \frac{1}{q}(qr)^{2^n}, \quad n = 0, 1, 2, \ldots \tag{12}$$

in this case. Denote by $\bar{U}(x, r)$ the closed ball centered at x and of radius $r > 0$.

We will need the following corollary of the induction theorem [see e.g. [6], [7], [8]].

THEOREM 1. (a) *Suppose that we can attach to the pair (F, x_0) a rate of convergence w on an interval T and a family of sets $Q(r) \subset E$, $r \in T$ such that the conditions*

$$x_0 \in Q(r_0) \quad \textit{for a certain } r_0 \in T \quad \textit{and} \quad (r \in T \textit{ and } x \in Q(r)) \tag{13}$$

$$\Rightarrow P(x) \in \bar{U}(x, r) \cap Q(w(r)). \tag{14}$$

Then the iteration generated by (7) converges to a fixed point x^* of equation (1), in such a way that the following estimates are satisfied:

$$x_n \in Q(w^{(n)}(r_0)), \tag{15}$$

$$\|x_n - x_{n-1}\| \leqq w^{(n-1)}(r_0), \tag{16}$$

and

$$\|x_n - x^*\| \leqq \sigma(w^{(n)}(r_0)), \qquad n = 0, 1, 2, \ldots . \tag{17}$$

(b) *Suppose that for a certain* $n \in \{1, 2, \ldots\}$ *the condition*

$$x_{n-1} \in Q(\|x_n - x_{n-1}\|) \tag{18}$$

is satisfied then for this n,

$$\|x_n - x^*\| \leqq d(\|x_n - x_{n-1}\|) \tag{19}$$

where we have denoted

$$d(r) = \sigma(r) - r. \tag{20}$$

(c) *Moreover, suppose that for a certain* N, *equality is attained in* (17), *then equality will be attained in* (16) *and* (17) *for all* $n \geqq N$.

We will need the following theorems on the convergence of Stirling's method ([10], Theorem 4 and Theorem 4', respectively).

THEOREM 2. *If* F' *is Lipschitz continuous with constant* γ *and* $\|F'(x)\| \leqq \alpha < 1$ *for all* $x \in E$, *then iteration* (7) *converges to* x^* *starting from any* $x_0 \in E$ *such that*

$$h_s = q_0 \frac{\|x_0 - F(x_0)\|}{1 - \alpha} < 1. \tag{21}$$

Moreover, the convergence is quadratic, with

$$\|x_n - x^*\| \leqq (h_s)^{2^n - 1} \frac{\|x_0 - F(x_0)\|}{1 - \alpha}, \qquad n = 0, 1, 2, \ldots . \tag{22}$$

Theorem 2 in bounded regions can be stated as follows.

THEOREM 3. *Suppose that* F' *is Lipschitz continuous with constant* γ *and uniformly bounded by a non-negative constant* $\alpha < 1$ *in the ball*

$$\bar{U}(x_0, r_0^*) = \left\{ x \in E / \|x - x_0\| \leqq r_0^* \equiv \frac{2\|x_0 - F(x_0)\|}{1 - \alpha} \right\}.$$

If (21) *holds, then Stirling's method converges to the unique fixed point* x^* *of* F *in* $\bar{U}$ *at the rate given by* (22).

We will now prove a theorem, concerning the convergence of Stirling's method in the class $C(\alpha, \gamma, r_0)$.

THEOREM 4. *If* $(F, x_0) \in C(\alpha, \gamma, r_0)$, *then the iterative algorithm* (7) *is well-defined, the sequence* $\{x_n\}_{n\geq 0}$ *produced by it converges to a fixed point* x^* *of equation* (1) *and the following estimates are true:*

$$\|x_n - x_{n-1}\| \leqq w^{(n-1)}(r_0), \tag{23}$$

$$\|x_{n-1} - x_0\| \leqq \sigma(r_0) - \sigma(w^{(n-1)}(r_0)) \leqq \sigma(r_0) - \sigma(\|x_n - x_{n-1}\|) \tag{24}$$

$$\|x_n - x^*\| \leqq \sigma(w^{(n)}(r_0)) \tag{25}$$

and

$$\|x_n - x^*\| \leqq d(\|x_n - x_{n-1}\|) \tag{26}$$

where w, σ, d *are given by* (10), (11) *and* (20), *respectively.*

PROOF. The proof uses Theorems 1 and 2. Since $0 < \alpha < 1$ the linear operator $I - F'(F(x))$ is invertible for all $x \in E$. Therefore, the sequence generated by (7) is well-defined for all $n = 0, 1, 2, \ldots$. We can now attach to iteration (7) the rate of convergence w given by (10) and the family of sets

$$Q(r) = \{x \in E; \|x - x_0\| \leqq \sigma(r_0) - \sigma(r), \|x - F(x)\| \leqq (1 - \alpha)r\} \tag{27}$$

where σ is given by (11). The hypotheses of the theorem imply that $Q(r_0) = \{x_0\}$, so that (15) is satisfied. Now let x be an element of $Q(r)$ and denote y by

$$y = x - (I - F'(F(x)))^{-1}(x - F(x)). \tag{28}$$

We want to show that $y \in Q(w(r))$. Using (27) and (28) we get

$$\|y - x_0\| \leqq \|y - x\| + \|x - x_0\| \leqq r + \sigma(r_0) - \sigma(r) = \sigma(r_0) - \sigma(w(r)). \tag{29}$$

Using the identities

$$y - F(y) = F(x) - F(y) - F'(F(x))(x - y), \tag{30}$$

$$\begin{aligned} &F(x) - F(y) - F'(F(x))(x - y) = \\ &= \int_0^1 [F'(\theta x + (1 - \theta)y) - F'(F(x))](x - y)d\theta \end{aligned} \tag{31}$$

we obtain

$$\begin{aligned} &\|(I - F'(F(y)))^{-1}(y - F(y))\| \leqq \\ &\leqq \frac{\gamma}{2(1 - \alpha)}(\|x - F(x)\| + \|y - F(x)\|)\|x - y\| \leqq \\ &\leqq \frac{\gamma}{2(1 - \alpha)}[(1 - \alpha)r + r + (1 - \alpha)r]r \leqq \\ &\leqq qr^2 = w(r). \end{aligned} \tag{32}$$

Thus, we have proved that condition (14) is also satisfied. Hence, the first part of Theorem 1 assures the fact that the sequence generated by (7) converges to a point x^* and that the relations (25) and (36) are satisfied. By continuity, iteration (7) gives $x^* = F(x^*)$. From the fact $x_{n-1} \in Q(w^{(n-1)}(r_0))$ and from the monotonicity of σ we get

$$\|x_n - x_{n-1}\| = \|(I - F'(F(x_{n-1})))^{-1}(x_{n-1} - F(x_{n-1}))\| \leqq w^{(n-1)}(r_0)$$

and

$$\|x_{n-1} - x_0\| \leqq \sigma(r_0) - \sigma(w^{(n-1)}(r_0)) \leqq \sigma(r_0) - \sigma(\|x_n - x_{n-1}\|).$$

Thus, the relation (15) is also verified for $n = 1, 2, \ldots$. The rest of the theorem follows from Theorem 1 immediately and that completes the proof.

REMARK. (a) It can easily be seen that our estimates given by (25) are eventually sharper than the ones given by (22) if

$$qr_0 < q_0 r_0$$

that is if

$$\frac{1}{2} < \alpha < 1.$$

(b) Estimates of the form (23), (24) or (26) are not given in [10].

Moreover, we can show

PROPOSITION 1. *Under the hypotheses of Theorem 4 the following are true:*

$$\|x_n - x^*\| \leqq \frac{1}{q} \sum_{k=0}^{\infty} (q\|x_n - x_{n-1}\|)^{2^{k+n}} \leqq \beta_1(n, r_0), \quad n = 1, 2, 3, \ldots \tag{33}$$

and

$$\|x_n - x^*\| \leqq \frac{1}{q} \sum_{k=0}^{\infty} \left(\frac{q}{1-\alpha}\|x_n - F(x_{n-1})\|\right)^{2^{k+n}}, \quad n = 1, 2, 3, \ldots \tag{34}$$

where

$$\beta_1(n, r) = \frac{1}{q} \sum_{k=0}^{\infty} (qw^{(n-1)}(r))^{2^{k+n}}.$$

PROOF. From $0 < \alpha < 1$ it follows that $I - F'(F(x_n))$ is invertible for all n and

$$\|(I - F'(F(x_n)))^{-1}\| \leqq \frac{1}{1-\alpha}, \qquad n = 0, 1, 2, \ldots. \tag{35}$$

Consider a pair $(F, x_0) \in C(\alpha, \gamma, r_0)$ and denote

$$r_{n-1} = \|(I - F'(F(x_{n-1})))^{-1}(x_{n-1} - F(x_{n-1}))\|. \tag{36}$$

We will show that $(F, x_{n-1}) \in C(\alpha, \gamma, r_{n-1})$. By (6) it suffices to show that

$$\frac{r_{n-1}}{\bar{q}} < 1.$$

But by (7), (23) and (36) we have

$$r_{n-1} = \|x_n - x_{n-1}\| \leqq w^{(n-1)}(r_0) \leqq \frac{1}{q}(qr_0)^{2^{n-1}} \leqq \frac{1}{q}(qr_0) = r_0 < \bar{q}.$$

Applying Theorem 3 to the pair $(F, x_{n-1}) \in C(\alpha, \gamma, r_{n-1})$ we deduce (33). Finally, noting that

$$\|x_n - x_{n-1}\| = \|(I - F'(F(x_{n-1})))^{-1}(x_{n-1} - F(x_{n-1}))\| \leqq$$
$$\leqq \frac{1}{1-\alpha}\|x_{n-1} - F(x_{n-1})\|$$

and using (33) we obtain (34). That completes the proof of the proposition.

We can now prove the following:

PROPOSITION 2. (a) *Under the hypotheses of Theorem 4 the following estimates are true:*

$$\|x_n - x^*\| \geqq$$
$$\geqq 2\left[1 + \left(1 + \frac{2(3+2\alpha)}{(1-\alpha)}\|x_n - x_{n+1}\|\right)^{1/2}\right]^{-1}\|x_n - x_{n+1}\| \tag{37}$$

and

$$\|x_n - x^*\| \geqq \left[1 + \frac{3+2\alpha}{2(1-\alpha)}\sigma(w^{(n)}(r_{n-1}))\right]^{-1}\|x_n - x_{n+1}\|, \quad n = 0, 1, 2, \ldots. \tag{38}$$

(b) *Suppose that in addition to the hypotheses of Theorem 4, the condition*

$$r_0 < \min(\bar{q}, \bar{r}_0) \tag{39}$$

is true, where $\bar{r}_0$ is the maximum positive number such that

$$(3+2\alpha)\gamma q^2 r^3 + 2(1-\alpha)qr - 2(1-\alpha) < 0 \quad \textit{for all } r \in [0, r_0].$$

Then the following estimate holds

$$\|x_n - x^*\| \leqq 2[2(1-\alpha) - (3+2\alpha)\gamma\sigma(w^{(n)}(r_{n-1}))]^{-1}\|x_n - F(x_n)\| \leqq$$
$$\leqq \beta_2(n, r_0, \|x_n - F(x_n)\|), \qquad n = 0, 1, 2, \ldots, \tag{40}$$

where

$$\beta_2(n,r,t)=2t[2(1-\alpha)-(3+2\alpha)\gamma\sigma(w^{(n(n-1))}(r))]^{-1}.$$

Furthermore, if

$$r_0<\min\left(\bar{r}_0,\frac{1-\alpha}{(3+2\alpha)\gamma}\right),$$

then

(41)

$$\|x_n-x^*\|\leqq\{(1-\alpha)-[(1-\alpha)^2-(3+2\alpha)\gamma\|x_n-F(x_n)\|]^{1/2}\}((3+2\alpha)\gamma)^{-1},$$
$$n=0,1,2,\ldots.$$

PROOF. (a) By taking norms in both sides of the identity

$$x_{n+1}-x_n=x^*-x_n+(I-F'(F(x_n)))^{-1}[F(x^*)-F(x_n)-F'(x_n))(x^*-x_n)],$$

we obtain

(42)
$$\|x_{n+1}-x_n\|\leqq$$

$$\leqq\|x_n-x^*\|+\frac{1}{1-\alpha}\left\|\int_0^1(F'(\theta x_n+(1-\theta)x^*)-F'(F(x_n)))(x^*-x_n)d\theta\right\|\leqq$$

$$\leqq\|x_n-x^*\|+\frac{1}{1-\alpha}\int_0^1\|(\theta x_n+(1-\theta)x^*-F(x_n))\|\|x_n-x^*\|d\theta\leqq$$

$$\leqq\|x_n-x^*\|+\frac{1}{1-\alpha}\int_0^1\|(1-\theta)(x^*-x_n)+(x_n-F(x_n))\|\|x_n-x^*\|d\theta\leqq$$

$$\leqq\|x_n-x^*\|+\frac{1}{1-\alpha}\left(\frac{3}{2}+\alpha\right)\|x_n-x^*\|^2.$$

The result (37) follows now immediately from (42). Furthermore, by (42) and (25) we get

$$\|x_{n+1}-x_n\|\leqq\|x_n-x^*\|+\frac{3+2\alpha}{2(1-\alpha)}\sigma(w^{(n)}(r_{n-1}))\|x_n-x^*\|$$

from which the estimate (38) follows.

(b) The linear operator defined by

$$L=\int_0^1F'(\theta x_n+(1-\theta)x^*)d\theta$$

is such that the operator $I-L$ is invertible if (39) holds. Indeed, from the identity

$$I-(I-F'(F(x_n)))^{-1}(I-L)=(I-F'(F(x_n)))^{-1}(L-F'(F(x_n))),$$

(35) and (5) it follows that

$$\|I-(I-F'(F(x_n)))^{-1}(I-L)\| \leqq \frac{\gamma}{1-\alpha}\int_0^1 \|\theta x_n+(1-\theta)x^*-F(x_n)\|d\theta \leqq$$

$$\leqq \frac{(3+2\alpha)\gamma}{2(1-\alpha)}\|x_n-x^*\| \leqq \frac{(3+2\alpha)\gamma}{2(1-\alpha)}\sigma(w^{(n)}(r_0)) \leqq$$

$$\leqq \frac{(3+2\alpha)\gamma}{2(1-\alpha)q}(qr_0)^{2^n}\sum_{k=0}^{\infty}(qr_0)^{2^k} \leqq$$

$$\leqq \frac{(3+2\alpha)\gamma}{2(1-\alpha)}q^2r_0^3\sum_{k=0}^{\infty}\frac{1-(qr_0)^k}{1-qr_0} \leqq \frac{(3+2\alpha)\gamma q^2r_0^3}{2(1-\alpha)(1-qr_0)}<1,$$

by the choice of r_0.

According to Banach's lemma it follows that the linear operator $I-L$ is invertible and

$$\|[(I-F'(F(x_n)))^{-1}(I-L)]^{-1}\| \leqq \left[1-\frac{(3+2\alpha)\gamma}{2(1-\alpha)}\|x_n-x^*\|\right]^{-1}. \tag{43}$$

Finally, using the identity

$$x_n-x^*=((I-F'(F(x_n)))^{-1}(I-L))^{-1}(I-F'(F(x_n)))^{-1}(x_n-F(x_n)),$$

(25), (35) and (43) we obtain (40) and (41) under the corresponding hypotheses. That completes the proof of the proposition.

We now complete this paper with two applications.

3. Applications

We provide an example where Newton's method fails to converge where Stirling's method converges to the fixed point x^* of a certain operator F.

EXAMPLE 1. Let us consider the real function F given by

$$F(x)=\begin{cases}\frac{1}{5}x & x\leqq 2\\ \frac{1}{10}(13x^2-29x+10) & 2\leqq x\leqq 3\\ x+1 & x\geqq 3.\end{cases}$$

Using Stirling's method given by (7) with $x_0 = 2$ we obtain

$$x_1 = 0 = x^*.$$

But, using Newton's method with $x_0 = 2$, the method fails to converge since $F'(2)$ does not exist.

We now provide a simple example of a two-boundary value problem with a trivial solution to compare our error estimates with the ones obtained in [10].

EXAMPLE 2. Consider the differential equation

$$\frac{1}{4}y'' + y^2 = y, \qquad y(0) = y(1) = 0. \tag{44}$$

We divide the interval $[0, 1]$ into n subintervals and we set $h = \frac{1}{n}$. Let $\{t_k\}$ be the points of subdivision with

$$0 = t_0 < t_1 < \cdots < t_n = 1.$$

A standard approximation for the second derivative is given by

$$y_i'' = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}, \quad y_i = y(t_i), \quad i = 1, 2, \ldots, n-1.$$

Take $y_0 = y_n = 0$ and define the operator $F: \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ by

$$F(y) = \frac{1}{4}H(y) + h^2\varphi(y),$$

$$H = \begin{bmatrix} -2 & 1 & & & \\ & \ddots & \ddots & 0 & \\ 1 & -2 & \ddots & \ddots & \\ 0 & \ddots & \ddots & \ddots & 1 \\ & & & 1 & -2 \end{bmatrix},$$

$$\varphi(x) = [y_1^2, y_2^2, \ldots, y_{n-1}^2]^{\mathrm{tr}} \quad \text{and} \quad y = [y_1, y_2, \ldots, y_{n-1}]^{\mathrm{tr}}.$$

Then

$$F'(x) = \frac{1}{4}H + 2h^2 \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_{n-1} \end{bmatrix}.$$

The solution of (44) can now be obtained as the fixed point of the equation

$$F(y) = y. \tag{45}$$

Let $y \in \mathbb{R}^{n-1}$, $H \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ and define the norms of y and H by

$$\|y\| = \max_{1 \leqq j \leqq n-1} |y_j| \quad \text{and} \quad \|H\| = \max_{1 \leqq j \leqq n-1} \sum_{k=1}^{n-1} |h_{jk}|.$$

For all $y, z \in \mathbb{R}^{n-1}$ for which $|y_i| > 0$, $|z_i| > 0$, $i = 1, 2, \ldots, n-1$ we obtain

$$\|F'(y) - F'(z)\| = \|\operatorname{diag} 2h^2(y_j - z_j)\| = 2h^2 \max_{1 \leqq j \leqq n-1} |y_j - z_j| \leqq 2h^2 \|y - z\|.$$

That is $\gamma = 2h^2$. We choose $n = 3$. Since a solution would vanish at the end points a reasonable choice of initial approximation seems to be $y(x) = \frac{1}{10^4} \sin \pi x$. This gives us the following vector:

$$v_0 = [8.66025403.10^{-5}, 8.66025403.10^{-5}]^{\mathrm{tr}}. \tag{46}$$

Let us choose the ball $\bar{U}(v_0, .1)$. Using Theorem 3 we get the following results

$$\alpha = .772202977$$

$$\frac{\|v_0 - F(v_0)\|}{1 - \alpha} = 4.752140333.10^{-4} = r_0, \quad r_0^* = 9.504280667.10^{-7},$$

$$q = .70998588, \quad q_0 = 1.241068778, \quad \bar{q} = q_0^{-1} \text{ and } h_s = 5.897732996.10^{-4}.$$

Set $w = 1$ to obtain $x_0 = 1.000475374$. With the above values it can easily be seen that the hypotheses of Theorem 4 are satisfied in the ball $\bar{U}(v_0, .1) \supset \bar{U}(x_0, r_0^*)$.

Therefore, there exists a unique fixed point $v^* = v^*(t) = 0$ for all $t \in [0, 1]$ of equation (44) which can be obtained as the limit of iteration (7) with v_0 given by (46). That is $(F, v_0) \in (\alpha, \gamma, r_0)$ and $(F, v_0) \in C(\alpha, \gamma, r_0, w)$.

We can now tabulate the following results

n	Rall (22)	ARG. (25)	ARG. (33)
0	$4.752140333.10^{-4}$	$4.75374.10^{-4}$	–
1	$2.802685484.10^{-7}$	$1.603349543.10^{-7}$	$1.621601384.10^{-7}$
2	$9.748652245.10^{-4}$	$1.825181975.10^{-14}$	$1.825181902.10^{-14}$
3	0	0	0

Moreover,

$$v_1 = \begin{bmatrix} 6.66185.10^{-10} \\ 6.66185.10^{-10} \end{bmatrix}, \quad v_2 = \begin{bmatrix} -6.10^{-19} \\ -6.10^{-19} \end{bmatrix} \quad \text{and } v_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The above table indicates that our results are better when compared with the corresponding ones in (22). This fact strongly recommends their usefulness in numerical applications.

REFERENCES

[1] BARTLE, R. G., Newton's method in Banach spaces, *Proc. Amer. Math. Soc.* **6** (1955), 827–831. *MR* **17**-176.

[2] DEUFLHARD, P. and HEINDL, G., Affine invariant convergence theorems for Newton's method and extensions to related methods, *SIAM J. Numer. Anal.* **16** (1979), 1–10. *MR* **88i**:65068

[3] GRAGG, W. B. and TAPIA, R. A., Optimal error bounds for the Newton–Kantorovich theorem, *SIAM J. Numer. Anal.* **11** (1974), 10–13. *MR* **49** #8334

[4] MIEL, G. J., Majorizing sequences and error bounds for iterative methods, *Math. Comp.* **34** (1980), 185–202. *MR* **81h**:65056

[5] ORTEGA, J. M. and RHEINBOLDT, W. C., *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970. *MR* **42** #8686

[6] POTRA, F.-A. and PTÁK, V., Sharp error bounds for Newton's process. *Numer. Math.* **34** (1980), 63–72. *MR* **81c**:65027

[7] POTRA, F.-A., On the a posteriori error estimates for Newton's method, *Beiträge Numer. Math.* **12** (1984), 125–138. *MR* **85h**:65128

[8] POTRA, F.-A. and PTÁK, V., *Nondiscrete induction and iterative processes*, Research Notes in Mathematics, 103, Pitman, Boston, Mass.–London, 1984. *MR* **86i**: 65003

[9] PTÁK, V., The rate of convergence of Newton's process, *Numer. Math.* **25** (1975/76), 279–285. *MR* **57** #18064

[10] RALL, L. B., Convergence of Stirling's method in Banach spaces, *Aequationes Math.* **12** (1975), 12–20. *MR* **51** #2281

[11] STIRLING, J., *Methodus differentialis: sive tractatus de summatione et interpolatione serierum infinitarum*, W. Boyer Publ., London, 1730.

[12] YAMAMOTO, T., A convergence theorem for Newton-like methods in Banach spaces, *Numer. Math.* **51** (1987), 545–557. *MR* **88i**:65081

(Received March 7, 1991)

DEPARTMENT OF MATHEMATICS
CAMERON UNIVERSITY
LAWTON, OK 73505
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 217–229

DREHKEGEL DES ZWEIFACH ISOTROPEN RAUMES DURCH VIER GEGEBENE PUNKTE

S. MICK

Zur Lösung eines kinematischen Problems hat H. Schaal in [5] die einparametrige Menge der Drehzylinder durch vier Punkte des dreidimensionalen euklidischen Raumes E_3 bestimmt. Im Anschluß daran wurde von U. Strobel in [6] und [7] die zweiparametrige Menge der Drehkegel des E_3 durch vier Punkte untersucht. Die Drehkegel des zweifach isotropen Raumes $I_3^{(2)}$ durch vier Punkte sind im projektiven Einbettungsraum P_3 des $I_3^{(2)}$ die Kegel 2. Ordnung durch vier gegebene Punkte, die im absoluten Punkt die absolute Gerade berühren. Von Weddle wurden in [8] die Kegel 2. Ordnung durch sechs Punkte des P_3 bestimmt und von Hierholzer in [2] die Gleichung der Kegelscheitelfläche erneut dargestellt. In dieser Arbeit werden zwei der Punkte durch ein Linienelement ersetzt und die dabei auftretende Kegelscheitelfläche im zweifach isotropen Raum beschrieben.

1. Einleitung

Es sei A_3 der dreidimensionale reelle affine Raum. Durch die Festlegung der Absolutfigur $\{\omega, f, F\}$, bestehend aus der Fernebene ω des A_3, einer Ferngeraden f und einem auf f liegenden Fernpunkt F, wird der affine Raum A_3 zum zweifach isotropen Raum $I_3^{(2)}$. Jede Untergruppe der affinen Gruppe, die gleichzeitig Automorphismengruppe der Absolutfigur ist, bestimmt eine zweifach isotrope Geometrie. Wählen wir die sechsgliedrige Gruppe G_6 der isotropen Bewegungen, die die isotropen Abstände und Winkel invariant läßt, so erhalten wir die *zweifach isotrope Bewegungsgeometrie* (siehe [1], 120 ff). H. Brauner hat in [1] die zweifach isotrope Geometrie entwickelt und auch bezüglich weiterer Begriffsbildungen der zweifach isotropen Geometrie wird auf diese Arbeiten verwiesen. Zur analytischen Darstellung des Raumes wählen wir ein affines Koordinatensystem $\{O, x, y, z\}$. Jeder Punkt des $I_3^{(2)}$ kann dann sowohl durch inhomogene Koordinaten $(x, y, z)^t$ als auch durch seine homogenen Koordinaten $(x_0 : x_1 : x_2 : x_3)^t$ dargestellt werden, wobei für $x_0 \neq 0$ $(x_0 : x_1 : x_2 : x_3) = (1 : x : y : z)$ gilt. Die absolute Ebene ω wird durch die Gleichung

$$x_0 = 0 \tag{1}$$

1991 *Mathematics Subject Classification.* Primary 51N25.
Key words and phrases. Doubly isotropic space, isotropic cone of revolution.

festgelegt. Wählen wir das Koordinatensystem so, daß die absolute Gerade Ferngerade der yz-Ebene und der absolute Punkt Fernpunkt der z-Achse ist, dann ist die absolute Gerade f durch das Gleichungssystem

$$x_0 = x_1 = 0 \tag{2}$$

festgelegt und für den absoluten Punkt F gilt

$$F = (0:0:0:1)^t. \tag{3}$$

Metrisch dual zu den isotropen Kreisen sind die als Ebenenmannigfaltigkeiten aufgefaßten Drehkegel definiert (siehe [1], 124 f). Ein *nichtisotroper Drehkegel, kurz Drehkegel des zweifach isotropen Raumes* $I_3^{(2)}$ ist ein Kegel 2. Klasse mit eigentlicher Spitze, der eine vollisotrope Erzeugende mit vollisotroper Tangentialebene enthält. Da bekanntlich die Fernkreise des $I_3^{(2)}$ die Fernkegelschnitte sind, die das absolute Linienelement $\{f, F\}$ enthalten, umhüllen die Ferngeraden der Ebenen eines Drehkegels einen Fernkreis. Im folgenden fassen wir die Drehkegel als Punktmengen und ihre Fernkreise als Punktkegelschnitte auf. Als Punktkegelschnitt kann ein Fernkreis auf zwei Arten zerfallen, entweder in ein Paar isotroper Ferngeraden oder in die absolute Gerade und eine nichtisotrope Ferngerade. Dem entsprechend bezeichnen wir als *zerfallende Drehkegel* Paare isotroper Ebenen und Ebenenpaare, bestehend aus einer vollisotropen Ebene und einer nichtisotropen Ebene. Anstelle der Scheitel treten die Kerngeraden, wobei ein Paar isotroper Ebenen eine vollisotrope Kerngerade, eine vollisotrope und eine nichtisotrope Ebene eine isotrope Kerngerade bestimmen. Die weiteren Typen von Drehkegeln des $I_3^{(2)}$ heißen *Drehzylinder, Punktkugeln* und *Punktgrenzkugeln.* Ein Drehzylinder hat einen isotropen Scheitel und berührt die Fernebene längs einer isotropen Ferngeraden, eine Punktkugel hat einen vollisotropen Scheitel und berührt die Fernebene längs der absoluten Geraden und eine Punktgrenzkugel schließlich hat ihren Scheitel im absoluten Punkt des $I_3^{(2)}$ und berührt die Fernebene längs der absoluten Geraden.

2. Synthetische Überlegungen zur Kegelscheitelfläche Φ

Zunächst betrachten wir die Menge der Drehkegel durch drei gegebene Punkte des $I_3^{(2)}$. Für ein Dreieck ABC des $I_3^{(2)}$, von dem keine Seite auf einer isotropen oder vollisotropen Geraden liegt und dessen Trägerebene eine nichtisotrope Ebene ist, gilt: Die Menge der Drehkegel durch die Punkte A, B und C ist dreiparametrig, denn es gibt zu jedem Punkt S des $I_3^{(2)}$, der weder auf einer Dreiecksseite noch auf den vollisotropen Geraden durch eine der Ecken liegt, genau einen Drehkegel durch die Punkte A, B und C mit S als Scheitel. Wenn S nicht in der Ebene ABC oder in der vollisotropen Ebene

durch eine der Ecken liegt, trägt der Kegel die Erzeugenden SA, SB, SC und längs der vollisotropen Erzeugenden SF berührt die vollisotrope Ebene durch SF. Für Punkte S der Ebene ABC, die auf keiner Dreiecksseite liegen, zerfällt der Drehkegel in die Ebene ABC und die vollisotrope Ebene durch S und für Punkte S, die in der vollisotropen Ebene durch eine der Ecken, aber nicht auf der vollisotropen Geraden durch diese Ecken liegen, zerfällt der Kegel in die vollisotrope Ebene durch S und die Verbindungsebene von S mit den restlichen beiden Ecken. Liegt S hingegen auf einer Seite des Dreiecks ABC, oder auf der vollisotropen Geraden durch eine der Ecken, so gibt es eine einparametrige Schar von Drehkegeln mit Scheitel S durch die gegebenen Punkte.

Es sei nun $ABCD$ ein Tetraeder des $I_3^{(2)}$, von dem keine Kante auf einer isotropen oder vollisotropen Geraden und keine Seitenfläche in einer isotropen Ebene liegt. Die Menge der Drehkegel durch A, B, C und D ist zweiparametrig und die Punkte S des $I_3^{(2)}$, die Scheitel eines Drehkegels sind, liegen auf einer Fläche Φ, der Kegelscheitelfläche. Um Geraden auf Φ zu bestimmen (Fig. 1), wählen wir zunächst einen Punkt S auf einer Tetraederkante oder auf einer der vollisotropen Geraden durch die Tetraederecken. Für einen solchen Punkt S stimmen zwei der Erzeugenden SA, SB, SC, SD und f_S — das ist die vollisotrope Gerade durch S — des Drehkegels überein, es bleiben also nur vier wesentliche Bestimmungsstücke und mit der vollisotropen Ebene durch f_S als Tangentialebene ist ein Drehkegel eindeutig bestimmt.

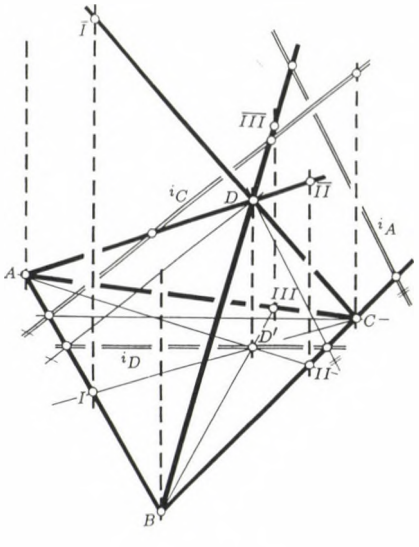


Fig. 1

Also liegen die Tetraederkanten und die vollisotropen Geraden durch die Ecken des Tetraeders auf Φ. Nun betrachten wir die zerfallenden Drehkegel durch A, B, C und D. Jeder Punkt einer so bestimmten Kerngeraden gehört

dann als Scheitel eines zerfallenden Drehkegels zu Φ. Zu den zerfallenden Drehkegeln durch A, B, C und D gehören die drei Ebenenpaare, die aus den isotropen Ebenen durch die Paare von Gegenkanten bestehen. Die vollisotropen Schnittgeraden sind gleichzeitig die Gemeinlote der Gegenkantenpaare und die Kerngeraden dieser zerfallenden Drehkegel. Weiters sind die Ebenenpaare, bestehend aus einer Tetraederseitenfläche und der vollisotropen Ebene durch die Gegenecke, zerfallende Drehkegel. Die vier Schnittgeraden solcher Ebenenpaare sind isotrope Geraden, die ebenfalls auf Φ liegen. Damit gilt

BEMERKUNG 1. *Auf der Kegelscheitelfläche Φ liegen 6 nichtisotrope, 4 isotrope und 7 vollisotrope Geraden, nämlich die 6 nichtisotropen Kanten des Tetraeders, die 4 isotropen Kerngeraden der zerfallenden Drehkegel, die eine Tetraederseitenfläche enthalten, die 4 vollisotropen Geraden durch die Ecken und die 3 (vollisotropen) Gemeinlote der Gegenkantenpaare des Tetraeders.*

Schneiden wir die vollisotropen Geraden mit der Fläche Φ, so erhalten wir

BEMERKUNG 2. *Im allgemeinen liegt auf jeder vollisotropen Geraden genau ein Punkt, der Scheitel eines Drehkegels durch die Punkte A, B, C und D ist.*

BEWEIS. Es sei S' ein Punkt der Ebene $\varepsilon = ABC$ und f_S die vollisotrope Gerade durch S'. Wir suchen nun Drehkegel durch die Punkte A, B, C und D, dessen Scheitel auf f_S liegen. Es sei φ_S die vollisotrope Ebene durch S' und $\varphi_S \cap \varepsilon = s'$ die isotrope Gerade von ε durch S'. Bestimmen A, B, C und das Linienelement (S', s') einen Kegelschnitt k, so ist der Punkt S auf f_S genau dann Scheitel eines Drehkegels durch die Punkte A, B, C und D, wenn der Durchstoßpunkt $\bar{D}$ der Geraden SD mit der Ebene ε auf dem Kegelschnitt k liegt (Fig. 2).

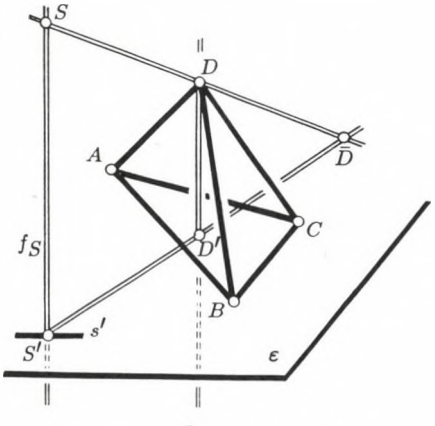


Fig. 2

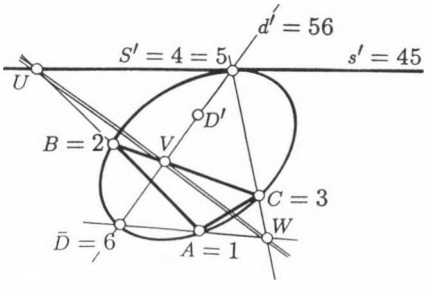


Fig. 3

Bezeichnen wir mit D' den Schnittpunkt der vollisotropen Geraden durch D mit der Ebene ε, so ist $S'D' = d'$ die Schnittgerade der isotropen Ebene $f_S D$ mit der Ebene ε und $\bar{D}$ der zweite Schnittpunkt von k mit d'. Konstruieren wir $\bar{D}$ mit Hilfe des Satzes von Pascal (Fig. 3), so ist $S = \bar{D}D \cap f_S$ der gesuchte Kegelscheitel. Diese Konstruktion liefert genau einen Kegelscheitel S auf f_S, außer die Geraden f_S und $\bar{D}D$ stimmen überein oder sind parallel.

Für die vollisotropen Geraden, auf denen kein Punkt Scheitel eines Drehkegels durch A, B, C und D ist, gilt

BEMERKUNG 3. *Bis auf endlich viele Ausnahmen liegt auf den Erzeugenden des vollisotropen Zylinders 3. Ordnung durch die Brennpunktskurve k_3 des Kegelschnittsbüschels mit den Grundpunkten A, B, C und D' kein Punkt, der Scheitel eines Drehkegels durch A, B, C und D ist. Auf den Erzeugenden durch die Grundpunkte und durch die Ecken des Diagonaldreiecks ist jeder Punkt Scheitel eines Drehkegels durch A, B, C und D. Ausnahmen sind also die 7 vollisotropen Geraden von Bemerkung 1.*

BEWEIS. Wir betrachten das Kegelschnittsbüschel mit den Grundpunkten A, B, C und D'. Die Tangenten durch den absoluten Punkt $F_1 = f \cap \varepsilon$ von ε an jeden Kegelschnitt des Büschels sind seine zwei isotropen Tangenten, die in den isotropen Brennpunkten berühren (siehe [4], 72). Wählt man einen Brennpunkt als Punkt S', so liefert die obige Konstruktion $\bar{D} = D'$, also ist DD eine vollisotrope Gerade und für $\bar{D}D' \neq f_S$ gibt es keinen Kegelscheitel auf der vollisotropen Geraden durch S'. $\bar{D}D' = f_S$ liefert die 7 Ausnahmegeraden. Die Brennpunkte eines Büschelkegelschnittes liegen also auf den Polaren des absoluten Punktes F_1 der Ebene ε. Alle diese Polaren gehen durch den doppeltkonjugierten Punkt $\bar{F}_1$ von F_1 bezüglich des Kegelschnittsbüschels. Das Strahlbüschel und das Kegelschnittsbüschel sind projektiv aufeinander bezogen. Die Projektivität zwischen einem Kegelschnittsbüschel und einem Geradenbüschel erzeugt eine Kurve k_3 3. Ordnung, was zu bewiesen war. Für verschiedene Typen von Kegelschnittsbüschels hat V. Šćurić diese Brennpunktskurven untersucht. Unter den vollisotropen Zylindern, die durch das Kegelschnittsbüschel mit den Grundpunkten A, B, C und D' bestimmt sind, wird im allgemeinen keine Punktgrenzkugel vorkommen.

3. Gleichung der Kegelscheitelfläche Φ

Um die Gleichung der Fläche Φ zu bestimmen, verwenden wir folgende Bedingung für die Kegelscheitel: S ist genau dann Scheitel eines Drehkegels durch die Punkte A, B, C und D, wenn die Fernpunkte A_u, B_u, C_u und D_u der Geraden SA, SB, SC und SD auf einem Fernkreis liegen. Da ein Fernkreis das absolute Linienelement enthält, lautet seine Gleichung

$$x_0 = 2\varrho x_1 x_3 + \alpha x_1 x_2 + \beta x_1^2 + \gamma x_2^2 = 0. \tag{4}$$

Wie man aus der Gleichung abliest, zerfällt der Fernkreis für $\varrho = 0$ in ein Paar isotroper Ferngeraden und für $\gamma = 0$ in die absolute Gerade und eine nichtisotrope Ferngerade. Sind

$$\begin{aligned} S &= (1 : x : y : z)^t \\ A &= (1 : a_1 : a_2 : a_3)^t \\ B &= (1 : b_1 : b_2 : b_3)^t \\ C &= (1 : c_1 : c_2 : c_3)^t \\ D &= (1 : d_1 : d_2 : d_3)^t \end{aligned} \tag{5}$$

die Koordinaten der Punkte, so sind die Koordinaten der Fernpunkte der Erzeugenden

$$\begin{aligned} A_u &= (0 : a_1 - x : a_2 - y : a_3 - z)^t \\ B_u &= (0 : b_1 - x : b_2 - y : b_3 - z)^t \\ C_u &= (0 : c_1 - x : c_2 - y : c_3 - z)^t \\ D_u &= (0 : d_1 - x : d_2 - y : d_3 - z)^t \end{aligned} \tag{6}$$

und wir erhalten durch Einsetzen von (6) in (4) ein homogenes lineares Gleichungssystem für die Koeffizienten $\varrho, \alpha, \beta, \gamma$ des isotropen Fernkreises:

$$\begin{aligned} 2\varrho(a_1 - x)(a_3 - z) + \alpha(a_1 - x)(a_2 - y) + \beta(a_1 - x)^2 + \gamma(a_2 - y)^2 &= 0 \\ 2\varrho(b_1 - x)(b_3 - z) + \alpha(b_1 - x)(b_2 - y) + \beta(b_1 - x)^2 + \gamma(b_2 - y)^2 &= 0 \\ 2\varrho(c_1 - x)(c_3 - z) + \alpha(c_1 - x)(c_2 - y) + \beta(c_1 - x)^2 + \gamma(c_2 - y)^2 &= 0 \\ 2\varrho(d_1 - x)(d_3 - z) + \alpha(d_1 - x)(d_2 - y) + \beta(d_1 - x)^2 + \gamma(d_2 - y)^2 &= 0. \end{aligned} \tag{7}$$

Dieses Gleichungssystem hat genau dann eine nichttriviale Lösung, wenn seine Determinante Null ist. Entwickelt man diese Determinante und ordnet nach den Monomen in x, y und z, so erhält man die Gleichung der Kegelscheitelfläche in den Unbestimmten x, y und z. Um die Berechnung der Determinante zu vereinfachen, führen wir eine isotrope Koordinatentransformation so durch, daß im neuen Koordinatensystem die gegebenen Punkten Koordinaten

$$\begin{aligned} A &= (1 : 0 : 0 : 0)^t \\ B &= (1 : a : 0 : 0)^t \\ C &= (1 : b : c : 0)^t \\ D &= (1 : d : e : f)^t \end{aligned} \tag{8}$$

haben. Damit keine Kante des Tetraeders auf einer isotropen oder vollisotropen Geraden und keine Seitenfläche in einer isotropen Ebene liegt und das Tetraeder auch nicht entartet, müssen alle Zahlen a, b, c, d, e und f von

Null verschieden und zusätzlich a, b und d paarweise verschieden sein. Mit (8) vereinfachen sich auch die Koordinaten (6) und das Gleichungssystem (7) und seine Determinante lautet

$$\text{(9)} \qquad \text{Det} = \begin{vmatrix} xz & xy & x^2 & y^2 \\ (a-x)z & (a-x)y & (a-x)^2 & y^2 \\ (b-x)z & (b-x)(c-y) & (b-x)^2 & (c-y)^2 \\ (d-x)(f-z) & (d-x)(e-y) & (d-x)^2 & (e-y)^2 \end{vmatrix}.$$

Wir erhalten als Gleichung der Kegelscheitelfläche Φ

$$\text{(10)} \qquad \text{Det} = zT_3(x,y) + T_4(x,y) = 0$$

mit

$$\begin{aligned} T_3(x,y) = {} & ce(c-e)x^3 + 2ce(d-b)x^2y + \\ & + [cd(a-d) + be(b-a)]xy^2 + bd[c(d-a) + e(a-b)]y^2 + \\ & + ce[e(a+b) - c(a+d)]x^2 + 2ace(b-d)xy + ace(cd-be)x \end{aligned} \qquad \text{(11)}$$

und

$$\text{(12)} \qquad T_4(x,y) = fy(cx-by)[-cx + (b-a)y + ac](x-d).$$

Flächen 4. Ordnung mit einer Gleichung der Gestalt (10), in der $T_3(x,y)$ ein Polynom 3. Grades und $T_4(x,y)$ ein Polynom 4. Grades bezeichnen, haben einen dreifachen Punkt im Fernpunkt der z-Achse. Φ ist also eine Fläche 4. Ordnung mit F als dreifachen Punkt. Φ ist rational, genauer gibt es zu jeden Punkt $P' = (x,y,0)^t$ der Ebene $\pi_1 = ABC$ mit der Gleichung $z = 0$ einen Punkt $P = \left(x, y, \frac{-T_4(x,y)}{T_3(x,y)}\right)^t$ der Fläche Φ, soferne $T_3(x,y) \neq 0$ ist, was in Übereinstimmung mit den Bemerkungen 2 und 3 steht. Wegen (10) und (11) ist

$$\begin{aligned} T_3(x,y) = {} & ce(c-e)x^3 + 2ce(d-b)x^2y + \\ & + [cd(a-d) + be(b-a)]xy^2 + bd[c(d-a) + e(a-b)]y^2 + \\ & + ce[e(a+b) - c(a+d)]x^2 + 2ace(b-d)xy + ace(cd-be)x = 0 \end{aligned} \qquad \text{(13)}$$

die Gleichung des Tangentialzylinder 3. Ordnung Γ_3 von Φ in F und in der Ebene π_1 seiner Spurkurve k_3 3. Ordnung, die nach Bemerkung 3 gleichzeitig die Brennpunktskurve des Kegelschnittsbüschels mit den Grundpunkten A, B, C und D' ist. Speziell erkennt man, daß wegen $\frac{\partial T_3}{\partial x} \neq 0$, $\frac{\partial T_3}{\partial y} = 0$, in den Punkten A, B, C und D' die Tangenten von k_3 isotrop sind. Schreibt man (13) homogen, so erkennt man, daß der Fernpunkt F_1 der y-Achse der Restschnittpunkt dieser Tangenten mit der Kurve ist. Damit ist auch gezeigt, daß k_3 eine Kurve 6. Klasse ist. Aus der Gleichung

$$\text{(14)} \qquad T_4(x,y) = fy(cx-by)[-cx + (b-a)y + ac](x-d) = 0$$

des Zylinders Γ_4 erkennt man, daß der Zylinder Γ_4 in 4 Ebenen zerfällt. Seine Spurkurve k_4 in π_1 besteht aus den Seiten des Dreiecks ABC und der isotropen Geraden der Ebene π_1 durch D'. Die Schnittgeraden der beiden Zylinder Γ_3 und Γ_4 liegen auf Φ, es sind die vollisotropen Geraden durch A, B, C und D, sowie die 3 vollisotropen Geraden durch die Punkte

$$
\begin{aligned}
I &= \Big(\frac{cd-be}{c-e}, 0\Big) \\
II &= \Big(\frac{acd}{-ae+be-cd}, \frac{ace}{-ae+be-cd}\Big) \\
III &= \Big(\frac{abe}{-ac-be+cd}, \frac{ace}{-ac-be+cd}\Big),
\end{aligned} \tag{15}
$$

die gleichzeitig die Gemeinlote der Gegenkantenpaare des Tetraeders $ABCD$ sind. Das bestätigt eine Teilaussage von Bemerkung 1. Darüber hinaus erkennt man aus der Gleichung von Φ, daß die vollisotropen Geraden durch A, B, C und D Torsalgeraden der Fläche sind, die Torsalebenen sind die vollisotropen Ebenen, die längs dieser Geraden Φ berühren. A, B, C und D sind Doppelpunkte von Φ und der Tangentialkegel in einem solchen Doppelpunkt ist der Drehkegel des $I_3^{(2)}$ durch die restlichen Doppelpunkte. Als nächstes untersuchen wir die Fernkurve von Φ. Aus (10), (11) und (12) erhalten wir die Gleichung der Fernkurve der Fläche Φ. Sie lautet

$$
\begin{aligned}
&czx((ce(c-e)x^2 + 2ce(d-b)xy + (cd(a-b) + be(b-a))y^2) + \\
&\quad + fxy(-c^2x^2 - b(b-a)y^2 + (c(b-a)+bc)xy) = 0.
\end{aligned} \tag{16}
$$

Ergänzend zu Bemerkung 1 erkennen wir, daß die absolute Gerade als Bestandteil der Fernkurve von Φ eine Gerade durch den dreifachen Punkt F von Φ ist. Der zweite Bestandteil der Fernkurve ist eine rationale Kurve l_3 mit dem Doppelpunkt F, der für spezielle Koordinaten von A, B, C und D zu einer Spitze werden kann. l_3 ist die Scheitelmenge der Drehzylinder des $I_3^{(2)}$ durch A, B, C und D. Dieser Menge gehört im allgemeinen genau eine Punktkugel an (siehe [3], 47). Als Ergebnis erhalten wir den folgenden

Satz. *Sind vier Punkte A, B, C und D des $I_3^{(2)}$ gegeben, die Ecken eines Tetraeders mit lauter nichtisotropen Kanten und nichtisotropen Seitenflächen sind, so liegen die Kegelscheitel der Drehkegel durch diese Punkte auf einer rationalen Fläche Φ 4. Ordnung. Der absolute Punkt F ist dreifacher Punkt, die Punkte A, B, C und D sind konische Doppelpunkte von Φ. Auf Φ liegen 18 Geraden, speziell sind die vollisotropen Geraden durch A, B, C und D Torsalgeraden von Φ mit vollisotropen Torsalebenen. Die Tangentialkegel 2. Ordnung in den Doppelpunkten sind die Drehkegel des $I_3^{(2)}$ durch die restlichen Doppelpunkte der Fläche, der Tangentialkegel 3. Ordnung im absoluten Punkt F geht durch die Brennpunktskurve k_3 des*

Kegelschnittsbüschels mit den Grundpunkten A, B, C und D', wobei D' der Schnittpunkt der vollisotropen Geraden durch D mit der Ebene ABC ist.

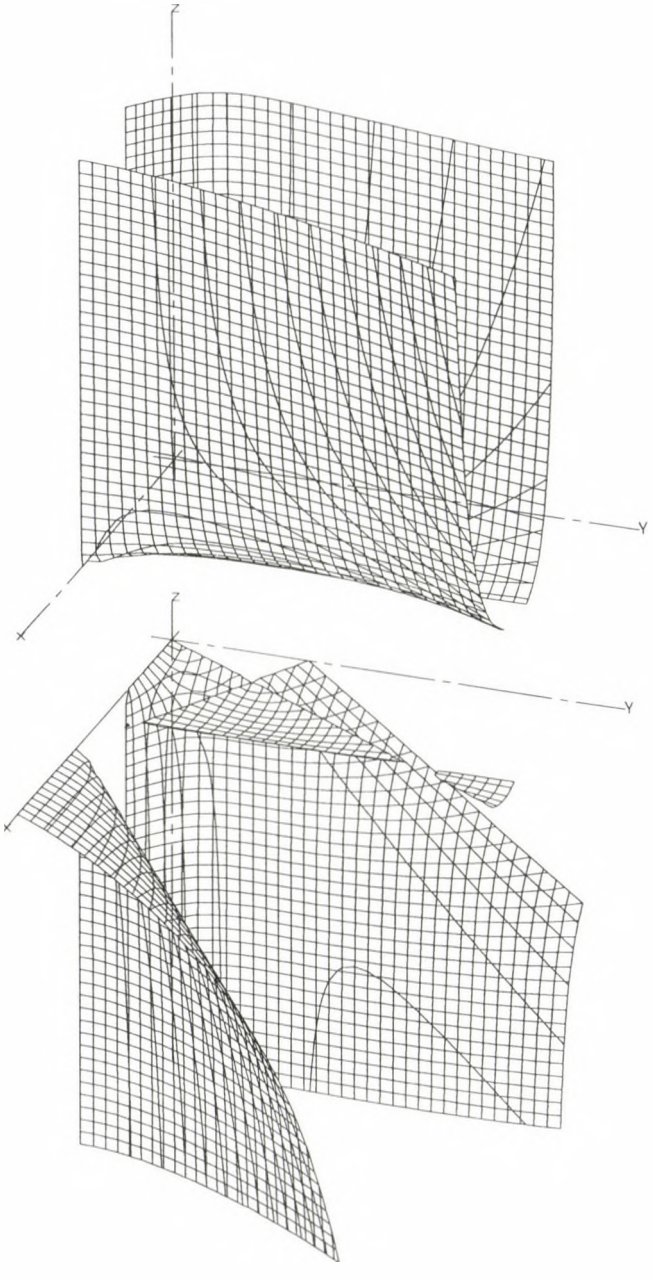


Fig. 4

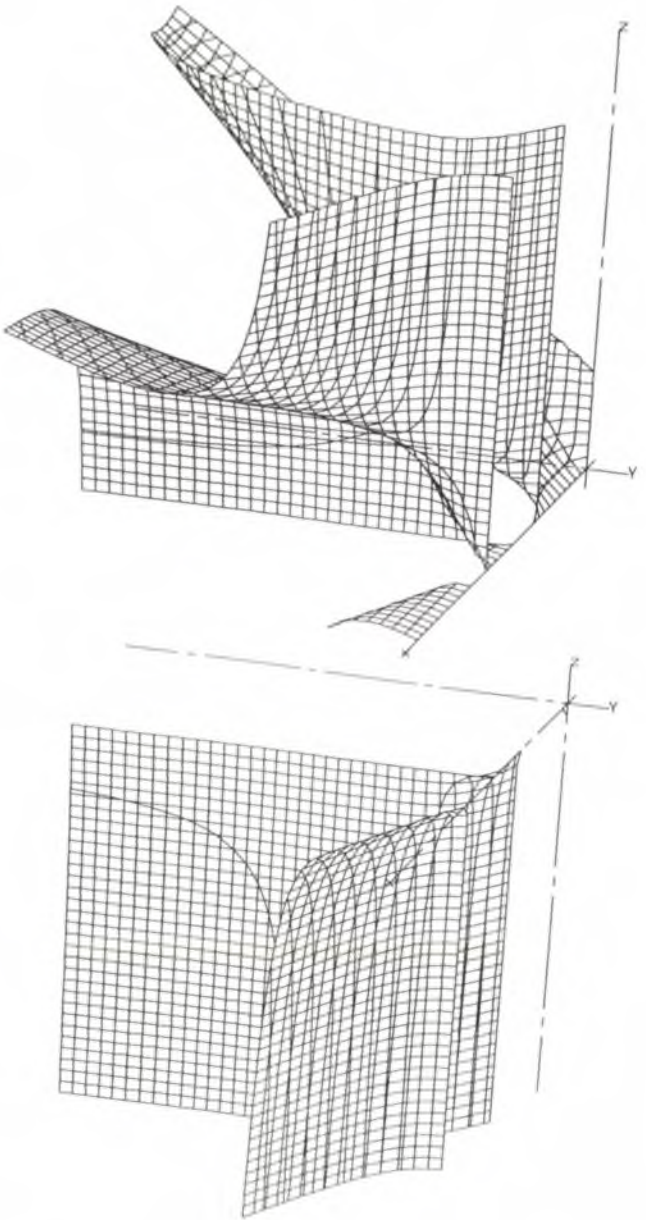


Fig. 4a

Die Figuren 4, 4a, 5, 5a zeigen die Kegelscheitelflächen für $a = 12$, $b = 3$, $c = -e = f = 5$, $d = 9$, bzw. $a = 12$, $b = 3$, $c = f = 5$, $d = 9$, $e = 2{,}7$.

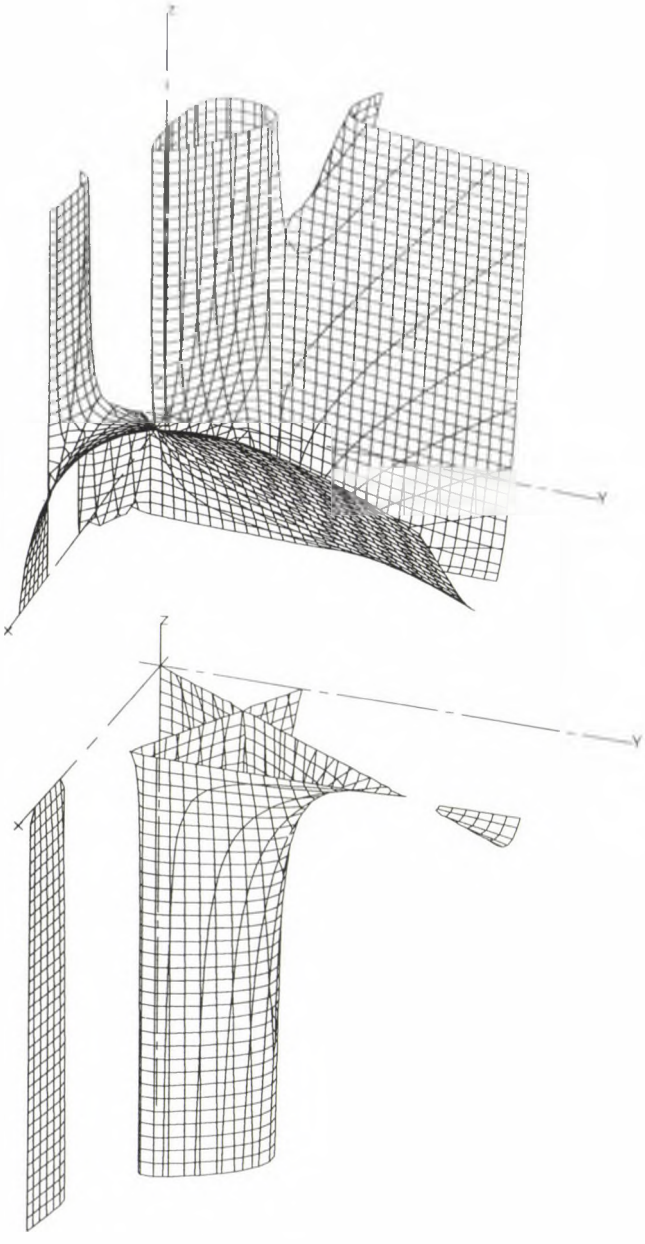


Fig. 5

Die Flächen wurden von Herrn Michael Schmidt von der TH Darmstadt auf einer VAX 8530 generiert und dargestellt, wofür ich ihm an dieser Stelle danken möchte.

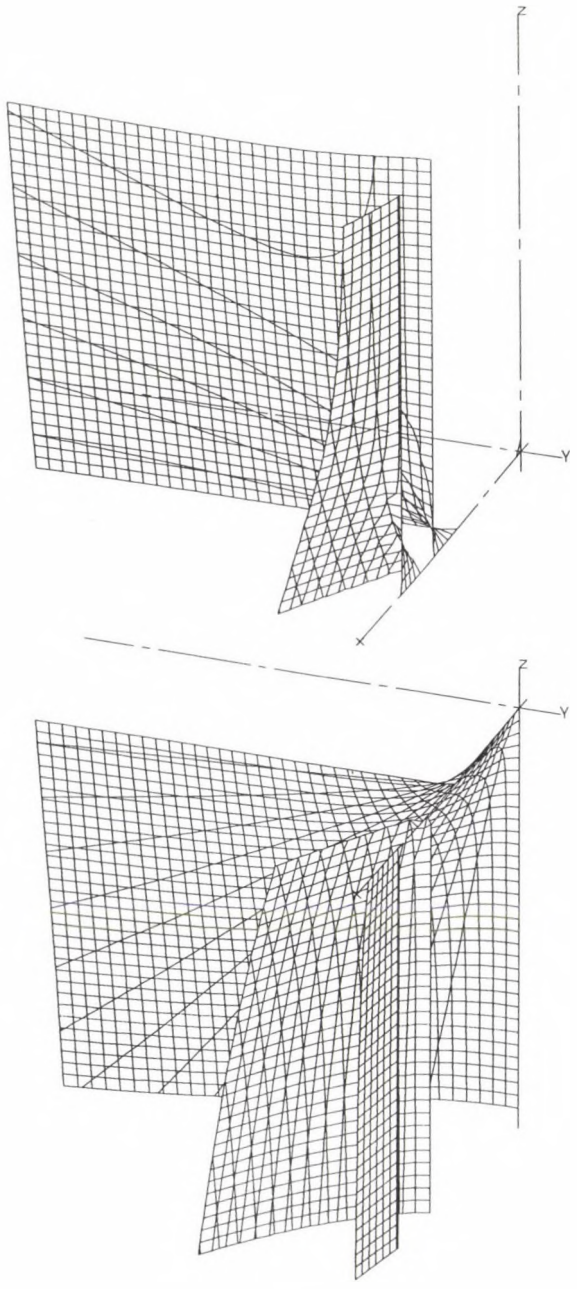


Fig. 5a

Bei besonderer Lage der gegebenen vier Punkte treten Modifikationen der Fläche vierter Ordnung ein, die wir mit Hilfe der Gleichungen (10), (11) und (12) leicht verfolgen können. Zum Abschluß soll ein interessanter Son-

derfall beschrieben werden. Sind zwei der Tetraederkanten im Gegensatz zur allgemeinen Voraussetzung isotrop, so zerfällt die Kegelscheitelfläche in die beiden vollisotropen Ebenen durch diese Kanten und eine Fläche zweiter Ordnung durch den absoluten Punkt F. Das von den vier nichtisotropen Kanten des Tetraeders gebildete windschiefe Vierseit ist ein Erzeugendenvierseit dieser Fläche zweiter Ordnung, die dadurch bereits bestimmt ist. Dies erkennt man, indem man in (10), (11) und (12) $d = a - b = 0$ setzt. Im allgemeinen wird die Fläche zweiter Ordnung ein Hyperboloid sein, wenn allerdings zwei der nichtisotropen Kanten des Tetraeders zu einer isotropen Ebene parallel sind, stellt sich ein hyperbolisches Paraboloid ein.

LITERATURVERZEICHNIS

[1] Brauner, H., Geometrie des zweifach isotropen Raumes I, II, III, *J. Reine Angew. Math.* **224** (1966), 118–146; **226** (1967), 132–158; **228** (1967), 38–70. *MR* **34** #1903; *MR* **35** #4831; *MR* **36** #2072

[2] Hierholzer, C., Ueber eine Fläche der vierten Ordnung, *Math. Ann.* **4** (1871), 172–180. *Jb. Fortschritte Math.* **3**, 394

[3] Pottmann, H., Über Scheitel von Normalrissen einer Raumkurve, *Österreich Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **196** (1987), 39–48. *MR* **89c**:53007

[4] Sachs, H., *Ebene isotrope Geometrie*, Friedr. Vieweg und Sohn, Braunschweig, 1987. *MR* **88j**:51036

[5] Schaal, H., Ein geometrisches Problem der metrischen Getriebesyntese, *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **194** (1985), 39–53. *MR* **87j**:53015

[6] Strobel, U., Die Drehkegel durch 4 Punkte, Dissertation, Stuttgart, 1990.

[7] Strobel, U., Über die Drehkegel durch vier Punkte, *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **198** (1989), 281–293. *MR* **91g**:53014

[8] Weddle, *Cambr. Dubl. Math. J.* **5** (1850).

(Eingegangen am 20. April, 1991)

INSTITUT FÜR GEOMETRIE
TECHNISCHE UNIVERSITÄT GRAZ
KOPERNIKUSGASSE 24
A–8010 GRAZ
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 231–241

A COMMUTATIVE NEUTRIX CONVOLUTION OF DISTRIBUTIONS ON $\mathbb{R}^m$

CHENG LIN-ZHI

Abstract

Let $\tau(x) \in \mathbf{C}^{\infty}(\mathbb{R})$ having the properties: (i) $\tau(x) = \tau(-x)$, (ii) $0 \leqq \tau(x) \leqq 1$, (iii) $\tau(x) = 1$ for $|x| \leqq \frac{1}{2}$, (iv) $\tau(x) = 0$ for $|x| \geqq 1$ be fixed. The unit sequence $\{\tau_n(x)\}$, $x \in \mathbb{R}^m$, and $n \in \mathbf{I}^m$, is defined by $\tau_n(x) = \tau(x_1/n_1) \ldots \tau(x_m/n_m)$ for $n_1, \ldots, n_m = 1, 2, \ldots$. The neutrix convolution $f \tilde{\circledast} g$ of two distributions f and g in $D'(\mathbb{R}^m)$ is then defined to be the neutrix limit of the sequence $\{f_n * g_n\}$, where $f_n = f \cdot \tau_n$ and $g_n = g \cdot \tau_n$. Several results are given.

1. Introduction

The following definition for the convolution of distributions compatible with the convolution of summable functions was first given by Schwartz [8].

DEFINITION 1. Let f and g be distributions in $D'(\mathbb{R}^m)$. The convolution product $f * g = g * f$ is defined by

$$\langle f * g, \Phi \rangle = \langle f(x) \times g(y), \Psi(x, y) \rangle \tag{1}$$

where

$$\Psi(x, y) = \Phi(x + y), \qquad \forall \Phi \in D(\mathbb{R}^m)$$

if the right-hand side of (1) has meaning.

Either of the following conditions, see [8],

(a) either f or g has compact support,
(b) the supports of f and g are limited on the same side,

is sufficient for the convolvability of two distributions under Definition 1.

However, the convolution $f * g$ under Definition 1 is not usually defined, since $\Phi \in D(\mathbb{R}^m)$ does not imply that Φ, defined as a function on $\mathbb{R}^{2m}$ by $(x, y) \mapsto \Phi(x + y)$, is in $D(\mathbb{R}^{2m})$.

The method of the sequential completion is another way to define the convolution of distributions that is also compatible with the convolution of summable functions. This was first used by Mikusiński [7].

To deal with the sequential approach we now let τ be a fixed function in $D(\mathbb{R})$ having the properties

1991 *Mathematics Subject Classification.* Primary 46F10

Key words and phrases. Convolution of distributions, unit sequence, sequential approach, decomposition theorem.

0081–6906/95/$ 4.00

(i) $\tau(x) = \tau(-x)$,
(ii) $0 \leqq \tau(x) \leqq 1$,
(iii) $\tau(x) = 1$ for $|x| \leqq \frac{1}{2}$,
(iv) $\tau(x) = 0$ for $|x| \geqq 1$.

The function τ_n is defined by $\tau_n(x) = \tau(x/n)$ for $n = 1, 2, \ldots$. It is obvious that $\{\tau_n\}$ is a sequence of functions in $D(\mathbb{R})$ converging to the unit function on $\mathbb{R}$ in the sense that

$$\lim_{n\to\infty} \langle \tau_n(x), \Phi(x) \rangle = \langle 1, \Phi(x) \rangle, \qquad \forall \Phi \in D(\mathbb{R}).$$

For an arbitrary distribution f we will define the truncating f_n by $f_n(x) = f(x)\tau_n(x)$ for $n = 1, 2, \ldots$. It follows that $\{f_n\}$ is a sequence of distributions with compact supports converging to f.

The following definition was given by Jones [6].

DEFINITION 2. Let f and g be distributions in $D'(\mathbb{R})$. Then the convolution product $f \circledast g$ is defined as the limit of the sequence $\{f_n * g_n\}$, providing the limit h exists in the sense that

$$\langle f \circledast g, \Phi \rangle = \lim_{n\to\infty} \langle f_n * g_n, \Phi \rangle = \langle h, \Phi \rangle$$

for all Φ in $D(\mathbb{R})$.

The convolution $f_n * g_n$ in this definition is in the sense of Definition 1 since the supports of f_n and g_n being compact. It is clear that the convolution $f \circledast g$ is commutative if it exists under Definition 2.

In the following we are going to generalize Definition 2 to define the neutrix convolution $f \bar{\circledast} g$ of two convolutions f and g in $D'(\mathbb{R}^m)$ on applying the special technique of the separation of variables to deal with the product of distributions, see [1] and [4].

2. The neutrix convolution product in $D'(\mathbb{R}^m)$

First of all we define the function $\tau_n(x)$, where $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $n = (n_1, \ldots, n_m) \in I^m$, by

$$\tau_n(x) = \tau(x_1/n_1) \ldots \tau(x_m/x_n)$$

for $n_1, \ldots, n_m = 1, 2, \ldots$. It is obvious that $\{\tau_n\}$ is a sequence of functions in $D(\mathbb{R}^m)$ converging to the unit function of $\mathbb{R}^m$ in the sense that

$$\lim_{n_1\to\infty} \cdots \lim_{n_m\to\infty} \langle \tau_n(x), \Phi(x) \rangle = \langle 1, \Phi(x) \rangle$$

or more briefly

$$\lim_{n\to\infty} \langle \tau_n(x), \Phi(x) \rangle = \langle 1, \Phi(x) \rangle$$

for all Φ in $D(\mathbb{R}^m)$.

For an arbitrary distribution f in $D'(\mathbb{R}^m)$ we define the truncating $f_n(x)$ by

$$f_n(x) = f(x)\tau_n(x).$$

It follows that $\{f_n\}$ is a sequence of distributions with compact supports converging to f in the sense that

$$\lim_{n\to\infty} \langle f_n(x), \Phi(x)\rangle = \langle f(x), \Phi(x)\rangle$$

for all Φ in $D(\mathbb{R}^m)$.

Following [5], but with a different unit sequence $\tau_n(x)$, we have

DEFINITION 3. Let f and g be distributions in $D'(\mathbb{R}^m)$ and let $f_n(x) = f(x)\tau_n(x)$ and $g_n(x) = g(x)\tau_n(x)$. Then the neutrix convolution product $f \bar{\circledast} g$ is defined as the limit of sequence $\{f_n * g_n\}$, providing the limit h exists in the sense that

$$\text{N-}\lim_{n_1\to\infty} \cdots \text{N-}\lim_{n_m\to\infty} \langle f_n * g_n, \Phi\rangle = \langle h, \Phi\rangle \tag{3}$$

or more briefly

$$\text{N-}\lim_{n\to\infty} \langle f_n * g_n, \Phi\rangle = \langle h, \Phi\rangle \tag{4}$$

for all Φ in $D(\mathbb{R}^m)$, provided h is independent of the order in which the neutrix limits are taken, where N is the neutrix, see van der Corput [2], having domain $N' = \{1, 2, \ldots, n, \ldots\}$ and range the real numbers with negligible functions finite linear sums of the functions

$$n^{\mu} \ln^{r-1} n, \ln^r n \qquad (\mu > 0;\ r = 1, 2, \ldots)$$

and all functions which converge to zero in the normal sense as n tends to infinity. In particular, if

$$\lim_{n_1\to\infty} \cdots \lim_{n_m\to\infty} \langle f_n * g_n, \Phi\rangle = \langle h, \Phi\rangle \tag{5}$$

for all Φ in $D(\mathbb{R}^m)$, we simply say that the convolution product $f \circledast g$ exists and write

$$f \circledast g = f \bar{\circledast} g.$$

The convolution $f_n * g_n$ in this definition is again in the sense of Definition 1. Clearly the neutrix convolution $f \bar{\circledast} g$ is commutative if it exists under Definition 3. The following theorems can be proved by modifying the corresponding proofs of theorems in [6].

THEOREM 1. *Let f and g be functions in $L_p(\mathbf{R}^m)$ and $L_q(\mathbf{R}^m)$, respectively, with $p^{-1}+q^{-1}=1$ and $1 \leqq p,q \leqq \infty$. Then the neutrix convolution $f \tilde{\circledast} g$ exists and*

$$(f \tilde{\circledast} g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\,dt,$$

the classical definition of the convolution product.

THEOREM 2. *Let f and g be distributions in $D'(\mathbf{R}^m)$. Suppose that either the condition* (a) *or the condition* (b) *in Definition 1 is true. Then the neutrix convolution $f \tilde{\circledast} g$ exists and*

$$f \tilde{\circledast} g = f * g.$$

In order to prove our main results we need the following lemmas (see Schwartz [8]).

LEMMA 1. *The convolution product of two direct products is equal to the direct product of the convolution products*

$$(A_x \times C_y) * (B_x \times D_y) = (A_x * B_x) \times (C_y * D_y)$$

if $A_x,\ B_x \in D'(\mathbf{R}^p)$ and $C_y,\ D_y \in D'(\mathbf{R}^q)$ such that each of A_x and C_y has compact support.

LEMMA 2. *The vector subspace of the functions $\Phi(x)$ in the form*

$$\Phi(x) = \Phi(x_1, \ldots, x_m) = \sum_i \Phi_{1i}(x_1) \ldots \Phi_{mi}(x_m)$$

is dense in $D(\mathbf{R}^m)$, where $\Phi_{1i}, \ldots, \Phi_{mi} \in D(\mathbf{R})$.

THEOREM 3 (DECOMPOSITION THEOREM). *Let f and g be distributions in $D'(\mathbf{R}^m)$ such that*

$$f(x) = f_1(x_1) \times \cdots \times f_m(x_m), \quad g(x) = g_1(x_1) \times \cdots \times g_m(x_m)$$

with $f_1, \ldots, f_m,\ g_1, \ldots, g_m$ in $D'(\mathbf{R})$. Suppose that the neutrix convolutions $f_1 \tilde{\circledast} g_1, \ldots, f_m \tilde{\circledast} g_m$ exist and equal $h_1, \ldots, h_m$, respectively. Then the neutrix convolution $f \tilde{\circledast} g$ exists and $f \tilde{\circledast} g = h_1 \times \cdots \times h_m$, i.e.

$$(f_1(x_1) \times \cdots \times f_m(x_m)) \tilde{\circledast} (g_1(x_1) \times \cdots \times g_m(x_m)) = (f_1 \tilde{\circledast} g_1) \times \cdots \times (f_m \tilde{\circledast} g_m).$$

In particular, if the convolutions $f_1 \circledast g_1, \ldots, f_m \circledast g_m$ exist, then

$$(f_1(x_1) \times \cdots \times f_m(x_m)) \circledast (g_1(x_1) \times \cdots \times g_m(x_m)) = (f_1 \circledast g_1) \times \cdots \times (f_m \circledast g_m).$$

PROOF. Putting

$$f_n(x) = f(x)\tau_n(x), \qquad g_n(x) = g(x)\tau_n(x)$$

and

$$f_{in_i}(x_i) = f_i(x_i)\tau_{n_i}(x_i), \qquad g_{in_i}(x_i) = g_i(x_i)\tau_{n_i}(x_i)$$

for $i = 1, \dots, m$. Then

$$f_n(x) = f_{1n_1}(x_1) \times \cdots \times f_{mn_m}(x_m) = \prod_{i=1}^{m} f_{in_i}(x_i),$$

$$g_n(x) = g_{1n_1}(x_1) \times \cdots \times g_{mn_m}(x_m) = \prod_{i=1}^{m} g_{in_i}(x_i).$$

By Lemma 1 we have

$$\begin{aligned}
&\langle (f_n * g_n)(x), \Phi_1(x_1) \dots \Phi_m(x_m) \rangle = \\
&= \langle \prod_{i=1}^{m} (f_{in_i} * g_{in_i})(x_i), \ \prod_{i=1}^{m} \Phi_i(x_i) \rangle \\
&= \prod_{i=1}^{m} \langle (f_{in_i} * g_{in_i})(x_i), \ \Phi_i(x_i) \rangle
\end{aligned}$$

for all $\Phi_1, \dots, \Phi_m$ in $D(\mathbf{R})$. Now since the neutrix convolution $f_i \tilde{\circledast} g_i$ exists and equals h_i for $i = 1, \dots, m$, it follows that

$$\begin{aligned}
&\underset{n\to\infty}{\text{N-lim}} \langle (f_n * g_n)(x), \ \Phi_1(x_1) \dots \Phi_m(x_m) \rangle \\
&= \prod_{i=1}^{m} \underset{n_i\to\infty}{\text{N-lim}} \langle (f_{in_i} * g_{in_i})(x_i), \ \Phi_i(x_i) \rangle \\
&= \prod_{i=1}^{m} \langle h_i, \Phi_i \rangle = \langle h_1 \times \cdots \times h_m, \Phi \rangle,
\end{aligned}$$

and by Lemma 2 it follows that

$$\underset{n\to\infty}{\text{N-lim}} \langle f_n * g_n, \ \Phi \rangle = \langle h_1 \times \cdots \times h_m, \Phi \rangle$$

for all Φ in $D(\mathbf{R}^m)$. The result of the theorem follows.

An analogical decomposition theorem has been proved for a different unit sequence $\tau_n(x)$ in [5].

In the following, several neutrix convolutions of distributions in $D'(\mathbf{R})$ by Definition 3 are given, and then we will extend them from $D'(\mathbf{R})$ to $D'(\mathbf{R}^m)$ on using the decomposition theorem.

3. Several neutrix convolutions of distributions in $D'(\mathbb{R})$

THEOREM 4. *The neutrix convolution product* $x^r \circledast x^s$ *exists in* $D'(\mathbb{R})$ *and*

$$x^r \circledast x^s = 0 \tag{6}$$

for $r, s = 0, 1, 2, \ldots$.

PROOF. Putting

$$(x^r)_n = x^r \tau(x/n), \qquad (x^s)_n = x^s \tau(x/n)$$

then

$$(x^r)_n * (x^s)_n = \int_{-n}^{n} (x-t)^r \tau((x-t)/n) t^s \tau(t/n)\, dt =$$

$$= n^{r+s+1} \int_{-1}^{1} v^s \tau(v)((x/n)-v)^r \tau(v-(x/n))\, dv$$

for $r, s = 0, 1, 2, \ldots$, where the substitution $nv = t$ has been made.

By Taylor's theorem we have

$$(x^r)_n * (x^s)_n = n^p \int_{-1}^{1} v^s \tau(v) J_1(n, x, v) J_2(n, x, v)\, dv$$

where $p + r + s = 1$ and

$$J_1(n, x, v) = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} (x/n)^j v^{r-j},$$

$$J_2(n, x, v) = \sum_{k=0}^{p} \frac{(-1)^k}{k!} \tau^{(k)}(v)(x/n)^k + o((x/n)^p).$$

Then

$$\underset{n\to\infty}{\text{N-lim}} \langle (x^r)_n * (x^s)_n, \Phi \rangle = \sum_{\substack{j+k=p \\ j=0,1,\ldots,r}} \langle h_{jk}, \Phi \rangle x^p$$

where

$$h_{jk} = \frac{(-1)^{s+1}}{k!} \binom{r}{j} \int_{-1}^{1} v^{r+s-j} \tau(v) \tau^{(k)}(v)\, dv$$

for $j = 0, 1, \ldots, r$ and $k = r + s - j + 1$.

Note that $r+s-j-k=-1$ the integrands are odd, then $h_{jk}=0$ for $j=0,1,\dots,r$ and $k=r+s-j+1$. Hence

$$\underset{n\to\infty}{\text{N-lim}}\langle (x^r)_n * (x^s)_n, \Phi\rangle = 0$$

for all Φ in $D(\mathbf{R})$. This completes the proof of the theorem.

THEOREM 5. *The neutrix convolution product $x_+^r \tilde{\circledast} x_-^s$ in the sense of principal value exists in $D'(\mathbf{R})$ and*

$$x_+^r \tilde{\circledast} x_-^s = B(r+1,s+1)\frac{(-1)^{s+1}x_+^{r+s+1}+(-1)^{r+1}x_-^{r+s+1}}{2} \tag{7}$$

for $r,s=0,1,2\dots$, where B denotes the Beta function.

PROOF. It is clear that

$$x^r = x_+^r + (-1)^r x_-^r, \qquad x^s = x_+^s + (-1)^s x_-^s \tag{8}$$

for $r,s=0,1,2\dots$. Suppose the neutrix convolution

$$x_+^r \tilde{\circledast} x_-^s = \underset{n\to\infty}{\text{N-lim}}\{(x_+^r)_n * (x_-^s)_n\}$$

exists for $r,s=0,1,2\dots$. It implies that the convolution $x_-^r \tilde{\circledast} x_+^s$ exists by the commutativity of the neutrix convolution product. We then have by (8)

$$\begin{aligned} 0 &= x^r \tilde{\circledast} x^s \\ &= x_+^r \tilde{\circledast} x_+^s + (-1)^{r+s} x_-^r \tilde{\circledast} x_-^s + (-1)^s x_+^r \tilde{\circledast} x_-^s + (-1)^r x_-^r \tilde{\circledast} x_+^s \\ &= B(r+1,s+1)x^p + (-1)^{r+s}B(r+1,s+1)x^p + (-1)^s x_+^r \tilde{\circledast} x_-^s + \\ &\quad + (-1)^r x_-^r \tilde{\circledast} x_+^s \end{aligned} \tag{9}$$

exists for $r,s=0,1,2\dots$, where $p=r+s+1$ and B denotes the Beta function, on using Theorem 4 and the distributivity of the neutrix convolution product. Let

$$x_+^r \tilde{\circledast} x_-^s = B(r+1,s+1)\{h(r,s)x_+^p + k(r,s)x_-^p\}$$

in which h and k are undetermined coefficients. Then

$$x_-^r \tilde{\circledast} x_+^s = x_+^s \tilde{\circledast} x_-^r = B(r+1,s+1)\{h(s,r)x_+^p + k(s,r)x_-^p\} \tag{10}$$

by the commutativity of the neutrix convolution and the symmetry of the Beta function. Now equation (9) can be deduced as

$$\begin{aligned} x_+^p + (-1)^{r+s}x_-^p + (-1)^s\{h(r,s)x_+^p + k(r,s)x_-^p\} + \\ + (-1)^r\{h(s,r)x_+^p + k(s,r)x_-^p\} = 0. \end{aligned}$$

To take account of that functions x_+^p and x_-^p are linear independent we have

$$1+(-1)^s h(r,s)+(-1)^r h(s,r)=0$$
$$1+(-1)^r k(r,s)+(-1)^s k(s,r)=0,$$

or

$$H(r,s)+H(s,r)=0, \qquad K(r,s)+K(s,r)=0$$

where

$$H(r,s)=\frac{1}{2}+(-1)^s h(r,s), \qquad K(r,s)=\frac{1}{2}+(-1)^r k(r,s).$$

It is natural to put

$$H(r,s)=0 \text{ and } K(r,s)=0$$

which implies that $H(s,r)=K(s,r)=0$. Then

$$h(r,s)=k(s,r)=\frac{(-1)^{s+1}}{2}, \qquad h(s,r)=k(r,s)=\frac{(-1)^{r+1}}{2}.$$

Hence we define the convolution products $x_+^r \tilde{\circledast} x_-^s$ and $x_-^r \tilde{\circledast} x_+^s$, called 'in the sense of principal value', as follows:

$$x_+^r \tilde{\circledast} x_-^s = B(r+1,s+1)\frac{(-1)^{s+1}x_+^{r+s+1}+(-1)^{r+1}x_-^{r+s+1}}{2}, \tag{11}$$

$$x_-^r \tilde{\circledast} x_+^s = B(r+1,s+1)\frac{(-1)^{r+1}x_+^{r+s+1}+(-1)^{s+1}x_-^{r+s+1}}{2}, \tag{12}$$

for $r,s=0,1,2,\ldots$ and employ the same symbol as the neutrix convolution product.

It is easy to verify that equations (9) and (10) are true under this definition. Note that equality (11) implies equality (12) by the commutativity of the neutrix convolution, the result of the theorem follows.

THEOREM 6. *The neutrix convolution products* $x^r \tilde{\circledast} x_+^s$, $x^r \tilde{\circledast} x_-^s$, $x^r \tilde{\circledast} |x|^s$ *and* $x^r \tilde{\circledast}(|x|^s \operatorname{sgn} x)$ *in the sense of principal value exist in* $D'(\mathbb{R})$ *and*

$$x^r \tilde{\circledast} x_+^s = \frac{B(r+1,s+1)x^{r+s+1}}{2}, \tag{13}$$

$$x^r \tilde{\circledast} x_-^s = \frac{B(r+1,s+1)(-1)^{s+1}x^{r+s+1}}{2}, \tag{14}$$

$$x^r \tilde{\circledast} |x|^s = \begin{cases} B(r+1,s+1)x^{r+s+1}, & s \text{ odd} \\ 0, & s \text{ even}, \end{cases} \tag{15}$$

$$x^r \tilde{\circledast}(|x|^s \operatorname{sgn} x) = \begin{cases} 0, & s \text{ odd} \\ B(r+1, s+1)x^{r+s+1}, & s \text{ even} \end{cases} \tag{16}$$

for $r, s = 1, 2, \ldots$, *where* B *denotes the Beta function.*

PROOF. By equalities (8) and (12) we have

$$\begin{aligned} x^r \tilde{\circledast} s_+^s &= x_+^r \tilde{\circledast} x_+^s + (-1)^r x_-^r \tilde{\circledast} x_+^s \\ &= B(r+1, s+1)\frac{x_+^{r+s+1} + (-1)^r((-1)^{r+1}x_+^{r+s+1} + (-1)^{s+1}x_-^{r+s+1})}{2} \\ &= B(r+1, s+1)\frac{x_+^{r+s+1} + (-1)^{r+s+1}x_-^{r+s+1}}{2} \\ &= B(r+1, s+1)\frac{x^{r+s+1}}{2}. \end{aligned}$$

Replace x by $-x$ in (13) we obtain

$$x^r * x_-^s = \frac{B(r+1, s+1)(-1)^{s+1}x^{r+s+1}}{2}.$$

The equalities (15) and (16) follow by the addition and the subtraction, respectively, from (13) and (14) on noting that

$$|x|^s = x_+^s + x_-^s, \qquad |x|^s \operatorname{sgn} x = x_+^s - x_-^s,$$

and the neutrix convolution being distributive with respect to addition.

COROLLARY 1. *The neutrix convolution product* $x^r \tilde{\circledast}(x^s \operatorname{sgn} x)$ *in the sense of principal value exists in* $D'(\mathbb{R})$ *and*

$$x^r \tilde{\circledast}(x^s \operatorname{sgn} x) = \frac{r!s!}{(r+s+1)!}x^{r+s+1} \tag{17}$$

for $r, s = 0, 1, 2 \ldots$. *In particular, if* $r = s$ *then*

$$x^r \tilde{\circledast}(x^r \operatorname{sgn} x) = \frac{(r!)^2}{(2r+1)!}x^{2r+1} = x^r \circledast (x^r \operatorname{sgn} x). \tag{18}$$

The last equality was given by Fisher [3]. If $r = 0$ then (18) becomes

$$1 \tilde{\circledast} \operatorname{sgn} x = x = 1 \circledast \operatorname{sgn} x. \tag{19}$$

The last equality was given by Jones [6].

PROOF. Let s be odd in (15) and let s be even in (16), respectively, we have

$$\begin{aligned} x^r|x|^s &= B(r+1, s+1)x^{r+s+1}, \qquad s \text{ odd} \\ x^r(|x|^s \operatorname{sgn} x) &= B(r+1, s+1)x^{r+s+1}, \qquad s \text{ even}. \end{aligned}$$

Equality (17) follows by the combination of the above two equalities on noting that

$$|x|^s = x^s \operatorname{sgn} x, \quad s \text{ odd}, \quad |x|^s \operatorname{sgn} x = x^s \operatorname{sgn} x, \quad s \text{ even}.$$

Equalitites (18) and (19) follow immediately.

The compatibility mentioned above shows that the selection of the principal values of the neutrix convolution products are reasonable.

4. Several neutrix convolutions of distributions in $D'(\mathbb{R}^m)$

Let $r = (r_1, \ldots, r_m)$ and $s = (s_1, \ldots, s_m)$. We put

$$x^r = x_1^{r_1} \ldots x_m^{r_m}, \qquad x^{r+s+1} = x_1^{r_1+s_1+1} \ldots x_m^{r_m+s_m+1},$$
$$x_+^s = (x_1)_+^{s_1} \ldots (x_m)_+^{s_m}, \qquad x_-^s = (x_1)_-^{s_1} \ldots (x_m)_-^{s_m},$$

$$|x|^s \operatorname{sgn} x = (|x_1|^{s_1} \operatorname{sgn} x_1) \ldots (|x_m|^{s_m} \operatorname{sgn} x_m),$$

$$\|S\| = S_1 + \cdots + S_m, \qquad r! = (r_1)! \ldots (r_m)!,$$
$$(r+s+1)! = (r_1+s_1+1)! \ldots (r_m+s_m+1)!, \qquad |x|^s = |x_1|^s \ldots |x_m|^s.$$

THEOREM 7. *Let f and g be distributions in $D'(\mathbb{R})$ of the form*

$$f(x) = x_1^r \times f_1(x_2, \ldots, x_m), \qquad g(x) = x_1^s \times g_1(x_2, \ldots, x_m)$$

and suppose that the neutrix convolution product $f_1 \tilde{\circledast} g_1$ exists in $D'(\mathbb{R}^{m-1})$. Then the neutrix convolution product $f \tilde{\circledast} g$ exists in $D'(\mathbb{R}^m)$ and

$$f \tilde{\circledast} g = 0$$

for $r, s = 1, 2, \ldots$. In particular, the neutrix convolution product $x^r \tilde{\circledast} x^s$ exists in $D'(\mathbb{R}^m)$ and

$$x^r \tilde{\circledast} x^s = 0 \tag{20}$$

for $r = (r_1, \ldots, r_m)$, $s = (s_1, \ldots, s_m)$, and $r_1, \ldots, r_m, s_1, \ldots, s_m = 0, 1, 2, \ldots$.

PROOF. It follows immediately from Theorem 4 on using the decomposition theorem.

THEOREM 8. *The neutrix convolution product $x_+^r \tilde{\circledast} x_-^s$ in the sense of the principal value exists in $D'(\mathbb{R}^m)$ and*

$$x_+^r \tilde{\circledast} x_-^s = 2^{-m} \prod_{i=1}^{m} B(r_i+1, s_i+1)\{(-1)^{s_i+1}(x_i)_+^{p_i} + (-1)^{r_i+1}(x_i)_-^{p_i}\} \tag{21}$$

for $r = (r_1, \ldots, r_m)$, $s = (s_1, \ldots, s_m)$, *and* $r_1, \ldots, r_m, s_1, \ldots, s_m = 0, 1, 2, \ldots$, *where* B *denotes the Beta function and* $p_i = r_i + s_i + 1$ *for* $i = 1, \ldots, m$.

PROOF. It follows immediately from Theorem 5 on using the decomposition theorem.

THEOREM 9. *The neutrix convolution products* $x^r \tilde{\circledast} x_+^s$, $x^r \tilde{\circledast} x_-^s$, $x^r \tilde{\circledast} |x|^s$ *and* $x^r \tilde{\circledast} (|x|^s \operatorname{sgn} x)$ *in the sense of the principal value exist in* $D'(\mathbf{R}^m)$ *and*

$$x_+^r \tilde{\circledast} x_+^s = 2^{-m} \frac{r!s!}{(r+s+1)!} x^{r+s+1}, \tag{22}$$

$$x_+^r \tilde{\circledast} x_-^s = 2^{-m} (-1)^{\|s\|+m} \frac{r!s!}{(r+s+1)!} x^{r+s+1}, \tag{23}$$

$$x_+^r \tilde{\circledast} (x^s \operatorname{sgn} x) = \frac{r!s!}{(r+s+1)!} x^{r+s+1} \tag{24}$$

for $r = (r_1, \ldots, r_m)$, $s = (s_1, \ldots, s_m)$, *and* $r_1, \ldots, r_m, s_1, \ldots, s_m = 0, 1, 2, \ldots$.

PROOF. Equalities (22), (23) and (24) follow from equalities (13), (14) and (17), respectively, on using the decomposition theorem.

REFERENCES

[1] CHENG, LIN-ZHI and FISHER, B., Several products of distributions on $\mathbf{R}^m$, *Proc. Roy. Soc. London Ser. A* **426** (1989), 425–439. *MR* **91a**:46037

[2] CORPUT, J. G. VAN DER, Introduction to the neutrix calculus, *Analyse Math.* **7** (1959–60), 281–399. *MR* **23**#A1989

[3] FISHER, B., A result on the convolution of distributions, *Proc. Edinburgh Math. Soc.* (2) **19** (1975), 393–395. *MR* **53**#3607

[4] FISHER, B. and CHENG, LIN-ZHI, The product of distributions on $\mathbf{R}^m$, *Comment. Math. Univ. Carolin.* (4) **33** (1992), 605–614. *MR* **94h**:46059

[5] FISHER, B., CHENG, LIN-ZHI and JONES, D. S., A commutative neutrix convolution product of distributions on $\mathbf{R}^m$ (submitted).

[6] JONES, D. S., The convolution of generalized functions, *Quart. J. Math. Oxford Ser.* (2) **24** (1973), 145–163. *MR* **49**#1100

[7] MIKUSIŃSKI, J., Irregular operations on distributions, *Studia Math.* **20** (1961), 163–169. *MR* **23**#A4007

[8] SCHWARTZ, L., *Théorie des distributions*, Tome I, Actualités Sci. Ind., no. 1091 = Publ. Inst. Math. Univ. Strasbourg, 9, Hermann & Cie., Paris, 1950. *MR* **12**–31; Tome II, Actualités Sci. Ind., no. 1122 = Publ. Inst. Math. Univ. Strasbourg, 10, Hermann & Cie., Paris, 1951. *MR* **12**–833

(Received May 3, 1991)

DEPARTMENT OF BASIC COURSE
HUAZHONG AGRICULTURAL UNIVERSITY
WUHAN 430070
PEOPLE'S REPUBLIC OF CHINA

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 243–259

EXPONENTIAL SERIES IN THE PROBLEMS OF INITIAL AND POINTWISE CONTROL OF A RECTANGULAR VIBRATING MEMBRANE

S. A. AVDONIN, S. A. IVANOV and I. JOÓ

1. Introduction

In the present paper we consider vibrations of a rectangular homogeneous membrane with Dirichlet boundary conditions in any finite time. We prove that arbitrary trajectories of any finite number of points of the membrane can be obtained choosing appropriate initial data T (Theorem 1(a)). It is shown that this problem is in some sense dual to a pointwise control problem. The reachability set of the system under any pointwise control in finitely many points is proved to be not dense in the phase space (Theorem 2(b)).

Our approach is based on the reduction of these problems (with the help of the Fourier method) to the investigation of the family $\mathcal{E}$ of the exponential vector-functions $\eta_{mn}\exp(\pm i\omega_{mn}t)$ where the vectors $\eta_{mn}\in\mathbb{C}^N$ are expressed by the eigenfunctions of the Laplace operator and the ω_{mn} are the eigenfrequencies of the membrane. It is proved that for any $T>0$ and for any integer $r\geqq 0$ there exists a subfamily $\mathcal{E}_r\subset\mathcal{E}$ which forms a Riesz basis in the Sobolev space $H^r(0,T;\mathbb{C}^N)$ (see Theorem 3).

