Studia Scientiarum Mathematicarum Hungarica

27/1992

13

VOLUME 27
NUMBERS 1–2
1992

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

Subscription information

Orders should be addressed to

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

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 1–24

THE FULL EMBEDDINGS OF THE CATEGORIES OF UNIFORM SPACES, PROXIMITY SPACES AND RELATED CATEGORIES INTO THEMSELVES AND EACH OTHER. II[1]

E. MAKAI, Jr.[2]

We use the same notations as in Part I. For the definitions we refer to Part I. Here we just recall some special notations and definitions (others are rather standard). $\mathcal{S}_0^-$ denotes the category with objects all pairs $(X, \mathcal{X})$, X a set, $\{\emptyset, X\} \subset \mathcal{X} \subset 2^X$, and morphisms $f\colon (X_1, \mathcal{X}_1) \to (X_2, \mathcal{X}_2)$ characterized by $f\colon X_1 \to X_2$, $f^{-1}(\mathcal{X}_2) \subset \mathcal{X}_1$. Coz, the category of cozero-spaces, is the full subcategory of $\mathcal{S}_0^-$ determined by $\mathrm{ObCoz} = \{(X, \mathcal{X}) \mid \exists$ uniformity on X, $\mathcal{X} = \{$ cozero-sets of all uniformly continuous real functions w.r.t. this uniformity$\}\}$. A uniform (etc.) space is called special if for any uniform (etc., resp.) space Y with $UY = UX$ (= underlying set of X) $U(Y,Y) = U(X,X)$ (=$\{$uniformly continuous functions $X \to X\}$) implies $Y = X$. For a concrete category $\mathcal{C}$ the underlying set functor is denoted by $U_{\mathcal{C}}$ (or U; this will be sometimes omitted). $J\colon \mathrm{Prox} \to \mathrm{Unif}$ is the concrete functor associating to a proximity the compatible precompact uniformity. *Subcategories* will always be assumed to be full.

§ 5. Embeddings of subcategories of Prox (Unif) in $\mathcal{S}_0^-$

For the case of full embeddings $\mathrm{Prox} \to \mathcal{S}_0^-$ we have a negative result. For analogous negative results, concerning full embeddings of categories connected with closure spaces into $\mathcal{S}_0^-$ cf. [3], [2].

The proof of the following proposition is related to the proof of [27], Proposition 13.

PROPOSITION 2. 1) *Let* $\mathcal{C} \subset \mathrm{Unif}$ *and let* $\mathrm{Ob}\,\mathcal{C}$ *contain a uniform space* C_0 *which does not have a base composed of all partitions of cardinality less*

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 54E15; Secondary 54E05, 54B30.

Key words and phrases. Categories of uniform (proximity, cozero, $T_{3\frac{1}{2}}$) spaces, category of set systems, full embeddings, special spaces, inductive (projective) generation, strongly rigid class of objects, spaces $Y_{\mathcal{F}}$, ultraproximities, free ultraspaces.

[1] Part I of this paper has appeared in the same journal, **25** (1990), 199–208.

[2] Research (partially) supported by Hungarian National Foundation for Scientific Research Grant no. 1807.

Akadémiai Kiadó, Budapest

than some cardinal. Let $F:\mathcal{C}\to\mathcal{S}_0^-$ *be a full embedding. Then* $\forall C\in\operatorname{Ob}\mathcal{C}$ *for* $FC=(X,\mathcal{X})$ *(where for simplicity we assume* $U_{\mathcal{C}}=U_{\mathcal{S}_0^-}F$*) we have* $\mathcal{X}\supset$ $\supset\{A\subset X \mid A,\ X\setminus A$ are far in $C\}$.

2) *Let* $\forall C\in\operatorname{Ob}\mathcal{C}$ $\forall A\subset UC,\ A,(UC)\setminus A$ *proximal in* C $\exists D\in\operatorname{Ob}\mathcal{C}$ $(\forall B,$ $B'\subset UD,\ B,(UD)\setminus B$ *near*, $B',(UD)\setminus B'$ *near* $\exists g\in U(D,D),\ g^{-1}(B)=B')$ $\exists f\in U(D,C),\ \exists B_1,B_2$ *near subsets of* $D,\ f(B_1)\subset A,\ f(B_2)\subset(UC)\setminus A$. (A space D has this property e.g. if, denoting by N^* the proximity on a countable discrete topological space corresponding to its one-point compactification, we have $B\subset UD,\ B,(UD)\setminus B$ near in $D\Rightarrow\exists$ embedded copy of N^* in D, which has infinite intersections both with B and $(UD)\setminus B$ and which is a retract of D by a retraction mapping B into B and $(UD)\setminus B$ into $(UD)\setminus B$ (e.g. D has a discrete topology and its completion is a zero-dimensional compact metric space). An above $f\in U(D,C)$ exists for such a D with non-discrete proximity if we have: $A\subset UC,\ A,\ (UC)\setminus A$ near in $C\Rightarrow\exists g\colon N^*\to C$, $\exists$ infinite disjoint subsets $N_1,\ N_2$ of N^* such that $g(N_1)\subset A,\ g(N_2)\subset(UC)\setminus A$ (e.g. the completion of the T_0-reflection of C is compact Fréchet–Urysohn).) *Let* $F:\mathcal{C}\to\mathcal{S}_0^-$ *be a concrete full embedding satisfying the conclusion of statement* 1), $\mathcal{C}\not\subset$ {indiscrete spaces}. *Then* $\forall C\in\operatorname{Ob}\mathcal{C}$ *for* $FC=(X,\mathcal{X})$ *we have* $\mathcal{X}=\{A\subset X\mid A,X\setminus A$ are far in $C\}$.

3) *If* $\exists C_1,C_2\in\operatorname{Ob}\mathcal{C}$ *such that for their reflections* rC_1,rC_2 *in* $\{C\in$ $\in\operatorname{Ob}\mathrm{Unif}|C$ has a basis consisting of finite partitions$\}$ *there holds* $U(C_1,C_2)\subsetneqq$ $\subsetneqq U(rC_1,rC_2)$ *(where for simplicity we assume the universal map* $C_i\to rC_i$ *has underlying function* 1_{UC_i}*) then there exists no concrete full embedding* $F:\mathcal{C}\to\mathcal{S}_0^-$ *satisfying the conclusion of statement* 2).

In particular $\{C\in\operatorname{Ob}\mathrm{Prox}\mid\delta dC=0,$the T_0-reflection of C has a metric completion$\}$ *has a unique full embedding into* $\mathcal{S}_0^-$ *— up to natural isomorphy — namely the one given under* 2), *but no subcategory of* $\{C\in\operatorname{Ob}\mathrm{Prox}\mid$ the T_0-reflection of C has a metric completion$\}$ *strictly containing the above subcategory admits one.*

PROOF. 1) By [22], Corollary to Lemma 2 $U_{\mathcal{C}}$ and $U_{\mathcal{S}_0^-}F$ are naturally isomorphic, thus we may assume $U_{\mathcal{C}}=U_{\mathcal{S}_0^-}F$. We have by [23], Remark 2 $U(C_0,C_0)\neq X_0^{X_0}$, where $X_0=UC_0$. Hence for $FC_0=(X_0\mathcal{X}_0)$ $\{\emptyset,X_0\}\neq\mathcal{X}_0\neq$ $\neq 2^{X_0}$. Let $\emptyset\neq A_0\subsetneqq X_0,\ A_0\in\mathcal{X}_0$ and let $C\in\operatorname{Ob}\mathcal{C},\ X=UC$. If $A,\ X\setminus A$ are far in C, define $f\in U(X,X_0)=\hom((X,\mathcal{X}),(X_0,\mathcal{X}_0))$ by $f(A)\subset\{x_0\}$, $f(X\setminus A)\subset\{y_0\}$, where $x_0\in A_0,\ y_0\in X_0\setminus A_0$. Then $A=f^{-1}(A_0)\in\mathcal{X}$.

2) For the second statement we first show for $FD=(UD,\mathcal{D})$ that $\mathcal{D}=$ $=\{B\subset UD\mid B,\ (UD)\setminus B$ are far in $D\}$. By the conclusion of statement 1) $\mathcal{D}\supset\{B\subset UD\mid B,(UD)\setminus B$ are far in $D\}$. If, however, $\exists B\in\mathcal{D},\ B,(UD)\setminus B$ are near in D then $|UD|>1$ and for any other set $B'\subset UD$ with $B',(UD)\setminus B'$ near $\exists g\in U(D,D)=\hom((UD,\mathcal{D}),(UD,\mathcal{D}))$ such that $B'=g^{-1}(B)$, hence $B'\in\mathcal{D}$. Therefore $\mathcal{D}=2^{UD}$, hence $U(D,D)=\hom((UD,\mathcal{D}),(UD,\mathcal{D}))=$ $=(UD)^{(UD)}$, hence by [23], Remark 2 D has for basis all partitions of UD

of cardinality less than some cardinal. Thus D is either indiscrete or is finer than a discrete proximity. However, $B, (UD)\setminus B$ are near in D, therefore the first possibility must hold. Now for any $C' \in \mathrm{Ob}\,\mathcal{C}$ and $FC' = (X', \mathcal{X}')$ we have $\hom(D, C') = \hom((UD, \mathcal{D}), (X', \mathcal{X}')) = (X')^{(UD)}$, hence by $|UD| > 1$ C' is indiscrete as well. This is a contradiction to our hypothesis. This contradiction proves $\mathcal{D} = \{B \subset UD \mid B, (UD)\setminus B$ are far in $D\}$.

Now we show $\forall C \in \mathrm{Ob}\,\mathcal{C}$ $\mathcal{X} = \{A \subset X \mid A, X \setminus A$ are far in $C\}$. Suppose the contrary, i.e. $\exists A \in \mathcal{X}$, $A, X \setminus A$ are proximal. Then $\exists f \in \hom(D, C) = \hom((UD, \mathcal{D}), (X, \mathcal{X}))$ with the property described in Proposition 2. Then $f^{-1}(A)$, $X \setminus f^{-1}(A)$ are proximal and $f^{-1}(A) \in \mathcal{D} = \{B \subset UD \mid B, (UD)\setminus B$ are far in $D\}$, a contradiction which proves $\mathcal{X} = \{A \subset X \mid A, X \setminus A$ are far in $C\}$.

The statement in brackets is mostly evident.

For a space D with the property described there and $B, B' \subset UD$ like above, a map $g\colon D \to D$ satisfying $g^{-1}(B) = B'$ is given as follows. We have embeddings $j, j'\colon N^* \to D$, associated with B, resp. B', and we have $k'\colon D \to N^*$ such that $k'j' = 1_{N^*}$ and $k'^{-1}j'^{-1}(B') = B'$. We may assume $j^{-1}(B) = j'^{-1}(B')$ ($=$ { odd natural numbers }, say). Let $g = jk'\colon D \to D$. Then $g^{-1}(B) = k'^{-1}j^{-1}(B) = k'^{-1}j'^{-1}(B') = B'$. We still have to show that if D has a discrete topology and its completion is a zero-dimensional compact metric space, $B \subset UD$, $B, (UD)\setminus B$ are near in D then there is an embedded copy of N^* satisfying the requirements of the statement in the brackets. Let $\bar{D}$ denote the completion of D. Then B and $(UD)\setminus B$ have a common accumulation point x in $\bar{D}\setminus D$, hence there is an embedded copy iN^* of N^* (with an embedding $i\colon N^* \to D$), $B \cap (UiN^*)$, $[(UD)\setminus B] \cap (UiN^*)$ being infinite, iN^* having the accumulation point x. There exists a bijection $h \in U(D, N^*)$. h is an isomorphism on iN^*, i.e. $hiN^* \subset N^*$. Evidently $\exists \varphi\colon N^* \to N^*$ which is a retraction onto hiN^* and which satisfies $\varphi((hB)\setminus(UhiN^*)) \subset (hB) \cap (UhiN^*)$, $\varphi([(UN^*)\setminus(hB)]\setminus(UhiN^*)) \subset [(UN^*)\setminus(hB)] \cap (UhiN^*)$. Then $(Uh)^{-1}U(\varphi h) = U\psi$, $\psi \in U(D, D)$ being the desired retraction.

3) We have by conclusion of statement 2) $U(rC_1, rC_2) = \{f\colon X_1 \to X_2 \mid (A \subset X_2, A, X_2 \setminus A$ far in $C_2) \Rightarrow (f^{-1}(A),\ X_1 \setminus f^{-1}(A)$ far in $C_1)\} = \hom((X_1, \mathcal{X}_1), (X_2, \mathcal{X}_2)) = U(C_1, C_2)$ $((X_i, \mathcal{X}_i) = FC_i)$, contradicting the hypothesis.

4) For the last statement observe that the categories in question contain spaces like D in 2), and each space contained by them has the property as C in 2). Therefore 2) applies. It remained to show that for $\{C \in \mathrm{Ob}\ \mathrm{Prox} \mid \delta dC = 0$, the T_0-reflection of C has a metric completion$\} \subsetneq \mathrm{Ob}\,\mathcal{C} \subset \{C \in \mathrm{Ob}\ \mathrm{Prox} \mid$ the T_0-reflection of C has a metric completion$\}$ there is no full embedding $F\colon \mathcal{C} \to \mathcal{S}_0^-$. Let $C \in \mathrm{Ob}\,\mathcal{C}$, $\delta dC \neq 0$. Then rC has for basis all finite uniform partitions of C, hence the compatible compactification $\widetilde{rC}$ of the T_0-reflection of rC is the 0-dimensional T_2-reflection of the compatible compactification $\tilde{C}$ of the T_0-reflection of C. However, $\widetilde{rC}$ is a T_2 continuous

image of a compact metric space, hence is compact metric as well ([4], Theorem 3.1.22). Therefore $rC \in \mathrm{Ob}\,\mathcal{C}$ and we have $U(rC,C) \not\ni 1_{UC} \in U(rC,rC)$, hence $U(rC,C) \subsetneqq U(rC,rC)$. By 3) this proves the statement. $\square$

COROLLARY 5. *Let $\mathcal{C} \subset$ Unif. Let there exist a non-indiscrete $C_0 \in \mathrm{Ob}\,\mathcal{C}$ such that the condition ($A \subset X_0 (= UC_0)$, $A, X_0 \setminus A$ far in C_0) implies ($A = \emptyset$ or $A = X_0$), or let there exist $C_1, C_2 \in \mathrm{Ob}\,\mathcal{C}$, $C_1 \not\cong C_2$, $UC_1 = UC_2 = X$, C_1, C_2 having discrete topologies, such that for the compatible compactifications $\bar{C}_i$ of pC_i and the canonical mappings $f_i \colon \beta X \to \bar{C}_i$ (X taken with discrete topology) we have* $\{f_1^{-1}(A_1) \mid A_1$ is clopen in the space $\bar{C}_1 \setminus C_1\} = \{f_2^{-1}(A_2) \mid A_2$ is clopen in the space $\bar{C}_2 \setminus C_2\}$. *Then the conclusion of statement* 3) *of Proposition 2 holds.*

PROOF. In the first case C_0 does not have a base composed of all partitions of cardinalities less than some cardinal, hence by [23], Remark 2 we have $U(C_0,C_0) \neq X_0^{X_0} = U(rC_0, rC_0)$.

In the second case let $B \subset X$, $B, X \setminus B$ be far in C_1, i.e. $\overline{B}^{\bar{C}_1} \cap \overline{(X \setminus B)}^{\bar{C}_1} = \emptyset$. Then $(\overline{B}^{\bar{C}_1} \setminus C_1) \cap (\overline{(X \setminus B)}^{\bar{C}_1} \setminus C_1) = \emptyset$, hence $f_1^{-1}(\overline{B}^{\bar{C}_1} \setminus C_1) \cap f_1^{-1}(\overline{(X \setminus B)}^{\bar{C}_1} \setminus C_1) = \emptyset$. By hypothesis $\exists A_2$ clopen in $\bar{C}_2 \setminus C_2$, $f_1^{-1}(\overline{B}^{\bar{C}_1} \setminus C_1) = f_2^{-1}(A_2)$, thus also $f_1^{-1}(\overline{(X \setminus B)}^{\bar{C}_1} \setminus C_1) = f_2^{-1}((\bar{C}_2 \setminus C_2) \setminus A_2)$. The disjoint closed subsets A_2, $(\bar{C}_2 \setminus C_2) \setminus A_2$ of $\bar{C}_2$ can be included in disjoint open sets U, V of $\bar{C}_2$. Since $U \cup V \supset \bar{C}_2 \setminus C_2$, therefore $U \cup V$ is cofinite in $\bar{C}_2$. Then $A_2' = U$ is clopen in $\bar{C}_2$, $A_2' \cap (\bar{C}_2 \setminus C_2) = A_2$. Therefore

$$f_1^{-1}(\overline{B}^{\bar{C}_1} \setminus C_1) \subset f_2^{-1}(A_2'),$$

$$f_1^{-1}(\overline{(X \setminus B)}^{\bar{C}_1} \setminus C_1) \subset f_2^{-1}(\bar{C}_2 \setminus A_2').$$

Since $f_1^{-1}(\overline{B}^{\bar{C}_1}) \supset \overline{B}^{\beta X}$, $f_1^{-1}(\overline{(X \setminus B)}^{\bar{C}_1}) \supset \overline{(X \setminus B)}^{\beta X}$ and the left-hand sets are disjoint, both of these inclusions are equalities. Therefore

$$\overline{B}^{\beta X} \setminus X = f_1^{-1}(\overline{B}^{\bar{C}_1} \setminus C_1) \subset f_2^{-1}(A_2') = f_2^{-1}\left(\overline{(A_2' \cap C_2)}^{\bar{C}_2}\right) = \overline{(A_2' \cap C_2)}^{\beta X}$$

and

$$\overline{(X \setminus B)}^{\beta X} \setminus X = f_1^{-1}\left(\overline{(X \setminus B)}^{\bar{C}_1} \setminus C_1\right) \subset f_2^{-1}(\bar{C}_2 \setminus A_2') =$$
$$= f_2^{-1}\left(\overline{(\bar{C}_2 \setminus A_2') \cap C_2}^{\bar{C}_2}\right) = \overline{(\bar{C}_2 \setminus A_2') \cap C_2}^{\beta X},$$

the last equalities following from farness of $A_2' \cap C_2$ and $(\bar{C}_2 \setminus A_2') \cap C_2$ in C_2 (analogously to the equality $f_1^{-1}(\overline{B}^{\bar{C}_1}) = \overline{B}^{\beta X}$). Hence $B \setminus (A_2' \cap C_2)$

and $(X \setminus B) \setminus [(\bar{C}_2 \setminus A_2') \cap C_2]$ are both finite, that is $B \triangle (A_2' \cap C_2)$ and $(X \setminus B) \triangle [(\bar{C}_2 \setminus A_2') \cap C_2]$ are finite. Since $A_2' \cap C_2$ and $(\bar{C}_2 \setminus A_2') \cap C_2$ are far in C_2, and both C_1, C_2 have discrete topologies, this implies $B, X \setminus B$ are far in C_2 as well.

Changing the role of the indices we obtain $(B, X \setminus B$ far in $C_1) \Leftrightarrow (B, X \setminus B$ far in $C_2)$. Hence 1_X is an isomorphism $FC_1 \to FC_2$, but $1_X \notin U(C_1, C_2)$ or $1_X \notin U(C_2, C_1)$. □

REMARKS. 1. The hypothesis in Proposition 2 about the space C_0 or non-indiscreteness of spaces is necessary. Namely if $\mathcal{C} \subset \mathcal{C}_1 = \{C \in \mathrm{Ob\ Unif} \mid C$ has a base composed of all partitions of UC of cardinality less than some cardinal (depending on C)$\} \subset$ Unif, $\mathcal{C} \neq \{$empty space$\}$, $\mathcal{C} \not\subset \{C \mid |UC| = 1\}$, then all the full embeddings $F : \mathcal{C} \to \mathcal{S}_0^-$ can be given — up to natural isomorphy — as follows.

Case 1. $\mathcal{C} \subset \{$discrete spaces$\}$ or $\exists$ cardinal α, $\mathcal{C} \subset \{C \in \mathrm{Ob\ Unif} \mid C$ has a base composed of all partitions of UC of cardinality less than $\alpha\}$; then there are two (concrete) full embeddings, given by $F_1 C = (X, 2^X)$, resp. $F_2 C = (X, \{\emptyset, X\})$ $(X = UC)$.

Case 2. $\mathcal{C} = \mathcal{C}' \cup \mathcal{C}''$, $\mathcal{C}'$ a category like $\mathcal{C}$ in Case 1, containing a non-indiscrete space, $\mathcal{C}''$ a non-empty class of indiscrete spaces with underlying sets of cardinalities > 1; then there is one (concrete) full embedding: $FC = (X, 2^X)$ for $C \in \mathrm{Ob}\,\mathcal{C}'$ and $FC = (X, \{\emptyset, X\})$ for $C \in \mathrm{Ob}\,\mathcal{C}''$.

Case 3. $\mathcal{C}$ is not of the above forms and then there is no full embedding. This follows from Proposition 5 in § 7. Namely $\forall C \in \mathrm{Ob}\,\mathcal{C}_1$ $U(C, C) = (UC)^{(UC)}$, hence, supposing F concrete, by [33], proof of Theorem 4.1, [27], Lemma 2 $FC = (X, 2^X)$ or $FC = (X, \{\emptyset, X\})$. Thus $F\mathcal{C} \subset \{(X, 2^X), (X, \{\emptyset, X\}) \mid X$ is a set $\} = \mathcal{C}_0 \subset \mathcal{S}_0^-$. However, $\mathcal{C}_0$ is isomorphic to the subcategory $\{$discrete proximities$\} \cup \{$indiscrete proximities$\}$ of Unif, hence we can apply Proposition 5 about full embeddings $\mathcal{C} \to$ Unif.

2. One easily finds a single proximity space C with metric completion such that $\{C\} (\subset \mathrm{Prox})$ cannot be fully embedded into $\mathcal{S}_0^-$. It suffices that C satisfies the property required for D in Proposition 2 and $U(C, C) \subsetneq U(rC, rC)$. This last relation holds, e.g., if C or rC is a special uniform space; thus it suffices that rC is infinite, has a discrete topology and a zero-dimensional metric completion ([23], Corollary 6). If rC is required to be any such space C_1 then we can find a space C as follows. Let $\bar{C}_1$ be the completion of C_1 and let C_2 be any connected compact metric space, $|C_2| > 1$. Let us consider $\bar{C}_1 \times C_2$ and select a subspace C of $C_1 \times C_2$ for which $\forall c_1 \in C_1$ $|(\{c_1\} \times C_2) \cap C| = 1$ and $\overline{C}^{\bar{C}_1 \times C_2} = C \cup \Big((\bar{C}_1 \setminus C_1) \times C_2\Big)$. (This can be done since C_2 has a countable dense subset $\{c_{2,n}\}$ and $\bar{C}_1 \setminus C_1$ has a countable dense subset $\{\bar{c}_{1,m}\}$ and in C_1 one can define by induction countably many disjoint sequences, each $\bar{c}_{1,m}$ being the limit in $\bar{C}_1$ of countably many of these sequences, say of $\{c_{1,m,n,i} \mid i \in N\}$, $n \in N$. Then we may

let $C = \{(c_{1,m,n,i}, c_{2,n}) \mid m,n,i \in N\} \cup \{(c_1, c_{2,1}) \mid c_1 \in C_1 \setminus \{c_{1,m,n,i} \mid m,n,i \in$ $\in N\}\}$, with the proximity inherited from $C_1 \times C_2$.) Then the far two-element partitions of C are the same as the (inverse images by the restriction of the projection $C_1 \times C_2 \to C_1$ to C of the) far two-element partitions of C_1. (In fact if $C = C' \cup C''$, C', C'' far in C then their closures $\overline{C'}$, $\overline{C''}$, taken in the compact space $\overline{C}^{\tilde{C}_1 \times C_2}$ form a far partition of $\overline{C}^{\tilde{C}_1 \times C_2}$. If the projections of $\overline{C'}, \overline{C''}$ on $\tilde{C}_1$ are not far then $\exists \tilde{c}_1 \in \tilde{C}_1 \setminus C_1$, $c_2', c_2'' \in C_2$, $(\tilde{c}_1, c_2') \in \overline{C'}, (\tilde{c}_1, c_2'') \in$ $\in \overline{C''}$. Then the connected set $\{\tilde{c}_1\} \times C_2$ intersects both $\overline{C'}$ and $\overline{C''}$, a contradiction.) Therefore in fact $C_1 \cong rC$. The retract property of suitable embedded copies of N^* in C follows from the corresponding property for C_1, shown in Proposition 2, statement 2).

Now we shall prove another statement on the full embeddings of subcategories of Prox into $\mathcal{S}_0^-$. Following J. Pelant and J. Reiterman [24A], let us call an *ultra-proximity* a proximity on a set, X, say, corresponding to a compactification of the set X (dense map into a compact T_2 space) which is the quotient of the Stone–Čech compactification βX of the set X (X taken with the discrete topology) under identification of two different ultrafilters (the dense map is obtained by composing the embedding $X \hookrightarrow \beta X$ with the quotient map; if both ultrafilters are fixed this map is not injective, if one is fixed and the other free it is not onto a topologically discrete subspace). Each proximity is the intersection of all ultraproximities finer than it (the same holding for compactifications) cf. [24A].

PROPOSITION 3. *Let $\mathcal{C} \subset$ Prox and let $\mathcal{C}$ contain each ultraproximity. Let $F\colon \mathcal{C} \to \mathcal{S}_0^-$ be a full embedding. Then $\forall C \in \mathrm{Ob}\,\mathcal{C}$ for $FC = (X, \mathcal{X})$ (where for simplicity we assume $U_{\mathcal{C}} = U_{\mathcal{S}_0^-} F$) we have $\mathcal{X} = \{A \subset X \mid A, X \setminus A$* are far in $C\}$. *If $\exists C_1, C_2 \in \mathrm{Ob}\,\mathcal{C}$ such that for their reflections rC_1, rC_2 in* $\{C \in \mathrm{Ob}$ Prox $\mid C$ has a basis consisting of finite partitions$\}$ *there holds $U(C_1, C_2) \subsetneqq U(rC_1, rC_2)$ (where for simplicity we assume the universal map $C_i \to rC_i$ has underlying function 1_{UC_i}) then there exists no full embedding $F\colon \mathcal{C} \to \mathcal{S}_0^-$. In particular the subcategory* $\{C \in \mathrm{Ob}$ Prox$\mid C$ has a basis consisting of finite partitions$\}$ *admits a unique full embedding into $\mathcal{S}_0^-$ (namely the above one) but no subcategory of* Prox *strictly containing this subcategory admits a full embedding into $\mathcal{S}_0^-$.*

The proof follows the lines of [27], Proposition 5. We begin with a lemma proved on the lines of [27], Lemma 5, and generalizing it (it dealt with the case of C defined by a free and a fixed ultrafilter, which is the fine proximity on a free ultraspace, cf. § 4 or § 6).

LEMMA 2. *Let C be an ultraproximity with underlying set X, let $\mathcal{X} \subset$ $\subset 2^X$ and let $U(C,C) \subset \hom((X,\mathcal{X}),(X,\mathcal{X}))$. If $p, q \in \beta X$ (X taken with the discrete topology) are the ultrafilters used for defining C then for $|X| > 2$ $\mathcal{X} = \{A \subset X \mid p \in \overline{A} \Leftrightarrow q \in \overline{A}\}$ or $\mathcal{X} = \{A \subset X \mid p \in \overline{A} \Rightarrow q \in \overline{A}\}$ or $\mathcal{X} = \{A \subset$*

$\subset X \mid q \in \overline{A} \Rightarrow p \in \overline{A}\}$ *(closure taken in* βX*) or* $\mathcal{X} = 2^X$ *or* $\mathcal{X} \subset \{\emptyset, X\}$*. If* $U(C,C) = \hom((X,\mathcal{X}),(X,\mathcal{X}))$ *then the last two possibilities cannot occur.*

PROOF. C is neither discrete nor indiscrete, thus by [23], Lemma 1 $X^X \neq U(C,C)$, thus in case $U(C,C) = \hom((X,\mathcal{X}),(X,\mathcal{X}))$ we have $\mathcal{X} \not\subset$ $\not\subset \{\emptyset, X\}$, $\mathcal{X} \neq 2^X$. Henceforward we will suppose these last two relations hold. Say $A \in \mathcal{X}$, $\emptyset \neq A \neq X$. Let $B \subset X$, $B, X \setminus B$ far in C. Define $f\colon X \to X$ by $f(B) \subset \{a_1\} \subset A$, $f(X \setminus B) \subset \{a_2\} \subset X \setminus A$. Then $f \in U(C,C) \subset$ $\subset \hom((X,\mathcal{X}),(X,\mathcal{X}))$, hence $B = f^{-1}(A) \in \mathcal{X}$. Hence $\mathcal{X} \supset \{A \subset X \mid p \in$ $\in \overline{A} \Leftrightarrow q \in \overline{A}\}$.

Suppose $\mathcal{X}$ does not equal this set system, i.e. $\exists A \in \mathcal{X}$, where e.g. $p \notin \overline{A}$, $q \in \overline{A}$. Let $A' \subset X$ be any other subset with $p \notin \overline{A'}$, $q \in \overline{A'}$. Then $p \notin \overline{A \cup A'}$, $q \in \overline{A \cap A'}$. Choose points $x \in X \setminus (A \cup A')$, $y \in A \cap A'$. Define $f\colon X \to X$ as identity on $X \setminus (A \cup A')$ and on $A \cap A'$, and let $f(A \setminus A') \subset \{x\}$, $f(A' \setminus A) \subset$ $\subset \{y\}$. Then $f \in U(C,C) \subset \hom((X,\mathcal{X}),(X,\mathcal{X}))$, hence $A' = f^{-1}(A) \in \mathcal{X}$. Thus $\mathcal{X} \supset \{A \subset X \mid p \in \overline{A} \Rightarrow q \in \overline{A}\}$. If here we do not have equality, i.e. $\exists A \in$ $\in \mathcal{X}$, $q \notin \overline{A}$, $p \in \overline{A}$, then similarly as above also $\mathcal{X} \supset \{A \subset X \mid q \in \overline{A} \Rightarrow p \in \overline{A}\}$ holds, therefore $\mathcal{X} = 2^X$. □

The last two cases in the lemma can evidently occur. Even for $U(C,C) =$ $= \hom((X,\mathcal{X}),(X,\mathcal{X}))$ each of the first three possibilities can occur, for the case of a free and a fixed ultrafilter ([27], proof of Proposition 8, since these are the systems of all clopen, open, resp. closed sets of a free ultraspace). For both p, q fixed the second and third possibilities cannot occur; use bijective functions $f \in U(C,C)$, $f(p) = q$, $f(q) = p$. The same holds for both p, q free if $\exists f \in U(C,C)$ bijection, for whose Stone–Čech extension $f^\beta\colon \beta X \to \beta X$ we have $f^\beta(p) = q$, $f^\beta(q) = p$; however, the general case is not clear.

The next lemma follows the lines of [27], Proposition 4.

LEMMA 3. *Let* C *be an ultraproximity on a set* X*, defined by using the ultrafilters* p, q*. Let* D *be an ultraproximity on* $X \times X$*, defined by using some ultrafilters* r, s *such that denoting by* $\pi_i^\beta\colon \beta(X \times X) \to \beta X$ *(*$X \times X$*,* X *taken with discrete topology) the Stone–Čech extension of the projections* $\pi_i\colon X \times X \to X$ *we have* $(\pi_1^\beta(r), \pi_2^\beta(r)) = (p,q)$*,* $(\pi_1^\beta(s), \pi_2^\beta(s)) = (q,p)$*. Then for any concrete functor* $F\colon \{C, D\}(\subset \mathrm{Prox}) \to \mathcal{S}_0^-$ *satisfying* $FC = (X, \{A \subset$ $\subset X \mid p \in \overline{A} \Rightarrow q \in \overline{A}\})$ *we have* $FD = (X \times X, 2^{X \times X})$*. In particular* F *cannot be a full embedding.*

PROOF. Let $FC = (X, \mathcal{X})$, $\mathcal{X} = \{A \subset X \mid p \in \overline{A} \Rightarrow q \in \overline{A}\}$. Let $A \subset X$, $p \notin \overline{A}$ (i.e. $p \in \overline{X \setminus A}$), $q \in \overline{A}$, thus $A \in \mathcal{X}$. Denote FD by $(X \times X, \mathcal{X}')$. Then by $\pi_1 \in U(D,C)$ we have $\mathcal{X}' \ni \pi_1^{-1}(A) = A \times X$. However, $(p,q) \in$ $\in \overline{X \setminus A} \times \beta X$, $(q,p) \in \overline{A} \times \beta X$, hence (from now on closure taking in $\beta(X \times X)$) $r \in \overline{(X \setminus A) \times X}$ (i.e. $r \notin \overline{A \times X}$), $s \in \overline{A \times X}$. Similarly, by using

π_2 we see $\mathcal{X}' \ni X \times A$, $r \in \overline{X \times A}$, $s \notin \overline{X \times A}$. This implies by Lemma 2 and $|X \times X| \geqq 4$ the statement. $\square$

LEMMA 4. *Let $\mathcal{C} \subset$ Unif (Prox) contain a uniform (proximity) space C and some uniformities (proximities) C_α on the same underlying set whose intersection C' is finer than C. Let $F: \mathcal{C} \to \mathcal{S}_0^-$ be a concrete functor. Suppose $\forall \alpha$ $FC_\alpha = (X, \mathcal{X}_\alpha)$, where $\mathcal{X}_\alpha \subset \{A \subset X \mid A, X \setminus A$ are far in $C_\alpha\}$. Then for $FC = (X, \mathcal{X})$ we have $\mathcal{X} \subset \{A \subset X \mid A, X \setminus A$ are far in $C'\}$.*

PROOF. Since $\forall \alpha$ $1_X \in U(C_\alpha, C)$ we have $\mathcal{X} \subset \bigcap_\alpha \mathcal{X}_\alpha$. However, if $A, X \setminus A$ are far in each C_α then they are far in C' as well. In fact $\{A, X \setminus A\}$ is a uniform partition in each C_α, it is a star refinement of itself. Thus it is normal ([17], p. 6) with respect to the (finite) coverings which are uniform in each C_α, i.e. is a uniform cover of C' as well. Therefore $\mathcal{X} \subset \bigcap_\alpha \mathcal{X}_\alpha \subset$ $\subset \{A \subset X \mid \forall \alpha\ A, X \setminus A$ are far in $C_\alpha\} \subset \{A \subset X \mid A, X \setminus A$ are far in $C'\}$. $\square$

PROOF of Proposition 3. We have for each $C \in \mathrm{Ob}\,\mathcal{C}$ $\mathcal{X} \supset \{A \subset X \mid A, X \setminus A$ are far in $C\}$ by Proposition 2. The converse inclusion follows from Lemmas 2, 3 and 4 (in Lemma 4 choosing $\{C_\alpha\} = \{$ all ultraproximities on UC finer than $C\}$, thus having $C' = C$), except for $C =$ two-point indiscrete proximity. In this case, however, $U(C,C) = X^X$, thus $\mathcal{X} = \{\emptyset, X\}$ or $\mathcal{X} = 2^X$ ([33], proof of Theorem 4.1). However, in the second case for any non-indiscrete ultraproximity C' and $FC' = (X', \mathcal{X}')$ we have $(X')^X =$ $= \hom((X, \mathcal{X}), (X', \mathcal{X}')) = \hom(C, C') \neq (X')^X$, a contradiction. The second statement of Proposition 3 follows like in Proposition 2.

Lastly, let $\mathcal{C} \supset \{C \in \mathrm{Ob}$ Prox $| C$ has a basis consisting of finite partitions$\}$ ($\supset \{$ultraproximities$\}$). If here equality holds the functor F given by $FC =$ $= (X, \{A \subset X \mid A, X \setminus A$ are far in $C\})$ is a full embedding. Otherwise $\exists C \in$ $\in \mathrm{Ob}\,\mathcal{C}$ whose reflection $rC \in \mathrm{Ob}\,\mathcal{C}$ given in Proposition 3 (with universal map having underlying function 1_X) is not isomorphic to C, thus $1_X \notin U(rC, C) \subset$ $\subset U(rC, rC) \ni 1_X$, hence $U(rC, C) \subsetneqq U(rC, rC)$. $\square$

§ 6. Embeddings of Coz into $\mathcal{S}_0^-$

We will prove a theorem on full embeddings Coz $\to \mathcal{S}_0^-$. The definition of Coz cf. in § 4 or the beginning of Part II of this paper. We will use the properties of Coz listed in § 4 without further reference. A *free ultraspace* is a space X for which $D \subset X \subset \beta D$, $|X \setminus D| = 1$ for some discrete space D. We note that $\{(U[0,1], \{\text{open sets of } [0,1]\})\} \subset \mathrm{Ob}\,\mathcal{S}_0^-$ has infinitely many not naturally isomorphic full embeddings into $\mathcal{S}_0^-$ ([26]). Therefore the method used in Theorem 3 (§3) is not directly applicable.

First we prove a statement on inductive generation of cozero-spaces.

PROPOSITION 4. *Let $C \in \mathrm{Ob\ Coz}$. Then there is a set of morphisms $f_\alpha : C_\alpha \to C$ in* Coz *inductively generating C (i.e. C has the final structure w.r.t. them) with the following property: for each non-cozero subset A of UC there is an α such that $f_\alpha^{-1}(A)$ is not a cozero-set in C_α, but it is a zero-set in C_α.*

PROOF. Inductive generation follows from the property mentioned in Proposition 4, thus it suffices to prove the mentioned property.

Let $A \subset UC$ ($C \in \mathrm{Ob\ Coz}$) be a non-cozero subset in C. Let $D \neq (\emptyset, \{\emptyset\})$ be a cozero-space, which has a partition $UD = \bigcup_1^\infty A_n$, the union of any cofinitely many A_n-s being dense in the topology generated by D. Consider the product $C' = C \times D \in \mathrm{Ob\ Coz}$. Let $A' =$ inverse image of A by the projection $C \times D \to C$. Let $A'_n = ((UC) \setminus A) \times A_n$. Then C is a retract of C' (with injection $i: C \to C'$ and retraction $r: C' \to C$, say), the inverse image of A under this retraction being A', and any cozero-set in C' intersecting some A'_n intersects infinitely many of the A'_n-s. These will be the only properties of C' used further.

We define the cozero-structure C'' on UC' as the inverse image of the usual topological cozero-structure of $\left\{\frac{1}{n} \mid n \in N\right\} \cup \{0\}$ by the map $g: UC' \to$ $\to \left\{\frac{1}{n} \mid n \in N\right\} \cup \{0\}$ defined by $g(A') = \{0\}$, $\forall n\, g(A'_n) \subset \left\{\frac{1}{n}\right\}$. That is, the cozero-sets of C'' are arbitrary unions of the sets A'_n and the unions of A' and an arbitrary cofinite family of the A'_n-s. Let $C^* = C' \vee C''$ denote the supremum of the cozero-structures C' and C''. The composition of the identical set-map $C^* \to C'$ and the above retraction $r: C' \to C$ gives a map $f: C^* \to C$. Then $(UC^*) \setminus f^{-1}(A) = (UC^*) \setminus A'$ is a cozero-set in C'', hence in C^* as well.

It remained to prove that $f^{-1}(A) = A'$ is not a cozero-set in C^*. C^* is the initial structure w.r.t. the identical set-maps on UC^* to C' and C''. Hence the cozero-sets of C^* are of the form $\bigcup_1^\infty (B'_k \cap B''_k)$, B'_k a cozero-set in C', B''_k a cozero-set in C''.

Suppose on the contrary that A' is a cozero-set in C^*, i.e. is of the form $A' = \bigcup_1^\infty (B'_k \cap B''_k)$ (B'_k, B''_k like above). Since each B''_k either contains A' or is disjoint to it and we also have $A' = \bigcup_1^\infty (B'_k \cap B''_k \cap A')$, in the above representation of A' we may suppose that each B''_k contains A', i.e. B''_k is of the form $B''_k = A' \cup (\cup\{A'_n \mid n \in N \setminus N_k\})$, $N_k \subset N$ finite. We have $B'_k \cap$ $\cap B''_k \subset A'_n$, i.e. $B'_k \cap (\cup\{A'_n \mid n \in N \setminus N_k\}) \subset A'$, or equivalently $B'_k \subset A' \cup (\cup$ $\cup\{A'_n \mid n \in N_k\})$. However, if B'_k intersected $\cup\{A'_n \mid n \in N_k\}$ then it would intersect infinitely many of the A'_n-s, a contradiction. Therefore $B'_k \subset A'$, thus $A' = \bigcup_1^\infty (B'_k \cap B''_k) \subset \bigcup_1^\infty B'_k \subset A'$, hence $A' = \bigcup_1^\infty B'_k$. Thus A' is a cozero-set in C'. Then $i^{-1}A' = i^{-1}r^{-1}A = A$ is a cozero-set in C, a contradiction. $\square$

Unfortunately, our proof does not give a nice inductively generating class for Coz (while $\{[0,1]\}$ is a projectively generating class, since for each cozero-space C {cozero-sets of C} $= \{f^{-1}(0,1] \mid f\colon C \to [0,1]$ is a cozero-map$\}$). Here and also in Theorem 4 $[0,1]$ denotes the cozero-space ($U[0,1]$, {topological cozero-sets of $[0,1]$}). Evidently, Proposition 4 holds for a subcategory $\mathcal{C}$ of Coz, closed under the operation $C \to C^*$. E.g. the following is sufficient: $\mathrm{Ob}\,\mathcal{C}$ contains a space D like in the proof, and all (onto) inverse images of $\left\{\frac{1}{n} \mid n \in N\right\} \cup \{0\}$, and is closed under products (in Coz) of two spaces and suprema (in Coz) of two cozero-structures.

THEOREM 4. 1) *Let $\mathcal{C} \subset$ Coz. Let $[0,1] \subset C_0 \in \mathrm{Ob}\,\mathcal{C}$, and let for any set $\emptyset \neq G \subsetneqq UC_0$ open in the topology generated by C_0 and for any cozero-set A of C_0 there exist $f \in \hom(C_0, C_0)$ such that $f^{-1}(G) = A$.* (E.g. for any set $\emptyset \neq G \subsetneqq UC_0$ open in the topology generated by C_0 there is a copy of $[0,1]$ in C_0 intersecting both G and $(UC_0) \setminus G$.) *Let $\mathcal{C}$ contain all cozero-spaces of the form (UC_0,* {topological cozero-sets of a free ultraspace D with $UC_0 = UD$}) *or* $(UC_0, \{A \cup B \mid A \subset (UC_0) \setminus \{p,q\},\ (B = \emptyset$ *or* $B = \{p,q\})\})$ *where $p, q \in UC_0$, $p \neq q$. Let $F\colon \mathcal{C} \to \mathcal{S}_0^-$ be a full embedding. Then, supposing F concrete, we have for each $C \in \mathrm{Ob}\,\mathcal{C}$ $FC = (X, \mathcal{X})$, where $\mathcal{X} \supset$* {cozero-sets of C}, *or for each $C \in \mathrm{Ob}\,\mathcal{C}$ $FC = (X, \mathcal{X})$, where $\mathcal{X} \supset$* {zero-sets of C}.

2) Let $C \in \mathrm{Ob}\,\mathcal{C}$, $A \in UC$, A no cozero-set in C imply $\exists C' \in \mathrm{Ob}\,\mathcal{C}$, $\exists f \in \hom(C', C)$, $f^{-1}(A)$ is not a cozero-set in C', but it is a zero-set in C'. Let $F\colon \mathcal{C} \to \mathcal{S}_0^-$ be a concrete full embedding satisfying the conclusion of the preceeding statement. Then $Fr\colon \mathcal{C} \to \mathcal{S}_0^-$ is naturally isomorphic to the concrete functor given by $FC = (UC$, {cozero-sets of C}) or to the one given by $FC = (UC$, {zero-sets of C}).

3) Consider Coz as (fully) embedded into Prox by the concrete functor F' given by all finite cozero-covers. Let $\mathcal{C} \subset$ Coz satisfy the conclusion of the last statement (for any full embedding $F\colon \mathcal{C} \to \mathcal{S}_0^-$), let $\mathrm{Ob}\,\mathcal{C} \ni C_0 \supset [0,1]$, let $F'\mathcal{C} \subset \mathcal{C}' \subset$ Prox and suppose $\exists C'_0 \in \mathrm{Ob}\,\mathcal{C}'$ ($\subset$ Ob Prox) whose coreflection cC'_0 in F'Coz satisfies $C'_0 \not\cong cC'_0 \in F'(\mathrm{Ob}\,\mathcal{C})$. Then there is no full embedding $F\colon \mathcal{C}' \to \mathcal{S}_0^-$.

In particular Coz admits just the above two full embeddings into $\mathcal{S}_0^-$ — up to natural isomorphy — but no subcategory of Prox, strictly containing F'Coz admits one.

PROOF. 1) We assume, as we may, for simplicity $U_{\mathcal{C}} = U_{\mathcal{S}_0^-}F$. We remind that $T_{3\frac{1}{2}}$ can be considered as a full subcategory of Coz, by the concrete full embedding $\mathrm{Ob}\,T_{3\frac{1}{2}} \ni D \to (UD$, {topological cozero-sets of D}). Therefore by [27], Proposition 4 for each free topological ultraspace C on UC_0, with the cozero-structure of all topological cozero-sets we have $FC = (X_0, \mathcal{X})$ ($X_0 = UC_0$), where $\mathcal{X} \subset \{$ open sets of the topology generated by $C\}$, or for each of these spaces C we have $FC = (X_0, \mathcal{X})$, where $\mathcal{X} \subset \{$closed sets of the topology generated by $C\}$. In the second case consider instead of F the full embedding $iF\colon \mathcal{C} \to \mathcal{S}_0^-$, where $i\colon \mathcal{S}_0^- \to \mathcal{S}_0^-$ is the concrete isomorphism introduced at the end of § 4 ($i(X,\mathcal{X}) = (X, \{A \subset X \mid X \setminus A \in \mathcal{X}\})$). Thus we may suppose we have the first case (note that the statement of the theorem for iF implies the statement of the theorem for F). For a cozero-space of the form $C = (UC_0, \{A \cup B \mid A \subset (UC_0) \setminus \{p,q\}, (B = \emptyset \text{ or } B = \{p,q\})\})$, $p, q \in UC_0$, $p \neq q$ we have by the remark following Lemma 2 in § 5 $\mathcal{X} = \{$cozero-sets of $C\} = \{$open sets of the topology generated by $C\} = \{$closed sets of the topology generated by $C\}$.

Thus we have for all the above considered spaces C $FC = (X_0, \mathcal{X})$, where $\mathcal{X} \subset \{$open sets of the topology generated by $C\}$. This implies by the proof of [27], Corollary 4 that for $FC_0 = (X_0, \mathcal{X}_0)$ we have $\mathcal{X}_0 \subset \{$open sets of the topology generated by $C_0\}$. (Note that by the coreflectivity of (an embedded copy of) $\mathrm{T}_{3\frac{1}{2}}$ in Coz, mentioned in § 4 and in the beginning of this proof, for C as a topological space finer than the topology generated by C_0, 1_{X_0} is a cozero-map $C \to (X_0, \{$all topological cozero-sets of the topology generated by $C_0\})$, hence also is a cozero-map $C \to C_0$.)

If $\mathcal{X}_0 = \{\emptyset, X_0\}$ then $\hom(C_0, C_0) = \hom((X_0 \mathcal{X}_0), (X_0, \mathcal{X}_0))$ implies $C_0 = (X_0, \{\emptyset, X_0\})$ or $C_0 = (X_0, 2^{X_0})$ ([33]), proof of Theorem 4.1, [27], Lemma 2), contradicting to $[0,1] \subset C_0$. Therefore $\exists G$, $\emptyset \neq G \neq UC_0$, G open in the topology generated by C_0, $G \in \mathcal{X}_0$. By the assumption on C_0 for any cozero-set A of C_0 $\exists f \in \hom(C_0, C_0) = \hom((X_0, \mathcal{X}_0), (X_0, \mathcal{X}_0))$, $f^{-1}(G) = A$. Hence $\{$cozero-sets of $C_0\} \subset \mathcal{X}_0$.

Since the restriction of C_0 to $U[0,1]$ is the cozero-space $[0,1]$, there is a cozero-set A_0 of C_0 for which $A_0 \cap [0,1] = (0,1]$. By what has been shown above, $A_0 \in \mathcal{X}_0$. Take now any $C \in \operatorname{Ob} \mathcal{C}$ and any cozero-set A of C. Then $\exists f \in \hom(C, [0,1]) \subset \hom(C, C_0)$, $A = f^{-1}(0,1] = f^{-1}(A_0)$. Hence for $FC = (X, \mathcal{X})$ we have $A \in \mathcal{X}$, i.e. $\mathcal{X} \supset \{$cozero-sets of $C\}$.

Still we have to show that if C_0 satisfies the hypothesis in brackets in 1) then for any set $\emptyset \neq G \subsetneq UC_0$ open in the topology generated by C_0 and for any cozero set A of C_0 $\exists f \in \hom(C_0, C_0)$, $f^{-1}(G) = A$. C_0 contains a copy of $[0,1]$ intersecting both G and $(UC_0) \setminus G$, and we may assume $[0,1] \cap G = (0,1]$, $[0,1] \cap ((UC_0) \setminus G) = \{0\}$. Then there is a cozero-map $f\colon C_0 \to [0,1]\,(\subset C_0)$ such that $A = f^{-1}(0,1] = f^{-1}(G)$.

2) Similarly like above, by passing to iF if necessary, we may suppose without loss of generality that for each $C \in \operatorname{Ob} \mathcal{C}$ we have $FC = (X, \mathcal{X})$, where $\mathcal{X} \supset \{$cozero-sets of $C\}$. Suppose for some $C \in \operatorname{Ob} \mathcal{C}$ $\mathcal{X} \neq \{$cozero-sets of $C\}$, i.e. $\exists A \in \mathcal{X}$, A is not a cozero-set of C. By hypothesis $\exists C' \in \operatorname{Ob} \mathcal{C}$, $\exists f \in \hom(C', C)$ such that $A' = f^{-1}(A)$ is not a cozero-set of C' (thus, denoting UC' by X', $\emptyset \neq A' \neq X'$) but $X' \setminus A'$ is a cozero-set of C' (thus $C' \neq (X', \{\emptyset, X'\}), C' \neq (X', 2^{X'})$). Let $FC' = (X', \mathcal{X}')$. Then by $f \in$

$\in \hom(C', C) = \hom((X', \mathcal{X}'), (X, \mathcal{X}))$ we have $A' \in \mathcal{X}'$ and by the hypothesis on F we have $X' \setminus A' \in \{$cozero-sets of $C'\} \subset \mathcal{X}'$. Thus $\{A', X' \setminus A'\} \subset \mathcal{X}'$. This implies by $\hom(C', C') = \hom((X', \mathcal{X}), (X', \mathcal{X}'))$, $\{\emptyset, X'\} \neq \{$cozero-sets of $C'\} \neq 2^{X'}$ and [27], Lemma 2 that $\{A', X' \setminus A'\} \subset \{$cozero-sets of $C'\}$, contradicting to $A' \notin \{$cozero-sets of $C'\}$. This contradiction shows $\forall C \in \in \mathrm{Ob}\, \mathcal{C}$ $\mathcal{X} = \{$cozero-sets of $C\}$.

3) By $C_0' \not\cong \mathrm{c}C_0'$ we have $|UC_0'| > 1$, hence by [22], Corollary to Lemma 2 any full embedding $F: \mathcal{C}' \to \mathcal{S}_0^-$ satisfies $U_{\mathcal{C}'} \sim U_{\mathcal{S}_0^-} F$. Thus we may suppose F concrete. Then $FF'I: \mathcal{C} \to \mathcal{S}_0^-$ is a concrete full embedding ($I: \mathcal{C} \hookrightarrow \mathrm{Coz}$ is the inclusion functor), and similarly like above we may suppose $\forall C \in \mathrm{Ob}\, \mathcal{C}$ $FF'C = FF'IC = (UC, \{$cozero-sets of $C\})$.

Let $UC_0' = X_0'$. We have $\mathrm{c}C_0' \in F'(\mathrm{Ob}\, \mathcal{C}) \subset \mathrm{Ob}\, \mathcal{C}'$, $\mathrm{c}C_0' = F'C_0^*$, say, where $C_0^* \in \mathrm{Ob}\, \mathcal{C} \subset \mathrm{Ob}\ \mathrm{Coz}$. We have, considering $[0,1]$ as a cozero-space (and recalling that also c is concrete) $F\mathrm{c}C_0' = FF'C_0^* = (X_0', \{$cozero-sets of $C_0^*\}) = = (X_0', \{f^{-1}(0,1] | f \in \hom(C_0^*, [0,1])\}) = (X_0', \{f^{-1}(0,1] | f \in U(F'C_0^*, F'[0,1]) = = U(\mathrm{c}C_0', [0,1])\})$ (the equality holding since F' is a full embedding, and the last time considering $[0,1]$ as a proximity space, which is the image of the cozero-space $[0,1]$ by F') $= F'^{-1}\mathrm{c}\ \mathrm{c}C_0' = F'^{-1}\mathrm{c}C_0' = (X_0', \{f^{-1}(0,1] \mid f \in \in U(C_0', [0,1])\})$ (by the definition of the functor c, cf. § 4). We have by hypothesis $U(\mathrm{c}C_0', C_0') \ni 1_{X_0'} \notin U(C_0', \mathrm{c}C_0')$. Therefore for $FC_0' = (X_0', \mathcal{X}_0')$ and $F\mathrm{c}C_0' = (X_0', \{f^{-1}(0,1] \mid f \in U(C_0', [0,1])\})$ we have $\hom(F\mathrm{c}C_0', FC_0') \ni \ni 1_{X_0'} \notin \hom(FC_0', F\mathrm{c}C_0')$, i.e. $\mathcal{X}_0' \subsetneqq \{f^{-1}(0,1] \mid f \in U(C_0', [0,1])\}$.

However, by $C_0 \in \mathrm{Ob}\, \mathcal{C}$ we have $FF'C_0 = (UC_0, \{$cozero-sets of $C_0\})$, and since $C_0 \supset [0,1]$, we have $\{$cozero-sets of $C_0\} \mid U[0,1] = \{$cozero-sets of $[0,1]\} \ni (0,1]$. Hence similarly as in 1) we conclude that for each $C' \in \mathrm{Ob}\, \mathcal{C}'$ with $FC' = (X', \mathcal{X}')$ we have $\mathcal{X}' \supset \{f^{-1}(0,1] \mid f \in U(C', [0,1])\}$. For $C' = C_0'$ this is a contradiction, hence such an F does not exist.

Finally, the last statement follows from the above ones, taking into consideration Proposition 4. □

§ 7. Embeddings of Prox and Unif to Unif

There are still two cases missing, those of the full embeddings Prox $\to$ $\to$ Unif and Unif $\to$ Unif. These conjectures of the author have been settled by M. Hušek, J. Pelant (the first one) and by J. Pelant – J. Reiterman (the second one). They have sent a complete proof of the first one and a sketchy proof of the second one and have kindly agreed that these should be included.

THEOREM 5 (M. Hušek – J. Pelant). *Let $\mathcal{C} \subset$ Prox and let $\mathcal{C}$ contain a space C_0 which is a special uniform space containing $[0,1]$. Let further $pN, (pN)^n \in \mathrm{Ob}\, \mathcal{C}$ (N is a countable discrete uniform space, $n > 1$ a fixed integer). Let $F: \mathcal{C} \to$ Unif be a full embedding. Then F is naturally isomorphic* *to the composite inclusion functor* $\mathcal{C} \hookrightarrow \mathrm{Prox} \hookrightarrow \mathrm{Unif}$. *In particular each full embedding* $\mathrm{Prox} \to \mathrm{Unif}$ *is naturally isomorphic to* J.

Here the word "special" can be deleted, cf. § 7, Remark 1.

For the proof we need two lemmas.

LEMMA 5. *Let* $\mathcal{C} \subset \mathrm{Prox}$ *and let* $F\colon \mathcal{C} \to \mathrm{Unif}$ *be a functor with* $pF =$ *= the restriction of* J *to* $\mathcal{C}$. *Let* $X \in \mathrm{Ob}\,\mathcal{C}$ *and* $FX \supset D$, D *a discrete uniform space. Then* $\forall Y \in \mathrm{Ob}\,\mathcal{C}$, $\forall f\colon Y \to pD$ $\{f^{-1}(d) \mid d \in UD\}$ *is a uniform cover of* FY.

PROOF. By $FX \supset D$ we have $X = pFX \supset pD \xleftarrow{f} Y$, since p preserves subspaces. Applying the functor F we have a map $FX \xleftarrow{g} FY$. Thus we have the following commutative diagram, where the underlying functions of the vertical arrows are identities.

$$\begin{tikzcd}
FX \arrow[d] & \supset & D \arrow[d] & \\
X=pFX & \supset & pD & Y=pFY \arrow[l, "f"'] \\
FX \arrow[u] & & & FY \arrow[lll, "g"] \arrow[u]
\end{tikzcd}$$

Hence Ug factors across the underlying function of the embedding of the discrete space D into FX, thus induces a map $FY \to D$, and the lemma follows. $\square$

LEMMA 6. *Let* $\mathcal{C} \subset \mathrm{Unif}$, $X, Y_1, \dots, Y_n \in \mathrm{Ob}\,\mathcal{C}$ *(n a natural number), and let* $f\colon X \to \prod_1^n Y_i$. *Let* $F\colon \mathcal{C} \to \mathrm{Unif}$ *be a concrete functor with* $FY_1, \dots, FY_n$ *discrete. Then* $\{f^{-1}(y_1, \dots, y_n) \mid y_i \in UY_i\}$ *is a uniform cover of* FX.

PROOF. Let π_i denote the projections $\prod_1^n Y_i \to Y_i$. Then $F(\pi_i f)\colon FX \to$ $\to FY_i$, hence we have a map $(F(\pi_1 f), \dots, F(\pi_n f))\colon FX \to \prod_1^n FY_i$ whose underlying map is Uf. Since FY_i is discrete the statement follows. $\square$

PROOF of Theorem 5. We suppose, as we may, for simplicity of notation $U_{\mathcal{C}} = U_{\mathrm{Unif}}F$. Then by § 3, Theorem 3 we have that pF is naturally isomorphic to the composite inclusion $\mathcal{C} \hookrightarrow \mathrm{Prox} \hookrightarrow \mathrm{Unif}$. Thus it suffices to show $F = pF$, i.e., $\forall X \in \mathrm{Ob}\,\mathcal{C}$ FX is precompact. Suppose on the contrary that FX contains a countable discrete subspace N. Applying Lemma 5 with $Y = pN$, $f = 1_{pN}$ we see $FpN = N$. Applying Lemma 6 with $X = (pN)^n$, $Y_i = pN$, $f = 1_{(pN)^n}$ we see $F[(pN)^n] = N^n$. Thus

$(pN)^n = pF[(pN)^n] = p(N^n)$, contradicting to $(pN)^n \neq p(N^n)$ ([17], II. 40). $\square$

THEOREM 6 (J. Pelant – J. Reiterman). *Let F: Unif $\to$ Unif be a full embedding. Then F is naturally isomorphic to the identity functor on* Unif.

In their letter the authors of this theorem have hinted to that the below $Y_{\mathcal{F}}$'s are proximally minimal ([16], p. 410, (2)), which implies $F(Y_{\mathcal{F}}) \cong Y_{\mathcal{F}}$, and since $\{Y_{\mathcal{F}}\}$ inductively generates Unif, $F \sim 1_{\text{Unif}}$. (Proximal minimality means any strictly finer uniformity induces a strictly finer proximity; this can be used together with Theorem 3 if we know that $F(Y_{\mathcal{F}})$ is finer than $Y_{\mathcal{F}}$.) The proof given below is possibly a bit different from theirs, and we give their theorem in a slightly modified form. Before stating it we need some

NOTATIONS. 1) For a set Y and α an infinite cardinal or $\alpha = 2$ $Y(\alpha)$ denotes the uniformity on Y having as base all partitions of Y of cardinality $< \alpha$ (thus $Y(2)$ is indiscrete).

2) For a set Y and $\mathcal{F}$ a (possibly improper) filter on Y $Y_{\mathcal{F}}$ denotes the uniformity on $Y \times \{0,1\}$ having as base all covers $\{\{(y,0)\},\{(y,1)\} \mid y \in \in Y \setminus F\} \cup \{\{(y,0),(y,1)\} \mid y \in F\}$, where $F \in \mathcal{F}$.

The spaces $Y_{\mathcal{F}}$ are useful in several problems, their significance is given by the following

LEMMA 7 ([17], Ch. III, Exercise 3, [16], pp. 410–411). *Each uniform space X is the quotient in* Unif *of some space $Y_{\mathcal{F}}$ (where $Y = (UX)^2$, $\mathcal{F} = \{\text{entourages}\}$). Thus the class of all spaces $Y_{\mathcal{F}}$ inductively generates* Unif. *Moreover, this last statement holds even for the class of all spaces $Y_{\mathcal{F}}$, with $\mathcal{F}$ an ultrafilter.* $\square$

THEOREM 6'. *Let $\mathcal{C} \subset$ Unif and let $\mathcal{C}$ contain non-discrete spaces $Y(\alpha)$ with α arbitrarily large, and a class of spaces $Y_{\mathcal{F}}$ inductively generating* Unif. *Let $F: \mathcal{C} \to$ Unif be a full embedding. Then F is naturally isomorphic to the inclusion $\mathcal{C} \hookrightarrow$ Unif.*

For the proof we need two propositions, the first being a generalization of [23], Proposition 8.

PROPOSITION 5. *Let $\mathcal{C}$ be a subcategory of the category consisting of all uniform spaces $Y(\alpha)$ ($\subset$ Unif), where for convenience we assume $\alpha = 2$ for $|Y| \leqq 1$, $\alpha = \aleph_0$ for $2 \leqq |Y| < \aleph_0$, $\alpha \leqq |Y|^+$ otherwise, and let $\mathcal{C} \neq \{$empty space$\}$, $\mathcal{C} \not\subset \{X \mid |UX| = 1\}$. Let $F: \mathcal{C} \to$ Unif be a full embedding. Then one of the following possibilities holds. 1): F is naturally isomorphic to the inclusion $\mathcal{C} \hookrightarrow$ Unif. 2): Denoting $A = \{\alpha \mid \exists Y, Y(\alpha) \in \operatorname{Ob} \mathcal{C}\}$ the class $A_0 = \{\alpha \in A \mid \alpha$ is not a maximal element of A and $\exists Y$, $Y(\alpha)$ is not discrete, $Y(\alpha) \in \operatorname{Ob}\mathcal{C}\}$ is a set. Let β be the smallest cardinal (β infinite or $\beta = 2$) greater than all $\alpha \in A_0$. F is naturally isomorphic to the concrete functor G defined by $G(Y(\alpha)) = Y(\alpha)$ if α is not a maximal element of A and $Y(\alpha)$*

is not discrete, and $G(Y(\alpha)) = Y(\min(|Y|^+, \beta^*))$ *where* $\beta^* \geqq \beta$ *(and* $\beta^* = 2$ *or* β^* *is an infinite cardinal or a symbol following all cardinals) — and for* $|Y| \leqq 1$ *or* $2 \leqq |Y| < \aleph_0$*, resp.,* $|Y|^+$ *is to be replaced by 2 or by* $\aleph_0$*, resp., — if* α *is a maximal element of* A *or if* $Y(\alpha)$ *is discrete.*

PROOF. Denoting by $U_{\mathcal{C}}: \mathcal{C} \to \mathrm{Set}$ and $U_{\mathrm{Unif}}: \mathrm{Unif} \to \mathrm{Set}$ the underlying set functors, by [22], Corollary to Lemma 2 $U_{\mathcal{C}}$ and $U_{\mathrm{Unif}}F$ are naturally isomorphic; for simplicity of notation we assume $U_{\mathcal{C}} = U_{\mathrm{Unif}}F$. By [23], Remark 2 $U(FY(\alpha), FY(\alpha)) = U(Y(\alpha), Y(\alpha)) = Y^Y$ implies $FY(\alpha) = Y(\alpha')$ for some α'.

Let $Y_1(\alpha_1), Y_2(\alpha_2) \in \mathrm{Ob}\,\mathcal{C}$. Then $U(Y_1(\alpha_1), Y_2(\alpha_2)) = Y_2^{Y_1}$ if $\alpha_2 \leqq \alpha_1$, and $U(Y_1(\alpha_1), Y_2(\alpha_2)) = \{f \in Y_2^{Y_1} \mid |f(Y_1)| < \alpha_1\}$ if $\alpha_1 < \alpha_2$. In the second case for $Y_1(\alpha_1)$ non-discrete (i.e. $\alpha_1 \leqq |Y_1|$) $U(Y_1(\alpha_1), Y_2(\alpha_2)) \neq Y_2^{Y_1}$, since $\alpha_1 < \alpha_2 \leqq \leqq |Y_2|^+$ implies $\alpha_1 \leqq |Y_2|$ hence $\exists f \in Y_2^{Y_1}$, $|f(Y_1)| = \alpha_1$. (Throughout this proof we use the notation $|Y|^+$ in the modified sense given in the statement of the Proposition.) Hence if $Y_1(\alpha_1)$ is non-discrete and α_1 is not maximal in A we have for some $\alpha_2 > \alpha_1$ $U(FY_1(\alpha_1), FY_2(\alpha_2)) = U(Y_1(\alpha_1), Y_2(\alpha_2)) = \{f \in \in Y_2^{Y_1} \mid |f(Y_1)| < \alpha_1\} \neq Y_2^{Y_1}$, which implies $FY_1(\alpha_1) = Y_1(\alpha_1')$, $FY_2(\alpha_2) = = Y_2(\alpha_2')$ with $\alpha_1' < \alpha_2'$. Thus $\{f \in Y_2^{Y_1} \mid |f(Y_1)| < \alpha_1\} = U(Y_1(\alpha_1'), Y_2(\alpha_2')) = = \{f \in Y_2^{Y_1} \mid |f(Y_1)| < \alpha_1'\}$ and $\exists f \in Y_2^{Y_1}$, $|f(Y_1)| = \alpha_1$ imply $\alpha_1 = \alpha_1'$ $(< \alpha_2')$, i.e. $FY_1(\alpha_1) = Y_1(\alpha_1)$.

If, on the other hand, $Y_1(\alpha_1) \in \mathrm{Ob}\,\mathcal{C}_0 = \{Y(\alpha) \in \mathrm{Ob}\,\mathcal{C} \mid Y(\alpha)$ is discrete or α is maximal in $A\}$ ($\mathcal{C}_0$ a subcategory of $\mathcal{C}$) then $\forall Y_2(\alpha_2) \in \mathrm{Ob}\,\mathcal{C}$ $U(Y_1(\alpha_1), Y_2(\alpha_2)) = Y_2^{Y_1}$, hence for $FY_1(\alpha_1) = Y_1(\alpha_1')$, $FY_2(\alpha_2) = Y_2(\alpha_2')$ we have $U(Y_1(\alpha_1'), Y_2(\alpha_2')) = Y_2^{Y_1}$ as well. This implies in turn that $Y_1(\alpha_1')$ is discrete or α_1' is maximal in $A' = \{\alpha' \mid \exists Y, Y(\alpha') \in F(\mathrm{Ob}\,\mathcal{C})\}$. Thus either for all $Y_1(\alpha_1) \in \mathrm{Ob}\,\mathcal{C}_0$ we have $FY_1(\alpha_1)$ discrete, and then we are done, or there is a smallest α_1 such that this does not hold. In the second case we have for this smallest α_1 $FY_1(\alpha_1) = Y_1(\beta^*)$ with β^* ($\leqq |Y_1|$) maximal in A' and also for any $Y(\alpha) \in \mathrm{Ob}\,\mathcal{C}_0$ with $|Y| \geqq |Y_1|$ $FY(\alpha) = Y(\beta^*)$ ($FY(\alpha) = Y(|Y|^+)$ being impossible by $|Y|^+ \geqq |Y_1|^+ > \beta^*$).

If the class A_0 is a proper class then noting $Y(\alpha) \in \mathrm{Ob}\,\mathcal{C} \setminus \mathrm{Ob}\,\mathcal{C}_0 \Rightarrow \Rightarrow FY(\alpha) = Y(\alpha)$ we see $A_0' = \{\alpha' \in A' \mid \alpha'$ is not a maximal element of A' and $\exists Y$, $Y(\alpha')$ is not discrete and $Y(\alpha') \in F(\mathrm{Ob}\,\mathcal{C})\}$ is a proper class as well. Thus the above β^* cannot exist, therefore $F =$ the inclusion $\mathcal{C} \hookrightarrow \mathrm{Unif}$. If A_0 is a set but the above β^* does not exist we are also done. Suppose now A_0 is a set and the above β^* exists. If $\mathcal{C} = \mathcal{C}_0$ then the above considerations prove the statement. Otherwise $\exists Y_1(\alpha_1) \in \mathrm{Ob}\,\mathcal{C} \setminus \mathrm{Ob}\,\mathcal{C}_0$; thus $\exists Y_2(\alpha_2) \in \mathrm{Ob}\,\mathcal{C}$, $\alpha_2 > \alpha_1$ and we have shown above $FY_1(\alpha_1) = Y_1(\alpha_1)$, $FY_2(\alpha_2) = Y_2(\alpha_2')$ with $\alpha_1 < \alpha_2'$. Therefore $\alpha_1 < \beta^*$ and the statement follows. □

PROPOSITION 6. *Let* $\mathcal{C} = \{E, Y_{\mathcal{F}}\} \subset \mathrm{Unif}$. *Let* $F: \mathcal{C} \to \mathrm{Unif}$ *be a concrete full embedding, and let* E *contain a discrete subspace* D *such that the restriction of* FE *to* UD *is discrete as well. Let further* $|D| \geqq |Y|$ $(|D| \geqq 2$ *for* Y

finite). Then $FY_{\mathcal{F}} = Y_{\mathcal{F}}$.

For the proof we need

LEMMA 8. *Let* X *be a uniform space,* $UX = Y \times \{0,1\}$, *and let* $\{\{(y,0),(y,1)\} \mid y \in Y\}$ *be a uniform partition of* X. *Then* $X = Y_{\mathcal{G}}$ *for some (possibly improper) filter* $\mathcal{G}$ *on* Y.

PROOF. It will be convenient to use entourages. Thus $\cup\{\{(y,0),(y,1)\}^2 \mid y \in Y\}$ is an entourage. X has a base consisting of symmetric entourages contained in $\cup\{\{(y,0),(y,1)\}^2 \mid y \in Y\}$, i.e. of entourages of the form $\mathcal{U}_G = (\cup \cup\{\{(y,0),(y,1)\}^2 \mid y \in G\}) \cup (\cup\{\{((y,0),(y,0)), ((y,1),(y,1))\} \mid y \in Y \setminus G\})$, for some $G \subset Y$. Then for each above G_1, G_2 there is an above G_3 such that $\mathcal{U}_{G_1} \cap \mathcal{U}_{G_2} \supset \mathcal{U}_{G_3}$, i.e. $G_1 \cap G_2 \supset G_3$, and for each above G and any $G \subset H \subset Y$ also H is of the form of the above G's. Thus $\mathcal{G} =$ [the set of the above G's] $(\ni Y)$ is a (possibly improper) filter, and $X = Y_{\mathcal{G}}$. □

PROOF of Proposition 6. We have $\{f \in U(Y_{\mathcal{F}}, E) \mid f(Y_{\mathcal{F}}) \subset UD\} = \{f \in \in U(F(Y_{\mathcal{F}}), F(E)) \mid f(F(Y_{\mathcal{F}})) \subset UD\}$. However, $\{\{f^{-1}(x) \mid x \in D\} \mid f \in \in U(Y_{\mathcal{F}}, E), f(Y_{\mathcal{F}}) \subset UD\}$ is a subbase of {uniform partitions of $Y_{\mathcal{J}}$}, and similarly for $F(Y_{\mathcal{F}})$, hence {uniform partitions of $Y_{\mathcal{F}}$} = {uniform partitions of $F(Y_{\mathcal{F}})$}. Thus in particular $\{\{(y,0),(y,1)\} \mid y \in Y\}$ is a uniform partition of $F(Y_{\mathcal{F}})$. Hence by Lemma 8 $F(Y_{\mathcal{F}}) = Y_{\mathcal{G}}$ for some (possibly improper) filter $\mathcal{G}$ on Y. However, both $Y_{\mathcal{F}}$ and $F(Y_{\mathcal{F}}) = Y_{\mathcal{G}}$ have bases consisting of all their uniform partitions, and we have seen above that these bases are equal. Hence $F(Y_{\mathcal{F}}) = Y_{\mathcal{F}}$. □

REMARK 1. One can show Proposition 6 in another way, too, only using $\{\{(y,0),(y,1)\} \mid y \in Y\}$ is a uniform partition of $F(Y_{\mathcal{F}}) = Y_{\mathcal{G}}$ (and $U(Y_{\mathcal{F}}, Y_{\mathcal{F}}) = U(Y_{\mathcal{G}}, Y_{\mathcal{G}})$). Namely one sees easily that $Y \setminus A \in \mathcal{F} \Leftrightarrow U(Y_{\mathcal{F}}, Y_{\mathcal{F}}) = = U(Y_{\mathcal{F}} \setminus (A \times \{0,1\}), Y_{\mathcal{F}}) \times Y_{\mathcal{F}}^{A \times \{0,1\}}$ (i.e., each $f \in U(Y_{\mathcal{F}}, Y_{\mathcal{F}})$ can be changed arbitrarily on $A \times \{0,1\}$, still obtaining an element of $U(Y_{\mathcal{F}}, Y_{\mathcal{F}}) \Leftrightarrow Y \setminus A \in \mathcal{G}$. Actually one can characterize similar subsets of any uniform space X. Let $\emptyset \neq B \subset X$, then $U(X,X) = U(X \setminus B, X) \times X^B$ iff X is indiscrete or X is the uniform sum of B and $X \setminus B$ and there is an infinite cardinal α such that B has as basis all covers of B of cardinality $< \alpha$ and $X \setminus B$ has a basis of uniform covers, each of cardinality $< \alpha$. More generally for $\emptyset \neq B \subset X$ $U(X,Y) = U(X \setminus B, Y) \times Y^B$ iff Y is indiscrete or X is the uniform sum of B and $X \setminus B$ and there is an infinite cardinal α such that every cover of B of cardinality $< \alpha$ is uniform and Y has a basis consisting of covers of cardinality $< \alpha$. In fact the hypothesis implies $U(B,Y) = Y^B$, which implies by [23], Remark 2 the statements about B and Y. If Y is not indiscrete, Y contains a two-point discrete subspace $\{y_1, y_2\}$, say, then $f \in Y^X$ defined by $f(B) = \{y_1\}$, $f(X \setminus B) \subset \{y_2\}$ is uniformly continuous by the hypothesis, hence X is the sum of B and $X \setminus B$. For proximity (topological) spaces by the proof of Theorem 4.1 in [33] we have analogously for $\emptyset \neq B \subset X$

$U(X,Y) = U(X \setminus B, Y) \times Y^B$ $(C(X,Y) = C(X \setminus B, Y) \times Y^B) \Leftrightarrow [\forall y_1, y_2 \in \in Y, (y_1, y_2) \in \mathcal{R} \subset Y^2$ (where the transitive and symmetric (resp. transitive) closure of $\mathcal{R}$ is Y^2) $\forall B' \subset B$ $\{B', B \setminus B'\} \in$ a covering subbase for the discrete proximity on B ($B' \in$ a subbase of the discrete topology on B) $\exists f \in U(B,Y)$ $(C(B,Y))$ $f(B') \subset \{y_1\}$, $f(B \setminus B') \subset \{y_2\}$ and $\forall y_1, y_2 \in \in Y$, $(y_1, y_2) \in \mathcal{R}$ $\exists g \in U(X,Y)$ $g(B) = \{y_1\}$, $g(X \setminus B) \subset \{y_2\}$ $(\forall y_1, y_2 \in Y$, $(y_1, y_2) \in \mathcal{R}$ $\forall b \in B$ $\forall x \in X \setminus B$ $\exists g, h \in C(X,Y)$, $g(b) = y_1$, $g(X \setminus B) \subset \{y_2\}$, $h(x) = y_1$, $h(B) = \{y_2\}] \Leftrightarrow [Y$ is indiscrete or X is the sum of B and $X \setminus B$, B being discrete].

REMARK 2. There is an evident connection between $U(Y_{\mathcal{F}}, Y_{\mathcal{F}})$ and $C'(Y(\mathcal{F}), Y(\mathcal{F}))/\sim$. Here $Y(\mathcal{F})$ is a topological space $Y \cup \{\infty\}$ $(\infty \notin Y)$ with Y open and discrete and neighbourhoods of ∞ are of the form $F \cup \{\infty\}$, $F \in \mathcal{F}$. Further $C'(Y(\mathcal{F}), Y(\mathcal{F})) = \{f \in C(Y(\mathcal{F}), Y(\mathcal{F})) \mid f^{-1}(\infty) = \{\infty\}\}$ and $\sim$ is the equivalence relation on $C'(Y(\mathcal{F}), Y(\mathcal{F}))$ defined by $f \sim g$ iff f and g are identical on a neighbourhood of ∞. This raises the following question. Let us consider pointed topological spaces (X, x_0), $x_0 \in X$. Let $C'(X,X) = \{f \in C(X,X) \mid f^{-1}(x_0) = \{x_0\}\}$ and define $\sim$ as above. Prove speciality results for $C'(X,X)/\sim$ rather than $C(X,X)$, i.e., for some spaces X show that for any space Y (possibly from some restricted class of spaces) the isomorphism of the semigroups $C'(X,X)/\sim$ and $C'(Y,Y)/\sim$ implies e.g. local homeomorphism — or some weaker equivalence property — of X and Y. E.g., for $X = Y = [0,1]$ (or $\{0\} \cup \{\frac{1}{n} \mid n \in N\}$ with the usual topology), $x_0 = y_0 = 0$ is each semigroup isomorphism of $C'(X,X)/\sim$ and $C'(Y,Y)/\sim$ induced by a local homeomorphism $X \to Y$? For X $\mathrm{T}_{3\frac{1}{2}}$ a weaker equivalence property can be, e.g., the homeomorphism type of $X^* = (\beta i)^{-1}(x_0)$, where βi is the Stone–Čech extension of the inclusion $i\colon X \setminus \{x_0\} \to X$ (cf. [19]) and one can ask if the isomorphism of the above semigroups implies homeomorphism of these spaces, which moreover renders the following diagram commutative:

$$\begin{tikzcd}
C'(X,X)/\sim \arrow[r] \arrow[d] & C'(Y,Y)/\sim \arrow[d] \\
C(X^*,X^*) \arrow[r] & C(Y^*,Y^*).
\end{tikzcd}$$

Here the vertical arrows are determined by letting to correspond to an $f \in \in C'(X,X)/\sim$ first its restriction $X \setminus \{x_0\} \to X \setminus \{x_0\}$, then taking Stone–Čech extension, lastly restriction to X^* (and similarly for Y), while the lower horizontal arrow is induced by the homeomorphism $X^* \to Y^*$. Of course one can ask all these questions for categories of spaces rather than for single spaces, and for full embeddings.

PROOF of Theorem 6'. By [22], Corollary to Lemma 2 $U_{\mathrm{Unif}}F \sim U_{\mathcal{C}}$. For simplicity of notation we suppose $U_{\mathrm{Unif}}F = U_{\mathcal{C}}$. Apply now Proposition 5 to the subcategory $\mathcal{C}'$ of $\mathcal{C}$, where $\mathrm{Ob}\,\mathcal{C}' = \{X \in \mathrm{Ob}\,\mathcal{C} \mid X = Y(\alpha)$ for some Y and α, X is non-discrete$\}$. Thus the restriction of F to $\mathcal{C}'$ equals the embedding

$\mathcal{C}' \hookrightarrow$ Unif. Applying Proposition 6 to $\{Z(\alpha), Y_{\mathcal{F}}\} \subset \mathcal{C}$ where $Z(\alpha) \in \operatorname{Ob} \mathcal{C}'$, $\alpha > |Y_{\mathcal{F}}|$ we see $FY_{\mathcal{F}} = Y_{\mathcal{F}}$ for each $Y_{\mathcal{F}} \in \operatorname{Ob} \mathcal{C}$.

By hypothesis $\{Y_{\mathcal{F}} \mid Y_{\mathcal{F}} \in \operatorname{Ob} \mathcal{C}\}$ inductively generates Unif, i.e. $\forall X \in$ $\in \operatorname{Ob}$ Unif has the finest uniformity making each function $f \in U(Y_{\mathcal{F}}, X), Y_{\mathcal{F}} \in$ $\in \operatorname{Ob} \mathcal{C}$, uniformly continuous. Now, by the technique developed in [15] (only applied dually as in Theorem 3) we see for each $C \in \operatorname{Ob} \mathcal{C}$ $FC = C$. $\square$

It is not clear if Unif is either projectively or inductively generated by special uniform spaces. (For topological — resp. $T_{3\frac{1}{2}}$ — spaces both analogous statements are true. For projective generation take any space X having a two-point subspace with exactly three open sets, and such that not all intersections of open sets of X are open (cf. [27], Corollary 1), resp. [0, 1]. For inductive generation — actually representation as a quotient of special spaces — cf. [25], Corollary 4 — where "space" means T_1-space — and [31], Theorem.) One has projective generation of Unif by unit balls of ℓ^∞-spaces ([17], II. 21) and inductive generation by the spaces $Y_{\mathcal{F}}$, $\mathcal{F}$ a filter on Y. However, the former ones are metric and in general not precompact and the same holds for $Y_{\mathcal{F}}$ if Y is infinite and $\mathcal{F}$ has a countable base. By proximal fineness of metric spaces (cf. [17], II. 38) for any such uniform space X we have $U(X,X) = U(pX,pX)$ $(= U(p_\alpha X, p_\alpha X)$, p_α denoting reflection in spaces having bases consisting of coverings of cardinalities smaller than α, [17], p. 52 and II. 33), hence X is not special. It is not clear how can one describe for the above mentioned projectively/inductively generating spaces X all uniform spaces X' with $UX = UX'$, $U(X,X) = U(X',X')$.

We can settle for Prox the above question. We use the term ultraproximity as given in § 5.

LEMMA 9. *Let C be an ultraproximity, $|UC| > 2$. Then C is a special proximity space.*

PROOF. An ultraproximity C satisfies $\delta dC = 0$. Hence by [23], Corollary 3 for any proximity spaces C', D, D' satisfying $UC' = UC$, $UD' = UD$ we have $U(C,D) \subset U(C',D') \Rightarrow D'$ is indiscrete or C' is finer than C, i.e. $C' = C$ or C' is discrete. In particular $U(C,C) = U(C',C') \Rightarrow C = C'$ since by [23], Lemma 1 $U(C,C) = U(C',C')$, C non-discrete, non-indiscrete proximity imply C' is not discrete or indiscrete either. $\square$

The following lemma is proved analogously to [25], Theorem 2 (on sums of special topological spaces).

LEMMA 10. *Let C_α $(\alpha \in A)$ be special proximity (uniform) spaces, $\sup\limits_\alpha |UC_\alpha| > 1$, and let $C = \coprod\limits_{\alpha \in A} C_\alpha$ be their sum. Let D be another proximity (uniform) space with underlying set UC and let $U(D,D) = U(C,C)$. Then D is coarser than C and is finer than the proximity on $\coprod\limits_{\alpha \in A} UC_\alpha$ defined by $B_1 \bar{\delta} B_2 \Leftrightarrow [\forall \alpha \ (B_1 \cap C_\alpha) \bar{\delta} (B_2 \cap C_\alpha)$ in C_α, and with the exception of finitely many α's one of $B_1 \cap C_\alpha$ and $B_2 \cap C_\alpha$ is empty] (or the uniformity on*

$\coprod_{\alpha\in A} UC_\alpha$ *having a base* $\Big\{\mathcal{B}_{\alpha_1}\cup\ldots\cup\mathcal{B}_{\alpha_n}\cup\{\bigcup_{\alpha\in A_i} UC_\alpha \mid i=1,\ldots,m\} \mid n,m\in N,$ $\alpha_1,\ldots,\alpha_n\in A$, $\mathcal{B}_{\alpha_i}$ is a uniform cover of C_{α_i} and $\{A_1,\ldots,A_m\}$ is a finite partition of $A\setminus\{\alpha_1,\ldots,\alpha_n\}\Big\}$*). If, moreover,* $\forall\alpha\ \delta dC_\alpha=0$ *or* $\exists\alpha\ C_\alpha\supset[0,1]$ *then* $D=C$ $(pD=pC)$.

Proof. We treat the case of uniform spaces (the other case is similar). Supposing D indiscrete by [23] Remark 2 C is indiscrete or has for basis all partitions of UC of cardinality less than some infinite cardinal β. If C is indiscrete then $|A|=1$, $A=\{\alpha_0\}$, say, and since $|UC|\geqq\sup_\alpha|UC_\alpha|>1$ we have that C_{α_0} is not special (the indiscrete and discrete uniformities on a set or the uniformity on that set having for base all partitions of cardinality less than some infinite cardinal all have the same set of self-maps). If C has a basis as given above then each C_α has an analogously defined basis consisting of partitions and is special, thus, like above, $\forall\alpha\ |UC_\alpha|\leqq 1$ which has been excluded.

Thus $\exists\{c_1,c_2\}\subset UC=UD$, $\{c_1,c_2\}$ is a discrete subspace of D. Then for any partition $\{A_1,A_2\}$ of A the function f defined by $f(\bigcup_{\alpha\in A_1}C_\alpha)\subset$ $\subset\{c_1\}$, $f(\bigcup_{\alpha\in A_2}C_\alpha)\subset\{c_2\}$ is uniformly continuous from C to C. Hence $\bigcup_{\alpha\in A_1}C_\alpha$ and $\bigcup_{\alpha\in A_2}C_\alpha$ are far in D as well. In particular C_α and $\bigcup_{\alpha'\neq\alpha}C_{\alpha'}$ are far in D. Denote by D_α the subspace of D with $UD_\alpha=UC_\alpha$. We have $U(C_\alpha,C_\alpha)=\{(f\mid C_\alpha)\mid f\in U(C,C),\ f(C_\alpha)\subset C_\alpha\}=\{(g\mid D_\alpha)\mid g\in$ $\in U(D,D), g(D_\alpha)\subset D_\alpha\}=U(D_\alpha,D_\alpha)$. By hypothesis this implies $D_\alpha=C_\alpha$. Hence D is coarser than $\coprod_{\alpha\in A}D_\alpha=\coprod_{\alpha\in A}C_\alpha=C$. D is finer than the uniformity with the base given in the lemma since $D_\alpha=C_\alpha$ and for each partition $\{A_1,A_2\}$ of A $\bigcup_{\alpha\in A_1}D_\alpha$ and $\bigcup_{\alpha\in A_2}D_\alpha$ are far in D.

Suppose now $\forall\alpha\ \delta dC_\alpha=0$, thus $\delta dC=0$, or $\exists\alpha\ C_\alpha\supset[0,1]$, thus $C\supset$ $\supset[0,1]$. Then by [23], Corollary 3 $U(D,D)=U(C,C)$ implies D is indiscrete or D is finer than pC. The first possibility having been excluded above we have $pD=pC$. □

The second statement of Theorem 7 is proved on the lines of its topological analogue [25], Corollary 4.

Theorem 7. *The category* Prox *is inductively generated by special proximity spaces (e.g. by the ultraproximities with underlying sets* X *satisfying* $|X|>2$*). Moreover, each proximity space is the quotient of a special proximity space* D *satisfying* $\delta dD=0$*. Let* $\mathcal{C}\subset$ Prox *contain a subclass consisting of special proximity spaces which inductively generates* Prox *and let* $F:\mathcal{C}\to$ Prox *be a full embedding. Then* F *is naturally isomorphic to the inclusion* $\mathcal{C}\hookrightarrow$ Prox.

Proof. Taking in account Lemma 9 inductive generation follows since each proximity is the intersection of all ultraproximities finer than it (resp. any proximity on a two-element set is a quotient of an ultraproximity on a three-element set). The statement about the full embeddings follows like in Theorem 6'.

Let now C be any non-discrete proximity space, $|UC|>2$ and let $\{C_\alpha\}$ be the set of all ultraproximities on UC finer than C. Following [25], Corollary 4 we put $C^* = \coprod_\alpha C_\alpha$ and $\iota_\alpha : C_\alpha \to C^*$ the canonical injections. Define $f: C^* \to C$ by $U(f\iota_\alpha) = 1_{UC}$. Since C is inductively generated by the maps $f_\alpha : C_\alpha \to C$ with $Uf_\alpha = 1_{UC}$, f is a quotient map. Evidently $\delta dC^* = 0$. Also by the last statement of Lemma 10 C^* is a special proximity space. Further, if C is discrete then either $C = \emptyset$ which is special or C is the quotient of a special proximity space D satisfying $\delta dD = 0$, e.g. of an ultraproximity defined with the help of two fixed ultrafilters, where $|UD|>2$. Lastly, a two-point indiscrete space is a quotient of an ultraproximity on a three-element set. □

Thus we have examples of special proximity spaces which are not special uniform spaces. Namely for an ultraproximity C of an infinite set X, defined with the ultrafilters $p,q \in \beta X$, where $p \in X$, $q \in \beta X \setminus X$, we have, using the remark before § 5, Lemma 2, $U(C,C) = C(\tau C, \tau C) = U(C',C')$, where C' is the fine uniformity on τC (or C' is any $G\tau C$ from § 4, Corollary 2 or the remarks at the beginning of § 4).

A question dual to the second statement of Theorem 7 is the following: is every proximity space the subspace of a special proximity space? (For T_1, resp. topological spaces there is an analogous statement, cf. [25] Corollary 3, resp. [27], Corollary 1.) To answer this question we first prove a proposition which is an analogue of [32] Theorem 4 (it dealt with the case C was a $T_{3\frac{1}{2}}$ space). Also our proposition sharpens some results of [23].

PROPOSITION 7. *Let C be a uniform (proximity, cozero) space, $C \supset [0,1]$. Let further D be another uniform (proximity, cozero) space with $UD = UC$, $U(D,D) = U(C,C)$. Then $pD = pC$ (resp. $D = C$). If C is precompact and has the finest uniformity compatible with its proximity* (i.e. by [18] C has no subspace which is a countable discrete proximity space and also is a retract of a proximal neighbourhood of itself) *then $D = C$.*

LEMMA 11 ([33], Proof of Theorem 3.1). *Let $\mathcal{C}$ be a concrete category, $C_0, C \in \operatorname{Ob}\mathcal{C}$, $i: C_0 \to C$, $r: C \to C_0$, $ri = 1_{C_0}$. Then $\{Uf \mid f \in \hom(C,C), (Uf)(UC) \subset UC_0\} \mid (UC_0) = \{Ug \mid g \in \hom(C_0,C_0)\}$.* □

PROOF of Proposition 7. $[0,1]$ is a retract of C by [17], III. 9, III. 17. Denote D_0 the subspace of D with $UD_0 = U[0,1]$. By Lemma 11 we have $U(D_0,D_0) = U([0,1],[0,1])$, hence by [23], Corollary 5 $D_0 = [0,1]$. Therefore both pC and pD are projectively generated by $\{Uf \mid f \in \hom(C,C), f(C) \subset [0,1]\} = \{Uf \mid f \in \hom(D,D),\ f(D) \subset [0,1]\}$, hence $pC = pD$. The rest is obvious. □

COROLLARY 6. *Each proximity (cozero, $T_{3\frac{1}{2}}$) space is a subspace of a proximity (cozero, $T_{3\frac{1}{2}}$) space, which is special as a uniform space, namely of a product of some power of* $[0,1]$ *and an indiscrete space.* □

Embeddings of proximity spaces into special proximity spaces (even as a subspace far from its complement) are even simpler; using Proposition 7 embed X into the sum $Y = X \coprod [0,1]$. However, Y is not a special uniform space in general; if $X = pX'$, $X \neq X'$, $U(X,X) = U(X',X')$ then $X \coprod [0,1] \neq \neq X' \coprod [0,1]$, $U(X \coprod [0,1], X \coprod [0,1]) = U(X' \coprod [0,1], X' \coprod [0,1])$. (In fact $p(X' \coprod [0,1]) = X \coprod [0,1]$, and if $f\colon X' \coprod [0,1] \to X' \coprod [0,1]$ is proximally continuous then it is uniformly continuous as well since 1) its restriction to $[0,1]$ is uniformly continuous; 2) its restriction to $X' \cap f^{-1}[0,1] \subset X'$ is uniformly continuous, since $p(X' \cap f^{-1}[0,1]) = X \cap f^{-1}[0,1]$ and $f(X' \cap f^{-1}[0,1]) \subset \subset [0,1]$; 3) its restriction g to $X' \cap f^{-1}(X')$ is uniformly continuous since we can extend g to a map $h\colon X' \to X'$ by defining h on $X' \cap f^{-1}[0,1]$ as constant, h is uniformly continuous by hypothesis, hence its restriction g is uniformly continuous as well.)

REMARKS. 1. Proposition 7 implies that in § 3, Theorem 3 (hence also in § 3, Corollary 1 and § 7, Theorem 5) the word "special" can be deleted. In fact, in the proof of Theorem 3, we have by Proposition 7, without speciality, instead of $(*)$ $pC_0 = pFC_0$ (supposing i_{C_0} identity); this, however, suffices to finish the proof of Theorem 3.

2. Almost all the statements of our paper remain valid if we everywhere assume the T_0-axiom. (Possible exceptions are e.g. the last statements of Propositions 2 and 3, § 5, and $T_{3\frac{1}{2}}$ being a maximal subcategory of Top, admitting a full embedding in Prox, second Remark in § 4, where the given proofs do not work in the T_0 case.)

3. In [23], Remark 5 it was asked if e.g. for $X_0 = [0,1]$ or $X_0 =$ Cantor set with the usual uniformity $U(X_0, X_0) \subset U(X,X)$, X some uniformity on the underlying set UX_0, inducing a discrete topology (proximity) implies X has a basis composed of all covers of cardinality less than some cardinal. The answer is no. Namely take any σ-algebra $\Sigma \subset 2^{UX_0}$ containing each one-element set such that $\forall f \in C(X_0, X_0)$ $f^{-1}(\Sigma) \subset \Sigma$ and consider the cozero-space (UX_0, Σ) and remind there is a full embedding Coz $\to$ Prox. (Or consider the uniform space with subbase $\{$countable Borel partitions$\} \cup \cup \{$finite partitions$\}$, which is separable. However, in case $X_0 = [0,1]$ the countable partition $\{P_n \mid n \in N\} = \mathcal{P} = \{\{0,1\}\} \cup \{h(Y + r) \mid r \text{ rational}\}$ is not a uniform cover, where $h\colon R \to (0,1)$ is a homeomorphism and $Y \subset R$ is a maximal set with all $y_1 - y_2$ $(\neq 0)$, $y_1, y_2 \in Y$ irrational, with Y bounded. In fact there is no finite subset $N_1 \subset N$ such that $\cup\{h^{-1}(P_n) \mid n \in N_1\}$ contains a Borel set of positive measure. For the case of the Cantor set C let $g\colon C \to [0,1]$ be the quotient mapping identifying neighbouring points of C in its usual order. Since g induces a homeomorphism of $C \setminus (\{$neighbouring

points$\} \cup \{0,1\})$ onto $[0,1] \setminus \{$dyadic rationals$\}$, $g^{-1}(\mathcal{P})$ is not a uniform cover.)

4. In [22], following [27] we have posed some questions about the determination of all full embeddings of Unif (Prox) into some larger categories. E.g. let $\mathcal{M}$ be the category with objects all pairs (X,μ), X a set, $\mu \subset 2^{2^X}$, with morphisms $f\colon (X_1,\mu_1) \to (X_2,\mu_2)$ characterized by $f\colon X_1 \to X_2, f^{-1}\mu_2 \subset \mu_1$. J. Reiterman has kindly drawn the attention to the fact that there are many more full embeddings Unif $\to \mathcal{M}$ than those given in [22] (in fact at least a proper class). Such are given e.g. by all open (closed, G_δ) uniform covers (or more generally by those consisting of sets which are intersections of $< \alpha$ open sets, α a cardinal). By Proposition 1, using Koubek's strongly rigid class with the fine uniformity, with the concrete functors F' given by the open uniform covers, F'' given by all uniform covers and H given by the uniformity generated by μ as a base we obtain as many not naturally isomorphic full embeddings as there are subclasses of a proper class. However, these examples are still bases of the uniformity. Alternatively, let $\mathcal{N}$ be the category with objects all pairs (X,μ), X a set, $\mu \subset 2^{2^X}$, with morphisms $f\colon (X_1,\mu_1) \to (X_2,\mu_2)$ characterized by $f\colon X_1 \to X_2$, $f\mu_1 \subset \mu_2$. A proper class of not naturally isomorphic concrete full embeddings $F_\alpha\colon \mathrm{Unif} \to \mathcal{N}$ ($2 < \alpha$ is a cardinal) is given by $F_\alpha C = (UC, \{\{A_\lambda\} \mid \{A_\lambda\} (\subset 2^X)$ contains arbitrarily small sets of C (i.e. $\forall$ uniform cover $\mathcal{W}$ of C $\exists W \in \mathcal{W}$, $\exists \lambda$, $A_\lambda \subset W$) and $\forall \lambda\ 1 \leqq |A_\lambda| < \alpha\})$, as follows from [28]. Actually one can give as many not naturally isomorphic concrete full embeddings as there are subclasses of a proper class. Use Koubek's strongly rigid class with the fine uniformity, with the concrete functors $F' = F_3$, $F'' = F_4$, but using dually inductive generation rather than projective generation in the proof of Proposition 1. (Note that this amounts to using the statement about topological categories over any base category in the Remark following § 4, Proposition 1, since topological category $\Leftrightarrow$ cotopological category. Choose Ob $\mathcal{T}_0 = \{(X,\mu) \in$ $\in \mathrm{Ob}\,\mathcal{N} \mid \forall \{A_\lambda\} \in \mu\ \forall \lambda\ |A_\lambda| \leqq 2\}$.)

REFERENCES

[1] Alexandrov, A. D., Additive set functions in abstract spaces, I, II, III, *Rec. Math. (Mat. Sbornik) (N.S.)* **8 (50)** (1940), 307–348, **9 (51)** (1941), 563–628, **13 (55)** (1943), 169–238. *MR* **2**-315, **3**-207, **6**-275

[2] Chvalina, J., Concerning a non-realizability of connected compact semi-separated closure operations by set systems, *Arch. Math. (Brno)* **12** (1976), 45–51. *MR* **55** #11184

[3] Chvalina, J. and Sekanina, M., Realisations of closure spaces by set systems, *Gen. Top. Relations to Modern Anal. Alg.* III. (Proc. III. Prague Top. Symp. 1971), Academia, Prague, 1972, 85–87. *MR* **50** #8415

[4] Engelking, R., *General topology*, PWN–Polish Scientific Publishers, Warsaw, 1977. *MR* **58** #18316b

[5] Gordon, H., Rings of functions determined by zero-sets, *Pacific J. Math.* **36** (1971), 133–157. *MR* **47** #9529

[6] Hager, A. W., Some nearly fine uniform spaces, *Proc. London Math. Soc.* (3) **28** (1974), 517–546. *MR* **53** #1528

[7] Hager, A. W., Perfect maps and epi-reflective hulls, *Canad. J. Math.* **27** (1975), 11–24. *MR* **51** #1751

[8] Hager, A. W., Cozero fields, *Confer. Sem. Mat. Univ. Bari* **175** (1980). *MR* **82g:** 54003

[9] Hausdorff, F., *Mengenlehre*, 2. Auflage, Walter de Gruyter und Co., Berlin–Leipzig, 1927. *Jahrbuch Fortschritte Math.* **53**, 169

[10] Hedrlín, Z., Pultr, A. and Trnková, V., Concerning a categorial approach to topological and algebraic theories, *Gen. Top. Relations to Modern Anal. Alg. II* (Proc. II. Prague Top. Sympos. 1966), Academia, Prague, 1967, 176–181. *MR* **37** #6344

[11] Herrlich, H., Wann sind alle Abbildungen in Y konstant?, *Math. Z.* **90** (1965), 152–154. *MR* **32** #3029

[12] Herrlich, H., Topologische Reflexionen und Coreflexionen, Lecture notes in mathematics. 78, Springer-Verlag, Berlin–New York, 1968. *MR* **41** #988

[13] Herrlich, H., Topological functors, *General Topology Appl.* **4** (1974), 125–142. *MR* **49** #7970

[14] Herrlich, H. and Strecker, G. E., *Category theory*, Allyn and Bacon, Boston, 1973. *MR* **50** #2284

[15] Hušek, M., Construction of special functors and its applications, *Comment. Math. Univ. Carolinae* **8** (1967), 555–566. *MR* **36** #5188

[16] Hušek, M., Lattices of reflections and coreflections in continuous structures, *Categorical Topology* (Proc. Conf. Mannheim, 1975), Lecture notes in mathematics, 540, Springer-Verlag, Berlin–Heidelberg–New York, 1976, 404–424. *MR* **56** #3798

[17] Isbell, J. R., *Uniform spaces*, Providence, R. I., 1964. *MR* **30** #561

[18] Isbell, J. R., Spaces without large projective subspaces, *Math. Scand.* **17** (1965), 89–105; Correction: *Math. Scand.* **22** (1968), 310. *MR* **33** #4882; **41** #989

[19] Jerison, M., Sur l'anneau des germes des fonctions continues, *C. R. Acad. Sci. Paris* **260** (1965), 6507-6509. *MR* **31** #3884

[20] Koubek, V., Each concrete category has a representation by T_2 paracompact topological spaces, *Comment. Math. Univ. Carolinae* **15** (1974), 655–664. *MR* **50** #7283

[21] Magill, Jr., K. D., A survey of semigroups of continuous self-maps, *Semigroup Forum* **11** (1975/76), 189–282. *MR* **52** #14140

[22] Makai, Jr., E., The isomorphisms of the category of uniform spaces and related categories, *Acta Math. Acad. Sci. Hungar.* **32** (1–2) (1978), 121–128. *MR* **80a:** 54051

[23] Makai, Jr., E., Uniformities uniquely determined by their uniformly continuous self-maps, *Studia Sci. Math. Hungar.* **19** (1984), 1–12. *MR* **86e:** 54034

[24] Nyikos, P. J., Epireflective categories of Hausdorff spaces, *Categorical Topology* (Proc. Conf. Mannheim, 1975), Lecture notes in mathematics, 540, Springer-Verlag, Berlin–Heidelberg–New York, 1976, 452–481. *MR* **56** #6589

[24A] Pelant, J. and Reiterman, J., Atoms in uniformities and proximities, *Gen. Top. Relations to Modern Anal. Alg. IV* (Proc. IV. Prague Top. Sympos. 1976), Soc. Czechoslovak Mathematicians and Physicists, Prague, 1977, Part B, 353–356. *MR* **54** #30987

[25] Rosický, J., Remarks on topologies uniquely determined by their continuous self-maps, *Czechoslovak Math. J.* **24 (99)** (1974), 373–377. *MR* **50** #1194

[26] Rosický, J., The topology of the unit interval is not uniquely determined by its continuous self-maps among set systems, *Colloq. Math.* **31** (1974), 179–188. *MR* **51** #1698

[27] ROSICKÝ, J. and SEKANINA, M., Realisations of topologies by set-systems, *Topics in topology* (Proc. Coll. Topology, Keszthely, 1972), North-Holland, Amsterdam–London; J. Bolyai Math. Soc. Budapest; American Elsevier, New York, 1974, 535–555. *MR* **50** #8418

[28] SANDBERG, V. JU., A new definition of uniform spaces, *Dokl. Akad. Nauk SSSR* **135** (1960), 535–537 (in Russian), (transl. *Soviet Math. Dokl.* **1** (1961), 1292–1294). *MR* **23** #A1348

[29] SCHLOMIUK, D. I., An elementary theory of the category of topological spaces, *Trans. Amer. Math. Soc.* **149** (1970), 259–278. *MR* **41** #3559

[30] SEKANINA, M., Embedding of the category of partially ordered sets into the category of topological spaces, *Fund. Math.* **66** (1969/70), 95–98. *MR* **40** #5505

[31] SHIMRAT, M., Decomposition spaces and separation properties, *Quarterly J. Math. Oxford* (2) **7** (1956), 128–129. *MR* **20** #3519

[32] ŠNEPERMAN, L. B., Semigroups of continuous mappings of topological spaces, *Sibirsk. Mat. Ž.* **6** (1965), 221–229, (in Russian). *MR* **31** #1660

[33] WARNDOF, J. C., Topologies uniquely determined by their continuous self-maps, *Fund. Math.* **66** (1969/70), 25–43. *MR* **42** #1048

(Received January 10, 1988)

MTA MATEMATIKAI KUTATÓINTÉZETE
P.O. BOX 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 25-35

ON THE DENSITY OF FLOATING BALLS

L. HÁRS

We have an infinite supply of balls whose radii belong to the fixed interval $[r, R]$, and if we throw these balls into water, they float with half of their body under the water. We deal with the problem to find the densest configuration of the balls on the infinite plane ocean.

Formally: We consider a system of circles (the great circles of the balls) drawn on the plane (the "ocean"), such that their interiors are pairwise disjoint and their radii belong to the fixed interval $[r, R]$. We will prove the following

THEOREM. *There exists a number $q \approx 5.88$, such that if $R/r \leqq q$ then the maximum volume density of the circle packing is attained when all the circles have radius R, and they form a honeycomb system (i.e. all the circles are as large as possible, and each one is tangent to 6 others, like the cells of the honeycomb).*

If $R/r = 3 + 2\sqrt{3} \approx 6.46$ then the maximum volume density circle packing consists of two kinds of circles, having radius r or R. The large circles form a honeycomb system and the small ones are put in the holes between them, touching the 3 large circles that pairwise touch each other. (See Figure 1.)

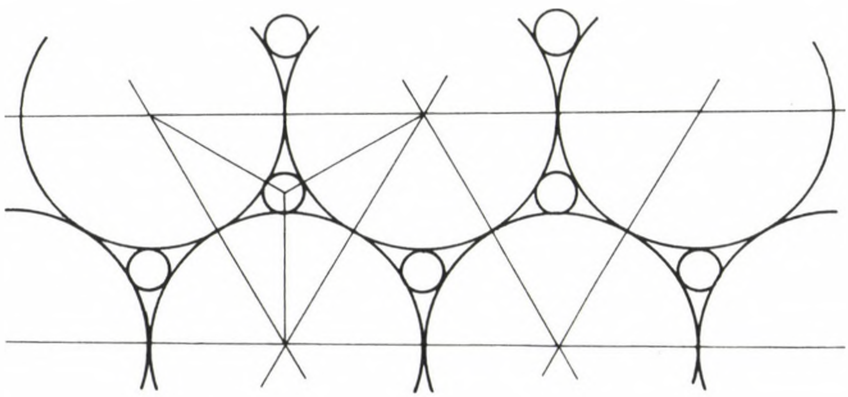


Fig. 1

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 52C17.
Key words and phrases. Packing, circles, density.