The important role of exponential families in the control theory of d.p.s. (distributed parameter systems) is widely known (see e.g. the review of Russell [1]). The minimality and the Riesz basis property of exponential families are the most essential tools in the investigations of such kinds of problems. Riesz basis criterion for exponential families obtained by B. S. Pavlov [2], cf. also [9]. Riesz bases from exponentials in the Sobolev spaces $H^r(0,T)$ have been investigated by Russell in [4]. In the problems where the system is controlled in finitely many points, exponential families in the space of vector-functions $L^2(0,T;\mathbb{C}^N)$ and $H^r(0,T;\mathbb{R}^N)$ naturally arise. Properties of such families and their applications in controllability problems of d.p.s. have been considered in [6–11], [16–19], [21]. Some results of these works are used in the present paper.

Some results of the present paper have been obtained during the visit of I. Joó in Leningrad, May 1989. Due to the difficulties in private contacts and in changing information these are prepared for publication only by now.

1991 *Mathematics Subject Classification.* Primary 49B22; Secondary 49B27.
Key words and phrases. Wave equation, Riesz bases from exponentials.

2. Statement of the results

Let

$$\Omega = (0,a) \times (0,b)$$

be a rectangle and denote A the operator $-\Delta$ with domain

$$D(A) := H^2(\Omega) \cap H_0^1(\Omega).$$

Let further

$$\varphi_{mn}(p) := \frac{2}{\sqrt{ab}} \sin \frac{\pi m}{a} x \sin \frac{\pi n}{b} y, \qquad p := (x,y);\ m,n \in \mathbb{N}.$$

These functions form in $L^2(\Omega)$ a complete orthonormal sequence and are eigenfunctions of the operator A:

$$A\varphi_{mn} = \omega_{mn}^2 \varphi_{mn}, \qquad \omega_{mn} := \sqrt{\left(\frac{\pi}{a}m\right)^2 + \left(\frac{\pi}{b}n\right)^2}.$$

Introduce the spaces $W_\beta, \beta \in \mathbb{R}$ as follows:

$$W_\beta := \left\{ w = \sum_{m,n} d_{mn}\varphi_{mn} : \|w\|_\beta^2 := \sum_{m,n} |d_{mn}|^2 \omega_{mn}^{2\beta} < \infty \right\}.$$

In case $\beta > 0$ the space W_β is the domain of the operator A^β. If $\beta < 0$ then W_β is the dual of the space $W_{-\beta}$ by the duality

$$\langle g, h \rangle = \sum_{m,n} g_{mn}\overline{h_{mn}}, \quad g = \sum g_{mn}\varphi_{mn} \in W_\beta, \quad h = \sum h_{mn}\varphi_{mn} \in W_{-\beta}.$$

We know [12] that $W_1 = H_0^1$, $W_{-1} = H^{-1}$. Denote

$$\mathcal{W}_\beta := W_\beta \oplus W_{\beta-1}.$$

Fix N distinct points $p_1, \dots, p_N \in \Omega$, $p_k = (x_k, y_k)$ and let $T > 0$ be arbitrary. Consider the initial-boundary problem

$$u_{tt} = \Delta u \quad \text{in } \Omega \times (0,T),$$

$$u\Big|_{\partial\Omega \times (0,T)} = 0, \tag{1}$$

$$u\Big|_{t=0} = u_0, \qquad u_t\Big|_{t=0} = u_1,$$

and let

$$\Phi(t, u_0, u_1) := (u(p_1,t), \dots, u(p_N,t)).$$

THEOREM 1. *Let $r \geqq 0$ be an integer. Then*

(a) *For any vector function $F \in H^r(0,T;\mathbf{R}^N)$ there exists an initial state $(u_0,u_1) \in \mathcal{W}_r$ such that $\Phi(t,u_0,u_1) = F(t)$.*

(b) *The dimension of the set of $(u_0,u_1) \in \mathcal{W}_r$ with $\Phi(t,u_0,u_1) = 0$, $0 \leqq t \leqq T$ is finite.*

To formulate Theorem 2 consider a membrane controlled at the points p_k, $i \leqq k \leqq N$, where the motion satisfies

$$v_{tt}(p,t) = \Delta v(p,t) + \sum_{k=1}^{N} \delta(p-p_k) f_k(t) \quad \text{in} \quad \Omega \times (0,T),$$

$$v\Big|_{\partial\Omega\times(0,T)} = 0, \tag{2}$$

$$v\Big|_{t=0} = v_t\Big|_{t=0} = 0.$$

THEOREM 2.

(a) *If $f_k \in L^2(0,T)$, $k = 1,\dots,N$ then $(v,v_t) \in C\left([0,T];\mathcal{W}_{\frac{1}{2}}\right)$.*

(b) *For the reachability set*

$$R(T) := \{(v(.\,,T), v_t(.\,,T)) : f_k \in L^2(0,T),\ k = 1,\dots,N\}$$

is not dense in $\mathcal{W}_{\frac{1}{2}}$ and codim $R(T) = \infty$.

3. The proof of Theorem 1

We use Fourier method and the theory of vector exponentials. The initial data $u_0 \in W_r$, $u_1 \in W_{r-1}$ have the expansions

$$u_0 = \sum a_{mn}\varphi_{mn}, \qquad u_1 = \sum b_{mn}\varphi_{mn} \tag{3}$$

$$\sum |a_{mn}|^2 \omega_{mn}^{2r} < \infty, \qquad \sum |b_{mn}|^2 \omega_{mn}^{2r-2} < \infty. \tag{4}$$

By the Fourier method we obtain the following expansion of the solution $u(p,t)$ of (1):

$$u(p,t) = \sum_{m,n} \left[a_{mn} \cos \omega_{mn} t + \frac{b_{mn}}{\omega_{mn}} \sin \omega_{mn} t \right] \varphi_{mn}(p).$$

Using the vector functions

$$c_{mn}(t) := \eta_{mn} \cos \omega_{mn} t, \qquad s_{mn}(t) = \eta_{mn} \sin \omega_{mn} t,$$

$$\eta_{mn} := \omega_{mn}^{-r} \left(\varphi_{mn}(p_1), \dots, \varphi_{mn}(p_N)\right) \in \mathbb{R}^N$$

we get the following expression for $\Phi(t)$:

$$\Phi(t, u_0, u_1) = \sum_{m,n} \left[a_{mn}\omega_{mn}^r c_{mn}(t) + b_{mn}\omega_{mn}^{r-1} s_{mn}(t)\right]. \tag{5}$$

Introduce the notation

$$\mathcal{F} := \{c_{mn}, s_{mn}\}_{m,n\in\mathbb{N}}.$$

LEMMA 1. *If there exists a subset $M \subset \mathbb{N}\times\mathbb{N}$ such that the family $\mathcal{F}_r := \{c_{mn}, s_{mn}\}_{(m,n)\in M} \subset \mathcal{F}$ forms a Riesz basis in $H^r(0,T;\mathbb{R}^N)$, then statement* (a) *of Theorem 1 follows.*

PROOF. Expand an arbitrary function $F \in H^r(0,T;\mathbb{R}^N)$ in the basis $\mathcal{F}_r$:

$$F(t) = \sum_{(m,n)\in M} \left[f_{mn}^0 c_{mn}(t) + f_{mn}^1 s_{mn}(t)\right], \tag{6}$$

$$\sum_{(m,n)\in M} \left[|f_{mn}^0|^2 + |f_{mn}^1|^2\right] < \infty. \tag{7}$$

In (5) we define

$$a_{mn} = b_{mn} = 0 \qquad \text{if } (m,n) \notin M, \tag{8}$$

$$a_{mn} = f_{mn}^0 \omega_{mn}^{-r}, \qquad b_{mn} = f_{mn}^1 \omega_{mn}^{1-r} \qquad \text{if } (m,n) \in M. \tag{9}$$

Then (4) holds and the initial data (u_0, u_1) defined by (3) belong to $\mathcal{W}_r$. Finally, by (5), $\Phi(t, u_0, u_1) = F(t)$ which completes the proof. □

For the construction of the subsystem $\mathcal{F}_r$ it is convenient to introduce the following family $\mathcal{E}$ of vector exponentials:

$$\mathcal{E} := \left\{e_{mn}^{\pm}\right\}_{m,n\in\mathbb{N}}, \qquad e_{mn}^{\pm}(t) := \eta_{mn} e^{\pm i\omega_{mn}t}.$$

LEMMA 2. *A family*

$$\mathcal{F}_* = \{e_{mn}, s_{mn}\}_{(m,n)\in M}, \qquad M \subset \mathbb{N}\times\mathbb{N}$$

forms a Riesz basis in $H^r(0,T;\mathbb{R}^N)$ if and only if the family

$$\mathcal{E}_* := \{e_{mn}^{\pm}\}_{(m,n)\in M}$$

forms a Riesz basis in $H^r(0,T;\mathbb{C}^N)$.

PROOF. $\mathcal{F}_*$ is Riesz basis in $H^r(0,T;\mathbf{R}^N)$ if and only if it is Riesz basis in $H^r(0,T;\mathbf{C}^N)$. By a classical theorem of N. K. Bari [3] a system (φ_n) is Riesz basis in a Hilbert space H if and only if (φ_n) is complete and there exist constants $0 < c_1 < c_2 < \infty$ satisfying

$$c_1 \sum |c_n|^2 \leqq \left\| \sum c_n \varphi_n \right\|_H^2 \leqq c_2 \sum |c_n|^2 \tag{10}$$

for every finite sum with arbitrary (complex) coefficients. Since the concrete value of the constants is not interesting for us, we use the abbreviation

$$\sum |c_n|^2 \asymp \left\| \sum c_n \varphi_n \right\|_H^2$$

instead of (10). So $\mathcal{F}_*$ is Riesz basis if and only if it is complete and

$$\sum_{(m,n)\in M} \left(|a_{mn}^+|^2 + |a_{mn}^-|^2\right) \asymp \left\| \sum_{(m,n)\in M} (a_{mn}^+ c_{mn} + a_{mn}^- s_{mn}) \right\|_H^2 \tag{11}$$

holds for every finite sum; $\mathcal{E}_*$ is Riesz basis if and only if it is complete and

$$\sum_{(m,n)\in M} \left(|\tilde{a}_{mn}^+|^2 + |\tilde{a}_{mn}^-|^2\right) \asymp \left\| \sum_{(m,n)\in M} (\tilde{a}_{mn}^+ e^+{}_{mn} + \tilde{a}_{mn}^- e^-{}_{mn}) \right\|^2. \tag{12}$$

By the Euler formulae $e_{mn}^{\pm} = c_{mn} \pm i s_{mn}$ we see that $\mathcal{F}_*$ is complete if and only if $\mathcal{E}_*$ is. Furthermore the linear hull of $\mathcal{F}_*$ and $\mathcal{E}_*$ is the same, namely

$$\sum_{(m,n)\in M} (a_{mn}^+ c_{mn} + a_{mn}^- s_{mn}) = \sum_{(m,n)\in M} (\tilde{a}_{mn}^+ e^+{}_{mn} + \tilde{a}_{mn}^- e^-{}_{mn})$$

if

$$a_{mn}^+ = \tilde{a}_{mn}^+ + \tilde{a}_{mn}^-, \qquad a_{mn}^- = i(\tilde{a}_{mn}^+ - \tilde{a}_{mn}^-)$$

and then

$$|a_{mn}^+|^2 + |a_{mn}^-|^2 = 2\left(|\tilde{a}_{mn}^+|^2 + |\tilde{a}_{mn}^-|^2\right),$$

which shows that (11) and (12) are also equivalent. The proof is complete. □

4. Existence of a basis family

We shall prove

THEOREM 3. *For any nonnegative integer r and any $T > 0$ there exists a family $\mathcal{E}_r \subset \mathcal{E}$ which forms a Riesz basis in $H^r(0,T;\mathbf{C}^N)$.*

We begin the proof with

LEMMA 3 ([18, 19]). *The family* $(\eta_{mn})_{m,n=1}^{2N}$ *is complete in* $\mathbb{C}^N$.

Choose a basis $\{\eta^j\}_{j=1}^N \subset \{\eta_{mn}\}_{m,n=1}^{2N}$ in $\mathbb{C}^N$. Then

LEMMA 4 ([18, 19]). *For any* $\mathcal{E} > 0$ *and* $T > 0$ *there exists a family* $\tilde{\mathcal{E}} \subset \mathcal{E}$ *which can be represented in the form*

$$\tilde{\mathcal{E}} = \mathcal{E}^1 \cup \cdots \cup \mathcal{E}^N, \qquad \mathcal{E}^j := \left\{ \eta_m^j e^{\pm i\lambda'_{mj} t} \right\}_{m=R}^{\infty},$$

where the following inequalities hold:

$$\|\eta_m^j - \eta^j\| < \varepsilon, \tag{13}$$

$$|\lambda_{mj} - \frac{2\pi}{T} m| < \varepsilon \tag{14}$$

for all $j = 1, \ldots, N$; $m = R, R+1, \ldots$.

LEMMA 5. *Let the family*

$$\Xi := \left\{ \mathcal{H}_n e^{i\mu_n t} \right\}_{n \in \mathbb{N}, \mathcal{H}_n \in \mathbb{C}^N}$$

be a Riesz basis in $L^2(0, T; \mathbb{C}^N)$. *Then there exists* $\varepsilon > 0$ *such that any family*

$$\tilde{\Xi} := \left\{ \tilde{\mathcal{H}}_n e^{i\tilde{\mu}_n t} \right\}_{n \in \mathbb{N}}$$

with

$$\|\mathcal{H}_n - \tilde{\mathcal{H}}_n\| < \varepsilon, \qquad |\mu_n - \tilde{\mu}_n| < \varepsilon, \quad n \in \mathbb{N}$$

is necessarily a Riesz basis in $L^2(0, T; \mathbb{C}^N)$.

PROOF. Perturbate first the vectors only. In [8], Ch. III, §5.2 the authors proved that for small $\varepsilon_1 > 0$ there exist constants $0 < c < C < \infty$ such that the family $\left\{ \tilde{\mathcal{H}}_n e^{i\mu_n t} \right\}$ is a Riesz basis in $L^2(0, T; \mathbb{C}^N)$ and

$$c \sum |d_n|^2 \leqq \left\| \sum d_n \tilde{\mathcal{H}}_n e^{i\mu_n t} \right\|_{L^2}^2 \leqq C \sum |d_n|^2 \tag{15}$$

whenever $\|\mathcal{H}_n - \tilde{\mathcal{H}}_n\| < \varepsilon_1$ for all n; the constants c, C do not depend on $(\tilde{\mathcal{H}}_n)$.

Next perturbate the spectrum. In this part we use the estimation technique of Duffin and Eachus [20]. Fix $\varepsilon_1 > 0$ and define an operator L on the linear hull of the family $\left\{ \tilde{\mathcal{H}}_n e^{i\mu_n t} \right\}$ by the formula

$$L\tilde{\mathcal{H}}_n e^{i\mu_n t} := \tilde{\mathcal{H}}_n e^{i\tilde{\mu}_n t}.$$

Define

$$\delta_n := \tilde{\mu}_n - \mu_n, \qquad |\delta_n| \leqq \varepsilon_2$$

for all n, and consider a finite linear combination

$$h := \sum d_n \tilde{\mathcal{H}}_n e^{i\mu_n t}.$$

Expanding $e^{i\delta_n t}$ in a Taylor series we obtain

$$\begin{aligned}
\|(I-L)h\| &= \left\| \sum d_n \tilde{\mathcal{H}}_n e^{i\mu_n t}(1 - e^{i\delta_n t}) \right\| = \\
&= \left\| \sum_n d_n \tilde{\mathcal{H}}_n e^{i\mu_n t} \sum_{k=1}^{\infty} \frac{(i\delta_n)^k}{k!} t^k \right\| = \\
&= \left\| \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_n d_n \tilde{\mathcal{H}}_n e^{i\mu_n t} (i\delta_n)^k \right\| \leqq \\
&\leqq \sum_{k=1}^{\infty} \frac{T^k}{k!} \left\| \sum_n d_n \tilde{\mathcal{H}}_n (i\delta_n)^k e^{i\mu_n t} \right\| \leqq \\
&\leqq \sqrt{C} \sum_{k=1}^{\infty} \frac{T^k}{k!} \left(\sum_n |d_n (i\delta_n)^k|^2 \right)^{\frac{1}{2}} \leqq \\
&\leqq \sqrt{C} \sum_{k=1}^{\infty} \frac{(T\varepsilon_2)^k}{k!} \left(\sum_n |d_n|^2 \right)^{\frac{1}{2}} = \\
&= \sqrt{C}\,(e^{T\varepsilon_2} - 1) \left(\sum_n |d_n|^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Comparing with (15) we see that

$$\|(I-L)h\| \leqq \sqrt{\frac{C}{c}}\,(e^{T\varepsilon_2} - 1)\,\|h\|.$$

For sufficiently small $\varepsilon_2 > 0$ it implies $\|I - L\| < 1$ and then L can be continued to an isomorphism; hence we showed the statement of Lemma 5 for $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$. □

LEMMA 6. *For any $T > 0$ there exists a family $\tilde{\mathcal{E}} \subset \mathcal{E}$ such that*

$$\Xi := \left\{ \eta^j e^{\frac{2\pi i n t}{T}} \right\}_{j=1,|n|<R}^{N} \cup \tilde{\mathcal{E}}$$

forms a Riesz basis in $L^2(0,T;\mathbb{C}^N)$ (namely the family $\tilde{\mathcal{E}}$ constructed in Lemma 4 is appropriate).

PROOF. The family

$$\Xi := \left\{ \eta^j e^{\frac{2\pi i n t}{T}} \right\}_{j=1,n\in\mathbb{Z}}^{N}$$

is of course a Riesz basis in $L^2(0,T;\mathbb{C}^N)$. For small $\varepsilon > 0$ the estimations (13), (14) show the statement by Lemma 5. □

PROPOSITION 1 ([18]). *If the family $\{\mathcal{H}_n e^{i\mu_n t}\}_1^\infty$ is complete in $L^2(0,T;\mathbb{C}^N)$ then for any $R > 0$ and $0 < T_1 < T$ the family $\{\mathcal{H}_n e^{i\mu_n t}\}_{n=R}^\infty$ is complete in $L^2(0,T_1;\mathbb{C}^N)$.*

PROOF OF THEOREM 3 IN THE CASE $r = 0$. As Lemma 6 and Proposition 1 show, the family $\mathcal{E}$ is complete in $L^2(0,T;\mathbb{C}^N)$ for any $T > 0$. Again by Lemma 6 the family $\tilde{\mathcal{E}}$ is also a Riesz basis in its closed linear hull $\vee\tilde{\mathcal{E}}$ (which has finite codimension in L_2). Since $\mathcal{E}$ is complete, we can take $e \in \mathcal{E} \setminus \vee\tilde{\mathcal{E}}$. Then $\{e\} \cup \tilde{\mathcal{E}}$ is again a Riesz basis in $\vee(e,\tilde{\mathcal{E}})$ whose codimension is smaller. Starting from $\{e\} \cup \tilde{\mathcal{E}}$ instead of $\tilde{\mathcal{E}}$ we can diminish further the codimension by adding elements of $\mathcal{E}$. In finitely many steps a complete Riesz basis can be constructed in this way which finishes the proof of the case $r = 0$.

REMARK. In [19] I. Joó proved the existence of a basis from $\mathcal{E}$ via the following

PROPOSITION 2 ([19]). *Let $\{\mathcal{H}_n e^{i\mu_n t}\}_{n\in\mathbb{Z}}$ be a Riesz basis in $L^2(0,T;\mathbb{C}^N)$ and let $\mu_0' \in \mathbb{C}^N$, $\mu_0' \neq \mu_n$. Then the new family*

$$\left\{ \mathcal{H}_0 e^{i\mu_0' t}, \mathcal{H}_n e^{i\mu_n} : n \in \mathbb{Z} \setminus \{0\} \right\}$$

is also a Riesz basis in $L^2(0,T;\mathbb{C}^N)$.

For the proof of Theorem 3 in the case $r > 0$ we need the generalization of the classical (one dimensional) theorem of Russell [4] stating that if the family $\{e^{i\mu_n t}\}$ forms a Riesz basis in $L\ (0,T)$ then the family

$$\left\{ (1+|\mu_n|)^{-M} e^{i\mu_n t} \right\}_{n=1}^{\infty} \cup \left\{ e^{i\nu_j t} \right\}_{j=1}^{M}; \qquad \{\nu_j\}_1^M \cap \{\mu_n\}_1^\infty = \emptyset$$

forms a Riesz basis in $H^r(0,T)$. The generalization for vector exponentials is given in the book [8] Ch II, § 5.3, and by another method in [19] by the third author (it was obtained at the same time, independently, with different proof). Namely

PROPOSITION 3 ([8, 19]). *Suppose that*

(i) *the family* $\{\mathcal{H}_n e^{i\mu_n t}\}_{n=1}^{\infty}$ *forms a Riesz basis in* $L^2(0,T;\mathbb{C}^N)$ *(and* $\mu_n \in \mathbb{C}^N$*),*

(ii) *the family* $\{\mathcal{H}_{lj} e^{i\mu_{lj} t} : j=1,\dots,N;\ l=1,\dots,r\}$ *is linearly independent and* $\{\mu_{lj}\} \cap \{\mu_n\} = \emptyset$*,*

(iii) *there exists a matrix polynomial* $P(z) = z^r + A_1 z^{r-1} + \cdots + A_r$ *such that the zeros of* $\det P(z)$ *coincide with* $\{\mu_{lj}\}$*, the zeros are semi-simple (the geometrical and algebraical multiplicities are equal) and* $P(\mu_{lj})\mathcal{H}_{lj} = 0$*.*

Then the family

$$\left\{\frac{1}{(1+|\mu_n|)^N}\mathcal{H}_n e^{i\mu_n t}\right\}_{n=1}^{\infty} \cup \left\{\mathcal{H}_{lj} e^{i\mu_{lj} t}\right\}_{j=1,l=1}^{N,\ r} \tag{16}$$

forms a Riesz basis in $H^r(0,T;\mathbb{C}^N)$*.*

REMARK. In contrast to the scalar case the additional condition (iii) is necessary to the Riesz basis property as the following example shows. Consider the orthonormal basis

$$\Xi_0 := \left\{\frac{1}{\sqrt{2\pi}} h_j e^{int}\right\}_{j=1,n\in\mathbb{Z}}^{N}$$

in $L^2(0,2\pi;\mathbb{C}^N)$, where $\{h_j\}_{j=1}^N$ is an orthonormal basis in $\mathbb{C}^N$. The family $\Xi_0 \cup \{h_1 e^{i\nu_n t}\}_{n=1}^N$ satisfies (i) and (ii), however, it is neither minimal nor complete in $H^1(0,2\pi,;\mathbb{C}^N)$. We can verify also directly that property (iii) fails. Indeed, for $r=1$ the polynomial $P(z)$ has the form $z + A_1$ and the equalities $(\nu_n I + A_1)h_1 = 0$ cannot be fulfilled for different values of ν_n.

LEMMA 7. *Let* $\nu_l \in \mathbb{R}$, $l=1,\dots,r$ *be distinct numbers and fix a basis* $\{\eta^j\}_{j=1}^N$ *in* $\mathbb{C}^N$*. Then for sufficiently small* $\epsilon > 0$ *the system* $\{\mathcal{H}_{lj} e^{i\mu_{lj} t}\}_{j=1,l=1}^{N,\ r}$ *satisfies condition* (iii) *of Proposition 3 whenever*

$$|\mu_{lj} - \nu_l| < \epsilon, \qquad \|\mathcal{H}_{lj} - \eta^j\| < \epsilon, \qquad l=1,\dots,r;\ j=1,\dots,N. \tag{17}$$

PROOF. We use induction on r.

(a) The case $r=1$.

For small $\epsilon > 0$ the vectors $\mathcal{H}_{1j}$ give a basis in $\mathbb{C}^N$. Denote B_1^ϵ the matrix whose eigenvectors are $\mathcal{H}_{1j}$ with eigenvalues μ_{1j}. Now it is trivial that the matrix polynomial $P_1(z) = z - B_1^\epsilon$ satisfies (iii).

(b) The case $r=2$.

As ϵ approaches zero, we have

$$B_1^\epsilon = \nu_1 I + \overline{\overline{o}}(1) \qquad \mathcal{H}_{2j} = \eta^j + \overline{\overline{o}}(1), \qquad \mu_{2j} = \nu_2 + \overline{\overline{o}}(1).$$

$$\tilde{\mathcal{H}}_{2j} := P_1(\mu_{2j})\mathcal{H}_{2j} = \left\{(\nu_2 - \nu_1)I + \bar{\bar{o}}(1)\right\}\left(\eta^j + \bar{\bar{o}}(1)\right) = (\nu_2 - \nu_1)\eta^j + \bar{\bar{o}}(1).$$

Consequently, for small ϵ the vectors $\tilde{\mathcal{H}}_{2j}$ form a basis in $\mathbb{C}^N$. Denote B_2^ϵ the matrix with eigenvectors $\tilde{\mathcal{H}}_{2j}$ and eigenvalues μ_{2j} and let $P_2(z) := (z - B_2^\epsilon)P_1(z)$. Then

$$P_2(\mu_{1j})\mathcal{H}_{1j} = (\mu_{1j}I - B_2^\epsilon)\,P_1(\mu_{1j})\mathcal{H}_{1j} = 0,$$
$$P_2(\mu_{2j})\mathcal{H}_{2j} = (\mu_{2j}I - B_2^\epsilon)\,\tilde{\mathcal{H}}_{2j} = 0,$$

hence (iii) fulfills.

(c) In case $r \geqq 3$ we use induction

$$P_r(z) := (z - B_r^\epsilon)P_{r-1}(z),$$

$\tilde{\mathcal{H}}_{rj} = P_{r-1}(\mu_{rj})\mathcal{H}_{rj}$ are the corresponding eigenvectors with eigenvalues μ_{rj}. The details are similar to case (b).

PROOF OF THEOREM 3 FOR $r \geqq 1$. Define $\tilde{T} := 2T$, $\nu_l := \frac{\pi}{T}(2l+1)$, $l = 1, \ldots, r$. Take a basis $\{\eta^j\}_{j=1}^N$ by Lemma 3 and fix $\epsilon > 0$ corresponding to Lemma 7. Now by Lemma 4 there exists a subsystem

$$\bar{\mathcal{E}}_{2T} = \left\{\eta_m^j e^{\pm i\lambda_{mj}t}\right\}_{j=1,\ m=R}^{N}$$

satisfying

$$\left|\lambda_{mj} - \frac{\pi}{T}m\right| < \epsilon, \qquad \|\eta_m^j - \eta^j\| < \epsilon. \tag{18}$$

Consider all elements of $\bar{\mathcal{E}}_{2T}$ with even m; as we have seen in the proof of the case $r = 0$, we can join finitely many elements with odd index m to obtain a Riesz basis in $L^2(0,T;\mathbb{C}^N)$. Let $Q \geqq R$ be so large that all used odd indices m are $\leqq 2Q+1$. Introduce

$$\mu_{lj} := \lambda_{2(l+Q)+1,j} - \frac{\pi}{T}2Q \qquad l = 1, \ldots, r$$
$$\mathcal{H}_{lj} := \eta^j_{2(l+Q)+1} \qquad j = 1, \ldots, N.$$

We apply Lemma 7; the estimates (18) show that property (iii) holds. Hence, using Proposition 3, we can extract a Riesz basis in $H^r(0,T;\mathbb{C}^N)$ from the elements of $\bar{\mathcal{E}}_{2T}$. Theorem 3 is completely proved. □

PROOF OF THEOREM 1. Lemmas 1,2 and Theorem 3 imply statement (a) of Theorem 1. To see (b) take the Riesz basis $\mathcal{E}_r \subset \bar{\mathcal{E}}_{2T}$ constructed in the proof of Theorem 3. Clearly, the set $\mathcal{E} \setminus \mathcal{E}_r$ is infinite. Denote $\mathcal{F}_r$ the set corresponding to $\mathcal{E}_r$:

$$\mathcal{F}_r := \{c_{mn}, s_{mn}\}_{(m,n)\in M}, \qquad M := \{(m,n) : e^{\pm}_{mn} \in \mathcal{E}_r\}.$$

By Lemma 2, $\mathcal{F}_r$ is a Riesz basis in $L^2(0,T;\mathbb{C}^N)$. Take an element $c_{kl} \notin \mathcal{F}_r$ and expand it in $\mathcal{F}_r$:

$$c_{kl}(t) = \sum_{(m,n)\in M} [\alpha_{mn} c_{mn}(t) + \beta_{mn} s_{mn}(t)]. \tag{19}$$

In analogy with (8) and (9), define

$$\begin{aligned} a_{mn} &:= b_{mn} := 0 \qquad \text{if } (m,n) \notin M \cup \{(k,l)\}, \\ a_{mn} &:= \alpha_{mn}\omega_{mn}^{-1}, \qquad b_{mn} := \beta_{mn}\omega_{mn}^{-1} \qquad \text{if } (m,n) \in M, \\ a_{kl} &:= -1, \qquad b_{kl} := 0. \end{aligned}$$

Then by (5)

$$\begin{aligned} \Phi(t, u_0, u_1) &= 0, \qquad t \in [0,T], \\ u_0 &:= \sum a_{mn}\varphi_{mn} \in W_r, \\ u_1 &:= \sum b_{mn}\varphi_{mn} \in W_{r-1}. \end{aligned}$$

There are infinitely many elements $c_{kl} \notin \mathcal{F}_r$, and the corresponding coefficient sequences $\{a_{mn}, b_{mn}\}$ are linearly independent (it is enough to consider the indices $(m,n) \notin M$). Consequently, the space of pairs $(u_0, u_1) \in W_r$ keeping $\Phi(t, u_0, u_1)$ to be zero, is infinite. The proof of Theorem 1 is complete. □

5. The proof of Theorem 2

For the proof of the statement (a) we use the transposition method [12]. Denote $\omega(p,t)$ the solution of the initial-boundary problem

$$\begin{aligned} &\omega_{tt}(p,t) = \Delta\omega(p,t), \qquad t \in (0,T'), \ p \in \Omega, \\ &\omega\Big|_{t=T} = \omega_0, \qquad \omega_t\Big|_{t=T} = \omega_1, \qquad \omega\Big|_{\partial\Omega\times(0,T)} = 0. \end{aligned} \tag{20}$$

LEMMA 8. *If $\omega_0 \in W_{\frac{1}{2}}$, $\omega_1 \in W_{-\frac{1}{2}}$ then for any point $p_0 \in \Omega$ we have $\omega(p_0, .) \in L^2(0,T)$ and*

$$\|\omega(p_0, .)\|^2_{L^2(0,T')} \leqq c(T') \left[\|\omega_0\|^2_{\frac{1}{2}} + \|\omega_1\|^2_{-\frac{1}{2}} \right]. \tag{21}$$

The constant $c(T')$ remains bounded if T is bounded.

PROOF. Take the expansions

$$\omega_0 = \sum a_{mn}\varphi_{mn}, \qquad \omega_1 = \sum b_{mn}\varphi_{mn},$$

$$\|\omega_0\|^2_{\frac{1}{2}} = \sum |a_{mn}|^2 \omega_{mn} < \infty, \tag{22}$$

$$\|\omega_1\|^2_{-\frac{1}{2}} = \sum |b_{mn}|^2 \omega_{mn} < \infty.$$

Then the Fourier method gives

$$\omega(p_0, t) = \sum \left[a_{mn} \cos \omega_{mn}(T' - t) + \frac{b_{mn}}{\omega_{mn}} \sin \omega_{mn}(T' - t) \right] \varphi_{mn}(p_0). \tag{23}$$

Recall the following result of Y. Meyer:

PROPOSITION 4. *Let $\mu_{-m} = -\mu_m$, $q_{-m} = q_m$ for $m \in \mathbb{Z}$. Then the estimate*

$$\left\| \sum_{m \in \mathbb{Z}} a_m q_m e^{i\mu_m t} \right\|^2_{L^2(0,T)} \leqq c(T) \sum |a_m|^2$$

holds for all sequences $\{a_m\}$ if and only if

$$\sup_{l \in \mathbb{N}} \sum_{l \leqq \mu_m \leqq l+1} q_m^2 < \infty. \qquad \square$$

We transform the expansion (23) in exponential form:

$$\omega(p_0, t) = \sum \left[\alpha^+_{mn} \omega_{mn}^{-\frac{1}{2}} e^{i\omega_{mn} t} + \alpha^-_{mn} \omega_{mn}^{-\frac{1}{2}} e^{-i\omega_{mn} t} \right] \varphi_{mn}(p_0),$$

$$\sum \left(|\alpha^+_{mn}|^2 + |\alpha^-_{mn}|^2 \right) \asymp \|\omega_0\|^2_{\frac{1}{2}} + \|\omega_1\|^2_{-\frac{1}{2}}.$$

By Proposition 4 the necessary and sufficient condition of (21) is:

$$\sup_{l \in \mathbb{N}} \sum_{l \leqq \omega_{mn} \leqq l+1} (\varphi_{mn}(p_0))^2 \omega_{mn}^{-1} < \infty. \tag{24}$$

The numbers $|\varphi_{mn}(p_0)|$ are bounded by $2/\sqrt{ab}$, hence it remains to check that

$$\sup_{l \in \mathbb{N}} \sum_{l \leqq \sqrt{m^2+n^2} \leqq l+1} \frac{1}{\sqrt{m^2+n^2}} < \infty,$$

or equivalently

$$\sup_{l \in \mathbb{N}} \frac{1}{l} \sum_{l \leqq \sqrt{m^2+n^2} \leqq l+1} 1 < \infty.$$

By symmetry it is enough to estimate the pairs (m,n) with $n \leqq m$, $l \leqq \sqrt{m^2+n^2} \leqq l+1$. In this case to any n there exists only $\underline{\underline{O}}(1)$ many m satisfying these inequalities, hence

$$\sum_{\substack{n \leqq m \\ l \leqq \sqrt{m^2+n^2} \leqq l+1}} 1 \leqq cl,$$

which completes the proof of Lemma 8. □

Returning to the proof of Theorem 2, we apply the Fourier method to (2) to obtain

$$\begin{aligned} v(p,t) &= \sum v_{mn}(t)\varphi_{mn}(p), \\ v_{mn}(t) &:= \sum_{j=1}^{N} \int_0^t \varphi_{mn}(p_j) f_j(\tau) \frac{\sin \omega_{mn}(t-\tau)}{mn} d\tau. \end{aligned} \tag{25}$$

It follows that

$$\dot{v}_{mn}(t) = \sum_{j=1}^{N} \int_0^t \varphi_{mn}(p_j) f_j(\tau) \cos \omega_{mn}(t-\tau) d\tau. \tag{26}$$

Fix any $T' \in [0,T]$; then by (22) and (23) we get

$$\begin{aligned} \langle v(.,T'), \omega_1 \rangle + \langle v_t(.,T'), \omega_0 \rangle = \\ = \sum_{j=1}^{N} \sum_{m,n} \int_0^{T'} \left[b_{mn} \varphi_{mn}(p_j) f_j(t) \frac{\sin \omega_{mn}(T'-t)}{\omega_{mn}} + \right. \\ \left. + a_{mn} \varphi_{mn}(p_j) f_j(t) \cos \omega_{mn}(T'-t) \right] dt \\ = \sum_{j=1}^{N} \int_0^{T'} f_j(t) \omega(p_j, t) dt. \end{aligned} \tag{27}$$

We know from Lemma 8 that the linear mapping $\{\omega_0, \omega_1\} \to \omega(p_0, .)$ is continuous from $\mathcal{W}_{\frac{1}{2}}$ to $L^2(0,T')$ hence the right-hand side of (27) is a continuous functional on $\mathcal{W}_{\frac{1}{2}}$; consequently

$$v(.,T') \in W_{\frac{1}{2}}, \qquad v_t(.,T') \in W_{-\frac{1}{2}}$$

and

$$\left\| \{v(.,T'), v_t(.,T')\} \right\|_{\mathcal{W}_{\frac{1}{2}}} \leqq c(T') \|f\|^2_{L^2(0,T';\mathbb{R}^N)}, \qquad f := \{f_j\}_{j=1}^N.$$

To prove the continuity of (v, v_t) in T' introduce the notation

$$z_{mn}^{\pm}(t) := \pm i\omega_{mn} v_{mn}(t) + \dot{v}_{mn}(t).$$

Then

$$\sum_{m,n} \left(|z_{mn}^{+}(T')|^2 + |z_{mn}^{-}(T')|^2\right) \omega_{mn}^{-1} \asymp \|(v(.,T'), v_t(.,T'))\|^2_{\mathring{W}_2^1} \tag{28}$$

and we have to show that

$$\sum_{m,n} \left|z_{mn}^{\pm}(T'+h) - z_{mn}^{\pm}(T')\right|^2 \omega_{mn}^{-1} \longrightarrow 0 \qquad \text{if } h \to 0.$$

By (25) and (26)

$$z_{mn}^{\pm}(t) = \sum_{j=1}^{N} \int_0^t \varphi_{mn}(p_j) f_j(\tau) e^{\pm i\omega_{mn}(t-\tau)} d\tau \tag{29}$$

and then

$$\begin{aligned} & z_{mn}^{\pm}(T'+h) - z_{mn}^{\pm}(T') = \\ & = \sum_{j=1}^{N} \varphi_{mn}(p_j) \left\{ \int_{T'}^{T'+h} f_j(\tau) e^{\pm i\omega_{mn}(T'+h-\tau)} d\tau + \right. \\ & \left. + \int_0^{T'} f_j(\tau) \left[e^{\pm i\omega_{mn} h} - 1\right] e^{\pm i\omega_{mn}(T'-\tau)} d\tau \right\}. \end{aligned} \tag{30}$$

Here the first member can be integrated in the way that the function $f_j(t)$ is zero unless $t \in [T', T'+h]$. Now we can apply (28)

$$\sum_{m,n,\pm} \omega_{mn}^{-1} \left| \varphi_{mn}(p_j) \int_{T'}^{T'+h} f_j(\tau) e^{\pm i\omega_{mn}(T'+h-\tau)} d\tau \right|^2 \leqq$$

$$\leqq c(T'+1) \sum_{j=1}^{N} \|f_j\|^2_{L^2(T', T'+h)} \longrightarrow 0 \qquad \text{if } h \to 0.$$

Next we estimate the second term of (30):

$$\sum_{m,n,\pm} \omega_{mn}^{-1} \left| \sum_{j=1}^{N} \varphi_{mn}(p_j) \int_0^{T'} f_j(t) \left[e^{\pm i\omega_{mn}h} - 1 \right] e^{\pm i\omega_{mn}(T'-t)} dt \right|^2 \leqq$$

$$\leqq c \left\{ \sum_{m+n \leqq R} |e^{\pm i\omega_{mn}h} - 1| \|f\|^2_{L^2(0,T';\mathbb{R}^N)} + \right.$$

$$\left. + \sum_{m+n>R} \omega_{mn}^{-1} \left| \sum_{j=1}^{N} \varphi_{mn}(p_j) \int_0^{T'} f_j(t) e^{\pm i\omega_{mn}(T'-t)} dt \right|^2 \right\}.$$

By (28) and (29) the second series converges, hence for large R it becomes as small as we want. If we fix such an R, the first sum tends obviously to zero. This finishes the proof of Theorem 2 (a). To see (b) consider the problem (20) in case $T' = T$. Taking the change of variable $t \to T - t$ it becomes problem (1). The linear set

$$V := \left\{ (\omega_0, \omega_1) \in \mathcal{W}_{\frac{1}{2}} : \omega(p_j, t) = 0,\ j = 1, \ldots, N;\ t \in [0, T] \right\} \subset \mathcal{W}_{\frac{1}{2}}$$

is infinite dimensional by Theorem 1 (b). From (27) we see that

$$\left(v(.,T), v_t(.,T) \right) \in \mathcal{W}_{\frac{1}{2}}$$

is orthogonal to this subspace V which proves also Theorem 2 (b). □

REMARK 1. The statement (a) of Theorem 2 is true for any bounded domain in $\mathbf{R}^2$ for which the spectral function of the Laplace operator satisfies the Weil's asymptotics. Indeed, in this case the key inequality (24) is the well-known square-sum estimate for the eigenfunctions.

REMARK 2. In [1], p. 180, Russell proved that $\cup_{T>0} R(T)$ is dense in $\mathcal{W}_{\frac{1}{2}}$ if and only if a "range condition" is fulfilled. If the multiplicity of eigenvalues is not bounded, then this condition fails, so Theorem 2 (b) follows directly from Russell's theorem in case of a square membrane. If $a^2/b^2 \notin \mathbb{Q}$ in a rectangular membrane then the multiplicity of every eigenvalue is 1 (the spectrum is simple) and the range condition holds for almost all points $p_j \in$ $\in \Omega$. Then $\cup_{T>0} R(T)$ is dense but, by Theorem 2, $R(T)$ itself is not dense for any $T > 0$.

REFERENCES

[1] RUSSELL, D. L., Controllability and stabilizability theory for linear partial differential equations: recent progress and open questions, *SIAM Rev.* **20** (1978), 639–739. *MR* **80c**:93032

[2] PAVLOV, B. S., The basis property of a system of exponentials and the condition of Muckenhoupt, *Dokl. Akad. Nauk SSSR* **247**(1979), No.1, 37–40 (in Russian). *Zbl* **429**:30004; *MR* **84j**:42042

[3] NIKOL'SKII, N. K., PAVLOV, B. S. and HRUŠČEV, S. V., Unconditional bases of exponentials and of reproducing kernels, *Complex analysis and spectral theory* (Leningrad, 1979/1980), Lecture Notes in Math., 864, Springer, Berlin-New York, 1981, 214–335. *MR* **84k**:46019

[4] RUSSELL, D. L., On exponential bases for the Sobolev spaces over an interval, *J. Math. Anal. Appl.* **87**(1982), 528–550. *MR* **83g**:46035

[5] AVDONIN, S. A., On the controllability of singular strings, *Mechanical problems of controllable motions*, 1982, 3–8 (in Russian).

[6] AVDONIN, S. A. and IVANOV, S. A., A generating matrix function in problems of the control of the vibrations of connected strings, *Dokl. Akad. Nauk. SSSR* **307**(1989), no.5, 1033–1037 (in Russian). (Translation in *Soviet Math. Dokl.* **40**(1990), No.1, 179–183.) *MR* **90k**:35145

[7] AVDONIN, S. A. and IVANOV, S. A., Riesz bases of exponentials in a space of vector-functions and controllability of an inhomogeneous string, *Operator Theory and function theory*, No.1, Leningrad Univ., Leningrad, 1983, 62–68 (in Russian). *MR* **86b**:46058

[8] AVDONIN, S. A. and IVANOV, S. A., Controllability of distributed parameter systems and families of exponentials, Kiev, 1989 (in Russian).

[9] AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., On Riesz bases from vector exponentials I, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **32**(1989), 101–114. *MR* **92m**:42034

[10] AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., On Riesz bases from vector exponentials II, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **32**(1989), 115–126. *MR* **92m**:42034

[11] AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., Initial and pointwise control of a rectangular membrane, *Avtomatika* **6**(1990) (in Russian).

[12] LIONS, J. L. and MAGENES, E., *Problèmes aux limites non homogènes et applications*, Vol. I–II, Travaux et Recherches Mathématiques, Nos. 17–18, Dunod, Paris, 1968. *MR* **40** #512, 513

[13] CASSELS, J. W. S., *An introduction to Diophantine approximation*, Cambridge Tracts in Mathematics and Math. Physics, No.45, Cambridge University Press, New York, 1957. *MR* **19**–396

[14] MEYER, Y., Étude d'un modèle mathématique issu du contrôle des structures spatiales déformables, *Nonlinear Partial Differential Equations and their Applications*, Collège de France séminaires, Vol. VII (Paris, 1983–1984), Res. Notes in Math., 122, Pitman, Boston, Mass.–London, 1985, 234–242. *MR* **88e**:73021

[15] HÖRMANDER, L., The spectral function of an elliptic operator, *Acta Math.* **121**(1968), 193–218. *MR* **58** #29418

[16] AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., *Families of exponentials and controllability of a rectangular membrane, Studia Sci. Math. Hungar.* **25**(1990), 291–306 (in Russian). *MR* **92c**:93011

[17] AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., Applications of exponential bases in the pointwise control of rectangular membranes, Math. Inst. Hungar. Acad. Sci., Budapest, Preprint No.64/1990.

[18] HORVÁTH, M., The vibration of a membrane in different points, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **33**(1990), 31–38. *MR* **92j**:35108

[19] JOÓ, I., On some Riesz bases, *Period. Math. Hungar.* **22** (1991), 187–196. *MR* **93a**: 47025

[20] DUFFIN, R. J. and EACHUS, J. J., Some notes on an expansion theorem of Paley and Wiener, *Bull. Amer. Math. Soc.* **48**(1942), 850–855. *MR* 4–97

[21] Joó, I., On the control of a circular membrane I, *Acta Math. Hungar.* **61** (1993), 303–325. *MR* **94a**:93010

(Received June 17, 1991)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 261–263

ON RANGE CHARACTERIZATION OF ADJOINT OPERATORS ON HILBERT SPACE

Z. SEBESTYÉN and L. KAPOS

We prove in this note that the inverse image of a finite dimensional subspace under an adjoint map can be characterized as the orthogonal complement of the image under the original linear operator of the subspace which is just the intersection of the domain of the map with the orthocomplement of the given finite subspace in a Hilbert space.

This is an improvement of the first named author's results in this direction [1, Theorem],[2]. The proof given here is a refinement of the one used in [2, Theorem] the trick owed to Riesz and known by his representation theorem of continuous linear functionals on Hilbert spaces.

THEOREM. *Let A be a densely defined, not necessarily bounded, linear operator on its domain $D(A)$ in a Hilbert space H. Let K be a closed subspace in H that has finite dimensional orthocomplement $K^\perp$. Then the orthocomplement L of the linear manifold $\{Ah\colon h \in D(A) \cap K\}$ belongs to $D(A^*)$, the domain of the adjoint operator A^* of A and one has*

$$A^*(L) = R(A^*) \cap K^\perp. \tag{1}$$

PROOF. To see the inclusion $R(A^*) \cap K^\perp \subset A^*(L)$ let $y \in D(A^*)$ so that $A^*y \in K^\perp$, that is $(A^*y, x) = 0$ holds for any $x \in K$, moreover for each of $x \in D(A) \cap K$. Hence we have that $(y, Ax) = (A^*y, x) = 0$ for any $x \in D(A) \cap$ $\cap K$. This implies that $y \in L$ and thus $A^*y \in A^*(L)$ indeed. We check now that the reverse inclusion $A^*(L) \subset R(A^*) \cap K^\perp$ also holds together with the fact to be proved $L \subset D(A^*)$. To this end let $y_1, \ldots, y_n$ be base in $K^\perp$ and let

$$H_i = \left\{x \in H : (x, y_j) = 0 \quad (j \neq i, 1 \leqq j \leqq n)\right\} \quad (i = 1, 2, \ldots, n).$$

Then H_i's are finite codimensional closed subspaces in H such that $D(A) \cap$ $\cap H_i$ remains dense in H_i via denseness of $D(A)$ in H. For the reader's convenience we show this fact by an easy induction on the codimension. For if M is one codimensional closed subspace in H then $K = \{x \in H : (x, e) = 0\}$ for some unit vector e in H. Let $x \in K$ be fixed and let $x_n \in D(A)$, $f \in D(A)$

1991 *Mathematics Subject Classification.* Primary 47A20; Secondary 47B15.
Key words and phrases. Hilbert space operators, closed operators, restrictions.

0081–6906/95/$ 4.00

with the requirements that $\|x_n - x\| \to 0$ $(n \to \infty)$ and $\|e - f\| < 1$. By the last assumption we have that

$$|(f,e)| = |(f-e,e)+(e,e)| \geqq (e,e) - |(f-e,e)| \geqq$$
$$\geqq 1 - \|e-f\|\,\|e\| > 0.$$

Let

$$y_n := x_n - \frac{(x_n,e)}{(f,e)} f \in D(A) \qquad (n = 1,2,\ldots),$$

then $y_n \in D(A) \cap K$ because $(y_n,e) = (x_n,e) - \frac{(x_n,e)}{(f,e)}(f,e) = 0$ for any $n = 1,2,\ldots$ and moreover

$$\|y_n - x\| \leqq \|x_n - x\| + \frac{\|(x_n,e)\|\,\|f\|}{|(f,e)|} = \|x_n - x\| + \frac{|(x_n - x,e)|\,\|f\|}{|(f,e)|} \leqq$$
$$\leqq \|x_n - x\| \left(1 + \frac{\|f\|}{|(e,f)|}\right) \longrightarrow 0 \qquad (n \to \infty).$$

Since $\overline{D(A) \cap H_i} = H_i$ there exists $v_i \in D(A) \cap H_i$ with the requirements $(v_i, y_i) = 1$, $(v_i, y_j) = 0$, $(j \neq i,\ j = 1,\ldots,n)$ for any $i = 1,2,\ldots,n$. Then we have

$$D(A) \cap K = \left\{ x - \sum_{i=1}^{n} (x,y_i) v_i : x \in D(A) \right\}.$$

On one side, for any $x \in D(A)$ we see

$$\left(x - \sum_{i=1}^{n} (x,y_i) v_i, y_j \right) = (x,y_j) - \sum_{i=1}^{n} (x,y_i)(v_i,v_j) =$$
$$= (x,y_j) - (x,y_j) = 0$$

so that

$$x - \sum_{i=1}^{n} (x,y_i) v_i \in D(A) \cap K.$$

Otherwise, for each $x \in D(A) \cap K$ we have $(x,y_i) = 0$ for $i = 1,2,\ldots,n$ so that $x = x - \sum_{i=1}^{n} (x,y_i) v_i$.

Now, if $y \in L$ then have that

$$\left(A\left(x - \sum_{i=1}^{n} (x,y_i) v_i \right), y \right) = 0 \qquad \left(x \in D(A) \right),$$

that is

$$(Ax,y) = \sum_{i=1}^{n} (x,y_i)(Av_i,y) = \left(x, \sum_{i=1}^{n} (y,Av_i) y_i \right) \qquad \left(x \in D(A) \right),$$

therefore that $y \in D(A^*)$ and

$$A^*y = \sum_{i=1}^{n} (y, Av_i)y_i \in K^{\perp}, \qquad A^*y \in R(A^*) \cap K^{\perp}.$$

The proof is complete.

REFERENCES

[1] SEBESTYÉN, Z., On ranges of adjoint operators in Hilbert space, *Acta Sci. Math. (Szeged)* **46**(1983), 295–298. *MR* **85i**:47003a

[2] SEBESTYÉN, Z., Least norm solution of linear equation in Hilbert space, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **34** (1991), 69–71. *MR* **93h**:47010

(Received July 7, 1991)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 265–274

POLYNOMIAL APPROXIMATION ON LOCALLY COMPACT ABELIAN GROUPS II

R. WINKLER

Abstract

The object of this paper is to describe the topological closure of the set of polynomial functions on a locally compact abelian group with respect to the topologies of pointwise convergence, uniform convergence on compact subsets and uniform convergence. In every case one gets again locally compact abelian groups which can be classified up to isomorphism.

1. Notations and abbreviations

Let G be a locally compact abelian (LCA-) group with operation $+$. (All groups considered are supposed to satisfy the Hausdorff separation axiom.) We shall use the following abbreviations:

$H \leqq G$	This means that H is a closed subgroup of G.
$\hat{G}$	The dual group of G, also written additively.
$\mathcal{U}(G)$	The system of open neighbourhoods of $0 \in G$.
Id_M	The identity map on a set M.
$\cong$	This symbol as well as the words "isomorphic" and "isomorphism" etc. will always be used not only in their algebraic but also in their topological meaning.
$\overline{M}^{(t)}$	The topological closure of the set M with respect to the topology t.
$a_\nu \to a$	The limit of the net $(a_\nu)_{\nu \in N}$ where $(N, \leqq)$ is a directed set.
$\mathbf{N}=\{0,1,2,\dots\}$	The set of nonnegative integers.
$\mathbf{P}=\{2,3,5,7\dots\}$	The set of all prime numbers.
$\mathbf{Z}$	The additive (discrete) group of integers.
$\mathbf{R}$	The additive group of reals with the natural topology.
$\mathbf{T}=\mathbf{R}/\mathbf{Z}\cong\hat{\mathbf{Z}}$	The one-dimensional torus with the natural topology.

1991 *Mathematics Subject Classifications.* Primary 22B05; Secondary 08A40.
Key words and phrases. LCA-groups, polynomial function, approximation.

I would like to thank Wolfgang Herfort for several stimulating discussions and his continuous encouragement.

$\mathbf{T}_d = \mathbf{R}_d/\mathbf{Z}$ The one-dimensional torus with the discrete topology.

$\widehat{\mathbf{T}_d} \cong \mathbf{Z}^{(B)}$ The dual group of the discrete torus, isomorphic to the Bohr-compactification $\mathbf{Z}^{(B)}$ of $\mathbf{Z}$ by the embedding $k \in \mathbf{Z} \mapsto k\mathrm{Id}_{\mathbf{T}_d} \in \widehat{\mathbf{T}_d}$.

$C(n) = \{t \in \mathbf{T} \mid nt = 0\} \leqq \mathbf{T}$ The finite cyclic group of order $n \in \mathbf{N}$, $n \geqq 1$, considered as a subgroup of the torus whenever convenient.

$C(p^\infty) = \bigcup_{n \in \mathbf{N}} C(p^n) \leqq \mathbf{T}_d$, $p \in \mathbf{P}$ The p-Prüfer group with discrete topology.

$\widehat{C(p^\infty)}$, $p \in \mathbf{P}$ The dual group of the p-Prüfer group which may be identified with the compact group of p-adic integers.

We are going to consider the following sets of mappings:

$\mathcal{F}_n(G) = \{f \mid f\colon G^n \to G\}$,

$\mathcal{P}_n(G) = \{f \in \mathcal{F}_n(G) \mid f(x_1, \ldots, x_n) = g + \sum_{i=1}^{n} k_i x_i,\ g \in G,\ k_i \in \mathbf{Z}\}$,

the set of polynomial functions in n variables on the abelian group G,

$\mathcal{E}(G) = \{f \in \mathcal{F}_1(G) \mid f$ is algebraic endomorphism$\}$,

$\mathcal{K}(G) = \{f \in \mathcal{E}(G) \mid f(H) \subseteq H$ for all $H \leqq G\}$,

$\mathcal{P}(G) = \{f \in \mathcal{F}_1(G) \mid f(x) = kx$ for some $k \in \mathbf{Z}\}$.

(Note that the considered maps in $\mathcal{F}_n(G)$, $\mathcal{E}(G)$ and $\mathcal{K}(G)$ are not necessarily continuous.)

pw, co, u By taking the operation pointwise the sets defined above are abelian groups, too. On these sets we are going to consider the topology pw of pointwise convergence, the topology co of uniform convergence on compact subsets (compact-open topology) and the topology u of uniform convergence (on the whole set G).

2. Introduction and auxiliary results

G. Kowol, cf. [2], was the first to consider polynomial approximation on topological universal algebras. The problem is to determine the topological closure of the set of polynomial functions. In [3] the case of locally compact abelian groups has been studied. This paper will complete the investigations of [3] and give a description and classification of the topological group $\overline{\mathcal{P}}_n(G)^{(t)}$ in terms of the underlying topology $t = pw, co$ and u and of the structure of G.

Beside some standard theory of LCA-groups (cf. Proposition 1) we shall use the main results of [4] where the topological group $\mathcal{K}(G)_{pw}$ has been described (cf. Proposition 2) and some auxiliary results from [3] (cf. Proposition 3).

PROPOSITION 1. *Let G be an arbitrary LCA-group. Then it is exactly of one of the following types:*

a) *Type 1: $\mathbf{Z} \cong D \leqq G$ for some D.*

b) *Type 2: G is periodic — i.e. every compactly generated (closed) subgroup is compact — and not totally disconnected.*

c) *Type 3: G is periodic and totally disconnected. In this case $0 \in G$ has a topological neighbourhood base of compact open subgroups. If $\exp(G/H) < \infty$ for every compact open $H \leqq G$ we say that G is of Type 3a, otherwise of Type 3b.*

PROOF. These facts are well-known (cf. for instance [1]).

PROPOSITION 2. *Let G be an LCA-group. Consider pw-topology. Then for the topological group $\mathcal{K}(G)_{pw}$ the following classification holds (cf. Proposition 1):*

a) *If G is of Type 1 then*

$$\mathcal{K}(G)_{pw} \cong \mathbf{Z}$$

by the isomorphism $k\mathrm{Id}_G \mapsto k$.

b) *If G is of Type 2 or 3 then*

$$\mathcal{K}(G)_{pw} \cong \widehat{T(G)} \cong \widehat{\mathbf{T}_d}/T(G)^{\perp}.$$

Here $T(G)$ denotes the set $\{\chi(g) \mid \chi \in \hat{G}, g \in G\}$ with discrete topology which is in fact a subgroup of $\mathbf{T}_d$ and $T(G)^{\perp}$ its annihilator. An isomorphism $\phi\colon \widehat{T(G)} \to \mathcal{K}(G)_{pw}$, $\alpha \mapsto \varepsilon_\alpha$ is characterized by the relation

$$\alpha \circ \chi = \chi \circ \varepsilon_\alpha \quad \text{for all} \quad \chi \in \hat{G}.$$

If G is of Type 2, this means

$$\mathcal{K}(G)_{pw} \cong \widehat{\mathbf{T}_d} = \mathcal{K}(\mathbf{T}) = \mathcal{E}(\mathbf{T}) \cong \mathbf{Z}^{(B)}.$$

If G is of Type 3, this implies

$$\mathcal{K}(G)_{pw} \cong \prod_{p \in \mathbf{P}} \widehat{C(p^{e_p})}$$

(topological direct product). Here $p^{e_p} = \exp(G_p) \in \mathbf{N} \cup \{\infty\}$ $(p^\infty = \infty)$ denotes the p-exponent of G which is the exponent of the topological p-Sylow subgroup G_p of G defined by

$$G_p = \{g \in G \mid \lim_{k \to \infty} p^k g = 0 \in G\}.$$

If $e_p < \infty$ then $\widehat{C(p^{e_p})} \cong C(p^{e_p})$ is a cyclic group of order p^{e_p}, otherwise $\widehat{C(p^\infty)}$ may be considered as the compact group of p-adic integers.

PROOF. Cf. assertions a) and b) of the Theorem and the remarks at the end of Section 4 in [4].

From [3] we use the following results:

PROPOSITION 3. a) *Suppose that the LCA-group G is not totally disconnected. Then the topological group $\overline{\mathcal{P}}(G)^{(u)}$ is discrete, i.e. $\overline{\mathcal{P}}(G)^{(u)} = \mathcal{P}(G)_u$, and $\overline{\mathcal{P}}(G)^{(u)} \cong \mathbf{Z}$ via the isomorphism*

$$\Phi\colon \overline{\mathcal{P}}(G)^{(u)} \to \mathbf{Z}, \quad \Phi\colon k\,\mathrm{Id}_G \mapsto k.$$

b) $\overline{\mathcal{P}}(G)^{(pw)} \subseteq \mathcal{K}(G)$ *for every LCA-group G.*

PROOF. Cf. [3] a) follows from Theorem 1 and b) is contained in Theorem 2.

3. The main results

THEOREM 1.

$$\overline{\mathcal{P}}_n(G)^{(t)} \cong G \times \left(\overline{\mathcal{P}}(G)^{(t)}\right)^n$$

holds for every abelian topological Hausdorff group G (not necessarily LCA) and each topology $t = pw, co, u$. An isomorphism

$$\Phi\colon G \times \left(\overline{\mathcal{P}}(G)^{(t)}\right)^n \to \overline{\mathcal{P}}_n(G)^{(t)}$$

is given by the definition

$$\Phi(g, \varepsilon_1, \ldots, \varepsilon_n)(x_1, \ldots, x_n) = g + \varepsilon_1(x_1) + \ldots + \varepsilon_n(x_n).$$

PROOF. Section 4.

By Theorem 1 our problem reduces to describe the topological group $\overline{\mathcal{P}}^{(t)}$ with respect to the topologies $t = pw, co, u$. This will be done in Theorem 2 and Theorem 3. Theorem 2 is a rather easy consequence of Proposition 2 and treats pw-topology, Theorem 3 describes the structure of the set $\overline{\mathcal{P}}(G)^{(t)}$ also for $t = co$ and $t = u$.

THEOREM 2.

$$\overline{\mathcal{P}}(G)^{(pw)} = \mathcal{K}(G)_{pw}$$

holds for every LCA-group G. For the structure of $\mathcal{K}(G)_{pw}$ cf. Proposition 2.

PROOF. Section 5.

THEOREM 3. *For the structure of the topological group $\overline{\mathcal{P}}(G)^{(t)}$, $t = pw, co, u$, the following two situations (which coincide if G is of Type 1) are possible (we use the notations of Proposition 1):*

Situation a):

$$\overline{\mathcal{P}}(G)^{(t)} \cong \mathbf{Z} \text{ via the isomorphism } k\,\mathrm{Id}_G \mapsto k$$

if G is of Type 1 and $t = pw, co, u$ or
if G is of Type 2 and $t = co, u$ or
if G is of Type 3b and $t = u$.
Situation b):

$$\overline{\mathcal{P}}(G)^{(t)} = \mathcal{K}(G)_t = \mathcal{K}(G)_{pw}$$

and the topologies t and pw coincide on $\mathcal{K}(G)$
if G is of Type 1 and $t = pw, co, u$ or

if G is of Type 2 and $t = pw$ or
if G is of Type 3a and $t = pw, co, u$ or
if G is of Type 3b and $t = pw, co$.
For further information about the structure of $\mathcal{K}(G)_{pw}$ cf. Proposition 2.

PROOF. Section 6.
If G is of Type 1 cf. Lemma 1.
If G is of Type 2 cf. Lemma 2.
If G is of Type 3a cf. Lemma 3.
If G is of Type 3b and $t = pw, co$ cf. Lemma 4.
If G is of Type 3b and $t = u$ cf. Lemma 6.

4. The proof of Theorem 1

Let t be one of the topologies pw, co and u. The following proof runs the same way for each of the three cases. We have to prove the following statements:

(i) $\Phi\left(G \times (\overline{\mathcal{P}}(G)^{(t)})^n\right) \subseteqq \overline{\mathcal{P}}_n(G)^{(t)}$.
(ii) $\overline{\mathcal{P}}_n(G)^{(t)} \subseteqq \Phi(G \times (\overline{\mathcal{P}}(G)^{(t)})^n)$, i.e. Φ is onto.
(iii) Φ is injective.
(iv) Φ is an algebraic homomorphism.
(v) Φ is continuous.
(vi) Φ^{-1} is continuous.

(i) Pick $g \in G$, $\varepsilon_i \in \overline{\mathcal{P}}(G)^{(t)}$ and $U \in \mathcal{U}(G)$ arbitrarily. To prove (i) it suffices to find $k_1, \ldots, k_n \in \mathbf{Z}$ such that

$$\Phi(g, \varepsilon_1, \ldots, \varepsilon_n)(g_1, \ldots, g_n) - (g + k_1 g_1 + \ldots + k_n g_n) \in U \tag{1}$$

for all $g_i \in M_i \subseteqq G$, $i = 1, \ldots, n$, where the sets M_i are arbitrary but fixed and finite (in the case of pw-topology) or compact (co-topology) or $M_i = G$ (u-topology). Take $V \in \mathcal{U}(G)$ with

$$nV = \{v_1 + \ldots + v_n \mid v_i \in V\} \subseteqq U.$$

Since $\varepsilon_i \in \overline{\mathcal{P}}(G)^{(t)}$ there are $k_i \in \mathbf{Z}$ with $k_i g_i - \varepsilon_i(g_i) \in V$ for all $g_i \in M_i$, $i = 1, \ldots, n$. Of course this implies (1).

(ii) Pick $\varepsilon \in \overline{\mathcal{P}}_n$ and define $g = \varepsilon(0, \ldots, 0) \in G$ and

$$\varepsilon_i(x) = \varepsilon(0_1, \ldots, 0_{i-1}, x, 0_{i+1}, \ldots, 0_n) - g$$

for $i = 1, \ldots, n$. We have to prove

$$\varepsilon_i \in \overline{\mathcal{P}}(G)^{(t)}, \quad i = 1, \ldots, n, \quad \text{and} \tag{2}$$

$$\varepsilon(x_1, \ldots, x_n) = g + \varepsilon_1(x_1) + \ldots + \varepsilon_n(x_n). \tag{3}$$

For doing this take any M_i as in (i) (w.l.o.g. assume $M_1 = \ldots = M_n = M$, $0 \in M$) and $U \in \mathcal{U}(G)$. Now take $V \in \mathcal{U}(G)$ with $(2n+2)V \subseteq U$ and $-V = V$. Since $\varepsilon \in \overline{\mathcal{P}}_n$ there are $g' \in G$ and $k_i \in \mathbf{Z}$, $i = 1, \ldots, n$, such that

$$\varepsilon(g_1, \ldots, g_n) - (g' + k_1 g_1 + \ldots + k_n g_n) \in V \tag{4}$$

for all $g_i \in M$. In the case $g_j = 0$ for $j \neq i$ this means

$$\varepsilon(0, \ldots, 0, g_i, 0, \ldots, 0) - (g' + k_i g_i) \in V \tag{5}$$

for all $g_i \in M$ and, if also $g_i = 0$, we have

$$g' - g \in -V = V. \tag{6}$$

Summation of (5) and (6) yields

$$\varepsilon_i(g_i) - k_i g_i \in V + V \subseteq U \tag{7}$$

for all $g_i \in M$ proving (2). Furthermore

$$\begin{aligned} &\varepsilon(g_1, \ldots, g_n) - (g + \varepsilon_1(g_1) + \ldots + \varepsilon_n(g_n)) = \\ &= \Big(\varepsilon(g_1, \ldots, g_n) - (g' + \sum_{i=1}^{n} k_i g_i)\Big) + (g' - g) + \\ &\quad + \sum_{i=1}^{n} (k_i g_i - \varepsilon_i(g_i)) \in (2n+2)V \subseteq U \end{aligned}$$

for all $g_i \in M$ by (4), (6) and (7). Since $U \in \mathcal{U}(G)$ was arbitrary and G is Hausdorff this proves (3).

(iii) Suppose

$$\Phi(g, \varepsilon_1, \ldots, \varepsilon_n) = \Phi(g', \varepsilon'_1, \ldots, \varepsilon'_n)$$

for $g, g' \in G$ and $\varepsilon_i, \varepsilon'_i \in \overline{\mathcal{P}}$, i.e.

$$g + \varepsilon_1(g_1) + \ldots + \varepsilon_n(g_n) = g' + \varepsilon'_1(g_1) + \ldots + \varepsilon'_n(g_n)$$

for all $g_i \in G$, $i = 1, \ldots, n$. Choosing $g_1 = \ldots = g_n = 0$ and using $\varepsilon_i(0) = \varepsilon_i'(0) = 0$ for $i = 1, \ldots, n$ we get $g = g'$ and hence, now choosing $g_j = 0$ only for $j \neq i$, also $\varepsilon_i(g_i) = \varepsilon_i'(g_i)$ for all $g_i \in G$, i.e. $\varepsilon_i = \varepsilon_i'$.

(iv) Trivial.

(v) This is an immediate consequence of continuity of group operation with respect to the topologies *pw*, *co* and *u*.

(vi) By (iv) it suffices to prove continuity at the point $0 \in \overline{\mathcal{P}}_n(G)^{(t)}$. Let therefore

$$\Phi(g^{(\nu)}, \varepsilon_1^{(\nu)}, \ldots, \varepsilon_n^{(\nu)}) \to 0$$

with respect to the topology t. In every case we have

$$g^{(\nu)} = \Phi(g^{(\nu)}, \varepsilon_1^{(\nu)}, \ldots, \varepsilon_n^{(\nu)})(0, \ldots, 0) \to 0$$

in G, implying also

$$\varepsilon_i^{(\nu)}(g) = \Phi(g^{(\nu)}, \varepsilon_1^{(\nu)}, \ldots, \varepsilon_n^{(\nu)})(0_1, \ldots, 0_{i_1}, g, 0_{i+1}, \ldots, 0_n) - g^{(\nu)} \to 0$$

with respect to t for $i = 1, \ldots, n$. Now the proof of Theorem 1 is complete.

5. The proof of Theorem 2

We have to prove two inclusions:

"$\subseteqq$": Proposition 3 b).

"$\supseteqq$": a) G is of Type 1, i.e., not periodic: Then, by Proposition 2 a), every $\varepsilon \in \mathcal{K}(G)$ is of the form $\varepsilon = k \operatorname{Id}_G$, $k \in \mathbf{Z}$, hence $\varepsilon \in \mathcal{P}(G) \subseteqq \overline{\mathcal{P}}(G)^{(pw)}$.

b) G is of Type 2 or 3, i.e. periodic: By Proposition 2 b),

$$K(G)_{pw} \cong \widehat{T(G)} \cong \widehat{\mathbf{T}_d}/T(G)^{\perp}.$$

$\widehat{\mathbf{T}_d} \cong \mathbf{Z}^{(B)}$, isomorphic to the Bohr-compactification of $\mathbf{Z}$, is monothetic and generated by $\operatorname{Id}_{\mathbf{T}_d}$. Thus the isomorphism ϕ described in Proposition 2 b) shows that $\mathcal{K}(G)_{pw} \cong \widehat{T(G)}$ is generated by $\operatorname{Id}_G \in \mathcal{P}(G)$, thus $\mathcal{K}(G) \subseteqq \overline{\mathcal{P}}(G)^{(pw)}$.

6. The proof of Theorem 3

LEMMA 1. *Theorem 3 holds if G is of Type 1, $t = pw, co, u$.*

PROOF. Proposition 2 a) and Theorem 2 imply

$$\mathcal{P}(G) \subseteqq \overline{\mathcal{P}}(G)^{(u)} \subseteqq \overline{\mathcal{P}}(G)^{(co)} \subseteqq \overline{\mathcal{P}}(G)^{(pw)} = \mathcal{K}(G) = \mathcal{P}(G),$$

proving Lemma 1.

LEMMA 2. *Theorem 3 holds if G is of Type 2 and $t = pw, co, u$.*

PROOF. $t = pw$: Theorem 2.

$t = co, u$: Let G_0 be the component of connectedness of $0 \in G$. On compact sets co-topology and u-topology coincide, G_0 is compact (G periodic), hence by Proposition 3 a)

$$\overline{\mathcal{P}}(G_0)^{(co)} = \overline{\mathcal{P}}(G_0)^{(u)} = \mathcal{P}(G_0) \cong \mathbf{Z} \text{ via the isomorphism } k\,\mathrm{Id}_{G_0} \mapsto k.$$

Now it is clear that

$$\Phi\colon \overline{\mathcal{P}}(G)^{(co)} = \overline{\mathcal{P}}(G)^{(u)} \to \mathbf{Z}, \quad k\,\mathrm{Id}_G \mapsto k$$

is isomorphism as well.

LEMMA 3. *Let G be of Type 3a. Then*

$$\overline{\mathcal{P}}(G)^{(pw)} = \overline{\mathcal{P}}(G)^{(co)} = \overline{\mathcal{P}}(G)^{(u)} = \mathcal{K}(G)$$

and the topologies pw, co and u coincide on this set. Hence Theorem 3 holds in this case.

PROOF. Assume $\varepsilon_\nu \to 0$ pointwise, $\varepsilon_\nu \in \mathcal{K}(G) = \overline{\mathcal{P}}(G)^{(pw)}$ (cf. Theorem 2). The Lemma is proved when we can show that $\varepsilon_\nu \to 0$ uniformly. To do this take any $U \in \mathcal{U}(G)$. Since G is of Type 3 there is a compact open $H \subseteqq U$, $H \leqq G$. $H' = G/H$ is discrete and, by definition of Type 3a, $\exp(H') = n < \infty$. Thus we can find an $h + H \in H'$ with order n. If we define $\varepsilon'_\nu(g+H) = \varepsilon_\nu(g) + H$ (this is well defined) we have $\varepsilon'_\nu \in \mathcal{K}(H') = \overline{\mathcal{P}}(H')^{(pw)} = \mathcal{P}(H') \cong C(n)$ (cyclic group of order n). $\varepsilon_\nu \to 0$ pointwise in G obviously implies $\varepsilon'_\nu \to 0$ pointwise in H' and therefore $\varepsilon'_\nu = 0$ or, equivalently, $\varepsilon_\nu(G) \subseteqq H \subseteqq U$ for all $\nu \geqq \nu_0$. Thus Lemma 3 is proved.

LEMMA 4. *Theorem 3 holds if G is of Type 3 and $t = pw, co$.*

PROOF. $t = pw$: Theorem 2.

$t = co$: Assume $\varepsilon_\nu \to 0$ pointwise with $\varepsilon_\nu \in \overline{\mathcal{P}}(G)^{(pw)}$. For an arbitrary compact $K \subseteqq G$ we have to show $\varepsilon_\nu \to 0$ uniformly on K. Since the closed subgroup generated by the set K is compact, too, w.l.o.g. we may assume $K \leqq G$. It is clear that $\varepsilon_\nu|_K \to 0$ in $\overline{\mathcal{P}}(K)^{(pw)}$. Consider any compact open subgroup $H \leqq K$, then, by compactness of K, K/H is finite. Hence K is of Type 3a and Lemma 3 applies proving Lemma 4.

For G of Type 3b and $t = u$ we need

LEMMA 5. *Let G be of Type 3. Then for*

$$\varepsilon \in \mathcal{K}(G) = \overline{\mathcal{P}}(G)^{(pw)} \supseteqq \overline{\mathcal{P}}(G)^{(u)}$$

the statements (i) *and* (ii) *are equivalent.*

(i) $\varepsilon \in \overline{\mathcal{P}}^{(u)}(G)$.

(ii) *For every compact open* $H \leqq G$ *there exists a* $k_H \in \mathbf{Z}$ *such that* $\varepsilon_H = k_H \operatorname{Id}_{G/H}$ *if we put* $\varepsilon_H(g+H) = \varepsilon(g) + H$. *(Note that* $\varepsilon_H \in \mathcal{K}(G/H)$ *is well defined.)*

PROOF. (i) $\Rightarrow$ (ii). Suppose $\varepsilon \in \overline{\mathcal{P}}(G)^{(u)}$ and take any compact open $H \leqq G$. By definition there is a $k_H \in \mathbf{Z}$ such that $\varepsilon(g) - k_H g \in H$ for all $g \in G$. But this immediately yields

$$\varepsilon_H(g+H) = \varepsilon(g) + H = k_H(g+H).$$

(ii) $\Rightarrow$ (i). To show (i) take any $U \in \mathcal{U}(G)$. Proposition 1 c) guarantees the existence of a compact open subgroup $H \subseteq U$. By (ii) there is a $k_H \in \mathbf{Z}$ such that $\varepsilon_H = k_H \operatorname{Id}_{G/H}$, hence

$$\varepsilon(g) - k_H g \in H \subseteq U$$

for all $g \in G$. $U \in \mathcal{U}(G)$ was arbitrary, thus Lemma 5 is proved.

LEMMA 6. *Theorem 3 holds if* G *is of Type 3b and* $t = u$.

PROOF. For

$$\Phi\colon \mathbf{Z} \to \overline{\mathcal{P}}(G)^{(u)},\ k \mapsto k \operatorname{Id}_G$$

we have to prove the following statements:

(i) $\Phi(\mathbf{Z}) \subseteq \overline{\mathcal{P}}(G)^{(u)}$.

(ii) $\overline{\mathcal{P}}(G)^{(u)} \subseteq \Phi(\mathbf{Z})$, i.e. Φ is onto.

(iii) Φ is injective.

(iv) Φ is an algebraic homomorphism.

(v) Φ is continuous.

(vi) Φ^{-1} is continuous.

Let $H \leqq G$ be a compact open subgroup such that $\exp(G/H) = \infty$ which exists since G is of Type 3b.

(i) Trivial.

(ii) Take any $\varepsilon \in \overline{\mathcal{P}}(G)^{(u)}$ and any compact open $H' \leqq H$. By Lemma 5 we have $k_H, k_{H'} \in \mathbf{Z}$ such that

$$(k_H \operatorname{Id}_G - \varepsilon)(G) \subseteq H \quad \text{and} \quad (k_{H'} \operatorname{Id}_G - \varepsilon)(G) \subseteq H'.$$

First we prove $k_{H'} = k_H$. For every $g \in G$ we get

$$(k_H - k_{H'})(g+H) =$$
$$= (k_H \operatorname{Id}_G - \varepsilon)(g+H) + (\varepsilon - k_{H'} \operatorname{Id}_G)(g+H) \subseteq H + H' \subseteq H.$$

Since $\exp(G/H') = \exp(G/H) = \infty$ this yields $k_H = k_{H'} = k$. $H' \leqq H$ was arbitrary, hence

$$\varepsilon(g) - kg \in \bigcap \{H' \leqq H \mid H' \text{ compact and open}\} = \{0\}$$

by Proposition 1 c) and indeed $\varepsilon = \Phi(k)$.

(iii) As in the proof of (ii) $k \neq k'$ implies $k \operatorname{Id}_{G/H} \neq k' \operatorname{Id}_{G/H}$ and

$$\Phi(k) = k \operatorname{Id}_G \neq k' \operatorname{Id}_G = \Phi(k').$$

(iv) Trivial.

(v) Trivial, since $\Phi\colon \mathbf{Z} \to \overline{\mathcal{P}}(G)^{(u)}$ acts on a discrete space.

(vi) Suppose $\Phi(k^{(\nu)})(g) \to 0$ uniformly. Since H is open this implies $k^{(\nu)}G \subseteqq H$ or $k^{(\nu)}(g+H) = 0$ in G/H for all $g \in G$ and all $\nu \geqq \nu_0$. Together with $\exp(G/H) = \infty$ this yields $k^{(\nu)} = 0$ for $\nu \geqq \nu_0$, i.e. $k^{(\nu)} \to 0$ in $\mathbf{Z}$.

REFERENCES

[1] HEWITT, E. and ROSS, K. A., *Abstract harmonic analysis*, Vol. I: Structure of topological groups. Integration theory, group representations, Die Grundlehren der mathematischen Wissenschaften, Bd. 115, Academic Press, Inc., New York; Springer-Verlag, Berlin–Göttingen–Heidelberg, 1963. *MR* **28** #158

[2] KOWOL, G., Approximation durch Polynomfunktionen auf universellen Algebren, *Monatsh. Math.* **93** (1982), 15–32. *MR* **83e**:08007

[3] WINKLER, R., Polynomial approximation on locally compact abelian groups, *Studia Sci. Math. Hungar.* **28** (1993), 129–138.

[4] WINKLER, R., Algebraic endomorphisms of LCA-groups that preserve closed subgroups, *Geom. Dedicata* **43** (1992), 307–320.

(Received July 16, 1991)

INSTITUT FÜR ALGEBRA UND DISKRETE MATHEMATIK
TECHNISCHE UNIVERSITÄT WIEN
WIEDNER HAUPTSTRASSE 8–10
A–1040 WIEN
AUSTRIA

or

KOMMISSION FÜR MATHEMATIK
ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN
DR. IGNAZ SEIPELPLATZ 2
A–1010 WIEN
AUSTRIA

e-mail: rwin@ lezvax.oeaw.ac.at

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 275–281

ON THE SURFACE AREA OF CONVEX POLYTOPES

A. BEZDEK* and T. ÓDOR

Abstract

In this paper we prove that the surface area of a given three dimensional convex polyhedron is less then the sum of all products ef where e and f are disjoint edges of the polyhedra and ef denotes the product of lengths of the edges e and f. We generalize this statement for n-dimensional convex polytopes and state a conjecture.

The second author motivated by a result of H. G. Eggleston, B. Grünbaum and V. Klee [2] asked the following question:

Is it possible to estimate the surface area of a given three dimensional convex polyhedron by the sum of all products ef where e and f are disjoint edges (have no common point) of the polyhedron and ef denotes the product of lengths of the edges e and f? In this paper we give an affirmative answer (Theorem 1) and prove an analogous theorem for n-dimensional convex polytopes (Theorem 2).

Let k be an integer such that $0 < k \leqq d$. Denote the set of the k-dimensional faces of the given convex polytopes P by P_k. If F is a k-dimensional face of P, then $V(F)$ will denote the k-dimensional volume of F. Although the notation $V_k(F)$ would be more precise for $V(F)$, we believe that omitting the index k does not cause any confusion in this paper. In accordance with this notation $V(P)$ will denote the volume of the polytope P. Finally, $S(P)$ will denote the surface area of the polytope P, i.e. $S(P)$ is the sum of the $d-1$-dimensional volumes of the $d-1$-dimensional faces.

We are going to prove the following two theorems.

THEOREM 1. *If P is a 3-dimensional convex polyhedron with edge set P_1 and surface area $S(P)$, then*

$$S(P) \leqq \sum_{\substack{e,f \in P_1 \\ e \cap f = \emptyset}} ef. \tag{1}$$

1991 *Mathematics Subject Classification.* Primary 52A45.
Key words and phrases. Convex polytopes, volume, surface area.

*Partially supported by the Hungarian National Foundation for Scientific Research Grant No. 1238.

THEOREM 2. *If P is an n-dimensional convex polytope with edge set P_1, volume $V(P)$ and surface area $S(P)$, then*

$$V(P) \leqq v_n \left(\sum_{\substack{e,f \in P \\ e \cap f = \emptyset}} ef \right)^{\frac{n}{2}}, \tag{2}$$

where v_n denotes the volume of the n-dimensional unit sphere, and

$$S(P) \leqq \frac{3}{n} \left(\sum_{\substack{e,f \in P \\ e \cap f = \emptyset}} ef \right)^{\frac{n-1}{2}}. \tag{3}$$

REMARK 1. We do not believe that (1) and (2) are sharp, probably both in (1) and (2) $S(P)$ and $V(P)$ can be replaced by $xS(P)$ and $xV(P)$ with $x > 1$.

REMARK 2. It is known that v_n is maximal for $n = 5$ and v_n tends to zero as n increases.

REMARK 3. For tetrahedra we will prove a stronger inequality than (1). If T is a tetrahedron with pairs of opposite sides (a, a'), (b, b') and (c, c') and with surface area $S(T)$, then

$$\frac{3}{2} S(T) \leqq aa' + bb' + cc'. \tag{4}$$

REMARK 4. We believe that besides (2) the following general inequality holds. Given positive integers $a_1, a_2, \ldots, a_k$ and s such that $d + 1 \geqq k + a_1 + \cdots + a_k$ and $s \geqq a_1 + \cdots + a_k$, then there is a constant c not depending on the choice of P, such that

$$S_s(P) = \sum_{F \in P_S} V(F) \leqq c \left(\sum_{\substack{F_j \in P_{a_j} \\ F_i \cap F_j = \emptyset}} V(F_1) \ldots V(F_k) \right)^{\frac{s}{a_1 + \cdots + a_k}}. \tag{5}$$

Note that if $d = 3$; $s = k = 2$; $a_1 = a_2 = 1$ then (1) is a special case of (5). It would be interesting to prove (5) for the special case when $k = 2$; $a_1 + a_2 = d - 1$ and $s = a_1 + a_2$.

The second author has an argument to show that once the integers $a_1, \ldots, a_k$, s are given such that $s \geqq a_1 + \cdots + a_k$ and (5) holds for any poly-

eder P of the $d_0 = a_1 + \cdots + a_k + k - 1$-dimensional space, then it also holds for any polyeder P of the $d \geqq d_0$-dimensional space. He also claims that from these inequalities it follows that the famous conjecture "$S_s(P) \leqq (S_r(P))^{s/r}$ $(s > r)$" is true in some previously not considered cases, too.

We start with three lemmas.

LEMMA 1. *If the quadrangle q has pairs of opposite sides (a, a') and (b, b'), then*

$$2 \operatorname{area}(q) \leqq aa' + bb'. \tag{6}$$

PROOF. Denote the vertices of q by A, B, C and D so that $a = AB$ and $b = BC$. One of the diagonals, say DB, lies in q. Let A' be the image of the vertex A under the reflection around the perpendicular bisector of the diagonal DB. The quadrangle $ABCD$ has the same area as that of $A'BCD$ which is either the sum or the difference of the areas of the triangles $A'DC$ and $A'BC$. Thus $\operatorname{area}(q) = \operatorname{area}(A'BCD) \leqq (aa' + bb')/2$. □

For our purposes we need to extend the class of convex polygons of the Euclidean plane. Reader familiar with manifolds will recognize that we essentially define 2-manifolds using multiple layers of the Euclidean plane.

Members of the extended class C are either convex polygons or are generated by gluing together a finite number of convex polygons. Now we explain how and when can we glue together the convex polygons $p_1, \ldots, p_n$ and will also define the edge set and the area of the new polygons. Start with a polygon p_1. Glue the second convex polygon p_2 to p_1 along a complete edge, say e, so that p_1 and p_2 are on different side of e. We consider $q_2 = p_1 \cup p_2$ a member of the class C with edge set consisting of all edges of p_1 and p_2 excluding e. Two edges of q_2 are said to be adjacent edges, if they are adjacent edges of p_1 or p_2 or if they share an endpoint with the edge e. The area of q_2 is defined as the sum of the areas of p_1 and p_2. Suppose $p_1, \ldots, p_i$, $i \leqq n$ are already glued together and they determine the member q_i of C. Then we glue the convex polygon p_{i+1} to q_i, along a complete edge, say e, so that at least in a small neighbourhood of e the polygons q_i and p_{i+1} are on different side. Although some points of q_i and p_{i+1} may lie over each other, we do not consider them identical (we say that they belong to different layers of the plane). We consider $q_{i+1} = q_i \cup p_{i+1}$ a polygon with edge set consisting of all edges of q_i and p_{i+1} excluding the edge e. Two edges of q_{i+1} are said to be adjacent edges, if they are adjacent edges of q_i or p_{i+1} or if they share an endpoint with the edge e. The area of q_{i+1} is defined as the sum of the areas q_i and q_{i+1}.

We show that

LEMMA 2. *If p is a generalized polygon with edges $e_1, \ldots, e_n$ $(n \geqq 4)$, then*

$$2 \operatorname{area}(p) \leqq \sum_{e_i \cap e_j = \emptyset} e_i e_j. \tag{7}$$

PROOF. We prove (7) by induction on n. If $n = 4$, then (7) is the same as (6). Suppose $n > 4$ and we know (7) for all generalized polygons of at most $n - 1$ sides.

The generalized polygon p must have four adjacent sides, say e_1, e_2, e_3 and e_4 such that

$$e_1 + e_4 \geqq e_2 + e_3 . \tag{8}$$

Indirectly suppose the opposite is true for all consecutive quadruples of sides of p. Adding together the corresponding inequalities we have that $2\Sigma e_i < 2\Sigma e_i$, a contradiction. Let V_1, V_2 and V_3 be vertices of p such that $V_1V_2 = e_2$ and $V_2V_3 = e_3$. In view of (8) there is a point X such that the triangle $V_1V_2V_3$ do not overlap V_1XV_3 and $|V_1X| \leqq e_1$ and $|V_3X| \leqq e_4$.

Applying Lemma 1 for the quadrangle $V_1V_2V_3X$ we get

$$e_1e_3 + e_2e_4 \geqq |V_1X|e_3 + |V_3X|e_2 \geqq 2\,\text{area}\,(V_1V_2V_3X) \geqq 2\,\text{area}\,(V_1V_2V_3) . \tag{9}$$

We show that the segments e_1, V_1V_3, $e_4, \ldots, e_n$ are the edges of a generalized polygon p' of $n-1$ sides such that $\text{area}\,(p) \leqq \text{area}\,(p') + \text{area}\,(V_1V_2V_3)$. Without loss of generality we may assume that p is obtained by gluing together triangles only and let $t_1, \ldots, t_{n-2}$ be these triangles. Finally let i be the index so that $t_i = XV_1V_2$. The vertex X is either identical to V_3 or there is an index j such that $t_j = XV_2V_3$. In the first case p' is obtained by gluing together the triangles t_k, $1 \leqq k \leqq n-2$, $k \neq i$, in the second case p' is obtained by gluing together the triangles t_k, $1 \leqq k \leqq n-2$, $k \neq i, j$ and the triangle $t_{XP_1P_3}$. It is apparent that in both cases $\text{area}\,(p) \leqq \text{area}\,(p') + \text{area}\,(V_1V_2V_3)$.

According to our inductive assumption (7) holds for p'. Thus

$$\left(\sum_{\substack{e_i \cap e_j = \emptyset \\ i,j \notin \{2,3\}}} e_ie_j \right) + |V_1V_3|(e_5 + \cdots + e_n) \geqq 2\,\text{area}\,(p') . \tag{10}$$

By adding (9) and (10), and replacing $|V_1V_3|$ with the larger $e_2 + e_3$ we get (7). □

Let P be a convex 3-dimensional polyhedron with surface S. The union of some 2-faces of P is called a polygonal piece of the surface S (in short a polygonal shell) if it is simply connected subset of the surface S. We say the polygonal shell has n sides if its boundary on the surface S consists of n edges. A polygonal shell is said to be primitive if it does not contain vertices of P other then those belonging to its boundary.

LEMMA 3. *The surface of a convex polyhedron can be cut along its edges into primitive polygonal shells each having at least* 4 *sides.*

PROOF. We construct a desired partition. First form single element groups using all faces having at least 4 sides. Then turn to the triangular

faces and form two element groups so that triangles of one group have a common side. After exhausting the set of triangular faces in this respect we start to enlarge our groups by adding triangular faces to them maintaining the property that the union of the faces in one group form a primitive polygonal shell. The rule is that we add a triangular face t to a group if t shares an edge with one of the faces of the group and besides this edge t is disjoint from the group elements.

It is not hard to see that the procedure terminates only if all faces are used. □

PROOF OF THEOREM 1. According to Lemma 3 we can cut the surface of P along its edges into primitive polygonal shells $p_1 \dots p_k$ each having at least 4 sides. Each p_i $(i = 1, \dots, k)$ can be flattened in the following sense. Consider p_i as a framework where the faces are rigid plates so that the neighbouring faces can rotate freely around their common edge. It is easy to see that this framework can be brought into a unique position where all faces lie in the same plane and adjacent faces do not overlap (non adjacent faces might, but it is not our concern).

Notice that the flattened polygonal shells p_i $(i = 1, \dots, k)$ are generalized polygons. Applying Lemma 2 we get k inequalities. Since each product $e_i e_j$ occurs in at most two of them, adding these inequalities together we get (1). □

PROOF OF THEOREM 2. First we recall two well-known theorems:

1. For the volume $V(P)$ and the surface area $S(P)$ of an n-dimensional convex polytope P holds the isoperimetric inequality ([1], p.109):

$$\left(\frac{V(P)}{v_n}\right)^{n-1} \leqq \left(\frac{S(P)}{s_n}\right)^n , \tag{11}$$

where v_n and s_n are the volume and the surface area of the n-dimensional ball with radius one. Note that for any n

$$s_n = n v_n . \tag{12}$$

2. Let u be a unit vector and $\Pi_u(P)$ be the projection of the convex polytope P onto the hyperplane with normal vector u. Then the surface area $S(P)$ can be computed by the Cauchy formula ([1], p.48):

$$S(P) = \frac{1}{v_{n-1}} \int_B V(\Pi_u(P))\, du , \tag{13}$$

where B denotes the n dimensional unit sphere and du denotes the spherical area element.

We prove the inequality (2) by induction on n. If n is 3, then raising both sides of the inequality (1) to $\frac{3}{2}$ and applying (11) with $n = 3$ we get the desired inequality, with even a better constant what we need.

Suppose that (2) holds for any $n-1$ or less dimensional convex polytope. Using the inequalities (11), (12) and (13) we have that

$$(V(P))^{\frac{n-1}{n}} \leqq \frac{1}{v_{n-1}\sqrt[n]{v_n n}} \int_B V(\Pi_u(P))\, du\,.$$

Applying the inductive hypothesis for the quantity under the integral sign we get that:

$$(V(P))^{\frac{n-1}{n}} \leqq \frac{1}{\sqrt[n]{v_n n}} \int_B \left(\sum_{\substack{e,f \text{ are edges of } \Pi_u(P);\\ e\cap f=\emptyset}} ef \right)^{\frac{n-1}{2}} du\,.$$

Since the edges of $\Pi_u(P)$ are projections of certain edges of P the function to be integrated can be estimated from above by the constant $\sum ef$, where $e, f \in E$, $e \cap f = \emptyset$ and thus we end up with the inequality:

$$(V(P))^{\frac{n-1}{n}} \leqq \frac{s_n}{\sqrt[n]{v_n n}} \left(\sum_{\substack{e,f\in E,\\ e\cap f=\emptyset}} ef \right)^{\frac{n-1}{2}}\,.$$

In view of (12) $\frac{s_n}{\sqrt[n]{v_n n}} = (v_n)^{\frac{n-1}{n}}$. Raising both sides of the above inequality to the power $\frac{n}{n-1}$ we get (2).

If $n=3$ then (3) is the same as (1). If $n>3$, then using (13) we have

$$S(P) = \frac{1}{v_{n-1}} \int_B V(\Pi_u(P))\, du\,.$$

Applying (2) for the convex polytope $\Pi_u(P)$ we have

$$S(P) \leqq \frac{1}{v_{n-1}} \int_B \left(\sum_{\substack{e,f \text{ are edges of } \Pi_u(P)\\ e\cap f=\emptyset}} ef \right)^{\frac{n-1}{2}} du\,.$$

Since the edges of $\Pi_u(P)$ are projections of certain edges of P the function to be integrated can be estimated from above by the constant $(\sum ef)^{\frac{n-1}{n}}$, where $e, f \in E$, $e \cap f = \emptyset$ and thus we end up with the inequality:

$$S(P) \leqq \frac{S_n}{v_{n-1}} \left(\sum_{\substack{e,f\in E\\ e\cap f=\emptyset}} ef \right)^{\frac{n-1}{2}}\,. \qquad \square$$

PROOF OF REMARK 2. Cut the surface of the tetrahedron T into two pieces along the edges b, c', b' and c. By flattening the pieces out and using Lemma 1 we have that

$$bb' + cc' \geqq S .$$

Similarly we have that

$$bb' + aa' \geqq S$$

and

$$aa' + cc' \geqq S .$$

Adding these inequalities together we get (2). □

REFERENCES

[1] BONNESEN, T. and FENCHEL, W., *Theorie der konvexen Körper*, Springer-Verlag, Berlin–Heidelberg–New York, 1974. *MR* **49** #9736

[2] EGGLESTON, H. G., GRÜNBAUM, B. and KLEE, V., Some semicontinuity theorems for convex polytopes and cell complexes, *Comment. Math. Helv.* **39** (1964), 165–188. *MR* **30** #5217

[3] FIREY, W. and SCHNEIDER, R., The size of skeletons of convex bodies, *Geom. Dedicata* **8** (1979), 99–103. *MR* **80f**:52013

(Received July 31, 1991)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIAI TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 283–287

HOMOMORPHISMS OF DISTRIBUTIVE LATTICES AS RESTRICTION OF CONGRUENCES: THE PLANAR CASE

E. T. SCHMIDT

Given a lattice L and a sublattice L', then the map of $\operatorname{Con} L$ to $\operatorname{Con} L'$ determined by restriction is a meet-homomorphism preserving 0 and 1. If L' is a convex sublattice, then this map is a lattice homomorphism. G. Grätzer and H. Lakser [1] proved that any $\{0,1\}$-preserving homomorphism of finite distributive lattices can be realized by restricting the congruence lattice of some finite planar lattice L to the congruence lattice of an ideal L' of L. In this note we give a short proof of this result.

THEOREM. *Let D and D' be finite distributive lattices and let $\Psi: D \to D'$ be a $\{0,1\}$-preserving lattice homomorphism. Then there exist a finite planar lattice L, an ideal L' of L and lattice isomorphisms*

$$\rho: D \to \operatorname{Con} L\,, \quad \rho': D' \to \operatorname{Con} L'$$

such that $\Psi\rho'$ is the composition of ρ with the restriction of $\operatorname{Con} L$ to $\operatorname{Con} L'$. Moreover, the lattices L and L' have no nontrivial automorphisms (see Figure 1).

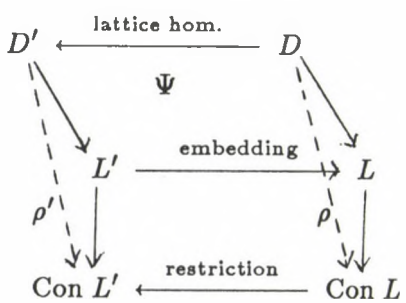


Fig. 1

PROOF. Let $\Psi: D \to D'$ be the given $\{0,1\}$-preserving homomorphism. By the duality between finite distributive lattices and finite posets Ψ determines an isotone map $\varphi: \mathcal{J}(D') \to \mathcal{J}(D)$. Conversely, Ψ is determined by φ.

1991 *Mathematics Subject Classification.* Primary 06B10; Secondary 08A30.
Key words and phrases. Restriction of congruences, lattice homomorphisms.

0081–6906/95/$ 4.00

Let T be the set $\mathcal{J}(D) \cup \mathcal{J}(D')$. We can extend φ to T by setting $x\varphi = x$ for $x \in \mathcal{J}(D)$. Denote $p_1, p_2, \ldots, p_m$ resp. $p_{m+1}, \ldots, p_n$ the elements of $\mathcal{J}(D')$ resp. $\mathcal{J}(D)$. φ can be characterized by a quasi-ordering $\preceq$ on T:

$$(*) \qquad p_i \preceq p_j \text{ if and only if } \begin{cases} p_i \leqq p_j \text{ in } \mathcal{J}(D'), & i, j \leqq m \text{ and} \\ p_i\varphi \leqq p_j\varphi \text{ in } \mathcal{J}(D) & \text{otherwise.} \end{cases}$$

It is easy to check that $\preceq$ is a quasi-ordering. Let Θ be the equivalence relation of T induced by this relation, i.e., $p_i \Theta p_j$ iff $p_i \preceq p_j$ and $p_j \preceq p_i$. Then T/Θ is a poset. By $(*)$ if $0 < i \leqq m$ then $p_i \preceq p_i\varphi$ and $p_i\varphi \preceq p_i$, i.e., $p_i \Theta p_i\varphi$. This implies $T/\Theta \cong \mathcal{J}(D)$.

We define two types of lattices A_{ij} and B_{ij}, $0 < j < i \leqq n$ by the diagrams illustrated in Figure 2. Let $\underline{n} = \{0 < 1 < \cdots < n\}$ be an $n+1$-element chain. A_{ij} is the direct product $\underline{n} \times \underline{2}$ augmented with the elements $c_i, c_j, c_{j-1}, \ldots, c_0$. B_{ij} is $\underline{n} \times \underline{2}$ augmented with the elements $d_j, d_i, d_{i+1}, \ldots, d_n$.

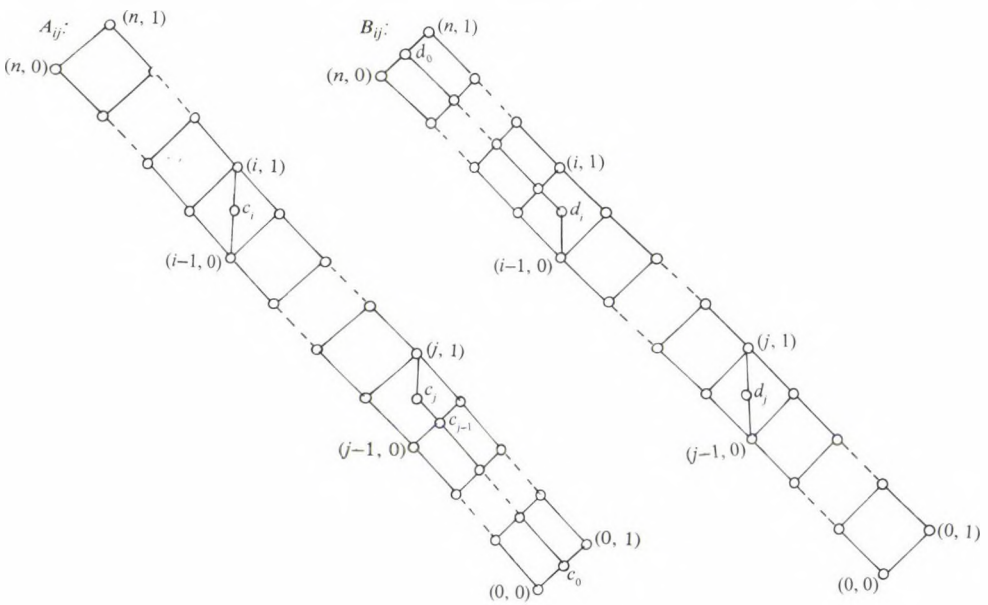


Fig. 2

Let $\mathcal{J}_{ij}$ be the ideal of A_{ij} (resp. B_{ij}) generated by $(n, 0)$ and let F_{ij} be the filter of A_{ij} (resp. B_{ij}) generated by $(0, 1)$. Then $\mathcal{J}_{ij} \cong F_{ij} \cong \underline{n}$. An easy computation shows that the following hold:

(i) Every congruence relation of A_{ij} (resp. B_{ij}) is determined by its restriction to $\mathcal{J}_{ij}$ and similarly to F_{ij}.

(ii) If $l \neq k$ then $(l-1, 0) \equiv (l, 0)$ forces $(k-1, 0) \equiv (k, 0)$ in A_{ij} iff $l = i$ and $k = j$.

(iii) For $l \neq k$ $(l-1,0) \equiv (l,0)$ forces $(k-1,0) \equiv (k,0)$ in B_{ij} iff $l = j$, $k = i$.

We define the lattice L. Consider the bijection $\sigma : [i-1, i] \to p_i$ between the prime intervals of $\underline{n}$ and the elements of T (σ is called a coloring of $\underline{n}$).

For $1 \leqq j < i \leqq n$ define

$$R_{ij} \cong \begin{cases} A_{ij} & \text{if } p_j \prec p_i \text{ in } T, \\ B_{ij} & \text{if } p_i \prec p_j \text{ in } T. \end{cases}$$

Order the pairwise disjoint R_{ij}-s say $R_{i_0j_0}, R_{i_1j_1}, \ldots, R_{i_sj_s}, \ldots$ such that $(i_0, j_0), (i_1, j_1), \ldots, (i_s, j_s)$ are exactly the pairs which satisfy $1 \leqq i_k, j_k \leqq m$ and either $p_{i_k} < p_{j_k}$ or $p_{j_k} < p_{i_k}$ in $\mathcal{J}(D')$. Now we apply the Hall–Dilworth gluing: the filter $F_{i_0j_0}$ of $R_{i_0j_0}$ is isomorphic to the ideal $\mathcal{J}_{i_1j_1}$ of $R_{i_1j_1}$. Identify $F_{i_0j_0}$ and $\mathcal{J}_{i_1j_1}$ via the isomorphism, we obtain the lattice $R_{i_0j_0} \cup$ $\cup R_{i_1j_1}$ which contains $F_{i_1j_1}$ as a filter. Then take $R_{i_2j_2}$ and its ideal $\mathcal{J}_{i_2j_2}$. We apply again the gluing construction, by identifying $F_{i_1j_1}$ and $\mathcal{J}_{i_2j_2}$. We continue this procedure, the resulting lattice is L. Then $\mathcal{J} = \mathcal{J}_{i_0j_0}$ is an ideal of L. $R_{i_0j_0}$ is isomorphic to one of the A_{ij}-s or B_{ij}-s, let k^* be the element of $\mathcal{J}_{i_0j_0} \subseteq R_{i_0j_0}$ which corresponds to the element $(k,0)$ of A_{ij} (or B_{ij}).

The properties (i), (ii) and (iii) imply that every congruence relation of L is determined by its restriction to $\mathcal{J}$ and $(j-1)^* \equiv j^*$ forces $(i-1)^* \equiv i^*$ $(i \neq j)$ in L iff $p_i \preceq p_j$ in T. Consequently, $\mathcal{J}(\operatorname{Con} L) \cong T/\Theta$ which proves $\operatorname{Con} L \cong D$.

Finally, we define the ideal L' of L. If D' is a Boolean lattice then $L' = \{0^* < 1^* < \cdots < s^*\}$.

$R_{i_sj_s}$ is isomorphic to one of the A_{ij}-s or B_{ij}-s. Denote $t \in R_{i_sj_s} \subseteq L$ the element which corresponds to $(0,1) \in A_{ij}$ (or B_{ij}) by this isomorphism, $m^* \in$ $\in \mathcal{J}_{i_0j_0}$ (m is the cardinality of $\mathcal{J}(D')$) and consider the ideal L' generated by the element $m^* \vee t$ (see Figure 3).

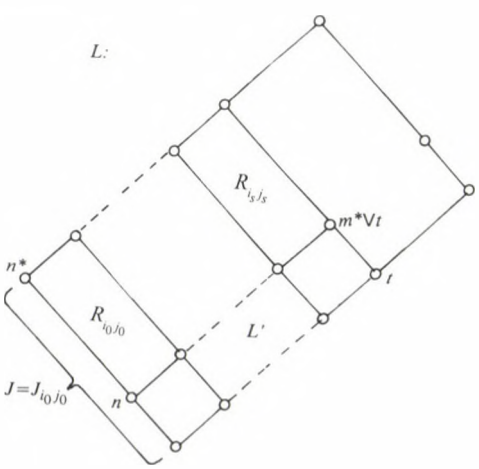


Fig. 3

By the given ordering of the "rows" R_{ij} we obtain that in $\mathcal{J} \cap L' = \mathcal{J}_{i_0 j_0} \cap L'$, $(j-1)^* \equiv j^*$ forces $(i-1)^* \equiv i^*$ $(i \neq j)$ iff $p_i < p_j$ in $\mathcal{J}(D')$, i.e., $\mathcal{J}(\operatorname{Con} L') \cong \mathcal{J}(D')$. This is equivalent to $\operatorname{Con} L' \cong D'$. It is clear that the restriction of $\operatorname{Con} L$ to $\operatorname{Con} L'$ is just the given $\{0,1\}$-preserving homomorphism Ψ.

If α is an arbitrary automorphism of L (and similarly of L') then its restriction to a "row" $R_{i_k j_k}$ is an automorphism of $R_{i_k j_k}$. Therefore we have only two special cases if α is a nontrivial homomorphism of $R_{i_k j_k}$. In these cases we modify the construction slightly.

(1) If $R_{i_0 j_0} \cong A_{ij}$ and $i = n$. Then the interval $[(n-1,0),(n,1)]$ is isomorphic to M_3. We replace this block by the lattice illustrated in Figure 4.

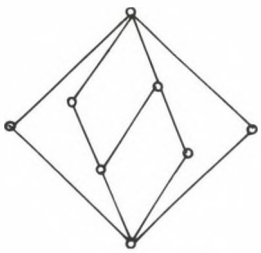


Fig. 4

If $[(n-1,0),(n,1)]$ is isomorphic to the lattice illustrated by Figure 5a, then replace this lattice defined by Figure 5b.

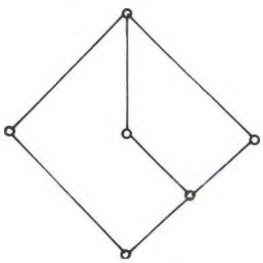


Fig. 5a

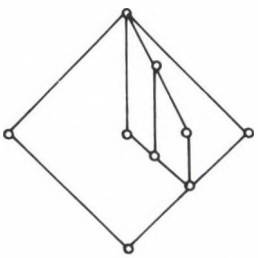


Fig. 5b

(2) We use the same modificated construction if $R_{i_k j_k}$ is the least row and $R_{i_k j_k} \cong A_{ij}$ or B_{ij} where $j = 1$ then the first block is again M_3 or the lattice illustrated by the dual of the lattice defined by Figure 5a.

REFERENCE

[1] GRÄTZER, G. and LAKSER, H., Homomorphisms of distributive lattices as restrictions of congruences II. Planarity and automorphisms, *Canadian J. Math.* **38** (1986), 1122–1134.

(Received August 2, 1991)

BUDAPESTI MŰSZAKI EGYETEM
KÖZLEKEDÉSMÉRNÖKI KAR
MATEMATIKA TANSZÉK
EGRY JÓZSEF U. 1
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 289–301

DENSE SUBSPACES OF QUASI-UNIFORM SPACES

H.-P. A. KÜNZI and A. LÜTHY

Abstract

A subspace D of a quasi-uniform space $(X,\mathcal{U})$ is said to be *doubly dense* in $(X,\mathcal{U})$ provided that it is dense both in $(X,\mathcal{U})$ and $(X,\mathcal{U}^{-1})$ and it is said to be *supdense* in $(X,\mathcal{U})$ provided that it is dense in $(X,\mathcal{U}^*)$. (Here, as usual, $\mathcal{U}^*$ denotes the coarsest uniformity on X finer than $\mathcal{U}$.) For various properties $\mathcal{O}$ we study variants of the following problem which originates in the theory of completing quasi-uniform spaces. If $(X,\mathcal{U})$ is a quasi-uniform space having a doubly dense (resp. supdense) subspace with a given property $\mathcal{O}$, does $(X,\mathcal{U})$ necessarily have property $\mathcal{O}$, too? Of course, the answers to these questions are negative in general. We show, however, that they are positive for several important properties $\mathcal{O}$ of quasi-uniform spaces.

1. Introduction and preliminary results

Given a quasi-uniform T_0-space $(X,\mathcal{U})$ having some nice property $\mathcal{O}$ it is often interesting to know, whether its bicompletion $(\widetilde{X},\widetilde{\mathcal{U}})$ has property $\mathcal{O}$, too. Simple examples show that our problem has a negative solution in general:

EXAMPLE 1. (a) Recall that a quasi-uniform space $(X,\mathcal{U})$ is said to be *point-symmetric* (resp. *locally symmetric* [12, pages 36 and 37]) provided that for each $U \in \mathcal{U}$ and each $x \in X$ there exists a symmetric $V \in \mathcal{U}$ such that $V(x) \subseteq U(x)$ (resp. $V^2(x) \subseteq U(x)$). Consider an arbitrary non-discrete (topological) T_3-space equipped with its Pervin quasi-uniformity $\mathcal{P}$. Then $(X,\mathcal{P})$ is locally symmetric [12, p. 37]. However, the following argument shows that the bicompletion $(\widetilde{X},\widetilde{\mathcal{P}})$ of $(X,\mathcal{P})$ is not point-symmetric, although it contains $(X,\mathcal{P})$ as a $\mathcal{T}(\widetilde{\mathcal{P}}^*)$-dense subspace [12, Theorem 3.33]: Point-symmetry of $(\widetilde{X},\widetilde{\mathcal{P}})$ implies that $\mathcal{T}(\widetilde{\mathcal{P}})$ is a T_1-topology [12, p. 36]. It follows that $\mathcal{P}$ is a uniformity [13, Proposition 2.1] and that X is a discrete topological space [12, Corollary 2.35] — a contradiction.

1991 *Mathematics Subject Classification.* Primary 54E15, 54E05, 54D35.

Key words and phrases. Quasi-uniformity, bicomplete, Lebesgue property, quiet, complete, precompact, transitive, symmetric, stable.

The first author completed the final version of this paper at the University of Oxford (GB) while he was supported by the Swiss National Science Foundation under Grant 8220-028387.

(b) Recall that a quasi-uniform space $(X,\mathcal{U})$ is called *uniformly regular* (see e.g. [14]) if for each $U \in \mathcal{U}$ there is a $V \in \mathcal{U}$ such that $\mathrm{cl}_{\mathcal{T}(\mathcal{U})} V(x) \subseteq U(x)$ whenever $x \in X$. Consider the totally bounded quasi-uniform T_0-space $(X,\mathcal{V})$ described in Example 8 of [20]. Both $\mathcal{V}$ and $\mathcal{V}^{-1}$ are uniformly regular, but $\mathcal{V}$ is not a uniformity. By [13, Proposition 2.1] the bicompletion $(\widetilde{X},\widetilde{\mathcal{V}})$ of $(X,\mathcal{V})$ is not a T_1-space. Hence $(\widetilde{X},\widetilde{\mathcal{V}})$ cannot be (uniformly) regular.

It is the aim of the present paper to show that — nevertheless — our questions have (at least partial) positive answers for surprisingly many important properties of quasi-uniform spaces, among them various conditions of completeness and precompactness and, astonishingly, even some conditions of symmetry.

Notation and terminology of this note coincide with that of reference [12]. Additionally we shall make use of the following concepts. A subset D of a bitopological space $(X,\mathcal{T}_1,\mathcal{T}_2)$ is called *doubly dense* (compare [6]) in X if D is dense both in $(X,\mathcal{T}_1)$ and $(X,\mathcal{T}_2)$. It is said to be *supdense* in X if it is dense in $(X,\sup\{\mathcal{T}_1,\mathcal{T}_2\})$. Similarly, a subspace D of a quasi-uniform space $(X,\mathcal{U})$ is said to be *doubly dense* in $(X,\mathcal{U})$, if D is both $\mathcal{T}(\mathcal{U})$-dense and $\mathcal{T}(\mathcal{U}^{-1})$-dense in X. If D is $\mathcal{T}(\mathcal{U}^*)$-dense in X, it is called *supdense* in $(X,\mathcal{U})$. Clearly, each supdense subspace of a quasi-uniform space $(X,\mathcal{U})$ is doubly dense. Examples show that the converse does not obtain in general. Of course the converse holds if the two topologies $\mathcal{T}(\mathcal{U})$ and $\mathcal{T}(\mathcal{U}^{-1})$ are comparable. The following lemma contains an elementary fact about supdense subspaces of quasi-uniform spaces. It is related to Lemma 3 of [18] and should be compared with the remark made in [17] after the proof of Lemma 6.

LEMMA 1. *Let $\mathcal{U}$ and $\mathcal{V}$ be quasi-uniformities on a set X such that $\mathcal{T}(\mathcal{U}) = \mathcal{T}(\mathcal{V})$ and $\mathcal{T}(\mathcal{U}^{-1}) = \mathcal{T}(\mathcal{V}^{-1})$. If $\mathcal{U}|D = \mathcal{V}|D$ where D is supdense in $(X,\mathcal{U})$, then $\mathcal{U} = \mathcal{V}$.*

PROOF. See [4, §11]. (In the proof of [6, Lemma 2.5] a similar idea is used.)

REMARK 1 (added during revision). The authors would like to thank Professor J. Deák for informing them about the extensive work done on extensions of quasi-uniform spaces in Hungary recently [1,3,4,5,6,7]. At several places in this note we shall make use of the notation and terminology introduced in these papers. In particular, the reader might wish to study §11 of [4] (dealing with firm extensions) before reading on.

Finally let us note that the following simple construction can be used to show that for many properties $\mathcal{O}$ our problem has a negative solution in the case of doubly dense subspaces.

EXAMPLE 2. Let $(E,\mathcal{V})$ be an arbitrary quasi-uniform space and let $-\infty$ and ∞ be two points not contained in the set E. Let $X = E \cup \{-\infty,\infty\}$ and

let $\mathcal{U}$ be the quasi-uniformity on X that is generated by $\left\{[\{-\infty\}\times X]\cup V\cup \cup[X\times\{\infty\}]: V\in\mathcal{V}\right\}$. Then $\{-\infty,\infty\}$ is doubly dense in $(X,\mathcal{U})$.

2. Completeness conditions

We recall that a filter $\mathcal{F}$ on a quasi-uniform space $(X,\mathcal{U})$ is called a $\mathcal{U}$-*Cauchy filter* on X if for each $U\in\mathcal{U}$ there is an $x\in X$ such that $U(x)\in\mathcal{F}$ [12, p. 47]. A quasi-uniform space $(X,\mathcal{U})$ is said to be *(convergence) complete* if each $\mathcal{U}$-Cauchy filter on X has a $\mathcal{T}(\mathcal{U})$-cluster point (a $\mathcal{T}(\mathcal{U})$-limit point) in X [12, p. 50].

In [8,9] D. Doitchinov introduces a different notion of completeness for a quasi-uniform space $(X,\mathcal{U})$. As usual, let us call a filter $\mathcal{G}$ on X a D-*Cauchy filter* provided that there exists a filter $\mathcal{F}$ on X so that for each $U\in\mathcal{U}$ there are $F\in\mathcal{F}$ and $G\in\mathcal{G}$ such that $F\times G\subseteqq U$. (In this case one writes $(\mathcal{F},\mathcal{G})\to 0$ and calls $(\mathcal{F},\mathcal{G})$ a *Cauchy pair of filters*.) The space $(X,\mathcal{U})$ is called D-*complete* provided that each D-Cauchy filter on X converges in $(X,\mathcal{U})$. Furthermore, $(X,\mathcal{U})$ is said to be *strongly* D-*complete* provided that if $(\mathcal{F},\mathcal{G})\to 0$, then $\mathcal{F}$ has a $\mathcal{T}(\mathcal{U})$-cluster point in X [15]. A quasi-uniform space $(X,\mathcal{U})$ is said to have the *Lebesgue property* [12, p. 97] if for each $\mathcal{T}(\mathcal{U})$-open cover $\mathcal{C}$ of X there is $U\in\mathcal{U}$ such that $\{U(x): x\in X\}$ is a refinement of $\mathcal{C}$.

PROPOSITION 1. *Let D be a supdense subspace of a quasi-uniform space $(X,\mathcal{U})$. Then $(X,\mathcal{U})$ is (convergence) complete if $(D,\mathcal{U}|D)$ is (convergence) complete.*

PROOF. Assume that $(D,\mathcal{U}|D)$ is (convergence) complete. Let $\mathcal{F}$ be a $\mathcal{U}$-Cauchy filter on X and let $\mathcal{F}_0$ be the filter on D generated by the filterbase $\{F\cap D: F\in\mathcal{T}(\mathcal{U}^*)\cap\mathcal{F}\}$. Consider an arbitrary $U\in\mathcal{U}$. Choose $W\in\mathcal{U}$ such that $W^2\subseteqq U$. Since $\mathcal{F}$ is a $\mathcal{U}$-Cauchy filter on X, there is an $x\in X$ such that $W(x)\in\mathcal{F}$. Because D is $\mathcal{T}(\mathcal{U}^*)$-dense in X, there exists $y\in D\cap W^{-1}(x)$. Then $W(x)\subseteqq W^2(y)\subseteqq U(y)\in\mathcal{F}$ and $(U|D)(y)=U(y)\cap D\in\mathcal{F}_0$. Thus $\mathcal{F}_0$ is a $\mathcal{U}|D$-Cauchy filter on D. Since $(D,\mathcal{U}|D)$ is (convergence) complete, there is a $z\in D$ such that z is a $\mathcal{T}(\mathcal{U}|D)$-cluster point (a $\mathcal{T}(\mathcal{U}|D)$-limit point) of $\mathcal{F}_0$.

In the first case let us show that z is a $\mathcal{T}(\mathcal{U})$-cluster point of $\mathcal{F}$. Assume the contrary. Then there is an $F\in\mathcal{F}$ such that $z\notin \mathrm{cl}_{\mathcal{T}(\mathcal{U})}F$. Choose $V\in\mathcal{U}$ such that $V(z)\cap V^{-1}(F)=\emptyset$. Since $V^{-1}(F)\cap D\in\mathcal{F}_0$ and z is a $\mathcal{T}(\mathcal{U}|D)$-cluster point of $\mathcal{F}_0$, we have reached a contradiction. Hence $(X,\mathcal{U})$ is complete provided that $(D,\mathcal{U}|D)$ is complete.

In the second case let us show that z is a $\mathcal{T}(\mathcal{U})$-limit point of $\mathcal{F}$. Let $V\in\mathcal{U}$. Choose $W\in\mathcal{U}$ such that $W^2\subseteqq V$. Since $(W|D)(z)\in\mathcal{F}_0$, there is a $G\in\in\mathcal{T}(\mathcal{U}^*)\cap\mathcal{F}$ such that $G\cap D\subseteqq W(z)$. Hence $G\subseteqq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}G=$ $=\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}(G\cap D)\subseteqq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}W(z)\subseteqq W^2(z)\subseteqq V(z)$, since D is $\mathcal{T}(\mathcal{U}^*)$-dense in

X and G is $\mathcal{T}(\mathcal{U}^*)$-open (see [11, Theorem 1.3.6]). Thus $V(z) \in \mathcal{F}$ and $\mathcal{F}$ converges to z in $(X,\mathcal{U})$. We have shown that $(X,\mathcal{U})$ is convergence complete provided that $(D,\mathcal{U}|D)$ is convergence complete.

The referee points out that our proof shows (in the light of Lemma 1.1 of [1]) that the statement on convergence completeness in Proposition 1 remains valid if one assumes only that D is doubly dense in $(X,\mathcal{U})$ and $\mathcal{T}(\mathcal{U})$ is a strict extension of $\mathcal{T}(U|D)$ (compare [4, 11.1]). He also observes that Proposition 2(a) can be generalized in the same way (see [1, Theorem 1.3]).

PROPOSITION 2. (a) *Let D be a supdense subspace of a quasi-uniform space $(X,\mathcal{U})$. If $(D,\mathcal{U}|D)$ is* D*-complete, then $(X,\mathcal{U})$ is* D*-complete.*

(b) *Let D be a doubly dense subspace of a quasi-uniform space $(X,\mathcal{U})$. If $(D,\mathcal{U}|D)$ is strongly* D*-complete, then $(X,\mathcal{U})$ is strongly* D*-complete.*

PROOF. We prove these two results simultaneously. Suppose that $(D,\mathcal{U}|D)$ is (strongly) D-complete. Let $(\mathcal{F},\mathcal{G})$ be a Cauchy pair of filters on $(X,\mathcal{U})$, let $\mathcal{F}_0$ be the filter on D generated by the filterbase $\{F \cap D: F \in \in \mathcal{T}(\mathcal{U}^{-1}) \cap \mathcal{F}\}$ and let $\mathcal{G}_0$ be the filter on D generated by the filterbase $\{G \cap D: G \in \mathcal{T}(\mathcal{U}) \cap \mathcal{G}\}$.

First we note that $(\mathcal{F}_0, \mathcal{G}_0) \to 0$: Let $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^3 \subseteq U$. Since $(\mathcal{F},\mathcal{G}) \to 0$, there exist $F \in \mathcal{F}$ and $G \in \mathcal{G}$ such that $F \times \times G \subseteq W$. Thus $[W^{-1}(F) \cap D] \times [W(G) \cap D] \subseteq U|D$, $W^{-1}(F) \cap D \in \mathcal{F}_0$ and $W(G) \cap D \in \mathcal{G}_0$. Hence $(\mathcal{F}_0, \mathcal{G}_0) \to 0$.

Since $(D,\mathcal{U}|D)$ is (strongly) D-complete, $\mathcal{G}_0$ (resp. $\mathcal{F}_0$) has a $\mathcal{T}(\mathcal{U}|D)$-limit point (resp. a $\mathcal{T}(\mathcal{U}|D)$-cluster point) z in D. An argument similar to the one given in the final two paragraphs of the proof of Proposition 1 shows that z is a $\mathcal{T}(\mathcal{U})$-limit point of $\mathcal{G}$ (resp. a $\mathcal{T}(\mathcal{U})$-cluster point of $\mathcal{F}$) in X. It follows that $(X,\mathcal{U})$ is (strongly) D-complete.

EXAMPLE 3. Let $X = \mathbf{R} \cup \{\infty\}$. Furthermore, let $\mathcal{U}$ be the quasi-uniformity on X generated by $\{U_\epsilon: \epsilon > 0\}$ where $U_\epsilon = (\bigcup_{x \in \mathbf{R}} \{x\} \times [x, x+\epsilon[) \cup \cup]-\epsilon, 0[\times\{\infty\}) \cup (\{\infty\}\times]0,\epsilon[) \cup \{(\infty,\infty)\}$ whenever $\epsilon > 0$. Clearly $D = \mathbf{R}$ is a doubly dense subspace of $(X,\mathcal{U})$ which is D-complete (for a proof see for instance Example 1(b) of [21]). However, the filter $\mathcal{F}$ on X generated by $\{[0,\epsilon[\cup\{\infty\}: \epsilon > 0\}$ is a D-Cauchy filter on $(X,\mathcal{U})$ that is not convergent. Hence $(X,\mathcal{U})$ is not D-complete.

REMARK 2. Note that a quasi-uniform T_2-space $(X,\mathcal{U})$ cannot have *proper* doubly dense complete (resp. D-complete) subspaces: Assume that D is a doubly dense complete (resp. D-complete) subspace of $(X,\mathcal{U})$ such that there exists $x \in X \setminus D$. Since $(U^{-1}(x) \cap D) \times (U(x) \cap D) \subseteq U^2$ whenever $U \in \mathcal{U}$, it is clear that the filter $\mathcal{F}$ generated on D by $\{U(x) \cap D: U \in \mathcal{U}\}$ is a D-Cauchy filter (and thus a $\mathcal{U}|D$-Cauchy filter) on D. Obviously, however, $\mathcal{F}$ cannot have a cluster point in $(D,\mathcal{U}|D)$, because $(X,\mathcal{U})$ is a T_2-space — a contradiction.

PROPOSITION 3. *Let D be a supdense subspace of a quasi-uniform space*

$(X,\mathcal{U})$. *If* $(D,\mathcal{U}|D)$ *has the Lebesgue property, then* $(X,\mathcal{U})$ *has the Lebesgue property.*

PROOF. We show that as in the preceding results on convergence completeness, it is sufficient in Proposition 3 to assume only that D is doubly dense in $(X,\mathcal{U})$ and $\mathcal{T}(\mathcal{U})$ is a strict extension of $\mathcal{T}(\mathcal{U}|D)$. (Then the sets $s(G)=\{x\in X$: There is $U\in\mathcal{U}$ such that $U(x)\cap D\subseteqq G\}$ whenever G is open in $\mathcal{T}(\mathcal{U}|D)$ form a base for $\mathcal{T}(\mathcal{U})$; see e.g. [1].): Suppose that $(D,\mathcal{U}|D)$ has the Lebesgue property. Let $\mathcal{C}$ be a $\mathcal{T}(\mathcal{U})$-open cover of X. Set $\mathcal{H}=\{H\in \in\mathcal{T}(\mathcal{U}|D)\colon s(H)\subseteqq C$ for some $C\in\mathcal{C}\}$. Clearly $\mathcal{H}$ is a $\mathcal{T}(\mathcal{U}|D)$-open cover of D. Since $(D,\mathcal{U}|D)$ has the Lebesgue property, there is a $U\in\mathcal{U}$ such that $\{(U|D)(x)\colon x\in D\}$ refines $\{H\colon H\in\mathcal{H}\}$. Let $W\in\mathcal{U}$ be such that $W^2\subseteqq U$. Without loss of generality we assume that $U(y)$ is $\mathcal{T}(\mathcal{U})$-open for each $y\in \in X$ (cf. [12, p. 3]). Fix $x\in X$. Since D is doubly dense in X, there is a $y\in D\cap W^{-1}(x)$. Then $W(x)\subseteqq W^2(y)\subseteqq U(y)\subseteqq s(U(y)\cap D)\subseteqq s(H)\subseteqq C$ for some $H\in\mathcal{H}$ and some $C\in\mathcal{C}$ by definition of $\mathcal{H}$. Hence $\{W(x)\colon x\in X\}$ is a refinement of $\mathcal{C}$. We have shown that $(X,\mathcal{U})$ has the Lebesgue property.

EXAMPLE 4. Let X be an orthocompact topological T_0-space and let $\mathcal{U}$ be the fine transitive quasi-uniformity of X. Then the bicompletion $(\tilde{X},\tilde{\mathcal{U}})$ is orthocompact: Since, by [12, Theorem 5.6], $(X,\mathcal{U})$ and thus, by Proposition 3, also $(\tilde{X},\tilde{\mathcal{U}})$ have the Lebesgue property, and since, by an argument similar to Corollary 5 of [18], $\tilde{\mathcal{U}}$ is the fine transitive quasi-uniformity of $(\tilde{X},\mathcal{T}(\tilde{\mathcal{U}}))$, the assertion follows from [12, Theorem 5.6].

EXAMPLE 5. Let $[0,1]$ be the unit interval of real numbers equipped with its usual (unique) uniformity $\mathcal{U}$ and let $D=\mathbf{Q}\cap[0,1]$ where $\mathbf{Q}$ denotes the set of rationals. Then D equipped with the subspace uniformity induced by $\mathcal{U}$ on D has none of the completeness properties studied in this section, although D is $\mathcal{T}(\mathcal{U})$-dense in $[0,1]$ and $([0,1],\mathcal{U})$ satisfies all the five completeness properties considered above.

3. Compactness conditions

In this section we consider several variants of precompactness. While all these conditions are equivalent in the realm of uniform spaces, they may differ considerably in the class of quasi-uniform spaces. Recall that a quasi-uniform space $(X,\mathcal{U})$ is called *Cauchy bounded* [16] if each ultrafilter on X is a D-Cauchy filter. It is called *precompact* [12, p. 51] if for each $V\in\mathcal{U}$ there is a finite subset F of X such that $V(F)=X$ and it is called *totally bounded* if for each $V\in\mathcal{U}$ there is a finite cover $\{A_i\colon i=1,\dots,n\}$ of X such that $(A_i\times A_i)\subseteqq V$ whenever $i=1,\dots,n$ [12, p. 12]. It is well known (and easy to see) that each totally bounded quasi-uniform space is Cauchy bounded and that each Cauchy bounded quasi-uniform space is precompact.

PROPOSITION 4. (a) *Let D be a supdense subspace of a quasi-uniform space $(X,\mathcal{U})$. Then $(D,\mathcal{U}|D)$ is hereditarily precompact (totally bounded) if and only if $(X,\mathcal{U})$ is hereditarily precompact (totally bounded).*

(b) *Let D be a $\mathcal{T}(\mathcal{U}^{-1})$-dense subspace of a quasi-uniform space $(X,\mathcal{U})$. Then $(D,\mathcal{U}|D)$ is Cauchy bounded if and only if $(X,\mathcal{U})$ is Cauchy bounded.*

(c) *Let D be a $\mathcal{T}(\mathcal{U}^{-1})$-dense subspace of a quasi-uniform space $(X,\mathcal{U})$. Then $(D,\mathcal{U}|D)$ is precompact if and only if $(X,\mathcal{U})$ is precompact.*

PROOF. (a) If $(X,\mathcal{U})$ is hereditarily precompact (totally bounded), then $(D,\mathcal{U}|D)$ is hereditarily precompact (totally bounded), because each subspace of a hereditarily precompact (totally bounded) quasi-uniform space is hereditarily precompact (totally bounded) [12, p. 12]. Since a quasi-uniform space $(X,\mathcal{V})$ is totally bounded if and only if both $(X,\mathcal{V})$ and $(X,\mathcal{V}^{-1})$ are hereditarily precompact [19, Lemma 1.1], it remains to prove only that the quasi-uniform space $(X,\mathcal{U})$ is hereditarily precompact provided that it has a supdense hereditarily precompact subspace D: Let $A \subseteq X$ and $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^3 \subseteq U$. For each $a \in A$ there exists $d_a \in (W \cap W^{-1})(a) \cap D$. Since $B = \{d_a : a \in A\}$ is a precompact subspace of D, there is a finite subset F of A such that $B \subseteq \bigcup_{f \in F} W(d_f)$. Let $a \in A$. Then $a \in W(d_a) \subseteq W^2(d_f) \subseteq$ $\subseteq W^3(f) \subseteq U(f)$ for some $f \in F$. Hence $(X,\mathcal{U})$ is hereditarily precompact.

(b) Suppose that $(D,\mathcal{U}|D)$ is Cauchy bounded. Let $\mathcal{G}$ be an ultrafilter on X and let $\mathcal{G}_0$ be an ultrafilter on D containing the filterbase $\{G \cap D : G \in$ $\in \mathcal{T}(U^{-1}) \cap \mathcal{G}\}$. Since $(D,\mathcal{U}|D)$ is Cauchy bounded, there is a filter $\mathcal{F}$ on D such that $(\mathcal{F},\mathcal{G}_0) \to 0$. Let $\mathcal{E}$ be the filter on X generated by the filterbase $\mathcal{F}$ on X. Consider an arbitrary entourage $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^2 \subseteq$ $\subseteq U$. Since $(\mathcal{F},\mathcal{G}_0) \to 0$, there are $F \in \mathcal{F}$ and $H \in \mathcal{G}_0$ such that $F \times H \subseteq W|D$. If $X \setminus W(H) \in \mathcal{G}$, then $W^{-1}(X \setminus W(H)) \cap D \in \mathcal{G}_0$, but $W^{-1}(X \setminus W(H)) \cap$ $\cap H = \emptyset$ — a contradiction. Since $\mathcal{G}$ is an ultrafilter on X, we conclude that $W(H) \in \mathcal{G}$. Furthermore, $F \times W(H) \subseteq U$ and $F \in \mathcal{E}$. It follows that $(\mathcal{E},\mathcal{G}) \to 0$. Hence $(X,\mathcal{U})$ is Cauchy bounded.

In order to prove the converse assume that $(X,\mathcal{U})$ is Cauchy bounded. Let $\mathcal{G}$ be an ultrafilter on D and denote by $\mathcal{H}$ the ultrafilter on X generated by the filterbase $\mathcal{G}$ on X. Since $(X,\mathcal{U})$ is Cauchy bounded, there is a filter $\mathcal{F}$ on X such that $(\mathcal{F},\mathcal{H}) \to 0$. Let $\mathcal{F}_0$ be the filter on D generated by the filterbase $\{F \cap D : F \in \mathcal{T}(\mathcal{U}^{-1}) \cap \mathcal{F}\}$. Consider an arbitrary $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^2 \subseteq U$. Since $(\mathcal{F},\mathcal{H}) \to 0$, there exist $F \in \mathcal{F}$ and $H \in \mathcal{H}$ such that $F \times H \subseteq W$. Then $(W^{-1}(F) \cap D) \times (H \cap D) \subseteq U|D$, $W^{-1}(F) \cap D \in$ $\in \mathcal{F}_0$ and $H \cap D \in \mathcal{G}$. Thus $(\mathcal{F}_0,\mathcal{G}) \to 0$. Hence $(D,\mathcal{U}|D)$ is Cauchy bounded.

(c) Assume that $(X,\mathcal{U})$ is precompact. Let $U \in \mathcal{U}$ and choose $W \in \mathcal{U}$ such that $W^2 \subseteq U$. There is a finite subset F of X such that $\bigcup\{W(f) : f \in F\} =$ $= X$. For each $f \in F$ choose $d_f \in D \cap W^{-1}(f)$. Thus $X \subseteq \bigcup\{W^2(d_f) : f \in$ $\in F\} \subseteq \bigcup\{U(d_f) : f \in F\}$. It follows that $(D,\mathcal{U}|D)$ is precompact.

In order to prove the converse we assume that $(D,\mathcal{U}|D)$ is precompact. Let $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^2 \subseteq U$. Since $(D,\mathcal{U}|D)$ is precompact,

there is a finite subset F of D such that $D \subseteq \bigcup\{W(f): f \in F\}$. Then $X = \mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})} \bigcup\{W(f): f \in F\} = \bigcup\{\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})} W(f): f \in F\} \subseteq \bigcup\{W^2(f): f \in F\} \subseteq \subseteq \bigcup\{U(f): f \in F\}$. Hence $(X,\mathcal{U})$ is precompact.

PROPOSITION 5. *Assume that $(X,\mathcal{U})$ is a quasi-uniform space with a $\mathcal{T}(\mathcal{U}^{-1})$-dense compact subspace D. Then $(X,\mathcal{U})$ is compact.*

PROOF. Let $\mathcal{C}$ be a $\mathcal{T}(\mathcal{U})$-open cover of X. Set $\mathcal{H}=\{H \in \mathcal{T}(\mathcal{U}): \mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})} H \subseteq \subseteq C \text{ for some } C \in \mathcal{C}\}$. Since $\mathcal{H}$ is a $\mathcal{T}(\mathcal{U})$-open cover of X and since $(D,\mathcal{U}|D)$ is compact, there is a finite subcollection $\mathcal{H}'$ of $\mathcal{H}$ such that $D \subseteq \cup\mathcal{H}'$. Thus $X = \mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}(\cup\mathcal{H}') = \bigcup\{\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})} H': H' \in \mathcal{H}'\} \subseteq \cup\mathcal{C}'$ for some finite subcollection $\mathcal{C}'$ of $\mathcal{C}$. Hence $(X, \mathcal{T}(\mathcal{U}))$ is compact.

4. Symmetry conditions

In this section we show that the studied questions have positive answers even for some kinds of symmetry conditions. Of course, in view of the examples presented in the introduction we cannot expect too much.

We begin by discussing some auxiliary results that seem to be of independent interest. Let us recall that the *weight* (cf. [11, p. 427]) of a quasi-uniform space $(X,\mathcal{U})$ is the minimal cardinal number of a base for $\mathcal{U}$.

PROPOSITION 6. *Let D be a supdense subspace of a quasi-uniform space $(X,\mathcal{U})$. Then the two spaces $(D,\mathcal{U}|D)$ and $(X,\mathcal{U})$ have the same weight.*

PROOF. The assertion is an immediate consequence of [4, Theorem 11.2]. Let us mention that the nontrivial part of the statement also follows from the fact that $\{\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} B: B \in \mathcal{B}\}$ is a base for $(X,\mathcal{U})$ whenever $\mathcal{B}$ is a base for $(D,\mathcal{U}|D)$: Consider $U \in \mathcal{U}$. Choose $V \in \mathcal{U}$ such that $V^3 \subseteq U$. There is a $B \in \mathcal{B}$ such that $B \subseteq V|D$, because $\mathcal{B}$ is a base for $\mathcal{U}|D$. Since $\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)}(V|D) \subseteq V^3$, we have that $\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} B \subseteq U$. It remains to show that $\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} B \in \mathcal{U}$. Choose $L \in \mathcal{U}$ such that $L|D \subseteq B$. Furthermore, choose a $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open entourage $H \in \mathcal{U}$ such that $H \subseteq L$ [12, Corollary 1.17]. Then $H \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} H \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)}(H|D) \subseteq \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)}(L|D) \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} B$ by [11, Theorem 1.3.6] and thus $\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)\times\mathcal{T}(\mathcal{U}^*)} B \in \mathcal{U}$.

A quasi-uniformity $\mathcal{U}$ on X is called *Smyth symmetric* [15] provided that whenever A and B are $\mathcal{T}(\mathcal{U})$-open sets and $U \in \mathcal{U}$ such that $U(A) \subseteq B$, there are $\mathcal{T}(\mathcal{U})$-open sets A' and B' and $V \in \mathcal{U}$ such that $A \cap A' = \emptyset$, $B \cup B' = X$ and $V(B') \subseteq A'$. It is known that a quasi-uniformity $\mathcal{U}$ is Smyth symmetric if and only if its quasi-proximity is a proximity [15].

LEMMA 2. *Let D be a doubly dense subspace of a quasi-uniform space $(X,\mathcal{U})$ such that $\mathcal{U}|D$ is Smyth symmetric. Then D is supdense in $(X,\mathcal{U})$.*

PROOF. Assume the contrary. Then there exist $a \in X$ and $U \in \mathcal{U}$ such that $(U \cap U^{-1})(a) \cap D = \emptyset$. Choose a $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open entourage W of $\mathcal{U}$ [12, Corollary 1.17] such that $W^2 \subseteq U$. Then $(W|D) \cap [(W(a) \cap D) \times (W^{-1}(a) \cap D)] = \emptyset$ because $W(a) \cap W^{-2}(a) \cap D = \emptyset$. Since $\mathcal{U}|D$ is Smyth symmetric, there exists a $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open entourage $H \in \mathcal{U}$ such that $(H|D) \cap [(W^{-1}(a) \cap D) \times (W(a) \cap D)] = \emptyset$. Hence $H \cap [\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}(W^{-1}(a) \cap D) \times \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(W(a) \cap D)] = \emptyset$, because H is $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open. However, since D is doubly dense in $(X, \mathcal{U})$, we have $(a, a) \in H \cap [\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}(W^{-1}(a) \cap D) \times \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(W(a) \cap D)]$ — a contradiction. We conclude that D is supdense in $(X, \mathcal{U})$. Let us note that our argument shows that it is sufficient to assume that $(D, \mathcal{U}|D)$ is semi-symmetric (= closed symmetric) (see e.g. [15]) provided that $\mathcal{T}(\mathcal{U})$ is regular.

Some readers may prefer the following less computational proof of Lemma 2 suggested by the referee (using the terminology of [3,4]): Assume that $\mathcal{U}|D$ is symmetric. Any trace filter pair $(\mathbf{f}, \mathbf{g})$ is round and Cauchy. It follows from the symmetry that $\mathbf{f} \cap \mathbf{g}$ is a Cauchy filter. Hence $\mathbf{f} = \mathbf{g}$, since round Cauchy filters in a uniform space are minimal Cauchy. Thus D is supdense. The statement on Smyth symmetry follows by applying this observation to the totally bounded reflection.

PROPOSITION 7. *Let $(X, \mathcal{U})$ be a quasi-uniform space and let D be a doubly dense subspace of $(X, \mathcal{U})$. Then $\mathcal{U}|D$ is a uniformity if and only if $\mathcal{U}$ is a uniformity.*

PROOF. Assume that $\mathcal{U}|D$ is a uniformity. Note first that by Lemma 2 D is supdense in $(X, \mathcal{U})$. Let $\mathcal{B}$ be a base for $\mathcal{U}|D$ consisting of symmetric entourages. The second argument given in the proof of Proposition 6 shows that $\mathcal{H} = \{\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*) \times \mathcal{T}(\mathcal{U}^*)} B : B \in \mathcal{B}\}$ is a base for $\mathcal{U}$. Since the members of $\mathcal{H}$ are symmetric, we have shown that $\mathcal{U}$ is a uniformity. The converse is obvious. (Of course, we could also use [4, §11].)

A quasi-uniformity $\mathcal{U}$ on a set X is called *open symmetric* [15] provided that whenever A and B are $\mathcal{T}(\mathcal{U})$-open sets, $A \delta_{\mathcal{U}} B$ if and only if $B \delta_{\mathcal{U}} A$. Here $\delta_{\mathcal{U}}$ denotes the quasi-proximity induced by $\mathcal{U}$ on X. It is known (and easy to see) [15, Proposition 4.2] that a quasi-uniformity $\mathcal{U}$ on a set X is open symmetric if and only if, whenever A is a $\mathcal{T}(\mathcal{U})$-open set, B is a $\mathcal{T}(\mathcal{U})$-closed set and $U^{-1}(A) \subseteq B$ for some $U \in \mathcal{U}$, there is a $V \in \mathcal{U}$ such that $V(A) \subseteq B$.

PROPOSITION 8. (a) *Let D be a supdense subspace of a quasi-uniform space $(X, \mathcal{U})$. If $\mathcal{U}|D$ is open symmetric, then $\mathcal{U}$ is open symmetric.*

(b) *Let D be a $\mathcal{T}(\mathcal{U})$-dense subspace of an open symmetric quasi-uniform space $(X, \mathcal{U})$. Then $(D, \mathcal{U}|D)$ is open symmetric.*

PROOF. (a) Suppose that $\mathcal{U}|D$ is open symmetric. Let A be a $\mathcal{T}(\mathcal{U})$-open subset of X and let B be a $\mathcal{T}(\mathcal{U})$-closed subset of X such that $U^{-1}(A) \subseteq B$ for some $U \in \mathcal{U}$. Hence $(U|D)^{-1}(D \cap A) \subseteq D \cap B$. Since $\mathcal{U}|D$ is open symmetric,

there is a $V \in \mathcal{U}$ such that $(V|D)(D \cap A) \subseteq D \cap B$. Without loss of generality we assume that V is $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open. Since $V \cap [(D \cap A) \times (D \setminus B)] = \emptyset$, we have that $V \cap [\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}(D \cap A) \times \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(D \setminus B)] = \emptyset$. Furthermore, $A \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)} A = \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}(D \cap A) \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}(D \cap A)$ and $X \setminus B \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(X \setminus B) = \mathrm{cl}_{\mathcal{T}(\mathcal{U})}[(X \setminus B) \cap D] = \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(D \setminus B)$. Consequently, $V(A) \subseteq B$. Hence $\mathcal{U}$ is open symmetric.

(b) Let A be open in $(D, \mathcal{T}(\mathcal{U}|D))$, let B be closed in $(D, \mathcal{T}(\mathcal{U}|D))$ and let U be a $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open entourage of $\mathcal{U}$ such that $(U|D)^{-1}(A) \subseteq B$. Choose a $\mathcal{T}(\mathcal{U})$-open set G of X such that $G \cap D = A$. If $G \cap U(D \setminus B) \neq \emptyset$, then $G \cap D \cap U(D \setminus B) \neq 0$, because D is $\mathcal{T}(\mathcal{U})$-dense in X, and we see that $(U|D)^{-1}(A) \not\subseteq B$ — a contradiction. Thus $G \cap U(D \setminus B) = \emptyset$. Choose a $\mathcal{T}(\mathcal{U}^{-1} \times \mathcal{U})$-open entourage $W \in \mathcal{U}$ such that $W^2 \subseteq U$. Then $W^{-1}(G) \subseteq X \setminus W(D \setminus B)$. Since $(X, \mathcal{U})$ is open symmetric, there exists $V \in \mathcal{U}$ such that $V(G) \subseteq X \setminus W(D \setminus B)$. Thus $(V|D)(A) \subseteq B$. We have shown that $(D, \mathcal{U}|D)$ is open symmetric.

PROPOSITION 9. *Let $(X, \mathcal{U})$ be a quasi-uniform space and let D be a doubly dense subspace of $(X, \mathcal{U})$. Then $\mathcal{U}|D$ is Smyth symmetric if and only if $\mathcal{U}$ is Smyth symmetric.*

PROOF. Obviously, Smyth symmetry is a hereditary property of quasi-uniform spaces (cf. [12, Proposition 1.30]). The assertion follows by applying Proposition 7 to the totally bounded reflection.

We finish this section by exhibiting two further symmetry properties for which the studied question has a negative answer.

EXAMPLE 6. A quasi-uniform space $(X, \mathcal{U})$ is said to be *small-set symmetric* [15] provided that for each $U \in \mathcal{U}$ and each $\mathcal{T}(\mathcal{U})$-open set A we have that $\mathrm{cl}_{\mathcal{T}(\mathcal{U})} A \subseteq U(A)$ [15, Lemma 3.1(c)]. It is known that a quasi-uniform space $(X, \mathcal{U})$ is small-set symmetric if and only if its conjugate $(X, \mathcal{U}^{-1})$ is point-symmetric [20, Lemma 4]. In particular, small-set symmetry is a hereditary property.

Arguing as in Example 1(a) we see that if $(X, \mathcal{P})$ is a non-discrete T_3-space equipped with its Pervin quasi-uniformity, then the space $(\widetilde{X}, \widetilde{\mathcal{P}}^{-1})$ is not small-set symmetric, although it contains the supdense small-set symmetric subspace $(X, \mathcal{P}^{-1})$.

A quasi-uniform space $(X, \mathcal{U})$ is called *equinormal* [12, p. 37] provided that for each $\mathcal{T}(\mathcal{U})$-closed subset F of X and each $\mathcal{T}(\mathcal{U})$-open set G of X containing F there is a $U \in \mathcal{U}$ such that $U(F) \subseteq G$. Note that uniform spaces need not be equinormal.

EXAMPLE 7. Let X be a normal topological T_2-space whose Hewitt realcompactification is not normal (e.g. take for X the Σ-product in $\mathbf{R}^{\omega_1}$ with base point 0; see [2]). Consider the completion $(\widetilde{X}, \widetilde{\mathcal{C}(X)})$ of the uniform space $(X, \mathcal{C}(X))$ where $\mathcal{C}(X)$ is the uniformity on X initial with respect to the

family of all continuous real-valued functions on X (see [11, Example 8.1.19 and Example 8.3.19]). Then $(\widetilde{X}, \mathcal{T}(\widetilde{\mathcal{C}(X)}))$ is the Hewitt realcompactification of X. Since $(\widetilde{X}, \mathcal{T}(\widetilde{\mathcal{C}(X)}))$ is not normal, the uniformity $\widetilde{\mathcal{C}(X)}$ cannot be equinormal. However, since X is normal, it is obvious by Urysohn's Lemma that the $\mathcal{T}(\widetilde{\mathcal{C}(X)})$-dense subspace $(X, \mathcal{C}(X))$ of $(\widetilde{X}, \widetilde{\mathcal{C}(X)})$ is equinormal.

On the other hand, let X be a non-normal Tychonoff space and let $\mathcal{C}^*(X)$ be the uniformity initial with respect to the family of all continuous bounded real-valued functions on X [11, Example 8.1.19]. Since $\mathcal{T}(\widetilde{\mathcal{C}^*(X)})$ is compact [11, Example 8.3.18], the completion $(\widetilde{X}, \widetilde{\mathcal{C}^*(X)})$ of $(X, \mathcal{C}^*(X))$ is equinormal, although $\mathcal{C}^*(X)$ is not equinormal, because X is not normal.

It seems worthwhile, however, to point out that for the property of equinormality our problem has a positive solution in an important special case.

PROPOSITION 10. *A normal quasi-uniform space $(X, \mathcal{U})$ with a supdense equinormal subspace D is equinormal.*

PROOF. Let F_1 and F_2 be disjoint $\mathcal{T}(\mathcal{U})$-closed subsets of X. Because $(X, \mathcal{U})$ is normal, there are $\mathcal{T}(\mathcal{U})$-open sets G_1 and G_2 of X such that $F_1 \subseteq G_1$, $F_2 \subseteq G_2$ and $\mathrm{cl}_{\mathcal{T}(\mathcal{U})} G_1 \cap \mathrm{cl}_{\mathcal{T}(\mathcal{U})} G_2 = \emptyset$. Since $(D, \mathcal{U}|D)$ is equinormal, there exists $H \in \mathcal{U}$ such that $H^3((\mathrm{cl}_{\mathcal{T}(\mathcal{U})} G_1) \cap D) \cap ((\mathrm{cl}_{\mathcal{T}(\mathcal{U})} G_2) \cap D) = \emptyset$. Thus $H(F_1) \cap F_2 \subseteq H(\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)} G_1) \cap \mathrm{cl}_{\mathcal{T}(\mathcal{U})} G_2 \subseteq H(\mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}(G_1 \cap D)) \cap \mathrm{cl}_{\mathcal{T}(\mathcal{U})}(G_2 \cap D) \subseteq H^2(G_1 \cap D) \cap H^{-1}(G_2 \cap D) = \emptyset$. We have shown that $(X, \mathcal{U})$ is equinormal.

5. Miscellaneous results

We conclude this paper by considering the properties of transitivity, quietness and stability. Let us recall that a quasi-uniform space $(X, \mathcal{U})$ is called *transitive* [12, p. 27] if it has a base consisting of transitive entourages.

PROPOSITION 11. *Let D be a supdense subspace of a quasi-uniform space $(X, \mathcal{U})$. Then $(D, \mathcal{U}|D)$ is transitive if and only if $(X, \mathcal{U})$ is transitive.*

PROOF. The nontrivial part of the assertion is a consequence of [4, §11] and [5, Theorem 3.14].

A quasi-uniform space $(X, \mathcal{U})$ is said to be *quiet* [8] provided that for each $U \in \mathcal{U}$ there is a $V \in \mathcal{U}$ such that whenever $(\mathcal{F}, \mathcal{G}) \to 0$ and x and y are points of X such that $V(x) \in \mathcal{G}$ and $V^{-1}(y) \in \mathcal{F}$, then $(x, y) \in U$. Note that quietness is a hereditary property of quasi-uniform spaces.

PROPOSITION 12. *Let D be a supdense subspace of a quasi-uniform space $(X, \mathcal{U})$. Then $(D, \mathcal{U}|D)$ is quiet if and only if $(X, \mathcal{U})$ is quiet.*

PROOF. The statement is a consequence of [6, 2.2] and [4, 11.2].

Let us mention in connection with Proposition 12 that Lemma 1 (see Section 1) can be strengthened in the case of doubly uniformly regular quasi-uniformities. (A quasi-uniformity $\mathcal{U}$ is called *doubly uniformly regular* [6] if both $\mathcal{U}$ and $\mathcal{U}^{-1}$ are uniformly regular. It is well known that quiet quasi-uniformities are doubly uniformly regular.)

LEMMA 3. *Let $\mathcal{U}$ and $\mathcal{V}$ be doubly uniformly regular quasi-uniformities on a set X such that $\mathcal{T}(\mathcal{U}) = \mathcal{T}(\mathcal{V})$ and $\mathcal{T}(\mathcal{U}^{-1}) = \mathcal{T}(\mathcal{V}^{-1})$. If $\mathcal{U}|D = \mathcal{V}|D$ where D is doubly dense in $(X,\mathcal{U})$, then $\mathcal{U} = \mathcal{V}$.*

PROOF. See [6, Lemma 2.5].

Using the results of [6] the reader will readily verify that if D is doubly dense in a doubly uniformly regular quasi-uniform space $(X,\mathcal{U})$, then $(X,\mathcal{U})$ is quiet whenever the subspace $(D,\mathcal{U}|D)$ is quiet, and the quasi-uniformities $\mathcal{U}$ and $\mathcal{U}|D$ have the same weight. In fact, whenever $\mathcal{B}$ is a base for $\mathcal{U}|D$, then the sets $\bigcup\{\mathrm{cl}_{\mathcal{T}(\mathcal{U}^{-1})}B_1 \times \mathrm{cl}_{\mathcal{T}(\mathcal{U})}B_2 : B_1 \times B_2 \subseteq B\}$ where $B \in \mathcal{B}$ form a base for $\mathcal{U}$.

A filter $\mathcal{F}$ on a quasi-uniform space $(X,\mathcal{U})$ is called *stable* if for any $U \in \mathcal{U}$ there is an $F \in \mathcal{F}$ such that $F \subseteq U(F')$ whenever $F' \in \mathcal{F}$. The space $(X,\mathcal{U})$ is called *stable* [10] if every D-Cauchy filter on it is a stable filter.

PROPOSITION 13. *A quasi-uniform space $(X,\mathcal{U})$ is stable if and only if it has a supdense stable subspace.*

PROOF. One easily checks that each subspace of a stable quasi-uniform space is stable [21, Corollary 4(a)]. For the converse assume that $(D,\mathcal{U}|D)$ is a supdense stable subspace of $(X,\mathcal{U})$. Let $\mathcal{G}$ be a D-Cauchy filter on $(X,\mathcal{U})$ and let $\mathcal{G}_0$ be the filter on D generated by $\{G \cap D : G \in \mathcal{T}(\mathcal{U}) \cap \cap \mathcal{G}\}$. According to the proof of Proposition 2, $\mathcal{G}_0$ is a D-Cauchy filter on $(D,\mathcal{U}|D)$. Let $U \in \mathcal{U}$. Choose $W \in \mathcal{U}$ such that $W^3 \subseteq U$. Since $(D,\mathcal{U}|D)$ is stable, there is $G \in \mathcal{T}(\mathcal{U}) \cap \mathcal{G}$ such that $G \cap D \subseteq (W|D)(G')$ whenever $G' \in \in \mathcal{G}_0$. Let $E \in \mathcal{G}$. Then $W(E) \cap D \in \mathcal{G}_0$. Hence we have $G \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}G = = \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}(G \cap D) \subseteq \mathrm{cl}_{\mathcal{T}(\mathcal{U}^*)}W(W(E)) \subseteq W^3(E) \subseteq U(E)$. Consequently $\mathcal{G}$ is stable in $(X,\mathcal{U})$ and thus $(X,\mathcal{U})$ is stable.

The referee observes that it is not possible to generalize Proposition 13 like the results on convergence completeness (in the remark following the proof of Proposition 1), since in [7, Example 8.5] Deák constructs a doubly stable quasi-uniformity such that its D-completion (in the sense of [10]) is not stable, although it is doubly strict.

PROBLEM. Find a (categorical ?) characterization of those quasi-uniform properties $\mathcal{O}$ such that each quasi-uniform space containing a supdense subspace with property $\mathcal{O}$ has property $\mathcal{O}$.

We would like to thank the referee for his helpful comments.

REFERENCES

[1] CSÁSZÁR, Á., D-complete extensions of quasi-uniform spaces, *Acta Math. Hungar.* **64** (1994), 41–54.

[2] CORSON, H. H., Normality in subsets of product spaces, *Amer. J. Math.* **81** (1959), 785–796. *MR* **21**#5947

[3] DEÁK, J., Extensions of quasi-uniformities for prescribed bitopologies. I, *Studia Sci. Math. Hungar.* **25** (1990), 45–67. *MR* **92b**:54058

[4] DEÁK, J., Extensions of quasi-uniformities for prescribed bitopologies. II, *Studia Sci. Math. Hungar.* **25** (1990), 69–91. *MR* **92b**:54058

[5] DEÁK, J., A survey of compatible extensions (presenting 77 unsolved problems), *Topology, Theory and Applications II* (Proc. Sixth Colloq., Pécs, 1989), Colloq. Math. Soc. János Bolyai, **55**, North-Holland, Amsterdam, 1993, 127–175.

[6] DEÁK, J., Extending and completing quiet quasi-uniformities, *Studia Sci. Math. Hungar.* **29** (1994), 349–362.

[7] DEÁK, J., A bitopological view of quasi-uniform completeness II, *Studia Sci. Math. Hungar.* (to appear).

[8] DOITCHINOV, D., On completeness of quasi-uniform spaces, *C.R. Acad. Bulgare Sci.* **41** (1988), No.7, 5–8. *MR* **89j**:54028

[9] DOITCHINOV, D., A concept of completeness of quasi-uniform spaces, *Topology Appl.* **38** (1991), 205–217. *MR* **92b**:54061

[10] DOITCHINOV, D., E-completions of quasi-uniform spaces (preprint).

[11] ENGELKING, R., *General topology, Second edition*, Sigma Series in Pure Mathematics, **6**, Heldermann, Berlin, 1989. *MR* **91c**:54001

[12] FLETCHER, P. and LINDGREN, W. F., *Quasi-uniform spaces*, Lecture Notes in Pure and Applied Mathematics, **77**, Marcel Dekker, Inc., New York, 1982. *MR* **84h**:54026

[13] FLETCHER, P. and LINDGREN, W. F., Compactifications of totally bounded quasi-uniform spaces, *Glasgow Math. J.* **28** (1986), 31–36. *MR* **87f**:54037

[14] FLETCHER, P. and HUNSAKER, W., Uniformly regular quasi-uniformities, *Topology Appl.* **37** (1990), 285–291. *MR* **92b**:54062

[15] FLETCHER, P. and HUNSAKER, W., Symmetry conditions in terms of open sets, *Topology Appl.* **45** (1992), 39–47.

[16] KOPPERMAN, R. D., Total boundedness and compactness for filter pairs, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **33** (1990), 25–30.

[17] KÜNZI, H.-P. A. and BRÜMMER, G. C. L., Sobrification and bicompletion of totally bounded quasi-uniform spaces, *Math. Proc. Cambridge Philos. Soc.* **101** (1987), 237–247. *MR* **88h**:54040

[18] KÜNZI, H.-P. A. and FERRARIO, N., Bicompleteness of the fine quasi-uniformity, *Math. Proc. Cambridge Philos. Soc.* **109** (1991), 167–186. *MR* **92e**:54026

[19] KÜNZI, H.-P. A., Functorial admissible quasi-uniformities on topological spaces, *Topology Appl.* **43** (1992), 39–47.

[20] KÜNZI, H.-P. A., MRŠEVIĆ, M., REILLY, I. L. and VAMANAMURTHY, M. K., Convergence, precompactness and symmetry in quasi-uniform spaces, *Math. Japonica* **38** (1993), 239–253.

[21] KÜNZI, H.-P. A. and JUNNILA, H. J. K., Stability in quasi-uniform spaces and the inverse problem, *Topology Appl.* **49** (1993), 175–189.

(Received August 14, 1991)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BERN
SIDLERSTRASSE 5
CH–3012 BERN
SWITZERLAND

e-mail: kunzi@ math-stat.unibe.ch

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 303–311

IN THE MAX-SEMIGROUP OF PROBABILITY DISTRIBUTIONS OVER THE PLANE THERE IS NO KHINCHINE-TYPE DECOMPOSITION THEOREM

A. ZEMPLÉNI[1]

Abstract

We consider the multiplicative semigroup of probability distribution functions on $\mathbf{R}^2$, which corresponds to the coordinatewise maximum of $\mathbf{R}^2$-valued independent random vectors. Irreducible and anti–irreducible distributions with given marginals are constructed. These turn out to be the random vectors with minimal and maximal correlations, respectively. For a class of distribution functions with independent components over separate rectangles the nonexistence of a decomposition into irreducibles and an anti–irreducible is proved.

1. Introduction

Let the multiplication of distribution functions (d.f.'s) F and G over $\mathbf{R}^2$ be defined as

$$F \cdot G(\underline{x}) = F(\underline{x}) \cdot G(\underline{x}) .$$

The d.f. $F \cdot G$ is the d.f. of the coordinatewise maximum of independent random vectors with d.f. F and G. We denote this semigroup by $D(\mathbf{R}^2, \vee)$, emphasizing that it can be considered as the convolution semigroup defined over the maximum-semigroup $(\mathbf{R}^2, \vee)$ of $\mathbf{R}^2$ (we use the notations $\vee$, $\wedge$ for the coordinatewise maximum and minimum of vectors in $\mathbf{R}^2$, respectively). All d.f.'s are assumed to be right-continuous. Furnishing $D(\mathbf{R}^2, \vee)$ with the weak topology we get a commutative semigroup which has been widely investigated. Balkema and Resnick [1] characterized the infinitely divisible distributions in this semigroup, de Haan [3] gave a representation for max-stable distributions via point processes. The most important field for its application is in the multivariate extreme-value theory.

Decompositions of d.f.'s were first investigated in case of the usual convolution semigroup of d.f.'s over the real line. Khinchine (see e.g. [4]) proved

1991 *Mathematics Subject Classification.* Primary 60E05; Secondary 60B99.

Key words and phrases. Anti-irreducible distributions, convolutions of probability measures, extremal correlation, irreducible distributions, probabilities over topological semigroups.

[1]Research partially supported by Hungarian Scientific Research Grant No. 1405.

the following theorem. (The precise definitions of the notions we use will be given in Section 2.)

THEOREM 1. *Any d.f. F can be decomposed into the form*

$$F = \prod_{i=1}^{\omega}{}^{*} G_i * H \tag{1}$$

where the d.f's G_i are irreducible, H is anti-irreducible, $\omega \in \mathbf{N} \cup \{\infty\}$ and the limit of an infinite convolution product is meant in the weak topology.

Meanwhile several generalizations of this theorem were proved. One interesting direction is to prove decomposition theorems for general commutative semigroups. The work of Davidson and Kendall [2] on Delphic semigroups and the recent monograph of Ruzsa and Székely [7] on Hun and Hungarian semigroups are the most important works in this direction. These theories imply the existence of decomposition theorems for several semigroups of probability measures. As an example we present the following theorem (Zempléni, [11]).

THEOREM 2. *Let S be a commutative, Hausdorff topological semigroup such that*

(i) *there exists a unit element u in S;*

(ii) *S is associate-free (there are no elements $s, t \in S$: $s \neq t : s \mid t, t \mid s$);*

(iii) *$T_s = \{t \in S : t \mid s\}$ (the set of divisors of s) is compact for every $s \in S$;*

(iv) *to every $s \in S$ and open set U with $U \supset T_s$ there exists an open neighbourhood V of s such that $y \in V, x \mid y$ implies $x \in U$.*

(Conditions (i) *to* (iii) *characterize the so-called Hun semigroups while* (iv) *is the stability property.) Then $D(S)$ is Hun again, thus by the results of Ruzsa and Székely* [7] *there is an analogous decomposition to* (1) *in $D(S)$.*

Theorem 2 is applicable to $(\overline{\mathbf{R}_+^2}, \vee)$, thus we have decomposition theorem in $D(\overline{\mathbf{R}_+^2}, \vee)$ (the maximum-semigroup of bivariate distribution functions, corresponding to nonnegative random vectors). Unfortunately neither (i) nor (iii) holds for $(\mathbf{R}^2, \vee)$.

In Section 3 we give a negative answer to the problem posed in Zempléni [9] whether despite the unapplicability of the general theory, Theorem 1 remains valid for $D(\mathbf{R}^2, \vee)$. The construction is based on results of Section 2. These results are interesting on their own, since here irreducible and anti-irreducible d.f.'s in $D(\mathbf{R}^2, \vee)$ and $D(\overline{\mathbf{R}_+^2}, \vee)$ with given margins are constructed.

2. Arithmetical properties and correlation

Now we give the precise definitions of the notions we use in the sequel. $(*, \vee)$ can denote any subsemigroup of $(\mathbf{R}^2, \vee)$.

DEFINITION 1. $F \in D(*, \vee)$ is called irreducible ($F \in \mathrm{Irr}\,(D(*, \vee))$), if $F = G \cdot H$ for $G, H \in D(*, \vee)$ implies $G = F$ or $H = F$.

DEFINITION 2. $F \in D(*, \vee)$ is called anti-irreducible ($F \in \mathrm{Air}\,(D(*, \vee))$), if F is not irreducible and $F = G \cdot H$ for $G, H \in D(*, \vee)$ with an H irreducible d.f. implies $G = F$.

(F is anti-irreducible if a decomposition with an irreducible term is not effective in the sense that F appears as remainder term, too.)

DEFINITION 3. $F \in D(*, \vee)$ is called infinitely divisible ($F \in I(D(*, \vee))$), if for every $n \in \mathbf{N}$ $\exists F_n \in D(*, \vee)$ such that $F = (F_n)^n$.

The notions in Definitions 1 and 2 coincide with the classes "effective irreducible", "effective anti-irreducible" in Ruzsa and Székely [7] but we do not use different classes, thus we omit the adjective.

Throughout the paper we identify $D(\mathbf{R}^2, \vee)$ with its homeomorphic image $D(\mathbf{R}_+^2, \vee)$ under

$$\phi : \mathbf{R}^2 \to \mathbf{R}_+^2$$

$$\phi(x, y) = (\exp(-x), \exp(-y))$$

$D(\mathbf{R}_+^2, \vee)$ corresponds to the d.f.'s of strictly positive random vectors. The following remarks are obvious.

REMARK 1. $F \in L(D(\mathbf{R}^2, \vee))$ (where L is any of the classes introduced in the definitions above) if and only if its image $\phi(F) \in L(D(\mathbf{R}_+^2, \vee))$.

REMARK 2. If $F \in D(\mathbf{R}_+^2, \vee)$ is in $\mathrm{Irr}\,(D(\overline{\mathbf{R}_+^2}, \vee))$ then $F \in \mathrm{Irr}\,(D(\mathbf{R}_+^2, \vee))$.

Thus we can concentrate on $D(\overline{\mathbf{R}_+^2}, \vee)$ and its subsemigroup $D(\mathbf{R}_+^2, \vee)$. The irreducible d.f.'s in these semigroups were investigated in Zempléni [9], the characterization of $\mathrm{Air}\,(D(\overline{\mathbf{R}_+^2}, \vee))$ can be found in Zempléni [10]. We cite this result since we need it to our construction of anti-irreducible distributions.

THEOREM 3. $F \in \mathrm{Air}\,(D(\overline{\mathbf{R}_+^2}, \vee))$ *if and only if*

$$\mathrm{supp}\,(F) = \{(x(t), y(t)) : t \in I\},$$

where $I = [0, 1]$ *or* $I = [0, 1)$, x *and* y *are nondecreasing functions such that if exactly one of its components is constant over the interval* $[0, \varepsilon)$ *then the constant is zero.*

COROLLARY 1. *If for* $F \in D(\mathbf{R}_+^2, \vee)$ *the conditions of Theorem 3 hold, then* $F \in \mathrm{Air}\,(D(\mathbf{R}_+^2, \vee))$, *too.*

PROOF. It is easily seen that $F \in I(D(\mathbf{R}_+^2, \vee))$ (since such an F is essentially a real d.f. and any $F \in D(\mathbf{R}, \vee)$ is infinitely divisible; $F \in D(\mathbf{R}_+^2, \vee)$

implies $F^{1/n} \in D(\mathbf{R}_+^2, \vee))$ thus $F \notin \operatorname{Irr}(D(\mathbf{R}_+^2, \vee))$. From this point on the proof of Theorem 3 in [10] can be repeated. □

Let us introduce the following notations: for a rectangle $T = [a, b] \times \times [c, d]$ in $\mathbf{R}_+^2$ let $S^{(1)} = [0, a] \times [c, d]$ and $S^{(2)} = [a, b] \times [0, c]$ denote two related rectangles to the left and below the original one, respectively. (By writing rectangles, we always mean parallel ones to the coordinate axes.)

To the proof of our main result we need the following lemma, which is a slightly modified version of Theorem 2.1 in Zempléni [9]. Its proof is identical to the original one (the difference in its formulation is caused by the fact that now noneffective decompositions of irreducible distributions are allowed, see Remark 1 in [9]).

LEMMA 1. *Suppose that to* $F \in D(\overline{\mathbf{R}_+^2}, \vee)$ *one can construct a sequence of rectangles* T_n *(*$n \in \mathbf{I}$*, where* $\mathbf{I}$ *denotes* $\mathbf{N}$ *or* $\mathbf{Z}$*) with the following properties:*

(i) T_n *and* T_{n+1} *have exactly two common vertices for all* n *where* n, $n+1 \in \mathbf{I}$*;*

(ii) $P_F(T_n) = 0$ *(*P_F *denotes the distribution corresponding to* F*) while* $P_F(S_n^{(i)}) > 0$ *for* $i = 1, 2$ *and for all* $n \in \mathbf{I}$*;*

(iii) $\operatorname{supp}(F) \subset \overline{\bigcup_{i \in \mathbf{I}} S_n^{(j)}}$ *for either* $j = 1$ *or* $j = 2$.

(See Fig. 1.) Then $F \in \operatorname{Irr}(D(\overline{\mathbf{R}_+^2}, \vee))$.

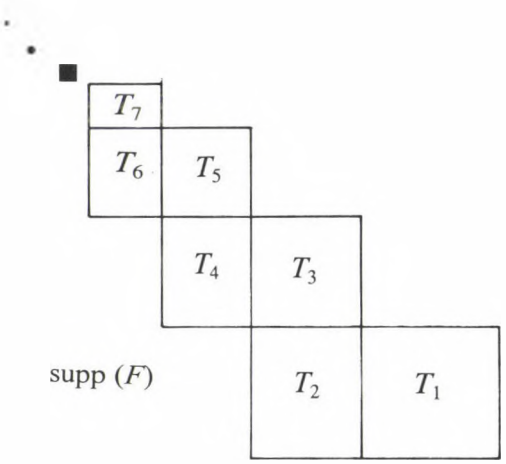


Fig. 1. An example for irreducible d.f. ($\mathbf{I} = \mathbf{N}$, $j = 2$)

One needs an infinite sequence of rectangles in Lemma 1 because of (iii). It is possible to prove an analogous lemma for finite number of rectangles,

but it would be not enough simply weaken (iii) by $\operatorname{supp}(F) \subset \overline{\bigcup_{j=1}^{2} \bigcup_{i \in \mathbf{I}} S_n^{(j)}}$. See the class of almost independent d.f.'s in Section 3 as a counterexample.

DEFINITION 4. We call the real d.f.'s F_1, F_2 left-coincident, if for $x_i = \min(\operatorname{supp} F_i)$ we have $F_1(x_1) = F_2(x_2)$ and cross-coincident if with $y_i = \sup(\operatorname{supp} F_i)$ $F_1(x_1) = 1 - F_2(y_2)$ and $F_2(x_2) = 1 - F_1(y_1)$.

Obviously continuous d.f.'s are always both left- and cross-coincident. Now we are in position to present the main result of this Section.

COROLLARY 2. *Let F_1 and F_2 be real d.f.'s with*

$$F_1(0) = F_2(0) = 0\,. \tag{2}$$

(a) *If F_1 and F_2 are left-coincident, then there exists an $F_{air} \in \operatorname{Air}(D(\overline{\mathbf{R}_+^2}, \vee)) \cap \operatorname{Air}(D(\mathbf{R}_+^2, \vee))$ with marginals F_1 and F_2;*

(b) *if F_1 and F_2 are cross-coincident, and there exist neighbourhoods S_{x_1} and S_{y_2} of x_1 and y_2 such that F_1 and F_2 are continuous in S_{x_1} and S_{y_2}, respectively (or an analogous condition holds for x_2 and y_1) then one can construct an $\exists F_{irr} \in \operatorname{Irr}(D(\overline{\mathbf{R}_+^2}, \vee)) \cap \operatorname{Irr}(D(\mathbf{R}_+^2, \vee))$ with marginals F_1 and F_2.*

Especially

$$F_{air}(x, y) = F_1(x) \wedge F_2(y) \tag{3}$$

and

$$F_{irr}(x, y) = |\, F_1(x) + F_2(y) - 1\,|_+ \tag{4}$$

is a suitable choice.

By the results of Hoeffding [5] we have that (if F_1 and F_2 have finite variances) (3) and (4) are the d.f.'s with maximal and minimal correlations among those with margins F_1 and F_2. The irreducible and anti-irreducible d.f.'s with given marginals are not unique in general.

PROOF. (a) By Corollary 1 it is enough to prove that for F_{air} the conditions of Theorem 3 hold. $F_{air} \in D(\mathbf{R}_+^2, \vee)$ immediately follows from (2).

$$\operatorname{supp}(F_{air}) = \overline{\{(F_1^{-1}(t), F_2^{-1}(t)) : t \in [0, 1]\}}$$

where $F^{-1}(y) = \inf\{x : F(x) > y\}$ (see Whitt [8]). F_{irr}^{-1} are obviously monotonic functions. $F_1^{-1}(t) = c$ for $t \in [0, \varepsilon]$ implies $F_1(x_1) = \varepsilon$ but then by the left-coincidence we have $F_2(x_2) = \varepsilon$, too ensuring that neither component of the reparametrized version can start as constant.

(b) In this case

$$\operatorname{supp}(F_{air}) = \overline{\{(F_1^{-1}(t), F_2^{-1}(1-t)) : t \in [0,1]\}}.$$

It is a graph of a monotonically decreasing function. By the cross-coincidence and our condition it is continuous in a neighbourhood of one of its ends. This implies that one can construct such sequence of rectangles above $\operatorname{supp}(F)$ as required in Lemma 1 (see Figure 1) which ensures that F_{air} is irreducible. □

3. A class of nondecomposable distribution functions in $D(\mathbf{R}_+^2, \vee)$

DEFINITION 5. Let $F \in D(\mathbf{R}_+^2, \vee)$ be called almost independent if

(i) there are rectangles $T_1 = [0,a] \times [b_1, b_2]$ and $T_2 = [a_1, a_2] \times [0,b]$ where $(0 < a < a_1 < a_2,\ 0 < b < b_1 < b_2)$ such that $\operatorname{supp}(F) = T_1 \cup T_2$. (Let us denote $[0,a] \times [0,b]$ by $T_1 \wedge T_2$);

(ii) there exist real d.f.'s $F_j^{(i)}$ $(i,j = 1,2)$ such that $F \mid_{T_i}(x,y) = p_i \cdot F_1^{(i)}(x) \cdot F_2^{(i)}(y)$ for $i = 1,2$ where $p_i = P_F(T_i)$ (thus $F_j^{(i)}$ is concentrated to T_i);

(iii) $\min\{\operatorname{supp}(F_1^{(1)})\} = \min\{\operatorname{supp}(F_2^{(2)})\} = 0$.

(I.e. F has independent coordinates separately both on T_1 and T_2; $p_1 + p_2 = 1$ by (i).)

THEOREM 4. *Let F be continuous and almost independent. Then*

(1) *Its decomposition in $D(\overline{\mathbf{R}_+^2}, \vee)$ has the form*

$$F = (p_1 \cdot F_2^{(1)} + p_2 \cdot F_1^{(2)}) \prod_{i=1}^{\infty} G_j, \tag{5}$$

where $\operatorname{supp}(G_j) \subset T_1 \wedge T_2$.

(2) *F has no decomposition of the form*

$$F = \prod_{i=1}^{\omega} G_j \cdot H,$$

where $G_j \in \operatorname{Irr}(D(\mathbf{R}_+^2, \vee))$, $H \in \operatorname{Air}(D(\mathbf{R}_+^2, \vee))$.

PROOF. We start with (1). First we show that the d.f. F is neither irreducible nor anti-irreducible even in $D(\mathbf{R}_+^2, \vee)$. Let G be a d.f. such that

$$\operatorname{supp}(G) = T_1 \wedge T_2 \tag{6}$$

with marginals $G^{(1)}(x) = \sqrt{F_1^{(1)}(x)}$ and $G^{(2)}(y) = \sqrt{F_2^{(2)}(y)}$. It is easy to check that $G \mid F$ and F/G is continuous, almost independent d.f. again: T_1, T_2 are the same as for F and the marginals are

$$\sqrt{F_1^{(1)}}\,,\ F_2^{(1)}\,,\ F_1^{(2)}\,,\ \sqrt{F_2^{(1)}}, \tag{7}$$

respectively. By the continuity of $F_1^{(1)}$ and $F_2^{(2)}$ Corollary 2 can be applied ensuring the existence of an irreducible $G \mid F$. We can continue the process of decomposition and finally as $n \to \infty$ we get the decomposition (5) since $(F_i^{(i)})^\alpha \to 1$ as $\alpha \to 0$. Finally we show that (5) is a Khinchine-type decomposition, which is done by showing that $H = (p_1 \cdot F_2^{(1)} + (1 - p_1) \cdot F_1^{(2)})$ is in $\operatorname{Irr}(D(\overline{\mathbf{R}_+^2}, \vee))$. By the straightforward equality

$$\operatorname{supp}(F \cdot G) = \overline{\operatorname{supp}(F) \vee \operatorname{supp}(G)}$$

(Lemma 3 in [9]) we have that in a decomposition $H = H_1 \cdot H_2$ where $H_1 \neq \delta_0$ (the degenerate d.f. in 0) $\operatorname{supp}(H_1) \cap \{(x, 0) : x > 0\} \neq \emptyset$. Thus $\operatorname{supp}(H_2) \cap \cap \{(0, y) : y > 0\} = \emptyset$ which implies $\operatorname{supp}(H_1) \supset \{(0, y) : b_1 < y < b_2\}$ and hence $H_2 = \delta_0$.

(2) Let

$$F = F' \cdot G \tag{8}$$

be an arbitrary decomposition of F in $D(\mathbf{R}_+^2, \vee)$. By Lemma 3 in [9] again we have that in decomposition (8) one of the components is concentrated to $T_1 \wedge T_2$, let this be G. It follows from condition (iii) of Definition 5 that G has to be continuous, thus F' is continuous as well. Let $(x, y) \in T_1$.

$$F(x, y) = p_1 \cdot F_1^{(1)}(x) \cdot F_2^{(1)}(y) = F'(x, y) \cdot G_1(x),$$

where G_1 is the first marginal of G. Thus

$$F'(x, y) = \left(\frac{F_1^{(1)}(x)}{G_1^{(1)}(x)} \right) \cdot F_2^{(1)}\,, \tag{9}$$

and an analogous decomposition to (9) holds over T_2 as well. By these decompositions of F' it is easy to check that it is almost independent;

$$F'^{(1)}_2 = F_2^{(1)}\,, \quad F'^{(2)}_1 = F_1^{(2)}\,. \tag{10}$$

By Part 1 of this Theorem we get that F has irreducible components, thus it is not anti-irreducible, meaning that the Khintchine-type decomposition procedure can be started. F' is almost independent, thus it is not irreducible

(see Part 1 again), so G has to be irreducible. But F' is not anti-irreducible either, so the procedure can be continued. As nothing was assumed about the properties of $G \in D(\mathbf{R}^2, \vee)$ in decomposition (8), the almost independence and (10) follows for a general decomposition $F = F' \cdot \prod_{i=1}^{\omega} G_i$. Thus there is no way of completing the sequence of decompositions, proving our assertion.

It is worth to mention that the only chance of finishing the procedure would be by achieving $F'^{(1)}_1 = F'^{(2)}_2 = \delta_0$ as it was done in (5) but $F = (p_1 \cdot F_2^{(1)} + p_2 \cdot F_1^{(2)})$ does not belong to $D(\mathbf{R}^2, \vee)$. □

The unusual lack of Khinchine-type decomposition and the fact that

$$\operatorname{Air}(D(\mathbf{R}^2_+, \vee)) \not\subset I(D(\mathbf{R}^2_+, \vee))$$

(see Zempléni [12]) shows that this semigroup has surprising properties despite its relative simplicity. This gives one more reason to consider its compactification $D(\overline{\mathbf{R}^2_+}, \vee)$ instead, as it is done in some related works, for example in Pancheva [6].

ACKNOWLEDGEMENT. The paper was written while the author held a fellowship of the Royal Society visiting the Department of Probability and Statistics at the University of Sheffield.

REFERENCES

[1] BALKEMA, A. A. and RESNICK, S. I., Max-infinite divisibility, *J. Appl. Probability* **14** (1977), 309–319. *MR* **55** #11338

[2] KENDALL, D. G., Delphic semigroups, infinite divisible regenerative phenomena and the arithmetic of p-functions, *Z. Wahrsch. Verw. Gebiete* **9** (1968), 163–195. *MR* **37** #5320

[3] HAAN, L. DE, A spectral representation for max-stable processes, *Ann. Probab.* **12** (1984), 1194–1204. *MR* **86b**:60089

[4] LINNIK, JU. V. and OSTROVSKII, I. V., *Decomposition of random variables and vectors*, Translations of Mathematical Monographs, Vol 48, American Mathematical Society, Providence, R. I., 1977. *MR* **55** #1404

[5] HOEFFDING, W., Maßstabinvariante Korrelationstheorie, *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* **5** (1940), 181–233. *MR* **3**–5

[6] PANCHEVA, E., General limit theorems for the maximum of independent random variables, *Teor. Veroyatnost. i Primenen.* **31** (1986), 730–744 (in Russian). (English translation: *Theory Probab. Appl.* **31**, 645–657.) *MR* **88f**:60044

[7] RUZSA, I. and SZÉKELY, G. J., *Algebraic probability theory*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, Wiley, New York, 1988. *MR* **89j**:60006

[8] WHITT, W., Bivariate distributions with given marginals, *Ann. Statist.* **4** (1976), 1280–1289. *MR* **54** #14045

[9] ZEMPLÉNI, A., On the arithmetical properties of the multiplicative structure of probability distribution functions, *Probability theory and mathematical statistics with applications* (Visegrád, 1985), Reidel, Dordrecht–Boston, MA, 1988, 221–233. *MR* **89m**:60034

[10] Zempléni, A., The description of the class I_0 in the multiplicative structure of distribution functions, *Mathematical statistics and probability theory*, Vol. A (Bad Tatzmannsdorf, 1986), Reidel, Dordrecht–Boston, MA–London, 1987, 291–303. *MR* **89b**:60046

[11] Zempléni, A., On the heredity of Hun and Hungarian property, *J. Theoret. Probab.* **3** (1990), 599–609. *MR* **91j**:60016

[12] Zempléni, A., Counterexamples in algebraic probability theory, *Proc. Probability Measures on Groups*, X, (ed. H. Heyer), Plenum Press, New York, 1991, 459–465.

(Received September 5, 1991)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
VALÓSZÍNŰSÉGELMÉLETI TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

e-mail: zempleni @ ludens.elte.hu

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 313–316

EINE EXTREMALEIGENSCHAFT DES REGULÄREN MOSAIKS $\{p,4\}$ IN DER HYPERBOLISCHEN EBENE UND IHRE VERALLGEMEINERUNG FÜR DIE MOSAIKE $\{p,2k\}$

Á. ILLÉS und I. VERMES

Herrn Professor László Fejes Tóth zum 80. Geburtstag gewidmet

Es wird in dieser Arbeit die Parkettierungsmöglichkeit der hyperbolischen Ebene durch kongruente rechtwinklige p-Ecke ($p \geqq 5$) gezeigt (Lemma 1) und eine Extremalaufgabe für diese Vielecke gelöst (Lemma 2). Daraus folgt unser Ergebnis über die Extremaleigenschaft der Mosaikklasse $\{p,4\}$.

Betrachte man ein p-Eck, dessen benachbartete Seiten sich rechtwinklig schneiden, und sei es kurz Recht-p-Eck genannt. Zum Beispiel ist es leicht zu sehen, daß ein Recht-p-Eck für $p=5$ durch die Seitenlängen zweier nichtbenachbarteten Seiten von ihm eindeutig bestimmt werden kann. Für $p>5$ können immer Recht-p-Ecke durch geeigneten Seitenstrecken bestimmt werden. Es hat noch keine reguläre Eigenschaften, aber ist es streng konvex, mit dem konstanten Flächeninhalt $\kappa^2\pi\frac{p-4}{2}$. Betrachte man einen willkürlichen inneren Punkt P eines Recht-p-Eckes $A_1 \dots A_p$. Wir fällen von P aus das Lot PT_i zur Seite A_iA_{i+1} ($i=1,2,\dots,p$ und $A_{p+1} := A_1$). Die Fußpunkte T_i sind die innere Punkte von A_iA_{i+1}, und der Kreis um P mit dem Radius $r := \min\{PT_i\}$ ist ein Inkreis des Recht-p-Eckes, folglich hat es zu jedem inneren Punkt einen Inkreis mit geeigneten Radien.

LEMMA 1. *Falls ein willkürliches Recht-p-Eck* ($p \geqq 5$) *in der hyperbolischen Ebene gegeben ist, so gibt es eine Parkettierung der Ebene, deren Elemente die kongruenten Exemplare dieses Recht-p-Eckes sind.*

BEWEIS. Betrachte man das Recht-p-Eck $A_1 \dots A_p$ ($p \geqq 5$), und seine Ecke A_i. Spiegelt man es an die Seite $A_{i-1}A_i$ bzw. A_iA_{i+1}, so erhält man die Spiegelbilder $A_1' \dots A_p'$ bzw. $A_1'' \dots A_p''$. Eine weitere Spiegelung an die Seite $A_i'A_{i+1}'$ bzw. $A_i''A_{i-1}''$ dasselbe Exemplar des Recht-p-Eckes ergibt. Damit ist die Ecke A_i lückenlos umgelegt. Auf derselben Weise können die anderen Ecken durch die Spiegelbilder an die geeigneten Seiten lückenlos umgelegt werden. Diese gespiegelten Exemplare bilden einen Gürtel um das vorgegebene Recht-p-Eck. Auf Grund des Alexandrow–Poincaréschen Satzes (vgl. Z. Lučić und E. Molnár [2] und [3]) kann die ganze Ebene durch die kongruenten Exemplare des betrachteten Recht-p-Eckes parkettiert werden. □

1991 *Mathematics Subject Classification.* Primary 52C20; Secondary 52A40.
Key words and phrases. Tilings in 2 dimensions, inequalities and extremum problems.

Unterstützt von der Ungarischen Akademie der Wissenschaften im Projekt OTKA Nr. 1615 (1991).