Akadémiai Kiadó, Budapest

The cases of other intervals remain unsolved, but Lemma 0 always gives good upper bound for the density.

Roughly speaking, by *volume density* or for short by *density*, we mean the sum of the volumes of the balls in a very large circular "sea", divided by the area of this sea. More precisely, let v_i denote the volume of ball B_i of great circle C_i of the packing, and D_ϱ the circle of radius ϱ centered at some fixed point O on the plane, then the *volume density* is defined by

$$\liminf_{\varrho\to\infty} \frac{\sum_{i:\, C_i\subset D_\varrho} v_i}{\varrho^2\pi}.$$

A similar theorem is proved for perimeter density in [4], while for the usual (area) density only weaker results are known (see [3]).

PROOF. To prove our theorem, we start with the reduction procedure of L. Fejes Tóth and J. Molnár (see [1]):

First we make the circle system saturated by adding new circles to it until no room remains for any other one. This obviously does not decrease the density.

Second we construct the hyperbola cells around the circles, where each cell consists of the points closest to the given circle. It is proved in [1] that the dual of this cellulation consists of triangles, with the following properties:

1. The vertices of these triangles are centers of some circles of our system;

2. These circles do not intersect the opposite sides of the triangles to which they belong;

3. Only the circles centered around the vertices of a certain triangle can have points inside that triangle.

Since these cells tile the plane, the maximum density relative to the triangles gives an upper bound for the density of the whole circle packing. Naturally, the (volume) density relative to a triangle means the sum of the angles weighted by $2/3$ times the third power of the radii of the circles around the vertices, divided by the area of the triangle.

Third step. It is proved in [2] that in a triangle the maximum weighted density occurs when the circles centered at the vertices touch each other, while we move the circles, but do not change their size. In particular, this theorem applies to the volume density as well, if the circles are weighted by $2/3$ times the cube of their radius. Thus we reduce our problem to consider the density in triangles of the type shown by Figure 2, where unity is chosen equal to the longest radius, and so $a \leqq x \leqq 1$.

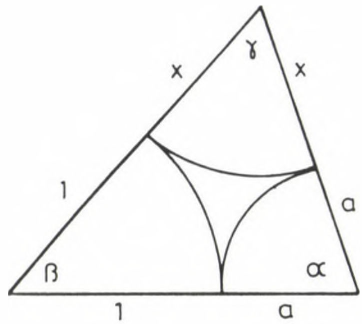


Fig. 2

Fourth step. We consider the density in this triangle when the side length x varies. It will turn out that this density is a quasi-convex function of x if $x \in [a, 1]$, i.e. it is maximal, when $x = a$ or $x = 1$.

Fifth step. Comparing these two values, we find that $x = 1$ gives the larger density.

Sixth step. We fix the two congruent circles of radius 1, and let the third circle vary. We will show that the density first decreases then increases as the smallest radius increases from 0 to 1. Therefore there exists a longest interval $[r_0, 1]$, where the density is maximal, if the third radius also equals to 1. Increasing the three equal radii increases the density, therefore three pairwise touching circles of radius R yield the only "best" cell, where the density is maximum, if an interval $[r, R]$, $R/r \leqq q$ is to be used. Fortunately, the honeycomb circle system can be cut into such cells, thus it is of maximum volume density.

Seventh step. If $R/r = 3 + 2\sqrt{3}$ in the sixth step of the proof we see that the density is maximal if we chose $x = r$, so the "best" cell is formed now of one small and two large circles. The circle system of Figure 1 can be cut into such cells, therefore in this case it is of maximal density.

Proof of Step 4. Using the notations of Figure 2 the density to be considered is given by

$$S(x) = \frac{4}{3} \frac{a^3\alpha + \beta + x^3\gamma}{(a+1)(x+1)\sin\beta}. \tag{1}$$

Disregarding the constant factor $\frac{4}{3(a+1)}$, later on we use

$$S_1(x) = \frac{a^3\alpha + \beta + x^3\gamma}{(x+1)\sin\beta}. \tag{2}$$

From the cosine theorem in the triangle of Figure 2 we get

$$(3)\qquad \begin{aligned} \cos\gamma &= 1-\frac{2a}{(x+1)(x+a)}, & \sin\gamma &= \frac{2\sqrt{ax(x+a+1)}}{(x+1)(x+a)},\\ \cos\beta &= 1-\frac{2ax}{(x+1)(a+1)}, & \sin\beta &= \frac{2\sqrt{ax(x+a+1)}}{(x+1)(a+1)},\\ \cos\alpha &= 1-\frac{2x}{(x+a)(a+1)}, & \sin\alpha &= \frac{2\sqrt{ax(x+a+1)}}{(x+a)(a+1)}. \end{aligned}$$

Here α,β and γ are functions of x. Their derivatives are

$$(4)\qquad \begin{aligned} \gamma' &= \frac{(\cos\gamma)'}{\sin\gamma} = -\frac{a(2x+a+1)}{(x+1)(x+a)}\frac{1}{\sqrt{ax(x+a+1)}},\\ \beta' &= \frac{a}{x+1}\frac{1}{\sqrt{ax(x+a+1)}},\\ \alpha' &= \frac{a}{x+a}\frac{1}{\sqrt{ax(x+a+1)}}. \end{aligned}$$

We have

$$\operatorname{sgn} S_1'(x) = \operatorname{sgn}\frac{S_2(x)}{(x+1)^2\sin^2\beta} = \operatorname{sgn} S_2(x),$$

where

$$(5)\qquad \begin{aligned} S_2(x) = {} & (a^3\alpha' + \beta' + 3x^2\gamma + x^3\gamma')(x+1)\sin\beta - \\ & -(a^3\alpha + \beta + x^3\gamma)[\sin\beta + (x+1)\cos\beta\cdot\beta']. \end{aligned}$$

Applying (3) and (4) the term above in square brackets can be written in the form

$$\frac{2a(2x+a+1)}{\sin\gamma(x+1)(x+a)(a+1)}.$$

It is positive, therefore we can divide S_2 by it without changing the sign. Let us denote the quotient by S_3,

$$(6)\qquad \begin{aligned} S_3(x) = {} & 3x^2\gamma\frac{2x(x+a+1)}{2x+a+1} - x^3(\gamma+\sin\gamma) + \\ & +\Big(\frac{a+1}{2x+a+1}\sin\beta - \beta\Big) + a^3\Big(\frac{a+1}{2x+a+1}\sin\alpha - \alpha\Big). \end{aligned}$$

We need the derivative of $S_3(x)$, too. For this we calculate

$$(7)\qquad \begin{aligned} & \Big[\gamma\frac{6x^3(x+a+1)}{2x+a+1} - x^3(\gamma+\sin\gamma)\Big]' = \\ & = \gamma\Big[6x^2\frac{4x+3a+3-2x(x+a+1)}{(2x+a+1)^2} - 3x^2\Big] + \\ & +\gamma'\Big[\frac{6x^3(x+a+1)}{2x+a+1} - x^3(1+\cos\gamma)\Big] - 3x^2\sin\gamma, \end{aligned}$$

$$\text{(8)} \qquad \left[\frac{a+1}{2x+a+1}\sin\beta-\beta\right]' = -\frac{2(4x+a+3)ax(x+a+1)}{(2x+a+1)^2(x+1)^2\sqrt{ax(x+a+1)}},$$

$$\text{(9)} \qquad \left[\frac{a+1}{2x+a+1}\sin\alpha-\alpha\right]' = -\frac{2(4x+3a+1)ax(x+a+1)}{(2x+a+1)^2(x+a)^2\sqrt{ax(x+a+1)}}.$$

Putting these together we get
(10)

$$S_3'(x) = \gamma x^2 \frac{24x^2+36(a+1)x+15(a+1)^2}{(2x+a+1)^2} - \frac{\sin\gamma}{(2x+a+1)^2(x+1)(x+a)} \times$$
$$\times[16x^6+36(a+1)x^5+(24a^2+72a+24)x^4+(9a^3+39a^2+39a+9)x^3+$$
$$+(3a^4+15a^3+12a^2+15a+3)x^2+(6a^4+6a^3+6a^2+6a)x+3a^4+2a^3+3a^2].$$

Let us denote $24x^2+36(a+1)x+15(a+1)^2$ by q_2, and the polynomial in square brackets by p_6. Let us divide S_3' in (10) by the positive coefficient of γ and denote the result by S_4, with

$$\text{(11)} \qquad S_4(x) = \gamma - \frac{2p_6\sqrt{ax(x+a+1)}}{x^2q_2(x+1)^2(x+a)^2}.$$

We need its derivative, too.

$$S_4'(x) = -\frac{a(2x+a+1)}{(x+1)(x+a)\sqrt{ax(x+a+1)}} -$$

$$\text{(12)} \qquad -\frac{\left[2p_6'\sqrt{ax(x+a+1)}+p_6\frac{a(2x+a+1)}{\sqrt{ax(x+a+1)}}\right]x^2q_2(x+1)^2(x+a)^2}{x^4q_2^2(x+1)^4(x+a)^4}+$$

$$+\frac{2p_6\sqrt{ax(x+a+1)}[(2xq_2+x^2q_2')(x+1)^2(x+a)^2+2x^2q_2(x+1)(x+a)(2x+a+1)]}{x^4q_2^2(x+1)^4(x+a)^4}.$$

If we multiply it with $\sqrt{ax(x+a+1)}(x+1)^3(x+a)^3x^2q_2^2/a$, we get the polynomial S_5. Its coefficients are

(13)

	a^0	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8
x^0	0	0	0	135	495	810	810	495	135
x^1	0	0	405	2460	5355	6600	5355	2460	405
x^2	0	405	4050	12084	17589	17589	12084	4050	405
x^3	135	3060	12366	17460	15462	17460	12366	3060	135
x^4	855	6840	3834	−22857	−22857	3834	6840	855	0
x^5	1485	−2124	−50877	−96456	−50877	−2124	1485	0	0
x^6	−1233	−31941	−109530	−109530	−31941	−1233	0	0	0
x^7	−7518	−53784	−97716	−53784	−7518	0	0	0	0
x^8	−10608	−41232	−41232	−10608	0	0	0	0	0
x^9	−7200	−15552	−7200	0	0	0	0	0	0
x^{10}	−2496	−2496	0	0	0	0	0	0	0
x^{11}	−384	0	0	0	0	0	0	0	0

Here each column is a polynomial of x multiplied by a power of a. The polynomials corresponding to the last three columns are always positive. In each of the other columns the sequences of coefficients have only one change of sign, therefore the polynomials corresponding to them have exactly one positive root according to the Descartes' rule (see e.g. [5]). If we substitute $x = 0.4$ into these polynomials we get the following (positive) results

19.24921	163.6838	540.952
2053.3248	5191.5404	7453.03

Each of these polynomials are negative for sufficiently large values of x, so their roots are larger than 0.4 and $S_5(x) > 0$ for $0 < x \leqq 0.4$. Consequently S_4 and S_3' are increasing functions of x in this interval, thus they have at most one root here.

Now we show that S_3' is positive for $x > 0.4$. For this we need the following

LEMMA 1.

$$1 + \frac{1 - \cos x}{3} < \frac{x}{\sin x} \quad \textit{for } 0 < x < \pi.$$

PROOF of Lemma 1. Multiplying by $\sin x$ we have

$$g(x) := \frac{4}{3}\sin x - \frac{1}{3}\cos x \sin x - x;$$
$$g'(x) = -\frac{2}{3}\cos^2 x + \frac{4}{3}\cos x - \frac{2}{3};$$
$$g''(x) = \frac{4\sin x}{3}(\cos x - 1).$$

Here $g''(x) < 0$, so $g(x)$ is concave and $g'(0) = 0$, the x axis is the tangent of $g(x)$, so the graph of it is under the x axis, what was to be proved.

Applying this Lemma in (10) for γ, $\left(\gamma < \left(1 + \frac{1-\cos\gamma}{3}\right)\sin\gamma\right)$, and multiplying the result with the positive quantity $(2x+a+1)^2(x+a)(x+1)/\sin\gamma$ we get the polynomial q_6, with $q_6 < S_3'$.

$$q_6 = 8x^6 + 24(a+1)x^5 + (27a^2+70a+27)x^4 + (6a^3+66a^2+66a+6)x^3 +$$
$$+ (-3a^4+10a^3+38a^2+10a-3)x^2 - 6a(a^3+a^2+a+1)x - (3a^4+2a^3+3a^2).$$

Here the first four coefficients are positive, the last two ones are negative for all positive a, so with any sign of the fifth coefficient $q_6(x)$ has one positive root. Since $q_6(0) < 0$, if we show that $q_6(0.4) > 0$ with any $0 < a \leqq 1$, it proves that $S_3'(x) > q_6(x) > 0$ here. But

$$q_6(0.4) = -5.88a^4 - 2.416a^3 + 5.5952a^2 + 5.46176a + 0.873728,$$

again a polynomial with one positive root. This polynomial is negative for sufficiently large values of a and for $a = 1$ it takes 3.634688, therefore it is positive for $0 < a \leqq 1$.

These show that S_3' is either positive or first negative then positive for $x \in [0,1]$, i.e. $S(x)$ is quasi-convex here.

PROOF of Step 5. For proving $S(1) < S(a)$ we need

LEMMA 2.

$$\frac{\arcsin x}{x} \geqq 1 + \frac{x^2}{6} \quad \text{if } x \in [0,1],$$
$$\frac{\arcsin x}{x} \leqq 1 + \frac{x^2}{c} \quad \text{if } x \in [0,1/2],$$

where c is defined by the equation $\frac{\arcsin 1/2}{1/2} = 1 + \frac{(1/2)^2}{c}$ $(c = 5.29688498\ldots)$.

PROOF of Lemma 2. Let us define in this case $g(x) = \arcsin x - \left(x + \frac{x^3}{b}\right)$ with the nonzero constant b. Now $g'(0) = 0$ and

$$g''(x) = x\left(\frac{1}{\sqrt{(1-x^2)^3}} - \frac{6}{b}\right).$$

If we chose $b=6$ then $g''(x)>0$ for $x\in[0,1)$, so $g(x)$ is convex, its graph is above the x axis, i.e. its tangent at $x=0$. It proves the first inequality.

If $b=c$ then $g(x)$ is first concave then convex in $[0,1]$ since the function $\frac{1}{\sqrt{(1-x^2)^3}}$ is increasing. The concave segment of $g(x)$ is under the x axis (its tangent at $x=0$), and the convex segment is under its chords, which chords are under the x axis because of the choice of c. This proves the second inequality.

Turning back to the proof of $S(1)<S(a)$, we have to show

$$\frac{a^3(\pi-2\beta_1)+2\beta_1}{(a+1)\frac{\sqrt{a(a+2)}}{a+1}}>\frac{2a^3\alpha_a+(\pi-2\alpha_a)}{(a+1)^2\frac{a\sqrt{2a+1}}{(a+1)^2}}. \tag{14}$$

Here we have

If $x=1$: $\beta_1=\gamma_1$, $\sin\beta_1=\frac{\sqrt{a(a+2)}}{a+1}$, $\alpha_1=\pi-2\beta_1$;

If $x=a$: $\alpha_a=\gamma_a$, $\cos\alpha_a=\frac{a}{a+1}$, $\beta_a=\pi-2\alpha_a$, $\sin\beta_a=\frac{2a\sqrt{2a+1}}{(a+1)^2}$.

Let us substitute $\beta_1=\arcsin\frac{\sqrt{a(a+2)}}{a+1}$ and $\alpha_a=\frac{\pi}{2}-\arcsin\frac{a}{a+1}$ into (14) and multiply the result by $\frac{(a+1)\sqrt{2a+1}}{2(1-a^3)}$ we get

(15)

$$s\Big(\frac{\sqrt{a(a+2)}}{a+1}\Big)\sqrt{2a+1}-s\Big(\frac{a}{a+1}\Big)>\frac{\pi a^2(a+1)^2}{(a^2+a+1)[a+2+\sqrt{a(a+2)(2a+1)}]},$$

where we define $s(z)=\frac{\arcsin z}{z}$. We apply now the first and second inequality of Lemma 2 for the first and second term of (15), respectively, noting that

$$\frac{a(a+2)}{6(a+1)^2}\sqrt{2a+1}-\frac{a^2}{c(a+1)^2}>\frac{1}{(a+1)^2}\Big(\frac{a}{3}-\Big(\frac{1}{c}-\frac{1}{6}\Big)a^2\Big)>0.$$

Consequently it is enough to prove that

$$\sqrt{2a+1}-1=\frac{2a}{\sqrt{2a+1}+1}>\frac{\pi a^2(a+1)^2}{(a^2+a+1)[a+2+\sqrt{a(a+2)(2a+1)}]}, \tag{16}$$

or the equivalent inequality

$$\frac{2}{\pi}>\frac{a(a+1)^2[\sqrt{2a+1}+1]}{(a^2+a+1)[a+2+\sqrt{a(a+2)(2a+1)}]}. \tag{17}$$

The function

$$\frac{(a+1)^2}{a^2+a+1}=1+\frac{1}{(\frac{1}{a}+a)+1} \tag{18}$$

is increasing since $\frac{1}{a}+a$ is a decreasing function in $(0,1]$. If we divide the right-hand side of (17) by this, the result

$$\frac{a(\sqrt{2a+1}+1)}{2+a+\sqrt{a(a+2)(2a+1)}}=\frac{\sqrt{2a+1}+1}{\frac{2}{a}+1+\sqrt{2(a+\frac{1}{a})+5}}. \tag{19}$$

Here the denominator of the right-hand side expression is decreasing, the numerator is increasing, the whole expression is an increasing function of a, thus (17) is only to be verified for $a=1$

$$0.6366197724\ldots=\frac{2}{\pi}>\frac{4(\sqrt{3}+1)}{18}=0.6071224017\ldots.$$

This completes the proof of Step 5.

PROOF of Step 6. Now we have the situation depicted in Fig. 3.

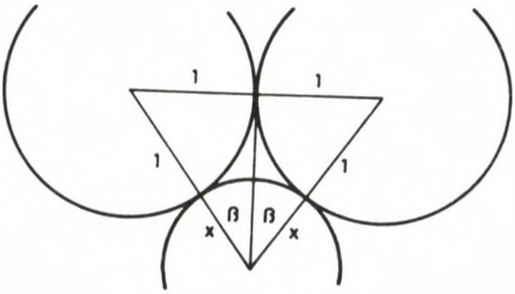


Fig. 3

Here $\beta=\arcsin\frac{1}{1+x}$, and the density to be considered is

$$T(x)=\frac{\frac{\pi}{2}+\beta(x^3-1)}{\sqrt{x^2+2x}}. \tag{20}$$

We need its derivative, too

$$T'(x)=\frac{\left[-\frac{x^3-1}{(x+1)\sqrt{x^2+2x}}+3x^2\beta\right]\sqrt{x^2+2x}-\left[\frac{\pi}{2}+\beta(x^3-1)\right]\frac{x+1}{\sqrt{x^2+2x}}}{x^2+2x}. \tag{21}$$

Here the coefficient of β is positive, because $x\leqq 1$, thus dividing T' by it, the sign does not change. The result is

$$T_1(x)=\beta-\frac{(x^3-1)\sqrt{x^2+2x}+\frac{\pi}{2}(x+1)^2}{(x+1)(2x^4+5x^3+x+1)}. \tag{22}$$

We calculate the derivative T_1' and, multiply it by $(x+1)^2(2x^4+5x^3+$ $+x+1)^2\sqrt{x^2+2x}$ we obtain

$$(23)\qquad \begin{aligned} T_2(x) &= \frac{\pi}{2}\sqrt{x^2+2x}(x+1)^2(6x^4+18x^3+15x^2)- \\ &-x(x+2)(2x^7+14x^6+15x^5+14x^4+38x^3+21x^2+2x+2). \end{aligned}$$

Its sign does not change if we replace both terms by their squares (i.e. multiplying by their positive sum). The result could be divided by $x^2(x+2)$ giving the polynomial

$$(25)\qquad \begin{aligned} &T_3(x)= \\ &=\frac{9\pi^2}{4}(4x^{11}+4x^{10}+176x^9+444x^8+701x^7+708x^6+446x^5+160x^4+25x^3)- \\ &-(4x^{15}+64x^{14}+368x^{13}+988x^{12}+1721x^{11}+3106x^{10}+5068x^9+5622x^8+ \\ &+5664x^7+6008x^6+4073x^5+1534x^4+560x^3+184x^2+20x+8)\approx \\ &\approx -4x^{15}-64x^{14}-368x^{13}-988x^{12}-1632.173x^{11}-2217.7356x^{10}- \\ &-1159.6367x^9+4237.7348x^8+9902.833x^7+9714.279x^6+5831.148x^5+ \\ &+2019.0576x^4-4.83475x^3-184x^2-20x-8. \end{aligned}$$

In this case there are two changes of sign in the sequence of coefficients, so $T_3(x)$ has 0 or 2 positive roots. $T_3(x)$ is negative for $x=0$ and for very large values of x, but $T_3(1)=25054.671>0$. Thus $T_3(x)$ has one root say x_0, in $[0,1]$, where its sign changes from "$-$" to "$+$", and so $T_1(x)$ is decreasing in $[0,x_0]$ and increasing in $[x_0,1]$. Since $T_1(0)=0$ and $T_1(1)=\pi/18$, $T_1(x)$ itself is first negative then positive in $[0,1]$, and so $T(x)$ is first decreasing then increasing here, what was to be proved in Step 6. Numerical calculations for the value x_1, where $T(x_1)=T(1)$ holds give $x_1=0.1701271803\ldots$, $q=1/x_1=5.877955529\ldots$.

PROOF of Step 7. The proof of Step 6 also gives the following

LEMMA 0. *The volume density of the packing of circles of radius from the interval $[r,R]$ never exceeds the volume density relative to the triangle of the centers of three pairwise touching circles*

a) if $R/r\leqq q$ then all three radii are R;

b) if $R/r\geqq q$ then two of the radii are R, the third is r.

Step 7 is a corollary of Lemma 0, and our proof is complete.

REMARK. If we give weights to the circles of the packing, such that it is the pth $(p\geqq 3)$ power of their radius, then part a) of Lemma 0 remains true. Namely, changing the weight of a circle of radius ϱ to $R^{p-3}\varrho^3$ will decrease neither the weight nor the weighted density, and with this new weight-function we can enlarge the circles until their radius equals to R

further increasing the density. But with three pairwise touching circles of radius R the density is the same as with the original weight-function.

Part a) of Lemma 0 is obviously true for linear combinations of the above weight-functions with positive coefficients, so we have the following

THEOREM 1. *The weighted density of a packing of circles of radius from interval* $[r, R]$ *never exceeds the weighted density relative to the triangle of the centers of three pairwise touching circles of radius* R, *if* $R/r \leqq q$ *and the weight of a circle of radius* $\varrho \in [r, R]$ *is given by*

$$\sum_i c_i \varrho^{p_i} \qquad (c_i > 0, \ \ p_i \geqq 3).$$

REFERENCES

[1] FEJES TÓTH, L. and MOLNÁR, J., Unterdeckung und Überdeckung der Ebene durch Kreise, *Math. Nachr.* **18** (1958), 235–243. *MR* **20** #2669

[2] FLORIAN, A., HÁRS, L. and MOLNÁR, J., On the ϱ-systems of circles, *Acta Math. Acad. Sci. Hungar.* **34** (1979), 205–221. *MR* **80j**: 52011

[3] FLORIAN, A., Ausfüllung der Ebene durch Kreise, *Rend. Circ. Mat. Palermo* (2) **9** (1960), 300–312. *MR* **27** #5171

[4] HÁRS, L., Circle packing with maximum total perimeter, *Studia Sci. Math. Hungar.*

[5] PÓLYA, G. and SZEGŐ, G., *Problems and theorems in analysis*, Vol. II, Springer-Verlag, Berlin–Heidelberg–New York, 1976. *MR* **53** #2; *MR* **57** #5529

(Received May 23, 1988)

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

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 37–42

A CONTRIBUTION TO KELLER'S CONJECTURE

K. CORRÁDI and S. SZABÓ

Abstract

The algebraic form of Keller's conjecture holds for the direct sum of cyclic groups of orders $p^e, q, \dots, q$, respectively, where p and q are different primes.

Introduction

The so-called Keller's conjecture is a geometrical problem originally. Namely, in 1930 O. H. Keller [3] conjectured that in a cube tiling which consists of translates of a closed n-dimensional cube there exist two cubes having a common $(n-1)$-dimensional face. The algebraic form of Keller's conjecture is the following. If G is a finite additive abelian group and

$$G = H + [g_1, r_1] + \dots + [g_n, r_n] \tag{1}$$

is a factorization, then $(H - H) \cap \{r_1 g_1, \dots, r_n g_n\} \neq \emptyset$. Here $H - H = \{h - h' : h, h' \in H\}$ and $[g_i, r_i] = \{0, g_i, 2g_i, \dots, (r_i - 1)g_i\}$. For the geometrical background see [9].

If G is the direct sum of the cyclic groups of orders $m_1, \dots, m_k$, respectively, then we will say that the k-tuples of integers $(m_1, \dots, m_k)$ is the type of the group G. Keller's conjecture has been proved for the groups of types (p^e, q^f) in [7] and $(p^e, p, \dots, p)$ in [1], where p and q are distinct primes. Further, independently of the structure of the group G it is proved for $n \leqq 6$ in [4].

In the rest of this paper we prove Keller's conjecture for groups of type $(p^e, q, \dots, q)$. This can be viewed as a step towards the complete solution since according to an argumentation of [10] the verification of Keller's conjecture can be reduced to (p, q)-groups with elementary Sylow subgroups that is whose type is $(p, \dots, p, q, \dots, q)$.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 20K01; Secondary 10E30, 52A45.

Key words and phrases. Factorization of abelian groups, cube tilings.

The result

A subset H of G is said to be periodic if its stabilizer subgroup $\operatorname{stab}(H)$ is not $\{0\}$. $\langle H\rangle$ and $|g|$ denote the generatum of H and the order of g, respectively. We need the next two lemmata.

LEMMA 1 (Proposition 3 of [5] p. 370). *If $G = A + [g,r]$ is a factorization of the finite abelian group G, then $\langle rg\rangle \subset \operatorname{stab}(A)$; and if in addition v is prime to r, then $G = A + [vg, r]$ is a factorization of G.*

LEMMA 2 (Theorem 2 of [2] p. 374). *If p and q are different primes $e \geqq 1$, $f \geqq 1$ and the $m = p^e q^f$-th cyclotomic polynomial divides polynomial $H(x)$ whose coefficients are non-negative integers and whose degree is less than m, then there exist polynomials $P(x)$ and $Q(x)$ with non-negative integer coefficients such that*

$$H(x) = P(x)((1-x^m)/(1-x^{m/p})) + Q(x)((1-x^m)/(1-x^{m/q})).$$

If M_i is the i-th character and g_j is the j-th element of G, then the matrix $M_i(g_j)$ is nonsingular as it may be shown by the standard orthogonality relations. We will use the independence of the columns.

THEOREM. *Keller's conjecture holds for groups of type $(p^e, q, \ldots, q)$, where p and q are distinct primes.*

PROOF. Let G be a group of type $(p^e, q, \ldots, q)$ with basis elements $t, s_1, \ldots, s_u$ of orders $p^e, q, \ldots, q$, respectively. As we have already seen in [1] we may suppose that each r_i is prime and $H \neq \{0\}$ in the factorization (1). We will prove that one of the factors is periodic in (1) which is, in the well-known way, enough to prove our result. To prove it suppose that none of the factors is periodic. We may arrange the factors such that

$$r_1 = \ldots = r_k = p \text{ and } r_{k+1} = \ldots = r_n = q.$$

According to [1] Keller's conjecture holds if the p-component of G is cyclic and either $|H|$ or $r_1 \ldots r_n$ is a power of p. Thus we may assume that $k < n$.

Now we shall prove that $k > 0$. Let $g_i = a_i t + b_{i1}s_1 + \ldots + b_{iu}s_u$, where $0 \leqq a_i \leqq p^e - 1$ and $0 \leqq b_{i1}, \ldots, b_{iu} \leqq q - 1$. Note that $a_i \neq 0$ for $i > k$ since otherwise $r_i g_i = q g_i = q a_i t = 0$ which means that the factor $[g_i, r_i]$ is a cyclic group, that is, a periodic subset. Thus $\langle r_i g_i\rangle = \langle q a_i t\rangle \supset \langle p^{e-1} t\rangle$ for each $k+1 \leqq i \leqq n$.

According to Lemma 1

$$\langle r_i g_i\rangle \subset \operatorname{stab}\Big(H + \sum_{j=1, j\neq i}^{n} [g_j, r_j]\Big).$$

Hence

$$\langle p^{e-1}\rangle \subset \bigcap_{i=1}^{n}\langle r_i g_i\rangle \subset \bigcap_{i=1}^{n}\operatorname{stab}\Big(H+\sum_{j=1,j\neq i}^{n}[g_j,r_j]\Big)=$$
$$=\operatorname{stab}\Big(\bigcap_{i=1}^{n}\Big(H+\sum_{j=1,j\neq i}^{n}[g_j,r_j]\Big)\Big)=\operatorname{stab}(H).$$

Thus H is periodic, unless $k>0$.

If $|a_i t| \geqq p^2$ for each $1\leqq i\leqq k$, then since $\langle r_i g_i\rangle \supset \langle p a_i t\rangle \supset \langle p^{e-1}t\rangle$ using the previous consideration we have that H is periodic. Thus $|a_i t|\leqq p$ for some i, $1\leqq i\leqq k$. We may assume that $|a_1 t|\leqq p$. According to Lemma 1 in the factorization

(2) $$G=H+[g_1,p]+\ldots+[g_k,p]+[g_{k+1},q]+\ldots+[g_n,q]$$

the factors $[g_1,p],\ldots,[g_k,p]$ can be replaced by $[qg_1,p],\ldots,[qg_k,p]$, that is, by $[qa_1t,p],\ldots,[qa_kt,p]$, respectively. So $a_i\neq 0$ for $1\leqq i\leqq k$. Consequently, $|a_1t|=p$ and since $[g_1,p]$ is not a subgroup of G there exists a non-zero term among $b_{11},\ldots,b_{1u}$. We may suppose that $b_{11}\neq 0$. Moreover, we may suppose that $b_{12}=\ldots=b_{1u}=0$. Indeed, hitherto the basis $t,s_1,\ldots,s_u$ was arbitrary. It is clear that $b_{11}s_1+\ldots+b_{1u}s_u$ can be augmented to a basis for the group $\langle s_1,\ldots,s_u\rangle$ and $s_1,\ldots,s_u$ may denote this new basis as well.

Now we prove that $|a_2t|\geqq p^2,\ldots,|a_kt|\geqq p^2$. Assume the contrary, say $|a_2t|\leqq p$. Then $|a_2t|=p$. From $|a_1t|=|a_2t|=p$ it follows that $a_1=p^{e-1}c_1$ and $a_2=p^{e-1}c_2$, where $1\leqq c_1,c_2\leqq p-1$. Let x be the solution of the system of congruences

$$xc_2\equiv qc_1 (\bmod\ p),\quad x\equiv 0 (\bmod\ q).$$

Clearly, this x is prime to p and satisfies the congruences

$$xa_2\equiv qa_1 (\bmod\ p^e),\quad x\equiv 0 (\bmod\ q).$$

In the factorization (2) factors $[g_1,p]$ and $[g_2,p]$ can be replaced by $[qg_1,p]=[qa_1t,p]$ and $[xg_2,p]=[qa_1t,p]$, respectively. But this is a contradiction since in a factorization no factor occurs twice. Thus $|a_2t|\geqq p^2,\ldots,|a_kt|\geqq p^2$.

Summing up our information about $pg_2,\ldots,pg_k,qg_{k+1},\ldots,qg_n$ we have

$$\langle p^{e-1}\rangle \subset \bigcap_{i=2}^{n}\langle r_i g_i\rangle \subset \bigcap_{i=2}^{n}\operatorname{stab}\Big(H+\sum_{j=1,j\neq i}^{n}[g_j,r_j]\Big)=$$
$$=\operatorname{stab}\Big(\bigcap_{i=2}^{n}\Big(H+\sum_{j=1,j\neq i}^{n}[g_j,r_j]\Big)\Big)=\operatorname{stab}(H+[g_1,r_1]).$$

In other words $H+[g_1,r_1]$ has a factorization in the form

(3) $$H+[g_1,r_1]=A+\langle p^{e-1}t\rangle.$$

Let $T=\langle t,s_1\rangle$, $S=\langle s_2,\dots,s_u\rangle$ and let ϱ be a primitive p^eq-th root of unity and let M be a character of G defined by $M(t)=\varrho^q$ and $M(s_1)=\varrho^{p^e}$. Obviously, if M runs over all of these characters then its restriction to S runs over the characters of S.

Applying these characters to the factorization (3) we have

$$\Big(\sum_{h\in H}M(h)\Big)\Big(\sum_{i=0}^{r_1-1}(M(g_1))^i\Big)=\Big(\sum_{a\in A}M(a)\Big)\Big(\sum_{i=0}^{p-1}(M(p^{e-1}t))^i\Big).$$

Note that

$$\sum_{i=0}^{p-1}(M(p^{e-1}t))^i=0$$

and

$$\sum_{i=0}^{r_1-1}(M(g_1))^i=\sum_{i=0}^{p-1}(M(a_1t+s_1))^i=\sum_{i=0}^{p-1}(M(p^{e-1}c_1t+s_1))^i\neq 0.$$

Hence

(4) $$\sum_{h\in H}M(h)=0.$$

Let $H=\{d_1(t+s_1)+h_1,\dots,d_v(t+s_1)+h_v\}$, where $0\leqq d_i\leqq p^eq-1$ and $h_i\in S$. From (4) it follows that

(5) $$\varrho^{d_1}M(h_1)+\ldots+\varrho^{d_v}M(h_v)=0$$

for each character of S.

If $h'_1,\dots,h'_w$ are all the different elements among $h_1,\dots,h_v$, then $H=$ $=H_1+h'_1\cup\ldots\cup H_w+h'_w$ is a partition of H, where $H_j=\{d_i(t+s_1)\colon h_i=$ $=h'_j, h_i\in H\}$. From (5) we have

$$\Big(\sum\nolimits_1\varrho^{d_i}\Big)M(h'_1)+\ldots+\Big(\sum\nolimits_w\varrho^{d_i}\Big)M(h'_w)=0$$

for each character of S. The notation $\sum_j$ means that the summation is taken for $i\in\{i\colon h_i=h'_j\}$. The non-singularity of the character matrix gives that $\left(\sum_1\varrho^{d_i}\right)=\ldots=\left(\sum_w\varrho^{d_i}\right)=0$. Since this holds for each p^eq-th primitive root of unity the $p^e q$-th cyclotomic polynomial divides polynomials $H_j(x) = \sum_j x^{d_i}$ for each $1 \leqq j \leqq w$. According to Lemma 2 there exist polynomials $P_j(x)$ and $Q_j(x)$ with non-negative integer coefficients such that

$$H_j(x) = P_j(x)\left(\left(1 - x^{p^e q}\right) / \left(1 - x^{p^{e-1}q}\right)\right) + Q_j(x)\left(\left(1 - x^{p^e q}\right) / \left(1 - x^{p^e}\right)\right).$$

If $P_j(x)$ is the zero polynomial for each j, $1 \leqq j \leqq w$, then

$$\langle s_1 \rangle \subset \mathrm{stab}(H_j); \text{ and so } \langle s_1 \rangle \subset \mathrm{stab}(H).$$

Otherwise for some j H_j contains a coset modulo the subgroup $\langle p^{e-1}t \rangle$. Thus $H \supset \langle p^{e-1}t \rangle + h$ with a suitable $h \in H$. In the factorization (2) factors H and $[g_1, r_1]$ can be replaced by $H - h$ and $[xg_1, r_1]$, respectively, where x is the solution of the congruences

$$xc_1 \equiv 1 (\mathrm{mod}\ p), \quad x \equiv 0 (\mathrm{mod}\ q).$$

This x satisfies the congruences

$$xa_1 \equiv p^{e-1} (\mathrm{mod}\ p^e), \quad x \equiv 0 (\mathrm{mod}\ q)$$

as well. Now $H - h \supset \langle p^{e-1}t \rangle = [xg_1, r_1]$ violates the factorization. This contradiction completes the proof.

REFERENCES

[1] Corrádi, K. and Szabó, S., Some special cases of Keller's conjecture, *J. Pure Appl. Algebra* **49** (1987), 247–252. *MR* **89b:** 20113

[2] Bruijn, N. G. de, On the factorization of cyclic groups, *Nederl. Akad. Wetensch. Proc. Ser. A* **56** = *Indag. Math.* **15** (1953), 370–377. *MR* **15**–503

[3] Keller, O. H., Über die lückenlose Erfüllung des Raumes mit Würfeln, *J. Reine Angew. Math.* **163** (1930), 231–248. *Jahrbuch Fortschr. Math.* **56**, 1120

[4] Perron, O., Über lückenlose Ausfüllung des n-dimensionalen Raumes durch kongruente Würfeln, I-II, *Math. Z.* **46** (1940), 1–26, 161–180. *MR* **2**-153; **2**-11

[5] Rédei, L., *Algebra*. Vol. 1, Akadémiai Kiadó, Budapest; Pergamon Press, Oxford, 1967. *MR* **35** #2697

[6] Sands, A. D., Factorization of cyclic groups, *Proc. Coll. Abelian Groups* (Tihany, 1963), Akadémiai Kiadó, Budapest, 1964, 139–146. *MR* **30** #3915

[7] Sands, A. D., On Keller's conjecture for certain cyclic groups, *Proc. Edinburgh Math. Soc.* **22** (1979), 17–21. *MR* **81c:** 20013

[8] Seitz, K., Investigations in the Hajós–Rédei theory of finite abelian groups, DM 75–6, Department of Mathematics, Karl Marx University of Economics, Budapest, 1975. *MR* **53** #655

[9] Stein, S. K., Algebraic tiling, *Amer. Math. Monthly* **81** (1974), 445–462. *MR* **49** #4819

[10] Szabó, S., Reductions of a problem of G. Hajós, *Notes on algebraic systems, IV*, DM 85–3, Karl Marx Univ. Econom., Budapest, 1985 *MR* **87d:** 20005

(Received June 19, 1988)

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

BUDAPESTI MŰSZAKI EGYETEM
ÉPÍTŐMÉRNÖKI KAR
MATEMATIKA TANSZÉK
H–1521 BUDAPEST
STOCZEK U. 2
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 43–50

ON A UNIFIED THEORY OF ITERATION METHODS FOR SOLVING NONLINEAR OPERATOR EQUATIONS, III

B. JANKÓ

The present work is a continuation of the papers [5], [6], [7]. Here we have tried to offer a unified theory for certain classes of iteration methods, applied for the solving of nonlinear equations defined in the classical Banach spaces.

We shall resume the above problem in the conditions of B_k-spaces, i.e. linear semiordered complete spaces, normed in a general sense, namely in L. V. Kantorovič's sense. (Here the axiom V of the linear semiordered space is satisfied only for a numerable upper bounded subset [13], p. 21.)

We shall show also in this system of conditions that the concept of convergence order, defined in B_k-spaces, implies a direct influence on the structure of iteration methods. Thus based on a certain principle of construction of iteration methods — still in the conditions of derivability — our purpose is to generate systematically new, large classes of iteration methods. In this way we can generate step by step and classify as well as the obtained iteration methods and the present circumstances enable the common treatment of these methods. We shall give at the same time common conditions for convergence. Obviously, the common treatment of all known iteration methods of higher order is a necessity in the development of this domain, and in this way that represents a fundamental question.

REMARK. Next we shall study only the case of simple solutions of the given nonlinear equation $P(x)=0$, i.e. existence of $[P'(x^*)]^{-1}$ is assumed, x^* being the solution. Elsewhere, the present work is not dealing with the optimality and complexity of the generated methods [25].

1. Let us consider the equation

$$P(x)=\Theta \tag{1}$$

where P is a nonlinear operator defined in a given domain D of a B_k-space X, having — for simplicity — his range also in X, without essentially restricting the conditions.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 65J15; Secondary 47H10.

Key words and phrases. Linear semiordered complete space, convergence order, generating and construction of classes of iteration methods.

Akadémiai Kiadó, Budapest

Our basic problem is to replace in a suitable manner the given operator equation (1), by another equivalent one

$$x - \Psi(x) = \Theta, \tag{1'}$$

in such a way that the convergence order should be $k \geqq 2$ for the iteration method

$$x_{n+1} = \Psi(x_n), \tag{2}$$

where k is a natural number. For this purpose we shall use the following concept of convergence order:

DEFINITION. Let x^* be a solution of the operator equation (1). We say that the above considered iteration method (2) possesses the order of convergence k, if

(i) the generalized norm $|x^* - x_n|$ tends in Kantorovič's sense to the null-element of Y, when $n \to \infty$, where Y is a linear semiordered space, and the generalized norm has its value in this spaces;

(ii) the derivatives of the iteration operator Ψ (introduced in the conditions of the B_k-spaces X [29], pp. 369) satisfy the following equalities:

$$\Psi'(x^*) = O_1,\ \ \Psi''(x^*) = O_2, \dots, \Psi^{(k-1)}(x^*) = O_{k-1},\ \ \Psi^{(k)}(x^*) \neq O_k, \tag{A}$$

where O_i $(i = 1, 2, \dots, k)$ are i-linear null-operators.

In the above mentioned work [7], we have considered already at the first time the following iteration operator

$$\Psi(x) := x - [P'(x) + \mu_1(x)P(x)]^{-1}P(x) + \lambda_2(x)[P(x)]^2, \tag{3}$$

where $\Psi(x)$ is a nonlinear operator with domain $D \subset X$ and range in X; moreover $\mu_1(x)$ and $\lambda_2(x)$ are bilinear operators for fixed x, being defined in the domain $D \times D \subset X \times X$ and having range also in X.

Using the above iteration operator Ψ, we have constructed the corresponding iteration method (2), and we have applied it for solving the equation (1). In such a way we have generated two essential classes of iteration methods of second and third order, respectively. So as particular cases we have obtained the well-known Newton–Kantorovič method, the Tchebycheff method and the method of tangent hyperbolas. Moreover, we have shown that besides these there exist a class of transfinite number of methods of second and of third order, respectively [7], [6], [5], [4].

In the case that the iteration operator was chosen in the form

$$\Psi(x) := x - \left[P'(x) + \mu_1(x)P(x)\right]^{-1}\left(P(x) + \lambda_2(x)[P(x)]^2\right) \tag{3'}$$

we have recovered L. K. Vohandu's method and the method of Ü. Kaasik, which are included as particular case in the class of methods $x_{n+1} = \Psi(x_n)$ constructed by (3') [28], [10], [8].

Some more general iteration operators may be constructed in the following form

$$\Psi(x) := x - [P'(x) + R(x)]^{-1} P(x) + Q(x) \tag{4}$$

and

$$\Psi(x) := x - [P'(x) + R(x)]^{-1} (P(x) + Q(x)), \tag{4'}$$

respectively, where

$$R(x) := \mu_1(x)P(x) + \ldots + \mu_{i+1}(x)[P(x)]^{i+1}$$

and

$$Q(x) := \lambda_2(x)[P(x)]^2 + \ldots + \lambda_{j+1}(x)[P(x)]^{j+1}, \quad (i + j = k - 1).$$

We mention here the operators Q and R that possess certain "multilinear" or "polynomial" character, being constructed by the above multilinear operators λ_i and μ_j.

On the other hand we can notice that the iteration method constructed by (4) contains among others the Tchebycheff-type methods, indicated by a formal development of the inverse of the nonlinear operator P (treated only under the conditions of classical Banach space) [17], [19 p. 72].

It is well-known that certain iteration methods have already been treated — in isolated manner — in semiordered spaces [2], [12], [16].

2. Now we are going to present more general classes of iteration operators and we shall generate certain interesting iteration methods, without limiting, of course, the above mentioned multilinear character. The supposed problem is formally similar to the case of Banach space. However, in the conditions of B_k-space our problem needs an other concept of convergence and convergence order and, of course, moreover we shall use an other notion of differentiability — as in the case of the classical Banach space [29], [13].

Let us now consider the following large class of iteration operators

$$\begin{aligned}\Psi(x) := x - \Big\{\alpha U(x) + \sum_i a_i U(x + R_i(x)) + r(x)\Big\} \cdot \\ \cdot \Big\{\beta P(x) + \sum_{i>2} b_i P(x + Q_i(x)) + q(x)\Big\},\end{aligned} \tag{5}$$

where

$$R_i(x) := \sum_j \mu_j^{(i)}(x)[P(x)]^i, \qquad r(x) := \sum_k \nu_k(x)[P(x)]^k,$$

$$Q_i(x) := \sum_j \lambda_j^{(i)}(x)[P(x)]^j,$$

$$q(x) := \sum_k \chi_k(x)[P(x)]^k,$$

then α, β, a_i, b_i denote real numbers and $\mu_j^{(i)}(x), \nu_k(x), \lambda_j^{(i)}(x), \chi_k(x)$ are also multilinear operators for fixed x. In the next step we shall give the nonlinear operator U of certain concrete expressions.

A. As a first particular case we shall consider the above iteration operator (5) in the following particular form

$$\Psi(x) := x - U(x)P(x + \lambda_2(x)[P(x)]^2).$$

In this case the condition $\Psi'(x^*) = O_1$ leads to the relation

$$\Psi'(x^*) := I - U(x^*)P'(x^*) = O_1,$$

which implies $U(x^*) = \Gamma(x^*)$. This means that the operator U may be chosen more generally — for any x — in the following form

$$U(x) := \Gamma\Big(x + \sum_j \mu_j(x)[P(x)]^j + \sum_k \nu_k(x)[P(x)]^k\Big).$$

In order to choose $\lambda_2(x)$ we have to use the condition $\Psi''(x^*) = O_2$, i.e. $\Psi''(x^*)(\Delta x)^2 = \Theta$ for any $\Delta x \in D$, where Θ is the null-element of X; so we have

$$\begin{aligned}(6)\qquad \Psi''(x^*)(\Delta x)^2 &:= -2U'(x^*)P'(x^*)(\Delta x)^2 - U(x^*)P''(x^*)(\Delta x)^2 - \\ &-U(x^*)P'(x^*)\left\{2\lambda_2(x^*)[P'(x^*)(\Delta x)]^2\right\} = O_2(\Delta x)^2.\end{aligned}$$

Based on the above expression of U and on his first derivative of the form

$$U'(x^*)(\Delta x) := -\Gamma(x^*)P''(x^*)(\Delta x)\Gamma(x^*)(\Delta x)$$

we get from (6) the relation

$$\Gamma(x^*)P''(x^*)(\Delta x)^2 - 2\lambda_2(x^*)[P'(x^*)\Delta x]^2 = \Theta.$$

Thus we have

$$\lambda_2(x^*)(\Delta x)^2 = \frac{1}{2}\Gamma(x^*)\left[P''(x^*)(\Gamma(x^*)(\Delta x)\right]^2$$

and for any x may be chosen

$$\lambda_2(x)(\Delta x)^2 = \frac{1}{2}\Gamma(x)P''(x)[\Gamma(x)\Delta x]^2 + \sum_1 \varrho_1(x)[P(x)]^2.$$

For the generality here we can pose instead of the variable x a multilinear form of P, too.

B. Let us here observe that in the particular case when we have

$$\Psi(x) := x - U(x + \mu(x))P(x)$$

and $U = P^{-1}$ then we obtain from $\Psi'(x^*) = O_1$ and from $\Psi''(x^*) = O_2$ the following relations

$$U(x^*) := \Gamma(x^*), \quad \mu(x^*) := \frac{1}{2}\Gamma(x^*)P(x^*).$$

This iteration method $x_{n+1} = \Psi(x_n)$ of third order constructed in this way needs two inverses but having only a single derivative of first order (it is well-known that sometimes the derivative of second order can be very complicated). This particular method was treated only in case of Banach spaces [15], [4] p. 171.

C. The generalized Traub method

$$x_{n+1} = x_n - \Gamma(x_n)\left\{P(x_n) + P(x_n - \Gamma(x_n)P(x_n))\right\}$$

treated in [26] belongs also to our class of iteration methods given by (5). For this purpose we consider $\alpha = 1$, $a_i = 0$, $r(x) = \Theta$, $\beta = 1$, $b_1 = 1$, $b_i = 0$, $(i = 2, \dots)$, $q(x) = \Theta$. For simplicity we can put $Q_1(x) = U(x)P(x)$. Thus we can use the following special iteration operator

$$\Psi(x) := x + U(x)\{P(x) + P[x + U(x)P(x)]\}.$$

From $\Psi'(x^*)\Delta x = \Theta$ we get $U(x^*) = -\Gamma(x^*)$ and if we choose $U(x) := -\Gamma(x)$ for any $x \in D$, then the condition $\Psi''(x^*)(\Delta x)^2 = \Theta$ shall be satisfied.

The most important are those concrete iteration methods and algorithms, — of course — which can be adapted directly and can be applied effectively at digital computers.

Referring to the construction of classes of iteration methods, in the conditions of B_k-spaces, we mention that in the definition of convergence's order we should have posed instead of (A) the conditions

$$\text{(A')} \qquad \Psi^{(\nu)}(x_n) = 0, \quad (\nu = 1, 2, \dots, k-1), \quad \Psi^{(k)}(x_n) \neq 0.$$

Of course these definitions may be considered equivalents in Banach spaces. However, using this second definition for the construction of iteration methods we shall obtain certain systems of differential equations, which are inconvenient and represent difficulties in the construction of iteration methods.

3. Now we are going to give common conditions of the convergence of order k for the general class of iteration method $x_{n+1} = \Psi(x_n)$, where $\Psi(x)$ is defined by (5). For this purpose let Ψ be uniformly differentiable of order k and we shall use the generalized Taylor's formula, established in the conditions of B_k-space, [2], [23],

$$\Psi(x_n) = \Psi(x^*) + \Psi'(x^*)(x_n - x^*) + \frac{1}{2}\Psi''(x^*)(x_n - x^*)^2 + \ldots$$
$$(7) \qquad \ldots + \frac{1}{(k-1)!}\int\limits_{x^*}^{x_n} \Psi^{(k)}(x)(x_n - x)^{k-1}dx$$

where $x := x^* + t(x_n - x^*)$, $0 \leqq t \leqq 1$ and the notion of integral is considered in the sense [29], [2].

Let us assume that the following conditions are fulfilled:

1°. Let P be defined on the order-segment D, given by the elements x satisfying the inequalities $\underline{x}_0 < x < \overline{x}_0$ where $\underline{x}_0$, $\overline{x}_0$ are certain known initial approximate solutions of the equation (1). There exist some additive and homogeneous operators Λ and Γ with positive inverses Λ^{-1}, Γ^{-1}, such that

a. $\Lambda x = P(x + \Delta x) - P(x) < \Gamma\Delta x$

for any positive Λx and for any $x, x + \Delta x \in [\underline{x}_0, \overline{x}_0]$;

b. $\Gamma^{-1}P(x)$ is monotone (isotone) and (0)-continuous; [29];

c. $P(\underline{x}_0) \leqq \Theta \leqq P(x_0)$

for the initial approximate solutions $\underline{x}_0, \overline{x}_0$, where $x_0 < \overline{x}_0$;

2°. The iteration operator Ψ defined by (5) is uniformly differentiable of order k [2], [23].

3°. The conditions (A) are satisfied for odd k and for $\Psi^{(k)}(x) > O_k$ for any $x \in [\underline{x}_0, \overline{x}_0]$.

These circumstances permit to use our comprehensive theorem established in [7], i.e. the following

THEOREM. *Let us assume that the above conditions* 1°–3° *are satisfied. The operator equation* (1) *possesses a unique solution* x *and the general iteration method* $x_{n+1} = \Psi(x_n)$, *given by* (5) *is convergent of order* k *to* x^*. *The monotone increasing lower iterations* $\{\overline{\underline{x}}_n\}$, *defined by the algorithm*

$$\overline{x}_{n+1} = \Psi(\overline{x}_n) \quad \textit{and} \quad \underline{x}_{n+1} = \Psi(\underline{x}_n)$$

respectively, converge to the solution x^*,

$$x^* := (B_k) - \lim \underline{x}_n = (B_k) - \lim \overline{x}_n,$$

where $\underline{x}_n$ *and* $\overline{x}_n$ *satisfy the following inequalities*

$$(8) \qquad \underline{x}_0 \leqq \underline{x}_1 \leqq \ldots \leqq \underline{x}_{n-1} \leqq x_n \leqq x^* \leqq \overline{x}_n \leqq \overline{x}_{n-1} \leqq \ldots \leqq \overline{x}_1 \leqq \overline{x}_0.$$

REMARK. At last we mention that a similar theorem can be used for the case when k is even by imposing the condition $\Psi^{(k)}(x) < O_k$, instead of $\Psi^{(k)}(x) > O_k$, for any $x \in [\underline{x}_0, \overline{x}_0]$.

REFERENCES

[1] BALÁZS, M. and MUNTEAN, I., A unification of Newton's method for solving equations, *Mathematica (Cluj)* **21 (44)** (1979), 117–122. *MR* **81k**: 65062

[2] BALUEV, A. N., On the method of Čaplygin, *Vestnik Leningr. Univ.* **11** (1956), No. 13, 27–42 (in Russian).

[3] GHINEA, M., Sur la résolution des équations opérationnelles dans les espaces de Banach, *Rev. Française Traitement Information Chiffres* **8** (1965), 3–22. *MR* **32** #593

[4] JANKÓ, B., *Rezolvarea ecuaţilor operaţionale neliniare în spaţii Banach* [The solution of nonlinear operator equations in Banach spaces], Editura Academii Republicii Socialiste România, Bucharest, 1969. *MR* **41** #7497

[5] JANKÓ, B., Sur la théorie unitaire des méthodes d'itération pour la résolution des équations opérationnelles non-linéaires, I, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 301–311. *MR* **27** # 1828

[6] JANKÓ, B., On a unified theory of iteration methods for solving nonlinear operator equations, II, *Ann. Univ. Sci. Budapest. Eötvös Sect. Comput.* **6** (1985), 183–189. *MR* **89h**: 47102

[7] JANKÓ, B., On the unitary theory of iteration methods for solving nonlinear operator equations considered in semiordered spaces, *Ann. Univ. Sci. Budapest. Eötvös Sect. Comput.* **2** (1979), 85–90. *MR* **83h**: 65071

[8] JANKÓ, B. and OLTEANU, D., On the Kaasik's method for solving nonlinear operator equations, *Buletinul Stiintific Institut*, Baia-Mare, Ser. B. **1** (1969), 315–317 (in Romanian).

[9] JARRATT, P., Some efficient fourth order multipoint methods for solving equations, *Nordisk Tidskr. Informationsbehandling (BIT)* **9** (1969), 119–124. *MR* **39** #7809

[10] KAAZIK, YU. YA., On a class of iteration processes for the solution of the operation equations, *Dokl. Akad. Nauk SSSR (N. S.)* **112** (1957), 579–582. *MR* **20** #234

[11] KALABA, R. E., On nonlinear differential equations, the maximum operation, and monotone convergence, *J. Math. Mech.* **8** (1959), 519–574. *MR* **21** #6453

[12] KANTOROVIČ, L. V., The principle of the majorant and Newton's method, *Doklady Akad. Nauk SSSR (N.S.)* **76** (1951), 17–20 (in Russian). *MR* **12**–835

[13] KANTOROVIČ, L. V., VULIH, B. Z. and PINSKER, A. G., *Functional analysis in partially ordered spaces*, Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow–Leningrad, 1950. *MR* **12**–340

[14] KING, R. F., A family of fourth order methods for nonlinear equations, *SIAM J. Numer. Anal.* **10** (1973), 876–879. *MR* **49** # 8326

[15] LONGAVER, L., On solving nonlinear operator equations, communicated at the National Session of the Romanian Students, Cluj, 1972 (unpublished).

[16] MIRAKOV, V. E., The majorant principle for Čebyšev's method, *Uspehi Mat. Nauk (N. S.)* **11** (1956), no. 3, 171–174 (in Russian). *MR* **20** #1932

[17] NEČEPURENKO, M. T., On Čebyšev's method for functional equations, *Uspehi Mat. Nauk (N. S.)* **9** (1954), no. 2, 163–170. *MR* **15**–801

[18] ORTEGA, J. M. and RHEINBOLDT, C. W., *Iterative solution of nonlinear equations in several variables*, Academic Press, New York–London, 1970. *MR* **42** #8686

[19] PAVALOIU, I., *Introduction to the approximation theory of the solving of equations*, Dacia Publishing House, Cluj, 1976 (in Romanian).

[20] PTÁK, V., The rate of convergence of Newton's process, *Numer. Math.* **25** (1975), 279–285. *MR* **57** #18064

[21] RHEINBOLDT, W. C. and VANDERGRAFT, J. S., On the local convergence of update methods, *SIAM J. Numer. Anal.* **11** (1974), 1069–1085. *MR* **52** #2188

[22] SCHRÖDER, J., Inclusion statements for operator equations by a continuity principle, *Manuscripta Math.* **21** (1977), 135–171. *MR* **56** #9343

[23] Slugin, S. N., Approximate solution of operator equations on the basis of S. A. Čaplygin's method, *Dokl. Akad. Nauk SSSR (N. S.)* **103** (1955), 565–568 (in Russian). *MR* **17**–387

[24] Szabó, B., On a method of fourth convergence's order for solving of nonlinear equations defined in Banach spaces, *Ann. Univ. Sci. Budapest. Eötvös Sect. Comput.*

[25] Traub, J. F., Wasilkowski, G. W. and Wozniakowski, H., *Information, uncertainty, complexity*, Addison–Wesley Publishing Co, Reading–London, 1983. *Zbl* **522**: 68041; *MR* **85g**: 68031

[26] Traub, J. F., *Iterative methods for the solution of equations*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964.

[27] Vandergraft, J., Newton's method for convex operators in partially ordered spaces, *SIAM J. Numer. Anal.* **4** (1967), 406–432.

[28] Vohandu, L. K., On the iterative methods for solving equations, Thesis, Univ. Tartu, 1955.

[29] Vulih, B. Z., *Introduction to the theory of partially ordered spaces*, Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1961 (in Russian). *MR* **24** #A3494

[30] Voler, R., Iterative Einschliessungen von Lösungen nichtlinearer Differentialgleichungen durch Newton-ähnliche Iterationsverfahren, *Aplikace Mat.* **31** (1986), no. 1, 1–18.

[31] Wolfe, M. A., Extended iterative methods for the solution of operator equations, *Numer. Math.* **31** (1978), 153–174. *MR* **80a**: 65126

[32] Woźniakowski, H., Maximal stationary iterative methods for the solution of operator equations, *SIAM J. Numer. Anal.* **11** (1974), 934–949. *MR* **51** #4667

(Received June 16, 1988)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
NUMERIKUS ANALÍZIS TANSZÉK
H-1088 BUDAPEST
MÚZEUM KRT. 6–8
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 51–63

ИНТЕРПОЛЯЦИОННЫЕ ПРОЦЕССЫ ЭРМИТА–ФЕЙЕРА С ГРАНИЧНЫМИ УСЛОВИЯМИ

Д. Л. БЕРМАН

Пусть

$$x_k = x_{kn} = \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \dots, n, \quad n = 1, 2, \dots \tag{1}$$

и C-множество всех функций $f(x)$, непрерывных в $[-1,1]$. Обозначим через $H_n(f,x)$ многочлен степени $2n-1$, однозначно определяемый из условий

$$H_n(f, x_{kn}) = f(x_{kn}), \quad H_n'(f, x_{kn}) = 0, \quad k = 1, 2, \dots, n,$$

где $f \in C$. Как известно, процесс $\{H_n(f,x)\}$ называется интерполяционным процессом Эрмита–Фейера. Л. Фейер [1] доказал, что для любой $f \in C$ выполняется равномерно в $[-1,1]$ соотношение

$$H_n(f,x) \to f(x), \quad n \to \infty. \tag{2}$$

Н. М. Крылов и И. Я. Штаерман [2] удлинили процесс Эрмита–Фейера. Они заменили полином $H_n(f,x)$ степени $2n-1$ на полином $p_n(f,x)$ степени $4n-1$, однозначно определяемый из условий $p_n(f,x_{kn}) = f(x_{kn})$, $p_n^{(j)}(f,x_{kn}) = 0$, $j = 1,2,3$, $k = 1,2,\dots,n$ где $p_n^{(j)}(f,x)$ — производная порядка j от $p_n(f,x)$. В [2] доказано, что для любой $f \in C$ выполняется равномерно в $[-1,1]$ соотношение

$$p_n(f,x) \to f(x), \quad n \to \infty. \tag{3}$$

Удлиним процесс $\{p_n(f,x)\}$ еще на один шаг. Обозначим через $p_{n,1}(f,x)$ многочлен степени $6n-1$, однозначно определяемый из условий $p_{n,1}(f,x_{kn}) = f(x_{kn})$, $p_{n,1}^{(j)}(f,x_{kn}) = d_{kn}^{(j)}$, $j=1,2,\dots,5$, $k=1,2,\dots,n$, где $\{x_{kn}\}$ — узлы (1) и $f \in C$. $\{d_{kn}^{(j)}\}$ наперед заданные вещественные числа. Можно доказать, что для любой $f \in C$ выполняется равномерно в $[-1,1]$ соотношение

$$p_{n,1}(f,x) \to f(x), \quad n \to \infty, \tag{4}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 41A05.
Key words and phrases. Approximation and expansions, convergence, divergence.

Akadémiai Kiadó, Budapest

если $|d_{kn}^{(j)}| \leqq A$, $j=1,2,3,4,5$, $k=1,2,\dots,n$, $n=1,2,\dots$, где A — абсолютная константа. Рассмотрим частный случай $K_n(f,x)$ полинома $p_{n,1}(f,x)$, когда $d_{kn}^{(j)}=0$, $j=1,2,3,4,5$, $k=\overline{1,n}$, $n=1,2,\dots$. Можно доказать, что при узлах (1) $K_n(f,x)$ имеет вид:

$$K_n(f,x)=\sum_{k=1}^{n} f(x_{kn})[\ell_k(x)]^6\Big(\sum_{j=1}^{n} a_{ks}(x-x_k)^{6-j}+1\Big),$$

$$\ell_K(x)=\frac{T_n(x)}{(x-x_K)T_n'(x_k)}, \qquad T_n(x)=\cos n\arccos x,$$

$$a_{k1}=\frac{1935}{16}\Big(\frac{x_k}{1-x_k^2}\Big)^5+\frac{9}{40}\frac{x_k^2}{(1-x_k^2)^4}+\frac{(1-n^2)(49x_k^3+3n^2-27)}{40(1-x_k^2)^3}+ \tag{5}$$

$$+\frac{x_k}{(1-x^2)^3}\Big[\Big(\frac{619n^2}{60}-\frac{2393}{40}\Big)(n^2-1)-\frac{(n^2-16)(n^2-4)}{120}\Big]+$$

$$+\frac{x^3}{(1-x_k^2)^4}\Big[\frac{2341}{48}(1-n^2)+\frac{77(4-n^2)}{16}-\frac{21}{40}(n^2-9)+\frac{n^2-16}{8}\Big],$$

$$a_{k2}=\frac{7}{(1-x_k^2)^2}\Big[\frac{13-10n^2}{16}\frac{x_k^2}{1-x_k^2}+\frac{(n^2-1)x^2}{4}+(n^2-1)\Big(\frac{n^2-1}{12}-\frac{n^2-9}{140}\Big)\Big],$$

$$a_{k3}=-\frac{x_k}{4(1-x_k^2)^2}\Big(\frac{x_k}{(1-x_k^2)}+33n^2-9\Big),$$

$$a_{k4}=\frac{1}{1-x_k^2}\Big(n^2-1+\frac{9x_k^2}{1-x_k^2}\Big),$$

$$a_{k5}=-\frac{3x_k}{1-x_k^2}.$$

В 1965 г. появилась статья автора [3], в которой изучался процесс $\{H_n(f,x)\}$ для узлов

$$x_0=1,\quad x_k=\cos(2k-1)\pi\,/\,2n,\quad k=1,2,\dots,n,\quad x_{n+1}=-1,\quad n=1,2,\dots, \tag{6}$$

полученных расширением узлов (1) добавлением в качестве узлов точек ±1. Оказалось, что этот процесс, построенный для $f(x)=|x|$, $f(x)=x^2$ расходится всюду в $(-1,1)$. При $f(x)=x$ этот процесс расходится во всех точках $x\neq0$ из $(-1,1)$.[1] Эти результаты неожиданные, если учесть результат (2) Л. Фейера. Итак, при расширении матрицы узлов (1) добавлением в качестве узлов точек ±1 для очень простых функций процесс $\{H_n(f,x)\}$ становится расходящимся всюду в $(-1,1)$. Будем это явление называть явлением U (от unexpected). Вполне естественно возник вопрос имеет ли место явление U

---
[1] См. [4], [5], [6].

для процесса Крылова–Штаермана? Л. Кук и Т. М. Миллс [7] доказали, что процесс Крылова–Штаермана, построенный при узлах (6) для $f(x)=(1-x^2)^3$ расходится при $x=0$. Недавно Р. Б. Саксена и С. Р. Мисра [8] доказали, что этот процесс расходится всюду в $(-1,1)$. Результат Кука–Миллса–Саксена–Мисра может быть усилен. Имеет место

Теорема 1. *Пусть четная функция $f(x)$ имеет ограниченную четвертую производную в $(-1,1)$ и пусть $f'(1)=f''(1)=0$. Тогда для того чтобы процесс $\{p_n(f,x)\}$ равномерно сходился в $[-1,1]$ необходимо и достаточно, чтобы $f^{(3)}(1)=0$. Если $f^{(3)}(1)\neq 0$, то процесс $\{p_n(f,x)\}$, построенный при узлах (6) для $f(x)$ расходится всюду в $(-1,1)$ [10].*

В связи с соотношением (4) возникает следующий вопрос: имеет ли место явление U для процесса $\{K_n(f,x)\}$? Этому вопросу и посвящена эта статья. Введем следующие полиномы: $C_{n,0}(f,x)$ многочлен степени $6n+1$, однозначно определяемый из условий $C_{n,0}(f,x_k)=$ $=f(x_k)$, $C^{(i)}_{n,0}(f,x_k)=0$, $i=1,2,3,4,5$, $k=1,2,\dots,n$, $C_{n,0}(f,\pm 1)=f(\pm 1)$, $n=1,2,\dots$. $C_{n,1}(f,x)$ — многочлен степени $6n+3$, однозначно определяемый из условий $C_{n,1}(f,x_k)=f(x_k)$, $C^{(i)}_{n,1}(f,x_k)=0$, $i=1,2,3,4,5$, $k=1,2,\dots,n$, $C_{n,1}(f,\pm 1)$, $C_{n,1}(f,\pm 1)=f(\pm 1)$, $C'_{n,1}(f,\pm 1)=0$. $C_{n,2}(f,x)$ — многочлен степени $6n+5$, однозначно определяемый из условий $C_{n,2}(f,x_k)=f(x_k)$, $C^{(i)}_{n,2}(f,x_k)=0$, $i=1,2,3,4,5$, $C_{n,2}(f,\pm 1)=f(\pm 1)$, $C^{(j)}_{n,2}(f,\pm 1)=0$, $j=1,2$, $k=1,2,\dots,n$, $n=1,2\dots$. Относительно процессов $\{C_{n,i}(f,x)\}_{n=0}^{\infty}$, $i=0,1,2$, справедливы следующие теоремы.

Теорема 2. *Интерполяционный процесс $\{C_{n,0}(f,x)\}$, построенный при узлах (1) для $f\in C$ удовлетворяет равномерно в $[-1,1]$ соотношению $C_{n,0}(f,x)\to f(x)$, $n\to\infty$.*

Доказательство. Из определения полиномов $K_n(f,x)$ и $C_{n,0}(f,x)$ следует, что

$$\begin{aligned}(7)\qquad C_{n,0}(f,x)-K_n(f,x)=\frac{T_n^6(x)}{2}[(f(1)-K_n(f,1)-f(-1)+\\ +K_n(f,-1))x+f(1)-K_n(f,1)+f(-1)-K_n(f,-1)].\end{aligned}$$

Отсюда и из (4) следует Теорема 2.

Теорема 3. *Пусть $f(x)$ имеет ограниченную вторую производную в $(-1,1)$. Для того чтобы в $[-1,1]$ выполнялось равномерно соотношение $C_{n,1}(f,x)\to f(x)$, $n\to\infty$, необходимо и достаточно, чтобы $f'(\pm 1)=0$. Если хоть одно из чисел $f'(\pm 1)$ отлично от нуля, то процесс $\{C_{n,1}(f,x)\}$ расходится всюду в $(-1,1)$.*

Доказательство. Из определения полиномов $C_{n,0}(f,x)$ и $C_{n,1}(f,x)$ выводим, что

$$C_{n,1}(f,x)-C_{n,0}(f,x)=T_n^6(x)(1-x^2)(ax+b),\ T_n(x)=\cos n\arccos x, \tag{7'}$$

где a и b находятся из системы уравнений:

$$[T_n^6(x)(1-x^2)(ax+b)]'_{x=\pm 1} = -C'_{n,0}(f,\pm 1).$$

Отсюда получим, что

$$a=\frac{1}{4}(C'_{n,0}(f,1)+C'_{n,0}(f,-1)),\quad b=\frac{1}{4}(C'_{n,0}(f,1)-C'_{n,0}(f,-1)). \tag{8}$$

Из (7) находим

$$\begin{gathered} C'_{n,0}(f,1)=\phi_n(K_n)+\varepsilon_{n,1},\ \ C'_{n,0}(f,-1)=\psi_n(K_n)+\varepsilon_{n,2}, \\ \varepsilon_{n,i}\to 0,\ \ n\to\infty,\ \ i=1,2, \end{gathered} \tag{9}$$

где

$$\phi_n(K_n)=K'_n(f,1)+6n^2(f(1)-K_n(f,1)) \text{ и}$$
$$\psi_n(K_n)=K'_n(f,-1)-6n^2(f(-1)-K_n(f,-1)).$$

В силу (8) и (9) имеем

$$a=\frac{1}{4}(\phi_n(K_n)+\psi_n(K_n))+\varepsilon_{n,3},$$
$$b=\frac{1}{4}(\phi_n(K_n)-\psi_n(K_n))+\varepsilon_{n,4},\quad \varepsilon_{n,j}\to 0,\ \ n\to\infty,\ \ j=3,4.$$

Отсюда и из (7) заключаем, что

$$\begin{aligned} &C_{n,1}(f,x)-C_{n,0}(f,x)= \\ &=\frac{T_n^6(x)(1-x^2)}{4}[(\phi_n(K_n)+\psi_n(K_n))x+\phi_n(K_n)-\psi_n(K_n)]+\varepsilon_{n,5}, \end{aligned} \tag{10}$$

$\varepsilon_{n,5}\to 0$, $n\to\infty$. Запишем полином $K_n(f,x)$ в виде

$$K_n(f,x)=\frac{1}{n^6}\sum_{k=1}^{n} f(x_k)A_k(x)B_k(x), \tag{11}$$

где

$$\begin{gathered} A_k(x)=\left(\frac{T_n(x)}{x-x_k}\right)^6, \\ B_k(x)=(1-x_k^2)^3\Big(\sum_{s=1}^{5} a_{ks}(x-x_k)^{6-s}+1\Big). \end{gathered} \tag{12}$$

Очевидно, что

$$A'_k(1) = \Big(\frac{6n^2}{(1-x_k)^6} - \frac{6}{(1-x_k)^7}\Big). \tag{13}$$

Ради простоты вычислений положим, что $f(-1) = f(1) = 0$.[1] Вычислим $\phi_n(K_n)$. Согласно (13) имеем

$$\phi_n(A_k B_k) = \frac{B'_k(1)}{(1-x_k)^6} - \frac{6B_k(1)}{(1-x_k)^7}. \tag{14}$$

Из формулы (11) получаем

$$\phi_n(K_n) = \frac{1}{n^6}\sum_{k=1}^{n} f(x_k)\phi_n(A_k B_k). \tag{15}$$

По формуле Тейлора

$$f(x_k) = f'(1)(x_k - 1) + \frac{f''(c_k)}{2}(x_k-1)^2, \quad x_k < c_k < 1. \tag{16}$$

из (14), (15), и (16) выводим

$$\phi_n(K_n) = \tau_{1,n} + \tau_{2,n}, \tag{17}$$

где

$$\tau_{1,n} = \frac{f'(1)}{n^6}\sum_{k=1}^{n}\Big(\frac{6B_k(1)}{(1-x_k)^6} - \frac{B'_k(1)}{(1-x_k)^5}\Big),$$

$$\tau_{2,n} = \frac{1}{2n^6}\sum_{n=1}^{n} f''(c_k)\left(\frac{B'_k(1)}{(1-x_k)^4} - \frac{6B_k(1)}{(1-x_k)^5}\right).$$

Лемма 1. *Выполняется равенство* $\lim\limits_{n\to\infty}\tau_{2,n}^{(1)} = 0$, *где*

$$\tau_{2,n}^{(1)} = -\frac{3}{n^6}\sum_{k=1}^{n} f''(c_k)\frac{B_k(1)}{(1-x_k)^5}.$$

Доказательство. Ясно, что

$$\begin{aligned} |\tau_{2,n}^{(1)}| \leqq \frac{3\|f''\|}{n^6}\sum_{k=1}^{n}(1-x_k^2)^3\Big(|a_{k1}| + \frac{|a_{k2}|}{1-x_k} + \frac{|a_{k3}|}{(1-x_k)^2} + \\ + \frac{|a_{k4}|}{(1-x_k)^3} + \frac{|a_{k5}|}{(1-x_k)^4} + \frac{1}{(1-x_k)^5}\Big) \equiv \sum_{i=1}^{6}\sigma_i, \end{aligned} \tag{18}$$

[1] Это ограничение не является существенным.

где $\|f''\| = \max\limits_{-1\leqq x\leqq 1} |f''(x)|$. Все слагаемые из правой части (18) оцениваются одинаковым образом. Поэтому рассмотрим лишь σ_1 и σ_6. Согласно формулам (5) имеем, что

$$\sigma_1 \leqq \frac{3\|f''\|}{6n^6} \sum_{k=1}^{n} (1-x_k^2)^3 \left(\frac{1935}{16(1-x_k^2)^5} + \frac{9}{40(1-x_k^2)^4} + \frac{O(n^2)}{(1-x_k^2)^3} + \right.$$
$$\left. + \frac{O(n^4)}{(1-x_k^2)^3} + \frac{O(n^2)}{(1-x_k^2)^4} \right) \equiv \sum_{i=1}^{5} \delta_i.$$

Все слагаемые рассматриваются одинаково. Поэтому ограничимся рассмотрением δ_1, δ_4 и δ_5. Имеем

$$\delta_1 \leqq \frac{C}{n^6} \sum_{k=1}^{n} \frac{1}{(1-x_k^2)^2}. \tag{19}$$

Известно, что [9]

$$\sum_{k=1}^{n} \frac{1}{(1-x_k^2)^2} = \frac{n^4+2n^2}{3}. \tag{20}$$

Отсюда и из (19) выводим, что $\delta_1 = O\left(\frac{1}{n^2}\right)$. Очевидно, что

$$\delta_4 = \frac{1}{n^6} \sum_{k=1}^{n} O(n^4) = O(\frac{1}{n}).$$

Имеем:

$$\delta_5 \leqq \frac{c}{n^6} \sum_{k=1}^{n} \frac{O(n^2)}{1-x_k^2}.$$

Воспользуемся теперь тождеством [9]

$$\sum_{k=1}^{n} 1 / (1-x_k^2) = n^2. \tag{21}$$

Тогда получим, что $\delta_5 = O\left(\frac{1}{n^2}\right)$. Итак, доказано, что $\lim\limits_{n\to\infty} \sigma_1 = 0$. Рассмотрим σ_6. Очевидно, что

$$\sigma_6 \leqq \frac{24\|f''\|}{n^6} \sum_{k=1} \frac{1}{(1-x_k)^2},$$

но [9]

$$\sum_{k=1}^{n} 1/(1-x_k)^2 = (2n^4+n^2)/3. \tag{22}$$

Следовательно, $\sigma_6 = O\left(\frac{1}{n^2}\right)$. Поэтому $\lim\limits_{n\to\infty} \tau_{2,n}^{(1)} = 0$.

Лемма 2. *Выполняется равенство* $\lim\limits_{n\to\infty} \tau_{2,n}^{(2)} = 0$, *где*

$$\tau_{2,n}^{(2)} = \frac{1}{2n^6} \sum_{k=1}^{n} f''(c_k) \frac{B_k'(1)}{(1-x_k)^4}.$$

Доказательство. Очевидно, что

$$|\tau_{2,n}^{(2)}| \leqq \frac{c}{n^6} \sum_{k=1}^{n} (1-x_k^2)^3 \left(5|a_{k1}| + \frac{4|a_{k2}|}{1-x_k} + \frac{3|a_{k3}|}{(1-x_k)^2} + \frac{2|a_{k2}|}{(1-x_k)^2} + \frac{|a_{k5}|}{(1-x_k)^4}\right).$$

Доказательство остальной части Леммы 2 не отличается от доказательства Леммы 1.

Следствие 1. *Выполняется равенство* $\lim\limits_{n\to\infty} \tau_{2,n} = 0$.

Действительно, $\tau_{2,n} = \tau_{2,n}^{(1)} + \tau_{2,n}^{(2)}$. Стало быть, следствие вытекает из Лемм 1 и 2. Вычислим теперь $\lim\limits_{n\to\infty} \tau_{1,n}$.

Лемма 3. *Справедливо равенство*

$$\lim_{n\to\infty} \frac{6f'(1)}{n^6} \sum_{k=1}^{n} \frac{B_k(1)}{(1-x_k)^6} = \frac{11067}{40} f'(1). \tag{23}$$

Доказательство. Согласно (12) имеем, что

$$\begin{aligned} \frac{B_k(1)}{(1-x_k)^6} = (1-x_k^2)^3 \Big(&\frac{a_{k1}}{1-x_k} + \frac{a_{k2}}{(1-x_k)^2} + \frac{a_{k3}}{(1-x_k)^3} + \\ &+ \frac{a_{k4}}{(1-x_k)^4} + \frac{a_{k5}}{(1-x_k)^5} + \frac{1}{(1-x_k)^6} \Big). \end{aligned} \tag{24}$$

Подставляя в правую часть значения коэффициентов $\{a_{ki}\}_{i=1}^{5}$ согласно формулам (5) и производя соответствующие вычисления, получим (23). В ходе вычислений использовались тождества (20), (21), (22) и тождество [9]

$$\sum_{k=1}^{n} \frac{1}{\sin^6 \theta_k/2} = \frac{8n^2}{3}(8n^4+5n^2+2), \qquad \theta_k = (2k-1)\pi/2n. \tag{25}$$

Аналогичным образом доказывается

Лемма 4. *Имеет место равенство*

$$\lim_{n\to\infty}\frac{f'(1)}{n^6}\sum_{k=1}^{n}\frac{B_k'(1)}{(1-x_k)^5}=\frac{57227}{240}f'(1).$$

Следствие 2. *Имеет место равенство* $\lim\limits_{n\to\infty}\tau_{1,n}=\frac{1835}{48}f'(1)$.

Действительно, Следствие 2 вытекает из Лемм 3 и 4. Из Следствий 1 и 2 следует.

Следствие 3. *Справедливо равенство*

$$\lim_{n\to\infty}\phi_n(K_n)=\frac{1835}{48}f'(1). \tag{26}$$

Действительно, равенство (26) вытекает из (17), Следствий 1 и 2. Аналогичным образом выводится, что

$$\lim_{n\to\infty}\psi_n(K_n)=\frac{1835}{48}f'(-1). \tag{27}$$

Из равенств (10), (26), (27) вытекает, что

$$\begin{gathered}C_{n,1}(f,x)-C_{n,0}(f,x)=\frac{1835}{192}T_n^6(x)(1-x^2)[(f'(1)+\\+f'(-1))x+f'(1)-f'(-1)]+\varepsilon_{n,6},\quad n\to\infty,\quad \varepsilon_{n,6}\to 0.\end{gathered} \tag{28}$$

Отсюда и из равенства (7') следует, что процесс сходится равномерно в $[-1,1]$ тогда и только тогда, когда $f'(1)=f'(-1)=0$. Если в какой то точке $(f'(1)+f'(-1))x+f'(1)-f'(-1)\neq 0$, $x\in(-1,1)$, то в этой точке процесс $\{C_{n,1}(f,x)\}$ расходится, ибо известно [5], что для любой $x\in(-1,1)$ можно найти такую последовательность натуральных чисел $\{n_k\}$, $n_1<n_2<\dots$, что $\lim\limits_{k\to\infty}T_{n_k}^2(x)=1$. Отсюда и из (28) следует, что процесс $\{C_{n,1}(f,x)\}$ в точке x расходится. При этом учитывается Теорема 2, согласно которой процесс $\{C_{n,0}(f,x)\}$ сходится равномерно в $[-1,1]$.

Переходим к исследованию процесса $\{C_{n,2}(f,x)\}$. Для простоты вычислений положим, что функция $f(x)$ — четная. Это ограничение не является существенным.

Теорема 4. *Пусть $f(x)$ имеет ограниченную третью производную в $(-1,1)$. Если $f(x)$ — четная функция и $f'(1)=0$, то для того чтобы в $[-1,1]$ выполнялось равномерно соотношение $C_{n,2}(f,x)\to f(x)$, $n\to\infty$,* *необходимо и достаточно, чтобы* $f''(1)=0$. *Если* $f''(1)\neq 0$, *то процесс* $\{C_{n,2}(f,x)\}$ *расходится всюду в* $(-1,1)$.

Доказательство. Из определения полиномов $C_{n,2}(f,x)$ и $C_{n,1}(f,x)$ и четности $f(x)$ следует, что

$$C_{n,2}(f,x)-C_{n,1}(f,x)=T_n^6(x)(1-x^2)^2d.$$

Учтем, что при $x=1$, $C''_{n,2}(f,1)=0$. Поэтому $-C''_{n,1}=[T_n^6(x)(1-x^2)]''_{x=1}d$. Стало быть, $d=-\frac{C''_{n,1}(f,1)}{8}$. Итак,

$$C_{n,2}(f,x)-C_{n,1}(f,x)=\frac{-T_n^6(x)(1-x^2)^2}{8}C''_{n,1}(f,1). \tag{29}$$

Вычислим функционал $Q_n(f)=C''_{n,1}(f,1)$. Из (7) имеем,[2] что $C'_{n,0}(f,1)=K'_n-6n^2K_n$. Далее из (7) выводим, что $C''_{n,0}(f,1)=K''_n-K_n(T_n^6)''_{x=1}$. Так как $(T_n^6)''_{x=1}=2n^2(16n^2-1)$, то получим, что $C''_{n,0}(f,1)=K''_n-2K_nn^2(16n^2-1)$.

Из равенства

$$C_{n,1}(f,x)-C_{n,0}(f,x)=\frac{1}{2}(1-x^2)T_n^6(x)C'_{n,0}(f,1)$$

выводим

$$\begin{gathered}Q_n(K_n)=K''_n-(12n^2+1)K'_n+8n^2(5n^2+1)K_n,\\ K_n^{(j)}=(K_n(f,x))^{(j)}_{x=1},\quad j=0,1,2.\end{gathered} \tag{30}$$

Для дальнейшего нужно вычислить $Q_n(A_kB_k)$. В силу (30) имеем

$$\begin{aligned}Q_n(A_kB_k)={}&A''_k(1)B_k(1)+2A'_k(1)B'_k(1)+A_k(1)B''_k(1)-\\ &-(12n^2+1)(A'_k(1)B_k(1)+A_k(1)B'_k(1))+8n^2(5n^2+1)A_k(1)B_k(1).\end{aligned} \tag{31}$$

Заметим, что

$$\begin{gathered}A''_k(1)=6\Big[\frac{5(n^2(1-x_k)-1)^2}{(1-x_k)^8}+\frac{n^2(n^2-1)}{3(1-x_k)^6}-\frac{2(n^2(1-x_k)-1)}{(1-x_k)^8}\Big],\\ B_k(1)=(1-x_k^2)^3\Big(\sum_{s=1}^{5}a_{ks}(1-x_k)^{6-s}+1\Big),\\ B'_k(1)=(1-x_k^2)^3\sum_{s=1}^{5}(6-s)(1-x_k)^{5-s},\\ B''_k(1)=(1-x_k^2)^3\sum_{s=1}^{4}(6-s)(5-s)(1-x_k)^{4-s}.\end{gathered} \tag{32}$$

---

[2] Считаем, что $f(1)=0$.

Лемма 5. *Справедливо равенство*

$$Q_n(A_kB_k)=(1-x_k^2)^3\Big(\sum_{s=1}^{5}a_{ks}\gamma_{ks}+\gamma_{k6}\Big), \tag{33}$$

где

$$\begin{gathered}\gamma_{k1}=\frac{2}{(1-x)^3}+\frac{11}{(1-x)^2},\quad \gamma_{k2}=\frac{24}{(1-x)^4}-\frac{2}{(1-x)^3},\\ \gamma_{k3}=\frac{12}{(1-x)^5}+\frac{3}{(1-x)^4},\quad \gamma_{k4}=\frac{20}{(1-x)^6}+\frac{4}{(1-x)^5},\\ \gamma_{k5}=\frac{30}{(1-x)^2}-\frac{5}{(1-x)^6},\quad \gamma_{k6}=\frac{42}{(1-x)^8}+\frac{6}{(1-x)^7},\quad x=x_k.\end{gathered} \tag{33'}$$

При этом коэффициенты $\{a_{ks}\}_{s=1}^5$ *задаются равенствами* (5).

Доказательство. Подставляем в правую часть равенства (31) значения $\{A_k^{(j)}(1)\}$, $\{B_k^{(j)}(1)\}$ согласно формулам (13), (32) и собираем сперва члены содержащие a_{k1}, затем собираем члены содержащие a_{k2} и т.д. После элементарных преобразований получим равенство (33).

Лемма 6. *Выполняются равенства*

$$\begin{gathered}\lim_{n\to\infty}\frac{1}{n^6}\sum_{k=1}^{n}(1-x_k)^3(1-x_k^2)^3|\gamma_{ki}a_{ki}|=\\ =\lim_{n\to\infty}\frac{1}{n^6}\sum_{k=1}^{n}(1-x_k)^3|\gamma_{k6}|(1-x_k^2)^3=0,\quad i=1,2,3,4,5.\end{gathered} \tag{34}$$

Доказательство. Ограничимся доказательством лишь первого и последнего из этих равенств. Имеем согласно (5)

$$|a_{k1}|\leqq\frac{C}{(1-x_k^2)^5}+\frac{O(n^2)}{(1-x_k^2)^4}+\frac{O(n^4)}{(1-x_k^2)^3}.$$

Стало быть,

$$(1-x_k)^3(1-x_k^2)^3|\gamma_{k1}a_{k1}|\leqq\sum_{i=1}^{3}\beta_i^{(k)}$$

где

$$\begin{gathered}\beta_1^{(k)}=\frac{C(1-x_k)\gamma_{k1}}{(1+x_k)^2},\quad \beta_2^{(k)}=\frac{O(n^2)|\gamma_{k1}|}{1+x_k},\\ \beta_3^{(k)}=O(n^4)(1-x_k)^3|\gamma_{k1}|,\\ \gamma_{k1}=\frac{2}{(1-x_k)^3}+\frac{11}{(1-x_k)^2}.\end{gathered} \tag{35}$$

Ограничимся лишь оценкой $\sum\limits_{k=1}^{n} \beta_1^{(k)}$ и $\sum\limits_{k=1}^{n} \beta_3^{(k)}$. Ясно из (35), что

$$n^{-6} \sum_{k=1}^{n} \beta_1^{(k)} |\gamma_{k1}| \leqq \sum_{k=1}^{n} 24 \,/\, (1 - x_k^2)^2 n^{-6}. \tag{36}$$

Из (20) и (36) выводим, что $n^{-6} \sum\limits_{k=1}^{n} \beta_i^{(k)} |\gamma_{k1}| \to 0$, $n \to \infty$. Из (35) видно, что

$$n^{-6} \sum_{k=1}^{n} \beta_3^{(k)} = O\Big(\frac{1}{n^2}\Big) \sum_{k=1}^{n} (2 + 11(1 - x_k)). \tag{37}$$

Но

$$\sum_{k=1}^{n} (2 + 11(1 - x_k)) = 13n. \tag{38}$$

Поэтому из (37) заключаем, что $n^{-6} \sum\limits_{k=1}^{n} \beta_3^{(k)} = O(\frac{1}{n})$. Итак, первое из равенств (34) доказано. Докажем последнее из равенств (34). В силу (33′) имеем, что

$$\begin{aligned} &\lim_{n \to \infty} n^{-6} \sum_{k=1}^{n} (1 - x_k)^3 (1 - x_k^2)^3 |\gamma_{k6}| \leqq \\ &\leqq \lim_{n \to \infty} C n^{-6} \sum_{k=1}^{n} 1/(1 - x_k). \end{aligned} \tag{39}$$

Воспользуемся тождеством $\sum\limits_{k=1}^{n} 1/(1 - x_k) = n^2$. Тогда из (39) имеем

$$\lim_{n \to \infty} \frac{1}{n^6} \sum_{k=1}^{n} (1 - x_k)^3 (1 - x_k^2)^3 |\gamma_{k6}| = 0.$$

Следствие 4. *Имеет место равенство*

$$\lim_{n \to \infty} \frac{1}{n^6} \sum_{k=1}^{n} (1 - x_k)^3 |Q_n(A_k B_k)| = 0. \tag{40}$$

Лемма 7. *Имеет место равенство*

$$\lim_{n \to \infty} \sum_{k=1}^{n} \frac{1}{4n^6} (1 - x_k)^2 Q_n(A_k B_k) = 256 \frac{29}{32}. \tag{41}$$

Доказательство. Подставляем в левую часть (41) значение $Q_n(A_k B_k)$ согласно Лемме 5, а затем в полученном равенстве подставляем значения коэффициентов $\{a_{ki}\}_{i=1}^{5}$ согласно (5). После элементарных преобразований и предельного перехода, когда $n \to \infty$, получим (41). При этом нужно будет воспользоваться тождествами (21), (22), (25).

Доказательство Теоремы 4. Без ограничения общности можно считать, что $f(1) = 0$. По формуле Тейлора с учетом, что $f(1) = f'(1) = 0$ имеем

$$f(x_k) = \frac{f''(1)}{2}(x_k - 1)^2 + \frac{f^{(3)}(c_k)}{6}(x_k - 1)^3, \quad x_k < c_k < 1. \tag{42}$$

Так как по условию Теоремы 4 $f'(1) = 0$, то согласно Теореме 3 процесс $\{C_{n,1}(f, x)\}$ сходится равномерно в $[-1, 1]$. Поэтому согласно (29) нужно доказать, что $C_n''(f, 1) \to 0$, $n \to \infty$, где $C_n''(f, 1)$ определяется по формуле (30). Из (11), (42) следует, что

$$Q_n(K_n) = \alpha_{n,1} + \alpha_{n,2}, \tag{43}$$

где

$$\alpha_{n,1} = \frac{f''(1)}{2n^6} \sum_{k=1}^{n} |x_k - 1|^2 Q_n(A_k B_k),$$

$$\alpha_{n,2} = \frac{1}{6n^6} \sum_{k=1}^{n} f''(c_k)(x_k - 1)^3 Q_n(A_k B_k).$$

Очевидно, что $|\alpha_{n,2}| \leqq \frac{\|f''\|}{6n^6} \sum\limits_{k=1}^{n} |x_k - 1|^3 |Q_n(A_k B_k)|$. Поэтому из следствия 4 выводим, что $\lim\limits_{n\to\infty} \alpha_{n,2} = 0$. Согласно Лемме 7 $\lim\limits_{n\to\infty} \alpha_{n,1} = 256\frac{29}{32} f''(1)$. Стало быть, из (43) заключаем, что $\lim\limits_{n\to\infty} C_n''(f1) = 256\frac{29}{32} f''(1)$. Заключительная часть доказательства Теоремы 4 не отличается от заключительной части доказательства Теоремы 3.

ЛИТЕРАТУРА

[1] Fejér, L., Über Interpolation, *Gött. Nachr.* (1916), 66–91. *Jahrbuch. Fortschr. Math.* **46**, 419.

[2] Крылов, Н. М. и Штаерман, И. Я., Sur quelques formules d'interpolation convergentes pour toute fonction continue, *Zapiski fiz-mat. otd. AN USSR* **1** (1922), 12–13; *Kiev Bull. Ac. Sc.* **1** (1922/23), 13–16. *Jahrbuch Fortschr. Math.* **49**, 207.

[3] Берман, Д. Л., К теории интерполяции, *Dokl. Akad. Nauk SSSR* **163** (1965), 551–554. *MR* **33** #4530

[4] Берман, Д. Л., Исследование интерполяционного процесса Эрмита–Фейера, *Dokl. Akad. Nauk SSSR* **187** (1969), 241–244. *MR* **40** #3127

[5] Берман, Д. Л., Об одном всюду расходящемся интерполяционном процессе Эрмита–Фейера, *Izv. Vysš. Učebn. Zaved. Matematika* **1970**, no. 1 (92), 3–8. *MR* **41** #7343

[6] Берман, Д. Л., Всюду расходящийся расширенный интерполяционный процесс Эрмита–Фейера, *Izv. Vysš. Učebn. Zaved. Matematika* **1975**, no. 9 (160), 84–87. *MR* **55** #10904

[7] Cook, W. L. and Mills, T. M., On Berman's phenomenon in interpolation theory, *Bull. Austral. Math. Soc.* **12** (1975), 457–465. *MR* **51** #13520

[8] Saxena, R. B. and Misra, S. R., An everywhere divergent Hermite–Fejér type interpolation process of higher order, *Acta Math. Hungar.* **48** (1986), 79–85. *MR* **87k**: 41004

[9] Берман, Д. Л., Решение одной экстремальной задачи теории операторов и аналогии тождества М. Рисса (Solution of an extremal problem of operator theory and analogues of the M. Riesz identity), *Uspekhi Mat. Nauk* **37** (1982), 187–188. *MR* **84d**: 42001

[10] Берман, Д. Л., Обобщение одной теоремы Р. Б. Саксена и С. Р. Мисра, *Acta Math. Hungar.* **57** (1991), 29–31.

(Поступило 25-ого апреля 1988 г.)

БАССЕЙНАЯ 68, КВ. 90
192238 ЛЕНИНГРАД
SOVIET UNION

Current address:

1357 W. ESTES # Q2
CHICAGO, IL 60626
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 65–69

THE VARIETY OF DOUBLE HEYTING ALGEBRAS IS CONGRUENCE UNIFORM

M. E. ADAMS and R. BEAZER

1. Introduction

A congruence relation on an algebra is said to be *uniform* if all its congruence classes have the same cardinality and an algebra is called *congruence uniform* if all its congruences are uniform. It is particularly advantageous to know that an algebra having a lattice reduct is congruence uniform because such algebras can often be represented pictorially and congruences on them can be visualized as suitable partitions of their diagram. Boolean lattices are, of course, congruence uniform and countable congruence uniform distributive lattices are characterized in [1]. Congruence uniform algebras in the varieties of pseudocomplemented semilattices, lattices with pseudocomplementation (alias, p-algebras), bounded relatively complemented lattices (alias, Heyting algebras), and double p-algebras are described in [4]. In that paper it is shown that every finite double Heyting algebra is congruence uniform and the problem: "Which infinite double Heyting algebras are congruence uniform?" is posed. The purpose of this note is to answer this question in the best possible way. We show that every such algebra is congruence uniform; in other words; the variety of double Heyting algebras is congruence uniform.

2. Preliminaries

A *double Heyting algebra* is an algebra $(L; \vee, \wedge, *, +, 0, 1)$ of type $(2,2,2,2,0,0)$ in which $(L; \vee, \wedge, 0, 1)$ is a bounded (distributive) lattice, the binary operation $*$ is given by $a \wedge x \leqq b$ iff $x \leqq a * b$, and the binary operation $+$ is characterized dually.

For any x in a double Heyting algebra L the element $x^* = x * 0$ ($x^+ = x + 1$) is the *pseudocomplement* (*dual pseudocomplement*) of x in L and so any double Heyting algebra L gives rise to a *distributive double p-algebra*

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 06D20; Secondary 06D15.

Key words and phrases. Double Heyting algebras, congruence relations.

Akadémiai Kiadó, Budapest

$(L; \vee, \wedge, ^*, ^+, 0, 1)$ of type $(2,2,1,1,0,0)$. For any x in an arbitrary distributive double p-algebra, we have

$$x^{++} \leqq x \leqq x^{**} \text{ and } x^{+*} \leqq x \leqq x^{*+}.$$

The *centre of* L, denoted $\mathrm{Cen}(L)$, is given by

$$\mathrm{Cen}(L) = \{x \in L : x = x^{+*}\} = \{x \in L : x = x^{*+}\}.$$

Elements $x^{n(+*)}$ are defined inductively by

$$x^{0(+*)} = x \text{ and } x^{(n+1)(+*)} = x^{n(+*)+*}.$$

If, for all $x \in L$, $x^{(n+1)(+*)} = x^{n(+*)}$, for some $n < \omega$, then L is said to have *finite range;* equivalently, if, for all $x \in L$, $x^{n(+*)} \in \mathrm{Cen}(L)$, for some $n < \omega$. A *normal filter* in L is a lattice filter F having the property that $x \in L$ implies $x^{+*} \in L$. The lattice of all normal filters of L is denoted $\mathcal{F}^n(L)$.

By a *congruence* on a double Heyting algebra (or double p-algebra) L we mean a lattice congruence preserving the operations $*$ and $+$ (or * and $^+$) and the congruence lattice of L is denoted $\mathrm{Con}(L)$. If $F \in \mathcal{F}^n(L)$ and $\theta \in \mathrm{Con}(L)$, then it is known (see [8], for example) that $\Theta_{\mathrm{lat}}(F)$, the smallest lattice congruence on L collapsing F, is a congruence on L and $\theta = \Theta_{\mathrm{lat}}([1]\theta)$ so that $\mathcal{F}^n(L) \cong \mathrm{Con}(L)$ under the mapping $F \mapsto \Theta_{\mathrm{lat}}(F)$ (see also [6]).

For the standard rules of computation in (double) Heyting algebras and (double) p-algebras we refer the reader to [2], [3], and [7]. All other unexplained notation and terminology can be found in [5].

3. Congruence uniformity

We begin with some elementary results about double Heyting algebras which are pertinent to the proof of our main theorem.

LEMMA 1. *If L is a double Heyting algebra then the identities:*

(i) $a = ((b \vee a^+) \wedge a) \vee (b \vee a^+)^+$,

(ii) $a = (a \vee b) \wedge (b * a)$,

(iii) $a \vee b = a \vee (b \wedge (b * a)^+)$

hold in L.

PROOF. (i) $((b \vee a^+) \wedge a) \vee (b \vee a^+)^+ = a \vee (b \vee a^+)^+ = a$, since $(b \vee a^+)^+ \leqq a^{++} \leqq a$.

(ii) $(a \vee b) \wedge (b * a) = (a \wedge (b * a)) \vee (b \wedge (b * a))$

$= a \vee (b \wedge (b * a))$, since $b * a \geqq a$,

$= a$, since $b \wedge (b * a) \leqq a$.

(iii) Obviously, $a \vee b \geqq a \vee (b \wedge (b * a)^+)$ and for the reverse inclusion it is enough to show that $a \vee (b * a)^+ \geqq b$. However, $a \geqq b \wedge (b * a)$ and so $a \vee (b * a)^+ \geqq (b \wedge (b * a)) \vee (b * a)^+ = b \vee (b * a)^+ \geqq b$. □

Henceforth, L will denote a double Heyting algebra, θ will be a congruence on L and $\theta_{\mathrm{lat}}(x, y)$ will denote the principal lattice congruence collapsing the pair $x, y \in L$.

LEMMA 2. *If* $[1]\theta = [a)$, *for some* $a \in L$, *then* $a \in \mathrm{Cen}(L)$ *and* $\theta = \theta_{\mathrm{lat}}(a, 1)$.

PROOF. If $[1]\theta = [a)$, then $a^{+*} \equiv 1(\theta)$. Therefore, $a^{+*} \geqq a$ and it follows that $a \in \mathrm{Cen}(L)$. Finally, $\theta = \theta_{\mathrm{lat}}([1]\theta) = \theta_{\mathrm{lat}}(a, 1)$. □

COROLLARY 3. *If* $[1]\theta = [a)$, *for some* $a \in L$, *and* $b \in L$ *is arbitrary, then* $[b]\theta$ *and* $[1]\theta$ *are isomorphic lattices.*

PROOF. By Lemma 2, $a \in \mathrm{Cen}(L)$. Clearly, $x \mapsto a \vee x$ is a lattice homomorphism from $[b]\theta$ to $[a)$. Furthermore, if $a \vee x = a \vee y$ and $x, y \in [b]\theta$, then $x \equiv y(\theta_{\mathrm{lat}}(a, 1))$, by Lemma 2, so that $x \wedge a = y \wedge a$, by the well-known description of principal congruences on distributive lattices, and therefore $x = y$. Thus, the mapping is injective. To see that the mapping is surjective, let $y \in [a)$ and $x = (b \vee a') \wedge y$ where a' is the complement of a. Then $a \vee x = a \vee y = y$ and $x \wedge a = b \wedge a$, so that $x \equiv b(\theta_{\mathrm{lat}}(a, 1))$. Therefore, $x \in [b]\theta$ by Lemma 2. □

The following simple observation will be used repeatedly. Let κ be an infinite cardinal T, L be a lattice, $C = \{c_i : i < \kappa\} \subseteqq L$, $D = \{d_i : i < \kappa\} \subseteqq L$, and $B = \{b_i : i < \kappa\} \subseteqq L$, with $b_i = b_j$ iff $i = j$, whenever $i, j < \kappa$. If either $b_i = c_i \vee d_i$, for all $i < \kappa$, or $b_i = c_i \wedge d_i$ for all $i < \kappa$, then at least one of the sets C or D has cardinality κ.

LEMMA 4. *If* $[a]\theta$ *is infinite, then* $|[1]\theta| \geqq |[a]\theta|$.

PROOF. We begin by showing that either $|[0]\theta| \geqq |[a]\theta|$ or $|[1]\theta| \geqq |[a]\theta|$. Suppose $|[a]\theta| = \kappa$ which, by hypothesis, is an infinite cardinal and write $[a]\theta = \{b_i : i < \kappa\}$ where $b_i = b_j$ iff $i = j$, whenever $i, j < \kappa$. By Lemma 1 (ii), $b_i = (a \vee b_i) \wedge (a * b_i)$, for any $i < \kappa$, and so $\{a \vee b_i : i < \kappa\}$ or $\{a * b_i : i < \kappa\}$ has cardinality κ. In the event that $\{a \vee b_i : i < \kappa\}$ has cardinality κ, then so does $\{b_i \wedge (b_i * a)^+ : i < \kappa\}$, by Lemma 1 (iii). But $b_i \wedge (b_i * a)^+ \equiv a \wedge (a * a)^+(\theta)$ and $a \wedge (a * a)^+ = a \wedge 1^+ = 0$ so that $b_i \wedge (b_i * a)^+ \in [0]\theta$. Thus, $|[0]\theta| \geqq \kappa = |[a]\theta|$. In the event that $\{a * b_i : i < \kappa\}$ has cardinality κ, we have $|[1]\theta| \geqq \kappa = |[a]\theta|$, since $a * b_i \in [1]\theta$.

To complete the proof we need only to show that if $|[0]\theta| \geqq |[a]\theta|$ then $|[1]\theta| \geqq |[0]\theta|$. Since $|[0]\theta| = \kappa'$, where $\kappa' \geqq \kappa$, we can write $[0]\theta = \{b_i : i < \kappa'\}$ and assume that, for $i, j < \kappa'$, $b_i = b_j$ iff $i = j$. Now, $b_i = b_i^{**} \wedge (b_i \vee b_i^*)$ and so $\{b_i^{**} : i < \kappa'\}$ or $\{b_i \vee b_i^* : i < \kappa'\}$ has cardinality κ'. In the event that $\{b_i^{**} : i < \kappa'\}$ has cardinality κ', then so does $\{b_i^* : i < \kappa'\}$ and, therefore, $|[1]\theta| \geqq \kappa' = |[0]\theta|$, since $b_i^* \in [1]\theta$. In the event that $\{b_i \vee b_i^* : i < \kappa'\}$ has cardinality κ' we can again conclude that $|[1]\theta| \geqq |[0]\theta|$, since $b_i \vee b_i^* \in [1]\theta$. In any case, $|[1]\theta| \geqq |[0]\theta|$ and the proof is complete. □

LEMMA 5. *If* $[1]\theta$ *is infinite, then* $|[b]\theta| \geqq |[1]\theta|$, *for any* $b \in L$.

PROOF. Suppose $|[1]\theta| = \kappa$ where, by hypothesis, κ is an infinite cardinal and $[1]\theta = \{b_i : i < \kappa\}$ with $b_i = b_j$ iff $i = j$, whenever $i, j < \kappa$. By Lemma 1 (i), $b_i = ((b \vee b_i^+) \wedge b_i) \vee (b \vee b_i^+)^+$, for any $i < \kappa$, and so $\{(b \vee b_i^+) \wedge b_i : i < \kappa\}$ or

$\{(b \vee b_i^+)^+ : i < \kappa\}$ has cardinality κ. In the event that $\{(b \vee b_i^+) \wedge b_i : i < \kappa\}$ has cardinality κ, we have $|[b]\theta| \geqq \kappa = |[1]\theta|$, since $(b \vee b_i^+) \wedge b_i \in [b]\theta$. In the event that $\{(b \vee b_i^+)^+ : i < \kappa\}$ has cardinality κ, then so does $\{b \vee b_i^+ : i < \kappa\}$ and again $|[b]\theta| \geqq |[1]\theta|$, since $b \vee b_i^+ \in [b]\theta$. ☐

Clearly, if every class of θ is finite then θ is uniform, by Corollary 3. On the other hand, if some class of θ is infinite then θ is uniform, by the conjunction of Lemmas 4 and 5. Thus, we have:

THEOREM 6. *Every double Heyting algebra is congruence uniform.* ☐

REMARK. In [4], it was shown that a double p-algebra L is congruence uniform iff it is congruence regular; by which is meant that any congruence on L is uniquely determined by any one of its classes. The bulk of the proof entails showing that the congruence regularity implies congruence uniformity and is heavily dependant on the validity of the, so called, determination principle (namely, $x^* = y^*$ and $x^+ = y^+$ imply $x = y$) which is known to characterize congruence regular double p-algebras; see [6] and [9]. However, T. Katriňák [6] has shown that any regular double p-algebra L is, in fact, a double Heyting algebra and that a binary relation on L is a double Heyting algebra congruence iff it is a double p-algebra congruence. Consequently, Theorem 6 yields an entirely new proof of the main result of [4].