Die Extremalaufgabe bezüglich des Recht-p-Eckes ist das folgende

LEMMA 2. *Es sei ein Recht-p-Eck ($p \geqq 5$) gegeben, das einen Inkreis $K(P,r)$ (mit dem Mittelpunkt P und mit dem Radius r) hat. Das Recht-p-Eck kann so umgeformt werden, daß es auch ein Recht-p-Eck — also mit konstantem Flächeninhalt — wird, dessen Seite einen Kreis $K(P,\varrho)$ berühren. Das umgeformte Recht-p-Eck ist regulär, wobei $\varrho \geqq r$ ist und folglich soll ϱ der maximale Halbmesser von Inkreisen der Recht-p-Ecke sein.*

BEWEIS. Betrachte man ein Recht-p-Eck $A_1 \dots A_p$ und die gefällten Lote PT_i, wobei PT_i auf A_iA_{i+1} senkrecht stehen, und T_i die inneren Punkte von A_iA_{i+1} sind. Diese Lote zerlegen das Recht-p-Eck in die Lambertschen Vierecke $A_iT_iPT_{i-1}$ (mit $T_{1-1} := T_p$ und $A_{p+1} := A_1$), deren spitze Winkel $\sphericalangle T_iPT_{i-1} := \beta_i$ sind. Bezeichnen $r_1, r_2, \dots, r_p$ die Strecken $PT_1, PT_2, \dots, PT_p$ (Abb. 1). Das Viereck $A_iT_iPT_{i-1}$ vom Winkel β_i kann durch ein Lambertsches Viereck $A_i'T_i'PT_{i-1}'$ mit demselben Spitzwinkel β_i — also mit konstantem Flächeninhalt — so ersetzt werden, daß es an die Winkelhalbierende PA_i' symmetrisch wird ($PT_{i-1}' = PT_i' = r_i'$). Dazu soll man beweisen, daß $r_i' \geqq \min(r_i, r_{i-1})$ auf Grund $r_i \geqq r$ immer besteht. Wir können $r_{i-1} > r_i$ voraussetzen. In diesem Fall schneidet die Winkelhalbierende von β_i die Strecke A_iT_i in einem Punkt D, denn das Spiegelbild T von T_i an die Winkelhalbierende wird ein innerer Punkt von $PT_{i-1} = r_{i-1}$. Ebenso folgt, daß $\sphericalangle T_iDT := 2\delta > \frac{\pi}{2} = \sphericalangle T_{i-1}A_iT_i$ ist, daß das Dreieck $PA_i'T_i'$ das Dreieck PDT_i enthält, und daß die Ungleichungen $r_{i-1} > r_i' > r_i$ gelten. Die Gleichheit in $r_i' \geqq \min(r_i, r_{i-1})$ kann nur im Falle bestehen, falls das Viereck $A_iT_iPT_{i-1}$ schon symmetrisch ist. Für alle vorkommende Vierecke können solche Umformungen konstruiert werden, und die Ungleichungen $r_i' \geqq \min(r_i, r_{i-1})$ für $i = 1, \dots, p$ sich ergeben. Für die Dreiecke $PA_i'T_i'$ kann man aufschreiben:

$$\operatorname{ch}\frac{r_i'}{\kappa}\sin\frac{\beta_i}{2} = \cos\frac{\pi}{4}. \tag{1}$$

Daraus folgt, daß die Funktion $\beta_i \longmapsto r_i'(\beta_i)$ in β_i streng abnehmend ist. Dem Makarowschen Lemma gemäß (vgl. [1] §7. 16 und [4]) existiert genau ein ϱ, wobei zu jedem β_i ein β_i' entspricht so, daß $\sum_{i=1}^{p} \beta_i' = 2\pi$ besteht, und $\varrho \geqq r := \min\{r_i\}$ ist. Falls $\varrho < r$ wäre, so bestände $\beta_i < \beta_i'$ wegen (1) für alle $i \in \{1, \dots, p\}$, was aber den Voraussetzungen $\sum_{i=1}^{p} \beta_i = 2\pi = \sum_{i=1}^{p} \beta_i'$ widerspricht. Damit ist es gezeigt, daß ein Recht-p-Eck existiert, dessen Seite einen Kreis $K(P,\varrho)$ berühren. Daraus folgt unmittelbar, daß es ein reguläres Recht-p-Eck ist. Weiter ist ϱ der maximale Halbmesser der Inkreisen für alle Recht-p-Ecke. Wenn aber ein Inkreis vom Radius $\varrho' > \varrho$ wäre, so wäre ein anderes reguläres Recht-p-Eck von einem größeren Inkreis, was aber wegen des konstanten Flächeninhalt unmöglich ist. □

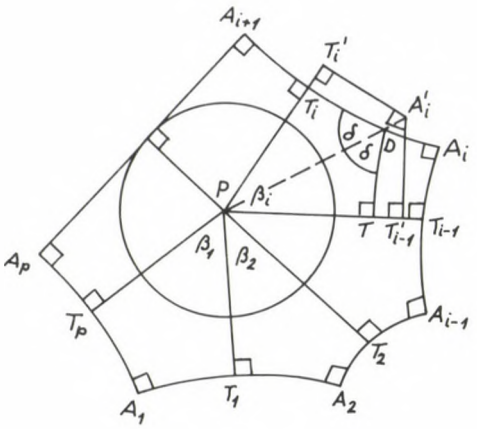


Abb. 1

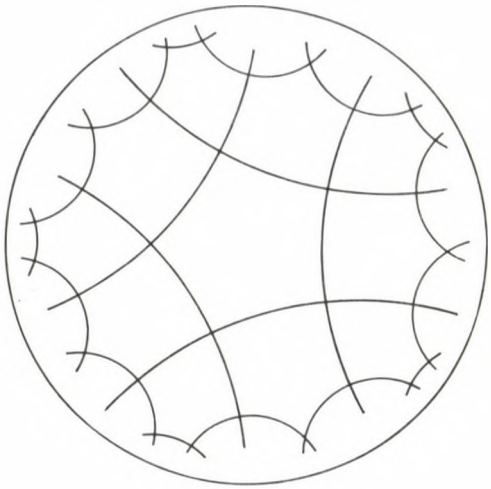


Abb. 2

ANMERKUNG. Das Lemma 2 gilt auch in dem Falle, wenn einige Winkel von $\alpha_1, \ldots, \alpha_p$ bei den Eckpunkten $A_1, \ldots, A_p$ vorgegeben sind, wobei das p-Eck konvex ist und $\sum_{i=1}^{p} \alpha_i < (p-2)\pi$ besteht. Natürlich soll man hier ersetzlich voraussetzen, daß die Fußpunkte T_i zu A_iA_{i+1} gehören. Daraus folgt die Existenz eines Tangentvieleckes um $K(P, \varrho)$ mit gleichen Winkeln außer den vorgegebenen. Der Beweis geht ganz ähnlich, wie vorher.

Die Extremaleigenschaft der Mosaikklasse $\{p, 4\}$: Die Parkettierung der hyperbolischen Ebene durch reguläre (und folglich kongruente) Recht-p-Ecke ist dadurch charakterisiert, daß sie unter den Parkettierungen durch Recht-p-Ecke das größte Inkreisradius hat. Diese Behauptung ist ein unmittelbarer Korollar unser beiden Lemma. Die Abbildung 2 zeigt einen Teil des Mosaiks $\{5, 4\}$ im Poincaréschen Kreismodell.

VERALLGEMEINERUNG. Die Parkettierungsmöglichkeit der hyperbolischen Ebene durch kongruente gleichwinklige p-Ecke des Winkels $\frac{\pi}{k}$ $(k \geqq 2)$ ist auf ähnlicher Weise beweisbar, wie das Lemma 1. Diese Vielecke sind streng konvex und das Lemma 2 gilt auch für solche Vielecke. Für die gleichwinklige Vielecke gilt die folgende Ungleichung

$$(p-2)\pi > p\frac{\pi}{k} \qquad \text{bzw.}$$

$$1 - \frac{2}{p} > \frac{1}{k}.$$

Daraus folgt zum Beispiel, daß $k = 2$ und $p \geqq 5$; $k = 3$ und $p \geqq 4$ bzw. $k \geqq 4$ und $p \geqq 3$ bestehen. Die gleichwinkligen Dreiecke sind reguläre Dreiecke, aber die gleichwinklige vier-, fünf-, $\ldots$, p-Ecke sind im allgemeinen

nicht-reguläre Vielecke. Das größte Inkreisradius charakterisiert das reguläre Mosaik $\{p, 2k\}$ für $k \geqq 3$ und $p \geqq 4$.

Ganz besonderen Dank sind wir Herrn Prof. E. Molnár für seine wertvollen Bemerkungen beim Lesen des Manuskriptes dieser Arbeit verpflichtet.

LITERATURVERZEICHNIS

[1] BEARDON, A. F., *The geometry of discrete groups*, Graduate Texts in Mathematics, 91, Springer-Verlag, Berlin–New York, 1983. *MR* **85d**:22026

[2] LUČIĆ, Z. und MOLNÁR, E., Fundamental domains for planar discontinuous groups and uniform tilings, *Geom. Dedicata* **40** (1991), 125–143. *MR* **92m**:52046

[3] LUČIĆ, Z. und MOLNÁR, E., Combinatorial classification of fundamental domains of finite area for planar discontinuous isometry groups, *Arch. Math. (Basel)* **54** (1990), 511–520. *MR* **92a**:20057

[4] MAKAROV, W. S., Eine Klasse der zwei-dimensionalen Fedorow-Gruppen, *Izv. Akad. Nauk SSSR Ser. Mat.* **31** (1967), 531–542 (russisch). *MR* **35** #6763

(Eingegangen am 5. September 1991)

ILLYÉS GYULA GIMNÁZIUM
H–7200 DOMBÓVÁR
HUNGARY

BUDAPESTI MŰSZAKI EGYETEM
GÉPÉSZMÉRNÖKI KAR
GEOMETRIA TANSZÉK
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 317–327

A NON-INTERACTION MODEL OF COMPLEMENT-MEDIATED LYSIS DIRECTED AGAINST TWO POPULATIONS OF SENSITIZED ERYTHROCYTES

T. BAKÁCS, K. BOGNÁR and G. TUSNÁDY

Introduction

The complement system, a set of proteins, defends the host from potentially dangerous microorganisms or other antigens by eliminating them from the blood and tissues. This is done either by the complement components alone or in collaboration with antibodies and/or cells [7]. The complement system is activated either by antibodies which acquire complement fixing properties as a consequence of their interaction with antigens (the classical activation pathway, [11]), or in the absence of antibodies by the surface of certain microorganisms (the alternative activation pathway, [4]). Although its exact mechanism is not completely understood, the function of the complement is exerted through a large protein complex termed the membrane attack complex (MAC) which has the ability to form a hole in surface membranes, thus lyse microorganisms and cells [3], [5]. In contrast to antibodies whose binding to antigens is strictly specific, complement can bind to various antibodies [9]. The binding of C1q, the first complement component, to antibody sensitized target cells activates a cascade process which is independent from the sensitizing antibody [7]. Since lysis of a cell by complement does not necessarily depend on activation of the system on the surface of that same cell [6] interactions between competing targets by using a common pool of activated molecules can be envisaged. We have earlier observed that in some combinations of antibodies interaction-like phenomena could indeed occur [2]. The purpose of the present paper is, therefore, to investigate the classical activation pathway of complement under condition when lysis is directed against two different populations of sensitized targets. Our central question was to confirm or reject the presence of genuine interactions between the lytic processes of the competing populations of erythrocytes by using a mathematical model.

1991 *Mathematics Subject Classifications.* Primary 92A07; Secondary 65U05.

Key words and phrases. Mathematical immunology, antibodies and complement, competitive systems, differential equations.

This work was partly supported by the Hungarian National Foundation for Scientific Research Grant No. 1818.

Human blood group A and B erythrocytes sensitized by anti-A and anti-B IgM monoclonal antibodies (mabs) were mixed (either the A or the B cells were labelled in the mixture by ^{51}chromium) and incubated with human serum (the source of complement) in the so called cold target competition assay where labelled and non-labelled erythrocytes (target and competitor respectively) competed for the complement [2]. The relative lysis was determined from the released isotope with a measurement error of size 5%. Three concentrations of target and competitor erythrocytes were used (24, 48 and $72*10^6$/ml). Since the fluid-phase concentrations of anti-A and anti-B mabs (BRIC.131 and BRIC.30, obtained from P. Judson, South Western Regional Transfusion Centre, Bristol, U.K.) were found to be 195 and 110 μg/ml respectively, similar sensitization conditions were provided by using the anti-A mab in 1/256, 1/512 and 1/1024, the anti-B in 1/128, 1/256 and 1/512 dilutions. (Under the experimental conditions used in the absence of sensitizing antibodies no lysis was detected, combinations without mabs therefore were not carried out.) Three dilutions of complement were employed (1/12, 1/20 and 1/28) and the assays were incubated for 0.33, 0.66 and 1.33 hours. Multiplying the number of possibilities we can see that 1215 individual measurements can be done, these were all carried out.

We have eight variables with the following measurement units:

T_A, T_B	target A and B in units of $1 * 10^6$/ml,
E	the effector (complement) in units of 100 times dilution,
A_A, A_B	anti-A and anti-B mabs, in units of 1000 times dilution on $50 * 10^6$ target/ml,
t	time in hours,
L_A, L_B	the relative lysis of target A and B (this quantity is a ratio, and does not need a unit).

The dynamics of competitive lysis, employing heterogeneous targets

At the beginning of our investigation of experimental results we were trying to decide if any form of interactions was present. After a while we realized that we have no usable general concept of non-interacting processes. In the particular model that we are using there is a natural way to define the lack of interaction, as it will be described below.

Let us suppose that the dynamics of the process for one target can be described by a pair of equations

$$E' = f(E,T),$$
$$T' = g(E,T),$$

with functions f_A, g_A for target A and f_B, g_B for target B, i.e. for target A

the equations

$$E' = f_A(E, T_A),$$
$$T'_A = g_A(E, T_A)$$

hold true, and for target B the equations

$$E' = f_B(E, T_B),$$
$$T'_B = g_B(E, T_B)$$

hold true (E' and T' denote derivatives with respect to time t). How to apply these dynamics for a competing situation?

Let us imagine that at a given moment t the total amount E of effector present in the process is cut into two parts: one is going to act on target A, denoted by E_A, and the other part E_B is going to act on target B. Then

$$E = E_A + E_B,$$

and the ratio $E_A : E_B$ may depend on the present value of all factors.

Our basic assumption is the following. The part E_A of effector attacking target A behaves as if all target in the process were type A. Since the dynamics is not linear, we have to blow up everything to the size E with the enlarging factor E/E_A. By doing so we figure out how much effector and target would be eliminated in the process in a short time period Δt. The hypothetical reaction with target $T^*_A = T_A * E/E_A$ and effector E has the infinitesimal quantities

$$\Delta E^*_A = f_A(E, T^*_A)\Delta t,$$
$$\Delta T^*_A = g_A(E, T^*_A)\Delta t,$$

and similarly

$$\Delta E^*_B = f_B(E, T^*_B)\Delta t,$$
$$\Delta T^*_B = g_B(E, T^*_B)\Delta t,$$

where $T^*_B = T_B * E/E_B$. These quantities would be measurable only in a virtual, artificially separated world. In the real process we have to recalibrate the infinitesimal quantities for real sizes:

$$\Delta E_A = (E_A/E)\Delta E^*_A, \quad \Delta T_A = (E_A/E)\Delta T^*_A,$$
$$\Delta E_B = (E_B/E)\Delta E^*_B, \quad \Delta T_B = (E_B/E)\Delta T^*_B.$$

This argument leads to the following system of equations

$$E'_A = (E_A/E) f_A(E, ET_A/E_A),$$
$$T'_A = (E_A/E) g_A(E, ET_A/E_A),$$

and a similar pair of equations holds true for target B:

$$E'_B = (E_B/E) f_B(E, ET_B/E_B),$$
$$T'_B = (E_B/E) g_B(E, ET_B/E_B)$$

resulting for E the combined equation

$$E' = (E_A/E) f_A(E, ET_A/E_A) + (E_B/E) f_B(E, ET_B/E_B).$$

The dynamics of non-competitive lysis, employing homogeneous target

The functions f_A, f_B, g_A, g_B will be approximated by a product of four terms:

- a constant,
- a profile function of effector,
- a profile function of target,
- a profile function of antibody.

The profile functions are normalized in such a way that for some specific value they are equal to 1. This value for target was $T_0 = 300$, for effector $E_0 = 5$, for antibody $A_0 = 500/128$ (i.e. for effector it corresponds to 1/20 dilution, for antibody 1/256 dilution on $50 * 10^6$/ml).

The individual profile functions are approximated by the function $Y(X, a)$ determined by the equations

$$\ln Y(s) = (e^{as} - e^{-as})/2,$$
$$\ln X(s) = (s/2) + (e^{2s} - 1)/4,$$

where a is the shape parameter of the family. This family of functions was chosen because of its flexibility: for $a = 0.1$ the function $Y(x)$ is similar to the square root of x, for $a = 1$ it is a sigmoid, for $a = 2$ it is J-shaped (see Fig. 1).

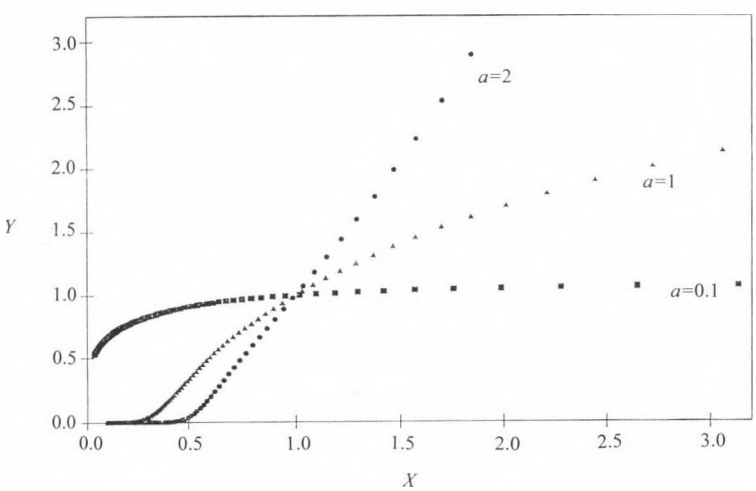


Fig. 1. Theoretical profile curves for the values 0.1, 1 and 2 of the shape parameter a

The parameters of the profile curves were fitted together with the parameters giving the ratio $E_A : E_B$. This ratio was approximated in the following way:

$$\ln(E_A/E_B) = k_0 + k_1 \ln(A_A/A_B) + k_2 \ln(T_A/T_B) + \\ + k_3 T_A + k_4 T_B + k_5 A_A + k_6 A_B + k_7 t + k_8 E$$

with estimated values for constants

$$\begin{aligned} k_0 &= -2.0202, & k_3 &= \ \ 0.1004, & k_6 &= \ \ 0.3391, \\ k_1 &= \ \ 5.0404, & k_4 &= -0.0117, & k_7 &= -6.0358, \\ k_2 &= \ \ 3.9625, & k_5 &= -0.3396, & k_8 &= \ \ 0.4851. \end{aligned}$$

Here the main terms are k_1 and k_2 resulting in the approximation

$$(E_A/E_B) \sim \text{const}(A_A/A_B)^5(T_A/T_B)^4.$$

The terms k_3 and k_4 give little effect, k_5 and k_6, however, seem to be important, it is remarkable how close they are to each other. The actual value of constant k_7 is seemingly large but the lysis is fast, thus the corresponding time value is small. The sign of k_7 means that lysis starts with T_A then the effector turns to T_B. The same effect is shown by k_8. The logarithms of the estimated shape parameters are given in Table 1, the logarithms of scaling constants in Table 2, where the logarithm of the main constant is also given. E.g., $h_A(E)$, the profile function part of $f_A(E,T)$, has the form

$$h_A(E) = H_A(E)/H_A(E_0),$$

where $E_0 = 5$ and

$$H_A(E) = Y(E/b, a),$$

where $Y(X,a)$ is the function defined above and

$$\ln a = 0.3437, \qquad \ln b = 1.6761.$$

The function $f_A(E,T_A)$ has the form

$$f_A(E,T_A) = cp_A(A_A)h_A(E)q_A(T_A),$$

Table 1
Logarithms of the estimated shape parameters

	Target A		Target B	
	Equation for effector	Equation for target	Equation for effector	Equation for target
Profile of target	−0.4246	0.3962	−0.3618	0.4074
Profile of effector	0.3437	0.1425	0.2990	0.1642
Profile of antibody	0.5234	0.4992	0.5651	0.5036

Table 2
Logarithms of scaling constants and the main constant

	Target A		Target B	
	Equation for effector	Equation for target	Equation for effector	Equation for target
Profile of target	3.4763	2.1074	3.7303	2.4202
Profile of effector	1.6761	1.5473	2.0123	1.9376
Profile of antibody	−5.9654	−3.6355	−1.4201	−1.9494
Main constant	5.2201	8.0014	4.9880	7.8880

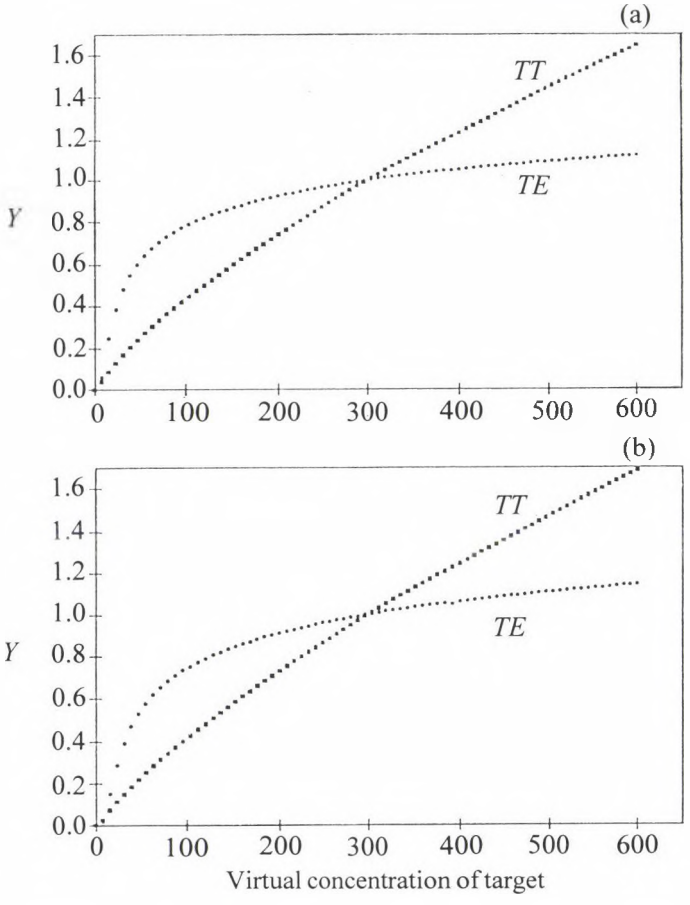


Fig. 2. Target profile curves (A erythrocytes Fig. 2a, B erythrocytes Fig. 2b)

where $\ln c = 5.2201$ and the form of functions p_A, q_A is similar to h_A. The profile functions are shown in Figure 2.a,b, 3.a,b, and 4.a,b, respectively.

The standard error of this model is 8.5% which is somewhat larger than the 5% measurement error but the individual errors exhibit no pronounced tendency. Perhaps with more flexible profile functions and regression form for the effector distribution between E_A and E_B the difference could be eliminated.

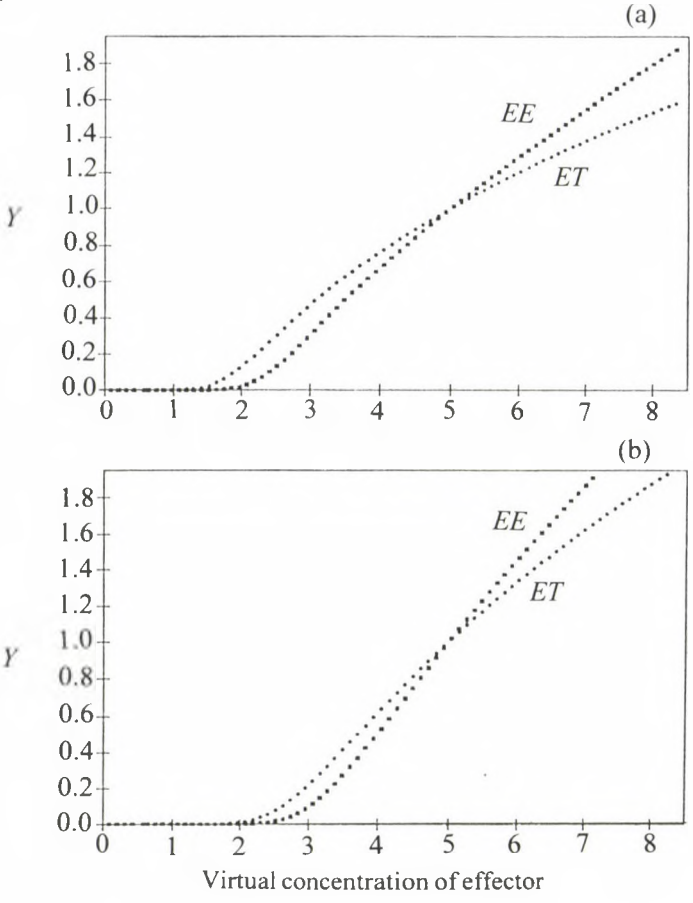


Fig. 3. Effector profile curves (A erythrocytes Fig. 3a, B erythrocytes Fig. 3b). Labels as in Fig. 2

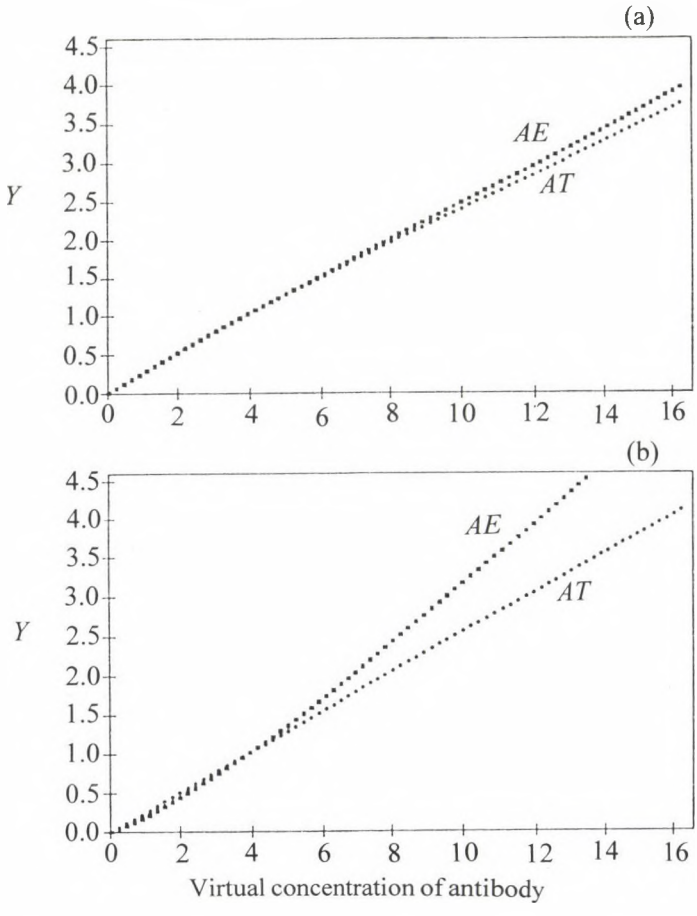


Fig. 4. Antibody profile curves (A erythrocytes Fig. 4a, B erythrocytes Fig 4b). Labels as in Fig. 2

As we can see, we fitted altogether 37 parameters to 1215 data points. The experimental data contained the input values T_A, T_B, E, A_A, A_B, t with outputs L_A, L_B. This latter pair was approximated with the pair K_A, K_B calculated from differential equations. Our system of differential equations is rather cumbersome (observe that the ratio $E_A : E_B$ depends on time, too). There is no hope for having explicit solutions for a system of equations like this. That is why the equations were integrated numerically by the Runge–Kutta method, then the sum of squares

$$(L_A - K_A)^2 + (L_B - K_B)^2$$

was minimized by a gradient method. This body of numerical work is close to be unmanageable with present day computers. All details were balanced between reliability and feasibility. One may ask of course whether our parameters or functions can be identified. On the one hand, perhaps under

certain conditions one may prove it. On the other hand, we are quite sure that the same experimental results are approximable with quite different models.

Discussion

We have shown in this paper that our non-interaction model describes the cold target competition experiment with a remarkable accuracy (the error of the model was only slightly larger than the error of the measurements). This result was interpreted to mean that under the conditions used, no stimulation and/or inhibition between the lysis of the sensitized A and B erythrocytes could be proven. Furthermore we have demonstrated that using the formula describing the effector distribution between the sensitized A ad B cells, the heterogeneous lytic process could be transformed into a sum of two homologous lytic processes. It is of importance to note, however, that in the presence of different mabs stimulation and/or inhibition were experimentally demonstrated (data not shown). Calculations using more complex models with interaction effects had indeed shown that it was possible to fit those models with similar standard error as in the present paper (manuscript in preparation).

The profile functions, describing the influence of a single component on the lytic process, were closely or practically linear (all but one shape parameters were larger than one). Since these are the corner stones of our model, it is of importance to note that similar relationships between the reaction components and lysis using sheep erythrocytes, rabbit antibodies and guinea pig complement has been previously demonstrated [9]. There were, however, two exceptions which are worth to consider. Firstly, it was the departure from linearity of the effector curve in the antibody equation (Fig. 4b). This suggested that in the presence of the ongoing A cell lysis the E_B fraction of the effector was less efficient than it would have been alone, i.e. the A cells exerted an inhibitory effect on the lysis of B cells. Secondly, the profile function of target cells in the effector equation approached a horizontal line (Fig. 2a,b), suggesting that on a per cell basis the complement worked more efficiently at high than at low target cell concentration. This phenomenon was interpreted to mean that a relatively low number of target cells were able to accumulate a higher concentration of complement on themselves than what was necessary for lysis, while addition of further target cells to the system resulted in a more 'economical' complement distribution, i.e. more erythrocytes were lysed.

All effector profile functions demonstrated a characteristic lag phase (Fig. 3a,b). It seems, therefore, that there is a threshold value (E_t) of the complement under which practically no lysis occurs, above this threshold, lysis is closely linear in the surplus ($E - E_t$) of effector. A molecular analysis by Kitamura and Inai [8] of the complement mediated lysis revealed that lysis of

sensitized sheep erythrocytes, opsonized by C1-8 components of the human complement (EAC1-8hu) proceeded with lag phase in the presence of C9hu. Their results suggested that in the case of human complement components C1-8hu and C9hu haemolysis was a multi-hit process, therefore more than one molecule of C9hu was necessary to produce a membrane lesion responsible for haemolysis of EAC1-8hu. Takeda et al. [12] confirmed that C9hu step was indeed a multi-hit process, but found that the whole human complement serum followed a one-hit process. The lag phase of our effector profile functions suggested, however, that the whole human complement serum followed a multi-hit process in the case of human blood group A and B erythrocytes target cells.

The relative concentration of the sensitizing antibodies regulate the lytic potential of complement by determining its distribution among competing erythrocytes [1], [2]. A mathematical approximation of this phenomenon can be expressed by a general function of form

$$\ln(E_A/E_B) = U(T_A/T_B) + V(A_A/A_B)$$

without any effect of time and effector. Although there is no theoretical reason to suggest that E_A/E_B does not depend on E, considering the very short half life of the activated complement components compared to the kinetics of the haemolysis, it is reasonable to suggest that this experimental system is not sufficiently 'sensitive' to reveal the kinetic of the complement distribution between the competing erythrocyte populations. It is important to note, however, that according to this approximation effector distribution is proportional to the 5th and 4th power of the ratios of competing antibody and target cell concentration, respectively, i.e. even a small difference in the absolute concentrations of the sensitizing antibodies and target cells will result in a strongly asymmetrical accumulation of complement between the two populations of sensitized cells. Our results therefore suggest that the sensitizing antibody is not only a receptor site for the complement but also a "driving force" which determines its "recruitment" among potential targets. A detailed discussion of the biological consequences of the presented results is given in Tusnády et al. [13].

ACKNOWLEDGEMENT. We thank József Fritz, János Komlós and László Varga for critical discussion, Katalin Kőkuti for expert technical assistance and Irén Imre and Norbert Bognár for the preparation of the manuscript.

REFERENCES

[1] BAKÁCS, T., RINGWALD, G. and KIMBER, I., Differential susceptibility of haemolysin-sensitized erythrocytes to complement-mediated citotoxicity: a biological manifestation of the relationship between antibody availability and complement fixation, *J. Clin. Immunol.* **30** (1989), 61–67.

[2] BAKÁCS, T., RINGWALD, G., TUSNÁDY, G., VÉGH, Zs. and KLEIN, E., Differences in the competitive capacity of monoclonal antibody sensitized human A and B erythrocytes in complement-mediated citotoxicity, *J. Clin. Immunol.* **32** (1990), 167–175.

[3] Bhakdi, S. and Tranum-Jensen, J., Complement lysis: a hole is a hole, *Immunol. Today* **12** (1991), 318–320.

[4] Cooper, N. R., Complement evasion strategies of microorganisms, *Immunol. Today* **12** (1991), 327–331.

[5] Esser, A. F., Big MAC attack: complement proteins cause leaky patches, *Immunol. Today* **12** (1991), 316–318.

[6] Hänsch, G. M., Hammer, C., Rother, U. and Shin, M. L., Studies on activation of C5 and C6 by C5-convertase or low pH, *Fed. Proc.* **39** (1980), 700–709.

[7] Kinoshita, T., Biology of complement: the overture, *Immunol. Today* **12** (1991), 291–295.

[8] Kitamura, H. and Inai, S., Molecular analysis of the reaction of C9 with EAC1-8, *J. Immunol.* **113** (1974), 1992–2003.

[9] Mayer, M. M., Complement and complement fixation, In: *Experimental Immunochemistry*, second edition. Edited by Kabat and Mayer, C. C. Thomas, 1961, 133–241.

[10] Novotny, J., Protein antigenecity: a thermodynamic approach, *Molecular Immunology* **28** (1991), 201–207.

[11] Sim, R. B. and Reid, K. B. M., C1: molecular interactions with activating systems, *Immunol. Today* **12** (1991), 307–311.

[12] Takeda, J., Kozono, H., Takata, Y., Hong, K., Kinoshita, T., Sayama, K., Tamaba, E. and Inoue, K., Number of hits necessary for complement-mediated haemolysis, *Microbiol. Immunol.* **30** (1986), 461–468.

[13] Tusnády, G., Erdei, A. and Bakács, T., Cold target competition analysis of the classical activation pathway of complement-mediated cytotoxicity, *Molecular Immunology* **29** (1992), 1347–1355.

(Received September 12, 1991)

ORSZÁGOS ONKOLÓGIAI INTÉZET
IMMUNOLÓGIAI OSZTÁLY
RÁTH GYÖRGY U. 7–9
H–1122 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
VALÓSZÍNŰSÉGELMÉLETI TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 329–344

COVERABLE GRAPHS

S. MILICI[1] and ZS. TUZA[2]

Abstract

We introduce and study the concept of F-coverable graphs. For two graphs F and G, we say that G is F-coverable if some weights $f(H)$ can be assigned to the subgraphs H isomorphic to F in G in such a way that $0 \leqq f(H) \leqq 1$ and $\sum_{e \in H \subset G} f(H) = 1$ holds for every edge e of G. We concentrate on the case where F is the complete graph on three vertices. We prove necessary conditions, reduction theorems, and non-existence results concerning K_3-coverable graphs. Our theorems can be applied in the theory of triple systems with index $\lambda > 1$, and further applications are expected in design theory.

1. Introduction

Let F and G be two finite undirected graphs, and let $\mathcal{F}$ be the collection of those subgraphs of G which are isomorphic to F. We say that G is F-*coverable* if there is a real function $f: \mathcal{F} \to [0,1]$ such that $\sum_{e \in H \in \mathcal{F}} f(H) = 1$ holds for every edge e of G.

The concept of F-coverable graphs is a common generalization of various structures much studied in combinatorics. Let us show two examples.

EXAMPLE 1. If G has an edge decomposition into subgraphs isomorphic to F (sometimes this relation is denoted by $F \mid G$, "F divides G"), then G has an integer-valued covering function $f: \mathcal{F} \to \{0,1\}$.

EXAMPLE 2. Let $\mathcal{S}$ be a triple system of index $\lambda \geqq 2$, i.e., a multiset of 3-element subsets — called blocks — of an n-element set X such that each pair $x, x' \in X$ is contained in precisely λ blocks. The *leave graph*, $G(\mathcal{S})$, of $\mathcal{S}$ has vertex set X, with two vertices x, x' being adjacent if and only if the pair $\{x, x'\}$ is contained in blocks of multiplicities less than λ. Then G has

1991 *Mathematics Subject Classification.* Primary 05C70, 05B40; Secondary 05B07.
Key words and phrases. Graph, triangle, edge cover, triple system, leave graph.

[1]Research supported by G.N.S.A.G.A and M.U.R.S.T.

[2]This research of the second author was done while under a C.N.R. Grant, visiting Università di Catania, November 1990.

a K_3-covering, e.g. we can assign $m(S)/\lambda$ to each block S of multiplicity $m(S) < \lambda$.

Motivated by the second example, in this paper we restrict our attention to K_3-coverings f such that f assigns a value strictly smaller than 1 to each triangle of G. More formally, assuming that G is a graph with edge set E, we denote by $\mathcal{T} = \mathcal{T}(G)$ the collection of "triangles" of G, i.e., 3-element *edge* sets $\{e, e', e''\} \subseteq E$ forming a complete subgraph K_3 on three vertices. A real function

$$f: \mathcal{T} \longrightarrow [0, 1]$$

is called a *uniform covering* of G if

$$\sum_{\substack{T \in \mathcal{T} \\ e \in T}} f(T) = 1$$

for every edge $e \in E$. A uniform covering f is *strict* if

$$f(T) < 1 \text{ for all } T \in \mathcal{T}.$$

We say that G is *coverable* if it has a strict uniform covering (i.e., "coverable" is a slightly stronger requirement than "K_3-coverable"). By *covering* we shall always mean a strict uniform covering.

In Section 2 we show how a coverable graph G can be reduced to a smaller one when some local structures with small vertex degrees occur in G, while in Section 3 we present necessary conditions for a graph to be coverable. Those observations are then combined to prove that a coverable graph must have a relatively large number of edges (Sections 4 and 5). Beside giving general bounds, we pay considerable attention to coverable graphs having all degrees even. An application of those results is given in [3].

Although we restrict our investigations to the case $F = K_3$ throughout, most results of Section 3 can be generalized for F-coverable graphs, for any F.

NOTATION. Let $G = (V, E)$ be a graph with vertex set V and edge set E, and let $x \in V$ be any vertex. We denote by $N(x)$ the set of vertices adjacent to x (= the set of *neighbours* of x); by $\Gamma(x)$ and $\Gamma[x]$ the subgraph of G induced by $N(x)$ and by $N(x) \cup \{x\}$, respectively (= the *open* resp. the *closed* neighbourhood of x); by K_p and by C_p the complete graph and the cycle on p vertices; and $d(x) := |N(x)|$ stands for the *degree* of x. If H is any graph, $V(H)$ and $E(H)$ denote the vertex set resp. the edge set of H. Finally, by definition, the vertices of a triangle T are those of the edges $e \in T$.

REMARK 1. As an immediate consequence of the definitions, each edge $e \in E$ of a coverable graph $G = (V, E)$ is contained in at least two triangles $T \in \mathcal{T}(G)$. In particular, $d(x) \geqq 3$ for every non-isolated vertex $x \in V$, $\Gamma(x)$ has minimum degree at least 2, and if $d(x) = 3$ then $\Gamma(x) = K_3$ and $\Gamma[x] = K_4$.

2. Reducibility

In this section we point out how a coverable graph $G=(V,E)$ can be reduced to a smaller one when the neighbourhood of a vertex of degree 3, 4, or 5 satisfies some properties. Those reductions will be referred to as <R3>, <R4>, and <R5>.

<R3> (Degree-3 reducibility)

If two adjacent vertices $x, x' \in V$ have degree 3, then $G \backslash E(\Gamma[x])$ is coverable if and only if so is G.

The other two reductions deal with the cases where <R3> cannot be applied.

<R4> (Degree-4 reducibility)

If a vertex x with $d(x)=4$ has two non-adjacent neighbours of degree 3, then $G \backslash E(\Gamma[x])$ is coverable if and only if so is G.

<R5> (Degree-5 reducibility)

If a vertex x with $d(x)=5$ has two non-adjacent neighbours of degree 3 and at least two neighbours x', x'' of degree 5, then $G \backslash E(\Gamma[x] \cup \Gamma[x'] \cup \Gamma[x''])$ is coverable if and only if so is G.

In order to prove the validity of those reductions, we shall need the following simple observation.

LEMMA 2. *If $\Gamma(x)$ is an odd cycle (and, in particular, if $d(x)=3$) then in every covering f of G, $f(T)=1/2$ holds for all triangles T incident to x.*

PROOF. Let $T_1, T_2, \ldots, T_d = T_0$ $(d=d(x))$ be the triangles incident to x, taken in a cyclic order; i.e., T_i and T_j $(i \neq j)$ share an edge if and only if $|i-j|=1$. Then T_i and T_{i+1} are the only triangles incident to their common edge, therefore

$$f(T_i) + f(T_{i+1}) = 1 \quad \text{for} \quad 0 \leqq i < d.$$

Taking the sum of these d inequalities divided by two, we obtain

$$\sum_{i=1}^{d} f(T_i) = d/2.$$

On the other hand, summing up $f(T_{2j}) + f(T_{2j+1})$ for $1 \leqq j < d/2$,

$$\sum_{i=2}^{d} f(T_i) = (d-1)/2$$

follows for d odd. We conclude that $f(T_1)=1/2$, and similarly $f(T_i)=1/2$ for all i, by symmetry.

In the particular case where $d(x)=3$ the property is implied by the fact that $\Gamma(x)$ is the cycle of length 3. □

PROOF of <R3>. If x and x' are adjacent vertices with $d(x)=d(x')=3$, then $N(x)\cup\{x\}=N(x')\cup\{x'\}$ and $\Gamma[x]=K_4$. Moreover, by Lemma 2, each triangle T sharing a vertex with $\{x,x'\}$ has $f(T)=1/2$, in every covering f of G. Hence, the four triangles meeting $\{x,x'\}$ provide a uniform covering of $\Gamma[x]$ (as each edge of $\Gamma[x]$ is contained in precisely two of those triangles). Consequently, f is a coverings of G if and only if it is a covering of $G\backslash E(\Gamma[x])$ and assigns weight $1/2$ to each triangle meeting $\{x,x'\}$. In particular, there is a one-to-one correspondence between the coverings of G and those of $G\backslash E(\Gamma[x])$. □

PROOF of <R4>. Let $d(y)=d(y')=3$, $y,y'\in N(x)$. Then $\Gamma[y]\cap\Gamma[y']$ is a triangle T_0 and $\Gamma[y]\cup\Gamma[y']=\Gamma[x]$ (since x has just two neighbours distinct from y and y').

Each edge of T_0 is contained in one triangle T_y meeting y and one triangle $T_{y'}$ meeting y'. By Lemma 2, $f(T_y)=f(T_{y'})=1/2$ holds in any covering f of G, so that the triangles meeting $\{y,y'\}$ provide a uniform covering of $\Gamma[x]$. Thus, f is a covering on $G\backslash E(\Gamma[x])$ if and only if it can be extended to a covering of G. □

PROOF of <R5>. Let $d(x)=5$, $N(x)=\{x',x'',y,y',z\}$, $d(x')=d(x'')=$ $=5$, $d(y)=d(y')=3$. Observe first that $|N(y)\cap N(y')|=2$. Indeed, $N(y)$ and $N(y')$ are 3-element subsets of the 4-element set $\{x,x',x'',z\}$, i.e., they share at least two vertices. On the other hand, if $\Gamma[y]\cap\Gamma[y']=T_0$ were a triangle, then the argument described in the proof of <R4> would yield that $G\backslash E(\Gamma[y]\cup\Gamma[y'])$ is a coverable graph G'. In this G', however, x would have degree 1, a contradiction.

Let e_0 be the unique edge in $\Gamma[y]\cap\Gamma[y']$. Since $x\in e_0$, we can assume without loss of generality that $x'\notin e_0$. Then the six triangles meeting $\{y,y'\}$ have total weight 1 on e_0, as well as on each of the six edges meeting $\{y,y'\}$; moreover, they have weight $1/2$ on four other edges of $\Gamma[x]$, two of them incident to x, the other two not containing x, as shown in Figure 1. The wavy lines at each vertex of the figure correspond to the edges covered uniformly (i.e., with total weight 1) by the triangles considered so far, while a label $1/2$ on an edge indicates demand for the total weight of the other triangles sitting on the edge marked.

So far, three edges are covered at x, and $d(x)=5$, therefore the two edges of weight $1/2$ have to induce a triangle T_1 with $f(T_1)=1/2$. This T_1 completes the covering at x, and yields demand $1/2$ on its third edge incident to x'. At this point, however, x' can have just one further neighbour z' (for $d(x')=5$), with demand 1 on the edge $x'z'$. Consequently, $x'z'$ induces triangles of weight $1/2$ with each of the two edges of weight $1/2$ incident to x'. In particular, z is adjacent to $e_0\backslash\{x\}$, yielding $x''\notin e_0$, $e_0=xz$, $d(z)\geqq 6$.

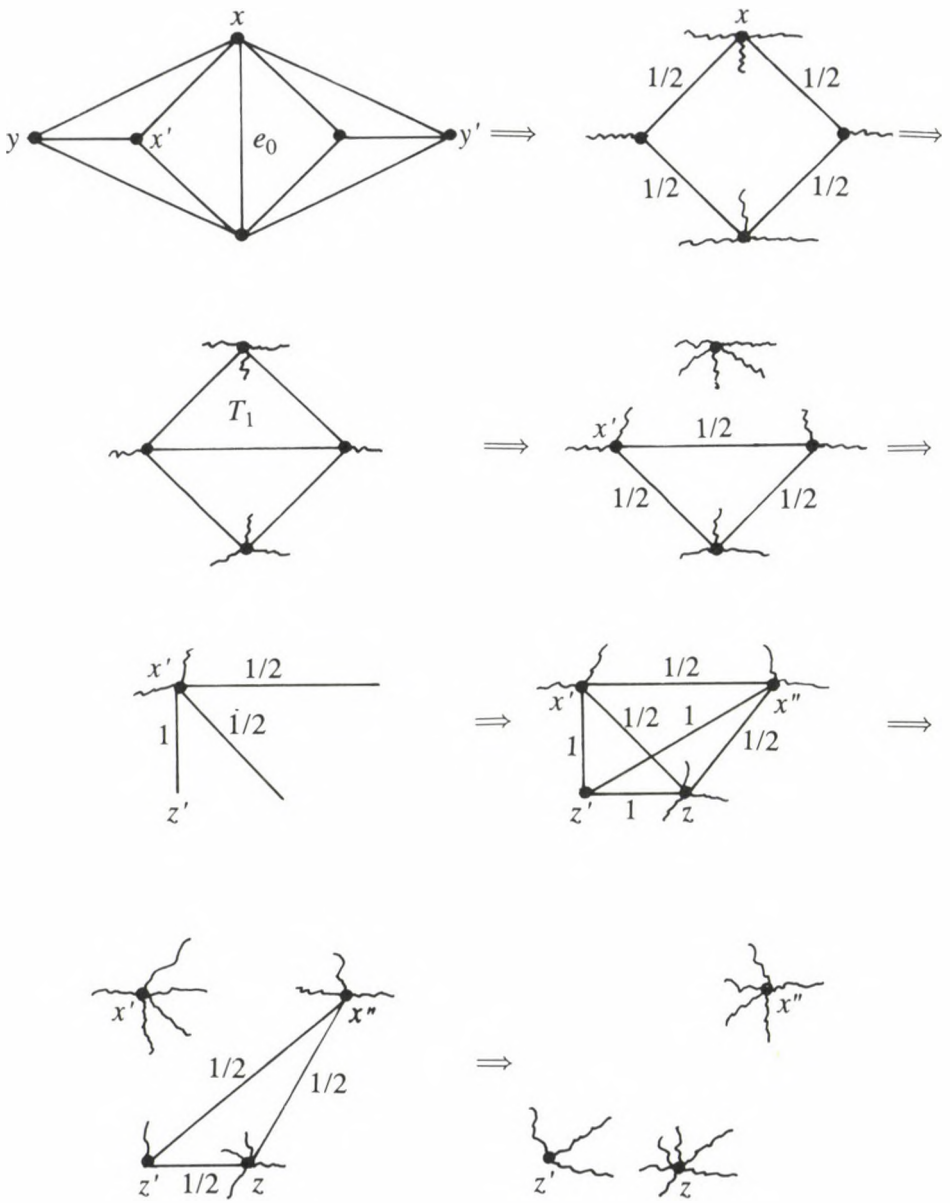


Fig. 1

The two triangles containing $x'z'$ complete the covering of the edges incident to x', leaving demand $1/2$ on the edges $x''z'$ and zz'. Since $d(x'') = 5$, it follows that $f(x''zz') = 1/2$, and this triangle completes the covering

of all edges in $\Gamma[x] \cup \Gamma[x'] \cup \Gamma[x'']$. (This graph does not have any further edge, since z and z' are the only vertices in the neighbourhood which might be incident to non-covered edges, but the edge zz' has already been covered).

Hence, we conclude that f covers G if and only if it assigns the weights given above to the triangles in $\Gamma[x] \cup \Gamma[x'] \cup \Gamma[x'']$, and its restriction to $G \setminus E(\Gamma[x] \cup \Gamma[x'] \cup \Gamma[x''])$ is a covering as well. □

The effects of the above reductions on the degree sequence of G are as follows.

<R3> deletes two vertices of degree 3 and decreases the degrees of two further vertices by 3 (some or both of them may become isolated).

<R4> deletes two vertices of degree 3 and decreases the degrees of three further vertices by 4 (at least one of them becomes isolated).

<R5> deletes two vertices of degree 3, deletes three vertices of degree 5, and decreases the degrees of two further vertices by 3 and 6, respectively (some or both of them may become isolated).

It is convenient to give a name to graphs which cannot be reduced to smaller ones by those operations.

DEFINITION. A graph is called *irreducible* if <R3> cannot be applied in it. Call a graph *strongly irreducible* if none of <R3>, <R4>, and <R5> can be applied in it.

3. Cut and degree constraints

In this section we provide three necessary conditions for a graph to be coverable. Those properties will be useful in the non-existence proofs later on.

For $X \subset V$, in a graph $G = (V, E)$, denote by $N(X)$ the set of those vertices of $V \setminus X$ which have a neighbour in X, by $e(X)$ the number of edges induced by X in G, and by $t(X)$ the number of triangles $T \in \mathcal{T}(G)$ belonging entirely to X and having at least one vertex of degree 3. For two disjoint sets $X, Y \subset V$, $e(X, Y)$ denotes the number of edges joining X with Y (called X–Y edges, for short).

<C> (Cut Constraint)

Let $G = (V, E)$ be a coverable graph. If $X \cup Y = V$ and $X \cap Y = \emptyset$, then

$$e(X, Y) + 3t(X) + 3t(Y) \leqq 2e(X) + 2e(Y).$$

PROOF. Every triangle containing an X–Y edge has precisely two X–Y edges and one edge induced by X or by Y. Thus, the total weight of those triangles is $\frac{e(X,Y)}{2}$, i.e., their contribution to a covering of the edges contained in X or in Y is exactly $f_1 = \frac{e(X,Y)}{2}$. Moreover, by Lemma 2, all

triangles incident to degree-3 vertices must have weight 1/2. Hence, the total contribution of such triangles contained in X or in Y to the covering is $f_2 = \frac{3}{2}(t(X)+t(Y))$. Certainly, the sum $f_1 + f_2$ of weights cannot exceed the demand $e(X)+e(Y)$. □

Property <C> gives an upper bound on $e(X,Y)$, the number of edges joining the partition classes. The following related inequality provides a lower bound under some assumptions.

<C*> (Dual Cut Constraint)

Let the graphs F_i $(1 \leqq i \leqq t)$ be triangle free induced subgraphs of a coverable graph G, with $e(F_i)$ edges and with mutually disjoint vertex sets $V(F_i)$. If G has e^* edges which meet some $V(f_i)$ but are not contained in any $V(F_i)$, then

$$e^* \geqq 2e(F_1) + \ldots + 2e(F_t).$$

PROOF. The total demand on the edges of the F_i is $\sum e(F_i)$, and each triangle sharing an edge with some F_i has exactly two edges in e^*. Thus, in any covering f of G, the total weight of the e^* edges is at least $2\sum e(F_i)$ (just provided by triangles covering $E(F_1) \cup \ldots \cup E(F_t)$). This sum cannot exceed their demand e^*. □

Recall that adjacent vertices of degree 3 can be eliminated from any coverable graph G, applying <R3>. One can see that every G can be transformed to a *unique* irreducible graph G_0 in this way.

DEFINITION. The *kernel* G^* of a graph G is the subgraph induced by the vertices of degree $\geqq 4$ in the (unique) irreducible graph G_0 obtained from G by repeated application of <R3>. The δ^*-sequence (or *reduced degree sequence*) $\delta^*(G)$ of G is the *multiset* $\{\delta_1, \ldots, \delta_k\}$, $k = |V(G^*)|$, in which each vertex of G^* is represented by its degree in G_0.

REMARK 3. (i) Since <R3> does not change the degrees modulo 3, $\delta^*(G)$ is non-empty whenever G has a vertex x with $d(x) \equiv 1$ or 2 (mod 3).

(ii) If all degrees are multiples of 3, then $\delta^*(G)$ may or may not be empty. For instance, if G is the graph with $3k+1$ vertices composed from edge-disjoint K_4's sharing one vertex, then $G_0 = G^* = \emptyset$, $\delta^*(G) = \emptyset$, while $(K_7 \backslash K_4)_0 = K_7 \backslash K_4$, $(K_7 \backslash K_4)^* = K_3$, $\delta^*(K_7 \backslash K_4) = \{6,6,6\}$.

(iii) The δ^*-sequence may be non-empty even if G is decomposable into K_4's, e.g. $\delta^*(K_{13}) = \{12, \ldots, 12\}$.

<D> (Degree Constraint)

If $\delta^*(G) = \{\delta_1, \ldots, \delta_k\} \neq \emptyset$ in a coverable graph G, then

$$k \geqq \left\lceil \frac{1}{2} \max_{1 \leqq i \leqq k} \delta_i \right\rceil + 1.$$

Moreover, a vertex of degree δ_i in G_0 has degree at least $\delta_i/2$ in G^*.

PROOF. Since every vertex of G^* can have at most $k-1$ neighbours in G^*, the lower bound on k is a consequence of the second part of the assertion. To prove the latter, let $x \in V(G^*)$ be a vertex of degree d in G_0, adjacent to t vertices $v_1, \ldots, v_t$ of $V(G_0) \backslash V(G^*)$ and hence having $d-t$ neighbours in G^*. Note first that each v_i has degree 3 by definition. Moreover, since G_0 is irreducible, any triangle containing an edge xv_i has its third vertex in G^*. Thus, the total weight t of triangles containing edges xv_i cannot exceed the demand $d-t$ on the edges incident to x in G^*. This fact implies $t \leqq d/2$. □

From <D> one can deduce the following observations.

LEMMA 4. *For the class of coverable graphs,*

(i) *there is no δ^*-sequence of length 1;*

(ii) *there is no δ^*-sequence of length 2;*

(iii) *the unique δ^*-sequence of length 3 is $\{4,4,4\}$, with G_0 shown in Fig. 2(a);*

(iv) *there are precisely three δ^*-sequences of length 4, namely $\{5,5,4,4\}$, $\{6,5,5,5\}$, and $\{6,6,6,6\}$, with their corresponding G_0 shown in Fig. 2(b), 2(c), and 2(d), respectively.*

PROOF. (i), (ii), and (iii) Since each degree in a δ^*-sequence is at least 4, we must have $k \geqq 3$ vertices in G^* by the Degree Constraint. In case of equality, no degree can exceed 4, i.e., $\delta^*(G) = \{4,4,4\}$; and the three vertices of G^* must be pairwise adjacent, each having precisely two neighbours of degree 3.

(iv) Assuming $k=4$, it follows from <D> that $\delta_i \leqq 6$ holds for $1 \leqq i \leqq k$. Also, we have $\delta_i \geqq 4$, and consequently each vertex of G^* has a neighbour of degree 3 in $G_0 \backslash G^*$. This also implies that G_0 has at least two vertices of degree 3.

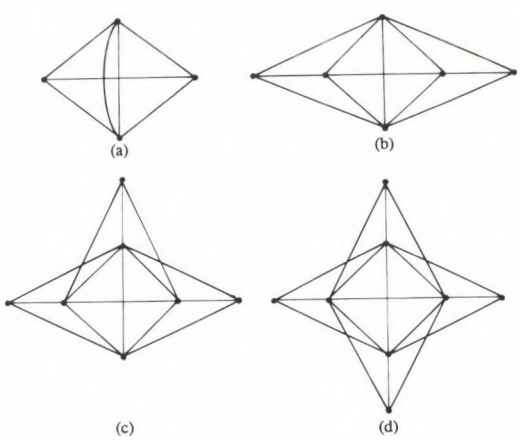


Fig. 2. The four coverable and irreducible graphs with δ^*-sequences of lengths at most four.

Let $y, y' \in V(G_0)\backslash V(G^*)$. If $\Gamma(y) = \Gamma(y')$ held, Lemma 2 would imply that the graph $G' = G_0\backslash E(\Gamma[y] \cup \Gamma[y'])$ is coverable as well. This edge deletion decreases precisely three degrees of G^* by 4, leading to a coverable graph of minimum degree $\leqq 2$, a contradiction. Hence, we obtain $\Gamma(y) \neq \Gamma(y')$ for all $y \neq y'$, $d(y) = d(y') = 3$. This property yields $2 \leqq |V(G_0)\backslash V(G^*)| \leqq 4$. If there are three or four vertices in $G_0\backslash G^*$, then their neighbourhoods induce K_4 in G^*. Also, though the neighbourhoods of two vertices form K_4 minus an edge only, the missing edge has to be added for otherwise G_0 were not coverable.

Conversely, it is easily seen that for each graph G_0 in Fig. 2 there is a covering $f\colon \mathcal{T}(G_0) \to \{0, 1/2\}$. □

4. Graphs with small excess

The results of this section (and also of the next one) are valid for all graphs, not only for irreducible ones.

Let $G = (V, E)$ be a graph on n vertices, $V = \{v_1, \ldots, v_n\}$. The degree sequence $\delta(G) = \{d_1, \ldots, d_n\}$ is defined as $d_i = d(v_i)$. We introduce the concept of *excess sequence* (ε-sequence) that consists of the non-zero elements of the multiset $\{\varepsilon_1, \ldots, \varepsilon_n\}$, where

$$\begin{aligned} \varepsilon_i &= d_i && \text{for } d_i \text{ even,} \\ \varepsilon_i &= d_i - 3 && \text{for } d_i \text{ odd.} \end{aligned}$$

The *excess* $\varepsilon(G)$ of G is now defined as

$$\varepsilon(G) = \frac{1}{2} \sum_{i=1}^{n} \varepsilon_i.$$

Since the excess ε_i at v_i is even by definition, the excess of a graph is an integer. Note further that if G is coverable and a vertex v_i has the smallest positive excess, $\varepsilon_i = 2$, then $d_i = 5$ holds.

Let us begin with the observation that every ε with $\varepsilon \equiv 0 \pmod 3$ is the excess of some graph. Indeed, take $2k+1$ vertex-disjoint triangles ($k \geqq 0$) and join all their vertices to a new vertex w (in other words, $2k+1$ edge-disjoint K_4's are incident to w). Then w has degree $6k+3$, with excess $6k$, while the degree-3 vertices have excess 0. Hence, the graph has excess $3k$. (The same excess is obtained if we take only $2k$ blocks isomorphic to K_4.) Part (iii) of the following result shows that the other two residue classes modulo 3 do not provide us with graphs of a very small excess.

THEOREM 5. *Let G be a coverable graph without isolated vertices.*

(i) *If $\varepsilon(G) = 0$, then $|V(G)| \equiv 0 \pmod 4$ and all connected components of G are isomorphic to K_4.*

(ii) *If $\varepsilon(G) = 3$, then all but one connected components of G are isomorphic to K_4, and the last component is one of the two graphs shown in Fig. 3. In particular, if all degrees of G are odd, then $|V(G)| \equiv 2 \pmod 4$.*

(iii) *If $\varepsilon(G)$ is not a multiple of 3, then $\varepsilon(G) \geqq 8$.*

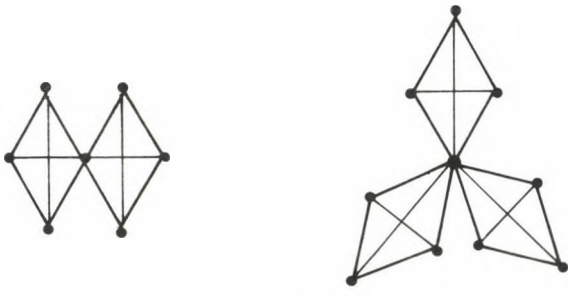


Fig. 3

PROOF. (i) The assumption $\varepsilon(G) = 0$ implies that G is regular of degree 3. Hence, $\Gamma[x] = K_4$ for all $x \in V(G)$, and G cannot have larger components.

(ii) Now $\varepsilon(G) = 3$ implies $\sum \varepsilon(x) = 6$. Since $\varepsilon(x)$ is even, there cannot occur more than one vertex of degree $\geqq 4$, for otherwise G would have a δ^*-sequence of length 2 or of length 3 but not $\{4,4,4\}$, contradicting Lemma 4. Thus, $d(x) = 6$ or 9, and $\delta^*(G) = \emptyset$. Then the degree of x uniquely determines G, yielding the graphs shown in Fig. 3.

(iii) Note first that the reductions <R3>, <R4>, and <R5> do not increase the excess (usually they decrease it), and the number of edges deleted in them is a multiple of 3. Hence, in order to prove the statement by contradiction, we may assume that G is a *strongly irreducible* graph with $7 \geqq \varepsilon(G) \equiv 1$ or 2 (mod 3).

The ε-sequence cannot be shorter than the δ^*-sequence. (In fact, for irreducible graphs they have the same length, as the excess-0 vertices are omitted from both of them). By Lemma 4, the δ^*-sequences of lengths $\leqq 4$ yield ε-sequences $\{4,4,4\}$, $\{2,2,4,4\}$, $\{6,2,2,2\}$, and $\{6,6,6,6\}$. In each of these four cases the excess of the graph in question is a multiple of 3, and consequently the initial graph G cannot have $\varepsilon(G) \equiv 1$ or 2 (mod 3). Hence, $|V(G^*)| \geqq 5$ and, since $\varepsilon(x) \geqq 2$ for each $x \in V(G^*)$, $\varepsilon(G) \geqq 5$ also holds.

If $\varepsilon(G) = 5$, then we must have $|V(G^*)| = 5$, $\varepsilon(x) = 2$ and $d(x) = 5$ for every $x \in V(G^*)$, $d(y) = 3$ for every $y \in V(G) \setminus V(G^*)$, and $V(G) \setminus V(G^*)$ has to be an independent set. Then each $x \in V(G^*)$ has a neighbour of degree 3, i.e., $V(G^*)$ meets at least five edges not belonging to G^*. On the other hand, the total degree sum for the vertices of G^* is 25, therefore the number of edges not in G^* is odd. Finally, their number is a multiple of 3, so that there are exactly 9 or at least 15 such edges. Then G^* has 8 or at most 5 edges, respectively; hence the latter is ruled out by <C>, as well as by <D>.

Assuming $|E(G^*)| = 8$, we have three vertices y, y', y'' of degree 3. Then

some $x \in V(G^*)$ has just one neighbour of degree 3. Set $X = V(G^*)\backslash\{x\}$ and $Y = \{x, y, y', y''\}$. Now $e(X) = 4$, $e(Y) = 1$, and $e(X, Y) = 12$, contradicting <C>.

Suppose that $\varepsilon(G) = 7$, $\sum_{X \in V(G^*)} \varepsilon(x) = 14$, $|V(G^*)| \geqq 5$. The four possibilities to decompose 14 into at least five positive even numbers are:

$$6+2+2+2+2, \quad 4+4+2+2+2, \quad 4+2+2+2+2+2, \quad 2+2+2+2+2+2+2.$$

We are going to show that none of them corresponds to the excess sequence of a coverable graph. Recall that a vertex of excess k has degree k or $k+3$, and that $\varepsilon(x) = 2$ implies $d(x) = 5$ if G is coverable. Note further that excess 4 and 6 may correspond to degree 4 or 7 and 6 or 9, respectively.

Case 1. The ε-sequence is $\{6, 2, 2, 2, 2\}$.

Now the δ^*-sequence is $\{6, 5, 5, 5, 5\}$, for $d(x) = 9$ is ruled out by <D>. For the same reason, G^* has minimum degree $\geqq 3$, so that each vertex of degree 5 is adjacent to at least two other vertices of degree 5. Since <R5> cannot be applied in G (by the assumption of strong irreducibility), the degree-5 vertices can have no more than one neighbour in $G\backslash G^*$.

Thus, $|V(G)\backslash V(G^*)| = 2$, $G^* = K_5$, and G is the graph exhibited in Fig. 4(a). Let X be the set of the four marked vertices, and set $Y = V(G)\backslash X$. With the notation of <C> we have $e(X, Y) = 10$, $t(X) = 1$ (the triangle in question is shaded), and $e(X) = e(Y) = 3$. Consequently,

$$e(X, Y) + 3t(X) = 13 > 12 = 2e(X) + 2e(Y),$$

implying the contradiction that G is not coverable, by <C>.

Case 2. The ε-sequence is $\{4, 4, 2, 2, 2\}$.

The three possible degree sequences belonging to this particular ε-sequence are $\{7, 7, 5, 5, 5\}$, $\{7, 4, 5, 5, 5\}$, and $\{4, 4, 5, 5, 5\}$. As described in <D>, vertices of degree $\geqq 5$ have at least 3 neighbours in G^*. Moreover, if a degree-4 vertex were adjacent to more than one vertices of $G\backslash G^*$, then <R4> could be applied, contradicting our assumption. Hence, in each of the three subcases, G^* has minimum degree $\geqq 3$, i.e., has degree sum 16, 18, or 20.

Taking into account that the number of edges joining G^* with $G\backslash G^*$ is a multiple of 3, we obtain that $G^* = K_5$, and its vertices of degree 7, 5, 4 have 3, 1, 0 neighbours in $G\backslash G^*$, respectively. This fact rules out each of the three degree sequences as follows.

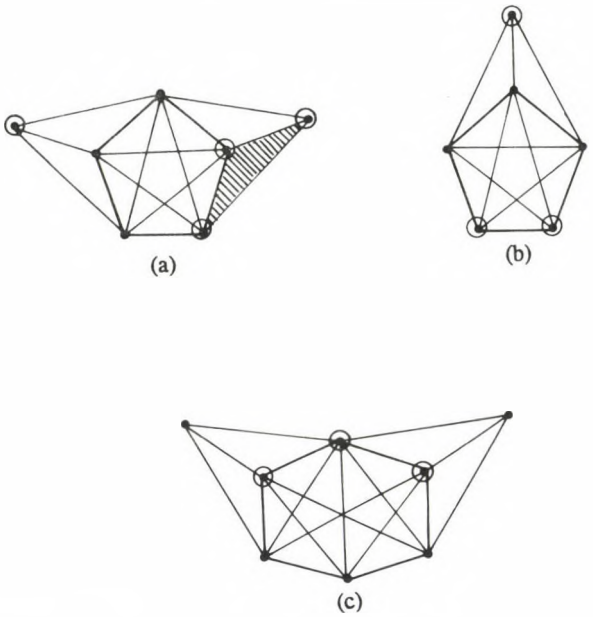


Fig. 4. Some graphs ruled out by the Cut Constraint; the vertices of the partition class X are marked.

In $\{7,7,5,5,5\}$ we have $|V(G)\backslash V(G^*)| = 3$. Moreover, the two degree-7 vertices, say x and x', are adjacent to the three vertices of degree 3, yielding the contradiction that the triangles containing the edge xx' should have total weight at least $3/2$, by Lemma 2.

In $\{7,4,5,5,5\}$, the degree-7 vertex should have three neighbours of degree 3. Thus, at least 9 edges should join G^* and $G\backslash G^*$. However, the degree sequence admits just 6 such edges.

In $\{4,4,5,5,5\}$, $G\backslash G^*$ has one vertex, adjacent to the degree-5 vertices of G^*, i.e., G is the graph shown in Fig. 4(b). Then the set X of the marked vertices, with $Y = V(G)\backslash X$, has $e(X,Y) = 9 > 8 = 2+6 = 2e(X)+2e(Y)$, contradicting <C>.

Case 3. The ε-sequence is $\{4,2,2,2,2,2\}$.

Now there are two possibilities for the degree sequence, namely $\{7,5,5,5,5,5\}$ and $\{4,5,5,5,5,5\}$. As in Case 1, each degree-5 vertex can have at most one neighbour in $G\backslash G^*$, so that $|V(G)\backslash V(G^*)| \leqq 2$.

In the first subcase, the degree-7 vertex forces that indeed there are two vertices of degree 3, and the graph is the one in Fig. 4(c) (since $\Gamma(y) = K_3$ for $d(y) = 3$). Again, we have a contradiction to <C> as $e(X,Y) = 13 > 12 = 4+8 = 2e(X)+2e(Y)$.

In $\{4,5,5,5,5,5\}$, the number of odd degrees in the subgraph G^* must be even, therefore $G\backslash G^*$ has precisely one vertex y, and G^* has 13 edges. The three neighbours of y are adjacent to each other and to y, so that they have at most two neighbours not in $N(y)$, yielding at most 6 edges from

$N(y)$ to $V(G^*)\backslash N(y)$ and just 3 edges in $N(y)$. Hence, the 3-element set $V(G^*)\backslash N(y)$ should induce more than 3 edges, a contradiction.

Case 4. The ε-sequence is $\{2,2,2,2,2,2,2\}$.

Of course, the degree sequence must be $\{5,5,5,5,5,5,5\}$. Again, by the assumption that $<$R5$>$ cannot be applied, each vertex has at most one neighbour in $G\backslash G^*$. Since the degree sum in G^* is even and the number of edges meeting $G\backslash G^*$ is a multiple of 3, $G\backslash G^*$ consists of just one vertex y and G^* has 16 edges.

Now each $x \in N(y)$ has 2 neighbours in $V(G^*)\backslash N(y)$, and $N(y)$ induces 3 edges; that is, 9 edges of G^* meet $N(y)$. Thus, the 4-element set $V(G)\backslash N(y)$ should induce 7 edges, a contradiction. □

We note that part (iii) of Theorem 5 is sharp for both residue classes $\neq 0$ modulo 3; namely, K_5 has excess 10 while $K_6 - e$ (an edge deleted from K_6) has excess 8, and one can see that both graphs are coverable.

5. Graphs with all degrees even

Let us say that G is an *even graph* if the degree of every vertex $x \in V(G)$ is even. Below we show how the results of the previous section can be improved for such graphs. (Those stronger bounds have proved to be useful in the study of triple systems, see [3].) Note that in coverable even graphs without isolated vertices, the minimum degree is at least 4, and the ε-sequence is identical to the degree sequence; in particular, $\varepsilon(G) = |E(G)|$ holds.

THEOREM 6. *Suppose that an even graph G is coverable.*

(i) *If $\varepsilon(G) \equiv 0 \pmod 3$, then either $\varepsilon(G) \geqq 18$ or $\varepsilon(G) = 12$ and G is isomorphic to $K_6 - 3K_2$, the complete graph of order 6 minus three pairwise disjoint edges.*

(ii) *If $\varepsilon(G) \equiv 1 \pmod 3$, then either $\varepsilon(G) \geqq 19$ or $\varepsilon(G) = 10$ and $G = K_5$.*

(iii) *If $\varepsilon(G) \equiv 2 \pmod 3$, then $\varepsilon(G) \geqq 20$.*

PROOF. By $d(x) \geqq 4$ we have $|V(G)| \geqq 5$ and $\varepsilon(G) \geqq 2|V(G)| \geqq 10$. Moreover, if G has a vertex of degree $\geqq 6$, then $|V(G)| \geqq 7$ holds, implying $\sum \varepsilon(x) \geqq$ $\geqq 6+4(|V(G)|-1) \geqq 30$, so that $\varepsilon(G) \geqq 15$ unless G is 4-regular. In particular, $\varepsilon(G) \geqq 15$ holds whenever $\varepsilon(G)$ is odd.

(i) If G is 4-regular, then $\varepsilon(G) = 2|V(G)|$ is even. The unique even multiple of 3 between 10 and 17 is 12, hence in this case we have $|V(G)| = 6$, and G is the unique 4-regular graph, $K_6 - 3K_2$, on 6 vertices.

If G has maximum degree $\geqq 6$ and $\varepsilon(G) < 18$, then $\varepsilon(G) = 15$, $|V(G)| = 7$, G has a vertex y with $d(y) = 6$, and $d(x) = 4$ for all $x \in V(G)\backslash\{y\}$. Then the complement of $G - y$ is a regular graph of degree 2 on 6 vertices, i.e., either the cycle C_6 or two disjoint triangles, $2K_3$.

Let x, x' be either a pair of antipodal points on C_6 or two vertices from distinct triangles of $2K_3$ (according to the structure of $G-y$). The choice of x' implies that x' has degree 1 in $\Gamma(x)$ in G, and therefore G cannot be coverable, by Remark 1.

(ii) Since G has to be 4-regular for $\varepsilon(G) < 15$, the case $\varepsilon(G) = 13$ is impossible, and K_5 is the unique graph having $\varepsilon(G) \equiv 1 \pmod 3$ with $\varepsilon(G) < 16$. For $\varepsilon(G) = 16$, there are two possibilities for the degree sequence; namely, $\{6,6,4,4,4,4,4\}$ and $\{4,4,4,4,4,4,4,4\}$.

In the first case the five degree-4 vertices have to induce a 2-regular graph F_1 which, therefore, is the cycle C_5. Moreover, there is an edge F_2 joining the two degree-6 vertices. Here F_1 and F_2 are vertex-disjoint triangle-free graphs with 6 edges in all, and the graph has $10 < 2 \cdot 6$ further edges. Thus, G is not coverable, according to <C*>.

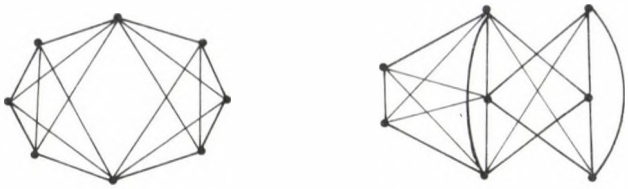
Hence, suppose that G is a (coverable) 4-regular graph on 8 vertices. The neighbourhood $\Gamma(x)$ of any vertex x has minimum degree ≥ 2. Thus, there exist at most 4 edges which are incident to, but are not contained in, $N[x] = N(x) \cup \{x\}$. Consequently, some vertex $y \in V(G) \setminus N[x]$ has at most one neighbour in $N[x]$. Since it has at most two further neighbours in $V(G) \setminus N[x]$, G cannot be 4-regular, a contradiction.

(iii) We have to show that $\varepsilon(G) \neq 11, 14, 17$. We have already seen that $\varepsilon(G) < 15$ could hold only if G were a 4-regular graph on 7 vertices. Hence, the complement of G should be either C_7 or the vertex-disjoint union of C_3 and C_4. The former graph, $G = \overline{C}_7$, is not coverable since it has some edges contained in just one triangle, while the latter is ruled out by <C>, setting $X = V(C_3)$ and $Y = V(C_4)$. (Then $e(X) = 0$, $e(Y) = 2$, and $e(X,Y) = 12$.)

Suppose that $\varepsilon(G) = 17$. The maximum degree in G is 6, for otherwise G would have at least 9 vertices and more than 17 edges. Hence, the degree sequence is either $(6,6,6,4,4,4,4)$ or $(6,4,4,4,4,4,4,4)$. The former corresponds to a unique graph whose degree-6 vertices are adjacent to every vertex and hence they induce a triangle K_3, while each degree-4 vertex has just one degree-4 neighbour, i.e., those four vertices induce $2K_2$. Setting $X = V(K_3)$ and $Y = V(2K_2)$, we obtain $e(X) = 3$, $e(Y) = 2$, and $e(X,Y) = 12$, a contradiction to <C>.

Suppose that the degree sequence is $(6,4,4,4,4,4,4,4)$. Then the degree-6 vertex y is non-adjacent to just one degree-4 vertex x. A vertex adjacent to y but not adjacent to x must have a common neighbour z with x. This z can have at most one common neighbour with x, so that G cannot be coverable. □

The bounds in Theorem 6 are best possible, as shown by the following examples. The graphs in Figs. 5(a) and 5(b) have excess 18 and 19, respectively, and they are easily seen to be coverable. (The former has been obtained from two copies of the graph of Fig. 2(a) by identifying two pairs of degree-3 vertices.) Moreover, a simple coverable graph with excess 20 is $2K_5$.



(a) (b)

Fig. 5. Coverable graphs with excess 18 and 19

6. Concluding remarks

(1) A main motivation of this paper is the study of triple systems in which each pair of points is contained in the same number, $\lambda \geqq 2$, of triples. Applying the results of Sections 4 and 5, in [3] we could settle almost completely the problem of how many blocks of multiplicity λ a triple system $TS(v, \lambda)$ can have. (The missing cases are where $\lambda \equiv 0 \pmod 6$ and $v \equiv 2 \pmod{12}$.)

(2) In the context of leave graphs of Steiner systems, several necessary conditions have been established, and their proof techniques can be applied to prove some of the results given above. In this sense, part of this work may be viewed as inherent in some previous papers, e.g. in [4, 1, 2]. We have to note, however, that *not* every coverable graph is a leave graph (consider $2K_5$, for example) and at several points, in order to keep our more general problem under control, we had to introduce new ideas different from the ones in the papers cited.

(3) It seems to be an important advantage of our approach that it can handle the relatively small structures without the use of a computer. Therefore we can expect that some statements which have been verified so far only by computer search (see e.g. Lemmas 2.15 and 2.16 in [2]) will have fairly simple mathematical proofs.

(4) An open problem, perhaps solvable by a refinement of our method, is to find all irreducible, coverable graphs G with excess $\varepsilon(G) = 10$, or to prove that K_5 is the unique such graph. Probably any answer to this problem (affirmative or negative) would have interesting consequences concerning repeated blocks in triple systems of index λ. (It can be proved that there is a unique graph G, shown in Fig. 6, such that $\varepsilon(G) = 10$ and $G^* = K_5$, with $d(x)$ odd for all $x \in V(G)$. In this way, if K_5 were unique, then the spectrum of blocks of multiplicity λ in $TS(v, \lambda)$ would completely be characterized. On the other hand, if there were some graphs $G^* \neq K_5$ with an extension G such that all degrees in G are odd, $\varepsilon(G) = 10$, and $|V(G)| \equiv 2 \pmod{12}$, then some attempts could be done to find a construction that fills the gap in the spectrum problem.)

(5) Because of the generality of F-coverable graphs, one can expect further applications of them in design theory as well as in other branches of combinatorics. We also note that the concept of coverable structures can be extended to other classes such as e.g. directed graphs, finite set systems (hypergraphs), sequences over a given alphabet, matrices over the set of integers, etc., offering wide areas of challenging open problems.

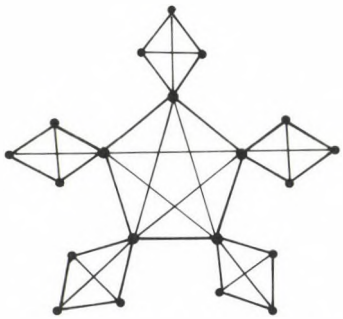


Fig. 6

REFERENCES

[1] COLBOURN, C. J. and MAHMOODIAN, E. S., The spectrum of support sizes for threefold triple systems, *Discrete Math.* **83** (1990), 9–19. *MR* **91f**:05016

[2] COLBOURN, C. J. and MAHMOODIAN, E. S., Support sizes of sixfold triple systems, *Discrete Math.* (to appear).

[3] MILICI, S. and TUZA, Zs., The spectrum of λ-times repeated blocks for $TS(v, \lambda)$, *Discrete Math.* **129** (1994), 159–166.

[4] ROSA, A. and HOFFMAN, D., The number of repeated blocks in twofold triple systems, *J. Combin. Theory Ser. A* **41** (1986), 61–88. *MR* **87g**:05062

(Received September 16, 1992; in revised form December 28, 1992)

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI CATANIA
VIALE A. DORIA 6
I–95125 CATANIA
ITALY

MTA SZÁMÍTÁSTECHNIKAI ÉS
AUTOMATIZÁLÁSI KUTATÓINTÉZETE
KENDE U. 13–17
H–1111 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 345–353

SOME DISTRIBUTION RESULTS ON RANK ORDER STATISTICS

J. SARAN and S. RANI

Abstract

This paper deals with the derivation of the null joint and marginal probability distributions of some rank order statistics by using Dwass technique. The rank order statistics considered include the number of positive reflections, the index of the i^{th} positive reflection and the interval between the i^{th} and the l^{th} positive reflections.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ denote random samples drawn from populations with unknown continuous distribution functions $F(x)$ and $G(x)$, respectively. Let $F_n(x)$ and $G_n(x)$ be the corresponding empirical distribution functions. Denote by $Z_1 < Z_2 < \ldots < Z_{2n}$ the ordered combined sample and let $Z_0 = -\infty$. On replacing each X_k in this ordered set by $+1$ and each Y_k by -1, there results a sequence of rank order indicators. A random variable defined as a function of the X_k and the Y_k only through these indicators is called a rank order statistic. Such statistics are often expressed in terms of

$$H_n(u) = n[F_n(u) - G_n(u)], \quad -\infty < u < \infty.$$

Dwass [4] developed a new technique (other than the combinatorial one) based on simple random walk with independent steps, in order to determine the distributions of some rank order statistics for the case of equal sample size. By using the Dwass technique Aneja and Sen [2], [3], Aneja [1], Mahendra Pratap [7] and Kaul [6] have derived the joint and marginal distributions of various rank order statistics. In this paper we derive the null joint and marginal distributions of two-sample rank order statistics viz., the number of positive reflections, the index of the i^{th} positive reflection and the length of the interval between the i^{th} and the l^{th} positive reflections by using Dwass technique.

1991 *Mathematics Subject Classification.* Primary 62G30.

Key words and phrases. Dwass technique, simple random walk, rank order statistics — positive reflection, the index of the i^{th} positive reflection, the length of the interval between the i^{th} and the l^{th} positive reflections, probability generating function.

2. The method

In deriving these results we use the Dwass technique which is based on the simple random walk

$$\left\{ S_j : S_j = \sum_{i=1}^{j} W_i,\ S_0 = W_0 = 0 \right\}$$

generated by a sequence $\{W_i\}$ of independent random variables with common probability distribution

$$P(W_i = +1) = p, \quad P(W_i = -1) = q; \quad q = 1 - p, \quad 1 \leqq i < \infty.$$

The assumption that $p < 1/2$ implies that the random walk $\{S_j\}$ is transient so that with probability one, $S_j = 0$ for only finitely many values of j. Let T be the largest value of j for which $S_j = H_n(Z_j) = 0$ and let U be a function defined on the random walk. Then U is said to satisfy assumption A when its value is completely determined by $W_1, W_2, \ldots, W_T$. The main theorem used for finding the distributions of rank order statistics is quoted from Dwass [4].

THEOREM 1. *Suppose U_n is a rank order statistic for every n and U is the related function satisfying assumption A. Define*

$$E(U) = h(p), \quad 0 \leqq p < 1/2.$$

Then the following power series in powers of pq is valid for $0 \leqq p < 1/2$:

$$h(p)/(1-2p) = \sum_{n=0}^{\infty} E(U_n) \binom{2n}{n} (pq)^n.$$

If ϕ is a function defined over the possible values of U then $\phi(U_n)$ is also a rank order statistic. In particular if ϕ is the set indicator function of B then $E(\phi(U_n)) = P(U_n \text{ in } B)$. While applying the theorem we shall let the symbols U, U_n represent $\phi(U)$, $\phi(U_n)$ for the various versions of ϕ that may be convenient to the problem at hand. This implies that the coefficient of $(pq)^n$ in the power series expansion of $P(U = k)/(1-2p)$ equals $\binom{2n}{n} P(U_n = k)$.

3. Definitions of rank order statistics

The following is the list of rank order statistics whose distributions will be derived. In what follows, we shall use the dual notation U, U_n for these rank order statistics as suggested in Theorem 1.

I. *Return to the origin.* A 'return' to the origin occurs at an index j for which $S_j = 0$.

II. *Positive and negative sojourns.* A 'sojourn' is defined as the segment between two consecutive returns to the origin. The segment between the origin and the first return point is also regarded as a sojourn. Let $0 < i_1 < i_2 < \ldots < 2n$ be the indices for which $H_n(Z_i) = 0$. If $H_n(Z_i) > 0$ for $i_{j-1} < i < i_j$, we say that the j^{th} sojourn is positive and if $H_n(Z_i) < 0$ for $i_{j-1} < i < i_j$, we say that the j^{th} sojourn is negative.

III. *Positive and negative reflections of height a.* A reflection at height a occurs at an index j when $S_j = a$ and $S_{j-1} = S_{j+1} = a - 1 = S_j - 1$ or $S_{j-1} = S_{j+1} = a + 1 = S_j + 1$, the reflection being positive or negative according as $S_{j-1} = S_{j+1} = a + 1$ or $S_{j-1} = S_{j+1} = a - 1$. Let $R_n(a)$ denote the total number of reflections of height a of which $R_n^+(a)$ are positive and $R_n^-(a)$ are negative with

$$R_n(a) = R_n^+(a) + R_n^-(a).$$

IV. *The index of the i^{th} positive reflection of height a.* Let $R_n^{+i}(a)$ denote the index of the i^{th} positive reflection at height a. Then $R_n^{+i}(a) =$ the index j where $S_j = a$ and $S_{j-1} = S_{j+1} = a + 1$ for the i^{th} time, $1 \leqq i \leqq R_n^+(a)$.

V. *The length of the interval between the i^{th} and the l^{th} positive reflections of height a.* Let $M_n^{+(i,l)}(a)$ denote the length of the interval between the i^{th} and the l^{th} positive reflections of height a $(1 \leqq i < l \leqq R_n^+(a))$, then

$$M_n^{+(i,l)}(a) = R_n^{+l}(a) - R_n^{+i}(a).$$

The above mentioned statistics with respect to the origin (i.e. for $a = 0$) are denoted by the same symbols without parentheses for a, e.g., $R_n^+(0) \equiv R_n^+$, $R_n^{+i}(0) \equiv R_n^{+i}$, $M_n^{+(i,l)}(0) \equiv M_n^{+(i,l)}$, etc.

4. Some basic results

Some of the results we list below concerning simple random walk appear in Feller [5] and the rest are easily derived from elementary considerations. The following list covers what is needed in the sequel.

(i) The probability generating function (PGF) for the first return time to the origin is

$$f(t) = 1 - (1 - 4pqt^2)^{1/2},$$

from which the probability of ever returning to the origin is $f(1) = 2p$.

(ii) The PGF of the length of the first passage through k is $(f(t)/2qt)^k$.

(iii) If the PGF of the length of a positive sojourn is denoted by $F^+(t)$ and that of a negative sojourn by $F^-(t)$ then

$$F^+(t) = F^-(t) = f(t)/2.$$

(iv) The PGF of the path segment between the origin and the first positive reflection is given by

$$\sum_{i=0}^{\infty}[F^-(t)]^i \sum_{j=0}^{\infty}\left[F^+(t)\sum_{i=1}^{\infty}(F^-(t))^i\right]^j F^+(t) =$$
$$= F^+(t)/(1-F^-(t)-F^+(t)F^-(t)).$$

For the proof of this result, one may observe that in the requisite path there may be some negative sojourns in the beginning. After that there may be segments each beginning with a positive sojourn and ending with at least one negative sojourn. Lastly there is one positive sojourn.

(v) The PGF of the path segment between any two consecutive positive reflections is

$$\sum_{j=0}^{\infty}\left[F^+(t)\sum_{i=1}^{\infty}(F^-(t))^i\right]^j F^+(t) =$$
$$= F^+(t)(1-F^-(t))/(1-F^-(t)-F^+(t)F^-(t)).$$

(vi) The probability of the path segment between any two consecutive positive reflections is

$$\sum_{j=0}^{\infty}\left(p\sum_{i=1}^{\infty}p^i\right)^j p = pq/(q-p^2).$$

(vii) The probability of the path segment between the last positive reflection and the last return point to the origin is

$$p\sum_{j=0}^{\infty}\left(p\sum_{i=1}^{\infty}p^i\right)^j\sum_{i=0}^{\infty}p^i = p/(q-p^2).$$

(viii) The following power series expansions in powers of pq valid for positive integers i, j and k which follow immediately from Dwass [4], (14) and (16) are frequently used in the sequel:

(a) $$\alpha^i = \left\{\frac{1}{2}[1-(1-4pqt^2)^{1/2}]\right\}^i = \sum_{r=i}^{\infty}A_{r-i}(i,2)(pqt^2)^r$$

where

$$A_a(b,c) = \frac{b}{b+ac}\binom{b+ac}{a}.$$

(b) $$p^j/(1-2p) = \sum_{s=j}^{\infty} \binom{2s-j}{s-j} (pq)^s.$$

(c) $$p^k = \sum_{t=k}^{\infty} A_{t-k}(k,2)(pq)^t.$$

For ease in expression, while dealing with bivariate PGF's we will abbreviate $f(s)/2$ and $f(t)/2$ by α and β, respectively, where $f(\cdot) = 1 - [1 - 4pq(\cdot)^2]^{1/2}$.

5. Joint distribution of $R_n^+(a)$, $R_n^{+i}(a)$ and $M_n^{+(i,l)}(a)$, $a \geqq 0$

THEOREM 2. *The bivariate probability generating function of the joint distribution of $R^{+i}(a)$, the index of the i^{th} positive reflection of height a and $M^{+(i,l)}(a)$, the length of the interval between the i^{th} and the l^{th} positive reflections of height a $(1 \leqq i < l \leqq r)$ when $R^+(a)$ equals $r \geqq 0$ is given by, for $a \geqq 0$,*

$$h(p) = E\left(s^{R^{+i}(a)} t^{M^{+(i,l)}(a)}; R^+(a) = r\right) =$$

$$(1) \quad = \sum_{j=i}^{\infty} \sum_{u=l-i}^{\infty} P(R^{+i}(a) = a + 2j, M^{+(i,l)}(a) = 2u, R^+(a) = r) s^{a+2j} t^{2u} =$$

$$= \alpha^{a+i}(1-\alpha)^{i-1}(1-\alpha-\alpha^2)^{-i}\beta^{l-i}(1-\beta)^{l-i}(1-\beta-\beta^2)^{-(l-i)} \times$$

$$\times p^{r-l+1} q^{r-l-a}(q-p^2)^{-(r-l+1)} s^{-a}(1-2p).$$

PROOF. Let $OABCDEF$ be a random walk path envisaged in the theorem where A and F be the first and the last return points of height a and, B, C, D and E be the first, i^{th}, l^{th} and the r^{th} positive reflection points of height a, respectively (Fig. 1). Then the path comprises seven segments viz. OA, AB, BC, CD, DE, EF and a segment beyond F. Of these, the first segment OA is a first passage through a with its length having PGF $(\alpha/qs)^a$, by (ii) of Section 4. The segment AB has a PGF $\alpha(1-\alpha-\alpha^2)^{-1}$, by (iv). The segments BC and CD involve exactly $(i-1)$ and $(l-i)$ positive reflections with PGF's $(\alpha(1-\alpha)/(1-\alpha-\alpha^2))^{i-1}$ and $(\beta(1-\beta)/(1-\beta-\beta^2))^{l-i}$, respectively, by (v). The segment DE involves exactly $(r-l)$ positive reflections with probability $(pq/(q-p^2))^{r-l}$, by (vi). The segment EF has no positive reflection with probability $p/(q-p^2)$, by (vii). The last segment beyond F does not involve a return to the height a with probability $(1-2p)$.

Thus

$$E\left(s^{R^+(a)}t^{M^+(i,l)(a)};\, R^+(a)=r\right)=$$
$$=(\alpha/qs)^a\left(\frac{\alpha}{1-\alpha-\alpha^2}\right)\left(\frac{\alpha(1-\alpha)}{1-\alpha-\alpha^2}\right)^{i-1}\left(\frac{\beta(1-\beta)}{1-\beta-\beta^2}\right)^{l-i}\times$$
$$\times\left(\frac{pq}{q-p^2}\right)^{r-l}\left(\frac{p}{q-p^2}\right)(1-2p)$$

which leads to (1).

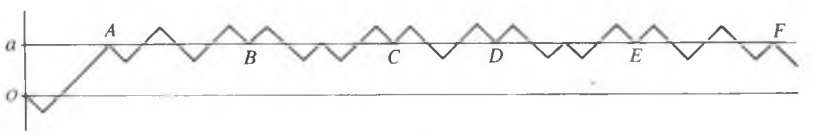


Fig. 1

Deductions. (i) Putting $t=1$ and $s=1$ in (1), we get, respectively,

$$E\left(s^{R^+(a)};\, R^+(a)=r\right)=$$
(2)
$$=\alpha^{i+a}(1-\alpha)^{i-1}(1-\alpha-\alpha^2)^{-i}p^{r-i+1}q^{r-i-a}(q-p^2)^{-(r-i+a)}s^{-a}(1-2p)$$

and

$$E\left(t^{M^+(i,l)(a)};\, R^+(a)=r\right)=$$
(3)
$$=\beta^{l-i}(1-\beta)^{l-i}(1-\beta-\beta^2)^{-(l-i)}p^{r-l+i+a+1}q^{r-l+i-a-1}(q-p^2)^{-(r-l+i+1)}(1-2p).$$

(ii) Putting $s=t=1$ in (1), we get

$$P(R^+(a)=r)=p^{r+a+1}q^{r-a-1}(q-p^2)^{-(r+1)}(1-2p) \tag{4}$$

(equivalent to [1], ch. II (93d)).

(iii) Summation of (1) over r from l to ∞ gives

$$E\left(s^{R^+(a)}t^{M^+(i,l)(a)}\right)=$$
(5)
$$=\alpha^{i+a}(1-\alpha)^{i-1}(1-\alpha-\alpha^2)^{-i}\beta^{l-i}(1-\beta)^{l-i}(1-\beta-\beta^2)^{-(l-i)}p(1/qs)^a.$$

(iv) Summing (2) over r from i to ∞ and (3) over r from l to ∞, we get, respectively,

$$E\left(s^{R^+(a)}\right)=\alpha^{i+a}(1-\alpha)^{i-1}(1-\alpha-\alpha^2)^{-i}p(qs)^{-a} \tag{6}$$

and

$$E\left(t^{M^{+(i,l)}(a)}\right) = \tag{7}$$
$$= \beta^{l-i}(1-\beta)^{l-i}(1-\beta-\beta^2)^{-(l-i)}p^{i+a+1}q^{i-a-1}(q-p^2)^{-i}.$$

Probability distributions. The following probability distributions corresponding to the PGF's (1) to (7) can be derived with the help of Theorem 1 and the power series expansions ((viii), Section 4).

$$\binom{2n}{n} P\left(R_n^{+i}(a) = a+2j, M_n^{+(i,l)}(a) = 2u, R_n^{+}(a) = r\right) =$$
$$= \sum_{k=0}^{i-1}\sum_{h=0}^{\infty}\sum_{m=0}^{l-i}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{f=0}^{h}\sum_{g=0}^{b}(-1)^{k+m+c+h+b}\times \tag{8}$$
$$\times\binom{i-1}{k}\binom{-i}{h}\binom{l-i}{m}\binom{-(l-i)}{b}\binom{-(r-l-1)}{c}\binom{h}{f}\binom{b}{g}\times$$
$$\times A_{\psi_1}(j-\psi_1+a,2)A_{\psi_2}(u-\psi_2,2)A_{\psi_3}(r-l+2+3c+a,2)$$

where

$$\psi_1 = j-i-k-h-f,$$
$$\psi_2 = u-l+i-m-b-g \quad \text{and} \quad \psi_3 = n-2c-j-u-r-a+l-1.$$

$$\binom{2n}{n} P\left(R_n^{+i}(a) = a+2j, R_n^{+}(a) = r\right) =$$
$$= \sum_{k=0}^{i-1}\sum_{h=0}^{\infty}\sum_{c=0}^{\infty}\sum_{f=0}^{h}(-1)^{k+h+c}\binom{i-1}{k}\binom{-i}{h}\binom{-(r-i+1)}{c}\binom{h}{f}\times \tag{9}$$
$$\times A_{\psi_1}(j-\psi_1+a,2)A_{\psi_2}(r-i+3c+a+2,2),$$

where

$$\psi_1 = j-i-k-h-f \quad \text{and} \quad \psi_2 = n-2c-r-j-a+i-1.$$

$$\binom{2n}{n} P\left(M_n^{+(i,l)}(a) = 2u, R_n^{+}(a) = r\right) = \sum_{m=0}^{l-i}\sum_{b=0}^{\infty}\sum_{g=0}^{b}\sum_{c=0}^{g}(-1)^{m+b+c}\times \tag{10}$$
$$\times\binom{l-i}{m}\binom{-(l-i)}{m}\binom{-(r-l+i+1)}{c}\binom{b}{g}A_{\psi_2}(j-\psi_2,2)\times$$
$$\times A_{\psi_3}(n+c+a-u+2-\psi_3,2),$$

where

$$\psi_2 = u - l + i - m - b - g \quad \text{and} \quad \psi_3 = n - 2c - a - u - r + l - i - 1.$$

(11)
$$\binom{2n}{n} P\left(R_n^+(a) = r\right) = \sum_{c=0}^{\infty} (-1)^c \binom{-(r+1)}{c} A_{n-a-r-2c-1}(r + 2a + 3c + 3, 2)$$

(equivalent to [1], ch. II (101)).

(12)
$$\begin{aligned} \binom{2n}{n} P\left(R_n^{+i}(a) = a + 2j, M_n^{+(i,l)}(a) = 2u\right) = \\ = \sum_{k=0}^{i-1} \sum_{h=0}^{\infty} \sum_{m=0}^{l-i} \sum_{b=0}^{\infty} \sum_{f=0}^{h} \sum_{g=0}^{b} (-1)^{k+m+h+b} \binom{i-1}{k} \binom{-i}{h} \times \\ \times \binom{l-i}{m} \binom{-(l-i)}{b} \binom{h}{f} \binom{b}{g} A_{\psi_1}(j - \psi_1 + a, 2) \times \\ \times A_{\psi_2}(u - \psi_2, 2) \binom{2n - 2j - 2u - 2a - 1}{n - j - u - a - 1}, \end{aligned}$$

where

$$\psi_1 = j - i - k - h - f \quad \text{and} \quad \psi_2 = u - l + i - m - b - g.$$

(13)
$$\begin{aligned} \binom{2n}{n} P\left(R_n^{+i}(a) = a + 2j\right) = \sum_{k=0}^{i-1} \sum_{h=0}^{\infty} \sum_{f=0}^{h} (-1)^{h+k} \binom{i-1}{k} \binom{-i}{h} \times \\ \times \binom{h}{f} A_{\psi_1}(j - \psi_1 + a, 2) \binom{2n - 2j - 2a - 1}{n - j - a - 1}, \end{aligned}$$

where $\psi_1 = j - i - k - h - f$.

(14)
$$\begin{aligned} \binom{2n}{n} P\left(M_n^{+(i,l)}(a) = 2u\right) = \sum_{m=0}^{l-i} \sum_{b=0}^{\infty} \sum_{g=0}^{b} \sum_{c=0}^{\infty} (-1)^{m+b+c} \times \\ \times \binom{l-i}{m} \binom{-(l-i)}{b} \binom{b}{g} \binom{-i}{c} A_{\psi_2}(j - \psi_2, 2) \binom{2n - 2u - c - i}{n - u + c + a + 1}, \end{aligned}$$

where $\psi_2 = u - l + i - m - b - g$.

REFERENCES

[1] ANEJA, K. G., Random walk and rank order statistics, Ph. D. Thesis, University of Delhi, Delhi, 1975.

[2] ANEJA, K. G. and SEN, K. Random walk and distributions of rank order statistics, *SIAM J. Appl. Math.* **23** (1972), 276–287. *MR* **47** #9770

[3] ANEJA, K. G. and SEN, K., Maxima in random walk and related rank order statistics, *Studia Sci. Math. Hungar.* **7** (1972), 425–428. *MR* **48** #7384

[4] DWASS, M., Simple random walk and rank order statistics, *Ann. Math. Statist.* **38** (1967), 1042–1053. *MR* **35** #6304

[5] FELLER, W., *An introduction to probability theory and its applications*, Vol. I, 3rd edition, Wiley, New York, 1968. *MR* **37**#3604

[6] KAUL, C. L., Contributions in rank order statistics, Ph. D. Thesis, University of Delhi, Delhi, 1982.

[7] PRATAP, M., Contribution to random walk, Ph. D. Thesis, Garhwal University, Srinagar (Garhwal), India, 1982.

(Received September 20, 1991)

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELHI
IND–110 007 DELHI
INDIA

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 355–373

STRAIGHT LEFT ORDERS

VICTORIA GOULD

1. Introduction

A subsemigroup S of a semigroup Q is a *left (right) order* in Q if every element of Q can be written as $a^{\#}b$ ($ba^{\#}$) where $a, b \in S$ and $a^{\#}$ is a group inverse of a and if, in addition, every square-cancellable element of S lies in a subgroup of Q. For the definitions of group inverse and square-cancellable element we refer the reader to Section 2, where further details of the terms used in this introduction may be found. If S is a left (right) order in Q then Q is a *semigroup of left (right) quotients* of S; if S is both a left order and a right order in Q then S is an *order* in Q and Q is a *semigroup of quotients* of S. These notions were introduced in [FP], although only in the context of orders in completely 0-simple semigroups.

Our approach to orders in semigroups is obviously inspired by the concept of a classical order in ring theory. Ore's theorem tells us that a ring R is a (classical) left order in some ring Q if and only if R contains a non-zero divisor and satisfies the left Ore condition, that is, given any $a, b \in R$, where a is a non-zero divisor, there exist elements $c, d \in R$ where c is a non-zero divisor such that $cb = da$. Orders in various special classes of rings have also been described; perhaps the best known example of a result of this sort is Goldie's celebrated theorem characterising left orders in simple artinian rings.

Fountain and Petrich give a description in [FP] of orders in completely 0-simple semigroups. A number of subsequent papers have characterised left orders in other well known classes of semigroups. But what of an analogue of Ore's theorem? What are the semigroups that can occur as left orders in any semigroup? This question is much more complex than the corresponding one for rings, essentially because we consider group inverses in *any* subgroup of a semigroup of left quotients, whereas for rings one concentrates on the group of units. A first step in the direction of an answer is made in [G4]. In that paper we characterise left orders in semigroups in the class of regular

1991 *Mathematics Subject Classification.* Primary 20M10; Secondary 20M17.

Key words and phrases. Square-cancellable, group inverse, semigroup of left quotients, straight left order.

0081–6906/95/$ 4.00

$\mathcal{H}$-semigroups, where a semigroup Q is an *$\mathcal{H}$-semigroup* if Green's relation $\mathcal{H}$ is a congruence on Q. A sequel, [G5], shows how this result yields as corollaries the previous characterisations of (left) orders in semigroups in particular classes, for all those classes considered had been classes of regular $\mathcal{H}$-semigroups.

If S is a left order in a regular $\mathcal{H}$-semigroup Q, then as shown in [G1], any element of Q may be written as $a^{\#}b$ where $a, b \in S$ and $a\mathcal{R}b$ in Q, a property which facilitates greatly consideration of products of elements in Q. In view of this we say that a left order S in a semigroup Q is *straight* if every element of Q can be written as $a^{\#}b$ where $a, b \in S$ and $a\mathcal{R}b$ in Q. *Straight right orders* and *straight orders* are defined in the obvious way. The aim of this paper is to characterise those semigroups that are straight left orders.

In Section 2 we give a number of definitions and preliminary results. Section 3 introduces and investigates the notion of a **-pair* $\mathcal{P} = (\leqq_l, \leqq_r)$ of preorders on a semigroup S. If S is a subsemigroup of a semigroup Q then defining $\leqq_l$ ($\leqq_r$) on S by $a \leqq_l b$ ($a \leqq_r b$) if and only if $Q^1a \subseteq Q^1b$ ($aQ^1 \subseteq bQ^1$), then $\mathcal{P} = (\leqq_l, \leqq_r) = \mathcal{P}(Q)$ is always a *-pair, the **-pair induced by* Q. Additionally, if S is a straight left order in Q then $\mathcal{P}(Q)$ satisfies a number of properties: any *-pair with these properties we call an *embeddable *-pair*. Using the concept of an embeddable *-pair we describe in Theorem 4.1 of Section 4 straight left orders.

Given a straight left order S in Q, if $\mathcal{L}^Q \cap (S \times S) = \mathcal{L}^{*S}$ and $R^Q \cap \cap (S \times S) = \mathcal{R}^{*S}$ then S is *stratified* in Q; stratified left orders have been investigated in a number of papers. If $\mathcal{P}(Q) = (\leqq_{\mathcal{L}^*}, \leqq_{\mathcal{R}^*})$, then we shall say that S is *fully stratified* in Q. It is immediate from the definitions that a fully stratified left order is stratified, however, the converse is not always true. In Section 5 we apply Theorem 4.1 to characterise fully stratified left orders.

The final section shows how the description of left orders in regular $\mathcal{H}$-semigroups given in [G4] may also be deduced from Theorem 4.1.

2. Preliminaries

We assume a familiarity with the basic notions of semigroup theory, in particular, Green's relations. As far as possible we follow standard notation and terminology, as can be found in [H].

The relation $\leqq_{\mathcal{L}}$ is defined on a semigroup S by the rule that for any elements a, b of S, $a \leqq_{\mathcal{L}} b$ if and only if $S^1a \subseteq S^1b$. Clearly $\leqq_{\mathcal{L}}$ is a preorder that is right compatible with multiplication and whose associated equivalence relation is $\mathcal{L}$. The dual relation to $\leqq_{\mathcal{L}}$ is denoted by $\leqq_{\mathcal{R}}$.

The relation $\leqq_{\mathcal{L}^*}$ is defined on a semigroup S by the rule that for any elements a, b of S, $a \leqq_{\mathcal{L}^*} b$ if and only if

$$bx = by \text{ implies that } ax = ay$$

for all $x, y \in S^1$. Then $\leqq_{\mathcal{L}^*}$ is a preorder that is right compatible with multiplication, as is the associated equivalence relation $\mathcal{L}^*$. The preorder $\leqq_{\mathcal{R}^*}$ and the equivalence relation $\mathcal{R}^*$ are defined dually. The relations $\mathcal{L}^*$ and $\mathcal{R}^*$ have the following alternative description.

LEMMA 2.1. [F] *Let S be a semigroup and let $a, b \in S$. Then $a\mathcal{L}^*b$ ($a\mathcal{R}^*b$) if and only if $a\mathcal{L}b$ ($a\mathcal{R}b$) in some oversemigroup of S.*

The intersection of $\mathcal{L}^*$ and $\mathcal{R}^*$ on any semigroup S is denoted by $\mathcal{H}^*$. It is easily seen that $\leqq_{\mathcal{L}} \subseteq \leqq_{\mathcal{L}^*}$ and $\leqq_{\mathcal{R}} \subseteq \leqq_{\mathcal{R}^*}$, so that $\mathcal{L} \subseteq \mathcal{L}^*$, $\mathcal{R} \subseteq \mathcal{R}^*$ and $\mathcal{H} \subseteq \mathcal{H}^*$. If S is a regular semigroup then $\leqq_{\mathcal{L}} = \leqq_{\mathcal{L}^*}$ and $\leqq_{\mathcal{R}} = \leqq_{\mathcal{R}^*}$ so that $\mathcal{L} = \mathcal{L}^*$, $\mathcal{R} = \mathcal{R}^*$ and $\mathcal{H} = \mathcal{H}^*$.

An element a of a semigroup S is *square-cancellable* if $a\mathcal{H}^*a^2$. Thus a is square-cancellable if and only if for all $x, y \in S^1$,

$$a^2x = a^2y \text{ implies that } ax = ay$$

and

$$xa^2 = ya^2 \text{ implies that } xa = ya.$$

Note that if S is a subsemigroup of Q and $a \in S$ lies in a subgroup of Q, then a is square-cancellable in S. By definition of (left, right) order, all such elements must lie in subgroups of any semigroup of (left, right) quotients. The set of square-cancellable elements of a semigroup S is denoted by $\mathcal{S}(S)$.

Let a be an element of a semigroup S. If a lies in a subgroup of S, then H_a is a subgroup, the maximum subgroup containing a. Thus if $a^\#$ is the inverse of a, in the sense of group theory, in a subgroup of S, then $a^\#$ is the inverse of a in H_a and so is well-defined. Given a semigroup S and $a \in S$, by writing $a^\#$ it will be implicit that a lies in a subgroup of S.

The notion of a left order in a group is much older than that of a left order in an arbitrary semigroup. We recall here that a semigroup is *right (left) reversible* if $Sa \cap Sb \neq \emptyset$ ($aS \cap bS \neq \emptyset$) for any $a, b \in S$. A theorem of Ore and Dubreil (Theorem 1.24 of [CP]) states that a semigroup S is a left order in a group G if and only if S is right reversible and cancellative.

This paper is concerned with a special sort of left orders: *straight* left orders. We summarize here some of their properties.

PROPOSITION 2.2. [G1] *Let S be a straight left order in a semigroup Q.*

(i) *If $a \in \mathcal{S}(S)$, then $H_a^Q \cap S$ is a left order in H_a^Q.*

(ii) *Every $\mathcal{H}$-class of Q contains an element of S; if $q \in Q$ and $q = a^\#b$ where $a \in \mathcal{S}(S)$, $b \in S$ and $a\mathcal{R}b$ in Q, then $q\mathcal{H}b$ in Q.*

Considering only left orders that are straight imposes regularity on the semigroups of left quotients.

LEMMA 2.3. *The following conditions are equivalent for a semigroup Q:*

(i) *Q is regular;*

(ii) *Q is a straight left order in Q;*

(iii) *there is a straight left order in* Q.

PROOF. That (i) implies (ii) is easy and that (ii) implies (iii) is immediate. Suppose that (iii) holds and S is a straight left order in Q. Then if $q \in Q$, $q = a^{\#}b$ for some $a \in \mathcal{S}(S)$, $b \in S$ with $a\mathcal{R}b$ in Q. By Proposition 2.2, $q\mathcal{H}b$ in Q so that $qRaHaa^{\#}$ in Q. Thus every $\mathcal{R}$-class of Q contains an idempotent and so Q is regular.

Finally in this section we make a remark on notation. If $(X)(l)$ denotes a condition having an obvious left-right dual, then $(X)(r)$ will denote the dual condition. In this case, '(X)' is shorthand for '$(X)(l)$ and $(X)(r)$'.

3. Embeddable *-pairs

An ordered pair $\mathcal{P} = (\leqq_l(\mathcal{P}), \leqq_r(\mathcal{P}))$ of preorders on a semigroup S is a **-pair* if $\leqq_l(P)$ is right compatible with multiplication, $\leqq_r(\mathcal{P})$ is left compatible with multiplication, $\leqq_l(\mathcal{P}) \subseteq \leqq_{\mathcal{L}^*}$ and $\leqq_r(\mathcal{P}) \subseteq \leqq_{\mathcal{R}^*}$. Clearly $(\leqq_{\mathcal{L}^*}, \leqq_{\mathcal{R}^*})$ is a *-pair for any semigroup. Given a *-pair $\mathcal{P} = (\leqq_l(\mathcal{P}), \leqq_r(\mathcal{P}))$ for a semigroup S, then we denote by $\mathcal{L}'(\mathcal{P})$ and $\mathcal{R}'(\mathcal{P})$ the equivalence relations associated with $\leqq_l(\mathcal{P})$ and $\leqq_r(\mathcal{P})$, respectively. We remark that $\mathcal{L}'(\mathcal{P})$ is a right congruence and $\mathcal{R}'(\mathcal{P})$ is a left congruence. Where there is no danger of ambiguity $\leqq_l(\mathcal{P})$, $\leqq_r(\mathcal{P})$, $\mathcal{L}'(\mathcal{P})$ and $\mathcal{R}'(\mathcal{P})$ are written more simply as $\leqq_l$, $\leqq_r$, $\mathcal{L}'$ and $\mathcal{R}'$. The notation for the equivalence relation $\mathcal{L}'(\mathcal{P}) \cap \mathcal{R}'(\mathcal{P})$ is $\mathcal{H}'(\mathcal{P})$ or $\mathcal{H}'$ and the $\mathcal{H}'(\mathcal{P})$-class of an element a is $H'_a(\mathcal{P})$ or H'_a. If $a\mathcal{H}'(\mathcal{P})a^2$ then a is $\mathcal{P}$-*good* or simply *good*.

If S is a subsemigroup of a semigroup Q then

$$\mathcal{P}(Q) = (\leqq_{\mathcal{L}^Q} \cap (S \times S), \leqq_{\mathcal{R}^Q} \cap (S \times S))$$

is clearly a *-pair for S: it is called the **-pair for S induced by Q*.

We wish to consider the *-pair for a semigroup S induced by a semigroup Q in which S is a straight left order. To describe such *-pairs it will be convenient to use the following notions.

Let $\mathcal{P} = (\leqq_l, \leqq_r)$ be a *-pair for a semigroup S. For a good element a of S, put

$$L(a, \mathcal{P}) = L(a) = \{b \in S : b \leqq_l a\}$$

and

$$\overline{L}(a, \mathcal{P}) = \overline{L}(a) = \{[b]_{\mathcal{L}'} : b \in L(a)\}$$

where $[b]_{\mathcal{L}'}$ is the $\mathcal{L}'$-equivalence class of b. Note that $L(a)$ is a union of $\mathcal{L}'$-classes and if $b \in L(a)$ then $ba \leqq_l a^2\mathcal{H}'a$ so that $ba \in L(a)$. Thus one may define a map $\rho_a(\mathcal{P}) = \rho_a : L(a) \to L(a)$ by $b\rho_a = ba$. Further, as $\mathcal{L}'$ is a right congruence, $\overline{\rho}_a(\mathcal{P}) = \overline{\rho}_a : \overline{L}(a) \to \overline{L}(a)$ given by $([b]_{\mathcal{L}'})\overline{\rho}_a = [ba]_{\mathcal{L}'}$ is a map. One defines in a dual manner $R(a, \mathcal{P}) = R(a)$, $\overline{R}(a, \mathcal{P}) = \overline{R}(a)$ and maps $\lambda_a(\mathcal{P}) = \lambda_a : R(a) \to R(a)$, $\overline{\lambda}_a(\mathcal{P}) = \overline{\lambda}_a : \overline{R}(a) \to \overline{R}(a)$.

The main theorem of this paper gives necessary and sufficient conditions on a *-pair $\mathcal{P}$ for a semigroup S such that S is a straight left order in a semigroup Q where $\mathcal{P} = \mathcal{P}(Q)$. Given such a *-pair $\mathcal{P}$, we know that Q is uniquely determined up to isomorphism.

PROPOSITION 3.1. *Let S be a straight left order in semigroups Q and Q', where $\mathcal{P}(Q) = \mathcal{P}(Q')$. Then Q is isomorphic to Q' under an isomorphism which restricts to the identity map on S.*

PROOF. This is an immediate consequence of Theorem 3.1 of [G2].

A subsemigroup S of a semigroup Q is *very large in Q* if S has non-empty intersection with every $\mathcal{H}$-class of Q. From Proposition 2.2 we know that if S is a straight left order in Q then S must be very large in Q.

PROPOSITION 3.2. *Let S be a very large subsemigroup of a regular semigroup Q and let $\mathcal{P}(Q) = \mathcal{P} = (\leqq_l, \leqq_r)$. Then S satisfies the following conditions with respect to $\mathcal{P}$.*

(Ei) $\mathcal{L}' \circ \mathcal{R}' = \mathcal{R}' \circ \mathcal{L}'$.

(Eii)(l) *For all $b, c \in S$, $b \leqq_l c$ if and only if $b\mathcal{L}'dc$ for some $d \in S$;*

(Eii)(r)

(Eiii) *Every $\mathcal{L}'$-class and every $\mathcal{R}'$-class contains a good element.*

(Eiv) *For all good elements a, ρ_a is one-one and preserves $\mathcal{R}'$-classes, $\overline{\rho}_a$ is one-one, λ_a is one-one and preserves $\mathcal{L}'$-classes and $\overline{\lambda}_a$ is one-one.*

PROOF. (Ei)(Eiii) These follow directly from the fact that S is very large in Q and $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$ in Q.

(Eii)(l) If $b\mathcal{L}'dc$ then $Qb = Qdc \subseteqq Qc$ so that $b \leqq_l c$. Conversely, if $b \leqq_l c$ then $b = qc$ for some $q \in Q$. Now $q\mathcal{H}h$ for some $h \in S$ so that $b = qc\mathcal{L}hc$ in Q and $b\mathcal{L}'hc$ in S.

(Eiv) Given $b, c \in L(a)$, we know $Qb \subseteqq Qa$ and $Qc \subseteqq Qa$. Since a is good, $a^\# a = aa^\# \mathcal{H} a$ for some $a^\# \in Q$. Certainly $bQ = baa^\# Q \subseteqq baQ \subseteqq bQ$ so that $b\mathcal{R}'ba = b\rho_a$. If $b\rho_a = c\rho_a$ then $baa^\# = caa^\#$ from which we have that $b = c$. The proof that $\overline{\rho}_a$ is one-one is similar.

Dual arguments now complete the proof.

A *-pair $\mathcal{P}$ for a semigroup S is an *embeddable *-pair* if $\mathcal{P}$ satisfies conditions (Ei), (Eii), (Eiii) and (Eiv). We give below an equivalent formulation of these conditions and deduce a number of further properties held by embeddable *-pairs. It is convenient at this point to list the conditions under consideration.

Let $\mathcal{P}$ be a *-pair for a semigroup S.

(Ev)(l) For all $a, b \in S$ where a is good, if $b \leqq_l a$ then $ba\mathcal{R}'b$.

(Evi)(l) For all $a, b, c \in S$ where a is good, if $b, c \leqq_l a$ and $ba = ca$, then $b = c$.

(Evii)(l) For all $a, b, c \in S$ where a is good, if $b, c \leqq_l a$ and $ba\mathcal{L}'ca$, then $b\mathcal{L}'c$.

(Fi)(l) For all $a, b, c \in S$ where a is good, if $b, c \leqq_l a$ and $ba \leqq_l ca$, then $b \leqq_l c$.

(Fii)(l) For all $a, b \in S$ where a is good, if $a\mathcal{L}'b$ then $ba\mathcal{H}'b$.

(Fiii) If $b \in S$ then there exists an $s \in S$ with

$$b\mathcal{R}'bs\mathcal{L}'s\mathcal{R}'sb\mathcal{L}'b;$$

further, given such an s, bs and sb are good, $bsb\mathcal{H}'b$ and $sbs\mathcal{H}'s$.

(Fiv) If a is good, then H'_a is a subsemigroup all of whose elements are good.

(Fv) (l) For all $a, b \in S$ where a is good, if $a \leqq_l b$ then $ba\mathcal{L}'a$.

LEMMA 3.3. *Let $\mathcal{P}$ be a *-pair for a semigroup S. Then $\mathcal{P}$ is an embeddable *-pair if and only if $\mathcal{P}$ satisfies conditions* (Ei), (Eii), (Eiii), (Ev), (Evi) *and* (Evii).

PROOF. This is just a matter of rewording.

LEMMA 3.4. *Let $\mathcal{P} = (\leqq_l, \leqq_r)$ be an embeddable *-pair for a semigroup S. Then $\mathcal{P}$ satisfies conditions* (Fi), ..., (Fv).

PROOF. (Fi)(l) If $b, c \leqq_l a$ where a is good and $ba \leqq_l ca$, then we know from (Eii)(l) that $ba\mathcal{L}'dca$ for some $d \in S$ and further, $dc \leqq_l c \leqq_l a$. Thus by (Evii)(l), $b\mathcal{L}'dc$ and it follows that $b \leqq_l c$.

(Fii)(l) If $a\mathcal{L}'b$ where a is good then certainly by (Ev)(l) we have $ba\mathcal{R}'b$. In addition, $ba\mathcal{L}'a^2\mathcal{H}'a\mathcal{L}'b$ so that $ba\mathcal{L}'b$.

(Fiii) If $b \in S$ then using condition (Eiii) there are good elements c, d in S with $c\mathcal{R}'b\mathcal{L}'d$. Then from (Ei) we may choose $s \in S$ with $c\mathcal{L}'s\mathcal{R}'d$. We claim that s is the element required. For using (Fii), $b\mathcal{H}'bd\mathcal{R}'bs$ and so by symmetry we obtain

$$b\mathcal{R}'bs\mathcal{L}'s\mathcal{R}'sb\mathcal{L}'b.$$

It is then easy to see that bs and sb are good. The last part of the condition is immediate from (Fii).

(Fiv) If a is good and $b, c \in H'_a$ then $bc\mathcal{R}'ba\mathcal{H}'b\mathcal{H}'a$, using (Fii). Dually, $bc\mathcal{L}'a$ so that $bc \in H'_a$ and H'_a is a subsemigroup, which clearly can consist only of good elements.

(Fv)(l) If $a, b \in S$ where a is good and $a \leqq_l b$ then $a\mathcal{H}'a^2 \leqq_l ba \leqq_l a$ so that $a\mathcal{L}'ba$.

4. Straight left orders

Given a straight left order S in Q, then by Proposition 2.2 and Lemma 2.3, Q is regular and S is very large in Q. Hence Proposition 3.2 gives that the *-pair $\mathcal{P}(Q)$ is an embeddable *-pair. In this section we give necessary and sufficient conditions on an embeddable *-pair $\mathcal{P}$ for a semigroup S such that S is a straight left order in a semigroup Q where $\mathcal{P} = \mathcal{P}(Q)$.

THEOREM 4.1. *Let S be a semigroup having an embeddable *-pair $\mathcal{P} = (\leqq_l, \leqq_r)$. Then S is a straight left order in a semigroup Q such that $\mathcal{P} = \mathcal{P}(Q)$ if and only if S satisfies the following conditions with respect to $\mathcal{P}$.*

(Gi) $\mathcal{S}(S) = \{a \in S : a \text{ is good}\}$.

(Gii) *If $a \in \mathcal{S}(S)$ then H'_a is right reversible.*

(Giii) *If $b, c \in S$ then $b \leqq_l c$ if and only if there exist $h \in \mathcal{S}(S)$ and $k \in S$ with $h\mathcal{R}'k$, $b \leqq_r h$ and $hb = kc$.*

(Giv) *If $a, c \in \mathcal{S}(S)$ and $b \in S$ with $a\mathcal{R}'b$, then there exist $u, h \in \mathcal{S}(S)$, $v, k \in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $au\mathcal{R}'bc$, $v \leqq_l c$, $u \leqq_r a$, $k \leqq_l a$, $hua = ka^2$ and $hvc^2 = kbc$.*

REMARK. Given a *-pair $\mathcal{P} = (\leqq_l, \leqq_r)$ satisfying (Gi) and (Giii), condition (Evii)(l) in the list of conditions $\mathcal{P}$ must satisfy to be an embeddable *-pair becomes superfluous. For in this case, if $b, c \in S$, $a \in \mathcal{S}(S)$, $b, c \leqq_l a$ and $ba\mathcal{L}'ca$, then using (Giii), $hba = kca$ for some $h \in \mathcal{S}(S)$ and $k \in S$ with $ba \leqq_r h$ and $h\mathcal{R}'k$. By (Eii)(l) and (Evi)(l) we have $hb = kc$ and by (Ev)(l) $b\mathcal{R}'ba \leqq_r h$. Again by (Giii), $b \leqq_l c$. Dually one obtains $c \leqq_l b$ so that $b\mathcal{L}'c$.

PROOF of Theorem 4.1. Let S be a straight left order in a semigroup Q where $\mathcal{P} = \mathcal{P}(Q)$. Then $a \in \mathcal{S}(S)$ if and only if $a\mathcal{H}a^2$ in Q. So $a \in \mathcal{S}(S)$ if and only if $a\mathcal{H}'a^2$ in S, that is, a is good. Given $a \in \mathcal{S}(S)$ then H_a^Q is a group and by Proposition 2.2, $H'_a = S \cap H_a^Q$ is a left order in H_a^Q. Theorem 1.24 of [CP] tells us that H'_a is right reversible.

To see that (Giii) holds, take $b, c \in S$ where $b \leqq_l c$. Then $b = h^\# kc$ for some $h \in \mathcal{S}(S)$ and $k \in S$ where $h\mathcal{R}^Q k$. Then $bQ = h^\# kcQ \subseteq h^\# Q = hQ$ and $hb = hh^\# kc = kc$. Conversely, given elements $b, c, k \in S$ and $h \in \mathcal{S}(S)$ with $b \leqq_r h$ and $hb = kc$, then $b = h^\# hb = h^\# kc$ so that $Qb \subseteq Qc$ and $b \leqq_l c$ as required.

Finally we consider (Giv). Let $a, c \in \mathcal{S}(S)$ and $b \in S$ with $a\mathcal{R}'b$. Then $a^\# bc^\# = u^\# v$ for some $u \in \mathcal{S}(S)$ and $v \in S$ with $u\mathcal{R}'v$. This gives that

$$Qv = Qu^\# v = Qa^\# bc^\# \subseteq Qc^\# = Qc$$

and

$$uQ = u^\# vQ = a^\# bc^\# Q \subseteq aQ.$$

Now $bc^\# = au^\# v$ so that $au\mathcal{R}^Q au^\# v = bc^\# \mathcal{R}^Q bc$. We also have that $ua^\# bc = vc^2$ and $ua^\# = h^\# k$ for some $h \in \mathcal{S}(S)$ and $k \in S$ with $h\mathcal{R}'k$. Since $u \leqq_r a$ and u is good we have by (Fv) that $ua\mathcal{R}'u$, so that

$$u\mathcal{R}^Q ua\mathcal{R}^Q ua^\# \mathcal{R}^Q h\mathcal{R}^Q k$$

and u, v, h and k are $\mathcal{R}'$-related. Also,

$$Qk = Qh^\# k = Qua^\# \subseteq Qa^\# = Qa.$$

From $ua^{\#} = h^{\#}k$ we have $hua = ka^2$ and from $h^{\#}kbc = vc^2$ we have $kbc = hvc^2$.

Conversely, we suppose that S satisfies conditions (Gi), ..., (Giv) with respect to $\mathcal{P}$. Our aim is to construct via equivalence classes of ordered pairs of elements of S a semigroup Q in which S is embedded as a straight left order. This we do using a series of lemmas.

Let

$$\sum = \{(a, b) \in S \times S : a \in \mathcal{S}(S), a\mathcal{R}'b\}.$$

We note that by (Eiii), $\sum$ is a non-empty set. Define a relation $\sim$ on $\sum$ by

$$(a, b) \sim (c, d) \text{ if and only if } a\mathcal{R}'c \text{ and there exist } h, k \in H'_a \text{ with}$$
$$ha^2 = kca, \quad hb = kd.$$

If $(a, b) \sim (c, d)$ then by (Fii) and (Fiv) we have $b\mathcal{H}'hb = kd\mathcal{H}'d$.

LEMMA 4.2. *The relation $\sim$ is an equivalence relation on $\sum$.*

PROOF. Let $(a, b) \in \sum$. Putting $h = k = a$ we see that $(a, b) \sim (a, b)$.

Suppose now that $(a, b) \sim (c, d)$ so that $a\mathcal{R}'c$ and $ha^2 = kca$, $hb = kd$ for some $h, k \in H'_a$. We have that $ac\mathcal{H}'c\mathcal{H}'c^2$ so that by (Gii), $uac = vc^2$ for some $u, v \in H'_c$. Then $uac, hac \in H'_c$ so that $suac = thac$ for some $s, t \in H'_c$. Since $\mathcal{R}' \subseteq \mathcal{R}^*$ we have $hac = kc^2$ so that from

$$svc^2 = suac = thac = tkc^2,$$

$c^2\mathcal{R}'d$ and $ac\mathcal{R}'b$ we have $svd = tkd$ and $sub = thb$. Thus

$$sub = thb = tkd = svd$$

and as $ub\mathcal{R}'vd\mathcal{R}'s$, condition (Evi) gives that $ub = vd$ and so $(c, d) \sim (a, b)$.

It remains to show that $\sim$ is transitive. Consider $(a, b), (c, d), (m, n) \in \sum$ where $(a, b) \sim (c, d)$ and $(c, d) \sim (m, n)$. Then $a\mathcal{R}'b\mathcal{R}'c\mathcal{R}'d\mathcal{R}'m\mathcal{R}'n$ and there are elements $h, k \in H'_a$, $u, v \in H'_c$ with

$$ha^2 = kca, \quad hb = kd, \quad uc^2 = vmc, \quad ud = vn.$$

Since H'_c is right reversible there are elements $u', v' \in H'_c$ with $u'c^2 = v'amc$ and then one may pick $s, t \in H'_c$ with $su = tu'$. Thus

$$svmc = suc^2 = tu'c^2 = tv'amc$$

and as $mc\mathcal{R}'n$, $svn = tv'an$. Also,

$$tu'd = sud = svn = tv'an$$

gives by (Evi) that $u'd = v'an$. Now pick $p, q \in H'_c$ with $pkc^2 = qu'c^2$. Then

$$(ph)a^2 = pkca = qu'ca = (qv'a)ma$$

and

$$(ph)b = pkd = qu'd = (qv'a)n$$

using the fact that $a\mathcal{R}^* c\mathcal{R}^* c^2\mathcal{R}^* d$. But $ph, qv'a \in H'_a$ and so $\sim$ is transitive.

Put $Q = \sum / \sim$ and denote the $\sim$-equivalence class of $(a, b) \in \sum$ by $[a, b]$. Define a multiplication on Q by

$$[a, b][c, d] = [u, vd]$$

where there are elements $u, h \in \mathcal{S}(S)$, $v, k \in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $au\mathcal{R}'bc$, $v \leqq_l c$, $u \leqq_r a$, $k \leqq_l a$, $hua = ka^2$ and $hvc^2 = kbc$. Certainly by (Giv) such elements always exist. By (Ev), $vd\mathcal{R}'vc\mathcal{R}'v\mathcal{R}'u$ so that $(u, vd) \in \sim$.

LEMMA 4.3. *The given multiplication on Q is well-defined.*

PROOF. Suppose that $[a, b] = [s, t]$ and $[c, d] = [x, y]$ in Q. Then $a\mathcal{R}'b\mathcal{R}'s\mathcal{R}'t$ and there exist $m, n \in H'_a$ with

$$ma^2 = nsa, \quad mb = nt. \tag{1}$$

Also, $c\mathcal{R}'d\mathcal{R}'x\mathcal{R}'y$ and there exist $p, q \in H'_c$ with

$$pc^2 = qxc, \quad pd = qy. \tag{2}$$

By the definition of multiplication,

$$[a, b][c, d] = [u, vd]$$

where there are elements $u, h \in \mathcal{S}(S)$, $v, k \in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $au\mathcal{R}'bc$, $v \leqq_l c$, $u \leqq_r a$, $k \leqq_l a$ and

$$hua = ka^2, \quad hvc^2 = kbc \tag{3}$$

and

$$[s, t][x, y] = [u', v'y]$$

where there are elements $u', h' \in \mathcal{S}(S)$, $v', k' \in S$ with $h'\mathcal{R}'k'\mathcal{R}'u'\mathcal{R}'v'$, $su'\mathcal{R}'tx$, $v' \leqq_l x$, $u' \leqq_r s$, $k' \leqq_l s$ and

$$h'u's = k's^2, \quad h'v'x^2 = k'tx. \tag{4}$$

To show that the multiplication is well-defined we must prove $(u, vd) \sim \sim (u', v'y)$.

Since $u \leqq_r a$ and $ma^2 = nsa$ we know that $mau = nsu$. Then

$$nsu\mathcal{R}'mbc = ntc\mathcal{R}'ntx\mathcal{R}'nsu'.$$

Certainly $ns \in \mathcal{S}(S)$ and $u \leqq_r a\mathcal{R}'ns$, $u' \leqq_r s\mathcal{R}'ns$ so that by (Evii), $u\mathcal{R}'u'$. Then as H'_u is right reversible, there are elements $e, f \in H'_u$ with

$$eu^2 = fu'u. \tag{5}$$

The elements hu^2, $h'u'u$ lie in H'_u so there are elements g, i in H'_u with

$$ghu^2 = ih'u'u. \tag{6}$$

We can then pick $j, l \in H'_u$ with

$$jeu^2 = lghu^2.$$

Now $u \leqq_r ma$ and $ma \in \mathcal{S}(S)$ so by (Ev), $mau\mathcal{L}'u\mathcal{H}'lghu^2$. From (Giii) there are elements $o \in \mathcal{S}(S)$ and $r \in S$ with $o\mathcal{R}'r$, $lghu^2 \leqq_r o$ and $olghu^2 = rmau$.

This gives that

$$rnsu = rmau = olghu^2 = olgkau \tag{7}$$

using (3) and $u \leqq_r a$. But $au\mathcal{R}^*bc$ so that $rmbc = olgkbc$. Then using (1) and (3),

$$rntc = rmbc = olgkbc = olghvc^2.$$

Since $c\mathcal{R}^*x$ this gives that $rntx = olghvcx$. We use (7) again to obtain

$$rnsu = olghu^2 = olih'u'u = olik'su,$$

from (4) and the fact that $u\mathcal{R}'u' \leqq_r s$. But $su\mathcal{R}'su'\mathcal{R}'tx$ and so

$$olghvcx = rntx = olik'tx.$$

From $l\mathcal{R}'lghu^2 \leqq_r o$ and $o \in \mathcal{S}(S)$, (Eii) and (Eiv) give that $lghvcx = lik'tx$. The same two conditions then yield $ghvcx = ik'tx$ and so by (4) we have that $ghvcx = ih'v'x^2$. Now from $jeu^2 = lghu^2$ we have $jev = lghv$ and so $jevcx = lih'v'x^2$. From (6),

$$jfu'u = jeu^2 = lghu^2 = lih'u'u,$$

so that $jfv' = lih'v'$. Thus $jevcx = jfv'x^2$ and so $evcx = fv'x^2$. Since $evcx \leqq_l x\mathcal{H}'pcx$ there are elements $w \in \mathcal{S}(S)$ and $z \in S$ with $w\mathcal{R}'z$, $evcx \leqq_r \leqq_r w$ and $wevcx = zpcx$. This gives by (2) that

$$wfv'x^2 = wevcx = zpcx = zqx^2$$

and as $cx\mathcal{R}'d$ and $x^2\mathcal{R}'y$ we have that

$$wevd = zpd = zqy = wfv'y.$$

But $ev'yd\mathcal{R}'evcx \leqq_r w$ and

$$fv'y\mathcal{R}'fv'x\mathcal{R}'fv'\mathcal{R}'v'\mathcal{R}'ev\mathcal{R}'evc\mathcal{R}'evd$$

so by (Evi), $evd = fv'y$. Thus $(u, vd) \sim (u', v'y)$ as required.

LEMMA 4.4. *If* $[a,b] \in Q$ *and* $x \in \mathcal{S}(S)$ *where* $x\mathcal{R}'a$, *then* $[xa, xb] \in Q$ *and* $[a,b] = [xa, xb]$.

PROOF. Since $xa\mathcal{H}'a$ and $xb\mathcal{H}'b$, certainly $[xa, xb] \in Q$. By the right reversibility of H'_a, there are elements $p, q \in H'_a$ with $pa^2 = qxa^2 = q(xa)a$, and as $a^2\mathcal{R}^*b$ one has $pb = qxb$.

LEMMA 4.5. *The given multiplication on* Q *is associative.*

PROOF. Given $[a,b], [c,d], [h,k] \in Q$ let

$$X = ([a,b][c,d])[h,k]$$

and

$$Y = [a,b]([c,d][h,k]).$$

Then $X = [u_1, v_1 d][h,k]$ where there are elements $u_1, h_1 \in \mathcal{S}(S)$, $v_1, k_1 \in S$ with $h_1\mathcal{R}'k_1\mathcal{R}'u_1\mathcal{R}'v_1$, $au_1\mathcal{R}'bc, v_1 \leqq_l c$, $u_1 \leqq_r a$, $k_1 \leqq_l a$ and

$$h_1u_1a = k_1a^2, \quad h_1v_1c^2 = k_1bc. \tag{8}$$

Now by Lemmas 4.3 and 4.4,

$$X = [h_1u_1, h_1v_1d][h,k] = [u_2, v_2k]$$

where there are elements $u_2, h_2 \in \mathcal{S}(S)$, $v_2, k_2 \in S$ with $h_2\mathcal{R}'k_2\mathcal{R}'u_2\mathcal{R}'v_2$, $h_1u_1u_2\mathcal{R}'h_1v_1dh$, $v_2 \leqq_l h$, $u_2 \leqq_r h_1u_1$, $k_2 \leqq_l h_1u_1$ and

$$h_2u_2h_1u_1 = k_2(h_1u_1)^2, \quad h_2v_2h^2 = k_2h_1v_1dh. \tag{9}$$

Considering Y, we have that

$$Y = [a,b][u_3, v_3k]$$

where there are elements $h_3, u_3 \in \mathcal{S}(S)$, $k_3, v_3 \in S$ with $h_3\mathcal{R}'k_3\mathcal{R}'u_3\mathcal{R}'v_3$, $cu_3\mathcal{R}'dh$, $v_3 \leqq_l h$, $u_3 \leqq_r c$, $k_3 \leqq_l c$ and

$$h_3u_3c = k_3c^2, \quad h_3v_3h^2 = k_3dh. \tag{10}$$

Again by Lemmas 4.3 and 4.4,

$$Y = [a,b][h_3u_3, h_3v_3k] = [u_4, v_4h_3v_3k]$$

where there are elements $u_4, h_4 \in \mathcal{S}(S)$, $v_4, k_4 \in S$ with $h_4\mathcal{R}'k_4\mathcal{R}'u_4\mathcal{R}'v_4$, $au_4\mathcal{R}'bh_3u_3$, $v_4 \leqq_l h_3u_3$, $u_4 \leqq_r a$, $k_4 \leqq_l a$ and

$$h_4u_4a = k_4a^2, \quad h_4v_4(h_3u_3)^2 = k_4bh_3u_3. \tag{11}$$

Using Lemma 4.4 once more gives that $X = [h_2u_2, h_2v_2k]$ and $Y = [h_4u_4, h_4v_4h_3v_3k]$. Since $h_3u_3\mathcal{H}'u_3 \leqq_r c$ we have from (8) that $h_1v_1ch_3u_3 = k_1bh_3u_3$ and so as $u_4 \leqq_r a$,

$$h_1u_1u_4 = k_1au_4\mathcal{R}'k_1bh_3u_3 = h_1v_1ch_3u_3\mathcal{R}'h_1v_1cu_3\mathcal{R}'h_1v_1dh\mathcal{R}'h_1u_1u_2$$

so that by (Evii), $u_1u_4\mathcal{R}'u_1u_2$. Now $au_4\mathcal{R}'bh_3u_3\mathcal{R}'bu_3 \leqq_r bc\mathcal{R}'au_1$ so as $u_4, u_1 \leqq_r a$ (Fi) gives that $u_4 \leqq_r u_1$. In addition, $u_2 \leqq_r h_1u_1\mathcal{R}'u_1$ so that by (Evii), $u_4\mathcal{R}'u_2$. Then $h_2u_2\mathcal{R}'h_4u_4$ and we can pick $m, n \in H'_{h_2u_2}$ with

$$mh_2u_2h_2u_2 = nh_4u_4h_2u_2.$$

We can now make the following sequence of deductions. From $h_2u_2\mathcal{R}'u_2$,

$$mh_2u_2^2 = nh_4u_4u_2$$

and so as $u^2 \leqq_r h_1u_1$ and $u^2 \leqq_r a$ we have from (9) and (11) that

$$mk_2h_1u_1u_2 = mh_2u_2^2 = nh_4u_4u_2 = nk_4au_2.$$

Now using (8) and $u_2 \leqq_r a$,

$$mk_2k_1au_2 = mk_2h_1u_1u_2 = nk_4au_2$$

and so as $u_2\mathcal{R}'u_4$,

$$mk_2k_1au_4 = nk_4au_4.$$

But $au_4\mathcal{R}'bh_3u_3\mathcal{R}'bu_3$ and so

$$mk_2k_1bu_3 = nk_4bu_3.$$

We now use the fact that $u_3 \leqq_r c$, $h_3u_3\mathcal{R}'u_3$, (8), (10) and (11) to obtain

$$mk_2h_1v_1cu_3 = nk_4bu_3 = nh_4v_4h_3u_3^2 = nh_4v_4k_3cu_3.$$

We know that $cu_3\mathcal{R}'dh$ and so

$$mk_2h_1v_1dh = nh_4v_4k_3dh$$

from which (9) and (10) give

$$mh_2v_2h^2 = nh_4v_4h_3v_3h^2.$$

Finally as $h^2\mathcal{R}^*k$ we have

$$mh_2v_2k = nh_4v_4h_3v_3k.$$

We have constructed a semigroup Q, which we now show is the semigroup of left quotients for which we are looking. First we show that S is embedded in Q.

Let $s \in S$: from (Eiii) we know that $s\mathcal{R}'a$ for some $a \in \mathcal{S}(S)$. Define $\phi\colon S \to Q$ by

$$s\phi = [a, as] \quad \text{where} \quad a \in \mathcal{S}(S),\ a\mathcal{R}'s.$$

Since $as\mathcal{H}'s\mathcal{R}'a$ this definition makes sense. Moreover, if $b \in \mathcal{S}(S)$ and $b\mathcal{R}'s$, then $b\mathcal{R}'a$ and there are elements $h, k \in H'_a$ with $ha^2 = kba$. But then $has = kbs$ and so $[a, as] = [b, bs]$, showing that ϕ is well-defined.

LEMMA 4.6. *The function ϕ embeds S in Q.*

PROOF. Suppose that $s,t\in S$ and $s\phi=t\phi$, so that $(a,as)\sim(b,bt)$ where $a,b\in\mathcal{S}(S)$, $a\mathcal{R}'s$ and $b\mathcal{R}'t$. Thus $a\mathcal{R}'b$ and there are elements $h,k\in H'_a$ with $ha^2=kba$ and $has=kbt$. Since $a\mathcal{R}'t$,

$$hat=kbt=has$$

so by (Evi), $s=t$ and ϕ is one-one.

To show that ϕ is a homomorphism, again consider $s,t\in S$ where $s\mathcal{R}'a$, $t\mathcal{R}'b$ and $a,b\in\mathcal{S}(S)$. Then

$$s\phi t\phi=[a,as][b,bt]=[u,vbt]$$

where there are elements $u,h\in\mathcal{S}(S)$, $v,k\in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $au\mathcal{R}'asb$, $v\leqq_l b$, $u\leqq_r a$, $k\leqq_l a$, $hua=ka^2$ and $hvb^2=kasb$.

Choose $c\in\mathcal{S}(S)$ with $st\mathcal{R}'c$, so that $(st)\phi=[c,cst]$. Now $u\leqq_r a$, $sb\leqq_r\leqq_r s\mathcal{R}'a$ and so from $au\mathcal{R}'asb$, (Evii) gives $u\mathcal{R}'sb$. Then $u\mathcal{R}'sb\mathcal{R}'st\mathcal{R}'c$ and we may choose $p,q\in H'_u$ with $pu^2=qcu$.

From $hua=ka^2$ we have $hus=kas$ so that

$$hvb^2=kasb=husb$$

and so by (Evi), $vb^2=usb$. But $b\mathcal{R}'t$ and so

$$pvbt=pust=qcst$$

as required.

In view of the preceding lemma we may where convenient identify an element s of S with its image under ϕ in Q, and S with $S\phi$.

LEMMA 4.7. *Let $x\in\mathcal{S}(S)$ and $[a,b]\in Q$. Then*
(i) *if $a\leqq_r x$ then $[x,x][a,b]=[a,b]$,*
(ii) *if $b\leqq_l x$ then $[a,b][x,x]=[a,b]$.*

PROOF. (i) By definition of multiplication,

$$[x,x][a,b]=[u,vb]$$

where there are elements $h,u\in\mathcal{S}(S)$, $k,v\in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $xu\mathcal{R}'xa$, $v\leqq_l a$, $u\leqq_r x$, $k\leqq_l x$, $hux=kx^2$ and $hva^2=kxa$. Hence $u\mathcal{R}'a$ and there are elements $p,q\in H'_a$ with $pa^2=qua$. From $hux=kx^2$ and $a\leqq_r x$ we have $hua=kxa=hva^2$ and so $ua=va^2$. Thus $pa^2=qva^2$ and as $a^2\mathcal{R}'b$, $pb=qvb$.

(ii) By definition of multiplication,

$$[a,b][x,x]=[u,vx]$$

where there are elements $h,u\in\mathcal{S}(S)$, $k,v\in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $au\mathcal{R}'bx$, $v\leqq_l x$, $u\leqq_r a$, $k\leqq_l a$, $hua=ka^2$ and $hvx^2=kbx$. Since $b\leqq_l x$ by assumption,

we have that $au\mathcal{R}'bx\mathcal{R}'b\mathcal{R}'ab$ and so $u\mathcal{R}'b\mathcal{R}'a$. Thus we may pick $p, q \in H'_a$ with $pa^2 = qua$.

From $hvx^2 = kbx$ and $hvx, kb \leqq_l x$ we have $hvx = kb$. We now pick $m, n \in H'_u$ with $mhu = nqu$, from which we obtain

$$npa^2 = nqua = mhua = mka^2.$$

Then

$$npb = mkb = mhvx = nqvx$$

since $ua\mathcal{R}'v\mathcal{R}'vx$, and then by (Evi), $pb = qvx$ as required.

LEMMA 4.8. *Let $b, c, d \in H'_a$ where $a \in \mathcal{S}(S)$. Then $[b, c][c, b] = [d, d]$.*

PROOF. Certainly $[x, y] \in Q$ for any $x, y \in H'_a$. Now $[b, c][c, b] = [u, vb]$ where there are elements $h, u \in \mathcal{S}(S)$, $k, v \in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $bu\mathcal{R}'c^2$, $v \leqq_l c$, $u \leqq_r b$, $k \leqq_l b$, $hub = kb^2$ and $hvc^2 = kc^2$.

Since $bu\mathcal{R}'c^2\mathcal{R}'bd$ and $u \leqq_r b$, we have $u\mathcal{R}'d$ and so there are elements $p, q \in H'_d$ with $pd^2 = qud$. Now $hub = kb^2 = hvb^2$ and so $ub = vb^2$. Thus $pd^2 = qvbd$ and so $pd = qvb$ as required.

The following lemma is an easy consequence of Lemmas 4.7 and 4.8.

LEMMA 4.9. *Every square-cancellable element of S lies in a subgroup of Q. If $a \in \mathcal{S}(S)$ then the group $\mathcal{H}$-class H_a of Q has identity $[a, a]$ and the inverse of $a\phi = [a, a^2]$ in H_a is $[a^2, a]$. Moreover, if $b \in H'_a$ then $[a, a] = [b, b]$.*

LEMMA 4.10. *Let $q = [a, b] \in Q$. Then $q = a^{\#}b$.*

PROOF. Given $q = [a, b]$ by definition we have that $a \in \mathcal{S}(S)$ and $a\mathcal{R}'b$. Thus a lies in a subgroup of Q with group inverse $a^{\#} = [a^2, a]$ and $b\phi = [a, ab]$. Then

$$a^{\#}b = [a^2, a][a, ab] = [u, vab]$$

where there are elements $h, u \in \mathcal{S}(S)$, $k, v \in S$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $a^2u\mathcal{R}'a^2$, $v \leqq_l a$, $u \leqq_r a^2$, $k \leqq_l a^2$, $hua^2 = ka^4$ and $hva^2 = ka^2$. From $a^2u\mathcal{R}'a^2\mathcal{R}'a^3$ and $u \leqq_r a^2$ we have $u\mathcal{R}'a$ and so there are elements $p, q \in H'_a$ with $pa^2 = qua$. Since $v \leqq_l a^2$ and $k \leqq_l a^2$ we have from $hva^2 = ka^2$ that $hv = k$. Thus $hua^2 = hva^4$ and so $ua^2 = va^4$, giving $ua = va^3$. Now $pa^2 = qva^3$ so that $pb = qvab$ and $[a, b] = [u, vab] = a^{\#}b$ as required.

From Lemmas 4.9 and 4.10 we have

COROLLARY 4.11. *The semigroup S is a left order in Q.*

We now proceed to show that $\mathcal{P} = \mathcal{P}(Q)$.

LEMMA 4.12. *Let $b \in S$ and as in* (Fiii) *let $s \in S$ where*

$$bs\mathcal{R}'b\mathcal{L}'sb\mathcal{R}'s\mathcal{L}'bs$$

so that $bs, sb \in \mathcal{S}(S)$, $bsb\mathcal{H}'b$ and $sbs\mathcal{H}'s$. Then $b\phi = [bs, bsb]$ and has (semigroup) inverse $[sb, s]$ in Q. Further,

$$[bs, bsb][sb, s] = [bs, bs]$$

and

$$[sb, s][bs, bsb] = [sb, sb].$$

PROOF. By definition of multiplication,

$$[bs, bsb][sb, s] = [u, vs]$$

where there are elements $h, u \in \mathcal{S}(S)$, $k, v \in \mathcal{S}$ with $h\mathcal{R}'k\mathcal{R}'u\mathcal{R}'v$, $bsu\mathcal{R}'bsbsb$, $v \leqq_l sb$, $u \leqq_r bs$, $k \leqq_l bs$, $hubs = kbsbs$ and $hvsbsb = kbsbsb$. Since $u \leqq_r bs$ and $bsb\mathcal{H}'b\mathcal{R}'bs$ we have from $bsu\mathcal{R}'bsbsb$ that $u\mathcal{R}'bsb\mathcal{R}'bs$. Thus there are elements $p, q \in H'_{bs}$ with $pbsbs = qubs$. We have

$$hvsbsb = kbsbsb = hubsb$$

and so $vsbsb = ubsb$ and $vsb = ub$. Then $pbsbs = qvsbs$ so that $pbs = qvs$ and $[bs, bsb][sb, s] = [bs, bs]$.

Computing now $[sb, s][bs, bsb]$ we have that it equals $[u', v'bsb]$ where there are elements $h', u' \in \mathcal{S}(S)$, $k', v' \in S$ with $h'\mathcal{R}'k'\mathcal{R}'u'\mathcal{R}'v'$, $sbu'\mathcal{R}'sbs$, $v' \leqq_l bs$, $u' \leqq_r sb$, $k' \leqq_l sb$, $h'u'sb = k'sbsb$ and $h'v'bsbs = k'sbs$. From $u', s \leqq_r sb$ and $sbu'\mathcal{R}'sbs$ we have $u'\mathcal{R}'s\mathcal{R}'sb$ and so there are elements $p', q' \in \in H'_{sb}$ with $p'sbsb = q'u'sb$. We have

$$h'u'sb = k'sbsb = h'v'bsbsb$$

giving that $u'sb = v'bsbsb$ and $u's = v'bsbs$. Then $p'sbsb = q'v'bsbsb$ from which we deduce $p'sb = q'v'bsb$ so that $[sb, s][bs, bsb] = [sb, sb]$.

To see that $[sb, s]$ is the inverse of b is now a straightforward application of Lemma 4.7.

LEMMA 4.13. *The embeddable *-pair $\mathcal{P}$ is induced by Q.*

PROOF. Let $b, c \in S$ and suppose that $s, t \in S$ are chosen so that $b\mathcal{R}'bs\mathcal{L}'s\mathcal{R}'sb\mathcal{L}'b$ and $c\mathcal{R}'ct\mathcal{L}'t\mathcal{R}'tc\mathcal{L}'c$. Put $b' = [sb, s]$ and $c' = [tc, t]$ so that b' and c' are inverses in Q of b, c, respectively.

Assume first that $b \leqq_r c$. Then $bs\mathcal{R}'b \leqq_r c\mathcal{R}'ct$ so that $cc'b = = [ct, ct][bs, bsb] = [bs, bsb] = b$ by Lemmas 4.7 and 4.12, so that $bQ^1 \subseteq cQ^1$. Conversely, if $bQ^1 \subseteq cQ^1$ then either $b = c$ (and certainly $b \leqq_r c$) or $b = cq$ for some $q = [m, n]$ in Q. Now $cq = [ct, ctc][m, n] = [u, vn]$ where in particular $u \in \mathcal{S}(S)$, $v \in S$ and $u \leqq_r ct\mathcal{R}'c$. Then from $[bs, bsb] = [u, vn]$ we have $b\mathcal{R}'bs\mathcal{R}'u \leqq_r c$. The proof that $b \leqq_l c$ if and only if $Q^1b \subseteq Q^1c$ is similar and so we deduce that $\mathcal{P} = \mathcal{P}(Q)$.

LEMMA 4.14. *The semigroup S is a straight left order in Q.*

PROOF. If $q \in Q$ then by Lemma 4.10, $q = [a, b] = a^{\#}b$ where $a\mathcal{R}'b$ in S. But $\mathcal{P} = \mathcal{P}(Q)$ so that $a\mathcal{R}b$ in Q and S is straight in Q.

5. Fully stratified left orders

If S is a straight left order in a semigroup Q and $\mathcal{P}(Q) = (\leqq_l, \leqq_r)$, then S is *stratified* if $\mathcal{L}' = \mathcal{L}^*$ and $\mathcal{R}' = \mathcal{R}^*$, and *fully stratified* if $\leqq_l = \leqq_{\mathcal{L}^*}$ and $\leqq_r = \leqq_{\mathcal{R}^*}$. Clearly a fully stratified left order is stratified.

Stratified left orders occur naturally in the theory of semigroups of quotients. It is known, for example, that every left order in a bisimple inverse ω-semigroup is stratified [G2] and it is easy to see they are indeed fully stratified. Moreover from Lemma 1.7 of [FG] it is immediate that a ring R that is a straight order in a ring Q is also fully stratified in Q. However, not every stratified left order is fully stratified, as is shown in the following example.

EXAMPLE 5.1. Let S be a right but not left reversible, cancellative semigroup. Thus there are elements a, b in S with $aS \cap bS = \emptyset$. Pick $s \in S$ arbitrarily and let P be the 2×2 matrix over S given by $p_{11} = a$, $p_{12} = b$, $p_{21} = s$, $p_{22} = 0$. Put $T = \mathcal{M}^0(2, S, 2; P)$ and $Q = \mathcal{M}^0(2, G, 2; P)$, where G is the group of left quotients of S. As shown in [G3], T is a stratified left order in Q. We claim that T is not fully stratified.

Let $\alpha = (1, s, 2)$ and $\beta = (1, s, 1)$. If $x, y \in T^1$ and $\beta x = \beta y$ then either $x = y = 0$ and $\alpha x = \alpha y$ or $x \neq 0$ and $y \neq 0$. Clearly if $x, y \neq 0$ and $x \mathcal{R} y$ in Q then $x = y$ and $\alpha x = \alpha y$. But we have now covered all possibilities for x and y, for if $(1, s, 1)(1, t, j) = (1, s, 1)(2, u, j)$ then $sat = sbu$, giving $at = bu$ and $aS \cap bS \neq \emptyset$, a contradiction. Further, if $(1, s, 1)(1, t, j) = (1, s, 1)$ then $sat = s$ gives $satb^2 = sb^2$ and $atb^2 = b^2$, again a contradiction: similarly, $(1, s, 1)(2, t, j) = (1, s, 1)$ gives a contradiction. Thus $\alpha \leqq_{\mathcal{L}^*} \beta$ in T, but $Q\alpha$ is not contained in $Q\beta$.

We now use Theorem 4.1 to describe fully stratified left orders.

THEOREM 5.2. *A semigroup S is a fully stratified left order in some semigroup Q if and only if the *-pair $\mathcal{P} = (\leqq_{\mathcal{L}^*}, \leqq_{\mathcal{R}^*})$ satisfies conditions* (Ei), (Eii)(r), (Eiii), (Evi), (Evii)(r) *and* (Gii), (Giii) *and* (Giv).

PROOF. If S is a fully stratified left order in Q then the *-pair $\mathcal{P} = (\leqq_{\mathcal{L}^*}, \leqq_{\mathcal{R}^*}) = \mathcal{P}(Q)$ is an embeddable *-pair. Thus by Lemma 3.3 and Theorem 4.1 $\mathcal{P}$ satisfies the given conditions.

Conversely we suppose that $\mathcal{P} = (\leqq_{\mathcal{L}^*}, \leqq_{\mathcal{R}^*})$ satisfies the stated conditions. Clearly (Gi) holds and so by Theorem 4.1 it is enough to show that $\mathcal{P}$ is an embeddable *-pair.

Suppose that $b \in S$ and $a \in \mathcal{S}(S)$ where $b \leqq_{\mathcal{L}^*} a$. Then if $x, y \in S^1$ and $xba = yba$ we have by (Evi)(l) that $xb = yb$, for it is clear from the definition of $\leqq_{\mathcal{L}^*}$ that $xb, yb \leqq_{\mathcal{L}^*} b$. This gives that $b \mathcal{R}^* ba$ and so (Ev)(l) holds: dually, (Ev)(r) holds. Considering now (Eii)(l), suppose that $b, c \in S$ and $b \leqq_{\mathcal{L}^*} c$. Then from (Giii) $hb = kc$ for some $h \in \mathcal{S}(S)$ and $k \in S$ with $b \leqq_{\mathcal{R}^*} h$. But then by (Ev)$(r)$ $b \mathcal{L}^* hb = kc$ and it follows that (Eii)(l) holds. By the comment at the beginning of the proof of Theorem 4.1 we also have that $\mathcal{P}$ is an embeddable *-pair as required.

6. Left orders in regular $\mathcal{H}$-semigroups

We recall from the introduction that an *$\mathcal{H}$-semigroup* is a semigroup on which Green's relation $\mathcal{H}$ is a congruence. Any left order in a regular $\mathcal{H}$-semigroup is straight [G1]. The aim of this final section is to specialise Theorem 4.1 to obtain Theorem 3.1 of [G4], which describes those semigroups that are left orders in regular $\mathcal{H}$-semigroups.

LEMMA 6.1. *Let S be a straight left order in a semigroup Q. Then Q is an $\mathcal{H}$-semigroup if and only if $\mathcal{H}^Q \cap (S \times S)$ is a congruence on S.*

PROOF. It is clear that if Q is an $\mathcal{H}$-semigroup then $\mathcal{H}^Q \cap (S \times S)$ is a congruence on S. The proof of the converse is just as in Lemma 3.20 of [G4].

The approach of [G4] is via consideration of suitable pairs of equivalence relations on a semigroup. If $\mathcal{O} = (\mathcal{L}', \mathcal{R}')$ where $\mathcal{L}'$, $\mathcal{R}'$ are equivalence relations on a semigroup S, then $\mathcal{O}$ is a *suitable pair* if $\mathcal{L}'$ is a right congruence contained in $\mathcal{L}^*$, $\mathcal{R}'$ is a left congruence contained in $\mathcal{R}^*$ and for any $a \in S$, $a \in \mathcal{S}(S)$ if and only if $a\mathcal{H}'a^2$, where $\mathcal{H}' = \mathcal{L}' \cap \mathcal{R}'$.

The following corollary makes use of two lemmas from the proof of Theorem 3.1 of [G4]. However, our object is to avoid the 'constructive' part of the proof of that theorem.

COROLLARY 6.2. *Let S be a semigroup and let $\mathcal{O} = (\mathcal{L}', \mathcal{R}')$ be a suitable pair for S. The following conditions are equivalent:*

(i) *S is a left order in a regular $\mathcal{H}$-semigroup Q such that $\mathcal{L}^Q \cap (S \times S) = \mathcal{L}'$ and $\mathcal{R}^Q \cap (S \times S) = \mathcal{R}'$;*

(ii) *S satisfies conditions* (A), (B), (C), (D) *and* (E) *with respect to $\mathcal{O}$:*

(A) *$\mathcal{H}'$ is a congruence on S and $S/\mathcal{H}'$ is regular,*

(B) *if $a \in \mathcal{S}(S)$ then H'_a is right reversible,*

(C)(l) *if $a \in \mathcal{S}(S)$, $b, c \in S$, $a\mathcal{L}'b\mathcal{L}'c$ and $ba = ca$, then $b = c$,*

(D)(l) *if $a, b \in \mathcal{S}(S)$ and $a\mathcal{L}'b$ then $ba\mathcal{H}'b$,*

(E)(l) *if $a, b, c \in S$ and $a\mathcal{L}'cba$, then $a\mathcal{L}'ba$.*

PROOF. (i) $\Rightarrow$ (ii) Certainly S is a straight left order in Q and so by Theorem 4.1 $\mathcal{P}(Q) = \mathcal{P} = (\leqq_l, \leqq_r)$ is an embeddable *-pair satisfying (Gi), (Gii), (Giii) and (Giv). Note that there is no ambiguity in the notation $\mathcal{L}'$, $\mathcal{R}'$, $\mathcal{H}'$. From Proposition 2.6 of [G4], $\mathcal{H}'$ is a congruence on S and $S/\mathcal{H}' \cong Q/\mathcal{H}$ so is regular. Condition (B) is just (Gii), (C) follows from (Evi) and (D) from (Fii). To see that (E) holds, let $a, b, c \in S$ where $a\mathcal{L}'cba$. Then $Qa = Qcba \subseteqq Qba \subseteqq Qa$ so that $a\mathcal{L}'ba$; dually one can show that (E)(r) holds.

(ii) $\Rightarrow$ (i) Let $T = S/\mathcal{H}'$ and let $\phi\colon S \to T$ be an onto homomorphism with kernel $\mathcal{H}'$. From Lemma 3.3 of [G4],

$$a\phi\mathcal{L}b\phi \text{ in } T \text{ if and only if } a\mathcal{L}'b \text{ in } S$$

and

$$a\phi\mathcal{R}b\phi \text{ in } T \text{ if and only if } a\mathcal{R}'b \text{ in } S,$$

for any $a, b \in S$, from which it follows that T is $\mathcal{H}$-trivial.

Define relations $\leqq_l$, $\leqq_r$ on S by

$$a \leqq_l b \text{ if and only if } Ta\phi \subseteq Tb\phi$$

and

$$a \leqq_r b \text{ if and only if } a\phi T \subseteq b\phi T$$

where $a, b \in S$. Then $\leqq_l, \leqq_r$ are preorders on S with associated equivalence relations $\mathcal{L}'$, $\mathcal{R}'$, respectively. Let $\mathcal{P} = (\leqq_l, \leqq_r)$: we show that $\mathcal{P}$ is an embeddable *-pair satisfying (Gi), (Gii), (Giii) and (Giv). Note that it is built into the definition of a suitable pair that an element a of S is square-cancellable if and only if $a\mathcal{H}'a^2$ and we remark that this is also equivalent to $a\phi$ being idempotent.

It is easy to see that $\leqq_l$ is right compatible with multiplication and $\leqq_r$ is left compatible. Given $b, c \in S$ with $b \leqq_l c$, then $b\phi = d\phi c\phi = (dc)\phi$ for some $d \in S$, so that $b\mathcal{H}'dc$ and certainly $b\mathcal{L}'dc$. Thus $b\mathcal{L}^*dc \leqq_{\mathcal{L}^*} c$ by the assumption that $\mathcal{L}' \subseteq \mathcal{L}^*$. So $\leqq_l \subseteq \leqq_{\mathcal{L}^*}$ and dually, $\leqq_r \subseteq \leqq_{\mathcal{R}^*}$. We now have that $\mathcal{P}$ is a *-pair for S. Again, there is no ambiguity in using the notation $\mathcal{L}'$, $\mathcal{R}'$, $\mathcal{H}'$.

Using the above paragraph it is immediate that (Eii) holds. Moreover, it is straightforward to show that (Ei), (Eiii), (Ev) and (Evii) hold. Considering (Evi)(l), suppose that $b, c \in S$, $a \in \mathcal{S}(S)$, $b, c \leqq_l a$ and $ba = ca$. Now from (Evii)(l) we certainly have that $b\mathcal{L}'c$ and so $b\mathcal{L}'c\mathcal{L}'d$ for some $d \in \mathcal{S}(S)$. In T, $Td\phi \subseteq Ta\phi$ so that $Td\phi \subseteq Ta\phi d\phi \subseteq Td\phi$ and $d\mathcal{L}'ad$. Further, $adad\mathcal{H}'ad^2\mathcal{H}'ad$ so that $ad \in \mathcal{S}(S)$, $bad = cad$ and $b\mathcal{L}'c\mathcal{L}'ad$. Condition (C)$(l)$ gives that $b = c$. The dual argument shows that (Evi)(r) holds. Thus $\mathcal{P}$ is an embeddable *-pair satisfying (Gi) and (Gii).

Let $b, c \in S$ where $b \leqq_l c$. Then $b\mathcal{H}'dc$ for some $c \in S$. Let s be the element of S associated with b whose existence is guaranteed by (Fiii). Since $\mathcal{H}'$ is a congruence on S, $sb\mathcal{H}'sdc$ and since $\mathcal{H}'_{sb}$ is right reversible there are elements p, q in H'_{sb} with $psb = qsdc$ so that $bpsb = bqsdc$. It is routine to show that $b\mathcal{R}'bps\mathcal{R}'bqsd$ and $bps \in \mathcal{S}(S)$ so that putting $h = bps$ and $k = bqsd$, h and k are the elements required for (Giii). Conversely, given elements h and k as in (Giii), then $b\mathcal{L}'hb = kc \leqq_l c$.

To apply Theorem 4.1 it remains to show that (Giv) holds. Let $a, c \in \in \mathcal{S}(S)$ and $b \in S$, where $a\mathcal{R}'b$. We use Lemma 3.4 of [G4] to find elements $s \in \mathcal{S}(S)$, $t \in S$ with $s\mathcal{R}'t\mathcal{H}'abc^2$ and $sabc^2 = tc^4$. From (Evi)(l) we have that $sab = tc^3$. Put $h = s$, $k = s^2a$, $u = sa^2$ and $v = tc$. We claim that h, k, u and v are the elements required by (Giv). For from $s\mathcal{R}'abc^2$ we have $s \leqq_r a$ and as $s, a \in \mathcal{S}(S)$ it follows that $as\mathcal{H}'s\mathcal{R}'sa\mathcal{H}'u\mathcal{R}'k$ and $u \in \mathcal{S}(S)$. Also,

$$v = tc\mathcal{H}'abc^3\mathcal{H}'abc^2\mathcal{H}'t\mathcal{R}'s$$

so that h, k, u and v are $\mathcal{R}'$-related. Since $s \leqq_r a$ we have

$$au = asa^2\mathcal{H}'sa^2 = u\mathcal{R}'abc^2\mathcal{H}'bc.$$

Clearly $v \leqq_l c$, $u \leqq_r a$ and $k \leqq_l a$. Finally,

$$hua = s(sa^2)a = (s^2a)a^2 = ka^2$$

and

$$hvc^2 = s(tc)c^2 = (s^2a)bc = kbc.$$

By Theorem 4.1, S is a straight left order in a semigroup Q such that $\mathcal{P} = \mathcal{P}(Q)$ and Lemma 2.3 gives that Q is regular. In addition, $\mathcal{H}' = \mathcal{H}^Q \cap \cap (S \times S)$ is a congruence on S so that by Lemma 6.1, Q is an $\mathcal{H}$-semigroup.

REFERENCES

[CP] Clifford, A. H. and Preston, G. B., *The algebraic theory of semigroups*, Vol. I, Mathematical Surveys, No. 7, American Mathematical Society, Providence, RI, 1961. *MR* **24** #A2627; Vol. II, 1967. *MR* **36** #1558

[F] Fountain, J. B., Abundant semigroups, *Proc. London. Math. Soc.* (3) **44** (1982), 103–129. *MR* **83e**:20070

[FG] Fountain, J. B. and Gould, V. A. R., Orders in rings without identity, *Comm. Algebra* **18** (1990), 3085–3110. *MR* **91h**:16031

[FP] Fountain, J. B. and Petrich, M., Completely 0-simple semigroups of quotients, *J. Algebra* **101** (1986), 365–402. *MR* **87j**:20103

[G1] Gould, V. A. R., Orders in semigroups, *Contributions to General Algebra*, **5**, Verlag Hölder–Pichler–Tempsky, Wien, 1987, 163–169. *MR* **89c**:20089

[G2] Gould, V. A. R., Semigroups of quotients, *Proceedings of the International Symposium for the Theory of Regular Semigroups and their Applications*, University of Kerala, 1986.

[G3] Gould, V. A. R., Absolutely flat completely 0-simple semigroups of left quotients, *J. Pure Appl. Algebra* **55** (1988), 261–288. *MR* **89k**:20090

[G4] Gould, V. A. R., Left orders in regular $\mathcal{H}$-semigroups. I, *J. Algebra* **141** (1991), 11–35. *MR* **92e**:20042

[G5] Gould, V. A. R., Left orders in regular $\mathcal{H}$-semigroups. II, *Glasgow Math. J.* **32** (1990), 95–108. *MR* **91g**:20089

[H] Howie, J. M., *An introduction to semigroup theory*, L. M. S. Monographs, No. 7, Academic Press, London–New York, 1976. *MR* **57** #6235

(Received October 1, 1991)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF YORK
HESLINGTON, YORK
YO1 5DD
ENGLAND

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 375–378

ON THE SPECTRUM OF THE LAPLACIAN IN NEGATIVELY CURVED MANIFOLDS

A. BORBÉLY

Abstract

Let M^n be an n-dimensional, complete, simply connected Riemannian manifold with sectional curvature $K \leqq -k \leqq 0$ and Ricci curvature $Ric \leqq -\alpha < 0$. Then the spectrum of the Laplacian on M^n is bounded below by $\left((n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}\right)^2/4$. This improves a previous result due to A. G. Setti and H. P. McKean.

1. Introduction

Let M^n be an n-dimensional, complete, simply connected Riemannian manifold with sectional curvature $K \leqq -k \leqq 0$ and Ricci curvature $Ric \leqq -\alpha < 0$. A. G. Setti [4] improved an earlier result of McKean showing that the bottom of the spectrum λ_0 of Δ (considered as a positive selfadjoint operator on $L^2(M^n)$) is bounded below by $(\alpha + (n-1)(n-2)k)/4$.

The proof of this theorem hinges upon a geometric estimate for the trace of the second fundamental form of geodesic spheres. This, however, is not sharp, unless M^n has constant sectional curvature. We improve this estimate (Lemma 2) such that it becomes sharp when M^n has constant sectional curvature or M^n is the complex hyperbolic space. In the case of the real hyperbolic space our estimate reduces to that of McKean and Setti.

Using the same method as in [4], [3], [5] and [1], this estimate gives us the following generalization of Setti's result:

THEOREM 1. *Let M^n be an n-dimensional, complete, simply connected Riemannian manifold with sectional curvature $K \leqq -k \leqq 0$ and Ricci curvature $Ric \leqq -\alpha < 0$. Then for the bottom of the spectrum λ_0 of the Laplace operator Δ we have*

$$\lambda_0 \geqq \left((n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}\right)^2/4.$$

1980 *Mathematics Subject Classification.* Primary 58G25; Secondary 35P15.
Key words and phrases. Spectrum of the Laplacian.

2. Notation

Let $p \in M$ be fixed. Denote by ST_pM the unit sphere in the tangent plane at p and by $S_p(t)$ the geodesic sphere of radius t around p. Following Chavel [1, pp. 64–67], consider the map $\mathbf{A}_t : ST_pM \to S_p(t)$ defined by

$$\mathbf{A}_t(v) = \exp_p(tv).$$

After identifying T_pM and $T_{\exp_p(tv)}M$ via parallel translations along geodesics we can regard the derivative of this map $A(t,v) : T_pM \to T_pM$ as an endomorphism of the tangent plane at $p \in M$.

Therefore in geodesic spherical coordinates at p, $(t,v) \in \mathbf{R}^+ \times ST_pM$ we can write the metric in the form

$$ds^2 = (dt)^2 + |A(t,v)dv|^2$$

and the volume element of M as (cf. [1, p. 67])

$$dV(\exp_p tv) = g(t,v)dtd\mu_p(v),$$

where $g(t,v) = \det A(t,v)$ and $d\mu_p$ denotes the $(n-1)$-dimensional measure on the unit sphere ST_pM.

Define also by $U(t,v)$ the second fundamental form of the geodesic sphere $S_p(t)$ at $\exp_p tv$. It is well-known [1, p. 72] that $U(t,v) = A(t,v)'A^{-1}(t,v)$: $v^\perp \to v^\perp$, where again we identified the tangent planes at p and at $\exp_p tv$ via parallel translation. That is

$$\operatorname{tr} U = (\ln(\det A))' = g'/g, \tag{1}$$

where "$'$" = "$\frac{\partial}{\partial t}$". It is also known [1, p. 72] that $U(t,v)$ satisfies the Riccati equation

$$U' + U^2 + R = 0, \tag{2}$$

where $R\xi = R(v,\xi)v$ is the curvature tensor at $\exp_p tv$ and $\xi \in v^\perp$ (again we identified the tangent planes via parallel translation along $\exp_p tv$).

3. Proof of Theorem 1

Following the idea of [3], [1], [4] and [5] we need the following geometric estimate:

LEMMA 2. *Let M^n be an n-dimensional, complete, simply connected Riemannian manifold with sectional curvature $K \leqq -k \leqq 0$ and Ricci curvature $Ric \leqq -\alpha < 0$. Then*

$$\operatorname{tr} U = g'/g \geqq (n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}.$$

The proof of Lemma 2 follows from two simple observations.

PROPOSITION 3. *Let* $f: \mathbf{R}^+ \to \mathbf{R}$ *be a* C^1*-function with* $\lim_{t\to 0+} f(t) = +\infty$ *and* $\inf f(t) = \beta > -\infty$. *Then for every* $\varepsilon > 0$ *there is a* t_0 *such that* $|f'(t_0)| < \varepsilon$ *and* $|f(t_0) - \beta| < \varepsilon$.

PROOF. The proof is elementary and we will leave it to the reader.

PROPOSITION 4. *Let* U *be a positive definite* $(n-1)\times(n-1)$*-matrix with eigenvalues* $\sqrt{k} \leqq \lambda_1 \leqq \cdots \leqq \lambda_{n-1}$ *and* $\operatorname{tr} U^2 \geqq \alpha$. *Then*

$$\operatorname{tr} U \geqq (n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}.$$

PROOF. Again it is easy to see that the minimum is assumed when $\sqrt{k} = \lambda_1 = \cdots = \lambda_{n-2}$ and $\lambda_{n-1} = \sqrt{\alpha - (n-2)k}$. Suppose this is not true. Then let λ_i be the first eigenvalue such that $\lambda_i > \sqrt{k}$, $i < n-1$ and consider a symmetric matrix V with eigenvalues $\lambda_1, \ldots, \lambda_{i-1}, \sqrt{k}, \sqrt{\lambda_i^2 + \lambda_{i+1}^2 - k}, \lambda_{i+2}, \ldots, \lambda_{n-1}$. Obviously, every eigenvalue of V is larger than or equal to $\sqrt{k}$ and

$$\operatorname{tr} V^2 = \operatorname{tr} U^2.$$

However, an elementary computation shows that

$$\operatorname{tr} U > \operatorname{tr} V,$$

which completes the proof of the proposition.

PROOF OF LEMMA 2. Let $v \in ST_pM$ be fixed and $\beta = \inf \operatorname{tr} U$. The comparison principle for the matrix Riccati equation (2) shows (cf. [2]) that every eigenvalue of U is larger than or equal to $\sqrt{k}$. Taking the trace of (2) we have

$$(\operatorname{tr} U)' + \operatorname{tr} U^2 = -\operatorname{tr} R \geqq \alpha. \tag{3}$$

Obviously, the function $\operatorname{tr} U$ satisfies the condition of Proposition 3, so for every $\varepsilon > 0$ there is a t_0 such that

$$\operatorname{tr} U^2(t_0, v) > \alpha - \varepsilon \quad \text{and} \quad \beta > \operatorname{tr} U(t_0, v) - \varepsilon.$$

From Proposition 4 we know that

$$\beta > (n-2)\sqrt{k} + \sqrt{\alpha - \varepsilon - (n-2)k} - \varepsilon.$$

Letting ε go to zero proves the Lemma.

Now, the proof of Theorem 1 is the same as in [4, p. 281], [1, p. 47] or [5, pp. 67–69] but for the sake of completeness we include it. By Rayleigh's Theorem, it suffices to show that for every $f \in C_0^\infty(M)$,

$$\int_M |\nabla f|^2 dV \geqq \left((n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}\right)^2 / 4 \int_M f^2 dV.$$

For convencience, set $\beta = (n-2)\sqrt{k} + \sqrt{\alpha - (n-2)k}$. Using spherical coordinates $(t, v) \in \mathbf{R}^+ \times ST_pM$, from Lemma 2, we have for every $v \in ST_pM$,

$$\int_0^\infty f^2(t,v)g(t,v)dt \leqq 1/\beta \int_0^\infty f^2(t,v)g'(t,v)dt =$$

$$= -2/\beta \int_0^\infty f(t,v)f'(t,v)g(t,v)dt \leqq$$

$$\leqq 2/\beta \left(\int_0^\infty f^2(t,v)g(t,v)dt \right)^{1/2} \times \left(\int_0^\infty (f'(t,v))^2 g(t,v)dt \right)^{1/2}.$$

Using the fact that $|\nabla f|^2 \geqq (f'(t,v))^2$ and integrating over $v \in ST_pM$, the Theorem follows immediately.

REFERENCES

[1] CHAVEL, I., *Eigenvalues in Riemannian geometry*, Pure and Applied Mathematics, 115, Academic Press, Orlando, Fla., 1984. *MR* **86g**:58140

[2] ESCHENBURG, J.–H. and HEINTZE, E., Comparison theory for Riccati equations, *Manuscripta Math.* **68** (1990), 209–214. *MR* **91d**:34034

[3] MCKEAN, H. P., An upper bound to the spectrum of Δ on a manifold of negative curvature, *J. Differential Geometry* **4** (1970), 359–366. *MR* **42** #1009

[4] SETTI, A. G., A lower bound for the spectrum of the Laplacian in terms of sectional and Ricci curvature, *Proc. Amer. Math. Soc.* **112** (1991), 277–282. *MR* **91h**:58120

[5] STRICHARTZ, R. S., Analysis of the Laplacian on the complete Riemannian manifold, *J. Funct. Anal.* **52** (1983), 48–79. *MR* **84m**:58138

(Received October 8, 1991)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NOTRE DAME
NOTRE DAME, IN 46556
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 379–388

REGULAR LATTICES

D. D. ANDERSON and C. JAYARAM

Abstract

Let L be a compactly generated multiplicative lattice with greatest element compact. We show that L is a finite Boolean algebra if and only if (i) L is reduced, (ii) every proper compact element is a zero divisor, and (iii) 0 is a product of compact primary elements. We define a lattice L to be *regular* if each compact element is complemented. Regular lattices are investigated and several conditions equivalent to a lattice being regular are given. For example, we show that L is regular if and only if L is reduced and every prime element is maximal.

1. Introduction

A multiplicative lattice is a complete lattice in which there is defined a commutative, associative multiplication which distributes over arbitrary joins (i.e., $a(\vee_\alpha b_\alpha) = \vee_\alpha (ab_\alpha)$) and has greatest element 1 (least element 0) as a multiplicative identity (zero) (see [1]). In this paper, we prove that a compactly generated multiplicative lattice L with compact identity element is a finite Boolean algebra if and only if L satisfies the following three conditions: (i) L is reduced, (ii) every proper compact element is a zero divisor and (iii) 0 is the product of a finite number of compact primary elements. Next we introduce the concept of a regular lattice and establish some equivalent conditions for an r-lattice L to be a regular lattice. It is shown that a compactly generated multiplicative lattice L in which 1 is a compact element and every finite product of compact elements is a compact element is a regular lattice if and only if L is reduced and every prime element is a maximal element. This result is used to show that L is regular if and only if every primary element is a maximal element. Finally, we prove that L is Noetherian regular if and only if L is reduced and every radical element has a unique representation as a meet of prime elements.

2. Preliminaries

Let L be a multiplicative lattice. An element p different from 1 is called *prime* if $ab \leqq p$ implies either $a \leqq p$ or $b \leqq p$. An element $p(\neq 1)$ is said to be

1980 *Mathematics Subject Classification* (1985 Revision). Primary 06E99, 06C15, 06F05, 06F10.

Key words and phrases. Multiplicative lattice, Boolean algebra.

primary if for every pair of compact elements $a, b \in L$, $ab \leqq p$ implies either $a \leqq p$ or $b^n \leqq p$ for some $n \in \mathbf{Z}^+$. A proper element m of L is said to be a *maximal element* if $m \nleqq a$ for any other proper element a of L. An element $a \in L$ is called *compact* whenever $a \leqq \vee X$, $X \subseteq L$, implies the existence of a finite number of elements $x_1, x_2, \ldots, x_n$ of X such that $a \leqq x_1 \vee x_2 \vee \cdots \vee x_n$. L is said to be *compactly generated* if every element of L is the join of compact elements.

3. Finite Boolean algebras

Throughout this section, L denotes a compactly generated multiplicative lattice with 1 a compact element. Since 1 is a compact element of L, maximal elements exists in L and every maximal element is a prime element.

An element $a \in L$ is said to be *complemented* if $a \wedge b = 0$ and $a \vee b = 1$ for some $b \in L$. If $a \vee b = 1$, then $a \wedge b = ab$. Thus in the definition of a being complemented we can replace the condition $a \wedge b = 0$ by $ab = 0$. Let $C(L) = \{a \in L | a \text{ is a complemented element of } L\}$. It can be easily verified that $C(L)$ is a Boolean algebra with $ab = a \wedge b$ for every $a, b \in C(L)$. An element $a \in L$ is said to be *nilpotent* if $a^n = 0$ for some $n \in \mathbf{Z}^+$, while a is called a *zero divisor*, if $ab = 0$ for some nonzero element $b \in L$. L is said to be *reduced* if 0 is the only nilpotent element of L. A nonzero element $a \in L$ is said to be an *atom* if $0 \leqq b \leqq a$ implies either $0 = b$ or $b = a$. For undefined terms from lattice theory, the reader is referred to [2].

Now we shall begin with the following lemmas.

LEMMA 1. *If $a \vee b = a \vee c = 1$, then $a \vee (bc) = 1$.*

PROOF. Obvious.

LEMMA 2. *Suppose L is reduced and every proper compact element of L is a zero divisor. Then every compact primary element is a maximal element.*

PROOF. Suppose x is a compact primary element. Let $x \leqq y < 1$ for some $y \in L$. As L is compactly generated, we have $y = \vee_\alpha a_\alpha$, where the a_α's are compact elements. If $y \leqq x$, then we are through. Suppose not. Then there is some a_α such that $a_\alpha \nleqq x$. So $x < x \vee a_\alpha < 1$. Since $x \vee a_\alpha$ is a proper compact element, by hypothesis $(x \vee a_\alpha)b = 0$ for some nonzero element $b \in L$ which may be chosen to be compact. Since $a_\alpha b = 0 \leqq x$ and x is primary, it follows that $b^n \leqq x$ for some $n \in \mathbf{Z}^+$ and so $b^{n+1} \leqq xb = 0$. As L is reduced, we get $b = 0$, a contradiction. Therefore every compact primary element is a maximal element.

LEMMA 3. *Suppose L is reduced. If x is a maximal element and $xy = 0$ $(y \neq 0)$, then y is an atom.*

PROOF. Suppose $0 \leqq z \leqq y$ for some $z \in L$. As x is a maximal element, we have either $x \vee z = 1$ or $z \leqq x$. If $x \vee z = 1$, then $y = y(x \vee z) = yz \leqq z$

and so $y = z$. If $z \leqq x$, then $zy \leqq xy = 0$ and therefore $z^2 \leqq zy = 0$. As L is reduced, we get $z = 0$. Hence y is an atom.

LEMMA 4. *L is a finite Boolean algebra with $xy = x \wedge y$ if and only if every maximal element is a complemented element.*

PROOF. The 'only if' part is obvious. We now prove the 'if' part. Using Zorn's Lemma, it can be easily proved that every element is a complemented element and hence $L = C(L)$ is a Boolean algebra with $xy = x \wedge y$. For each maximal element m_α, let y_α be its complement. By Lemma 3, y_α is an atom. Now $1 = \vee y_\alpha$. Since 1 is compact, $1 = y_{\alpha_1} \vee \cdots \vee y_{\alpha_n}$. This shows that L contains only a finite number of atoms and hence L is a finite Boolean algebra.

LEMMA 5. *L is a finite Boolean algebra with $xy = x \wedge y$ if and only if for every maximal element $m \in L$, there is some complemented atom $n \in L$ such that $n \nleqq m$.*

PROOF. The 'only if' part is obvious. We now prove the 'if' part. By Lemma 4, it is enough to prove that every maximal element is a complemented element. Let m be a maximal element. By hypothesis, there is some complemented atom $e \in L$ such that $e \nleqq m$. We claim that $m = e'$ where e' is a complement of e. Since $ee' = 0 \leqq m$ and m is a prime element, it follows that $e' \leqq m$. Again since $0 \leqq em \leqq e$, $e \nleqq m$, and e is an atom, it follows that $em = 0$ so that $m = m1 = m(e \vee e') = me' \leqq e'$. This shows that $m = e'$ and hence every maximal element is a complemented element. This completes the proof of the lemma.

We now characterize finite Boolean algebras as follows.