4. Strong congruence uniformity

Motivated by Corollary 3, we will say that a congruence θ on a double Heyting algebra L is *strongly uniform* if all its classes are isomorphic lattices and call L *strongly congruence uniform* if all its congruences are strongly uniform. Our next objective is to characterize strongly congruence uniform double Heyting algebras.

THEOREM 7. *For a double Heyting algebra L, the following are equivalent:*

(i) *L is strongly congruence uniform;*
(ii) *every class of any congruence on L has a least element;*
(iii) $\mathrm{Con}(L)$ *is Boolean;*
(iv) *L has finite range and finite centre.*

PROOF. Suppose that L is strongly congruence uniform. Let θ be a congruence on L. Then, for any $a \in L$, $[a]\theta \cong [0]\theta$ and so $[a]\theta$ has a least element. Thus, (i) implies (ii).

Suppose, now, that (ii) holds. Let F be a normal filter of L and $\theta = \Theta_{\mathrm{lat}}(F)$. Then θ is a congruence on L and $F = [1]\theta$. Hence, $F = [a)$ for some $a \in L$ which, by Lemma 2, must belong to $\mathrm{Cen}(L)$. Thus, every normal filter of L is of the form $[z)$ with $z \in \mathrm{Cen}(L)$. It is now clear that $\mathrm{Cen}(L)$ is isomorphic to $\mathcal{F}^n(L)$ under the mapping $z \mapsto [z')$ where z' denotes the complement of z and, therefore, $\mathrm{Con}(L)$ is Boolean, since $\mathrm{Con}(L) \cong \mathcal{F}^n(L)$. Thus, (ii) implies (iii).

The equivalence of (iii) and (iv) has been established by H. P. Sankappanavar in [8] and so it remains only to show that (iv) implies (i). Suppose then that (iv) holds. Let $F \in \mathcal{F}^n(L)$. Then $F = \bigvee(N(a) \colon a \in F)$, where $N(a)$ is the principal normal filter in L generated by $a \in F$. Now, it is straightforward to verify that $N(a) = \{x \in L \colon x \geq a^{m(+*)}$ for some $m < \omega\}$ and, since L has finite range, the existence of a smallest integer n such that $a^{n(+*)} \in \mathrm{Cen}(L)$ is guaranteed. As a direct consequence, $N(a) = N(a^{n(+*)})$ and so $F = \bigvee(N(a) \colon a \in F \cap \mathrm{Cen}(L))$. However, $F \cap \mathrm{Cen}(L)$ is finite, so that $z = \bigwedge(F \cap \mathrm{Cen}(L))$ exists and $z \in \mathrm{Cen}(L)$ from which it follows easily that $F = N(z) = [z)$. Consequently, if θ is a congruence on L, then $[1]\theta \in \in \mathcal{F}^n(L)$ and $[1]\theta = [z)$, for some $z \in \mathrm{Cen}(L)$, so that θ is strongly uniform, by Corollary 3. $\square$

REMARK. All the concepts and results about double Heyting algebras that we have employed or proven are, in fact, meaningful and valid in the context of Heyting algebras with dual pseudocomplementation; in particular, Theorems 6 and 7 are valid for such algebras.

REFERENCES

[1] ADAMS, M. E. and BEAZER, R., Congruence uniform distributve lattices, *Acta Math. Hungar.* **57** (1991), 41–52.

[2] BALBES, R. and DWINGER, PH., *Distributive lattices*, University of Missouri Press, Columbia, Missouri, 1974. *MR* **51** #10185

[3] BEAZER, R., The determination congruence on double p-algebras, *Algebra Universalis* **6** (1976), 121–129. *MR* **54** #7341

[4] BEAZER, R., Congruence uniform algebras with pseudocomplementation, *Studia Sci. Math. Hungar.* **20** (1985), 43–48. *MR* **88d**: 06012

[5] GRÄTZER, G., *General lattice theory*, Birkhäuser Verlag, Basel and Stuttgart, 1978. *MR* **80c:** 06001a

[6] KATRIŇÁK, T., The structure of distributive double p-algebras. Regularity and congruences, *Algebra Universalis* **3** (1973), 238–246. *MR* **48** #10924

[7] KATRIŇÁK, T., Subdirectly irreducible distributive double p-algebras, *Algebra Universalis* **10** (1980), 195–219. *MR* **81g**: 06006

[8] SANKAPPANAVAR, H. P., Heyting algebras with dual pseudocomplementation, *Pacific J. Math.* **117** (1985), 405–415. *MR* **86m:** 06021

[9] VARLET, J. C., A regular variety of type $\langle 2, 2, 1, 1, 0, 0\rangle$, *Algebra Universalis* **2** (1972), 218–223. *MR* **48** #3824

(Received May 18, 1988)

DEPARTMENT OF MATHEMATICS
STATE UNIVERSITY OF NEW YORK
NEW PALTZ, NY 12561
U.S.A.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GLASGOW
UNIVERSITY GARDENS
GLASGOW G12 8QW
SCOTLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 71–81

ON THE INVERSE FUNCTION THEOREM IN F-SPACES

B. SLEZÁK

The well-known Inverse Function Theorem says that if X and Y are Banach spaces, $f \in X \to Y$ is a function which is strictly differentiable at the point x and the derivative $Df(x)\colon X \to Y$ is an isomorphism then f is a local homeomorphism, i.e. there exists an open neighbourhood V of the point x such that $f_{|V}$ is a homeomorphism between the set V and the open set $f(V)$. However, there are examples (see [5]) when f is continuously differentiable function between Fréchet-spaces, $Df(x)$ is an isomorphism at each point x but f is not a local homeomorphism because its range is nowhere dense in Y.

The following question arises: under which assumptions is it true that the derivative "locally characterizes" the function in the sense that if the derivative is an isomorphism (an injection or a surjection) then f is a local isomorphism (injection or surjection, respectively). In this paper I give a sufficient condition in the case that X is a complete metric space, Y is a metrizable topological vector space with an invariant metric and the functions $A \in X \to Y$ and $f \in X \to Y$ are "close enough" to each other (Theorem 1). In Theorem 2 I give a generalization of the open mapping theorems given in [2] and [8] to p-homogeneous pseudonormable vector spaces as a special case of Theorem 1.

T. Szilágyi called my attention to some results which are closely related to this problem (private communication). Indeed, the Miliutin Theorem (see [3] or [4]) is equivalent to Theorem 1 (i). The Dmitruk Theorem (see [3]) is more general than the Miliutin Theorem and the Open Mapping Theorem of Phan Quoc Khanh (see [6]) includes some general open mapping theorems as consequences, for instance the Dmitruk theorem and the Ptak's closed graph theorem.

Definition 1. Let (X, d_1) be a metric space, Y a metrizable topological vector space with the metric $d_2, U \subset X$ a nonempty set. Suppose that $A\colon U \to Y$ and $f\colon U \to Y$ are functions such that $A - f$ satisfies Lipschitz condition on the set $V \subset U$. We define the *semidistance* of the functions A

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 46A06; Secondary 46A30.

Key words and phrases. Complete system, a-covering function, b-compressed function, strictly tangential functions, p-normed topological vector space.

Akadémiai Kiadó, Budapest

and f in the following way:

$$d_V(A,f):=\min\{k \mid \forall\{u,v\}\subset V: d_2((A-f)(u),(A-f)(v))\leqq kd_1(u,v)\}.$$

If $A-f$ is not Lipschitzian on V then let $d_V(A,f):=\infty$.

The function A and f are *strictly tangential* at the point $x\in U$ if $A(x)=$ $=f(x)$ and for each positive number ε there is an open neighbourhood V of the point x such that $d_V(A,f)<\varepsilon$.

If X is a topological vector space, A is continuous affine mapping and A and f are strictly tangential at x then f is *strictly differentiable* at x and $Df(x)=A-A(0)$.

REMARK. It is easy to see that if $(X,d_1),(Y,d_2)$ and (Z,d_3) are metric linear spaces with invariant metrics, A and f are strictly tangential at $x\in$ $\in X$, B and g are strictly tangential at x, C and h are strictly tangential at $f(x)$ then $A+B$ and $f+g$ are strictly tangential at x and $C\circ A$ and $h\circ f$ are strictly tangential at x. Furthermore, A,B,f and g are pairwise strictly tangential at x. But we give a counterexample after Theorem 1 for two different continuous linear mappings which are strictly tangential at a point. It means that the derivative defined above is not necessarily unique.

NOTATION. If r is a positive number, (X,d) is a metric space and $x\in X$ then

$$B(x;r):=\{u\in X \mid d(x,u)\leqq r\}$$

denotes a closed ball of radius r.

DEFINITION 2. A family Σ of closed balls in the metric space X is called a *complete system* if $B(x;r)\in\Sigma$ implies $B(x';r')\in\Sigma$ whenever $r'+d(x,x')\leqq r$ (see [4] and [3]). Let a and b be positive numbers, X and Y metric spaces. The function $T\colon X\to Y$ is said to be *a-covering* on the system Σ if

$$B(T(x);a\cdot r)\subset T(B(x;r)) \text{ whenever } B(x;r)\in\Sigma.$$

The function $S\colon X\to Y$ is called *b-compressed* on the system Σ if $S(B(x;r))\subset B(S(x);b\cdot r)$, whenever $B(x;r)\in\Sigma$. (If, for instance, S is a Lipschitzian function on every ball belonging to Σ with constant b.)

Let X' be a non-empty subset of X. If the set

$$\Sigma':=\{B(x;r)\cap X' \mid x\in X' \text{ and } B(x;r)\in\Sigma\}$$

is not empty then Σ' is called the system in the metric space $(X',d_{|X'\times X'})$, inherited from the system Σ.

LEMMA 1. *Let T and S be functions between the metric spaces (X,d_1) and (Y,d_2), Σ a system of closed balls in X.*

(i) *If T is invertible then T is a-covering on Σ if and only if T^{-1} is $\frac{1}{a}$-compressed on the system $\Sigma_2:=\{B(T(x);ar)) \mid B(x,r)\in\Sigma\}$.*

(ii) *If Σ is complete and S is b-compressed on Σ then S satisfies Lipschitz-condition on the balls $B(x;r)$ with the constant b, whenever $B(x;3r)\in\Sigma$.*

(iii) *Let X' be an arbitrary nonempty subset of X. If Σ is complete and the system Σ' in X' inherited from Σ exists then Σ' is complete.*

PROOF. (i) The two properties are equivalent because $B(T(x);ar)\subset \subset T(B(x;r))$ if and only if $T^{-1}(B(T(x);ar))\subset B(x;r)$.

(ii) Let us suppose that $B(x_0,3r)\in\Sigma$ and $u,x\subset B(x_0;r)$. Then $B(x;d_1(x,u))\in\Sigma$ since $d_1(x,u)\leqq 2r\leqq 3r-d_1(x_0,x)$. So

$$S\left(B\left(x;d_1(x,u)\right)\right)\subset B\left(S(x);bd_1(x,u)\right),$$

hence $d_2(S(x),S(u))\leqq bd_1(x,u)$.

(iii) Suppose that $B'(x';r)\in\Sigma'$, $u\in X'$ and $r'+d(x',u)\leqq r$. Then $B(x';r)\in\Sigma$ and $B(u;r')\in\Sigma$ since Σ is quasicomplete. Hence $B'(u;r')\in \in\Sigma'$. □

LEMMA 2. *Let (X,d_1) be a metric space, Y metrizable topological vector space with the translation invariant metric d_2, $A\colon X\to Y$ and $f\colon X\to Y$ continuous functions, Σ complete system in X. Let us suppose that A is a-covering and $A-f$ is b-compressed on Σ, $0<b<a$. If $B(x_0;r)\in\Sigma$ and $y\in B(f(x_0);(a-b)r)$ then there exists a sequence $\{x_n\}_{n=0}^{\infty}\subset B(x_0;r)$ with the following properties:*

(i) *$A(x_{n+1})=(A-f)(x_n)+y$, $n\in N$;*

(ii) *x_n and $A(x_n)$ are Cauchy-sequences in X and in Y, respectively;*

(iii) *$\lim d_2(A(x_n),A(x_0))\leqq ra$.*

PROOF. Put $k:=\frac{b}{a}$. We show by induction on n that if x_n is defined then we can give x_{n+1} such that $A(x_{n+1})=(A-f)(x_n)+y$ and

$$d_1(x_\ell,x_m)\leqq rk^m(1-k^{\ell-m}), \tag{1}$$

furthermore

$$B(x_m;d_1(x_n,x_m))\in\Sigma$$

and

$$B(x_n;kd_1(x_n,x_{n-1}))\in\Sigma \tag{2}$$

whenever $0\leqq m<\ell<n+1$.

If $n=0$ then

$$\begin{aligned}&(A-f)(x_0)+y\in(A-f)(x_0)+B(f(x_0);r(a-b))=\\ &=B(A(x_0);r(a-b))\subset A\left(B\left(x_0;r\frac{a-b}{a}\right)\right)=A\left(B\left(x_0;r(1-k)\right)\right),\end{aligned}$$

because d_2 is translation invariant, A is a-covering on Σ and $B(x_0; r(1-k)) \in \in \Sigma$, as $B(x_0; r) \in \Sigma$ and $r(1-k) \leqq r$. Hence there exists a point $x_1 \in \in B(x_0; r(1-k))$ such that $A(x_1) = (A-f)(x_0) + y$.

Also, it is easy to see that the properties (1) and (2) hold.

Suppose that x_n is defined. Then

$$\begin{aligned}(A-f)(x_n)+y \in &(A-f)(B(x_{n-1}; d_1(x_n, x_{n-1})) + A(x_n) - (A-f)(x_{n-1}) \subset \\ &\subset B((A-f)(x_{n-1}); bd_1(x_n, x_{n-1})) + A(x_n) - (A-f)(x_{n-1}) \subset \\ &\subset B(A(x_n); bd_1(x_n, x_{n-1})) \subset \\ &\subset A(B(x_n; kd_1(x_n, x_{n-1})),\end{aligned}$$

as $A-f$ is b-compressed on Σ, d_2 is translation invariant, $B(x_{n-1}; d_1(x_n, x_{n-1})) \in \Sigma$, A is a-covering on Σ and $B(x_n; kd(x_n, x_{n-1})) \in \Sigma$.

It follows that there exists a point $x_{n+1} \in B(x_n; kd_1(x_n, x_{n-1}))$ such that

$$A(x_{n+1}) = (A-f)(x_n) + y$$

and if $0 \leqq m < n+1$ then

$$\begin{aligned}d_1(x_{n+1}, x_m) \leqq d_1(x_{n+1}, x_n) + d_1(x_n, x_m) \leqq \\ \leqq kd_1(x_n, x_{n-1}) + rk^m(1-k^{n-m}) \leqq rk^m(1-k^{n+1-m}).\end{aligned}$$

$B(x_{n+1}; kd_1(x_{n+1}, x_n)) \in \Sigma$, since

$$kd_1(x_{n+1}, x_n) + d_1(x_{n+1}, x_0) \leqq r$$

and $B(x_0; r) \in \Sigma$.

Furthermore, $B(x_m, d_1(x_{n+1}, x_m)) \in \Sigma$ because

$$d_1(x_{n+1}, x_m) + d_1(x_m, x_0) \leqq rk^m(1-k^{n+1-m}) + r(1-k^m) \leqq r.$$

Hence the definition of the sequence $\{x_n\}$ is correct and the inequalities (1) show that $\{x_n\}$ is a Cauchy sequence. Now we show that $\{A(x_n)\}$ is a Cauchy sequence, too. Indeed, if $0 \leqq m < n$ then

$$\begin{aligned}d_2(A(x_{n+1}), A(x_{m+1})) = d_2((A-f)(x_n), (A-f)(x_m)) \leqq \\ \leqq bd_1(x_n, x_m)\end{aligned}$$

because $x_n \in B(x_m; d_1(x_n, x_m)) \in \Sigma$ by (2), so

$$(A-f)(x_n) \in (A-f)B(x_m; d_1(x_n, x_m)) \subset B((A-f)(x_m); bd_1(x_n, x_m)).$$

In the end, let us see the sequence $\{d_2(A(x_n), A(x_0))\}$.

$$d_2(A(x_n), A(x_0)) \leqq \sum_{i=1}^{n-1} d_2(A(x_{i+1}), A(x_i)) + d_2(A(x_1), A(x_0)) \leqq$$

$$\leqq \sum_{i=1}^{n-1} bd_1(x_i, x_{i-1}) + (a-b)r \leqq br \sum_{i=1}^{n-1} k^{i-1}(1-k) + (a-b)r \leqq ar.$$

These inequalities show that if $d_2(A(x_n), A(x_0))$ converges then $\lim d_2(A(x_n), A(x_0)) \leqq ra$. But

$$d_2(A(x_n), A(x_0)) \leqq d_2(A(x_n), A(x_m)) + d_2(A(x_m), A(x_0)),$$

so

$$d_2(A(x_n), A(x_0)) - d_2(A(x_m), A(x_0)) \leqq bd_1(x_{n-1}, x_{m-1}),$$

whenever $0 < m < n$. It follows that $d_2(A(x_n), A(x_0))$ is a Cauchy sequence. □

PROPOSITION 1. *Let (X, d_1) be a metric space, (Y, d_2) metrizable topological vector space with the translation invariant metric d_2, $W \subset X$ a nonempty set, $A\colon X \to Y$ and $f\colon W \to Y$ functions. Suppose that A is continuous and $d_W(A, f) = b < \infty$. Then*

(i) *f is continuous on W.*

Suppose furthermore that A is injective, A^{-1} is Lipschitzian on $A(X)$ with the constant $\frac{1}{a}$ for some $a > b$. Then we have

(ii) *f_1 is injective and f^{-1} is Lipschitzian with the constant $\frac{1}{a-b}$.*

(iii) *If the set $H := A(X) \cap f(W)$ is not empty then $d_H(A^{-1}, f^{-1}) \leqq \frac{b}{a}\frac{1}{a-b}$.*

PROOF. (i) If $\{u, v\} \subset W$ then

$$\begin{aligned} d_2(f(u), f(v)) &= d_2(f(u) - A(u), f(v) - A(u)) \leqq \\ &\leqq d_2((f - A)(u), (f - A)(v)) + d_2((f - A)(v), f(v) - A(u)) \leqq \\ &\leqq bd_1(u, v) + d_2(A(u), A(v)). \end{aligned}$$

This shows that f is continuous.

(ii) Suppose that $\{u, v\} \subset W$, $u \neq v$ and $f(u) = f(v)$. Then

$$\begin{aligned} d_2(A(u), A(v)) &= d_2((A - f)(u), (A - f)(v)) \leqq \\ &\leqq bd_1(u, v) = bd_1(A^{-1}(A(u)), A^{-1}(A(v))) \leqq \frac{b}{a} d_2(A(u), A(v)), \end{aligned}$$

which is a contradiction. Let us suppose that $y, z \subset f(W)$

$$\begin{aligned} d_1(f^{-1}(y), f^{-1}(z)) &= d_1(A^{-1} \circ A \circ f^{-1}(y), A^{-1} \circ A \circ f^{-1}(z)) \leqq \\ &\leqq \frac{1}{a} d_2(A \circ f^{-1}(y), A \circ f^{-1}(z)) = \\ &= \frac{1}{a} d_2\left((A - f) \circ f^{-1}(y) + y, (A - f) \circ f^{-1}(z) + z\right) \leqq \\ &\leqq \frac{1}{a}\Big[d_2\Big((A - f) \circ f^{-1}(y), (A - f) \circ f^{-1}(z)\Big) + \\ &+ d_2\left((A - f) \circ f^{-1}(z) + y, (A - f) \circ f^{-1}(z) + z\right)\Big] \leqq \end{aligned}$$

$$\leqq \frac{b}{a} d_1\big(f^{-1}(y), f^{-1}(z)\big) + \frac{1}{a} d_2(y,z).$$

Hence

$$d_1(f^{-1}(y), f^{-1}(z)) \frac{1}{a} \frac{1}{1-\frac{b}{a}} d_2(y,z) = \frac{1}{a-b} d_2(y,z).$$

(iii) If $\{y,z\} \subset H$ then

$$d_1((A^{-1} - f^{-1})(y), (A^{-1} - f^{-1})(z)) =$$

$$= d_1(A^{-1}(y - A \circ f^{-1}(y)), A^{-1}(z - A \circ f^{-1}(z))) \leqq$$

$$\leqq \frac{1}{a} d_2(y - A \circ f^{-1}(y), z - A \circ f^{-1}(z)) =$$

$$= \frac{1}{a} d_2((f - A) \circ f^{-1}(y), (f - A) \circ f^{-1}(z)) \leqq$$

$$\leqq \frac{b}{a} d_1(f^{-1}(y), f^{-1}(z)) \leqq \frac{b}{a} \frac{1}{a-b} d_2(y,z). \qquad \square$$

THEOREM 1. *Let (X, d_1) be a metric space, Y a metrizable topological vector space with the translation invariant metric d_2, $W \subset X$ be a nonempty set, $\Sigma \subset \Sigma(W)$ a complete system, $A \colon X \to Y$ and $f \colon W \to Y$ functions. Suppose that A and f are continuous on every ball belonging to Σ, A is a-covering and $A - f$ is b-compressed on Σ, $0 < b < a$. Then*

(i) *If (X, d_1) is complete then f is $a-b$-covering on Σ.*

Moreover, let us suppose that (Y, d_2) is complete, i.e. (Y, d_2) is an F-space. Then

(ii) *If A is homeomorphism then f is $a-b$-covering on Σ.*

(iii) *If A has a continuous right inverse A_r which is $\frac{1}{a}$-compressed on the system $\Sigma_2 := \{B(A(x); a\varrho) \mid B(x;\varrho) \in \Sigma\}$ and the system Σ' in $X' := \operatorname{im}(A_r)$, inherited from Σ, exists then $f_{|X'}$ is $b-a$-covering on Σ'.*

REMARK. In fact, the continuity of f at the interior points of the balls follows from Proposition 1 (i) and from Lemma 1 (ii).

PROOF. (i) We have to show that $B(f(x_0); r(a-b)) \subset f(B(x_0;r))$, whenever $B(x_0;r) \in \Sigma$. By Lemma 2 if $B(x_0;r) \in \Sigma$ and $y \in B(f(x_0); r(a-b))$ then we can give a Cauchy sequence $\{x_n\}$ in $B(x_0;r)$ such that

$$A(x_{n+1}) = (A - f)(x_n) + y.$$

Since X is complete and A and $A - f$ are continuous, so x_n converges to a point x belonging to $B(x_0;r)$ and $A(x) = (A - f)(x) + y$. Hence $f(x) = y$. This means that $B(f(x_0); r(a-b)) \subset f(B(x_0;r))$.

(ii) We only have to show that if A is a homeomorphism and (Y, d_2) is complete then the sequence $\{x_n\}$ given in the point (i) above converges in X.

By Lemma 2 $\{A(x_n)\}$ is a Cauchy sequence and $\lim d_2(A(x_0), A(x_n)) \leqq ar$, hence $A(x_n)$ converges to a point $z \in B(A(x_0); ar)$. It follows from the continuity of A^{-1} that x_n converges to the point $x := A^{-1}(z) \in B(x_0; r)$. Now we can follow the proof on the same way as in the point (i) above.

(iii) The function $A_{|X'}$ is a homeomorphism. The system Σ' is complete in X' by Lemma 1 (iii). Furthermore, $A_{|X'}$ is an a-covering on Σ' by Lemma 1 (i). We are done by (ii). □

REMARK. Let f and g be functions mapping an F-space to another. If they are strictly tangential at a point x and f is a covering on a complete system Σ, then g is obviously locally open at x, provided $B(x; r) \in \Sigma$ for some $r > 0$. By the well-known Open Mapping Theorem every continuous linear mapping between F-spaces is open if it is onto. But it depends on the metrics whether it is covering or not. In Proposition 2 we show that every continuous linear mapping between p-normed spaces (see Definition 3) is open if and only if it is covering. The following question arises:

Suppose that A and f are functions between F-spaces, they are strictly tangential at a point x and A is continuous and linear. The question arises whether it is true that f is locally open at x. The answer is the negative. If A is not covering then it is possible that f is not locally injective and not locally open even if A is a linear homeomorphism, as it is shown in the following

COUNTEREXAMPLE. Let two metrics be given on R:

$$d_1(x, y) := |x - y|^{1/2} \text{ and } d_2(x, y) := |x - y|.$$

Then they are linear functions mapping from (R, d_1) to (R, d_2) strictly tangential to each other at the point zero. Indeed, if

$$\begin{aligned} f &:= C_1 \mathrm{id}_R : (R, d_1) \to (R, d_2) \\ g &:= C_2 \mathrm{id}_R : (R, d_1) \to (R, d_2), \quad C_1 \neq C_2, \end{aligned}$$

then

$$\begin{aligned} d_2((f - g)(x), (f - g)(y)) = |C_1 - C_2||x - y| \leqq \\ \leqq \varepsilon d_1(x, y) = \varepsilon |x - y|^{1/2} \end{aligned}$$

whenever

$$|x - y| \leqq \delta = \left(\frac{\varepsilon}{|C_1 - C_2|} \right)^2 .$$

It follows that the function $\mathbf{0}: R \to R$, $x \mapsto 0$ and the function id_R are strictly tangential at 0, however, $\mathbf{0}$ is neither injective nor locally open.

DEFINITION 3. Let X be a metrizable topological vector space with the translation invariant metric d. The function $\|.\| : X \to R$, $x \mapsto d(x, 0)$ is said to be a p-homogeneous pseudonorm if there is a positive number p such that for every point $x \in X$ and for every scalar t the equality $\|tx\| = |t|^p \|x\|$ holds.

Now we give a generalization of the theorems given in [2] and [8]: if the function f between two Banach-spaces is strictly differentiable at a point x_0 and the derivative $Df(x_0)$ is surjective then f is locally open; if $Df(x_0)$ is injective and its range is closed then f is locally injective.

PROPOSITION 2. *Let X and Y be topological vector spaces with the p-homogeneous pseudonorms $\|.\|_1$ and $\|.\|_2$, respectively, and $A\colon X \to Y$ be a linear mapping. Then*

(i) *A is continuous if and only if there exists a positive number K such that for every $x \in X$:*

$$\|A(x)\|_2 \leqq K\|x\|_1$$

(i.e. A is Lipschitzian on X with the constant K).

(ii) *Let $B(0;1)$ and $B(0;a)$ be closed balls in X and in Y, respectively. If $B(0;a) \subset A(B(0;1))$ then A is a-covering on the complete system $\Sigma(X)$.*

PROOF. If $r > 0$ then

$$B(0;r) = \{x \mid \|x\| \leqq r\} = \left\{x \mid \left\|\frac{1}{\sqrt[p]{r}}x\right\| = \frac{1}{r}\|x\| \leqq 1\right\} =$$
$$= \{\sqrt[p]{r}x \mid \|x\| \leqq 1\} = \sqrt[p]{r}B(0;1).$$

Hence

$$tB(0;r) = t\sqrt[p]{r}B(0;1) = B(0;t^p r), \tag{4}$$

whenever the scalar t is positive.

(i) It is clear that the condition is sufficient. Let us suppose that A is continuous. Then it is bounded, i.e. if the set H is bounded in X then $A(H)$ is bounded in Y. The equalities (4) show that every ball in X and Y is bounded, so there is a number $s > 0$ such that

$$A(B(0;\|x\|_1)) = A(\sqrt[p]{\|x\|_1}B(0;1)) = \sqrt[p]{\|x\|_1}A(B(0;1)) \subset$$
$$\subset \sqrt[p]{\|x\|_1}sB(0;1) = B(0;s^p\|x\|_1).$$

It means that $\|A(x)\|_2 \leqq K\|x\|_1$, whenever $x \in X$, $A(B(0;1)) \subset sB(0;1)$ and $K \geqq s^p$.

(ii) For every point $x \in X$ and for every positive number r

$$B(0;ar) = \sqrt[p]{r}B(0;a) \subset A(\sqrt[p]{r}B(0;1)) = A(B(0;r))$$

holds, hence

$$B(A(x);ar) \subset A(B(0;r)) + A(x) = A(B(0;r) + x) = A(B(x;r)). \qquad \square$$

The following proposition shows that the strict derivative of a function between p-normed spaces is unique. Also, it is easy to see that if A and f are strictly tangential at a point x_0 then tA and tf are strictly tangential at x_0 for every scalar t.

PROPOSITION 3. *Let X and Y be p-normed topological vector spaces, A and B be continuous linear mappings from X into Y. If A and B are (strictly) tangential at a point x_0 then $A = B$.*

PROOF. Suppose that for every positive number ε there exists a number δ such that $\|x - x_0\|_1 \leqq \delta$ implies

$$\|(A - B)(x) - (A - B)(x_0)\|_2 \leqq \varepsilon \|x - x_0\|_1.$$

Since for every point $u \in x$ there exists a number t such that $\|tu\|_1 \leqq \delta$ and

$$\|(A - B)(tu + x_0) - (A - B)(x_0)\|_2 = |t|^p \|(A - B)(u)\|_1 \leqq \\ \leqq \varepsilon \|tu\|_1 = |t|^p \varepsilon \|u\|_1,$$

we have that $\|(A - B)(u)\|_2 = 0$ for every $u \in x$. Hence $A = B$. □

THEOREM 2. *Suppose that $(X, \|.\|_1)$ and $(Y, \|.\|_2)$ are complete p-homogeneous pseudonormed topological vector spaces and the function $f \in X \to Y$ is strictly differentiable at the point $x_0 \in X$.*

(i) *If $Df(x_0)$ is surjective then f is locally open on a neighbourhood of x_0.*

(i') *If $Df(x_0)$ has a continuous linear right inverse R on Y (i.e. the kernel of $Df(x_0)$ is complemented in X) with Lipschitz constant $\frac{1}{a}$ then f has a local right inverse f_r which is strictly differentiable $f(x_0)$ and $Df_r(f(x_0)) = R$.*

(ii) *If $Df(x_0)$ is injective and its range is closed in Y then f is locally injective on a neighbourhood of the point x_0.*

(ii') *If $Df(x_0)$ has a continuous linear left inverse L, i.e. the range of $Df(x_0)$ is complemented, then f has a local left inverse $f_\ell \in Y \to X$, which is strictly differentiable at the point $f(x_0)$ and $Df_\ell(f(x_0)) = L$.*

(iii) *If $Df(x_0)$ is an isomorphism then f is locally homeomorphism and the local inverse f^{-1} of f is strictly differentiable at $f(x_0)$ and*

$$Df^{-1}(f(x_0)) = [Df(x_0)]^{-1}.$$

PROOF. It is clear that (iii) follows from (i), (i'), (ii) and (ii'). Let $A := Df(x_0)$. It follows from the Open Mapping Theorem (see for instance in [7]) that A is an open mapping from X onto the range of A, because by our assumptions the range is an F-space. Hence there is a positive number a such that $B(0; a) \subset A(B(0; 1))$. So A is a-covering on the complete system $\Sigma(X)$ by Proposition 2 (ii).

Let W be a neighbourhood of the point x_0 and ϱ be a positive number such that $b := d_W(f, A) < a$ and $B(x_0; \varrho) \subset W$. By Proposition 1 (i) f is continuous on W. It is clear that $A - f$ is b-compressed on the complete system $\Sigma(W)$.

(i) It follows from the point (i) of Theorem 1 that f is $a-b$-covering on $\Sigma(W)$, so it is open on the ball $B(x_0;\varrho)$. Hence the statement (i) holds.

(i') Put $A_r := R - R(A(x_0)) + x_0$ and $X' := \mathrm{im}(A_r) \cap W$. Obviously, $X' \neq \emptyset$ because $x_0 \in X'$. Since $A - f$ and A_r are Lipschitzian on X' with the constants b and $\frac{1}{a}$, respectively, $f_{|X'}$ is invertable by Proposition 1 (ii). Denote $(f_{|X'})^{-1}$ by f_r. By Lemma 1 (i) $A_{|X'}$ is an a-covering on the system Σ' in X' inherited from $\Sigma(W)$ and Σ' is complete by Lemma 1 (iii). We have that $B(f(x_0);(a-b)\varrho) \subset f(X')$, implying Theorem 1 (ii).

As

$$H := A(X') \cap B(f(x_0);(a-b)\varrho) = B(f(x_0);(a-b)\varrho),$$

it follows from Proposition 1 (iii) that $d_H(A_r, f_r) \leqq b\frac{a}{a-b}$.

As the definition of f_r does not depend on the choice of b (if $b < a$), it follows that f_r is strictly differentiable at $f(x_0)$ and $Df_r(f(x_0)) = R$. Indeed, the number b can be arbitrary small because f is strictly differentiable at x_0.

(ii) The function $A\colon X \to A(X)$ is open by the Open Mapping Theorem. So A^{-1} is continuous linear mapping and Lipschitzian with some constant "a" by Proposition 2 (ii). Since A is the strict derivative of f at x_0, we can choose the neighbourhood W of x_0 such that $b := d_w(A;f) < a$. We have that $f_{|W}$ is injective, applying Proposition 1 (ii).

(ii') If A has a continuous linear left inverse L then $Y = E \oplus F$, where E denotes the range of A and F denotes the kernel of L. The function $\mathrm{pr}_E\colon E \oplus F \to E$, $u + v \mapsto u (u \in E, v \in F)$ is continuous and linear. Hence $\mathrm{pr}_E \circ f$ is strictly differentiable at x_0 and

$$D(\mathrm{pr}_E \circ f)(x_0) = \mathrm{pr}_E \circ A = A.$$

By the point (ii) $\mathrm{pr}_E \circ f_{|W}$ is invertable and

$$D((\mathrm{pr}_E \circ f_{|W})^{-1})(\mathrm{pr}_E(f(x_0))) = A^{-1}, \text{ if } d_W(\mathrm{pr}_E \circ f, A) \leqq b < a,$$

where $\frac{1}{a}$ is the Lipschitz constant of L. So the function $f_\ell := \left(\mathrm{pr}_E \circ f_{|W}\right)^{-1} \circ$ $\circ \mathrm{pr}_E$ is strictly differentiable at $f(x_0)$ and $Df_\ell(f(x)) = A^{-1} \circ \mathrm{pr}_E = L$. Furthermore, $f_\ell \circ f(x) = (\mathrm{pr}_E \circ f_{|W})^{-1} \circ \mathrm{pr}_E \circ f(x) = x$, whenever $x \in W$. □

In Theorem 1 (iii) we assumed that the function A has a right inverse A_r which is compressed on the system Σ_2. When A is a covering linear mapping then $\Sigma_2 = \Sigma(Y)$ and A_r is compressed if and only if it is Lipschitzian, as it follows from Lemma 1 (iii). But we do not know whether a Lipschitzian linear function has a Lipschitzian right inverse, even if it is between Banach spaces. This question is equivalent to the following

PROBLEM. Let E be a Banach space and N be a closed subspace of it. Does the canonical projection $\pi\colon E \to E/N$ between E and the factor-space E/N have a locally Lipschitzian right inverse?

The Michael's selection theorem says that there is continuous right inverse, but it is not Lipschitzian, and we cannot apply the so called Approximate Selection Theorem, because the relation π^{-1} is not upper semicontinuous (see [1]).

REFERENCES

[1] Aubin, J. P. and Cellina, A., *Differential inclusions*, Springer-Verlag, Berlin–Heidelberg–New York, 1984.

[2] Bourbaki, N., *Éléments de mathématique. Fasc. XXXIII, Variétés différentiels et analytiques*, Fascicule de résultats (Paragraphes 1 à 7), Actualités Scientifiques et Industrielles, No. 1333, Hermann, Paris, 1967. *MR* **36** #2161

[3] Dmitruk, A. V., Milyutin, A. A. and Osmolovskiĭ, N. P., Ljusternik's theorem and the theory of the extremum, *Uspekhi Mat. Nauk* **35** (1980), No. 6, 11–46 (in Russian). *MR* **82c**: 58010

[4] Levitin, E. S., Milyutin, A. A. and Osmolovskiĭ, N. P., Higher order conditions for local minima in problems with constraints, *Uspekhi Math. Nauk.* **33**, (1978), no. 6, 85–148 (in Russian). *MR* **80f**: 49001

[5] Łojasiewicz, S., An example of continuous injective polynomial with nowhere dense range, whose differential at each point is an isomorphism, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **24** (1976), 1109–1111. *MR* **55** #11292

[6] Phan, Quoc Khánh, An induction theorem and general open mapping theorems, *J. Math. Anal. Appl.* **118** (1986), 519–534. *MR* **87m**: 46014

[7] Rudin, W., *Functional analysis*, McGraw-Hill series in higher mathematics, McGraw-Hill, New York, 1973. *MR* **51** #1315

[8] Szilágyi, T., Generalization of the implicit function theorem and of Banach's open mapping theorem, *Acta Sci. Math. (Szeged)* **39** (1977), 391–396. *MR* **58** #18527

(Received March 10, 1988)

VESZPRÉMI EGYETEM
MATEMATIKA TANSZÉK
EGYETEM ÚT 12
H–8201 VESZPRÉM
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 83–95

A NOTE ON THE ARITHMETIC FORM OF THE LARGE SIEVE

A. SÁRKÖZY

1. Throughout this paper, we use the following notation: We write $e^{2\pi i\alpha} = e(\alpha)$. The distance from the real number x to the nearest integer is denoted by $\|x\|$. $p(n)$ denotes the least prime factor of the positive integer n, while $P(n)$ denotes the greatest prime factor of n. $\nu(n)$ is the total number of prime factors of n so that $\nu(p_1^{\alpha_1}\dots p_r^{\alpha_r}) = \alpha_1 + \dots + \alpha_r$. The cardinality of the finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$.

2. The analytical form of the large sieve is the following:

If K is an integer, N is a positive integer, $a_{K+1}, a_{K+2}, \dots, a_{K+N}$ are complex numbers, $\mathcal{X}$ is a set of real numbers for which $\|x - x'\| \geqq \delta > 0$ whenever x and x' are distinct members of $\mathcal{X}$, and we write

$$S(x) = \sum_{n=K+1}^{K+N} a_n e(nx),$$

then

$$\sum_{x\in X} |S(x)|^2 \ll (\delta^{-1} + N) \sum_{n=K+1}^{K+N} |a_n|^2. \tag{1}$$

In fact, Montgomery [9] proved this with a constant factor 1 on the right-hand side:

$$\sum_{x\in X} |S(x)|^2 \leqq (\delta^{-1} + N) \sum_{n=K+1}^{K+N} |a_n|^2. \tag{2}$$

(See also [8], pp. 12–13.)

To derive the arithmetic form of the large sieve, assume that $\mathcal{N} \in \{K+1, K+2, \dots, K+N\}$, and write

$$Z = |\mathcal{N}|, \quad Z(q,h) = |\{n\colon n \in \mathcal{N},\ n \equiv h \pmod q\}|. \tag{3}$$

1991 *Mathematics Subject Classifications.* Primary 11N35.
Key words and phrases. Large sieve, distribution of integers in residue classes.

Research supported by Hungarian National Foundation for Scientific Research Grant no. 1811.

Akadémiai Kiadó, Budapest

If q is a positive integer, $b_0, b_1, \dots, b_{q-1}$ are any complex numbers and we write

$$F(\alpha) = \sum_{h=0}^{q-1} b_h e(h\alpha),$$

then by a well-known Parseval-formula type identity we have

$$q \sum_{h=0}^{q-1} |b_h|^2 = \sum_{a=0}^{q-1} \left| F\left(\frac{a}{q}\right) \right|^2. \tag{4}$$

Choosing here $b_h = Z(q,h) - \frac{Z}{q}$ and writing

$$T(\alpha) = \sum_{n \in \mathcal{N}} e(n\alpha), \tag{5}$$

clearly we have $F(0) = 0$ and $F(\frac{a}{q}) = T(\frac{a}{q})$ for $q \nmid a$, hence (4) can be rewritten in the form

$$q \sum_{h=0}^{q-1} \left(Z(q,h) - \frac{Z}{q} \right)^2 = \sum_{a=1}^{q-1} \left| T\left(\frac{a}{q}\right) \right|^2. \tag{6}$$

Let Q be a positive real number. Then writing up this identity for every prime $p \leqq Q$, and adding the formulas obtained in this way, we get

$$\sum_{p \leqq Q} p \sum_{h=0}^{p-1} \left(Z(p,h) - \frac{Z}{p} \right)^2 = \sum_{p \leqq Q} \sum_{a=1}^{p-1} \left| T\left(\frac{a}{p}\right) \right|^2. \tag{7}$$

To estimate the right-hand side, we use (1) with $T(\alpha) = S(\alpha)$ and $\mathcal{X} = \left\{ \frac{a}{p} : p \leqq Q,\, 1 \leqq a \leqq p-1 \right\}$. In this way, we get the standard arithmetic form of the large sieve:

$$\sum_{p \leqq Q} p \sum_{h=0}^{p-1} \left(Z(p,h) - \frac{Z}{p} \right)^2 \ll (Q^2 + N)Z. \tag{8}$$

In the most applications we use this large sieve inequality for giving an upper bound for Z under the assumption that for many primes $p \leqq Q$, the set $\mathcal{N}$ must not meet many (say, $> cp$) "forbidden" residue classes modulo p.

3. In several problems studied recently one would need an extension of the arithmetic form (8) to composite moduli. In other words, one would like to give an upper bound for

$$\sum_{m \in \mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 \tag{9}$$

for any set $\mathcal{M}$ of sifting moduli (and for any set $\mathcal{N}$).

Extensions of the large sieve to composite module have been given by Montgomery [7], Johnsen [6], Gallagher [5], Salerno and Viola [11], Elliott [2], Erdős and Sárközy [4] and Elliott and Sárközy [3]. Montgomery's idea was to replace the identity (6) by

$$q\sum_{h=0}^{q-1}\left|\sum_{d\mid q}\frac{\mu(d)}{d}Z\left(\frac{q}{d},h\right)\right|=\sum_{\substack{0\leqq a\leqq q-1\\(a,q)=1}}\left|T\left(\frac{a}{q}\right)\right|^2. \tag{10}$$

If the sifting moduli are primes and there are "many" ($\sim p/2$ or more) "forbidden" residue classes for every $p\leqq Q$ (including the small primes), then by using (10) he could sharpen (by saving a $\log Q$ factor) on the estimate for Z that can be derived from (8). On the other hand, in the general case one cannot give an upper bound for (9) in this way. Johnsen [6] and Gallagher [5] have studied the more general case when the "forbidden" residue classes belong to prime powers (including the primes). Salerno and Viola [11] studied sifting by almost primes under quite strong conditions on both the sifting moduli and the forbidden residue classes. Elliott [2] estimated a sum quite close to (9) in the case when the sifting moduli are small. Erdős and Sárközy [4] gave an upper bound for (9) in the special case when $\mathcal{M}$ consists of prime squares. Finally, Elliott and Sárközy [3] estimated a sum like (9) in a special situation.

In general, one cannot expect a non-trivial estimate for the sum (9) without any condition on $\mathcal{M}$. This can be illustrated by the following example: Let $\mathcal{M}=\mathcal{N}=\{n\colon n\leq N, 2\mid n\}$. Then both Z and $|\mathcal{M}|$ are large (in terms of N), however, also the "variance" (9) is large. In this paper our goal is to show that this difficulty can be avoided by certain mild assumptions on the prime factor structure of the elements of $\mathcal{M}$. In fact, we will derive a non-trivial and nearly best possible estimate for (9) from the composite moduli analogue of (7) in each of the important special situations when the sifting set $\mathcal{M}$ consists of (i) prime powers; (ii) almost primes; (iii) integers all whose prime factors are of medium size; (iv) highly composite numbers (i.e., numbers having many small prime factors).

4. We will prove the following theorem:

THEOREM. *Let K be an integer, let N be a positive integer, assume that $\mathcal{N}\subset\{K+1,K+2,\dots,K+N\}$, and define Z, $Z(q,h)$ and $T(\alpha)$ by (3) and (5), respectively. Let $\mathcal{M}$ be a set of positive integers, and write $M=|\mathcal{M}|$, $\mathcal{M}_d=\{m\colon m\in\mathcal{M},\ d\mid m\}$, $M_d=|\mathcal{M}_d|$. Then we have*

$$\sum_{m\in\mathcal{M}}m\sum_{h=0}^{m-1}\left(Z(m,h)-\frac{Z}{m}\right)^2=\sum_{d=2}^{+\infty}M_d\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2. \tag{11}$$

Furthermore,

(i) *If k is a positive integer and v is a real number, then we have*

$$\sum_{p\leqq v} p^k \sum_{h=0}^{p^k-1} \left(Z(p^k,h) - \frac{Z}{p^k} \right)^2 \leqq (v^{2k}+N)Z. \tag{12}$$

(ii) *If $k \geqq 2$ is a positive integer, u, t are real numbers with $2 \leqq u$, $u^k \leqq t$ and $\mathcal{M} = \{m\colon m \leqq t,\ p(m) \geqq u,\ \nu(m) = k\}$, then we have*

(13)

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 < c\left(t + \frac{N}{u\log t}(\log\log t - \log\log u)^{k-2} \right) tZ$$

where $c = c(k)$ depends on k (but it is independent of all the other parameters).

(iii) *If u, v, t are real numbers with $2 \leqq u < v \leqq t \leqq N$, $u^2 \leqq t$, and $\mathcal{M} =$ $= \{m\colon 1 < m \leqq t,\ u \leqq p(m) < P(m) \leqq v\}$, then we have*

(14)

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 \ll \begin{cases} \left(t^2 + \frac{Nt}{u\log u} \right) Z & \text{for } uv \geqq t \\ \left(t^2 + \frac{Nv}{\log u} + \frac{Nt\log v}{u\log u\log t} \right) Z & \text{for } uv < t. \end{cases}$$

(iv) *If u, v are real numbers with $2 \leqq u < v$, k is a positive integer and $\mathcal{M}$ denotes the set of all the square-free integers that are products of exactly k prime factors p satisfying $u < p \leqq v$, then we have*

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 < \left(2v^{2k} + N\binom{\pi(v)-\pi(u)}{k-1} \right) Z. \tag{15}$$

REMARKS 1. One can use (11) for deriving a good upper bound for (9) only if the numbers M_d are small, say, $M_d = o(M)$ for all d. The example $\mathcal{M} = \mathcal{N} = \{n\colon n \leqq N, 2 \mid n\}$ given in Section 3 shows the necessity of this condition; in fact, in this case the problem is that M_2 is large. Thus to get a nontrivial upper bound for (9), one needs an assumption which implies that the numbers M_d are small. The simplest assumption of this type is to assume (as in (ii), (iii) and (iv)) that $p(m)$ is large for every $m \in \mathcal{M}$.

2. Both (12) and (15) involve (8) as a special case.

3. In the special situations studied by Montgomery [7], Gallagher [5] and Elliott [2], their estimates give an upper bound for Z better by a logarithm factor than the one that can be derived from our estimates above. However, as also Montgomery's and Gallagher's comments show, it seems hopeless to save this logarithmic factor also in the general case (e.g., when u is large), and in the most applications this loss is irrelevant.

4. (11) generalizes (14) in [4], and (12) in (i) generalizes the inequality used in [4]. (14) in (iii) covers the situation studied in [3] while applications of (15) in (iv) (to attack problems like the ones in [1] and [10]) will be given by G. N. Sárközy and by Stewart and me.

5. Proof of the Theorem. First we are going to prove (11). By (6), we have

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1}\left(Z(m,h)-\frac{Z}{m}\right)^2 = \sum_{m\in\mathcal{M}}\sum_{a=1}^{m-1}\left|T\left(\frac{a}{m}\right)\right|^2. \tag{16}$$

If $1\in\mathcal{M}$, then the contribution of $m=1$ to both sides of (11) and (16) is 0, thus we may assume that $1\notin\mathcal{M}$. Let us write every fraction $\frac{a}{m}$ in the sum on the right-hand side in the reduced form $\frac{b}{d}$ where $d>1$, $(b,d)=1$ and $d\mid m$. Then we get

$$\begin{aligned}\sum_{m\in\mathcal{M}}\sum_{a=1}^{m-1}\left|T\left(\frac{a}{m}\right)\right|^2 &= \sum_{m\in\mathcal{M}}\sum_{\substack{d\mid m\\ d>1}}\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2 = \\ &=\sum_{d=2}^{+\infty}\left(\sum_{m\in\mathcal{M}_d}1\right)\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2=\sum_{d=2}^{+\infty}M_d\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2.\end{aligned} \tag{17}$$

(11) follows from (16) and (17).

6. Now we are going to derive (i)–(iv) from (11).

(i) We use (11) with $\mathcal{M}=\{p^k\colon p \text{ prime},\ p\leqq v\}$. Then clearly,

$$M_d=\begin{cases}1 & \text{if } d=p^j,\ p\leqq v,\ 1\leqq j\leqq k\\ 0 & \text{otherwise.}\end{cases}$$

Thus it follows from (11) that

$$\begin{aligned}\sum_{p\leqq v}p^k\sum_{h=0}^{p^k-1}\left(Z(p^k,h)-\frac{Z}{p^k}\right)^2 &= \sum_{m\in\mathcal{M}} m\sum_{h=0}^{m-1}\left(Z(m,h)-\frac{Z}{m}\right)^2 = \\ &=\sum_{p\leqq v}\sum_{j=1}^{k}\sum_{\substack{0<b<p^j\\(b,p)=1}}\left|T\left(\frac{b}{p^j}\right)\right|^2.\end{aligned}$$

To estimate the inner sum, we use (2) with $a_n=1$ if $n\in\mathcal{N}$, $a_n=0$ if $n\notin\mathcal{N}$, $a_n=0$ if $n\notin\mathcal{N}$ and $\mathcal{X}=\left\{\frac{k}{p^j}\colon p\leqq v,\ 1\leqq j\leqq k,\ 0<b<p^j,\ (b,p)=1\right\}$. If $\frac{b}{p^j}\in\mathcal{X}$, $\frac{a}{q^i}\in\mathcal{X}$ and $\frac{b}{p^j}\neq\frac{a}{q^i}$, then clearly

$$\left\|\frac{b}{p^j}-\frac{a}{q^i}\right\|\geqq\frac{1}{p^jq^i}\geqq\frac{1}{(pq)^k}\geqq\frac{1}{v^{2k}}.$$

Thus (2) yields

$$\sum_{p \leqq v} \sum_{j=1}^{k} \sum_{\substack{0<b<p^j \\ (b,p)=1}} \left| T\left(\frac{b}{p^j}\right) \right|^2 \leqq (v^{2k}+N) \sum_{n \in \mathcal{N}} 1 = (v^{2k}+N)Z.$$

7. (ii) Let us write

$$\pi(y,z,j) = |\{n \colon n \leqq y,\, p(n) \geqq z,\, \nu(n) = j\}|.$$

Clearly, for $j = 2, 3, \dots$ we have

$$\pi(y,z,j) = \sum_{z \leqq p \leqq y^{1/j}} \pi\left(\frac{y}{p}, p, j-1\right).$$

By using this recursion and the prime number theorem, it can be shown easily by induction that for fixed j ($\geqq 1$) and $2 \leqq z \leqq y$, $y \to +\infty$ we have

$$\pi(y,z,j) = \frac{1}{(j-1)!} \frac{y}{\log y} (\log\log y - \log\log z)^{j-1} + \tag{18}$$
$$+ O\left(\frac{y}{\log y} (\log\log(1+y))^{j-2}\right)$$

(where the implicit constant in the error term may depend on j).

Let us write

$$\mathcal{D}_r = \{n \colon n \leqq t,\, p(n) \geqq u,\, \nu(n) = r\}.$$

Then by (11), we have

$$\sum_{m \in \mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m}\right)^2 = \sum_{d=2}^{+\infty} M_d \sum_{\substack{0<b<d \\ (b,d)=1}} \left| T\left(\frac{b}{d}\right) \right|^2 = \tag{19}$$
$$= \sum_{r=1}^{k} \sum_{d \in \mathcal{D}_r} M_d \sum_{\substack{0<b<d \\ (b,d)=1}} \left| T\left(\frac{b}{d}\right) \right|^2.$$

Clearly,

$$d \in \mathcal{D}_r \text{ implies } d \geqq u^r, \tag{20}$$

thus in view of (18) for $d \in \mathcal{D}_r$ we have

$$M_d = |\{m \colon m \leqq t,\, p(m) \geqq u, \nu(m) = k,\, d \mid m\}| =$$

$$= |\{n\colon n \leqq t/d,\ p(n) \geqq u,\ \nu(n) = \nu(m) - \nu(d) = k - r\}| = \pi(t/d, u, k-r) <$$

(21)

$$< \frac{1}{(k-r-1)!}\frac{t}{d\log t/d}(\log\log t/d - \log\log u)^{k-r-1} +$$

$$+c\frac{t}{d\log t/d}(\log\log(1+t/d))^{k-r-2} \quad \text{for} \quad 1 \leqq r \leqq k-1,\ t/d \geqq u^{k-r}$$

and clearly,

(22) $$M_d \leqq 1 \text{ for } r = k$$

(23) $$M_d = 0 \text{ for } 1 \leqq r \leqq k-1,\ t/d < u^{k-r}.$$

Define the positive integer L by

(24) $$t < 2^L u^k \leqq 2t,$$

and for $r = 1, 2, \dots, k$, $j = 1, 2, \dots, L$, write $I(r,j) = (t/2^j u^{k-r}, t/2^{j-1}u^{k-r}]$. Then in view of (1), (20), (21), (22) and (23), it follows from (19) that

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1}\left(z(m,h) - \frac{Z}{m}\right)^2 =$$

$$= \sum_{d\in\mathcal{D}_k} M_d \sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2 + \sum_{r=1}^{k-1}\sum_{d\in\mathcal{D}_r} M_d \sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2 \leqq$$

$$\leqq \sum_{d\in\mathcal{D}_k}\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2 + \sum_{r=1}^{k-1}\sum_{j=1}^{L}\left(\max_{d\in\mathcal{D}_r\cap I(r,j)} M_d\right)\sum_{d\in\mathcal{D}_r\cap I(r,j)}\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2 \ll$$

$$\ll (t^2+N)Z + \sum_{r=1}^{k-1}\sum_{j=1}^{L}\left(\frac{1}{(k-r-1)!}\frac{2^j u^{k-r}}{\log 2^j u^{k-r}}(\log\log 2^j u^{k-r} - \log\log u)^{k-r-1} +\right.$$

(25)

$$\left.+\frac{2^j u^{k-r}}{\log 2^j u^{k-r}}\left(\log\log(1+2^j u^{k-r})\right)^{k-r-2}\right)\left(\frac{t^2}{2^{2j}u^{2(k-r)}} + N\right)Z =$$

$$= (t^2+N)Z + Z\sum_{r=1}^{k-1}\sum_{j=1}^{L}\left(\frac{1}{(k-r-1)!}\frac{t^2}{2^j u^{k-r}\log 2^j u^{k-r}}(\log\log 2^j u^{k-r} -\right.$$

$$- \log\log u)^{k-r-1} + \frac{t^2}{2^j u^{k-r} \log 2^j u^{k-r}} (\log\log(1+2^j u^{k-r}))^{k-r-2} \Bigg) +$$

$$+ NZ \sum_{r=1}^{k-1} \sum_{j=1}^{L} \left(\frac{1}{(k-r-1)!} \frac{2^j u^{k-r}}{\log 2^j u^{k-r}} (\log\log 2^j u^{k-r} - \log\log u)^{k-r-1} + \right.$$

$$\left. + \frac{2^j u^{k-r}}{\log 2^j u^{k-r}} (\log\log(1+2^j u^{k-r}))^{k-r-2} \right).$$

An easy computation shows that it suffices to keep the first term $(t^2+N)Z$ and the first half of the $r=1$, $j=L$ term of the last double sum, since the total contribution of all the other terms is less than a constant multiple of these terms. Thus, in view of $u^k \leqq t$ and (24), it follows from (25) that

$$\sum_{m \in \mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 \ll$$

$$\ll (t^2+N)Z + NZ \frac{1}{(k-2)!} \frac{2^L u^{k-1}}{\log 2^L u^{k-1}} (\log\log 2^L u^{k-1} - \log\log u)^{k-2} \ll$$

$$\ll \left(t^2 + N + \frac{tN}{u \log t} (\log\log t - \log\log u)^{k-2} \right) Z \ll$$

$$\ll \left(t + \frac{N}{u \log t} (\log\log t - \log\log u)^{k-2} \right) tZ.$$

8. (iii) If $d>1$ and $M_d>0$, then clearly, $u \leqq d \leqq t$ and, writing $\mathcal{D} = \{n\colon 1<n \leqq t,\ u \leqq p(n) \leqq P(n) \leqq v\}$, we have $d \in \mathcal{D}$. Let

$$\mathcal{D}_1 = \mathcal{D} \cap \left(\frac{t}{u}, \right], \quad \mathcal{D}_2 = \mathcal{D} \cap \left[\max\left(u, \frac{t}{v}\right), \frac{t}{u} \right],$$

$$\mathcal{D}_3 = \begin{cases} \mathcal{D} \cap [u, \frac{t}{v}) & \text{for } uv < t \\ \emptyset & \text{for } uv \geqq t. \end{cases}$$

By (11), we have

$$\sum_{m \in \mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 =$$

$$= \sum_{d \in \mathcal{D}} M_d \sum_{\substack{0<b<d \\ (b,d)=1}} \left| T\left(\frac{b}{d}\right) \right|^2 = \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3, \tag{26}$$

where

$$\sum\nolimits_i = \sum_{d\in\mathcal{D}_i} M_d \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2 \quad \text{(for } i=1,2,3).$$

If $d\in\mathcal{D}_1$ and $M_d>0$, then there is a positive integer k with $dk\in\mathcal{M}$. This implies that

$$k=1 \text{ or } p(k)\geqq u \tag{27}$$

and

$$k \leqq \frac{t}{d} < \frac{t}{t/u} = u. \tag{28}$$

It follows from (27) and (28) that $k=1$, hence $M_d=1$. Thus

$$\sum\nolimits_1 = \sum_{d\in\mathcal{D}_1} \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2.$$

To estimate this sum we use (1) with $\mathcal{X}=\left\{\frac{b}{d}\colon d\in\mathcal{D}_1,\ 0<b<d,\ (b,d)=1\right\}$. If $\frac{b}{d}\in\mathcal{D}_1$, $\frac{b'}{d'}\in\mathcal{D}_1$ and $\frac{b}{d}\neq\frac{b'}{d'}$, then

$$\left\|\frac{b}{d}-\frac{b'}{d'}\right\| \geqq \frac{1}{dd'} \geqq \frac{1}{t^2}.$$

Thus we obtain

$$\sum\nolimits_1 \ll (t^2+N)Z. \tag{29}$$

Assume now that $d\in\mathcal{D}_2$. Then

$$\begin{aligned} M_d &= |\{m\colon m\in\mathcal{M},\ d\mid m\}| = |\{k\colon dk\in\mathcal{M}\}| = \\ &= |\{k\colon u\leqq p(k)\leqq P(k)\leqq v,\ k\leqq t/d\}|. \end{aligned}$$

By $d\in\mathcal{D}_2$, here we have

$$\frac{t}{d}\leqq\frac{t}{t/v}=v$$

and

$$\frac{t}{d}\geqq\frac{t}{t/u}=u.$$

Thus by Brun's sieve,

$$M_d=|\{k\colon u\leqq p(k), k\leqq t/d\}| < c\frac{t}{d\log u}$$

so that

$$\sum\nolimits_2 \ll \sum_{d\in\mathcal{D}_2} \frac{t}{d\log u} \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2 \leqq$$

(30) $$\leqq \frac{t}{\log u} \sum_{\max(u,t/v)\leqq d\leqq t/u} \frac{1}{d} \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2 \ll$$

$$\ll \frac{t}{\log u} \sum_{j=j_1}^{j_2} \frac{1}{2^j} \sum_{2^{j-1}<d\leqq 2^j} \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2$$

where the integers j_1, j_2 are defined by $2^{j_1-1} < \max(u, \frac{t}{v}bigr) \leqq 2^{j_1}$, $2^{j_2} \leqq 2\frac{t}{u} < 2^{j_2+1}$, respectively. The last double sum can be estimated by using (1) with $\delta \geqq (2^{-j})^2 = 2^{-2j}$. Then we obtain from (30) that

$$\sum\nolimits_2 \ll \frac{t}{\log u} \sum_{j=j_1}^{j_2} \frac{1}{2^j}(2^{2j}+N)Z \ll$$

(31) $$\ll \frac{t}{\log u}\left(2^{j_2}+N2^{-j_1}\right)Z \ll \frac{t}{\log u}\left(\frac{t}{u}+N\left(\max\left(u,\frac{t}{v}\right)\right)^{-1}\right)Z =$$

$$= \frac{t}{\log u}\left(\frac{t}{u}+N\min\left(\frac{1}{u},\frac{v}{t}\right)\right)Z.$$

Finally, assume that $uv<t$ and $d\in\mathcal{D}_3$. Then

$$M_d = |\{m:\, m\in\mathcal{M},\, d\mid m\}| = |\{k:\, dk\in\mathcal{M}\}| =$$
$$= |\{k:\, k\leqq t/d,\, u\leqq p(k)\leqq P(k)\leqq v\}|.$$

Here we have $d\leqq t/v$, hence $v\leqq t/d$. Thus by Brun's sieve,

$$M_d < c_1\frac{t}{d}\prod_{p<u}\left(1-\frac{1}{p}\right)\prod_{v<p\leqq t/d}\left(1-\frac{1}{p}\right) <$$
$$< c_2\frac{t}{d}\frac{\log v}{\log t/d\,\log u}.$$

As in the estimate of $\sum_2$, we define the integers j_3, j_4 by $2^{j_3-1} < u \leqq 2^{j_3}$ and $2^{j_4} \leqq 2\frac{t}{v} < 2^{j_4+1}$, respectively, and we use (1):

$$\sum\nolimits_3 \ll \sum_{d\in\mathcal{D}_3} \frac{t\log v}{d\log t/d\,\log u} \sum_{\substack{0<b<d\\(b,d)=1}} \left|T\left(\frac{b}{d}\right)\right|^2 \leqq$$

$$\leqq \frac{t \log v}{\log u} \sum_{u<d\leqq t/v} \frac{1}{d \log t/d} \sum_{\substack{0<b<d \\ (b,d)=1}} \left| T\left(\frac{b}{d}\right)\right|^2 \ll$$

$$\ll \frac{t \log v}{\log u} \sum_{j=j_3}^{j_4} \frac{1}{2^j \log t/2^j} \sum_{2^{j-1}<d\leqq 2^j} \sum_{\substack{0<b<d \\ (b,d)=1}} \left| T\left(\frac{b}{d}\right)\right|^2 \ll$$

$$\ll \frac{t \log v}{\log u} \sum_{j=j_3}^{j_4} \frac{1}{2^j \log t/2^j} (2^{2j} + N) Z \ll$$

$$\ll \frac{t \log v}{\log u} \left(\frac{2^{j_4}}{\log t/2^{j_4}} + \frac{N}{2^{j_3} \log t/2^{j_3}} \right) Z \ll$$

$$\ll \frac{t \log v}{\log u} \left(\frac{t/v}{\log v} + \frac{N}{u \log t/u} \right) Z,$$

hence, in view of $u^2 \leqq t$,

$$\sum\nolimits_3 \ll \left(\frac{t^2}{v \log u} + \frac{Nt \log v}{u \log u \log t} \right) Z. \tag{32}$$

It follows from (26), (29), (31) and (32) that for $uv \geqq t$ we have

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 = \sum\nolimits_1 + \sum\nolimits_2 \ll$$

$$\ll (t^2 + N)Z + \frac{t}{\log u} \left(\frac{t}{u} + \frac{N}{u} \right) Z \ll \left(t^2 + \frac{Nt}{u \log u} \right) Z$$

(since $t \leqq N$ and $u^2 \leqq t$), while for $uv < t$ we have

$$\sum_{m\in\mathcal{M}} m \sum_{h=0}^{m-1} \left(Z(m,h) - \frac{Z}{m} \right)^2 = \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 \ll$$

$$\ll (t^2 + N)Z + \frac{t}{\log u} \left(\frac{t}{u} + \frac{Nv}{t} \right) Z + \left(\frac{t^2}{v \log u} + \frac{Nt \log v}{u \log u \log t} \right) Z =$$

$$= \left(\left(t^2 + \frac{t^2}{u \log u} + \frac{t^2}{v \log u} \right) + \left(N + \frac{Nv}{\log u} \right) + \frac{Nt \log v}{u \log u \log t} \right) Z \ll$$

$$\ll \left(t^2 + \frac{Nv}{\log u} + \frac{Nt \log v}{u \log u \log t} \right) Z$$

and this completes the proof of (14).

9. (iv) Let us write $\pi(v)-\pi(u)=s$. If $k>s$, then $\mathcal{M}=\emptyset$ so that (15) is trivial. Thus we may assume that $k\leqq s$. We write

$$\mathcal{D}_r=\left\{d\colon d\mid \prod_{u<p\leqq v} p,\ \nu(d)=r\right\}.$$

Then clearly,

$$d\leqq v^r \text{ for } d\in\mathcal{D}_r. \tag{33}$$

Furthermore, for $d\in\mathcal{D}_r$ $(1\leqq r\leqq k)$ we have

$$\begin{aligned}M_d&=|\{m\colon d\mid m, m\in\mathcal{M}\}|=|\{n\colon dn\in\mathcal{M}\}|=\\&=\left|\left\{n\colon n\mid d^{-1}\prod_{u<p\leqq v} p,\ \nu(n)=k-\nu(d)=k-r\right\}\right|.\end{aligned}$$

Here n has $k-r$ distinct prime factors which can be selected from the $s-r$ distinct prime factors of $d^{-1}\prod\limits_{u<p\leqq v} p$, so that

$$M_d=\binom{s-r}{k-r}=\binom{s-r}{s-k}\quad(\text{for } d\in\mathcal{D}_r).$$

Thus by (11), we have

$$\begin{aligned}\sum_{m\in\mathcal{M}} m\sum_{h=0}^{m-1}\left(Z(m,h)-\frac{Z}{m}\right)^2&=\sum_{d=2}^{+\infty} M_d\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2=\\&=\sum_{r=1}^{k}\sum_{d\in\mathcal{D}_r}\binom{s-r}{s-k}\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2=\\&=\sum_{r=1}^{k}\binom{s-r}{s-k}\sum_{d\in\mathcal{D}_r}\sum_{\substack{0<b<d\\(b,d)=1}}\left|T\left(\frac{b}{d}\right)\right|^2.\end{aligned}$$

In view of (33), the last double sum can be estimated by (2) with $\delta=v^{-2r}$ so that

$$\sum_{m\in\mathcal{M}} m\sum_{h=0}^{m-1}\left(Z(m,h)-\frac{Z}{m}\right)^2\leqq\sum_{r=1}^{k}\binom{s-r}{s-k}(v^{2r}+N)Z=$$

$$(34)\qquad = \left(\sum_{r=1}^{k}\binom{s-r}{s-k}v^{2r} + N\sum_{j=0}^{k-1}\binom{s-k+j}{s-k}\right) Z =$$

$$= \left(\sum_{r=1}^{k}\binom{s-r}{s-k}v^{2r} + N\binom{s}{s-k+1}\right) Z.$$

Writing $A_r = \binom{s-r}{s-k}v^{2r}$, for $2 \leqq r \leqq k$ we have

$$\frac{A_{r-1}}{A_r} = \frac{(s-r+1)}{(k-r+1)v^2} < \frac{s}{v^2} < \frac{1}{v},$$

hence

$$(35)\qquad \sum_{r=1}^{k} A_r < A_k \sum_{j=0}^{+\infty} v^{-j} = v^{2k}\frac{1}{1-1/v} < 2v^{2k}.$$

(15) follows from (34) and (35).

REFERENCES

[1] Balog, A. and Sárközy, A., On sums of sequences of integers, I, *Acta Arith.* **44** (1984), 73–86. *MR* **86g**: 11056a

[2] Elliott, P.D.T.A., Sums and differences of additive arithmetic functions in mean square, *J. Reine Angew. Math.* **309** (1979), 21–54. *MR* **80i**: 10071

[3] Elliott, P.D.T.A. and Sárközy, A., The distribution of the number of prime divisors of sums $a+b$, *J. Number Theory* **29** (1988), 94–99. *MR* **89f**: 11134

[4] Erdős, P. and Sárközy, A., On divisibility properties of integers of the form $a+a'$, *Acta Math. Hungar.* **50** (1987), 117–122. *MR* **88i**: 11065

[5] Gallagher, P. X., Sieving by prime powers, *Acta Arith.* **24** (1973/74), 491–497. *MR* **49** #2613

[6] Johnsen, J., On the large sieve method in $GF(q,x)$, *Mathematika* **18** (1971), 172–184. *MR* **46** #1761

[7] Montgomery, H. L., A note on the large sieve, *J. London Math. Soc.* **43** (1968), 93–98. *MR* **37** #184

[8] Montgomery, H. L., *Topics in multiplicative number theory*, Lecture notes in mathematics, Vol. 227, Springer-Verlag, Berlin–New York, 1971. *MR* **49** #2616

[9] Montgomery, H. L., The analytic principle of the large sieve, *Bull. Amer. Math. Soc.* **84** (1978), 547–567. *MR* **57** #5931

[10] Pomerance, C., Sárközy, A. and Stewart, C. L., On divisors of sums of integers, III, *Pacific J. Math.* **133** (1988), 363–379. *MR* **89j**: 11090

[11] Salerno, S. and Viola, C., Sieving by almost-primes, *J. London Math. Soc.* (2) **14** (1976), 221–234. *MR* **58** #16568

(Received February 1, 1988)

MTA MATEMATIKAI KUTATÓINTÉZETE
P.O. BOX 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 97–118

APPROXIMATION OF CONVEX DISCS BY POLYGONS: THE PERIMETER DEVIATION

A. FLORIAN

Dedicated to my friend Professor L. Fejes Tóth on his 75^{th} birthday

Abstract

We consider the class of all convex discs with areas and perimeters bounded by given constants. Which disc of this class has the least possible perimeter deviation from a convex k-gon? By perimeter deviation we mean always the metric defined in [5, p. 135] and denoted there by $\delta^1(C, D)$. Some of its properties are developed in [7] and, independently and in a different way, in [4]. We shall answer this question and discuss some related results.

1. Introduction

By a *convex disc* we mean a convex compact subset of the Euclidean plane with interior points. We shall concern ourselves with the approximation of convex discs by convex polygons. One method of measuring the deviation between two discs is given by their perimeter deviation. Let $p(M)$ denote the perimeter of the set M. If X and Y are convex discs, the *perimeter deviation between X and Y* is defined by

(1) $$\delta^P(X,Y) = 2p([X,Y]) - p(X) - p(Y),$$

where $[X,Y]$ stands for the convex hull of $X \cup Y$. This distance function, introduced in a different way in [5, p. 135], is a metric on the class of all compact convex non-empty subsets of the plane. Its properties are discussed to a certain extent in [4] and [7]. It should be noted that δ^P differs substantially from the perimeter deviation used in [1] which is not a metric. $\delta^P(X,Y)$ may be considered a counterpart to the area deviation between X and Y, defined by

(2) $$\begin{aligned}\delta^A(X,Y) &= a(X \cup Y) - a(X \cap Y) = \\ &= a(X) + a(Y) - 2a(X \cap Y),\end{aligned}$$

where $a(M)$ denotes the area of the set M.

In Sections 2 and 3 we shall use a and p to denote positive constants satisfying the isoperimetric inequality

(3) $$\frac{p^2}{a} \geqq 4\pi.$$

1991 *Mathematics Subject Classifications.* Primary 52A40; Secondary 52A10.
Key words and phrases. Approximation of convex discs, perimeter deviation.

Akadémiai Kiadó, Budapest

Let $\mathcal{C}(a,p)$ be the class of all convex discs with area not less than a and perimeter not greater than p. A convex polygon with at most k sides is called a k-gon. By a regular k-gon we mean a regular polygon with exactly k sides. Let $\mathcal{P}_k$ denote the class of all k-gons. A measure for the closeness of the approximation of k-gons to discs from $\mathcal{C}(a,p)$ is given by

$$\Delta^P(a,p,k) = \inf \delta^P(C,P), \tag{4}$$

where the infimum is taken over all $C \in \mathcal{C}(a,p)$ and all $P \in \mathcal{P}_k$. This function is interesting only in the case when

$$\frac{p^2}{a} < 4k \tan\frac{\pi}{k}, \tag{5}$$

which means that p is less than the perimeter of a regular k-gon of area a. Otherwise we have $\mathcal{C}(a,p) \cap \mathcal{P}_k \neq \emptyset$, so that $\Delta^p(a,p,k) = 0$.

In Section 2 we shall find the infimum of $p([C,P])$, taken over all $C \in \mathcal{C}(a,p)$ and all k-gons P of given perimeter. In Section 3 this result will be applied to determine those members of $\mathcal{C}(a,p)$ and $\mathcal{P}_k$ for which $\delta^P(C,P)$ is minimal. The corresponding problem involving the area deviation in place of the perimeter deviation is solved in [3].

2. A minimum problem

Let $\mathcal{P}_k(p_0)$ denote the class of all k-gons of perimeter p_0. In this section we shall deal with the following

PROBLEM. *Find a member of $\mathcal{C}(a,p)$ and a member of $\mathcal{P}_k(p_0)$ such that the convex hull of their union has the least possible perimeter.*

Accordingly we introduce the function

$$m(a,p;k,p_0) = \min p([C,P]), \tag{6}$$

where the minimum is to be taken over all discs $C \in \mathcal{C}(a,p)$ and all k-gons $P \in \mathcal{P}_k(p_0)$. The existence of the minimum follows from the Blaschke selection theorem. Obviously, m is an increasing function of p_0.

In certain cases, the solution to our problem can be deduced from two previous results, which we are going to recollect for the reader's convenience.

Let P^* be a regular k-gon and let C^* be a convex disc that is bounded by k congruent circular arcs, each joining two consecutive vertices of P^*. We call C^* a *regular arc-sided k-gon* with *kernel* P^*. If (5) is satisfied it can be shown (see [2]) that there is exactly one regular arc-sided k-gon C^* with area a and perimeter p. Let 2φ be the central angle of the circular arcs bounding C^*, where $0 < \varphi \leqq \pi/k$. We define a function $\Phi(q)$ by the parametric equations

$$(7) \qquad \Phi(q) = \frac{\varphi^2}{\sin^2 \varphi}, \quad q = \frac{\varphi - \sin\varphi \cos\varphi}{\sin^2 \varphi} \quad (0 < \varphi \leqq \pi/k)$$

for $0 < q \leqq \bar{q}$, where $\bar{q}$ corresponds to $\varphi = \pi/k$, and put $\Phi(0) = 1$. Elementary calculation yields the equation

$$(8) \qquad \frac{p^2}{a} = 4k \sin\frac{\pi}{k} \frac{\Phi(q)}{\cos(\pi/k) + q\sin(\pi/k)},$$

which has a single root $q \in (0, \bar{q}]$. Note that $p(P^*) = (p \sin\varphi)/\varphi = p(\Phi(q))^{-1/2}$. It will be convenient to consider P^* itself as a degenerate regular arc-sided k-gon with kernel P^*.

Let the function $g_1(a,p,k)$ be defined by

$$(9) \qquad g_1(a,p,k) = \begin{cases} \frac{p}{\sqrt{\Phi(q)}} & \text{if } \frac{p^2}{a} < 4\pi t \\ p & \text{if } \frac{p^2}{a} \geqq 4\pi t, \end{cases}$$

where $t = t(k) = (k/\pi)\tan(\pi/k)$, and $\Phi(q)$ is given by (7). Let C be a disc from $\mathcal{C}(a,p)$ and P a k-gon with $P \subset C$. It was proved in [2] that

$$(10) \qquad p(P) \leqq g_1(a,p,k).$$

If $p^2/a \leqq 4\pi t$ and $p(P) = g_1(a,p,k)$, then C is a (possibly degenerate) regular arc-sided k-gon of area a and perimeter p, and P is the kernel of C.

Let C be a convex disc of area not less than a, and let P be a k-gon from $\mathcal{P}_k(p_0)$ with $P \subset C$. Since g_1 is a strictly increasing function of p, for $p \geqq \sqrt{4\pi a}$, it follows from (9) and (10) that

$$(11) \qquad p(C) \geqq G(a,k,p_0),$$

where

$$(12) \qquad G(a,k,p_0) = \begin{cases} \sqrt{4\pi a} & \text{if } p_0 < \sqrt{\frac{4a}{\pi}}k \sin\frac{\pi}{k} \\ p_0\sqrt{\Phi(q)} & \text{if } \sqrt{\frac{4a}{\pi}}k \sin\frac{\pi}{k} \leqq p_0 < \sqrt{4\pi t a} \\ p_0 & \text{if } \sqrt{4\pi t a} \leqq p_0, \end{cases}$$

and $q = 4ka/p_0^2 - \cot(\pi/k)$. For $p_0 \in [\sqrt{4a/\pi}k\sin(\pi/k), \sqrt{4\pi t a}]$ equality holds in (11) if and only if C is a (possibly degenerate) regular arc-sided k-gon of area a, and P is the kernel of C. From (9) and (10) we further conclude that

$$(13) \qquad g_1(a,p,k) \leqq p$$

and

$$m(a,p;k,p_0) \leqq p \quad \text{if} \quad p_0 \leqq g_1(a,p,k). \tag{14}$$

Let $\overline{P}$ be a regular k-gon, and let $\overline{C}$ be a convex disc obtained from $\overline{P}$ by rounding off the corners of $\overline{P}$ by k congruent circular arcs. We call $\overline{C}$ a *smooth regular k-gon* with *case* $\overline{P}$. It will be convenient to consider $\overline{P}$ itself as a degenerate smooth regular k-gon with case $\overline{P}$.

Let C be a disc from $\mathcal{C}(a,p)$ and $P \in \mathcal{P}_k$ with $C \subset P$. By applying the isoperimetric inequality for k-gons to the formulae (17) and (18) in [3] we obtain

$$p(P) \geqq g_2(a,p,k), \tag{15}$$

where

$$g_2(a,p,k) = \begin{cases} t\left(p - \sqrt{(p^2 - 4a\pi)(1-t^{-1})}\right) & \text{if } \frac{p^2}{a} < 4\pi t \\ \sqrt{4\pi t a} & \text{if } \frac{p^2}{a} \geqq 4\pi t. \end{cases} \tag{16}$$

Equality holds in (15) if and only if C is a smooth regular k-gon of area a and perimeter p, and P is the case of $C(p^2/a < 4\pi t)$, or $C = P$ is a regular k-gon of area $a(p^2/a \geqq 4\pi t)$. By (15) we have

$$\begin{aligned} m(a,p;k,p_0) &= p_0 \quad \text{if } p_0 \geqq g_2(a,p,k), \text{ and} \\ m(a,p;k,p_0) &> p_0 \quad \text{if } p_0 < g_2(a,p,k). \end{aligned} \tag{17}$$

Returning to our problem, we distinguish the cases $p^2/a \geqq 4\pi t$ and $p^2/a < < 4\pi t$.

Case (a). $p^2/a \geqq 4\pi t$.
From (9) and (16) we see that

$$g_2(a,p,k) \leqq g_1(a,p,k).$$

Let $p_0 < g_2(a,p,k)$, and let C and P be members of $\mathcal{C}(a,p)$ and $\mathcal{P}_k(p_0)$ such that

$$p([C,P]) = m(a,p;k,p_0). \tag{18}$$

Inequalities (15) and (14) imply that $C \not\subset P$ and

$$p([C,P]) \leqq p. \tag{19}$$

We shall see in the proof of Theorem 1 (Lemma 6) that, whenever (18) together with the assumptions $P \not\subset C$ and $C \not\subset P$ is satisfied, then $p(C) = p$, whence $p([C, P]) > p$. Still, by (19) this is impossible. Thus we conclude that $P \subset C$, and from (11) and (18) it follows that

$$p([C, P]) = G(a, k, p_0). \tag{20}$$

The results in (17) and (20) can be summarized in

REMARK 1. If $p^2/a \geqq 4\pi t$, we have

$$m(a, p; k, p_0) = G(a, k, p_0). \tag{21}$$

From the suppositions (18) and $\sqrt{4a/\pi k}\, \sin(\pi/k) \leqq p_0 \leqq \sqrt{4\pi t a}$ it follows that C is a (possibly degenerate) regular arc-sided k-gon of area a, and P is the kernel of C.

Case (b). $p^2/a < 4\pi t$.

In view of (9) and (16) we have the inequality

$$g_1(a, p, k) < g_2(a, p, k), \tag{22}$$

which is contrary to case (a). Let $p_0 \leqq g_1(a, p, k)$, and let $C \in \mathcal{C}(a, p)$ and $P \in \in \mathcal{P}_k(p_0)$ satisfy (18). By repeating the argument used in case (a) we again come to the conclusion that P is contained in C and that (20) holds in case (b) as well. Taking into account that $g_2(a, p, k) \geqq \sqrt{4\pi t a}$ we see from (12) and (17) that (21) continues to hold for $p_0 \geqq g_2(a, p, k)$.

REMARK 2. If $p^2/a < 4\pi t$ and either $p_0 \leqq g_1(a, p, k)$ or $p_0 \geqq g_2(a, p, k)$ we have

$$m(a, p; k, p_0) = G(a, k, p_0). \tag{23}$$

If C and P satisfy (18) and if $p_0 = g_2(a, p, k)$, then C is a smooth regular k-gon of area a and perimeter p, and P is the case of C. If C and P satisfy (18) and if $\sqrt{4a/\pi k}\, \sin(\pi/k) \leqq p_0 \leqq g_1(a, p, k)$, then C is a regular arc-sided k-gon of area a, and P is the kernel of C.

We now proceed to find the minimum of $p([C, P])$ with $C \in \mathcal{C}(a, p)$ and $P \in \mathcal{P}_k(p_0)$ in the more complicated case when

$$g_1(a, p, k) < p_0 < g_2(a, p, k).$$

To describe the extremal configuration we consider the outer parallel domain C of a regular arc-sided k-gon C^* at some distance r_1. If C^* is not a circle, then C is bounded by k equal circular arcs of radius r_1 and k equal circular arcs of radius $r_2 > r_1$. The tangents at the end points of the i-th arc of radius r_1 intersect at a point, say, A_i. The points $A_1, \dots, A_k$ are the vertices of a regular k-gon P which we call the *k-gon associated with C* (Fig. 1). By a k-gon associated with a circle C we mean any regular k-gon concentric with C.

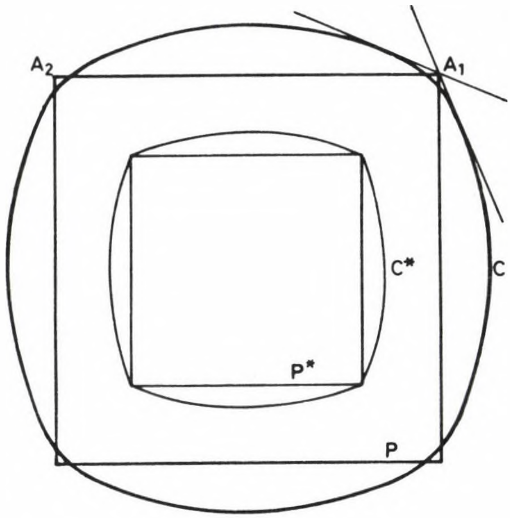


Fig. 1

THEOREM 1. *Let*

(i) $$g_1(a,p,k) < p_0 < g_2(a,p,k),$$

and let $C \in \mathcal{C}(a,p)$ and $P \in \mathcal{P}_k(p_0)$ be such that

(ii) $$p([C,P]) = m(a,p;k,p_0).$$

Then C is an outer parallel domain of a regular arc-sided k-gon and P is the k-gon associated with C. Furthermore, C has area a and perimeter p.

We shall, in fact, prove Theorem 1, making the weaker assumptions (ii) and

(iii) $$C \not\subset P, \quad P \not\subset C$$

instead of (i) and (ii). Theorem 1, in union with Remarks 1 and 2, solves the problem raised at the beginning of this section. Observe that a regular arc-sided k-gon and its kernel (as well as a smooth regular k-gon and its case) may be regarded as a degenerate parallel domain of a regular arc-sided k-gon and the associated k-gon.

PROOF of Theorem 1. Let us assume that $C \in \mathcal{C}(a,p)$ and $P \in \mathcal{P}_k(p_0)$ satisfy the conditions (ii) and (iii). We will develop the properties of C and P in a series of nineteen lemmas, the last showing that C and P correspond with the statement of our theorem.

By (iii), there is a vertex of P outside C. We now show

LEMMA 1. *In the exterior of C there is a vertex of P which is an extreme point of $[C,P]$.*

PROOF. From assumption (iii) and the Krein–Milman theorem it follows that there is an extreme point of $[C,P]$ outside C. Such a point is necessarily a vertex of P. $\square$

LEMMA 2. *If P' is a k-gon with $p(P') > p(P)$, then $p([C, P']) > p([C, P])$.*

PROOF. Since $p([C, P']) < p([C, P])$ clearly contradicts (ii) we can suppose that $p([C, P']) = p([C, P])$. Thus $P' \not\subset C$, so that there exists a vertex A_1' of P' outside C, which is an extreme point of $[C, P']$. Let P'' be a k-gon obtained from P' by a small displacement of A_1' into P'. Then $p(P'') > p(P)$, yet $p([C, P'']) < p([C, P])$, which is impossible. □

LEMMA 3. *P has exactly k vertices (and therefore interior points).*

PROOF. Otherwise we choose a point $U \in C \setminus P$ and consider the k-gon $P' = [P, \{U\}]$. Since $p(P') > p(P)$ and $p([C, P']) = p([C, P])$ we have a contradiction to Lemma 2. □

We denote the vertices of P in the anti-clockwise sense by $A_1, \ldots, A_k$, $A_{k+1} = A_1$ and write $A \vee B$ for the line joining the distinct points A and B.

LEMMA 4. *Every side $A_i A_{i+1}$ of P contains the intersection of C with $A_i \vee A_{i+1}$.*

PROOF. Otherwise we choose a point $U \in (C \cap (A_i \vee A_{i+1})) \setminus A_i A_{i+1}$ and conclude as before. Observe that $P' = [P, \{U\}]$ is a k-gon. □

COROLLARY 1. *No vertex of P lies in the interior of C.*

LEMMA 5. *C and P have interior points in common.*

PROOF. Suppose that this is not true. Then there is a line t supporting P and separating P from C. t intersects $[C, P]$ in a segment UV and divides $[C, P]$ into two convex discs C' and P' containing C and P, respectively. Hence

$$[C, P] = C' \cup P', \quad C' \cap P' = UV. \tag{24}$$

We shall denote the length of a segment or an arc by $|.|$. Because UV is a proper subset of P', there exists a point W on t such that $2|UW| = p(P')$, and V is between U and W. The points U and V divide the boundary of C' into the segment UV and an arc denoted by $\overset{\frown}{VU}$. From the definition of C' it follows that

$$C' = [C, UV].$$

Thus

$$\begin{aligned}[C, UW] &= [C, UV \cup VW] = [C', VW] = [UV \cup \overset{\frown}{VU}, VW] = \\ &= \operatorname{conv}(WV \cup \overset{\frown}{VU} \cup UW).\end{aligned} \tag{25}$$

The segment $UW = Q$ may be considered a k-gon of perimeter

$$p(Q) = 2|UW| = p(P') \geqq p(P). \tag{26}$$

By (24) we have

$$\begin{aligned} p([C,P]) &= p(C') + p(P') - 2|UV| = \\ &= |UV| + |\overset{\frown}{VU}| + 2|UW| - 2|UV| = \\ &= |WV| + |\overset{\frown}{VU}| + |UW|, \end{aligned}$$

and this is the length of the closed curve composed of WV, $\overset{\frown}{VU}$ and UW. This length is not less than $p(\mathrm{conv}(WV \cup \overset{\frown}{VU} \cup UW))$ (see [6]). Using (25) we conclude that

$$p([C,P]) \geqq p([C,Q])$$

which together with (26) contradicts either Lemma 2 or Lemma 3. □

Supposition (iii) and Lemma 5 imply

COROLLARY 2. *The boundary of P contains interior points of C.*

LEMMA 6. *$a(C) = a$ and $p(C) = p$.*

PROOF. Because $C \not\subset P$, there is an extreme point E of $[C,P]$ outside P. Obviously, E is a boundary point of C. Suppose now that $a(C) > a$. Let the origin be an inner point of C and choose a positive $\lambda < 1$ so that $\lambda C = C'$ satisfies $a(C') > a$. From $E \notin C'$ it follows that $[C',P] \neq [C,P]$, so that $p([C',P]) < p([C,P])$. Since $C' \in \mathcal{C}(a,p)$, we have a contradiction to assumption (ii).

Supposing $p(C) < p$ we choose $\varrho > 0$ so that $P \not\subset C_\varrho$ and $p(C_\varrho) < p$, where C_ϱ is the outer parallel domain of C at distance ϱ. The set $C' = C_\varrho \cap [C,P]$ is a member of $\mathcal{C}(a,p)$. Since C' contains C we see that $[C',P] = [C,P]$. Let A_1 be a vertex of P outside C. There are points of $[C,\{A_1\}]$ belonging to $C_\varrho \setminus C$. Thus $C' \neq C$ and $a(C') > a$. Since C' and P satisfy condition (iii), we have a contradiction to the part of Lemma 6 proved before. □

LEMMA 7. *If a side of P intersects the interior of C, then both adjacent sides contain points of C.*

PROOF. Suppose that the side $A_{i-1}A_i$ intersects the interior of C and that $A_iA_{i+1} \cap C = \emptyset$. By Lemma 4, C and A_{i-1} are contained in the same open half-plane H bounded by $A_i \vee A_{i+1}$. Let b denote the line bisecting the outer angle of P at the vertex A_i. A small displacement along b carries A_i into a point A_i' in H. Denoting the k-gon $A_1 \ldots A_i' \ldots A_k$ by P' we have

$$p(P') > p(P), \tag{27}$$

because A_i' is outside the ellipse with the foci A_{i-1}, A_{i+1} and passing through A_i. We distinguish two cases:

(a) $C \cap b = \emptyset$. t and t' are half-lines starting from A_i and A_i' and supporting C at points B and B' that are separated from P by $A_{i-1} \vee A_i$ (Fig. 2). Then

$$p([C,P]) - p([C,P']) = |A_{i+1}A_i| + |A_iB| + |\overset{\frown}{BB'}| - \\ - |A_{i+1}A_i'| - |A_i'B'|.$$

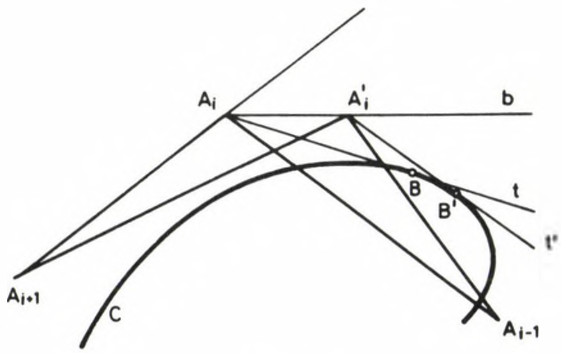


Fig. 2

Since $|A_iB| + |\overset{\frown}{BB'}| \geqq |A_iB| + |BB'| \geqq |A_iB'|$, we have

(28) $$p([C,P']) \leqq p([C,P]),$$

provided that

$$|A_{i+1}A_i| + |A_iB'| > |A_{i+1}A_i'| + |A_i'B'|.$$

But this inequality is true because A_i' is an inner point of the ellipse with the foci A_{i+1}, B' and passing through A_i.

(b) $C \cap b \neq \emptyset$. Since $A_i' \in [C,P]$, inequality (28) holds also in this case.

(27) together with (28) contradicts Lemma 2, and Lemma 7 is established. □

LEMMA 8. *Every side of P intersects C.*

PROOF. By Corollary 2 we may assume that A_1A_2 intersects the interior of C. We suppose that $A_{i-1}A_i$ contains points of C and that A_iA_{i+1} does not. In view of Lemma 7, $A_{i-1}A_i$ contains no inner point of C, so that $A_{i-1} \vee A_i$ supports C. We displace A_i on this line toward A_{i-1} into A_i' such that $|A_iA_i'|$ is small and denote the k-gon $A_1 \dots A_i' \dots A_k$ by P'. Then $p(P') < p(P)$ and

(29) $$p([C,P]) - p([C,P']) = p(P) - p(P').$$

Lemma 7 implies that $C \cap A_kA_1 \neq \emptyset$. We may assume without loss of generality that $A_k \vee A_1$ supports C; otherwise we consider $A_{k-1}A_k$ instead of

A_kA_1. By displacing A_1 on $A_k \vee A_1$ into the point $A_1' \notin P'$ we obtain the new k-gon $A_1' \dots A_i' \dots A_k = P''$ with $p(P'') > p(P')$. We choose A_1' such that $p(P'') = p(P)$. Let B and B' be two points on the boundary of C, but not on $A_k \vee A_1$, such that $A_1 \vee B$ and $A_1' \vee B'$ support C (if $A_1 \in \operatorname{bd} C$, take $B = A_1$). Then

$$p(P'') - p(P') = |A_1A_1'| + |A_1'A_2| - |A_1A_2| \tag{30}$$

and

$$\begin{aligned} p([C,P'']) - p([C,P']) &= |A_1A_1'| + |A_1'B'| - |A_1B| - |\overset{\frown}{BB'}| \leqq \\ &\leqq |A_1A_1'| + |A_1'B'| - |A_1B'|. \end{aligned} \tag{31}$$

Note that the quadrangle $A_1A_1'B'A_2$ is convex if $|A_iA_i'|$ is sufficiently small. The triangle inequality shows that

$$|A_1'A_2| + |A_1B'| > |A_1A_2| + |A_1'B'|,$$

so that by (30) and (31)

$$p(P'') - p(P') > p([C,P'']) - p([C,P']). \tag{32}$$

By combining (32) with (29) and observing that $p(P'') = p(P)$ we obtain

$$p([C,P]) > p([C,P'']).$$

This contradicts assumption (ii) and the lemma is proved. □

Let A_i be a vertex of P outside C. Let T_i and U_i be distinct points on the boundary of C such that $A_i \vee T_i$ and $A_i \vee U_i$ support C, and $A_iT_i \cap \cap C = \{T_i\}$, $A_iU_i \cap C = \{U_i\}$. Lemma 8 implies that A_iT_i and A_iU_i are boundary segments of $[C,P]$. T_i and U_i divide the boundary of C into two arcs. Let $\overset{\frown}{T_iU_i}$ denote that arc which lies between A_i and the segment T_iU_i.

Lemma 9. *$\overset{\frown}{T_iU_i}$ is either a circular arc or a line segment.*

Proof. It suffices to show that any subarc $\overset{\frown}{VW}$ of $\overset{\frown}{T_iU_i}$, where V and W are different from T_i and U_i, is either a circular arc or a segment. Since $\overset{\frown}{VW}$ lies in the interior of $[C,P]$, it has a positive distance ϱ from $\operatorname{bd}[C,P]$. We cover $\overset{\frown}{VW}$ by a finite number of subarcs, say $b_1, \dots, b_r$, such that each overlaps the following and has a length less than ϱ. Lemma 9 is proved if we can show that every b_i is either a circular arc or a segment. Suppose that $b_i = \overset{\frown}{V'W'}$ is neither the one nor the other. b_i is replaced by the circular arc $\overline{b_i}$ with the same length and the same endpoints and lying on the same side of the chord $V'W'$. As $|\overline{b_i}| < \varrho$, $\overline{b_i}$ lies in the interior of $[C,P]$. Thus the (possibly non-convex) disc $D = (C \setminus \operatorname{conv} b_i) \cup \operatorname{conv} \bar{b}_i$ is a subset of $[C, P]$ with $p(D) = p$. In view of the well-known isoperimetric property of the circular segments we note that $a(D) > a$. Hence

$$C' = \operatorname{conv} D \in \mathcal{C}(a, p), \quad [C', P] \subset [C, P]. \tag{33}$$

By (ii) we have $[C', P] = [C, P]$. Using (iii) and Lemma 6, we see that $C' \not\subset P$ and $P \not\subset C'$, so that $a(C') = a$. The contradiction to $a(D) > a$ completes the proof. □

The following lemma states that C has a smooth boundary.

LEMMA 10. *Through every boundary point of C there passes exactly one support line.*

PROOF. Suppose that B is a non-regular boundary point of C. Let s be a support line through B different from the limiting support lines. Let b be a subarc of $\operatorname{bd} C$ covering B and such that its chord is parallel to s. On a circular arc (possibly a segment) of type $\overset{\frown}{T_i U_i}$ considered in Lemma 9 we choose a subarc b' such that b and b' have chords of the same length. By cutting the sets $\operatorname{conv} b$ and $\operatorname{conv} b'$ off from C and interchanging their positions we obtain a set D with $p(D) = p(C)$ and $a(D) = a(C)$. If b is sufficiently small, D is non-convex and contained in $[C, P]$. Thus $C' = \operatorname{conv} D$ satisfies (33) and, moreover, $a(C') > a$. Then we can conclude as in the proof of Lemma 9. □

Lemma 10 shows that in Lemma 9 the arc $\overset{\frown}{T_i U_i}$ cannot be a line segment.

COROLLARY 3. *The subarcs of type $\overset{\frown}{T_i U_i}$ of the boundary of C, considered in Lemma 9, are circular arcs. We shall refer to them as arcs of type I.*

LEMMA 11. *All arcs of type I have the same radius.*

PROOF. Suppose that two arcs of type I have different radii. Let c and c' be proper subarcs of them with chords of equal lengths. By exchanging the positions of the circular segments $\operatorname{conv} c$ and $\operatorname{conv} c'$ we obtain from C a non-convex disc D. Obviously, $p(D) = p$, $a(D) = a$, so that $C' = \operatorname{conv} D \in \mathcal{C}(a, p)$. If c and c' are sufficiently small, then $D \subset [C, P]$. The rest of the proof is the same as in Lemma 10. □

By Corollary 2, at least one side of P, say $A_{i-1}A_i$, contains interior points of C. A_{i-1} and A_i may be outside or on the boundary of C. If A_i is outside C, let T_i and U_i be two points on the boundary of C such that $A_i \vee T_i$ and $A_i \vee U_i$ support C and $A_i T_i \cap C = \{T_i\}$, $A_i U_i \cap C = \{U_i\}$ (see the notation used in Lemma 9). If $A_i \in \operatorname{bd} C$, put $T_i = U_i = A_i$. We consider that boundary arc of C which is separated from P by $A_{i-1} \vee A_i$. This arc contains the points U_{i-1} and T_i, and the subarc $\overset{\frown}{U_{i-1}T_i}$. Since the boundary of C is smooth, $\overset{\frown}{U_{i-1}T_i}$ is not a line segment.

LEMMA 12. *$\overset{\frown}{U_{i-1}T_i}$ is a circular arc.*

PROOF. The line t supporting C at a point X of $\overset{\frown}{U_{i-1}T_i}$ does not intersect the interior of the segment $A_{i-1}A_i$. Thus t is a support line of the convex polygon $A_1 \dots A_{i-1}XA_i \dots A_k$. That means that C and P lie in the same closed half-plane determined by t. Hence t is also a support line of $[C, P]$ and $\overset{\frown}{U_{i-1}T_i}$ is part of the boundary of $[C, P]$. Let c be the circular arc with the same length and the same endpoints as $\overset{\frown}{U_{i-1}T_i}$ and lying on the same side of $U_{i-1} \vee T_i$. Suppose that $\overset{\frown}{U_{i-1}T_i} \neq c$. Similarly as in the proof of Lemma 9, we form the (possibly non-convex) set

$$D = (C \setminus \operatorname{conv} \overset{\frown}{U_{i-1}T_i}) \cup \operatorname{conv} c.$$

Then $p(D) = p(C)$ and $a(D) > a(C)$, so that $C' = \operatorname{conv} D \in \mathcal{C}(a,p)$ and $a(C') > a$. Since P and $\operatorname{conv} \overset{\frown}{U_{i-1}T_i}$ do not overlap, the set E defined by

$$E = ([C, P] \setminus \operatorname{conv} \overset{\frown}{U_{i-1}T_i}) \cup \operatorname{conv} c$$

has the properties

$$p(E) = p([C, P]), \quad D \subset E, \quad P \subset E.$$

The required contradiction follows from $[C', P] \subset \operatorname{conv} E$ and

$$p([C', P]) \leqq p(\operatorname{conv} E) \leqq p(E) = p([C, P])$$

in a similar way as in the proof of Lemma 9. □

The circular arcs of type $\overset{\frown}{U_{i-1}T_i}$ will be called *arcs of type II.* We denote the centre of the corresponding circle by N_{i-1}.

LEMMA 13. *All arcs of type II have the same radius.*

PROOF. Suppose that two arcs of type II have different radii. Let c and c' be small subarcs of them with chords of equal lengths. By exchanging the positions of $\operatorname{conv} c$ and $\operatorname{conv} c'$ we obtain from C a non-convex disc D with $p(D) = p(C)$ and $a(D) = a(C)$. Since the further argument is very similar to that used in the proof of Lemma 12, we omit the details. □

LEMMA 14. *Let r_1 and r_2 denote the radius of the arcs of type I and type II, respectively. Then*

$$r_1 \leqq r_2. \tag{34}$$

PROOF. Suppose that $r_1 > r_2$. Let c_1 and c_2 be small subarcs, with chords of equal lengths, of two arcs of type I and type II, respectively. By interchanging the positions of $\operatorname{conv} c_1$ and $\operatorname{conv} c_2$ we obtain a non-convex disc D with $a(D) = a(C)$ and $p(D) = p(C)$. Thus $C' = \operatorname{conv} D \in \mathcal{C}(a,p)$. Note that $[D,P] = [C',P]$. In view of $r_1 > r_2$, $[D,P]$ is a proper subset of $[C,P]$, which implies $p([C',P]) < p([C,P])$. But this is impossible. □

LEMMA 15. *C is strictly convex.*

PROOF. Suppose that a segment s is part of the boundary of C. We may assume that s lies either on the boundary or in the interior of $[C,P]$. Let c be a small subarc of an arc of type II. We cut $\operatorname{conv} c$ off from C and join it to s, so obtaining from C a non-convex set D with $p(D) = p(C)$ and $a(D) = a(C)$. Thus $C' = \operatorname{conv} D \in \mathcal{C}(a,p)$ and $a(C') > a$. By this process we obtain from $[C,P]$ a (possibly non-convex) set E, where $D \cup P \subset E$ and $p(E) \leqq p([C,P])$. The contradiction follows from

$$p([C',P]) \leqq p(\operatorname{conv} E) \leqq p(E) \leqq p([C,P])$$

and $a(C') > a$. □

From Corollary 3, Lemma 12 and Lemma 15 we infer

COROLLARY 4. *The boundary of C is composed of circular arcs of type I and type II.*

Let A_i be a vertex of P on the boundary of C. At least one of the sides A_iA_{i-1} and A_iA_{i+1} intersects the interior of C, and the other touches C.

LEMMA 16. *The normal to the boundary of C at A_i bisects the angle $\sphericalangle A_{i-1}A_iA_{i+1}$.*

PROOF. Otherwise the ellipse with the foci A_{i-1} and A_{i+1} and passing through A_i would intersect the interior of C. Thus we could displace A_i into the interior of C without changing the perimeter of P or $[C,P]$. But this contradicts Corollary 1. In particular we see that both $A_{i-1}A_i$ and A_iA_{i+1} intersect the interior of C. □

The Lemmas 12, 13 and 16 imply

COROLLARY 5. *If A_i is a vertex of P on the boundary of C, then A_i is an inner point of a circular arc of radius r_2 that forms part of* $\operatorname{bd} C$ *and* $\operatorname{bd}[C,P]$.

Let A_i be a vertex of P outside C. Following the notation used in Lemma 9 we consider two points T_i and U_i on the boundary of C such that $A_i \vee T_i$ and $A_i \vee U_i$ are tangents of C. $\overset{\frown}{T_iU_i}$ is a (circular) arc of type I. We denote the centre of the corresponding circle by M_i.

LEMMA 17. *The line $A_i \vee M_i$ bisects the angle $\sphericalangle A_{i-1}A_iA_{i+1}$.*

PROOF. To simplify the notation, we take $i = 1$. Suppose, on the contrary, that M_1 and A_2 are on the same side of the line bisecting $\sphericalangle A_kA_1A_2$.

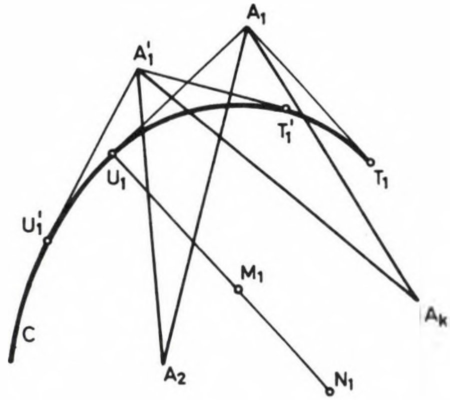


Fig. 3

Thus A_1A_2 intersects the interior of C. The line through A_1 normal to M_1A_1 determines two open half-planes, one of them containing C. We displace A_1 into this half-plane and write A_1' for the new position. The support lines to C containing A_1' intersect the boundary of C at, say, T_1' and U_1'. We can choose A_1' so that (Fig. 3)

(α) $|A_kA_1'|+|A_1'A_2|=|A_kA_1|+|A_1A_2|$,

(β) $|M_1A_1'|<|M_1A_1|$, and

(γ) T_1' lies on $\overset{\frown}{T_1U_1}$, and U_1' on $\overset{\frown}{U_1T_2}$, i.e. the arc of type II adjacent to $\overset{\frown}{T_1U_1}$.