THEOREM 1. *L is a finite Boolean algebra with $xy = x \wedge y$ if and only if L satisfies the following three conditions.*

(i) *L is reduced.*

(ii) *Every proper compact element of L is a zero divisor.*

(iii) *0 is the product of a finite number of compact primary elements.*

PROOF. The 'only if' part is obvious. We now establish the 'if' part. By hypothesis and Lemma 2, $0 = a_1 a_2 \ldots a_n$, where the a_i's are maximal elements. By (ii), for each $i \in \{1, 2, \ldots, n\}$ there exists a nonzero $b_i \in L$ with $a_i b_i = 0$. By Lemma 3, each b_i is an atom. Also $a_i \vee (b_1 \vee \cdots \vee b_n) \geqq a_i \vee b_i = 1$ for $i = 1, 2, \ldots, n$; so that by Lemma 1, $\bigvee_{i=1}^{n} b_i = 1$. Now the result follows from Lemma 5. This completes the proof of the theorem.

REMARK 1. It is not hard to show that the conditions (i), (ii) and (iii) of the above theorem are independent.

LEMMA 6. *Every complemented element of L is a compact element.*

PROOF. Let $a \in L$ be a complemented element of L. Suppose $a \leqq \vee X$. Then $1 = a \vee a' \leqq \vee X \vee a'$ (a' is a complement of a). Since 1 is compact,

it follows that $a \vee a' = 1 \leqq (a_1 \vee a_2 \vee \cdots \vee a_n) \vee a'$ for some $a_1, a_2, \ldots, a_n \in X$. Now $a = a1 = a((a_1 \vee a_2 \vee \cdots \vee a_n) \vee a') = a(a_1 \vee a_2 \vee \cdots \vee a_n \vee a') = a(\bigvee_{i=1}^{n} a_i) \leqq \bigvee_{i=1}^{n} a_i$ and hence a is compact. This completes the proof of the lemma.

Observe that L is a Boolean algebra with $xy = x \wedge y$ if and only if every element of L is a complemented element of L. Using this fact, we prove the following theorem.

THEOREM 2. *The following statements on L are equivalent.*
(i) *L is a finite Boolean algebra with $xy = x \wedge y$.*
(ii) *L is a Boolean algebra with $xy = x \wedge y$.*
(iii) *L is reduced and every proper element of L is a zero divisor.*

PROOF. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) is obvious. We now prove (iii) $\Rightarrow$ (i). Suppose (iii) holds. First we show that every element of L is a complemented element. Let $a \in L$. Put $a^* = \vee\{x \in L | ax = 0\}$. Obviously $aa^* = 0$. We claim that $a \vee a^* = 1$. Suppose $a \vee a^* \neq 1$. Then by hypothesis $(a \vee a^*)b = 0$ for some $b \neq 0$. Observe that $ab = 0$ and $a^*b = 0$. Since $ab = 0$, we get $b \leqq a^*$ and so $b^2 \leqq a^*b = 0$. As L is reduced, $b = 0$, a contradiction. Therefore $a \vee a^* = 1$ and hence every element of L is a complemented element. Consequently, by Lemma 4, L is a finite Boolean algebra with $xy = x \wedge y$.

4. Regular lattices

Throughout this section, L denotes a compactly generated multiplicative lattice with 1 as a compact element. We also assume that every finite product of compact elements of L is a compact element. For any $a \in L$, let $\sqrt{a} = \vee\{x \in L | x$ is compact and $x^n \leqq a$ for some $n \in \mathbf{Z}^+\}$. It can be easily shown that $\sqrt{a} = \wedge\{p \in L | a \leqq p$ and p is a prime element$\}$ (see also Theorem 3.6 of [6]). For any $a, b \in L$, let $(a : b) = \vee\{x \in L | bx \leqq a\}$. According to [4], an element $m \in L$ is said to be *meet (join) principal* if $a \wedge mb = m((a : m) \wedge b)$ $(a \vee (b : m) = ((am \vee b) : m))$ for all $a, b \in L$. An element $m \in L$ is called *weak meet (join) principal* if $a \wedge m = m(a : m)$ $(a \vee (0 : m) = (ma : m))$ for all $a \in L$, and m is said to be *(weak) principal* if m is both (weak) meet and (weak) join principal. A multiplicative lattice L is called an r-lattice ([1]) if it is modular, principally generated, compactly generated, and has 1 compact. Note that in an r-lattice, every finite product of compact elements is a compact element (see [1]). For details on principal elements the reader is referred to [1] and [4].

In this section, we introduce the concept of a regular lattice and obtain some equivalent conditions for L to be a regular lattice. Next Noetherian regular lattices are characterized.

We shall begin with the following lemma.

LEMMA 7. *Let $a \in L$ be a complemented element. Suppose a' is a complement of a. Then*
(i) $a' = (0 : a)$, *and*
(ii) a *is weak principal.*

PROOF. (i) Since $aa' = 0$, we get $a' \leqq (0 : a)$. Now $(0 : a) = (0 : a)1 = (0 : a)(a \vee a') = (0 : a)a \vee (0 : a)a' = (0 : a)a' \leqq a'$, so $(0 : a) \leqq a'$ and hence $a' = (0 : a)$.

(ii) Obviously a is weak meet principal. Now we show that a is weak join principal. Let $ba \leqq ca$. By (i) and Proposition 1.1 of [1], it is enough if we show that $b \leqq c \vee a'$. Now $b = b1 = b(a \vee a') = ba \vee ba' \leqq c \vee a'$ since $ba \leqq ca \leqq c$ and $ba' \leqq a'$. Thus a is weak principal.

LEMMA 8. *Suppose L is an r-lattice. Then an element $a \in L$ is a complemented element if and only if a is an idempotent principal element.*

PROOF. Suppose a is a complemented element. Obviously a is an idempotent. By Lemma 7 (ii), a is weak principal. Since L is modular, by Proposition 1.1 (6) of [1], a is principal and hence a is an idempotent principal element.

The converse part is obvious.

We now introduce the concept of a regular lattice and characterize them.

DEFINITION 1. L is said to be a *regular lattice* if every compact element of L is a complemented element of L.

A commutative ring R with identity is called (von Neumann) regular if for each $a \in R$, there exists $x \in R$ such that $axa = a$. The lattice of all ideals of a commutative regular ring with identity is a regular lattice. The lattice of all ideals of a Boolean algebra is also a regular lattice.

THEOREM 3. *An r-lattice L is regular if and only if every compact element is an idempotent principal element.*

PROOF. Follows from Lemma 8.

LEMMA 9. *If L is regular, then every element of L is an idempotent.*

PROOF. Let $a \in L$. As L is compactly generated, we have $a = \bigvee_\alpha a_\alpha$, where each a_α is a complemented element. Note that each a_α is an idempotent. Now $a^2 = a(\bigvee_\alpha a_\alpha) = \bigvee_\alpha aa_\alpha = \bigvee_\alpha a_\alpha = a$ since $aa_\alpha = a_\alpha$ for each α.

THEOREM 4. *Let L be an r-lattice. Then the following statements are equivalent.*
(i) L *is a regular lattice.*
(ii) *Every element of L is an idempotent.*
(iii) $a \wedge b = ab$ *for every* $a, b \in L$.
(iv) *For any $a \in L$, there is some $x \in L$ such that $a = axa$.*
(v) $a = \sqrt{a}$ *for every* $a \in L$.

PROOF. (i) $\Rightarrow$ (ii) follows from Lemma 9 and (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (ii) is obvious. Now we show that (ii) $\Rightarrow$ (i). By (ii) every principal element is an idempotent and so by Lemma 8, every principal element is a complemented element. Consequently every compact element is a complemented element. That is, L is regular. This completes the proof of the theorem.

REMARK 2. It can easily be shown that an r-lattice L is regular if and only if every principal element is a complemented element.

DEFINITION 2. L is called a *Noetherian regular lattice* if it is regular and satisfies the ascending chain condition.

REMARK 3. It can be easily verified that L is Noetherian regular if and only if L is a Boolean algebra.

It is well known that a commutative ring R with identity is von Neumann regular if and only if R is semiprime (0 is the only nilpotent element of R) and every prime ideal is a maximal ideal. We now establish the abstract version of the above result.

THEOREM 5. *L is regular if and only if L is reduced and every prime element is a maximal element.*

PROOF. Suppose L is regular. By Lemma 9, L is reduced. Let p be a prime element. We claim that p is maximal. Suppose not. Then $p < m$ for some maximal element. As L is compactly generated, there is some compact element a such that $a \not\leqq p$ and $a \leqq m$. Let a' be a complement of a. Then $a'a = 0 \leqq p$ and so $a' \leqq p \leqq m$. Therefore $1 = a \vee a' \leqq m$, a contradiction. Hence every prime element is a maximal element.

Conversely, assume that a is a compact element which is not a complemented element. Let $D = \{b \in L | b$ is compact and $a \vee b = 1\}$ and $D_1 = \{fa^n | f \in D$ and $n \in \mathbf{Z}^+ \cup \{0\}\}$. Observe that $0 \notin D_1$, so by the Separation Lemma of [6], there exists a prime element $p \in L$ such that $t \not\leqq p$ for all $t \in D_1$. Obviously $p \vee a < 1$ and hence $p \vee a \leqq m$ for some maximal element m of L. Again since $p < m$, this contradicts the fact that every prime element is a maximal element. Therefore every compact element is a complemented element and hence L is regular.

DEFINITION 3. L is said to be *semisimple* if $0 = \wedge\{a \in L | a$ is a maximal element of $L\}$.

THEOREM 6. *Let L be an r-lattice. Then the following statements are equivalent.*

(i) *L is Noetherian regular.*
(ii) *L is semisimple and satisfies the descending chain condition.*
(iii) *L is the direct product of a finite number of two element Boolean algebras.*

PROOF. (i) $\Rightarrow$ (ii). By Remark 3 and Theorem 2, L is finite Boolean algebra and hence L is semisimple and satisfies the descending chain condition.

(ii) $\Rightarrow$ (iii). First we show that L contains only a finite number of maximal elements.

Let $S=\{a \in L | a$ is the meet of a finite number of maximal element of $L\}$. Since L satisfies d.c.c., it follows that S contains a minimal element, say $a \in \in S$. Suppose $a = \bigwedge_{i=1}^{n} p_i$, where p_i's are maximal elements of L. We claim that $a = 0$. Suppose $a \neq 0$. Then $a \not\leqq p$ for some maximal element p of L, since L is semisimple. So $b = (\bigwedge_{i=1}^{n} p_i) \wedge p \in S$ and $b < a$ which contradicts the minimality of a. Hence $0 = \bigwedge_{i=1}^{n} p_i$. Now it can be easily shown that $p_1, \ldots, p_n$ are the only maximal elements of L, and $L \cong L/p_1 \times \cdots \times L/p_n$ where each L/p_i is a two element Boolean algebra. Thus (iii) holds. The implication (iii) $\Rightarrow$ (i) follows from Remark 3. This completes the proof of the theorem.

5. Primary elements in lattices

Throughout this section, L denotes a compactly generated multiplicative lattice in which 1 is a compact element and every finite product of compact elements is a compact element. An element a of L is said to be a *radical* element if $a = \sqrt{a}$. An element a of L is called *completely irreducible* if whenever $a = \wedge_\alpha a_\alpha$, then $a = a_\alpha$ for some α. In this section, regular lattices are characterized in terms of primary elements and also it is shown that L is Noetherian regular if and only if L is reduced and every radical element has a unique representation as a meet of primary elements.

Now we need some lemmas.

LEMMA 10. *Every element of L is the meet of completely irreducible elements.*

PROOF. Follows from 6.1 of [3, page 43].

LEMMA 11. *Let $m \in L$ and $\{a_\alpha\} \subseteq L$. Then $\bigwedge_\alpha (m : a_\alpha) = (m : \bigvee_\alpha a_\alpha)$.*

PROOF. It is easily established that this identity holds in any multiplicative lattice.

Let $a \in L$. An element $b \in L$ is called *prime to a* if whenever $bc \leqq a$ then $c \leqq a$. For any $a \in L$, we denote $p_a = \bigvee \{x \in L | x$ is non-prime to $a\}$.

LEMMA 12. *Suppose a is a completely irreducible element of L and let $\{a_\alpha\} \subseteq L$. If a_α is non-prime to a, then $\bigvee_\alpha a_\alpha$ is non-prime to a.*

PROOF. By Lemma 11, we have $\bigwedge_\alpha (a : a_\alpha) = (a : \bigvee_\alpha a_\alpha)$. Since each a_α is non-prime to a, for each α, there exists $b_\alpha \in L$ such that $a_\alpha b_\alpha \leqq a$ and $b_\alpha \not\leqq a$. So $a < (a : a_\alpha)$ for every α. As a is completely irreducible, we get $a < \bigwedge_\alpha (a : a_\alpha) = (a : \bigvee_\alpha a_\alpha)$. Therefore $\bigvee_\alpha a_\alpha$ is non-prime to a.

LEMMA 13. *If a is a completely irreducible element of L, then p_a is a prime element.*

PROOF. Suppose $xy \leqq p_a$. As L is compactly generated, we have $xy = \bigvee_{\alpha\in\Delta} b_\alpha$, where the b_α's are compact elements. We claim that each b_α is non-prime to a. Let $\alpha \in \Delta$. Then $b_\alpha \leqq p_a$ and so $b_\alpha \leqq a_1 \vee \cdots \vee a_n$ for some $a_1, a_2, \ldots, a_n \in \{c \in L | c \text{ is non-prime to } a\}$. By Lemma 12, $a_1 \vee \cdots \vee a_n$ is non-prime to a and hence b_α is non-prime to a. Again by Lemma 12, $xy = \bigvee_\alpha b_\alpha$ is non-prime to a and hence either x or y is non-prime to a. This shows that p_a is a prime element.

LEMMA 14. *If every element of L is a minimal prime element, then every completely irreducible element is primary.*

PROOF. Let x be a completely irreducible element. Then by Lemma 13, p_x is a prime element. We claim that $p_x = \sqrt{x}$. Obviously $\sqrt{x} \leqq p_x$. Suppose a is a compact element and let $a \nleqq \sqrt{x}$. Then $a^n \nleqq x$ for all $n \in \mathbf{Z}^+$. Put $S = \{ba^n | b \text{ is compact, } b \nleqq p_x \text{ and } n \in \mathbf{Z}^+ \cup \{0\}\}$. Observe that S is a multiplicative subset of L. Also $1 \in S$ and $0 \notin S$. So by the Separation Lemma of [6], there is a prime element $p \in L$ such that $t \nleqq p$ for all $t \in S$. We show that $p \leqq p_x$. If $p \nleqq p_x$, then there is a compact element $b \in L$ such that $b \leqq p$ and $b \nleqq p_x$. As $b \nleqq p_x$, it follows that $b \in S$ and so $b \nleqq p$, a contradiction. Therefore $p \leqq p_x$ and hence by hypothesis $p = p_x$. Again since $a \in S$, we have $a \nleqq p = p_x$. Thus, for every compact element $a \in L$, $a \nleqq \sqrt{x}$ if and only if $a \nleqq p_x$. Consequently, $\sqrt{x} = p_x$. Now we prove that x is primary. Suppose a and b are compact elements of L such that $ab \leqq x$ and $a \nleqq x$. Then b is non-prime to x, so $b \leqq p_x = \sqrt{x}$. As b is compact we have $b \leqq a_1 \vee \cdots \vee a_n$, where $a_i^{n_i} \leqq x$ for some $n_i \in \mathbf{Z}^+$ $(i = 1, 2, \ldots, n)$. Again by Lemma 1 of [2, page 336] $b^m \leqq x$ for some $m \in \mathbf{Z}^+$ and hence x is primary. This completes the proof of the lemma.

We now characterize regular lattices as follows.

THEOREM 7. *The following statements on L are equivalent.*

(i) *L is a regular lattice.*

(ii) *Every primary element of L is a maximal element.*

(iii) *Every primary element of L is a minimal prime element.*

PROOF. (i) $\Rightarrow$ (ii). Suppose (i) holds. Let p be a primary element of L. Suppose p is not a maximal element. Then $p < q$ for some maximal element q of L. As L is compactly generated, there is a compact element $b \in L$ such that $b \leqq q$ and $b \nleqq p$. Since L is regular, b is a complemented element. Let b' be a complement of b. As $bb' = 0 \leqq p$ and p is primary, we have either $b' \leqq p$ or $b^n \leqq p$. But b is idempotent and so $b' \leqq p < q$. Consequently $q = 1$, a contradiction. Therefore, every primary element is a maximal element.

(ii) $\Rightarrow$ (iii). Obvious.

(iii) $\Rightarrow$ (i). Suppose (iii) holds. Then by (iii), every prime element is a minimal prime element and so by Lemma 10, Lemma 14 and by (iii),

every element is the meet of prime elements and hence every element is an idempotent element. Consequently L is reduced and hence, by Theorem 5, L is regular. This completes the proof of the theorem.

It is well-known that a commutative ring R with identity is Noetherian regular if and only if R is semiprime and every radical ideal has a unique representation as an intersection of prime ideals (see Theorem 2 of [5]). Our last theorem gives an abstract version of the above result.

THEOREM 8. *L is Noetherian regular if and only if L is reduced and every radical element has a unique representation as a meet of prime elements.*

PROOF. Suppose L is Noetherian regular. Then by Remark 3 and Theorem 2, L is a finite Boolean algebra. Obviously, prime elements coincide with maximal elements in L and also it can be easily shown that in a finite Boolean algebra, every element has a unique representation as the meet of a finite number of maximal elements.

Conversely, assume that L is reduced, and every radical element has a unique representation as a meet of prime elements. First, we show that every nonzero radical element is a complemented element. Let a be a nonzero radical element. Put $b = \wedge\{p \in L | a \not\leqq p,\ p \text{ is a prime element}\}$. As L is reduced, $ab = \wedge\{p \in L | p \text{ is a prime element}\} = 0$. We claim that $a \vee b = 1$. If $a \vee b < 1$, then there is a prime element p_0 such that $a \vee b \leqq p_0$. Then $b = \wedge\{p \in L | a \not\leqq p, p \text{ is a prime element}\} = \wedge\{p \in L | a \not\leqq p, p \text{ is a prime element}\} \wedge \wedge p_0$. So that b has two representations, a contradiction. Therefore $a \vee b = 1$. Thus every nonzero radical element is complemented.

Now we show that every nonzero element is complemented. Let x be a nonzero element. Since $\sqrt{x}$ is a complemented element, by Lemma 6, $\sqrt{x}$ is a compact element. Again since $\sqrt{x}$ is compact, we have $\sqrt{x} \leqq a_1 \vee \cdots \vee a_n$, where $a_i^{n_i} \leqq x$ for some $n_i \in \mathbf{Z}^+$ $(i = 1, 2, \ldots, n)$; so that $(\sqrt{x})^k \leqq x$ for some $k \in \mathbf{Z}^+$. As $\sqrt{x}$ is idempotent, we get $\sqrt{x} \leqq x$ and therefore $x = \sqrt{x}$ is a complemented element. Consequently, by Lemma 6, L is a regular lattice in which every element is compact. Hence L is Noetherian regular. This completes the proof of the theorem.

REFERENCES

[1] ANDERSON, D. D., Abstract commutative ideal theory without chain condition, *Algebra Universalis* **6** (1976), 131–145. *MR* **54** #7332

[2] BIRKHOFF, G., *Lattice theory*, Third edition, Amer. Math. Soc. Colloquium Publications, Vol. XXV, American Mathematical Society, Providence, R. I., 1967. *MR* **37** #2638

[3] CRAWLEY, P. and DILWORTH, R. P., *Algebraic theory of lattices*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973. *Zbl* **494**.06001

[4] DILWORTH, R. P., Abstract commutative ideal theory, *Pacific J. Math.* **12** (1962), 481–498. *MR* **26** # 1333

[5] JAYARAM, C. and MANNEPALLI, V. L., Noetherian regular rings, *Indian J. Pure Appl. Math.* **20** (1989), 554–559. *MR* **90g**:13029

[6] THAKARE, N. K., MANJAREKAR, C. S. and MAEDA S., Abstract spectral theory. II. Minimal characters and minimal spectrums of multiplicative lattices, *Acta Sci. Math. (Szeged)* **52** (1988), 53–67. *MR* **89k**:06013

(Received October 18, 1991)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF IOWA
IOWA CITY, IA 52242
U.S.A.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF KWALUSENI
SWAZILAND

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 389–409

A BITOPOLOGICAL VIEW OF QUASI-UNIFORM COMPLETENESS. I

J. DEÁK

Abstract

In this series, we look for notions of quasi-uniform completeness and corresponding completions that are symmetric in the following sense: (i) $\mathcal{U}$ is complete iff $\mathcal{U}^{-1}$ is so; (ii) the completion of $\mathcal{U}^{-1}$ is isomorphic with the conjugate of the completion of $\mathcal{U}$.

The present paper only contains introductory material. The main results will follow in Parts II and III.

The problem mentioned in the Abstract will be described in more detail in § 1. The reader is recommended to glance through § 0, because we use some unconventional notations and terminology. The notes in parentheses in the Contents should perhaps be read only after § 1.

CONTENTS

Part I

1991 *Mathematics Subject Classification*. Primary 54E15, 54D35; Secondary 54E55, 54G20.

Key words and phrases. Quasi-uniformity, bitopology, complete (in several senses), (half-)extension, completion, open/round/Cauchy filter (pair), linked filter pair.

Research supported by the Hungarian National Foundation for Scientific Research Grant No. 2114.

Part II

INDEX

(The notations containing no constant letter can be found at the end of the index.)

§ 0 Preliminaries

A. Notations and terminology

0.1 *Set theoretic notations.* There is in general a fundamental set X, which is assumed, as a rule, to be non-empty. For $\mathfrak{a} \subset \exp X$, $\sec \mathfrak{a} = \sec_X \mathfrak{a}$ consists of the subsets of X meeting each element of $\mathfrak{a}$. For $\mathfrak{a}, \mathfrak{b} \subset \exp X$, $\mathfrak{a}(\cap)\mathfrak{b} = \{A \cap B \colon A \in \mathfrak{a}, B \in \mathfrak{b}\}$. In order to get along with as few parentheses as possible, let us agree that (i) $\times$ precedes the other set theoretic operations; (ii) $\cup, \cap$ and $\setminus$ precede $|$, which denotes the restriction (i.e. trace on a subset) of structures, relations, filters, etc.; (iii) ${}^bA^c_a$ (with A, a, b, c standing for any

letter, number or other symbol) is to be understood as $\left({}^{b}(A_a)\right)^{c}$ (when b and c denote operations).

0.2 *Filters and filter pairs.* A *filter* is usually a proper filter; it will be explicitly stated when $\exp X$ is allowed as a filter. A *filter (base) pair* is an ordered pair of filters (filter bases) in X. For $\mathfrak{a} \subset \exp X$, $\operatorname{fil}\mathfrak{a} = \operatorname{fil}_X \mathfrak{a}$ is the filter generated by $\mathfrak{a}$ (or $\exp X$ if $\mathfrak{a}$ is not centred). If a filter (base) pair is denoted by $\mathfrak{f}^0$ then it is understood that $\mathfrak{f}^0 = (\mathfrak{f}^{-1}, \mathfrak{f}^1)$, and similarly with other letters (also with letters having indices); conversely, if systems $\mathfrak{f}^{-1}$ and $\mathfrak{f}^1$ are defined then $(\mathfrak{f}^{-1}, \mathfrak{f}^1)$ is denoted by $\mathfrak{f}^0$. Given centred systems $\mathfrak{a}^{-1}$ and $\mathfrak{a}^1$, $\operatorname{Fil}\mathfrak{a}^0 = \operatorname{Fil}_X \mathfrak{a}^0 = (\operatorname{fil}\mathfrak{a}^{-1}, \operatorname{fil}\mathfrak{a}^1)$ is the *filter pair generated by* $\mathfrak{a}^0$. We say that a filter base pair $\mathfrak{f}^0$ has a property defined for filter pairs if $\operatorname{Fil}\mathfrak{f}^0$ has the property. The filter pair $\mathfrak{f}^0$ is *finer* than the filter pair $\mathfrak{g}^0$ ($\mathfrak{g}^0$ is *coarser* than $\mathfrak{f}^0$) if $\mathfrak{f}^i \supset \mathfrak{g}^i$ $(i = \pm 1)$. A filter pair is *minimal/maximal* with respect to a given property if it has the property, but no strictly coarser/finer filter pair has it. If $\mathfrak{F}$ is a family of filters or filter pairs then $\mathfrak{F}^{\mathrm{m}}$ denotes the family of the minimal elements of $\mathfrak{F}$, and $\mathfrak{F}^{\mathrm{M}}$ that of the maximal ones. A filter pair $\mathfrak{f}^0$ is *linked* if $S_{-1} \cap S_1 \neq \emptyset$ whenever $S_i \in \mathfrak{f}^i$ (equivalently: $\mathfrak{f}^{-1}(\cap)\mathfrak{f}^1$ is a filter). For a filter pair $\mathfrak{f}^0$, $\mathfrak{f}^{\times} = \{S_{-1} \times S_1 \colon S_i \in \mathfrak{f}^i \ (i = \pm 1)\}$. (Similarly with other letters.) If $K \in \mathfrak{f}^{\times}$ then we write $K = K_{-1} \times K_1$; conversely, given $K_i \in \mathfrak{f}^i$ $(i = \pm 1)$, $K_{-1} \times K_1$ is denoted by K. (Again, similarly with any letter replacing K.) For families $\mathfrak{F}$ and $\mathfrak{G}$ of filters (filter pairs), $\mathfrak{G}$ *overlays* $\mathfrak{F}$ if for any $\mathfrak{f} \in \mathfrak{F}$ ($\mathfrak{f}^0 \in \mathfrak{F}$) there is a $\mathfrak{g} \in \mathfrak{G}$ ($\mathfrak{g}^0 \in \mathfrak{G}$) coarser than $\mathfrak{f}$ (than $\mathfrak{f}^0$); $\mathfrak{G}$ *overlaps* $\mathfrak{F}$ if for any $\mathfrak{f} \in \mathfrak{F}$ ($\mathfrak{f}^0 \in \mathfrak{F}$) there is a $\mathfrak{g} \in \mathfrak{G}$ ($\mathfrak{g}^0 \in \mathfrak{G}$) such that $(\mathfrak{f}, \mathfrak{g})$ is linked ($(\mathfrak{f}^i, \mathfrak{g}^i)$ is linked for $i = \pm 1$). $\mathfrak{f}^0 \mid S = (\mathfrak{f}^{-1} \mid S, \mathfrak{f}^1 \mid S)$ is the *trace* of $\mathfrak{f}^0$ on $S \subset X$.

0.3 *(Bi)topological spaces.* In a bitopological space $(X; \mathcal{T}^{-1}, \mathcal{T}^1)$, cl^i denotes the $\mathcal{T}^i$-closure, and cl^{s} the $\sup\{\mathcal{T}^{-1}, \mathcal{T}^1\}$-closure. cl^i_S is the closure in the subspace S $(i = \pm 1, \mathrm{s})$; if a superspace of X is also given then the closures in it are denoted by Cl^i. The notations cl and Cl are used in topological spaces. The bitopology $(\mathcal{T}^{-1}, \mathcal{T}^1)$ is *regular* if each point has a $\mathcal{T}^i$-neighbourhood base consisting of $\mathcal{T}^{-i}$-closed sets $(i = \pm 1)$. The *$\mathcal{T}_0$-reflexion* of a bitopological space is defined just as for topological spaces: those points are identified that have the same neighbourhood filter pair (equivalently: the same $\sup\{\mathcal{T}^{-1}, \mathcal{T}^1\}$-neighbourhood filter). The filter pair $\mathfrak{f}^0$ in a bitopological space *converges/clusters* to x if $\mathfrak{f}^i$ $\mathcal{T}^i$-converges/clusters to x $(i = \pm 1)$; the following expressions will also be used: is *convergent*, is a *cluster* point of, has a cluster point. $\mathfrak{f}^0$ is *open* if $\mathfrak{f}^i$ is $\mathcal{T}^i$-open for $i = \pm 1$ (i.e. it has a base consisting of $\mathcal{T}^i$-open sets).

0.4 *Extensions of (bi)topological spaces.* The topological space $(Y, \mathcal{S})$ is an *extension* of the topological space $(X, \mathcal{T})$ ($\mathcal{S}$ is an extension of $\mathcal{T}$) if $(X, \mathcal{T})$ is a dense subspace of $(Y, \mathcal{S})$. The *trace filter* $\mathfrak{f}(a)$ of $a \in Y$ is the trace on X

of the $\mathcal{S}$-neighbourhood filter of a. If $x \in X$ then $\mathfrak{f}(x)$ is the $\mathcal{T}$-neighbourhood filter of x; for any $a \in Y$, $\mathfrak{f}(a)$ is $\mathcal{T}$-open. Conversely, if we prescribe $\mathcal{T}$-open filters $\mathfrak{f}(a)$ for each $a \in Y \supset X$ then there are extensions of $\mathcal{T}$ with just these trace filters. (We shall also use other expressions: extension for $\mathfrak{f}(a)$, inducing $\mathfrak{f}(a)$, etc.) There exists a finest and a coarsest one among such extensions, called the *loose extension*, respectively the *strict extension* (for the given trace filters). The neighbourhood filter of $a \in Y$ is $\mathrm{fil}_Y\{S \cup \{a\} : S \in \mathfrak{f}(a)\}$ in the loose extension, $\mathrm{fil}_Y\{\{b : S \in \mathfrak{f}(b)\} : S \in \mathfrak{f}(a)\}$ in the strict extension.

The bitopological space $(Y; \mathcal{S}^{-1}, \mathcal{S}^1)$ is an *extension* of the bitopological space $(X; \mathcal{T}^{-1}, \mathcal{T}^1)$ if X is doubly dense in Y. (For a topological property P, a bitopological space, a subset of it, or a bitopological extension is *doubly P* if it is P in both topologies separately.) The *trace filter pair* of $a \in Y$, denoted by $\mathfrak{f}^0(a)$, is the trace on X of the neighbourhood filter pair of a, i.e. $\mathfrak{f}^i(a)$ is the trace filter of a in the extension $\mathcal{S}^i$ of $\mathcal{T}^i$. Each $\mathfrak{f}^0(a)$ is open; for $a \in X$, it coincides with the neighbourhood filter pair of a. Conversely, if we are given trace filter pairs with these two conditions then there are extensions belonging to them; the doubly loose is the finest, and the doubly strict the coarsest one. If there exist regular extensions for some trace filter pairs then there is a finest one among them, called *fine regular extension* ([De3] 2.1, 2.2).

The extension $(Y, \mathcal{S})$ of $(X, \mathcal{T})$ is *reduced* [Cs3] if $p \in Y \setminus X$, $a \in Y$, $p \neq a$ imply that p and a have different $\mathcal{S}$-neighbourhood filters; it is *weakly basic* (a stronger condition) if the same holds for the trace filters instead of the neighbourhood filters; it is *basic* if the trace filters of the new points are non-convergent and different. A basic extension is clearly weakly basic. Reduced and (weakly) basic extensions of bitopological spaces are defined similarly, with filter pairs instead of filters.

0.5 *Quasi-uniformities.* See [FL2] for fundamental information on quasi-uniformities and quasi-proximities. $\mathcal{U}^{\mathrm{t}}$ is the quasi-proximity induced by the quasi-uniformity $\mathcal{U}$; δ^{p} is the topology induced by the quasi-proximity δ; thus $\mathcal{U}^{\mathrm{tp}}$ denotes the topology of $\mathcal{U}$. The bitopology of $\mathcal{U}$ or δ is $(\mathcal{U}^{-\mathrm{tp}}, \mathcal{U}^{\mathrm{tp}})$, respectively $(\delta^{-\mathrm{p}}, \delta^{\mathrm{p}})$, where $\mathcal{U}^{-\mathrm{tp}} = (\mathcal{U}^{-1})^{\mathrm{tp}}$, $\delta^{-\mathrm{p}} = (\delta^{-1})^{\mathrm{p}}$. The (bi)topological notions in a quasi-uniform or quasi-proximity space (open filter (pair), convergence, clustering, closures etc.) are to be understood with respect to the induced (bi)topology. If a map f between the quasi-uniform spaces $(X, \mathcal{U})$ and $(Y, \mathcal{V})$ is quasi-uniformly continuous then we shall shortly say that f is *uniformly continuous*, or that f is $(\mathcal{U}, \mathcal{V})$-continuous. $\mathcal{U}^{\mathrm{s}}$ denotes the uniformity $\sup\{\mathcal{U}^{-1}, \mathcal{U}\}$. (Observe that $\mathcal{U}^{\mathrm{stp}} = \sup\{\mathcal{U}^{-\mathrm{tp}}, \mathcal{U}^{\mathrm{tp}}\}$.) A quasi-uniformity $\mathcal{U}$ will be called (i) T_0 if U^{tp} is T_0 (equivalently: $\mathcal{U}^{-\mathrm{tp}}$ is T_0, U^{stp} is T_0); (ii) T_1 if $\mathcal{U}^{\mathrm{tp}}$ is T_1 (equivalently: $\mathcal{U}^{-\mathrm{tp}}$ is T_1). The T_0 *reflexion* of a quasi-uniformity is defined in the same way as for bitopologies. The T_0 reflexion clearly commutes with taking the induced (bi)topology.

Three important quasi-uniformities on $\mathbf{R}$: $\mathcal{U}_{\mathbf{eu}}$ is the Euclidean unifor-

mity; $\mathcal{U}_{so}$ is the Sorgenfrey quasi-uniformity, i.e. $\{U_{(\varepsilon)}: \varepsilon > 0\}$ is a base for it, where $U_{(\varepsilon)}x = [x, x+\varepsilon[$; $\mathcal{U}_{se}$, the quasi-uniformity of semicontinuities, is defined similarly, with $U_{(\varepsilon)}x =]\leftarrow, x+\varepsilon[$.

0.6 *Filters and filter pairs in a quasi-uniform space.* Let $(X,\mathcal{U})$ be a quasi-uniform space, $\mathfrak{f}$ and $\mathfrak{g}$ filters, $\mathfrak{f}^0$ a filter pair in it. $\mathfrak{f}$ is *round* ([Kow], [Cs4]) if for any $S \in \mathfrak{f}$ there are $U \in \mathcal{U}$ and $T \in \mathfrak{f}$ such that $U[T] \subset S$. (Round filters are open.) $\mathfrak{f}^0$ is *round* [De3] if $\mathfrak{f}^i$ is $\mathcal{U}^i$-round ($i = \pm 1$). $\mathfrak{f}$ is *Cauchy* [SP] if for any $U \in \mathcal{U}$ there is an $x \in X$ with $Ux \in \mathfrak{f}$. $\mathfrak{f}^0$ is *Cauchy* [De3] if for any $U \in \mathcal{U}$ there is a $K \in \mathfrak{f}^\times$ with $K \subset U$. $\mathfrak{g}$ is a *cofilter* [Do3] of $\mathfrak{f}$ if $(\mathfrak{g}, \mathfrak{f})$ is Cauchy. $\mathfrak{f}$ is *D-Cauchy* [Do3] if it has a cofilter. The *envelope* of $\mathfrak{f}$ [Sam] (of $\mathfrak{f}^0$) is the finest round filter (filter pair) coarser than $\mathfrak{f}$ (than $\mathfrak{f}^0$); it will be denoted by $\mathfrak{f}^E$, respectively $\mathfrak{f}^{0E}$; $\mathfrak{f}$ or $\mathfrak{f}^0$ is allowed here to be only a filter base (pair). Clearly, $\mathfrak{f}^E = \{U[S]: S \in \mathfrak{f}, U \in \mathcal{U}\}$ and $\mathfrak{f}^{0E} = (\mathfrak{f}^{-1E^{-1}}, \mathfrak{f}^{1E})$, where E^{-1} denotes the $\mathcal{U}^{-1}$-envelope. If a superspace $(Y,\mathcal{V})$ of $(X,\mathcal{U})$ is given then E stands occasionally for the $\mathcal{V}$-envelope instead of the $\mathcal{U}$-envelope; this will be either explicitly mentioned, or clear from the context (e.g. because $\mathfrak{f}$ or $\mathfrak{f}^0$ is not in X). We introduce notations for some families of filters or filter pairs in a quasi-uniform space:

$\mathfrak{F}_C$ = the Cauchy filters,
$\mathfrak{F}_D$ = the D-Cauchy filters,
$\mathfrak{F}_O$ = the open Cauchy filters,
$\mathfrak{F}_R$ = the round Cauchy filters,
$\mathfrak{P}_C$ = the Cauchy filter pairs,
$\mathfrak{P}_O$ = the open Cauchy filter pairs,
$\mathfrak{P}_R$ = the round Cauchy filter pairs,
$\mathfrak{P}_L$ = the linked Cauchy filter pairs.

Given a family $\mathfrak{F}$ of filters, $\mathfrak{F}^n$ and $\mathfrak{F}^N$ consist of the non-convergent, respectively non-clustering elements of $\mathfrak{F}$, while $\mathfrak{F}^E = \{\mathfrak{f}^E: \mathfrak{f} \in \mathfrak{F}\}$. $\mathfrak{P}^n$, $\mathfrak{P}^N$ and $\mathfrak{P}^E$ are defined analogously for a family $\mathfrak{P}$ of filter pairs.

0.7 *Extensions of quasi-uniform spaces.* Let $(Y,\mathcal{V})$ be a superspace of $(X,\mathcal{U})$. We say that $\mathcal{V}$ is an *extension* of $\mathcal{U}$ if X is doubly dense in Y (i.e. if $(\mathcal{V}^{-tp}, \mathcal{V}^{tp})$ is an extension of $(\mathcal{U}^{-tp}, \mathcal{U}^{tp})$); $\mathcal{V}$ is a *half-extension* of $\mathcal{U}$ if X is $\mathcal{U}^{tp}$-dense in Y (i.e. if $\mathcal{V}^{tp}$ is an extension of $\mathcal{V}^{tp}$). An extension (or half-extension) of a quasi-uniformity is reduced, basic or weakly basic if the corresponding bitopological (or topological) extension has this property. (Caution! A quasi-uniform extension can be (weakly) basic without having the same property when regarded as a half-extension.)

There exist half-extensions of $\mathcal{U}$ for prescribed trace filters iff they are round; if so then the finest of them can be described as follows ([Cs4] 2.5): $\{V(f,U): f \in \Phi,\ U \in \mathcal{U}\}$ is a base for ${}^{0)}\mathcal{U}$ (called the *uniformly loose* half-extension for the given trace filters), where Φ denotes the family of all those

functions $f: Y \to \exp X$ for which $f(a) \in \mathfrak{f}(a)$ $(a \in Y)$, and $V(f, U)a = \{a\} \cup \cup U[f(a)]$. Similarly, there is an extension of $\mathcal{U}$ for prescribed trace filter pairs iff they are round and Cauchy; if so then there is a finest extension ${}^0\mathcal{U}$ for these trace filter pairs, where

$$\{V(f^{-1}, f^1, U): f^i \in \Phi^i \ (i = \pm 1),\ U \in \mathcal{U}\}$$

is a base for ${}^0\mathcal{U}$, Φ^i denotes the family of the functions $f: Y \to \exp X$ for which $f(a) \in \mathfrak{f}^i(a)$, and

$$a\, V(f^{-1}, f^1, U)\, b \text{ iff either } a = b \text{ or } \exists x \in f^1(a),\ \exists y \in f^{-1}(b), x\, U\, y$$

([De3] 3.3 and [De4] 6.1). ${}^{0)}\mathcal{U}$ induces the loose extension of $\mathcal{U}^{\mathrm{tp}}$ ([Cs4] 2.1), and ${}^0\mathcal{U}$ the fine regular extension of $(\mathcal{U}^{-\mathrm{tp}}, \mathcal{U}^{\mathrm{tp}})$ ([De3] 3.1).

Let now $\mathfrak{F}$ be a family of filters in $(X, \mathcal{U})$. Then ${}^{0)}\mathcal{U}(\mathfrak{F})$ denotes ${}^{0)}\mathcal{U}$ taken with

(1) $\quad Y \setminus X = \{\mathfrak{f} \in \mathfrak{F}: \mathfrak{f}$ is round, $\mathfrak{f}$ is not a $\mathcal{U}^{\mathrm{tp}}$-neighbourhood filter$\}$,

and $\mathfrak{f}(p) = p$ for $p \in Y \setminus X$. The notation ${}^*\mathcal{U}(\mathfrak{F})$ will be used in the same way, where ${}^*\mathcal{U}$ is some other construction yielding a half-extension for prescribed trace filters, possibly only under some additional assumption on the filters (in which case we shall only use this notation if the filters in (1) do satisfy this assumption). Analogously, if $\mathfrak{P}$ is a family of Cauchy filter pairs then ${}^0\mathcal{U}(\mathfrak{P})$ denotes ${}^0\mathcal{U}$ taken with

$$Y \setminus X = \{\mathfrak{f}^0 \in \mathfrak{P}: \mathfrak{f}^0 \text{ is round, } \mathfrak{f}^0 \text{ is not a neighbourhood filter pair}\},$$

and $\mathfrak{f}^0(p) = p$ $(p \in Y \setminus X)$. ${}^*\mathcal{U}(\mathfrak{P})$ will be used in the same way if $*$ denotes a quasi-uniform extension for prescribed trace filter pairs. Clearly, ${}^*\mathcal{U}(\mathfrak{F})$ is always a weakly basic half-extension, and ${}^*\mathcal{U}(\mathfrak{P})$ a weakly basic extension.

B. Notations used in the counterexamples

0.8 *Distances.* A real valued function d defined on a subset of X^2 is a *distance* [De3] on X if the Triangle Inequality

$$d(x, y) + d(y, z) \geqq d(x, z) \tag{1}$$

is satisfied in the sense that if $d(x, y)$ and $d(y, z)$ are both defined then so is $d(x, z)$, and (1) holds. If d is a distance then $\{U_{(\varepsilon)}: \varepsilon > 0\}$ is a base for a quasi-uniformity $\mathcal{U}(d)$, where

$$U_{(\varepsilon)} = U_{(\varepsilon)}(d) = \{(x, x): x \in X\} \cup \{(x, y): d(x, y) < \varepsilon\}.$$

For example, $\mathcal{U}_{\mathrm{eu}} = \mathcal{U}(d_{\mathrm{eu}})$, $\mathcal{U}_{\mathrm{se}} = \mathcal{U}(d_{\mathrm{se}})$ and $\mathcal{U}_{\mathrm{so}} = \mathcal{U}(d_{\mathrm{so}})$, where d_{eu} is the Euclidean distance on $\mathbb{R}$,

$$d_{\mathrm{se}}(x, y) = y - x,$$
$$d_{\mathrm{so}}(x, y) = y - x \quad \text{if} \quad x < y.$$

When a distance d already defined is used in the definition of another distance, a condition like

$$= d(x, y) \quad \textit{if}\ldots$$

is to be understood as follows:

$$= d(x, y) \quad \text{if} \quad d(x, y) \quad \textit{is defined and}\ldots.$$

For a distance d, let

$$\bar{d}(x, y) = \begin{cases} d(x, y) & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

$\mathcal{U}(\bar{d}) = \mathcal{U}(d)$. If d is a distance on X and e on Z then a distance $f = d \times e$ can be defined on $X \times Z$ as follows:

$$f((x, z), (y, w)) = \max\{\bar{d}(x, y), \bar{e}(z, w)\},$$

such that the left-hand side is not defined when either of the two distances on the right-hand side is not defined. $d^2 = d \times d$.

0.9 *Sets and filters in* $\mathbf{R}$. The set of the rationals is denoted by $\mathbf{Q}$. For $a > b$, $]a, b[$ means $]b, a[$.

$$-\mathbf{N} = \{-n : n \in \mathbf{N}\}, \quad 1/\mathbf{N} = \{1/n : n \in \mathbf{N}\}, \quad -1/\mathbf{N} = \{-1/n : n \in \mathbf{N}\},$$
$$\mathbf{R}_0 = \mathbf{R} \setminus \{0\}, \quad \mathbf{R}_{-1} =]\leftarrow, 0[, \quad \mathbf{R}_1 =]0, \rightarrow[,$$
$$\overline{\mathbf{R}}_{-1} =]\leftarrow, 0], \quad \overline{\mathbf{R}}_1 = [0, \rightarrow[.$$

$\mathfrak{e}(x)$ is the Euclidean neighbourhood filter of x; $\mathfrak{e}_0(x) = (\mathfrak{e}_{-1}(x), \mathfrak{e}_1(x))$ is the $\mathcal{U}_{so}$-neighbourhood filter pair of x. (Lower indices are used exceptionally in this notation.) $\mathfrak{e} = \mathfrak{e}(0)$, $\mathfrak{e}_0 = (\mathfrak{e}_{-1}, \mathfrak{e}_1) = \mathfrak{e}_0(0)$.

0.10 *Products of filters.* If $\mathfrak{f}$ is a filter, $\mathfrak{f}^0$ a filter pair in X, $\mathfrak{h}$ is a filter, $\mathfrak{h}^0$ a filter pair in Z then a filter $\mathfrak{f} \times \mathfrak{h}$ and filter pairs $\mathfrak{f}^0 \times \mathfrak{h}^0$, $\mathfrak{f}^0 \times \mathfrak{h}$ and $\mathfrak{f} \times \mathfrak{h}^0$ are defined in $X \times Z$ as follows:

$$\mathfrak{f} \times \mathfrak{h} = \operatorname{fil}\{F \times H : F \in \mathfrak{f}, H \in \mathfrak{h}\},$$
$$\mathfrak{f}^0 \times \mathfrak{h}^0 = (\mathfrak{f}^{-1} \times \mathfrak{h}^{-1}, \mathfrak{f}^1 \times \mathfrak{h}^1),$$
$$\mathfrak{f}^0 \times \mathfrak{h} = \mathfrak{f}^0 \times (\mathfrak{h}, \mathfrak{h}), \quad \mathfrak{f} \times \mathfrak{h}^0 = (\mathfrak{f}, \mathfrak{f}) \times \mathfrak{h}^0.$$

$\mathfrak{f}^2 = \mathfrak{f} \times \mathfrak{f}$, $(\mathfrak{f}^0)^2 = \mathfrak{f}^0 \times \mathfrak{f}^0$. The product of three filters and/or filter pairs is defined analogously, such that it is a filter pair whenever at least one of the factors is a filter pair.

0.11 *Points and filters in* $\mathbf{R}^2$ *and* $\mathbf{R}^3$. $0^2 = (0, 0)$, $0^3 = (0, 0, 0)$. If $x \in \mathbf{R}^2$ or $x \in \mathbf{R}^3$ then the coordinates of x are denoted by x', x'' (and x'''), and similarly with other letters. Given $x \in \mathbf{R}^2$, $\mathfrak{e}(x') \times \mathfrak{e}(x'')$ will also be written as $\mathfrak{e}^2(x', x'')$ or $\mathfrak{e}^2(x)$, and similarly for the products of other filters and filter pairs defined in 0.9. Analogous notations will be used for products of filters and/or filter pairs in $\mathbf{R}^3$.

§ 1 The problem

1.1 We are looking for a notion of quasi-uniform completeness that satisfies the conditions below, or at least most of them; cf. [De4] 12.6. (The last sentence in 12.6 b) is mistaken; for the last two lines of it, read: "there may appear in Y new non-convergent stable Cauchy filter pairs.")

a) It is *symmetric* in the sense that $\mathcal{U}$ is complete iff $\mathcal{U}^{-1}$ is so.

This condition is certainly satisfied if a quasi-uniformity is defined to be complete iff the filter pairs of some kind or other (having a property shared by all the neighbourhood filter pairs, such that if $(\mathfrak{f}^{-1}, \mathfrak{f}^1)$ has this property in $(X,\mathcal{U})$ then $(\mathfrak{f}^1, \mathfrak{f}^{-1})$ has it in $(X,\mathcal{U}^{-1})$) are convergent, or have a cluster point; such a completeness will be called *bitopological*. Most of the symmetric notions of completeness considered in this paper are bitopological. Similarly, a notion of quasi-uniform completeness is *topological* provided that it can be described with the $\mathcal{U}^{\mathrm{tp}}$-convergence or -clustering of certain filters.

b) For uniformities, it coincides with the usual completeness.

c) A closed subspace of a complete space is complete.

For non-symmetric notions, the meaning of "closed" is clear: it has to be understood in the topology $\mathcal{U}^{\mathrm{tp}}$. For symmetric notions, the appropriate definition of closedness runs as follows: a subset A in the bitopological space $(X; \mathcal{T}^{-1}, \mathcal{T}^1)$ is *quasi-closed* ([Da]; see also [De], [De2]) if it is the intersection of a $\mathcal{T}^{-1}$-closed and a $\mathcal{T}^1$-closed set (equivalently $A = \mathrm{cl}^{-1}A \cap \mathrm{cl}^1 A$). Assume namely that $\mathcal{U}$ is complete in the sense that each filter pair having some property P is convergent (clustering), and let A be a quasi-closed subset of X; trying to prove the completeness of A, take a filter pair $\mathfrak{f}^0$ in A with property P; now if $\mathrm{Fil}_X\, \mathfrak{f}^0$ also has property P then it $\mathcal{U}$-converges ($\mathcal{U}$-clusters) to some $x \in X$, and the quasi-closedness of A implies $x \in A$. Thus c) holds with quasi-closedness for bitopological notions of completeness, assuming P is good enough to allow the conclusion above that $\mathrm{Fil}_X\, \mathfrak{f}^0$ has property P.

d) $\mathcal{U}_{\mathrm{so}}$ is complete, $\mathcal{U}_{\mathrm{so}}|\mathbf{R}_0$ is incomplete.

e) $\mathcal{U}_{\mathrm{se}}$ is complete, $\mathcal{U}_{\mathrm{se}}|\mathbf{R}_0$ is incomplete.

These conditions are required on the analogy of $\mathcal{U}_{\mathrm{eu}}$ being complete and $\mathcal{U}_{\mathrm{eu}}|\mathbf{R}_0$ incomplete; $\mathcal{U}_{\mathrm{so}}$ and $\mathcal{U}_{\mathrm{se}}$ are the natural non-symmetric counterparts of $\mathcal{U}_{\mathrm{eu}}$, and it is difficult to tell which is the more natural one: $(\mathcal{U}_{\mathrm{se}}^{-\mathrm{tp}}, \mathcal{U}_{\mathrm{se}}^{\mathrm{tp}})$ takes the role of the Euclidean topology in the embedding theorem for completely regular spaces (see [B]), while $\mathcal{U}_{\mathrm{so}}$ has better separation properties.

One could also require that a space should be "compact" iff it is "complete" and "precompact", or at least that compact spaces should be complete. The meaning of "precompact" is not clear here, and that of "compact" is even less so in the bitopological case.

1.2 By a *completion* we mean a prescription that assigns to each quasi-uniformity $\mathcal{U}$ a complete extension ${}^*\mathcal{U}$ (only half-extension in the non-symmetric case). In the case of a (bi)topological notion of completeness, the

filters (filter pairs) in question clearly have to be overlaid or overlapped (according as convergence or clustering is assumed) by the trace filters (filter pairs). If the notion of completeness depends only on the T_0 reflexion of the quasi-uniformity then we may assume (discarding the superfluous points) that ${}^*\mathcal{U}$ is a reduced extension, respectively half-extension, of $\mathcal{U}$. Some authors prefer to consider only T_0 spaces, or they define a completion to be a complete (half-)extension of the T_0 reflexion rather than of $\mathcal{U}$ itself.

We would like the conditions below to hold. The loose notation ${}^*X$ will be used for the fundamental set of ${}^*\mathcal{U}$. (The same convention applies to other symbols instead of $*$.)

f) If $\mathcal{U}$ is a uniformity then ${}^*\mathcal{U}$ is its usual completion.

g) ${}^*\mathcal{U}^{-1} = {}^*(\mathcal{U}^{-1})$.

More precisely, g) means that there is an isomorphism f between the two quasi-uniformities such that $f(x) = x$ $(x \in X)$. A completion satisfying g) will be called *bitopological.* Similarly, f) also means the existence of an isomorphism. A completion will be called *(weakly) basic* if ${}^*\mathcal{U}$ is always a (weakly) basic extension of $\mathcal{U}$ (respectively a (weakly) basic half-extension for non-symmetric notions of completeness). Assuming b), any weakly basic bitopological completion satisfies f). A completion is *natural* if any isomorphism between two spaces can be extended to an isomorphism between their completions. All completions in this paper are natural and weakly basic.

h) $({}^*X, {}^*\mathcal{U})$ is a *complete hull* of $(X, \mathcal{U})$.

This means that no proper subspace of ${}^*X$ containing X can be complete. A weaker assumption: if $\mathcal{U}$ is complete then ${}^*\mathcal{U} = \mathcal{U}$.

i) ${}^*\mathcal{U}$ is a *finest* complete extension of $\mathcal{U}$.

More precisely, if $(Z, \mathcal{V})$ is another complete extension then there exists a $({}^*\mathcal{U}, \mathcal{V})$-continuous map f such that $f(x) = x$ $(x \in X)$. We cannot speak about the finest extension, because there may exist different, even non-isomorphic, finest ones. It would be an essentially weaker assumption that ${}^*\mathcal{U}$ is the finest one among the complete (half-)extensions belonging to some trace filters or filter pairs. [E.g. ${}^{0)}\mathcal{U}(\{\{X\}\})$ is complete in any conceivable topological notion of completeness; cf. [CnH] Construction 1.] One could require several stronger versions of i), e.g. that any uniformly continuous map from $(X, \mathcal{U})$ into a complete space can be extended to a uniformly continuous map from $({}^*X, {}^*\mathcal{U})$ ("extension theorem for maps"), or that the extension of the identity in i) has some additional property (injective; an embedding; surjective if $\mathcal{V}$ is a basic (half-)extension; essentially unique), or that ${}^*\mathcal{U}$ is determined by i), i.e. it is basic, and if $\mathcal{V}$ is another finest complete basic (half-)extension then there is an isomorphism f between ${}^*\mathcal{U}$ and $\mathcal{V}$ such that $f(x) = x$ $(x \in X)$. If this last condition holds for a quasi-uniformity $\mathcal{U}$ (i.e. it is not assumed that each quasi-uniformity has a completion) then we shall say that $\mathcal{U}$ has a unique finest basic complete (half-)extension. The extension theorem for maps is sometimes valid ([Cs2] (16.76), [Cs5] 3.5); in other

cases, it holds only for a special class of spaces (completion is defined only for certain quasi-uniformities, and $\mathcal{V}$ is also taken from a special, possibly different, class, see [Do3] Th. 4, [Do5] Th. 2, [Cs6] § 3).

Assuming that c) holds, i) is equivalent to the following stronger statement:

i′) If $(X,\mathcal{U})$ is a subspace of the complete space $(Z,\mathcal{V})$ then there is a $({}^*\mathcal{U},\mathcal{V})$-continuous map f such that $f(x)=x$ $(x\in X)$.

Requiring f in i′) to be an embedding, we obtain an even stronger condition equivalent to:

i″) If $(X,\mathcal{U})$ is a subspace of $(Z,\mathcal{V})$ then there is an embedding f of $({}^*X,{}^*\mathcal{U})$ into $({}^*Z,{}^*\mathcal{V})$ such that $f(x)=x$ $(x\in X)$.

Doitchinov [Do4] regards i″) as a very important property of a reasonable notion of completion. This condition is, however, too strong, since it does not allow a space to have two different complete basic (half-)extensions with the same system of trace filters or filter pairs. We only have one notion of completion defined for all quasi-uniform spaces that satisfies i″) (see ${}^{\mathrm{L}}\mathcal{U}$ later; cf. [Kr], [Cs2], [Sal], [Kr2], [LF], [FL2], [De4]). The completion satisfying i″) introduced in [Do4] is defined only for a special class of spaces; extending the construction to a somewhat larger class (see [De7]), i″) is lost.

1.3 We shall also look for notions of completeness and completions for quasi-proximities, such that conditions analogous to a) to i) above are satisfied. In terms of totally bounded quasi-uniformities, this means that completeness has to be defined only for totally bounded spaces, and ${}^*\mathcal{U}$ has also to be totally bounded. Given a notion of completeness for quasi-uniformities, it yields one for quasi-proximities; but a completion for quasi-uniformities does not necessarily yield one for quasi-proximities, since it can happen that $\mathcal{U}$ is totally bounded, but ${}^*\mathcal{U}$ is not so, while ${}^*\mathcal{U}^{\mathrm{t}}$ is not complete (see ${}^{\mathrm{S}}\mathcal{U}$ later).

§ 2 Non-symmetric notions of completeness

2.1 In this section we look over the non-symmetric notions of completeness from the literature, ignoring the ones that do not satisfy b) ([FL], [Cs5] §§ 1–2, [Sz]). Where not stated otherwise, properties of filters are to be understood with respect to $\mathcal{U}$ and $\mathcal{U}^{\mathrm{tp}}$. A quasi-uniformity is

SP-complete ("complete" in [SP], "convergence complete" in [FL2]; present terminology in [Cs5]) if each Cauchy filter is convergent;

MN-complete ("complete" in [MN]) if each Cauchy filter has a cluster point (equivalently: each Cauchy ultrafilter is convergent);

FN-complete ("almost complete" in [FN]) if each open Cauchy filter has a cluster point (equivalently: each maximal open Cauchy filter is convergent);

W-complete ("P-complete" in [W]) if each round Cauchy filter has a cluster point (equivalently: each maximal round Cauchy filter is convergent);

D-complete ("complete" in [Do3]–[Do6]; present terminology in [Kp] and several other papers) if each D-Cauchy filter is convergent;

K-complete (based on an idea from [Kp], introduced in [FH] as "strongly *D*-complete") if each $\mathcal{U}^{-1}$-D-Cauchy filter has a $\mathcal{U}^{tp}$-cluster point (equivalently: each $\mathcal{U}^{-1}$-D-Cauchy ultrafilter is $\mathcal{U}^{tp}$-convergent);

Cs-complete (an equivalent notion is called "complete" in [Cs], "*C*-complete" in [CrH]; "half-complete" with the present definition in [De5]) if the second member of any linked Cauchy filter pair is convergent.

On the analogy of MN-, FN- and W-completeness, we could, but do not, consider spaces in which each (open/round) D-Cauchy filter has a cluster point, respectively the second member of any open/round linked Cauchy filter pair has one; it is namely not the purpose of this paper to introduce new non-symmetric notions of completeness. It does not change the definition of Cs-completeness if only the existence of a cluster point is required: if $\mathfrak{f}^0$ is a linked Cauchy filter pair then so is $(\mathfrak{f}^{-1}(\cap)\mathfrak{f}^1, \mathfrak{f}^{-1}(\cap)\mathfrak{f}^1)$, and then $\mathfrak{f}^1$ converges to any cluster point of $\mathfrak{f}^{-1}(\cap)\mathfrak{f}^1$. (Cf. [Cs] (15.49) and [CrH] p. 42.)

2.2 The next table shows some properties of the above notions; and of the corresponding completions. a) never holds, b) always does; e) and g) are irrelevant for non-symmetric notions (as to e): $\mathcal{U}_{se}$ is a good space only from the bitopological point of view). Some comments and remarks will follow the table (including the explanation of the symbols $\mathfrak{F}$ and $\mathfrak{P}$ in the last two lines).

completeness	c)	d)	completion	f)	h)	i)
SP	+	−	$^{SP}\mathcal{U} = {}^{0)}\mathcal{U}(\mathfrak{F}_C^n)$	−	−	+
MN	+	−	$^{MN}\mathcal{U} = {}^{0)}\mathcal{U}(\mathfrak{F}_C^{MnE})$	−	−	+
FN	−	−	$^{FN}\mathcal{U} = {}^{0)}\mathcal{U}(\mathfrak{F}_O^{MnE})$	−	−	+
W	−	+	$^{W}\mathcal{U} = {}^{0)}\mathcal{U}(\mathfrak{F}_R^{Mn})$	−	+	+
D	+	+	$^{D}\mathcal{U} = {}^{0)}\mathcal{U}(\mathfrak{F}_D^n)$	−	−	+
K	+	−	?			
Cs	+	−	$^{0)}\mathcal{U}(\mathfrak{F})$	−	−	+
			$^{0}\mathcal{U}(\mathfrak{P})$	+	−	−

To SP: $^{SP}\mathcal{U}$ is given in [Cs5]. i) (in fact, the extension theorem for maps) is proved in [Cs5] 3.5.

To MN: [W] § 6 introduces $^{MN}\mathcal{U}$, with the unessential difference that the trace filters are taken with multiplicity (the non-convergent Cauchy ultrafilters, and not their envelopes, are the new points). A worse construction (essentially $^{0)}\mathcal{U}(\mathfrak{F}_C^{NE})$, but again with multiplicity) is considered in [FL2] 3.40. We prove that $^{MN}\mathcal{U}$ is indeed MN-complete (since the construction in

[W] is within the proof of a theorem that contains additional assumptions): let $\mathfrak{f}$ be a Cauchy ultrafilter in ${}^{\mathrm{MN}}X$; if $X \in \mathfrak{f}$ then $\mathfrak{f} \mid X$ is Cauchy in $(X, \mathcal{U})$ (the proof given in [Cs5] 3.1 with ${}^{\mathrm{SP}}\mathcal{U}$ works for uniformly loose extensions with arbitrary systems of C a u c h y trace filters), thus if $\mathfrak{f} \mid X$ is not $\mathcal{U}^{\mathrm{tp}}$-convergent then $(\mathfrak{f} \mid X)^{\mathrm{E}} \in \mathfrak{F}_{\mathsf{C}}^{\mathrm{MnE}}$, implying that $\mathfrak{f}$ converges to a new point in ${}^{\mathrm{MN}}X$. The extension theorem for maps can be proved in the same way as in the case of ${}^{\mathrm{SP}}\mathcal{U}$; one has only to observe that if $\mathfrak{f}$ is an ultrafilter then so is $f(\mathfrak{f})$, and $f(\mathfrak{f})^{\mathrm{E}} \subset f(\mathfrak{f}^{\mathrm{E}})$. Concerning h), see Example 2.3 c) (which deals with h) for SP-completeness, too).

To FN: If $\mathfrak{f}$ is an open Cauchy filter in ${}^{\mathrm{FN}}X$ then $\mathfrak{f} \mid X$ is $\mathcal{U}$-open and $\mathcal{U}$-Cauchy (just as above), so there is a maximal open Cauchy filter finer than it that is either convergent in X, or it ${}^{\mathrm{FN}}\mathcal{U}^{\mathrm{tp}}$-converges to a new point; thus $\mathfrak{f} \mid X$, and also $\mathfrak{f}$, has a cluster point in ${}^{\mathrm{FN}}X$, i.e. ${}^{\mathrm{FN}}\mathcal{U}$ is FN-complete indeed. Concerning h), see again Example 2.3 c). i) is clear: Let $(Y, \mathcal{V})$ be an FN-complete extension, $f(x) = x$ $(x \in X)$. For $p \in {}^{\mathrm{FN}}X \setminus X$, $\mathfrak{f}(p)$ is of the form $\mathfrak{h}^{\mathrm{E}}$ with $\mathfrak{h} \in \mathfrak{F}_{\mathsf{O}}^{\mathrm{Mn}}$. As X is dense in Y, there is a maximal $\mathcal{V}^{\mathrm{tp}}$-open filter $\mathfrak{h}'$ such that $\mathfrak{h}' \mid X = \mathfrak{h}$. The $\mathcal{V}$-envelope of $\mathfrak{h}$ is coarser than $\mathfrak{h}'$, thus $\mathfrak{h}'$ is Cauchy, and so it is convergent. Let now $f(p)$ be any of the limit points of $\mathfrak{h}'$; then the filter base $\mathfrak{f}(p)$ $\mathcal{V}^{\mathrm{tp}}$-converges to $f(p)$. Now f is $({}^{\mathrm{FN}}\mathcal{U}^{\mathrm{tp}}, \mathcal{V}^{\mathrm{tp}})$-continuous, which implies uniform continuity, since f is an extension of a uniformly continuous map (namely the identity) to a map from a uniformly loose half-extension (cf. [Cs5] 2.3). i′) is, however, false (hence the extension theorem for maps cannot hold either):

Example. Let $X =]0, 1] \times \{0\}$, $X_* =]0, 1] \times [0, 1]$, $\mathcal{U}_* = \mathcal{U}(d_{\mathrm{eu}} \times d_0) \mid X_*$ where the distance d_0 is defined on $\mathbf{R}$ by

$$d_0(x, y) = |y - x| \quad \text{if} \quad y \neq 0. \tag{1}$$

Now $(X, \mathcal{U}) = (X, \mathcal{U}_* \mid X)$ is a subspace of the FN-complete space $(Y, \mathcal{V}) = ({}^{\mathrm{FN}}X_*, {}^{\mathrm{FN}}\mathcal{U}_*)$. Assume that the identity of X has an $({}^{\mathrm{FN}}\mathcal{U}, \mathcal{V})$-continuous extension f, and consider a maximal open filter $\mathfrak{f}$ in X finer than $\mathfrak{e}^2 \mid X$. Let $p \in {}^{\mathrm{FN}}X \setminus X$ be the corresponding point. As the filter base $\mathfrak{f}$ is not convergent in X_*, we have $f(p) \in {}^{\mathrm{FN}}X_* \setminus X_*$, implying that $\mathfrak{f}$ is finer than the envelope of a maximal $\mathcal{U}^{\mathrm{tp}}$-open filter, a contradiction, since such a filter can have no trace on X. □

To W: It is proved in [W] that ${}^{\mathrm{W}}\mathcal{U}$ is W-complete. ${}^{\mathrm{W}}\mathcal{U}$ is the unique finest basic W-complete half-extension of $\mathcal{U}$ (the proof is straightforward). i′) is false: $\mathcal{U}$ and $\mathcal{U}_*$ from the above example will do again (consider now the minimal round Cauchy filter $\mathfrak{f} = \mathfrak{e}^2 \mid X$; or check that $\mathcal{F}_{\mathsf{O}}^{\mathrm{MnE}} = \mathcal{F}_{\mathsf{R}}^{\mathrm{Mn}}$ in X as well as in X_*).

To D: See [Cs6] 2.1 and 3.2. Concerning h), see Example 2.3 b).

To K: Strictly speaking, K-completeness is not a topological notion in the sense of 1.1, since the $\mathcal{U}^{\mathrm{tp}}$-neighbourhood filters are not always $\mathcal{U}^{-1}$-D-

Cauchy. This shortcoming can be remedied by considering a more complicated class of filters, namely those that are either $\mathcal{U}^{-1}$-D-Cauchy or $\mathcal{U}^{\mathrm{tp}}$-convergent (see [CrH] for a similar technical modification of the property $\mathcal{U}^{\mathrm{s}}$-Cauchy). There exists a K-complete half-extension with only one new point (see in 1.2 after i)); the question mark in the table means that we do not know whether there is a less crude K-completion, e.g. one in which the trace filters of the new points are $\mathcal{U}$-envelopes of non-$\mathcal{U}^{\mathrm{tp}}$-convergent $\mathcal{U}^{-1}$-D-Cauchy ultrafilters. ${}^{0)}\mathcal{U}$ with these trace filters will not do: let $X =]0,1] \times \times [0,1]$, $\mathcal{U} = \mathcal{U}(d^2_{\mathrm{eu}}) \mid X$. For each $p \in [0,1] \times \{0\}$, $\mathfrak{f}(p) = \mathfrak{e}^2(p) \mid X$ belongs to the class of filters in question. Denote by Y the fundamental set of the extension. Now

$$\mathfrak{f}^0 = \mathrm{Fil}_Y(\mathfrak{e}^2|]0,1] \times \{0\}, \mathfrak{e}^2|X)$$

is a Cauchy filter pair, but $\mathfrak{f}^{-1}$ has no ${}^{0)}\mathcal{U}^{\mathrm{tp}}$-cluster point.

To Cs: $\mathfrak{P}$ is the family of the linked round Cauchy filter pairs whose second member is not convergent; $\mathfrak{F}$ consists of these second members. To prove that ${}^{0)}\mathcal{U}(\mathfrak{F})$ is Cs-complete, take a linked Cauchy filter pair $\mathfrak{f}^0$ in ${}^{0)}X$. Now if $\mathfrak{f}^1$ has a trace on ${}^{0)}X \setminus X$ then this trace is of the form $\mathrm{fil}\{\{p\}\}$, implying $\mathfrak{f}^{-1} = \mathrm{fil}\{\{p\}\}$, so, $\mathfrak{f}^0$ being Cauchy, $\mathfrak{f}^1$ converges to p. Otherwise, $\mathfrak{f}^0 \mid X$ is linked and Cauchy, and the proof can be concluded in the usual way. ${}^{0)}\mathcal{U}(\mathfrak{F})$ is not a Cs-complete hull, since it can occur that $\mathfrak{F}$ is overlayed by a proper subfamily of it (see Examples 2.3 a) and b)), and ${}^{0)}\mathcal{U}$ taken with such a subfamily is also Cs-complete (the proof is the same). The extension theorem for maps is valid: for $p \in {}^{0)}X \setminus X$, let $\mathfrak{f}^0(p)$ be the round linked Cauchy filter pair for which $\mathfrak{f}^1(p)$ is the trace filter of p; now $\mathfrak{g}^0 = f(\mathfrak{f}^0(p))$ is also linked and Cauchy; hence $\mathfrak{g}^1$ converges to some point $f(p)$, and the uniform continuity of the map f defined this way can be proved as in the case of FN-completeness.

${}^0\mathcal{U}(\mathfrak{P})$ is also Cs-complete: If $\mathfrak{f}^0$ is a linked round Cauchy filter pair in ${}^0X$ then $\mathfrak{f}^0 \mid X$ has the same properties (to prove that it is linked, take ${}^0\mathcal{U}^{\mathrm{itp}}$-open sets $S_i \in \mathfrak{f}^i$, and pick $a \in S_{-1} \cap S_1$; the trace filter pair $\mathfrak{f}^0(a)$ is linked, and $S_i \cap X \in \mathfrak{f}^i(a)$, so $S_{-1} \cap S_1 \cap X \neq \emptyset$). Now $\mathfrak{f}^1 \mid X$ is ${}^0\mathcal{U}^{\mathrm{tp}}$-convergent, which implies that $\mathfrak{f}^1$ is convergent, too, provided that the topological extension is strict and $\mathfrak{f}^1$ is open ([Cs6] 1.1); but a quasi-uniform extension with linked trace filter pairs is always doubly strict (see [De4] Theorem 11.2, which gives also several alternative descriptions of ${}^0\mathcal{U}$ taken with linked trace filter pairs). If $\mathcal{U}$ is a uniformity then $\mathfrak{P}$ consists of the filter pairs $(\mathfrak{f}, \mathfrak{f})$ where $\mathfrak{f}$ is a non-convergent round Cauchy filter; as there is only one quasi-uniform extension for prescribed linked trace filter pairs (again [De4] Theorem 11.2), ${}^0\mathcal{U}(\mathfrak{P})$ has to coincide with the usual uniform completion. h) does not hold, for the same reason as in the case of ${}^{0)}\mathcal{U}(\mathfrak{F})$.

2.3 Examples. a) On $X = [-1,1] \setminus \{0\}$, consider the distance

$$d(x,y) = |y - x| \quad \text{if} \quad xy > 0 \quad \text{or} \quad x < 0 < y.$$

$\mathcal{U}(d)$ is totally bounded. $\mathfrak{f}^0_{-1} = (\mathfrak{e}_{-1}|X, \mathfrak{e}|X)$ and $\mathfrak{f}^0_1 = (\mathfrak{e}|X, \mathfrak{e}_1|X)$ are linked round Cauchy filter pairs. As $\mathfrak{e}|X \subset \mathfrak{e}_1|X$, both Cs-completions remain Cs-complete if we drop $\mathfrak{e}_1|X$ from $\mathfrak{F}$, respectively $\mathfrak{f}^0_1$ from $\mathfrak{P}$.

b) In $\mathbf{R}^2$, take halflines H_n $(n \in \mathbf{N})$ starting from 0^2, and let

$$A_n = \{x_{nk} : k \in \mathbf{N}\}, \qquad X = \bigcup_{n \in \mathbf{N}} A_n$$

where $x_{nk} \in H_n$ and $d^2_{\mathrm{eu}}(x_{nk}, 0^2) = 1/k$. Define $\mathcal{U} = \mathcal{U}(d)$,

$$d(x_{nk}, x_{mj}) = \begin{cases} |1/k - 1/j| & \text{if } n = m, \\ 1/k + 1/j & \text{if } n > m. \end{cases}$$

Now the filter pairs

$$\mathfrak{f}^0_n = \Big(\mathrm{fil}(\mathfrak{e}^2 \mid \bigcup_{s=n}^{\infty} A_s), \mathrm{fil}(\mathfrak{e}^2 \mid \bigcup_{s=1}^{n} A_s)\Big) \quad (n \in \mathbf{N})$$

are linked, round and Cauchy, $\mathfrak{f}^1_1 \supset \mathfrak{f}^1_2 \supset \ldots$, but the filter $\bigcap_{n \in \mathbf{N}} \mathfrak{f}^1_n$ is not even Cauchy. (In fact, any Cauchy filter is finer than some $\mathfrak{f}^1_n$.) We shall see in Lemma 9.2 that there is no similar example with $\mathcal{U}$ totally bounded.

c) In the above space, let $\mathfrak{g}_n$ be a free ultrafilter with $A_n \in \mathfrak{g}_n$ $(n \in \mathbf{N})$. As $\mathcal{U}^{\mathrm{tp}}$ is discrete, $\mathfrak{g}_n$ is also a maximal open filter. $\mathfrak{g}^{\mathrm{E}}_n = \mathfrak{f}^1_n$, thus $\mathfrak{g}^{\mathrm{E}}_n \in \mathfrak{F}^{\mathrm{MnE}}_{\mathsf{C}} = \mathfrak{F}^{\mathrm{MnE}}_{\mathsf{O}}$. This shows that some of the filters can be superfluous in the definition of ${}^{\mathrm{SP}}\mathcal{U}$, ${}^{\mathrm{MN}}\mathcal{U}$ and ${}^{\mathrm{FN}}\mathcal{U}$, but we cannot take ${}^{0)}\mathcal{U}(\mathfrak{F}^{\mathrm{mn}}_{\mathsf{C}})$, ${}^{0)}\mathcal{U}(\mathfrak{F}^{\mathrm{Mn\,Em}}_{\mathsf{C}})$, respectively ${}^{0)}\mathcal{U}(\mathfrak{F}^{\mathrm{MnEm}}_{\mathsf{O}})$.

§ 3 Symmetric but not bitopological notions of completeness

3.1 There is an obvious way to introduce symmetric notions of completeness: assume that both $\mathcal{U}$ and $\mathcal{U}^{-1}$ have one of the non-symmetric properties from § 2. As the simplest example, let us consider doubly SP-complete spaces. There are quasi-uniformities that do not possess a doubly SP-complete extension, since the existence of such an extension implies that all the Cauchy filters are D-Cauchy. (Indeed, if $\mathfrak{f}$ is Cauchy in $(X, \mathcal{U})$ then it converges to a point a in the extension, thus $\mathfrak{f}^1(a)$ is a D-Cauchy filter coarser than $\mathfrak{f}$). So the appropriate question is whether or not each quasi-uniform space has a doubly SP-complete *quasi-extension*, which means that X is only *quasi-dense* in the larger space $(Y, \mathcal{V})$, i.e. $Y = \mathrm{Cl}^{-1} X \cup \mathrm{Cl}^{1} X$. (In [De6] § 3, we used the expressions "extension" and "proper extension" for quasi-extension, respectively extension.) Doubly closed subspaces have to be considered in Condition c).

Any quasi-uniform space $(X,\mathcal{U})$ has a very simple doubly SP-complete quasi-extension: with $p,q\notin X$, take on $Y=X\cup\{p,q\}$ the quasi-uniformity for which

$$\{U\cup\{p\}\times Y\cup Y\times\{q\}: U\in\mathcal{U}\}$$

is a base. We do not know whether there is a better double SP-completion, with finer trace filters. The following modification of ${}^{\mathrm{SP}}\mathcal{U}$ might seem to be a promising candidate:

Let $Y=X\cup P_{-1}\cup P_1$ where X, P_{-1} and P_1 are disjoint, and let $\mathfrak{f}^i$ be a bijection from P_i onto the family of the non-$\mathcal{U}^{i\mathrm{tp}}$-convergent $\mathcal{U}^i$-round $\mathcal{U}^i$-Cauchy filters. Denote by Φ^i the collection of the functions F defined on P_i satisfying $F(p)\in\mathfrak{f}^i(p)$ $(p\in P_i)$. Assign to any $U\in\mathcal{U}$ and $F_i\in\Phi^i$ $(i=\pm1)$ the entourage $V=V(U,F_{-1},F_1)$ defined by

$$\begin{aligned}
&x\,V\,y \text{ iff } x\,U\,y \quad (x,y\in X),\\
&p\,V\,y \text{ iff } p\in P_1,\ \ y\in U[F_1(p)] \quad (p\in P_{-1}\cup P_1,\ y\in X),\\
&x\,V\,q \text{ iff } q\in P_{-1},\ x\in U^{-1}[F_{-1}(q)] \quad (x\in X,\ q\in P_{-1}\cup P_1),\\
&p\,V\,q \text{ iff } U[F_1(p)]\cap F_{-1}(q)\neq\emptyset \quad (p\in P_1,\ q\in P_{-1}),\\
&p\,V\,q \text{ iff } p=q \quad (p,q\in P_{-1}\cup P_1,\ (p,q)\notin P_1\times P_{-1}).
\end{aligned}$$

Now the above entourages form a base for a quasi-uniformity $\mathcal{V}$ on Y; formally, it is ${}^0\mathcal{U}$ taken with the "filter pairs" $(\exp X,\mathfrak{f}^1(p))$ $(p\in P_1)$ and $(\mathfrak{f}^{-1}(q),\exp X)$ $(q\in P_{-1})$ (cf. [De6] § 3, where one of the members of a filter pair is allowed to be $\exp X$). This $\mathcal{V}$ is, however, not always doubly SP-complete:

EXAMPLE. Let $X=\mathbf{N}\times\mathbf{R}_0\cup H$ where H consists of all the functions $\mathbf{N}\to\mathbf{R}_1$. Consider the distance

$$d(x,y)=\begin{cases} y''-x'' & \text{if } x,y\in\mathbf{N}\times\mathbf{R}_0,\ x'=y',\ x''<0<y'',\\ 1/x' & \text{if } x\in\mathbf{N}\times\mathbf{R}_0,\ y\in H,\ x''\in\mathbf{R}_{-1}\cup\{y(x')\}.\end{cases}$$

If we define $(Y,\mathcal{V})$ for $\mathcal{U}=\mathcal{U}(d)$ as above then $\mathcal{V}^{-1}$ is not SP-complete: the filter $\mathfrak{f}$ generated by the sequence p_n $(n\in\mathbf{N})$ is $\mathcal{V}^{-1}$-Cauchy but not $\mathcal{V}^{-\mathrm{tp}}$-convergent, where $p_n\in P_1$ and $\mathfrak{f}^1(p_n)=\mathrm{fil}(\mathfrak{e}_1^2(n,0)\mid\mathbf{N}\times\mathbf{R}_0)$.

To prove that $\mathfrak{f}$ is $\mathcal{V}^{-1}$-Cauchy, take a $V=V(U_{(t)},F_{-1},F_1)$, and pick $h\in H$ such that $(n,h(n))\in F_1(p_n)$ $(n\in\mathbf{N})$. Now $p_n\,V\,h$ holds for $n>1/t$, thus $V^{-1}h\in\mathfrak{f}$.

$\mathfrak{f}$ clearly does not converge to any point of $P_1\cup\mathbf{N}\times\mathbf{R}_0$. For $h\in H$ fixed, take $F_1\in\Phi^1$ with

$$F_1(p_n)=\{n\}\times]0,h(n)[\qquad (n\in\mathbf{N}), \tag{1}$$

and let $F_{-1}\in\Phi^{-1}$ be arbitrary. Then $V^{-1}h\notin\mathfrak{f}$ where

$$V=V(U_{(1)},F_{-1},F_1); \tag{2}$$

thus $\mathfrak{f}$ does not $\mathcal{V}^{-\mathrm{tp}}$-converge to h either. Finally, let $q \in P_{-1}$, and choose $x \in X$ such that $U_{(1)}^{-1}x \in \mathfrak{f}^{-1}(q)$. If $x \in \mathbb{N} \times \mathbb{R}_0$ then it is evident that $\mathfrak{f}$ does not $\mathcal{V}^{-\mathrm{tp}}$-converge to q; if $x = h \in H$ then $V^{-1}q \notin \mathfrak{f}$ where V is defined by (2), with F_1 satisfying (1) and $F_{-1}(q) = U_{(1)}^{-1}x$. □

§ 4 C-completeness

4.1 A quasi-uniformity will be called L-*complete* (equivalent notions introduced as "complete" in [Kr], "doubly complete" in [Cs], "bicomplete" in [FL3]; present definition, with the name "pair complete" in [LF], "complete" in [De4]; see also [Cs2], [Kr2], [Sal]) if each linked Cauchy filter pair is convergent (equivalently: each linked round Cauchy filter pair is convergent; each linked Cauchy filter pair has a cluster point; see [De4] Lemma 12.2). Any quasi-uniformity has a unique finest basic L-complete extension, namely ${}^0\mathcal{U}(\mathfrak{P}_\mathsf{L})$, which can be described in several simpler ways, see [De4] Theorem 11.2, or, with $\mathcal{U}^\mathrm{s}$-Cauchy filters instead of filter pairs, see in some of the references cited above. $\mathcal{U}$ is L-complete iff the uniformity $\mathcal{U}^\mathrm{s}$ is complete. It is more appropriate to consider now *firm extensions*, which means that the original space is $\mathcal{V}^{\mathrm{stp}}$-dense in the extension $(Y, \mathcal{V})$; ${}^\mathsf{L}\mathcal{U} = {}^0\mathcal{U}(\mathfrak{P}_\mathsf{L})$ is essentially the only reduced L-complete firm extension of $\mathcal{U}$, cf. [Cs2] (16.76) or [LF] Corollary 17. The theory of L-completeness and L-completion is well-known, so we do not go into details; see [De4] § 12 and the references therein.

For all its good properties, L-completeness has a disadvantage: there are too many complete spaces, e.g. any subspace of $(\mathbb{R}, \mathcal{U}_{\mathrm{so}})$ is complete (as $\mathcal{U}_{\mathrm{so}}^\mathrm{s}$ is discrete). The class of the complete spaces can be easily made smaller: let us use less filter pairs in the definition.

4.2 Making the most obvious choice, we call a quasi-uniformity
C-*complete* [De7] if each Cauchy filter pair is convergent;
A-*complete* if each Cauchy filter pair has a cluster point.

Both notions satisfy Conditions a) to e). To b): A complete uniform space is C-complete, because if $\mathfrak{f}^0$ is a Cauchy filter pair in it then $\mathfrak{f}^{-1\mathrm{E}} = \mathfrak{f}^{1\mathrm{E}}$ is a Cauchy filter. To e): $\mathcal{U} = \mathcal{U}_{\mathrm{se}} \mid \mathbb{R}_0$ is not even L-complete, since $\mathcal{U}^\mathrm{s} = \mathcal{U}_{\mathrm{eu}} \mid \mathbb{R}_0$ is not complete. Given a Cauchy filter pair $\mathfrak{f}^0$ in $(\mathbb{R}, \mathcal{U}_{\mathrm{se}})$, take $K(1) \supset K(2) \supset \ldots \supset \in \mathfrak{f}$ with $K(n) \subset U_{(1/n)}$; now with $x_n = \inf K_{-1}(n)$, $y_n = \sup K_1(n)$ we have $y_n \leqq x_n + 1/n$, thus, the two sequences being monotone, $\lim_n y_n \leqq \lim_n x_n$, and $\mathfrak{f}^0$ converges to any point between these limits.

We do not know whether each quasi-uniformity has a C-complete, or at least an A-complete, extension. ${}^0\mathcal{U}$ is namely the only construction available for arbitrary trace filter pairs in arbitrary spaces, and it is not A-complete in the next example, independently of how we choose the trace filter pairs.

Example. Consider on $(-1/\mathbb{N}) \times \mathbb{R}_0 \cup \mathbb{R}_0 \times (1/\mathbb{N})$ the quasi-uniformity $\mathcal{U} = \mathcal{U}(d_{\mathrm{so}}^2) \mid X$. Assume that $({}^0X, {}^0\mathcal{U})$ is A-complete with a suitable system

of trace filter pairs. The filter pairs

$$\mathfrak{f}^0_{-n} = \epsilon^2_0(-1/n, 0) \mid X, \quad \mathfrak{f}^0_n = \epsilon^2_0(0, 1/n) \mid X \qquad (n \in \mathbf{N})$$

are Cauchy, so $\mathfrak{f}^0_k$ has a cluster point p_k in ${}^0X$ $(k \in \pm\mathbf{N})$. Now $\mathfrak{f}^0(p_k)(\cap)\mathfrak{f}^0_k$ is round and Cauchy, implying $\mathfrak{f}^0(p_k) = \mathfrak{f}^0_k$, since $\mathfrak{f}^0_k$ is a minimal as well as maximal round Cauchy filter pair. Take the filters $\mathfrak{f}^i$ in ${}^0X$ generated by the sequence $\langle p_{in}\rangle$ $(i = \pm 1)$. $\mathfrak{f}^0$ is Cauchy: for $V = V(U_{(t)}, F_{-1}, F_1)$ we have $p_{-m} V p_n$ whenever $m, n > 1/t$. The assumption that $\mathfrak{f}^0$ has a cluster point b will lead to a contradiction.

Pick $K \in \mathfrak{f}^\times(b)$ with $K \subset U_{(1)}$, and let $V = V(U_{(1)}, F_{-1}, F_1)$ where $F_i(b) \subset \subset K_i$ $(i = \pm 1)$,

$$F_1(p_{-n}) \subset \{-1/n\} \times \mathbf{R}_1, \ F_{-1}(p_n) \subset \mathbf{R}_{-1} \times \{1/n\} \qquad (n \in \mathbf{N}).$$

As b is a $\mathcal{V}^{-\mathrm{tp}}$-cluster point of $\mathfrak{f}^{-1}$, there is a $k \in \mathbf{N}$ with $p_{-k} V b$, $p_{-k} \neq \neq b$, implying the existence of $x \in F_1(p_{-k})$ and $y \in K_{-1}$ such that $x\, U_{(1)}\, y$. From $x'' > 0$ we get $y'' > 0$. Now $K_1 \subset U_{(1)} y \subset [y', \to [\times [y'', \to [$, so $U_{(1)}[K_1] \cap \cap F_{-1}(p_n) = \emptyset$ if $n > 1/y''$, i.e. $b V p_n$ does not hold, contradicting the assumption that b is a $\mathcal{V}^{\mathrm{tp}}$-cluster point of $\mathfrak{f}^1$. □

We can, however, hope for a complete extension if only some special Cauchy filter pairs are used in the definition of completeness (preferably, a class essentially larger than that of the linked Cauchy filter pairs): there exist several other constructions with special trace filter pairs, and, on the other hand, it seems to be possible that even ${}^0\mathcal{U}$ will do in such situations. Before introducing the new notions of completeness, some properties of filter pairs will be defined in the next section, and some constructions will be given in § 6 and compared in § 7. These three sections continue investigations from [De4] §§ 7 and 8; some of the results, included for the sake of completeness, will not be needed in the sequel.

REFERENCES

[B] Bîrsan, T., Sur les espaces bitopologiques complètement réguliers, *An. Şti. Univ. "Al. I. Cuza" Iaşi Secţ. Ia Mat. (N. S.)* **16** (1970), No. 1, 29–34. *MR* **42** #8436

[CnH] Carlson, J. W. and Hicks, T. L., On completeness in quasi-uniform spaces, *J. Math. Anal. Appl.* **34** (1971), No. 3, 618–627. *MR* **43** #6868

[CrH] Carter, K. S. and Hicks, T. L., Some results on quasi-uniform spaces, *Canad. Math. Bull.* **19** (1976), No. 1, 39–51. *MR* **54** #11284

[Cs] Császár, Á., *Foundations of general topology*, Pergamon Press, Oxford, 1963. *MR* **28** #575

[Cs2] Császár, Á., *Grundlagen der allgemeinen Topologie*, Akadémiai Kiadó, Budapest, 1963. *MR* **26** #6917

[Cs3] Császár, Á., *General topology*, Akadémiai Kiadó, Budapest and Adam Hilger Ltd., Bristol, 1978. *MR* **57** #13812

[Cs4] Császár, Á., Extensions of quasi-uniformities, *Acta Math. Acad. Sci. Hungar.* **37** (1981), No. 1–3, 121–145. *MR* **82f**:54039

[Cs5] Császár, Á., Complete extensions of quasi-uniform spaces, *General topology and its relations to modern analysis and algebra V* (Proc. Fifth Prague Topological Symp., 1981), Sigma Series in Pure Math. **3**, Heldermann, Berlin, 1983, 104–113. *MR* **84e**:54030

[Cs6] Császár, Á., D-complete extensions of quasi-uniform spaces, *Acta Math. Hungar.* **64** (1994), 41–54.

[CsM] Császár, Á. and Matolcsy, K., Syntopogenous extensions for prescribed topologies, *Acta Math. Acad. Sci. Hungar.* **37** (1981), No. 1–2, 59–75. *MR* **82i**:54006

[Da] Datta, M. C., Projective bitopological spaces, *J. Austral. Math. Soc.* **13** (1972), No. 3, 327–334. *MR* **46** #4496

[De] Deák, J., On bitopological spaces I, *Studia Sci. Math. Hungar.* **25** (1990), No. 4, 457–481. *MR* **92m**:54054

[De2] Deák, J., On bitopological spaces III, *Studia Sci. Math. Hungar.* **26** (1991), No. 1, 19–33. *MR* **93g**:54043

[De3] Deák, J., Extensions of quasi-uniformities for prescribed bitopologies I, *Studia Sci. Math. Hungar.* **25** (1990), No. 1–2, 45–67. *MR* **92b**:54058

[De4] Deák, J., Extensions of quasi-uniformities for prescribed bitopologies II, *Studia Sci. Math. Hungar.* **25** (1990), No. 1–2, 69–91. *MR* **92b**:54058

[De5] Deák, J., On the coincidence of some notions of quasi-uniform completeness defined by filter pairs, *Studia Sci. Math. Hungar.* **26** (1991), 411–413. *MR* **94b**:54077

[De6] Deák, J., A survey of compatible extension (presenting 77 unsolved problems), *Topology, theory and applications II* (Proc. Sixth. Colloq., Pécs, 1989), Colloq. Math. Soc. János Bolyai **55**, North-Holland, Amsterdam, 1993. *MR* **94i**:54057

[De7] Deák, J., Extending and completing quiet quasi-uniformities, *Studia Sci. Math. Hungar.* **29** (1994), 349–362.

[De8] Deák, J., A counterexample on completeness in relator spaces, *Publ. Math. Debrecen* **41** (1992), 307–309. *MR* **93f**:54039

[De9] Deák, J., Quasi-uniform completeness and neighbourhood filters, *Studia Sci. Math. Hungar.* (to appear).

[Do] Doitchinov, D., Completeness and completions of quasi-metric spaces, Third National Conf. on Topology (Trieste, 1986), *Rend. Circ. Mat. Palermo* (2) Suppl. No. 18 (1988), 41–50. *MR* **89g**:54066

[Do2] Doitchinov, D., Completeness in quasi-metric spaces, *Topology Appl.* **30** (1988), No. 2, 127–145. *MR* **90e**:54068

[Do3] Doitchinov, D., On completeness of quasi-uniform spaces, *C. R. Acad. Bulg. Sci.* **41** (1988), No. 7, 5–9. *MR* **89j**:54028

[Do4] Doitchinov, D., A concept of completeness of quasi-uniform spaces, *Topology Appl.* **38** (1991), No. 3, 205–217.

[Do5] Doitchinov, D., Another class of completable quasi-uniform spaces, *C. R. Acad. Bulg. Sci.* **44** (1991), No. 3, 5–6.

[Do6] Doitchinov, D., Stable quasi-uniform spaces and their completions.

[F] Fletcher, P., On completeness of quasi-uniform spaces, *Arch. Math. (Basel)* **22** (1971), No. 2, 200–204. *MR* **46** #4489

[FH] Fletcher, P. and Hunsaker, W., Completeness using pairs of filters, *Topology Appl.*

[FL] Fletcher, P. and Lindgren, W. F., *C*-complete quasi-uniform spaces, *Arch. Math. (Basel)* **30** (1978), No. 2, 175–180. *MR* **58** #7562

[FL2] Fletcher, P. and Lindgren, W. F., *Quasi-uniform spaces*, Lecture Notes in Pure Appl. Math. **77**, Marcel Dekker, New York, 1982. *MR* **84h**:54026

[FN] Fletcher, P. and Naimpally, S. A., On almost complete and almost precompact quasi-uniform spaces, *Czechoslovak Math. J.* **21** (96) (1971), No. 3, 383–390. *MR* **44** #4708

[HHC] HUFFMAN, S. M., HICKS, T. L. and CARLSON, J. W., Complete quasi-uniform spaces, *Canad. Math. Bull.* **23** (1980), No. 4, 297–498. *MR* **82f**:54040

[I] ISBELL, J. R., *Uniform spaces*, Math. Surveys **12**, Amer. Math. Soc., Providence, 1964. *MR* **30** #561

[JB] JAS, M. and BAISNAB, A. P., Bitopological spaces and associated q-proximity, *Indian J. Pure Appl. Math.* **13** (1982), No. 10, 1142–1146. *MR* **83m**:54048

[KJ] KÜNZI, H.-P. A. and JUNNILA, H. J. K., Stability in quasi-uniform spaces and the inverse problem, *Topology Appl.* **49** (1993), 175–189. *MR* **94b**:54081

[KMRV] KÜNZI, H.-P. A., MRŠEVIĆ, M., REILLY, I. L. and VAMANAMURTHY, M. K., Convergence, precompactness and symmetry in quasi-uniform spaces.

[KP] KOPPERMAN, R. D., Total boundedness and compactness for filter pairs, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*

[KR] KRISHNAN, V. S., On additive, asymmetric, semi-uniform spaces and semigroups, *J. Madras Univ.* B **32** (1962), 175–198. *MR* **28** #5422

[KR2] KRISHNAN, V. S., Semiuniform spaces, and seminorms, semimetrics, semiécarts in apo-semigroups, *General topology and its relations to modern analysis and algebra III* (Proc. Conf. Kanpur, 1968), Academia, Prague, 1971, 163–171. *MR* **44** #3934

[KÜ] KÜNZI, H.-P. A., Complete quasi-pseudo-metric spaces, *Acta Math. Hungar.*

[KÜ2] KÜNZI, H.-P. A., Functorial admissible quasi-uniformities on topological spaces, *Topology Appl.*

[KW] KOWALSKY, H.-J., *Topologische Räume*, Birkhäuser, Basel, 1961. *MR* **22** #12502

[LF] LINDGREN, W. F. and FLETCHER, P., A construction of the pair completion of a quasi-uniform space, *Canad. Math. Bull.* **21** (1978), No. 1, 53–59. *MR* **58** #7562

[ME] MEENAKSHI, K. N., Completion of bitopological spaces, *J. Madras Univ.* **35–36** (1965–1966), 27–31. *MR* **42** #5210

[MN] MURDESHWAR, M. G. and NAIMPALLY, S. A., *Quasi-uniform topological spaces*, Nordhoff, Groningen, 1966. *MR* **35**#2267

[SAL] SALBANY, S., *Bitopological spaces, compactifications and completions*, Math. Monographs Univ. Cape Town **1**, Department Math., Univ. Cape Town, Cape Town, 1974. *MR* **54** #13869

[SAM] SAMUEL, P., Ultrafilters and compactifications of uniform spaces, *Trans. Amer. Math. Soc.* **64** (1948), 100–132. *MR* **10**, 54

[SM] SMYTH, M. B., Completeness of quasi-uniform spaces in terms of filters, Unpublished manuscript, London, 1987.

[SP] SIEBER, J. L. and PERVIN, W. J., Completeness in quasi-uniform spaces, *Math. Ann.* **158** (1965), No. 2, 79–81. *MR* **30** #2449

[ST] STOLTENBERG, R. A., A completion for a quasi-uniform space, *Proc. Amer. Math. Soc.* **18** (1967), No 5, 864–867. *MR* **35** #6124

[SÜ] SÜNDERHAUF, P., The Smyth-completion of a quasi-uniform space, Preprint No. 1427 (1991), Technische Hochschule, Darmstadt.

[SZ] SZÁZ, A., Lebesgue relators, *Monatshefte Math.* **110** (1990), No. 3–4, 315–319.

[W] WARD, A. J., A generalization of almost compactness, with an associated generalization of completeness, *Czechoslovak Math. J.* **25** (**100**) (1975), No. 4, 514–530. *MR* **52** #11851

(Received November 12, 1991)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 411–431

A BITOPOLOGICAL VIEW OF QUASI-UNIFORM COMPLETENESS. II

J. DEÁK

Abstract

A quasi-uniformity is S-complete if each stable Cauchy filter pair is convergent. Using an extension for stable trace filter pairs (which is different from the one given in [De4]), we show that each quasi-uniformity has an S-completion. The following modification of D-completeness will also be considered: each stable D-Cauchy filter is convergent.

§ 5 Special properties of filters and filter pairs*

5.1 Let $(X,\mathcal{U})$ be a quasi-uniform space. A filter $\mathfrak{f}$ in X is *stable* ([I], [Cs4]) if for any $U \in \mathcal{U}$, $\bigcap_{S\in\mathfrak{f}} U[S] \in \mathfrak{f}$. A filter pair $\mathfrak{f}^0$ in X is

stable [De4] if $\mathfrak{f}^i$ is $\mathcal{U}^i$-stable $(i = \pm 1)$;

concentrated [De4] if for any $K \in \mathfrak{f}^\times$ there is a $U \in \mathcal{U}$ such that $L \subset K$ whenever $L \in \mathfrak{f}^\times$ and $L \subset U$;

weakly concentrated [De4] if for any $U \in \mathcal{U}$ there is a $U_0 \in \mathcal{U}$ such that $K, L \in \mathfrak{f}^\times$, $K, L \subset U_0$ imply $K_{-1} \times L_1 \subset U$.

Some simple results from [De4] § 7: A filter pair coarser than a weakly concentrated one has the same property. $\mathfrak{f}^{0E}$ is stable iff $\mathfrak{f}^0$ is so. (It is easy to check that a similar statement holds for weak concentratedness.) The following implications (and no more) are valid:

$$\begin{array}{ccccc} \text{non-Cauchy} & & \text{linked} & \overset{2}{\Longrightarrow} & \text{stable} \\ & \searrow & \overset{1}{\not\swarrow} & \searrow & \\ & & \text{concentrated} & \Longrightarrow & \text{weakly concentrated} \end{array}$$

1991 *Mathematics Subject Classification.* Primary 54E15, 54D35; Secondary 54E55, 54G20.

Key words and phrases. Quasi-uniformity, bitopology, S-complete, D-complete, SD-complete, (half-)extension, completion, round/Cauchy/stable filter (pair), D-Cauchy filter, (doubly) stable quasi-uniformity.

*For §§ 0–4 and the references see Part I in this Studia, pp. 389–409.

Research supported by Hungarian National Foundation for Scientific Research Grant No. 2114.

0081–6906/95/$ 4.00

where $\overset{1}{\Longrightarrow}$ holds for round and $\overset{2}{\Longrightarrow}$ for Cauchy filter pairs. A Cauchy filter pair is concentrated iff it is weakly concentrated and minimal Cauchy. A Cauchy filter pair $\mathfrak{f}^0$ is weakly concentrated iff

(1) for any $U \in \mathcal{U}$ there is a $U_0 \in \mathcal{U}$ such that xUy whenever $U_0 x \in \mathfrak{f}^1$ and $U_0^{-1} y \in \mathfrak{f}^{-1}$.

For any weakly concentrated Cauchy filter pair $\mathfrak{f}^0$ there exists a unique concentrated Cauchy filter pair $\mathfrak{m}^0$ coarser than it that can be defined as follows:

$$\mathfrak{m}^i = \operatorname{fil}\{M_i(U) : U \in \mathcal{U}\} \quad (i = \pm 1), \tag{2}$$

$$M_i(U) = M_i(U, \mathfrak{f}^0) = \bigcup\{K_i : K \in \mathfrak{f}^\times,\ K \subset U\}. \tag{3}$$

The fact that a certain class of Cauchy filter pairs is overlayed by its minimal elements can be of use when looking for complete extensions, since the filter pairs in this class (at least, the ones in X), can be made convergent using only the minimal elements as trace filter pairs; in this way we have a chance of obtaining a completion that is a complete hull. We are going to show that for any stable Cauchy filter pair there is a (not necessarily unique) minimal stable Cauchy filter pair coarser than it. Instead of giving now a direct proof, we choose a roundabout way in order to point out a connexion between the two statements on minimal filter pairs (for weakly concentrated, respectively for stable ones).

5.2 Given a filter $\mathfrak{f}$ in $(X, \mathcal{U})$, we define

$$\mathfrak{m}^i(\mathfrak{f}) = \{M^i(U) : U \in \mathcal{U}\},$$
$$M^i(U) = M^i(U, \mathfrak{f}) = \{x \in X : U^{-i} x \in \mathfrak{f}\}.$$

If $U \subset V$ then $M^i(U) \subset M^i(V)$; if $S \supset M^i(U)$ then $S = M^i(V)$ with $V = U \cup (S \times X)^{-i}$; thus $\mathfrak{m}^i(\mathfrak{f})$ is either a filter or $\exp X$; clearly, it is a filter iff $\mathfrak{f}$ is $\mathcal{U}^i$-Cauchy. Thus if $\mathfrak{f}^0$ is a Cauchy filter pair then

$$\mathfrak{m}^0(\mathfrak{f}^0) = (\mathfrak{m}^{-1}(\mathfrak{f}^1), \mathfrak{m}^1(\mathfrak{f}^{-1}))$$

is a filter pair. In fact, $\mathfrak{m}^0(\mathfrak{f}^0) = \mathfrak{m}^0$ (with $\mathfrak{m}^0$ from 5.1), since if $x \in M_i(U)$ (cf. 5.1 (3)) then $x \in K_i$ with some $K \in \mathfrak{f}^\times$, $K \subset U$, so $U^{-i} x \supset K_{-i} \in \mathfrak{f}^{-i}$, i.e., $x \in M^i(U, \mathfrak{f}^{-i})$, and conversely, if $x \in M^i(U, \mathfrak{f}^{-i})$ then, taking a $K \in \mathfrak{f}^\times$ with $K \subset U$, we have $L \in \mathfrak{f}^\times$, $L \subset U$ where $L_i = \{x\} \cup K_i$, $L_{-i} = U^{-i} x \cap K_{-i}$; hence $M_i(U) = M^i(U, \mathfrak{f}^{-i})$. (Thus "fil" was superfluous in 5.1 (2).)

Let $\mathfrak{f}^0$ be a Cauchy filter pair. $\mathfrak{m}^0(\mathfrak{f}^0)$ is coarser than $\mathfrak{f}^0$ (since $M_i(U) \in \mathfrak{f}^i$ is clear); $\mathfrak{f}^0$ is concentrated iff $\mathfrak{f}^0 = \mathfrak{m}^0(\mathfrak{f}^0)$ (a simple rewording of the definition, already used in [De4] 7.13).

If $\mathfrak{f}$ is a D-Cauchy filter then (by the first statement of the preceding paragraph) $\mathfrak{m}^{-1}(\mathfrak{f})$ is coarser than any cofilter of $\mathfrak{f}$; hence if $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ is Cauchy then $\mathfrak{m}^{-1}(\mathfrak{f})$ is the coarsest cofilter of $\mathfrak{f}$. These observations make the following definition plausible:

DEFINITION. A filter $\mathfrak{f}$ in a quasi-uniform space is
a) *cominimal* if it has a minimal cofilter;
b) *strongly cominimal* if $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ is a Cauchy filter pair. □

If $\mathfrak{g}$ and $\mathfrak{h}$ are cofilters of $\mathfrak{f}$ then so is $\mathfrak{g} \cap \mathfrak{h}$; thus a minimal cofilter of $\mathfrak{f}$ is the coarsest cofilter of it.

LEMMA. *For a filter in a quasi-uniform space, each of the following conditions implies the next:*
(i) *stable D-Cauchy,*
(ii) *strongly cominimal,*
(iii) *cominimal,*
(iv) *D-Cauchy.*
If $\mathfrak{f}$ *is strongly cominimal then* $\mathfrak{m}^{-1}(\mathfrak{f})$ *is its coarsest cofilter.*

PROOF. (iii) $\Longrightarrow$ (iv) is evident. (ii) $\Longrightarrow$ (iii) and the second sentence have already been proved. To show (i) $\Longrightarrow$ (ii), let $\mathfrak{f}$ be stable and D-Cauchy, and $U \in \mathcal{U}$. With $F = \bigcap_{S \in \mathfrak{f}} U[S] \in \mathfrak{f}$ we have $M^{-1}(U) \times F \subset U^2$, thus $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ is Cauchy indeed. □

REMARK (Á. Császár). The above proof only uses that $\mathfrak{f}$ is Cauchy instead of D-Cauchy. Therefore any stable Cauchy filter is D-Cauchy. In particular, the Cauchy and the D-Cauchy filters are the same in totally bounded spaces (since all filters are stable in such a space, see [Cs4] 4.5). See also 5.7 and 5.8.

5.3 The next lemma shows that there is indeed a kinship between stability and weak concentratedness. (We have already cited stronger results from [De4].)

LEMMA. *If* $\mathfrak{f}^0$ *is weakly concentrated Cauchy then* $\mathfrak{f}^1$ *is strongly cominimal.* □

REMARK. If $\mathfrak{f}^0$ is weakly concentrated Cauchy then either of the filters $\mathfrak{f}^i$ alone determines $\mathfrak{m}^0(\mathfrak{f}^0)$, since $\mathfrak{m}^{-i}(\mathfrak{f}^i) = \mathfrak{m}^{-i}(\mathfrak{m}^i(\mathfrak{f}^{-i}))$; this explains the notation $\mathfrak{m}^i(\mathfrak{f}^i)$ used in [De7]. (A different construction for $\mathfrak{m}^0$ can be found at the end of [De7] 0.3.)

EXAMPLES. a) *A strongly cominimal filter that is neither stable nor the second member of a weakly concentrated Cauchy filter pair.* On $X = \mathbf{R}_0 \times \times \{0\} \cup \mathbf{R}_1 \times \{1\}$, let $\mathcal{U} = \mathcal{U}(d)$ be defined by the distance

$$d(x,y) = \begin{cases} y' - x' & \text{if} \quad x' < 0 < y', \quad y'' = 0, \\ -x'y' & \text{if} \quad x' < 0 < y', \quad y'' = 1. \end{cases}$$

$\mathfrak{f} = \mathfrak{e}_1^2|X$ is clearly not stable. $\mathfrak{m}^{-1}(\mathfrak{f}) = \mathfrak{e}_{-1}^2|X$, so $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ is Cauchy, i.e., $\mathfrak{f}$ is strongly cominimal. If there existed a weakly concentrated Cauchy filter pair $(\mathfrak{g}, \mathfrak{f})$ then the coarser filter pair $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ would also be weakly concentrated; but it is not: no entourage $U_{(t)}$ is good for $U_{(1)}$ in 5.1 (1), since we have $U_{(t)}(-t/2, 0) \in \mathfrak{f}$, $U_{(t)}^{-1}(2/t, 1) \in \mathfrak{m}^{-1}(\mathfrak{f})$, but the distance of these points is 1.

b) *Cominimal, not strongly cominimal.* On $X = \mathbf{R}_0$, let

$$d(x, y) = y - x \quad \text{if} \quad x < 0 < y, \quad (x \notin \mathbf{Q} \Rightarrow -x > y).$$

With $\mathfrak{f} = \mathfrak{e}_1|X$, $(\mathfrak{m}^{-1}(\mathfrak{f}), \mathfrak{f}) = \mathfrak{e}_0|X$ is not Cauchy, but $\mathfrak{f}$ has a coarsest cofilter, namely $\mathrm{fil}\,(\mathfrak{e}_{-1}|\mathbf{Q})$.

c) *D-Cauchy, not cominimal.* The filter $\mathfrak{e}^2|X$ in $(X, \mathcal{U}(d))$, where $X = \mathbf{R}_{-1} \times \mathbf{N} \cup \mathbf{R}_1 \times \{0\}$, and

$$d(x, y) = y' - x' \quad \text{if} \quad x' < 0 < y' < 1/x''. \qquad \square$$

5.4 LEMMA. *The intersection of a non-empty collection of stable filters is stable, too.* $\square$

5.5 LEMMA. *For any stable Cauchy filter pair there is a minimal stable Cauchy filter pair coarser than it.*

PROOF. Let $\mathfrak{f}^0$ be stable and Cauchy. Denote by $\mathfrak{g}^{-1}$ the intersection of the $\mathcal{U}^{-1}$-stable cofilters of $\mathfrak{f}^1$. $\mathfrak{g}^{-1}$ is $\mathcal{U}^{-1}$-stable (Lemma 5.4), and $(\mathfrak{g}^{-1}, \mathfrak{f}^1)$ is Cauchy (Lemma 5.2). Thus $\mathfrak{g}^{-1}$ is the coarsest $\mathcal{U}^{-1}$-stable cofilter of $\mathfrak{f}$. Similarly, there is a coarsest $\mathcal{U}$-stable $\mathcal{U}^{-1}$-cofilter $\mathfrak{g}^1$ of $\mathfrak{g}^{-1}$. Now $\mathfrak{g}^0$ is a stable Cauchy filter pair coarser than $\mathfrak{f}^0$. Assume that $\mathfrak{h}^0$ has the same properties, and it is coarser than $\mathfrak{g}^0$. Then $(\mathfrak{h}^{-1}, \mathfrak{f}^1)$ is stable Cauchy, thus $\mathfrak{h}^{-1} \supset \mathfrak{g}^{-1}$ by the choice of $\mathfrak{g}^{-1}$, i.e., $\mathfrak{h}^{-1} = \mathfrak{g}^{-1}$. Therefore $\mathfrak{h}^1 \supset \mathfrak{g}^1$, now by the choice of $\mathfrak{g}^1$. $\square$

The minimal stable Cauchy filter pair furnished by this lemma is in general neither unique, nor minimal Cauchy:

EXAMPLES. a) In Example 2.3 a), $\mathfrak{f}_{-1}^0$ and $\mathfrak{f}_1^0$ are minimal (stable) Cauchy filter pairs coarser than $\mathfrak{e}_0|X$.

b) On $X = \mathbf{R}_0 \times \{0\} \cup \mathbf{R}_{-1} \times \{1\}$, let

$$d(x, y) = y' - x' \quad \text{if} \quad \begin{array}{l} \text{either } x'' = y'' = 0 \\ \text{or } x'' = 1, \ y' > 0. \end{array}$$

$\mathfrak{f}^0 = \mathfrak{e}_0^2|X$ is a minimal stable Cauchy filter pair, but it is not minimal Cauchy, since it remains Cauchy if $\mathfrak{e}_{-1}^2$ is replaced by $\mathfrak{e}_{-1}^2 \cap \mathfrak{e}_{-1}^2(0, 1)$. $\square$

NOTATIONS. $\mathfrak{F}_S$ denotes the family of the stable Cauchy (stable D-Cauchy, cf. Remark 5.2) filters; $\mathfrak{P}_S$ is the family of the stable Cauchy filter pairs. $\square$

PROBLEM. Does $\mathfrak{P}_C^m$ overlay $\mathfrak{P}_S$? A positive answer would be surprising, but probably not of much use in the theory of complete extensions.

5.6 LEMMA (in [De6] 3.12 without proof). *Each stable minimal Cauchy filter pair is concentrated.*

PROOF. If $\mathfrak{f}^0$ is stable Cauchy then $(\mathrm{m}^{-1}(\mathfrak{f}^1), \mathfrak{f}^1)$ and $(\mathfrak{f}^{-1}, \mathrm{m}^1(\mathfrak{f}^{-1}))$ are Cauchy by Lemma 5.2. Thus the minimality of $\mathfrak{f}^0$ implies that $\mathrm{m}^0(\mathfrak{f}^0) = \mathfrak{f}^0$, i.e., $\mathfrak{f}^0$ is concentrated. □

A minimal stable Cauchy filter pair is in general not weakly concentrated, see [De4] Example 7.17 b).

COROLLARY. *In a totally bounded space, $\mathfrak{P}_C$ is overlayed by the concentrated Cauchy filter pairs.*

PROOF. Lemmas 5.5 and 5.6, recalling that all the filter pairs are stable. □

Observe that the quasi-uniformity in Example 5.5 a) is totally bounded; thus unicity does not hold in this corollary either.

5.7 REMARK. A filter $\mathfrak{f}$ is called *strictly tame* [De6] (introduced in [Cs4] as "weakly Cauchy") if for any $U \in \mathcal{U}$ there is a $U_0 \in \mathcal{U}$ such that $\bigcap\{Ux : U_0x \in \mathfrak{f}\} \in \mathfrak{f}$. A filter is strongly cominimal iff it is strictly tame and D-Cauchy; this is straightforward from the definitions, and D-Cauchy can be replaced by Cauchy, just as in Remark 5.2.

5.8 According to [KMRV] Proposition 9, if $\mathcal{U}^{-1}$ is hereditarily precompact then each $\mathcal{U}$-Cauchy filter is $\mathcal{U}$ – D-Cauchy. The proof given there shows more: the filters in question are strongly cominimal. We are going to prove a yet stronger result (found independently by Künzi and Junnila, see [KJ] Proposition 1): $\mathcal{U}^{-1}$ is hereditarily precompact iff each filter is $\mathcal{U}$-stable. (Knowing this, one can apply Lemma and Remark 5.2.)

DEFINITION. A filter $\mathfrak{f}$ in $(X, \mathcal{U})$ is *hereditarily Cauchy* if for any $S \in \sec \mathfrak{f}$, $\mathfrak{f}|S$ is $\mathcal{U}|S$-Cauchy. □

LEMMA. *An ultrafilter is $\mathcal{U}$-stable iff it is hereditarily $\mathcal{U}^{-1}$-Cauchy.*

PROOF. Assume first that the ultrafilter $\mathfrak{f}$ is not $\mathcal{U}$-stable. Then there is a $U \in \mathcal{U}$ such that $\bigcap_{S \in \mathfrak{f}} U[S] \notin \mathfrak{f}$, therefore $H = X \setminus \bigcap_{S \in \mathfrak{f}} U[S] \in \mathfrak{f}$. If $x \in H$ then $x \notin U[S]$ for some $S \in \mathfrak{f}$, thus $U^{-1}x \cap S = \emptyset$, $U^{-1}x \notin \sec \mathfrak{f}$, $U^{-1}x \notin \mathfrak{f}$. Hence $\mathfrak{f}|H$ is not $\mathcal{U}^{-1}|H$-Cauchy, i.e., $\mathfrak{f}$ is not hereditarily $\mathcal{U}^{-1}$-Cauchy.

Conversely, assume that $\mathfrak{f}$ is not hereditarily $\mathcal{U}^{-1}$-Cauchy, and take $H \in \in \sec \mathfrak{f} = \mathfrak{f}$ such that $\mathfrak{f}|H$ is not $\mathcal{U}^{-1}|H$-Cauchy. Then there is a $U \in \mathcal{U}$ with $U^{-1}x \notin \mathfrak{f}$ $(x \in H)$. Now $S_x = X \setminus U^{-1}x \in \mathfrak{f}$ $(x \in H)$. For each $x \in H$, $x \notin U[S_x]$, so $T = \bigcap_{x \in H} U[S_x] \subset X \setminus H$, thus $T \notin \mathfrak{f}$, implying that $\mathfrak{f}$ is not $\mathcal{U}$-stable. □

PROPOSITION. *Each filter is stable in $(X,\mathcal{U})$ iff $\mathcal{U}^{-1}$ is hereditarily precompact.*

PROOF. Each of the following statements is equivalent to the next one:
(i) each filter is stable;
(ii) each ultrafilter is stable (Lemma 5.4);
(iii) each ultrafilter is hereditarily $\mathcal{U}^{-1}$-Cauchy;
(iv) for any $S \subset X$, each ultrafilter in S is $\mathcal{U}^{-1}|S$-Cauchy;
(v) any $S \subset X$ is $\mathcal{U}^{-1}|S$-precompact (since a quasi-uniformity is precompact iff each ultrafilter is Cauchy, [FL2] 3.14). □

COROLLARY (improves [Cs4] 4.5). *A quasi-uniformity is totally bounded iff each filter pair is stable.*

PROOF. A quasi-uniformity is totally bounded iff it is doubly hereditarily precompact ([Kü2] Lemma 1). □

REMARK. Hereditarily Cauchy filters (or a weaker version, with $S \in \sec \mathfrak{f}$ replaced by $S \in \mathfrak{f}$ in the definition) could perhaps also be used for defining quasi-uniform completeness (in [Sm], a quasi-uniformity is called *complete* if each round filter satisfying the weaker condition is the neighbourhood filter of a unique point, see also [De9]).

5.9 According to [KMRV] Theorem 3, a quasi-pseudometric space is hereditarily precompact iff each countable subspace of it is precompact. The same holds for quasi-uniformities, too, because any quasi-uniformity is the supremum of quasi-pseudometrizable quasi-uniformities ([Cs] (13.47)), and the supremum of hereditarily precompact quasi-uniformities has the same property ([KMRV] Corollary 8; or use the lemma below). This result can also be proved with the help of stable filters: Assume that $\mathcal{U}^{-1}$ is not hereditarily precompact, and take a non-stable filter $\mathfrak{f}$ (Proposition 5.8). Then there is a $U \in \mathcal{U}$ with $Z = \bigcap_{S \in \mathfrak{f}} U[S] \notin \mathfrak{f}$. Define $S_n \in \mathfrak{f}$ and $x_n \in X \setminus Z$ $(n \in \mathbb{N})$ by recursion as follows: $S_1 = X$; $x_n \in S_n \setminus Z$; $S_{n+1} \subset S_n$ and $x_n \notin U[S_{n+1}]$. Now the filter generated in $C = \{x_n : n \in \mathbb{N}\}$ by the sequence $\langle x_n \rangle$ is not $\mathcal{U}|C$-stable, thus $\mathcal{U}^{-1}|C$ is not hereditarily precompact (again Proposition 5.8).

LEMMA (similar to [KJ] Lemma 2). *Assume that $\mathfrak{f}$ is an ultrafilter in $(X,\mathcal{U})$, $\mathcal{S}$ is a subbase for $\mathcal{U}$, and $\bigcap_{S \in \mathfrak{f}} U[S] \in \mathfrak{f}$ for each $U \in \mathcal{S}$. Then $\mathfrak{f}$ is stable.*

PROOF. Let $U = \bigcap_1^n U_k$, where each $U_k \in \mathcal{S}$. Then $T_k = \bigcap_{S \in \mathfrak{f}} U_k[S] \in \mathfrak{f}$. We claim that $T = \bigcap_1^n T_k \subset \bigcap_{S \in \mathfrak{f}} U[S]$. Assume the contrary, and take $S \in \mathfrak{f}$ such that $T \not\subset U[S]$. Pick a $y \in T \setminus U[S]$. Let $S_k = \{x \in S : y \notin U_k x\}$. For each

$x \in S$, $y \notin Ux$, implying that there is a k with $y \notin U_k x$. Thus $S = \bigcup_1^n S_k$, and so there is a k such that $S_k \in \mathfrak{f}$. Now $y \notin U_k[S_k]$, hence $y \notin T_k \supset T$, a contradiction. □

§ 6 Extensions with stable trace filter pairs

6.1 Let round Cauchy trace filter pairs $\mathfrak{f}^0(a)$ $(a \in Y)$ be prescribed in the quasi-uniform space $(X,\mathcal{U})$. In [De4] § 8 we proved that $\{{}^1U : U \in \mathcal{U}\}$ is a base for a quasi-uniformity ${}^1\mathcal{U}$ on Y, where

$$a\,{}^1U b \quad \text{iff} \quad \text{either } a = b$$
$$\text{or } U[A] \cap B \neq \emptyset \quad \text{for any} \quad A \in \mathfrak{f}^1(a),\ B \in \mathfrak{f}^{-1}(b);$$

${}^1\mathcal{U}|X = \mathcal{U}$; ${}^1\mathcal{U}$ induces the prescribed trace filter pairs iff they are stable; if so then ${}^1\mathcal{U} = {}^0\mathcal{U}$. Thus ${}^1\mathcal{U}$ is not a new extension, but only a simplified description of ${}^0\mathcal{U}$ in the case of stable filter pairs. It is the aim of this section to define an essentially different extension for stable trace filter pairs. We begin with recalling a construction with stable trace f i l t e r s:

6.2 Let us be given round trace filters $\mathfrak{f}(a)$ $(a \in Y)$ in $(X,\mathcal{U})$. For $U \in \mathcal{U}$, define

$$a\,{}^5U b \quad \text{iff} \quad U[S] \in \mathfrak{f}(b) \quad \text{whenever} \quad S \in \mathfrak{f}(a),$$

cf. [Cs3] 6.3, [CsM] 6.3, [De6] 1.7, [Do5], [Do6], [Cs6] § 1. Parts of the next lemma are contained in these papers.

LEMMA. *$\{{}^5U : U \in \mathcal{U}\}$ is always a base for a quasi-uniformity ${}^5\mathcal{U}$ on Y; ${}^5\mathcal{U}|X = \mathcal{U}$. The ${}^5\mathcal{U}^{\mathrm{tp}}$-trace filter of $a \in Y$ is finer than $\mathfrak{f}(a)$ (possibly $\exp X$). ${}^5\mathcal{U}$ is a half-extension inducing the prescribed trace filters iff they are stable; if so then ${}^5\mathcal{U}^{\mathrm{tp}}$ is a strict extension of $\mathcal{U}^{\mathrm{tp}}$. X is ${}^5\mathcal{U}^{-\mathrm{tp}}$-dense in Y iff the filters $\mathfrak{f}(a)$ are Cauchy; if so then $\mathfrak{m}^{-1}(\mathfrak{f}(a))$ is the ${}^5\mathcal{U}^{-\mathrm{tp}}$ trace filter of $a \in Y$.*

PROOF. If $V \subset U$ then ${}^5V \subset {}^5U$, thus $\mathcal{B} = \{{}^5U : U \in \mathcal{U}\}$ is a filter base consisting of entourages. ${}^5V^2 \subset {}^5U$ whenever $V^2 \subset U$, so $\mathcal{B}$ is indeed a base for a quasi-uniformity. If $V^2 \subset U$ then ${}^5V|X \subset U$ and $V \subset {}^5U|X$, implying that ${}^5\mathcal{U}|X = \mathcal{U}$.

For $a \in Y$ fixed, define

$$H_U = \cap\{U[S] : S \in \mathfrak{f}(a)\} \quad (U \in \mathcal{U}).$$

If $V \subset U$ then $H_V \subset H_U$, thus $\{H_U : U \in \mathcal{U}\}$ is a base for a filter $\mathfrak{h}(a)$ (allowing $\mathfrak{h}(a) = \exp X$). It follows from the roundness of $\mathfrak{f}(a)$ that $\mathfrak{f}(a) \subset \mathfrak{h}(a)$. Now

$\mathfrak{f}(a)$ is stable iff $\mathfrak{f}(a) = \mathfrak{h}(a)$ (since, by definition, it is stable iff $\mathfrak{h}(a) \subset \mathfrak{f}(a)$). We claim that $\mathfrak{h}(a)$ is the ${}^{5)}\mathcal{U}^{\mathrm{tp}}$-trace filter of a.

Indeed, ${}^{5)}Ua \cap X \subset H_U$ is clear. Conversely, assuming that $V \in \mathcal{U}$, $V^2 \subset U$, we have $H_V \subset {}^{5)}Ua \cap X$, since if $x \in H_V$ and $S \in \mathfrak{f}(a)$ then $x \in V[S]$, thus $U[S] \in \mathfrak{f}(x)$.

Let each $\mathfrak{f}(a)$ be stable. We have just seen that ${}^{5)}\mathcal{U}$ is a half-extension with the trace filters $\mathfrak{h}(a) = \mathfrak{f}(a)$. ${}^{5)}\mathcal{U}^{\mathrm{tp}}$ is strict, since $\{b \in Y : H_U \in \mathfrak{f}(b)\} \subset {}^{5)}Ua$.

For $a \in Y$, $V \in \mathcal{U}$ and $V^2 \subset U$ we have

$$ {}^{5)}V^{-1}a \cap X \subset M^{-1}(U, \mathfrak{f}(a)) \subset {}^{5)}U^{-1}a \cap X $$

(straightforward from the definitions), thus $\mathfrak{m}^{-1}(\mathfrak{f}(a))$ is the ${}^{5)}\mathcal{U}^{-\mathrm{tp}}$-trace filter of a; it is a proper filter iff $\mathfrak{f}(a)$ is Cauchy (see 5.2). □

COROLLARY. *If the trace filters $\mathfrak{f}(a)$ are stable and (D-)Cauchy then ${}^{5)}\mathcal{U}$ is an extension for the trace filter pairs $(\mathfrak{m}^{-1}(\mathfrak{f}(a)), \mathfrak{f}(a))$; ${}^{5)}\mathcal{U}^{\mathrm{tp}}$ is a strict extension of $\mathcal{U}^{\mathrm{tp}}$.* □

Under the assumptions of the corollary, ${}^{5)}\mathcal{U}$ is not necessarily a doubly strict extension of $\mathcal{U}$:

EXAMPLE. Let $X = \mathbf{R}_0 \times \mathbf{R} \cup \{0^2\}$, $\mathcal{U} = \mathcal{U}(d_{\mathrm{so}} \times d_{\mathrm{eu}})|X$, $Y = \mathbf{R}^2$,

$$ \mathfrak{f}(0, s) = \mathfrak{e}_1 \times \mathfrak{e}(0, s)|X \quad (s \in \mathbf{R}_0). \tag{1} $$

The trace filters are round, stable and D-Cauchy, and

$$ \mathfrak{m}^{-1}(\mathfrak{f}(0, s)) = \mathfrak{e}_{-1} \times \mathfrak{e}(0, s)|X \quad (s \in \mathbf{R}_0). \tag{2} $$

${}^{5)}\mathcal{U}^{-\mathrm{tp}}$ is not a strict extension of $\mathcal{U}^{-\mathrm{tp}}$, because $(0, s)\, {}^{5)}U_{(t)}0^2$ does not hold for any $s \in \mathbf{R}_0$ and $t > 0$, implying ${}^{5)}U_{(t)}^{-1}0^2 \subset X$, although $\mathfrak{e}_{-1} \times \mathfrak{e}$ is the neighbourhood filter of 0^2 in the strict extension of $\mathcal{U}^{-\mathrm{tp}}$ with the trace filters (2). □

6.3 Let us return now to the bitopological case where round Cauchy trace filter pairs $\mathfrak{f}^0(a)$ $(a \in Y)$ are prescribed. ${}^{5)}\mathcal{U}$ taken with $\mathfrak{f}^1(a)$ is not an extension for the given trace filter pairs, even when they are stable, since $\mathfrak{m}^{-1}(\mathfrak{f}^1(a))$ can be strictly coarser than $\mathfrak{f}^{-1}(a)$. Nevertheless, ${}^{5)}\mathcal{U}(\mathfrak{P}_S)$ or ${}^{5)}\mathcal{U}(\mathfrak{P}_S^{\mathfrak{m}})$ could perhaps be used for making the stable Cauchy filter pairs converge since ${}^{5)}\mathcal{U}$ induces in these cases trace filter pairs overlaying $\mathfrak{P}_S$. ${}^{5)}\mathcal{U}$ is, however, not bitopological in the sense of 1.2 g). A bitopological modification ${}^{5}\mathcal{U}$ of ${}^{5)}\mathcal{U}$ can be obtained as follows:

$$ {}^{5}\mathcal{U} = \sup\{ {}^{5)}\mathcal{U}, {}^{(5}\mathcal{U} = {}^{5)}(\mathcal{U}^{-1})^{-1}\} $$

where ${}^{5)}(\mathcal{U}^{\mathfrak{i}})$ is taken with the trace filters $\mathfrak{f}^{\mathfrak{i}}(a)$ $(a \in Y)$. In other words, $\{{}^5U : U \in \mathcal{U}\}$ is a base for ${}^5\mathcal{U}$ where

$$a\,{}^5U b \quad \text{iff} \quad \begin{array}{ll} U[S] \in \mathfrak{f}^1(b) & \text{whenever} \quad S \in \mathfrak{f}^{-1}(a), \\ U^{-1}[S] \in \mathfrak{f}^{-1}(a) & \text{whenever} \quad S \in \mathfrak{f}^{-1}(b). \end{array}$$

We shall write $a\,({}^5U b$ iff the condition in the second line holds. (Thus ${}^5U = {}^{5)}U \cap {}^{(5}U$.) If $\mathcal{B}$ is a base for $\mathcal{U}$ then $\{{}^5U : U \in \mathcal{B}\}$ is a base for ${}^5\mathcal{U}$; the analogous statement for subbases is false.

LEMMA. ${}^5\mathcal{U} \subset {}^1\mathcal{U}$.

PROOF. If $V^2 \subset U$ then ${}^1V \subset {}^5U$. Indeed, if $a\,{}^1V b$, $a \neq b$ and $S \in \mathfrak{f}^1(a)$ then, taking $K \in \mathfrak{f}^{\times}(b)$ with $K \subset V$, we have $V[S] \cap K_{-1} \neq \emptyset$, hence $U[S] \supset \supset K_1 \in \mathfrak{f}^1(b)$, thus $a\,{}^{5)}U b$, i.e., ${}^1V \subset {}^{5)}U$; analogously, ${}^1V \subset {}^{(5}U$. □

6.4 LEMMA. *${}^5\mathcal{U}$ is always a quasi-uniformity such that ${}^5\mathcal{U}|X = \mathcal{U}$. ${}^5\mathcal{U}$ is an extension for the prescribed trace filter pairs iff they are stable.*

PROOF. By Lemma 6.2, ${}^{5)}\mathcal{U}$ and ${}^{(5}\mathcal{U}$ are quasi-uniformities such that their trace on X is $\mathcal{U}$; thus the same holds for their supremum. If the prescribed trace filter pairs are stable then ${}^{5)}\mathcal{U}$ and ${}^{(5}\mathcal{U}$ induce the trace filter pairs $(\mathfrak{m}^{-1}(\mathfrak{f}^1(a)), \mathfrak{f}^1(a))$, respectively $(\mathfrak{f}^{-1}(a), \mathfrak{m}^1(\mathfrak{f}^{-1}(a)))$ (Lemma 6.2). $\mathfrak{m}^{-\mathfrak{i}}(\mathfrak{f}^{\mathfrak{i}}(a))$ is coarser than $\mathfrak{f}^{-\mathfrak{i}}(a)$ (Lemma 5.2), thus ${}^5\mathcal{U}$ induces $\mathfrak{f}^0(a)$. If, say, $\mathfrak{f}^1(a)$ is not $\mathcal{U}$-stable then the ${}^{5)}\mathcal{U}^{\mathrm{tp}}$-trace filter is strictly finer than $\mathfrak{f}^1(a)$ (Lemma 6.2), and the ${}^5\mathcal{U}^{\mathrm{tp}}$-trace filter of a is even finer than that. □

For stable trace filter pairs, ${}^1\mathcal{U}$ and ${}^5\mathcal{U}$ can be different. In fact, ${}^1\mathcal{U}$ induces the fine regular extension associated with the given trace filter pairs ([De4] Theorem 8.7), but ${}^5\mathcal{U}$ need not do so:

EXAMPLE. Let $X = \mathbf{R}_0 \times \mathbf{R}$, $Y = \mathbf{R}^2$, $\mathcal{U} = \mathcal{U}(d_{\mathrm{so}} \times d_{\mathrm{eu}})|X$, and consider the stable trace filter pairs $\mathfrak{e}_0 \times \mathfrak{e}(0,s)|X$ $(s \in \mathbf{R})$. Now ${}^1\mathcal{U}$ is doubly loose, while ${}^5\mathcal{U} = \mathcal{U}(d_{\mathrm{so}} \times d_{\mathrm{eu}})$ is a doubly strict extension (and the two bitopologies are evidently different). □

6.5 Assume now that $\mathcal{U}$ is totally bounded. Then each filter (pair) is stable, thus ${}^1\mathcal{U}$ and ${}^5\mathcal{U}$ are always extensions for the prescribed trace filter pairs. Even now, the bitopologies of ${}^1\mathcal{U}$ and ${}^5\mathcal{U}$ can be different, as the following modification of Example 6.4 shows:

EXAMPLE. Let $X = ([-1,1] \setminus \{0\}) \times [-1,1]$, $Y = [-1,1]^2$, $\mathcal{U} = \mathcal{U}(d \times \times d_{\mathrm{eu}})|X$ with d from Example 2.3 a), $\mathfrak{f}^0(0,s) = \mathfrak{e}_0 \times \mathfrak{e}(0,s)|X$ $(-1 \leqq s \leqq 1)$. □

${}^1\mathcal{U}$ is not totally bounded in this example, since ${}^1\mathcal{U}|Y \setminus X$ is discrete; on the other hand:

PROPOSITION. *If $\mathcal{U}$ is totally bounded then so are ${}^{5)}\mathcal{U}$ and ${}^{5}\mathcal{U}$ taken with arbitrary trace filters (filter pairs).*

PROOF. It is enough to deal with ${}^{5)}\mathcal{U}$, because that result applied to $\mathcal{U}^{-1}$ yields that ${}^{(5}\mathcal{U}$ is totally bounded, too, and the supremum of totally bounded quasi-uniformities has the same property.

Let $\mathfrak{f}(a)$ $(a \in Y)$ be the trace filters. We have to show that for $U \in \mathcal{U}$ fixed, there is a finite cover $\mathfrak{p}$ of Y such that

$$P^2 \subset {}^{5)}U \quad (P \in \mathfrak{p}). \tag{1}$$

As $\mathcal{U}$ is totally bounded, there is a cover $\{A_1, \dots, A_n\} \not\ni \emptyset$ of X such that $A_j^2 \subset U$ $(1 \leqq j \leqq n)$. X being dense in Y, $\mathfrak{c} = \{\mathrm{Cl}^1 A_1, \dots, \mathrm{Cl}^1 A_n\}$ is a cover of Y. We claim that the partition $\mathfrak{p}$ of Y generated by $\mathfrak{c}$ satisfies (1).

It can be assumed without loss of generality that

$$P = \bigcap_1^k \mathrm{Cl}^1 A_j \setminus \bigcup_{k+1}^n \mathrm{Cl}^1 A_j$$

where $1 \leqq k \leqq n$. Let $a, b \in P$, and $S \in \mathfrak{f}(a)$; to prove (1), it is enough to check that $U[S] \in \mathfrak{f}(b)$. For $j \leqq k$, pick $x_j \in S \cap A_j$; then $A_j \subset Ux_j \subset U[S]$, thus $U[S] \supset \bigcup_1^k A_j$, and this union belongs to $\mathfrak{f}(b)$, since if $j > k$ then $b \notin \mathrm{Cl}^1 A_j$ implies that $Y \setminus A_j$ is a neighbourhood of b, hence $\mathfrak{f}(b) \ni \bigcap_{k+1}^n (X \setminus A_j) \subset \bigcup_1^k A_j$. □

6.6 In Example 6.5, the trace filter pairs are not minimal Cauchy, while the minimal Cauchy filter pairs in that space are all linked, and such filter pairs are in general useless in counterexamples (as too many positive results are valid for them). We give here a general method for constructing totally bounded counterexamples starting from not totally bounded ones (in this case: from Example 6.4). This method was already used in [De7] Examples 3.2 and 3.3.

Assume that we have an example built in some way or other on $\mathcal{U}_{\mathrm{eu}}$, $\mathcal{U}_0 = \mathcal{U}_{\mathrm{so}} | \mathbf{R}_0$ and the minimal $\mathcal{U}_0$-Cauchy filter pair $\mathfrak{e}_0 | \mathbf{R}_0$. $\mathcal{U}_{\mathrm{eu}}$ clearly has to be replaced by its trace on an interval. On the other hand, the totally bounded space $(X, \mathcal{U}(d))$ from Example 2.3 a) is not a good substitute for $(\mathbf{R}_0, \mathcal{U}_0)$, since $\mathfrak{e}_0 | X$ is not minimal Cauchy. It is, however, easy to remedy this shortcoming:

EXAMPLE. We give a quasi-uniform space $(X^\circ, \mathcal{U}^\circ)$ and a subspace $(X^{\circ\circ}, \mathcal{U}^{\circ\circ})$ of it such that they are totally bounded, and there is a non-linked minimal $\mathcal{U}^{\circ\circ}$-Cauchy filter pair $\mathfrak{h}^0$. ($\mathcal{U}_0$ will be replaced by $\mathcal{U}^{\circ\circ}$ in the counterexamples, and, occasionally, $\mathcal{U}_{\mathrm{so}}$ by $\mathcal{U}^\circ$). Let

$$X^\circ = \{x \in \mathbf{R}^2 : |x'| = |x''| \leqq 1\}, \quad X^{\circ\circ} = X^\circ \setminus \{0^2\},$$

$$d^\circ(x,y) = d^2_{\mathrm{eu}}(x,y) \quad \text{if} \quad \begin{array}{l}\text{either } x' \leqq 0 \leqq y' \\ \text{or } x'y' > 0 < x''y'',\end{array}$$

$d^{\circ\circ} = d^\circ | X^{\circ\circ}$. $\mathcal{U}^\circ = \mathcal{U}(d^\circ)$ is clearly totally bounded. Let $\mathcal{U}^{\circ\circ} = \mathcal{U}(d^{\circ\circ})$.

$$\mathfrak{h}^0 = \mathfrak{e}_0 \times \mathfrak{e} | X^{\circ\circ} \tag{1}$$

is a minimal $\mathcal{U}^{\circ\circ}$-Cauchy filter pair, which is not linked. There are also other (linked) non-convergent minimal Cauchy filter pairs in $X^{\circ\circ}$, namely

$$\mathfrak{h}^0_{1,-1} = (\mathrm{fil}\,(\mathfrak{e}^2 | X^{\circ\circ} \setminus \mathbb{R}^2_1),\ \mathfrak{e}_1 \times \mathfrak{e}_{-1} | X^{\circ\circ})$$

and three similar ones, denoted by $\mathfrak{h}^0_{1,1}$, $\mathfrak{h}^0_{-1,-1}$ and $\mathfrak{h}^0_{-1,1}$ (the indices show in which quarter-plane the elements of the two filters meet). Check that there are no more non-convergent minimal Cauchy filter pairs. □

6.7 Example (improving Examples 6.4 and 6.5). With X°, $X^{\circ\circ}$ and $d^{\circ\circ}$ from Example 6.6, let $X = X^{\circ\circ} \times [-1,1]$, $Y = X^\circ \times [-1,1]$, $\mathcal{U} = \mathcal{U}(d^{\circ\circ} \times \times d_{\mathrm{eu}} | X)$, $\mathfrak{f}^0(0^2, s) = \mathfrak{h}^0 \times \mathfrak{e}(0^2, s)$ (for $s \in [-1,1]$, with $\mathfrak{h}^0$ from 6.6 (1)). The trace filter pairs are minimal Cauchy. ${}^1\mathcal{U}$ and ${}^5\mathcal{U}$ induce different bitopologies. □

6.8 The following modification of Example 6.2 shows that ${}^5\mathcal{U}$ is not necessarily a doubly strict extension, even if $\mathcal{U}$ is totally bounded and the trace filter pairs are minimal Cauchy:

Example. With the notations of Example 6.6, let $X = X^{\circ\circ} \times [-1,1] \cup \cup \{0^3\}$, $Y = X^\circ \times [-1,1]$, $\mathcal{U} = \mathcal{U}(d^\circ \times d_{\mathrm{eu}}) | X$, $\mathfrak{f}^0(p) = \mathfrak{e}_0 \times \mathfrak{e}^2(p) | X$ $(p \in \in Y \setminus X)$. ${}^{5)}\mathcal{U}^{-\mathrm{tp}}$ is not a strict extension of $\mathcal{U}^{-\mathrm{tp}}$, for the same reason as in Example 6.2. Similarly, ${}^{(5}\mathcal{U}^{\mathrm{tp}}$ is not a strict extension of $\mathcal{U}^{\mathrm{tp}}$; hence neither topology of ${}^5\mathcal{U}$ is a strict extension. □

§ 7 Comparing ${}^5\mathcal{U}$ with other constructions

We shall compare ${}^5\mathcal{U}$ with ${}^2\mathcal{U}$ and ${}^4\mathcal{U}$ introduced in [De4]. This section can be skipped without breaking the continuity of the paper.

7.1 Recall from [De4] § 8 that for $U \in \mathcal{U}$ and trace filter pairs $\mathfrak{f}^0(a)$ $(a \in Y)$, entourages ${}^2U$ and ${}^4U$ on Y are defined as follows:

$a\,{}^2U b$ iff either $a = b$
or there are $K \in \mathfrak{f}^\times(a)$, $L \in \mathfrak{f}^\times(b)$ with $K, L \subset U$, $K_1 \cap L_{-1} \neq \emptyset$,[1]

$a\,{}^4U b$ iff there are $A \in \mathfrak{f}^{-1}(a)$ and $B \in \mathfrak{f}^1(b)$ such that $A \times B \subset U$.

For $k = 2, 4$, $\{{}^kU : U \in \mathcal{U}\}$ is a base for a quasi-uniform extension with the given trace filter pairs iff they are *uniformly concentrated* (which means that they are concentrated, and also *uniformly weakly concentrated,* i.e., the condition in the definition of weak concentratedness holds with U_0 depending only on U, and not on the filter pair); see [De4] Theorems 8.11 and 8.13. In the trace filter pairs are uniformly concentrated then ${}^4\mathcal{U}$ is the coarsest extension for them, and it is doubly strict ([De4] Theorem 8.13).

Let us assume now that ${}^1\mathcal{U}$, ${}^2\mathcal{U}$, ${}^4\mathcal{U}$ and ${}^5\mathcal{U}$ are all extensions for the prescribed trace filter pairs, i.e., that these are stable and uniformly concentrated ; in other words: stable, minimal Cauchy and uniformly weakly concentrated. (The other conditions imply that the trace filter pairs are (weakly) concentrated, cf. Lemma 5.6, but here they have to be uniformly so. A family of Cauchy filter pairs is uniformly weakly concentrated iff 5.1 holds with U_0 depending only on U, see [De4] Lemma 7.15 b).)

We already know that ${}^5\mathcal{U}$ can be different from ${}^1\mathcal{U}$ (Example 6.7), and also from ${}^4\mathcal{U}$ (in Example 6.8, the trace filter pairs are uniformly concentrated, ${}^4\mathcal{U}$ is a doubly strict extension, but ${}^5\mathcal{U}$ is not so). The next example shows that

(i) ${}^2\mathcal{U}$ and ${}^5\mathcal{U}$ are incomparable,

(ii) the bitopology of ${}^5\mathcal{U}$ is not determined by that of $\mathcal{U}$ (i.e., there are two quasi-uniformities inducing the same bitopology such that the trace filter pairs are round, Cauchy and stable with respect to both, but the extensions induce different bitopologies);

(iii) the bitopology of ${}^2\mathcal{U}$ is not determined by that of $\mathcal{U}$ either (this fact was mentioned in [De6] 3.12 without proof).

7.2 EXAMPLE. Let $\mathcal{U} = \mathcal{U}(d^2_{so})|X$ on

$$X = [-1, 0[\times\{0\} \cup \bigcup_{n\in\mathbb{N}} ([-1/n, 1/n] \setminus \{0\}) \times \{1/n\}.$$

On the same set, we consider $\mathcal{V} = \mathcal{U}(d)$, too, where

$$d(x, y) = \begin{cases} y' - x' & \text{if} \quad x'' = y'',\ x' < y', \\ y'' - x'' & \text{if} \quad x' > 0 < y',\ x'' < y'', \\ y'' - x'' - x' & \text{if} \quad x' < 0 < y',\ x'' < y''. \end{cases}$$

$\mathcal{U}$ and $\mathcal{V}$ induce the same bitopology (the sum of the Sorgenfrey bitopologies of the intervals). Let $Y = X \cup \{0^2\} \cup \{0\} \times 1/\mathbb{N}$,

$$\mathfrak{f}^0(p) = \mathfrak{e}^2_0(p)|X \quad (p \in Y \setminus X). \tag{1}$$

[1] In other words, $a\,{}^2U\,b$ iff either $a = b$ or there is an $x \in X$ with $U^{-1}x \in \mathfrak{f}^{-1}(a)$, $Ux \in \mathfrak{f}^1(b)$.

These filter pairs are stable, minimal Cauchy and uniformly weakly concentrated with respect to both quasi-uniformities ($U_{(t/2)}$ is good for $U_{(t)}$ in 5.1 (1)). Hence ${}^2\mathcal{U}$, ${}^5\mathcal{U}$, ${}^2\mathcal{V}$ and ${}^5\mathcal{V}$ are extensions with the trace filter pairs (1). Now the sequence $\langle(0, 1/n)\rangle_{n\in\mathbf{N}}$ converges to 0^2 in ${}^2\mathcal{U}^{\mathrm{tp}}$ and ${}^2\mathcal{V}^{\mathrm{tp}}$, but neither in ${}^2\mathcal{U}^{\mathrm{tp}}$ nor in ${}^5\mathcal{V}^{\mathrm{tp}}$ (because it is not convergent in ${}^{(5}\mathcal{V}^{\mathrm{tp}}$). □

7.3 We conclude this section with an addition to [De4] § 10; see there the definition of the operation $**$.

PROPOSITION. *If the trace filter pairs are stable then* ${}^5\mathcal{U}**\{\{X\}\} = {}^1\mathcal{U}$.

PROOF. Let $U \in \mathcal{U}$ be fixed. Pick $V \in \mathcal{U}$ with $V^2 \subset U$, and assume that $a\,{}^5V**Xb$, $a \neq b$. Then there is an $x \in X$ such that $a\,{}^5Vx\,{}^5Vb$. Now $a\,{}^{5)}Vx$, so for $A \in \mathfrak{f}^1(a)$, $x \in V[A]$; similarly, $x\,{}^{(5}Vb$ implies that for any $b \in \mathfrak{f}^{-1}(b)$, $x \in V^{-1}[B]$. This means that $V^2[A] \cap B \neq \emptyset$, i.e., $U[A] \cap B \neq \emptyset$. Hence ${}^5V**X \subset {}^1U$, and so ${}^1\mathcal{U} \subset {}^5\mathcal{U}**\{\{X\}\}$; the latter is also an extension for the prescribed trace filter pairs ([De4] Theorem 10.5), and ${}^1\mathcal{U}$ is the finest extension, so equality holds. □

Compare this result with [De4] Theorem 10.7 stating that ${}^4\mathcal{U}**\{\{X\}\} = {}^2\mathcal{U}$. Observe also that ${}^2\mathcal{U}**\{\{X\}\} = {}^2\mathcal{U}$. (And evidently ${}^1\mathcal{U}**\{\{X\}\} = {}^1\mathcal{U}$.) In [De4] Example 10.7 b), ${}^5\mathcal{U}$ does not give anything new, since ${}^5\mathcal{U} = {}^4\mathcal{U}$ there (although described with different notations, $\mathcal{U}$ and the trace filter pairs are the same as in Example 6.4 of the present paper.)

§ 8 S-completeness

8.1 DEFINITION. A quasi-uniformity is S-*complete* if any stable Cauchy filter pair is convergent. □

It is enough to consider open, or even round, filter pairs, since if $\mathfrak{f}^0$ is stable then so is $\mathfrak{f}^{0E}$. The next example shows that there is no S-completion of the form ${}^1\mathcal{U}$ (or ${}^0\mathcal{U}$, since the example is totally bounded, so any filter pair is stable, implying ${}^0\mathcal{U} = {}^1\mathcal{U}$).

EXAMPLE. Let A_j, B_j and C_j $(1 \leqq j \leqq 4)$ be disjoint half-open intervals in $\mathbf{R}^2$, with their open ends at 0^2, $A = \bigcup_1^4 A_j$, $B = \bigcup_1^4 B_j$, $C = \bigcup_1^4 C_j$. On $X = A \cup B \cup C$, let $\mathcal{U} = \mathcal{U}(d)$ with

$$d(x, y) = d^2_{\mathrm{eu}}(x, y) \quad \text{if} \quad \begin{array}{l} \text{either } x \text{ and } y \text{ are in the same interval,} \\ \text{or } x \in A_j,\ y \in B_k \text{ with } j \neq k, \\ \text{or } x \in B_j,\ y \in C_k \text{ with } j \neq k, \\ \text{or } x \in A,\ y \in C. \end{array}$$

$\mathcal{U}$ is totally bounded.

$$\mathfrak{h}^0 = (\text{fil}\,(\mathfrak{e}^2|A_1 \cup A_2), \text{fil}\,(\mathfrak{e}^2|B_3 \cup B_4 \cup C))$$

is a non-convergent stable minimal Cauchy filter pair. Assuming that ${}^1\mathcal{U}$ is S-complete with suitable trace filter pairs, there has to be a $p_{12} \in {}^1X \setminus X$ such that $\mathfrak{f}^0(p_{12}) = \mathfrak{h}^0$. Similarly, there is a $q_{13} \in {}^1X \setminus X$ such that

$$\mathfrak{f}^0(q_{13}) = (\text{fil}\,(\mathfrak{e}^2|A \cup B_2 \cup B_4),\ \text{fil}\,(\mathfrak{e}^2|C_1 \cup C_3)),$$

and $p_{34}, q_{24} \in {}^1X \setminus X$ with the role of the indices interchanged. For any $t > 0$, $p_{jk}\,{}^1U_{(t)}q_{mn}$, since there is a $u \in \{1,2,3,4\}$ different from all the indices j,k,m,n, and then arbitrary elements of $\mathfrak{f}^1(p_{jk})$ and $\mathfrak{f}^{-1}(q_{mn})$ meet in B_u. Therefore

$$\mathfrak{k}^0 = (\text{fil}\,\{\{p_{12}, p_{34}\}\}, \text{fil}\,\{\{q_{13}, q_{24}\}\})$$

is a ${}^1\mathcal{U}$-Cauchy filter pair, which has to converge to some $c \in {}^1X$. Take a $K \in \mathfrak{f}^\times(c)$ with $K \subset {}^1U_{(1)}$. Now $p_{12}\,{}^1U_{(1)}c\,{}^1U_{(1)}q_{13}$ and $p_{12} \neq c \neq q_{13}$ (because e.g., $p_{34}\,{}^1U_{(1)}p_{12}$ does not hold), so there are

$$x \in B_3 \cup B_4 \cup C \in \mathfrak{f}^1(p_{12}), \quad y_i \in K_i \in \mathfrak{f}^i(c), \quad z \in B_2 \cup B_4 \cup A \in \mathfrak{f}^{-1}(q_{13})$$

such that $xU_{(1)}y_{-1}U_{(1)}y_1U_{(1)}z$. $x \in B_3 \cup C$ is impossible, since $xU_{(3)}z$ would imply $z \in B_3 \cup C$. Thus $x \in B_4$ and similarly $z \in B_4$; hence $y_{-1}, y_1 \in B_4$, too. Replacing q_{13} by q_{24}, the same reasoning furnishes a $y_1' \in B_3 \cap K_1$, a contradiction, since $y_{-1}U_{(1)}y_1'$ cannot hold. □

8.2 It is our next aim to show that ${}^5\mathcal{U}$, taken with suitable trace filter pairs, is S-complete.

LEMMA. *If the trace filter pairs are stable, and* $\mathfrak{f}^0$ *is an open Cauchy filter pair in* $({}^5X, {}^5\mathcal{U})$ *such that* $\mathfrak{f}^0|X = \mathfrak{f}^0(a)$ *for some* $a \in {}^5X$ *then* $\mathfrak{f}^0$ *converges to* a.

PROOF. For reasons of symmetry, it is enough to show that $\mathfrak{f}^1$ ${}^5\mathcal{U}^{\text{tp}}$-converges to a, i.e., that ${}^{5)}Ua \in \mathfrak{f}^1$ and ${}^{(5}Ua \in \mathfrak{f}^1$ for any $U \in \mathcal{U}$.

1° Let $T = \bigcap\{U[F] : F \in \mathfrak{f}^1(a)\}$. Then $T \in \mathfrak{f}^1(a)$, thus $T \in \mathfrak{f}^1|X$, implying that there is a ${}^5\mathcal{U}^{\text{tp}}$-open $S \in \mathfrak{f}^1$ such that $S \cap X \subset T$. Now $S \subset {}^{5)}Ua$, because if $b \in S$ then $T \supset S \cap X \in \mathfrak{f}^1(b)$, and $U[F] \supset T$ $(F \in \mathfrak{f}^1(a))$, so $a\,{}^{5)}Ub$. Hence ${}^{5)}Ua \in \mathfrak{f}^1$.

2° $\mathfrak{f}^0$ being Cauchy, there is a $K \in \mathfrak{f}^\times$ such that $K \subset {}^{(5}U$. We claim that $K_1 \subset {}^{(5}Ua$; hence ${}^{(5}Ua \in \mathfrak{f}^1$, too. Indeed, assume that $b \in K_1$. Take a $T \in \mathfrak{f}^{-1}(b)$. Then for any $x \in K_{-1} \cap X$, $x\,{}^{(5}Ub$, thus $U^{-1}[T] \in \mathfrak{f}^{-1}(x)$, $x \in U^{-1}[T]$; so $K_{-1} \cap X \subset U^{-1}[T]$, i.e., $U^{-1}[T] \in \mathfrak{f}^{-1}|X = \mathfrak{f}^{-1}(a)$. As $T \in \mathfrak{f}^{-1}(b)$ was arbitrary, we have $a\,{}^{(5}Ub$, so $K_1 \subset {}^{(5}Ua$. □

REMARKS. a) [Cs6] 1.1 (cited in the last paragraph of 2.2) could have been used in 1°, but the application is not straightforward, because ${}^{5)}\mathcal{U}^{\mathrm{tp}}$ can be different from ${}^{5)}\mathcal{U}^{\mathrm{tp}}$.

b) It is used in 1° that $\mathfrak{f}^0|X$ is finer than $\mathfrak{f}^0(a)$, and in 2° the converse. It is in fact not enough to assume that $\mathfrak{f}^0|X$ is finer than $\mathfrak{f}^0(a)$, see later in Example 8.4.

8.3 LEMMA. *Let $(Y,\mathcal{V})$ be a half-extension of $(X,\mathcal{U})$ for stable trace filters such that $\mathcal{V}\subset {}^{5)}\mathcal{U}$ where ${}^{5)}\mathcal{U}$ is taken with the same trace filters, and assume that $\mathfrak{f}$ is a $\mathcal{V}$-stable $\mathcal{V}^{\mathrm{tp}}$-open filter in Y. Then $\mathfrak{f}|X$ is $\mathcal{U}$-stable.*

PROOF. It is enough to show that

$$T\bigcap\{U[S\cap X]: S\in\mathfrak{f},\ S \text{ is } \mathcal{V}^{\mathrm{tp}}\text{-open}\}\in\mathfrak{f}|X, \tag{1}$$

since the sets $S\cap X$ considered here form a base for $\mathfrak{f}|X$. Let a $\mathcal{V}^{\mathrm{tp}}$-open set $S\in\mathfrak{f}$ be fixed. We claim that

$$U[S\cap X]\supset {}^{5)}U[S]\cap X. \tag{2}$$

Indeed, take a y from the right-hand side, and choose $a\in S$ with $a\,{}^{5)}U y$; S being open, $S\cap X\in\mathfrak{f}(a)$, so $U[S\cap X]\in\mathfrak{f}(y)$, therefore y is in the left-hand side of (2). Now $T\supset \bigcap\limits_{S\in\mathfrak{f}} {}^{5)}U[S]\cap X$, and $\bigcap\limits_{S\in\mathfrak{f}} {}^{5)}U[S]\in\mathfrak{f}$, since ${}^{5)}U\in{}^{5)}\mathcal{U}\subset\mathcal{V}$ and $\mathfrak{f}$ is $\mathcal{V}$-stable. Thus $T\in\mathfrak{f}|X$, proving (1). □

COROLLARY. *If the trace filter pairs are stable, and $\mathfrak{f}^0$ is stable and open in $({}^5X,{}^5\mathcal{U})$ or in $({}^1X,{}^1\mathcal{U})$ then $\mathfrak{f}^0|X$ is also stable in $(X,\mathcal{U})$.*

PROOF. ${}^{5)}\mathcal{U}\subset{}^5\mathcal{U}\subset{}^1\mathcal{U}$, so the lemma gives that $\mathfrak{f}^1|X$ is $\mathcal{U}$-stable; $\mathfrak{f}^{-1}|X$ is $\mathcal{U}^{-1}$-stable by the lemma applied to $\mathcal{U}^{-1}$ and the trace filters $\mathfrak{f}^{-1}(a)$. □

8.4 NOTATION. ${}^{\mathsf{S}}\mathcal{U}={}^5\mathcal{U}(\mathfrak{P}_{\mathsf{S}})$.

THEOREM. *${}^{\mathsf{S}}\mathcal{U}$ is always* S*-complete.*

PROOF. Let $\mathfrak{f}^0$ be a round stable Cauchy filter pair in $({}^{\mathsf{S}}X,{}^{\mathsf{S}}\mathcal{U})$. $\mathfrak{f}^0|X$ is round and Cauchy in $(X,\mathcal{U})$; it is also stable by Corollary 8.3. Thus if $\mathfrak{f}^0$ is not a neighbourhood filter pair then it is the trace filter pair of a new point. Hence $\mathfrak{f}^0|X=\mathfrak{f}^0(a)$ with some $a\in{}^{\mathsf{S}}X$, and $\mathfrak{f}^0$ converges to a by Lemma 8.2. □

This theorem shows that any quasi-uniformity has an S-complete extension, but it does not give a good completion. Take e.g., the space $(\mathbf{R},\mathcal{U}_{\mathrm{so}})$. For any $x\in\mathbf{R}$, there are three round stable Cauchy filter pairs different from $\mathfrak{f}^0(x)$ and converging to x; they all have to be taken as trace filter pairs, although $\mathcal{U}_{\mathrm{so}}$ is already S-complete. One would expect on the analogy of some results from [Cs6] and [De7] that ${}^5\mathcal{U}(\mathfrak{P})$ is S-complete whenever $\mathfrak{P}$ overlays

$\mathfrak{P}_S$ (in particular when $\mathfrak{P} = \mathfrak{P}_S^m$), or at least when $\mathfrak{P} = \mathfrak{P}_S^n$. This is, however, not the case:

EXAMPLE. Let X, Y and $\mathcal{U}$ be as in Example 6.2, and take ${}^5\mathcal{U}$ with the trace filter pairs

$$\mathfrak{f}^0(p) = \mathfrak{e}_0 \times \mathfrak{e}(p)|X \quad (p \in Y \setminus X). \tag{1}$$

Each $\mathfrak{f}^0(p)$ is stable and minimal Cauchy. ${}^5\mathcal{U}$ is not S-complete, because

$$\mathfrak{g}^0 = \mathrm{Fil}_Y(\mathfrak{e}^2|Y \setminus X, \mathfrak{e}^2|Y \setminus X) \tag{2}$$

is a stable Cauchy filter pair that does not converge to 0^2, and it clearly cannot converge to anywhere else. It is straightforward to check that all the non-convergent round Cauchy filter pairs figure in (1). Thus ${}^5\mathcal{U}(\mathfrak{P}_S^n) = {}^5\mathcal{U}(\mathfrak{P}_S^m)$ is not S-complete. □

PROBLEM. Does every quasi-uniformity have an S-complete basic extension? For a positive answer, we would need a construction different from both ${}^1\mathcal{U}$ and ${}^5\mathcal{U}$ (Example 8.1 and the one above).

8.5 A quasi-uniformity is called *stable* ([Do5], [Do6]) if each D-Cauchy filter is stable (equivalently: each round D-Cauchy filter is stable, since if $\mathfrak{f}^E$ is stable then so is $\mathfrak{f}$). Similarly to [Do5] Theorem 1:

PROPOSITION. *If $\mathcal{U}$ is doubly stable then ${}^5\mathcal{U}(\mathfrak{P}_C)$ is C-complete.*

PROOF. $\mathfrak{P}_C = \mathfrak{P}_S$ in this case, thus ${}^5\mathcal{U}$ is an extension of $\mathcal{U}$, and all the round Cauchy filter pairs are trace filter pairs. If $\mathfrak{f}^0$ is round and Cauchy in ${}^5X$ then $\mathfrak{f}^0|X = \mathfrak{f}^0(a)$ for some $a \in {}^5X$, and $\mathfrak{f}^0$ converges to a by Lemma 8.2. □

It is important in the proof that $\mathfrak{f}^0$ in Lemma 8.2 was not required to be stable, since it can occur that $\mathcal{U}$ is doubly stable, but ${}^5\mathcal{U}(\mathfrak{P}_C)$ is not so:

EXAMPLE. On X from Example 7.2, let

$$d(x,y) = d_{so}^2(x,y) \quad \text{if} \quad x'' \leqq y'', \quad x' + x''(y'' - x'') \leqq y'. \tag{1}$$

$d_{so}^2(x,y)$ was defined for $x'' \leqq y''$, $x' \leqq y'$; the conditions in (1) are stronger than that, since $x'' \geqq 0$ in X, and $y'' - x'' \geqq 0$ by the first condition. d_{so}^2 is a distance, so in order to prove that d is distance, it is enough to check that if $d(x,y)$ and $d(y,z)$ are defined then so is $d(x,z)$. $x'' \leqq z''$ is evident;

$$x' + x''(z'' - x'') = x' + x''(y'' - x'') + x''(z'' - y'') \leqq y' + y''(z'' - y'') \leqq z'.$$

Let $\mathcal{U} = \mathcal{U}(d)$ and $\mathcal{V} = \mathcal{U}(d_{so}^2|X)$; they induce the same bitopology, because for x fixed there is a $t > 0$ such that if $d_{so}^2(x,y) < t$ then $x'' = y''$, and so $d(x,y) = d_{so}^2(x,y)$ (and the same can be said about the pairs (y,x)). $\mathcal{V} \subset \mathcal{U}$, so any $\mathcal{U}$-Cauchy filter pair $\mathfrak{f}^0$ is also $\mathcal{V}$-Cauchy, therefore it converges to

some $w \in \mathbf{R}^2$ in $\mathcal{U}(d^2_{so})$. If $w \neq 0^2$ then $\mathfrak{f}^0$ is stable because it behaves just like a Cauchy filter pair in $(\mathbf{R}_0, \mathcal{U}_{so}|\mathbf{R}_0)$. So let us assume that $w = 0^2$. Then $\mathfrak{f}^0$ is finer than $\mathfrak{h}^0 = \mathfrak{e}^2_0|X$. We are going to show that $\mathfrak{f}^1$ is $\mathcal{U}$-stable (it is much simpler to check that $\mathfrak{f}^{-1}$ is $\mathcal{U}^{-1}$-stable).

It is enough to know that for any $t > 0$,

$$\bigcap\{U_{(t)}[S] : S \in \mathfrak{f}^1\} \supset]0, t[^2 \cap X, \tag{2}$$

because the right-hand side is in $\mathfrak{h}^1 \subset \mathfrak{f}^1$. To prove (2), let $S \in \mathfrak{f}^1$ be fixed. For $0 < s < t/2$, take $x \in S$ such that $x', x'' < s$ (this is possible, since $\mathfrak{f}^1$ $\mathcal{U}(d^2_{so})^{tp}$-converges to 0^2). Now

$$U_{(t)}x \supset]2s, t[^2 \cap X, \tag{3}$$

since if y is in the right-hand side then $x'' \leqq s \leqq y''$ and $x' + x''(y'' - x'') \leqq s + {} + s \cdot 1 \leqq y'$. From (3) applied to each $0 < s < t/2$ we obtain $U_{(t)}[S] \supset]0, t[^2 \cap X$, which holds for any $S \in \mathfrak{f}^1$, proving (2).

Consider now the filter pairs

$$\mathfrak{f}^0(p) = \mathfrak{e}^2_0(p)|X \quad (p \in \{0\} \times 1/\mathbf{N}), \tag{4}$$

and take ${}^5\mathcal{U}$ such that ${}^5X \supset H = \{0\} \times 1/\mathbf{N}$, with the stable minimal Cauchy filter pairs in (4). We claim that ${}^5\mathcal{U}$ is not stable. Let $\mathfrak{f}$ be the filter in ${}^5X$ generated by the sequence $\langle (0, 1/n)\rangle_{n \in \mathbf{N}}$. $\mathfrak{f}$ is D-Cauchy, with the cofilter $\mathrm{fil}\,(\mathfrak{e}^2_{-1}|X)$. If $p, q \in H$, $p \neq q$ then $p\,{}^{5)}\mathcal{U}q$ never holds, because

$$S =]0, p''[\times \{p''\} \in \mathfrak{f}^1(p), \quad U_{(1)}[S] \notin \mathfrak{f}^1(q).$$

Thus $\mathfrak{f}$ is not ${}^{5)}\mathcal{U}$-stable, hence not ${}^5\mathcal{U}$-stable either. □

8.6 Let us call $\mathcal{U}$ *substable* if $\mathfrak{F}_C$ is overlayed by $\mathfrak{F}_S$ (in the terminology of [Cs6]: $(X, \mathcal{U})$ is a D-space). According to [Cs6] 1.7, if $\mathcal{U}$ is substable then ${}^{5)}\mathcal{U}(\mathfrak{F}_S)$ is D-complete, and so is ${}^{5)}\mathcal{U}(\mathfrak{F})$ with any $\mathfrak{F} \subset \mathfrak{F}_S$ overlaying $\mathfrak{F}_S$. The analogous generalization of Proposition 8.5 is false: in the next example, $\mathfrak{P}_C$ is overlayed by $\mathfrak{P}_S$, and yet ${}^5\mathcal{U}(\mathfrak{P})$ is not C-complete with any $\mathfrak{P} \subset \mathfrak{P}_S$. (This is not at all surprising after Example 8.4.)

EXAMPLE (a modification of Example 8.4). On $X = \mathbf{R}_0 \times \mathbf{R} \cup \{0^2\}$, let $\mathcal{U} = \mathcal{U}(d)$ where

$$d(x, y) = d^2_{eu}(x, y) \quad \text{if} \quad x' \leqq y',\ |x''| \geqq x'.$$

$\mathfrak{P}_C$ is overlayed by $\mathfrak{P}_S$, since any non-stable Cauchy filter pair is convergent.

($\mathcal{U}$ is finer than the restriction of the C-complete quasi-uniformity $\mathcal{U}(d_{so} \times d_{eu})$, so each Cauchy filter pair $\mathfrak{f}^0$ is finer than $\mathfrak{e}_0 \times \mathfrak{e}(a)|X$ with some $a \in \mathbf{R}^2$. If $\mathfrak{f}^0$ is not stable then $|a''| = a'$; the Cauchy property implies

now that $\mathfrak{f}^{-1}$ has no trace on the set $\{x \in X : |x''| < x'\}$; hence $\mathfrak{f}^0$ converges to a.)

Assume that ${}^5\mathcal{U}$ is C-complete. The filter pairs 8.4 (1) are again stable and minimal Cauchy, so there are points in ${}^5X$ with these trace filter pairs; as in 8.4, we identify the corresponding new point with the elements of $\{0\} \times \times \mathbb{R}_0$. The filter pair $\mathfrak{g}^0$ from 8.4 (2) is Cauchy, so it has to converge to some $c \in {}^5X$. $\mathfrak{g}^{0E}$ also converges to c, and

$$\mathfrak{g}^{0E}|X = \mathrm{Fil}_X(\mathfrak{e}_0 \times \mathfrak{e}|\mathbb{R}_0 \times \mathbb{R}) = \mathfrak{h}^0.$$

Now $\mathfrak{h}^0$ is finer than $\mathfrak{f}^0(c)$, a contradiction, since $\mathfrak{f}^0(0^2)$ is the only stable Cauchy filter pair coarser than $\mathfrak{h}^0$, and $\mathfrak{g}^0$ does not converge to 0^2. □

§ 9 A modification of D-completeness

This section can again be skipped; it deals with the non-symmetric analogue of S-completeness, and contains also some remarks on D-completeness and on stable spaces.

9.1 It is proved [Do6] that if $\mathcal{U}$ is stable then ${}^{5)}\mathcal{U}(\mathfrak{F}^n_S)$ is D-complete; it is in fact enough to know that $\mathcal{U}$ is substable, see [Cs6]. Although any quasi-uniformity has D-complete half-extensions, cf. 2.2, ${}^{5)}\mathcal{U}$ is of interest because it has two additional properties: it is an extension, not just a half-extension (to avoid ambiguity, we shall use the expression *double extension* in this section), and ${}^{5)}\mathcal{U}^{tp}$ is a strict extension of $\mathcal{U}^{tp}$. It is not known whether each quasi-uniformity has a D-complete double extension; concerning strictness, see 9.4. A complete double extension can be obtained, however, if we change the definition of D-completeness:

DEFINITION. A quasi-uniformity is SD-*complete* if each stable (D-)Cauchy filter is convergent.

NOTATION. ${}^{SD}\mathcal{U} = {}^{5)}\mathcal{U}(\mathfrak{F}^n_S)$.

PROPOSITION. *${}^{SD}\mathcal{U}$ is a basic* SD*-completion; it is a double extension, and ${}^{SD}\mathcal{U}^{tp}$ is a strict extension of $\mathcal{U}^{tp}$. If $\mathcal{U}$ is a uniformity then ${}^{SD}\mathcal{U}$ is its usual completion.*

PROOF. ${}^{SD}\mathcal{U}^{tp}$ is strict and $\mathcal{U}$ is a double extension by Lemma 6.2. Let $\mathfrak{f}^1$ be a stable D-Cauchy filter in ${}^{SD}X$, $\mathfrak{f}^{-1}$ a cofilter of $\mathfrak{f}^1$. Then $\mathfrak{h}^0 = \mathfrak{f}^{0E}|X$ is a Cauchy filter pair, and $\mathfrak{h}^1$ is $\mathcal{U}$-stable by Lemma 8.3. Thus $\mathfrak{h}^1$ converges to some $a \in {}^{SD}X$, and then so does $\mathfrak{f}^1$, see in the second paragraph of "To Cs" in 2.2. If $\mathcal{U}$ is a uniformity then it is stable, thus ${}^{SD}\mathcal{U} = {}^{5)}\mathcal{U}(\mathfrak{F}^n_D)$, which is the usual uniform completion ([Do5] Proposition 6). □

The following can be added to the table in 2.2:

$$\mathsf{SD} \quad + \quad + \quad \begin{cases} {}^{\mathrm{SD}}\mathcal{U} & + & - & - \\ {}^{0)}\mathcal{U}(\mathfrak{F}_{\mathsf{S}}^{\mathrm{n}}) & - & - & + \end{cases}$$

${}^{0)}\mathcal{U}(\mathfrak{F}_{\mathsf{S}}^{\mathrm{n}})$ is SD-complete, because if $\mathfrak{f}$ is a non-convergent stable (D-)Cauchy filter in ${}^{0)}X$ then $X \in \mathfrak{f}$, implying that $\mathfrak{f}|X$ is $\mathcal{U}$-stable; it is also Cauchy (see the reference to [Cs5] in 2.2 "To MN"), hence D-Cauchy. (It was possible to avoid here the more complicated reasoning from [Cs6] 2.3. Lemma 8.3 could also be used when establishing that $\mathfrak{f}|X$ is stable.)

REMARK. Stable analogues of MN-, FN- and W-completeness could also be introduced, cf. the last paragraph of 2.1. The following implications hold between the non-symmetric notations of completeness (we do not investigate whether there hold others, too; squares stand for the notions having no name):

$$\begin{array}{ccccccccc} & & \mathsf{SP} & \Rightarrow & \mathsf{MN} & \Rightarrow & \mathsf{FN} & \Rightarrow & \mathsf{W} \\ & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\ \mathsf{K} & \Rightarrow & \mathsf{D} & \Rightarrow & \square & \Rightarrow & \square & \Rightarrow & \square \\ \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\ \square & \Rightarrow & \mathsf{SD} & \Rightarrow & \square & \Rightarrow & \square & \Rightarrow & \square \\ \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\ \square & \Leftrightarrow & \mathsf{Cs} & \Leftrightarrow & \square & \Rightarrow & \square & \Rightarrow & \square \end{array}$$

The reverse implication in the left bottom corner is valid, because if $\mathfrak{f}^0$ is linked and $\mathfrak{f}^1$ $\mathcal{U}^{\mathrm{tp}}$-converges to x then x is clearly a $\mathcal{U}^{\mathrm{tp}}$-cluster point of $\mathfrak{f}^{-1}$. Below K, it is not clear whether we should consider first members of stable Cauchy filter pairs, or rather the filters having a $\mathcal{U}$-stable $\mathcal{U}^{-1}$-cofilter.

9.2 Both in ${}^{5)}\mathcal{U}$ and ${}^{0)}\mathcal{U}$ above, it is enough to take trace filters overlaying $\mathfrak{F}_{\mathsf{S}}$ (the same proof; stability has to be assumed in the case of ${}^{5)}\mathcal{U}$). We cannot, however, take ${}^{5)}\mathcal{U}(\mathfrak{F}_{\mathsf{S}}^{\mathrm{m}})$ or ${}^{0)}\mathcal{U}(\mathfrak{F}_{\mathsf{S}}^{\mathrm{m}})$, since $\mathfrak{F}_{\mathsf{S}}^{\mathrm{m}}$ does not necessarily overlay $\mathfrak{F}_{\mathsf{S}}$: in Example 2.3 b), $\mathfrak{f}_1^1 \supset \mathfrak{f}_2^1 \supset \dots$ are stable and Cauchy, but their intersection is not Cauchy. We are going to show that $\mathfrak{F}_{\mathsf{S}}^{\mathrm{m}}$ overlays $\mathfrak{F}_{\mathsf{S}}$ in a special class of spaces.

DEFINITION. A quasi-uniformity is *(strongly) cominimal* if each D-Cauchy filter is (strongly) cominimal. □

For a quasi-uniformity $\mathcal{U}$, we have the following properties, each implying the next one: $\mathcal{U}$ is totally bounded, $\mathcal{U}^{-1}$ is hereditarily precompact, $\mathcal{U}$ is stable (5.8), $\mathcal{U}$ is strongly cominimal (5.2), $\mathcal{U}$ is cominimal (5.2).

LEMMA. *If $\mathcal{U}^{-1}$ is cominimal then $\mathfrak{F}_{\mathsf{S}}^{\mathrm{m}}$ overlays $\mathfrak{F}_{\mathsf{S}}$.*

PROOF. We intend to use Zorn's Lemma. Let $\mathfrak{F} \subset \mathfrak{F}_{\mathsf{S}}$ be ordered by inclusion. By Lemma 5.2, $(\mathrm{m}^{-1}(\mathfrak{f}), \mathfrak{f})$ is a Cauchy filter pair ($\mathfrak{f} \in \mathfrak{F}$), and

if $\mathfrak{f} \subset \mathfrak{g}$ then $\mathfrak{m}^{-1}(\mathfrak{f}) \supset \mathfrak{m}^{-1}(\mathfrak{g})$. Thus $\mathfrak{H} = \{\mathfrak{m}^{-1}(\mathfrak{f}) : \mathfrak{f} \in \mathfrak{F}\}$ is also ordered by inclusion, $\mathfrak{h} = \bigcup \mathfrak{H}$ is a filter. $\mathfrak{h}$ is $\mathcal{U}^{-1}$-D-Cauchy because it is finer than such filters. $\mathcal{U}^{-1}$ being cominimal, there is a coarsest filter $\mathfrak{k}$ for which $(\mathfrak{h}, \mathfrak{k})$ is Cauchy. $\mathfrak{h} \supset \mathfrak{m}^{-1}(\mathfrak{f})$ implies that $(\mathfrak{h}, \mathfrak{f})$ is also Cauchy, thus $\mathfrak{k} \subset \mathfrak{f}$ $(\mathfrak{f} \in \mathfrak{F})$, and so $(\mathfrak{h}, \bigcap \mathfrak{F})$ is Cauchy, too. $\bigcap \mathfrak{F}$ is also stable (Lemma 5.4). □

PROPOSITION. *If $\mathcal{U}^{-1}$ is cominimal (in particular, if $\mathcal{U}$ is hereditarily precompact) then $\mathcal{U}$ has an SD-complete hull, which is a double extension as well as a basic strict half-extension; for uniformities, it coincides with the usual completion.*

PROOF. ${}^{5)}\mathcal{U}(\mathfrak{F}_{\mathrm{S}}^{\mathrm{m}})$ has the required properties: use the lemma, and the observation before the definition. If $\mathcal{U}$ is a uniformity then $\mathfrak{F}_{\mathrm{S}}^{\mathrm{m}} = \mathfrak{F}_{\mathrm{C}}^{\mathrm{m}} = \mathfrak{F}_{\mathrm{R}}$. □

Similarly, if $\mathcal{U}$ is substable and $\mathcal{U}^{-1}$ is cominimal (in particular, if $\mathcal{U}$ is doubly stable) then ${}^{5)}\mathcal{U}\mathfrak{F}_{\mathrm{D}}^{\mathrm{m}}$ is a D-complete hull with the properties mentioned in the proposition. In Example 8.5, $\mathcal{U}$ is doubly stable, but ${}^{5)}\mathcal{U}(\mathfrak{F}_{\mathrm{D}}^{\mathrm{m}})$ is not stable. Let us also note that a quasi-uniformity can be stable and hereditarily precompact without being totally bounded, see [KJ], after the proof of Corollary 3.

9.3 The existence of a D-complete double extension can be guaranteed, besides for substable spaces, in another special case, too:

PROPOSITION. *Assume that there is in $(X, \mathcal{U})$ a finite family of D-Cauchy filters overlaying $\mathfrak{F}_{\mathrm{D}}^{\mathrm{n}}$. Then $\mathcal{U}$ has a D-complete double extension which is also a strict half-extension.*

PROOF. Let $\mathfrak{F}$ denote the finite family of filters. Taking the envelopes and then discarding the superfluous filters, we may assume that the elements of $\mathfrak{F}$ are round and incomparable (i.e., none of them is finer than any other). Let $\mathfrak{f}^{-1}$ be a $\mathcal{U}^{-1}$-round cofilter of $\mathfrak{f}^1$ $(\mathfrak{f}^1 \in \mathfrak{F})$, and $\mathfrak{P} = \{\mathfrak{f}^0 : \mathfrak{f}^1 \in \mathfrak{F}\}$. Now ${}^{0}\mathcal{U}(\mathfrak{P})$ is a double extension. It is also strict by the Lemma below, thus ${}^{0}\mathcal{U}(\mathfrak{P})$ is D-complete ([Cs6] 1.3). □

LEMMA. *There belongs only one topological extension to any finite family of incomparable non-convergent open trace filters given in a topological space.*

PROOF. We show that the loose extension is strict. With the usual notations, let $a \in Y$, and N a neighbourhood of a in the loose extension. Then there is an open $S \in \mathfrak{f}(a)$ such that $\{a\} \cup S \subset N$. For each $p \in Y \setminus (X \cup \{a\})$, $\mathfrak{f}(a) \not\subset \mathfrak{f}(p)$ (either by the incomparability, or because $\mathfrak{f}(p)$ is not convergent), so it can be assumed that $S \notin \mathfrak{f}(p)$ $(a \neq p \in Y \setminus X)$, and therefore $\{a\} \cup S$ is a neighbourhood of a in the strict extension. □

REMARKS. a) If $Y \setminus X$ is finite then $\mathcal{U}^{-\mathrm{tp}}$-density is not needed in [Cs6] 1.3.

b) (Á. Császár.) The construction from [Cs4] 7.6 can also be used if the filters in $\mathfrak{F}$ satisfy the additional condition called there "weakly Cauchy"

(*strictly tame* in [De6]); in this case it is not necessary to make the elements of $\mathfrak{F}$ incomparable.

c) $\mathcal{U}$ from Example 9.4 restricted to $\mathbf{R}_0 \times \{0\} \cup \{0\} \times \mathbf{R}_0$ shows that the conditions of the proposition can hold without the filters being strictly tame.

PROBLEM. Is it true that if $\mathfrak{P}^n_C$ is overlayed by a finite family of Cauchy filter pairs then there is a C-complete extension?

9.4 Any quasi-uniformity has a D-complete (in fact, SP-complete) strict half-extension: if $(X,\mathcal{U})$ is not SP-complete then $\mathfrak{f} = \{X\}$ is not convergent, and so ${}^{0)}\mathcal{U}(\{\mathfrak{f}\})$ is strict. It would be better to have D-complete strict half-extensions with D-Cauchy trace filters; in the next example, there does not exist such a half-extension.

EXAMPLE. On $X = \mathbf{R}^2$, let $\mathcal{U} = \mathcal{U}(d)$, where

$$d(x,y) = d^2_{\mathrm{eu}}(x,y) \quad \text{if} \quad \begin{array}{l} \text{either } x'' \neq 0 = y'', \\ \text{or } x'' = y'' = 0,\ x' < y'. \end{array}$$

Assume that $(Y,\mathcal{V})$ is a D-complete strict half-extension of $(X,\mathcal{U})$ with D-Cauchy trace filters. Then there are points $p_t \in Y \setminus X$ $(t \in \mathbf{R})$ such that $\mathfrak{f}(p_t) = \mathfrak{e}(t) \times \mathrm{fil}_{\mathbf{R}}\{\{0\}\}$, since these filters are minimal D-Cauchy. For each $a \in \mathbf{R} \times \{0\} \cup \{p_t : t \in \mathbf{R}\} = Z$, $\mathbf{R} \times \{0\} \in \mathfrak{f}(a)$, therefore $(\mathcal{V}|Z)^{\mathrm{tp}}$ is a strict extension of $(\mathcal{U}|\mathbf{R} \times \{0\})^{\mathrm{tp}}$. Identifying $\mathbf{R} \times \{0\}$ with $\mathbf{R}$, we obtain that $\mathcal{U}_{\mathrm{so}}$ has a strict half-extension with the trace filters $\mathfrak{e}(t)$ $(t \in \mathbf{R})$, a contradiction according to the example after [Cs4] Theorem 6.1. □

(Received November 12, 1991)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 433–441

ZUR KONSTRUKTION DES REGULÄREN SIEBZEHNECKS

J. STROMMER

Meinem lieben Freund, Herrn Prof. H. Sachs zum 50. Geburtstag gewidmet

Abstract

At the construction of the regular 17-gon one has to solve a chain of dependent quadratic equations. All the authors of the various constructions have been concentrating on geometric representation of the roots of these equations. As F. Klein ([5], p. 19 and 26), F. Enriques ([2], p. 175), Th. Vahlen ([9], p. 155) and later also H. Lebesgue ([6], pp. 149–150) emphasized their wishes to construct the regular 17-gon on the base of purely geometric analysis. This is the intention of this paper giving such a discussion, which can be treated also in a textbook of plane geometry.

Von Alters her sind uns von den regulären Polygonen mit ungerader Seitenzahl nur die Konstruktion der Polygone von 3, 5 und 15 Seiten bekannt. Erst Gauss hat im Jahre 1796 bewiesen, daß auch die Konstruktion des regulären Siebzehnecks mit Zirkel und Lineal durchführbar ist. Man hat zu diesem Zweck eine Reihe zusammenhängender, quadratischer Gleichungen konstruktiv aufzulösen. Die Entdecker der verschiedenen Konstruktionen beschränkten sich darauf, die Wurzeln dieser Gleichungen geometrisch darzustellen. Wie schon F. Klein ([5], S. 19 u. 26), F. Enriques ([2], S. 175), Th. Vahlen ([9], S. 155) und später auch H. Lebesgue ([6], S. 149–150) bemerkt haben, wäre es wünschenswert eine nur aus geometrischen Erwägungen abgeleitete Konstruktion des Siebzehnecks zu haben.[1)] Im Folgenden geben wir eine *rein geometrische* Analyse, die in jedem Lehrbuch der Geometrie Platz finden könnte, und aus der sich die bekannten Konstruktionen leicht ableiten lassen.

1991 *Mathematics Subject Classifications.* Primary 51M15; Secondary 51M20.
Key words and phrases. Constructions, regular polygons.

[1)] Eine rein geometrische Analyse muß Erchinger gehabt haben, von dem Gauss in den *Göttingischen gelehrten Anzeigen* (19. Dezember 1825) berichtet (s. [3], S. 187): „Das eigentlich Verdienstliche der Abhandlung des Hrn. Erchinger beruht ... in der rein geometrischen Begründung ... und diese ist mit so musterhafter mühsamer Sorgfalt, alles nicht rein Elementarische zu vermeiden, durchgeführt, daß sie dem Verf. zur Ehre gereicht." Diese Abhandlung ist aber verlorengegangen (vgl. [4], S. 15 u. 68).

0081–6906/95/$ 4.00

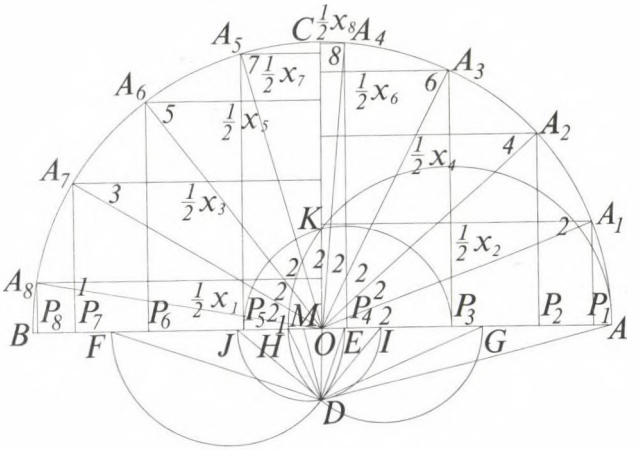


Fig. 1

Zu diesem Zweck sei O der Mittelpunkt eines Kreises mit dem Radius $OA = 1$ (Fig. 1), in den das reguläre 17-Eck $AA_1A_2 \ldots A_{16}$ einbeschrieben werden soll. Es seien $P_1, P_2, \ldots, P_8$ die Projektionen der Eckpunkte $A_1, A_2, \ldots, A_8$ auf den Durchmesser AB des Kreises. Die von denselben Eckpunkten auf den zu AB senkrechten Radius OC gefällten Lote sind gleich der Hälfte je einer der Diagonalen eines demselben Kreis einbeschriebenen regulären 34-Ecks. Wir bezeichnen diese Diagonalen in abnehmender Reihenfolge ihrer Größe mit $x_1, x_2, \ldots, x_8$. Dann gilt:

$$\begin{aligned} 2 \cdot OP_1 &= x_2, & 2 \cdot OP_5 &= x_7, \\ 2 \cdot OP_2 &= x_4, & 2 \cdot OP_6 &= x_5, \\ 2 \cdot OP_3 &= x_6, & 2 \cdot OP_7 &= x_3, \\ 2 \cdot OP_4 &= x_8, & 2 \cdot OP_8 &= x_1. \end{aligned}$$

In der Figur sind die Maßzahlen der einzelnen Winkel in Bezug auf $\frac{\pi}{17}$ als Winkeleinheit angegeben.