By (α), the k-gons P and $P'=A_1'A_2\dots A_k$ have the same perimeter. Observe that $A_1'\notin[C,P]$; otherwise $[C,P']$ would be a proper subset of $[C,P]$, which implies a contradiction to assumption (ii). Thus (β) and (γ) are satisfied if A_1' is sufficiently close to A_1. To prove the lemma it suffices to show that

$$\Delta=p([C,P])-p([C,P'])>0. \tag{35}$$

We shall use the following notations: $|M_1A_1|=d$; $|M_1A_1'|=b$; N_1 is the centre of the arc $\overset{\frown}{U_1T_2}$ (by Lemma 10, the points N_1,M_1,U_1 are collinear); $|N_1A_1'|=c$; $\sphericalangle N_1M_1A_1'=\varphi$ if $r_2>r_1$ (note that $\varphi>\sphericalangle N_1M_1A_1>\frac{\pi}{2}$).

The configuration in Figure 3 is determined by the five parameters r_1, r_2, d, b and φ. Keeping fixed the other parameters, we consider Δ as a function of r_2 only, where $r_2\geqq r_1$ by Lemma 14. Elementary calculation yields

$$\begin{aligned}\Delta=&2|T_1A_1|+|\overset{\frown}{U_1U_1'}|-|\overset{\frown}{T_1T_1'}|-|T_1'A_1'|-|A_1'U_1'|=\\ =&r_2\left(\arccos\frac{r_2}{c}-\arcsin\frac{b\sin\varphi}{c}\right)-\\ &-r_1\left(2\arccos\frac{r_1}{d}-\arccos\frac{r_1}{b}+\varphi-\pi\right)+\\ &+2\sqrt{d^2-r_1^2}-\sqrt{b^2-r_1^2}-\sqrt{c^2-r_2^2}=\Delta(r_2),\end{aligned} \tag{36}$$

where

$$c^2 = b^2 + (r_2 - r_1)^2 - 2b(r_2 - r_1)\cos\varphi. \tag{37}$$

Since $c \to b$ and $\arcsin(b \sin\varphi/c) \to \pi - \varphi$ as r_2 tends to r_1, we have

$$\lim_{r_2 \to r_1} \Delta(r_2) = 2r_1 \left(\arccos \frac{r_1}{b} - \arccos \frac{r_1}{d}\right) + \\ +2\left(\sqrt{d^2 - r_1^2} - \sqrt{b^2 - r_1^2}\right).$$

By condition (β) we have $b < d$, so that

$$\lim_{r_2 \to r_1} \Delta(r_2) > 0. \tag{38}$$

This proves (35) in the case $r_2 = r_1$. If $r_2 > r_1$, we make use of (37) and obtain from (36)

$$\frac{d\Delta}{dr_2} = \arccos \frac{r_2}{c} - \arcsin \frac{b \sin\varphi}{c} + \\ + \frac{r_2 b \sin\varphi}{c^2} - \frac{r_2 - r_1 - b\cos\varphi}{c}\sqrt{1 - \frac{r_2^2}{c^2}}.$$

Observe that $\arccos(r_2/c) - \arcsin(b \sin\varphi/c) = \psi$ is the central angle of the arc $\overset{\frown}{U_1 U_1'}$. Hence

$$\frac{d\Delta}{dr_2} = \psi - \sin\psi > 0. \tag{39}$$

(35) now follows from (38), (39) and Lemma 14. □

From Corollary 2 and the Lemmas 16 and 17 we infer

COROLLARY 6. *Every side of P intersects the interior of C.*

Let us denote by $2\alpha_i$ the angle formed by the two support lines of C passing through A_i, and by $2\beta_i$ the interior angle of P at A_i. The following lemma points out that the vertices of P satisfy a certain 'condition of equilibrium'.

LEMMA 18. $\dfrac{\cos\alpha_1}{\cos\beta_1} = \ldots = \dfrac{\cos\alpha_k}{\cos\beta_k}$.

PROOF. It suffices to show that the supposition

$$\frac{\cos\alpha_1}{\cos\beta_1} > \frac{\cos\alpha_2}{\cos\beta_2} \tag{40}$$

leads to a contradiction with the assumption (ii). (40) implies that A_1 is in the exterior of C. We displace A_1 through a small distance x_1 on the bisector of $\sphericalangle T_1A_1U_1$ into the interior of P and denote the new position by A_1'. Writing P' for the k-gon $A_1'A_2\dots A_k$ and $|M_1A_1| = d_1$ we have

$$p([C,P]) - p([C,P']) =$$
$$= 2\left[\sqrt{d_1^2 - r_1^2} - \sqrt{(d_1-x_1)^2 - r_1^2} - r_1 \arccos\frac{r_1}{d_1} + r_1 \arccos\frac{r_1}{d_1 - x_1}\right].$$

Observing that $r_1 = d_1 \sin\alpha_1$, we find

$$\lim \frac{p([C,P]) - p([C,P'])}{x_1} = 2\cos\alpha_1 \tag{41}$$

as x_1 tends to 0. Writing $|A_iA_{i+1}| = s_i$, we note that P' has the same side-lengths as P except for s_k and s_1, which are replaced by $s_k' = |A_kA_1'|$ and $s_1' = |A_1'A_2|$, where

$$s_k' = \left(s_k^2 + x_1^2 - 2s_kx_1 \cos\beta_1\right)^{1/2} \tag{42}$$

and

$$s_1' = (s_1^2 + x_1^2 - 2s_1x_1 \cos\beta_1)^{1/2}$$

respectively.

We now displace A_2 through a small distance x_2 on the bisector of $\sphericalangle T_2A_2U_2$ into the exterior of P' to a point A_2'. (If $A_2 \in \operatorname{bd} C$, displace A_2 on the normal of $\operatorname{bd} C$.) By this process we obtain from P' a new k-gon $A_1'A_2'A_3\dots A_k = P''$ (Fig. 4). We take x_2 so that $p(P'') = p(P)$. P'' has the same side-lengths as P' except for s_1' and s_2, which are replaced by $s_1'' = |A_1'A_2'|$ and $s_2'' = |A_2'A_3|$. Denoting the angle $\sphericalangle A_1A_2A_1'$ by $\Delta\beta_2$ we have

$$s_1' \sin\Delta\beta_2 = x_1 \sin\beta_1$$

and

$$s_1''^2 = s_1'^2 + x_2^2 - 2s_1'x_2\cos(\pi - \beta_2 + \Delta\beta_2).$$

Some trigonometrical calculation yields

$$\begin{aligned} s_1'' &= [(s_1 - x_1\cos\beta_1 + x_2\cos\beta_2)^2 + (x_1\sin\beta_1 + x_2\sin\beta_2)^2]^{1/2},\\ s_2'' &= (s_2^2 + x_2^2 + 2s_2x_2\cos\beta_2)^{1/2}. \end{aligned} \tag{43}$$

The above condition $p(P'') = p(P)$ is equivalent to the equation

$$s_k' + s_1'' + s_2'' = s_k + s_1 + s_2. \tag{44}$$

By means of (42) and (43) it is easy to show that (44) determines x_2 as a unique continuous function $x_2(x_1)$ on some interval $0 \leqq x_1 < \xi$ with $x_2(0) = 0$.

By differentiating the left-hand side of (44) with respect to x_1 and x_2, one proves that $x_2(x_1)$ is differentiable and that $\lim x_2'(x_1) = \cos\beta_1/\cos\beta_2$ as $x_1 \to 0$. Hence

$$\lim_{x_1\to 0}\frac{x_2}{x_1} = \frac{\cos\beta_1}{\cos\beta_2}. \tag{45}$$

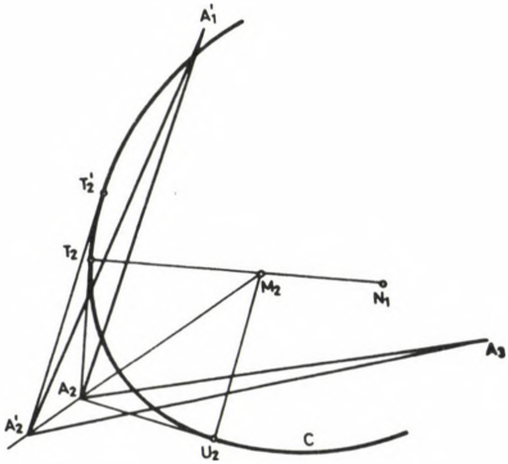


Fig. 4

Let $A_2' \vee T_2'$ be a support line of C, where $T_2' \in \overset{\frown}{U_1T_2}$. Using the notations $|M_2A_2| = d_2$ and $|N_1A_2'| = z$, we have

$$d_2 = \frac{r_1}{\sin\alpha_2}, \qquad z^2 = (d_2+x_2)^2 + (r_2-r_1)^2 + 2(r_2-r_1)(d_2+x_2)\sin\alpha_2$$

and

$$p([C,P'']) - p([C,P']) = 2|A_2'T_2'| - 2|A_2T_2| - 2|\overset{\frown}{T_2T_2'}| =$$
$$= 2\sqrt{z^2-r_2^2} - 2\sqrt{d_2^2-r_1^2} - 2r_2\left[\arccos\frac{r_2}{z} - \arcsin\frac{(d_2+x_2)\cos\alpha_2}{z}\right].$$

A straightforward calculation shows that

$$\lim \frac{p([C,P''])-p([C,P'])}{x_2} = 2\cos\alpha_2 \tag{46}$$

as x_2 tends to 0, and this continues to hold when $r_2 = r_1$ or $\alpha_2 = \pi/2$. The combination of (40), (41), (45) and (46) leads to the conclusion that $p([C,P]) > p([C,P''])$ for sufficiently small x_1. Since $p(P'') = p(P)$, this is impossible and the proof of the lemma is complete. □

Since at least one vertex of P is outside C, it follows from Lemma 18 that $\cos\alpha_i > 0$, for $i = 1, \ldots, k$. This implies

COROLLARY 7. *All vertices of P are outside C.*

We displace the single vertex A_1 of P on the bisector of the angle $\sphericalangle T_1A_1U_1$ through a small distance x_1 in one or the other direction, obtaining the k-gon $A_1'A_2\ldots A_k = P'$. From (42) and two similar equations we infer that $\lim[p(P)-p(P')]/x_1 = \pm 2\cos\beta_1$ as $x_1 \to 0$. Combining this with (41) and (46) we can note for later use

$$\lim_{x_1\to 0}\frac{p([C,P])-p([C,P'])}{p(P)-p(P')} = \frac{\cos\alpha_1}{\cos\beta_1}. \tag{47}$$

In the case $r_2 = r_1$, C is a circle. By Lemma 17, the k-gon P is circumscribing a circle concentric with C, and Lemma 18 shows that P is regular. This proves our theorem when $r_2 = r_1$ or, equivalently, when $p^2/a = 4\pi$. Thus we may now assume $r_2 > r_1$.

LEMMA 19. *C and P correspond to the statement of Theorem 1.*

PROOF.

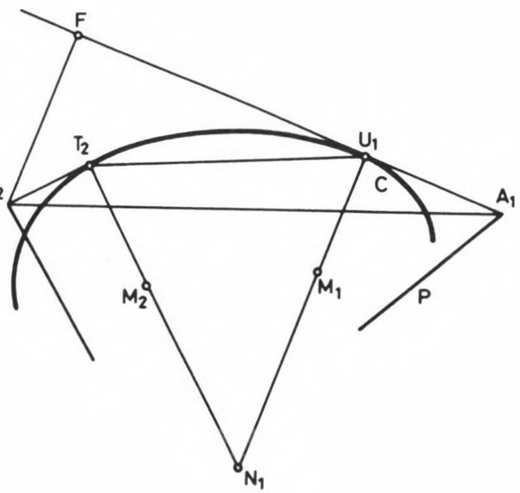


Fig. 5

Keeping up the notations used in Lemma 18, we recall that $|M_1U_1| = |M_2T_2| = r_1$, $|N_1U_1| = |N_1T_2| = r_2$, and N_1U_1 and N_1T_2 are orthogonal to A_1U_1 and A_2T_2, respectively (Fig. 5). A_1A_2 intersects M_1U_1 and M_2T_2.

We first show that

$$|A_1U_1| = |A_2T_2|. \tag{48}$$

Without loss of generality we may assume that $|A_1U_1| \geqq |A_2T_2|$. Let F be the orthogonal projection of A_2 onto $A_1 \vee U_1$, and set $|A_2F| = v$, $|FU_1| = w$ and $|A_1U_1| = x$. Keeping fixed the other parameters, we consider β_1, α_1 and β_2 as functions of x. By using

$$\cos\alpha_1 = \frac{x}{\sqrt{x^2+r_1^2}}, \quad \cos(\alpha_1-\beta_1) = \frac{w+x}{\sqrt{(w+x)^2+v^2}}$$

we obtain

$$\frac{\cos\beta_1}{\cos\alpha_1} = \left(w + x + \frac{vr_1}{x}\right)\left((w+x)^2 + v^2\right)^{-1/2}.$$

The derivation of this function differs from

$$v - \frac{v^2 r_1}{x^2} - \frac{r_1}{x}(w+x) - \frac{r_1}{x^2}(w+x)^2$$

only by a positive factor. Because $r_1/x = \tan\alpha_1 > \tan(\alpha_1 - \beta_1) = v/(w+x)$, we conclude that $\cos\beta_1/\cos\alpha_1$ is strictly decreasing in x. Obviously, $\cos\beta_2/\cos\alpha_2$ is strictly increasing in x. Since $x = |A_2T_2|$ implies the relation $\cos\beta_1/\cos\alpha_1 = \cos\beta_2/\cos\alpha_2$, which is to be satisfied by Lemma 18, equation (48) is proved. Hence $A_1A_2 \| M_1M_2$, so that $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, and more generally

(49) $$\alpha_1 = \ldots = \alpha_k, \quad \beta_1 = \ldots = \beta_k.$$

Thus P is an equiangular k-gon. Since, moreover,

$$|A_iA_{i+1}| = 2(r_2 - r_1)\sin(\alpha_i - \beta_i) + 2r_1\frac{\cos\beta_i}{\sin\alpha_i},$$

P is also equilateral. Hence P is regular and so is the k-gon $P^* = M_1M_2 \ldots M_k$. P^* is the kernel of a regular arc-sided k-gon C^* bounded by circular arcs of radius $r_2 - r_1$. C is the outer parallel domain of C^* at distance r_1, and P is the k-gon associated with C.

This completes the proof of Lemma 19 and Theorem 1. □

Let C be a parallel domain of a regular arc-sided k-gon, and let P be the k-gon associated with C. We conclude this section by showing that the parameters $a(C) = a$, $p(C) = p$ and $p(P) = p_0$ determine C up to isometry.

Let $C = (C^*)_{r_1}$, where C^* is a regular arc-sided k-gon, and let $a(C) = a$ and $p(C) = p$ be given. Let 2φ denote the central angle of the arcs bounding C^*. The discs C form a pencil joining the smooth regular k-gon, corresponding to $\varphi = 0$, with the regular arc-sided k-gon, corresponding to $\varphi = \varphi^*$, where $q = q(\varphi^*)$ is determined by (7) and (8). Let r and r^* be the in-radius of P and P^*, respectively, where P^* is the kernel of C^*. Then

(50) $$r(\varphi) = r^*(\varphi) + \frac{\cos(\pi/k)}{\cos(\pi/k - \varphi)}r_1(\varphi).$$

By applying (8) to C^* and Steiner's formulas to $(C^*)_{r_1}$ and setting $\cot(\pi/k) = u$, we find

$$r^* = \frac{u}{2k}\sqrt{\frac{p^2 - 4a\pi}{\Phi - \pi(u+q)/k}},$$

$$r_1 = \frac{1}{2\pi}\left(p - \sqrt{\frac{(p^2 - 4a\pi)\Phi}{\Phi - \pi(u+q)/k}}\right).$$

From (50) we obtain by differentiation

$$2k \tan\frac{\pi}{k}\sin^3\varphi\left[\Phi - \frac{\pi(u+q)}{k}\right]^{3/2}\left(r' + r_1\frac{\cos(\pi/k)\sin(\pi/k - \varphi)}{\cos^2(\pi/k - \varphi)}\right) =$$
$$= -\sqrt{p^2 - 4a\pi}(\sin\varphi - \varphi\cos\varphi)[\tan(\pi/k - \varphi) - (\pi/k - \varphi)],$$

which shows that $r'(\varphi) < 0$. Thus φ is uniquely determined by $p(P) = p_0$.

3. The perimeter deviation

We now turn to the problem of finding such members C of $\mathcal{C}(a,p)$ and $P \in \mathcal{P}_k$ for which $\delta^P(C,P)$ is minimal. In view of a remark made in Section 1, we can restrict ourselves to values of a, p and k satisfying condition (5).

For a disc $C \in \mathcal{C}(a,p)$ and a k-gon $P \in \mathcal{P}_k(p_0)$ we have by (1)

$$\delta^P(C,P) = 2p([C,P]) - p(C) - p(P). \tag{51}$$

Because $p(C) \leqq p$, it follows from (51) and (6) that

$$\delta^P(C,P) \geqq 2m(a,p;k,p_0) - p - p_0. \tag{52}$$

If $p_0 \geqq g_2(a,p,k)$, Remark 2 implies that

$$\delta^P(C,P) \geqq g_2(a,p,k) - p,$$

with equality if and only if C is a smooth regular k-gon of area a and perimeter p, and P is the case of C. Let P' be the k-gon obtained from P by displacing the vertex A_1 of P inwards on the bisector of the angle $\sphericalangle A_k A_1 A_2$ through a sufficiently small distance. By using (47) with $\alpha_1 = \beta_1$, it follows easily that $\delta^P(C,P') < \delta^P(C,P)$, which shows that $\delta^P(C,P)$ is not minimal. Thus we can assume in the following that $p_0 < g_2(a,p,k)$.

Since $p(C) \leqq p([C,P])$, we conclude from (51) and (6) that

$$\delta^P(C,P) \geqq m(a,p;k,p_0) - p_0. \tag{53}$$

If $p_0 \leqq g_1(a,p,k)$, we have by (53) and Remark 2

$$\delta^P(C,P) \geqq G(a,k,p_0) - p_0,$$

where G is given by (12). The function on the right-hand side is strictly decreasing in p_0, and thus attains its minimum for $p_0 = g_1(a,p,k)$. Therefore, we need to consider only such values of p_0 for which

$$g_1(a,p,k) \leqq p_0 < g_2(a,p,k). \tag{54}$$

We again make use of (52) and observe that, by Theorem 1 and Remark 2, equality takes place in (52) if and only if C is an outer parallel domain of a regular arc-sided k-gon of area a and perimeter p and P is the k-gon associated with C. If $p_0 = g_1(a,p,k)$, then C is degenerate, which means that C is a regular arc-sided k-gon with kernel P.

Let us first assume that $\delta^P(C,P)$ is minimal for some p_0 with $g_1(a,p,k) < < p_0 < g_2(a,p,k)$. Resuming the notation used in Lemma 18, we can state that

$$\frac{\cos\alpha_1}{\cos\beta_1} = \frac{1}{2} \quad \text{if } g_1(a,p,k) < p_0 < g_2(a,p,k). \tag{55}$$

Otherwise, we could reduce $\delta^P(C,P)$ by displacing the vertex A_1 of P on the bisector of the angle $\sphericalangle A_k A_1 A_2$. This follows from (47) and (51). Secondly, if we assume that $\delta^P(C,P)$ is minimal for $p_0 = g_1(a,p,k)$, a similar argument shows that

$$\frac{\cos\alpha_1}{\cos\beta_1} \geqq \frac{1}{2} \quad \text{if } p_0 = g_1(a,p,k). \tag{56}$$

Note that $\alpha_1 = \frac{\pi}{2} - \left(\frac{\pi}{k} - \varphi\right)$ and $\beta_1 = \frac{\pi}{2} - \frac{\pi}{k}$. Since

$$\frac{\cos\alpha_1}{\cos\beta_1} = \frac{\sin\left(\frac{\pi}{k} - \varphi\right)}{\sin\frac{\pi}{k}}$$

is a strictly decreasing function of φ, for $0 \leqq \varphi \leqq \varphi^*$, we have to consider two cases.

(i) If $\sin(\pi/k - \varphi^*) < (\sin\pi/k)/2$, (56) is impossible and the minimum of $\delta^P(C,P)$ is attained in the case indicated by (55). C is a (proper) parallel domain of a regular arc-sided k-gon, and P is the k-gon associated with C.

(ii) If $\sin(\pi/k - \varphi^*) \geqq (\sin\pi/k)/2$, (55) is impossible and the minimum of $\delta^P(C,P)$ is attained in the case indicated by (56). C is a regular arc-sided k-gon, and P is the kernel of C.

Note that

$$\varphi_0 = \frac{\pi}{k} - \arcsin\left(\frac{1}{2}\sin\frac{\pi}{k}\right)$$

is half the central angle of the arcs bounding a regular arc-sided k-gon C_0 with $\cos\alpha_1/\cos\beta_1 = \frac{1}{2}$. By (8), the isoperimetric ratio of C_0 is

$$\varrho(k) = 4k\sin\frac{\pi}{k}\frac{\Phi(q_0)}{\cos(\pi/k) + q_0\sin(\pi/k)},$$

where $q_0 = q(\varphi_0)$. Observing that the right-hand side of (8) is a strictly decreasing function of q (and also of φ), we can summarize the result of this section in

THEOREM 2. *Suppose that* $p^2/a < 4k \tan \pi/k$. *There is exactly one disc* C *from* $\mathcal{C}(a,p)$ *and one* k*-gon* P *such that*

$$\delta^P(C,P) = \Delta^P(a,p,k).$$

C *and* P *are characterized by the following properties:*

(i) $a(C) = a, p(C) = p$.

(ii) *if* $p^2/a < \varrho(k)$, C *is a parallel domain of a regular arc-sided* k*-gon and* P *is associated with* C. *Furthermore,*

$$\frac{\cos \alpha_1}{\cos \beta_1} = \frac{1}{2}.$$

(iii) *if* $p^2/a \geqq \varrho(k)$, C *is a regular arc-sided* k*-gon, and* P *is the kernel of* C.

ACKNOWLEDGEMENT. Part of this paper was written when I was visiting the University of Trieste in 1989. I would like to thank the Consiglio Nazionale delle Ricerche and the Dipartimento di Scienze Matematiche for making this visit possible and for their kind hospitality.

REFERENCES

[1] EGGLESTON, H. G., Approximation to plane convex curves, I. Dowker-type theorems, *Proc. London Math. Soc.* (3) **7** (1957), 351–377. *MR* **20** #1280

[2] FEJES TÓTH, G. and FLORIAN, A., Covering of the plane by discs, *Geom. Dedicata* **16** (1984), 315–333. *MR* **86a**: 52020

[3] FLORIAN, A., Approximation of convex discs by polygons, *Discrete Comput. Geom.* **1** (1986), 241–263. *MR* **88b**: 52002

[4] FLORIAN, A., On a metric for the class of compact convex sets, *Geom. Dedicata* **30** (1989), 69–80. *MR* **90h**: 52005

[5] GRUBER, P. M., Approximation of convex bodies, *Convexity and its Applications*, Birkhäuser-Verlag, Basel–Boston–Stuttgart, 1983, 131–162. *MR* **85d**: 52001

[6] KLÖTZLER, R. und ROTH, M., Zur Länge einer ebenen geschlossenen Kurve und des Randes ihrer konvexen Hülle, *Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-Naturwiss. Reihe* **37** (1988), 313–317. *MR* **90a**: 52021

[7] VITALE, R. A., L_p metrics for compact, convex sets, *J. Approx. Theory* **45** (1985), 280–287. *MR* **88a**: 52003

(Received February 13, 1990)

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT SALZBURG
HELLBRUNNERSTRASSE 34
A-5020 SALZBURG
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 119–123

О ПОДАЛГЕБРАХ КОНЕЧНОЙ КОРАЗМЕРНОСТИ

А. МЕКЕЙ

Пусть R алгебра над полем F, A её подалгебра. Если размерность пространства $(R/A,+)$ конечна над полем F, т.е. $\dim_F(R/A,+)<$ $<\infty$ то говорят, что подалгебра A имеет конечную коразмерность.

В работе [1] А. И. Мальцев показал, что всякий односторонний идеал конечной коразмерности, содержит двухсторонний идеал алгебры, так же конечной коразмерности. В работе [2] I. Lewin показал, что любое подкольца конечного индекса содержит идеал кольца, также конечного индекса. Основной результат настоящей работы является следующая

Теорема. *Пусть A подалгебра алгебры R над полем F имеющая конечную коразмерность: $n=\dim_F(R/A,+)$. Тогда в A содержится идеал $I\lhd R$ также конечной коразмерности, причем*

$$\dim_F(R/I)\leqq n(n^2+2n+2).$$

Доказательство. Рассмотрим два случая.

Случай 1. Пусть A — бесконечномерная подалгебра. Базис $a_1,a_2,\ldots,a_n,\ldots$ алгебры A дополним до базиса алгебры R, элементами $e_1,\ldots,e_n$. Тогда элементы $e_i+\mathcal{A}$, $i=\overline{1,n}$ образуют базис фактор-пространства $(R/A,+)=V_A$. Для простоты этот базис также будем обозначать через $e_1,e_2,\ldots,e_n$. Рассмотрим пространство $V=F\oplus V_A$ с базисом $e_0=1$, $e_1,\ldots,e_n$ где $1\in F$.

Определим линейные отображения из пространства $V\otimes V$ в V_A с помощью элементов алгебры A по следующему правилу.

Пусть $a\in\mathcal{A}$ и $e_i\otimes e_j$ $i,j=0,1,\ldots,n$ базис пространства $V\otimes V$ тогда, положим:

$$f_a(e_i\otimes e_j)=e_iae_j(\mathrm{mod}\,\mathcal{A})=\sum_{k=1}^{n}\alpha_k^{i,j}e_k(\mathrm{mod}\ A) \tag{1}$$

а для произвольного элемента v

$$v=\sum_{i,j=0}^{n}\beta_{ij}e_i\otimes e_j\in V\otimes V:$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 16P10; Secondary 16D15, 16P99.

Key words and phrases. Subalgebras, ideals, codimension.

$$f_a(v)=\sum_{i,j=0}^{n}\beta_{ij}f_a(e_i\otimes e_j)=\sum_{i,j=0}^{n}\beta_{ij}\alpha_k e_k(\operatorname{mod}\mathcal{A}).$$

Каждому линейному отображению f_a где $a\in\mathcal{A}$ определяемому по правилу (1) сопоставим матрицу

$$\mathcal{A}_a=(\alpha_k^{i,j}),\quad \alpha_k^{i,j}\in F\quad k=1,2,\dots,n\quad i,j=0,1,\dots,n.$$

Пусть $a_1,a_2,\dots,a_m,\dots$ базис алгебры A, $f_{a_1},f_{a_2},\dots,f_{a_m},\dots$ соответствующие им линейные отображения. Соответствующие им матрицы будем обозначать через $A_{a_s}=(\alpha_{a_s,k}^{i,j})$, $s=1,2,\dots,m,\dots$.

Нетрудно проверить, что

$$(2)\qquad \begin{gathered} f_{a_m}+f_{a_s}=f_{a_m+a_s},\quad \alpha f_{a_m}=f_{\alpha a_m}\\ A_{a_m}+A_{a_s}=A_{a_m+a_s},\quad \alpha A_{a_m}=A_{\alpha a_m},\quad \alpha\in F. \end{gathered}$$

Далее, $\dim(V\otimes V)=(n+1)^2$, $\dim V_A=n$, поэтому матрицы A_{a_s} имеют размерности $d=n\times(n+1)^2$. Множество всех матриц размерности $n\times(n+1)^2$ над полем F является векторным пространством размерности d. Поэтому среди матриц $\mathcal{A}_{a_s}$ существуют, не более чем d линейно независимых матриц. Пусть они есть $\mathcal{A}_{a_1},\dots,\mathcal{A}_{a_k}$ где $k\leqq d$, причем можно считать, что $a_1,a_2,\dots,a_k$ часть базисных элементов алгебры A.

Тогда все матрицы $\mathcal{A}_{a_j}$, $j>k$ соответствующие отображениям f_{a_j}, $j>k$ являются линейными комбинациями матриц $\mathcal{A}_{a_1},\dots,\mathcal{A}_{a_k}$, т.е.

$$\mathcal{A}_{a_j}=\sum_{m=1}^{k}\beta_m\mathcal{A}_{a_m}$$

где $j>k$. Это означает, что $f_{a_j}=\sum\limits_{m=1}^{k}\beta_m f_{a_m}$, $j>k$. Согласно (2), отсюда имеем:

$$0=f_{a_j}-\sum_{m=1}^{k}\beta_m f_{a_m}=f_{a_j}-f_{\sum\beta_m a_m}=f_{a_j-\sum\limits_{m=1}^{k}\beta_m a_m},\quad j>k.$$

Следовательно, отображение $f_{a_j-\sum\limits_{m=1}^{k}\beta_m a_m}$ является нулевым отображением по $\operatorname{mod}\mathcal{A}$, т.е.

$$(3)\qquad f_{a_j-\sum\limits_{m=1}^{k}\beta_m a_m}(e_r\otimes e_s)=0=e_r\Big(a_j-\sum_{m=1}^{k}\beta_m a_m\Big)e_s(\operatorname{mod}\mathcal{A})$$

или то же самое, что

$$e_r\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)e_s \in \mathcal{A}, \quad \text{где} \quad j > k$$

для любых $r, s = 0, 1, 2, \dots, n$.

В силу линейной независимости базисных элементов имеем, что $a_j - \sum\limits_{m=1}^{k} \beta_m a_m \neq 0$ где $j > k$.

Кроме того, из

$$a_j - \sum_{m=1}^{k} \beta_m a_m \in \mathcal{A}, \qquad f_{a_j - \sum\limits_{m=1}^{k} \beta_m a_m} = 0$$

следует, что

$$a_p\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)a_q, \;\; e_r\Big(a_j - \sum_{n=1}^{k} \beta_m a_m\Big)a_q, \;\; \Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)e_r \in \mathcal{A},$$

где $p, q = 1, 2, \dots, n, \dots, r = 0, 1, 2, \dots, n$, $j > k$.

Обозначим через $\overline{A}$ подалгебру алгебры A, порожденную всеми элементами вида:

$$a_j - \sum_{m=1}^{k} \beta_m a_m, \;\; e_r\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)e_s, \;\; a_p\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)e_r$$

$$e_r\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)a_q, \;\; a_p\Big(a_j - \sum_{m=1}^{k} \beta_m a_m\Big)a_q, \;\; \text{где} \;\; j > k,$$

$p, q = 1, 2, 3, \dots, n, \dots$; $r, s = 0, 1, 2, \dots, n$, и $a_j - \sum\limits_{m=1}^{k} \beta_m a_m$, $j > k$ элементы алгебры A обладающие свойством (3), т.е. $f_{a_j - \sum\limits_{m=1}^{k} \beta_m a_m} = 0$.

Отсюда ясно, что $\overline{A}$ является идеалом алгебры R содержащийся в A. Сумму всех идеалов алгебры R содержащихся в A обозначим через I.

Фактор алгебра R/I содержит подалгебру A/I причем пространства $(R/I)/(A/I) \cong R/A$ так же имеет размерность n. Если $\dim A/I > n(n+1)^2$ тогда рассуждая как и выше найдем нетривиальный идеал алгебры R/I содержащийся в A/I. Это противоречит максимальности идеала I, содержащегося в A. Следовательно

$$\dim A/I \leqq n(n+1)^2.$$

Отсюда следует, что

$$\dim R/I \leqq n + n(n+1)^2 = n(n^2+2n+2).$$

Случай 2. Пусть подалгебра A — конечномерна. Если $\dim \mathcal{A} \leqq \leqq n(n+1)^2$ и $\dim(R/\mathcal{A},+) = n$ тогда в качестве искомого идеала можно брать идеал (0). Тогда ясно, что $\dim R \leqq n(n^2+2n+2)$. Если $\dim \mathcal{A} > > n(n+1)^2$, тогда рассуждаем как в случае 1 и найдем нетривиальный идеал I удовлетворяющий условию теоремы и.т.д.

Естественно возникает вопрос: Нельзя ли снизить оценку? Она достигается следующими замечаниями.

Замечание 1. Если алгебра R обладает единицу, $1 \notin A$, тогда не нужно рассмотреть $V_A \oplus F = V$, а сразу V_A, тогда пространства $V_A \otimes V_A$ имеет размерность n^2 и мы рассмотрим матрицы размера $n \times n^2 = n^3$. Тогда $\dim R/I \leqq n^3 + n = n(n^2+1)$.

Замечание 2. Если алгебра R необладает единицей тогда с самого начала можно рассмотреть пространства $V_A \otimes V_A$ размерности n^2 и затем дополним его пространствами: $F \otimes V_A$, $V_A \otimes F$, т.е. рассмотрим $V_1 = (F \otimes V_A) \oplus (V_A \otimes F) \oplus (V_A \otimes V_A)$. Тогда пространство V_1 имеет размерность $n^2 + 2n$. Тогда нам нужно будет рассмотреть матрицы размера $n \times (n^2 + 2n)$. В этом случае

$$\dim R/I \leqq n + n(n^2+2n) = n(n+1)^2.$$

Последние оценки лучше чем данной, в основной теореме.

Следствие 1. *Пусть алгебра R над полем F обладает подалгеброй A конечной коразмерности и причем A аппроксирусма конечномерными алгебрами. Тогда алгебра R аппроксимируема конечномерными алгебрами.*

Доказательство. Так как A обладает системой идеалов $\mathcal{A}_\alpha$ $\alpha \in \mathcal{J}$, конечной коразмерности, т.е. $\dim A/\mathcal{A}_\alpha < \infty$ для всех $\alpha \in \mathcal{J}$. Каждая $\mathcal{A}_\alpha$ есть подалгебра алгебры R конечной коразмерности, поэтому $\mathcal{A}_\alpha$ содержит идеалы B_α алгебры R также конечной коразмерности. Из $\bigcap\limits_{\alpha \in \mathcal{J}} \mathcal{A}_\alpha = 0$ равенства следует, что $\bigcap\limits_{\alpha \in \mathcal{J}} B_\alpha = 0$. Отсюда вытекает что алгебра R аппроксимируется конечномерными алгебрами.

Следствие 2. *Алгебра R над полем F аппроксимируется конечномерными алгебрами тогда и только тогда, когда все её подалгебры конечной коразмерности пересекаются тривиальным образом.*

Следствие 3. *Бесконечномерная простая алгебра не обладает подалгебру конечной коразмерности.*

Следствие 4. *Пусть A подалгебра конечной коразмерности алгебры R над полем F. Если алгебра A представимо матрицами тогда и алгебра R представимо матрицами.*

Последнее следствие было доказано в [1], в случае когда A односторонний идеал конечной коразмерности, алгебры R. Доказательство следствие 3 сразу следует из основной теоремы и доказательства теоремы 10 из [1].

В заключение заметим, что основной результат работы был анонсирован в тезисах всемирного конгресса математиков 1986 года.

LITERATURA

[1] Мальцев, А. И., О представлении бесконечных алгебр, *Rec. Math. [Mat. Sbornik]* N. S. **13 (55)** (1943), 263–286. *MR* **6**-116

[2] Lewin, J., Subrings of finite index in finitely generated rings, *J. Algebra* **5** (1967), 84–88. *MR* **34** #196

(Поступило 25-ое января, 1988)

КАФЕДРА ВЫСШЕЙ АЛГЕБРЫ
МАТЕМАТИЧЕСКОГО ФАКУЛЬТЕТА
МОНГОЛЬСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
УЛААН БААТАР
П/Я 46/627
МОНГОЛИЯ

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 125–132

RADICALS AND 0-BANDS OF SEMIGROUPS

A. V. KELAREV

Introduction

The concept of radical plays a crucial role in the structure theory of rings (see [2]). It can be extended to various classes of algebras including semigroups, and is useful in investigating their structure, too.

In the structure theory of semigroups, however, there is another basic concept, namely, that of decomposition of a semigroup into a band of its subsemigroups (see [3]). In [22] Weissglass proposed a natural way to extend the notion of band to rings. In [10] we studied interconnections of radicals and bands of rings and described such situations when their interaction is good in some sense. The aim of the present paper is to consider interactions of radicals and bands of semigroups.

There are several analogues of ring radicals in semigroup theory. Defining radicals as some ideals of semigroups seems to be most similar to the ring approach. (See [6]–[9]. Preliminaries on such radicals are included in § 1 of the present paper.) These radicals, however, are defined only for semigroups with zero. And a band of semigroups with zero need not have a zero. Therefore we modify the notion of band slightly, making it applicable to semigroups with zero.

Let Ω be a band, i.e., a semigroup satisfying the identity $x^2 = x$. A semigroup S with zero 0 is said to be a 0-band of subsemigroups S_α, $\alpha \in \Omega$, if

(1) $S = \bigcup\limits_{\alpha\in\Omega} S_\alpha$;

(2) $S_\alpha \cap S_\beta = \{0\}$ when $\alpha \neq \beta$;

(3) $S_\alpha S_\beta \subseteq S_{\alpha\beta}$ for each $\alpha, \beta \in \Omega$.

If Ω is a semilattice, i.e., a commutative band, then we call S a 0-semilattice of subsemigroups S_α. If Ω is a left band, i.e., a band satisfying the identity $xy = x$, we call S a left 0-band of subsemigroups S_α. The semigroups S_α are called the components of the band. In case when S is a 0-band Ω of subsemigroups S_α, we write $S = \bigcup\limits_{\Omega} S_\alpha$.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 20M11.
Key words and phrases. Semigroups with zero, radicals, bands of semigroups.

Akadémiai Kiadó, Budapest

In what follows, the word "semigroup" will mean "semigroup with zero" unless stated otherwise. The concept of 0-band is likely to be first introduced in [11]. It is quite natural and is virtually a well-known object. One may consider 0-unions of semigroups as an example of 0-bands, see [3] § 6.3. Some authors used 0-bands without giving the definition explicitly. For instance, in [13], [14] and [15] left and rectangular 0-bands are applied when describing primitive regular and primitive inverse semigroups, and in [4] left 0-bands are used for characterizing semigroups having completely 0-simple semigroups of quotients. Rectangular 0-bands were investigated in [13]–[14], where they were called "0-matrix decompositions". The authors of [13]–[14] used the term "rectangular 0-bands" with a different meaning.)

It is well-known that a semigroup S is a band of subsemigroups if and only if there exists a homomorphism of S on Ω, see [3] § 1.8. Now we give an analogous characterization of 0-bands.

REMARK 1. *A semigroup S is a 0-band Ω of subsemigroups if and only if there is a mapping f of S onto Ω such that for each $x, y \in S$, $xy \neq 0$ implies $f(xy) = f(x) \cdot f(y)$.*

Proof is easy and we omit it.

Let us return to the question on interaction of radicals and 0-bands. How to formulate this in a concrete way? In the corresponding ring situation the following problem was posed in [5]: to describe the radicals ϱ such that the radical of a band of rings is equal to the sum of the radicals of the components. This problem is solved in [10]. Here we consider its semigroup equivalent: which are the radicals ϱ such that the radical of a 0-band is equal to the union of the radicals of the components? The following definition will be useful for discussing the results. Let Δ be a class of bands. We say that a radical ϱ commutes with 0-bands of Δ if, for every $\Omega \in \Delta$ and every $S = \bigcup_{\Omega} S_\alpha$, the equality $\varrho(S) = \bigcup_{\alpha \in \Omega} \varrho(S_\alpha)$ holds.

The present paper carries out a thorough investigation of radicals commuting with 0-bands. It consists of two sections. In § 1 we study some properties of semigroup radicals. In § 2 for each class Δ of bands all radicals commuting with 0-bands of Δ are described.

The author is grateful to Professor L. N. Shevrin for his kind supervision and to Drs M. V. Volkov and L. Márki for useful advices.

§ 1. Properties of semigroup radicals

We give the following definitions according to [9]. Let ϱ be a mapping which assigns to every semigroup S an ideal $\varrho(S)$ of S. A semigroup S is said to be a radical (semisimple) semigroup if $\varrho(S) = S$ ($\varrho(S) = \{0\}$). The mapping ϱ is called a radical if, for each semigroup S,

(1) $\varrho(S/I) \supseteqq (\varrho(S) \cup I)/I$ whenever I is an ideal of S;
(2) $\varrho(S)$ is the largest radical ideal of S;
(3) the quotient semigroup $S/\varrho(S)$ is semisimple.

We denote by $\mathcal{R}$ and $\mathcal{S}$ the classes of all radical and semisimple semigroups, respectively, belonging to the radical ϱ. We call a radical trivial if its radical or semisimple class consists of one-element semigroups only. Note that each radical is uniquely determined by fixing its semisimple or radical class. Thus, there are exactly two trivial radicals.

In [9] (see also [16]) the following description of radical classes has been given.

PROPOSITION 1. *A class $\mathcal{R}$ is a radical class if and only if*
(a) *$\mathcal{R}$ is closed under Rees quotient semigroups;*
(b) *$\mathcal{R}$ is closed under ideal extensions;*
(c) *the union of all $\mathcal{R}$-ideals in an arbitrary semigroup belongs to $\mathcal{R}$.*

Semisimple classes were described in [15]. We shall use this description in a corrected form (see [18], II.1.8). A semigroup S is said to be a Rees subdirect product of subsemigroups S_i, $i \in I$, if there are ideals T_i, $i \in I$, in S such that $\bigcap_{i\in I} T_i = 0$ and $S/T_i \cong S_i$ for all i.

PROPOSITION 2. *A class $\mathcal{S}$ is a semisimple class if and only if*
(d) *$\mathcal{S}$ is hereditary, i.e., closed under ideals;*
(b) *$\mathcal{S}$ is closed under ideal extensions;*
(e) *$\mathcal{S}$ is closed under Rees subdirect products.*

In [1] and [8] it has been shown that each semigroup S contains a least ideal $\eta(S)$ such that $S/\eta(S)$ has no nonzero nilpotent elements. Evidently, the mapping η is a radical. It is analogous to the generalized nil radical of rings (see [2] ch. 4 § 2), so in [7] η has been called the generalized nil radical.

LEMMA 1 ([6] and [8]). *A semigroup is η-semisimple if and only if it is a subdirect product of semigroups without zero divisors.*

Like in ring theory, we call a radical ϱ strict if all subsemigroups of any semisimple semigroup are also semisimple. A radical ϱ is said to be hereditary if for each ideal I of a semigroup S the equality $\varrho(I) = I \cap \varrho(S)$ holds. The strictness of η is obvious, so Remark 2 of [7] implies

LEMMA 2. *The generalized nil radical η is strict and hereditary.*

Now we can prove the following

LEMMA 3. *Let $\mathcal{R}$ be the radical class of a nontrivial strict radical ϱ. Then $\mathcal{R}$ contains the radical class $\mathcal{N}$ of η.*

PROOF. First we show that $\mathcal{R}$ contains a semigroup S with $S \neq S^2$. Choose $\{0\} \neq R \in \mathcal{R}$. Let $A = \{0, \alpha\}$ be a semigroup with zero multiplication. Denote by W the free product of A and R with common zero. Let

$V = W \setminus \{\alpha\}$. Since ϱ is strict, $\varrho(W) \supseteqq V$, as V is an ideal generated in W by R. If $\varrho(W) = V$, then V is the sought semigroup, because $V \neq V^2$. On the other hand if $\varrho(W) \neq V$, then $\varrho(W) = W$ and W is the sought one.

We have proved there is $S \in \mathcal{R}$, $\{0\} \neq S/S^2 \in \mathcal{R}$. By [6, Th. 3] this yields that all semigroups with zero multiplication belong to $\mathcal{R}$.

Now let $T \in \mathcal{N}$. We show that $T \in \mathcal{R}$. Suppose the contrary, let $F = T/\varrho(T)$ be nonzero. Since F is η-radical, it contains an $x \neq 0$ such that $x^2 = 0$. The subsemigroup of F generated by x has zero multiplication and belongs to $\mathcal{R}$. This contradicts the strictness of ϱ, and completes the proof.

We obtain a description of strict radicals.

THEOREM 1. *Let $\mathcal{M}$ be a nonempty class of semigroups, and denote by $\mathcal{R}_{\mathcal{M}}$ the class of semigroups S such that each nonzero Rees quotient semigroup S/I contains a nonzero subsemigroup from $\mathcal{M}$ or a nonzero subsemigroup from the radical class $\mathcal{N}$ of η. Then there exists a (clearly unique) strict radical $\varrho_{\mathcal{M}}$ with radical class $\mathcal{R}_{\mathcal{M}}$. Conversely, each strict radical coincides with $\varrho_{\mathcal{M}}$ for an appropriate $\mathcal{M}$.*

PROOF. The converse is evident: it suffices to choose $\mathcal{M}$ to be the class of all radical semigroups.

First we prove the following statement. Let A be a semigroup containing a subsemigroup R from $\mathcal{R}_{\mathcal{M}}$. Denote by I the ideal generated in A by R. We claim that $I \in \mathcal{R}_{\mathcal{M}}$. Indeed, for any proper ideal J of I we need to show that I/J has a nonzero subsemigroup from $\mathcal{M} \cup \mathcal{N}$. If J does not contain R, then I/J contains the subsemigroup $(P \cup J)/J = R/(R \cap J)$ which contains a nonzero subsemigroup from $\mathcal{M} \cup \mathcal{N}$. Therefore we may assume that $J \supseteqq R$. As $I = A^1RA^1$, we deduce $I^3 \subseteqq (A^1IA^1)J(A^1IA^1) \subseteqq J$. Hence I/J is a 3-nilpotent semigroup belonging to $\mathcal{N}$. We have proved that in both cases $I \in \mathcal{R}_{\mathcal{M}}$.

Using this statement it is routine to verify that $\mathcal{R}_{\mathcal{M}}$ satisfies all conditions of Proposition 1. Hence, there is a radical $\varrho_{\mathcal{M}}$ with radical class $\mathcal{R}_{\mathcal{M}}$. By the above statement $\varrho_{\mathcal{M}}$ obviously is strict, which completes the proof.

Note that our result is distinct from the description of strict ring radicals [21], which has the following form: Let $\mathcal{M}$ be a class of rings, $\mathcal{R}_{\mathcal{M}}$ be the class of all rings A such that every nonzero homomorphic image of A contains a nonzero subring which is in $\mathcal{M}$. A radical class is strict if and only if it coincides with $\mathcal{R}_{\mathcal{M}}$ for some $\mathcal{M}$.

It is impossible to get the semigroup description in such a form. The difference is that there exists a least nontrivial strict semigroup radical — the generalized nil radical, whereas in the case of rings or algebras over a field there is no least nontrivial radical.

The proof of Theorem 1 yields

COROLLARY 1. *For each strict radical and for each semigroup S, the ideal generated by any radical subsemigroup of S is also radical.*

We call a radical ϱ weakly hereditary if, for any semigroup S and any ideal I, the inclusion $\varrho(I) \supseteqq I\varrho(S) \cup \varrho(S)I$ holds.

LEMMA 4. *Let ϱ be a strict and weakly hereditary nontrivial radical. Then ϱ coincides with the generalized nil radical.*

PROOF. By Lemma 3 the radical class $\mathcal{R}$ of ϱ contains the radical class $\mathcal{N}$ of η. Suppose that there is a semigroup $S \in \mathcal{R}$, $S \notin \mathcal{N}$. Then the quotient semigroup $F = S/\eta(S)$ belongs to $\mathcal{R}$, and by Lemma 1 it has an ideal I such that $R = F/I$ is nonzero and has no zero divisors.

Fixing an arbitrary semigroup Q we show that $Q \in \mathcal{R}$. We may assume that $Q \cap R = \{0\}$. Defining on the union $W = R \cup Q$ a multiplication by $tq = qt = q$ for $t \in T$, $q \in Q$, Q becomes a semigroup which is a 0-semilattice of the subsemigroup R and the ideal Q. So $\varrho(Q) \supseteqq Q\varrho(W) \supseteqq QR = Q$, whence $Q \in \mathcal{R}$ and $\mathcal{R}$ is trivial. This contradiction completes the proof.

Now we shall investigate two properties of semigroup radicals. We call a radical ϱ right (left) weakly hereditary if, for each semigroup S and every right (left) ideal I of S, the inclusion $\varrho(I) \supseteqq I\varrho(S)$ (or $\varrho(I) \supseteqq \varrho(S)I$) holds. A right and left weakly hereditary radical is obviously weakly hereditary. We say that ϱ is right (left) strict if, for each semigroup S, the radical of every right (left) ideal is contained in $\varrho(S)$. Obviously, a strict radical is right and left strict.

LEMMA 5. *If a right (left) strict radical is right (or left) weakly hereditary then it is trivial.*

PROOF. Let ϱ be right strict and weakly hereditary. Suppose that there is a nonzero semigroup S belonging to the radical class $\mathcal{R}$ of ϱ. If $S^2 \neq S$ then $\mathcal{R}$ contains the nonzero quotient semigroup S/S^2 which has zero multiplication. If $S^2 = S$, then we denote by R a semigroup with zero multiplication such that $|R| = |S|$, $R \cap S = \{0\}$. We claim that $R \in \mathcal{R}$. Let π be one-to-one mapping of R onto S, $\pi(0) = 0$. Define on $W = R \cup S$ a multiplication by the rule $rs = \pi^1(\pi(r)s)$, $sr = 0$, where $r \in R$, $s \in S$. Then W becomes a semigroup with (right) ideal R. By right strictness $\varrho(R) \supseteqq RS = R$. Thus, inevitably, $\mathcal{R}$ contains a nonzero semigroup with zero multiplication. Hence the semigroup $A = \{0, \alpha\}$ with zero multiplication is radical.

For an arbitrary semigroup T we shall prove that $T \in \mathcal{R}$. Denote by W the free product of T and A with a common zero. Consider the subsemigroup B generated by $\{\alpha t \mid t \in T\}$ and $\{\alpha t\alpha \mid t \in T\}$, and put $C = B \cup A$. Let σ be the congruence on W generated by all pairs $(s\alpha t, st)$ with $s, t \in T$. Denote by f the natural homomorphism of W onto $\overline{W} = W/\sigma$. Let $\overline{C} = f(C)$, $\overline{B} = f(B)$, $\overline{T} = f(T)$, $\overline{\alpha} = f(\alpha)$. Evidently, $\overline{T} \cong T$.

The semigroup C is a right ideal in $\overline{W}$, and $\overline{A}$ is a right ideal in $\overline{C}$. By right strictness $\varrho(\overline{W}) \supseteqq \varrho(\overline{C}) \supseteqq \varrho(\overline{A}) = A$. Hence $\varrho(\overline{W}) \supseteqq \overline{W}^1\overline{AW}^1$. Since $\overline{T}\overline{\alpha}\overline{T} = \overline{T}^2$ and $\overline{W}^1\overline{AW}^1 = \overline{A} \cup \overline{T}\overline{\alpha} \cup \overline{\alpha}\overline{T} \cup \overline{\alpha}\overline{T}\overline{\alpha} \cup \overline{T}^2$, $\overline{W}/\overline{W}^1\overline{AW}^1$ is a semigroup with zero multiplication. This yields that $\overline{W}/\varrho(\overline{W})$ has zero multiplication. Since A is radical, it follows that all semigroups with zero multiplication are radical. Hence $\overline{W}$ is radical, because $\mathcal{R}$ is closed under ideal extensions.

By hereditariness we conclude $\varrho(\overline{B}) \supseteq \overline{B}\overline{W} = \overline{\alpha}\overline{T}^2 \cup \overline{\alpha}\overline{T}\overline{\alpha}$. Hence $\overline{B}/\varrho(\overline{B})$ has zero multiplication, which implies $\overline{B} \in \mathcal{R}$ and $T \cong \overline{B}/\overline{\alpha}\overline{T}\overline{\alpha} \in \mathcal{R}$. We have proved that $\mathcal{R}$ contains all semigroups, so ϱ is trivial. The lemma is proved.

§ 2. Radicals commuting with 0-bands

Firstly we consider radicals commuting with noncommutative 0-bands.

THEOREM 2. *Let Δ be a class of bands containing a noncommutative band Ω. Then every radical commuting with 0-bands of Δ is trivial.*

PROOF. Let a radical ϱ commute with 0-bands of Δ. The semigroup Ω is known to have a left or right zero subband Λ, $|\Lambda| = 2$. Assume that Λ is a left zero band. We claim that ϱ commutes with Λ. Indeed, for an arbitrary 0-band $S = \bigcup_{\Lambda} S_\alpha$ we put $S_\alpha = \{0\}$ if $\alpha \in \Omega \setminus \Lambda$, and obtain

$$\varrho(S) = \varrho\Big(\bigcup_{\Omega} S_\alpha\Big) = \bigcup_{\Omega} \varrho(S_\alpha) = \bigcup_{\Lambda} \varrho(S_\alpha).$$

We will show that ϱ is right strict and right weakly hereditary. Fix any semigroup S and a right ideal I in S, and take a semigroup A such that $A \cong I$, $A \cap S = 0$. Denote by π an isomorphism of A onto I. Define on $W = S \cup A$ a multiplication by $sa = s\pi(a)$, $as = \pi^{-1}(\pi(a)s)$. With this multiplication W is a semigroup which is a 0-band Λ of its subsemigroups S and A. So $\varrho(W) = \varrho(S) \cup \varrho(A)$. Since $\varrho(W)$ is an ideal in W, we get $A\varrho(S) \subseteq \varrho(A)$. Hence $I\varrho(S) \subseteq \varrho(I)$, so that ϱ is right strict.

Likewise we infer $\varrho(A)S \subseteq \varrho(A)$ and $S\varrho(A) \subseteq \varrho(S)$. So $V = \varrho(I) \cup \varrho(S)$ is an ideal in S. Hence V is an ideal extension of $\varrho(S)$ by the radical semigroup $\varrho(I)/(\varrho(I) \cap \varrho(S))$, which implies that V is a radical. It follows that $\varrho(I) \subseteq \varrho(S)$, i.e. ϱ is right weakly hereditary. By Lemma 5 ϱ is trivial, and the proof is complete.

THEOREM 3. *Let Δ be a nonempty class of nonzero semilattices. A nontrivial radical ϱ commutes with 0-bands of Δ if and only if ϱ coincides with the generalized nil radical η.*

PROOF. *Necessity.* Let ϱ be a nontrivial radical commuting with an Ω from Δ. Being a nonzero semilattice, Ω contains a subsemilattice $\Lambda = \{0, 1\}$. As in the proof of Theorem 2, it follows that ϱ commutes with Λ.

We show that ϱ is strict. Suppose the contrary, and let S be a semisimple semigroup containing a radical subsemigroup R. Take a semigroup $T = R$

such that $T \cap S = \{0\}$. Denote by π an isomorphism of T onto R. Define on $W = S \cup T$ a multiplication by $st = s\pi(t)$, $ts = \pi(t)s$, where $s \in S$, $t \in T$. Then W turns into a semigroup which is a 0-semilattice Λ of its subsemigroup T and ideal S. Since ϱ commutes with Λ we conclude that $\varrho(W) = T$. Because $\varrho(W)$ is an ideal in W, we have $ST \cup TS \subseteqq \varrho(S) = \{0\}$; so that $SR = RS = \{0\}$ and R is an ideal in S. This contradicts to the semisimplicity of S.

We show that ϱ is weakly hereditary. Fix an arbitrary semigroup S and an ideal I in S. Take a semigroup $T \cong S$, $T \cap S = \{0\}$. Denote by π an isomorphism of T onto I. Endow $W = T \cup S$ with multiplication by the rule $st = \pi^{-1}(s\pi(t))$, $ts = \pi^{-1}(\pi(t)s)$, where $s \in S$, $t \in T$. Then W becomes a semigroup which is a 0-semilattice Λ of its subsemigroup S and ideal T. Hence $\varrho(W)$ is equal to $U = \varrho(S) \cup \varrho(T)$. Since U is an ideal in W, it holds $\varrho(S)T \cup T\varrho(S) \subseteqq \varrho(T)$. Transforming this inclusion by the isomorphism π^{-1}, we get $\varrho(S)I \cup I\varrho(S) \subseteqq \varrho(I)$.

Thus ϱ is strict and weakly hereditary, hence by Lemma 4 it coincides with the generalized nil radical η.

Sufficiency. We will show that η commutes with all 0-semilattices. Take any semigroup $S = \bigcup_{\Omega} S_\alpha$ where Ω is a semilattice. Let $I_\alpha = \eta(S_\alpha)$, $I = \bigcup_{\Omega} I_\alpha$. We need to prove that $I = \eta(S)$.

The inclusion $I \subseteqq \eta(S)$ follows from Lemma 3. For proving the converse inclusion, we firstly show that I is an ideal in S. Choose any $s \in S$, $t \in I$. Let $t \in I_\alpha$, $s \in S_\beta$. We need to prove that $ts, st \in I_{\alpha\beta}$. Since η is hereditary, ts and st belong to $\varrho(Q)$, where $Q = \bigcup_{\gamma \leqq \alpha\beta} S_\gamma$. Because the radical class $\mathcal{N}$ of η is closed under Rees quotient semigroups, st and ts belong to $Q \,/ \bigcup_{\gamma < \alpha\beta} S_\gamma = S_{\alpha\beta}$. Hence I is an ideal in S.

Let $P = S/I$, $P_\alpha = S_\alpha / I_\alpha$. Obviously $P = \bigcup_{\Omega} P_\alpha$. Each P_α has no nonzero nilpotent elements, which implies that P is η-semisimple. Hence $\varrho(S) \subseteqq I$. The theorem is proved.

REFERENCES

[1] ADAMS, D. H., Semigroups with no non-zero nilpotent elements, *Math. Z.* **123** (1971), 168–176. *MR* **44** #6873

[2] ANDRUNAKIEVICH, V. A. and RJABUHIN, JU. M., *Radicals of algebras and structural theory*, Nauka, Moscow, 1979 (in Russian). *MR* **82a**: 16001

[3] CLIFFORD, A. H. and PRESTON, G. B., *The algebraic theory of semigroups*, vol. 1, Math. Surveys, No. 7, American Mathematical Society, Providence, RI, 1961. *MR* **24** #A2627; vol. 2, 1967. *MR* **36** #1558

[4] FOUNTAIN, J. and GOULD, V., Completely 0-simple semigroups of quotients, II, *Contributions to general algebra*, 3 (Vienna, 1984), Hölder–Pichler–Tempsky, Wien, 1985, 115–124. *MR* **87j**: 20104

[5] GARDNER, B. J., Radicals of supplementary semilattice sums of associative rings, *Pacific J. Math.* **58** (1975), 387–392. *MR* **51** #12912

[6] GRIGOR, R. S., On the theory of semigroup radicals, I, *Mat. Issled.* **6** (1971), No. 4(22), 37–55 (in Russian). *MR* **45** #2055

[7] GRIGOR, R. S., On the theory of radicals of semigroups, II, *Mat. Issled.* **8** (1973), No. 4(30), 28–46 (in Russian). *MR* **49** #10802

[8] GRIGOR, R. S., On semigroups without nilpotent elements, *I All-Union symposium on the theory of semigroups*, Sverdlovsk, 1969, 24 (in Russian).

[9] HOFFMAN, A. E., Radicals of semigroups with zero, *Notices Amer. Math. Soc.* **20** (1973), A 359.

[10] KELAREV, A. V., Radicals and bands of associative algebras, *Algebra i Logika* **28** (1989), 430–443 (in Russian).

[11] KOZHEVNIKOV, O. B., A generalization of the concept of complete regularity, *Associativnye dejstvija,* Leningrad. Gos. Ped. Inst., Leningrad, 1983, 50–56 (in Russian). *MR* **85k**: 20182

[12] LALLEMENT, G. and PETRICH, M., Some results concerning completely 0-simple semigroups, *Bull. Amer. Math. Soc.* **70** (1964), 777–778. *MR* **30** #3166

[13] LALLEMENT, G. and PETRICH, M., Décomposition *I*-matricielles d'un demi-groupe, *J. Math. Pures Appl.* (9) **45** (1966), 67–117. *MR* **33** #5765

[14] LALLEMENT, G. and PETRICH, M., A generalization of the Rees theorem in semigroups, *Acta Sci. Math. (Szeged)* **30** (1969), 113–132. *MR* **39** #2899

[15] LEEUVEN, L. C. A., ROOS, C. and WIEGANDT, R., Characterizations of semisimple classes, *J. Austral. Math. Soc. Ser. A* **23** (1977), 172–182. *MR* **56** #3045

[16] MÁRKI, L., Radical semisimple classes and varieties of semigroups with zero, *Algebraic theory of semigroups*, (Proc. Sixth Algebraic Conf., Szeged, 1976), Colloq. Math. Soc. J. Bolyai, **20**, North-Holland, Amsterdam, 1979, 357–369. *MR* **80h**: 20085

[17] MÁRKI, L., Structure theorems on certain regular and inverse semigroups, *Czechoslovak. Math. J.* **27 (102)** (1977), 388–393. *MR* **56** #3163

[18] MÁRKI, L., MLITZ, R. and WIEGANDT R., A general Kurosh–Amitsur radical theory, *Comm. Algebra* **16** (1988), 249–305.

[19] SHEVRIN, L. N., A band of semigroups, *Mathematical encyclopaedia*, vol. 4, 1090–1091.

[20] STEINFELD, O., On semigroups which are unions of completely 0-simple subsemigroups, *Czechoslovak. Math. J.* **27** (1977), 388–393.

[21] STEWART, P. N., Strict radical classes of associative rings, *Proc. Amer. Math. Soc.* **39** (1973), 273-278. *MR* **47** #1851

[22] WEISSGLASS, J., Semigroup rings and semilattice sums of rings, *Proc. Amer. Math. Soc.* **39** (1973), 471–478. *MR* **48** #456

(Received August 25, 1987)

POSADSKAYA 56/1–6
EKATHERINBURG 620 086
RUSSIA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 133–142

DIE REGELFLÄCHEN DES E_n, DIE EINE AUS EBENEN KURVEN BESTEHENDE KONGRUENZSCHAR TRAGEN. I

H. WRESNIK

1. Die EK-Regelflächen des E_3 waren schon seit langer Zeit Gegenstand geometrischer Untersuchungen. So zeigten A. Mannheim [1] und A. Schönflies [5], daß die Scharkurven der EK-Schar einer Regelfläche Ellipsen (speziell Geraden) sind.

Obwohl bereits 1970 von H. Sachs [3], [4] nichttriviale Gegenbeispiele gefunden wurden, bemerkte erst H. Vogler [8], daß der obige Satz in dieser Form nicht richtig ist. Falls nämlich die Trägerebenen der Scharkurven einer EK-Schar zu einer raumfesten Geraden parallel sind, so müssen, wie H. Vogler zeigte, die Scharkurven keinesfalls Ellipsen sein. In allen anderen Fällen stimmt jedoch die Aussage von Mannheim und Schönflies und die Scharkurven sind durch den Bewegungsparameter affin aufeinander bezogen.

In den letzten beiden Jahren behandelte H. Vogler in einer Reihe von (bisher unveröffentlichten) Arbeiten analoge Fragestellungen im n-dimensionalen, reellen, affinen Raum A_n. Er betrachtete Affinbewegungen einer Geraden g, deren Punkte ebene Bahnen durchlaufen. Dabei zeigte es sich, daß die Scharkurven durch den Bewegungsparameter affin aufeinander bezogen sind, wenn ihre Trägerebenen keinen gemeinsamen Fernpunkt besitzen.

Die folgende Arbeit behandelt die EK-Regelflächen des E_n. Dazu sei im E_n ein kartesisches Normalkoordinatensystem gegeben, das bei der projektiven Erweiterung des E_n zu einem projektiven Raum P_n in natürlicher Weise ein projektives Koordinatensystem induziert. Des weiteren wollen wir den P_n, falls nötig, komplex erweitern.

2. Sei nun Φ eine EK-Regelfläche mit der Parameterdarstellung

$$\vec{\mathbf{x}}(u,v)=\vec{\mathbf{l}}(u)+v\,\vec{\mathbf{e}}(u) \tag{1}$$

mit $u\in I\subset R$, $\vec{\mathbf{l}}, \vec{\mathbf{e}}\in C^3$, $\vec{\mathbf{e}}^2\equiv 1$, $v\in R$, wobei die Scharkurven der EK-

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 51N20; Secondary 53A05.

Key words and phrases. Ruled surfaces, family of plane congruences.

[1] Weitere Arbeiten zu diesem Themenkreis stammen von K. Meirer [2] und dem Verfasser [9], [10].

Schar durch $v = \text{const.}$ gegeben seien. Dann gilt identisch in u und v [2]

$$\dot{\vec{\mathbf{x}}} \wedge \ddot{\vec{\mathbf{x}}} \wedge \dddot{\vec{\mathbf{x}}} = \vec{\mathbf{0}}\,. \tag{2}$$

Setzt man (1) in (2) ein und beachtet die Multilinearität des äußeren Produktes, so erhält man eine kubische Form in v, die identisch in v verschwindet. Daher verschwinden ihre Koeffizienten (identisch in u) und man erhält so die Bedingungen

$$\begin{gathered}
\dot{\vec{\mathbf{l}}} \wedge \ddot{\vec{\mathbf{l}}} \wedge \dddot{\vec{\mathbf{l}}} = \vec{\mathbf{0}} \\
\dot{\vec{\mathbf{l}}} \wedge \ddot{\vec{\mathbf{l}}} \wedge \dddot{\vec{\mathbf{e}}} + \dot{\vec{\mathbf{l}}} \wedge \ddot{\vec{\mathbf{e}}} \wedge \dddot{\vec{\mathbf{l}}} + \dot{\vec{\mathbf{e}}} \wedge \ddot{\vec{\mathbf{l}}} \wedge \dddot{\vec{\mathbf{l}}} = \vec{\mathbf{0}} \\
\dot{\vec{\mathbf{l}}} \wedge \ddot{\vec{\mathbf{e}}} \wedge \dddot{\vec{\mathbf{e}}} + \dot{\vec{\mathbf{e}}} \wedge \ddot{\vec{\mathbf{l}}} \wedge \dddot{\vec{\mathbf{e}}} + \dot{\vec{\mathbf{e}}} \wedge \ddot{\vec{\mathbf{e}}} \wedge \dddot{\vec{\mathbf{l}}} = \vec{\mathbf{0}} \\
\dot{\vec{\mathbf{e}}} \wedge \ddot{\vec{\mathbf{e}}} \wedge \dddot{\vec{\mathbf{e}}} = \vec{\mathbf{0}}\,.
\end{gathered} \tag{3}$$

(3_1) ist nach Voraussetzung erfüllt und (3_4) bedeutet, daß das sphärische Bild von Φ eben ist.

Sind nun α und β zwei verschiedene Trägerebenen von Scharkurven der EK-Schar, so spannen sie einen höchstens 4-dimensionalen Raum auf. Daher liegt Φ in einem höchstens 5-dimensionalen Teilraum des E_n, weshalb die folgenden Ansätze gemacht werden können.

$$\vec{\mathbf{x}}(u,v) = \begin{pmatrix} a_1(u)\cos\beta \\ 0 \\ a_1(u)\sin\beta \\ a_2(u) \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{4}$$

bzw.

$$\vec{\mathbf{x}}(u,v) = \begin{pmatrix} 0 \\ 0 \\ a_1(u)\cos\beta \\ a_1(u)\sin\beta \\ a_2(u) \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \\ 0 \end{pmatrix} \tag{5}$$

mit $a_1, a_2 \in C^3$, $\beta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $u \in I \subset R$, $v \in R$.[3]

3. Untersuchen wir zunächst den Fall (4), so erhalten wir aus $(3_{2,3})$

$$\begin{gathered}
\cos\beta[A_{12}(u)\cos u - A_{13}(u)\sin u - A_{23}(u)\cos u] = 0 \\
\sin\beta[A_{12}(u)\sin u + A_{13}(u)\cos u - A_{23}(u)\sin u] = 0 \\
\sin\beta[-A_{12}(u)\cos u + A_{13}(u)\sin u + A_{23}(u)\cos u] = 0
\end{gathered} \tag{6}$$

[2] Mit "·" bezeichnen wir die Ableitungen nach u.

[3] Mit (4) ist der 4-dimensionale und mit (5) der 5-dimensionale Fall erfaßt. Der 3-dimensionale Fall wurde wie in **1.** erwähnt, bereits ausreichend behandelt.

mit $A_{12}(u) = \dot{a}_1\ddot{a}_2 - \ddot{a}_1\dot{a}_2$ usf. und

$$\sin\beta(\dot{a}_1 + a_1) = 0, \quad \dot{a}_2 + a_2 = 0. \tag{7}$$

Aus (7_2) folgt somit für a_2 die Darstellung

$$a_2(u) = A_2 \cos u + B_2 \sin u + C_2 \tag{8}$$

mit $A_2, B_2, C_2 \in R$. Wegen (7_1) müssen nun die beiden folgenden Fälle betrachtet werden.

a) Ist $\beta \neq 0$, dann erhält man

$$a_1(u) = A_1 \cos u + B_1 \sin u + C_1 \tag{9}$$

mit $A_1, B_1, C_1 \in R$. Ist hingegen

b) $\beta = 0$, so kann die Funktion $a_1(u)$ in (4) beliebig gewählt werden.

Wie man nun leicht nachrechnet, erfüllen die Ansätze (8) und (9) für $a_2(u)$ und $a_1(u)$ auch die Gleichungen (6), weshalb man im Fall $\beta \neq 0$ als Parameterdarstellung der Lösungsflächen

$$\vec{\mathbf{x}}(u,v) = \begin{pmatrix} \cos\beta(A_1 \cos u + B_1 \sin u) \\ 0 \\ \sin\beta(A_1 \cos u + B_1 \sin u) \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{10}$$

erhält. Da Φ nicht in einem dreidimensionalen Teilraum liegen soll, muß $(A_2, B_2) \neq (0,0)$ und für $k = 0$ noch zusätzlich $\beta \neq 0$ und $(A_1, B_1) \neq (0,0)$ vorausgesetzt werden.

Betrachtet man nun den Fall $\beta = 0$ und setzt die Darstellung (8) für $a_2(u)$ in (6) ein, so erhält man die Bedingung

$$A_2(\dot{a}_1 + a_2) = 0. \tag{11}$$

Ist $A_2 \neq 0$, so erhält man für a_1 die Darstellung (9) und die zugehörigen Lösungsflächen sind durch (10) erfaßt.

Für den Fall $A_2 = 0$ hingegen folgt als Parameterdarstellung der Lösungsflächen

$$\vec{\mathbf{x}}(u,v) = \begin{pmatrix} a_1(u) \\ 0 \\ 0 \\ B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix}, \tag{12}$$

wobei $a_1 \in C^3$, k und B_2 von Null verschieden sein müssen.

4. Die analoge Vorgangsweise wie in **3.** liefert für den Ansatz (5) mit (3_2) die Gleichungen

$$\sin\beta(\dot{a}_1 + a_1) = 0, \quad \cos\beta(\dot{a}_1 + a_1) = 0, \quad \dot{a}_2 + a_2 = 0, \tag{13}$$

weshalb $a_1(u)$ und $a_2(u)$ die Darstellungen (9) und (8) besitzen müssen. Da mit diesen Ansätzen auch die Bedingung (3_3) erfüllt ist, erhält man als Parameterdarstellung der Lösungsflächen

$$\vec{x}(u,v)=\begin{pmatrix}0\\0\\\cos\beta(A_1\cos u+B_1\sin u)\\\sin\beta(A_1\cos u+B_1\sin u)\\A_2\cos u+B_2\sin u\end{pmatrix}+v\begin{pmatrix}\cos u\\\sin u\\k\\0\\0\end{pmatrix},\tag{14}$$

wobei $(A_i,B_i)\neq(0,0)$, $i=1,2$, $\beta\neq 0$ und im Fall $k=0$ noch zusätzlich $\beta\neq$ $\neq\frac{\pi}{2}$ gelten muß. Des weiteren darf auch, wie im folgenden gezeigt wird, der Ausdruck $A_1B_2-A_2B_1$ nicht verschwinden.

5. Wir wollen uns nun mit der von den Trägerebenen der Scharkurven der EK-Schar gebildeten 3-Regelfläche Ψ (im Sinne von [7]) beschäftigen. Die im folgenden verwendeten Begriffe aus der Theorie der verallgemeinerten Regelflächen können in [6] und [7] nachgelesen werden.

Für die EK-Regelflächen vom Typ (10) erhält man für die 3-Regelfläche Ψ die folgende Parameterdarstellung

$$\vec{x}(v,\lambda,\mu)=\begin{pmatrix}0\\0\\vk\\0\end{pmatrix}+\lambda\begin{pmatrix}A_1\cos\beta+v\\0\\A_1\sin\beta\\A_2\end{pmatrix}+\mu\begin{pmatrix}B_1\cos\beta\\v\\B_1\sin\beta\\B_2\end{pmatrix}.\tag{15}$$

Setzt man

$$\vec{e}_1:=(A_1\cos\beta+v,0,A_1\sin\beta,A_2)^t,\quad \vec{e}_2:=(B_1\cos\beta,v,B_1\sin\beta,B_2)^t,$$

so besitzt das asymptotisches Bündel $A(v)$ von Ψ wegen

$$\det(\vec{e}_1,\vec{e}_2,\dot{\vec{e}}_1\dot{\vec{e}}_2)=(A_2B_1-A_1B_2)\sin\beta\tag{16}$$

genau dann die Dimension 4 (für alle $v\in R$), wenn (16) nicht verschwindet[4]. Insbesondere kann in diesem Fall keine Scharkurve Teil einer Geraden sein. Da somit das asymptotisches Bündel $A(v)$ den Gesamtraum aufspannt und maximale Dimension besitzt, muß (15) eine Gratregelfläche mit einer Kehllinie g sein. Um eine Parameterdarstellung von g zu finden, bedenkt man, daß g nach [7] die Menge der singulären Punkte von Ψ ist. Dabei ist ein Punkt von Ψ genau dann singulär, wenn

$$\vec{e}_1\wedge\vec{e}_2\wedge\dot{\vec{x}}=\vec{0}\tag{17}$$

---

[4] Von nun an bedeute "$\cdot$" die Ableitung nach v.

mit $\dot{\vec{\mathbf{x}}} = (\lambda, \mu, k, 0)^t$ gilt. Dies führt nach kurzer Rechnung auf

$$(18) \qquad \lambda = \frac{k[(A_2B_1 - A_1B_2)\cos\beta - B_2v]}{(A_2B_1 - A_1B_2)\sin\beta}, \quad \mu = \frac{kvA_2}{(A_2B_1 - A_1B_2)\sin\beta},$$

woraus man für g die Parameterdarstellung

$$(19) \qquad g(v) = \begin{pmatrix} \frac{kB_2v^2}{(A_2B_1 - A_1B_2)\sin\beta} + 2k\cot\beta v \\ \frac{kA_2v^2}{(A_2B_1 - A_1B_2)\sin\beta} \\ 2kv \\ 0 \end{pmatrix}$$

erhält. Wegen $(A_2, B_2) \neq (0,0)$ handelt es sich dabei für $k \neq 0$ um eine Parabel in der Ebene

$$(20) \qquad \vec{\mathbf{x}}(\lambda,\mu) = \lambda \begin{pmatrix} \cot\beta \\ 0 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -B_2 \\ A_2 \\ 0 \\ 0 \end{pmatrix},$$

für $k = 0$ hingegen ergibt sich eine fastkegelige Gratregelfläche, für die der Punkt $(0,0,0,0)^t$ der einzige kegelige Punkt ist.

Betrachten wir nun den Fall, für den (16) verschwindet und damit die Dimension des asymptotischen Bündels $A(v)$ von Ψ kleiner als 4 ist. Zunächst wollen wir allerdings untersuchen, ob es Scharkurven gibt, die Teil einer Geraden sind. Die dazu gehörenden v-Werte wollen wir dann in (15) ausschließen, da sonst die 3-Regelfläche Ψ nicht definiert ist. Obiges tritt genau dann ein, wenn $\vec{\mathbf{e}}_1$ und $\vec{\mathbf{e}}_2$ linear abhängig sind, was, wie man leicht nachrechnet, mit

$$(21) \qquad \begin{array}{l} (A_1\cos\beta + v)v = 0, \quad B_1\sin\beta v = 0, \quad A_2v = 0 \\ (A_1B_2 - A_2B_1)\cos\beta + B_2v = 0, \quad A_1\sin\beta v = 0 \end{array}$$

gleichwertig ist. Somit ist eine Scharkurve genau dann eben, wenn

a) $A_1B_2 - A_2B_1 = 0$ gilt — die entsprechende Scharkurve gehört zu $v = 0$ — oder wenn

b) $A_1B_2 - A_2B_1 \neq 0$ und $A_2 = 0$ gilt — die entsprechende Scharkurve gehört zu $v = -A_1$.

Nun untersuchen wir das asymptotisches Bündel $A(v)$ von Ψ, das von den Vektoren $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \dot{\vec{\mathbf{e}}}_1, \dot{\vec{\mathbf{e}}}_2$ aufgespannt wird. Da sie nach Voraussetzung linear abhängig sind, betrachten wir die beiden Vektortripel $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \dot{\vec{\mathbf{e}}}_1$ bzw. $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \dot{\vec{\mathbf{e}}}_2$. Wie man leicht nachrechnet, sind die Vektoren des ersten Tripels genau dann linear abhängig, wenn

$$(22) \qquad A_1 = A_2 = 0 \text{ oder } \beta = A_2 = 0$$

gilt, während man für die lineare Abhängigkeit der Vektoren des zweiten Tripels die folgenden Bedingungen erhält:

Ist $A_1B_2 - A_2B_1 \neq 0$, so sind die Vektoren höchstens für $v = \frac{(A_2B_1 - A_1B_2)\cos\beta}{B_2}$ linear abhängig, während für $A_2B_1 - A_1B_2 = 0$ die Bedingungen

$$B_1 = B_2 = 0 \text{ oder } \beta = B_2 = 0 \tag{23}$$

lauten.

Wir wollen, da das asymptotisches Bündel $A(v)$ konstante Dimension besitzen soll, auch $v = \frac{(A_2B_1 - A_1B_2)\cos\beta}{B_2}$ für die weiteren Betrachtungen ausschließen. Wegen $(A_2, B_2) \neq (0,0)$ erkennt man, daß nicht beide Vektortripel gleichzeitig linear abhängig sein können, weshalb die Dimension des asymptotischen Bündels $A(v)$ drei ist.

Betrachten wir nun zuerst den Fall, daß $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \dot{\vec{\mathbf{e}}}_1$ linear unabhängig sind, dann gilt wegen (22), (16) und $(A_2, B_2) \neq (0,0)$ auch $A_2 \neq 0$. Damit ist für $k \neq 0$ — im Fall $k = 0$ ist Ψ fastkegelig — auch $\det(\dot{\vec{\mathbf{l}}}, \vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \dot{\vec{\mathbf{e}}}_1) = vkA_2$ von Null verschieden, weshalb Ψ eine 2-Zentralregelfläche besitzt.[5] Um eine Parameterdarstellung derselben zu finden, berücksichtigt man, daß ein Punkt $\vec{\mathbf{x}}$ von Ψ genau dann Zentralpunkt ist, wenn der Tangentenvektor $\dot{\vec{\mathbf{x}}} = \dot{\vec{\mathbf{l}}} + \dot{\lambda}\vec{\mathbf{e}}_1 + \lambda\dot{\vec{\mathbf{e}}}_1 + \dot{\mu}\vec{\mathbf{e}}_2 + \mu\dot{\vec{\mathbf{e}}}_2$ aus der linearen Hülle der Vektoren $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \vec{\mathbf{n}}$ ist, wobei sich der zu $A(v)$ orthogonale Vektor $\vec{\mathbf{n}}$ zu

$$\vec{\mathbf{n}} = \left(0, 0, 1, -\frac{A_1}{A_2}\sin\beta\right)^t \tag{24}$$

berechnet. Dies zeigt insbesondere, daß Ψ (streng)konoidal ist.

Hiermit ergibt sich nach kurzer Rechnung für $\dot{\vec{\mathbf{l}}}$ die Zerlegung $\dot{\vec{\mathbf{l}}} = \alpha_1\vec{\mathbf{e}}_1 + \alpha_2\vec{\mathbf{e}}_2 + \alpha_3\dot{\vec{\mathbf{e}}}_1 + \alpha_4\vec{\mathbf{n}}$ mit

$$\begin{aligned} \alpha_1 &= \frac{kA - 2\sin\beta}{A_1^2\sin^2\beta + A_2^2} & \alpha_2 &= 0 \\ \alpha_3 &= -\frac{kA_1\sin\beta(A_1\cos\beta + v)}{A_1^2\sin^2\beta + A_2^2} & \alpha_4 &= \frac{kA_2^2}{A_1^2\sin^2\beta + A_2^2} \end{aligned} \tag{25}$$

und wegen $\dot{\vec{\mathbf{e}}}_2 = \beta_1\vec{\mathbf{e}}_1 + \beta_2\vec{\mathbf{e}}_2 + \beta_3\dot{\vec{\mathbf{e}}}_1$ mit

$$\beta_1 = -\frac{B_2}{A_2v}, \quad \beta_2 = \frac{1}{v}, \quad \beta_3 = \frac{\cos\beta}{A_2v}(A_2B_1 - A_1B_2) + \frac{B_2}{A_2} \tag{26}$$

[5] $\mathbf{l}$ ist dabei die in der Darstellung (15) von Ψ verwendete Leitgerade.

sind, wie man durch Einsetzen in die Darstellung von $\dot{\vec{x}}$ erkennt, die Zentralpunkte durch

$$\alpha_3 + \lambda + \beta_3 \mu = 0 \tag{27}$$

gekennzeichnet. Damit erhält man als Parameterdarstellung der 2-Zentralregelfläche von Ψ

(28)
$$\vec{x}(v,\mu) = \begin{pmatrix} \alpha_3(A_1 \cos\beta + v) \\ 0 \\ kv + \alpha_3 A_1 \sin\beta \\ \alpha_3 A_2 \end{pmatrix} + \mu \begin{pmatrix} B_1 \cos\beta - \beta_3(A - 1\cos\beta + v) \\ v \\ (B_1 - \alpha_3 A_1)\sin\beta \\ B_2 - \alpha_3 A_2 \end{pmatrix}.$$

Sind jedoch die Vektoren $\vec{e}_1, \vec{e}_2, \dot{\vec{e}}_1$ linear abhängig, dann müssen, wie bereits gezeigt, die Vektoren $\vec{e}_1, \vec{e}_2, \dot{\vec{e}}_2$ linear unabhängig sein. Daher berechnet sich die Determinante der Vektoren $\dot{\vec{l}}, \vec{e}_1, \vec{e}_2, \dot{\vec{e}}_2$ zu

$$\det(\dot{\vec{l}}, \vec{e}_1, \vec{e}_2, \dot{\vec{e}}_2) = k[(A_1 B_2 - A_2 B_1)\cos\beta + B_2 v]. \tag{29}$$

Für $k = 0$ ist Ψ fastkegelig und für $k \neq 0$ ist im Falle $A_2 B_1 - A_1 B_2 = 0$ wegen $B_2 \neq 0$ die Determinante (29) von Null verschieden. Ist aber $A_2 B_1 - A_1 B_2 \neq$ $\neq 0$, dann verschwindet (29) höchstens für $v = \frac{(A_2 B_1 - A_1 B_2)\cos\beta}{B_2}$, das wir aber bereits oben ausgeschlossen haben. Daher besitzt Ψ auch in diesem Fall eine 2-Zentralregelfläche.

Da sich der Normalvektor $\vec{n}$ des asymptotischen Bündels $A(v)$ zu

$$\vec{n} = \left(0, 0, 1 - \frac{B_1}{B_2}\sin\beta\right)^t \tag{30}$$

berechnet und nicht von v abhängt, ist Ψ auch hier (streng)konoidal. Mit den Zerlegungen $\dot{\vec{l}} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \dot{\vec{e}}_2 + \alpha_4 \vec{n}$ mit

$$\begin{aligned}
\alpha_1 &= -\frac{kB_2 \sin\beta\cos\beta}{(B_1^2 \sin^2\beta B_2^2)[(A_1 B_2 - A_2 B_1)\cos\beta + vB_2]} \\
\alpha_2 &= \frac{kB_1 B_2 \sin\beta(A_1\cos\beta + v)}{(B_1^2 \sin^2\beta + B_2^2)[(A_1 B_2 - A_2 B_1)\cos\beta + vB_2]} \\
\alpha_3 &= -\frac{vkB_1 B_2 \sin\beta(A_1\cos\beta + v)}{(B_1^2 \sin^2\beta + B_2^2)[(A_1 B_2 - A_2 B_1)\cos\beta + vB_2]} \\
\alpha_4 &= \frac{kB_2^2}{B_1^2 \sin^2\beta + B_2^2}
\end{aligned} \tag{31}$$

und $\dot{\vec{\mathbf{e}}}_1 = \beta_1 \vec{\mathbf{e}}_1 + \beta_2 \vec{\mathbf{e}}_2 + \beta_3 \dot{\vec{\mathbf{e}}}_2$ mit

$$\begin{aligned} \beta_1 &= \frac{B_2}{(A_1B_2 - A_2B_1)\cos\beta + vB_2} \qquad \beta_2 = -\frac{A_2}{(A_1B_2 - A_2B_1)\cos\beta + vB_2} \\ \beta_3 &= \frac{vA_2}{(A_1B_2 - A_2B_1)\cos\beta + vB_2} \end{aligned} \tag{32}$$

folgt analog zu obigem als Parameterdarstellung der 2-Zentralregelfläche von Ψ:

$$\vec{\mathbf{x}}(v,\lambda) = \begin{pmatrix} -\alpha_3 B_1 \cos\beta \\ -\alpha_3 v \\ vk - \alpha_3 B_1 \sin\beta \\ -\alpha_3 B_2 \end{pmatrix} + \lambda \begin{pmatrix} A_1\cos\beta + v - \beta_3 B_1 \cos\beta \\ -\beta_3 v \\ (A_1 - \beta_3 B_1)\sin\beta \\ A_2 - \beta_3 B_2 \end{pmatrix}. \tag{33}$$

Wenden wir uns nun den Regelflächen vom Typ (12) zu, so erhält man für die 3-Regelfläche Ψ die Parameterdarstellung

$$\vec{\mathbf{x}}(v,\lambda,\mu) = \begin{pmatrix} 0 \\ 0 \\ vk \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ v \\ 0 \\ B_2 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{34}$$

Man erkennt daraus unmittelbar, daß alle Trägerebenen zur Geraden $y = z = = t = 0$ parallel sind. Für den Normalvektor $\vec{\mathbf{n}}$ des asymptotischen Bündels $A(v)$, aufgespannt von den Vektoren $\vec{\mathbf{e}}_1 := (0, v, 0, B_2)^t$, $\vec{\mathbf{e}}_2 := (1,0,0,0)^t$, $\dot{\vec{\mathbf{e}}}_1$ findet man $\vec{\mathbf{n}} = (0,0,1,0)^t$, weshalb auch hier Ψ (streng)konoidal ist. Desweiteren ergibt sich daraus sofort die Zerlegung $\dot{\vec{\mathbf{l}}} = k\vec{\mathbf{n}}$ für den Tangentenvektor der Leitkurve $\vec{\mathbf{l}} := (0,0,vk,0)^t$ von Ψ.

Daher besitzt Ψ eine 2-Zentralregelfläche, für die man die Parameterdarstellung

$$\vec{\mathbf{x}}(v,\mu) = \begin{pmatrix} 0 \\ 0 \\ vk \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{35}$$

erhält. Es handelt sich dabei also um ein Parallelstrahlbüschel in der Ebene $y = t = 0$.

Zum Abschluß unserer Überlegungen betrachten wir die EK-Regelflächen mit der Parameterdarstellung (14). Für die 3-Regelfläche Ψ der Scharkurventrägerebenen ergibt sich daraus

$$\vec{\mathbf{x}}(v,\lambda,\mu) = \begin{pmatrix} 0 \\ 0 \\ vk \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} v \\ 0 \\ A_1\cos\beta \\ A_1\sin\beta \\ A_2 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ v \\ B_1\cos\beta \\ B_1\sin\beta \\ B_2 \end{pmatrix}. \tag{36}$$

Das asymptotisches Bündel $A(v)$, aufgespannt von den Vektoren $\vec{e}_1 :=$ $:= (v, 0, A_1 \cos\beta, A_1 \sin\beta, A_2)^t$, $\vec{e}_2 := (0, v, B_1 \cos\beta, B_1 \sin\beta, B_2)^t$, $\dot{\vec{e}}_1$, $\dot{\vec{e}}_2$ besitzt genau dann die Dimension vier, wenn

$$A_2 B_1 - A_1 B_2 \neq 0 \tag{37}$$

gilt.

Ist (37) nicht erfüllt, dann überzeugt man sich unschwer, daß alle Trägerebenen zur Richtung $(B_1, -A_1, 0, 0, 0)^t$ parallel sind, weshalb die EK-Regelfläche Φ sogar in einem 4-dimensionalen Teilraum liegt und bereits diskutiert wurde.

Gilt hingegen die Bedingung (37), so besitzt Ψ wegen

$$\det\left((0,0,k,0,0)^t, \vec{e}_1, \vec{e}_2, \dot{\vec{e}}_1, \dot{\vec{e}}_2\right) = k(A_2 B_1 - A_1 B_2)\sin\beta$$

für $k \neq 0$ eine Striktionslinie s, während Ψ im Fall $k = 0$ fastkegelig ist. Desweiteren ist Ψ, da der Normalenvektor $\vec{n} = (0, 0, \sin\beta, -\cos\beta, 0)^t$ des asymptotischen Bündels $A(v)$ nicht von v abhängt, (streng)konoidal.

Mit der Zerlegung $\dot{\vec{l}} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \dot{\vec{e}}_1 + \alpha_4 \dot{\vec{e}}_2 + \alpha_5 \vec{n}$ der Leitlinie l von Ψ mit

$$\begin{aligned} &\alpha_1 = \frac{k B_2 \cos\beta}{A_2 B_1 - A_1 B_2} \qquad \alpha_2 = -\frac{k A_2 \cos\beta}{A_1 B_2 - A_2 B_1} \\ &\alpha_3 = -\frac{v k B_2 \cos\beta}{A_1 B_2 - A_2 B_1} \qquad \alpha_4 = \frac{v k A_2 \cos\beta}{A_1 B_2 - A_1 B_2} \qquad \alpha_5 = k \sin\beta \end{aligned} \tag{38}$$

findet man für die Zentralpunkte von Ψ die Bedingungen

$$\lambda + \alpha_3 = 0, \quad \mu + \alpha_4 = 0 \tag{39}$$

und damit ergibt sich als Parameterdarstellung der Striktionslinie s

$$\vec{s}(v) = \begin{pmatrix} \frac{v^2 k B_2 \cos\beta}{A_1 B_2 - A_2 B_1} \\ -\frac{v^2 k A_2 \cos\beta}{A_1 B_2 - A_1 B_2} \\ vk(1+\cos^2\beta) \\ vk \sin\beta \cos\beta \\ 0 \end{pmatrix}, \tag{40}$$

die wegen $(A_2, B_2) \neq (0,0)$ für $\beta \neq \frac{\pi}{2}$ und $k \neq 0$ eine Parabel in der Ebene

$$\vec{x}(\lambda, \mu) = \lambda \begin{pmatrix} B_2 \\ -A_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 1+\cos^2\beta \\ \sin\beta \cos\beta \\ 0 \end{pmatrix} \tag{41}$$

ist und für $\beta = \frac{\pi}{2}$ mit der dritten Koordinatenachse zusammenfällt. Fassen wir die Ergebnisse zusammen, so erhält man

SATZ 1. *Eine EK-Regelfläche* Φ *des* E_n *liegt bereits in einem Teilraum* E_k *des* E_n, *wobei* $k \leqq 5$ *gilt. Ist* $k = 4$, *so besitzt* Φ *die Parameterdarstellungen* (10) *oder* (12) *und im Falle* $k = 5$ *die Parameterdarstellung* (14).

Die von den Trägerebenen der Scharkurven der EK-Schar gebildete 3-Regelfläche Ψ *ist für* $k = 0$ *fastkegelig und für* $k \neq 0$ *im Fall* (10) *eine (1,3)-Gratregelfläche mit einer Parabel als Gratlinie oder eine 3-Regelfläche mit einer 2-Zentralregelfläche.*

Im Fall (12) *ist* Ψ *eine 3-Regelfläche mit einem Parallelstrahlbüschel als 2-Zentralregelfläche und im Fall* (14) *eine 3-Regelfläche mit einer Parabel als Striktionslinie.*

Eine eingehende Diskussion der EK-Regelflächen (10), (12) und (14) und die Betrachtung spezieller EK-Regelflächen soll in einer zweiten Arbeit durchgeführt werden.

LITERATURVERZEICHNIS

[1] MANNHEIM, A., Propriétés relatives aux trajectoires des points d'une figure de forme invariable, *C. R. Acad. Sci. Paris* **76** (1873), 635–638.

[2] MEIRER, K., Über windschiefe Flächen mit ausgezeichneten Scharen ebener Kurven, *First Geometry Colloquium* (Graz, 1980), Ber. no. 164, Forschungszentrum Graz, Graz, 1981. *MR* **82j**: 53006

[3] SACHS, H., Die Strahlflächen mit durchwegs ebenen Orthogonaltrajektorien der Erzeugenden, *Arch. Math. (Basel)* **21** (1970), 437–445. *MR* **43** #1057

[4] SACHS, H., Über die Kurven konstanten Striktionsabstandes auf windschiefen Flächen, *Mh. Math.* **74** (1970), 445–461. *MR* **43** #3929

[5] SCHÖNFLIES, A., *Geometrie der Bewegung in synthetischer Darstellung*, Teubner, Leipzig, 1886. *Jahrbuch Fortschr. Math.* **18**, 817

[6] FRANK, H. und GIERING, O., Regelflächen mit Zentralräumen, *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **187** (1978), 139–163. *MR* **80k**: 53004

[7] FRANK, H. und GIERING, O., Verallgemeinerte Regelflächen, *Math. Z.* **150** (1976), 261–271. *MR* **54** #5984

[8] VOGLER, H., Räumliche Zwangläufe mit ebenen Bahnkurven, Second Zagreb–Graz Meeting of Mathematicians (Plitvice, 1980), Ber. no. 162, Forschungszentrum Graz, Graz, 1981. *MR* **82j**: 53008

[9] WRESNIK, H., *Regelflächen mit ebenen Kongruenzscharen*, Dissertation, Technische Universität Graz, Graz, 1988.

[10] WRESNIK, H., Regelflächen mit einem ebenen Fußpunktkurvenbüschel, welche die Erzeugenden nach kongruenten Punktreihen schneiden.

(Eingegangen am 22. Juni 1988)

INSTITUT FÜR GEOMETRIE
TECHNISCHE UNIVERSITÄT GRAZ
KOPERNIKUSGASSE 24
A–8010 GRAZ
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 143–146

COMMON FIXED POINT RESULTS FOR ITERATIONS IN METRIC LINEAR SPACES

LIAQAT ALI KHAN

Abstract

In this paper we obtain some common fixed point results for Mann iterates of two self-mappings on a metric linear space under various contractive conditions.

In ([1], [3], [4]), it has been shown that for a self-mapping T on a normed space X satisfying various contractive conditions, if the sequence of Mann iterates associated with T converges, it converges to a fixed point of T. These results have recently been extended by the author in [2] to the case of metric linear spaces. In this paper we consider two self-mappings S and T on a metric linear space X and show that if the sequence of Mann iterates associated with S or T converges, it converges to a common fixed point of S and T.

In the sequel we assume that the topology of X is generated by an *F-norm* q which has the following properties:

(a) $q(x) \geqq 0$, and $q(x) = 0$ iff $x = 0$;

(b) $q(x+y) \leqq q(x) + q(y)$;

(c) $q(rx) \leqq q(x)$ for all (real or complex) scalars r with $|r| \leqq 1$;

(d) If $r_n \to r$ and $x_n \to x$, then $q(r_n x_n - rx) \to 0$.

For any $x_0 \in X$, we consider the *Mann iterative process* associated with S as $x_{n+1} = (1-c_n)x_n + c_n S x_n$ for $n > 0$, where $\{c_n\}_{n=0}^{\infty}$ satisfies (i) $c_0 = 1$, (ii) $0 \leqq c_n \leqq 1$ for $n > 0$, (iii) there exists an integer $N > 1$ and a constant $r > 0$ such that $r \leqq c_n$ for all $n \geqq N$. We mention that the condition (iii) here is less restrictive than the corresponding conditions considered in [1], [3], [4].

THEOREM 1. *Let S and T be self-mappings on X satisfying at least one of the following conditions:*

(1) $q(x - Sx) + q(y - Ty) \leqq \alpha q(x - y)$, $\alpha > 0$;

(2) $q(x - Sx) + q(y - Ty) \leqq \beta[kq(x - y) + q(x - Ty) + q(y - Sx)]$; $0 \leqq \beta < 1$ *and* $k > 0$;

(3) $q(Sx - Ty) + q(x - Sx) + q(y - Ty) \leqq \gamma[q(x - Ty) + q(y - Sx)]$; $0 \leqq \gamma < 2$,

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 47H10; Secondary 46A06.

Key words and phrases. Metric linear space, F-norm, contractive mappings, Mann iterates, common fixed point.

(4) $q(Sx-Ty) \leqq \delta \max\{kq(x-y), q(x-Sx)+q(y-Ty), q(x-Ty)+ +q(y-Sx)\}$, $0 \leqq \delta < 1$ *and* $k > 0$,
for all $x, y \in X$. *If, for some* $x_0 \in X$, *the sequence* $\{x_n\}_{n=0}^{\infty}$ *of Mann iterates associated with* S *or* T, *converges to a point* $u \in X$, *then* u *is a common fixed point of* S *and* T.

PROOF. Suppose $x_{n+1} = (1-c_n)x_n + c_n S x_n$ for $n > 0$, with $\lim_{n\to\infty} x_n = u$. Choose an integer $N \geqq 1$ and a constant $r > 0$ such that $r \leqq c_n$ for all $n \geqq N$. Then, for $n \geqq N$,

$$\begin{aligned} q(Tu-u) &\leqq q(Tu-Sx_n)+q(Sx_n-x_n)+q(x_n-u) \leqq \\ &\leqq q(Tu-Sx_n)+q(r^{-1}(x_{n+1}-x_n))+q(x_n-u). \end{aligned} \tag{5}$$

If S and T satisfy (1), then

$$q(Tu-Sx_n) \leqq q(Tu-u)+q(u-x_n)+q(x_n-Sx_n) \leqq (\alpha+1)q(u-x_n).$$

If S and T satisfy (2), then

$$\begin{aligned} q(Tu-Sx_n) &\leqq q(Tu-u)+q(u-x_n)+q(x_n-Sx_n) \leqq \\ &\leqq \beta[kq(u-x_n)+q(u-Sx_n)+q(x_n-Tu)]+q(u-x_n) \leqq \\ &\leqq (\beta k+\beta+1)q(u-x_n)+\beta q(r^{-1}(x_{n+1}-x_n))+\beta q(x_n-Tu). \end{aligned}$$

If S and T satisfy (3), then

$$\begin{aligned} q(Tu-Sx_n) &\leqq \gamma[q(u-Sx_n)+q(x_n-Tu)]-q(u-Tu)-q(x_n-Sx_n) \leqq \\ &\leqq \gamma q(u-x_n)+(\gamma-1)q(r^{-1}(x_{n+1}-x_n))+\gamma q(x_n-Tu)-q(u-Tu). \end{aligned}$$

If S and T satisfy (4), then

$$\begin{aligned} q(Tu-Sx_n) &\leqq \delta \max\{kq(u-x_n), q(u-Tu)+q(x_n-Sx_n), \\ &\qquad q(u-Sx_n)+q(x_n-Tu)\} \leqq \\ &\leqq \delta \max\{kq(u-x_n), q(u-Tu)+q(r^{-1}(x_{n+1}-x_n)), \\ &\qquad q(u-x_n)+q(r^{-1}(x_{n+1}-x_n))+q(x_n-Tu)\}. \end{aligned}$$

Substituting the values of $q(Tu-Sx_n)$ in (5) and letting $n \to \infty$, we obtain

$$q(Tu-u) \leqq \lambda q(Tu-u),$$

where $\lambda = \max\{\beta, \gamma-1, \delta\} < 1$. Hence $Tu = u$.

Now, to show that $Su = u$, we apply again each of (1)–(4) and obtain

$$\begin{aligned} q(Su-u) = q(Su-u)+q(Tu-u) &\leqq \alpha q(u-u) = 0, \\ q(Su-u) = q(Su-u)+q(Tu-u) &\leqq \beta q(Su-u), \\ q(Su-u) = q(Su-Tu) &\leqq (\gamma-1)q(Su-u), \end{aligned}$$

and

$$q(Su-u) = q(Su-Tu) \leqq \delta q(Su-u),$$

respectively. Thus $Su = u$, and this completes the proof.

Regarding the uniqueness of the fixed point, we have the following results.

COROLLARY 2. *Under the hypothesis of Theorem 1, suppose that in place of conditions* (1)–(4), *S and T satisfy at least one of the following conditions:*
(6) $q(Sx - Ty) + q(x - Sx) + q(y - Ty) \leqq \gamma[q(x - Ty) + q(y - Sx)]$, $0 \leqq \gamma < \frac{1}{2}$,
(7) $q(Sx - Ty) \leqq \delta \max\{q(x - y), q(x - Sx) + q(y - Ty), q(x - Ty) + q(y - Sx)\}$, $0 \leqq \delta < 1$,
for all $x, y \in X$. Then u is the unique fixed point of S and T.

PROOF. If S and T satisfy (6) and (7), then they also satisfy conditions (3) and (4), respectively, of Theorem 1 with $k = 1$. Therefore $Su = u = Tu$. For uniqueness, suppose that $Sv = v = Tv$ for some v $(\neq u) \in X$. Using (6), we can write

$$\begin{aligned} q(u - v) = q(Su - Ty) &\leqq \\ \leqq \gamma[q(u - Tv) + q(v - Su)] - q(u - Su) - q(v - Tv) &= \\ = 2\gamma q(u - v). \end{aligned}$$

Since $0 \leqq 2\gamma < 1$, we have $u = v$. Similarly, using (7), we obtain

$$q(u - v) = q(Su - Tv) \leqq \delta q(u - v),$$

and so $u = v$, as required.

COROLLARY 3. *Under the hypothesis of Theorem 1, suppose, in addition, that at least one of the following strict inequality conditions holds:*
(A) $q(u - Sx) < q(u - x) + q(x - Sx)$;
(B) $q(u - Tx) < q(u - x) + q(x - Tx)$;
(C) $q(u - x) < q(u - Sx) + q(Sx - x)$;
(D) $q(u - x) < q(u - Tx) + q(Tx - x)$
for all x $(\neq u) \in X$. Then u is the unique common fixed point of S and T.

PROOF. By Theorem 1, $Su = u = Tu$. Suppose also that $Sv = v = Tv$ for some v $(\neq u) \in X$. Using (A), we have

$$q(u - v) = q(u - Sv) < q(u - v) + q(v - Sv) = q(u - v),$$

and so $u = v$. Similarly, the other conditions also imply that $u = v$.

REFERENCES

[1] DOTSON, JR., W. G., On the Mann iterative process, *Trans. Amer. Math. Soc.* **149** (1970), 65–73. *MR* **41** #2477

[2] KHAN, L. A., Fixed point theorems for Mann iterates in metrizable linear topological spaces, *Math. Japon.* **33** (1988), 247–251. *MR* **89g**: 47079

[3] RHOADES, B. E., Fixed point iterations using infinite matrices, *Trans. Amer. Math. Soc.* **196** (1974), 161–176. *MR* **50** #1063

[4] Rhoades, B. E., Extensions of some fixed point theorems of Ćirić, Maiti, and Pal, *Math. Sem. Notes Kobe Univ.* **6** (1978), 41–46. *MR* **58** #12985

(Received July 10, 1988)

DEPARTMENT OF MATHEMATICS
FEDERAL GOVERNMENT COLLEGE
H–8 ISLAMABAD
PAKISTAN

Present address:

DEPARTMENT OF MATHEMATICS
QUAID-I-AZAM UNIVERSITY
PAK–45320 ISLAMABAD
PAKISTAN

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 147–158

ON MAHLER'S APPROXIMATION FUNCTION FOR POWERS OF CERTAIN ALGEBRAIC NUMBERS

C. ELSNER

Abstract

Let $w_n(\zeta)$ denote Mahler's approximation function. The following theorems are proved:

i) For an integer m and some positive integers s and t with $(s,t)=1$, $t<s$ and $s \geqq 3$ we have

$$w_{s-2}(\zeta)=w_{s-2}(\zeta^t)$$

for $\zeta \in \mathbb{C}$ with $\zeta^s - m = 0$ and $\deg \zeta = s$.

ii) Let n be a positive integer greater than 2. If b and c are integers with $b \neq 0$ and if m is any positive integer with $m \mid n$, we prove

$$w_{n-2}(\zeta)=w_{n-2}(\zeta^m)$$

for each ζ satisfying $\zeta^n + b\zeta + c = 0$ and $\deg \zeta = \deg \zeta^m$.

Both equations are deduced from a theorem, which allows to prove more general results for $w_n(\zeta)$.

1. Introduction

If P is a polynomial with integer coefficients we write $H(P)$ to denote the usual height of the polynomial P.

In his famous papers [1] Mahler introduced a function to measure the approximation of complex numbers by algebraic numbers whose degree does not exceed a fixed integer n: For $\zeta \in \mathbb{C}$ and $n \in \mathbb{N}$ the positive real number $w_n(\zeta)$ denotes the supremum of all positive numbers w_n, such that there are infinitely many polynomials $P \in \mathbb{Z}[x]$ satisfying $\deg P \leqq n$ and $|P(\zeta)| \leqq H(P)^{-w_n}$. This is equivalent to the following definition of $w_n(\zeta)$: Let

$$M(n,H;\zeta)=\{P \in \mathbb{Z}[x]:\ \deg P \leqq n,\ \ H(P) \leqq H,\ \ P(\zeta) \neq 0\}.$$

Then

$$w_n(H,\zeta) := \min_{P \in M(n,H;\zeta)} |P(\zeta)|, \tag{1.1}$$

1980 *Mathematics Subject Classifications.* Primary 11J82; Secondary 11J68.

Key words and phrases. Mahler's approximation function, approximation by algebraic numbers.

Akadémiai Kiadó, Budapest

$$(1.2) \qquad w_n(\zeta) := \overline{\lim_H} \frac{-\log w_n(H,\zeta)}{\log H}.$$

The most important relations for $w_n(\zeta)$ are tabulated in (1.3).