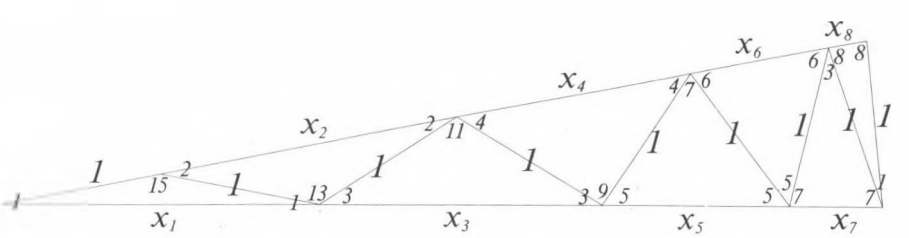


Fig. 2

Aus jenen gleichschenkeligen Dreiecken, deren Basen der Reihe nach gleich $x_1, x_2, \ldots, x_8$ und deren Schenkel alle gleich 1 sind, kann man ein

gleichschenkeliges Dreieck zusammensetzen (s. Fig. 2).[2] Aus diesem großen gleichschenkeligen Dreieck folgt unsere *Grundgleichung*:

$$x_1 + x_3 + x_5 + x_7 - x_2 - x_4 - x_6 - x_8 = 1.$$

Aus der Figur ergibt sich ferner die Relation $1 : \frac{1}{2}x_1 = x_1 : 1 + \frac{1}{2}x_2$, d.h.

$$x_1^2 = 2 + x_2;$$

ebenso $1 : \frac{1}{2}x_1 = x_2 : \frac{1}{2}(x_1 + x_3)$, d.h.

$$x_1 x_2 = x_1 + x_3;$$

usw. Die so erhaltenen Relationen für die Produkte von je zweien der Größen $x_1, x_2, \ldots, x_8$ stellen wir in der folgenden Multiplikationstabelle[3] zusammen:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$2 + x_2$	$x_1 + x_3$	$x_2 + x_4$	$x_3 + x_5$	$x_4 + x_6$	$x_5 + x_7$	$x_6 + x_8$	$x_7 - x_8$
x_2	$x_1 + x_3$	$2 + x_4$	$x_1 + x_5$	$x_2 + x_6$	$x_3 + x_7$	$x_4 + x_8$	$x_5 - x_8$	$x_6 - x_7$
x_3	$x_2 + x_4$	$x_1 + x_5$	$2 + x_6$	$x_1 + x_7$	$x_2 + x_8$	$x_3 - x_8$	$x_4 - x_7$	$x_5 - x_6$
x_4	$x_3 + x_5$	$x_2 + x_6$	$x_1 + x_7$	$2 + x_8$	$x_1 - x_8$	$x_2 - x_7$	$x_3 - x_6$	$x_4 - x_5$
x_5	$x_4 + x_6$	$x_3 + x_7$	$x_2 + x_8$	$x_1 - x_8$	$2 - x_7$	$x_1 - x_6$	$x_2 - x_5$	$x_3 - x_4$
x_6	$x_5 + x_7$	$x_4 + x_8$	$x_3 - x_8$	$x_2 - x_7$	$x_1 - x_6$	$2 - x_5$	$x_1 - x_4$	$x_2 - x_3$
x_7	$x_6 + x_8$	$x_5 - x_8$	$x_4 - x_7$	$x_3 - x_6$	$x_2 - x_5$	$x_1 - x_4$	$2 - x_3$	$x_1 - x_2$
x_8	$x_7 - x_8$	$x_6 - x_7$	$x_5 - x_6$	$x_4 - x_5$	$x_3 - x_4$	$x_2 - x_3$	$x_1 - x_2$	$2 - x_1$

Aus den Größen x können wir nun folgende, aus je zwei verschiedenen Faktoren bestehende Produkte zusammenstellen:

$$x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_1x_6, x_1x_7, x_1x_8;$$
$$x_2x_3, x_2x_4, x_2x_5, x_2x_6, x_2x_7, x_2x_8;$$
$$x_3x_4, x_3x_5, x_3x_6, x_3x_7, x_3x_8;$$
$$x_4x_5, x_4x_6, x_4x_7, x_4x_8;$$
$$x_5x_6, x_5x_7, x_5x_8;$$
$$x_6x_7, x_6x_8;$$
$$x_7x_8.$$

[2] Es ist leicht zu sehen, daß, wenn n eine natürliche Zahl der Form $2k+1$ ist ($k > 1$) und $\varphi = \frac{\pi}{n}$, aus jenen gleichschenkeligen Dreiecken, deren Basiswinkel $\varphi, 2\varphi, \ldots, k\varphi$ und deren Schenkel alle gleich 1 sind, sich immer ein gleichschenkeliges Dreieck zusammenstellen läßt. Die Basis des gleichschenkeligen Dreiecks mit dem Basiswinkel $k\varphi$, ist gleich der Seite des dem Kreis vom Radius 1 einbeschriebenen regulären $2n$-Ecks. Diese Figur verwendet ein anonymer Araber, der vor ungefähr tausend Jahren lebte, und später auch Vieta, um jene (kubische) Gleichung abzuleiten, von der die Konstruktion des regelmäßigen Siebenecks abhängt (vgl. [9], S. 83).

[3] Diese Tabelle verwendet Th. Vahlen ([9], S. 148–152), um jene quadratischen Gleichungen abzuleiten, auf deren Lösung sich die Teilung des Kreises in 17 Teile zurückführen läßt.

Wir gehen nun von dem Produkt x_1x_2 aus und suchen diejenigen Größen x auf, deren Summe bzw. Differenz mit diesem Produkt gleich ist. Dann suchen wir diejenigen Größen x auf, deren Summe bzw. Differenz mit dem Produkt der so erhaltenen Größen x gleich ist, usw. So kommen wir nach acht Schritten zu jenem Produkt zurück, von dem wir ausgegangen sind, während wir der Reihe nach die, in der Spalte I aufgeschriebenen Relationen erhalten. Wenn wir von dem Produkt x_1x_3, x_1x_8, x_2x_4, x_2x_6, x_4x_5, x_4x_8 oder x_7x_8 ausgehen, so bekommen wir dieselben Relationen in derselben zyklischen Reihenfolge. Wenn wir aber von dem ersten, in der Spalte I nicht vorkommenden Produkt x_1x_4 ausgehen, dann erhalten wir die in der Spalte II aufgeschriebenen Relationen. Auf dieselbe Weise erhalten wir die in der Spalte III und zuletzt die in der Spalte IV aufgeschriebenen Relationen.

I.	II.	III.	IV.
$x_1x_2 = x_1 + x_3,$	$x_1x_4 = x_3 + x_5,$	$x_1x_5 = x_4 + x_6,$	$x_1x_6 = x_5 + x_7,$
$x_1x_3 = x_2 + x_4,$	$x_3x_5 = x_2 + x_8,$	$x_4x_6 = x_2 - x_7,$	$x_5x_7 = x_2 - x_5,$
$x_2x_4 = x_2 + x_6,$	$x_2x_8 = x_6 - x_7,$	$x_2x_7 = x_5 - x_8,$	$x_2x_5 = x_3 + x_7,$
$x_2x_6 = x_4 + x_8,$	$x_6x_7 = x_1 - x_4.$	$x_5x_8 = x_3 - x_4,$	$x_3x_7 = x_4 - x_7,$
$x_4x_8 = x_4 - x_5,$		$x_3x_4 = x_1 + x_7,$	$x_4x_7 = x_3 - x_6,$
$x_4x_5 = x_1 - x_8,$		$x_1x_7 = x_6 + x_8,$	$x_3x_6 = x_3 - x_8,$
$x_1x_8 = x_7 - x_8,$		$x_6x_8 = x_2 - x_3,$	$x_3x_8 = x_5 - x_6,$
$x_7x_8 = x_1 - x_2.$		$x_2x_3 = x_1 + x_5.$	$x_5x_6 = x_1 - x_6.$

Die in einer und derselben Spalte stehenden Gleichungen drücken die unter acht bzw. vier über der Verbindungsstrecke von je zwei Punkten P als Durchmesser beschriebenen, einander in bestimmter, zyklischer Reihenfolge folgenden Kreisen bestehenden Relationen aus, nachdem die Potenz von O in Bezug auf einen dieser Kreise gleich der Entfernung des Mittelpunktes des folgenden Kreises von O ist. Beispielsweise kann man die Relation $x_2x_6 = x_4 + x_8$ auch so schreiben $4 \cdot OP_1 \cdot OP_3 = 2(OP_2 + OP_4)$, d.h. man hat

$$OP_1 \cdot OP_3 = \frac{1}{2}(OP_2 + OP_4);$$

hier ist das Produkt auf der linken Seite gleich der Potenz von O in Bezug auf den über P_1P_3 als Durchmesser beschriebenen Kreis und der Ausdruck auf der rechten Seite ist die Entfernung des Mittelpunktes des über P_2P_4 als Durchmesser beschriebenen Kreises von O.

Da die Gleichungen der Spalte II zwischen der Potenz von O in Bezug nur auf vier, die über P_3P_5, P_6P_7, P_1P_4, P_2P_8 als Durchmesser beschriebenen Kreise und der Entfernung der Mittelpunkte E, F, G, H derselben Kreise von O bestehenden Beziehungen ausdrücken und außerdem für diese Entfernungen

$$OE = \frac{1}{2}(OP_3 - OP_5) = \frac{1}{4}(x_6 - x_7),$$

$$OF = \frac{1}{2}(OP_6 + OP_7) = \frac{1}{4}(x_3 + x_5),$$

$$OG = \frac{1}{2}(OP_1 + OP_4) = \frac{1}{4}(x_2 + x_8),$$
$$OH = \frac{1}{2}(OP_8 - OP_2) = \frac{1}{4}(x_1 - x_4)$$

auch unsere Grundgleichung drückt eine Bedingung aus, so wird es unser Ziel sein, zwischen den Punkten E, F, G, H weitere Beziehungen zu finden, mit deren Hilfe diese Punkte konstruiert werden können. Wenn wir nämlich diese Punkte kennen, so können wir das gesuchte Polygon auf verschiedene Weise konstruieren. Gemäß der letzten Gleichung der Spalte II ist z.B.

$$OP_3 \cdot OP_5 = OH = OA \cdot OH$$

und daher schneiden sich die über AH und P_3P_5 als Durchmesser gezeichneten Kreise in einem Punkt K des Radius OC (s. Fig. 1). Kennt man also die Punkte E und H, so kann man die Punkte P_3 und P_5 und mit deren Hilfe auch das gesuchte Polygon konstruieren.

Indem wir zwischen den Punkten E, F, G, H neue Beziehungen suchen, bemerken wir zunächst, daß die über EF und GH als Durchmesser beschriebenen Kreise die Verlängerung des Halbmesser OC in demselben Punkt D schneiden. In der Tat gilt

$$\begin{aligned} OE \cdot OF &= \frac{1}{16}(x_6 - x_7)(x_3 + x_5) \\ &= \frac{1}{16}(x_3x_6 - x_3x_7 + x_5x_6 - x_5x_7) \\ &= \frac{1}{16}(x_3 - x_8 - x_4 + x_7 + x_1 - x_6 - x_2 + x_5) \end{aligned}$$

und so ist infolge unserer Grundgleichung

$$OE \cdot OF = \frac{1}{16}.$$

In ähnlicher Weise zeigt man, daß

$$OG \cdot OH = \frac{1}{16}(x_2 + x_8)(x_1 - x_4) = \frac{1}{16},$$

also

$$OE \cdot OF = OG \cdot OH$$

gilt. Außerdem ist $OD = \frac{1}{4}$.

Ferner bemerken wir, daß G die Strecke EF innerlich in demselben Verhältnis teilt, wie H äußerlich. Es ist nämlich

$$\begin{aligned} EG \cdot FH &= (OG - OE)(OF - OH) \\ &= \frac{1}{16}(x_2 + x_8 - x_6 + x_7)(x_3 + x_5 - x_1 + x_4) \\ &= \frac{1}{16}(-x_1 + x_2 + x_3 + x_4 + x_5 - x_6 + x_7 + x_8) \end{aligned}$$

und

$$\begin{aligned} FG \cdot EH &= (OF + OG)(OE + OH) \\ &= \frac{1}{16}(x_3 + x_5 + x_2 + x_8)(x_6 - x_7 + x_1 - x_4) \\ &= \frac{1}{16}(-x_1 + x_2 + x_3 + x_4 + x_5 - x_6 + x_7 + x_8), \end{aligned}$$

also $EG \cdot FH = FG \cdot EH$, bzw.

$$EG : FG = EH : FH.$$

Ferner folgt, wegen $\sphericalangle EDF = \sphericalangle GDH = R$, daß

$$\sphericalangle GDE = \sphericalangle EDH = \sphericalangle HDE = \frac{1}{2}R$$

gilt.

Es sei nun I derjenige Punkt der Geraden AB, der die Strecke EF äußerlich in demselben Verhältnis teilt, wie O innerlich. Dann gilt die Proportion $OE : OF = EI : FI$ und somit $OE \cdot FI = OF \cdot EI$ bzw. $OE(OF + OI) = OF(OI - OE)$, und folglich

$$OI = \frac{2 \cdot OF \cdot OE}{OF - OE} = \frac{(x_6 - x_7)(x_3 + x_5)}{2(x_3 + x_5 - x_6 + x_7)},$$

oder wegen $(x_6 - x_7)(x_3 + x_5) = 1$ schließlich

$$OI = \frac{1}{2(x_3 + x_5 - x_6 + x_7)}.$$

Da $\sphericalangle EDF = R$ ist, so hat man ferner $\sphericalangle EDI = \sphericalangle ODE$.

Es sei nun J derjenige Punkt der Geraden AB, der die Strecke GH äußerlich in demselben Verhältnis teilt, wie O innerlich. Dann gilt die Proportion $OG : OH = GJ : HJ$ und somit $OG \cdot HJ = OH \cdot GJ$ bzw. $OG(OJ - OH) = OH(OG + OJ)$, und man findet

$$OJ = \frac{2 \cdot OG \cdot OH}{OG - OH} = \frac{(x_2 + x_8)(x_1 - x_4)}{2(x_2 + x_8 - x_1 + x_4)}.$$

Beachtet man, daß $(x_2 + x_8)(x_1 - x_4) = 1$ ist, so ergibt sich

$$OJ = \frac{1}{2(x_2 + x_8 - x_1 + x_4)}.$$

Wegen $(x_2 + x_5 - x_6 + x_7)(x_2 + x_8 - x_1 + x_4) = 4$ erhält man

$$OI \cdot OJ = \frac{1}{16} = \overline{OD}^2$$

und daher

$$OI = \frac{1}{16 \cdot OJ} = \frac{1}{8}(x_2 + x_8 - x_1 + x_4)$$

bzw.

$$OJ = \frac{1}{16 \cdot OI} = \frac{1}{8}(x_3 + x_5 - x_6 + x_7).$$

Demnach ist I bzw. J die Mitte von GH bzw. EF.

Es sei nun M der Mittelpunkt von IJ. Dann gilt

$$OM = \frac{1}{16}(x_3 + x_5 - x_6 + x_7 - x_2 - x_8 + x_1 - x_4) = \frac{1}{16},$$

und hieraus folgt $\sphericalangle ADM = R$, wegen $OA \cdot OM = \frac{1}{16} = \overline{OD}^2$.

Wir bemerken noch, daß der Punkt A die Strecke IJ äußerlich in demselben Verhältnis teilt, wie O innerlich, und demnach gilt $\sphericalangle ODI = \sphericalangle IDA$, also $\sphericalangle 4 \cdot ODE = \sphericalangle ODA$.

In der Tat ist $OJ - OI = \frac{1}{8} = 2 \cdot OI \cdot OJ$, d.h. $OJ(1 - OI) = OI(1 + OJ)$, also $OJ(OA - OI) = OI(OA + OJ)$, und schließlich $OJ \cdot AI = OI \cdot AJ$, bzw.

$$AI : AJ = OI : OJ.$$

Aus unseren Überlegungen ergibt sich folgende Konstruktion des dem Kreise mit dem Mittelpunkt O und dem Radius OA einbeschriebenen regulären 17-Ecks (s. Fig. 1):

Wir zeichnen in dem Kreis den zu AB senkrechten Halbmesser OC und verlängern ihn um die Strecke $OD = \frac{1}{4}OA$. In D errichten wir die Senkrechte auf AD, welche OA in M schneidet. Dann beschreiben wir einen Kreis um M mit dem Radius MD, der OA in I und OB in J schneidet. Auf der Verlängerung von BJ und AI tragen wir das Stück $JE = JD$ und $IH = ID$ ab. Über AH als Durchmesser beschreiben wir einen Kreis, welcher OC in K schneidet. Dann beschreiben wir von E durch K einen Kreis, der AB in den Punkten P_3 und P_5 schneidet. Die in diesen Punkten auf AB errichteten Senkrechten schneiden den gegebenen Kreis in den Eckpunkten A_3, A_{14} und A_5, A_{12} des gesuchten Polygons.

Diese Konstruktion stammt allem Anschein nach von H. Lebesgue. Er leitete sie aus der allgemeinen algebraischen Theorie der Kreisteilung ab (s. [6], S. 145–148).

Aus den obigen Überlegungen folgt auch die folgende Konstruktion des 17-Ecks (s. Fig. 1):

Auf der Verlängerung des Halbmessers OC tragen wir wieder das Stück $OD = \frac{1}{4}OA$ ab. Dann bestimmen wir auf OA und OB die Punkte E und H so, daß $\sphericalangle EDO = \sphericalangle \frac{1}{4}ADO$ und $\sphericalangle EDH = \frac{1}{2}R$ wird. Mit Hilfe der Punkte E, H können wir dann die Ecken A_3, A_{14} und A_5, A_{12} des gesuchten Polygons ebenso finden wie oben.

Diese Konstruktion hat H. Richmond im Jahre 1893 angegeben. Der Beweis bei ihm (s. [7] und [8]) beruht auf der Bemerkung, daß die geometrische Lösung quadratischer Gleichungen, wenn die Koeffizienten gewisse Bedingungen erfüllen (u.a. die absoluten Glieder ganze Zahlen sind) auf Winkelhalbierung zurückgeführt werden kann.

Auf Grund unserer Erörterungen können wir die bekannten Konstruktionen des regulären 17-Ecks (s. z.B. [4]) leicht beweisen; so z.B. auch die ziemlich geläufige *Serret–Bachmann*sche Konstruktion, welche in den Lehrbüchern, die die Siebzehnteilung des Kreises behandeln, stets zu finden ist (s. z.B. [1], S. 216–217, und [2], S. 177–179).

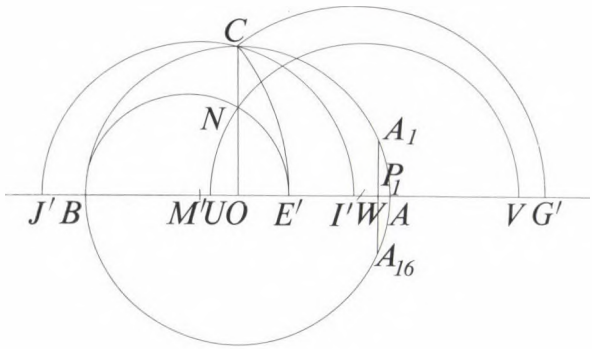


Fig. 3

Es sei $OA = 1$ der Radius des gegebenen Kreises, welcher in 17 gleiche Teile geteilt werden soll (Fig. 3). Senkrecht zum Durchmesser AB des Kreises zeichnen wir den Radius OC. Dann tragen wir auf OB das Stück $OM' = \frac{1}{4}OA$ ab und zeichnen um M' mit dem Radius $M'C$ einen Kreis, der OA in I' und OB in J' schneidet. Aus J' mit dem Radius $J'C$ beschreiben wir einen Kreisbogen, welcher AB in E' schneidet. Ferner beschreiben wir aus I' mit dem Radius $I'C$ einen Kreisbogen, welcher die Verlängerung von AB in G' schneidet. Dann beschreiben wir über BE' als Durchmesser einen Halbkreis, der OC in N schneidet. Wir stechen mit der Strecke $\frac{1}{2}OG'$ in N ein und schlagen einen Kreisbogen, welcher AB in W schneidet. Wenn dann der aus W durch N gezeichnete Kreis AB in U und deren Verlängerung in V schneidet, so ist OU gleich der Seite des dem gegebenen Kreis einbeschriebenen 34-Ecks und der Mittelpunkt von OV ist die Projektion P_1 der A benachbarten Punkte A_1 und A_{16} des gesuchten regulären 17-Ecks.

In der Tat gilt

$$UV = OG' = 4 \cdot OG = x_2 + x_8$$

und

$$OU \cdot OV = \overline{ON}^2 = OB \cdot OE' = OE' = 4 \cdot OE = x_6 - x_7 = x_2 x_8,$$

und somit folgt $OU = x_8$ und $OV = x_2$.

LITERATURVERZEICHNIS

[1] ADLER, A., *Theorie der geometrischen Konstruktionen*, Sammlung Schubert, LII, G. J. Göschen, Leipzig, 1906. *Jb. Fortschritte Math.* **37**, 511

[2] ENRIQUES, F., *Fragen der Elementargeometrie*, II. Teil: Die geometrischen Aufgaben, ihre Lösung und Lösbarkeit, Deutsch von H. Fleischer, B. G. Teubner, Leipzig, 1907. *Jb. Fortschritte Math.* **38**, 520

[3] GAUSS, C. FR., *Werke*, Hrsg. von der Königl. Gesellschaft der Wissenschaften zu Göttingen II, Höhere Arithmetik, Leipzig, 1863.

[4] GOLDENRING, R., *Die elementargeometrischen Konstruktionen des regelmäßigen Siebzehnecks*, B. G. Teubner, Leipzig und Berlin, 1915. *Jb. Fortschritte Math.* **45**, 752

[5] KLEIN, F., *Vorträge über ausgewählte Fragen der Elementargeometrie*, ausgearb. von F. Tägert, B. G. Teubner, Leipzig, 1895. *Jb. Fortschritte Math.* **26**, 546

[6] LEBESGUE, H., *Leçons sur les constructions géométriques*, Gauthier-Villars, Paris, 1950. *MR* **11**-678

[7] RICHMOND, H. W., A construction for a regular polygon of seventeen sides, *Quart. J. Pure Appl. Math.* **26** (1893), 206–207. *Jb. Fortschritte Math.* **25**, 909

[8] RICHMOND, H. W., To construct a regular polygon of 17 sides, *Math. Ann.* **67** (1909), 459–461. *Jb. Fortschritte Math.* **40**, 557

[9] VAHLEN, TH., *Konstruktionen und Approximationen in systematischer Darstellung*, Teubners Lehrbücher der mathematischen Wissenschaften, XXXIII, B. G. Teubner, Leipzig und Berlin, 1911. *Jb. Fortschritte Math.* **42**, 501

(Eingegangen am 15. November, 1991)

BUDAPESTI MŰSZAKI EGYETEM
GEOMETRIA TANSZÉKE
STOCZEK U. 4
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 443–446

EXTERNAL CHARACTERIZATION OF GENERALIZED MANIFOLDS

M. BOGNÁR

Dedicated to Professor Ákos Császár on his 70th birthday

We shall describe a type of generalized manifolds that is related to some fundamental properties of n-manifolds with boundary lying in the $(n+1)$-dimensional euclidean space R^{n+1}.

The starting point is a theorem of Schoenflies [6]. He proved in 1902 that if M is a compact set in R^2 such that $R^2 \setminus M$ has two components and every point of M is accessible from each of these components then M is a closed Jordan curve.

Since each Jordan curve in R^2 admits these properties above and the Jordan curve is the only compact connected 1-manifold, the Schoenflies theorem gives an external characterization of Jordan curves or of compact connected 1-manifolds lying in R^2.

The question whether the compact connected 2-manifolds lying in R^3 could be characterized in an analogous way was answered by Brouwer in 1911 in the negative [3].

The next important step was the result of Kaluzsay [5]. On the encouragement of Frédéric Riesz he stated and proved in 1915 that a compact subset M of R^3 must be homeomorphic to a 2-sphere if it satisfies the following three conditions:

(1) $R^3 \setminus M$ has two components;

(2) Each point of M is accessible from every component of $R^3 \setminus M$;

(3) Every closed polygon in $R^3 \setminus M$ is contractible continuously to a point in $R^3 \setminus M$.

However, while conditions (1) and (2) are fulfilled for each topological sphere in R^3 condition (3) fails to be always satisfied. Alexander constructed in 1924 a topological sphere in R^3 which does not satisfy the third condition of Kaluzsay [1].

Finally, Wilder characterized in 1929 and 1930 the 2-spheres [7] and the connected compact 2-manifolds [8] in R^3 in the following manner:

A compact set M in R^3 is a 2-sphere if and only if

(1′) $R^3 \setminus M$ has two components and M is the common boundary of each of these components.

1991 *Mathematics Subject Classification.* Primary 57P99; Secondary 57N05.
Key words and phrases. Generalized manifolds, k-manifolds.

(2′) The components of $R^3 \setminus M$ are uniformly locally connected, i.e., for each component D of $R^3 \setminus M$ and for every positive real ε there is a positive real δ such that for any two points x and y of D which are nearer to each other than δ x and y may be connected by an arc in D whose diameter is less than ε.

(3′) The 1-dimensional Betti number (mod 2) of $R^3 \setminus M$ vanishes.

M is a connected 2-manifold if and only if besides the preceding first and second conditions the 1-dimensional Betti number (mod 2) of $R^3 \setminus M$ is finite.

Starting from this latter theorem Wilder has defined $(n-1)$-dimensional generalized manifolds such that those embedded in R^n may be characterized as compact sets in R^n possessing analogous external properties as the 2-manifolds in R^3 [9].

However, these external properties are also of algebraic character so as the third condition in the case $n=3$.

We want to escape from the algebraic conditions. Generally, we are satisfied if our figures will be manifolds in the 1-dimensional case. We expect only in the triangulable case that the 2-dimensional figures should be manifolds.

We shall define objects in arbitrary T_2-space which could be considered as generalized manifolds.

Let R be a T_2-space and (X, A) a compact pair in R, i.e., X is a compact set in R and A is a closed subset of X.

A domain (a connected nonvoid open set) V in R is said to be k-regular mod (X, A) if the following conditions hold:

(a) $V \cap A = \emptyset$;

(b) $V \cap X$ is a domain in X;

(c) $V \setminus X$ consists of two components;

(d) the closure of each component of $V \setminus X$ contains $V \cap X$.

The compact pair (X, A) itself is called a *k-manifold in R* if it satisfies the following two conditions:

(a′) $X \setminus A$ is a nonempty connected space;

(b′) for every $q \in X \setminus A$ the k-regular domains that contain the point q form a basis for the neighbourhood system of the point q in R.

Now if R is the 2-euclidean space R^2 and (X, A) is a k-manifold in R^2 then $X \setminus A$ is either a closed Jordan curve or it is homeomorphic to the real line R^1 (see [2] 4.7). Hence $X \setminus A$ is always a 1-manifold. If $A = \emptyset$ then X is a closed Jordan curve. The proof of these facts depends on a theorem of Á. Császár [4] which says: A separable connected locally connected complete metric space which is not a singleton and fails to contain any triode is homeomorphic either to a circle or to the line R^1 or to a segment of R^1 or to a closed halfline of R^1.

If $R = R^3$ then we can find a k-manifold (X, A) in R^3 such that $X \setminus A$ is not a 2-manifold. This compact pair (X, A) can be constructed as follows:

Let A be the frame of the unit square

$$Q^2 = \{(x, y, z) \in R^3;\ 0 \leqq x \leqq 1,\ 0 \leqq y \leqq 1,\ z = 0\},$$

i.e.,

$$A = Q^2 \setminus \{(x, y, z) \in R^3;\ 0 < x < 1,\ 0 < y < 1,\ z = 0\}.$$

For $q \in R^3$ and $\lambda \in \mathbf{R}$ let $\psi(q, \lambda)$ be the central similarity of R^3 with the center q and ratio λ, i.e., $\psi(q, \lambda)$ is the map $\psi(q, \lambda) : R^3 \to R^3$ defined by the formula

$$\psi(q, \lambda)(w) = \lambda w + (1 - \lambda)q \qquad (w \in R^3).$$

Let $q_1 = (0, 0, 0)$, $q_2 = \left(\frac{1}{2}, 0, 0\right)$, $q_3 = (1, 0, 0)$, $q_4 = \left(1, \frac{1}{2}, 0\right)$, $q_5 = (1, 1, 0)$, $q_6 = \left(\frac{1}{2}, 1, 0\right)$, $q_7 = (0, 1, 0)$ and $q_8 = \left(0, \frac{1}{2}, 0\right)$. Let

$$B = \left\{(x, y, z) \in R^3;\ \frac{1}{3} \leqq x \leqq \frac{2}{3},\ \frac{1}{3} \leqq y \leqq \frac{2}{3},\ z = \frac{1}{3}\right\},$$

$$C = \left\{(x, y, z) \in R^3;\ x \in \left\{\frac{1}{3}, \frac{2}{3}\right\},\ \frac{1}{3} \leqq y \leqq \frac{2}{3},\ 0 \leqq z \leqq \frac{1}{3}\right\},$$

$$D = \left\{(x, y, z) \in R^3;\ \frac{1}{3} \leqq x \leqq \frac{2}{3},\ y \in \left\{\frac{1}{3}, \frac{2}{3}\right\},\ 0 \leqq z \leqq \frac{1}{3},\ \left(x - \frac{1}{2}\right)^2 + \left(z - \frac{1}{6}\right)^2 \geqq \frac{1}{50}\right\},$$

$$E = \left\{(x, y, z) \in R^3;\ \left(x - \frac{1}{2}\right)^2 + \left(z - \frac{1}{6}\right)^2 = \frac{1}{50},\ \frac{1}{3} \leqq y \leqq \frac{2}{3}\right\}$$

and let $X_1 = B \cup C \cup D \cup E$ (see Figures 1 and 2).

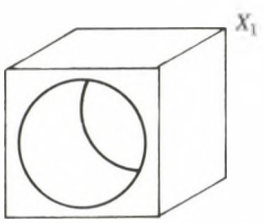


Fig. 1

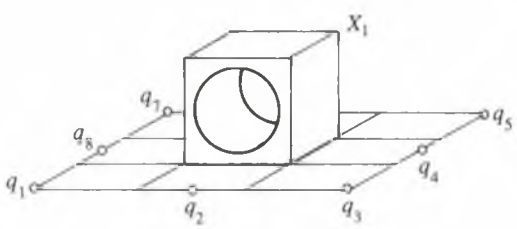


Fig. 2

For $i = 1, 2, \ldots, 8$ let $\psi_i = \psi\left(q_i, \frac{1}{3}\right)$. We construct the sets $X_1, \ldots, X_k, \ldots$ in a recursive way. X_1 is already defined. For $k \geqq 2$ let

$$X_k = X_{k-1} \cup \bigcup_{i=1}^{8} \psi_i(X_{k-1}) = X_1 \cup \bigcup_{i=1}^{8} \psi_i(X_{k-1}).$$

Let X be the closure of $\bigcup_{k=1}^{\infty} X_k$, i.e., $X = \overline{\bigcup_{k=1}^{\infty} X_k}$. Obviously, X is a compact set and $A \subset X$. It is easy to see that (X, A) is a k-manifold in R^3 and $X \setminus A$ is not a 2-manifold. We omit the proof of these facts.

If Y is the cone over A with the center $(0, 0, -1)$ and $Z = X \cup Y$ then $(Z, \emptyset)$ is a k-manifold in R^3, too. Observe that the counterexample of Brouwer is not a k-manifold in R^3.

Notice that in R^4 we can find triangulable k-manifolds (Y, B) where the $Y \setminus B$-s are not 3-manifolds. Indeed, let B be a torus in $R^3 = \{(x, y, z, w) \in \in R^4;\ w = 0\}$, and let Y be the cone over B with the center $(0, 0, 0, 1)$. Then (Y, B) is a triangulable compact pair, it is a k-manifold in R^4 but $Y \setminus B$ is obviously not a 3-manifold.

However, if (X, A) is a triangulable compact pair, which is a k-manifold in R^3 then $X \setminus A$ is a 2-manifold.

The proof of this last statement is also omitted.

REFERENCES

[1] ALEXANDER, J. W., An example of a simply connected surface bounding a region which is not simply connected, *Proc. Nat. Acad. USA.* **10** (1924), 8–10. *Jb. Fortschritte Math.* **50**, 661

[2] BOGNÁR, M., Cohomological pseudomanifolds, *Acta Math. Sci. Hungar.* **57** (1991), 91–109. *MR* **93b**:57019

[3] BROUWER, L. E. J., Über Jordansche Mannigfaltigkeiten, *Math. Ann.* **71** (1911), 320–327. *Jb. Fortschritte Math.* **42**, 418

[4] CSÁSZÁR, Á., Sur les courbes atriodiques, *Acta Math. Acad. Sci. Hungar.* **9** (1958), 329–332. *MR* **21** # 2964

[5] KALUZSAY, K., A felületre vonatkozó Jordan tétel megfordítása, *Math. és Phys. Lapok* **24** (1915), 101–104. *Jb. Fortschritte Math.* **45**, 1383

[6] SCHOENFLIES, A., Über einen grundlegenden Satz der Analysis situs, *Gött. Nachr.* (1902), 185–192. *Jb. Fortschritte Math.* **33**, 502

[7] WILDER, R. L., A converse of the Jordan–Brouwer theorem in three dimensions, *Bull. Amer. Math. Soc.* **35** (1929), 771. *Jb. Fortschritte Math.* **55**, 332

[8] WILDER, R. L., A converse of the theorem regarding the separations of E_3 by a closed two dimensional manifold of genus p, *Bull. Amer. Math. Soc.* **36** (1930), 219. *Jb. Fortschritte Math.* **56**, 516

[9] WILDER, R. L., Generalized closed manifolds in n-space, *Ann. of Math.* (2) **35** (1934), 876–903. *Jb. Fortschritte Math.* **60**, 535 and 1221

(Received June 10, 1994)

EÖTVÖS LÓRÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 447–459

WHY IS THE POTENTIAL LOGARITHMIC IN THE PLANE?

J. HORVÁTH

Dedicated to Ákos Császár on the occasion of his seventieth birthday

1. George Green and Carl Friedrich Gauss

The first hero of my tale is George Green (1793–1841), a self-taught scientist and a baker by trade in Nottingham, England. In 1828 he published a booklet under the title "An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism". The Essay was dedicated to the Duke of Newcastle, Lord Lieutenant of the County of Nottingham, with whose financial help it was published, and there were also fifty-one subscribers; regrettably the dedication disappeared from the collected edition of Green's work [7].

In a long footnote at the beginning of the Essay, Green recalls that when two small bodies are charged with electricity, the force of repulsion will be proportional to the product of the charges and inversely proportional to the n-th power of their distance. He adds that Charles Augustin Coulomb (1736–1806) has shown in 1785 that "n is equal to two".

The Introductory Observations of the Essay consider a body carrying a charge specified by a density. Green says that the force which the body exerts on some charged point "will be expressed by a partial differential of a certain function of the coordinates which serve to define the point's position in space. The consideration of this function is of great importance ... and [we] will call it *potential function*...".

With modern notation we say that if the body is the domain Ω in $\mathbf{R}^3$ and f is the density of the mass or charge on Ω, then the value of the potential at the point $x = (x_1, x_2, x_3)$ is

$$U(x) = \iiint_{\Omega} \frac{f(y)}{|x-y|}\, dx,$$

and the force exerted on the unit charge placed at x is $\mathbf{F}(x) = \operatorname{grad} U(x)$. Actually the function U was already introduced by Pierre Simon marquis

1991 *Mathematics Subject Classification.* Primary 35J05, 46F10; Secondary 31B10.

Key words and phrases. Potential, Riesz kernel, Schwartz distribution, analytic continuation.

0081–6906/95/$ 4.00

de Laplace (1749–1827), who proved that it satisfies the equation $\Delta U = 0$, where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$$

is the operator named after him. Laplace omitted to specify that $\Delta U = 0$ holds only outside Ω. This was done by Denis Poisson (1781–1840) who stated in 1813 that U satisfies the Poisson equation $\Delta U = -4\pi f$, without, however, giving a rigorous justification for it [6, p.54].

Let me immediately introduce the (Newtonian) *potential* in $\mathbb{R}^n$ for any integer $n \geqq 3$. If f is a twice continuously differentiable function whose support

$$\operatorname{Supp} f = \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}$$

is a compact subset of $\mathbb{R}^n$, then we set

$$U(x) = \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-2}}\, dy,$$

where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ and $|x| = (x_1^2 + \ldots + x_n^2)^{\frac{1}{2}}$. The function U satisfies the Poisson equation

$$\Delta U = \frac{\partial^2 U}{\partial x_1^2} + \ldots + \frac{\partial^2 U}{\partial x_n^2} = -\frac{4\pi^{n/2}}{\Gamma\left(\dfrac{n-2}{2}\right)} f.$$

The deep reason for this relation and the significance of the numerical factor on the right should become clear before the end of this lecture. Observe that these formulas are meaningless when $n = 2$.

In no. 3 of his Essay Green proves what is now called "Green's theorem" of integral calculus:

$$\iiint U \Delta V\, dx + \iint U \frac{\partial V}{\partial n}\, d\sigma = \iiint V \Delta U\, dx + \iint V \frac{\partial U}{\partial n}\, d\sigma,$$

and puts the calculations of Laplace and Poisson on solid grounds. He also proves (no. 2) that the charge on a conductor Ω is carried by its surface Σ, and (no. 4) that if the potential U is given in the interior and the exterior of Ω, then the charge on Σ is determined by the jump of the normal derivative of U at Σ. In no. 5 he considers what we, following B. Riemann, now call *Dirichlet's problem*: Given f on Σ, he looks for a function V such that $\Delta V = 0$ inside Ω and $V = f$ on Σ. He introduces the Green function

$$G(x, y) = \frac{1}{|x-y|} + h(x, y)$$

which is such that for all $x \in \Omega$ one has $G(x, y) = 0$ if $y \in \Sigma$ and $\Delta_y h(x, y) = 0$ if $y \in \Omega$. He gives the formula

$$V(x) = \frac{1}{4\pi} \iint_{\Sigma} f(y) \frac{\partial G}{\partial n}(x, y)\, d\sigma(y)$$

for the solution. As to the touchy question of the existence of G, Green offers the following famous argument: "To convince ourselves that there exists such a function ... G ...; conceive the surface to be a perfect conductor put in communication with the earth, and a unit of positive electricity to be concentrated in the point x, then the total potential function arising from x and from the electricity it will induce on the surface will be the required value of G." It took about seventy years to give a mathematical proof for the existence of G.

In England Green's talent was soon recognized. He entered Gonville and Caius College (Cambridge) in 1833 and was elected a fellow in 1839. He wrote several more papers on mathematical physics of which "On the determination of the exterior and interior attractions of ellipsoids of variable densities" [7, pp. 185–222] has the particular interest that it is one of the very first places where n-dimensional space is considered throughout. According to the magnificent history of mathematics in the 19-th century of Felix Klein [17, p.19], the discovery of his talent did not benefit Green. Once in Cambridge, he succumbed to alcohol which caused his untimely death.

On the Continent Green's Essay became known only much later. It is reasonable to assume that when in 1839 C. F. Gauss (1777–1855) wrote his "Allgemeine Lehrsätze in Beziehung auf die im verkehrten Verhältnisse des Quadrates der Entfernung wirkenden Anziehungs- und Abstoßungs-Kräfte" [6] he was totally unaware of Green's contributions, which he rediscovers. He introduces the potential with the following words [6, p. 6]: "Zur bequemeren Handhabung ... werden wir uns erlauben, dieses V mit einer besondern Benennung zu belegen, und diese Größe das *Potential* der Massen, worauf sie sich bezieht, nennen".

Gauss tried to prove the existence of an *equilibrium distribution*, i.e. of a charge whose potential is constant on the whole conductor, by minimizing the integral which expresses the energy of the charge [6, no. 30]. He writes: "... offenbar muß für Eine solche Vertheilungsart ein Minimumwerth dieses Integrals stattfinden." Later generations did not find the existence of a minimum obvious. Also Gauss could not have given a correct proof because the equilibrium distribution may not have a density and he did not have the Stieltjes integral at his disposition.

To conclude this section, let me mention that Otto Hölder proved in his 1882 Stuttgart dissertation that if f only satisfies the condition $|f(x) - f(y)| = O(|x - y|^{\alpha})$, with $\alpha > 0$, named after him, then U is twice differentiable and satisfies the Poisson equation [16, pp. 152–156].

2. Marcel Riesz and Otto Frostman

The existence of an equilibrium distribution was proved only in 1935, almost one-hundred years after Gauss' Algemeine Lehrsätze, by the Swedish mathematician Otto Frostman in his Lund thesis "Potentiels d'Equilibre et Capacité des Ensembles".

For his proof Frostman uses the generalized potentials introduced a little earlier by his teacher Marcel Riesz (1886–1959) but published only later [22, nos. 40, 42, 47]. For $n \geqq 2$ and $0 < \operatorname{Re}\alpha < n$ the *Riesz potential of order* α is defined on $\mathbf{R}^n$ by

$$\mathcal{R}_\alpha f(x) = \frac{\Gamma\left(\frac{n-\alpha}{2}\right)}{2^\alpha \pi^{n/2} \Gamma\left(\frac{\alpha}{2}\right)} \int \frac{f(y)}{|x-y|^{n-\alpha}}\, dy,$$

where f is assumed to satisfy appropriate conditions. If $n \geqq 3$, then $\alpha = 2$ is in the range considered and yields the Newtonian potential. Among the properties of the operator $\mathcal{R}_\alpha$ Riesz emphasizes the composition formula

$$\mathcal{R}_\alpha(\mathcal{R}_\beta f) = \mathcal{R}_{\alpha+\beta} f \tag{1}$$

and the relations

$$\Delta \mathcal{R}_\alpha f = -\mathcal{R}_{\alpha-2} f \quad \text{and} \quad \lim_{\alpha \to 0+} \mathcal{R}_\alpha f = f. \tag{2}$$

If we define $\mathcal{R}_0 f = f$, we get formally the Poisson equation $\Delta \mathcal{R}_2 f = -\mathcal{R}_0 f = -f$ and we see that $-\Delta$ behaves like what one would expect from $\mathcal{R}_{-2}$. Using to-day's terminology, we can say that $\mathcal{R}_\alpha f$ is the convolution $R_\alpha * f$ of f with the Riesz kernel of order α defined by

$$R_\alpha(x) = \frac{\Gamma\left(\frac{n-\alpha}{2}\right)}{2^\alpha \pi^{n/2} \Gamma\left(\frac{\alpha}{2}\right)} |x|^{\alpha-n}.$$

For $0 < \operatorname{Re}\alpha < n$ the function R_α is locally integrable, and for $\frac{n}{2} < \operatorname{Re}\alpha < n$ it is locally square integrable.

Motivated by his research on the Cauchy problem for the wave equation, Riesz considered the analytic continuation into the half-plane $\operatorname{Re}\alpha \leqq 0$ of the holomorphic function $\alpha \mapsto \mathcal{R}_\alpha f(x)$, and proved that under appropriate conditions on f the formulas (1) and (2) remain valid for $\alpha_0 < \operatorname{Re}\alpha < n$, in particular $\mathcal{R}_{-2k} f = (-\Delta)^k f$. He also made the false statement that $\alpha \mapsto \mathcal{R}_\alpha f(x)$ is holomorphic in the whole half-plane $\operatorname{Re}\alpha > \alpha_0$ [22, p. 593].

Frostman considers potentials

$$U^\sigma(x) = \int \frac{d\sigma(y)}{|x-y|^{n-\alpha}}$$

of not necessarily positive charges (i.e. set functions or measures) σ and the mutual energy

$$(\sigma|\tau) = \int U^{\sigma}(x)d\tau(x) = \iint \frac{d\sigma(x)d\tau(y)}{|x-y|^{n-\alpha}}$$

of two such charges σ and τ. One of his main Lemmas (no. 16, p. 28) states that $(\sigma|\sigma) \geqq 0$, and that $(\sigma|\sigma) = 0$ only if $\sigma = 0$. The positivity is an immediate consequence of Riesz' composition formula (1) since

$$\frac{\Gamma\left(\frac{n-\alpha}{2}\right)}{2^{\alpha}\pi^{n/2}\Gamma\left(\frac{\alpha}{2}\right)}(\sigma|\sigma) = \int (R_{\alpha} * \sigma)(x)\, d\sigma(x) =$$

$$= \iint d\sigma(x)\, d\sigma(y) R_{\alpha/2}(z-x)R_{\alpha/2}(z-y)dz = \int (R_{\alpha/2} * \sigma)^2 dz.$$

Henri Cartan [1, vol. III, nos. 70, 74, 75] reformulates the ideas of Frostman in the framework of the space $\mathcal{E}$ of charges which have a finite energy. The mutual energy is an inner product on $\mathcal{E}$, and Cartan proves that the cone $\mathcal{E}_+$ of positive charges is complete with respect to the distance $||\sigma - \tau|| = (\sigma - \tau|\sigma - \tau)^{\frac{1}{2}}$. The existence of an equilibrium distribution then follows immediately from the Lemma of Frederick Riesz concerning the projection onto a convex closed set. Why is the whole space $\mathcal{E}$ not complete? The answer to this question will be seen in the next section.

3. Laurent Schwartz and Jacques Deny

In 1944 Schwartz discovered the theory of distributions. I will summarize the definitions and facts we shall need in the sequel. More details can be found in my expository article [9] or in the recent well-written book of R. Strichartz [25]. To study the theory in depth, the book of Schwartz himself [24] and the first volume of L. Hörmander's four-volume work [8] can be recommended.

Denote by $\mathcal{D}$ (or by $\mathcal{D}(\mathbf{R}^n)$ if it is necessary to indicate the dimension) the vector space of all *test functions*, i.e. functions φ defined on $\mathbf{R}^n$, with values in $\mathbf{R}$ or $\mathbf{C}$, which have compact support and continuous partial derivatives of all orders. A *distribution* T on $\mathbf{R}^n$ is a linear map $\varphi \mapsto \langle T, \varphi\rangle$ from $\mathcal{D}$ into $\mathbf{R}$ or $\mathbf{C}$ which satisfies the following condition: For every compact subset K of $\mathbf{R}^n$ there exist two constants $M > 0$ and $m > 0$ such that

$$|\langle T, \varphi\rangle| \leqq M \max_{|\rho| \leqq m} \max_{x} |\partial^{\rho}\varphi(x)| \tag{3}$$

for every $\varphi \in \mathcal{D}$ with $\operatorname{Supp}\varphi \subset K$. Here we used the notation which is becoming standard: the multiindex $\rho = (\rho_1, \ldots, \rho_n) \in \mathbf{N}^n$ is an n-tuple of positive

integers, its order is $|\rho| = \rho_1 + \ldots + \rho_n$; we set $\partial_j = \frac{\partial}{\partial x_j}$ $(1 \leqq j \leqq n)$ and $\partial^\rho = \partial_1^{\rho_1} \ldots \partial_n^{\rho_n}$. Inequality (3) expresses the fact that the linear form T is continuous for a certain topology on $\mathcal{D}$. The space of all distributions is denoted by $\mathcal{D}'$ (or, if necessary, by $\mathcal{D}'(\mathbb{R}^n)$).

EXAMPLE 1. Let f be a Lebesgue-measurable function on $\mathbb{R}^n$ which is integrable on every compact set. Such a function is said to be locally integrable. Then f yields the distribution which with $\varphi \in \mathcal{D}$ associates the number $\int_{\mathbb{R}^n} f(x)\varphi(x)dx$. In (3) we can choose $m = 0$ and $M = \int_K |f(x)|dx$. This distribution will also be denoted by f.

EXAMPLE 2. The Dirac distribution δ is defined by $\langle \delta, \varphi \rangle = \varphi(0)$. This time $M = 1$ and m is again zero.

Motivated by the special case when T is defined by an appropriate locally integrable function, Schwartz introduced the following operations on distributions:

Differentiation. For $T \in \mathcal{D}'$, $\varphi \in \mathcal{D}$ and $\rho \in \mathbb{N}^n$ one sets

$$\langle \partial^\rho T, \varphi \rangle = (-1)^{|\rho|} \langle T, \partial^\rho \varphi \rangle.$$

Thus $\langle \partial^\rho \delta, \varphi \rangle = (-1)^{|\rho|} (\partial\varphi)(0)$.

Multiplication. Let f be a function on $\mathbb{R}^n$ which has continuous partial derivatives of all orders. Then the distribution fT is defined by

$$\langle fT, \varphi \rangle = \langle T, f\varphi \rangle$$

for all $\varphi \in \mathcal{D}$.

If φ and ψ are two test functions on $\mathbb{R}^n$, we denote by $\varphi \otimes \psi$ the function $(x, y) \mapsto \varphi(x)\psi(y)$ on $\mathbb{R}^{2n}$. Clearly $\varphi \otimes \psi$ belongs to $\mathcal{D}(\mathbb{R}^{2n})$. Let now S and T be two distributions on $\mathbb{R}^n$. Their *tensor product* $S \otimes T$ is the distribution on $\mathbb{R}^{2n}$ defined by

$$\langle S \otimes T, \varphi \otimes \psi \rangle = \langle S, \varphi \rangle \langle T, \psi \rangle.$$

One proves that this defines $S \otimes T$ as a continuous linear form on all of $\mathcal{D}(\mathbb{R}^{2n})$, i.e. that the functions $\varphi \otimes \psi$ form a total subset of $\mathcal{D}(\mathbb{R}^{2n})$.

The definition of the *convolution* is more delicate. If f and g are two integrable functions on $\mathbb{R}^n$, their convolution is the integrable function

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x - y)g(y)dy.$$

The distribution it defines according to Example 1 is given by

$$\begin{aligned} \langle f * g, \varphi \rangle &= \iint f(x - y)g(y)\varphi(x)dx\,dy \\ &= \iint f(x)g(y)\varphi(x + y)dx\,dy, \end{aligned}$$

which suggests to define the convolution $S * T$ of two distributions S and T by

$$\langle S * T, \varphi \rangle = \langle S \otimes T, \varphi(x+y) \rangle.$$

The trouble is that the function $(x, y) \mapsto \varphi(x+y)$ on $\mathbf{R}^{2n}$ has compact support only if φ is identically zero.

To define $S * T$ we introduce the space $\mathcal{B}_0 = \mathcal{B}_0(\mathbf{R}^n)$ of all functions φ on $\mathbf{R}^n$ which have continuous derivatives $\partial^\rho \varphi$ of all orders and they all tend to 0 at infinity, i.e. given $\rho \in \mathbf{N}^n$ and $\varepsilon > 0$ there exists $R > 0$ such that $|\partial^\rho \varphi(x)| \leqq \leqq \varepsilon$ for $|x| \geqq R$. A linear map T from $\mathcal{B}_0$ into $\mathbf{R}$ or $\mathbf{C}$ is called an *integrable distribution* on $\mathbf{R}^n$ [5] if there exist constants $M > 0$ and $m > 0$ such that

$$|\langle T, \varphi \rangle| \leqq M \max_{|\rho| \leqq m} \max_x |\partial^\rho \varphi(x)|$$

for all $\varphi \in \mathcal{B}_0$. Clearly $\mathcal{D} \subset \mathcal{B}$ and if T is an integrable distribution it satisfies condition (3) with the same M and m for all compact sets K. Thus every integrable distribution is a distribution: if we denote by $\mathcal{B}_0' = \mathcal{B}_0'(\mathbf{R}^n)$ the vector space of all integrable distributions on $\mathbf{R}^n$, then $\mathcal{B}_0' \subset \mathcal{D}'$. If now φ is a bounded function on $\mathbf{R}^n$ with continuous and bounded derivatives of all orders, and if $T \in \mathcal{B}_0'$, we define $\langle T, \varphi \rangle$ as follows: Let $\psi \in \mathcal{D}$ be such that $\psi(x) = 1$ for $|x| \leqq 1$, $\psi(x) = 0$ for $|x| \geqq 2$ and $0 \leqq \psi(x) \leqq 1$ everywhere. Set $\psi_n(x) = \psi(x/n)$ for $n \in \mathbf{N}$. Then the sequence $\langle T, \psi_n \varphi \rangle$ converges and its limit will be $\langle T, \varphi \rangle$. In particular, if 1 denotes the function whose value is identically one, the expression $\langle T, 1 \rangle$ is defined (it is called the integral of T).

We say that two distributions S and T on $\mathbf{R}^n$ are *convolvable* if for every $\varphi \in \mathcal{D}(\mathbf{R}^n)$ the distribution $\varphi(x+y) S \otimes T$ is integrable on $\mathbf{R}^{2n}$. In that case $S * T$ is defined by

$$\langle S * T, \varphi \rangle = \langle \varphi(x+y) S \otimes T, 1 \rangle,$$

where now 1 is the function identically one on $\mathbf{R}^{2n}$ [4, 12, 23].

For any $T \in \mathcal{D}'$ one has $\delta * T = T$ and more generally $\partial^\rho \delta * T = \partial^\rho T$. If S and T are convolvable, then

$$\partial^\rho (S * T) = \partial^\rho S * T = S * \partial^\rho T.$$

N. Ortner [19, 20] observed that the last two may be equal even if $S * T$ is not defined.

The theory of distributions cleared up a concept which was implicit for a long time in the theory of partial differential operators $P(\partial) = \sum_{|\rho| \leqq m} c_\rho \partial^\rho$ with constant coefficients. A distribution E is a *fundamental* (or elementary) *solution* of $P(\partial)$ if $P(\partial) E = \delta$. If E and T are convolvable, then $E * T$ is a solution of the partial differential equation $P(\partial) X = T$ since

$$P(\partial)(E * T) = (P(\partial) E) * T = \delta * T = T.$$

A theorem due to B. Malgrange and L. Ehrenpreis, for which P. Wagner has given recently a remarkably simple proof, states that every $P(\partial) \neq 0$ has a fundamental solution. My friend Norbert Ortner called my attention to the fact that a definition of a fundamental solution, essentially equivalent to the above, was given by N. Zeilon in 1911 [29].

In one of the first works which used the new theory [2, 3], J. Deny considered the space of distributions which have a finite energy. Using the Fourier transformation he proved that this space is isometric to a certain L^2-space, hence by the Riesz–Fischer theorem it is complete. Thus the elements missing from Cartan's space $\mathcal{E}$ were distributions. The completeness of $\mathcal{E}_+$ follows because a positive distribution is a measure.

4. Holomorphic functions and Riesz distributions

Let Λ be some non-empty domain in $\mathbb{C}$. A function $\alpha \mapsto T_\alpha$ defined in Λ and with values in $\mathcal{D}'$ is said to be holomorphic if for every $\varphi \in \mathcal{D}$ and every $\alpha \in \Lambda$ the limit

$$\lim_{h \to 0} \frac{\langle T_{\alpha+h}, \varphi \rangle - \langle T_\alpha, \varphi \rangle}{h} \tag{4}$$

exists. It then defines a distribution $\frac{dT}{d\alpha}$ such that for $\varphi \in \mathcal{D}$ the value $\langle \frac{dT}{d\alpha}, \varphi \rangle$ is given by (4). Thus the distribution-valued function $\alpha \mapsto T_\alpha$ is holomorphic if and only if for every test function φ the complex-valued function $\alpha \mapsto \langle T_\alpha, \varphi \rangle$ is holomorphic.

Let now Λ be a domain contained in a larger domain $\Lambda_1 \subset \mathbb{C}$, and $\alpha \mapsto$ $\to T_\alpha$, a holomorphic function on Λ. Suppose that for every $\varphi \in \mathcal{D}$ there exists a scalar-valued holomorphic function F_φ on Λ_1 such that $F_\varphi(\alpha) =$ $= \langle T_\alpha, \varphi \rangle$ for $\alpha \in \Lambda$. Then by the Banach–Steinhaus theorem there exists for each $\alpha \in \Lambda_1$ a distribution T_α such that $F_\varphi(\alpha) = \langle T_\alpha, \varphi \rangle$, this time for $\alpha \in \Lambda_1$. The holomorphic function $\alpha \mapsto T_\alpha$ on Λ_1 is the *analytic continuation* of the original function into the larger domain Λ_1.

It most often happens that for $\alpha \in \Lambda$ the distribution T_α is associated with a locally integrable function f_α, i.e.

$$\langle T_\alpha, \varphi \rangle = \int_{\mathbb{R}^n} f_\alpha(x) \varphi(x) dx.$$

Then the analytic continuation of T_α is said to be a *pseudofunction.*

Suppose now that $\alpha \mapsto T_\alpha$ is holomorphic in Λ with the possible exception of the point $\alpha_0 \in \Lambda$ and that for $\alpha \neq \alpha_0$ one has

$$T_\alpha = \frac{S_{-1}}{\alpha - \alpha_0} + S_0 + S_\alpha,$$

where $S_{-1}, S_0 \in \mathcal{D}'$, the function $\alpha \mapsto S_\alpha$ is holomorphic in all of Λ and $S_{\alpha_0} = 0$. If the residue $\operatorname*{Res}_{\alpha=\alpha_0} T_\alpha = S_{-1}$ is different from zero, we say that $\alpha \mapsto T_\alpha$ has a pole of order 1 at α_0. The distribution

$$S_0 = \lim_{\alpha \to \alpha_0} \left(T_\alpha - \frac{S_{-1}}{\alpha - \alpha_0} \right) \tag{5}$$

is the finite part of T_α at $\alpha = \alpha_0$ denoted by $\operatorname*{Pf}_{\alpha=\alpha_0} T_\alpha$ or simply by $\operatorname{Pf} T_{\alpha_0}$ [10, 11]. If $S_{-1} = 0$, then $\operatorname{Pf} T_{\alpha_0}$ is simply the value of T_α at α_0. The limit in (5) is to be understood in the sense that $\langle S_0, \varphi \rangle$ is for any $\varphi \in \mathcal{D}$ the limit of

$$\langle T_\alpha, \varphi \rangle - \frac{\langle S_{-1}, \varphi \rangle}{\alpha - \alpha_0}$$

as $\alpha \to \alpha_0$. Let the distribution T_α be defined by the locally integrable function f_α when $\alpha \in \Lambda$. Let $\Lambda_1 \supset \Lambda$ and assume that $\alpha \mapsto T_\alpha$ is holomorphic in Λ_1 with the possible exception of poles in $\Lambda_1 \setminus \Lambda$. Then for $\alpha \in \Lambda_1 \setminus \Lambda$ the value $\langle \operatorname{Pf} T_\alpha, \varphi \rangle$ is the classical notion of the *Hadamard finite part* of the integral

$$\langle T_\alpha, \varphi \rangle = \int_{\mathbf{R}^n} f_\alpha(x) \varphi(x) dx,$$

which is defined when $\alpha \in \Lambda$. Both the residue and the finite part have an importance for partial differential equations. For instance, the fundamental solution of the wave operator

$$\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \ldots - \frac{\partial^2}{\partial x_n^2}$$

is the finite part when n is odd, and the residue when n is even, of a certain pseudofunction introduced by Marcel Riesz [11, pp. 52–54].

One has

$$\operatorname*{Pf}_{\alpha=\alpha_0} (\partial^\rho T_\alpha) = \partial^\rho \operatorname*{Pf}_{\alpha=\alpha_0} T_\alpha.$$

Schwartz [24, II.2;28] claims that this is false. He considers the distribution x_+^α on $\mathbf{R}$ defined by

$$\langle x_+^\alpha, \varphi \rangle = \int_0^\infty x^\alpha \varphi(x) dx \quad \text{for} \quad \operatorname{Re} \alpha > -1$$

and says that the derivative of $\operatorname*{Pf}_{\alpha=-l} x_+^\alpha$ is not $-l \operatorname*{Pf}_{\alpha=-l} (x_+^{\alpha-1})$. This is true. However, the derivative of $\operatorname*{Pf}_{\alpha=-l} x^\alpha$ is $\operatorname*{Pf}_{\alpha=-l} (\alpha x_+^{\alpha-1})$ and one has

$$\operatorname*{Pf}_{\alpha=-l} (\alpha x_+^{\alpha-1}) = -l \operatorname*{Pf}_{\alpha=-l} (x_+^{\alpha-1}) + \frac{(-1)^l}{l!} \partial^l \delta.$$

We are nearing the conclusion of this lecture. Let $n \geqq 2$ and on $\mathbb{R}^n$ let us consider the distribution $|x|^{\alpha-n}$ which for $\operatorname{Re}\alpha > 0$ is given by

$$\langle |x|^{\alpha-n}, \varphi \rangle = \int_{\mathbb{R}^n} |x|^{\alpha-n}\varphi(x)dx.$$

The function $\alpha \mapsto |x|^{\alpha-n} \in \mathcal{D}'$ has an analytic continuation into the whole plane with the exception of the points $\alpha = -2k$ $(k \in \mathbb{N})$ where it has simple poles with residues

$$\frac{2\pi^{n/2}}{2^{2k}k!\Gamma\left(\frac{n}{2}+k\right)}\Delta^k\delta.$$

The scalar-valued function $\alpha \to \Gamma\left(\frac{\alpha}{2}\right)$ has simple poles at the same points with residues $2(-1)^k k!$. Therefore the distribution

$$\frac{1}{\Gamma\left(\frac{\alpha}{2}\right)}|x|^{\alpha-n}$$

is a holomorphic function of α in the whole plane, i.e. an entire function on $\mathbb{C}$. Its value at is $\alpha = -2k$ is

$$\frac{2^{-2k}\pi^{n/2}}{\Gamma\left(\frac{n+2k}{2}\right)}(-\Delta)^k\delta.$$

This motivates the definition of the (elliptic) Riesz distribution R_α by

$$\operatorname{Pf}\frac{\Gamma\left(\frac{n-\alpha}{2}\right)}{2^\alpha\pi^{n/2}\Gamma\left(\frac{\alpha}{2}\right)}|x|^{\alpha-n}$$

for $\alpha \in \mathbb{C}$. In particular $R_{-2k} = (-\Delta)^k\delta$ for $k \in \mathbb{N}$. The distributions R_α and R_β are convolvable if either $\operatorname{Re}(\alpha+\beta) < n$ or if at least one of the values α or β is equal to $-2k$ [13, 14], and then we have [18, 20] $R_\alpha * R_\beta = R_{\alpha+\beta}$ which is the distributional form of M. Riesz's composition formula (1). In particular

$$(\Delta)^k R_{2k} = (-\Delta)^k\delta * R_{2k} = R_{-2k} * R_{2k} = R_0 = \delta,$$

i.e. $(-1)^k R_{2k}$ is a fundamental solution of the differential operator Δ^k. For $k = 1$ and $n \geqq 3$ the solution of the partial differential equation $\Delta X = T$ is therefore

$$-R_2 * T = -\frac{\Gamma\left(\frac{n-2}{2}\right)}{4\pi^{n/2}}|x|^{2-n} * T,$$

which is the solution of Poisson's equation mentioned at the beginning.

The factor $\Gamma\left(\frac{n-\alpha}{2}\right)$ we introduced in the numerator of the definition of R_α produces poles at the points $\alpha = n + 2k$ $(k \in \mathbb{N})$. The finite part which defines R_α at these points can be calculated to be

$$R_{n+2k} = \frac{2(-1)^k}{2^{n+2k}\pi^{n/2}\Gamma\left(\frac{n+2k}{2}\right)k!} \times$$

$$\times |x|^{2k}\left\{\log\frac{2}{|x|} + \frac{1}{2}\left(\sum_{\nu=1}^{k}\frac{1}{\nu} - \gamma + \psi\left(\frac{n+2k}{2}\right)\right)\right\},$$

where γ is the Euler–Mascheroni constant and $\psi(x) = \Gamma'(x)/\Gamma(x)$ [13, pp. 181–182]. Observe that if n is even, then R_{n+2k} is a fundamental solution of $(-\Delta)^{n/2+k}$. In particular for $n = 2$, $k = 0$ a solution of

$$\Delta u(x) = f(x)$$

is the function

$$u(x) = (-R_2 * f)(x) = \frac{1}{2\pi}\int_{R^2} f(y)\log|x-y|dy.$$

This is why the potential is logarithmic in the plane!

REFERENCES

[1] CARTAN, H., *Oeuvres*, Vol. I, II, III, Edited by R. Remmert and J.-P. Serre, Springer-Verlag, Berlin–New York, 1979. *MR* **81c**:01031

[2] DENY, J., Les potentiels d'énergie finie, *Acta Math.* **82** (1950), 107–183. *MR* **12**–98

[3] DENY, J., Sur la définition de l'énergie en théorie du potentiel, *Ann. Inst. Fourier Grenoble* **2** (1950), 83–99. *MR* **13**–459

[4] DIEROLF, P. and VOIGT, J., Convolution and S'-convolution of distributions, *Collect. Math.* **29** (1978), 185–196. *MR* **81d**:46041

[5] DIEROLF, P. and VOIGT, J., Calculation of the bidual for some function spaces. Integrable distributions, *Math. Ann.* **253** (1980), 63–87. *MR* **82d**:46060

[6] GAUSS, C. F., *Allgemeine Lehrsätze in Beziehung auf die im verkehrten Verhältnisse des Quadrats der Entfernung wirkenden Anziehungs- und Abstossungskräfte*, Ostwald's Klassiker der Exakten Wissenschaften, Nr. 2, Wilhelm Engelmann, Leipzig, 1912. *Jb. Fortschritte Math.* **43**, 893

[7] GREEN, G., *Mathematical papers*, N. M. Ferrers ed., Macmillan and Co., London, 1871; Chelsea, Bronx, N. Y., 1970. *Jb. Fortschritte Math.* **3**, 565; *MR* **40** # 4075

[8] HÖRMANDER, L., *The analysis of linear partial differential operators*. I. Distribution theory and Fourier analysis; II. Differential operators with constant coefficients; III. Pseudodifferential operators; IV. Fourier integral operators, Grundlehren der mathematischen Wissenschaften, **256**, **257**, **274**, **275**, Springer-Verlag, Berlin–New York, 1983, 1983, 1985, 1985. *MR* **85g**:35002a, 35002b; **87d**:35002a, 35002b

[9] HORVÁTH, J., An introduction to distributions, *Amer. Math. Monthly* **77** (1970), 227–240. *MR* **40** # 6261

[10] HORVÁTH, J., Finite parts of distributions, *Linear operators and approximation* (Proc. Conf., Oberwolfach, 1971), Internat. Ser. Numer. Math., Vol. 20, Birkhäuser, Basel, 1972, 142–158. *MR* **52** # 6409

[11] HORVÁTH, J., Distribuciones definidas por prolongación analítica, *Rev. Colombiana Mat.* **8** (1974), 47–95. *MR* **51** # 6405

[12] HORVÁTH, J., Sur la convolution des distributions, *Bull. Sci. Math.* (2) **98** (1974), 183–192. *MR* **55** # 8787

[13] HORVÁTH, J., Composition of hypersingular integral operators, *Applicable Anal.* **7** (1977/78), 171–190. *MR* **58** # 17830

[14] HORVÁTH, J., Convolution de noyaux hypersinguliers, *Initiation Seminar on Analysis:* G. Choquet–M. Rogalsky–J. Saint-Raymond, 19th Year: 1979/80, Exp. No. 8, Publ. Math. Univ. Pierre et Marie Curie, no. 41, Univ. Paris VI, Paris, 1980. *MR* **84f**:46051

[15] HORVÁTH, J., ORTNER, N. and WAGNER, P., Analytic continuation and convolution of hypersingular higher Hilbert–Riesz kernels, *J. Math. Anal. Appl.* **123** (1987), 429–447. *MR* **88c**:46048

[16] KELLOG, O. D., *Foundations of potential theory,* Die Grundlehren der mathematischen Wissenschaften im Einzeldarstellungen, Band 31, Springer, Berlin, 1929; Frederick Ungar, New York. *Jb. Fortschritte Math.* **55**, 282

[17] KLEIN, F., *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert,* Teil I, Die Grundlehren der mathematischen Wissenschaften im Einzeldarstellungen, Band 24, Springer, Berlin, 1926. *Jb. Fortschritte Math.* **52**, 22

[18] ORTNER, N., Faltung hypersingulärer Integraloperatoren, *Math. Ann.* **248** (1980), 19–46. *MR* **82b**:46047

[19] ORTNER, N., Sur la convolution des distributions, *C. R. Acad. Sci. Paris Sér A–B* **290** (1980), A533–A536. *MR* **81e**:46025

[20] ORTNER, N., Convolution des distributions et des noyaux euclidiens, *Initiation Seminar on Analysis:* G. Choquet–M. Rogalsky–J. Saint-Raymond, 19th Year: 1979/80, Exp. No. 12, Publ. Math. Univ. Pierre et Marie Curie, no. 41, Univ. Paris VI, Paris, 1980. *MR* **84f**:46052

[21] ORTNER, N., Analytic continuation and convolution of hypersingular higher Hilbert–Riesz kernels, *A. Haar Memorial Conference,* vol. I, II (Budapest, 1985), Colloq. Math. Soc. János Bolyai, **49**, North-Holland, Amsterdam–New York, 1987, 675–685. *MR* **88e**:00007

[22] RIESZ, M., *Collected papers,* ed. by Lars Gårding and Lars Hörmander, Springer-Verlag, Berlin–New York, 1988. *MR* **90a**:01109

[23] ROIDER, B., Sur la convolution des distributions, *Bull. Sci. Math.* (2) **100** (1976), 193–199. *MR* **57** # 3852

[24] SCHWARTZ, L., *Théorie des distributions,* Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. IX-X, Nouvelle édition, Hermann, Paris, 1966. *MR* **35** # 730

[25] STRICHARTZ, R., *A guide to distribution theory and Fourier transforms,* CRC Press, Boca Raton, Fla., 1994.

[26] WAGNER, P., Zur Faltung von Distributionen, *Math. Ann.* **276** (1987), 467–485. *MR* **88f**:46086

[27] WAGNER, P., Bernstein–Sato-Polynome und Faltungsgruppen zu Differentialoperatoren, *Z. Anal. Anwendungen* **8** (1989), 407–423. *Zbl* **694**.35025

[28] WAGNER P., On the multiplication and convolution of homogeneous distributions, *Rev. Colombiana Mat.* **24** (1990), 183–197. *MR* **92c**:46049

[29] ZEILON, N., Das Fundamentalintegral der allgemeinen partiellen linearen Differentialgleichung mit konstanten Koeffizienten, *Ark. Mat. Astr. Fys.* **6** (1911), Nr. 38, 1–32. *Jb. Fortschritte Math.* **43**, 456

(Received November 16, 1994)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND AT COLLEGE PARK
COLLEGE PARK, MD 20742
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 30 (1995), 461–467

A NOTE ON SET MAPPINGS WITH MEAGER IMAGES

P. KOMJÁTH

Dedicated to Ákos Császár on his seventieth birthday

Abstract

If a set of reals of cardinal at most $\aleph_2$ has a Sierpiński type decomposition then that can be obtained from an ordering but this is not true for $\aleph_3$.

0. Introduction

In this paper we consider set mappings on sets of reals, i.e., when a function f is defined on some $A \subseteq \mathbf{R}$ such that $x \notin f(x) \subseteq A$ and we try to find free sets, that is, subsets $B \subseteq A$ such that $x \notin f(y)$ holds for $x, y \in B$. The set $f(x)$ is sometimes called the *image* of x. A classical theorem in combinatorial set theory states that if A is uncountable and $f(x)$ is always finite then there is a free set $B \subseteq A$ with $|B| = |A|$. If, however, CH holds, then there is a set mapping on $\mathbf{R}$ with countable images with no two-element free sets. This is actually Sierpiński's famous decomposition theorem in disguise ([2,3]). We are interested in the question when it is possible to give a set mapping on some set A of reals with meager images, with no two-element free subsets. Now Sierpiński's theorem can be re-stated as a "yes" answer if A has cardinal $\aleph_1$. In this case one can simply take a well-ordering of length ω_1 and associate to every point the set of its predecessors. (In fact, this argument works if every subset $A' \subseteq A$ with $|A'| < |A|$ is meager.) Miklós Laczkovich (Budapest) asked if it is always the case that this is the sole reason for the existence of such a set mapping, i.e., if there exists a set mapping as above, then there is an ordering with all its initial segments meager. The answer is obviously "yes", if A has cardinal $\aleph_1$. We show — building on some arguments of C. Freiling — that this is also the case if $|A| = \aleph_2$. Also, a characterization is given, in terms of cardinal invariants of A, of the property that every meager set mapping has an n-element free subsets. This has the corollary that if set mappings on A with meager images have 3-element free subsets then they

1991 *Mathematics Subject Classification.* Primary 03E35; Secondary 04A20, 28A05.
Key words and phrases. Set mappings, free sets, 2nd category sets.

The author acknowledges the support of the Hungarian National Science Fund Grant No. 1908 and that of DIMACS (Rutgers University, New Jersey).

have arbitrarily large finite free subsets. For $A = \mathbf{R}$ we can improve this to the requirement of 2-element free subsets.

The main result of this paper, which will be given in the last two sections, is a consistency proof of the existence of a set A of cardinal $\aleph_3$ such that there exists on A a meager set mapping with no two-element free subsets but there is no meager ordering of A.

We mention some problems which we have been unable to settle. First, assuming that A is of cardinal $\aleph_2$, and every meager set mapping on A has a 3-element free set, can we conclude that they have infinite free sets? As for the consistency result, can we make $A = \mathbf{R}$? Is it possible to give a similar consistency proof for the analogous measure case?

1. Notation, definitions

We use the standard axiomatic set theory notation. Notably, the set $\mathbf{R}$ of reals is identified with the set ${}^{\omega}2$ of infinite zero-one sequences. As usual, a set $A \subseteq \mathbf{R}$ is *meager* (or of first category) if it is the union of countably many nowhere dense sets. ${}^{<\omega}2$ is the set of finite zero-one sequences. If A is a set, κ a cardinal, then $[A]^{\kappa} = \{X \subseteq A : |X| = \kappa\}$. With notions of forcing we follow the convention that smaller conditions give more information.

If $A \subseteq \mathbf{R}$ is a set, a *meager set mapping* is a function $f : A \to P(A)$ such that $f(x)$ is meager for every $x \in A$, and $x \notin f(x)$. A *Sierpiński decomposition* is $A \times A = B \cup C$ such that B meets every horizontal line in a meager set, C meets every vertical line in a meager set.

STATEMENT 1. *A set $A \subseteq \mathbf{R}$ has a Sierpiński decomposition iff it has a meager set mapping with no 2-element free subset.*

PROOF. If f is a set mapping with no 2-element free sets, then put $B = \{(x, y) \in A \times A : x \in f(y)\}$, $C = A \times A - B$. If $B \cup C$ is a decomposition, set $f(x) = \{y : (x, y) \in B \text{ or } (y, x) \in C\}$. □

A (well) ordering $\prec$ on A is *meager* if every initial segment $\{y \in A : y \prec x\}$ is a meager subset of A.

STATEMENT 2. *If a set has a meager ordering then it has a meager set mapping with no 2-element free subsets.*

PROOF. If $\prec$ is a meager ordering, set $f(x) = \{y \in A : y \prec x\}$. □

STATEMENT 3. *If a set has a meager ordering then it has a meager well ordering.*

PROOF. If $\prec$ is a meager ordering then by Hausdorff's theorem there is a well ordered cofinal subset of $\prec$. This gives rise to a decomposition of the underlying set into a well ordered union of some disjoint sets $\{A_\alpha : \alpha < \kappa\}$ such that $\bigcup\{A_\beta : \beta < \alpha\}$ is meager for every $\alpha < \kappa$. This can be transformed into a well ordering if one arbitrarily well orders the sets A_α. □

If X is a set, (X,T) is a *tournament* if $T \subseteq X \times X$ and for any $x, y \in X$, $x \neq y$ exactly one of $(x,y) \in T$, $(y,x) \in T$ holds. For $x \in X$ set $T(x) = \{y : (x,y) \in T\}$ the points *covered by* x. Similarly, for $Y \subseteq X$ set $T(Y) = \bigcup\{T(y) : y \in Y\}$. A subset $X' \subseteq X$ is *κ-covered* for some cardinal κ if there is a set $Y \subseteq X$, $|Y| \leq \kappa$ such that $X' \subseteq T(Y)$.

2. Sets of cardinal $\leq \aleph_2$

In this Section we assume that $A \subseteq \mathbf{R}$ is a set of cardinal $\aleph_2$. We analyse the situation using some ideas of Chris Freiling [1]. We consider the following properties.

(a) Every $B \subseteq A$ of cardinal $\aleph_1$ is meager.

(b) A is the union of $\aleph_1$ meager sets.

(c) $A = B \cup C$ where B is the union of $\aleph_1$ meager sets, and every subset of cardinal $\aleph_1$ of C is meager.

STATEMENT 4. *If* (a) *or* (b) *holds then A has a meager ordering.*

PROOF. If (a) holds take any well ordering of A in type ω_2. If (b) holds take the ω_1 type union of the $\aleph_1$ sets establishing (b). In this case every point is preceded by the union of countably many meager sets which is meager. □

STATEMENT 5. *If neither* (a) *nor* (b) *holds then every meager set mapping on A has a 2-element free set.*

PROOF. Let f be a meager set mapping on A. As (a) fails there is a non-meager $B \subseteq A$ of cardinality $\aleph_1$. As (b) fails there is a $y \in A$ such that $y \notin \bigcup\{f(x) : x \in B\}$. If f has no two-element free sets then necessarily $B \subseteq f(y)$ which is a contradiction as B is non-meager. □

STATEMENT 6. *If* (c) *holds there is a meager set mapping with no 3-element free sets.*

PROOF. Let f_B, f_C be meager set mappings on B, resp. C with no two-element free subsets. Now $f_B \cup f_C$ works. □

STATEMENT 7. *If* (c) *fails and a meager set mapping f is given on A then for every finite n there is a free set of size n.*

PROOF. By induction on $1 \leq i \leq n$ select the non meager $B_i \subseteq A - (B_1 \cup \cup \cdots \cup B_{i-1})$ such that $B_i \subseteq A - \bigcup\{f(x) : x \in B_1 \cup \cdots \cup B_{i-1}\}$, $|B_i| = \aleph_1$. This is possible, as otherwise we could find witnesses for (c). We can now select by reverse induction $x_i \in B_i$ $(1 \leq i \leq n)$ such that $x_i \notin \bigcup\{f(x_j) : i < j \leq n\}$. This is again possible as no B_i is meager. □

We notice that the last two statements have the following consequence. If every meager set mapping on A has a 3-element free set then there exist arbitrarily large finite free sets. In general, "three" here cannot be improved to "two" (i.e., in some appropriate models of set theory), but can be, if $A = \mathbf{R}$.

STATEMENT 8. *If $A = \mathbf{R}$ and both* (a) *and* (b) *fail for A then* (c) *fails, as well.*

PROOF. As (a) fails for $\mathbf{R}$ there is a non meager $B \subseteq \mathbf{R}$ of cardinal $\aleph_1$. To show that (c) is not true, it suffices to show that if $C \subseteq \mathbf{R}$ is the union of $\aleph_1$ meager sets, then $B + x \subseteq \mathbf{R} - C$ holds for some $x \in \mathbf{R}$. In order to argue that such an x exists it suffices to show that the set $\{x \in \mathbf{R} : B + x \not\subseteq \mathbf{R} - C\}$ is the union of $\aleph_1$ meager sets. Put $C = \bigcup\{C_\alpha : \alpha < \omega_1\}$ and enumerate B as $B = \{b_\beta : \beta < \omega_1\}$. If $(x + B) \cap C \neq \emptyset$ then $x + b_\beta \in C_\alpha$ for some α, $\beta < \omega_1$, and to any given pair α, β the set of these x's is $(-b_\beta) + C_\alpha$, a meager set. □

3. Construction of a tournament

THEOREM 1. *The following is consistent with GCH. There exists a tournament on ω_3 such that there is no ordering of ω_3 with ω-covered initial segments.*

PROOF. Assume GCH. Our notion of forcing with which we get the model of the Theorem will be the following $(P, \leqq)$. A condition (s, t) is a tournament with $s \in [\omega_3]^{\leqq \aleph_1}$. $(s', t') \leqq (s, t)$ iff $s' \supseteq s$ and $t = t' \cap (s \times s)$. It is easy to see that $(P, \leqq)$ is $\leqq \omega_1$-closed and $\aleph_3$-c.c. Therefore, forcing with $(P, \leqq)$ cardinals, cofinalities, and GCH are preserved. If $G \subseteq P$ is a generic set, put $T = \bigcup\{t : (s, t) \in G\}$. T will obviously be a tournament on ω_3.

Assume that some p forces that $\prec$ orders ω_3 in such a way that every initial segment is ω-covered. We consider three cases.

Case 1. Some $p' \leqq p$ forces that the cofinality of $\prec$ is $\leqq \omega_1$.

In this case a certain $(s, t) = q \leqq p'$ determines a set Y of some $\aleph_1$ elements which establishes that ω_3 is ω_1-covered. We can as well assume that $Y \subseteq s$. But then, if $x \in \omega_3 - s$, $q' = (s', t') \leqq (s, t)$ is the following $s' = s \cup \{x\}$, $t' = t \cup (\{x\} \times s)$ then q' forces that x is not covered by s, a contradiction.

Case 2. Some $p' \leqq p$ forces that the cofinality of $\prec$ is ω_2.

Then, p' forces that there are countable sets Y_α such that ω_3 is covered by the union of any $\aleph_2$ of the $\{T(Y_\alpha) : \alpha < \omega_2\}$. As CH holds in the enlarged model we can assume that these sets form a Δ-system, $Y_\alpha = U \cup V_\alpha$ where the sets $\{U, V_\alpha : \alpha < \omega_2\}$ are pairwise disjoint. We can as well assume that $p' = (s, t)$ determines what the elements of U are and in fact that $U \subseteq s$ holds. For $\xi \in \omega_3 - s$ set $p_\xi = (s \cup \{\xi\}, t \cup (\{\xi\} \times s))$. For every $\xi \in \omega_3 - s$ there is some $q_\xi \leqq p_\xi$ which determines a $\beta(\xi) < \omega_2$ so that $\xi \in T(Y_\alpha)$ for $\alpha \geqq \beta(\xi)$. For $\aleph_3$ of these ξ, $\beta(\xi) = \beta$ holds and we can also assume, again by the Δ-system lemma, that $q_\xi = (s' \cup s_\xi, t_\xi)$ where the sets s', s_ξ are disjoint, and $t' = t_\xi \cap (s \times s)$ is the same.

There is an $r = (s'', t'') \leqq (s', t')$ which determines a V_α with $\alpha \geqq \beta$ such that $V_\alpha \cap s' = \emptyset$ and $V_\alpha \subseteqq s''$. Select finally a $\xi < \omega_3$ so that $s_\xi \cap s'' = \emptyset$. Then, let $\bar{r}$ be the following condition

$$\bar{r} = (s'' \cup s_\xi, t'' \cup t_\xi \cup (s_\xi \times (s'' - s'))).$$

This condition forces that $U \cup V_\alpha \subseteqq T(\xi)$, so $\xi \notin T(Y_\alpha)$, a contradiction.

Case 3. Some $p' \leqq p$ forces that the cofinality of $\prec$ is ω_3.

In this case ω_2 is $\prec$-bounded so it is ω-covered. Some $q = (s, t) \leqq p'$ forces that $\omega_2 \subseteqq T(Y)$ for a certain countable $Y \subseteqq s$. But again, if $x \in \omega_2 - s$, then $q' = (s', t')$, $s' = s \cup \{x\}$, $t' = t \cup (\{x\} \times s)$ forces that x is uncovered by s, a contradiction. □

4. A set of cardinal $\aleph_3$

THEOREM 2. *If T is a tournament on κ, a cardinal, then there is a ccc forcing notion which adds a set $A = \{r_\alpha : \alpha < \kappa\}$ of reals such that exactly those subsets of A which are indexed by ω-covered subsets of κ, become meager.*

PROOF. A condition will be of the form $p = (t, f, h, g) \in P$ where $t \in [\kappa]^{<\omega}$, $\mathrm{Dom}\,(f) = t$, for $\alpha \in t$, $f(\alpha) \in {}^{<\omega}2$, $h(\alpha, \beta) < \omega$ for $\alpha \neq \beta \in t$, $\beta \in T(\alpha)$. $g(\alpha, n)$ is a set of finitely many functions $s \in {}^{<\omega}2$ for finitely many pairs $\langle \alpha, n \rangle \in t \times \omega$ such that $s \not\subseteqq f(\beta)$, $f(\beta) \not\subseteqq s$ hold whenever $\beta \in T(\alpha)$ and $h(\alpha, \beta) = n$.

$p' = (t', f', h', g') \leqq p = (t, f, h, g)$ iff $t' \supseteqq t$, $f'(\alpha) \supseteqq f(\alpha)$ $(\alpha \in t)$, $h' \supseteqq h$, $g'(\alpha, n) \supseteqq g(\alpha, n)$ if the latter is defined.

The intuition behind this definition is the following. $f(\alpha)$ gives the first several (binary) digits of r_α. For every α, to make $\{r_\beta : \beta \in T(\alpha)\}$ meager we decompose this set into countably many pieces (this decomposition is approximated by h) and finally the 0–1 functions in $g(\alpha, n)$ will approximate the collection of open binary intervals giving a dense open set disjoint from the n-th piece, therefore establishing that the set in question is meager.

CLAIM 1. *For every $\alpha < \kappa$, $n < \omega$, the set $D = \{(t, f, h, g) : \alpha \in t, \mathrm{Dom}\, f(\alpha) \geqq n\}$ is dense in $(P, \leqq)$.*

PROOF OF CLAIM. Let $p = (t, f, g, h)$ be an element of P and $\alpha < \kappa$. Assume first that $\alpha \notin t$. Let N be a natural number greater than any of the values in the range of h. If we now take $p' = (t \cup \{\alpha\}, f', h', g)$ where f' is an extension of f such that $f(\alpha) = \emptyset$, h' extends h in such a way that $h(\beta, \alpha) = N$ for every $\beta \in t \cap T(\alpha)$, then p' will be a condition, as no contradiction arises from the information we know so far about r_α and the approximations on the meager sets. If, however, $\alpha \in t$, we can arbitrarily extend $f(\alpha)$. □

If $G \subseteqq P$ is generic, set $r_\alpha = \bigcup \{f(\alpha) : (t, f, h, g) \in G\}$, a real, and $A = \{r_\alpha : \alpha < \kappa\}$ will be our set.

CLAIM 2. *Given* $s \in {}^{<\omega}2$, $\alpha < \kappa$, $n < \omega$, *the set*

$$\{(t, f, h, g) : \text{ some } s' \supseteq s \text{ has } s' \in g(\alpha, n)\}$$

is dense.

PROOF OF CLAIM. Take a long enough (say of length N) extension of $f(\beta)$ for every $\beta \in t \cap T(\alpha)$ such that $h(\alpha, \beta) = n$ and then extend s to an s' of length N, different from all those $f(\beta)$ values. The long extensions are possible by the previous Claim. Now we can add s' to the g-part of the condition. □

CLAIM 3. *In* $V[G]$, $\{r_\beta : \beta \in T(\alpha)\}$ *is meager.*

PROOF OF CLAIM. The set $\{r_\beta : h(\alpha, \beta) = n\}$ is nowhere dense as is witnessed by the intervals given by $\{g(\alpha, n) : (t, f, g, h) \in G\}$ by the previous Claim. □

CLAIM 4. $(P, \leqq)$ *is ccc.*

PROOF OF CLAIM. If p_ξ are given $(\xi < \omega_1)$ we can assume that they are of the form $p_\xi = (t \cup t_\xi, f \cup f_\xi, h_\xi, g \cup g_\xi)$ where the structures $(t, t_\xi; T, h_\xi)$ are isomorphic. Then, we can find a common extension $(t \cup t_{\xi_0} \cup t_{\xi_1}, f \cup f_{\xi_0} \cup \cup f_{\xi_1}, h, g \cup g_{\xi_0} \cup g_{\xi_1})$ where h extends $h_{\xi_0} \cup h_{\xi_1}$ in such a way that $h(\alpha, \beta) = N$ for $\alpha \in t_{\xi_0}$, $\beta \in t_{\xi_1}$ where N is bigger than every natural number occurring in the second coordinates of the domains of g_{ξ_0}, g_{ξ_1}. □

CLAIM 5. *In* $V[G]$, *if* $B \subseteq A$ *is a meager set then* $\{\alpha < \kappa : r_\alpha \in B\}$ *is* ω*-covered.*

PROOF OF CLAIM. Assume that $1 \Vdash \underline{B} \subseteq \underline{A}$ is nowhere dense. Let N be a countable elementary submodel of the structure $(H((2^\kappa)^+); P, \underline{B}, \Vdash, \ldots)$. Our intention is to show that $1 \Vdash$ if $r_\alpha \in \underline{B}$ then $\alpha \in T(N \cap \kappa)$. Assume that p' forces that $r_\alpha \in B$ and $N \cap \kappa \subseteq T(\alpha)$. Write p' as $p' = (t \cup t', f \cup f', h', g \cup \cup g')$ where $t \subseteq N$, $t' \cap N = \emptyset$, $\mathrm{Dom}(f) = t$, $\mathrm{Dom}(f') = t'$, $\mathrm{Dom}(g) \subseteq t \times \omega$, $\mathrm{Dom}(g') \subseteq t' \times \omega$. By elementarity, N has an isomorphic condition $p'' = (t \cup t'', f \cup f'', h'', g \cup g'')$ with $t'' \subseteq N$. Let $p^* = (t^*, f^*, h^*, g^*) \leqq p''$ be a condition, $p^* \in N$, $p^* \Vdash \underline{B} \cap I = \emptyset$ where $I = I(s)$ is the interval of those functions extending s for some $s \supseteq f'(\alpha)$. Such a p^* exists as B is nowhere dense and N is elementary. p^* and p' are compatible as if $\beta \in t'$, $\gamma \in t^*$, $\beta \in T(\gamma)$, $s \in g^*(\gamma, n)$ for some $n < \omega$ then $s \not\subseteq f'(\beta)$ as β has a twin β'' in t'' for which $\beta'' \in T(\gamma)$ and so $s \not\subseteq f''(\beta'') = f'(\beta)$. Moreover, as $N \cap \kappa \subseteq T(\alpha)$ we can even extend the common extension to one which forces that $r_\alpha \in I$ which is a contradiction as this would force $r_\alpha \in I \cap B = \emptyset$. □

THEOREM 3. *It is consistent that there exists a set* $A \subseteq \mathbf{R}$ *of cardinal* $\aleph_3$ *with a meager set mapping that has no two-element free set but* A *has no meager ordering.*

PROOF. Let V satisfy the statement of Theorem 1. Let $(P, \leqq)$ be the notion of forcing given in Theorem 2, for the tournament T, adding $A = \{r_\alpha : \alpha < \omega_3\}$. Then, $\{f(r_\beta) : \beta \in T(\alpha)\}$ is always meager and this set mapping has no 2-element free subsets (as T is a tournament). If in V^P, A is the increasing union of meager subsets, then ω_3 is the increasing union of ω-covered subsets, $\omega_3 = \bigcup\{B_\xi : \xi < \kappa\}$ for some cardinal κ. If now $C_\xi = \{x : \text{some } p \text{ forces } x \in B_\xi\}$, then, by ccc of P, C_ξ is ω-covered, and $\omega_3 = \bigcup\{C_\xi : \xi < \kappa\}$ is an increasing union of ω-covered subsets in V, a contradiction. □

ACKNOWLEDGEMENT. The author is thankful for the referee for a very careful job.

REFERENCES

[1] FREILING, C., Axioms of symmetry: throwing darts at the real number line, *J. Symbolic Logic* **51** (1986), 190-200. *MR* **87f**:03148

[2] SIERPIŃSKI, W., Sur un théorème équivalent à l'hypothèse du continu $(2^{\aleph_0} = \aleph_1)$, *Bull. Internat. Acad. Polon. Sci. et Lettr. Classe Sci. Math. Natur. Sér. A: Sci. Math. = Krak. Anz.*, 1919, 1-3. *Jb. Fortschritte Math.* **47**, 897

[3] SIMMS, J. C., Sierpiński's theorem, *Simon Stevin* **65** (1991), 69-163. *MR* **92j**:01052

(Received May 5, 1994)

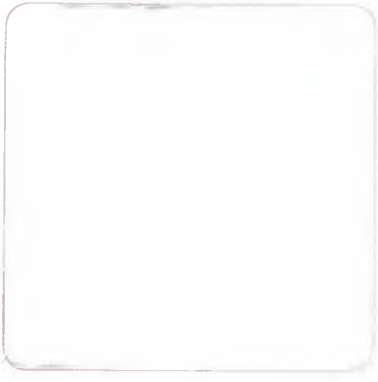
EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
SZÁMÍTÓGÉPTUDOMÁNYI TANSZÉK
MÚZEUM KRT. 6-8
H–1088 BUDAPEST
HUNGARY

e-mail: kope@cs.elte.hu

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Volume 30

CONTENTS



PRINTED IN HUNGARY

TypoTEX Kft, Budapest

Typeset by TypoTEX Ltd., Budapest
PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda, Budapest

CONTENTS

HU ISSN 0081-6906