If ζ is algebraic, we write $s = \deg \zeta$, otherwise $s = \infty$

	ζ without restrictions	ζ real	ζ not real
s without restrictions	$0 \leqq w_1(\zeta) \leqq w_2(\zeta) \leqq \leqq \ldots \leqq \infty$ (1)		$w_1(\zeta) = 0$ (2)
s and n without restrictions	$w_n(\zeta) \leqq s-1$ (3)		$w_n(\zeta) \leqq \frac{s-2}{2}$ (4)
$s < \infty$, n without restrictions	$w_n(\zeta) \leqq 2n-1$ (5)	$w_1(\zeta) = 1$ $(s > 1)$ (6)	$w_n(\zeta) \leqq n-1$ (7)
$n < s$		$w_n(\zeta) \geqq n$ (8)	$w_n(\zeta) \geqq \frac{n-1}{2}$ (9)
$n \geqq s-1 \geqq 1$		$w_n(\zeta) = s-1$ (10)	$w_n(\zeta) = \frac{s-2}{2}$ (11)

Table 1.3

A proof for (1), (3), (8) and (9) can be found in chapter III of [2]; (2) follows from $w_1(H,\zeta) \geqq \min\{1; |\mathrm{Im}\,\zeta|\} > 0$ $(H \in \mathbb{N})$; (6) can be deduced from (5), (8) or the theorem of Thue–Siegel–Roth; Theorem 2 and Theorem 2.2 in [3] imply (5), (7); and (10), (11) can be proved with (1), (3), (4) and two equations from [4], which were first deduced by E. Wirsing from some of his deep estimates for $w_n(\zeta)$.

With regard to Table (1.3) the following conjecture seems reasonable:

Let ζ and η be algebraic numbers, and furthermore we assume that they are either both real or both non-real. Then we have $w_n(\zeta) = w_n(\eta)$ for all $n \in \mathbb{N}$. From (2), (6) in Table 1.3 and from the condition $n \geqq s-1 \geqq 1$ for (10) and (11) it is easy to see that this conjecture will only be of interest if $2 \leqq n \leqq s-2$, $s \geqq 4$. Now let ζ and η be both transcendental numbers. If there are integers $a_0, \ldots, a_m$ and a positive integer m such that $\sum_{\mu=0}^{m} a_\mu \eta^\mu = \zeta$, we obtain from (14) and (15) in [2; chapter III]:

$$(1.4) \qquad w_{nm}(\eta) \geqq w_n(\zeta) \geqq \frac{w_n(\eta) - m + 1}{m} \quad \text{for all} \quad n \in \mathbb{N}.$$

2. Some lemmas and obvious applications

LEMMA 1. *Let $\zeta \in \mathbb{C} \setminus \{0\}$. Then we have*

$$w_n(\zeta) = w_n\left(\frac{1}{\zeta}\right)$$

for every positive integer n.

PROOF. If

$$P(x)=\sum_{\nu=0}^{n} a_\nu x^\nu \text{ and } Q(x)=\sum_{\nu=0}^{n} a_{n-\nu} x^\nu$$

are polynomials with integer coefficients, we obtain $|P(\zeta)| = |\zeta|^n \left|Q\left(\frac{1}{\zeta}\right)\right|$; and the assertion immediately follows from (1.1) and (1.2).

LEMMA 2. *Let $\zeta \in \mathbb{C}$; $q_1, q_2 \in \mathbb{Q}$; $q_1 \neq 0$. Then we have*

$$w_n(\zeta) = w_n(q_1\zeta + q_2)$$

for every positive integer n.

PROOF. Let $H \in \mathbb{N}$. Now we choose integers $a_0, \dots, a_n$ such that

$$w_n(H, q_1\zeta + q_2) = \left|\sum_{\nu=0}^{n} a_\nu (q_1\zeta + q_2)^\nu\right|.$$

Furthermore let $z \in \mathbb{N}$ satisfy $z(q_1q_2) \in \mathbb{Z}$ (if $q_2 \neq 0$) or $zq_1 \in \mathbb{Z}$ (if $q_2 = 0$). Then

$$z^n w_n(H, q_1\zeta + q_2) = \left|\sum_{\mu=0}^{n}\Bigl(\sum_{\nu=\mu}^{n} a_\nu \binom{\nu}{\mu} z^n q_1^\mu q_2^{\nu-\mu}\Bigr)\zeta^\mu\right|. \tag{2.1}$$

For every μ with $0 \leqq \mu \leqq n$ we have by $K := z^3|q_1q_2|$ ($q_2 \neq 0$) or $K := z^2|q_1|$ ($q_2 = 0$)

$$\left|\sum_{\nu=\mu}^{n} a_\nu \binom{\nu}{\mu} z^n q_1^\mu q_2^{\nu-\mu}\right| \leqq 2^{n+1} K^n H.$$

From (2.1) we conclude

$$z^n w_n(H, q_1\zeta + q_2) \geqq w_n((2^{n+1}K^n H), \zeta),$$

and by definition (1.2) this leads to

$$w_n(q_1\zeta + q_2) \leqq w_n(\zeta). \tag{2.2}$$

As before, we obtain for $\eta = q_1\zeta + q_2$ ($q_1 \neq 0$):

$$w_n(\zeta) = w_n\Bigl(\frac{1}{q_1}\eta - \frac{q_2}{q_1}\Bigr) \leqq w_n(\eta) = w_n(q_1\zeta + q_2),$$

and this (together with (2.2)) proves the lemma.

COROLLARY 1. *Let $b, c \in \mathbb{Z}$; $c \neq 0$; $n, s \in \mathbb{N}$. Furthermore let ζ be an algebraic number, defined as a root of one of the polynomials*
i) $m_\zeta(x) := x^s + bx + c$ *or*
ii) $m_\zeta(x) := x^s + bx^{s-1} + c$.
Then we have

$$w_n(\zeta) = w_n(\zeta^{s-1}).$$

PROOF. If ζ is a root of a polynomial from i), the assertion follows from $\zeta = \frac{-c}{\zeta^{s-1}+b}$ by use of Lemma 1 and Lemma 2; otherwise we know $\zeta^{s-1} = \frac{-c}{\zeta+b}$.

COROLLARY 2. *Let $a, b, c, d \in \mathbb{Z}$; $ad - bc \neq 0$; $\zeta \in \mathbb{C}$. Then we have for every positive integer n*

$$w_n(\zeta) = w_n\left(\frac{a\zeta + b}{c\zeta + d}\right).$$

PROOF. First assume $c = 0$. This reduces the corollary to Lemma 2. Now let $c \neq 0$. With

$$\frac{a\zeta + b}{c\zeta + d} = \frac{a}{c} - \frac{1}{c^2}\frac{ad - bc}{\zeta + \frac{d}{c}}$$

the assertion is proved.

3. A basic theorem to prove more general results

THEOREM 1. *Let ζ be any algebraic number of degree s, assume that $m_\zeta(x) \in \mathbb{Z}[x]$ is the minimal polynomial of ζ. By R we denote the ring of polynomials $\frac{\mathbb{Q}[x]}{m_\zeta}$. Let $P(x) \in \mathbb{Q}[x]$; $n \in \mathbb{N}$; $n \leqq s - 1$. For every k with $0 \leqq k \leqq s - 1$ we write*

$$P^k(x) = (P(x))^k = \sum_{\nu=0}^{s-1} a_{k,\nu} x^\nu,$$

where the rationals $a_{k,\nu}$ $(0 \leqq k \leqq s-1,\ 0 \leqq \nu \leqq s-1)$ define the irreducible representation of the polynomial $P^k(x)$ in the ring R. Furthermore we assume that the following conditions are satisfied:

i) $\det\left((a_{k,\nu})_{\substack{0 \leqq k \leqq s-1 \\ 0 \leqq \nu \leqq s-1}}\right) \neq 0$.

ii) *Let $b_0, \ldots, b_{s-1}$ be arbitrary integers, $b_0^2 + \ldots + b_{s-1}^2 > 0$. Then let $x_0, \ldots, x_{s-1}$ be the solution of the inhomogeneous system of equations*

$$\left(\sum_{k=0}^{s-1} x_k a_{k,\nu}\right)_{0 \leqq \nu \leqq s-1} = (b_\nu)_{0 \leqq \nu \leqq s-1}.$$

For all integers $b_0, \ldots, b_{s-1}$, $b_0^2 + \ldots + b_{s-1}^2 > 0$, *there exists a polynomial* $T(x) \in \mathbb{Q}[x]$ *such that we can assume for*

$$U(x) := T(x)(x_0 + x_1 x + \ldots + x_{s-1} x^{s-1}) :$$

The coefficients of U *are integers,*

$$H(U) = O(\max\{|b_0|, |b_1|, \ldots, |b_{s-1}|\}), \qquad \deg U \leqq n \tag{3.1}$$

iii) *At last*

$$|T(P(\zeta))| = O(1) \tag{3.2}$$

holds for all polynomials T *defined above. The constants in* $O(\ldots)$ *are independent of* $b_0, b_1, \ldots, b_{s-1}$ *resp.* $x_0, x_1, \ldots, x_{s-1}$. *Then we have*

$$w_n(\zeta) \leqq w_n(P(\zeta)).$$

PROOF. Let $H \in \mathbb{N}$. The identity

$$w_n(H, \zeta) = |Q(\zeta)| \tag{3.3}$$

holds for a certain polynomial $Q(x) = b_0 + b_1 x + \ldots + b_n x^n \in \mathbb{Z}[x]$. Now set $b_{n+1} = 0, \ldots, b_{s-1} = 0$; from i) we deduce the existence of the solution $x_0, \ldots, x_{s-1}$ of the system in ii). If we sum up all the equations of this system, we get

$$Q(x) = \sum_{\nu=0}^{s-1} b_\nu x^\nu = \sum_{k=0}^{s-1} x_k \Big(\sum_{\nu=0}^{s-1} a_{k,\nu} x^\nu \Big) = \sum_{k=0}^{s-1} x_k P^k(x)$$

(in R). Hence the polynomial $R(x) = \sum_{\mu=0}^{s-1} x_\mu x^\mu$ satisfies

$$Q(\zeta) = R(P(\zeta)). \tag{3.4}$$

From ii) we know that there is a polynomial $T(x) \in \mathbb{Q}[x]$ such that (3.1) holds for the polynomial $U(x) = T(x)R(x)$. Denote by C_0 the constant from the estimate of $H(U)$ in (3.1). From the definition of $Q(x)$ we know

$$\max\{|b_0|, \ldots, |b_n|\} \leqq H,$$

and therefore we obtain $|U(P(\zeta))| \geqq w_n(C_0 H, P(\zeta))$. With respect to (3.3) and (3.4) it is clear that

$$|T(P(\zeta))| w_n(H, \zeta) \geqq w_n(C_0 H, P(\zeta)).$$

Now let C_1 be the constant from the estimate of (3.2). Then

$$-\frac{\log C_1}{\log H}+\frac{-\log w_n(H,\zeta)}{\log H}\leqq\frac{-\log w_n(C_0H,P(\zeta))}{\log(C_0H)}\frac{\log(C_0H)}{\log H}.$$

By use of Definition (1.2) we finish the proof of the theorem, because C_0 and C_1 are both independent of $b_0,\dots,b_n$ (and therefore of H, too).

REMARK. a) Condition i) in Theorem 1 holds if and only if

$$\deg\zeta=\deg P(\zeta).$$

b) From the proof we know

$$\sum_{\nu=0}^{s-1}b_\nu x^\nu=\sum_{k=0}^{s-1}x_kP^k(x)$$

if condition i) of the theorem holds. Set $s=n$ and $b_{n-1}=0$. If we have $x_{n-1}=0$ for all $b_0,\dots,b_{n-2}\in\mathbb{Z}$, it follows that $w_{n-2}(\zeta)\leqq w_{n-2}(P(\zeta))$, because $T(x)$ becomes a constant polynomial consisting of the least common multiple of all denominators of $x_0,\dots,x_{n-2}$. Obviously, every x_i $(0\leqq i\leqq n-2)$ is only linearly dependent on all $b_0,\dots,b_{n-2}$, and since $b_0,\dots,b_{n-2}$ are integers we conclude that the denominator of x_i remains independent of $b_0,\dots,b_{n-2}$. Hence T depends only on $m_\zeta(x)$ and $P(x)$. We will make use of this principle in the proof of Theorem 3.

4. Two applications of Theorem 1

THEOREM 2. *Let $m\in\mathbb{Z}$; $s,t\in\mathbb{N}$; $(s,t)=1$; $t<s$; $s\geqq3$. With ζ we denote an arbitrary root of the polynomial x^s-m; and let* $\deg\zeta=s$. *Then*

$$w_{s-2}(\zeta)=w_{s-2}(\zeta^t).$$

PROOF. We have $m_\zeta(x):=x^s-m$, $P^k(x):=x^{kt}$ $(0\leqq k\leqq s-1)$. The reduction of P^k modulo m_ζ leads to

$$\begin{aligned}
P^k(x)&=m^0x^{kt-0s}, &&\text{if } 0\leqq kt\leqq s-1\\
P^k(x)&=m^1x^{kt-1s}, &&\text{if } s\leqq kt\leqq 2s-1\\
P^k(x)&=m^2x^{kt-2s}, &&\text{if } 2s\leqq kt\leqq 3s-1\\
&\ \vdots\\
P^k(x)&=m^{(t-1)}x^{kt-(t-1)s}, &&\text{if } (t-1)s\leqq kt\leqq ts-1.
\end{aligned}$$

Thus we have

$$a_{k,\nu} = 0, \qquad \text{if } \nu \neq kt - rs \ \ (0 \leqq r \leqq t-1)$$
$$a_{k,\nu} = m^r, \qquad \text{if } \nu = kt - rs$$

for the coefficients of the system of equations in Theorem 1. Hence we are able to compute the unique solution of this system:

$$\begin{aligned} x_k &= m^0 b_{kt-0s}, & &\text{if } 0 \leqq kt \leqq s-1\\ x_k &= m^{-1} b_{kt-1s}, & &\text{if } s \leqq kt \leqq 2s-1\\ &\vdots \\ x_k &= m^{-(t-1)} b_{kt-(t-1)s}, & &\text{if } (t-1)s \leqq kt \leqq ts-1, \end{aligned}$$

and therefore

$$b_0 + b_1\zeta + \ldots + b_{s-1}\zeta^{s-1} = x_0 + x_1\zeta^t + x_2\zeta^{2t} + \ldots + x_{s-1}\zeta^{(s-1)t} \tag{4.1}$$

holds. Furthermore we know

$$m^{t-1}x_i \in \mathbb{Z} \text{ for all } 0 \leqq i \leqq s-1. \tag{4.2}$$

Now assume $n = s-2$, which means $b_{s-1} = 0$ in Theorem 1. We want to reduce the exponents of ζ on the right-hand side of (4.1). Let $\nu \in \{0, 1, \ldots, s-1\}$ denote the uniquely determined subscript satisfying $x_\nu = b_{s-1} = 0$. With this number we determine $n_1 := (s-\nu-1)t$. From (4.1) we obtain

$$\zeta^{n_1}(b_0 + \ldots + b_{s-1}\zeta^{s-1}) = \sum_{\mu=0}^{s-1} x_\mu \zeta^{\mu t + n_1}. \tag{4.3}$$

Thus we know that every exponent of ζ on the right-hand side of (4.3) is divisible by t. Let $\mu \in \{0, 1, \ldots, s-1\}$. Then there exist uniquely determined nonnegative integers m_μ and r_μ such that

$$\mu t + n_1 = (s - \nu + \mu - 1)t = (m_\mu s + r_\mu)t,$$

and $0 \leqq r_\mu < s$. Particularly, we deduce from $0 \leqq \nu \leqq s-1$ and $0 \leqq \mu \leqq s-1$:

$$\begin{aligned} r_\mu = s-1 \quad &\Longleftrightarrow s - \nu + \mu - 1 = m_\mu s + (s-1)\\ &\Longleftrightarrow \mu = \nu. \end{aligned} \tag{4.4}$$

Notice that $\zeta^{\mu t + n_1} = m^{tm_\mu}\zeta^{tr_\mu}$, and if we substitute into (4.3):

$$\zeta^{n_1}(b_0 + \ldots + b_{s-2}\zeta^{s-2}) = \sum_{\mu=0}^{s-1} x_\mu m^{tm_\mu}(\zeta^t)^{r_\mu} = \sum_{\mu=0}^{s-2} x_\mu m^{tm_\mu} x^{r_\mu}\Big|_{x=\zeta^t}.$$

The last equality holds with respect to (4.4) and $x_\nu = 0$. From (4.2) we deduce that there are integers $c_0, \dots c_{s-2}$ such that

$$m^{t-1}\zeta^{n_1}(b_0 + \dots + b_{s-2}\zeta^{s-2}) = \sum_{\mu=0}^{s-2} c_\mu x^\mu \Big|_{x=\zeta^t}. \tag{4.5}$$

Now we specify the polynomial T and the constant C_1 from Theorem 1:

$$T(x) := m^{t-1}x^{s-\nu-1}, \quad C_1 := \left|m^{t-1}\zeta^{n_1}\right|.$$

It is easy to see that n_1, C_1 and $m_0, \dots , m_{s-1}$ are independent of $b_0, \dots , b_{s-2}$. By (4.5) we obtain

$$H\left(\sum_{\mu=0}^{s-2} c_\mu x^\mu\right) \leqq |m^{t-1}| \max\{|x_0|, \dots , |x_{s-2}|\} \leqq$$
$$\leqq C_0 \max\{|b_0|, \dots , |b_{s-2}|\},$$

because every x_i $(0 \leqq i \leqq s-2)$ is only linearly dependent on all $b_0, \dots , b_{s-2}$. Hence we have checked all the conditions of Theorem 1; and we conclude

$$w_{s-2}(\zeta) \leqq w_{s-2}(\zeta^t). \tag{4.6}$$

$(s,t) = 1$ implies that there is a positive integer w satisfying $w < s$ and $wt \equiv$ $\equiv 1(\operatorname{mod} s)$. Then

$$\left(m^{t/s}\right)^w = rm^{1/s}$$

for some positive integer r. Thus by analogous arguments as used to deduce (4.6) we obtain

$$w_{s-2}(\zeta^t) \leqq w_{s-2}\left(\left(\sqrt[s]{m^t}\right)^w\right) = w_{s-2}(r\sqrt[s]{m}) = w_{s-2}(\zeta)$$

(by use of Lemma 2). Together with (4.6) the theorem is proved.

THEOREM 3. *Let* $b, c \in \mathbb{Z}$; $b \neq 0$; $n, m \in \mathbb{N}$; $n \geqq 3$, $m \mid n$, $m < n$. *Assume* ζ *to be an algebraic number of degree* n, *which is a root of the polynomial* $m_\zeta(x) := x^n + bx + c$. *Furthermore let* $\deg \zeta^m = n$. *Then*

$$w_{n-2}(\zeta) = w_{n-2}(\zeta^m).$$

PROOF. For every k $(0 \leqq k \leqq n-1)$ we have $P^k(x) := x^{km}$ and we calculate two numbers ν and j defined by

$$\nu := \left[\frac{mk}{n}\right] \quad \text{and} \quad 0 \leqq j \leqq \frac{n}{m} - 1, \quad j \equiv k \left(\operatorname{mod}\left(\frac{n}{m}\right)\right).$$

In the first part of the proof we show that with

$$P^k(x) \equiv (-1)^\nu \sum_{\mu=0}^{\nu} \binom{\nu}{\mu} b^{\nu-\mu} c^\mu x^{jm+\nu-\mu} \mod (m_\zeta(x)) \tag{4.7}$$

we have a representation of $P^k(x)$, which is completely reduced in the ring $\frac{\mathbb{Q}[x]}{m_\zeta}$. From the construction of j we obtain:

i) $r := \frac{mk}{n} - \frac{m}{n}j$ for some non-negative integer r,

ii) $0 \leqq \frac{m}{n}j \leqq 1 - \frac{m}{n}$.

$m \mid n$ yields $0 \leqq \frac{m}{n}\ j < 1$, hence $\frac{mk}{n} - 1 < r \leqq \frac{mk}{n}$. We know $r \in \mathbb{N}_0$, and thus $r = \nu$; at last we gather $km = n\nu + jm$.

Now we compute

$$P^k(x) = x^{jm}(x^n)^\nu \equiv x^{jm}(-bx-c)^\nu =$$
$$= (-1)^\nu \sum_{\mu=0}^{\nu} \binom{\nu}{\mu} b^{\nu-\mu} c^\mu x^{jm+\nu-\mu} \mod (m_\zeta(x)).$$

To prove that this representation is completely reduced we have to check

$$0 \leqq jm + \nu - \mu \leqq n-1 \quad \text{for every} \quad k \in \{0, 1, \ldots, n-1\}. \tag{4.8}$$

The lower estimate is obvious, the upper one follows from

$$jm + \nu - \mu \leqq jm + \nu \leqq n - m + \nu < n,$$

because

$$\nu \leqq \frac{m}{n}k \leqq \frac{m}{n}(n-1) < m.$$

Now we use the principle described at the end of the proof of Theorem 1. To do this we prove: There are non-negative integers ν, μ and j as defined above with

$$jm + \nu - \mu = n - 1$$

if and only if

$$k = n - 1. \tag{4.9}$$

Without loss of generality we may put $\mu = 0$; notice (4.8) and $0 \leqq \mu \leqq \nu$. If we set $k = n-1$ we obtain from $\nu = \left[m - \frac{m}{n}\right]$ and $m \mid n$: $\nu = m - 1$; and from the relation mentioned above, $\nu \frac{n}{m} = k - j$, we deduce $j = \frac{n}{m} - 1$. This implies

$$jm + \nu = \left(\frac{n}{m} - 1\right) m + (m-1) = n - 1.$$

On the other hand we assume $k \leqq n-2$. With regard to $\nu \leqq \left[m - \frac{2m}{n}\right] \leqq m - 1$ we have to distinguish two cases:

First case: $\nu = m-1$.
From $\nu\frac{n}{m} = k-j$ we obtain $j \leqq \frac{n}{m}-2$, and this yields

$$jm+\nu \leqq \left(\frac{n}{m}-2\right)m+(m-1) = n-m-1 < n-1.$$

Second case: $\nu \leqq m-2$.

$$jm+\nu \leqq \left(\frac{n}{m}-1\right)m+(m-2) = n-2.$$

Notice $\deg\zeta = \deg\zeta^m$; condition i) in Theorem 1 holds. Hence we deduce from (4.7) and (4.9)

$$w_{n-2}(\zeta) \leqq w_{n-2}(\zeta^m). \tag{4.10}$$

Now we define $\eta := \zeta^m$ and $q := \frac{n}{m}$. At the beginning of the second part of the proof we show

$$R(\eta) := \left[\sum_{\nu=0}^{m}\binom{m}{\nu}c^\nu\eta^{n-\nu q}\right] - (-b)^m\eta = 0. \tag{4.11}$$

First we raise $\zeta^n + c = -b\zeta$ to the power m, and by use of the binomial theorem we get

$$\sum_{\nu=0}^{m}\binom{m}{\nu}c^\nu\zeta^{nm-n\nu} = (-b)^m\zeta^m.$$

If we substitute $\zeta^{nm-n\nu} = \zeta^{nm-\nu qm} = \eta^{n-\nu q}$ into this expression, we have proved (4.11). It is obvious that

$$\deg_\eta R = n. \tag{4.12}$$

$Q(x)$ denotes the polynomial x^q. Let $0 \leqq \nu \leqq n-1$. Then we compute the completely reduced representation of $Q^\nu(x)$ modulo $R(x)$; and we write down all exponents of x appearing in this reduced representation. This gives the following table:

$$\begin{array}{ll}
0 \leqq \nu \leqq m-1: & 0q \\
 & 1q \\
 & 2q \\
 & \vdots \\
 & (m-1)q \\
m \leqq \nu \leqq 2m-1: & (m-1)q, (m-2)q, \dots, 0;\ (0q+1) \\
 & (m-1)q, (m-2)q, \dots, 0;\ (1q+1), (0q+1) \\
 & \vdots \\
 & (m-1)q, (m-2)q, \dots, 0;\ (m-1)q+1, \dots, (0q+1) \\
2m \leqq \nu \leqq 3m-1: & (m-1)q, \dots, 0;\ (m-1)q+1, \dots, (0q+1);\ (0q+2) \\
 & \vdots \\
\vdots & \\
(k-1)m \leqq \nu \leqq & \\
\quad \leqq km-1: & (m-1)q, \dots, 0;\ (m-1)q+1, \dots, 1; \\
\quad = n-1 & \qquad (m-1)q+2, \dots, 2;\ 0q+(q-1) \\
 & \vdots \\
 & (m-1)q, \dots, 0;\ (m-1)q+1, \dots, 1; \\
 & (m-1)q+2, \dots;\ (m-1)q+(q-1), \dots, 0q+(q-1)
\end{array}$$

Every integer of the table can be rewritten as

$$rq + s \text{ with } 0 \leqq r \leqq m-1 \text{ and } 0 \leqq s \leqq q-1.$$

First case: $s < q-1$. Then

$$rq + s \leqq (m-1)q + (q-2) = mq - 2 < mq - 1.$$

Second case: $s = q-1$ and $r < m-1$. Then

$$rq + s \leqq (m-2)q + (q-1) = (mq-1) - q < mq - 1.$$

Third case: $s = q-1$ and $r = m-1$. Then

$$rq + s = (m-1)q + (q-1) = n - 1.$$

We notice that the third case arises if and only if $\nu = n-1$. Hence by the same argument as used to deduce (4.10) we obtain from (4.11) and (4.12)

$$w_{n-2}(\eta) \leqq w_{n-2}(\eta^q);$$

notice $\deg \zeta^m = \deg \zeta = \deg(-b\zeta - c) = \deg \zeta^n$. At last we remember (4.10), and by use of Lemma 2 we get

$$w_{n-2}(\zeta) \leqq w_{n-2}(\zeta^m) \leqq w_{n-2}(\zeta^{mq}) = w_{n-2}(\zeta^n) = w_{n-2}(-b\zeta - c) = w_{n-2}(\zeta),$$

and the proof is complete.

REFERENCES

[1] Mahler, K., Zur Approximation der Exponentialfunktion und des Logarithmus I, II, *J. Reine Angew. Math.* **166** (1931), 118–136, 137–150. *Zbl* **3**, 151; **3**, 388

[2] Schneider, Th., *Einführung in die transzendenten Zahlen*, Springer-Verlag, Berlin–Göttingen–Heidelberg, 1957. *MR* **19**-252

[3] Wirsing, E. A., On approximations of algebraic numbers by algebraic numbers of bounded degree. 1969 *Number Theory Institute* (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, NY, 1969), American Mathematical Society, Providence, R. I., 1971, 213–247. *MR* **47** #8470

[4] Wirsing, E., Approximation mit algebraischen Zahlen beschränkten Grades, *J. Reine Angew. Math.* **206** (1960), 67–77. *MR* **26** #79

(Received August 10, 1988)

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT HANNOVER
WELFENGARTEN 1
D–3000 HANNOVER 1
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 159–167

PARTIAL CORRECTNESS WITHOUT ACTUAL INFINITY[1]

M. A. SUCHENEK

Abstract

In this paper the problem of Andréka, Németi, Sain [1], and Csirmaz [4]: "Does there exist a set $S \subseteq F_{td}$, such that for every set $\Sigma \subseteq F_d$, and every statement $\varrho \in HF_d$, conditions $\Sigma \overset{\omega}{\models} \varrho$ and $S \cup \Sigma \models \varrho$ are equivalent?" where:

- F_d is the set of all first-order formulae of type d (describing data structure);
- F_{td} is the set of all first-order formulae of many sorted type td (describing temporary states of memory, and time and data structure);
- HF_d is the set of all Hoare–Floyd partial correctness statements of type d;
- $\overset{\omega}{\models}$ is the standard relation of semantical implication;
- $\models$ is the relation of semantical implication of the nonstandard dynamic logic of [1];

is solved negatively. For this reason a new concept of finitely approximative partial correctness is proposed. The relation $\overset{\infty}{\models}$ of semantical implication of the new concept is shown to be axiomatizable within nonstandard dynamic logic of [1]. Moreover, relations $\overset{\omega}{\models}$ and $\overset{\infty}{\models}$ are proven (Theorem 2.5) to coincide for a broad class of theories and partial correctness statements.

1. Non-standard dynamic logic vs. partial correctness

Following Andréka, Németi, and Sain [1], we use the notations and concepts of nonstandard dynamic logic. Notions from ordinary logic, such as structure, formula, sentence, satisfaction relation $\models$, etc., are standard and may be found in any textbook of mathematical logic, e.g. in Barwise [2], and also in [1]. In particular, if $\mathbf{M}$ is a first-order structure then we write $\mathbf{M} \models \varphi$ iff $\mathbf{M} \models \varphi[s]$ holds for all possible valuations s in $\mathbf{M}$; and if Σ is a set of formulae then we write $\Sigma \models \varphi$ iff for every first-order structure $\mathbf{M}$ the following implication holds: if $(\forall \sigma \in \Sigma)(\mathbf{M} \models \sigma)$ then $\mathbf{M} \models \varphi$.

We apply the following conventions:

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 68Q60; Secondary 03B45, 03B70, 03C13, 68Q55.

Key words and phrases. Dynamic logic, finite models, models with explicit time, partial correctness in EC_Δ classes, program semantics.

[1] This paper constitutes a revised version of [7].

- 0 is the empty set;
- ${}^{A}B$ is the set of all functions from A into B;
- $Dm(f)$ is the domain of f;
- d is a similarity type;
- F_d is the set of all first-order formulae of type d;
- P_d is the set of all finite effective program schemes over type d;
- X_d is the set of all variables of elements of P_d;
- HF_d is the set $F_d \times P_d \times F_d$ of all Floyd–Hoare partial correctness statements of type d;
- t is a similarity type, such that $Dm(d) \cap Dm(t) = 0$ and $\{0,1,+,\times,\leqq\} \subseteqq Dm(t)$;
- F_t is the set of all first-order formulae of type t;
- Int is a relation symbol of many-sorted type $\langle 0, t, d \rangle$;
- F_{td} is the set of all first-order many-sorted formulae, generated by $F_d \cup F_t$ and Int.

(Formulae of $F_d \cup F_t$ are treated as many sorted formulae here, e.g. the symbol $\leqq$ is of type $\langle t, t \rangle$, and the symbol $+$ is treated as a ternary relation symbol of type $\langle t, t, t \rangle$).

We call the quadruple $\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle$ a *time-model of type td* iff:
- $\mathcal{T}$ is a (time) structure of type t;
- $\mathcal{D}$ is a (data) structure of type d;
- $int \in {}^{X_d \times T}D$, where T is the universe of structure $\mathcal{T}$, and D is the universe of structure $\mathcal{D}$; and
- $0, 1, +, \times, \leq$ are interpreted in $\mathcal{T}$ as usual.

In particular, the standard structure N of all natural numbers must be embedded in $\mathcal{T}$.

We define *valuations* in $\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle$ as triples $\langle u, w, v \rangle$, where u is an infinite sequence of elements of T, w is an infinite sequence of elements of D, and v is an infinite sequence of elements of X_d; the first sequence corresponds to objects of type t, the second one to objects of type d, and the third one to objects of type 0. We assume int as the semantics of symbol Int. Definition of satisfaction is standard, e.g.

$$\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle \models Int(x_3, x_0, x_7)[\langle u, w, v \rangle]$$

holds iff $int(v_3, u_0) = w_7$, and so on.

We call a time-model $\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle$ of type td a *realization* of program $p \in P_d$ iff there exists an execution of p over data structure $\mathcal{D}$ within time structure $\mathcal{T}$, such that for every variable x of p, its value after ι-th step is $int(x, \iota)$, where ι runs through T. The semantics of Floyd–Hoare statements is defined as follows. Let $\langle \varphi, p, \psi \rangle \in HF_d$ and $\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle$ be a time-model of type td. We write $\langle \mathcal{T}, \mathcal{D}, X_d, int \rangle \models \langle \varphi, p, \psi \rangle$ iff:
- (i) the execution corresponding to int diverges (i.e. it does not halt),

or

- (ii) for some $\iota \in T$, this execution halts after ι-th step with the implication $\mathcal{D} \models \varphi[ext(0)] \Rightarrow \mathcal{D} \models \psi[ext(\iota)]$ being satisfied,

where $ext(\kappa)$ denotes the valuation in $\mathcal{D}$ defined by $ext(\kappa)(x) = int(x, \kappa)$ for all $x \in X_d$, and $\kappa \in T$. Note that the time structure T is allowed to contain transfinite numbers, thus infinite executions do not necessarily diverge .

We consider two relations $\models$ and $\overset{\omega}{\models}$ of entailment. Let $\Sigma \subseteq F_{td}$ and $\langle \varphi, p, \psi \rangle \in HF_d$. We write $\Sigma \models \langle \varphi, p, \psi \rangle$, iff for every realization $\mathbf{M}$ of p, $\mathbf{M} \models \langle \varphi, p, \psi \rangle$ holds whenever $\mathbf{M} \models \Sigma$. We write $\Sigma \overset{\omega}{\models} \langle \varphi, p, \psi \rangle$ iff the above implication holds at least for standard (i.e. with the time structure isomorphic to the standard structure n of all natural numbers) realizations of p. It is easily seen that $\Sigma \overset{\omega}{\models} \langle \varphi, p, \psi \rangle$ expresses the usual partial correctness of p in every data structure that satisfies Σ, with respect to the precondition φ and the postcondition ψ.

The relation $\overset{\omega}{\models}$ is less investigated than the relation $\models$. Therefore it would come handy to reduce $\Sigma \overset{\omega}{\models} \varrho$ to $S \cup \Sigma \models \varrho$ for some $S \subseteq F_{td}$ by proper axiomatization S of $\overset{\omega}{\models}$ within $\models$. Several attempts have been made in this direction (one of the most natural of them was taking Peano Arithmetic PA_t of type t as S), but they turned out to be unsatisfactory in some cases. It is already known from Bergstra and Tucker [3] that for every $\Sigma \subseteq F_d$ which has only infinite models, there exist a similarity type t, and $S \subseteq F_{td}$ conservative over Σ, such that for every $\varrho \in HF_d$, $\Sigma \overset{\omega}{\models} \varrho$ is equivalent to $S \cup \Sigma \models \varrho$. Unfortunately, [3] does not guarantee that there is an $S \subseteq F_{td}$ which works for all such $\Sigma \subseteq F_d$.

The problem of *uniform* axiomatization of $\overset{\omega}{\models}$ within $\models$ has been posed in Andréka, Németi, Sain [1] (problem 1, page 276), and next in Csirmaz [4]:

PROBLEM 1.1. Does there exist $S \subseteq F_{td}$, such that for every $\Sigma \subseteq F_d$ and $\varrho \in HF_d$, $\Sigma \overset{\omega}{\models} \varrho$ is equivalent to $S \cup \Sigma \models \varrho$?

We prove that the answer to this problem is negative (even after we relax the restriction on S and allow it to be an arbitrary subset of the set DF_d of formulae of nonstandard dynamic logic of [1], which is a proper superset of F_{td}). For that purpose we need some prerequisites.

DEFINITION 1.2. A binary relation $\mathbf{r}$ between sets Σ of formulae and formulae φ is said to be *compact* iff the following equivalence holds for every Σ and φ:

$$\Sigma \mathbf{r} \varphi \text{ iff } \Sigma_0 \mathbf{r} \varphi \text{ for some finite } \Sigma_0 \subseteq \Sigma. \qquad \square$$

FACT 1.3. *The relation $\models$ of entailment of nonstandard dynamic logic is compact.* (Proof in [1].) $\square$

FACT 1.4. *Every Hoare–Floyd statement* $\langle\varphi, p, \psi\rangle$ *is expressible by the formula* $\varphi \supset \Box(p, \psi)$ *of* DF_d *(see [1] for details).* $\square$

FACT 1.5. $\overset{\omega}{\models}$ *is not compact.*

PROOF. Take $\Sigma = \{\neg(\infty = 0 + \underbrace{1 + \ldots + 1}_{i \text{ times}}) \mid i \in \omega\}$, where ∞ is a constant symbol, and $\varrho = \langle true, \{\texttt{1 : x}\leftarrow\texttt{x+1; 2 : if}\neg(\infty = \texttt{x})\texttt{goto1;3 :halt}\}, false\rangle$. It follows that $\Sigma \overset{\omega}{\models} \varrho$, but for no finite $\Sigma_0 \subseteqq \Sigma$, $\Sigma_0 \overset{\omega}{\models} \varrho$. $\square$

Now suppose that S is the required set. Let $\varrho = \langle\varphi, p, \psi\rangle$ and let $\Sigma \overset{\omega}{\models} \varrho$. $\Sigma \overset{\omega}{\models} \varrho$ is, by this supposition, equivalent to $S \cup \Sigma \models \varrho$, which, by Fact 1.4, is equivalent to $S \cup \Sigma \models \varphi \supset \Box(p, \psi)$. By Fact 1.3, it follows that for some finite $S_0 \subseteqq S$ and finite $\Sigma_0 \subseteqq \Sigma$, $S_0 \cup \Sigma_0 \models \varphi \supset \Box(p, \psi)$, which by Fact 1.4, is equivalent to $S_0 \cup \Sigma_0 \models \varrho$. Hence we get $S \cup \Sigma_0 \models \varrho$, which is, by the above supposition, equivalent to $\Sigma_0 \overset{\omega}{\models} \varrho$. Thus we have shown that $\overset{\omega}{\models}$ is compact, which contradicts Fact 1.5.

2. Finitely approximative partial correctness

As we have demonstrated, $\overset{\omega}{\models}$ cannot be axiomatized uniformly within $\models$ in the usual sense. The reason for this is an infinitary character of relation $\overset{\omega}{\models}$, caused by its lack of compactness[2]. Consequently, $\overset{\omega}{\models}$ does not coincide with $\models$ on some infinite theories with infinite models. But do we really need them in Computer Science? Aristotle in *Physics* (*Φυσικὴ Ἀκρόασις*) refused to accept the actual infinity, postulating only the existence of potentially infinite objects. Mycielski has shown in [5] how ordinary analysis may be successfully handled within first-order logic without the assumption of actual infinity. Inspired by his ideas we introduce a relation $\overset{\infty}{\models}$ of finitely approximative partial correctness, which seems to be a reasonable approximation of $\overset{\omega}{\models}$.

Because of the finitistic structure of contemporary computers, it indeed does not seem practical to assume that they may contain infinite or infinitely axiomatizable data structures. On the other hand, imposing fixed upper bounds (say, $10^{10^{10}}$) on the number of their semantical or syntactical components would make a too restrictive, even if an incidentally satisfied assumption. Therefore we postulate, after Aristotle, that the potential infinity is the one and the only one which, at least for a time being, is needed in investigations of behavior of computers and their programs.

---

[2] This fact implies also a certain kind of non-effectiveness of $\overset{\omega}{\models}$; see [6] for details.

In this paper we consider the following circumstances the potential infinity may manifest its presence. A time and data structure $\mathbf{M}_i$ is finite but is allowed to grow unboundedly, and the same concerns its specification Σ_i. Their intention is to approximate some possibly infinite structure $\mathbf{M}$ with possibly infinite specification Σ. The Aristotelian potential infinity principle says that if $\mathbf{M}$ or Σ are infinite then their existence should not be postulated.

In what follows we apply the following conventions. Σ denotes a subset of F_{td} interpreted as a set of constraints on a time-model with *arbitrary* (in particular finite) time structure. If X is a set, we write $X \subseteq \mathbf{M}$ iff X is a subset of the union of the universes of time and data structures of $\mathbf{M}$. For time-models of type td we write $\langle \mathcal{T}, \mathcal{D}, X_d, int\rangle \subseteq \langle \mathcal{T}'\mathcal{D}', X_d, int'\rangle$ iff $\mathcal{T} \subseteq \mathcal{T}'$, $\mathcal{D} \subseteq \mathcal{D}'$ and $int = int' \restriction Dm(int)$ (recall that $\mathcal{T}$ and $\mathcal{D}$ are many-sorted models, thus their functions are treated as special cases of relations). In the sense of the defined inclusion, we apply the union operator $\cup$ to increasing sequences of time-models.

From now we allow all functions in time-models to be partial ones. E.g. given a and ι, the undefined value of $int(a, \iota)$ means that the corresponding computation has been aborted, which is never understood as a proper termination. Execution in time-models with finite time structure is said to *diverge* iff it has been aborted or it reached the last time point (i.e. the one, for which the successor function is undefined) before execution of a halt statement. In light of such a definition, *partial correctness in* $\mathbf{M}$ is understood as impossibility of yielding an incorrect output in the case of proper (i.e. by execution of a halt statement) termination of an execution.

DEFINITION 2.1. We write $\Sigma \overset{\infty}{\models} \langle \varphi, p, \psi\rangle$ iff for some finite $\Sigma_0 \subseteq \Sigma$ and every finite (i.e. with finite time and data structures) realization $\mathbf{M}$ of p, $\mathbf{M} \models \Sigma_0$ implies $\mathbf{M} \overset{\omega}{\models} \langle \varphi, p, \psi\rangle$. □

It is obvious that the relation $\overset{\infty}{\models}$ of finitely approximative partial correctness does not require the actual infinity. Of course, $\overset{\infty}{\models}$ cannot always coincide with $\overset{\omega}{\models}$, for $\overset{\infty}{\models}$ is compact while $\overset{\omega}{\models}$ is not. However, $\overset{\infty}{\models}$ seems to possess just the features of practical partial correctness, because the real execution of the real program is performed under finite number of explicit assumptions, on finite computer system, and in finite time.

The relation $\overset{\infty}{\models}$, in contrast to $\overset{\omega}{\models}$, is axiomatizable within nonstandard dynamic logic.

FACT 2.2. *There is $S \subseteq DF_d$ such that for every $\Sigma \subseteq F_d$ and every $\varrho \in \in HF_d$ conditions $\Sigma \overset{\infty}{\models} \varrho$ and $S \cup \Sigma \models \varrho$ are equivalent.*

PROOF. We define S as follows. Put $\langle \varphi, p, \psi\rangle \in S$ iff $0 \overset{\infty}{\models} \langle \varphi, p, \psi\rangle$. Moreover, for every sentence $\pi \in F_{td}$, $\pi \supset \langle \varphi, p, \psi\rangle \in S$ iff $\langle \pi \wedge \varphi, p, \psi\rangle \in S$. It is easy to check that S is the required axiomatization. $\square$

Now we show that $\overset{\infty}{\models}$ and $\overset{\omega}{\models}$ coincide in practically interesting cases.

Definition 2.3. Let $\Gamma \subseteqq F_{td}$. Σ is said to be Γ-*meaningful* iff for every $\varphi \in \Gamma$ the following implication holds:

if φ is true in all finite time-models of some finite subset of Σ then $\Sigma \models \varphi$.

(The converse implication follows from the compactness of first-order logic.) $\square$

It should be noted that for every Γ-meaningful theory Σ the intersection of Γ with the set of semantical consequences of Σ does not depend on whether actually infinite objects are allowed or not.

Definition 2.4. Let $\Gamma \subseteqq F_{td}$. Program $p \in P_d$ is said to be of the *property* Γ iff all tests and formulae that define substitutions of p are in Γ.

The following theorem gives a characterization of Σ's for which $\overset{\omega}{\models}$ and $\overset{\infty}{\models}$ prove exactly the same partial correctness statements.

Theorem 2.5. *Let Γ be closed under logical connectives and substitutions of variables. Σ is Γ-meaningful iff for every $\varphi, \psi \in \Gamma$ and every program p of property Γ, the conditions $\Sigma \overset{\omega}{\models} \langle \varphi, p, \psi \rangle$ and $\Sigma \overset{\infty}{\models} \langle \varphi, p, \psi \rangle$ are equivalent.*

Proof. First, let us note that $\Sigma \overset{\omega}{\models} \langle \varphi, p, \psi \rangle$ is equivalent to:

for every complete and finite path in p, corresponding path predicate π, and formula f of the resulting substitution, $\Sigma \models \neg(\varphi \wedge \pi \wedge \neg\psi[\overrightarrow{y}/\overrightarrow{x}] \wedge f(\overrightarrow{x}, \overrightarrow{y}))$,

where:

- $\overrightarrow{x}$ is the sequence of all variables of p;
- $\overrightarrow{y}$ is the sequence (of length of $\overrightarrow{x}$) of variables beyond φ, ψ and p;
- $\psi[\overrightarrow{y}/\overrightarrow{x}]$ is obtained from ψ by substitution $y_i \rightarrow x_i$; and
- $f(\overrightarrow{x}, \overrightarrow{y})$ says that after the execution of the path, variables $\overrightarrow{x}$ assumed values $\overrightarrow{y}$.

(To see why, let α abbreviate $\varphi \wedge \pi \wedge \neg(\psi[\overrightarrow{y}/\overrightarrow{x}] \wedge f(\overrightarrow{x}, \overrightarrow{y}))$. Suppose $\Sigma \overset{\omega}{\not\models} \langle \varphi, p, \psi \rangle$. Take a model $\mathbf{M}$ of Σ and a finite execution in $\mathbf{M}$ which does not satisfy the partial correctness specification $\langle \varphi, \psi \rangle$. Take π and f as the corresponding path predicate and resulting substitution. We obtain $\mathbf{M} \models \exists \overrightarrow{x} \exists \overrightarrow{y} \alpha$, hence $\Sigma \not\models \neg\alpha$. On the other hand, if $\Sigma \not\models \neg\alpha$ then for some model $\mathbf{M}$ of Σ, $\mathbf{M} \models \exists \overrightarrow{x} \exists \overrightarrow{y} \alpha$, or in other words, there is a sequence $\overrightarrow{s}$ of elements of $\mathbf{M}$, such that $\mathbf{M} \models \alpha[\overrightarrow{s}]$. Any such $\overrightarrow{s}$ defines the initial values of $\overrightarrow{x}$ that produce incorrect execution of p in $\mathbf{M}$. Thus $\Sigma \overset{\omega}{\not\models} \langle \varphi, p, \psi \rangle$.)

Now, let us turn to the proof of the theorem.

($\Rightarrow$) Let Σ be Γ-meaningful and let φ, ψ, p be as required in the theorem. As we have noted, $\Sigma \overset{\omega}{\models} \langle\varphi, p, \psi\rangle$ means that for every α as above, $\Sigma \models \neg\alpha$. Obviously $\alpha \in \Gamma$. Thus, by Γ-meaningfulness of Σ, this condition is equivalent to the existence of finite $\Sigma_0 \subseteqq \Sigma$, such that $\neg\alpha$ is true in all finite models of Σ_0, that is to say, once more applying the above observation, $\Sigma \overset{\infty}{\models} \langle\varphi, p, \psi\rangle$.

($\Leftarrow$) Let $\psi \in \Gamma$ and $\varrho\langle\neg\psi \vee \psi, \{1 : \mathit{halt}\}, \psi\rangle$. Suppose that $\Sigma \overset{\omega}{\models} \varrho$ and $\Sigma \overset{\infty}{\models} \varrho$ are equivalent. We have $\Sigma \models \psi$ iff $\Sigma \overset{\omega}{\models} \varrho$, which by the supposition is the same as $\Sigma \overset{\infty}{\models} \varrho$. The last condition means that there exists finite $\Sigma_0 \subseteqq \Sigma$, such that every finite model of Σ_0 satisfies ψ. □

3. Examples and open problems

In this section we give a few examples of sufficient conditions for Γ-meaningfulness. Such conditions, by Theorem 2.5, guarantee the interchangeability of $\overset{\omega}{\models}$ and $\overset{\infty}{\models}$. We start from a strictly finite case.

EXAMPLE 3.1. Take $\Gamma_0 = F_{td}$. If Σ is a finite theory only with finite models, then Σ is Γ_0-meaningful.

In case of potentially infinite structures the notion of absoluteness is useful.

DEFINITION 3.2. A formula φ is said to be Σ-absolute iff for every pair $\mathbf{M}, \mathbf{N}$ of models of Σ such that $\mathbf{M} \subseteqq \mathbf{N}$, and every valuation v in $\mathbf{M}$, conditions $\mathbf{M} \models \varphi[v]$ and $\mathbf{N} \models \varphi[v]$ are equivalent. □

In the process of approximating an infinite structure by an increasing sequence of finite ones, absolute formulae are exactly the ones which maintain their logical values independently of temporary state of approximation. E.g. all formulae with bound quantifiers (i.e. of the form $\forall x \in y$ or $\exists x \in y$) of Zermelo–Fraenkel set theory ZF, are ZF-absolute. Also every quantifier-free formula is absolute. Constructibility aspects may impose additional preferences on theories whose models are subjected to approximation. The following class seems particularly promising.

DEFINITION 3.3. Σ is said to be *asymptotically modellable* iff for every model $\mathbf{M}$ of Σ, and every finite $\Sigma_0 \subseteqq \Sigma$, there is finite model $\mathbf{N}$ of Σ_0, such that $\mathbf{N} \subseteqq \mathbf{M}$. □

The following example shows that absolute formulae form a Γ which guarantees the Γ-meaningfulness of asymptotically modellable theories.

EXAMPLE 3.4. Let Σ be asymptotically modellable. Take $\Gamma_\Sigma = \{\varphi \mid \varphi$ is Σ_0-absolute for some finite $\Sigma_0 \subseteqq \Sigma\}$. Then Σ is Γ_Σ-meaningful and, moreover, every program p is of property Γ_Σ.

PROOF. Let Σ be asymptotically modellable and φ be Σ_0-absolute, for some finite $\Sigma_0 \subsetneq \Sigma$. Let φ be true in all finite models of some finite $\Sigma_1 \subsetneq \Sigma$. Of course φ is $\Sigma_0 \cup \Sigma_1$-absolute and true in all finite models of $\Sigma_0 \cup \Sigma_1$. Let $\mathbf{M} \models \Sigma$. Take finite $\mathbf{N} \subsetneq \mathbf{M}$ such that $\mathbf{N} \models \Sigma_0 \cup \Sigma_1$. In particular, $\mathbf{N} \models \varphi$. But $\mathbf{M} \models \Sigma_0 \cup \Sigma_1$, therefore, by $\Sigma_0 \cup \Sigma_1$-absoluteness of φ, $\mathbf{M} \models \varphi$.

For every Σ, every program is of property Γ_Σ because only absolute (quantifier-free) formulae are allowed as its tests and definitions of substitutions. □

It turns out that quite weak assumption about complexity of the prenex form of formulas in Γ suffices for its Γ-meaningfulness.

DEFINITION 3.5. $\varphi \in F_{td}$ is said to be *inductive* iff it is of the form $\forall x_1 \ldots \forall x_n \psi$ or $\exists x_1 \ldots \exists x_n \psi$, where ψ is any quantifier-free formula. We denote the set of all inductive formulae by $\forall\exists$. □

REMARK. Since the usual meaning of a formula $\psi(\overrightarrow{x})$ with free variables $\overrightarrow{x}$ within a theory Σ is $\forall \overrightarrow{x} \varphi(\overrightarrow{x})$, the above definition is consistent with others known in the literature, e.g. in [2], Chap. A2, def. 2.15.

In practice, the inductive formulae are the most common cases of input-output assertions. It is also obvious that every program is of $\forall\exists$-property. So, $\forall\exists$-meaningfulness makes a practically sufficient condition for the equivalence of $\overset{\omega}{\models}$ and $\overset{\infty}{\models}$.

EXAMPLE 3.6. If Σ is asymptotically modellable then it is $\forall\exists$-meaningful.

PROOF. Let Σ be asymptotically modellable. Let $\mathbf{M} \models \Sigma$ and let φ be true in all finite models of some finite $\Sigma_0 \subsetneq \Sigma$. Construct an increasing sequence $\langle \mathbf{M}_i \mid i \in \omega \rangle$, such that for each i, $\mathbf{M}_i \models \Sigma_0$, and $\mathbf{M} = \cup\{\mathbf{M}_i \mid i \in \omega\}$. $\mathbf{M}_i \models \varphi$ for all i, hence by Chang–Łoś–Suszko Theorem (Barwise [2], chap. A2, thm. 3.13), taking into account that φ is an inductive formula, we obtain $\mathbf{M} \models \varphi$. □

PROBLEM 3.7. Find the greatest Γ such that every asymptotically modellable theory is Γ-meaningful.

We end this paper with a concrete example of asymptotically modellable, and hence $\forall\exists$-meaningful theory.

EXAMPLE 3.8. If Σ is universal (i.e. composed of formulae of the form $\forall x_1 \ldots \forall x_n \psi$, where ψ is quantifier-free) then it is asymptotically modellable.

PROOF. By the Łoś–Tarski Theorem (Barwise [2], chap. A2, thm. 3.11) every substructure of a model of Σ is a model of Σ. Thus if Σ is universal then it is asymptotically modellable. □

PROBLEM 3.9. Find the greatest Γ such that every universal theory is Γ-meaningful.

4. Acknowledgement

I would like to express my thankfulness to Hajnal Andréka, László Csirmaz, István Németi, and Ildikó Sain for introducing me to the sophisticated world of non-standard dynamic logic during my short scholarship at Mathematical Institute of Hungarian Academy of Sciences in 1982. Moreover, I would like to thank Winfried Just whose speech about meaningfulness at the Seminar on Philosophy of Mathematics, Warsaw 1983, inspired me to pursuing ideas of Jan Mycielski. Separate thanks are due to my friend Dr. John J. Hutchinson for numerous corrections of my SLE.

REFERENCES

[1] Andréka, H., Németi, I. and Sain, I., A complete logic for reasoning about programs via nonstandard model theory, I & II, *Theoret. Comput. Sci.* **17** (1982), 193–212, 259–278. *MR* **84d**: 68007a, 68007b

[2] *Handbook of Mathematical Logic,* Edited by Jon Barwise, Studies in logic and the foundations of mathematics, Vol. 90, North-Holland, Amsterdam, 1977. *MR* **56** #15351; *MR* **58** #10395

[3] Bergstra, J. A. and Tucker, J. V., Two theorems about the completeness of Hoare's logic, Technical Report, Matematisch Centrum, Amsterdam, 1981.

[4] Csirmaz, L., Theorems and problems about non-standard dynamic logic, 1982 (preprint).

[5] Mycielski, J., Analysis without actual infinity, *J. Symbolic Logic* **46** (1981), 625–633. *MR* **83b**: 03015

[6] Suchenek, M. A., Compactness in logic of programs, *Algebra, combinatorics and logic in computer science,* Vol. I, II (Győr, 1983), Colloq. Math. Soc. J. Bolyai, **42**, North-Holland, Amsterdam–New York, 1986, 803–810.

[7] Suchenek, M. A., *Floyd–Hoare implication is not uniformly axiomatizable within non-standard dynamic logic,* Research Reports 59/83, Institute of Computer Science, Warsaw Technical University, Nowowiejska 15/19, 00-665 Warszawa, Poland, 1983.

(Received September 5, 1988)

DEPARTMENT OF MATHEMATICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67208–1595
U.S.A.

Present address:

DEPARTMENT OF COMPUTER SCIENCE
CALIFORNIA STATE UNIVERSITY
DOMINGUEZ HILLS
CARSON, CA 90747
U.S.A.

e-mail: Suchenek @ filly. calstate.edu

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 169–172

ON A REPRESENTATION OF ALGEBRAIC INTEGERS

B. KOVÁCS[1] and A. PETHŐ[2]

1. Introduction

Let **R** be an integral domain (with unit element 1), $\alpha \in \mathbf{R}$ with $\alpha \neq 0, \mathcal{N} = \{b_0, b_1, \dots, b_m\}$, where b_i $(0 \leqq i \leqq m)$ and $m \geqq 1$ are fixed integers. $\{\alpha, \mathcal{N}\}$ is called a *number system* in **R** if every $\gamma \in \mathbf{R}$ can be uniquely represented as

$$\gamma = a_0 + a_1\alpha + \dots + a_k\alpha^k \tag{1.1}$$

where $a_i \in \mathcal{N}$ for $i = 0, 1, \dots, k$ and $a_k \neq 0$ if $k \neq 0$. If $\mathcal{N} = \mathcal{N}_0 = \{0, 1, \dots, m\}$ then the number system $\{\alpha, \mathcal{N}\}$ is called a *canonical number system.* The exponent k will be denoted by $L(\gamma, \alpha)$.

In the rings of integers of quadratic number fields (over **Q**) all canonical number systems are known ([2], [3]). E. H. Grossman [1] proved an asymptotic estimate and an upper bound for $L(\gamma, \alpha)$, if α belongs to an imaginary or to a real quadratic number field, respectively. We know from [4] that an integral domain **R** of characteristic 0 (with unit element) has a number system if and only if $\mathbf{R} = \mathbf{Z}[\alpha]$, where α is algebraic over **Q**.

The aim of our paper is to give a generalization and refinement of the estimates of [1] in case of $\mathbf{R} = \mathbf{Z}[\beta]$, where β is an algebraic integer. Namely we shall prove the following

THEOREM. *Let β be an algebraic integer of degree $n \geqq 1$, and let $\{\alpha, \mathcal{N}\}$ be a number system in $\mathbf{Z}[\beta]$. Then there exist effectively computable constants $C_1(\alpha, \mathcal{N})$, $C_2(\alpha, \mathcal{N})$, depending only on α and $\mathcal{N}$, such that*

$$\max_{1 \leqq i \leqq n} \frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} + C_1(\alpha, \mathcal{N}) \leqq L(\gamma, \alpha) \leqq \max_{1 \leqq i \leqq n} \frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} + C_2(\alpha, \mathcal{N}) \tag{1.2}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 11A63; Secondary 11R33.

Key words and phrases. Number fields, algebraic integers, number systems.

[1] Research supported in part by Grants 273 and 400 from the Hungarian National Foundation for Scientific Research.

[2] Research supported in part by Hungarian National Foundation for Scientific Research Grant No. 273/86.

[3] **Z** and **Q** denote the ring of rational integers and the field of rational numbers, respectively.

holds for every $0 \neq \gamma \in \mathbf{Z}[\beta]$, *where* $\alpha^{(i)}$ *and* $\gamma^{(i)}$ *denote the i-th conjugates of* α *and* γ, *respectively.*

REMARK. When β is an imaginary quadratic integer, (1.2) was proved by Grossman [1]. He gave also an upper bound for $L(\gamma, \alpha)$, if β is a real quadratic integer. It follows from our theorem that his upper bound is not sharp.

2. Proof of the Theorem

Let $\{\alpha, \mathcal{N}\}$ be a number system in $\mathbf{Z}[\beta]$. We can see from Theorem 3 [4], that $\mathcal{N}$ is a complete residue system mod $|N_{K/Q}(\alpha)|$, and $|\alpha^{(i)}| > 1$ holds for every conjugate of α. Let now $\gamma \in \mathbf{Z}[\beta]$ and

$$\gamma = a_0 + a_1\alpha + \ldots + a_k\alpha^k; \quad a_j \in \mathcal{N} \text{ for } 0 \leqq j \leqq k \text{ and } a_k \neq 0. \tag{2.1}$$

Taking conjugates this implies

$$\gamma^{(i)} = a_0 + a_1\alpha^{(i)} + \ldots + a_k(\alpha^{(i)})^k \text{ for every } 1 \leqq i \leqq n. \tag{2.2}$$

From (2.2) we get

$$|\gamma^{(i)}| = |a_0 + a_1\alpha^{(i)} + \ldots + a_k(\alpha^{(i)})^k| \leqq D\sum_{j=0}^{k}|\alpha^{(i)}|^j = D\frac{|\alpha^{(i)}|^{k+1}-1}{|\alpha^{(i)}|-1}, \tag{2.3}$$

where $D = \max\limits_{b\in\mathcal{N}}|b|$. Thus

$$\frac{(|\alpha^{(i)}|-1)|\gamma^{(i)}|}{D} < |\alpha^{(i)}|^{k+1}, \tag{2.4}$$

and

$$\frac{\log(|\alpha^{(i)}|-1) - \log D}{\log|\alpha^{(i)}|} + \frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} < k+1 \text{ for every } 1 \leqq i \leqq n. \tag{2.5}$$

But the first summand of (2.5) is an effectively computable constant, depending only on α and $\mathcal{N}$, and so it follows from (2.5) that

$$L(\gamma, \alpha) \geqq \max_{1\leqq i\leqq n}\frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} + C_1(\alpha, \mathcal{N}). \tag{2.6}$$

Let now k denotes the minimum of those natural numbers for which

$$|\alpha^{(i)}|^k \geqq |\gamma^{(i)}| \tag{2.7}$$

holds for every conjugate of α. Such a k exists because $|\alpha^{(i)}|>1$ for $1 \leqq i \leqq n$. By Lemma 4 of [4] there exist $a_j \in \mathcal{N}$, $0 \leqq j \leqq k-1$ and $\gamma_1 \in \mathbf{Z}[\beta]$ such that

$$\gamma = a_0 + a_1\alpha + \ldots + a_{k-1}\alpha^{k-1} + \alpha^k\gamma_1 \tag{2.8}$$

and

$$|\gamma_1^{(i)}| < \frac{|\gamma^{(i)}|}{|\alpha^{(i)}|^k} + \frac{D}{|\alpha^{(i)}|-1} \leqq 1 + \frac{D}{|\alpha^{(i)}|-1} = C_{2,i} \tag{2.9}$$

because $\frac{|\gamma^{(i)}|}{|\alpha^{(i)}|^k} \leqq 1$ and $|\alpha^{(i)}|>1$. Here $C_{2,i}$'s are constants depending only on α and $\mathcal{N}$. From the definition of k we can deduce that

$$k \leqq \max_{1 \leqq i \leqq n} \frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} + 1. \tag{2.10}$$

Since $\gamma_1 \in \mathbf{Z}[\beta] = \mathbf{Z}[\alpha]$, we can write

$$\gamma_1^{(i)} = x_0 + x_1\alpha^{(i)} + \ldots + x_{n-1}\left(\alpha^{(i)}\right)^{n-1}, \tag{2.11}$$

where $x_j \in \mathbf{Z}$ for $j \in \{0,1,\ldots,n-1\}$. Using (2.9) and (2.11) we obtain

$$|\gamma_1^{(i)}| = \left|x_0 + x_1\alpha^{(i)} + \ldots + x_{n-1}\left(\alpha^{(i)}\right)^{n-1}\right| \leqq C_{2,i}, \quad 1 \leqq i \leqq n. \tag{2.12}$$

But it is well-known that (2.12) has only finitely many solutions in $(x_0, x_1, \ldots, x_{n-1}) \in \mathbf{Z}^n$ and these solutions are effectively computable. Let Γ be the following subset of $\mathbf{Z}[\beta]$:

$$\Gamma = \{\delta \mid \delta \in \mathbf{Z}[\alpha] \text{ and } \delta \text{ is solution of } |\delta^{(i)}| \leqq C_{2,i}, 1 \leqq i \leqq n\},$$

and take $\Gamma(L) = \max_{\delta\in\Gamma} L(\delta,\alpha)$. Since Γ is a finite set and $\{\alpha, \mathcal{N}\}$ is a number system, hence $\Gamma(L)$ exists and it is an effectively computable constant. From (2.8) and (2.10) we get now that

$$L(\gamma,\alpha) \leqq \max_{1 \leqq i \leqq n} \frac{\log|\gamma^{(i)}|}{\log|\alpha^{(i)}|} + \Gamma(L) + 1. \tag{2.13}$$

(2.6) and (2.13) proves the theorem.

REFERENCES

[1] GROSSMAN, E. H., Number bases in quadratic fields, *Studia Sci. Math. Hungar.* **20** (1985), 55–58. *MR* **88m:** 11086

[2] KÁTAI, I. und KOVÁCS, B., Kanonische Zahlensysteme in der Theorie der quadratischen algebraischen Zahlen, *Acta Sci. Math. (Szeged)* **42** (1980), 99–107.

[3] KÁTAI, I. and KOVÁCS, B., Canonical number systems in imaginary quadratic fields, *Acta Math. Acad. Sci. Hungar.* **37** (1981), 159–164. *MR* **83a:** 12005

[4] KOVÁCS, B. and PETHŐ, A., Number systems in integral domains especially in orders of algebraic number fields, *Acta Sci. Math. (Szeged)* **55** (1991), 287–299.

(Received August 10, 1988)

KOSSUTH LAJOS TUDOMÁNYEGYETEM
MATEMATIKAI INTÉZETE
P.O.BOX 12
H–4010 DEBRECEN
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 173–182

AN ELEMENTARY PROOF OF SOME RESULTS CONCERNING SUMS OF DISTINCT TERMS FROM A GIVEN SEQUENCE OF INTEGERS

U. ZANNIER

Introduction

A number of papers have been devoted to the following additive problem: Given a sequence $S = \{s_1 < s_2 < \dots\}$ of natural numbers investigate its partition function

$$\varrho_S(n) = \text{ number of solutions of } n = s_{i_1} + s_{i_2} + \dots + s_{i_r}$$

where r is unrestricted and $i_1 > i_2 > \dots > i_r$.

We mention, among others, the papers of Birch [1], of Erdős [4], and the very general results of Cassels [2] and Roth–Szekeres [6].

In [6] the authors obtain an asymptotic expansion for $\varrho_S(n)$ under very general and natural conditions on S and Cassels shows $\varrho_S(n) > 0$ for large n under even less restrictive assumptions.

However, the method used by these authors is analytic, inspired by the circle method, and so it may be of some interest to have elementary combinatorial proofs of some results in this direction.

Three elementary papers on this subject are, for instance, the above mentioned ones of Birch and Erdős, and the paper of Perelli–Zannier [5].

In all these papers the conclusion is that $\varrho_S(n) > 0$ for large n, under various different conditions on the sequence S. Anyway, the theorems do not apply when S consists for example of the values of a polynomial at natural numbers, a situation which is covered as an extremely particular case by both Cassels and Roth–Szekeres' results.

The object of this article is to give, in a completely elementary way, a condition on S sufficient to imply $\varrho_S(n) > 0$ for $n > n_0$, which is fairly general and, moreover, works for polynomial sequences.

Define

$$C_S = \{n \in N, \varrho_S(n) > 0\}.$$

We shall prove the following

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 11BXX.
Key words and phrases. Sequences of integers, sums, additive problem.

THEOREM. *Assume $S = \{s_1 < s_2 < \ldots\}$ is an infinite sequence such that*
(i) $s_{u+1} - s_u = o(s_u)$;
(ii) *there exist integers $n_1, \ldots, n_k$ and a number B such that, for infinitely many indices u*

$$0 < |n_1(s_{u+1} - s_u) + \ldots + n_k(s_{u+k} - s_u)| < B;$$

(iii) *for every integer m, $1 \leqq m \leqq B$, C_S meets every arithmetic progression with difference m.*
Then C_S contains every sufficiently large integer.

REMARKS. (1) The inelegant condition (ii) may appear more natural observing that a special case occurs when the k-th difference of s_u is infinitely often bounded and nonzero: of course this fact suggests application to sequences which behave like a polynomial.

(2) A proof completely similar to the one given below gives the stronger Theorem in which condition (ii) is replaced by

(ii)* *There exist integers $n_1, \ldots, n_k$ and an infinity of sets of indices $\{u_1, v_1, \ldots, u_k, v_k\}$ with $u_i \leqq v_i$ and $u_1 < u_2 < \ldots < u_k$, such that*

$$(\alpha)\ s_{v_i} \sim s_{u_i} \qquad (\beta)\ s_{u_k} = O(s_{u_1})$$
$$(\gamma)\ 0 < |n_1(s_{v_1} - s_{u_1}) + \ldots + n_k(s_{v_k} - s_{u_k})| < B$$

where B is absolutely bounded.

For the sake of simplicity we have, however, preferred to prove only the weaker result.

(3) Our Theorem does not follow from Cassels' result in [2], his condition (*) being independent from our (i), and "generally" stronger. However, our main restriction is (ii) and this implies Cassels' (*) for every irrational θ.

As a simple instance we shall prove the following

COROLLARY. *Assume that either $S = \{f(x), x \in N\}$ where f is a nonconstant polynomial taking integral values on N which have no common prime divisor, or $S = \{[g(x)], x \in N\}$ where [ ] denotes integral part, and where g is a nonconstant polynomial with an irrational leading coefficient. Then the above Theorem applies.*

(This result follows, however, from both papers [2] and [5].)

Before commencing the proofs we introduce some notation.

If S is a sequence we let $S(x_1, x_2) = \{s \in S, x_1 < s \leqq x_2\}$ and $S(x) = S(0, x)$.

We define upper and lower asymptotic densities $\overline{d}(S), \underline{d}(S)$ as usual

$$\overline{d}(S) = \limsup |S(x)|/x, \qquad \underline{d}(S) = \liminf |S(x)|/x,$$

where, for a set A, $|A| = \mathrm{card}(A)$. When these numbers coincide we denote their common value by $d(S)$, the asymptotic density of S.

PROOFS. We shall need some simple lemmas.

LEMMA 1. *Assume $S = \{s_1 < s_2 < \ldots\}$ satisfies $s_{i+1} \leqq 2s_i$ for all $i \geqq i_0$. Then every positive integer n may be written as*

$$n = s_{i_1} + s_{i_2} + \ldots + s_{i_r} + c \tag{1}$$

where $i_1 > i_2 > \ldots > i_r \geqq i_0$, $0 \leqq c < s_{i_0}$ and

$$s_{i_\nu} > s_{i_{\nu+1}} + s_{i_{\nu+2}} + \ldots + s_{i_r} + c$$

for $\nu = 1, 2, \ldots, r-1$.

PROOF. Use induction on n, the result being obvious for $n < s_{i_0}$.

When $n \geqq s_{i_0}$ choose i_1 to satisfy $s_{i_1} \leqq n < s_{i_1+1}$, and apply the induction to $n - s_{i_1}$. We obtain (3) where $s_{i_2}, s_{i_3}, \ldots, s_{i_r}$, c satisfy the conclusions.

Now, since $i_1 \geqq i_0$, $0 \leqq s_{i_2} + \ldots + s_{i_r} + c = n - s_{i_1} < s_{i_1+1} - s_{i_1} \leqq s_{i_1}$ whence the result. $\square$

LEMMA 2. *Assumptions being as in Lemma 1, suppose C_S meets every congruence class* mod b, *for a certain positive b. Then there exists $T \geqq 1$ such that for every natural number n, at least one of the numbers*

$$n, n+b, \ldots, n+Tb$$

belongs to C_S.

PROOF. Let R be a finite subset of S such that C_R meets every congruence class mod b, and set

$$S' = S \backslash R = \{s'_1 < s'_2 < \ldots\}.$$

We may assume, increasing eventually R, that $s'_{i+1} < 2s'_i$ for all $i \geqq 1$. Let $T' = s'_1 + \sum_{r \in R} r$, and apply Lemma 1 with S' in place of S and $n + T'b$ in place of n, obtaining

$$n + T'b = s'_{i_1} + \ldots + s'_{i_h} + c, \quad i_1 > i_2 > \ldots > i_h \geqq 1 \quad 0 \leqq c < s'_1.$$

Let $n_0 \in C_R$ be such that $n_0 \equiv c (\bmod b)$. Then $n_0 = c + Ub$, where $|U| \leqq \leqq n_0 + c \leqq T'$, whence

$$n + (T' + U)b = s'_{i_1} + \ldots + s'_{i_h} + n_0 \in C_S$$

and $0 \leqq T' + U \leqq 2T'$, making Lemma 2 true with $T = 2T'$. $\square$

LEMMA 3. *Let A be a finite set of integers and assume that, for some $b>0$ and $T \geqq 1$, and for every integer n, not all the numbers n, $n+b, \ldots, n+Tb$ belong to A. Set $B = \{a \in A, a+b \in A\}$. Then $|B| \leqq |A|(1 - T^{-1})$.*

PROOF. Write $A = \bigcup_{i=1}^{k} P_i$, where the P_i are short arithmetic progressions with difference b: $P_i = \{a_i + mb, m = 0, 1, \ldots, d_i\}$. We may clearly assume them to be pairwise disjoint and maximal. We have $d_i + 1 = |P_i| \leqq T$ and

$$|A| = \sum |P_i| \leqq kT.$$

If $a = a_i + d_i b$ then $a + b \notin A$, whence

$$|B| \leqq \sum d_i = |A| - k \leqq |A|(1 - T^{-1}),$$

as claimed.

LEMMA 4. *Assume S satisfies* (i) *of Theorem. Then the asymptotic density of C_S exists.*

In a paper entitled "On the density of the set of sums" (to appear in *Acta Arithmetica*) I. Ruzsa has considerably weakened the assumption of this lemma, replacing (i) by $s_{i+1} \leqq 2s_i$ for large i.

PROOF. It is convenient to define a function $g\colon R^+ \to R$ as follows. If $s_h < x \leqq s_{h+1}$ set $g(x) = s_{h+1} - s_h$ $(= o(x))$. Let $\varepsilon > 0$ be given. Select ϱ so large to make $(\frac{2}{3})^\varrho < \varepsilon$ and then x_0 such that $g(x) < \frac{\varepsilon}{\varrho}x - 1$ for $x \geqq x_0$.

Set $d = \overline{d}(C_S)$, and find $m > x_0$ such that $|C(m)| > (d - \varepsilon)m$ $(C = C_S)$. If $s_{\nu-1} < m \leqq s_\nu$ then

$$|C(s_\nu)| \geqq |C(m)| > (d - \varepsilon)s_\nu - g(m) \geqq (d - \varepsilon - \frac{\varepsilon}{\varrho})s_\nu. \tag{2}$$

Construct a new sequence $R = \{r_1 < r_2 < \ldots\}$ as follows: set $r_1 = s_\nu$. Having constructed $r_1, r_2, \ldots, r_k$ assume

$$s_h < 2r_k \leqq s_{h+1} \quad \text{and define} \quad r_{k+1} = s_h.$$

From our conventions we derive

$$2r_k > r_{k+1} \geqq 2r_k - g(2r_k) \tag{3}$$

whence in particular

$$2r_k > r_{k+1} > (\frac{3}{2})r_k \quad \text{for } k \geqq 1 \text{ (if } \varepsilon < \frac{1}{2}). \tag{4}$$

Let $d_k = |C(r_k)|/r_k$.
Firstly, we have

$$\begin{gathered} |C(r_{k+1})| \geqq |C(2r_k)| - g(2r_k) \geqq 2|C(r_k)| - 1 - g(2r_k), \text{ whence} \\ d_{k+1} \geqq d_k - 2\frac{\varepsilon}{\varrho}. \end{gathered} \tag{5}$$

On the other hand Lemma 1 applied to the sequence

$$\{r_1, r_2, \ldots, r_k\} \cup \{2r_k, 2r_k + 1, 2r_k + 2, \ldots\}$$

gives an expression

$$r_{k+1} = r_{i_1} + r_{i_2} + \ldots + r_{i_s} + c \tag{6}$$

where $k \geqq i_1 > i_2 > \ldots > i_S$, $0 \leqq c < r_1$, and where

$$r_{i_\mu} > r_{i_{\mu+1}} + r_{i_{\mu+2}} + \ldots + r_{i_s} + c. \tag{7}$$

One trivially has

$$|C(r_j + T)| \geqq |C(r_j)| + |C(T)| \quad \text{for } 0 \leqq T < r_j$$

(remember that $r_j \in S$). Iterating this inequality and using (6) and (7) we get

$$\begin{aligned} |C(r_{k+1})| &\geqq |C(r_{i_1} + \ldots + r_{i_S})| \geqq |C(r_{i_1})| + \ldots + |C(r_{i_S})| \geqq \\ &\geqq (\min_{j \leqq k} d_j)(r_{i_1} + \ldots + r_{i_S}) \geqq (\min_{j \leqq k} d_j) r_{k+1} - r_1, \end{aligned}$$

whence

$$d_{k+1} \geqq \min_{j \leqq k} d_j - r_1/r_k \geqq \min_{j \leqq k} d_j - (\frac{2}{3})^k \quad \text{(by (4))}$$

and, for any $h \geqq 1$

$$d_{k+h} \geqq \min_{j \leqq k} d_j - (\frac{2}{3})^k \left(1 + \frac{2}{3} + (\frac{2}{3})^2 + \ldots\right) = \min_{j \leqq k} d_j - 3(\frac{2}{3})^k.$$

In particular, setting $k = \varrho$ and remembering that $(\frac{2}{3})^\varrho < \varepsilon$

$$d_{\varrho+h} \geqq \min_{j \leqq \varrho} d_j - 3\varepsilon \tag{8}$$

while for $j \leqq \varrho$ our first inequality (5) gives, after iteration,

$$d_j \geqq d_1 - j\frac{\varepsilon}{\varrho} \geqq d_1 - \varepsilon.$$

In conclusion $d_k \geqq d_1 - 4\varepsilon \geqq d - 6\varepsilon$ for all $k \geqq 1$ (we have used (2)).

Let n be a large integer. Apply Lemma 1 to n and the sequence R obtaining

$$n = r_{i_1} + \ldots + r_{i_S} + c \qquad 0 \leqq c < r_1 \tag{9}$$

$$r_{i_\mu} > r_{i_{\mu+1}} + \ldots + r_{i_S} + c.$$

Repeating the above argument we get

$$\begin{aligned} |C(n)| &\geqq |C(r_{i_1} + \ldots r_{i_S})| \geqq |C(r_{i_1})| + \ldots + |C(r_{i_S})| \geqq \\ &\geqq (d - 6\varepsilon)(r_{i_1} + \ldots + r_{i_S}) \geqq (d - 6\varepsilon)(n - r_1), \end{aligned}$$

whence

$$\liminf |C(n)|/n \geqq d - 6\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary the lemma follows.

Proof of the main theorem

First we show that $d = d(C_S) = 1$.

Let $C = C_S$ as above and $\tilde{C} = N \setminus C$. Select a sufficiently large u such that

$$|n_1(s_{u+1} - s_u) + \ldots + n_k(s_{u+k} - s_u)| = b \neq 0, \qquad b < B.$$

By Lemma 4

$$|C(s_u)| \sim ds_u, \quad |C(2s_u)| \sim 2ds_u \quad \text{and so } |C(s_u, 2s_u)| \sim ds_u.$$

Let m be such that

$$s_{u+k} - s_u < m < s_u, \quad m + s_u \in \tilde{C}.$$

There are at least

$$s_u - |C(s_u, 2s_u)| - (s_{u+k} - s_u) \sim (1-d)s_u + o(s_u)$$

such integers.

We contend that, for these integers m in fact

$$m - (s_{u+i} - s_u) \in \tilde{C} \text{ for } i = 0, 1, \ldots, k.$$

For, since $m-(s_{u+i}-s_u)<2s_u-s_{u+i} \leqq s_u \leqq s_{u+i}$, $m+s_u = m-(s_{u+i}-s_u)+s_{u+i}$ would otherwise clearly belong to C. This implies that, setting $t_i = s_{u+i} - s_u$,

$$(10) \qquad \begin{aligned} &|\{m, t_k < m < s_u, m - t_i \in \tilde{C} \text{ for } i = 0, 1, \ldots, k\}| \geqq \\ &\geqq (1-d)s_u + o(s_u) \geqq |\tilde{C}(s_u)| + o(s_u). \end{aligned}$$

Define now, for $\nu \geqq 0$, sets $\tilde{C}_\nu$ inductively as follows. Put $\tilde{C}_0 = \tilde{C}(s_u)$ and, for $\nu \geqq 1$,

$$(11) \qquad \tilde{C}_\nu = \{m, \nu t_k < m < s_u, m - t_i \in \tilde{C}_{\nu-1} \text{ for } i = 0, 1, \ldots, k\}.$$

We can rewrite (10) in the form

$$(12) \qquad |\tilde{C}_1| \geqq |\tilde{C}_0| + o(s_u).$$

Observe that, since $t_0 = 0$, $\tilde{C}_\nu \subseteq \tilde{C}_{\nu-1}$ for any $\nu \geqq 1$. Let now $\nu \geqq 2$ and pick $m \in \tilde{C}_{\nu-1} \setminus \tilde{C}_\nu$. Then, by definition, $m - t_i \notin \tilde{C}_{\nu-1}$ for some $i \neq 0$. Since anyway $m \in \tilde{C}_{\nu-1}$ we have that $m - t_j \in \tilde{C}_{\nu-2}$ for all j, whence

$$m - t_i \in \tilde{C}_{\nu-2} \setminus \tilde{C}_{\nu-1} \text{ for some } i.$$

Since the number $m - t_i$ can be counted correspondingly to at most k values of m as above, we get

$$|\tilde{C}_{\nu-1} \setminus \tilde{C}_\nu| \leqq k|\tilde{C}_{\nu-2} \setminus \tilde{C}_{\nu-1}| \text{ for } \nu \geqq 2.$$

Now, by (10),

$$|\bar{C}_0 \setminus \bar{C}_1| = o(s_u)$$

whence inductively, *for any fixed* ν

$$|\bar{C}_{\nu-1} \setminus \bar{C}_\nu| = o(s_u)$$

and

$$|\bar{C}_0 \setminus \bar{C}_\nu| = o(s_u)$$

implying that

(12) $$|\bar{C}_\nu| = |\bar{C}_0| + o(s_u) \quad \text{(for any fixed } s\text{)}.$$

Note that $m \in \bar{C}_\nu$ implies that

$$\nu t_k < m, s_u \text{ and } m - t_{i_1} - t_{i_2} - \ldots - t_{i_\nu} \in \bar{C}_0$$

for every choice of $i_1, \ldots, i_\nu$ ($\leqq k$).

Consider the sums $t_{i_1} + \ldots + t_{i_\nu}$. It is clear that, if $\nu > 2(|n_1| + \ldots + |n_k|)$, at least two of these sums will differ by $|n_1 t_1 + \ldots + n_k t_k| = b$. (For instance look at $|n_1|t_1 + \ldots + |n_k|t_k$ and $\{|n_1| + n_1\}t_1 + \ldots + \{|n_k| + n_k\}t$.) This leads to the existence of ϱ, of the form $t_{i_1} + \ldots + t_{i_\nu}$, such that

$$m \in \bar{C}_\nu \Rightarrow m - \varrho, \quad m - \varrho + b \in \bar{C}_0.$$

At this point we appeal to Lemmas 2 and 3 which, applied in an obvious way to our situation give

$$|\bar{C}_\nu| \leqq (1 - \frac{1}{T})|\bar{C}_0|$$

for some integer T bounded independently of u.

Comparing this inequality with (12) we get $|\bar{C}(s_u)| = o(s_u)$, i.e. $d = 1$, as asserted.

To complete the proof of the Theorem, let R be a finite subsequence of S such that C_R meets every arithmetical progression with difference bounded by B, and set $S' = S \setminus R$.

Choose then an infinite sequence of indices $\{u_j\}$ such that to satisfy assumption (ii), and moreover such that

(13) $$s_{u_{j+1}} > 4 s_{u_j}.$$

Extract from S' an infinite sequence $S'' = \{s_1'' < s_2'' < \ldots\}$ such that

$$\frac{3}{2} s_i'' < s_{i+1}'' < 2 s_i''$$

and $s_i'' \neq s_{u_j - \tau}$ for all i and j and $\tau \leqq k$.

This is certainly possible, in view of the relatively slow rate of growth of s and in view of (13).

It is now clear that the sequence $S_1 = S \setminus S''$ satisfies the same assumption as S, whence, by what has been proved above,

$$d(C_{S_1}) = 1. \tag{14}$$

Moreover, from the inequality $s''_{i+1} < 2s''_i$ and Lemma 2 (applied to S'' with $b = 1$) we obtain

$$d(C_{S''}) > 0. \tag{15}$$

Now the proof can be completed by means of a quite standard argument: Let n be a large integer and consider the integers

$$n - a, \quad 1 \leqq a < n, \quad a \in C_{S''}.$$

There are $|C_{S''}(n-1)|$ such numbers, and in the same interval there are $|C_{S_1}(n)|$ integers belonging to C_{S_1}. Since, by (14) and (15),

$$|C_{S_1}(n)| + |C_{S''}(n-1)| > n$$

for all $n > n_0$, there exists, for large n, some $a \in C_{S''}$ such that $n - a \in C_{S_1}$. But, in view of $S_1 \cap S'' = \emptyset$, this implies $n \in C_S$.

Proof of Corollary

(a) Take first $S = \{f(k)\}$ where f is as in the statement, and let us verify the three conditions of the Theorem.

Condition (i) is trivial. For (ii), assume $f(x) = ax^h +$ lower terms and take the h-th difference of $f(x)$. We obtain

$$\left|\sum_{\varrho=0}^{h}(-1)^{\varrho}\binom{h}{\varrho}f(x+\varrho)\right| = |ah!|$$

whence

$$\left|\sum_{\varrho=0}^{h}(-1)^{\varrho}\binom{h}{\varrho}\{f(x+\varrho) - f(x)\}\right| = |ah!| \neq 0$$

if $h = \deg f$ so (ii) is satisfied with $k = \deg f$ and any u.

(iii). Let m be any integer $\geqq 2$ and write $m = \varrho_1^{\alpha_1}\varrho_2^{\alpha_2}\dots\varrho_r^{\alpha_r}$, where the ϱ_i are distinct primes. By assumption, for $i = 1, 2, \dots, r$ there exists $x_i \in$ $\in N$ with $\varrho_i \nmid f(x_i)$. Now since f is integral valued on N, it has rational coefficients, whence, for sufficiently large e we have

$$f(x_i + t\varrho^e) \equiv f(x_i) \pmod{\varrho_i^{\alpha_i}} \quad \text{for all } t \in Z \text{ and all } i.$$

Find then $m_0 \equiv x_i (\bmod \varrho_i)$ for $i = 1, \ldots, r$. Then

$$f(m_0 + t \prod \varrho_i^c) \equiv f(m_0)(\bmod m)$$

for all t, and $(f(m_0), m) = 1$. It follows that, adding a suitable number of distinct terms of the sequence S, all congruent to $f(m_0) \bmod m$, we may represent every congruence class $\bmod m$.

(b) Take now $S = \{[g(x)]\}$ where g is a non-constant polynomial with irrational leading coefficient. (i) is again trivial. To verify (ii) write again $h = \deg g$, but take the $h+1$-th difference. As in part (a) we see that we are done provided

$$\sum_{\varrho} (-1)^{\varrho} \binom{h+1}{\varrho} [g(x+\varrho)] \neq 0$$

for infinitely many integers x. Assume the contrary. Then clearly

$$\sum_{\varrho} (-1)^{\varrho} \binom{h+1}{\varrho} ((g(x+\varrho))) = 0 \quad \text{for } x > x_0, \ x \in N$$

where $((\ ))$ denotes the fractional part.

But then $((g(x)))_{x \in N, x > x_0}$ would be a polynomial, necessarily constant.

This fact in turn means that the polynomial $g(x+1) - g(x)$ would assume, for integral $x > x_0$, integral values, a contradiction since its leading coefficient is irrational.

Let now n be any positive integer. Since the polynomial $\frac{g(x)}{m}$ has irrational leading coefficient it follows from known theorems (see for example [3] chapter 10) that the sequence $\left(\left(\frac{g(x)}{m}\right)\right)$ is dense in $(0,1)$, whence, for infinitely many integral x and any fixed b, $0 \leqq b < m$, $\frac{b}{m} < \left(\left(\frac{g(x)}{m}\right)\right) < \frac{b+1}{m}$.

It follows that

$$tm + b \leqq g(x) < tm + b + 1,$$

for some $t \in Z$, i.e.

$$[g(x)] = tm + b \equiv b(\bmod m).$$

Further remarks

We have given in the Corollary only a simple example of how the Theorem can be applied, giving known results.

It is clear, however, that the sequences there defined contain many more terms than what is needed for the assumptions to be satisfied.

By a slight modification of the proof, it is possible to make it effective, i.e. to calculate a number M, such that, for $n \geqq M$, $\varrho_S(n) > 0$. Such an M will depend on

(1) An effective measure for the rate of decay of $(s_{u+1} - s_u)/s_u$;
(2) The distribution of the values u satisfying (ii);
(3) The integer k, B and the magnitude of the n_i;
(4) The magnitude of the terms of S needed to satisfy (iii).

In both applications of the Corollary these parameters can be effectively calculated, substantially in terms of the height of f, and of an effective measure of irrationality for the leading term of g.

REFERENCES

[1] Birch, B. J., Note on a problem of Erdős, *Proc. Cambridge Philos. Soc.* **55** (1959), 370–373. *MR* **22** #26

[2] Cassels, J. W. S., On the representation of integers as the sums of distinct summands taken from a fixed set, *Acta Sci. Math. (Szeged)* **21** (1960), 111–124. *MR* **24** #A103

[3] Cassels, J. W. S., *An introduction to Diophantine approximation,* Cambridge Tracts in Mathematics and Mathematical Physics, No. 45, Cambridge University Press, New York, 1957. *MR* **19**-396

[4] Erdős, P., On the representation of large integers as sums of distinct summands taken from a fixed set, *Acta Arith.* **7** (1961/62), 345–354. *MR* **26** #2387

[5] Perelli, A. and Zannier, U., On sums of distinct integers belonging to certain sequences, *Acta Math. Hungar.* **41** (1983), 251–254. *MR* **84j:** 10063

[6] Roth, K. F. and Szekeres, G., Some asymptotic formulae in the theory of partitions, *Quart. J. Math. Oxford Ser. (2)* **5** (1954), 241–259. *MR* **16**-797

(Received September 26, 1988)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SALERNO
GUIDECCA 783
I–30133 VENEZIA
ITALY

Present address:

ISTITUTO UNIVERSITARIO DI ARCHITETTURA
DIPARTIMENTO DI SCIENZA E TECNICA DEL RESTAURO
S. CROCE, 191
I–30135 VENEZIA
ITALY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 183–187

UNIQUE ENVELOPE PROPERTY

E. FRIED and G. GRÄTZER*

1. Introduction

A partial lattice P has the Unique Envelope Property if it can be completed in only one way to a lattice. This concept is a generalization of the Unique Amalgamation Property we investigated in an earlier paper.

There are a number of natural ways to complete a partial lattice P to a lattice; all these completions have to be isomorphic if P has the Unique Envelope Property. In this note, we prove that if two of these completions (namely, the free completion and the sublattice of the MacNeille completion generated by P) are isomorphic, then P has the Unique Envelope Property.

2. Preliminaries

For the basic concepts and notation of lattice theory, in particular, for the basic concepts and results concerning partial lattices, the reader is referred to [4].

We start this note with two definitions related to partial lattices. The reader should note the subtle differences between the lattice theoretic and partial lattice theoretic versions.

DEFINITION 1. Let L be a partial lattice, and let X be a subset of L. By restricting the $\vee$ and $\wedge$ from L to X, we obtain the partial lattice X, called the partial lattice *induced by L on X*.

DEFINITION 2. Let X and L be partial lattices and let ϕ be a map of X into L. Then ϕ is an *embedding* of X into L iff ϕ is an isomorphism between X and the partial lattice $X\phi$, the partial lattice induced by L on the set $X\phi$.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 06B99; Secondary 06A10.

Key words and phrases. Partial lattice, freely generated lattice, ideal lattice, MacNeille completion, essential extension.

* The research of both authors was supported by the NSERC of Canada.

If ϕ is the identity map, then X is a subset of L, and X is called a *partial sublattice* of L.

Note that an embedding is a one-to-one mapping and a homomorphism; the converse is not true.

The next definition defines the central concept of this note.

DEFINITION 3. Let P be a partial lattice. P has the *Unique Envelope Property* (UEP) iff there is a lattice P^e (the *unique envelope* of P) such that P is a partial sublattice of P^e, and whenever P is embedded into a lattice L and L is generated by P, then L and P^e are isomorphic over P. Alternatively, P has an embedding ψ into P^e; moreover, whenever P has an embedding ϕ into a lattice L, then there is an embedding δ of the lattice P^e into L satisfying $\psi\delta = \phi$.

OBSERVATION. Take the lattices A and B sharing the sublattice S, and take the partial algebra $\mathrm{Part}(A,B,S)$ and the poset $P(A,B,S)$ as defined in [1]. Then we can make $\mathrm{Part}(A,B,S)$ into a partial lattice $\mathrm{Pl}(A,B,S)$ by defining $a \vee b = c$ ($a \in A$, $b \in B$, $c \in A \cup B$) iff $(a] \vee (b] = (c]$ in the ideal lattice of $\mathrm{Part}(A,B,S)$, and dually (see, e.g., [4]). Then $\mathrm{Pl}(A,B,S)$ has the UEP iff A and B over S have the Unique Amalgamation Property.

There are a number of lattices in which P is naturally embedded: the free lattice, $F(P)$, over P; the ideal lattice, $\mathrm{Id}(P)$, of P; the dual ideal lattice, $\mathrm{D}(P)$, of P; the MacNeille completion, P^c, of P. P has a natural embedding into each, denoted by ψ^F, ψ^{Id}, ψ^{D}, ψ^c, respectively.

The image of P under ψ^F generates $F(P)$. The analogous statement fails for the other three lattices, in general. So it is useful to introduce notation for the sublattices generated by the natural images of P: $\mathrm{Id}_{\mathrm{fd}}(P)$ is the sublattice of $\mathrm{Id}(P)$ generated by ψ^{Id}; $\mathrm{D}_{\mathrm{fd}}(P)$ is the sublattice of $\mathrm{D}(P)$ generated by ψ^{D}; P^c_{fd} is the sublattice of P^c generated by ψ^c ("fd"stands for "finitely defined"; this is not the same as "finitely generated").

Applying the definition of UEP twice, we conclude that if P has the UEP, then $F(P)$, $\mathrm{Id}_{\mathrm{fd}}(P)$, $\mathrm{D}_{\mathrm{fd}}(P)$, and P^c_{fd} are isomorphic; in fact, isomorphic over P, that is, the isomorphisms commute with the isomorphisms ψ^F, ψ^{Id}, ψ^{D}, ψ^c.

3. Results

Any one of the isomorphisms mentioned at the end of Section 2 yields useful information. Let us illustrate this point with the isomorphism, α, between $\mathrm{Id}_{\mathrm{fd}}(P)$ and $\mathrm{D}_{\mathrm{fd}}(P)$:

CLAIM 1. *Let the partial lattice P have the UEP. If* $\sup(u,v) = w$ *in P, then* $(u] \vee (v] = (w]$ *in* $\mathrm{Id}_{\mathrm{fd}}(P)$ *(and in* $\mathrm{Id}(P)$*).*

PROOF. If $\sup(u, v) = w$ in P, then $[u) \cap [v) = [w)$, and so $[u) \vee [v) = [w)$ in $\mathrm{D}(P)$. Applying the isomorphism α, we obtain $(u] \vee (v] = (w]$ in $\mathrm{Id}_{\mathrm{fd}}(P)$, as claimed.

Since $(u] \vee (v] = (w]$ in $\mathrm{Id}_{\mathrm{fd}}(P)$ can easily be described using the structure of P, Claim 1 states a very strong consequence of UEP.

To obtain a characterization of the Unique Amalgamation Property, in the papers [2] and [3], we used a number of such consequences. So it was a great surprise to us that one of the many such consequences characterizes UEP. It is the purpose of this note to prove this statement.

THEOREM 1. *Let P be a partial lattice. Then P has the Unique Envelope Property if, and only if, there is an isomorphism α over P between $F(P)$ and P^{c}_{fd} (that is, $\psi^{F} \alpha = \psi^{c}$).*

The proof of this result easily follows from the following theorem which seems to be of independent interest; but first we need one more definition:

DEFINITION 4. Let A and B partial lattices, and let B be a partial sublattice of A. Then A is said to be an *essential extension* of B provided that whenever Θ is a congruence relation of A such that the natural map ϕ of B into A/Θ: $b\phi = [b]\Theta$ is an embedding of B into A/Θ, then $\Theta = \omega$.

In this definition, $[b]\Theta$ is the congruence class of A under Θ containing b.

Note again the subtle distinction between lattice theory and partial lattice theory: If A and B are lattices, B a sublattice of A, then A is said to be an *essential extension* of B provided that whenever Θ is a congruence relation of A such that the restriction of Θ to B is ω, then $\Theta = \omega$. This same definition would not work for partial lattices.

THEOREM 2. *P^{c}_{fd} is an essential extension of P (or, of $P\psi^{c}$, to be more precise).*

4. Proofs

In this section we prove Theorems 1 and 2.

PROOF of Theorem 2. Let Θ be a nontrivial congruence relation of P^{c}_{fd} such that the map ϕ of $P\psi^{c}$ into $P^{c}_{\mathrm{fd}}/\Theta$ defined by $(p]\phi = [(p]]\Theta$ is an embedding. Let $X, Y \in P^{c}_{\mathrm{fd}}$, $X \subset Y$, $X \equiv Y(\Theta)$. Let $p \in Y - X$; then $X \wedge (p] \equiv Y \wedge (p](\Theta)$, that is, $X \wedge (p] \equiv (p](\Theta)$, and $X \wedge (p] \subset (p]$. In other words, without loss of generality, we can assume that $Y \in P\psi^{c}$. Let $Y = (p]$.

If X is also principal, then $x \in P\psi^{c}$, and we found two distinct elements of P^{c}_{fd} congruent under Θ, contradicting that ϕ is one-to-one. So, without loss of generality, we can assume that X is not principal.

p is an upper bound of X in P. If there is a smaller upper bound q of X in P, then $X \equiv Y(\Theta)$ implies that $X \vee (q] \equiv Y \vee (q](\Theta)$, that is, $(q] \equiv (p](\Theta)$, and we found two distinct elements of P^{c}_{fd} congruent under Θ, again contradicting that ϕ is one-to-one.

Thus we may assume that p is a minimal upper bound of X in P. Since X is closed, p is not the least upper bound of X in P. Therefore, there is an upper bound u of X in P which is incomparable with p.

Then $X \equiv Y(\Theta)$ implies that $X \vee (u] \equiv Y \vee (u](\Theta)$, that is, $(u] \equiv Y \vee \vee (u](\Theta)$. Thus, in P^c_{fd}/Θ, $[(p]]\Theta \leqq [(u] \vee (p]]\Theta = [(u]]\Theta$. Since ϕ is an embedding, we conclude that $(p] \leqq (u]$, a contradiction. This contradiction proves Theorem 2.

Proof of Theorem 1. The "only if" part is trivial by Definition 3. So let α be an isomorphism over P between $F(P)$ and P^c_{fd}; α satisfies $\Psi^F \alpha = \psi^c$. We choose P^e to be $F(P)$. To show that P has the UEP, whenever P has an embedding ϕ into a lattice L, we must find an embedding δ of the lattice $F(P)$ into L satisfying $\psi^F \delta = \phi$. So let ϕ be given. Since both ψ^F and ϕ are embeddings, there is an isomorphism μ between $P\psi^F$ and $P\phi$ satisfying $\psi^F \mu = \phi$. Since $P\psi^F$ freely generates $F(P)$, μ extends to a homomorphism δ of $F(P)$ into L satisfying $\psi^F \delta = \phi$. It remains to show that δ is an embedding.

Let Θ be the kernel of δ is nontrivial. Using the isomorphism α, we obtain an isomorphic copy Θ' of Θ. $P^c_{\mathrm{fd}}/\Theta'$ is isomorphic over P to $F(P)\delta$, hence, the natural map η of $P\psi^c$ into $P^c_{\mathrm{fd}}/\Theta'$: $q\eta = [q]\Theta'$ is an embedding of $P\psi^c$ into $P^c_{\mathrm{fd}}/\Theta'$. By Theorem 2, P^c_{fd} is an essential extension of $P\psi^c$. By Definition 4, we conclude that $\Theta = \omega$; thus δ is one-to-one, completing the proof of Theorem 1.

5. Concluding remarks

One can derive many consequences from Theorem 1, along the lines of Claim 1, by comparing the arithmetic in $F(P)$ with the arithmetic in P^C, as we do it in [3].

Theorem 1 is not as powerful as the characterization of the Unique Amalgamation Property in [2], which is structural. It is an interesting problem to find a structural characterization of the Unique Envelope Property.

Let us mention an interesting group of problems: one can take the set of natural completions of a partial lattice P into a lattice; each pair of such lattices have to be naturally isomorphic. Compare the effect of these isomorphisms. For instance, if $\mathrm{Id}_{\mathrm{fd}}(P)$, $\mathrm{D}_{\mathrm{fd}}(P)$, and P^c_{fd} are isomorphic, does that imply the Unique Envelope Property?

Finally, we should point out that there is a characterization of the Unique Envelope Property in a universal algebraic framework. Let $\mathbf{K}$ be a variety of universal algebras; let P be a partial algebra that can be embedded in an algebra in $\mathbf{K}$. Let $F(P)$ denote the free algebra in $\mathbf{K}$ generated by P. Then P has the Unique Envelope Property with respect to $\mathbf{K}$ if and only if $F(P)$ is an essential extension of P. The proof is very easy.

REFERENCES

[1] Fried, E. and Grätzer, G., Pasting infinite lattices, *J. Australian Math. Soc.* **47** (1989), 1–21.

[2] Fried, E. and Grätzer, G., The unique amalgamation property for lattices, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **33** (1990), 167–176.

[3] Fried, E. and Grätzer, G., Strong amalgamation of distributive lattices, *J. Algebra* **128** (1990), 446–455.

[4] Grätzer, G. *General lattice theory,* Pure and Applied Mathematics, **75**, Academic Press, New York, N. Y., 1978; Birkhäuser Verlag, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Mathematische Reihe, Band 52, Basel, 1978. *MR* **80c:** 06001a, 06001b.

(Received October 3, 1988)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANITOBA
WINNIPEG, MANITOBA
R3T 2N2
CANADA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 189–200

ON THE COHOMOLOGY $H^*(L_k, L_s)$

A. FIALOWSKI

Introduction

Let $W_1^{pol} = W_1$ be the infinite dimensional Lie algebra of vector fields $f(x)\frac{d}{dx}$ on the line with polynomial coefficients. The Lie algebra W_1 has an additive algebraic basis consisting of the fields $e_k = x^{k+1}\frac{d}{dx}$, $k \geqq 1$, in which the bracket is described by

$$[e_k, e_\ell] = (\ell - k)e_{k+\ell}.$$

Consider the subalgebras L_k, $k \geqq 0$ of W_1, consisting of the fields such that they and their first k derivatives vanish at the origin. The Lie algebra L_k is generated by the basis elements $\{e_k, e_{k+1}, \dots\}$. The algebras W_1 and L_k are naturally graded by $\deg e_i = i$. Obviously, the infinite dimensional subalgebras L_k of W_1 are nilpotent for $k \geqq 1$.

The cohomology theory of infinite dimensional Lie algebras is worked out in [6]. The cohomology rings $H^*(W_1)$ and $H^*(L_k)$, $k \geqq 0$, with trivial coefficients are computed in [7] and [8]. The main results are the following:

1 $\quad H^q(W_1) = \begin{cases} \mathbf{C} & \text{for } q = 0, 3 \\ 0 & \text{for all other } q, \end{cases}$

2 $\quad H^q(L_0) = \begin{cases} \mathbf{C} & \text{for } q = 0, 1 \\ 0 & \text{for } q > 1, \end{cases}$

3 $\quad \dim H^q(L_k) = \binom{q+k-1}{k-1} + \binom{q+k-2}{k-2}$ for $k \geqq 1$.

In particular,

$$\dim H^q(L_1) = 2 \text{ for } q \geqq 1,$$
$$\dim H^q(L_2) = 2q + 1,$$
$$\dim H^q(L_3) = (q+1)^2.$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 17B56; Secondary 17B65.

Key words and phrases. Lie algebra of vector fields, Lie algebra cohomology, spectral sequence.

The work was partly done during a fellowship of the Alexander von Humboldt-Stiftung at the Max-Planck-Institut für Mathematik, Bonn.

With the help of this last result, one can compute the cohomology of the Lie algebra W_1 with coefficients in different modules. For each $\lambda \in \mathbf{C}$, let F_λ denote the W_1-module of the tensor fields of the form $f(z)dz^{-\lambda}$, where $f(z)$ is a formal power series in z. Then the formula

$$(g\frac{d}{dx})f dx^{-\lambda} = (gf' - \lambda f g')dx^{-\lambda}$$

gives the action of W_1 on F_λ. The module F_λ has an additive basis $\{f_j \mid j = 0, 1, \dots\}$ where $f_j = x^j dx^{-\lambda}$ and the action on the basis elements is

$$e_i f_j = (j - (i+1)\lambda) f_{i+j}.$$

Denote by $\mathcal{F}_\lambda$ the W_1-module which is defined in the same way, except that the index j runs over all integers. Define the adjoint modules F'_λ, $\mathcal{F}'_\lambda$ as modules of linear functionals $F_\lambda \to \mathbf{C}$, $\mathcal{F}_\lambda \to \mathbf{C}$ which are finite in the sense that they take nonzero values only on a finite number of f_j-s. Obviously $\mathcal{F}'_\lambda = \mathcal{F}_{-1-\lambda}$ and $F'_\lambda = \mathcal{F}_{-1-\lambda}/F_{-1-\lambda}$.

Let us define now the L_0-module $F_{\lambda,\mu}$ as the subspace, generated — like F_λ — by the elements f_j, $j = 0, 1, \dots$, on which L_0 acts by

$$e_i f_j = (j + \mu - (i+1)\lambda) f_{i+j}.$$

In this definition μ can be an arbitrary complex number. Let $F'_{\lambda,\mu}$ denote the module, conjugate to $F_{\lambda,\mu}$. Finally define the modules $\mathcal{F}_{\lambda,\mu}$ over W_1 as $F_{\lambda,\mu}$ above, without requiring the positivity of j. Obviously, $\mathcal{F}'_{\lambda,\mu} = \mathcal{F}_{-1-\lambda,-\mu}$ and $F'_{\lambda,\mu} = \mathcal{F}_{-1-\lambda,-\mu}/F_{-1-\lambda,-\mu}$. The cohomology of the Lie algebra W_1 and L_1 with coefficients in the above mentioned tensor field modules are known (see [2], [3]). The computation reduces to that of the cohomology of the algebra L_1 with trivial coefficients. In the case of the Lie algebra W_1, the problem is also solved for cohomology with coefficients of the form $F_\lambda \otimes F_\mu$ [2].

Considering the adjoint representation as coefficient space, we get a very important application of the cohomology. The elements of the space $H^2(L, L)$ correspond to the infinitesimal deformations of the Lie algebra L (see e.g. [6]). In the case of the Lie algebra L_0, we get that $H^2(L_0; L_0) = 0$, consequently, L_0 is rigid (see [5] and [6]). As an L_1-module, L_1 is $F_{1,1}$, and we have the result $\dim H^q_{(-m)}(L_1; L_1) = \dim H^q_{(-m)}(L_2; \mathbf{C})$ (see [4]). In particular,

$$\dim H^2_{(-m)}(L_1; L_1) = \begin{cases} 1 & \text{for } m = 2, 3, 4 \\ 0 & \text{for } m \neq 2, 3, 4. \end{cases}$$

Here the cohomology space is defined in the graded sense:

$$H^q(L_1; L_1) = \bigoplus_m H^q_{(-m)}(L_1; L_1)$$

where for the cocycle ϕ, representing a class of $H^q_{(-m)}(L_1;L_1)$, the weight of $\phi(e_{i_1},\dots,e_{i_q})$ is $-m+i_1+\dots+i_q$.

The analogous problem for the Lie algebras L_k, $k>1$ seems to be very difficult and has not been solved. Nor has any other cohomology space for these Lie algebras been computed, with other than trivial coefficients. We naturally want to know first of all the cohomology with coefficients in the adjoint representation. In this paper, we prove the finiteness of these cohomology spaces, and also give lower bounds for their dimensions. In fact, we give estimates in the more general cases, where the coefficient module in an L_k-module L_s: we study the cohomology spaces $H^*(L_k;L_s)$ with $k \geqq 1$ and $s \geqq 1$. We make the computations for the lower bound for $k=2$.

Special attention is paid to the cohomology $H^2(L_2;L_2)$ which is, according to the general theory, the space of infinitesimal deformations of the Lie algebra L_2. The computation of $H^2(L_2;L_2)$ is the first step in determining the base of the versal deformation of the Lie algebra L_2; recall that the similar problem for L_1 is completely solved in [4].

I would like to thank Dmitrij Fuchs for stimulating discussions.

§1. Upper bounds. Finiteness

Suppose first that $s>k$. Then the L_k-module cohomology $H^q(L_k;L_s)$ may be estimated via the cohomology $H^q(L_{k+1})$. Consider the module $\mathrm{Ind}^{L_k}_{L_{k+1}}(1)$. It has the following structure: it has the basis $x_1,x_2,x_3,x_4,\dots$ and $e_kx_i=x_{i+1}$, $e_\ell x_i=0$ for $\ell>k$. The sum

$$\bigoplus_k \mathrm{Ind}^{L_k}_{L_{k+1}}(1)$$

is isomorphic to the module $N_{k,s}$ which has the following structure: it has the basis $y_s, y_{s+1}, y_{s+2},\dots$.

The action of the vector field e_k is shown by the arrows, while the fields e_ℓ with $\ell>k$ act trivially. One may assume that $e_ky_t=(t-k)y_{t+k}$. (It is important here that $s>k$, so that the difference $t-k$ cannot vanish.)

Now consider the complex $C^*(L_k;L_s)$. It has a natural (decreasing) filtration: $F^{m,q}=F^mC^q(L_k;L_s)$ consists of those $c\in C^q(L_k;L_s)$, for which $c(e_{i_1},\dots,e_{i_q})=0$ for $i_1+\dots+i_q<m$. Consider the groups

$$G^{m,q}=\frac{F^{m,q}\cap d^{-1}F^{m+r,q+1}}{(F^{m,q}\cap dF^{m-r,q-1})+(F^{m+1,q}\cap d^{-1}F^{m+r,q+1})}.$$

They lie between the groups $E_r^{m,q-m}$ and $E_{r+1}^{m,q-m}$ of the spectral sequence, corresponding to our filtration, and hence the sum $\oplus_m G^{m,q}$ gives the upper estimate for $H^q(L_k; L_s)$. On the other hand, for $r = k+1$ this sum is precisely the cohomology of the complex with all the summands corresponding to the action of e_t with $t > r$ removed from the formula for the differential. But this is nothing else but $H^q(L_k; N_{k,s})$.

Remark. The descending filtration $\{F^m\}$ of the complex $C^*(L_k; L_s)$ is infinite, and this could create some convergence problems for the spectral sequence, but in our case everything is good thanks to the finite dimensionality of $H^*(L_k)$. What we have to check is essentially the fact that $\bigcap_m \mathrm{Im}\{H^q(F^m) \to H^q(L_k; L_s)\} = 0$. But actually $H^q(F^m) = 0$ for m being large enough. Indeed, let $H^q_{(m')}(L_k) = 0$ for $m' \geqq m$ (the subscript is related to the grading in L_k). Let then $c = c_m \in F^{m,q}$ be a cocycle, also of some fixed degree d. Then $c(e_{i_1}, \dots, e_{i_q}) = 0$ for $i_1 + \dots + i_q < m$ and let $c(e_{i_1}, \dots, e_{i_q}) = c_{i_1 \dots i_q} e_{m-d}$ for $i_1 + \dots + i_q = m$. Then $\overline{c}$, $\overline{c}(e_{i_1}, \dots, e_{i_q}) = c_{i_1 \dots i_q}$ is a cocycle in $C^q_{(m)}(L_k)$, and hence $\overline{c} = \delta \overline{b}$, $\overline{b} \in C^{q-1}_{(m)}(L_k)$. Define $b_m \in F^{m,q-1}$ by the formula

$$b_m\left(e_{j_1}, \dots, e_{j_{q-1}}\right) = \begin{cases} \overline{b}\left(e_{j_1}, \dots, e_{j_{q-1}}\right) e_{m-d} & \text{for } j_1 + \dots + j_{q-1} = m, \\ 0 & \text{otherwise.} \end{cases}$$

Evidently, $c_m - \delta b_m \in F^{m+1}$; we set $c_{m+1} = c_m - \delta b_m$ and acting precisely as before find b_{m+1} with $c_{m+1} - \delta b_{m+1} \in F^{m+2}$; set $c_{m+2} = c_{m+1} - \delta b_{m+1}$, and so on. The series $\sum_{p=m}^{\infty} b_m$ obviously converges (b_m have disjoint supports), and $c = \delta b$ where b is the sum of the series.

Hence the following statement is valid.

Theorem 1. *If $s > k$,*

$$\dim H^q(L_k; L_s) \leqq k \dim H^q(L_k; \mathrm{Ind}_{L_{k+1}}^{L_k}(1)) = \\ = k \dim H^q(L_{k+1}) < \infty.$$

Remark. The last inequality follows from [8].

The general case (s is arbitrary natural number, may be less than or equal to k) can be traced back to the previous one. If $s \leqq k$, then consider the exact sequence

$$0 \to L_{k+1} + L_s \to L_s / L_{k+1} \to 0.$$

The cohomology spaces $H^q(L_k; L_s/L_{k+1})$, $q = 1, 2, \dots$ are finite dimensional. In particular, LEMMA.

$$\dim H^q(L_k; L_s/L_{k+1}) \leqq \dim H^q(L_k) \dim H^q(L_k; L_s/L_k) = \\ = (k+1-s) \dim H^q(L_k).$$

PROOF. Denote the module L_{k-i+1}/L_{k+1} by M_i. We have the exact sequence

$$0 \to M_{i-1} \to M_i \to \mathbf{C} \to 0$$

from which it follows that $M_i/M_{i-1} = C$. Then

$$L_s/L_{k+1} = M_{k+1-s} \supset M_{k+s} \supset \ldots \supset M_1 \supset M_0 = 0,$$

and we have the next cohomology sequence:

$$\ldots \to H^q(L_k; M_{i-1}) \to H^q(L_k; M_i) \to H^q(L_k; \mathbf{C}) \to \ldots .$$

From this it follows that

$$\dim H^q(L_k; M_i) \leqq \dim H^q(L_k; M_{i-1}) \oplus \dim H^q(L_k)$$

which gives $\dim H^q(L_k; M_i) \leqq i \dim H^q(L_k)$. The lemma is proved.

THEOREM 2. *If* $s \leqq k$,

$$\dim H^q(L_k; L_s) \leqq k \dim H^q(L_{k+1}) + (k+1-s) \dim H^q(L_k).$$

PROOF. Consider the next cohomology sequence:

$$\ldots \to H^q(L_k; L_{k+1}) \to H^q(L_k; L_s) \to H^q(L_k; L_s/L_{k+1}) \to \ldots .$$

From this it follows that

$$\dim H^q(L_k; L_s) \leqq \dim H^q(L_k; L_{k+1}) + \dim H^q(L_k; L_s/L_{k+1}).$$

Using now the lemma, we get the result.

EXAMPLE. For small values of q,

$$\dim H^q(L_k) = \begin{cases} 1 \text{ for} & q = 0 \\ k+1 \text{ for} & q = 1 \\ \frac{k(k+3)}{2} \text{ for} & q = 2. \end{cases}$$

Consequently, we have

$$\dim H^1(L_k; L_k) \leqq k(k+2) + (k+1) = k^2 + 3k + 1$$

and

$$\dim H^2(L_k; L_k) \leqq \frac{k(k+1)(k+4)}{2} + \frac{k(k+3)}{2} = \frac{k}{2}(k^2 + 6k + 7).$$

The first inequality shows that our upper bounds are too crude, because for $H^1(L_k; L_k)$ we have

PROPOSITION.
$$\dim H^1(L_k; L_k) = k.$$

PROOF. Consider the L_k-module L_0, and the following exact sequence:

$$0 \to L_k \to L_0 \to L_0/L_k \to 0.$$

Obviously, the dimension of the trivial module L_0/L_k is k. Then we have the next cohomology sequence:

$$H^0(L_k; L_0) \to H^0(L_k; L_0/L_k) \xrightarrow{\tau} H^1(L_k; L_k) \xrightarrow{\mu} H^1(L_k; L_0) \to .$$

Here $H^0(L_k, L_0) = 0$ as there are no invariants. Further, $\dim H^0(L_k; L_0/L_k) = k$ and τ is monomorphism. Easy to see that μ is zero, consequently $\dim H^1(L_k; L_k) = k$. Remark that the exterior derivations of L_k are the bracket operations with $e_0, e_1, \dots, e_{k+1}$.

For $k = 2$ the second inequality gives

$$\dim H^2(L_2; L_2) \leqq 23.$$

§2. Lower bounds

We can compute the cohomology $H^*(L_k; L_s)$, with the help of the spectral sequence associated to the filtration

$$L_s \supset L_{s+1} \supset L_{s+2} \supset \dots$$

in the coefficient module. As L_{t+1}/L_t is a one-dimensional trivial L_k-module, the first term of this spectral sequence consists of the group $H^q(L_k)$. The spectral sequence itself is a modification of the one introduced by Feigin and Fuks in [2] for computing the cohomology space of the Lie algebra L_1 with coefficients in the modules $F_{\lambda,\mu}$ (see Introduction). Let us give a more convenient construction of this spectral sequence. We have the grading in the cohomology spaces:

$$H^*(L_k; L_s) = \bigoplus_r H^*_{(r)}(L_k; L_s)$$

where $H^*_{(r)}(L_k; L_s)$ is the cohomology of the cochain complex of the form

$$(*) \qquad e_{i_1} \wedge \dots \wedge e_{i_q} \to \alpha_{i_1,\dots,i_q} e_{i_1+\dots+i_q+r}, \quad r \in \mathbf{Z}.$$

Denote the cochain groups of this complex by $C^q_{(r)}(L_k; L_s)$. These cochain groups are similar to the cochain groups $C^q(L_k)$, with the following difference: in $(*)$ one should have $i_1 + \ldots + i_q + r \geqq s$. So we have the following situation. The cochain groups $C^*(L_k)$ are also graded:

$$C^q(L_k) = \bigoplus_r C^q_{(r)}(L_k),$$

and we have the isomorphism

$$(**) \qquad C^q_{(r)}(L_k; L_s) \cong \bigoplus_{t \geqq s-r} C^q_{(t)}(L_k).$$

This isomorphism is given by the mapping

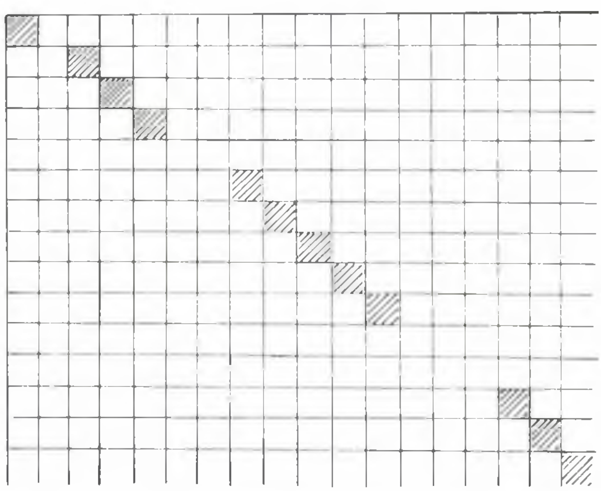
$$\begin{aligned}&\left[e_{i_1} \wedge \ldots \wedge e_{i_q} \to \alpha_{i_1 \ldots i_q} e_{i_1 + \ldots + i_q + r}\right] \in C^q_{(r)}(L_k; L_s) \leftrightarrow \\ \leftrightarrow &\left[e_{i_1} \wedge \ldots \wedge e_{i_q} \to \alpha_{i_1 \ldots i_q}\right] \in C^q_{(t)}(L_k), \qquad t = i_1 + \ldots + i_q.\end{aligned}$$

We assign to the summand $C^q_{(t)}(L_k)$ in the expansion $(**)$ the filtering index t. So we have the spectral sequence with:

$$E_0^{p,q} = \begin{cases} C^{p+q}_{(-p)}(L_k) & \text{for } p \geqq s - r, \\ 0 & \text{for } p < s - r; \end{cases}$$

$$E_1^{p,q} = \begin{cases} H^{p+q}_{(-p)}(L_k) & \text{for } p \geqq s - r, \\ 0 & \text{for } p < s - r. \end{cases}$$

The spectral sequence converges in usual sense to $H^*_{(q)}(L_k; L_s)$. For $r \geqq s$ the E^1 term of our spectral sequence does not depend on r; for example, if $k = 2$ it has the form



(Here ■ means a one-dimensional space; the differentials act by the usual way to the right and then down.)

Let $E_\alpha^n = \bigoplus\limits_{p+q=n} E_\alpha^{p,q}$ and $\varrho^n = \varrho^n(r)$ be equal to the sum of ranks of differentials $d_\alpha^n \colon E_\alpha^n \to E_\alpha^{n+1}$ (for $\alpha \geqq 1$). Clearly

$$\dim H_{(r)}^n(L_k; L_s) = \dim E_1^n - \varrho^{n-1} - \varrho^n.$$

For almost all $r \geqq s$, the numbers $\varrho^n(r)$ are the same (as the differential depends on r polynomially). These generic values of $\varrho^n(r)$ may be easily calculated since we know that $\dim H^q(L_k; L_s)$ is finite for any q (see Section 1) and thus $H_{(r)}^q(L_k; L_s) = 0$ for almost all r.

Consider the case $k = 2$. In this case the dimensions of E_1^α are the following:

r	E_1^0		E_1^1		E_1^2		E_1^3		$\ldots$
$s+2$	1	$\to$	3	$\to$	5	$\to$	7	$\to$	
$s+1$	1	$\to$	3	$\to$	5	$\to$	7	$\to$	$\ldots$
s	1	$\to$	3	$\to$	5	$\to$	7	$\to$	$\ldots$
$s-1$	0		3	$\to$	5	$\to$	7	$\to$	$\ldots$
$s-2$	0		2	$\to$	5	$\to$	7	$\to$	$\ldots$
$s-3$	0		1	$\to$	5	$\to$	7	$\to$	$\ldots$
$s-4$	0		0		5	$\to$	7	$\to$	$\ldots$
$s-5$	0		0		5	$\to$	7	$\to$	$\ldots$
$s-6$	0		0		5	$\to$	7	$\to$	$\ldots$
$s-7$	0		0		5	$\to$	7	$\to$	$\ldots$
$s-8$	0		0		4	$\to$	7	$\to$	$\ldots$
$s-9$	0		0		3	$\to$	7	$\to$	$\ldots$
$s-10$	0		0		2	$\to$	7	$\to$	$\ldots$
$s-11$	0		0		1	$\to$	7	$\to$	$\ldots$
$s-12$	0		0		0		7	$\to$	$\ldots$
$\ldots$	$\ldots$		$\ldots$		$\ldots$		$\ldots$		
$s-15$	0		0		0		7	$\to$	$\ldots$
$s-16$	0		0		0		6	$\to$	$\ldots$etc.

The complex

$$1 \to 3 \to 5 \to 7 \to 9 \to \ldots$$

(the numbers are the dimensions of the spaces) is acyclic iff the differentials are of ranks $1, 2, 3, 4, 5, \ldots$. Thus at a generic point our complex is of the form

$$1 \xrightarrow{1} 3 \xrightarrow{2} 5 \xrightarrow{3} 7 \xrightarrow{4} 9 \xrightarrow{5} \ldots$$

(the rank of the differential is shown above the arrow). At a finite number of exceptional points the ranks are *smaller*. In those cases the complex is truncated. If a matrix $A = \blacksquare$ has rank $\leqq r$ then the rank of the truncated matrix $\underbrace{\blacksquare}_{k}$ cannot be larger than $\max(r, k)$.

The cohomology space will be the smallest, if the ranks are maximal as we described (the smaller the rank, the larger the cohomology space). The maximal possible ranks are the following (the rank does not drop anywhere):

$$
\begin{array}{lllcccc}
 & & & \multicolumn{4}{c}{\text{cohomology}} \\
 & & & H^0 & H^1 & H^2 & H^3 \ \dots \\
r \geqq s & 1 \xrightarrow{1} 3 \xrightarrow{2} 5 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 0 \\
s-1 & & & 0 & 1 & 0 & 0 \\
 & \quad\ \ 3 \xrightarrow{2} 5 \xrightarrow{3} 7 \xrightarrow{4} \dots & & & & & \\
s-2 & & & 0 & 1 & 0 & 0 \\
s-3 & \quad\ \ 2 \xrightarrow{2} 5 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 0 \\
s-4 & \quad\ \ 1 \xrightarrow{1} 5 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 1 & 0 \\
s-5 & & & 0 & 0 & 2 & 0 \\
s-6 & \quad\ \ 0 \qquad 5 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 2 & 0 \\
s-7 & & & 0 & 0 & 2 & 0 \\
s-8 & \qquad\qquad 4 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 1 & 0 \\
s-9 & \qquad\qquad 3 \xrightarrow{3} 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 0 \\
s-10 & \qquad\qquad 2 \xrightarrow{2} 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 1 \\
s-11 & \qquad\qquad 1 \xrightarrow{1} 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 2 \\
s-12 & \qquad\qquad\qquad 7 \xrightarrow{4} \dots & & 0 & 0 & 0 & 3 \\
s-13 & \qquad\qquad\qquad 7 \xrightarrow{4} \dots & & & & & 3 \\
s-14 & \qquad\qquad\qquad 7 \xrightarrow{4} \dots & & & & & 3 \\
s-15 & \qquad\qquad\qquad 7 \xrightarrow{4} \dots & & & & & 3 \\
s-16 & \qquad\qquad\qquad 6 \xrightarrow{4} \dots & & & & & 2 \\
 & \qquad\qquad\qquad 5 \longrightarrow \dots & & & & & 1 \\
 & \qquad\qquad\qquad 4 \longrightarrow \dots & & & & & 0 \\
 & \qquad\qquad\qquad \dots & & & & & \dots \\
\hline
 & & & 0 & 2 & 8 & 18
\end{array}
$$

Here the ranks are maximal, but everywhere we must have

$$
\begin{aligned}
&\text{rank } \leqq 3 \text{ and} \\
&\text{rank } \leqq \text{ number of rows.}
\end{aligned}
$$

From these computations one obtains:

THEOREM 3.

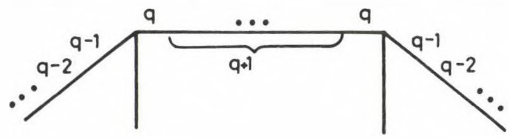
$$\dim H^q(L_k; L_s) \geqq 2q^2$$

for any q, not depending on s. In particular, we obtain

$$
\begin{aligned}
\dim H^2(L_2; L_2) &\geqq 8, \\
\dim H^3(L_2; L_2) &\geqq 18.
\end{aligned}
$$

PROOF. From the above computations we get

$$\dim H^q(L_2, L_s) = \frac{q(q+1)}{2} + q(q+1) + \frac{q(q-1)}{2} = 2q^2.$$



Comparing the above result with the estimation from Section 1:

$$\dim H^q(L_2; L_s) \leqq 2H^q(L_3) = 2(q+1)^2 \text{ for } s \geqq 2.$$

This means that the dimension is between $2q^2$ and $2(q+1)^2$. The lower bound is perhaps closer to the real value of the dimension.

Let us write out separately the results for the 2-dimensional cohomology space of the Lie algebra L_2 with coefficients in the adjoint representation:

$$8 \leqq \dim H^2(L_2; L_2) \leqq 23.$$

The computation in Section 2 for L_k with $k > 2$ is analogous, but, because of the unwieldy formulas for $\dim H^q_{(r)}(L_k)$, the answer would be very complicated. But it seems likely that the following is true.

CONJECTURE. *For any k and s,*

$$\dim H^q(L_k; L_s) = k \dim H^{q-1}(L_{k+1}).$$

If this is true, then the estimate of Theorem 3 is actually exact. But the similar estimates for L_k with $k > 2$ are only close to reality. For example, the procedure of this section gives for $q = 1, 2, 3, 4$

$$\dim H^q(L_3; L_s) \geqq 3, 15, 41, 87,$$

while the conjecture asserts that

$$\dim H^q(L_3; L_s) = 3, 15, 42, 90.$$

§3. Another method for computing $H^2(L_2; L_2)$

With the help of results from [3] we have another method of computing the cohomology $H^2(L_2; L_2)$. We have

$$H^2(L_2; L_2) = H^2(L_1; M)$$

where M is the "coinduced" module $\mathrm{Coind}_{L_2}^{L_1} L_2$. This module is generated by the elements

$$\begin{array}{c} e_2,\ e_3,\ e_4,\dots \\ xe_2,\ xe_3,\ xe_4,\dots \\ x^2e_2,\ x^2e_3,\ x^2e_4,\dots \\ \vdots \end{array}$$

The action of the Lie algebra is the following. The action of the field e_1 is multiplication by x, while in the upper row $e_2, e_3, \dots$ act in the natural way. The actions of the other elements are defined using the previous ones. For instance, for $n \geqq 2$, $k \geqq 2$ we have

$$(k-m)e_{m+k} = e_m(e_k) = \frac{1}{m-2}[xe_{m-1}(e_k) - e_{m-1}(xe_k)]$$

from which it follows that

$$e_m(xe_k) = (k-m)xe_{m+k} - (m-1)(k-m-1)e_{m+k+1}.$$

Analogously,

$$\begin{gathered} e_m(x^ne_k) = xe_m(x^{n-1}e_k) - (m-1)e_{m+1}(x^{n-1}e_k) = \\ = (k-m)x^ne_{m+k} + n(m-1)(k-m-1)x^{n-1}e_{m+k+1} + \\ + \text{lower degree terms at } x \end{gathered}$$

which gives a recurrent method of determining $e_m(x^ne_k)$.

Having the structure of the module M, we apply the complex

$$M \leftarrow M\overset{1}{\bigoplus}M \leftarrow M\overset{2}{\bigoplus}M \leftarrow M\overset{3}{\bigoplus}M \leftarrow \dots$$

(see [3]). To find the space $H^2(L_2; L_2)$, it is enough to know the singular vectors in the Verma module $V_{0,0}$ of degree 5, 7, 12, 15. We should remark that although the above computation can be worked out, and should give a precise answer, it leads to complicated formulas.

REFERENCES

[1] FEIGIN, B. L. and FIALOWSKI, A., Cohomology of the infinite-dimensional Lie algebra L_1 with nontrivial coefficients, *Bull. Amer. Math. Soc. (N.S.)* **17** (1987), 333–337. *MR* **89b**: 17016

[2] FEIGIN, B. L. and FUKS, D. B., Homology of the Lie algebra of vector fields on the line, *Funktsional. Anal. i Prilozhen.* **14** (1980), 45–60. *MR* **82 b**: 17017

[3] Feigin, B. L. and Fuks, D. B., Verma modules over the Virasoro algebra, *Topology* (Leningrad, 1982), Lecture notes in mathematics, vol. 1060, Springer–Verlag, Berlin–New York, 230–245. *MR* **86g**: 17004

[4] Fialowski, A., Deformations of a Lie algebra of vector fields on the line, *Uspekhi Mat. Nauk* **38** (1983), 201–202. *MR* **84m**: 17011; Russian Math. Surveys, 38 (1983), 201–202.

[5] Fialowski, A., An example of formal deformations of Lie algebras, Proceedings of the NATO-ASI "Deformation theory" Conference, Italy 1986, Kluwer, Dordrecht, 1988, 375–402.

[6] Fuks, D. B., *Cohomology of infinite-dimensional Lie algebras*, Contemporary Soviet Mathematics, Consultants Bureau, New York, N. Y., 1986. *MR* **88b**: 17001

[7] Gelfand, I. M. and Fuks, D. B., Cohomologies of the Lie algebra of formal vector fields, *Izv. Akad. Nauk. SSSR Ser. Mat.* **34** (1970), 322–337. *MR* **42** #1103

[8] Gončarova, L. V., Cohomology of Lie algebras of formal vector fields on the line, *Funkcional. Anal. i Priložen.* **7** (1973), no. 2, 6–14, no. 3, 33–44. *MR* **49** #4058a, 4058b

(Received November 17, 1988)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616–8633
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 201–205

ON THE LEFT HEREDITARINESS OF PSEUDOCOMPLEMENT OF HEREDITARY RADICALS

E. R. PUCZYŁOWSKI

In [7] Snider proved that if S_1, S_2 are hereditary radicals of associative rings then $(S_1 : S_2) = \{R \mid S_2(R') \subseteq S_1(R')$ for every homomorphic image R' of $R\}$ is a radical and is the pseudocomplement of S_2 relative to S_1 in the lattice of all radicals of associative rings. (Actually Snider uses $(S_1 : S_2)$ to denote the pseudocomplement in the lattice of hereditary radicals. We hope that this small change of notation will not lead to misunderstanding). In [2] Jaegermann and Sands proved that if S_1, S_2 are N-radicals then so is $(S_1 : S_2)$. Recall that S is an N-radical [6] if it is left strong, left hereditary and contains the prime radical β. In [4] it was shown that $(S_1 : S_2)$ is left strong whenever S_2 is left hereditary and S_1 is hereditary, left strong and contains β. In this paper we show that $(S_1 : S_2)$ is left hereditary whenever S_1 is left hereditary containing β and S_2 is hereditary and left strong. This result is in a sense dual to that of [4] and together they generalize the quoted Jaegermann and Sands' result. We also discuss necessity of the assumptions of the theorem.

Throughout all rings are associative. To denote that I is an ideal (left ideal) of a ring R, we write $I \triangleleft R$ $(I < R)$.

All considered radicals are radicals of associative rings. Recall that a radical S is called hereditary (left hereditary) if $I \triangleleft R$ $(I < R)$ and $R \in S$ imply $I \in S$. A radical S is called left strong if $I < R, I \in S$ imply $I \subseteq S(R)$. Fundamental definitions and properties of radicals may be found in [8].

The lower radical determined by the union of radicals S_1, S_2 is denoted $S_1 \vee S_2$. We use $0, 1$ and β to denote the radical consisting only of the trivial ring $\{0\}$, the radical of all rings and the prime radical, respectively. Observe that for every hereditary radicals $S_1, S_2, (S_1 : S_2) = \{R \mid$ for every homomorphic image R' of R if $S_1(R') = 0$ then $S_2(R') = 0\}$.

1. In this section we prove the following

THEOREM. *If S_1 is a left hereditary radical containing β and S_2 is a hereditary and left strong radical then the radical $(S_1 : S_2)$ is left hereditary.*

For the proof we need the following lemmas.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 16A21; Secondary 16A22.

Key words and phrases. Hereditary radical, left hereditary radical, strong radical, pseudocomplement of radicals.

Akadémiai Kiadó, Budapest

LEMMA 1. *For the hereditary radicals* $S, S_1, S_2, (S_1 : S_2) \subseteqq (S_1 \vee S : S_2 \vee S)$.

PROOF. We have to show that if $R \in (S_1 : S_2)$ and $(S_1 \vee S)(R) = 0$ then $(S_2 \vee S)(R) = 0$. However, $(S_1 \vee S)(R) = 0$ if and only if $S_1(R) = 0$ and $S(R) = 0$. Since $R \in (S_1 : S_2)$ and $S_1(R) = 0$ also $S_2(R) = 0$. Thus $(S_2 \vee S)(R) = 0$.

LEMMA 2 ([5], Corollary 10). *If* S *is a left strong and hereditary radical then so is the radical* $S \vee \beta$.

LEMMA 3. *Suppose that* S *is a radical,* $I \triangleleft L < R$ *and* $S(L/I) = J/I$. *Then* $LIR \triangleleft R$ *and* $J + LIR/LIR \in S \vee \beta$.

PROOF. Obviously, $LIR \triangleleft R$ and $J + LIR/LIR \approx J/U$, where $U = J \cap LIR$. Observe that $I^3 \subseteqq U$, $I^3 \triangleleft J$ and $J/I^3/U/I^3 \approx J/U$. Now $J/I \in S$, so $J/I \approx J/I^3/I/I^3 \in S$. However, $I/I^3 \in \beta$, so $J/I^3 \in S \vee \beta$. Thus J/U, being a homomorphic image of J/I^3, is in $S \vee \beta$.

PROOF of the theorem. Let $R \in (S_1 : S_2)$ and $L < R$. We have to show that if for an ideal I of L, $S_1(L/I) = 0$ then $S_2(L/I) = J/I = 0$. Putting in Lemma 3, $S = S_2$, one gets that $J + LIR/LIR \in S_2 \vee \beta$. Since the radical $S_2 \vee \beta$ is hereditary and $LJ + LIR/LIR \triangleleft J + LIR/LIR$, $LJ + LIR/LIR \in S_2 \vee \beta$. Observe that $LJ + LIR/LIR < R/LIR$. Hence by Lemma 2, $LJ + LIR/LIR \subseteqq (S_2 \vee \beta)(R/LIR)$. Since $R \in (S_1 : S_2)$, putting in Lemma 1 $S = \beta$, one gets $R \in (S_1 \vee \beta : S_2 \vee \beta) = (S_1, S_2 \vee \beta)$. Thus $LJ + LIR/LIR \subseteqq (S_2 \vee \beta)(R/LIR) \subseteqq S_1(R/LIR)$. Now, since S_1 is left hereditary and $LJ + LIR/LIR < R/LIR$, $LJ + LIR/LIR \in S_1$. Observe that $LJ + LIR/LIR \triangleleft J + LIR/LIR$ and $(J + LIR/LIR)^2 \subseteqq LJ + LIR/LIR$. Thus, since $\beta \subseteqq S_1$, $J/U \approx J + LIR/LIR \in S_1$, where $U = J \cap LIR$. Since $J/U^2/U/U^2 \in S_1$ and $\beta \subseteqq S_1$, $J/U^2 \in S_1$. Observe that $U^2 \subseteqq LIRJ \subseteqq LIL \subseteqq I$. Thus $J/I \approx J/U^2/I/U^2 \in S_1$. Hence $J/I \subseteqq S_1(L/I) = 0$ and the theorem is proved.

2. In this section we discuss necessity of the assumptions of the theorem. It is well-known [1, 7] that if S_1 or S_2 is not hereditary then usually the class $(S_1 : S_2)$ is not radical. Because of that in our discussion we consider hereditary radicals only.

Observe that for every radical S, $(S : S) = 1$. This shows that for given radicals S_1, S_2 the assumption that $(S_1 : S_2)$ is left hereditary says nothing about the properties of S_1 and S_2. However, for every radical S, $(S : 1) = S$. This shows that to get that for every left strong and hereditary radical S_2, $(S_1 : S_2)$ is left hereditary one has to assume that S_1 is left hereditary. The following proposition gives slightly more.

PROPOSITION 1. *If* $S_1 \subseteqq S_2$ *are hereditary radicals such that* S_2 *and* $(S_1 : S_2)$ *are left hereditary then* S_1 *is left hereditary.*

PROOF. Observe that $S_1 \subseteqq S_2$ implies $S_1 \subseteqq (S_1 : S_2)$. Let $L < R$ and $R \in S_1$. Since $(S_1 : S_2)$ is left hereditary and $S_1 \subseteqq (S_1 : S_2)$, $L \in (S_1 : S_2)$. Thus $S_2(L) \subseteqq S_1(L)$. However, since $S_1 \subseteqq S_2$ and S_2 is left hereditary, $L \in S_2$. Hence $L = S_2(L) \subseteqq S_1(L)$, which gives $L \in S_1$.

Observe that for every radical S, $(S:0)=1$. Hence one cannot omit in the proposition the assumption that $S_1 \subseteqq S_2$. Since for every radical S, $(S:S)=1$, the assumption that S_2 is left hereditary cannot be omitted as well.

The following proposition shows that one cannot omit in the theorem the assumption that $\beta \subseteqq S_1$.

PROPOSITION 2. *If S is a hereditary radical such that $(S:\beta)$ is left hereditary then $\beta \subseteqq S$.*

PROOF. Let $R=M_2(Q)$ be the ring of 2×2-matrices over rationals. Clearly, $R\in(S:\beta)$. Now if Z^0 is the zero ring on the additive group of the ring Z of integers then $Z^0 \approx \binom{00}{Z0} \triangleleft \binom{00}{Q0} \triangleleft \binom{Q0}{Q0} < R$. Hence, since $(S:\beta)$ is left hereditary, $Z^0\in(S:\beta)$. This gives $Z^0=\beta(Z^0)\subseteqq S(Z^0)$. Thus $Z^0\in S$ and consequently $\beta\subseteqq S$.

We have not been able to answer the question whether a given hereditary radical S_2 such that for every left hereditary radical S_1 containing β the radical $(S_1:S_2)$ is left hereditary, must be left strong. Although the general answer seems to be "not", the following two propositions show that some assumptions of such sort are necessary.

Let U be the Brown–McCoy radical. It is well-known that U is not left strong.

PROPOSITION 3. *The radical $(\beta:U)$ is not left hereditary.*

PROOF. Let R be a simple domain with 1 which is not a division ring. Obviously, $R\in(\beta:U)$. Let $a\neq 0$ be a non-invertable element of R. It is easy to check that Ra is a simple ring without 1. Hence $Ra\in U$ but $\beta(Ra)=0$, so $Ra\notin(\beta:U)$.

Let for a given ring R, $M(R)$ be the ring of all $N\times N$-matrices over R containing finitely many non-zero entries, where N is the set of positive integers.

LEMMA 4. $M(R)\approx M(M(R))$.

PROOF. Let $f\colon N\to N\times N$ be a bijection. Given a matrix $A\in M(R)$ define $\bar{f}(A)$ as the matrix of $M(M(R))$ which has at (k,m)-position the matrix of $M(R)$ with (l,n)-entry equal to (i,j)-entry of A, where $f(i)=$ $=(k,l)$, $f(j)=(m,n)$. The map $\bar{f}$ gives an isomorphism between $M(R)$ and $M(M(R))$.

LEMMA 5. *Let N be the nil radical. Then $N_1=\{R\mid M(R)\in N\}$ is a left hereditary radical contained in N.*

PROOF. Straightforward.

PROPOSITION 4. *The radical $(N_1:N)$ is left hereditary if and only if $(N_1:N)=1$, i.e. $N_1=N$.*

PROOF. If $N_1 = N$ then clearly $(N_1 : N) = 1$ and so $(N_1 : N)$ is left hereditary. Conversely, let R be any ring and R^1 the ring R with a unity adjoined. Given an ideal I of $M(R)$, let I^* be the ideal of $M(R^1)$ generated by I. By Andrunakievich lemma, $J = (I^*)^3 \subseteq I$. Since J is an ideal of $M(R^1)$, $J = M(K)$ for some ideal K of R. Now $N(M(R/K))$ and $N_1(M(R/K))$ are ideals of $M(R^1/K)$, so $N(M(R/K)) = M(K_1/K)$, $N_1(M(R/K)) = M(K_2/K)$ for some ideals K_1, K_2 of R. Now Lemma 4 gives $M(K_2/K) \in N_1$, so $K_1 = K_2$. Obviously, $N(M(R)/I) = M(K_1)/I$ and $N_1(M(R)/I) = M(K_2)/I$. Thus $N(M(R)/I) = N_1(M(R)/I)$ for every ideal I of $M(R)$. This shows that $M(R) \in (N_1 : N)$. Now let L be the set of all matrices of $M(R)$ whose entries outside the first column are equal to zero. Clearly L is a left ideal of $M(R)$ and R is a homomorphic image of L. Thus, since $(N_1 : N)$ is left hereditary, $R \in (N_1 : N)$. This proves that $(N_1 : N) = 1$.

REMARK. The same arguments show that if S is any radical containing β and $S_1 = \{R \mid M(R) \in S\}$ then $(S_1 : S)$ is left hereditary if and only if $(S_1 : S) = 1$.

Krempa [3] and Sands [6] proved that N is left strong (i.e. Koethe's problem has a positive solution) if and only if $N = N_1$. This and Proposition 4 give

COROLLARY. *The radical $(N_1 : N)$ is left hereditary if and only if the radical N is left strong.*

We close with the following

QUESTION. Is the radical $(\beta : N)$ left hereditary if and only if N is left strong?

ACKNOWLEDGEMENT. The author is grateful to Professor R. Wiegandt for fruitful discussions during preparation of this paper.

This research was supported by Hungarian National Foundation for Scientific Research Grant no. 1813.

REFERENCES

[1] DIVINSKY, N. and SULINSKI, A., Radical pairs, *Canad. J. Math.* **29** (1977), 1086–1091. *MR* **58** #16743

[2] JAEGERMANN, M. and SANDS, A. D., On normal radicals, N-radicals, and A-radicals, *J. Algebra* **50** (1978), 337–349. *MR* **58** #16745

[3] KREMPA, J., Logical connections between some open problems concerning nil rings, *Fund. Math.* **76** (1972), 121–130. *MR* **46** #5377

[4] OSŁOWSKI, B. J. and PUCZYŁOWSKI, E. R., On strong radical properties of alternative algebras, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **25** (1977), 845–850. *MR* **57** #6132

[5] PUCZYŁOWSKI, E. R., Hereditariness of strong and stable radicals, *Glasgow Math. J.* **23** (1982), 85–90. *MR* **83j**: 16015

[6] SANDS, A. D., Radicals and Morita context, *J. Algebra* **24** (1973), 335–345. *MR* **48** #6157

[7] Snider, R. L., Lattices of radicals, *Pacific J. Math.* **40** (1972), 207–220. *MR* **46** # 7290

[8] Wiegandt, R., *Radical and semisimple classes of rings,* Queen's Papers in Pure and Applied Mathematics, No. 37, Queen's University, Kingston, Ontario, 1974. *MR* **50** #2227

(Received November 28, 1988)

INSTYTUT MATEMATYKI
UNIWERSYTET WARSZAWSKI
BANACHA 2
PL–02–097 WARSZAWA
POLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 207–212

ON PRIME-ADDITIVE NUMBERS

P. ERDŐS and N. HEGYVÁRI

Throughout this paper $c_1, c_2, \dots$ will denote positive absolute constants, $A(x) = \sum_{a_i \leqq n} 1$; $B(x) = \sum_{b_i \leqq n} 1$; $\nu(n)$ denotes the number of the distinct prime factors of n. Let us observe that, e.g.

$$2 \times 3 \times 5 = 30 = 2 + 3 + 5^2$$
$$2^2 \times 3 \times 19 = 228 = 2^7 + 3^4 + 19$$
$$5 \times 7 \times 89 = 3115 = 5^4 + 7^4 + 89$$

and

$$3^2 \times 23 \times 919 = 190233 = 3^{11} + 23^3 + 919$$

(found by P. Massias).

We call a number n *strongly prime-additive* if

$$n = \sum_{p|n} p^{\alpha_p} \text{ with } \alpha_p > 0 \text{ and } p^{\alpha_p} < n \leqq p^{\alpha_p + 1}.$$

We conjecture that there are infinitely many strongly prime-additive numbers, but we could not prove this.

Assume that there are numbers $1 \leqq \beta_p \leqq \alpha_p$ so that $n = \sum_{p|n} p^{\beta_p}$ then we call n prime-additive. We could not prove that there are infinitely many prime-additive numbers. Also we suspect that the number of prime-additive numbers not exceeding x is for large x very much larger than the number of prime-additive numbers $\leqq x$, but we could not prove this (e.g. $42 = 2 \times 3 \times 7 = 7 + 3 + 2^5$ is prime-additive *but* not strongly prime-additive).

Perhaps the following remark is of some interest: $2 \times 3 \times 5$ is prime-additive, $2 \times 3 \times 5 \times 7 \times 11 = 2310 = 2^9 + 3 + 5^4 + 7^3 + 11^3$ is prime-additive, too. Perhaps the product of the first $2r + 1$ primes is always or at least infinitely often prime-additive.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 11P45; Secondary 11A41.

Key words and phrases. Other additive questions involving primes, primes, Hardy–Ramanujan theorem, de Bruijn's function, representation of a number by the powers of its prime divisors.

Akadémiai Kiadó, Budapest

If instead of all prime divisors, i.e. we require only

$$(*) \qquad n = \sum_{\text{some } p|n} p^{\alpha_p} \text{ with } \alpha_p > 0, \ \ \nu(n) > 1$$

then we can already show that there are infinitely many such values of n. Let us call these numbers (i.e. numbers satisfying $(*)$) *weakly prime-additive* (briefly w.p.a.).

THEOREM 1. *For every prime p there exists a w.p.a. number n for which $p \mid n$.*

Denote by $A = \{a_1, a_2, \dots\}$ the set of w.p.a. numbers. Then we shall prove the following

THEOREM 2. *Let c_1 and c_2 be sufficiently small. Then*

$$c_1(\log n)^3 < A(n) < \frac{n}{e^{(\log n)^{1/2-c_2}}}.$$

Proofs

PROOF of Theorem 1. Let $p > 2$ be an arbitrary prime number. We shall seek n in the form

$$(1.1) \qquad n = 2^\alpha + p^\beta + q^\gamma$$

where also q is a prime number.

The relation $2 \mid n$ is trivial. Let now s be a quadratic non-residue mod p. It is well-known that there exists a prime q which satisfies

$$(1.2) \qquad \begin{aligned} q &\equiv s (\operatorname{mod} p) \\ q &\equiv 1 (\operatorname{mod} 4). \end{aligned}$$

If $\alpha = r(p-1)$ and $\gamma = (2u+1)(p-1)/2$ then

$$2^\alpha + p^\beta + q^\gamma \equiv 1 + 0 + -1 \equiv 0 \, (\operatorname{mod} p)$$

so we have $p \mid n$.

Furthermore, by the law of reciprocity and because of (1.2) we obtain that

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) (-1)^{(p-1)(q-1)/4} = \left(\frac{q}{p}\right) = \left(\frac{s}{p}\right) = -1$$

i.e. p is a quadratic non-residue mod q.

Thus if $\alpha = t(p-1)$ and $\beta = (2v+1)(q-1)/2$ then we have

$$n = 2^\alpha + p^\beta + q^\gamma \equiv 1 + (-1) + 0 \equiv 0 \, (\operatorname{mod} q)$$

i.e. $q \mid n$.

Summing up: if $\alpha = z(p-1)(q-1)$, $\beta = (2v+1)(q-1)/2$, $\gamma = (2u+1)\times$ $\times(p-1)/2$ then $2pq \mid n$, i.e. n is a w.p.a. Q.E.D.

PROOF of Theorem 2.

LEMMA 1. *Let* $B_k = \{p_1^{\alpha_1} + p_2^{\alpha_2} + \ldots + p_k^{\alpha_k} : \alpha_i = td_i + r_i,\ i = 1, \ldots, k\}$. *Then*

$$B_k(x) > c_k(\log x)^k.$$

PROOF of Lemma 1. We shall apply mathematical induction. For $k = 1$ the statement is obvious. Now let $k > 1$ and let

$$B_{k-1}^{(\nu)} = \{p_2^{\alpha_2} + \ldots + p_k^{\alpha_k} : \alpha_i = td_i + r_i,\ \ i = 2, \ldots, k \text{ and}$$
$$p_2^{\alpha_2} + \ldots + p_k^{\alpha_k} < p_1^{\nu d_1 + r_1}(p_1^{d_1} - 1)\}.$$

By the inductive hypothesis we have

$$B_{k-1}^{(\nu)}(p_1^{\nu d_1 + r_1}(p_1^{d_1} - 1)) > c_{k-1}\{\log(p_1^{\nu d_1 + r_1}(p_1^{d_1} - 1))\}^{k-1} = c_k' \nu^{k-1}.$$

Because of $(p_1^{\nu d_1 + r_1} + B_{k-1}^{(\nu)}) \cap (p_1^{\mu d_1 + r_1} + B_{k-1}^{(\mu)}) = \emptyset$ for $\mu \neq \nu$ and

$$B_k \supset \bigcup_{\nu=1}^{\infty} (B_{k-1}^{(\nu)} + p_1^{\nu d_1 + r_1})$$

thus we have that

$$B_k(x) > \sum_{\nu=1}^{\log x/t} B_{k-1}^{(\nu)}(x) > c_k' \sum \nu^{k-1} = c_k(\log x)^k$$

(where $t = d_1 \log p_1$).

By Theorem 1 we have that $n = 2^{4k} + 3^{4v+2} + 5^{2u+1}$ is w.p.a. thus we established the *lower* bound by Lemma 1. Now we are going to show the *upper* bound. Let $a_1 < a_2 < \ldots < a_r \leqq x$ be the sequence of w.p.a. numbers up to x. Assume that

$$r \geqq x \,/\, e^{(\log x)^{1/2 - c_2}}. \tag{1.3}$$

We shall show that (1.3) leads to a contradiction. For this we need two lemmas.

LEMMA 2. *If* $c_2 > 0$ *is sufficiently small then for all but* $o\left(x / e^{(\log x)^{1/2 - c_2}}\right)$ *integers* $n \leqq x$ *we have*

$$\nu(n) < (\log x)^{1/2} / \log\log x.$$

It is known that

$$|\{\nu(n)=k,\ n \leqq x\}| \leqq c\frac{x}{\log x}\frac{(2\log\log x)^{k-1}}{(k-1)!}$$

(see [1]). Thus

$$\Omega(x) := \left|\{n\colon \nu(n) \geqq (\log x)^{1/2}/4\log\log x;\ n \leqq x\}\right| <$$

$$< c(x/\log x)\sum_{k>L}\frac{(2\log\log x)^k}{k!} < c(x/\log x)\sum_{k>L}(4\log\log x/k)^k$$

where $L=(\log x)^{1/2}/4\log\log x$, and because of $(4\log\log x/k)^k$ is a decreasing function for $k>L$ we have

$$\Omega(x) < c(x/\log x)(2\log\log x/L)^L\sum_{k>L}1 = o(x/e^{(\log x)^{1/2-c_2}}).$$

By Lemma 1 we can assume that there exists a subsequence $\{a'_1 < a'_2 < < \ldots < a'_s\}$ of $\{a_1 < \ldots < a_r\}$ for which

$$\nu(a'_i) < (\log x)^{1/2}/4\log\log x$$

and

$$s > 2x/3e^{(\log x)^{1/2-c_2}}.$$

LEMMA 3. *Let $\Psi(x,y)$ denote the number of integers not exceeding x whose all prime factors do not exceed y. Then*

$$\Psi(n, e^{(\log n)^{1/2}}) < n/2e^{(\log n)^{1/2-\varepsilon}}.$$

This is a consequence of the theorem of De Bruijn (see [2]). So we can select elements $\{\tilde{a}_1 < \ldots < \tilde{a}_z\}$ of $\{a'_1 < \ldots < a'_s\}$ for which

$$z > x/6e^{(\log x)^{1/2-c_2}}$$

and whose largest prime factors are bigger than $e^{(\log x)^{1/2}}$.

Denote the largest prime factor of $\tilde{a}_i$ by q_i and put $b_i = \tilde{a}_i/q_i$. Since $b_i \leqq x/e^{(\log x)^{1/2}}$, there exist at least

$$x/6e^{(\log x)^{1/2-c_2}} \Big/ x/e^{(\log x)^{1/2}} > e^{(\log x)^{1/2}/2}$$

b_{i_j} such that $b_{i_j} = b_{i_t}$ if $1 \leqq j \leqq t \leqq e^{(\log x)^{1/2}/2}$. Put

$$b_{i_j} = p_1^{\alpha_1}\ldots p_k^{\alpha_k} \quad j=1,\ldots,e^{(\log x)^{1/2}/2}.$$

The $\tilde{a}_{i_j}$ are w.p.a. numbers so

$$\tilde{a}_{i_j} = p_{t_1}^{\gamma_{t_1,j}} + \ldots + p_{t_r}^{\gamma_{t_r,j}} + \varepsilon_{i_j} q_i^{\beta_i}$$

where $\varepsilon_{i_j} = 0$ or 1 and $\{p_{t_1}, \ldots, p_{t_r}\} \subseteqq \{p_1, \ldots, p_k\}$. Here $k \leqq \leqq (\log x)^{1/2}/4 \log\log x$, hence the number of integers which can be written in the form

$$p_{t_1}^{\gamma_{t_1}} + \ldots + p_{t_r}^{\gamma_{t_r}}$$

is less than $(\log x)^{(\log x)^{1/2}/4 \log\log x} = e^{(\log x)^{1/2}/4}$. This implies that there exist at least

$$e^{(\log x)^{1/2}/2} / e^{(\log x)^{1/2}/4}$$

$\tilde{a}_{i_j}$ for which

$$\tilde{a}_{i_j} = p_{t_1}^{\gamma_1} + p_{t_2}^{\gamma_2} + \ldots + p_{t_r}^{\gamma_r} + \varepsilon_{i_j} q_j^{\beta_j}$$

$\varepsilon_{i_j} = 0$ or 1, where $p_{t_1}^{\gamma_1} + \ldots + p_{t_r}^{\gamma_r}$ is a fixed sum of prime powers. In this case $\tilde{a}_{i_j}$ is said to have property F.

The number of distinct values of the powers of q_j is at most $\log x$, so we obtain that there exist at least

$$e^{(\log x)^{1/2}/4} / \log x > e^{(\log x)^{1/2}/5}$$

$\tilde{a}_{i_j}$ with property F for which the largest prime factors are distinct. The $\tilde{a}_{i_j}$ are w.p.a. numbers, so $q_j \mid \tilde{a}_{i_j}$ and thus $q_j \mid p_{t_1}^{\gamma_1} + \ldots + p_{t_r}^{\gamma_r}$ must hold, as well. Thus we obtain that

$$x \geqq \tilde{a}_{i_j} \geqq p_{t_1}^{\gamma_1} + \ldots + p_{t_r}^{\gamma_r} \geqq \prod_{j=1}^{e^{(\log x)^{1/2}/5}} q_j \geqq \left(e^{(\log x)^{1/2}}\right)^{e^{(\log x)^{1/2}/5}} > x$$

if x is large enough. This contradiction proves the upper bound.

Added in proof to Theorem 2. The authors conjectured and very recently A. Balog and C. Pomerance proved that for every k there are $p_1, p_2, \ldots, p_k$ primes such that $p_1^{\alpha_1} + \ldots + p_k^{\alpha_k}$ is w.p.a. This argument and Lemma 1 show that $A(x) > (\log x)^k$ for every k. The proof goes as follows.

Fix $k > 2$. We show that for each $j \leq k$ there are distinct primes $p_1, \ldots, p_j$ and positive integers $\alpha_1, \ldots, \alpha_j$ with

$$(*) \qquad \sum_1^j p_i^{\alpha_i} \equiv j - k (\operatorname{mod} p_1 \ldots p_j).$$

We do this by induction on j, the assertion being the case $j = k$. Let p_1 be any prime factor of $k - 1$ and let $\alpha_1 > 0$ be arbitrary. Then $p_1^{\alpha_1} \equiv 1 - k (\operatorname{mod} p_1)$, so that we have $(*)$ for $j = 1$. Assume $(*)$ is true for j and $j < k$. Let

$m = \sum_1^j p_i^{\alpha_i}$. Then $m + k - j \equiv 0 (\bmod p_1 \dots p_j)$, so that if p_{j+1} is any prime factor of $m + k - j - 1$, then $p_{j+1} \notin \{p_1, \dots, p_j\}$. Let $\alpha_{j+1} = \varphi(p_1 \dots p_j)$. Then $p_{j+1}^{\alpha_{j+1}} \equiv 1 (\bmod p_1 \dots p_j)$, so that

$$S := m + p_{j+1}^{\alpha_{j+1}} + k - j - 1 \equiv m + k - j \equiv 0 (\bmod p_1 \dots p_j).$$

In addition, by the choice of p_{j+1}, we have $S \equiv 0 (\bmod p_{j+1})$. Thus

$$\sum_i^{j+1} p_i^{\alpha_i} \equiv j + 1 - k (\bmod p_1 \dots p_{j+1})$$

and we have $(*)$ for $j + 1$.

ACKNOWLEDGEMENT. The authors express their gratitude to R. Freud for his remarks concerning Theorem 1.

REFERENCES

[1] RAMANUJAN, S., *Collected papers*, Edited by G. H. Hardy [et al.], Cambridge Univ. Press, 1927, 262–275. *Jahrbuch Fortschritte Math.* **53**, 30

[2] BRUIJN, N. G. DE, On the number of positive integers $\leqq x$ and free of prime factors $> y$. II, *Nederl. Akad. Wetensch. Proc. Ser. A* **69** *Indag. Math.* **28** (1966), 239–247. *MR* **34** #5770

(Received December 22, 1988)

MTA MATEMATIKAI KUTATÓINTÉZETE
P.O. BOX 127
H–1364 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ÁLTALÁNOS ISKOLAI TANÁRKÉPZŐ FŐISKOLAI KAR
MARKÓ UTCA 29
H–1055 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 213–221

ON QUERMASSES OF SIMPLICES

H. MARTINI and B. WEISSBACH

Our considerations refer to Euclidean vector space R^d ($d \geqq 2$). We shall prove that the projection body and the difference set of a d-simplex S are polars with respect to a sphere of radius $\sqrt{d\,V(S)}$, where $V(S)$ denotes the volume of S. This is equivalent to the property that each Steiner symmetral of S has exactly two extreme points outside of the corresponding symmetrization space. As a consequence we obtain sharp bounds for outer $(d-1)$-quermasses of a d-simplex in terms of the volume and inner 1-quermasses of this polytope. By means of these bounds a characterization of regular d-simplices is observed.

1. Notation and definitions

Let R^d ($d \geqq 2$) denote the d-dimensional Euclidean vector space with scalar product $\langle \cdot, \cdot \rangle$ and unit sphere $S^{d-1} := \{u \in R^d \mid \langle u, u \rangle = 1\}$. We shall write K^d for the set of *convex bodies*, i.e. compact, convex subsets of R^d with interior points. For further notation and background material the reader should consult the books [1] and [6].

In particular, the area of the orthogonal image of $K \in K^d$ in the $(d-1)$-subspace $H = \{x \in R^d \mid \langle x, u \rangle = 0\}$ is called the *outer* $(d-1)$*-quermass* or *brightness* $\overline{V}_{d-1}(K,u)$ of K with respect to $u \in S^{d-1}$. Further on, the length of the longest chord of K in direction u is named the *inner* 1*-quermass* $\underline{V}_1(K,u)$ of K and u. The measure $\overline{V}_{d-1}(K,u)$ is the restriction to S^{d-1} of the *support function* $h(\Pi K, u)$ of a convex body ΠK, called the *projection body* of K, whereas $\underline{V}_1(K,u)$ is reciprocal to the correspondingly restricted *distance function* $g(DK,u)$ of the difference set $DK = K + (-1)K$. Thus we have the relations

$$\overline{V}_{d-1}(K,u) = h(\Pi K, u) = \max\{\langle x, u \rangle \mid x \in \Pi K\},$$

$$\frac{1}{\underline{V}_1(K,u)} = g(DK,u) = \min\{\varrho > 0 \mid u \in \varrho K\}, \quad u \in S^{d-1}.$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 52A20; Secondary 51N10.

Key words and phrases. Convex bodies, simplices, projection bodies, difference sets, Steiner symmetrization, outer and inner quermasses.

These two measures are invariant under Steiner symmetrization of $K \in K^d$ at H which we define in the following: Let G be a line orthogonal to H. If G meets $K \in K^d$, then let $S_H(G \cap K)$ be the line segment in G and of the same length as $G \cap K$, which is centred about $G \cap H$. The union of all such line segments $S_H(G \cap K)$ is the *Steiner symmetral* $S_H(K)$ of K and u, whereas the described transform is called the *Steiner symmetrization* of K at H.

Finally, $V(K)$ denotes the *volume* of K, which is also invariant under Steiner symmetrization.

2. Results

Our basic statement is the following

MAIN THEOREM. *For every direction $u \in S^{d-1}$ the quermasses $\overline{V}_{d-1}(S,u)$ and $\underline{V}_1(S,u)$ of an arbitrary d-simplex S $(d \geqq 2)$ satisfy*

$$\overline{V}_{d-1}(S,u) \cdot \underline{V}_1(S,u) = d\,V(S). \qquad \square \tag{1}$$

Let us denote by $h(K_i,x)$ and $g(K_i,x)$, $x \in R^d$, the support functions and the distance functions of $K_i \in K^d$ with $0 \in \operatorname{int} K_i$ $(i=1,2)$. The equality $h(K_1,x) = g(K_2,x)$ for every $x \in R^d$ implies that K_1 is the polar set regarding K_2 with respect to S^{d-1}. The converse statement is also true (cf. [6], 127–128). Moreover, with $0 \in \operatorname{int} Q_i$ the bodies $Q_i \in K^d$ $(i=1,2)$ are polars with respect to the sphere $\{y \in R^d \mid \|y\| = r\}$ if and only if for each $x \in R^d \setminus \{0\}$

$$\frac{h(Q_1,x)}{g(Q_2,x)} = r^2 \tag{2}$$

holds. It is further known that for this polarity of Q_1 and Q_2 the restriction to S^{d-1} of (2) is sufficient, i.e.

$$\frac{h(Q_1,u)}{g(Q_2,u)} = r^2, \quad u \in S^{d-1}. \tag{2*}$$

These remarks verify the equivalence of the main theorem and the following

COROLLARY. *The projection body ΠS and the difference set DS of a d-simplex S with volume $V(S)$ are polars with respect to a sphere of radius $\sqrt{d\,V(S)}$.* $\square$

To obtain bounds for $\overline{V}_{d-1}(S,u)$ we shall use also the *width* $\overline{V}_1(S,u)$ of S in direction u. This measure is called the *outer 1-quermass* of S and u, too.

Clearly, for a convex body K the extremal values of $\underline{V}_1(K,u)$ and $\overline{V}_1(K,u)$ coincide ([1],51). Hence, using (1) and the notations $\max \underline{V}_1(K,u) = \max \overline{V}_1(K,u) =: D$, $\min \underline{V}_1(K,u) = \min \overline{V}_1(K,u) =: \Delta$ for diameter and minimal width of the body $K \in K^d$, we get immediately

THEOREM 1. *For every d-simplex S $(d \geqq 2)$ the equalities*

$$\min \overline{V}_{d-1}(S,u) = \frac{d\,V(S)}{D}, \quad \max \overline{V}_{d-1}(K,u) = \frac{d\,V(S)}{\Delta} \tag{3}$$

hold, and these extremal values are attained if and only if the corresponding directions are normal to parallel supporting hyperplanes of S with maximal and minimal distances, respectively.

There exist various estimates for quantities of simplices where the cases of equality are characteristic properties of the regular representatives. By Theorem 1 we shall add such a characterization.

For an arbitrary d-simplex S we write r for the radius of the greatest inscribed ball and R for the radius of the smallest circumscribed ball. Then we have (cf. [2], 291–292)

$$\frac{(d+1)^{\frac{d+1}{2}}}{d!} d^{\frac{d}{2}} r^d \leqq V(S) \leqq \frac{(d+1)^{\frac{d+1}{2}}}{d!} d^{-\frac{d}{2}} R^d.$$

(From this we also obtain $dr \leqq R$.) Moreover, by theorems of Jung and Steinhagen (see [1], 77–79) the relations

$$D \geqq R \left(\frac{2d+2}{d}\right)^{\frac{1}{2}}, \quad \Delta \leqq r \frac{d^{\frac{1}{2}}(d+1)}{\left[\frac{d+1}{2}\right]^{\frac{1}{2}} \left[\frac{d+2}{2}\right]^{\frac{1}{2}}}$$

hold. In all cases equality holds if and only if S is regular. Hence, by (3) and simple transformations we get

THEOREM 2. *For an arbitrary d-simplex S $(d \geqq 2)$ we have*

$$\min \overline{V}_{d-1}(S,u) \leqq R^{d-1} \frac{(1+\frac{1}{d})^{\frac{d}{2}}}{(d-1)!} \left(\frac{d}{2}\right)^{\frac{1}{2}}, \tag{4}$$

$$\max \overline{V}_{d-1}(S,u) \geqq r^{d-1} \frac{(1+\frac{1}{d})^{\frac{d}{2}}}{(d-1)!} \cdot \frac{d^{d-\frac{1}{2}}}{(d+1)^{\frac{1}{2}}} \left[\frac{d+1}{2}\right]^{\frac{1}{2}} \left[\frac{d+2}{2}\right]^{\frac{1}{2}}, \tag{5}$$

$$\frac{\max \overline{V}_{d-1}(S,u)}{\min \overline{V}_{d-1}(S,u)} \geqq \frac{\left[\frac{d+1}{2}\right]^{\frac{1}{2}} \left[\frac{d+2}{2}\right]^{\frac{1}{2}}}{\left(\frac{d+1}{2}\right)^{\frac{1}{2}}}. \tag{6}$$

In each case equality holds if and only if S is regular. □

(It should be remarked that (5) was given already in [10].) Relations between outer $(d-1)$-quermasses, diameter, minimal width and volume are known for arbitrary convex bodies, too. They were observed by Firey (cf. [3]

and [4]). In the second paper it is shown that for a convex body K in every direction $u \in S^{d-1}$

$$(7) \qquad \overline{V}_{d-1}(K,u)\underline{V}_1(K,u) \leqq d\,V(K)$$

holds. To see this, one has to symmetrize the convex body K at a $(d-1)$-subspace H with normal direction u. Then $S_H(K)$ has the same volume as K and is intersected by H in a $(d-1)$-dimensional set of area $\overline{V}_{d-1}(K,u)$. Further $S_H(K)$ contains at least one line segment L, whose length is given by $\underline{V}_1(K,u)$. Since the convex hull of the $(d-1)$-dimensional intersection and L is contained in $S_H(K)$, (7) is observed, and equality holds if and only if this convex hull and $S_H(K)$ coincide. Clearly, in that case the only extreme points of $S_H(K)$ outside of H are the end-points of L. Therefore we can formulate a second statement which is equivalent to the main theorem.

THEOREM 3. *Each Steiner symmetral of an arbitrary d-simplex S $(d \geqq 2)$ has exactly two extreme points outside of the corresponding symmetrization space.* □

If $u \in S^{d-1}$ is assumed to be a direction of the diameter or of the maximal brightness of a convex body $K \in K^d$, then (7) implies the estimations

$$(8) \qquad d\,V(K) \geqq D \min \overline{V}_{d-1}(K,u), \quad d\,V(K) \geqq \Delta \max \overline{V}_{d-1}(K,u).$$

We want to correct the answer given by Firey to the question in which cases equality holds. Equality in (8) is not only obtained if K is degenerate, as follows by (3).

3. Proofs

The main theorem shall be verified by proving the equivalent corollary related to the sets ΠS and DS (H. Martini): It is trivial to show that these bodies have a finite set of extreme points, i.e. ΠS and DS are convex d-polytopes. Hence these sets are uniquely determined by their vertex sets and sets of support numbers, respectively. (Note that the support number in the outward normal direction of one $(d-1)$-face is defined to be the oriented distance of the corresponding facet hyperplane to the origin.)

Consequently, for the confirmation of the polarity we can restrict our attention to the sets of all facet hyperplanes of ΠS and of all vertices of DS. It is shown that all vertices of DS lie exactly on those lines through O which are parallel to edges of S. On the other hand these and only these lines are the facet normals of ΠS. (These proofs shall be given later.) Thus it is sufficient to show (1) only for all directions parallel to edges of S.

The direction of the joining edge regarding the vertices e_h, e_k of the simplex shall be denoted by u_{hk}. Let us consider the uniquely determined affinity α with e_h, e_k as fixed points, which transforms the remaining vertices of S into a hyperplane with normal direction u_{hk}, where the joining line segments of corresponding points with respect to α are (degenerate or) parallel to $e_h - e_k$. Obviously, this affinity is a shear with fixed lines of direction u_{hk}. Hence α preserves the measures $V(S)$, $\overline{V}_{d-1}(S, u_{hk})$, and $\underline{V}_1(S, u_{hk}) = \|e_h - e_k\|$. But for $\alpha(S)$

$$d\,V(\alpha(S)) = \overline{V}_{d-1}(\alpha(S), u_{hk})\underline{V}_1(\alpha(S), u_{hk})$$

holds by means of a classical formula.

Now the postponed proofs for the used statements on the boundary structure of the polytopes DS and ΠS shall be presented. The difference set of a convex body K is given by $DK = K + (-1)K$. Clearly, $x = x_1 + x_2$ is only an extreme point of the Minkowski sum $K' = K'_1 + K'_2$, if x_i is extremal in K'_i with $i = 1, 2$ (cf. [6], 90). Consequently vertices of the difference set of $S = \operatorname{conv}\{e_0, \dots e_d\}$ are at most the points $e_h - e_k$. On the contrary, every such point is a vertex of DS for $h \neq k$. We assume that $e_h - e_k$ $(h \neq k)$ is not a vertex of DS. Then there would exist two points $p_1 \neq p_2$ from DS with

$$e_h - e_k = \lambda_1 p_1 + \lambda_2 p_2; \quad \lambda_1 + \lambda_2 = 1; \quad \lambda_i > 0; \quad i = 1, 2.$$

Each point $p \in DS$ is representable by

$$p = \sum_{j=0}^{d} \mu_j e_j \quad \text{with} \quad \sum_{j=0}^{d} \mu_j = 0; \quad -1 \leqq \mu_j \leqq +1,$$

and $\mu_h = 1$, $\mu_k = -1$ would imply $p = e_h - e_k$. Suppose

$$p_i = \sum_{j=0}^{d} \mu_j^{(i)} e_j, \quad \sum_{j=0}^{d} \mu_j^{(i)} = 0, \quad -1 \leqq \mu_j^{(i)} \leqq 1.$$

By the assumption that $e_h - e_k$ lies between p_1 and p_2 we have $\sum\limits_{j=0}^{d} \tau_j e_j = 0$ and $\sum\limits_{j=0}^{d} \tau_j = 0$ with

$$\tau_j = \lambda_1 \mu_j^{(1)} + \lambda_2 \mu_j^{(2)}, \quad j \in \{0, \dots, d\} \setminus \{h, k\}$$
$$\tau_h = \lambda_1 \mu_h^{(1)} + \lambda_2 \mu_h^{(2)} - 1, \quad \tau_k = \lambda_1 \mu_k^{(1)} + \lambda_2 \mu_k^{(2)} + 1.$$

Since the points $e_0, \dots, e_d$ are affinely independent, the coefficients τ_j have to vanish in each case. By $\mu_h^{(i)} \leqq 1$ and $\lambda_i > 0$ (i.e. $\lambda_1 \mu_h^{(1)} \leqq \lambda_1$ and $\lambda_2 \mu_h^{(2)} \leqq \lambda_2$) the condition $\tau_h = 0$ (i.e. $\lambda_1 \mu_h^{(1)} + \lambda_2 \mu_h^{(2)} = 1$) can only be satisfied for $\mu_h^{(1)} = \mu_h^{(2)} = 1$. An analogous conclusion leads one from $\tau_k = 0$ to $\mu_k^{(1)} = \mu_k^{(2)} = -1$. But then the equality $p_1 = p_2 = e_h - e_k$ would be contradictory to the supposition.

The projection body ΠK of $K \in K^d$ can also be described in a simple manner, if K is a polytope. Let u_i ($\|u_i\| = 1$; $i = 0, \dots, n$) present the outward normal direction of the ith facet of K. Further $V_{d-1}(u_i)$ shall be the area of this $(d-1)$-face. Introducing $a_i := V_{d-1}(u_i)u_i$, we obtain

$$\sum_{i=0}^{n} a_i = 0 \quad \text{and} \quad \Pi K = \sum_{i=0}^{n} \operatorname{conv}\{0, a_i\}.$$

A supporting hyperplane of ΠK with $u \in S^{d-1}$ as its outward normal direction is determined by

$$\langle x, u \rangle = \frac{1}{2} \sum_{i=0}^{n} |\langle a_i, u \rangle| = \sum_{i \in J(u)} \langle a_i, u \rangle = \overline{V}_{d-1}(K, u)$$

with

$$J(u) = \{i \in \{0, \dots, n\} \mid \langle a_i, u \rangle > 0\}.$$

Only sets of the form

$$\operatorname{conv}\{0, a_{i_0}\} + \ldots + \operatorname{conv}\{0, a_{i_k}\} + \lambda_{i_{k+1}} a_{i_{k+1}} + \ldots + \lambda_{i_n} a_{i_n}$$

with $\lambda_{i_j} \in \{0; 1\}$ can be realized as $(d-r)$-dimensional faces ($r \in \{1, \dots, d-1\}$) of ΠK (see [8]). These faces are special translates of convex bodies

$$Z_{i_0, \dots, i_k} := \sum_{j=0}^{k} \operatorname{conv}\{0, a_{i_j}\}$$

with no more than $k+1$ dimensions.

For the generation of a facet of ΠK one has to assure $\dim Z_{i_0, \dots, i_k} = d-1$. Therefore only vector products of $d-1$ linearly independent points from $\{a_0, \dots, a_n\}$ can lead to outward facet normals of ΠK.

For a d-simplex we have $n = d$. As is well-known, each $(d-1)$-tuple from $\{a_0, \dots, a_d\}$ is linearly independent. Therefore always points $\overline{u}_{hk} \in S^{d-1}$ exist, which are uniquely determined up to reflection at the origin and satisfy $\langle a_i, \overline{u}_{hk} \rangle = 0$, $i \in \{0, \dots, d\} \setminus \{h, k\}$ with $0 \leqq k$, $k \leqq d$, and $h \neq k$.

Facets of ΠK can only lie orthogonally to the directions given by these points. On the other hand, this is valid in each case. The scalar products $\langle a_h, \overline{u}_{hk} \rangle$ and $\langle a_k, \overline{u}_{hk} \rangle$ cannot vanish, but we have $\langle a_h, \overline{u}_{hk} \rangle + \langle a_k, \overline{u}_{hk} \rangle = 0$. For example, if $\langle a_h, \overline{u}_{hk} \rangle > 0$ holds, then the supporting hyperplane of ΠK with $\overline{u}_{hk}$ as its outward normal direction is determined by $\langle x, \overline{u}_{hk} \rangle = \langle a_h, \overline{u}_{hk} \rangle$, $x \in R^d$. Clearly, a_h is a point of that hyperplane. But the points $a_h + a_i$, $i \in \{0, \dots, d\} \setminus \{h, k\}$, are also elements of this linear manifold, i.e. altogether d linearly independent points of the projection body. Consequently, the set of directions given by $\overline{u}_{hk}$ and the set of directions determined by edges of the simplex coincide.

One might prove Theorems 1 and 2 also without the main theorem, namely by means of statements in [7] or using equivalent assertions on the projection body of a simplex S (B. Weissbach):

The minimal outer $(d-1)$-quermass of S belongs to the set of distances between facet hyperplanes of ΠS and the origin. According to the preceding remarks these distances are determined by

$$d_{hk} := |\langle a_h, u_{hk}\rangle| = |\langle a_k, u_{hk}\rangle|,$$
$$u_{hk} = \frac{e_h - e_k}{\|e_h - e_k\|} \qquad (h \neq k).$$

Taking the centre of gravity of S as the origin, we obtain

$$-\frac{1}{d}e_h = \sum_{\substack{i=0\\ i\neq h}}^{d} \frac{1}{d}e_i.$$

Thus $-\frac{1}{d}e_h$ is a point from the convex hull of the points e_i with $i \in \{0,\dots,d\}\setminus\{h\}$. Further $-\frac{1}{d}e_h$ lies in that facet hyperplane of S which does not contain e_h. The direction determined by $a_h \neq 0$ is assumed to be normal to this facet hyperplane. Therefore $\langle a_h, e_i + \frac{1}{d}e_h\rangle = 0$, $i \in \{0,\dots,d\}\setminus\{h\}$, and we get

$$d_{hk} = \frac{1}{\|e_h - e_k\|}\frac{d+1}{d}|\langle a_h, e_h\rangle|, \quad h \neq k.$$

If e_h^* denotes the orthogonal projection of e_h onto the line through O and a_h, and if $b(u_h)$ represents the width of S in the direction of this line (i.e. the corresponding altitude of the simplex), then

$$\langle a_h, e_h\rangle = -\|a_h\|\|e_h^*\| = -V_{d-1}(u_h)\frac{d}{d+1}b(u_h) = -\frac{d^2}{d+1}V(S).$$

Accordingly we have

$$\min \overline{V}_{d-1}(S,u) = \min_{h\neq k} d_{hk} = \frac{1}{\max\|e_h - e_k\|}d\,V(S) = \frac{d\,V(S)}{D}.$$

The maximal outer $(d-1)$-quermass is contained in the set of distances between O and vertices of the projection body. Remember the statements on extremal points of a Minkowski sum of convex sets and on the representation of ΠS: Only points of the form

$$p(J) = \sum_{i\in J} a_i, \quad J \subset \{0,\dots,d\},$$

with $0 < \operatorname{card} J < d+1$ satisfy this condition. As an immediate consequence we observe

$$\max \overline{V}_{d-1}(S,u) = \max_J \Big\| \sum_{i\in J} a_i \Big\|.$$

Once more we assume the centre of gravity of S to be the origin. By a suitable numbering for every permissible i then the relations

$$\langle a_i, e_i\rangle = -\frac{d^2}{d+1}V(S), \quad \langle a_i, e_h\rangle = \frac{d}{d+1}V(S), \;\; h \neq i;$$

$$\langle a_i, e_h - e_k\rangle = 0, \quad i \neq h, k,$$

hold. We introduce $\overline{J} := \{0, \dots, d\} \setminus J$. By the supposition regarding card J every dissection $(J, \overline{J})$ generates a dissection of the vertex set of S, too. It is well-known that for an arbitrary dissection of vert S there exist two uniquely determined parallel hyperplanes $H(J)$ and $H(\overline{J})$ with the following property: $H(J)$ contains the vertices e_i $(i \in J)$ and $H(\overline{J})$ the remaining vertices. Hence these hyperplanes are given by

$$H(J) = \Big\{ x \;\Big|\; \Big\langle \sum_{i\in \overline{J}} a_i, \;\; x - e_h \Big\rangle = 0, \;\; h \in J \Big\},$$

$$H(\overline{J}) = \Big\{ x \;\Big|\; \Big\langle \sum_{i\in J} a_i, \;\; x - e_k \Big\rangle = 0, k \in \overline{J} \Big\}.$$

The distance of $H(J)$ and $H(\overline{J})$ shall be denoted by $q(J) = q(\overline{J})$. It is not difficult to show that the minimum of all such distances yields the minimal width of S. Therefore

$$q(J) = \frac{1}{\Big\|\sum\limits_{i\in J} a_i\Big\|} \Big| \Big\langle \sum_{i\in J} a_i, \;\; e_h - e_k \Big\rangle \Big| = \frac{1}{\Big\|\sum\limits_{i\in J} a_i\Big\|} \, |\langle a_h, e_h - e_k\rangle| = \frac{1}{\Big\|\sum\limits_{i\in J} a_i\Big\|} d\,V(S)$$

with $h \in J$ and $k \in \overline{J}$, and consequently we have

$$\max \overline{V}_{d-1}(S,u) = \frac{1}{\min q(J)} d\,V(S) = \frac{d\,V(S)}{\Delta}.$$

(For this convenient approach to the minimal width of a d-simplex we also refer to [11].)

4. Concluding remarks

At the Oberwolfach meeting on Convex Bodies in 1974, E. Heil posed the following problem (see also [5], no. 23): *Let K be the convex hull of d line segments in Euclidean d-space. Is the volume of K greater than or equal* *to the volume of a simplex S generated by translates of these line segments having a common endpoint?* In [9] McMullen gave an affirmative answer to this problem, moreover in a cleverly generalized form. It should be noticed that the affirmative answer to Heil's original problem can also be obtained as an immediate consequence of our main theorem. Using the measures $\overline{V}_{d-1}(\cdot,u)$, $\underline{V}_1(\cdot,u)$ of K and S in the direction u of one such generating line segment, we get $V(K) \geqq V(S)$ by a simple inductive argument regarding $\overline{V}_{d-1}(\cdot,u)$ and by the obvious inequality $\underline{V}_1(K,u) \geqq \underline{V}_1(S,u)$ (cf. (1), (7) and the assertions leading to Theorem 3).

Finally we remark that the first named author meanwhile could characterize d-simplices in K^d ($d \geqq 2$) by the properties given in the main theorem, its corollary and Theorem 3.

REFERENCES

[1] Bonnesen, T. and Fenchel, W., *Theorie der konvexen Körper*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Bd. 3, Springer, Berlin, 1934. *Zbl* **8**, 77

[2] Fejes Tóth, L., *Reguläre Figuren*, Akadémiai Kiadó, Budapest, 1965. *MR* **30** #3408

[3] Firey, W. J., Lower bounds for volumes of convex bodies, *Arch. Math.* **16** (1965), 69–74. *MR* **31** #5152

[4] Firey, W. J., Lower bounds for quermassintegrals of a convex body, *Portugal. Math.* **25** (1966), 141–146. *MR* **38** #607

[5] Gruber, P. M. and Schneider, R., Problems in geometric convexity, *Contributions to Geometry*, eds. J. Tölke and J. M. Wills (Proc. Geom. Sympos. Siegen, 1978), Birkhäuser, Basel, 1979, 255–278. *MR* **82d:** 52001

[6] Leichtweiss, K., *Konvexe Mengen*, Hochschulbücher für Mathematik, **81**, Deutscher Verlag der Wissenschaften, Berlin, 1980. *MR* **81b**: 52001

[7] Martini, H. and Weissbach, B., Zur besten Beleuchtung konvexer Polyeder, *Beiträge Algebra Geom.* **17** (1984), 151–168. *MR* **85j**: 52017

[8] McMullen, P., On zonotopes, *Trans. Amer. Math. Soc.* **159** (1971), 91–109. *MR* **43** #5410

[9] McMullen, P., The volume of certain convex sets, *Math. Proc. Cambridge Philos. Soc.* **91** (1982), 91–97. *MR* **83a**: 52008

[10] Weissbach, B., Eine Kennzeichnung der regulären Simplexe, *2nd Colloquium for geometry and combinatorics*, Part 1 and 2 (Karl-Marx-Stadt, 1983), Technische Hochschule Karl-Marx-Stadt, Karl-Marx-Stadt, 1983, 225–228. *MR* **87h**: 52023

[11] Weissbach, B., Schranken für die Dicke der Simplexe, *Beiträge Algebra Geom.* **26** (1988), 5–11.

(Received January 23, 1989)

SEKTION MATHEMATIK
PÄDAGOGISCHE HOCHSCHULE
WIGARDSTRASSE 17
D/O–8060 DRESDEN

SEKTION MATHEMATIK
TECHNISCHE UNIVERSITÄT
PSF 124
D/O–3010 MAGDEBURG
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 223–233

NONLOCAL AND STRONGLY NONLINEAR THIRD BOUNDARY VALUE PROBLEM

I. M. HASSAN

§1. Introduction

We shall consider the following third boundary value problem:

$$\sum_{|\alpha|\leqq 1}(-1)^{|\alpha|}\partial^{\alpha}[f_{\alpha}\circ(\mathrm{id},u,u')]+g\circ(\mathrm{id},u)=F\quad \text{in } \Omega \tag{1}$$

$$\partial_{\nu}^{*}(u)=h_1\circ(\mathrm{id},u)+h_2\circ(\mathrm{id},u\circ\Phi)\quad \text{on } \partial\Omega \tag{2}$$

where Ω is a possibly unbounded domain in $\mathbf{R}^n$,

$$\partial_{\nu}^{*}:=\sum_{|\alpha|=1}[f_{\alpha}\circ(\mathrm{id},u,u')]\nu_{\alpha},$$

ν_{α} denote the coordinates of the normal unit vector on $\partial\Omega$ ($|\alpha|=1$). Φ is a C^1-diffeomorphism in a neighbourhood of $\partial\Omega$ such that $S:=\Phi(\partial\Omega)\subset\overline{\Omega}$, $\partial\Omega$ is bounded, continuously differentiable.

It must be emphasized that in the terms $g\circ(\mathrm{id},u)$ and $h_1\circ(\mathrm{id},u)$ no growth restriction is imposed but it is supposed that g, h_1 satisfy the sign condition $g(x,\eta)\eta\geqq 0$, $h_1(x,\eta)\eta\leqq 0$. It will be proved the existence of weak solutions of (1), (2) by using arguments of [1] (see also [2]).

Weak solution of (1), (2) will be defined as follows. Assuming that u is a classical solution of (1), (2) by Gauss–Ostrogradskij theorem and by using an integral transformation we obtain

$$\begin{aligned}&\sum_{|\alpha|\leqq 1}\int_{\Omega}[f_{\alpha}\circ(\mathrm{id},u,u')]\partial^{\alpha}v-\int_{\Omega}[h_1\circ(\mathrm{id},u)]vd\sigma-\\ &-\int_{S}[h_2^{*}\circ(\mathrm{id},u)](v\circ\Phi^{-1})d\sigma+\int_{\Omega}g\circ(\mathrm{id},u)v=\int_{\Omega}Fv.\end{aligned}\tag{3}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 35J55.
Key words and phrases. Nonlinear third boundary value problem.

Thus weak solution of (1), (2) will be defined by (3).

Nonlocal linear boundary value problems have been considered e.g. in [4]–[7]. The importance of nonlocal boundary value problems in plasmaphysics has been emphasized in [8]. In [9], [10] nonlocal and nonlinear boundary value problems have been studied.

§2. Existence theorem

Denote by $W_p^1(\Omega)$ the Sobolev space of real valued functions u, whose distributional derivatives of order $\leqq 1$ belong to $L^p(\Omega)$ $(1 < p < \infty)$. The norm in $W_p^1(\Omega)$ is defined by

$$\|u\|_{W_p^1(\Omega)} = \Big\{ \sum_{|\alpha| \leqq 1} \int_\Omega |\partial^\alpha u|^p \Big\}^{1/p}.$$

The points $\xi \in \mathbf{R}^{n+1}$ will be written also in the form $\xi = (\eta, \zeta)$ where $\eta \in \mathbf{R}$ and $\zeta \in \mathbf{R}^n$.

Assume that

(a) f_α, h_1, h_2^* and g satisfy the Carathéodory condition, i.e. they are measurable in x for each ξ resp. η and continuous in ξ resp. η for a.e. $x \in \Omega$.

(b) There exist constants $c_1 > 0$, p $(1 < p < \infty)$, and a function $k_1 \in L^q(\Omega)$ where $\frac{1}{p} + \frac{1}{q} = 1$, such that

$$|f_\alpha(x, \xi)| \leqq c_1 |\xi|^{p-1} + k_1(x)$$

for all $|\alpha| \leqq 1$, $\xi \in \mathbf{R}^{n+1}$ and a.e. $x \in \Omega$.

(c) For all (η, ζ), $(\eta, \zeta') \in \mathbf{R}^{n+1}$ with $\zeta \neq \zeta'$ and a.e. x in Ω

$$\sum_{|\alpha|=1} \left[f_\alpha(x, \eta, \zeta) - f_\alpha(x, \eta, \zeta') \right] (\xi_\alpha - \xi'_\alpha) > 0.$$

(d) There exist a constant $c_2 > 0$ and a function $k_2 \in L^1(\Omega)$ such that for a.e. x in Ω and all $\xi \in \mathbf{R}^{n+1}$

$$\sum_{|\alpha| \leqq 1} f_\alpha(x, \xi) \xi_\alpha \geqq c_2 |\xi|^p - k_2(x).$$

(e) For each $s > 0$, there exists $g_s \in L^1(\Omega)$ such that for a.e. x in Ω

$$|g(x, \eta)| \leqq g_s(x) \text{ if } |\eta| \leqq s.$$

(f) For any $\eta \in \mathbf{R}$, a.e. x in Ω

$$g(x, \eta)\eta \geqq 0.$$

(g) For any fixed $s>0$, there exists $h_{1,s}\in L^1(\partial\Omega)$ such that for a.e. $x\in\partial\Omega$

$$|h_1(x,\eta)|\leqq h_{1,s}(x) \text{ if } |\eta|\leqq s.$$

(h) For any $\eta\in\mathbf{R}$, a.e. x in $\partial\Omega$ we have

$$h_1(x,\eta)\eta\leqq 0.$$

(i) There exist numbers $c>0$, ϱ with $0<\varrho<p-1$ and a function $k\in$ $\in L^{1+\frac{1}{\varrho}}(S)$ such that for any $\eta\in\mathbf{R}$, a.e. $x\in S$

$$|h_2^*(x,\eta)|\leqq c|\eta|^\varrho+k(x).$$

The main result of this paper is the following

THEOREM. *Assume that conditions* (a)–(i) *are fulfilled. Then for any* $F\in\left(W_p^1(\Omega)\right)^*$ *there exists* $u\in W_p^1(\Omega)$ *such that*

$$\begin{gathered}g\circ(\mathrm{id},u)\in L^1(\Omega),\ \ [g\circ(\mathrm{id},u)]u\in L^1(\Omega),\\ h_1\circ(\mathrm{id},u)\in L^1(\partial\Omega),\ \ [h_1\circ(\mathrm{id},u)]u\in L^1(\partial\Omega)\end{gathered}\tag{4}$$

and

$$\begin{gathered}\sum_{|\alpha|\leqq 1}\int_\Omega[f_\alpha\circ(\mathrm{id},u,u')]\partial^\alpha v-\int_S[h_2^*\circ(\mathrm{id},u)](v\circ\Phi^{-1})d\sigma-\\ -\int_{\partial\Omega}[h_1\circ(\mathrm{id},u)]vd\sigma+\int_\Omega g\circ(\mathrm{id},u)v=\langle F,v\rangle\end{gathered}\tag{5}$$

for all $v\in W_p^1(\Omega)$ *satisfying* $v\in L^\infty(\Omega)$, $v|_{\partial\Omega}\in L^\infty(\partial\Omega)$ *and for* $v=u$.

Suppose that the assumptions (a)–(i) are satisfied. For any $u,v\in W_p^1(\Omega)$ let

$$\begin{gathered}\langle T(u),v\rangle:=\sum_{|\alpha|\leqq 1}\int_{\partial\Omega}[f_\alpha\circ(\mathrm{id},u,u')]\,\partial^\alpha v-\\ -\int_S[h_2^*\circ(\mathrm{id},u)]\,(v\circ\Phi^{-1})d\sigma\end{gathered}\tag{6}$$

and for any number $\mu>0$ let

$$g_\mu(x,\eta):=\begin{cases}g(x,\eta) & \text{if } |g(x,\eta)|\leqq\mu,\ \ |x|\leqq\mu\\ \mu\frac{g(x,\eta)}{|g(x,\eta)|} & \text{if } |g(x,\eta)|>\mu,\ \ |x|\leqq\mu\\ 0 & \text{if } |x|>\mu\end{cases}\tag{7}$$

and

$$h_{1,\mu}(x,\eta) := \begin{cases} h_1(x,\eta) & \text{if } |h_1(x,\eta)| \leqq \mu, \quad x \in \partial\Omega \\ \mu \frac{h_1(x,\eta)}{|h_1(x,\eta)|} & \text{if } |h_1(x,\eta)| > \mu, \quad x \in \partial\Omega. \end{cases}$$

Define operator S_μ by

$$\langle S_\mu(u), v\rangle := \int_\Omega g_\mu \circ (\mathrm{id}, u)v - \int_{\partial\Omega} h_1 \circ (\mathrm{id}, u) v d\sigma.$$

Firstly we shall prove several lemmas (similar lemmas are proved in [1] and [2]).

LEMMA 1. *$T + S_\mu$ is pseudomonotone operator.*

PROOF. By (a), (b), (i), (6) and (7) $T + S_\mu$ is bounded. Suppose that (u_j) converges in $W_p^1(\Omega)$ weakly to u, $((T + S_\mu)(u_j))$ is convergent weakly in $\left(W_p^1(\Omega)\right)^*$ to y and

(9) $$\lim_{j\to\infty} \sup \langle (T + S_\mu)(u_j), u_j - u\rangle \leqq 0.$$

Then there is a subsequence (u_j') of (u_j) such that

$$\lim_{j\to\infty} (u_j') = u \quad \text{a.e. in } \Omega.$$

Thus by Lebesgue's dominated convergence theorem

(10) $$\begin{aligned} &\lim \|g_\mu \circ (\mathrm{id}, u_j') - g_\mu \circ (\mathrm{id}, u)\|_{L^q(\Omega)} = 0, \\ &\lim \|h_{1,\mu} \circ (\mathrm{id}, u_j') - h_{1,\mu} \circ (\mathrm{id}, u)\|_{L^q(\partial\Omega)} = 0, \end{aligned}$$

where q is defined by $\frac{1}{p} + \frac{1}{q} = 1$, whence

$$\lim_{j\to\infty} S_\mu(u_j') = S_\mu(u)$$

weakly in $\left(W_p^1(\Omega)\right)^*$ and so

$$\lim_{j\to\infty} T(u_j') = y - S_\mu(u)$$

weakly in $\left(W_p^1(\Omega)\right)^*$. From equality

$$\langle S_\mu(u_j'), u_j' - u\rangle = \langle S_\mu(u_j') - S_\mu(u), u_j' - u\rangle + \langle S_\mu(u), u_j' - u\rangle$$

it follows that

(12) $$\lim_{j\to\infty} \langle S_\mu(u_j'), u_j' - u\rangle = 0,$$

because by (10), the boundedness of $\|u_j' - u\|_{W_p^1(\Omega)}$ and Hölder's inequality

$$\lim_{j\to\infty} \langle S_\mu(u_j') - S_\mu(u), u_j' - u\rangle = 0.$$

Therefore (9) implies that

$$\limsup_{j\to\infty} \langle T(u_j'), u_j' - u\rangle \leqq 0. \tag{13}$$

Since T is pseudomonotone (see [4]) by (11) and (13) we have

$$T(u) = y - S_\mu(u),$$

i.e.

$$(T + S_\mu)(u) = y.$$

Further,

$$\lim_{j\to\infty} \langle T(u_j'), u_j' - u\rangle = 0$$

and so by (12)

$$\lim_{j\to\infty} \langle (T + S_\mu)(u_j'), u_j' - u\rangle = 0. \tag{14}$$

(14) is valid also for the sequence (u_j) (because else by the above arguments we get to a contradiction) and so the proof is complete.

LEMMA 2. *Assume that (u_j) converges weakly in $W_p^1(\Omega)$ to u and there is a constant c such that*

$$\int_\Omega [g_j \circ (\mathrm{id}, u_j)]u_j - \int_{\partial\Omega} [h_{1,j} \circ (\mathrm{id}, u_j)]u_j \leqq c. \tag{15}$$

Then

$$\begin{aligned} g \circ (\mathrm{id}, u) &\in L^1(\Omega), \quad [g \circ (\mathrm{id}, u)]u \in L^1(\Omega),\\ h_1 \circ (\mathrm{id}, u) &\in L^1(\partial\Omega), \quad [h_1 \circ (\mathrm{id}, u)]u \in L^1(\partial\Omega), \end{aligned}$$

and there exists a subsequence (u_{j_k}) of (u_j) such that

$$\lim_{k\to\infty} u_{j_k} = u \quad \text{a.e. in } \Omega \quad \text{and on } \partial\Omega, \tag{16}$$

$$\begin{aligned} \lim_{k\to\infty} \|g_{j_k} \circ (\mathrm{id}, u_{j_k}) - g \circ (\mathrm{id}, u)\|_{L^1(\Omega)} &= 0,\\ \lim_{k\to\infty} \|h_{1,j_k} \circ (\mathrm{id}, u_{j_k}) - h_1 \circ (\mathrm{id}, u)\|_{L^1(\partial\Omega)} &= 0. \end{aligned} \tag{17}$$

PROOF. As (u_j) tends to u weakly in $W_p^1(\Omega)$ thus there exists a subsequence (u_{j_k}) of (u_j) with the property (16). Since for the trace of u_j $\|u_j\|_{L^1(\partial\Omega)}$ is also bounded thus it may be assumed that

$$\lim_{k\to\infty} u_{j_k} = u \quad \text{a.e. on } \partial\Omega.$$

Thus, by (a) it is easy to show that

$$\lim_{k\to\infty} g_{j_k}(x, u_{j_k}(x)) = g(x, u(x)) \quad \text{for a.e. } x \in \Omega \tag{18}$$

and

$$\lim_{k\to\infty} h_{1,j_k}(x, u(x)) = h_1(x, u(x)) \quad \text{for a.e. } x \in \partial\Omega.$$

By (7), (15) and the assumption (f), (h) we have

$$\int_\Omega [g_j \circ (\mathrm{id}, u_j)]u_j \leqq c,$$

$$-\int_{\partial\Omega} [h_{1,j} \circ (\mathrm{id}, u_j)]u_j d\sigma \leqq c.$$

Therefore by (18), (f), (h) Fatou's lemma implies that

$$[g \circ (\mathrm{id}, u)]u \in L^1(\Omega), \quad [h_1 \circ (\mathrm{id}, u)]u \in L^1(\partial\Omega).$$

Thus we have proved the first part of Lemma 2.

Now we shall prove the second part of Lemma 2. For any $\delta > 0$ we have by (e)

$$|g_{j_k}(x, u_{j_k}(x))| \leqq g_{\delta^{-1}}(x) + \delta\, |g_{j_k}(x, u_{j_k}(x))u_{j_k}(x)|\,.$$

This implies that $g_{j_k}(x, u_{j_k}(x))$ is equiintegrable. Because, for any measurable set E in Ω

$$\int_E |g_{j_k}(x, u_{j_k}(x))|\, dx \leqq \int_E g_{\delta^{-1}}(x)dx + \delta c.$$

Given $\varepsilon > 0$, let $\delta = \frac{\varepsilon}{2c}$. Then for sufficiently small measure of E

$$\int_E |g_{j_k}(x, u_{j_k}(x))|\, dx < \varepsilon$$

and there is a set A_ε of finite measure with

$$\int_{\Omega\setminus A_\varepsilon} |g_{j_k}(x, u_{j_k}(x))|\, dx < \varepsilon.$$

By Vitali's theorem and (18) this shows that

$$g_{j_k} \circ (\mathrm{id}, u_{j_k}) \to g \circ (\mathrm{id}, u) \quad \text{in} \quad L^1(\Omega).$$

Similarly can be proved that

$$h_{1,j_k} \circ (\mathrm{id}, u_{j_k}) \to h_1 \circ (\mathrm{id}, u) \quad \text{in} \quad L^1(\partial\Omega).$$

LEMMA 3. *The operator $T + S_\mu$ is coercive, i.e.*

$$\lim_{\|u\| \to \infty} \left\langle \frac{\langle (T+S_\mu)(u), u\rangle}{\|u\|} \right\rangle = +\infty.$$

PROOF. From (f) and (h) we obtain

$$\int_\Omega g_\mu(x,u)u \geqq 0, \qquad \int_{\partial\Omega} h_{1,\mu}(x,u)u d\sigma \leqq 0.$$

This implies that $\langle S_\mu(u), u\rangle \geqq 0$. Thus by using conditions (d) and (i) we obtain

$$\begin{aligned} &\frac{\langle (T+S_\mu)u, u\rangle}{\|u\|} = \frac{\langle T(u), u\rangle + \langle S_\mu(u), u\rangle}{\|u\|} \geqq \\ &\geqq \frac{\langle T(u), u\rangle}{\|u\|} \geqq c_2 \|u\|^p_{W^1_p(\Omega)} - c_3 - c_4 \|u\|^{\varrho+1}_{W^1_p(\Omega)} - c_5 \|u\|_{W^1_p(\Omega)} \end{aligned} \tag{19}$$

($c_3 - c_5$ are constants). From this inequality and $\varrho + 1 < p$ it follows that $T + S_\mu$ is coercive.

The proof of the Theorem. By Lemmas 1 and 3 the operator $T + S_j$ is bounded, pseudomonotone and coercive for all $j = 1, 2, 3, \ldots$. By using the well-known theory of pseudomonotone operators in reflexive Banach space we obtain that for any F in $\left(W^1_p(\Omega)\right)^*$ there exists u_j in $W^1_p(\Omega)$ such that

$$(T + S_j)(u_j) = F. \tag{20}$$

By (19) (where the constants are independent of μ) the sequence (u_j) is bounded in $W^1_p(\Omega)$. T is a bounded operator and so the sequence $(T(u_j))$ is bounded in $\left(W^1_p(\Omega)\right)^*$. Since $W^1_p(\Omega)$ is a reflexive Banach space, there exists a subsequence (u_{j_k}) of (u_j) such that

$$\begin{aligned} &\lim_{k\to\infty} (u_{j_k}) = u \quad \text{weakly in} \quad W^1_p(\Omega) \\ &\lim_{k\to\infty} T(u_{j_k}) = y \quad \text{weakly in} \quad \left(W^1_p(\Omega)\right)^*. \end{aligned} \tag{21}$$

By definition of S_j and (21) we find

$$\int_{\Omega} [g_{j_k} \circ (\mathrm{id}, u_{j_k})]\, u_{j_k} - \int_{\partial\Omega} [h_{1,j_k} \circ (\mathrm{id}, u_{j_k})]\, u_{j_k} d\sigma =$$
$$= \langle S_{j_k}(u_{j_k}), u_{j_k} \rangle = \langle F, u_{j_k} \rangle - \langle T(u_{j_k}), u_{j_k} \rangle \leqq$$
$$\leqq \|F\|_{(W_p^1(\Omega))^*} \|u_{j_k}\|_{W_p^1(\Omega)} + \|T(u_{j_k})\|_{(W_p^1(\Omega))^*} \|u_{j_k}\|_{W_p^1(\Omega)} < c.$$

Thus by Lemma 2

$$[g \circ (\mathrm{id}, u)]u \in L^1(\Omega), \quad g \circ (\mathrm{id}, u) \in L^1(\Omega),$$
$$[h_1 \circ (\mathrm{id}, u)]u \in L^1(\partial\Omega), \quad h_1 \circ (\mathrm{id}, u) \in L^1(\partial\Omega)$$

and there is a subsequence $\left(u'_{j_k}\right)$ of (u_{j_k}) such that

$$\text{(22)} \qquad \begin{aligned} &\lim_{k\to\infty} u'_{j_k} = u \ \text{ a.e. in } \Omega, \\ &\lim_{k\to\infty} u'_{j_k} = u \ \text{ a.e. on } \partial\Omega \end{aligned}$$

and also

$$\text{(23)} \qquad \begin{aligned} &\lim_{k\to\infty} \left\| g_{j'_k} \circ \left(\mathrm{id}, u'_{j_k}\right) - g \circ (\mathrm{id}, u) \right\|_{L^1(\Omega)} = 0, \\ &\lim_{k\to\infty} \left\| h_{1,j'_k} \circ \left(\mathrm{id}, u'_{j_k}\right) - h_1 \circ (\mathrm{id}, u) \right\|_{L^1(\partial\Omega)} = 0. \end{aligned}$$

From (20) it follows that for any v in $W_p^1(\Omega)$ with $v \in L^\infty(\Omega)$ and $v|_{\partial\Omega} \in$ $\in L^\infty(\partial\Omega)$

$$\text{(24)} \qquad \left\langle \left(T + S'_{j_k}\right) \left(u'_{j_k}\right), v \right\rangle = \langle F, v \rangle .$$

By using (21) and (23), as $k \to \infty$ we find

$$\text{(25)} \qquad \langle y, v \rangle + \int_{\Omega} [g \circ (\mathrm{id}, u)]v - \int_{\partial\Omega} [h_1 \circ (\mathrm{id}, u)]\, v = \langle F, v \rangle.$$

Now we shall show that $y = T(u)$. Since T is pseudomonotone, it is sufficient to prove the inequality

$$\lim_{k\to\infty} \sup \left\langle T\left(u'_{j_k}\right), u'_{j_k} - u \right\rangle \leqq 0.$$

We have

$$\left\langle T\left(u'_{j_k}\right), u'_{j_k} - u \right\rangle = \left\langle T\left(u'_{j_k}\right), u'_{j_k} \right\rangle - \left\langle T\left(u'_{j_k}\right), u \right\rangle$$

and so by (20), (21) and Fatou's lemma

$$\limsup_{k\to\infty} \langle T(u'_{j_k}), u'_{j_k} - u\rangle = \limsup_{k\to\infty} \left\langle F - S_{j'_k}(u'_{j_k}), u'_{j_k}\right\rangle - \langle y, u\rangle \leqq$$
$$\leqq \langle F - y, u\rangle - \liminf\left\{\int_\Omega \left[g_{j'_k}\circ(\mathrm{id}, u'_{j_k})\right] u'_{j_k} - \int_{\partial\Omega}\left[h'_{1,j_k}\circ(\mathrm{id}, u'_{j_k})\, u'_{j_k}\right] d\sigma\right\} \leqq$$
$$\leqq \langle F - y, u\rangle - \int_\Omega [g\circ\langle(\mathrm{id}, u)]u + \int_{\partial\Omega}[h_1\circ(\mathrm{id}, u)]u d\sigma.$$

Thus, for any $w \in W^1_p(\Omega)\cap L^\infty(\Omega)$, by using (25)

$$\limsup_{k\to\infty} \langle T(u'_{j_k}), u'_{j_k} - u\rangle \leqq \langle F - y, u - w\rangle +$$
$$(26)\qquad + \int_\Omega [g\circ(\mathrm{id}, u)](w-u) - \int_{\partial\Omega}[h_1\circ(\mathrm{id}, u)](w-u)d\sigma.$$

Since $\partial\Omega$ is continuously differentiable thus $u \in W^1_p(\Omega)$ can be extended to $\mathbf{R}^n$ such that we obtain $u \in W^1_p(\mathbf{R}^n)$. We know (see [1]): there is a sequence (w_j) in $W^1_p(\mathbf{R}^n)\cap L^\infty(\mathbf{R}^n)$ such that (w_j) converges to u in $W^1_p(\mathbf{R}^n)$ and a.e. in $\mathbf{R}^n$; further,

$$(27)\qquad |w_j(x)| \leqq |u(x)| \quad \text{a.e. in } \mathbf{R}^n.$$

Now we show that for the trace of w_j and u we have

$$(28)\qquad |w_j|_{\partial\Omega}(x)| \leqq |u|_{\partial\Omega}(x)| \quad \text{for a.e. } x \in \partial\Omega.$$

We have in a.e. $y \in \mathbf{R}^n$

$$-|u(y)| \leqq w_j(y) \leqq |u(y)|.$$

Thus for any $\eta_\varepsilon \in C_0^\infty(\mathbf{R}^n)$ with the properties

$$\operatorname{supp}\eta_\varepsilon \subset B_\varepsilon = \{x \in \mathbf{R}^n : |x| \leqq \varepsilon\}, \quad \eta_\varepsilon \geqq 0 \quad \text{and} \quad \int \eta_\varepsilon = 1,$$

we have

$$-\int_{\mathbf{R}^n} |u(y)|\eta_\varepsilon(x-y)dy \leqq \int_{\mathbf{R}^n} w_j(y)\eta_\varepsilon(x-y)dy \leqq \int_{\mathbf{R}^n} |u(y)|\eta_\varepsilon(x-y)dy,$$

and so by using notation $v_\varepsilon(x) := \int_{\mathbf{R}^n} v(y)\eta_\varepsilon(x-y)dy$

$$(29)\qquad -|u|_\varepsilon|_{\partial\Omega} \leqq w_{j,\varepsilon}|_{\partial\Omega} \leqq |u|_\varepsilon|_{\partial\Omega}.$$

Since $(w_{j,\varepsilon}) \to w_j$ and $(|u|_\varepsilon) \to |u|$ in $W_p^1(\mathbf{R}^n)$ as $\varepsilon \to +0$ $\left(|u| \in W_p^1(\mathbf{R}^n)\right)$ thus

$$w_{j,\varepsilon}|_{\partial\Omega} \to w_j|_{\partial\Omega}$$

and

$$|u|_\varepsilon|_{\partial\Omega} \to |u||_{\partial\Omega} \text{ in } L^1(\partial\Omega)$$

as $\varepsilon \to +0$. Consequently, for a suitable sequence (ε_k) with $\lim_{k\to\infty}(\varepsilon_k) = 0$ we have

$$w_{j,\varepsilon_k}\Big|_{\partial\Omega} \to w_j|_{\partial\Omega}, \quad |u||_{\partial\Omega} \to |u||_{\partial\Omega}$$

a.e. on $\partial\Omega$ as $k \to \infty$. Therefore, from (29) we obtain

$$-|u||_{\partial\Omega}(x) \leqq w_j|_{\partial\Omega}(x) \leqq |u||_{\partial\Omega}(x)$$

which proves (28).

Now we have

$$\langle F - y, u - w_j \rangle \to 0$$

and

$$\int_\Omega [g \circ (\mathrm{id}, u)] w_j \to \int_\Omega [g \circ (\mathrm{id}, u)] u,$$

$$\int_{\partial\Omega} [h_1 \circ (\mathrm{id}, u)] w_j d\sigma \to \int_{\partial\Omega} [h_1 \circ (\mathrm{id}, u)] u d\sigma$$

by (27), (28) and Lebesgue's dominated convergence theorem, since

$$[g \circ (\mathrm{id}, u)] \in L^1(\Omega), \quad [h_1 \circ (\mathrm{id}, u)] u \in L^1(\partial\Omega).$$

Thus from (26) it follows that

$$\limsup_{k\to\infty} \left\langle T\left(u'_{j_k}\right), u'_{j_k} - u \right\rangle \leqq 0.$$

Consequently, $y = T(u)$, and $\langle T(u_{j_k}), u_{j_k} - u \rangle \to 0$. Therefore, from (25) we obtain (5) for all $v \in W_p^1(\Omega)$ with $v \in L^\infty(\Omega)$, $v \in L^\infty(\partial\Omega)$. Setting $v = w_j$ in (25) we find that (5) is true also for $v = u$. The proof of the existence theorem is complete.

REFERENCES

[1] WEBB, J. R. L., Boundary value problem for strongly nonlinear elliptic equations, *J. London Math. Soc.* (2) **21** (1980), 123–132. *MR* **82e**: 35039

[2] SIMON, L., On strongly nonlinear elliptic equations in unbounded domains, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **28** (1985), 241–252. *MR* **87m**: 35099

[3] BROWDER, F. E., Non-local elliptic boundary value problems, *Amer. J. Math.* **86** (1964), 735–750. *MR* **30** #2215

[4] Browder, F. E., Pseudo-monotone operators and nonlinear elliptic boundary value problems on unbounded domains, *Proc. Nat. Acad. Sci. U.S.A.* **74** (1977), 2659–2661. *MR* **56** #3469

[5] Carleman, T., Sur la théorie des équations intégrales et ses applications, *Verh. internat. Math.-Kongr.*, Zürich, 1932, Bd. 1, 138–151. *Zbl* **6**, 400

[6] Бицадзе, А. В. и Самарский, А. А., О некоторых простейших обобщениях линейных эллиптических краевых задач, *Dokl. Akad. Nauk SSSR* **185** (1969), 732–740. *MR* **40** #540

[7] Скубачевский, А. Л., Эллиптические задачи с нелокальными условиями вблизи границы, *Mat. Sb. (N.S.)* **129** (1986), 279–302. *MR* **87h**: 35089

[8] Самарский, А. А., О некоторых проблемах теории дифференциальных уравнений (Some problems of the theory of differential equations), *Differentsial'nye Uravneniya* **16** (1980), 1925–1935. *MR* **82d**: 35003

[9] Simon, L., Nonlinear elliptic differential equations with nonlocal boundary conditions, *Acta Math. Hungar.* **56** (1990), 343–352.

[10] Simon, L., Strongly nonlinear elliptic variational inequalities with nonlocal boundary conditions, *Differential Equations: Qualitative Theory,* Colloquia Math. Soc. J. Bolyai, Vol. 53, North-Holland, Amsterdam, 1990, 605–620.

(Received January 30, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 235–240

A NOTE ON J_2-RADICALS OF Γ-NEAR-RINGS

G. L. BOOTH

Abstract

The J_2-radical for zerosymmetric Γ-near-rings and the $J_{2(0)}$-radical for arbitrary Γ-near-rings have been defined by the author. In this paper, we characterise the left operator near-ring of a Γ-subnear-ring A of a Γ-near-ring M. If M is zero-symmetric and has a strong left unity, then for any invariant subgroup A of M, it holds that $J_2(A) = A \cap J_2(M)$. A similar equality holds for $J_{2(0)}$, where M is an arbitrary Γ-near-ring with a strong left unity.

1. Preliminaries

A Γ-*near-ring* is a triple, $(M, +, \Gamma)$, where

(i) $(M, +)$ is a (not necessarily abelian) group;

(ii) Γ is a nonempty set of binary operators on M, such that for each $\gamma \in \Gamma$, $(M, +, \gamma)$ is a (right distributive) near-ring;

(iii) For all $x, y, z \in M$, $\gamma, \mu \in \Gamma$, $x\gamma(y\mu z) = (x\gamma y)\mu z$.

If, in addition $x\gamma 0 = 0$ for all $x \in M$, $\gamma \in \Gamma$, then M is said to be *zerosymmetric*. Let A be a normal subgroup of the additive group of M. If for all $x, y \in M$, $\gamma \in \Gamma$ and $a \in A$, it holds that $x\gamma(a+y) - x\gamma y \in A$ and $a\gamma x \in A$, then A is called an *ideal* of M, denoted $A \triangleleft M$. The same notation will be used for ideals of near-rings.

A subgroup B of M for which $x\gamma a \in B$ and $a\gamma x \in B$ for all $a \in A$, $\gamma \in \Gamma$ and $x \in M$, is called an *invariant subgroup* of M. Note that all ideals of a zerosymmetric Γ-near-ring M are invariant. A subgroup of M which is itself a Γ-near-ring is called a Γ-subnear-ring of M.

If $A \triangleleft M$, then the factor group M/A is a Γ-near-ring with the operation $(x+A)\gamma(y+A) = x\gamma y + A$. If M and M' are Γ-near-rings (for the same Γ), and $f\colon M \to M'$ is a group homomorphism such that $f(x\gamma y) = f(x)\gamma f(y)$ for all $x, y \in M$, $\gamma \in \Gamma$, then f is called a Γ-*near-ring homomorphism*.

Let $x \in M$, $\gamma \in \Gamma$. If $y \in M$, we define $[x,\gamma]y = x\gamma y$. Let $\mathcal{L}$ be the near-ring of all mappings of M into itself, with pointwise addition and composition of mappings. The subnear-ring L of $\mathcal{L}$ generated by the set $L_0 = \{[x,\gamma]\colon x \in \in M, \gamma \in \Gamma\}$, is called the *left operator near-ring* of M. L_0 is called the

---

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 16A76; Secondary 16A78, 16A20, 16A21.

Key words and phrases. Γ-near-ring, left operator near-ring, J_2-radical, strong left unity.

generating set of L.

An L_0-*word* is an algebraically meaningful expression made up of symbols from the set $L_0 \cup \{(\,,), +, -\}$. We will use notation of the form $F(\gamma_1, \dots, \gamma_n)$ to denote an L_0-word which contains the elements $\gamma_1, \dots, \gamma_n$ of L_0 and no others. L_0 words may be added and multiplied in the obvious way. If $F(\gamma_1, \dots, \gamma_n)$ is an L_0-word, $n > 1$, then there exist L_0-words G and H such that either

$$F(\lambda_1, \dots, \lambda_n) = G(\mu_1, \dots, \mu_p) + H(\nu_1, \dots, \nu_q)$$

or

$$F(\lambda_1, \dots, \lambda_n) = G(\mu_1, \dots, \mu_p) + H(\nu_1, \dots, \nu_q),$$

where $p, q < n$ and $\{\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_q\} \subseteqq \{\lambda_1, \dots, \lambda_n\}$. Note that every element of L has at least one representation as an L_0-word. We will frequently identify an L_0-word with the element ℓ of L which it represents.

If $A \subseteqq L$, we define $A^+ = \{x \in M : [x, \gamma] \in A \ \forall \gamma \in \Gamma\}$. If $B \subseteqq M$, then $B^{+\prime} = \{\ell \in L : \ell x \in B \ \forall x \in M\}$.

It may be verified (cf. [2]) that these operators take ideals (invariant subgroups) onto ideals (invariant subgroups), and preserve intersections. If $\ell \in L$, $x \in M$, $\gamma \in \Gamma$, it may be shown that $\ell[x, \gamma] = [\ell x, \gamma]$. A *strong left unity* for M is an element (d, δ) of $M \times \Gamma$ such that $d\delta x = x$ for all $x \in M$. It is easily seen that, in this case, $[d, \delta]$ is the (two-sided) unity for L.

Proposition 1.1 ([3], Proposition 1.1). *Let M be a Γ-near-ring with left operator near-ring L. Then M is zerosymmetric if and only if L is zerosymmetric.*

All other notations, definitions and conventions concerning near-rings will be as in Pilz [7]. Let $\mathcal{V}$ be the variety of near-rings, zero-symmetric near-rings, Γ-near-rings (for a fixed Γ), or zerosymmetric Γ-near-rings. A *Kurosh–Amitsur radical (KA-radical)* is a subclass $\mathcal{R}$ of $\mathcal{V}$ satisfying:

(i) $\mathcal{R}$ is closed under homomorphic images.

(ii) Each element A of $\mathcal{V}$ contains a unique maximal $\mathcal{R}$-ideal, $\mathcal{R}(A)$, which contains all the $\mathcal{R}$-ideals of $\mathcal{V}$.

(iii) If $A \in \mathcal{V}$, $B \triangleleft A$, then $B, A/B \in \mathcal{R}$, implies $A \in \mathcal{R}$.

If, in addition, $A \in \mathcal{V}$, $B \triangleleft A$ implies that $\mathcal{R}(B) = B \cap \mathcal{R}(A)$, then $\mathcal{R}$ is said to be *ideal-hereditary*.

2. The J_2-radical

Throughout this section, let M be a Γ-near-ring with left operator near-ring L.

Let A be a Γ-subnear-ring of M, and let ϕ be a nonempty subset of Γ. We will denote by $[A, \phi]$ the subnear-ring of L generated by $\{[a, \gamma] : a \in A, \gamma \in \phi\}$. If $\gamma \in \Gamma$, $[A, \{\gamma\}]$ will be denoted $[A, \gamma]$. It is easily verified that $[A, \gamma] = \{[a, \gamma]\, a \in A\}$.

THEOREM 2.1 (cf. [1], Theorem 2.3). *Let A be a Γ-subnear-ring of M, and let L' be the left operator near-ring of A. Then L' is isomorphic to the factor near ring $[A,\Gamma]/K$ where $K=\{\ell\in[A,\Gamma]\colon \ell A=0\}$.*

PROOF. Let L'_0 be the generating set for L'. We will use the notation $\langle z,\gamma\rangle$ ($a\in A$, $\gamma\in\Gamma$) to denote elements of L'_0, in order to distinguish them from elements of the generating set L_0 of L. Let $A_0=\{[a,\gamma]\colon a\in A,\gamma\in\Gamma\}$. If $\lambda=[a,\gamma]\in A_0$, we define $\hat{\lambda}=\langle a,\gamma\rangle$. It may easily be verified that the operation $\lambda\to\hat{\lambda}$ is a well-defined mapping of A_0 onto L'_0. We define a mapping $f\colon[A,\Gamma]\to L'$ as follows:

If $\ell\in[A,\Gamma]$, there exist $\lambda_1,\dots,\lambda_n\in A_0$ such that $\ell=F(\lambda_1,\dots,\lambda_n)$ where F is an A_0-word. Let $f(\ell)=F(\hat{\lambda}_1,\dots,\hat{\lambda}_n)$, and let $a\in A$. It is clear that, if $\lambda\in A_0$, then $\lambda a=\hat{\lambda}a$. It follows by induction on n that

$$F(\lambda_1,\dots,\lambda_n)a=F(\hat{\lambda}_1,\dots,\hat{\lambda}_n)a. \tag{1}$$

Now suppose that $\mu_1,\dots,\mu_m\in A_0$ and G is an A_0-word such that

$$\ell=F(\lambda_1,\dots,\lambda_n)=G(\mu_1,\dots,\mu_m).$$

Then, if $a\in A$

$$F(\lambda_1,\dots,\lambda_n)a=G(\mu_1,\dots,\mu_m)a$$

whence

$$F(\hat{\lambda}_1,\dots,\hat{\lambda}_n)a=G(\hat{\mu}_1,\dots,\hat{\mu}_m)a$$

by (1). It follows that

$$F(\hat{\lambda}_1,\dots,\hat{\lambda}_n)=G(\hat{\mu}_1,\dots,\hat{\mu}_m)$$

whence the mapping $f\colon[A,\Gamma]\to L'$ is well-defined.

It is easily seen that f maps the near-ring $[A,\Gamma]$ homomorphically onto L'. Let K be the kernel of f. Then

$$\begin{aligned}\ell\in K&\Leftrightarrow f(\ell)=0\\&\Leftrightarrow f(\ell)a=0\quad\text{for all } a\in A\\&\Leftrightarrow \ell a=0\quad\text{for all } a\in A,\text{ by (1).}\end{aligned}$$

Hence $K=\{\ell\in L\colon \ell A=0\}$. The result now follows from the fundamental homomorphism theorem for near-rings.

An additive group G is called an *$M\Gamma$-group* if, for all $x,y\in M$, $\gamma,\mu\in\Gamma$ and $\gamma\in\Gamma$, it holds that:

(i) $x\gamma g\in G$

(ii) $x\gamma(y\mu g)=(x\gamma y)\mu g$

(iii) $(x+y)\gamma g=x\gamma g=y\gamma g$.

A subgroup H of G is called an *$M\Gamma$-subgroup* of G if $x\gamma h\in H$ for all $x\in M$, $\gamma\in\Gamma$, $g\in G$ such that $G=M\gamma g=\{x\gamma g\colon x\in M\}$, then G is called *monogenic.* The set $\{x \in M : x\gamma g = 0\}$ for all $\gamma \in \Gamma$, $g \in G$ is called the *annihilator* of G, and denoted $\mathrm{ann}_{M\Gamma} G$. It is easily seen that $\mathrm{ann}_{M\Gamma} G \triangleleft M$.

Let M be a zerosymmetric Γ-near-ring. An $M\Gamma$-group G is said to be of *type 2* if G is monogenic and contains no $M\Gamma$-subgroups other than 0 and G itself. An ideal P of M is called *2-primitive* if $P = \mathrm{ann}_{M\Gamma} G$, where G is an $M\Gamma$-group of type 2. The J_2-radical of M, $J_2(M)$ is the intersection of the 2-primitive ideals of M.

LEMMA 2.2 ([3] Proposition 3.3 (b) and [5] Theorem 2.2). *Let M be a zerosymmetric Γ-near-ring. Then*

(a) $J_2(L)^+ = J_2(M)$, where $J_2(L)$ denotes the J_2-radical of the near-ring L.

(b) If A is an invariant subgroup of M, then $J_2(A) \subseteq A \cap J_2(M)$.

LEMMA 2.3. *Let M be zerosymmetric, and let A be an invariant subgroup of M. Then*

$$\begin{aligned} J_2(A) &= J_2([A,\Gamma])^+ \\ &= \{a \in A : \langle a,\gamma\rangle \in J_2(L') \ \textit{for all}\ \gamma \in \Gamma\}. \end{aligned}$$

By Theorem 2.1, L' is isomorphic to $[A,\Gamma]/K$, where $K = \{\ell \in [A,\Gamma] : \ell a = 0 \text{ for all } a \in A\}$. Suppose that $\ell, \ell' \in K$, and that $x \in M$. It may be shown that, since A is an invariant subgroup of M, $\ell' x \in A$. Hence $\ell(\ell' x) = 0$ by definition of K. It follows that $\ell\ell' = 0$ for all $\ell, \ell' \in K$. Now L is zerosymmetric by Proposition 1.1, and J_2 is a KA-radical in the variety of zerosymmetric near-rings, which contains the nil near-rings. Hence, $K \subseteq J_2([A,\Gamma])$. It is easily shown from the definition of KA-radical that $J_2([A,\Gamma]/K) = J_2([A,\Gamma])/K$. Hence

$$\begin{aligned} J_2(A) &= \{a \in A : \langle a,\gamma\rangle \in J_2(L') \forall \gamma \in \Gamma\} \\ &= \{a \in A : [a,\gamma] + K \in J_2([A,\Gamma]/K \forall \gamma \in \Gamma\} \\ &= \{a \in A : [a,\gamma] \in J_2([A,\Gamma]) \forall \gamma \in \Gamma\} \\ &= J_2([A,\Gamma])^+. \end{aligned}$$

THEOREM 2.4. *Let M be a zerosymmetric Γ-near-ring with a strong left unity (d,δ). Let A be an invariant subgroup of M. Then*

$$J_2(A) = A \cap J_2(M).$$

PROOF. It is easily shown that $[d,\delta]$ is the (two-sided) unity for L. Hence, if $\ell \in L$, then $\ell = \ell[d,\delta] = [\ell d,\delta]$, whence every element of L is of the form $[x,\delta]$ for some $x \in M$. Suppose that $\ell \in [A,\Gamma]$. Since A is an invariant subgroup of M, $\ell d \in A$. Hence $\ell = \ell[d,\delta] = [\ell d,\delta]$. Thus, $[A,\delta]$ is an invariant subgroup of L. It follows from [6], Theorem 8.6, that

$$J_2([A,\delta]) = [A,\delta] \cap J_2(L). \tag{2}$$

Now if $a \in A$ and $\gamma \in \Gamma$, then $[a,\gamma] = [a,\gamma][d,\delta] = [a\gamma d,\delta] \in [A,\delta]$. Hence,

$$A \subseteqq [A,\delta]^+. \tag{3}$$

Now

$$\begin{aligned} J_2(A) &= J_2([A,\Gamma])^+ && \text{by Lemma 2.3} \\ &= J_2([A,\delta])^+ && \text{by (1)} \\ &= ([A,\delta] \cap J_2(L))^+ && \text{by (2)} \\ &= [A,\delta]^+ \cap J_2(L)^+ && \\ &= [A,\delta]^+ \cap J_2(M) && \text{by Lemma 2.2 (b)} \\ &\subseteqq A \cap J_2(M) && \text{by (3).} \end{aligned}$$

The reverse inclusion is Lemma 2.2(b). This completes the proof.

Let N be an arbitrary (i.e. not necessarily zerosymmetric) near-ring. Veldsman [8] defined an ideal A of N to be 2(0)-primitive if N/A is a 2-primitive, zerosymmetric near-ring. $J_{2(0)}(N)$ is the intersection of the 2(0)-primitive ideals of N. In [8], Theorem 4.2.4, it is shown that $J_{2(0)}$ is a KA-radical in the variety of all near-rings, and that if $A \triangleleft N$, then $J_{2(0)}(A) \subseteqq A \cap \cap J_{2(0)}(N)$, with equality if A is invariant.

Similarly, for an arbitrary Γ-near-ring M we define

$$J_{2(0)}(M) = \cap\{P \triangleleft M : M/P \text{ is zerosymmetric and 2-primitive}\}.$$

In [4], Propositions 3.7 and 3.10, it is shown that $J_{2(0)}(M) = J_{2(0)}(L)^+$ and that if $A \triangleleft M$, then $J_{2(0)}(A) \subseteqq J_{2(0)}(M) \cap A$. In view of these facts, and using the same arguments employed in the proofs of Lemma 2.3 and Theorem 2.4, we may prove the following result:

THEOREM 2.5. *Suppose that M has a strong left unity, and that A is an invariant ideal of M. Then*

$$J_{2(0)}(A) = A \cap J_{2(0)}(M).$$

REMARK. Let M be a zerosymmetric Γ-near-ring and let A be an invariant subgroup of M. It is an open question whether in general, it holds that $J_2(A) = A \cap J_2(M)$, as is the case for zerosymmetric near-rings. If it does hold, then J_2 is an ideal-hereditary KA-radical in the variety of zerosymmetric Γ-near-ring. In view of [4], Proposition 3.1, this would also imply that $J_{2(0)}$ is a KA-radical in the variety of all Γ-near-rings.

REFERENCES

[1] Booth, G. L., Operator rings of a gamma ring, *Math. Japon.* **31** (1986), No. 2, 175–183. *MR* **87e**: 16098

[2] Booth, G. L., A note of Γ-near-rings, *Studia Sci. Math. Hungar.* **23** (1988), 471–475. *MR* **90b**: 16043

[3] Booth, G. L., Jacobson radicals of Γ-near-rings, *Rings, radicals and modules* (Proc. Hobart conference, 1987), Research notes in mathematics series, Longmans, Harlow, 1989. *MR* **90m**: 16040

[4] Booth, G. L., Radicals in general Γ-near-rings, *Questiones Math.* **14** (1991), 117–127.

[5] Booth, G. L. and Groenewald, N. J., On radicals of gamma-near-rings, *Math. Japonica* **35** (1989), 417–425.

[6] Meldrum, J. D. P., *Near-rings and their links with groups,* Research notes in Mathematics, **134**, Pitman, Boston, 1985. *MR* **88a**: 16068

[7] Pilz, G., *Near-rings. The theory and its applications,* North-Holland, Amsterdam, 1977. *MR* **57** #9761

[8] Veldsman, S., Modulo-constant ideal-hereditary radicals of near-rings, *Quaestiones Math.* **11** (1988), No. 3, 253–278. *MR* **89k**: 16071

(Received February 23, 1989)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ZULULAND
PRIVATE BAG X1001
ZA-3886 KWADLANGEZWA
SOUTH AFRICA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 241–242

TRANSILLUMINATION OF LATTICE PACKING OF BALLS

T. HAUSEL

The paper of J. Horváth [1, see Theorem 2] contains the following statement: If a lattice packing of balls is given in E^n $(n \geqq 3)$, then there exists an affine subspace of E^n of dimension $n-2$ which is disjoint to the balls. In the proof of this statement (see [1], pp. 424–425) he uses the following (for simplicity we only take the special case $\ell = 2$ and $b_3 =$ a shortest non-zero lattice vector of a lattice). Let $a_1, a_2, b_3, b_4, \dots, b_n$ be an arbitrary system of linearly independent vectors from a lattice in E^n, that contains b_3. Then he considers the orthogonal projection of E^n to an orthogonal complement Σ_n^2 of $\operatorname{lin}\{b_3, \dots, b_n\}$. This Σ_n^2 is a 2-plane in E^n, and can be supposed to contain 0. However, for arbitrary $a_1, a_2, b_4, \dots, b_n$ (of length at least $\|b_3\|$) Σ_n^2 is in general no subspace of $\operatorname{lin}\{a_1, a_2, b_3\}$, hence this projection has no restriction to a projection of $\operatorname{lin}\{a_1, a_2, b_3\}$ into itself, which is however used in [1] further. Namely [1] applies to this restricted projection $\operatorname{lin}\{a_1, a_2, b_3\} \to \Sigma_n^2$ a theorem of I. Hortobágyi, that necessitates $\Sigma_n^2 \subset \operatorname{lin}\{a_1, a_2, b_3\}$.

In fact, [1], Theorem 2 itself is invalid, and here we actually prove the following

THEOREM. *There exists a lattice packing of balls in* E^n *intersecting every affine subspace of* E^n *of dimension* $n - [c\sqrt{n}]$, *where* c *is a positive absolute constant.*

PROOF. Throughout the proof we use the terminology, notations and results of the paper of R. Kannan and L. Lovász [2], in particular $\lambda_1(\mathsf{L}_n)$ denotes the minimal length of a non-zero vector of a lattice L_n, and $\mu_j(K, \mathsf{L}_n)$ is the j-th covering minimum of a convex body K with respect to a lattice L_n.

According to the result of Conway and Thompson [3, Chapter II, Theorem 9.5] there exists a lattice L_n of rank n with $\mathsf{L}_n = \mathsf{L}_n^*$ in E^n for which

$$\lambda_1(\mathsf{L}_n)\lambda_1(\mathsf{L}_n^*) \geqq c_1 n, \tag{1}$$

where c_1 is a positive absolute constant. Let us draw balls around all points of L_n with diameter $\lambda_1(\mathsf{L}_n)$. We show that this lattice packing of balls possesses the property claimed in the Theorem. Let P be one of the points of L_n. Let us consider the ball B which is drawn around P. Since $1/\lambda_1(\mathsf{L}_n^*)$ is the maximum of the distances of two parallel and neighbouring lattice

Akadémiai Kiadó, Budapest

hyperplanes in $\mathbf{L}_n$, (1) implies that the lattice width of $\mathbf{B}$ is not less than $c_1 n$. Thus, $\mu_1(\mathbf{B}, \mathbf{L}_n) \leqq 1/(c_1 n)$. Using Theorem (2.7) from [2] we get that

$$\mu_j(\mathbf{B}, \mathbf{L}_n) < c' j^2 \mu_1(\mathbf{B}, \mathbf{L}_n) \leqq \frac{c' j^2}{c_1 n}, \tag{2}$$

with a positive absolute constant c'. If we choose $j = [c\sqrt{n}]$ with $c^2 = c_1/c'$, then (2) proves our Theorem.

REMARK. If the conjecture $\mu_{j+1}(\mathbf{B}) \leqq \mu_j(\mathbf{B}) + \mu_1(\mathbf{B})$ (where $\mathbf{B}$ is ball) were true, see [2], then we could replace $n - [c\sqrt{n}]$ by $n - [cn]$ in the Theorem.

REFERENCES

[1] HORVÁTH, J., Über die Durchsichtigkeit gitterförmiger Kugelpackungen, *Studia Sci. Math. Hungar.* **5** (1970), 421–426. *MR* **45** #7615

[2] KANNAN, R. and LOVÁSZ, L., Covering minima and lattice-point-free convex bodies, *Ann. of Math.* **128** (1988), 577–602. *MR* **89i**: 52020

[3] MILNOR, J. and HUSEMOLLER, D., *Symmetric bilinear forms*, Ergebnisse der Mathematik un ihrer Grenzgebiete, Bd. 73, Springer-Verlag, New York–Heidelberg, 1973. *MR* **58** #22129

(Received April 27, 1992)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

Physics

Spacetime without Reference Frames

by
T. MATOLCSI

In the concept of this book spacetime is the fundamental notion; the points of spacetime are structured with the assumption of absolute time and absolute velocity of light resulting in the non-relativistic and special relativistic case, respectively. This gives the possibility of developing both the non-relativistic and the special relativistic chapters along the same notions: world line, observer, splitting of spacetime to space and time, reference frames, splitting of classical fields to spacelike and timelike components, the symmetry groups of spacetime (the Galilean and the Poincaré group).
The book contains lots of examples with detailed calculations through which the reader can clearly understand the connection between the traditional way of thinking and the new way of handling the problems presented in the book; the well-known special relativistic paradoxes are treated in detail. In the general relativistic case, only the basic thoughts are expressed.
The mathematics involved is rather simple and it is summarized in the second part of the book.

This book is an enlarged and revised version of "A concept of Mathematical Physics — Models for Space-Time" by T. Matolcsi

In English. 1992. Approx. 400 pages. Numerous figures. 17 x 25 cm. Hardbound.
Approx. $ 50.00

Akadémiai Kiadó, Budapest

PRINTED IN HUNGARY

TypoTEX Kft, Budapest

RECENTLY ACCEPTED PAPERS

ABUJABAL, H. A. S., BELL, H. E., KHAN, M. S. and KHAN, M. A., Commutativity of semiprime rings with power constraints

ANDERSON, D. D. and JAYARAM, C., Regular lattices

ANTOSIK, P. and SWARTZ, C., Boundedness and continuity for bilinear operators

AVDONIN, S. A., IVANOV, S. A. and JOÓ, I., Exponential series in the problems of initial and pointwise control of a rectangular vibrating membrane

BAKÁCS, T., BOGNÁR, K. and TUSNÁDY, G., On competitive killing activity of complement

BOYADZHIEV, K. N. and LEVAN, N., Strong stability of Hilbert space contraction semigroups

BÖHM, W. and MOHANTY, S. G., On random walks with barriers and their applications

CHENG, L.-Z., A commutative neutrix convolution of distribution on R^n

DEÁK, J., A bitopological view of quasi-uniform completeness I

DEÁK, J., A bitopological view of quasi-uniform completeness II

DEÁK, J., A bitopological view of quasi-uniform completeness III

DEÁK, J., A note on weak symmetry properties of quasi-uniformities

DEÁK, J., Quasi-uniform completeness and neighbourhood filters

DUFNER, G., Binäres Goldbachproblem in kurzen Intervallen. Teil I. Die explizite Formel

DUFNER, G., Binäres Goldbachproblem in kurzen Intervallen. Teil II. "Fast alle"-Ergebnisse

FÉNYES, T., On the Fourier transform of the modified Bessel function with respect to the order

GOULD, V., Straight left orders

ILLÉS, Á. and VERMES, I., Eine Extremaleigenschaft des regulären Mosaiks $\{p,4\}$ in der hyperbolischen Ebene und ihre Verallgemeinerung für die Mosaike $\{p,2k\}$

KÜNZLI, H.-P. A. and LÜTHY, A., Dense subspaces of quasi-uniform spaces

MICK, S., Drehkegel des zweifach isotropen Raumes durch vier gegebene Punkte

MIELOSZYK, E., Well defined solution problem for certain generalized boundary value problem

RAUSCHE, A., Investigation of the lower shadow in various posets

SARAN, J. and RANI, S., Some distribution results of rank order statistics

SCHMIDT, T., Homomorphisms of distributive lattices as restriction of congruences: the planar case

SEBESTYÉN, Z. and KAPOS, L., On range characterization of adjoint operators on Hilbert space

STROMMER, GY., Zur Konstruktion des regulären Siebzehnecks

TUMURBAT, S., Right Lie algebras and their multiplication algebras

VÁSÁRHELYI, É., Ein isoperimetrisches Problem bezüglich Zerlegungen der euklidischen Ebene

WISMATH, S. L., Joins of some varieties of semigroups

ZEMPLÉNI, A., In the max-semigroup of probability distributions over the plane there is no Khinchine-type decomposition theorem

CONTENTS

HU ISSN 0081-6906 Index: 26.748

Studia Scientiarum Mathematicarum Hungarica

EDITOR-IN-CHIEF

A. HAJNAL

13

VOLUME 27
NUMBERS 3–4
1992

1828

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

Subscription information

Orders should be addressed to

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

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 243–260

MULTIEXTREMAL ALLOCATION PROBLEMS
(models and solution methods)

V. R. KHACHATUROV

Abstract

In the paper models and methods are considered for solving various multiextremal allocation problems in combinatorial formulations.

1. The main theses of the methods which are used

Let a function $f(\omega)$ be defined on the set Ω consisting of a finite number of elements $\omega \in \Omega$. It is required to determine such an element

(1) $$\alpha \in \Omega$$

that

(2) $$f(\alpha) = \min f(\omega), \qquad \omega \in \Omega.$$

Practically any discrete programming problems may be written in such a form.

If the number of elements in the set $\Omega(|\Omega|)$ is small and the values $f(\omega)$ for $\omega \in \Omega$ are calculated simply enough, one can use the method of the total enumeration to determine α and $f(\alpha)$. Otherwise the necessity arises to elaborate methods excluding the total overselection. Among known methods which proved to be effective in solving problems of different types one can mention the methods of linear and dynamic programming [1, 2], consecutive calculations [3], branch and bound method [4] and some others [5–7]. However, all of them suffer from the common drawback, namely, "the sensitivity" to small alterations of problem conditions. For example, if in a linear programming problem the convexcave function is considered in place of a linear functional the simplex-method becomes inapplicable for finding its minimum although the set on which this minimum is accomplished (vertices of a polygon) remains the same; addition of a complementary constraint often makes inapplicable the dynamic programming method which is used successfully in solving the problem without the constraint; a violation of the condition $\mathcal{S}(\delta, \gamma) \leqq 0$ can make inapplicable the method of consecutive calculations; even the universal (in exposition) branch and bound method has this drawback: an insignificant alteration of the conditions of the problem

Akadémiai Kiadó, Budapest

under consideration may demand the construction of new rules of branching and estimate calculation.

The approximation-combinatorial method [8, 15], presented below, shows ways of modification of the methods to decrease their "sensitivity" to an alteration of the problem conditions and enlarges the class of problems which can be solved.

We describe the basic aspects of this method in solving the problem (1), (2).

1. On the set Ω an approximating function $P(\omega)$ is defined in such a way, that

$$f(\alpha) \geqq P(\alpha), \tag{3}$$

and for $P(\omega)$ there are effective methods and algorithms for determining not only $\alpha_0 \in \Omega$ with

$$P(\alpha_0) = \min_{\omega \in \Omega} P(\omega), \tag{4}$$

but all the elements $\omega \in \Omega$ with the values $P(\omega)$ differing from $P(\alpha_0)$ in not more than some fixed number $R \geqq 0$. In other words it becomes possible to determine such a subset $\Omega_0 \subset \Omega$, that

$$P(\alpha_0) \leqq P(\omega) \leqq P(\alpha_0) + R, \qquad \omega \in \Omega_0, \tag{5}$$

$$P(\omega) > P(\alpha_0) + R, \qquad \omega \in \Omega \setminus \Omega_0. \tag{6}$$

2. Certain number $\bar{c}$ is chosen in such a way that

$$P(\alpha_0) \leqq \bar{c}.$$

3. The subset $\Omega_0 \subset \Omega$ is determined by solving the problem (5), (6) under

$$P(\alpha_0) + R = \bar{c}. \tag{7}$$

4. The element $\tilde{\alpha} \in \Omega_0$ is determined in such a way that

$$f(\tilde{\alpha}) = \min_{\omega \in \Omega_0} f(\omega). \tag{8}$$

The value $\tilde{\alpha}$ and $f(\tilde{\alpha})$ are adopted as a solution.

A criterion of optimality. If

$$f(\tilde{a}) \leqq \bar{c}, \tag{9}$$

holds then

$$\tilde{\alpha} = \alpha, \qquad f(\tilde{\alpha}) = f(\alpha). \tag{10}$$

To prove the criterion of optimality, we show that $\alpha \in \Omega_0$, i.e. (5) is valid. Using (7), (9), (2)–(4) in succession, we get

$$P(\alpha_0) + R = \bar{c} \geqq f(\tilde{\alpha}) \geqq f(\alpha) \geqq P(\alpha) \geqq P(\alpha_0).$$

Afterwards taking into consideration (8) and (2) we get $f(\tilde{\alpha}) = f(\alpha)$, which was required to prove.

Approximate solution. If

$$f(\tilde{\alpha}) > \bar{c}, \tag{11}$$

holds then

$$\bar{c} < f(\alpha) \leqq f(\tilde{\alpha}), \tag{12}$$

i.e. $\tilde{\alpha}$ and $f(\tilde{\alpha})$ define an approximate solution.

The right-hand side of the inequality (12) follows from (2). The left-hand side will be proved from the contrary. Let $\bar{c} \geqq f(\alpha)$, then

$$P(\alpha_0) + R = \bar{c} \geqq f(\alpha) \geqq P(\alpha) \geqq P(\alpha_0). \tag{13}$$

Thus $\alpha \in \Omega_0$ (see (5)). From this and also from (2), (8), (13) $f(\alpha) = = f(\tilde{\alpha}) \leqq \bar{c}$ follows which contradicts (11). Hence, $\bar{c} < f(\alpha)$.

Corollary 1. *If we determine beforehand in some way or other the upper bound c for $f(\omega)$*

$$f(\alpha) \leqq c,$$

and take $\bar{c} = c$, then $f(\tilde{\alpha}) = f(\alpha)$ always.

The proof follows from (13), (5), (2), (8).

Using the relations (7), (5) and (6) consider Ω_0 as a function of the value $\bar{c}$, i.e. $\Omega_0 = \Omega_0(\bar{c})$. Then

Corollary 2. *If $\bar{c}_1 \leqq \bar{c}_2$, then*

$$\Omega_0(\bar{c}_1) \subset \Omega_0(\bar{c}_2) \quad \textit{and} \quad |\Omega_0(\bar{c}_1)| \leqq |\Omega_0(\bar{c}_2)|.$$

Corollary 3. *If $f(\omega) \geqq P(\omega)$ for each $\omega \in \Omega$ then all $\omega \in \Omega$ belong to Ω_0 with the value $f(\omega) \leqq \bar{c}$.*

In fact, since $P(\alpha_0) + R = \bar{c} \geqq f(\omega) \geqq P(\omega) \geqq P(\alpha_0)$, $\omega \in \Omega_0$. Thus if $\Omega_0^0 \subset \Omega$ is the set of all $\omega \in \Omega$ with $f(\omega) \leqq \bar{c}$ then $\Omega_0^0 \subset \Omega_0$. In this case the function $f(\omega)$ itself possesses the properties (4)–(6) of the approximating function and may be used in an approximation combinatorial method for solving more complicated problems.

In papers [8, 15] the approximation-combinatorial method is presented in more detail and various classes and examples of approximating functions used for solving various mathematical programming problems are given. Here we shall consider a wide class of approximating functions used in the approximation-combinatorial method for solving allocation problems, namely, functions satisfying the sufficient condition of application of the consecutive calculations method [3].

In the method of consecutive calculations the functions $P(\omega)$ defined on the set Ω of all subsets ω of the set $I = \{1, 2, \dots, m\}$ are treated. (The number of elements $\omega \in \Omega$ or $\omega \subset I$ is equal to $|\Omega| = 2^m$.)

The requirement of application of the condition

$$\mathcal{S}(\delta, \gamma) \equiv P(\delta) + P(\gamma) - P(\delta \cup \gamma) - P(\delta \cap \gamma) \leqq 0 \tag{14}$$

for any two subsets δ and γ of the set I is the sufficient condition of application of the method of consecutive calculations for determination $\min P(\omega) = P(\alpha_0)$ for all $\omega \in \Omega$.

There are three rules of rejection [3, 9, 20] of wittingly not optimal subsets, which are used in the algorithms of consecutive calculations [3, 9, 20] for the determination of the optimal subsets (global minimum) $\alpha_0 \subset I$. These rules of rejection allow to determine the global minimum of the function $P(\omega)$ by looking through about m^3 of variants from the total number 2^m of variants.

For solving the problem (5), (6) three generalized rules of rejection [9, 10, 20] are suggested from which as a particular case the rules of rejection for determination of the global minimum follow.

Any algorithm for seeking a global minimum [9] may simultaneously serve as an algorithm for seeking all close variants if instead of the rules of rejection used in them one will apply the corresponding generalized rules of rejection.

The modified algorithms of consecutive calculations [9], successfully passed the experimental test [11, 12, 13] are elaborated to improve the algorithms of solving the problems (5), (6).

The first generalized rule of rejection. If for any subsets $\omega_1 \subset \omega_2 \subset I$ the values $P(\omega_1)$ and $P(\omega_2)$ are known and if $P(\omega_1) + R < P(\omega_2)$, then one can neglect (exclude from the consideration) all $2^{m-|\omega_2|}$ subsets $\omega \supset \omega_2$ because for them wittingly

$$P(\omega) > P(\alpha_0) + R.$$

Let $P(\omega_1) + R < P(\omega_2)$. Take $\delta = \omega_2$ and $\gamma = \omega_1 \cup (\omega \setminus \omega_2)$, then $\delta \cup \gamma = \omega$, $\delta \cap \gamma = \omega_1$. Such values δ, γ, $\delta \cup \gamma$, $\delta \cap \gamma$ are possible as $\omega_1 \subset \omega_2 \subset \omega$. From the condition $\mathcal{S}(\delta, \gamma) \leqq 0$ get $P(\omega) \geqq P(\omega_2) + P(\omega_1 \cup (\omega \setminus \omega_2)) - P(\omega_1)$, and since $P(\omega_1 \cup (\omega \setminus \omega_2)) \geqq P(\alpha_0)$, $P(\omega_2) - P(\omega_1) > R$, then $P(\omega) > P(\alpha_0) + R$, which is asserted in the rule of rejection. For $R = 0$ this rule coincides with the first rule of rejection from [3].

The second generalized rule of rejection. If for any subsets $\omega_1 \subset \omega_2 \subset I$ the values $P(\omega_1)$ and $P(\omega_2)$ are known and if $P(\omega_1) > P(\omega_2) + R$, then one can neglect all $2^{|\omega_1|}$ subsets $\omega \subset \omega_1$, because for them wittingly

$$P(\omega) > P(\alpha_0) + R.$$

Let $P(\omega_1) > P(\omega_2) + R$. Then $\delta = \omega_1$ and $\gamma = \omega \cup (\omega_2 \setminus \omega_1)$ then $\delta \cup \gamma = \omega_2$, $\delta \cap \gamma = \omega$. Such values δ, γ, $\delta \cup \gamma$, $\delta \cap \gamma$ are possible as $\omega \subset \omega_1 \subset \omega_2$. As well as in the previous case, we get

$$P(\omega) \geqq P(\omega_1) + P(\omega \cup (\omega_2 \setminus \omega_1)) - P(\omega_2) \geqq P(\alpha_0) + R,$$

which is asserted in the rule of rejection. For $R = 0$ this rule coincides with the second rule of rejection from [3].

The third generalized rule of rejection. If for any subsets $\omega_1 \subset \omega_2$ it turns out that either $P_1(\alpha') > P(\tilde{\alpha}_0) + R$ or $P_2(\alpha') > P(\alpha_0) + R$ then there is no necessity of considering all $2^{|\omega_2 \setminus \omega_1|}$ subsets $\omega_1 \subset \omega \subset \omega_2$ because wittingly $P(\alpha') > P(\alpha_0) + R$, i.e. among all the subsets of this kind there is close to the optimal subset. Here it is assumed that:

$P(\tilde{\alpha}_0)$ is the known value of the function $P(\omega)$ for some $\tilde{\alpha}_0 \subset I$;

$$P(\alpha') = \min_{\omega_1 \subset \omega \subset \omega_2} P(\omega),$$

$$P_1(\alpha') \equiv P(\omega_1) - \sum_{i \in \omega_2 \setminus \omega_1} \Delta_1(i) \leqq P(\alpha'),$$

$$P_2(\alpha') \equiv P(\omega_2) - \sum_{i \in \omega_2 \setminus \omega_1} \Delta_2(i) \leqq P(\alpha'),$$

where

$$\Delta_1(i) = \begin{cases} P(\omega_1) - P(\omega_1 \cup i), & \text{if } P(\omega_1) - P(\omega_1 \cup i) \geqq 0 \\ 0, & \text{if } P(\omega_1) - P(\omega_1 \cup i) < 0. \end{cases}$$

$$\Delta_2(i) = \begin{cases} P(\omega_2) - P(\omega_2 \setminus i), & \text{if } P(\omega_2) - P(\omega_2 \setminus i) \geqq 0 \\ 0, & \text{if } P(\omega_2) - P(\omega_2 \setminus i) < 0. \end{cases}$$

The validity of this rule of rejection follows from evident inequalities

$$P(\alpha') \geqq P_1(\alpha') \quad (\text{or } P_2(\alpha')) > P(\tilde{\alpha}_0) + R \geqq P(\alpha_0) + R.$$

Give the proof of the estimate $P_1(\alpha')$. As $\omega_1 \subset \alpha'$ then it is obvious that $\alpha' = \omega_1 \cup \{\alpha' \setminus \omega_1\}$. Let $|\alpha' \setminus \omega_1| = r$ and $\alpha' \setminus \omega_1 = \{i_1, i_2, \dots, i_r\}$. Take $i_1 \in \alpha' \setminus \omega_1$ and define the subsets $\delta = \omega_1 \cup i_1$, $\gamma = \alpha' \setminus i_1$, then $\sigma = \delta \cup \gamma = \alpha'$, $\varepsilon = \delta \cap \gamma = \omega_1$. From the condition (14) we obtain

$$P(\alpha') \geqq P(\alpha' \setminus i_1) - [P(\omega_1) - P(\omega_1 \cup i_1)]. \tag{15}$$

For $P(\alpha' \setminus i_1)$ we obtain similarly

$$P(\alpha' \setminus i_1) \geqq P\left(\{\alpha' \setminus i_1\} \setminus i_2\right) - [P(\omega_1) - P(\omega_1 \cup i_2)],$$

where $i_2 \in \{\alpha' \setminus i_1\}\omega_1 = \{\alpha' \setminus \omega_1\} \setminus i_1$. Substitute to the previous inequality:

$$P(\alpha') \geqq P(\{\{\alpha' \setminus i_1\} \setminus i_2) - [P(\omega_1) - P(\omega_1 \cup i_1)][P(\omega_1) - P(\omega_1 \cup i_2)].$$

Making it r times for all $i_l \in \alpha' \setminus \omega_1 = \bigcup_{k=1}^{r} i_k$, we get:

$$P(\alpha') \geqq P(\omega_1) - \sum_{i \in \alpha' \setminus \omega_1} [P(\omega_1) - P(\omega_1 \cup i)].$$

From (15) it follows that

$$P(\omega_1) - P(\omega_1 \cup i_1) \geqq P(\alpha' \setminus i_1) - P(\alpha') \geqq 0.$$

Since (15) is valid for all $i_l \in \alpha' \setminus \omega_1$, then $P(\omega_1) - P(\omega_1 \cup i) \geqq 0$ for all $i \in \in \alpha' \setminus \omega_1$. Therefore one can write down

$$P(\alpha') \geqq P(\omega_1) - \sum_{i \in \alpha' \setminus \omega_1} \Delta(i).$$

According to the definition $\Delta(i) \geqq 0$ for all $i \in \omega_2 \setminus \omega_1$, therefore

$$P(\alpha') \geqq P(\omega_1) - \sum_{i \in \omega_2 \setminus \omega_1} \Delta(i) \equiv P_1(\alpha')$$

which was required to prove. The proof of the estimate $P_2(\alpha')$ is carried out similarly.

The condition (14) is satisfied by the function of the sufficiently general kind. We give examples of such functions.

EXAMPLE 1.

$$P(\omega) = P^1(\omega) + P^2(I \setminus \omega).$$

If $P^1(\omega)$ and $P^2(\omega), \omega \subset I$ satisfy the condition (14), then the function $P(\omega)$ satisfies (14) as well. This is settled by direct verification.

EXAMPLE 2.

$$P(\omega) = P^1(\omega) - P^2(\omega).$$

$P(\omega)$ satisfies (14) if the function $P^1(\omega)$ also satisfies (14) and for $P^2(\omega)$ the inverse inequality is valid, i.e.

$$\mathcal{S}(\delta, \gamma) \geqq 0.$$

EXAMPLE 3.

$$P(\omega) = P(\omega_1, \dots, \omega_n) = \sum_{t=1}^{T} \sum_{i=1}^{n} c_{it} P^{it}(\omega_i).$$

Let every function $P^{it}(\omega_i)$ satisfy the condition (14), $c_{it} \geqq 0$, $\omega_i \subset I = \{1,2,\dots,m\}$. Then $P(\omega_1,\dots,\omega_n)$ also satisfies (14) under the corresponding definition of notions $\delta \cup \gamma$ and $\delta \cap \gamma$.

Introduce for $P(\omega)$ the notions of the unification σ and intersection ε for two elements δ and γ.

If $\delta = (\delta_1,\dots,\delta_n)$, $\gamma = (\gamma_1,\dots,\gamma_n)$, then

$$\sigma = \delta \cup \gamma = (\delta_1 \cup \gamma_1,\dots,\delta_n \cup \gamma_n), \quad \text{and} \quad \varepsilon = \delta \cap \gamma = (\delta_1 \cap \gamma_1,\dots,\delta_n \cap \gamma_n).$$

Now (14) is observed which one can easily verify directly.

If any $c_{it} < 0$ and functions $P^{it}(\omega_i)$ corresponding to them satisfy the condition $\mathcal{S}\ (\delta,\gamma) \geqq 0$, then $P(\omega)$ satisfies (14).

EXAMPLE 4.

$$P(\omega) = P(\omega_1,\dots,\omega_r) = P^1(\omega_1) + P^2(\omega_1 \cup \omega^2) + \dots + P^r\Bigl(\bigcup_{i=1}^{r} \omega_i\Bigr).$$

Let every function $P^i(\eta)$ satisfy condition (14) for $\eta \subset I = \{1,2,\dots,m\}$ and σ and ε are defined as well as in the previous case.

The function $P(\omega)$ satisfies condition (14) if the functions $P^i(\eta)$ possess the property $P^i(\eta^1) \geqq P^i(\eta^2)$ for $\eta^1 \subset \eta^2$.

This statement is verified by directly substituting into (14) the corresponding values $P(\omega)$ and by using the property of functions $P^i(\eta)$.

2. Classes of allocation problems solved by the method of consecutive calculations

One can present numerous particular formulations of problems in which condition (14) is valid. Allocation problems of various kinds are mostly examined. Present some of them.

Problem 1. Find

$$\min_{x_{ij}} \Bigl[\sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij} + \sum_{i\in\omega} T_i \operatorname{sign} x_i\Bigr]$$

under the conditions

$$\sum_{i\in I} x_{ij} = b_j, \quad \sum_{j\in J} x_{ij}d_{ij} \leqq a_i \qquad x_{ij} \geqq 0,$$

where

$$x_i = \sum_{j\in J} x_{ij}, \quad \operatorname{sign} x_i = \begin{cases} 0, & x_i = 0, \\ 1, & x_i > 0, \end{cases}$$

$c_{ij} \geqq 0$, $T_i \geqq 0$, $b_j > 0$, $d_{ij} > 0$.

The proof is in [3, 9, 13].

Problem 2. Find

$$\min_{x_{ij}} \left[\sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{i \in I} g_i(x_i) \right]$$

under the conditions of Problem 1, $g_i(x_i)$ is piecewise linear, discontinuous function.

See the solution in [9].

Problem 3. Dynamic allocation problems without constraints on capacities of production points: find

$$\min_{x_{ij}} \left[\sum_{t \in \tau} \sum_{j \in J} \sum_{i \in I} c_{ij}^t (x_{ij}^t) + \sum_{t \in \tau} \sum_{i \in I} g_i^t \left(\sum_{s=1}^{t-1} x_i^S, x_i^t \right) \right]$$

under the conditions:

$$(16) \quad \sum_{i \in I} x_{ij}^t = b_j^t, \; j \in J, \; t \in \tau = \{1, \dots, T\}, \; x_{ij}^t \geqq 0, \; i \in I, \; j \in J, \; t \in \tau.$$

$$x_i^t = \sum_{j \in J} x_{ij}^t, \quad i \in I, \quad t \in \tau.$$

If

(17)

$$g_i^t \left(\sum_{S=1}^{t-1} x_i^S, x_i^t \right) = \begin{cases} (D_i^t + k_i^t x_i^t) \operatorname{sign}(x_i^t), & \text{if } \sum\limits_{S=1}^{t-1} x_i^S = 0, \\ k_i^t x_i^t, & \text{if } \sum\limits_{S=1}^{t-1} x_i^S > 0 \left(\sum\limits_{S=1}^{0} x_i^S \stackrel{\text{def}}{=} 0 \right), \end{cases}$$

$c_{ij}^t(x_{ij}^t)$ is the convex upward function, then it can be shown that the function $P(\omega_1, \dots, \omega_T)$ corresponding to this problem is presented in a form of the sum of the functions from Problems 3 and 4.

If $g_i^t \left(\sum\limits_{S=1}^{t-1} x_i^S, x_i^t \right) = (D_i^t + k_i^t x_i^t) \operatorname{sign}(x_i^t)$, then the corresponding function $P(\omega_1, \dots, \omega_T) = \sum\limits_{t \in \tau} P^t(\omega_t)$, where all $P^t(\omega_t)$ satisfy (14).

Thus in this case it is required to solve T problems $\min\limits_{\omega_t \subset I} P^t(\omega_t)$ by the method of consecutive calculations. If $D_i^t = D_i$, $t \in \tau$, $i \in I$, then not the vector $(\omega_1, \dots, \omega_T)$ is the allocation variant, but $\omega \subset J$, where ω is the subset of production points created for the time interval $(1, T)$. I.e. in this case it is required to solve one problem $\min_{\omega \subset I} P(\omega)$ by the method of consecutive calculations.

Problem 4. Dynamic allocation problems with constraints on capacities of production points and on communication capacities.

These problems differ from Problem 3 in the additional constraints:

$$x_i^t \leqq a_i^t, \quad i \in I, \ t \in \tau,$$
$$x_{ij}^t \leqq d_{ij}^t, \quad i \in I, \ j \in J, \ t \in \tau.$$

If $c_{ij}^t(x_{ij}^t) = c_{ij}^t x_{ij}^t$ then for solving these problems the method of consecutive calculations is used as well as in the previous case. However, for the calculation of values of the functions $P(\omega)$ it is required to solve transportation linear programming problems.

3. Classes of allocation problems solved by the approximation-combinatorial method

1. *Allocation problems with additional constraints.* An appearance of additional conditions in Problems 1–4 may lead to violation of condition (14) for them, and this may imply that the method of consecutive calculations will not be applicable to solve them.

For example, the following conditions are often met.

Condition 1. The constraint on the total extension of communications:

$$\sum_{i \in I} \sum_{j \in J} x_{ij} d_{ij} \leqq d.$$

Condition 2. An attachment of a customer only to one supplier:

$$x_{ij} \in \{0, b_j\}.$$

In the first case it is necessary to carry out an approximation by neglecting the additional condition. After determination of the subset Ω_0 the calculation of $f(\omega)$ for all $\omega \in \Omega_0$ is determined by solving transportation problems with the additional Condition 1.

In the second case Condition 2 at first is substituted by the condition $x_{ij} \geqq 0$. After the determination of the subset Ω_0 for the approximating function the calculation of $f(\omega)$ for $\omega \in \Omega_0$ (already with condition 2) leads to solving transportation problems with Boolean variables which are solved by various methods; in particular, the approximation combinatorial method [14] may be applied to such a problem, too. Other additional conditions are possible as well.

2. *Allocation problems with communications.* There is a set $J = \{1,2,\ldots,n\}$ of raw material sources with known raw material amounts $b_j > 0$. The set of possible points of raw materials work $I = \{1,2,\ldots,m\}$ with unknown amounts of work x_i, $a_i \geqq x_i \geqq 0$ is given. For each $i \in I$ the function $g_i(x_i)$ of raw materials work cost

$$g_i(x_i) = \begin{cases} 0, & x_i = 0, \\ k_i x_i + T_i, & x_i > 0 \end{cases}$$

is given.

Worked raw materials are to be delivered to customers located on the set of points Q.

Two total graphs of possible communications are given. The graph $u_1(J \cup I)$ connects the raw material sources J with the possible points of work I; the graph $u_2(I \cup Q)$ connects points of work with customer points. On ribs $(j;z) \subset u_1$ and $(i;t) \subset u_2$ the functions $r_{jz}(w_{jz})$ and $P_{it}(y_{it})$ equal to

$$r_{jz}(w_{jz}) = \begin{cases} 0, & w_{jz} = 0, \\ d_{jz} + l_{jz} w_{jz}, & w_{jz} > 0, \end{cases}$$

$$P_{it}(y_{it}) = \begin{cases} 0, & y_{it} = 0, \\ v_{it} + u_{it} y_{it}, & y_{it} > 0, \end{cases}$$

are given.

Here w_{jz} is the amount of raw materials conveyed along the rib (communication) (j,z); y_{it} is the amount of worked raw materials conveyed along the communication (i,t); l_{jz}, u_{it} are the conveyance cost of a unit of load along communications (j,z) and (i,t), respectively; d_{jz}, v_{it} are the construction costs of the communications (j,z), (i,t).

For each $\omega \subset I$ one can construct the network $\mathcal{S}_1(\omega \cup J)$ (subgraph of the graph u_1) connecting J with ω along which all raw materials are conveyed to the point ω, and the network $\mathcal{S}_2(\omega \cup Q)$ (subgraph of the graph u_2) connecting ω with Q along which all worked production is conveyed to the customers Q.

The costs of these networks determined by means of the functions r_{jz} and P_{it}, will be denoted by $c_1(\omega \cup J)$ and $c_2(\omega \cup Q)$, respectively. The cost of raw materials work at the points $\omega \subset I$ determined by means of the functions $g_i(x_i)$, $i \in \omega$, will be denoted by $g(\omega)$.

Then for each $\omega \subset I$

$$f(\omega) = \min_{\{\mathcal{S}_1(\omega \cup J)\},\{\mathcal{S}_2(\omega \cup Q)\}} \left(c_1(\omega \cup J) + g(\omega) + c_2(\omega \cup Q)\right).$$

The problem consists in determination of such $\alpha \subset I$, that $f(\alpha) = \min f(\omega)$ for each $\omega \subset I$.

Under the most general conditions it is highly difficult to obtain the exact solution of this problem as even with fixed ω the determination of $f(\omega)$ is not an easy problem. The application of the approximation-combinatorial method as it has been shown in practical calculations gives quite acceptable approximate solutions.

In the simplest case the exact solution can be found.

Let $d_{jz} = v_{it} = 0$. Then it is easy to show that for each $\omega \subset I$

$$f(\omega) = \min_{x_{ij}} \left[\sum_{i\in\omega} \sum_{j\in J} c_{ij} x_{ij} + \sum_{i\in\omega} (k_i + d_i x_i) \right] + \sum_{i\in\omega} T_i$$

under the conditions

$$\sum_{i\in\omega} x_{ij} = b_j, \quad x_i \equiv \sum_{j\in J} x_{ij} \leqq a_i, \quad x_{ij} \geqq 0,$$

where x_{ij} is the amount of the raw materials conveyance from the j-th source to the i-th point of work; $\|c_{ij}\|$ is the matrix of the shortest "distances" calculations on the basis of the variety of values $\{l_{jz}\}$ corresponding to arcs of the graph U_1 (the value c_{ij} implies the minimum conveyance cost of a unit of raw materials from the j-th point to the i-th); α_i is the minimum "distance" (the minimum conveyance cost of a unit of worked raw materials) from the point i to the set of customers Q. For such a function $f(\omega)$ the optimal solution $(\alpha, f(\alpha))$ can be found by the method of consecutive calculations.

Let $d_{jz} \geqq 0$, $v_{it} \geqq 0$ (for some arcs it is a strict inequality), $x_i \geqq 0$.

For each $j \in J$ and $i \in I$ find the values

$$\mathcal{S}_{ij} = \min_{k\in J\cup\{i\}, k\neq j} \left[d_{jk} + (l_{jk} + c_{ik}) b_j \right],$$

where c_{ik} is determined as in the last case and $c_{ii} = 0$.

For each $i \in I$ one can define the function

$$\Psi_i(x_i) = \begin{cases} 0, & x_i = 0, \\ r_i + l_i x_i, & x_i > 0. \end{cases}$$

The line $r_i + l_i x_i$ passes through two points:

$$\left(\underline{a}, \min_{t\in I\cup Q} (v_{it} + (u_{it} + \alpha_t)\underline{a}) \right),$$
$$\left(\overline{a}, \min_{t\in I\cup Q} (v_{it} + (u_{it} + \alpha_t)\overline{a}) \right),$$

where α_t is determined as in the previous case and if $\alpha_t = 0$ if $t \in Q$, $\underline{a} = \min_{j\in J} b_j$, $\overline{a} = \sum_{j\in J} b_j$.

Let us construct the function $P(\omega)$:

$$P(\omega) = \min_{x_{ij}} \left[\sum_{i\in\omega} \sum_{j\in J} \frac{\mathcal{S}_{ij}}{b_j} x_{ij} + \sum_{i\in\omega} (k_i + l_i) x_i \right] + \sum_{i\in\omega} (T_i + r_i)$$

under conditions

$$\sum_{i\in\omega} x_{ij} = b_j, \quad x_{ij} \geqq 0, \quad x_i = \sum_{j\in J} x_{ij}.$$

It can be shown that $f(\omega) \geqq P(\omega)$ for all $\omega \subset I$. By using the approximation-combinatorial method we find $\Omega_0 \subset \Omega$. Since it is difficult to calculate the values $f(\omega)$, some solution is determined with the value $\overline{f}(\omega) \geqq f(\omega)$ instead of $f(\omega)$, for example, as follows. The values $x_{ij} = x_{ij}(\omega)$ for which the value $P(\omega)$ is accomplished are determined simultaneously with calculating $P(\omega)$, and for each $i \in \omega$ the subsets of raw materials sources $\lambda_i \subset J$ attached to the i-th point of work are determined

$$\bigcup_{i\in\omega} \lambda_i = J, \qquad \lambda_i \bigcap_{i\neq k} \lambda_k = \emptyset.$$

By $x_{ij}(\omega)$

$$x_i(\omega) = \sum_{j\in J} x_{ij}(\omega), \qquad i \in \omega$$

are determined.

Then it can be determined

$$\overline{f}(\omega) = \sum_{i\in\omega} \min_{\{S_1(\{i\}\cup\lambda_i)\}} c_1(\{i\} \cup \lambda_i) + \sum_{i\in\omega}(k_i x_i(\omega) + T_i) + \min_{\{S_2(\omega\cup Q)\}} c_2(\omega \cup Q),$$
$$x_i = x_i(\omega),$$

by solving $|\omega| + 1$ problems of the network optimization with the discontinuous functions r_{jz} and P_{it} on ribs. There are efficient algorithms and programs for solving such problems.

It is accepted

$$\overline{\alpha} \in \Omega_0 \quad \text{with} \quad \overline{f}(\overline{\alpha}) = \min_{\omega\in\Omega_0} \overline{f}(\omega)$$

as an approximate solution of the original problem.

In this case $f(\alpha)$ is estimated as follows:

$$\overline{f}(\overline{\alpha}) \geqq f(\alpha) \geqq P(\alpha_0).$$

3. *The problem of allocation of territorial-industrial complexes.* There are n industries with different kinds of production. The set of points of possible facilities construction I is known. At each point $i \in I$ a construction of a facility of any industry is possible. The sets of points J^k, $k = 1, 2, \ldots, n$ in which customers of production of the k-th industry are located are known. Every j_k-th customer of the k-th industry consumes $b^k_{j_k} > 0$ production of the k-th industry. The matrices $\|c_{ij_k}\|$, $i \in I$, $j_k \in J^k$ of the conveyance cost of a unit of production of the k-th industry from the i-th facility to the

j_k-th customer are known. The upper bounds $a_i^k > 0$ for the amount x_i^k of production of the k-th industry at the i-th point are known. The functions

$$g_i(x_i^k) = \begin{cases} 0, & x_i^k = 0, \\ l_i^k x_i^k + T_i^k, & x_i^k > 0,\ (T_i^k \geqq 0), \end{cases}$$

of the cost of production x_i^k for the k-th industry at the i-th point are known, too. By assumption that a facilities allocation at one industry does not influence a facilities allocation at the other the total expenses $\overline{P}^k(\omega_k)$ associated with the facilities allocation of the k-th industry on the subset of the points $\omega_k \subset I$ and with the corresponding transport expenses are determined as follows:

$$\overline{P}^k(\omega_k) = \min_{x_{ij_k}} \sum_{i \in \omega_k} \sum_{j_k \in J^k} (c_{ij_k} + l_i^k) x_{ij_k} + \sum_{i \in \omega_k} T^k \equiv c^k(\omega_k) + \sum_{i \in \omega_k} T_i^k$$

under constraints

$$\sum_{i \in \omega_k} x_{ij_k} = b_{j_k}^k, \quad x_i^k \equiv \sum_{j_k \in J^k} x_{ij_k} \leqq a_i^k, \quad x_{ij_k} \geqq 0.$$

However, this assumption is inadmissible in planning a facilities allocation of various industries on the territory of any district, republic or country. A construction of different kind of facilities at one point leads as a rule by many reasons to a decrease of the total expenses due to common communications, fuller employment of labour resources, cheapening of housing and living facilities etc. The concentration of difficult kind of facilities at one point (the creation of territorial-industrial complexes) is the basis for the rise of new settlements and towns.

We denote by $\omega = (\omega_1, \dots, \omega_n)$ the totality of territorial-industrial complexes defined by an allocation of facilities of the k-th industry on the corresponding subsets of the points ω_k, $k = 1, 2, \dots, n$. Let $\sigma = \bigcup_{k=1}^{n} \omega_k$. For each $i \in \sigma$ one can determine the number n_i signifying the quantity of facilities of various industries located at one and the same i-th point. Thus, n_i, $1 \leqq n_i \leqq n$ is the amount of the subsets ω_k containing point i. The totality of facilities of various industries located at the i-th point defines the framework of the i-th territorial-industrial complex.

We consider the optimization problem of the allocation of territorial-industrial complexes assuming that with increasing by a unit the number n_i of facilities located at the i-th point the common expenses correspond to the i-th territorial industrial complex diminished by the value $D_i > 0$. Then for each ω one can determine

$$f(\omega) = f(\omega_1, \dots, \omega_n) = \sum_{k=1}^{n} \overline{P}^k(\omega_k) - \sum_{i \in \sigma} D_i(n_i - 1).$$

The problem consists in the determination of such $\alpha = (\alpha_1, \dots, \alpha_n)$ that $f(\alpha) = \min f(\omega_1, \dots, \omega_n)$ for all $\omega_k \subset I$, $k = 1, 2, \dots, n$. For $f(\omega)$ construct the approximating function $P(\omega)$:

$$f(\omega) \geqq \sum_{k=1}^{n} c_k(\omega_k) + \sum_{k=1}^{n} \sum_{i \in \omega_k} T_i^k - \sum_{i \in \sigma} D_i\left(n_i - \frac{n_i}{n}\right) =$$

$$= \sum_{k=1}^{n} c_k(\omega_k) + \sum_{k=1}^{n} \sum_{i \in \omega_k} T_i^k - \sum_{k=1}^{n} \sum_{i \in \omega_k} \frac{D_i(n-1)}{n} =$$

$$= \sum_{k=1}^{n} c_k(\omega_k) + \sum_{k=1}^{n} \sum_{i \in \omega_k} \left(T_i^k - \frac{D_i(n-1)}{n}\right) = \sum_{k=1}^{n} P^k(\omega_k) \equiv P(\omega),$$

$$P^k(\omega_k) = \overline{P}^k(\omega_k) - \sum_{i \in \omega_k} \frac{D_i(n-1)}{n}.$$

Thus $P(\omega)$ is one of the above given approximating function of the general kind for which the condition is valid. Determined for it Ω_0 we find

$$f(\tilde{\alpha}) = \min_{\omega \in \Omega_0} f(\omega).$$

Note that in this case the value $f(\omega)$ is determined by the simple formula:

$$f(\omega) = P(\omega) + \sum_{i \in \sigma} \frac{D_i(n - n_i)}{n}.$$

4. *Dynamic allocation problems with additional conditions*

$$\min_{x_{ij}^t} \left[\sum_{t \in \tau} \sum_{i \in I} \sum_{j \in J} c_{ij}^t \left(\sum_{S=1}^{t-1} x_{ij}^S, x_{ij}^t\right) + \sum_{t \in \tau} \sum_{i \in I} g_i^t \left(\sum_{S=1}^{t-1} x_i^S, x_i^t\right)\right]$$

under conditions (16).

If $g_i^t\left(\sum_{S=1}^{t-1} x_i^S, x_i^t\right)$ is of the form (17) and

$$c_{ij}^t\left(\sum_{S=1}^{t-1} x_{ij}^S, x_{ij}^t\right) = \begin{cases} \left(\bar{c}_{ij} + c_{ij}^t(x_{ij}^t)\right) \operatorname{sign}(x_{ij}^t), & \text{if } \sum\limits_{S=1}^{t-1} x_{ij}^S = 0 \\ c_{ij}^t(x_{ij}^t), & \text{if } \sum\limits_{S=1}^{t-1} x_{ij}^S > 0, \end{cases}$$

$c^t_{ij}(x^t_{ij})$ is the convex upward function, $f(\omega)$ will be of the form

$$f(\omega)=f(\omega_1,\dots,\omega_T)=\sum_{t\in\tau}\sum_{i\in\omega_t}D^t_i+\min_{x^t_{ij}}\Bigg\{\sum_{j\in J}\sum_{t\in\tau}\sum_{i\in\bigcup\limits_{S=1}^{t}\omega_S}(c^t_{ij}(x^t_{ij})+k^t_ix^t_{ij})+$$

$$+\sum_{j\in J}\sum_{i\in\bigcup\limits_{t\in\tau}\omega_t}\tilde{c}_{ij}\operatorname{sign}\left(\sum_{t\in\tau}x^t_{ij}\right)\Bigg\}$$

under conditions (16).

To determine $\min\limits_{\omega\in\Omega}f(\omega)$ (here Ω is the set of all $\omega=(\omega_1,\omega_2,\dots,\omega_T)$), $\omega_t\subset I$ for all $t\in\tau$ apply the approximation-combinatorial method. As an approximating function take the function of the form

$$P(\omega)=P(\omega_1,\dots,\omega_T)=\sum_{t\in\tau}\sum_{i\in\omega_t}D^t_i+$$

$$+\sum_{j\in J}\sum_{t\in\tau}\min_{i\in\bigcup\limits_{S=1}^{t}\omega_S}\left(c^t_{ij}(b^t_j)+k^t_ib^t_j+\frac{\tilde{c}_{ij}b^t_j}{\sum\limits_{t\in\tau}b^t_j}\right).$$

It is easy to note that $P(\omega)\leqq f(\omega)$ for all $\omega\in\Omega$ and $P(\omega)$ satisfies the condition $\mathcal{S}(\delta,\gamma)\leqq 0$ as $P(\omega)$ is of the form of the function from Problem 3. After the set Ω_0 is determined the question arises about the determination of the values of the function $f(\omega)$ for all $\omega\in\Omega_0$. Denote

$$\bar{\bar{c}}^t_{ij}(x^t_{ij})=c^t_{ij}(x^t_{ij})+k^t_ix^t_{ij},$$

(18)

$$f_j(\omega_1,\dots,\omega_T)=\min_{x^t_{ij}}\Bigg\{\sum_{t\in\tau}\sum_{i\in\bigcup\limits_{S=1}^{t}\omega_S}\bar{\bar{c}}^t_{ij}(x^t_{ij})+\sum_{i\in\bigcup\limits_{t\in\tau}\omega_S}\tilde{c}_{ij}\operatorname{sign}\Big(\sum_{t\in\tau}x^t_{ij}\Big)\Bigg\}$$

under conditions

$$\sum_{i\in\bigcup\limits_{S=1}^{t}\omega_S}x^t_{ij}=b^t_j,\ t\in\tau,\ x^t_{ij}\geqq 0,\ t\in\tau,\ i\in\bigcup_{S=1}^{t}\omega_S.$$

Then

$$f(\omega)=\sum_{t\in\tau}\sum_{i\in\omega_t}D^t_i+\sum_{j\in J}f_j(\omega).$$

The problem of finding the value of the function $f(\omega_1, \dots, \omega_T)$ reduces to the determination of the value of function $f_j(\omega_1, \dots, \omega_T)$ for all $j \in J$. Denote $v = \bigcup_{t \in \tau} \omega_t$, $v_t = \bigcup_{S=1}^{t} \omega_S$. It is evident that $v = \bigcup_{t \in \tau} v_t$. Suppose $d_{ij}^t(x_{ij}^t) = \bar{\bar{c}}_{ij}^t(x_{ij}^t)$ for all $t \in \tau$, $i \in v_t$; $d_{ij}(x_{ij}^t) = \infty$ for all $t \in \tau$, $i \in v \setminus v_t$. It can be proved that (18) is equivalent to the following problem: to find

$$\varphi_j(v) = \min_{x_{ij}^t} \left\{ \sum_{i \in v} \tilde{c}_{ij} \operatorname{sign}\left(\sum_{t \in \tau} x_{ij}^t \right) + \sum_{t \in \tau} \sum_{i \in v} d_{ij}^t(x_{ij}^t) \right\}$$

under conditions

$$\sum_{i \in v} x_{ij}^t = b_j^t, \ t \in \tau, \ x_{ij}^t \geqq 0, \ i \in v, \ t \in \tau.$$

Choose any subset $z \subset v$ in such a way that $\operatorname{sign}\left(\sum_{t \in \tau} x_{ij}^t \right) = 1$ for all $i \in z$, $\operatorname{sign}\left(\sum_{t \in \tau} x_{ij}^t \right) = 0$, for all $i \in v \setminus z$. Denote the minimum expenses in choosing z by

$$\Psi_j(z) = \sum_{i \in z} \tilde{c}_{ij} + \min_{x_{ij}^t} \sum_{t \in \tau} \sum_{i \in z} d_{ij}^t(x_{ij}^t)$$

under condition: $\sum_{i \in z} x_{ij}^t = b_j^t$ for all $t \in \tau$, $x_{ij}^t \geqq 0$. Then

$$f_j(\omega) = \varphi_j(v) = \min_{z \subset v} \Psi_j(z).$$

Note that $\Psi_j(z) = \sum_{i \in z} \tilde{c}_{ij} + \sum_{t \in \tau} \min_{i \in z} d_{ij}^t(b_j^t)$ and satisfies the condition $\mathcal{S}(\delta, \gamma) \leqq 0$. Thus the problem of determination of the value of function $f(\omega)$ on the fixed ω consists in solving n problems (18) by the method of consecutive calculations.

If there are constraints on capacities of production points the approximation-combinatorial method can be used, too. However, the calculation of the values of the function $f(\omega)$ for $\omega \subset \Omega_0$ gets significantly complicated and commensurable with the difficulty of solving the problems of Balinski [19].

In more detail these problems are described in the paper [16] where other cases are treated, too.

REFERENCES

[1] Gass, S. I., *Linear Programming: Methods and Applications*, 2nd ed., McGraw-Hill, New York, 1964. *MR* **33** #3772

[2] Bellman, R. E. and Dreyfus, S. E., *Applied Dynamic Programming*, Princeton University Press, Princeton, N. J., 1962. *MR* **25** #3791

[3] CHERENIN, V. P., *Solving Some Optimal Combinatorial Problems of the Optimal Planning by the Method of Consecutive Calculations*, Nauchno-metodicheskie materialy ekonomiko-matematicheskogo seminara LEMM AN SSSR, Vyp. 2, M., rotaprint Gipromez, 1962 (in Russian).

[4] LAWLER, E. L. and WOOD, D. E., Branch-and-bound methods: a survey, *Operations Res.* **14** (1966), 699–719. *MR* **34** #2338

[5] MURTY, K. C., Solving the fixed charge problems by ranking the extreme points, *Operations Res.* **16** (1968), 268–279.

[6] LEBEDEV, S. S., Methods of sequencing enumeration in linear programming problems with zero-one variables, *Matem. program.*, vyp. 5, Alma-Ata, Institut ekonomiki AN Kaz. SSR, 1969, 1–118 (in Russian).

[7] EMELICHEV, V. A. and KOMLIK, V. I., A solution of a discrete programming problem by the method of constructing a sequence of plans, *Dokl. Akad. Nauk SSSR* **188** (1969), 272–273 (in Russian). *MR* **40** #5254

[8] KHACHATUROV, V. R., On an approximative combinatorial method of solving the mathematical programming problems, Trudy III Zimnej shkoly po matematicheskomu programmirovaniyu, g. Drogobych., Vyp. 3 M., CEMI AN SSSR, 1970, 657–674 (in Russian).

[9] KHACHATUROV, V. R., Some matters of the method of consecutive calculations and its applications to the allocation problems. Dissertaciya kand. fiz.-mat. nauk, M., CEMI AN SSSR, 1968 (in Russian).

[10] KHACHATUROV, V. R., The general selection rules for solving a class of combinatorial problems. Materialy nauch. konf. po matem. i mekhan. AN Kaz. SSR, Alma-Ata, Institut ekonomiki AN Kaz. SSSR, 1968, 95–97 (in Russian).

[11] KHACHATUROV, V. R. and ZANCHEVA, F. N., Computerizing algorithms and program for determination the optimal and all of the variants close to it in the facilities allocation with unlimited capacities, Materialy nauch. konf. po matem. i mekhan. AN Kaz. SSR, Alma-Ata, Institut ekonomiki AN Kaz. SSR, 1968, 97–99 (in Russian).

[12] ASTACHOV, N. D., VESELOVSKI, V. E., SIGAL, I. KH. and KHACHATUROV, V. R., On the experience of solving the allocation problem by the modificated algorithm of consecutive calculations, Tezisy dokladov VI. Vsesoyuznoi Konferencii po ekstremalnym zadacham, Ch. I, Tallin, Izd-vo AN SSSR, 1973, 26–27 (in Russian).

[13] KHACHATUROV, V. R., Determination the optimal variant and all of variants close to it in the facilities allocation with limited capacities. *Izv. AN Kaz. SSR, Ser. fiz.-mat.*, 1976, N 3, 38-43 (in Russian).

[14] KHACHATUROV, V. R. and KALDYBAEV, S. U., Solving the distribution problem with zero-one variables by the approximative combinatorial method, Trudy v zimnej shkoly po matem. program. AN SSSR, 1973, 226–240 (in Russian).

[15] KHACHATUROV, V. R., The approximative combinatorial method and some of its applications, *USSR Computational Mathematics and Mathematical Physics*, v. 14, 1974, 1464–1487.

[16] KHACHATUROV, V. R. and ASTAKHOV, N. D., Dynamic allocation problems (models and solution methods), Matekon, Moscow, N 1, 1976.

[17] KHACHATUROV, V. R. and VESELOVSKY, V. E., Solution of large-scale allocation problems, Tezisy II. Vsesoyuznogo seminara "Chislennye metody nelinejnogo programmirovaniya", Charkov, 1976 (in Russian).

[18] VESELOVSKIĬ, V. E., Algorithms for solving of large-scale allocation problems, Matekon, Moscow, N 4, 1977. See *MR* **58** #4304

[19] BALINSKI, M. L., Fixed-cost transportation problems, *Naval. Res. Log. Quart.* **8** (1961), 41–54.

[20] KHACHATUROV, V. R., The stability of optimal values in problems of discrete programming, Optimization Techniques IFIP Technical Conference, Novosibirsk, 1974, Springer-Verlag, Berlin–Heidelberg–New York, 1975.

(Received January 18, 1984)

ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР
АКАДЕМИИ НАУК РОССИИ
УЛ. ВАВИЛОВА, 40
117 966 – МОСКВА ГСП-1
COMMONWEALTH OF INDEPENDENT STATES

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 261–266

ВНЕШНЕЕ ОСВЕЩЕНИЕ ДВУХСТОРОННЕГО КОНУСА

Е. БОННЕ и П. С. СОЛТАН

Пусть R^n — евклидово пространство размерности n с началом координат O. В этом пространстве нам понадобятся следующие конструкции.

Рассмотрим замкнутый выпуклый конус $C_1^n \subsetneqq R^n$ той же размерности n с вершиной O. Известно, что C_1^n представим в виде произведения $R^k \times C_1^\ell$, $k+\ell=n$, где R^k-подпространство пространства R^n, а C_1^ℓ-выпуклый конус с единственной вершиной O, обычно называемый выступающим конусом [1]. Легко проверить, что если C^n-выступающий конус, то в R^n существует гиперплоскость R^{n-1}, пересекающая C_1^n по ограниченному выпуклому множеству K^{n-1} размерности $n-1$. В случае, когда R^{n+1} — многогранник, то C_1^n назовем *многогранным* конусом. В частности, когда K^{n-1}-симплекс размерности $n-1$, то C_1^n назовем *n-гранным* конусом. Луч конуса C_1^n, исходящий из его вершины и проходящий через экстремальную точку выпуклого множества K^{n-1}, назовем *ребром* конуса C_1^n. Очевидно, у n-гранного конуса имеются в точности n ребер. Теперь, пусть C_1^n и C_2^n — выпуклые конусы пространства R^n с общей вершиной O, симметричные друг другу относительно этой вершины. Множество $C^n=C_1^n \cup C_2^n$ назовем двухсторонним конусом пространства R^n. Точку O назовем вершиной этого конуса. Для случая когда, например, C_1^n представим в виде $R_1^k \times C_1^\ell$, то для $C_2^n = R_2^k \times C_2^\ell$ — как симметричного конусу C_1^n, R_2^k просто совпадает с R_1^k. Поэтому $C^n=(R_1^k \cup R_2^k)\times(C_1^\ell \cup C_2^\ell)=R^k \times C^\ell$. Двусторонный конус C^ℓ назовем по аналогии с C_1^ℓ также выступающим. Через $\operatorname{int} C^n$ и $\operatorname{bd} C^n$ обозначим соответственно внутренность и границу конуса C^n.

Далее, пусть S — некоторая точка множества $R^n \setminus C^n$, а $x \in \operatorname{bd} C^n$ — произвольная граничная точка для C^n, отличная от вершины O. Будем говорить, что точка x внешне освещается точкой S, если открытый отрезок Sx дизъюнитем конусу C^n и луч Sx проходит через внутренность $\operatorname{int} C^n$ конуса C^n [2]. Точку O будем считать по определению внешне освещенной любой точкой из $R^n \setminus C^n$. Нас интересует следующая

1980 *Mathematics Subject Classification* (1985 Revision). Primary 52D20.
Key words and phrases. Theory of illumination, convex sets, outer illumination.

Akadémiai Kiadó, Budapest

Задача. Найти минимальное число точек пространства R^n, внешне освещающих в совокупности всю границу $\operatorname{bd} C^n$ конуса C^n.

Обозначим это число через $p(C^n)$.

Теорема. *Число $p(C^n)$ удовлетворяет неравенствам*

$$2 \leqq p(C^n) \leqq 2n,$$

причем для $n \geqq 3$ равенство $p(C^n) = 2n$ имеет место тогда и только тогда, когда C^n — двухсторонний n-гранный конус. Для $n = 2$ имеет место $p(C^n) = 2$.

Перед тем, как перейти к доказательству указанной теоремы, доказем несколько вспомогательных предложений.

Для формулировки этих предложений напомним следующее понятие [2, 4]. Будем говорить, что точка $x \in \operatorname{bd} C^n$ внутренне освещается точкой $S \neq x$ пространства R^n, эсли отрезок $[S, x]$ содержит внутренние точки конуса C^n.

Лемма 1. *Если точка $S \in R^n \setminus C^n$ внутренне (внешне) освещает точку $x_1 \in \operatorname{bd} C^n$, то S внешне (внутренне) освещает симметричную для x, точку $x_2 \in \operatorname{bd} C^n$.*

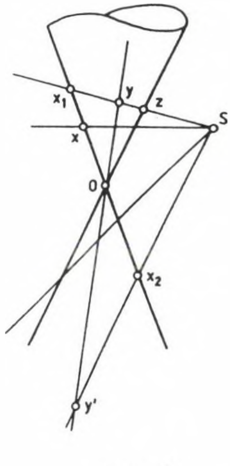


рис. 1

Доказательство. Покажем утверждение для первого случая, второй случай рассуждается аналогично. Пусть z такая точка отрезка $[S, x_1]$, что прямая zO параллельна прямой Sx_2 (рис. 1). Тогда любая прямая yO, где y находится между z и x_1, пересекает прямую Sx_2 в некоторой точке y' такой, что x_2 находится между S и y'. Если же точка y является внутренней точкой конуса C^n, то и точка y' также является внутренней точкой, ибо лучи Oy и Oy' — внутренние. Существование такой точки y легко следует из условия внутренней освещенности точки x_1.

Таким образом лемма 1 доказана.

Лемма 2. *Если некоторая точка x луча Ox, отличная от O, освещается (внутренне или внешне) точкой S, то этой же точкой S освещается (в том же смысле) и любая другая точка этого луча, отличная от O.*

Доказательство является очевидным (рис. 1).

Лемма 3. *Пусть K^m — ограниченное выпуклое тело линейного пространства R^m, а $S_1, S_2, \dots, S_p$ — точки границы $\operatorname{bd} K^m$ тела K^m,* *внутренне освещающие всю границу тела K^m. Тогда в $R^m \setminus K^m$ можно подобрать систему точек $S'_1, S'_2, \dots, S'_p$ также внутренне освещающие всю границу тела K^m.*

Доказательство. Обозначим через $L_1, L_2, \dots, L_p$ области освещенности границы $\operatorname{bd} K^m$ тела K^m соответственно источниками S_1, $S_2, \dots, S_p$. Очевидно, каждое L_i, $i = 1, 2, \dots, p$, является открытым множеством, а следовательно совокупность $\{L_1, L_2, \dots, L_p\}$ образует открытое покрытие компактного множества $\operatorname{bd} K^m$. Известно [3, A), стр. 98], что для такого покрытия существует такое замкнутое покрытие $L'_1, L'_2, \dots, L'_p$ множества $\operatorname{bd} K^m$, что $L'_i \subset L_i$, $i = 1, 2, \dots, p$. Далее, рассмотрим положительное число r, удовлетворяющее условию

$$r < \min_{1 \leqq i \leqq p} \min_{x \in L'_i} \|S_i - x\|.$$

Это число существует ввиду освещенности ($S_i \notin L'_i$) и замкнутости множества L'_i. Теперь для каждой точки $x \in L'_i$ рассмотрим интервал $]S_i, x[$. В силу определения внутренней освещенности, $]S_i, x[\subset \operatorname{int} K^m$. Пусть $y(x)$ — точка этого интервала, для которой $\|S_i - y(x)\| = r$, а $Y_i = \{y(x) : x \in L_i\}$ — совокупность всех таких точек. Множество Y_i есть подмножество сферы $\phi(S_i, r)$ с центром S_i и радиусом r (рис. 2). Ввиду замкнутости L'_i получаем, что и Y_i — замкнутое множество (как топологический образ компактного множества L'_i при центральном проектировании в сферу $\phi(S_i, r)$ с центра S_i). Заметим, что множества Y_i и $\operatorname{bd} K^m$ не имеют общих точек:

$$\operatorname{bd} K^m \cap Y_i \subset \operatorname{bd} K^m \cap \cup\{]S_i, x[: x \in L_i\} \subset \operatorname{bd} K^m \cap \operatorname{int} K^m = \emptyset.$$

Следовательно, из замкнутости множеств $Y_i, \operatorname{bd} K^m$ имеем, что расстояние между ними положительно. Пусть ε_i — число, выражающее это расстояние. Далее, рассмотрим шар $\sum_i(S_i, \varepsilon_i)$ радиуса ε_i с центром в точке S_i. Учитывая граничность точки S_i для множества K^m получаем, что множество $\sum_i(S_i \varepsilon_i)$ содержит точки, не принадлежащие телу K^n. Пусть S'_i — такая точка. Покажем, что S'_i внутренне освещает множество $L'_i \subset \operatorname{bd} K^m$.

В самом деле, пусть x — произвольная точка множества L'_i, и $y'(x)$ такая точка интервала $]S'_i, x[$, что $\|S_i - x\|/\|y(x) - x\| = \|S'_i - x\|/\|y'(x) - x\|$ (рис. 2). Пусть далее h' — общее значение этих частных. Тогда $\|y(x) - y'(x)\| < h' - \|S_i - S'_i\| \leqq h' \cdot \varepsilon_i < \varepsilon_i$. Следовательно, точка $y'(x)$ находится от множества Y_i на расстоянии меньше, чем ε_i. В силу выбора числа ε_i это означает, что $y'(x)$ — внутреняя точка тела K^m. Таким образом, S'_i внутренне освещает множество $L'_i \subset \operatorname{bd} K^m$, $i = 1, 2, \dots, p$. Лемма доказана.

Доказательство теоремы. Пусть C^n — произвольный двухсторонний конус пространства R^n с вершиной O. В силу отмечанного в самом начале имеем $C^n = R^k \times C^{n-k}$. Обозначим через R^{m+1}, линейное пространство, натянутое на C^{n-k}, $m+1=n-k$. Заметим, что для внешнего или внутреннего освещения границы C^n достаточно осветить границу (относительно R^{m+1}) конуса C^{m+1} в пространстве R^n. Ибо, если некоторая точка $x \in \operatorname{bd} C^{m+1}$ освещена какой-либо точкой S в пространстве R^{m+1}, то тогда той же точкой но уже в R^n, освещено и все линейное многообразие R^k_x, проходящее через точку x и лежащее на границе конуса C^n, где $R^k_x = R^k + x$.

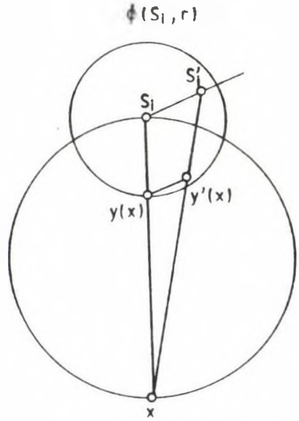


рис. 2

Теперь рассмотрим пространство R^{m+1} и проведем в нем гиперплоскость R^m, пересекающую одну из выпуклых компонент конуса C^n, например, конус C_1^{m+1}, по ограниченному выпуклому множеству K^m (рис. 3). В силу теоремы работы [4] тело K^m внутренне освещается не менее двумя и не более чем $m+1$ источниками света $S_1, S_2, \ldots, S_{m+1}$, лежащими на границе $\operatorname{bd} K^m$ тела K^m в пространстве R^m. Тогда, на основе леммы 3, для выпуклого тела $K^m \subset R^m$ существуют в $R^m \setminus K^m$, такие точки $S_1', S_2', \ldots, S_{m+1}'$, которые в совокупности внутренне освещают всю границу тела K^m. Далее, используя леммы 1, 2 и представление конуса $C^{m+1} = C_1^{m+1} \cup C_2^{m+1}$ получаем, что граница выпуклого конуса C_2^{m+1} (рис. 3) внешне освещается точками $S_1', S_2', \ldots, S_{m+1}'$. Повторяя рассуждения этого абзаца для выпуклого конуса C_2^{m+1}, равного C_1^{m+1}, получим, что столько же источников, пусть они будут $S_1'', S_2'', \ldots, S_n''$, достаточно для внешнего освещения границы конуса C_1^{m+1}. Таким образом, для конуса $C^{m+1} = C_1^{m+1} \cup C_2^{m+1}$ имеем неравенство

$$p(C^{m+1}) \leqq 2(m+1). \tag{1}$$

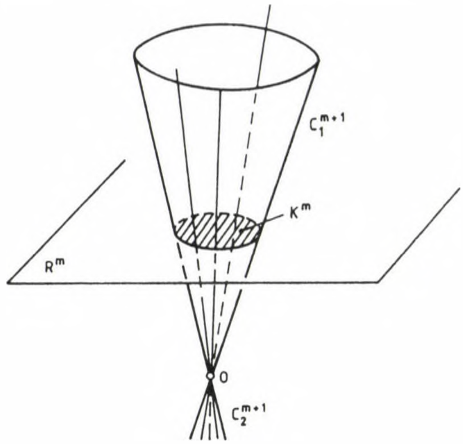


рис. 3

Далее покажем, что для $n=2$ неравенство (1) может быть улучшено. В самом деле, если конус C^{m+1} имеет размерность 2, т.е. $m=1$, то неравенство (1) получает вид: $p(C^2) \leqq 4$. Но в таком случае K^m — это просто отрезок, а S_1' и S_2' внешне освещает не только границу конуса C_2^2, но и границу конуса C_1^2, т.е. S_1'' и S_2'' — излишние источники.

Наконец, одна точка не может внешне освещать всей границы $\operatorname{bd} C^n$, потому что она внешне освещает некоторую точку $x_1 \in \operatorname{bd} C^n$, но тогда в силу леммы 1 не освещает симметрической для x_1 точки x_2. Итак, окончательно получаем

$$p(C^n) = 2 \text{ для } n=2, \text{ и } 2 \leqq p(C^n) \leqq 2(m+1) \leqq 2n \quad \text{для } \ n \geqq 3.$$

Наконец докажем, что для $n \geqq 3$ равенство $p(C^n) = 2n$ верно тогда и только тогда, когда C^n — двухсторонний n-гранный конус.

Пусть $C^n = C_1^n \cup C_2^n$ — двухсторонний n-гранный конус. Тогда $p(C^n) \leqq 2n$. Далее, для каждого ребра конусов $C_1^n\sqrt{nC_2^n}$ нужен один источник света и никакие два ребра этих конусов одним источником, размещенного в $R^n \setminus C^n$, внешне не освещаются. Но у конуса C_1^n имеется n ребер, а у объединения $C_1^n \cup C_2^n$ всего $2n$ ребер. Следовательно, $p(C^n) \geqq 2n$. Откуда получаем, что для двухстороннего n-гранника $p(C^n) = 2n$.

Обратно, пусть C^n — такой двухсторонний конус пространства R^n, что $p(C^n) = 2n$. Покажем, что C^n — двухсторонний n-гранник.

Первоначально заметим, что C^n — выступающий конус. В самом деле, если бы $C^n = R^k \times C^\ell$, где C^ℓ — выступающий конус, $k > 0$, то в силу формулы (1) мы имели бы

$$p(C^n) \leqq p(C^\ell) \leqq 2\ell < 2n,$$

что противоречит предложению $p(C^n) = 2n$.

Теперь пусть $C^n = C_1^n \cup C_2^n$ — как это оговорено выше. Рассмотрим выпуклый конус $C_1^n \subset R^n$. В R^n проведем гиперплоскость R^{n-1} таким образом, что пересечение $R^{n-1} \cap C_1^n$ есть ограниченное выпуклое множество K^{n-1}. Это возможно в силу того, что C^n — выступающий конус. Утверждается, что K^{n-1} — симплекс размерности $n-1$. В самом деле, если K^{n-1} не был бы симплексом, то в силу работы [4] вся граница тела K^{n-1} внутренне освещалась бы не более чем n источниками света, лежащими на границе тела K^{n-1} в пространстве R^{n-1}, а следовательно, согласно рассуждениям, проведенными выше при получении формулы (1), получили бы, что $p(C^n) < 2n$, что противоречит предложению. Итак, K^{n-1} симплекс размерности $n-1$, а следовательно, и C^n — двухсторонний n-гранник. Таким образом, теорема полностью доказана.

Замечание. Для случая двухстороннего гладкого конуса $C^n \subset R^n$ верно равенство $p(C^n) = 3$. Этот интересный факт был установлен М. Мюнхом (Дрезден, ГДР).

ЛИТЕРАТУРА

[1] Бурбаки, Н., *Топологические векторные пространства*, Издат. иностр. лит., Москва, 1959.

[2] Болтянский, В. Г. и Солтан, П. С., *Комбинаторная геометрия различных классов выпуклых множеств*, Штиинцэ, Кишинев, 1978.

[3] Понтрягин, Л. С., *Непрерывные группы*, Гос. издат. техн.-теорет. лит., Москва, 1954. *MR* **17**-171

[4] Солтан, П. С., Об освещении границы выпуклого тела изнутри [On the illumination of the boundary of a convex body from within], *Mat. Sb.* **57** (99) (1962), 443–448 (in Russian). *MR* **29** #518

[5] Bohne, E., Münch, M., Beleuchtungsgebiete auf singulären Drehquadriken bei geometrischer Zentralbeleuchtung aus mehreren Lichtquellen, *Beiträge Algebra Geom.* **18** (1984), 115–128. *MR* **85i**: 51032

(Поступило 14-ое апреля, 1986 г.)

Erhart Bohne

AM ANGER 22/15-06
D/O–8036 DRESDEN
FEDERAL REPUBLIC OF GERMANY

P. S. Soltan

MOLDAVIAN STATE UNIVERSITY
UL. MATEEVICIA 60
277 039 KISHINEV
MOLDOVA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 267–271

SUBDIRECTLY IRREDUCIBLE ALGEBRAS IN VARIETIES DEFINED BY EXTERNALLY COMPATIBLE IDENTITIES

J. PŁONKA

0. We consider algebras of a given type $\tau\colon F \to N$ where F is a set of fundamental operation symbols and N is the set of non-negative integers (see [1]). A term φ of type τ will be called non-trivial if it is different from a single variable. A term q will be called non-trivial unary if it is non-trivial and exactly one variable occurs in q. For a non-trivial term φ we denote by $\mathrm{ex}(\varphi)$ the most external fundamental operation symbol in φ, i.e. if $f \in F$ and $\varphi_0, \dots, \varphi_{\tau(f)-1}$ are terms of type τ then $\mathrm{ex}\big(f(\varphi_0, \dots, \varphi_{\tau(f)-1})\big) = f$.

An identity $\varphi = \psi$ of type τ is called externally compatible (see [2]) if it is of the form $x = x$ or both φ and ψ are non-trivial and $\mathrm{ex}(\varphi) = \mathrm{ex}(\psi)$. For every variety K of type τ we denote by $K_{\mathbf{ex}}$ the variety of type τ defined by all externally compatible identities from $\mathrm{Id}(K)$. Studying externally compatible identities seems to be interesting since if K is a variety, then the set of all externally compatible identities from $\mathrm{Id}(K)$ is an equational theory (closed under Birkhoff's derivation rules, see [6]). Put $F^* = \{f \in F : \tau(f) > 0\}$.

In this paper we find all subdirectly irreducible algebras from $K_{\mathbf{ex}}$ under the assumption that $F^* \neq \emptyset$ and for every $f \in F^*$ there exists a non-trivial unary term q_f such that $\mathrm{ex}(q_f) = f$ and the identity $q_f(x) = x$ belongs to $\mathrm{Id}(K)$.

This assumption is satisfied if K is a variety of groups, of rings with 1, of lattices, of Boolean algebras and many others. The results of this paper were announced on the conference in Bachotek, May 1987.

1. Preliminaries. Let us denote by $\mathrm{Ex}(\tau)$ the set of all externally compatible identities of type τ and denote by $K(\mathrm{Ex}(\tau))$ the variety defined by $\mathrm{Ex}(\tau)$. The following three properties (i)–(iii) were noticed in [2].

(i) the set $\mathrm{Ex}(\tau)$ is an equational theory. The following identities form an equational base of it:

$$f(x_0, \dots, x_{\tau(f)-1}) = f(y_0, \dots, y_{\tau(f)-1}) \qquad (f \in F).$$

(ii) $K(\mathrm{Ex}(\tau))$ is a non-degenerated variety such that in every algebra from $K(\mathrm{Ex}(\tau))$ the realization of each $f \in F$ is a constant.

(iii) For every variety K of type τ we have $K_{\mathbf{ex}} = K \vee K(\mathrm{Ex}(\tau))$.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 08BXX.

Key words and phrases. Subdirectly irreducible algebras, externally compatible identities, dispersion of an algebra, congruently non-isolated element.

In [4] and [5] a construction called a dispersion of an algebra was defined. Here we give a shortened definition of this construction (cf. [3]).

We say that an algebra $\mathfrak{A} = (A; F^{\mathfrak{A}})$ of type τ is a dispersion of an algebra $\mathfrak{I} = (I; F^{\mathfrak{I}})$ of type τ if there is a partition $\{A_i\}_{i \in I}$ of A and a family $\{0_f\}_{f \in F}$ of mappings from I into A such that for each $i \in I$ we have $0_f(i) \in A_i$; and if $a_k \in A_{i_k}$ $(k = 0, 1, \ldots, \tau(f) - 1)$ then

(iv) $f^{\mathfrak{A}}(a_0, \ldots, a_{\tau(f)-1}) = 0_f\big(f^{\mathfrak{I}}(i_0, \ldots, i_{\tau(f)-1})\big)$.

It was shown in [5] that

(v) If $\varphi(x_0, \ldots, x_{n-1})$ is a non-trivial term of type τ on variables $x_0, \ldots, x_{n-1}$, $\mathfrak{A}$ is a dispersion of $\mathfrak{I}$, $a_k \in A_{ik}$ $(k = 0, \ldots, n-1)$, then

$$\varphi^{\mathfrak{A}}(a_0, \ldots, a_{n-1}) = 0_{\mathrm{ex}(\varphi)}\big(\varphi^{\mathfrak{I}}(i_0, \ldots, i_{n-1})\big).$$

Assume that a variety K satisfies two conditions:

(a_1) There exists a non-trivial unary term q of type τ such that for each $f \in F$ the identity

$$q\big(f(x_0, \ldots, x_{\tau(f)-1})\big) = q\big(f(q(x_0), \ldots, q(x_{\tau(f)-1}))\big)$$

belongs to $\mathrm{Id}(K)$.

(a_2) For each $f \in F^*$ there exists a non-trivial unary term q_f of type τ such that $\mathrm{ex}(q_f) = f$ and the identity

$$q_f\big(q(f(x_0, \ldots, x_{\tau(f)-1}))\big) = f(x_0, \ldots, x_{\tau(f)-1})$$

belongs to $\mathrm{Id}(K)$.

It was proved in [5] (Corollary 4) that:

(vi) If K is a variety of type τ satisfying (a_1) and (a_2) then an algebra $\mathfrak{A}$ of type τ belongs to K_{ex} iff $\mathfrak{A}$ is a dispersion of an algebra from K.

2. Subdirectly irreducible algebras in K_{ex}. If A is a set and P is a partition of A, we shall denote by $E(P)$ the equivalence relation induced in A by P. Denote $\omega_A = E(\{\{a\} \mid a \in A\})$.

We have:

(vii) Let $\mathfrak{B} = (B; F^{\mathfrak{B}}) \in K(\mathrm{Ex}(\tau))$. Then $\mathfrak{B}$ is subdirectly irreducible iff $\mathfrak{B}$ has exactly two elements.

The proof is left to the reader.

Let $\mathfrak{A} = (A; F^{\mathfrak{A}})$ be an algebra. We denote by $\mathrm{Con}(\mathfrak{A})$ the set of all congruences on $\mathfrak{A}$.

We shall say that an element $a \in A$ is congruently non-isolated if for every congruence $\sim \in (\mathrm{Con}(\mathfrak{A}) \setminus \{\omega_A\})$ the congruence class $[a]_\sim$ of a is not 1-element.

Let K be a variety of type τ. Let us consider the following two conditions:

(c_1) $F^* \neq \emptyset$.

(c_2) For every $f \in F^*$ there exists a non-trivial unary term q_f such that $\mathrm{ex}(q_f) = f$ and the identity $q_f(x) = x$ belongs to $\mathrm{Id}(K)$.

THEOREM 1. *If a variety K satisfies conditions* (c_1) *and* (c_2) *then an algebra* $\mathfrak{A} \in K_{ex}$ *is subdirectly irreducible iff one of the following three conditions holds:*

(d_1) $\mathfrak{A}$ *belongs to* K *and is subdirectly irreducible.*

(d_2) $\mathfrak{A}$ *is a 2-element algebra from* $K(\mathrm{Ex}(\tau))$.

(d_3) $\mathfrak{A}$ *is a dispersion of an algebra* $\mathfrak{I} = (I; F^{\mathfrak{I}})$ *from* K, $|I| > 1$, *there exists* $i_0 \in I$ *such that* $|A_{i_0}| = 2$, *say* $A_{i_0} = \{0_1, 0_2\}$, i_0 *is congruently non-isolated in* $\mathfrak{I}$, $|A_i| = 1$ *for* $i \in (I \setminus \{i_0\})$ *and there exists a partition* $\{F_1, F_2\}$ *of* F^* *with* $F_1 \neq \emptyset \neq F_2$ *such that* $0_f(i_0) = 0_k$ *for* $f \in F_k$ $(k = 1, 2)$.

REMARK 1. Note that many varieties satisfy the assumptions of the theorem. For example we have $x \cdot (x \cdot x^{-1}) = (x^{-1})^{-1} = x$ in groups, $1 \cdot x = x$ in rings with 1, $x \vee x = x \wedge x = x$ in lattices, $(x')' = x$ in Boolean algebras.

PROOF OF THEOREM 1 (NECESSITY). Let $\mathfrak{A} = (A; F^{\mathfrak{A}}) \in K_{ex}$. By (vi), $\mathfrak{A}$ is a dispersion of an algebra $\mathfrak{I} = (I; F^{\mathfrak{I}}) \in K$. In fact, conditions ($c_1$) and ($c_2$) imply ($a_1$) and ($a_2$) since it is enough to take as q from (a_1) an arbitrary q_f $(f \in F^*)$.

1° If $|A_i| = 1$ for each $i \in I$ then $\mathfrak{A} \in K$ since each 0_f sets up an isomorphism of $\mathfrak{I}$ and $\mathfrak{A}$. So $\mathfrak{A}$ is subdirectly irreducible iff (d_1) holds.

2° If $|I| = 1$ then $\mathfrak{A} \in K(\mathrm{Ex}(\tau))$ and by (vii), $\mathfrak{A}$ is subdirectly irreducible iff (d_2) holds.

3° Assume $|I| > 1$ and there are $i \in I$ such that $|A_i| > 1$. For each such i we define a relation R_i in $\mathfrak{A}$ putting for $a, b \in A$:

$$aR_ib \quad \text{if} \quad a = b \quad \text{or} \quad a, b \in A_i.$$

By (iv), R_i is a congruence of $\mathfrak{A}$ different from ω_A.

Now if there are two different $i, j \in I$ such that $|A_i| \neq 1 \neq |A_j|$ then $\mathfrak{A}$ is subdirectly reducible since $R_i \cap R_j = \omega_A$.

4° Assume $|I| > 1$ and there is exactly one $i_0 \in I$ for which $|A_{i_0}| > 1$. If there is a congruence R on $\mathfrak{I}$ different from ω_I such that $|[i_0]_R| = 1$ then $\mathfrak{A}$ is subdirectly reducible. In fact, let us denote by a_i the unique element of A_i for $i \in I \setminus \{i_0\}$. We define in $\mathfrak{A}$ a relation R' putting

$$xR'y \text{ if } x = y \text{ or } x = a_i,\ y = a_j,\ i \neq i_0 \neq j \text{ and } iRj.$$

Then R' is a congruence on $\mathfrak{A}$ by (iv) and $R' \neq \omega_A$. Consequently $R_{i_0} \cap R' = \omega_A$. Thus in this case if $\mathfrak{A}$ is subdirectly irreducible then i_0 must be congruently non-isolated.

5° Let $|I| > 1$ and $|A_{i_0}| > 2$. For each $a \in A_{i_0}$ denote by $R(a)$ the relation on A defined by:

$$xR(a)y \quad \text{if} \quad x = y \quad \text{or} \quad x, y \in (A_{i_0} \setminus \{a\}).$$

Then every $R(a)$ is a congruence on $\mathfrak{A}$ different from ω_A and $\bigcap_{a \in A_{i_0}} R(a) = \omega_A$. So $\mathfrak{A}$ is subdirectly reducible.

6° Let $|I|>1$, $|A_{i_0}|=2$, $A_{i_0}=\{0_1,0_2\}$ and $0_f(i_0)=0_2$ for every $f\in F^*$. Denote $R''=E(\{A\setminus\{0_1\},\{0_1\}\})$. Then R'' is a congruence in $\mathfrak{A}$ different from ω_A and $R_{i_0}\cap R''=\omega_A$. So $\mathfrak{A}$ is subdirectly reducible.

Proof of the sufficiency. For (d_1) and (d_2) the proof is trivial. Assume that (d_3) holds. To prove that $\mathfrak{A}$ is subdirectly irreducible it is enough to show that R_{i_0} is the unique atom of $\mathrm{Con}(\mathfrak{A})$. Let $\sim\in\mathrm{Con}(\mathfrak{A})\setminus\{\omega_A\}$. It cannot happen that $|[0_1]_\sim|=|[0_2]_\sim|=1$, since then denote by S the relation in I defined by:

$$iSj \quad\text{if}\quad i=j \quad\text{or}\quad i\neq i_0\neq j \quad\text{and}\quad a_i\sim a_j.$$

So S is a congruence on $\mathfrak{I}$ different from ω_1 and $|[i_0]_S|=1$. Consequently i_0 is congruently isolated — a contradiction.

Assume $|[0_1]_\sim|>1$, $a_j\in[0_1]_\sim$ for some $j\neq i_0$ and $0_2\notin[0_1]_\sim$. By the assumptions, there exists $f\in F_2$. Then $q_f(0_1)=0_f(i_0)=0_2$ and $q_f(a_j)=0_f(j)=a_j$ by (v). So $\sim$ is not a congruence. Thus $0_1\sim 0_2$ for every $\sim\in\mathrm{Con}(\mathfrak{A})\setminus\{\omega_A\}$.

By Birkhoff's subdirect decomposition theorem we conclude:

Corollary 1. *If K is a variety of algebras satisfying (c_1) and (c_2) then every algebra from K_{ex} is isomorphic to a subdirect product of algebras of the form (d_1), (d_2) or (d_3).*

For two varieties K_1 and K_2 of type τ we denote by $K_1\otimes K_2$ the class of all algebras $\mathfrak{A}$ such that there exist $\mathfrak{A}_1\in K_1$ and $\mathfrak{A}_2\in K_2$ such that $\mathfrak{A}$ is isomorphic to a subdirect product of $\mathfrak{A}_1$ and $\mathfrak{A}_2$. Obviously we always have $K_1\otimes K_2\subsetneq K_1\vee K_2$.

Let $\tau_1: F\to N$ be a type of algebras such that $F=\{f\}\cup\{c_r\}_{r\in R}$, $\tau_1(f)>0$ and $\tau_1(c_r)=0$ for all $r\in R$.

Corollary 2. *If K is a variety of type τ_1 and there exists a non-trivial unary term q such that the identity $q(x)=x$ belongs to $\mathrm{Id}(K)$ then $K_{\mathrm{ex}}=K\otimes K(\mathrm{Ex}(\tau_1))$.*

Indeed, we can use Theorem 1 and there are no algebras of the form (d_3) in K_{ex}.

Example 1. If K is a variety of groups with fundamental operation symbols $\cdot$ and 1 satisfying $x^n=1$ then $K_{\mathrm{ex}}=K\otimes K(\mathrm{Ex}(\tau_1))$. In fact, we can take $q(x)=x^{n+1}$ in Corollary 2.

Example 2. If K is a variety of idempotent groupoids then $K_{\mathrm{ex}}=K\otimes K(\mathrm{Ex}(\tau_1))$.

Corollary 3. *If K is a variety of type τ such that $|F^*|>1$, K satisfies (c_1) and (c_2) and there exists an algebra $\mathfrak{I}$ in K such that $|I|>1$ and there is a congruently non-isolated element in $\mathfrak{I}$, then $K_{\mathrm{ex}}\neq K\otimes K(\mathrm{Ex}(\tau))$.*

Proof. In fact, we have in K_{ex} algebras of the form (d_3) belonging neither to K nor to $K(\mathrm{Ex}(\tau))$.

EXAMPLE 3. If K is a non-degenerated variety of lattices or the variety of Boolean algebras then $K_{\mathrm{ex}} \neq K \otimes K(\mathrm{Ex}(\tau))$.

In fact, we have two-element algebras in K and the assumption (c_1) and (c_2) are satisfied, by Remark 1.

EXAMPLE 4. If K is a non-degenerated variety of groups with fundamental operation symbols $\cdot$ and $^{-1}$ then $K_{\mathrm{ex}} \neq K \otimes K(\mathrm{Ex}(\tau))$.

Indeed, take as $\mathfrak{T}$ a group generated by one element different from the identity.

COROLLARY 4. *Let K be the variety of all distributive lattices. For every positive integer n there exists in K_{ex} a subdirectly irreducible algebra of cardinality $2^n + 1$.*

PROOF. Let $\mathcal{L}_2$ be a two-element lattice. Since every finite Boolean lattice is congruence uniform, so each element in the direct power $\mathcal{L} = \mathcal{L}_2^n$ is congruently non-isolated and we can construct a subdirectly irreducible algebra of the form (d_3) being a dispersion of $\mathcal{L}$.

REFERENCES

[1] BURRIS, S. and SANKAPPANAVAR, H. P., *A Course in Universal Algebra*, Graduate Texts in Mathematics, Vol. 78, Springer-Verlag, New York, 1981. *MR* **83k**: 08001

[2] CHROMIK, W., On externally compatible varieties of algebras (to appear).

[3] HAŁKOWSKA, K., On dispersion operator on classes of algebras, *Polish Acad. Sci. Inst. Philos. Sociol. Bull. Sect. Logic* **17** (1988), 67–69. (Not in *MR*.)

[4] PŁONKA, J., *P*-compatible identities of universal algebras, *Beiträge zur Algebra, Algebra-Tagung Halle* 1986, Wissensch. Beitr., No. 33, Martin-Luther-Univ. Halle-Wittenberg, Halle (Saale), 1987, 209–211. *MR* **88g**: 00020

[5] PŁONKA, J., *P*-compatible identities and their applications to classical algebras, *Math. Slovaca* (in print).

[6] TARSKI, A., Equational logic and equational theories of algebras, *Contributions to Math. Logic* (Colloquium, Hannover 1966), North-Holland, Amsterdam, 1968, 275–288. *MR* **38** #5692

(Received July 24, 1987)

MATHEMATICAL INSTITUTE OF THE
POLISH ACADEMY OF SCIENCES
UL. KOPERNIKA 18
PL–51–617 WROCŁAW
POLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 273–277

ON THE ISOMETRIC DISSECTION PROBLEM FOR CONVEX SETS

M. EDELSTEIN, K. JOHNSON and A. C. THOMPSON

§ 0. Introduction

A set S in a metric space X is said to be decomposable (under a given family of isometries) or, more precisely, 2-decomposable if it is the disjoint union of two subsets which are mutually isometric (by isometries from the family). Throughout this paper the family will be the group of all isometries of the space onto itself.

Recently E. Hertel proved [7] that every closed, bounded convex set in Euclidean (finite dimensional) space is indecomposable. He also asked whether this is true in more general spaces. The purpose of this paper is to show that this is so. We prove that every closed, bounded convex set in a strictly convex, reflexive Banach space is indecomposable; and that the same is true in all finite dimensional Banach spaces, whether strictly convex or not. Clearly, the result of Hertel and its generalization to all strictly convex, finite dimensional Banach spaces follow as corollaries. For several reasons, however, we include an independent proof of the latter fact. First, because we consider our proof simpler and more direct (no induction is involved) even in the Euclidean case. Secondly, because the proof for all finite dimensional spaces depends on the strictly convex (actually Euclidean) case and it seems preferable to us that this not rely on the infinite dimensional case. Thirdly, because it sheds light on the conceptual difference between the arguments used in the finite dimensional and the infinite dimensional situation. At a certain stage of both proofs we exhibit an invariant line. In the one case this is done by straightforward classical methods of linear algebra; in the other we need to employ nonlinear (and non-constructive) methods of functional analysis.

In the second section (dealing with the infinite dimensional case) we exhibit examples of decomposable sets in ℓ^1. Examples of a similar nature are possible in c_o, ℓ^∞ and $C[0,1]$. Finally, in §3 we give a fuller discussion of the history of the problem.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 52A45; Secondary 46B20, 52A20.

Key words and phrases. Dissection of convex sets, tiling of convex sets, isometry, strictly convex spaces, reflexive spaces.

§1. Finite dimensional space

PROPOSITION 1. *If S is a closed, bounded convex set in a strictly convex, finite dimensional, normed linear space E, then S is indecomposable.*

PROOF. Suppose that, contrary to the assertion, S is decomposable so that S is the disjoint union of sets A, B and that $F\colon E \to E$ is an isometry mapping A onto B. Then (cf. [2]) F is affine, i.e. $F(x) = T(x) + a$ $(x \in E)$, where T is a linear isometry and $a \in E$.

To prove the proposition it suffices to show that either a line or a point remains invariant under F. In the latter case, if $F(x_o) = x_o$, let s_o be the (necessarily unique by strict convexity) nearest point of S to x_o. Since A and B are disjoint, s_o cannot also be fixed under F. However, since F and F^{-1} preserve distance, $\|F(s_o) - x_o\| = \|s_o - x_o\| = \|F^{-1}(s_o) - x_o\|$. Thus neither $F(s_o)$ nor $F^{-1}(s_o)$ is in S which is absurd since s_o is in either A or B.

Similarly, if a line L in E exists such that $F[L] = L$ then the set S_o of nearest points of S to L is a line segment (parallel to L) and $F[A \cap S_o] = B \cap \cap S_o$. Here the known (cf. §3) one-dimensional version of the proposition applies to establish the truth of the assertion.

To complete the proof we now establish the claim that there is either a line or a point which is invariant under F. First, suppose that $T - I$ is invertible. In this case one verifies directly that $x_o = (T - I)^{-1}(-a)$ is fixed under F. Next, if $T - I$ is not invertible, let U be the null space of $T - I$, i.e. $U = \{x \mid Tx = x\}$, and let W be the range of $T - I$. Then $E = U \oplus W$ and so a can be represented as $a = u_o + w_o$ with $u_o \in U$ and $w_o \in W$. Now W is invariant under T and, restricted to W, $T - I$ is invertible. Hence, if we define $F_w\colon W \to W$ by setting $F_w(x) = Tx + w_o$, then, as above, F_w has the fixed point $x_o = (T - I)^{-1}(-w_o)$. Therefore, $F(x_o + \lambda u_o) = T(x_o + \lambda u_o) + a = = (Tx_o + w_o) + (\lambda T u_o + u_o) = F_w(x_o) + (\lambda + 1)u_o = x_o + (\lambda + 1)u_o$, showing that the line $L = \{x \in E \mid x = x_o + \mu u_o, \mu \in \mathbf{R}\}$ is invariant under F.

THEOREM 1. *If S is a closed, bounded, convex set in a finite dimensional normed linear space X then S is indecomposable.*

PROOF. Let C be the unit ball of E. Let G be the group of all isometries of E and G_0 the subgroup of those which leave the origin fixed.

It is well-known [1] that there exists a unique ellipsoid (the Loewner ellipsoid) D of minimal volume circumscribed about C. If $F \in G_0$ then it is necessarily linear and so $F(D)$ is again an ellipsoid circumscribed about $F(C) = C$. Now, since $F(C) = C$, F clearly preserves volume ($|\det F| = 1$) and so $F(D)$, having the same volume as D is, by the uniqueness of D, equal to D. Thus F is also an isometry of the Euclidean space $(\mathbf{R}^n, D)$ whose unit ball is D. In addition, all translations are isometries in both metrics. This shows that G is a subgroup of the group of isometries of $(\mathbf{R}^n, D)$.

Since Proposition 1 shows that S is indecomposable by the group of Euclidean isometries, *a fortiori* it is indecomposable by G.

§ 2. Banach space

In the infinite-dimensional case the situation is more complicated, as the following example illustrates. In the Banach space $\ell^1(\mathbf{R})$ of absolutely summable sequences in $\mathbf{R}$, let S be the closed convex hull of all the vectors in the standard basis of that space, and let F be the unilateral shift, i.e. $F((x_1, x_2, \dots, x_n \dots)) = (0, x_1, x_2, \dots x_n \dots)$. Let A be the subset of S consisting of vectors $(y_1, y_2, \dots, y_n, \dots)$ whose first nonzero coordinate is y_n with n odd and let $B = S \setminus A$. Clearly $F[A] = B$ so that S is indeed, decomposable.

The above example may lead one to suppose that the non-invertibility of the isometry is important for such an example. (In the Introduction we said we would be dealing with the *group* of onto isometries.) In the next Proposition, however, we show, by using a bilateral shift, that this is not so.

PROPOSITION 2. *With S denoting the same set as in the above example, there exists an invertible isometry under which S is decomposable.*

PROOF. Let $\{e_n : n = 1, 2, \dots\}$ be the standard basis in $\ell^1(\mathbf{R})$ and define F on ℓ^1 as follows: $F(e_{2i-1}) = e_{2i+1}$ $(i = 1, 2, 3, \dots)$, $F(e_{2i}) = e_{2i-2}$ $(i = 2, 3, 4, \dots)$, $Fe_2 = e_1$; extend F to the whole of ℓ^1 linearly and continuously. Now F maps S isometrically onto S. Thus if $x \in S$, the orbit $\{F^n(x) : n = 0, \pm 1, \pm 2, \dots\}$ lies in S. Form the set A_0 by choosing one point from each orbit. Finally let $A = \bigcup_{n=-\infty}^{\infty} F^{2n}(A_0)$ and $B = \bigcup_{n=-\infty}^{\infty} F^{2n+1}(A_0)$. Clearly $F[A] = B$, and, again, S is decomposable under this isometry.

REMARK. The above construction is based on a general one used in [4]. Since F is a free isometry (in the sense of [4]) S allows decompositions into m isometric pieces $2 \leqq m \leqq \aleph_o$.

The examples above show that we cannot expect Theorem 1 to extend to all Banach spaces. We do, however, have:

THEOREM 2. *Let X be a strictly convex, reflexive Banach space and let S be a closed, bounded, convex subset of X. If A is a nonempty subset of S and if F is an isometry of X such that $S = A \cup F(A)$, then $A \cap F(A)$ is nonempty; i.e. S is indecomposable.*

In this theorem, and its proof, as well as in the following lemmas, we employ standard terminology and facts from Banach space theory (cf. e.g. [2]).

LEMMA 1. *Under the hypotheses of Theorem 2 the set $D = (F - I)(S)$, where I is the identity map on X, is convex and weakly compact.*

PROOF. The isometry F is affine, i.e., $F(x) = Tx + a$ where T is linear and a is a fixed vector. (To see this one may appeal either to the strict convexity or, since F is onto, the Mazur–Ulam Theorem, cf. [2].) Thus D is a translate of $(T - I)[S]$. Since $T - I$ is a bounded linear operator it is obvious that $(T-I)[S]$ is bounded and convex and, therefore, so is D. It now suffices to show that $(T-I)[S]$ is closed. Let $\{y_n\}$ be a sequence in $(T-I)[S]$ converging to y in X, and let $\{x_n\}$ in S be a sequence such that $(T-I)(x_n)=y_n$ for $n=1,2,3,\ldots$. Choose $\{x_{n_i}\}$ as a subsequence of $\{x_n\}$ which converges weakly to x, say. Since x lies in S it suffices to show that $y= =(T-I)(x)$. Given any $f\in X^*$ we have that $f(Tx_{n_i})=(T^*f)(x_{n_i})$ converges to $(T^*f)(x)=f(Tx)$. Hence $f((T-I)x_{n_i})$ converges to $f((T-I)(x))$. This being true for all $f\in X^*$ it follows that $(T-I)x=y$.

LEMMA 2. *Under the hypotheses of Theorem 2 if $F(x)\neq x$ for all $x\in S$ then there is a straight line L which intersects S and is mapped into itself by F.*

PROOF. The weakly compact convex set D contains a member z of least norm; i.e. $\|z\|=\inf\{\|x\|:x\in D\}$. Let $x\in S$ be such that $z=F(x)-x$. We consider two cases 1) $F(x)\in S$ and 2) $F(x)\notin S$. In the first case let $u=F(x)$ and $v=(u+x)/2$. We have $\|F(v)-v\|\geqq\|F(x)-x\|$ since $v\in S$. On the other hand

$$\|F(v)-v\|\leqq\|F(v)-F(x)\|+\|F(x)-v\|=\|v-x\|+\|F(x)-v\|=\|F(x)-x\|.$$

Thus, by strict convexity, u, v and $F(v)$ are collinear. Similarly, in case 2), let $u=F^{-1}(x)$ and $v=(u+x)/2$. Here again $\|F(v)-v\|\geqq\|F(x)-x\|$ and, since $\|F(v)-x\|=\|F(v)-F(u)\|=\|v-u\|$, the reverse inequality, $\|F(v)-v\|\leqq \leqq\|F(x)-x\|$, readily follows. Thus, as in case 1), x, v and $F(v)$ must be collinear.

Let L be the line through u and v. Since $v=\frac{1}{2}(u+x)$, x is also on L. In case 1), $u=F(x)$ and $F(v)$ are on L and in case 2) $x=F(u)$ and $F(v)$ are on L; in both cases $F[L]\subseteqq L$.

PROOF OF THEOREM 2. We may assume that F has no fixed points in S. Choosing L as in Lemma 2, we see that the result follows from the one-dimensional case.

REMARK. Reflexivity is needed in Theorem 2 only because we prefer to confine ourselves to closed and bounded convex sets. The proof, of course, holds for any weakly compact set in a strictly convex Banach space.

§ 4. Historical notes and open problems

The question of the indecomposability of the unit ball in E^2 occurred as Problem 51 in Elemente der Mathematik posed by van der Waerden ([11]). A solution for the more general case of a compact, strictly convex subset of E^2 was presented by Puppe in the same journal ([9]). Another proof can be found in [6]. Hertel [7] deals with closed, convex sets in n-dimensional Euclidean space. The case $n=1$ was investigated by Gustin in 1951 ([5]), who established the more general result that $[0,1]$ cannot be decomposed into m mutually isometric subsets for any $m>1$. The proof of this more general result is quite involved, Hertel [7] gives a different, and shorter proof of the result. An easy proof of the case $m=2$ appeared in [10].

The question of what happens when 2-decomposability is replaced by m-decomposability (m finite) is largely unresolved. The one-dimensional case is solved, as mentioned above. For $n > 1$, however, there are only sporadic results. For $X = \mathsf{E}^2$, $m = 3$, indecomposability has been established when S is a square or equilateral triangle by J. Currie (unpublished), but not for any other choice of S. In particular the case in which S is the unit ball is unsolved. For $n > 2$ and S the unit ball in the Banach space being considered, S is known to be m-indecomposable provided $m \leqq n$ [4]. The Euclidean case of this result was first proved by Wagon [13] whose paper also gives an excellent survey of the status of the problem at that time.

If m is allowed to be infinite (and, in particular, countable) then the character of the problem changes considerably. For example, von Neumann [12] showed that an interval (closed, open or neither) is countably decomposable. This was extended in finite dimensional spaces by Dekker and deGroot [3] and by Mycielski [8]; and in certain infinite dimensional spaces by Edelstein [4].

REFERENCES

[1] Danzer, L., Laugwitz, D. and Lenz, H., Über das Löwnersche Ellipsoid und sein Analogon unter den einem Eikörper einbeschriebenen Ellipsoiden, *Arch. Math.* **8** (1957), 214–219. *MR* **20** #1283

[2] Day, M. M., *Normed Linear Spaces*, 3rd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 21, Springer-Verlag, New York, 1973. *MR* **49** #9588

[3] Dekker, Th. J. and de Groot, J., Decompositions of a sphere, *Fund. Math.* **43** (1956), 185–194. *MR* **19**-1068

[4] Edelstein, M., Isometric decompositions, *J. London Math. Soc.* (2) **37** (1988), 158–163. *MR* **89a**: 46046

[5] Gustin, W., Partitioning an interval into finitely many congruent parts, *Ann. of Math.* (2) **54** (1951), 250–261. *MR* **13**-217

[6] Hadwiger, H. and Debrunner, H., *Combinatorial Geometry in the Plane*, V. Klee trans., Holt, Rinehart and Winston, New York, 1964. *MR* **29** #1577

[7] Hertel, E., Disjunkte Pflasterungen konvexer Körper, *Studia Sci. Math. Hungar.* **21** (1986), 379–386. *MR* **88i**: 52020

[8] Mycielski, J., On the paradox of the sphere, *Fund. Math.* **42** (1955), 348–355. *MR* **18**-324

[9] Puppe, D., *Elem. Math.* **4** (1949), 141.

[10] Raindrop, M. E., Solution to problem 284, *Canad. Math. Bull.* **24** (1981), 382–383.

[11] van der Waerden, B. L., *Elem. Math.* **4** (1949), 18.

[12] von Neumann, J., Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen, *Fund. Math.* **11** (1928), 230–238. *Jb. Fortschritte Math.* **54**, 96

[13] Wagon, S., Partitioning intervals, spheres and balls into congruent pieces, *Canad. Math. Bull.* **26** (1983), 337–340. *MR* **84g**: 52025

(Received September 27, 1988)

DEPARTMENT OF MATHEMATICS, STATISTICS AND
COMPUTING SCIENCE
DALHOUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA
B3H 3J5
CANADA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 279–286

ON MEET-DISTRIBUTIVE LATTICES

M. STERN

Abstract

It is shown that meet-distributive lattices are both strong (in the sense of Faigle) and consistent (in the sense of Kung), but not balanced (in the sense of Reuter). In fact, we prove that a meet-distributive lattice is balanced if and only if it is distributive. These results are derived from more general theorems on lower semimodular lattices and on locally modular lattices.

1. Introduction

In recent times, there is a rebirth of a class of lattices which were discovered by Dilworth [6] in 1940 and subsequently rediscovered several times by several authors and given several names. In recent papers these lattices are called meet-distributive (see Edelman [7], [8], Edelman–Jamison [9]). Their duals are also called locally free lattices (see Crapo [3]). The story of the frequent rediscoveries of this concept is narrated in Monjardet [17] (see also Edelman [8]). The present revival of the subject matter is largely due to the fact that meet-distributive lattices or their duals (i.e. locally free lattices) are shown to be closely linked to two external objects, that is, objects outside of lattice theory: these are the abstract convex geometries (see Edelman [8], Edelman–Jamison [9]) and the greedoids in the sense of Korte–Lovász [13] (for the special greedoids related to locally free lattices, cf. Crapo [3]; for a general introduction to greedoid theory[1] we refer to Björner–Ziegler [2]).

Aside from these external links, meet-distributive lattices have also a number of interesting intrinsic lattice theoretic properties.

In the present paper, we clarify the connection of meet-distributive lattices to the recently introduced concepts of strongness (see Faigle [10]), consistence (see Kung [14]), and to the property of being balanced (see Reuter [18]).

In particular, we shall see that a meet-distributive lattice is always both strong (Corollary 5) and consistent (Corollary 8) but not balanced, in gen-

1980 *Mathematics Subject Classifications.* Primary 06C10; Secondary 06B99.

Key words and phrases. Meet-distributive lattice, locally free lattice, strong, balanced, consistent, upper semimodular, lower semimodular, locally modular.

[1]Cf. also B. Korte, L. Lovász and R. Schrader, *Greedoids*, Springer-Verlag, Berlin, 1991.

eral. We show that a meet-distributive lattice is balanced if and only if it is distributive (Corollary 12).

Let us recall some notions. All lattices are assumed to be of finite length, although there are several deep extensions of some of the subsequent assertions to broader classes of lattices.

If L is a lattice of finite length, we denote its dual by L^d. We write $c \prec d$ $(c, d \in L)$ if c is a lower cover of d.

An element $u \in L$ is called join-irreducible if it has exactly one lower cover u'. By $J(L)$ we denote the set of all join-irreducibles of L.

An element $s \in L$ is said to be meet-irreducible if it has exactly one upper cover which we denote by s^*. Thus, for our purpose, we do not consider the least element 0 (the greatest element 1) to be join-irreducible (meet-irreducible).

A lattice L of finite length is said to be (upper) semimodular if, for $x, y \in L$,

$$x \wedge y \prec y \quad \text{implies} \quad x \prec x \vee y.$$

L is called lower semimodular if, for $x, y \in L$,

$$x \prec x \vee y \quad \text{implies} \quad x \wedge y \prec y.$$

For general information on lattice theory we refer to Crawley–Dilworth [5] and Grätzer [12].

2. Dual local modularity and meet-distributive lattices

If L is a lattice of finite length and $b \in L$, we denote by b^+ the join of all elements covering b; dually, b_+ denotes the meet of all elements covered by b.

According to Crawley–Dilworth [5], Ch. 7, p. 50, we say that the lattice L is locally modular if the interval $[b, b^+]$ is a modular sublattice for all $b \in L$.

A finite lattice L is called locally free (see Crapo [3]) if the interval $[b, b^+]$ is a Boolean sublattice for all $b \in L$.

It is clear that any locally modular lattice is upper semimodular. Thus we have the following implications:

(+) locally free $\Longrightarrow$ locally modular $\Longrightarrow$ upper semimodular.

It is easy to see that these implications are non-reversible, in general. Dualizing the concept of a locally free lattice we obtain meet-distributive lattices: A finite lattice L is called meet-distributive if the interval $[b_+, b]$ is a Boolean sublattice for all $b \in L$. A lattice L of finite length is said to be dually locally modular if the dual lattice L^d is locally modular. Dually to (+) we obtain then the following non-reversible implications:

(++) meet-distributive $\Longrightarrow$ dually locally modular $\Longrightarrow$ lower semimodular

Meet-distributive lattices have a number of interesting lattice theoretic properties characterizing them. Among others, we have

THEOREM 1 (Dilworth [6], Avann [1], see also Edelman [8]). *For a finite lattice L, the following conditions are equivalent:*

(i) *L is meet-distributive;*

(ii) *L is lower semimodular and every modular sublattice of L is distributive;*

(iii) *Every $x \in L$ can be uniquely expressed as the join of a minimal set of join-irreducibles.*

In fact, it was the arithmetical property (iii) of finite distributive lattices which stimulated research to characterize the class of those finite lattices having this arithmetical property.

An example of a non-distributive meet-distributive lattice is exhibited in Figure 1.

3. Strongness, consistence, and the property of being balanced in meet-distributive lattices

Upper semimodularity and lower semimodularity are generalizations of modularity. There are other properties of modularity which have been isolated and proven to be fruitful concepts. These are, among others, the properties of being strong, balanced, and consistent which we consider here in connection with meet-distributive lattices.

The concept of a strong lattice is due to Faigle [10] and may be reformulated as follows:

A lattice L of finite length is called strong if, for all $u \in J(L)$ and for all $x \in L$,

$$u \leqq x \vee u' \quad \text{implies} \quad u \leqq x.$$

From the definition it is immediate that any atomistic lattice is strong.

We shall see (Corollary 3) that strongness may be viewed as a weakening of lower semimodularity. This will easily follow from the subsequent considerations.

For a lattice L of finite length, we define the so-called arrow relations (see Wille [21]) between $J(L)$ and $M(L)$ by

$$u \nearrow s \overset{\text{def}}{\Longleftrightarrow} u \vee s = s^* \quad \text{and} \quad u \swarrow s \overset{\text{def}}{\Longleftrightarrow} u \wedge s = u'$$

for $u \in J(L)$ and $s \in M(L)$ with $u \nleq s$. Next we split up the concept of being balanced due to Reuter [18]:

We say that a lattice of finite length is lower balanced if

$$u \nearrow s \quad \text{implies} \quad u \swarrow s$$

and that it is upper balanced if

$$u \swarrow s \quad \text{implies} \quad u \nearrow s.$$

A lattice of finite length is said to be balanced if it is both upper and lower balanced.

It is immediate that any upper semimodular lattice of finite length is upper balanced and that any lower semimodular lattice of finite length is lower balanced.

For an example of a balanced lattice being neither upper nor lower semimodular, we refer to Reuter [18]. Whereas a strong lattice is not balanced in general (see Figure 1), we can show that a balanced lattice is always strong. In fact, the property of being lower balanced characterizes strongness:

Theorem 2 (Stern [20], Theorem 18.3). *A lattice of finite length is strong if and only if it is lower balanced.*

An immediate consequence is

Corollary 3 (Faigle [11]). *Any lower semimodular lattice of finite length is strong.*

Thus we have the following implications for lattices of finite length:

$$\text{lower semimodular} \Longrightarrow \text{lower balanced} \Longleftrightarrow \text{strong}.$$

The first implication is non-reversible: for example, the lattice of flats of a finite affine incidence geometry is strong but not lower semimodular.

Corollary 3 implies also that any modular lattice of finite length is strong. Moreover it is a direct consequence of the isomorphism theorem for modular lattices that any modular lattice of finite length is balanced. In fact, it follows from Theorem 2 that, in an upper semimodular lattice of finite length, the property of being balanced is equivalent to strongness:

Corollary 4 (Reuter [18]). *An upper semimodular lattice of finite length is strong if and only if it is balanced.*

Proof. Any upper semimodular lattice of finite length is upper balanced. By Theorem 2 it is lower balanced (and hence balanced) if and only if it is strong.

Since a meet-distributive lattice is lower semimodular (see Theorem 1), Corollary 3 also yields

Corollary 5. *Any meet-distributive lattice is strong.*

Whereas the notions of strongness and balance coincide in upper semimodular lattices (see Corollary 4), the situation is different in lower semimodular lattices:

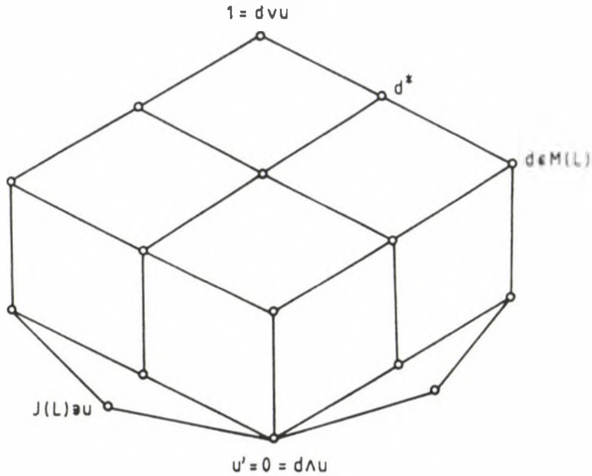


Fig. 1

As the dual of a locally free lattice (see Crapo [3], Figure 11), the lattice of Figure 1 is meet-distributive; hence it is (lower semimodular and) strong. However, we have (notation of Figure 1) $d \wedge u = u' \prec u$ but d is not a lower cover of $d \vee u$, that is, the lattice is not (upper) balanced.

Let us now turn to the property of consistence as introduced by Kung [14]. This property may be reformulated for our purpose as follows:

Let L be a lattice of finite length. A join-irreducible $u \in J(L)$ is said to be consistent if $x \vee u \in J([x, 1])$ holds for all $x \in L$ with $u \nleq x$. The lattice L is consistent if each of its join-irreducibles is consistent. We remark that this differs from Kung's original definition in that here we do not consider 0 to be join-irreducible. (In Kung's theory — leading to a solution of Rival's matching problem for modular lattices — it is essential to include 0, see Kung [14], [15], [16]).

A lattice L of finite length is dually consistent if the dual lattice L^d is consistent.

From the isomorphism theorem for modular lattices it is again clear that any modular lattice of finite length is both consistent and dually consistent.

A concept equivalent to dual consistence was already formulated and used by Crawley [4] to characterize lattices (of finite length) possessing the Kurosh–Ore replacement property for meet-decompositions:

For all $x, y \in L$, if the sublattice $[x, x \vee y]$ has exactly one atom, then the sublattice $[x \wedge y, y]$ has exactly one atom (see Crawley–Dilworth [5], p. 53).

In the presence of semimodularity, one has

THEOREM 6 (s. Crawley–Dilworth [5], Theorem 7.5, p. 53 and Theorem 7.6, p. 54). *Let L be an upper semimodular lattice of finite length. Then the following conditions are equivalent:*

(i) *L has the Kurosh–Ore replacement property for meet-decompositions;*

(ii) *L is locally modular;*

(iii) *L is dually consistent.*

Dualizing Theorem 6, we obtain

THEOREM 7. *Let L be a lower semimodular lattice of finite length. Then the following conditions are equivalent.*

(i) *L has the Kurosh–Ore replacement property for join decompositions;*
(ii) *L is dually locally modular;*
(iii) *L is consistent.*

From this we get immediately

COROLLARY 8 (Reuter [18]). *A meet-distributive lattice is consistent.*

PROOF. Follows from the fact that any meet-distributive lattice is dually locally modular (see Section 2) and from Theorem 7.

As we have seen, any lower semimodular lattice is strong (Corollary 3). On the other hand, a lower semimodular lattice is not consistent, in general: consider e.g. the dual of the lattice of flats of an affine incidence geometry. This is again in contrast to the situation in upper semimodular lattices: it can be shown that an upper semimodular lattice of finite length is strong if and only if it is consistent (see Faigle [10]; cf. also Reuter [18]).

We show now that if a meet-distributive lattice is balanced, then it is distributive (the converse holds trivially). This will be a consequence of the more general assertion of Theorem 10 below. As a preparation, we shall need

THEOREM 9 (Stern [19]). *A lattice L of finite length is upper semimodular if and only if, for all $u \in J(L)$ and for all $b \in L$,*

$$u \wedge b = u' \prec u \quad \textit{implies} \quad b \prec u \vee b.$$

Now we are in a position to prove

THEOREM 10. *Let L be a locally modular lattice of finite length. Then L is balanced if and only if it is modular.*

PROOF. Any locally modular lattice L of finite length is upper semimodular. We show that L is also lower semimodular, provided it is balanced. To prove lower semimodularity it suffices by the dual of Theorem 9 to show that for any meet-irreducible element $d \in M(L)$ and for any $y \in L$,

$$d \prec d \vee y = d^* \quad \text{implies} \quad d \wedge y \prec y.$$

From local modularity we get by Theorem 6 that L is dually consistent. Hence the interval $[d \wedge y, y]$ has exactly one atom, say p.

Consider now a join-irreducible $u \in J(L)$ for which $u \leqq y$ but $u \nleqq d \wedge y$. Then $u \nleqq d$ and from $u \leqq d^*$ it follows that $d^* = d \vee u$.

The property of being lower balanced implies

$$d \wedge u = u' \prec u.$$

Then also $(d \wedge y) \wedge u = u' \longrightarrow\!\!< u$ and semimodularity yields

$$d \wedge y \longrightarrow\!\!< (d \wedge y) \vee u \leqq y.$$

Hence $(d \wedge y) \vee u$ is an atom of the interval $[d \wedge y, y]$ and thus we have $(d \wedge y) \vee u = p$. Since u is a join-irreducible with the properties $u \leqq y$ and $u \nleqq d \wedge y$ but otherwise arbitrary, it follows that $p = y$.

In other words, we get $d \wedge y \longrightarrow\!\!< y$, that is, L is lower semimodular. Since L is also upper semimodular and of finite length, it is modular. The converse is obvious.

Since the properties of being balanced and of being modular are self-dual, dualizing Theorem 10 yields

THEOREM 11. *A dually locally modular lattice of finite length is balanced if and only if it is modular.*

This implies, in particular,

COROLLARY 12. *A meet-distributive lattice is balanced if and only if it is distributive.*

PROOF. Let L be a balanced meet-distributive lattice. Any meet-distributive lattice is dually locally modular (see Section 2). Hence L is modular by Theorem 11. From Theorem 1 (equivalence of conditions (i) and (ii)) it follows that L is even distributive. The converse is obvious.

We close with the observation that, for upper semimodular lattices of finite length, the property of being balanced (which is in this case equivalent to both strongness and consistence) does not imply modularity, in general.

REFERENCES

[1] AVANN, S. P., Application of the join-irreducible excess function to semimodular lattices, *Math. Ann.* **142** (1960/61), 345–354. *MR* **23** #A81

[2] BJÖRNER, A. and ZIEGLER, G. M., Introduction to Greedoids, in: N. White (ed.) *Combinatorial Geometries: Advanced Theory*, Cambridge University Press (in print).

[3] CRAPO, H. H., Selectors: a theory of formal languages, semimodular lattices, and branching and shelling processes, *Adv. in Math.* **54** (1984), 233–277. *MR* **86d**: 05029

[4] CRAWLEY, P., Decomposition theory for nonsemimodular lattices, *Trans. Amer. Math. Soc.* **99** (1961), 246–254. *MR* **22** #10930

[5] CRAWLEY, P. and DILWORTH, R. P., *Algebraic Theory of Lattices*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[6] DILWORTH, R. P., Lattices with unique irreducible decompositions, *Ann. of Math.* (2) **41** (1940), 771–777. *MR* **2**-120

[7] EDELMAN, P. H., Meet-distributive lattices and the anti-exchange closure, *Algebra Universalis* **10** (1980), 290–299. *MR* **81c**: 06006

[8] EDELMAN, P. H., Abstract convexity and meet-distributive lattices, *Combinatorics and ordered sets* (Arcata, Calif., 1985), Contemporary Mathematics, Vol. 57, Amer. Math. Soc., Providence, R. I., 1986, 127–150. *MR* **87m**: 52003

[9] EDELMAN, P. H. and JAMISON, R. E., The theory of convex geometries, *Geom. Dedicata* **19** (1985), 247–270. *MR* **87f**: 52002

[10] FAIGLE, U., Geometries on partially ordered sets, *J. Combin. Theory Ser. B* **28** (1980), 26–51. *MR* **81m**: 05054

[11] FAIGLE, U., written communication, 1987.

[12] GRÄTZER, G., *General Lattice Theory*, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Math. Reihe, Bd. 52, Birkhäuser Verlag, Basel, 1978. *MR* **80c**: 06001a

[13] KORTE, B. and LOVÁSZ, L., Greedoids — a structural framework for the greedy algorithm, *Progress in Combinatorial Optimization* (Waterloo, 1982), Academic Press, Toronto, Ont., 1984, 221–243. *MR* **86e**: 90091

[14] KUNG, J. P. S., Matchings and Radon transforms in lattices I. Consistent lattices, *Order* **2** (1985), 105–112. *MR* **88b**: 06010a

[15] KUNG, J. P. S., Radon transforms in combinatorics and lattice theory, *Combinatorics and ordered sets* (Arcata, Calif., 1985), Contemporary Mathematics, Vol. 57, Amer. Math. Soc., Providence, R. I., 1986, 33–74. *MR* **88d**: 05024

[16] KUNG, J. P. S., Matchings and Radon transforms in lattices II. Concordant sets, *Math. Proc. Cambridge Philos. Soc.* **101** (1987), 221–231. *MR* **88b**: 06010b

[17] MONJARDET, B., A use for frequently rediscovering a concept, *Order* **1** (1985), 415–417. *MR* **86e**: 06012

[18] REUTER, K., The Kurosh–Ore exchange property, *Acta Math. Hungar.* **53** (1989), 119–127. *MR* **90c**: 06012

[19] STERN, M., A characterization of semimodularity in lattices of finite length, *Acta Sci. Math. (Szeged)* **51** (1987), 217–219. *MR* **88m**: 06007

[20] STERN, M., *Semimodular lattices*, Teubner-Verlag, Stuttgart–Leipzig, 1991.

[21] WILLE, R., Subdirect decomposition of concept lattices, *Algebra Universalis* **17** (1983), 275–287. *MR* **85m**: 06012a

(Received September 27, 1988)

MARTIN-LUTHER-UNIVERSITÄT
FACHBEREICH MATHEMATIK UND INFORMATIK
D/O–4010 HALLE
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 287–294

EIN ISOPERIMETRISCHES PROBLEM BEZÜGLICH ZERLEGUNGEN DER EUKLIDISCHEN EBENE

ÉVA VÁSÁRHELYI

Wir betrachten ein konvexes Vieleck mit höchstens sechs Seiten und zerlegen es in endlich viele konvexe Teilvielecke $A_1, A_2, \dots, A_N$. Wir bezeichnen den Flächeninhalt, den Umfang und die Seitenzahl von A_i mit a_i, p_i bzw. s_i $(i = 1, 2, \dots, N)$.

G. Fejes Tóth [1] und L. Fejes Tóth [2] haben diesbezüglich die Ungleichung

$$\sum_{i=1}^{N} p_i \geqq \sqrt{8\sqrt{3}} \sum_{i=1}^{N} \sqrt{a_i}$$

mit der Nebenbedingung $\min \sqrt{a_i} / \max \sqrt{a_i} \geqq q_0 = 0.562\dots$ bewiesen. Die Ungleichung besagt, daß man aus annähernd gleichen Vielecken keine bessere Zerlegung erreichen kann als mit regelmäßigen Sechsecken.

G. Kertész [3] hat ein Mosaik konstruiert, für welches diese Nebenbedingung nicht erfüllt ist und für das die obige Abschätzung nicht gilt. Daher ist eine Einschränkung vom Typ $\min \sqrt{a_i} / \max \sqrt{a_i} \geqq q_0$ nötig.

L. Fejes Tóth hat die Frage nach dem Infimum $I(q)$ von $\sum p_i / \sum \sqrt{a_i}$ aufgeworfen, unter der Annahme, daß $A_1, A_2, \dots, A_N$ konvexe Vielecke sind, die in einem Vieleck mit höchstens sechs Seiten eine Zerlegung bilden und $\min \sqrt{a_i} / \max \sqrt{a_i} \geqq q$, $q \in (0; 1]$ ist.

In dieser Arbeit wird eine untere Schranke für $I(q)$ gegeben.

Die Homogenität q_H und der zu untersuchende Quotient σ_H der Vielecksmenge $H = \{A_i\}_{i=1}^{N}$ wird durch

$$q_H = \frac{\min \sqrt{a_i}}{\max \sqrt{a_i}}$$

und

$$\sigma_H = \frac{\sum\limits_{i=1}^{N} p_i}{\sum\limits_{i=1}^{N} \sqrt{a_i}}$$

definiert.

Aus dem Eulerschen Polyedersatz folgt, daß die mittlere Seitenzahl in den zu untersuchenden Mosaiken nicht größer als 6 ist.

Im folgenden wird die Forderung, nach der die Vielecke ein Mosaik bilden sollen, außer Acht gelassen und wir beschäftigen uns nur mit Vielecksmengen der mittleren Seitenzahl höchstens 6.

SATZ. *Es sei H eine endliche Vielecksmenge mit den Eigenschaften B_1, B_2 mit*
B_1: *Die mittlere Seitenzahl der zu H gehörigen Polygone ist höchstens* 6.
B_2: $q_H \geqq q$, *wobei q ein fest gegebenes Element im Intervall* $(0; 1]$ *ist.*

Weiters sind f_n, Q_k^m, Q^m und $F_k^m(q)$ gemäß (a), (b), (c) *und* (d) *erklärt:*

$$\text{(a)} \qquad f_n = 2\sqrt{n \tan\frac{\pi}{n}}; \qquad n \geqq 3$$

$$\text{(b)} \qquad Q_k^m = \frac{(6-k)(f_m - f_{m+1})}{f_k - f_m - (m-6)(f_m - f_{m+1})}; \quad m \geqq 6, \quad k = 4; 5$$

$$\text{(c)} \qquad Q^m = \frac{2f_5 - f_4 - f_m}{(m-6)(f_4 - f_5)};$$

$$\text{(d)} \qquad F_k^m(q) = \frac{q(m-6)f_k + (6-k)f_m}{q(m-6) + 6 - k}.$$

Bezeichne ferner $F(q)$ das Infimum des Quotienten σ_H solcher Mengen H, dann gilt die folgende Gleichung

$$(*) \qquad F(q) = \begin{cases} F_5^m(q), \ q \in (Q_5^m; Q_5^{m-1}], & m = 6; 7; \dots ; 11, Q_5^5 = 1 \\ F_5^{12}(q), \ q \in (q^*; Q_5^{11}], & q^* = Q^{12} \\ F_4^{12}(q), \ q \in (Q_4^{12}; q^*], & \\ F_4^m(q), \ q \in (Q_4^m; Q_4^{m-1}], & m = 13; 14; \dots . \end{cases}$$

BEMERKUNGEN. 1. Aus der Definition von $I(q)$ und $F(q)$ folgt, daß $F(q)$ eine untere Schranke für $I(q)$ darstellt. Der Satz gibt uns die exakte Bestimmung von $F(q)$.

2. Geometrisch läßt sich f_n als der Umfang eines regelmäßigen n-Ecks mit dem Flächeninhalt 1 interpretieren.

Die Werte $f_3, f_4, \dots , f_n$ bilden eine monotone abnehmende konvexe Folge mit dem Grenzwert $2\sqrt{\pi} = 3,55 \dots$.

3. $F_k^m(q)$ bezeichnet den Quotienten σ_H, wobei H aus regelmäßigen m-Ecken mit dem Einheitsflächeninhalt, sowie aus regelmäßigen k-Ecken mit dem Flächeninhalt q^2 besteht und die mittlere Seitenzahl der Menge 6 ist.

Durch äquivalentes Umformen erhalten wir die folgende Form von F_k^m

$$F_k^m(q) = f_k - (f_k - f_m)\frac{1}{\frac{m-6}{6-k}q + 1}; \quad m \geqq 6; \ k = 3; 4; 5.$$

Daraus ist es leicht ersichtlich, daß $F_m^k(q)$ für $0 < q \leqq 1$ eine monoton zunehmende Funktion ist. Ist $m = 6$, dann gilt

$$F_3^6(q) = F_4^6(q) = F_5^6(q) = f_6 = \sqrt{8\sqrt{3}} = 3,72\dots.$$

Für $m > 6$ ist $F_k^m(q)$ eine streng zunehmende, konkave Funktion. Weiterhin gilt

$$\lim_{q \to 0} F_k^m(q) = f_m.$$

Für den Beweis des Satzes verwenden wir noch folgende Aussagen:

Hilfssatz 1. *Ist $k = 3; 4; 5$, $m \geqq 6$ und $q \in [0; 1]$, dann gilt die Gleichung*

$$F(q) = \inf_{k,m} F_k^m(q).$$

Beweis. Es sei $H^* = \{A_i^*\}_{i=1}^N$ eine Menge von Vielecken. Wir bezeichnen den Flächeninhalt, den Umfang und die Seitenzahl von A_i^* mit a_i, p_i^*, bzw. s_i^* $(i = 1, 2, \dots, N)$. Setzen wir voraus, daß die Bedingungen B_1 und B_2 erfüllt sind. Dann gibt es dazu eine Menge $H = \{A_i\}_{i=1}^N$, die aus regelmäßigen Vielecken besteht, deren Summe von Seitenzahlen gleich $6N$ ist; ferner ist A_i ein regelmäßiges s_i-Eck mit dem Flächeninhalt a_i, wobei $s_i \geqq s_i^*$. Wegen der isoperimetrischen Ungleichung bekommen wir die Abschätzungen

$$p_i^* \geqq f_{s_i^*}\sqrt{a_i} \geqq f_{s_i}\sqrt{a_i} = p_i,$$

und dadurch für den zu untersuchenden Quotienten die Ungleichung

$$\frac{\sum\limits_{i=1}^N p_i^*}{\sum\limits_{i=1}^N \sqrt{a_i}} = \sigma_{H^*} \geqq \sigma_H = \frac{\sum\limits_{i=1}^N p_i}{\sum\limits_{i=1}^N \sqrt{a_i}}.$$

Wir wollen jetzt für σ_H eine untere Abschätzung angeben.

Es sei $\mathcal{H}$ die Vereinigungsmenge von n Exemplaren von H. Wegen der Voraussetzung $\sum s_i = 6N$, gibt es entsprechende natürliche Zahlen "n" so, daß wir die Menge $\mathcal{H}$ in folgende Teilmengen H_{ij} einteilen können: die Teilmengen H_{ij} bestehen aus höchstens zwei verschiedenen Vielecken und die mittlere Seitenzahl der Teilmenge ist 6. Wir wählen eine so erhaltene Teilmenge aus. Setzen wir voraus, daß diese Teilmenge aus u Examplaren von A_i und v Exemplaren von A_j besteht, dann gilt

$$q' = \min\left\{\frac{\sqrt{a_j}}{\sqrt{a_i}}; \frac{\sqrt{a_i}}{\sqrt{a_j}}\right\} \geqq q_H$$

und (für die ausgewählte Teilmenge)

$$\sigma_{ij} = \frac{up_i + vp_j}{u\sqrt{a_i} + v\sqrt{a_j}}.$$

Ist $s_i = s_j = 6$, dann bekommen wir die Gleichungen

$$\sigma_{ij} = F_k^6(q') = F_k^6(q) = f_6.$$

Falls die ausgewählte Teilmenge keine Sechsecke enthält, dann gelten $s_i < 6$ und $s_j > 6$ (mit passender Wahl der Numerierung). In diesem Fall gilt

$$\sigma_{ij} = \frac{(s_j - 6)f_{s_i}\sqrt{a_i} + (6 - s_i)f_{s_j}\sqrt{a_j}}{(s_j - 6)\sqrt{a_i} + (6 - s_i)\sqrt{a_j}} = F_{s_i}^{s_j}(q') \geqq F_{s_i}^{s_j}(q).$$

Wir bemerken, daß $\sigma_{\mathcal{H}}$ ein gewichtetes arithmetisches Mittel der Werte σ_{ij} mit den Gewichten $(s_j - 6)\sqrt{a_i} + (6 - s_i)\sqrt{a_j}$ ist. Andererseits ist $\sigma_H = \sigma_{\mathcal{H}}$. Da das gewichtete Mittel niemals kleiner als der Wert $\min \sigma_{ij}$ ist, kann man für beliebige Mengen H^* mit den Eigenschaften B_1 und B_2 sagen, daß

$$\sigma_{H^*} \geqq \sigma_H \geqq \min \sigma_{ij} \geqq \inf_{k,m} F_k^m(q).$$

Folglich gilt die Ungleichung

$$F(q) \geqq \inf_{k,m} F_k^m(q).$$

Die umgekehrte Ungleichung

$$F(q) \leqq \inf_{k,m} F_k^m(q)$$

ist nach der Definition von $F(q)$ evident. □

In dem Hilfssatz 2 sammeln wir die technischen Hilfsmittel.

HILFSSATZ 2. A) *Ist* $m \geqq 6$ *und* $q \in (0; 1]$, *dann gilt*

(1) $$F_3^m(q) \geqq F_4^m(q), \quad q \in (0; 1],$$

und die Ungleichungen

(2) $$F_k^{m+1}(q) \lesseqqgtr F_k^m(q), \qquad (k = 4; 5)$$

gelten dann und nur dann, wenn

(2') $$q \lesseqqgtr Q_k^m = \frac{(6 - k)(f_m - f_{m+1})}{f_k - f_m - (m - 6)(f_m - f_{m+1})}.$$

Die Folge Q_k^m $(m \geqq 6)$ *ist streng abnehmend mit dem Grenzwert* 0*:*

$$Q_k^{m+1} < Q_k^m, \quad \lim_{m\to\infty} Q_k^m = 0. \tag{3}$$

B) *Es seien* $m \geqq 7$ *und* $q \in [0;1]$. *Die Ungleichungen*

$$F_4^m(q) \lesseqqgtr F_5^m(q) \tag{4}$$

gelten dann und nur dann, wenn

$$q \lesseqqgtr Q^m = \frac{2f_5 - f_4 - f_m}{(m-6)(f_4 - f_5)}. \tag{4'}$$

C) *Für* $m \geqq 13$ *gilt die Ungleichung*

$$Q^m > Q_5^{m-1}. \tag{5}$$

Beweis. Da diese Relationen mit Hilfe der Eigenschaften von f_n und durch einfaches Umformen einzusehen sind, beweisen wir exemplarisch nur die Behauptung C).

Wir wenden vollständige Induktion über m an.

a) Nach numerischen Rechnungen folgt

$$Q^{13} > Q_5^{12}.$$

b) Durch äquivalentes Umformen bekommen wir aus (5) für $m = n$ die Ungleichung

$$2f_5 - f_4 - f_{n-1} > (n-7)(f_{n-1} - f_n). \tag{5'}$$

Addieren wir auf beiden Seiten von (5') den Wert $f_{n-1} - f_n$ und benutzen die Abschätzung $f_{n-1} - f_n > f_n - f_{n+1}$, dann erhalten wir die Ungleichung

$$2f_5 - f_4 - f_n > (n-6)(f_n - f_{n+1}). \tag{5''}$$

Andererseits ist (5") äquivalent mit (5) für $m = n + 1$. Das bedeutet aber, daß (5) für $m \geqq 13$ gültig ist. □

Der Beweis des Satzes. Wir betrachten die Funktionen

$$F_k(q) := \inf_m F_k^m(q), \qquad k = 3; 4; 5$$

und nennen diese Funktionen Dreiecksschranke, Vierecksschranke bzw. Fünfecksschranke. Wir können die Funktion $F(q)$ durch

$$F(q) = \min\{F_k(q) \mid k = 3; 4; 5\}$$

angeben. Wir zeigen, daß die Schranke $F(q)$ mit der Vierecksschranke im Fall $q \leqq q^*$ und mit der Fünfecksschranke im Fall $q > q^*$ koinzidiert. Nach der

Relation (1) spielt die Dreiecksschranke in $F(q)$ keine Rolle. Wir betrachten die Intervalle

$$I_m = (Q_5^m; Q_5^{m-1}] \quad \text{und} \quad J_m = (Q_4^m; Q_4^{m-1}],$$

wobei $Q_4^5 = Q_5^5 = 1$, $m \geqq 6$.

Aus (3) folgt die Tatsache, daß die kleinen Intervalle I_m und J_m das Intervall $(0;1]$ ausfüllen:

$$\bigcup_{m=6}^{\infty} I_m = \bigcup_{m=6}^{\infty} J_m = (0;1].$$

Nach (2) und (2') können wir die Funktionen F_4 und F_5 explizit angeben:

$$\begin{aligned} F_4(q) &= F_4^m(q) \quad \text{für} \quad q \in J_m, \\ F_5(q) &= F_5^m(q) \quad \text{für} \quad q \in I_m. \end{aligned}$$

Nach numerischen Rechnungen folgt

$$Q^{12} \in I_{12} \bigcap J_{12},$$

und dies bedeutet für $q^* = Q^{12} = 0,0329\ldots$,

$$F_4(q^*) = F_4^{12}(q^*) = F_5^{12}(q^*) = F_5(q^*) = F(q^*).$$

Wir wählen die Intervalle $J_m \subset (0; q^*]$ aus. Mit Rücksicht auf (4), (4') und (5) ergeben sich die Relationen

$$F_5(q) = F_5^m(q) > F_4^m(q) \geqq F_4(q), \qquad (q \in I;\ m \geqq 13), \tag{6}$$

$$F_5(q) = F_5^{12}(q) \geqq F_4^{12}(q) \geqq F_4(q), \quad (q \in (Q_5^{12}; q^*]), \tag{7}$$

$$F_5(q) \leqq F_4^{12}(q) = F_4(q), \quad (q \in (q^*; Q_4^{11})). \tag{8}$$

Wir betrachten nun die Intervalle $J_m \subset (q^*; 1]$. Dies bedeutet, daß $6 \leqq m \leqq 11$ gilt. Die Relationen

$$F_4(q) = F_4^6(q) = f_6 \geqq F_5(q), \qquad (q \in J_6), \tag{9}$$

sind evident. Wir bemerken, daß Q^m in (4') für $m = 7; 8$ negativ ist. Daraus folgt, daß die Ungleichungen

$$F_4(q) = F_4^m(q) > F_5^m(q) \geqq F_5(q), \qquad (q \in J_m;\ m = 7; 8) \tag{10}$$

gelten. Die numerischen Werte zeigen, daß $Q^m < Q_4^m$ für $m = 9; 10; 11$ gilt. Daraus und aus (4) und (4') bekommen wir die Relationen

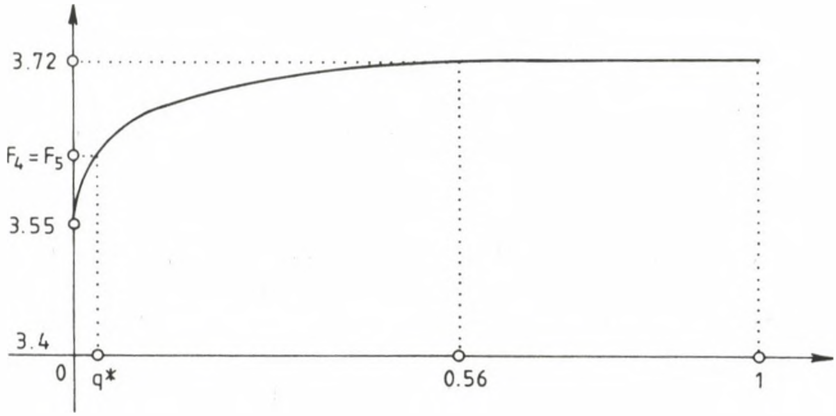
$$F_4(q) = F_4^m(q) > F_5^m(q) \geqq F_5(q) \qquad (q \in J_m,\ m = 9; 10; 11). \tag{11}$$

Aus den Relationen (6)–(11) ergibt sich die untere Schranke $F(q)$ in der Form $(*)$. □

Es sei bemerkt, daß die Gleichungen

$$\lim_{m\to\infty} F_4 = \lim_{m\to\infty} F_5 = \lim_{q\to 0} F(q) = 2\sqrt{\pi}$$

für die Grenzwerte gelten. Das Diagramm drückt den Verlauf der Schranke $F(q)$ aus.



Erweiterung für normale Konvexe Mosaike der euklidischen Ebene

Wir sagen, daß ein Mosaik normal ist, wenn ein Paar (ϱ_1, ϱ_2) positiver reeller Zahlen existiert mit der Eigenschaft, daß sich jeder Mosaikstein durch einen Kreis mit Radius ϱ_2 überdecken läßt und andererseits, in jedem solchen Mosaikstein ein Kreis mit Radius ϱ_1 enthalten ist.

Die Menge der normalen konvexen Mosaike, die Menge der konvexen Bereiche und die Menge der konvexen Vielecke werden durch $\mathcal{M}$, $\mathcal{B}$ bzw. $\mathcal{P}$ bezeichnet.

Es sei $M \in \mathcal{M}$ und wir bezeichnen mit A_i die Flächen des Mosaiks, ferner bezeichnen wir mit a_i, p_i bzw. s_i den Flächeninhalt, den Umfang und die Seitenzahl von A_i ($i \in \Lambda$, Λ ist eine Indexmenge). Wir betrachten einen Bereich $B \in \mathcal{B}$. Es sei O ein innerer Punkt von B; λB bedeutet das Bild von B bezüglich der Ähnlichkeit mit Zentrum O und Koeffizient λ. Jetzt nehmen wir diejenigen Flächen von M, die in λB enthalten sind. Die mittlere Seitenzahl $s_{\lambda B}$ und der Quotient $\sum p_i / \sum \sqrt{a_i} = \sigma_{\lambda B}$ dieser Flächen hängen davon ab, wie wir B, O und λ wählen.

Die mittlere Seitenzahl s_M, die Homogenität q_M und der isoperimetrische Mittelwert σ_M des Mosaiks M werden durch

$$s_M = \sup_{B\in\mathcal{B}} \sup_{O\in \operatorname{int} B} \limsup_{\lambda\to\infty} s_{\lambda B},$$

$$q_M = \inf \sqrt{a_i} \,/ \sup \sqrt{a_i}$$

und

$$\sigma_M = \inf_{B \in \mathcal{B}} \inf_{O \in \operatorname{int} B} \liminf_{\lambda \to \infty} \sigma_{\lambda B}$$

definiert.

Wie allgemein bekannt ist, gelten für Mosaike $M \in \mathcal{M}$ die folgenden Zusammenhänge:

$$s_M \leqq 6,$$
$$\sigma_M = \inf_{P \in \mathcal{P}} \inf_{O \in \operatorname{int} P} \liminf_{\lambda \to \infty} \sigma_{\lambda P}$$
$$q_M \in (0; 1].$$

Nutzt man diese Beziehungen, so kann man zeigen, daß $F(q)$ eine untere Schranke auch für normale konvexe Mosaike der euklidischen Ebene ist, d.h.

$$\sigma_M \geqq F(q_M).$$

Diese Schranke ist offensichtlich nicht genau, da $F(q)$ in einem Intervall mit $F_5^7(q)$ koinzidiert, obwohl aus regelmäßigen Fünfecken und Siebenecken in der euklidischen Ebene kein Mosaik gebaut werden kann.

LITERATURVERZEICHNIS

[1] Fejes Tóth, G., An isoperimetric problem for tessellations, *Studia Sci. Math. Hungar.* **10** (1975), 171–173. *MR* **56** #3748

[2] Fejes Tóth, L., Isoperimetric problems for tilings, *Mathematika* **32** (1985), 10–15. *MR* **87c**: 52023

(Eingegangen am 2. November 1988.)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 295–299

SOME REMARKS ON 2-GROUPS HAVING SOFT SUBGROUPS

L. HÉTHELYI

A maximal abelian subgroup A of a p-group in G is called a soft subgroup G if A is maximal in its normalizer. The basic properties of soft subgroups were investigated in [1]. In this paper we shall consider 2-groups with soft subgroups. We shall characterize those 2-groups in which every maximal abelian group is soft.

We shall investigate the special case when a 2-group G contains a soft subgroup of index 2. We shall show various conditions which guarantee that all maximal abelian subgroups are soft in this case.

Throughout the paper we use the notation of [1].

We shall need the following result which is valid for an arbitrary prime p.

LEMMA 1. *Suppose G is a non-abelian p-group. Let A be a maximal abelian subgroup of G with index p in G. Suppose that G' or $Z(G)$ is cyclic. Then every maximal abelian subgroup of G is soft in G.*

PROOF. By Statement 1 of [2] we only need to prove Lemma 1 in the case of G' being cyclic. Let $x \in G \setminus A$. Then $K_i(G) = [K_{i-1}(G), \langle x \rangle]$ for $i=3, \dots, \mathrm{cl}(G)+1$. As $x^p \in A$, $\exp(K_i(G)/K_{i+1}(G)) = p$. Thus $|K_i(G)/K_{i+1}(G)| = p$ for $2 \leqq i \leqq \mathrm{cl}(G)$. Then $G' \cap Z(G) = K_n(G)$, where $K_n(G)$ is the last nontrivial term of the lower central series of G. Let B be any maximal abelian subgroup of G different from A. Let $N = N_G(B)$. Then $N = (N \cap A)B$ and $\mathrm{cl}(N) = 2$. Thus $N' \cap Z(N) \leqq Z(G)$. Thus $N' \leqq K_n(G)$. However, as $|K_n(G)| = p$, $N' = K_n(G)$. As $|N : N \cap A| = p$ and as $|Z||N||N'| = |N \cap A|$, $|N : Z(N)| = p^2$ follows. Thus $|N : B| = p$. □

STATEMENT 1. *Suppose G is a non-abelian 2-group. Let A be a maximal abelian subgroup of G with index 2 in G. Suppose that $Z(G)$ is cyclic. Then G' is cyclic.*

PROOF. It is easy to see that for any $x \in G \setminus A$ and $y \in G'$ $x^{-1}yx = y^{-1}$. Thus $\Omega_1(G') \leqq Z(G)$, and so $|\Omega_1(G')| = 2$. Then as G' is abelian it is cyclic. □

PROPOSITION 1. *Suppose G is a 2-group and that every maximal abelian subgroup of G is soft in G. Then G' is cyclic.*

1980 *Mathematics Subject Classification* (1985 Revision). Primary 20D10.

Key words and phrases. p-groups, central series, abelian group, commutator series.

Akadémiai Kiadó, Budapest

PROOF. The assumption of the proposition guarantees an abelian subgroup A of index two in G. Moreover we can assume that $|\phi(G')\cap Z(G)|=1$. Otherwise let $Z_1=\phi(G')\cap Z(G)$. Let $\overline{G}=G/Z_1$. Let $\overline{B}$ be a maximal abelian subgroup of $\overline{G}, \overline{B}\neq\overline{A}$. Then $\overline{B}=\langle\overline{x}, Z(\overline{G})\rangle$ for some $x\in G\setminus A$. Now $C=\langle x, Z(G)\rangle$ is a maximal abelian subgroup in G, hence C is soft by assumption. For $B\geqq C$, Lemma 2 of [1] yields $|N_G(B):B|=2$, hence $\overline{B}$ is soft in $\overline{G}$. Then $\overline{G}'$ is cyclic by an induction, so G' is cyclic. Thus $\exp(G')=2$. However, $\Omega_1(G')\leqq Z(G)$. Then $G'\leqq Z(G)$. So every maximal abelian subgroup of G is of index two in G by assumption. Then $|G:Z(G)|=4$, and thus $|G'|=2$. □

PROPOSITION 2. *Suppose G is a non-abelian 2-group. Then every maximal abelian subgroup of G is soft in G if and only if $G/Z(G)$ is dihedral.*

PROOF. If every maximal abelian subgroup of G is soft then there exists an abelian subgroup A of G such that $|G:A|=2$. If $\mathrm{cl}(G)=2$ then $|G:Z(G)|$=4 and thus $|G'|$=2. So we can assume that $\mathrm{cl}(G)$>2. Let B be a maximal abelian subgroup of G which is different from A. Then as $A\cdot B=G$, $|B:A\cap B|=2$ follows. Thus $|B:Z(G)|=2$. Let T be a maximal abelian subgroup of G which is different from A and which is not contained in the same maximal subgroup of G as B. Then $\langle B,T\rangle=G$ by Theorem 1 of [1]. Thus $G/Z(G)$ is generated by two involutions and so it is dihedral.

Suppose now that $G/Z(G)$ is dihedral.

Then G has an abelian subgroup A of index 2. Moreover it is easy to see that for any maximal abelian subgroup B of G which is different from A we have $B=\langle x\rangle Z(G)$ for some $x\in G$. As $N_G(B)/Z(G)=4$ we have $|N_G(B):B|=2$. □

PROPOSITION 3. *Let G be a non-abelian p-group where p is an arbitrary prime. Suppose that G contains an abelian subgroup A of index p, and that $|G:\phi(G)|=p^2$. Then every maximal abelian subgroup of G is soft in G.*

PROOF. First we note that $K_i(G)=[K_{i-1}(G),\langle x\rangle]$ for $i=3,\dots,\mathrm{cl}(G)+$ $+1$ for any $x\in G\setminus A$. We claim that $\exp(K_i(G)/K_{i+1}(G))=p$.

Suppose $a_1\in K_1(G)\setminus K_{i+1}(G)$. Then $[a_1,x]\in K_{i+1}(G)$. Let $\overline{G}=$ $=G/K_{i+2}(G)$. Then $[\overline{a}_1,\overline{x}^p]=[\overline{a}_1,\overline{x}]^p=1$. Then $[a_1,x]^p\in K_{i+1}(G)$, which proves the claim. We now show that $|K_i(G)/K_{i+1}(G)|=p$ for $i=2,\dots,\mathrm{cl}(G)$. The claim is certainly true if $\mathrm{cl}(G)=2$.

Let $K_n(G)$ be the last nontrivial term of the lower central series of G. Let the bar denote homomorphic images in $G/K_n(G)$. Then $|K_i(G)/K_{i+1}(G)|=$ $=p$ for $i=2,\dots,\mathrm{cl}(G)$. Thus we only have to show that $|K_n(G)|=p$. Let a $K_{n-1}(G)\setminus K_n(G)$. Then $[K_{n-1}(G),x]=[\langle x\rangle K_n(G),x]=[a,x]$. However, as $\exp(K_n(G))=p$, $|K_n(G)|=p$ follows which proves the claim. Now $G'\cap$ $\cap Z/(G)=K_n(G)$ follows where $K_n(G)$ is the last nontrivial term of the lower central series of G. However, let B be any maximal abelian subgroup of G different from A. Then $N = N_G(B) = (A \cap N_G(B))B$ and $N' \leqq Z(N) \leqq Z(G)$. Then $N' \leqq K_n(G)$. Thus $|N'| = p$. As $|A \cap N| = |Z(N)|, |N'|, |N : Z(N)| = p^2$ follows, which proves Proposition 3. □

COROLLARY. *Suppose that G is a non-abelian 2-group generated by two elements. Assume that G contains an abelian subgroup A of index 2. Then every maximal abelian subgroup of G is soft in G, so $G/Z(G)$ is dihedral.*

PROOF. The corollary follows from Proposition 2 and 3. □

PROPOSITION 4. *Suppose G is a non-abelian 2-group. Let A be a maximal abelian subgroup of G of index two. Then the following conditions hold.*

a) $|G' : \phi(G')| < |G : \phi(G)|$.

b) $\exp(G') = 2^{\mathrm{cl}(G)-1}$.

c) *Every subgroup of A containing $Z(G)$ is normal in G. Moreover every subgroup of G' is normal in G.*

d) *$K_i(G)/K_{i+1}(G)$ and $Z_i(G)/Z_{i-1}(G)$ are elementary abelian for $i = 1, \dots, \mathrm{cl}(G)$.*

e) $\mathrm{cl}(G) = \max\{\mathrm{cl}(H); H \leqq G|H : \phi(H)| \leqq 4\}$.

PROOF. a) We can assume that $\exp(G') = 2$. Then $\phi(G) \leqq Z(G)$ follows. So $|G : \phi(G)| \geqq |G : Z(G)| > |A : Z(G)| = |G'|$, as $|A| = |G'||Z(G)|$.

For c) we note that $x^2 \in Z(G)$ for any $x \in G \setminus A$. Thus $|\overline{x} \cdot \overline{y}| = 2$ for any $\overline{x} \in G \setminus \overline{A} \cdot \overline{y} \in \overline{A}$ where the bar denotes homomorphic images in $G/Z(G)$. Thus for any $x \in G \setminus A$ x inverts every element of $G/Z(G)$ by conjugation.

d) is trivial.

For e) let y be an element of G such that $[x, y] = \exp G'$. Take $H = \langle x, y \rangle$. Then $\mathrm{cl}(H) = \mathrm{cl}(G)$. □

PROPOSITION 5. *Suppose G is a non-abelian two generated 2-group which contains an abelian subgroup of index 2. Then every subgroup of G can be generated by at most 3 elements.*

PROOF. By Proposition 1 and 3 G' is cyclic and every subgroup of G/G' can be generated by at most two elements. □

It is easy to see that if G is a non-abelian p-group which contains an abelian subgroup A of index p then for every $x \in G \setminus Z(G)$, $C_G(x)$ is abelian. The following proposition is a partial converse of this when $p = 2$.

PROPOSITION 6. *Suppose that G is a non-abelian 2-group and that for every $x \in G \setminus Z(G)$, $C_G(x)$ is abelian. If $\phi(G)$ is not contained in $Z(G)$ then G contains an abelian subgroup A of index 2.*

PROOF. We shall need the following lemma.

LEMMA 2. *Suppose that G is a non-abelian 2-group and that $C_G(x)$ is abelian for every $x \in G \setminus Z(G)$. Let A be a maximal abelian normal subgroup of G. If G/A is cyclic then $|G : A| = 2$.*

PROOF. We may assume that $|G:A|=4$. Let M be a maximal subgroup of G containing A. Let M_1 be another maximal subgroup of G. Then $M_1/A\cap \cap M_1$ is cyclic. If $A\cap M_1 \leqq Z(G)$, then it follows that M_1 is abelian. Then $\mathrm{cl}(G)=2$, $|G:Z(G)|=8$ and as $\frac{|G|}{4}=|A|=|G'||Z(G)|$, $|G'|=2$. On the other hand $|G'||Z(G)|=M_1=\frac{|G|}{2}$. This contradiction proves that $A\cap M_1 \nleqq Z(G)$. Hence by assumption $C_G(A\cap M_1)$ is abelian. It contains the maximal abelian subgroup A so $C_G(A\cap M_1)=A$. Thus $A\cap M_1$ is a maximal abelian subgroup in M_1, and we are done by induction. □

PROOF OF PROPOSITION 6. Let A be a maximal abelian normal subgroup of G. By Lemma 2 G/A is elementary abelian so $\phi(G)\leqq A$. As for any $x\in G\setminus A$ $C_G(x)\cap A=Z(G)$, $x^2\in Z(G)$ for any x outside A. Let the bar denote the homomorphic images in $G/Z(G)$. As $\phi(G)$ is not contained in $Z(G)$ there exists an u in A such that $|\overline{u}|\geqq 4$. However, $|\overline{v}|=2$ for any $\overline{v}\in \overline{G}\setminus\overline{A}$. Thus $\overline{v}\overline{u}\overline{v}=\overline{u}^{-1}$. Therefore $C_{\overline{G}}(\overline{u})=\overline{A}$. However, if v_1, v_2 are two elements in $G\setminus A$ such that v_1 is not contained in Av_2 then on the one hand $\overline{v}_1\cdot\overline{v}_2\notin A$ while on the other $\overline{v}_1\overline{v}_2\in C_{\overline{G}}(u)=\overline{A}$. This contradiction shows that $|G:A|=2$. □

PROPOSITION 7. *Suppose G is a non-abelian 2-group and that for every $x\in G\setminus Z(G)$, $C_G(x)$ is abelian. Suppose that every nonabelian factor of G inherits the above property. Then one of the following conditions hold:*

(i) *G contains an abelian subgroup A of index 2.*

(ii) $|G:Z(G)|\leqq 8$.

PROOF. Suppose that G does not contain an abelian subgroup of index 2. We first show that if Z_1 is a subgroup of order 2 in $\Omega_1(Z(G))$ then $Z(G/Z_1)=Z(G)/Z_1$. If not, then take an element $u\notin Z(G)$ such that $\overline{u}\in \in Z(G/Z_1)$ where the bar denotes the homomorphic image of u in G/Z_1. Then $C_G(u)$ is a normal subgroup of G and $|G:C_G(u)|=2$, which is not the case. This proves the claim. Next we observe that if $x\in G\setminus Z(G)$ then $|C_{\overline{G}}(\overline{x}):\overline{C_G(x)}|\leqq 2$ (where the bar denotes homomorphic images in G/Z_1). Now by induction either $\overline{G}$ contains an abelian subgroup of index 2 or $8=|\overline{G}:Z(\overline{G})|=|G:Z(G)|$. In the first case G contains an abelian subgroup B of index 4, such that $C_{\overline{G}}(\overline{B})$ is of index 2 in $\overline{G}$.

Next we show that $|B:Z(G)|=2$. Suppose the contrary. Let b_1 and b_2 be two elements of $B\setminus Z(G)$ such that b_2 is not contained in $b_1Z(G)$. Let a be an element in $G\setminus B$ such that $[a,B]\leqq Z_1$. Then as $|Z_1|=2$ and as $C_G(b_1)=C_G(b_2)=B$

$$a^{-1}b_1^{-1}ab_1=a^{-1}b_2^{-1}ab_2$$

follows. Thus $b_2b_1=b_1b_2\in C_G(a)$. So $\langle a,B\rangle\leqq C_G(b_1b_2)$ is nonabelian. Thus $b_1b_2\in Z(G)$ contrary to the choices of b_1 and b_2. This proves that $|B:Z(G)|=2$. Thus $|G:Z(G)|=8$. □

COROLLARY 4. *Suppose G is a non-abelian 2-group. Then $c \cdot d \cdot (G) = \{1,2\}$ if and only if for every $x \in G \setminus Z(G)$, $C_G(x)$ is abelian and this property is inherited by every factor of G.*

PROOF. See chapter 12 of [4]. □

REFERENCES

[1] HÉTHELYI, H., Soft subgroups of p-groups, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **27** (1984), 81–85. *MR* **87c**: 20044

[2] HÉTHELYI, H., On linear p-groups of degree p, *Studia Sci. Math. Hungar.* **23** (1988), 369–373. *MR* **90a**: 20032

[3] HUPPERT, B., *Endliche Gruppen I*, Die Grundlehren der mathematischen Wissenschaften, Band 134, Springer-Verlag, Berlin – New York, 1967. *MR* **37** #302

[4] ISAACS, I. M., *Character Theory of Finite Groups*, Pure and Applied Mathematics, No. 69, Academic Press, New York, 1976. *MR* **57** #417

(Received October 10, 1988)

BUDAPESTI MŰSZAKI EGYETEM
VEGYÉSZMÉRNÖKI KAR
MATEMATIKA TANSZÉK
SZTOCZEK U. 2
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 301–306

ON THE CHARACTERISTIC OF A PROJECTIVE GEOMETRY

H. LAKSER

1. Introduction. In Hermann and Huhn [4] lattice identities were given that determine the characteristic of a ring R in terms of the lattices of submodules of unitary R-modules. These results were generalized in Hutchinson and Czédli [5]. An important special case is when R is a division ring; the lattices are then the lattices of subspaces of a desarguesian projective geometry. (See Fried, Grätzer and Lakser [1] for an application to this case.)

In this note we establish identities determining the characteristic in the case of a projective geometry using a more elementary, geometric approach. Although our results are subsumed by [4] and [5] in the desarguesian case, certainly the most important one, our results also apply to non-desarguesian planes, where the above-mentioned results do not seem to apply.

We base our approach on the following lemma, announced in Grätzer and Lakser [3], useful for proving lattice identities in modular geometric lattices. A proof is given in [2, Lemma IV.5.11, p. 207].

LEMMA 1 [3]. *Let ε be a lattice identity of the form $p \leqq q$ where p and q are lattice polynomials and each variable occurs in p at most once. If ε holds for the atoms and the minimal element of a modular geometric lattice L, then ε holds for L.*

2. The coordinates. Any projective geometry of dimension $\geqq 3$ satisfies Desargues' Theorem, and the classical coordinatization of von Staudt applies. Since we wish to include non-desarguesian planes in our discussion, we present a modified approach to the coordinatization. We essentially follow the procedure given in Hughes and Piper [6, Chapter V], although our notation departs from theirs.

A *coordinate system* in a projective geometry G of dimension $\geqq 2$ is a sequence of four coplanar points, $\langle 0, i, \infty, \infty^* \rangle$, no three of which are collinear. Note that this notation precludes the use of "0" to denote the minimal element of a bounded lattice. We let R denote the set of those points on the line $0 \vee \infty$ distinct from ∞. In [6] a ternary ring structure is defined on R; this ternary ring structure determines two loop structures on R, an additive one, denoted $+$, with identity 0, and a multiplicative one with identity

$$1 = (i \vee \infty^*) \wedge (0 \vee \infty).$$

This research was supported by the NSERC of Canada.

We need only the additive structure, and so we recall only that part of the construction.

Set

$$1^* = (i \vee \infty) \wedge (0 \vee \infty^*)$$

and

$$1_\infty = (1 \vee 1^*) \wedge (\infty \vee \infty^*)$$

(see Fig. 1).

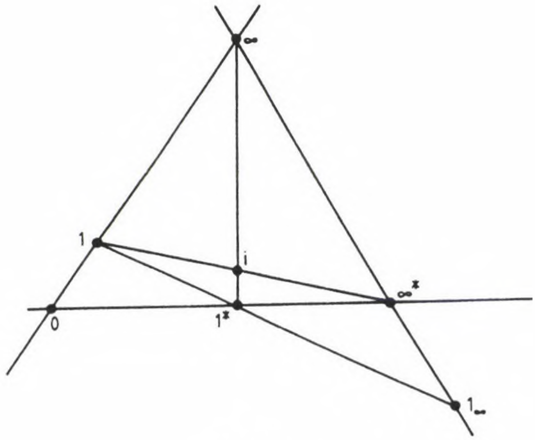


Fig. 1

For each $a \in R$ set

$$a^* = (a \vee 1_\infty) \wedge (0 \vee \infty^*).$$

Then, given $a, b \in R$, we set

$$a + b = [((a^* \vee \infty) \wedge (b \vee \infty^*)) \vee 1_\infty] \wedge (0 \vee \infty).$$

(See Fig. 2.) Then $\langle R, +, 0\rangle$ is a loop with 0 the identity; $0 + a = a + 0$ for all $a \in R$.

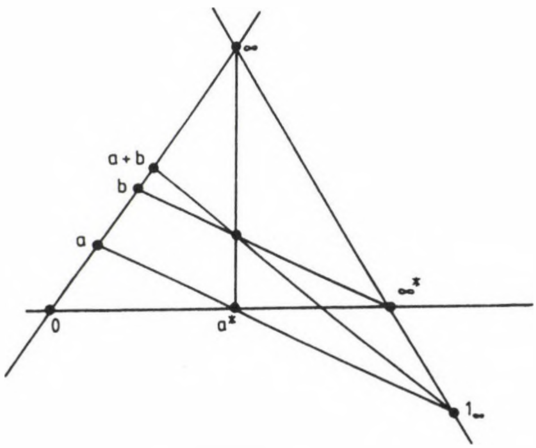


Fig. 2

For each positive integer n and each $a \in R$ we define na by setting

$$1a = a$$

and

$$(n+1)a = a + na$$

(see Fig. 3), that is, we associate on the right.

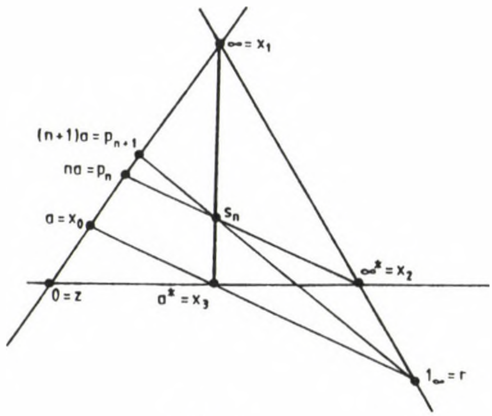


Fig. 3

If the projective geometry G satisfies Desargues' Theorem then R is a division ring under the additive and multiplicative operations, and any two such division rings determined by two different coordinate systems in G are isomorphic. Otherwise, two different coordinate systems generally yield non-isomorphic ternary rings.

3. The identities. We define the 4-ary lattice polynomials z, r, s_n, p_n by setting

$$\begin{gathered} z(x_0, x_1, x_2, x_3) = (x_0 \vee x_1) \wedge (x_2 \vee x_3), \\ r(x_0, x_1, x_2, x_3) = (x_0 \vee x_3) \wedge (x_1 \vee x_2), \\ p_1(x_0, x_1, x_2, x_3) = x_0, \\ s_n(x_0, x_1, x_2, x_3) = (p_n(x_0, x_1, x_2, x_3) \vee x_2) \wedge (x_1 \vee x_3), \\ p_{n+1}(x_0, x_1, x_2, x_3) = (r(x_0, x_1, x_2, x_3) \vee s_n(x_0, x_1, x_2, x_3)) \wedge (x_0 \vee x_1). \end{gathered}$$

We define two identities for $n > 1$,

$$\sigma_n: \quad z(x_0, x_1, x_2, x_3) \leqq p_n(x_0, x_1, x_2, x_3)$$

and

$$\tau_n: \quad z(x_0, x_1, x_2, x_3) \leqq x_0 \vee p_{n+1}(x_0, x_1, x_2, x_3).$$

LEMMA 2. *Let $\langle 0, i, \infty, \infty^* \rangle$ be a coordinate system in the projective geometry G, let $a \in R$, the associated ternary ring, and let $n > 1$. Then*

$$na = 0$$

if and only if

$$z(a, \infty, \infty^*, a^*) \leqq p_n(a, \infty, \infty^*, a^*)$$

and, if $a \neq 0$, then

$$na \neq 0$$

if and only if

$$z(a, \infty, \infty^*, a^*) \leqq a \vee p_{n+1}(a, \infty, \infty^*, a^*).$$

PROOF. By definition of a^*,

$$z = z(a, \infty, \infty^*, a^*) = 0$$

and

$$p_n = p_n(a, \infty, \infty^*, a^*) = na$$

(see Fig. 3).

Since both z and p_n are points,

$$z \leqq p_n \Leftrightarrow z = p_n \Leftrightarrow 0 = na$$

and, if $a \neq 0 = z$,

$$z \leqq a \vee p_{n+1} \Leftrightarrow a \neq p_{n+1} \Leftrightarrow a \neq (n+1)a = a + na \Leftrightarrow na \neq 0$$

since $\langle R, +, 0 \rangle$ is a loop, concluding the proof. □

It thus follows that if the lattice $L(G)$ of subspaces of a projective geometry G satisfies the lattice identity σ_n then any ternary coordinate ring of G satisfies the identity $nx = 0$ and that if $L(G)$ satisfies the identity τ_n then any ternary coordinate ring satisfies $x \neq 0 \Rightarrow nx \neq 0$. To prove the converse, we shall make use of Lemma 1. As usual, we are obliged to consider certain degenerate cases.

A set of points $\{a, b, c, d\}$ of the projective geometry G is said to be *non-degenerate* if all are distinct and no three are collinear. Otherwise, the set is said to be *degenerate*.

LEMMA 3. *If $\{a, b, c, d\}$ is a non-degenerate set of coplanar points in the projective geometry G, then G has a coordinate system $\langle 0, i, \infty, \infty^* \rangle$ with $\langle a, b, c, d \rangle = \langle 1, \infty, \infty^*, 1^* \rangle$.*

PROOF. We set $0 = (a \vee b) \wedge (c \vee d)$ (this is a point by coplanarity), and set $\infty = b$, $\infty^* = c$, and $i = (a \vee \infty^*) \wedge (d \vee \infty)$.

LEMMA 4. *If $\{a, b, c, d\}$ is a degenerate set of distinct points then*

$$z(a, b, c, d) \leqq p_n(a, b, c, d)$$

for all $n > 1$.

PROOF. If all a, b, c, d are collinear then $z = r = s_1 = s_n = p_n = a \vee b$ for all $n > 1$. If only a, b, c are collinear then $z = c$, $r = a$, $s_n = b$ for all $n \geqq 1$, and $p_n = a \vee b > c$ for all $n > 1$. If only a, b, d are collinear then $z = d$, $r = b$, $s_1 = a$, $s_n = a \vee b$, and $p_n = a \vee b > d$ for all $n > 1$. If only a, c, d are collinear then $z = a$, $r = c$, $s_n = d$, and $p_n = a$ for all $n \geqq 1$. Finally, if only b, c, d are collinear then $z = b$, $r = d$, $s_1 = c$, $s_n = b \vee c$, and $p_n = b$ for all $n > 1$.

Thus $z \leqq p_n$ in each case, proving the lemma. □

For any other degeneracy, that is, if either at least two of a, b, c, d coincide or if the minimal element $\emptyset \in \{a, b, c, d\}$, the result $z \leqq p_n$ follows from the modular identity alone:

LEMMA 5. *Let L be a modular lattice and let $a_0, a_1, a_2, a_3 \in L$. If either*
1) *there are distinct i, j with $a_i = a_j$,*
2) *there is an i with $a_i \leqq a_j$ for all j*

then

$$z(a_0, a_1, a_2, a_3) \leqq p_n(a_0, a_1, a_2, a_3)$$

for all $n > 1$.

PROOF. If 1) applies, then, for each of the six possibilities, we need only check $z \leqq p_n$ in the free modular lattice on three generators. Similarly, if 2) applies, we need only check each of the four cases in the free modular lattice on three generators, setting the relevant a_i equal to the minimal element of that free lattice. The details are left to the reader. □

THEOREM 1. *Let G be a projective geometry of dimension $\geqq 2$ and let $n > 1$. Then the following three conditions are equivalent:*
1) *The lattice of subspaces $L(G)$ satisfies the identity σ_n.*
2) *All ternary coordinate rings of G satisfy the identity $n1 = 0$.*
3) *All ternary coordinate rings of G satisfy the identity $nx = 0$.*

PROOF. 1) $\Rightarrow$ 3) by Lemma 2.

3) $\Rightarrow$ 2) is clear.

To establish 2) $\Rightarrow$ 1), let $n1 = 0$ in all ternary coordinate rings of G. Note that $z(x_0, x_1, x_2, x_3)$ has no repetition of variables. Thus, by Lemma 1, we need only verify σ_n for substitutions a_0, a_1, a_2, a_3 that are points or the null subspace $\emptyset$. If either $\emptyset \in \{a_0, a_1, a_2, a_3\}$ or $\{a_0, a_1, a_2, a_3\}$ is a degenerate set of points then it follows by Lemmas 4 and 5 that $z(a_0, a_1, a_2, a_3) \leqq$ $\leqq p_n(a_0, a_1, a_2, a_3)$.

So let $\{a_0, a_1, a_2, a_3\}$ be a non-degenerate set of points in G. If these points are non-coplanar then

$$z(a_0, a_1, a_2, a_3) = \emptyset \leqq p_n(a_0, a_1, a_2, a_3).$$

If they are coplanar then, by Lemma 3, there is a coordinate system $\langle 0, i, \infty, \infty^* \rangle$ with $\langle a_0, a_1, a_2, a_3 \rangle = \langle 1, \infty, \infty^*, 1^* \rangle$, and, by Lemma 2,

$$z(a_0, a_1, a_2, a_3) = 0 = n1 = p_n(a_0, a_1, a_2, a_3).$$

Consequently, 2) $\Rightarrow$ 1), concluding the proof of the theorem. □

By the obvious similar proof we get:

THEOREM 2. *Let G be a projective geometry of dimension $\geqq 2$ and let $n > 1$. Then the following three conditions are equivalent:*

1) *The lattice of subspaces $L(G)$ satisfies the identity τ_n.*
2) *All ternary coordinate rings of G satisfy $n1 \neq 0$.*
3) *All ternary coordinate rings of G satisfy $x \neq 0 \Rightarrow nx \neq 0$.*

In the desarguesian case we get the following:

COROLLARY. *Let G be a desarguesian projective geometry of dimension $\geqq 2$, and let R be the coordinate division ring of G.*

R has characteristic p for some prime p if and only if the lattice $L(G)$ satisfies the identity σ_p.

R has characteristic 0 if and only if $L(G)$ satisfies the set of identities $\{\tau_p \mid p \text{ prime}\}$.

The usual model-theoretic argument shows that characteristic 0 cannot be determined by a finite set of identities.

REFERENCES

[1] FRIED, E., GRÄTZER, G. and LAKSER, H., Projective geometries as cover-preserving sublattices, *Algebra Universalis* **27** (1990), 270–278. *MR* **91c**: 06018

[2] GRÄTZER, G., *General Lattice Theory*, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Math. Reihe, Bd. 52, Birkhäuser Verlag, Basel, 1978. *MR* **80c**: 06001a

[3] GRÄTZER, G., and LAKSER, H., Three remarks on the arguesian identity, Abstract, *Notices Amer. Math. Soc.* **20** (1973), A-253, A-313.

[4] HERRMANN, C. and HUHN, A. P., Zum Begriff der Charakteristik modularer Verbände, *Math. Z.* **144** (1975), 185–194. *MR* **52** #5503

[5] HUTCHINSON, G. and CZÉDLI, G., A test for identities satisfied in lattices of submodules, *Algebra Universalis* **8** (1978), 269–309. *MR* **57** #9621

[6] HUGHES, D. R. and PIPER, F. C., *Projective Planes*, Graduate Texts in Mathematics, Vol. 6, Springer-Verlag, New York, N. Y., 1973. *MR* **48** #12278

(Received September 5, 1988)

DEPARTMENT OF MATHEMATICS AND ASTRONOMY
UNIVERSITY OF MANITOBA
341 MACHRAY HALL
WINNIPEG, MANITOBA
R3T 2N2
CANADA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 307–312

CONTINUITY AND TRANSFINITE SEQUENCES OF MAPS

J. EWERT

Let X denote a topological space and let Ω be the first uncountable ordinal number.

A transfinite sequence $\{a_\xi : \xi < \Omega\}$ of elements of X is said to be convergent to $a \in X$ if for each neighbourhood U of a there exists $\xi_0 < \Omega$ such that $a_\xi \in U$ for each ξ, $\xi_0 \leqq \xi < \Omega$, [5].

In the sequel we will use the following

1. LEMMA [5]. *Let X be a first countable T_1-space. If a transfinite sequence $\{a_\xi : \xi < \Omega\}$ converges to $a \in X$, then there exists $\xi_0 < \Omega$ such that $a_\xi = a$ for every ξ, $\xi_0 \leqq \xi < \Omega$.*

A transfinite sequence $\{f_\xi : \xi < \Omega\}$ of maps of X into a topological space (Y, τ) is called convergent to a map $f : X \to Y$ if for every $x \in X$ the sequence $\{f_\xi(x) : \xi < \Omega\}$ converges to $f(x)$, [10]; then we write $f = \tau - \lim\limits_{\xi<\Omega} f_\xi$ or shortly $f = \lim\limits_{\xi<\Omega} f_\xi$.

A topological space X is called a sequential space if a set $A \subset X$ is closed if and only if together with any sequence it contains all its limits [1, p. 78]. It is known that every first countable space is a sequential space but the converse is not true [1, Examples 1.6.18 and 1.6.19].

2. LEMMA [1, Prop. 1.6.15]. *A map f of a sequential space X into a topological space Y is continuous if and only if $f(\lim\limits_{n\to\infty} x_n) \subset \lim\limits_{n\to\infty} f(x_n)$ for every sequence $\{x_n : n \geqq 1\}$ in the space X.*

A set Y with two topologies τ_1 and τ_2 is called a bitopological space and it is denoted by (Y, τ_1, τ_2), [3].

3. THEOREM. *Let X be a sequential space and (Y, τ_1, τ_2) a bitopological one in which (Y, τ_2) is a first countable T_1-space. If $f_\xi : X \to Y$ is a τ_1-continuous map for $\xi < \Omega$ and $f = \tau_2 - \lim\limits_{\xi<\Omega} f_\xi$, then f is τ_1-continuous.*

PROOF. Assume that f is not τ_1-continuous at a point $x_0 \in X$. According to Lemma 2 there exists a sequence $\{x_n : n \geqq 1\}$ of elements of X such that

---

1980 *Mathematics Subject Classification* (1985 Revision). Primary 54C60.

Key words and phrases. Multivalued map, upper and lower semicontinuity, transfinite sequence.

Akadémiai Kiadó, Budapest

$x_0 \in \lim\limits_{n\to\infty} x_n$ and $f(x_0) \notin \tau_1 - \lim\limits_{n\to\infty} f(x_n)$. Without loss of generality we can assume that there exists a τ_1-neighbourhood V of $f(x_0)$ such that $f(x_n) \notin V$ for $n \geqq 1$. From Lemma 1 we can choose numbers $\alpha_n < \Omega$ for $n=0,1,\ldots$, for which it holds $f_\xi(x_n)=f(x_n)$, $\xi \geqq \alpha_n$, $n=0,1,\ldots$. Using properties of ordinal numbers we can choose $\xi_0 < \Omega$ satisfying $\alpha_n < \xi_0$ for $n=0,1,\ldots$. So we have $f_\xi(x_n)=f(x_n)$ for $n=0,1,\ldots$, $\xi \geqq \xi_0$. Let $\xi \geqq \xi_0$ be established. Since f_ξ is τ_1-continuous we have $f_\xi(U) \subset V$ for some neighbourhood U of x_0. Thus we can take $n_0 \geqq 1$ such that $f_\xi(x_n) \in V$ for $n \geqq n_0$. Hence $f(x_n)=f_\xi(x_n) \in V$ for $n \geqq n_0$ what is the contradiction finishing the proof.

If X is first countable and $\tau_1=\tau_2$, then Theorem 3 gives Corollary 1 in [7]. For metric spaces X and Y it is Theorem 1 in [10], and if $X=Y=R$ we obtain the result contained in [9].

In a topological space (Y,τ) the symbol $\mathcal{C}(Y,\tau)$ or shortly $\mathcal{C}(Y)$ is used to denote the family of all non-empty closed subsets. For any $U \in \tau$ we denote

$$\langle U\rangle = \{A \in \mathcal{C}(Y) \colon A \subset U\}$$
$$\langle U,Y\rangle = \{A \in \mathcal{C}(Y) \colon A \cap U \neq \emptyset\}.$$

$\tilde{\tau}$ is the Vietoris topology in $\mathcal{C}(Y)$ induced by τ. Moreover τ^+ and τ^- denote topologies in $\mathcal{C}(Y)$ induced by the base $\{\langle U\rangle \colon U \in \tau\}$ and the subbase $\{\langle U,Y\rangle \colon U \in \tau\}$, respectively.

If F, F_ξ, $\xi < \Omega$ are multivalued maps defined on a topological space X with closed values in (Y,τ), we will write $F, F_\xi \colon X \to Y$. The transfinite sequence $\{F_\xi \colon \xi < \Omega\}$ of multivalued maps is called convergent to F if for each $x \in X$ the sequence $\{F_\xi(x) \colon \xi < \Omega\}$ is convergent to $F(x)$ in the space $(\mathcal{C}(Y),\tilde{\tau})$. As a consequence of Theorem 3 we have the following corollaries:

4. Corollary. *Let X be a sequential space and (Y,τ) a topological one such that $(\mathcal{C}(Y),\tilde{\tau})$ is a first countable T_1-space. If $F_\xi, F \colon X \to Y$ are multivalued maps with closed values, $F = \tilde{\tau} - \lim\limits_{\xi<\Omega} F_\xi$ and F_ξ are upper (lower) semicontinuous, then F is upper (lower) semicontinuous.*

Proof. A multivalued map F_ξ is upper (lower) semicontinuous if and only if the single valued map $F_\xi \colon X \to (\mathcal{C}(Y),\tau^+,\tilde{\tau})$ is τ^+-continuous (resp. $F_\xi \colon X \to (\mathcal{C}(Y),\tau^-,\tilde{\tau})$ is τ^--continuous). Thus the conclusion follows from Theorem 3.

A multivalued map $F \colon X \to Y$ is said to be upper (lower) c-continuous at $x_0 \in X$ if for each open set $V \subset Y$ with $Y \setminus V$ compact such that $F(x_0) \subset \subset V$ (resp. $F(x_0) \cap V \neq \emptyset$) there exists a neighbourhood U of x_0 such that $F(x) \subset V$ (resp. $F(x) \cap V \neq \emptyset$) for $x \in U$, [2]. A map F is called upper (lower) c-continuous if it is upper (lower) c-continuous at each point.

5. Corollary. *Let X be a sequential space and (Y,τ) a topological one such that $(\mathcal{C}(Y),\tilde{\tau})$ is a first countable T_1-space. If $F_\xi, F \colon X \to Y$ are*

multivalued maps with closed values, $F = \tilde{\tau} - \lim_{\xi<\Omega} F_\xi$ and F_ξ are upper (lower) c-continuous, then F is upper (lower) c-continuous.

PROOF. The family $\tau_c = \{W \in \tau : Y \setminus W \text{ is compact}\} \cup \{\emptyset\}$ is a topology on Y. A multivalued map F_ξ is upper (lower) c-continuous if and only if the single valued map $F_\xi : X \to (\mathcal{C}(Y,\tau),(\tau_c)^+)$, (resp. $F_\xi : X \to (\mathcal{C}(Y,\tau),(\tau_c)^-)$) is continuous. Thus it suffices to apply Theorem 3 to the single valued maps F_ξ with values in $(\mathcal{C}(Y,\tau),(\tau_c)^+,\tilde{\tau})$ or $(\mathcal{C}(Y,\tau),(\tau_c)^-,\tilde{\tau})$, respectively.

6. COROLLARY. *Let X be a sequential space and let $f_\xi : X \to R$, $\xi < \Omega$ be real functions. If $f = \lim_{\xi<\Omega} f_\xi$ and f_ξ are upper (lower) semicontinuous, then f is upper (lower) semicontinuous.*

PROOF. Let τ denote the natural topology on R and let $\tau_1 = \{\emptyset, R\} \cup \cup \{(-\infty,a) : a \in R\}$, $\tau_2 = \{\emptyset, R\} \cup \{(a,\infty) : a \in R\}$. Then it suffices to apply Theorem 3 to functions with values in (R,τ_1,τ) or (R,τ_2,τ), respectively.

If X is a metric space, then Corollary 6 gives Theorem 1″ in [10].

For a bitopological space (Y,τ_1,τ_2) we denote by $\mathcal{Z}(Y,\tau_i)$ the class of all non-empty τ_i-closed compact subsets of Y. If $F : X \to Y$ is a multivalued map, then the symbols $C^+(F,\tau_i)$ and $C^-(F,\tau_i)$ are used to denote the sets of points at which F is upper or lower τ_i-semicontinuous, respectively.

In a bitopological space (Y,τ_1,τ_2) the topology τ_2 is said to be regular with respect to τ_1 if for each τ_2-open set U and each $x \in U$ there exists a set $V \in \tau_2$ such that $x \in V \subset \overline{V}^{(1)} \subset U$, where $\overline{V}^{(1)}$ denotes the τ_1-closure of V, [3].

7. THEOREM. *Let X be a first countable locally separable space, (Y,τ_1,τ_2) a bitopological one such that τ_2 is regular with respect to τ_1 and $(\mathcal{Z}(Y,\tau_2),\tilde{\tau}_2)$ be a first countable T_1-space. Let $F, F_\xi : X \to Y$, $\xi < \Omega$ be multivalued maps with values in $\mathcal{Z}(Y,\tau_2)$ and $F = \tilde{\tau}_2 - \lim F_\xi$. If each F_ξ is lower τ_1-semicontinuous, then*

$$C^+(F,\tau_2) = \bigcap_{\xi<\Omega} \bigcup_{\substack{\alpha \geqq \xi \\ \alpha<\Omega}} C^+(F_\alpha,\tau_2) = \bigcup_{\xi<\Omega} \bigcap_{\substack{\alpha \geqq \xi \\ \alpha<\Omega}} C^+(F_\alpha,\tau_2).$$

PROOF. Evidently we have

$$\bigcup_{\xi<\Omega} \bigcap_{\substack{\alpha \geqq \xi \\ \alpha<\Omega}} C^+(F_\alpha,\tau_2) \subset \bigcap_{\xi<\Omega} \bigcup_{\substack{\alpha \geqq \xi \\ \alpha<\Omega}} C^+(F_\alpha,\tau_2). \tag{1}$$

Now, suppose $x_0 \in \bigcap_{\xi<\Omega} \bigcup_{\substack{\alpha \geqq \xi \\ \alpha<\Omega}} C^+(F_\alpha,\tau_2) \setminus C^+(F,\tau_2)$. Then we can choose a sequence $\{x_n : n \geqq 1\}$ in X converging to x_0 and a τ_2-open set V containing $F(x_0)$ such that $F(x_n) \cap (Y \setminus V) \neq \emptyset$ for $n = 1,2,\ldots$. Using analogous arguments as in the proof of Theorem 3 we can choose an $\alpha < \Omega$ such that $x_0 \in C^+(F_\alpha,\tau_2)$ and $F_\alpha(x_n) = F(x_n)$ for $n = 0,1,2,\ldots$, what is a contradiction. Thus we have shown

$$\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{+}(F_{\alpha},\tau_2)\subset C^{+}(F,\tau_2). \tag{2}$$

So it suffices to prove

$$C^{+}(F,\tau_2)\subset\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{+}(F_{\alpha},\tau_2). \tag{3}$$

Assume that $x_0\in C^{+}(F,\tau_2)\setminus\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{+}(F_{\alpha},\tau_2)$. Let U be a separable neighbourhood of x_0 and $\{x_n: n\geqq 1\}$ a dense subset of U. We can choose $\alpha<\Omega$ such that $F_\alpha(x_n)=F(x_n)$ for $n=0,1,2,\ldots$ and $x_0\notin C^{+}(F_\alpha,\tau_2)$. Hence there exists a τ_2-open set V containing $F_\alpha(x_0)$ such that each neighbourhood U' of x_0 contains a point x' for which $F_\alpha(x')\cap(Y\setminus V)\neq\emptyset$ holds. Since $F(x_0)$ is τ_2-compact and τ_2 is regular with respect to τ_1 we have $F(x_0)\subset\subset W\subset\overline{W}^{(1)}\subset V$ for some τ_2-open set W. The condition $x_0\in C^{+}(F,\tau_2)$ implies that there is a neighbourhood U_1 of x_0, $U_1\subset U$ such that $F(x)\subset W$ for $x\in U_1$. On the other hand for some $z\in U_1$ we have $F_\alpha(z)\cap(Y\setminus\overline{W}^{(1)})\neq\emptyset$. Using the lower τ_1-semicontinuity of F_α at z we can take a neighbourhood U_2 of z, $U_2\subset U_1$ such that $F_\alpha(x)\cap(Y\setminus\overline{W}^{(1)})\neq\emptyset$ for $x\in U_2$. Thus for $x_n\in\in U_2$ we have $F_\alpha(x_n)\cap(Y\setminus\overline{W}^{(1)})\neq\emptyset$ and $F_\alpha(x_n)=F(x_n)\subset W$ what is the contradiction finishing the proof.

By the same way we obtain

8. THEOREM. *Let X be a first countable locally separable space, (Y,τ_1,τ_2) a bitopological one such that τ_2 is regular with respect to τ_1 and $(\mathcal{C}(Y,\tau_2),\tilde{\tau}_2)$ a first countable T_1-space. Let $F,F_\xi\colon X\to Y$, $\xi<\Omega$ be multivalued maps with values in $\mathcal{C}(Y,\tau_2)$ and $F=\tilde{\tau}_2-\lim_{\xi<\Omega}F_\xi$. If F_ξ are upper τ_1-semicontinuous, then:*

$$C^{-}(F,\tau_2)=\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{-}(F_{\alpha},\tau_2)=\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{-}(F_{\alpha},\tau_2).$$

9. COROLLARY. *Let X be a first countable locally separable space, (Y,τ) a locally compact second countable T_2-space. Let $F,F_\xi\colon X\to Y$, $\xi<\Omega$ be multivalued maps with compact values and $F=\tilde{\tau}-\lim_{\xi<\Omega}F_\xi$.*

(a) *If F_ξ are lower c-continuous, then*

$$C^{+}(F,\tau)=\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{+}(F_{\alpha},\tau)=\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}} C^{+}(F_{\alpha},\tau).$$

(b) *If F_ξ are upper c-continuous, then*

$$C^-(F,\tau)=\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^-(F_\alpha,\tau)=\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^-(F_\alpha,\tau).$$

PROOF. Let us consider the topology $\tau_c=\{W\in\tau: Y\setminus W \text{ is } \tau\text{-compact}\}\cup\cup\{\emptyset\}$. If a set $A\subset Y$ is relatively τ-compact, then $\overline{A}=\overline{A}^{(c)}$, where $\overline{A}^{(c)}$ denotes the τ_c-closure of A. This fact implies that in the bitopological space (Y,τ_c,τ) the topology τ is regular with respect to τ_c. Under assumptions on (Y,τ), $(\mathcal{Z}(Y,\tau),\tilde{\tau})$ is a first countable T_1-space. Moreover the lower (upper) c-continuity means the lower (upper) τ_c-semicontinuity; thus the conclusions simply follow from Theorem 7 and 8, respectively.

Now let E be a normed space with the norm topology τ and the weak topology τ_w. Since for each open ball K we have $\overline{K}=\overline{K}^{(w)}$ where $\overline{K}^{(w)}$ is the τ_w-closure of K; in the bitopological space (E,τ_w,τ) the topology τ is regular with respect to τ_w. Moreover the space $(\mathcal{Z}(Y,\tau),\tilde{\tau})$ is T_1 first countable. So for upper or lower weakly semicontinuous (i.e. τ_w-semicontinuous) maps we have

10. COROLLARY. *Let X be a first countable locally separable space and (E,τ_w,τ) a normed space with the weak and the norm topology. Let $F,F_\xi: X\to\to E$, $\xi<\Omega$, be multivalued maps with compact values and $F=\tilde{\tau}-\lim_{\xi<\Omega}F_\xi$.*

(a) *If F_ξ are lower weakly semicontinuous, then*

$$C^+(F,\tau)=\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^+(F_\alpha,\tau)=\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^+(F_\alpha,\tau).$$

(b) *If F_ξ are upper weakly semicontinuous, then*

$$C^-(F,\tau)=\bigcap_{\xi<\Omega}\bigcup_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^-(F_\alpha,\tau)=\bigcup_{\xi<\Omega}\bigcap_{\substack{\alpha\geqq\xi\\ \alpha<\Omega}}C^-(F_\alpha,\tau).$$

Let X,Y be topological spaces; a multivalued map $F: X\to Y$ is said to be upper (lower) quasi-continuous at $x_0\in X$ if for each open set $V\subset Y$ with $F(x_0)\subset V$ (resp. $F(x_0)\cap V\neq\emptyset$) and for each neighbourhood U of x_0 there exists an open non-empty set $U_1\subset U$ such that $F(x)\subset V$ ($F(x)\cap V\neq\emptyset$) for $x\in U_1$, [6, 8]. F is called upper (lower) quasi-continuous if it is upper (lower) quasi-continuous at each point. If f is a single-valued map, then upper and lower quasi-continuity coincide with quasi-continuity defined by Kempisty [4].

11. REMARK. Theorems 7, 8 and Corollaries 9 and 10 remain true if the upper (lower) semicontinuity of F_ξ is replaced by the upper (lower) quasi-continuity with respect to suitable topologies.

In particular for single-valued maps we obtain

12. COROLLARY. *Let X be a first countable locally separable space and Y a first countable T_3-space. If $f_\xi: X\to Y$, $\xi<\Omega$, are quasi-continuous maps and $f=\lim_{\xi<\Omega}f_\xi$, then*

$$C(f) = \bigcap_{\xi < \Omega} \bigcup_{\substack{\alpha \geqq \xi \\ \alpha < \Omega}} C(f_\alpha) = \bigcup_{\xi < \Omega} \bigcap_{\substack{\alpha \geqq \xi \\ \alpha < \Omega}} C(f_\alpha),$$

where $C(f)$ is the set of all points at which f is continuous.

Finally let us observe that the last result is not true for usual sequences (even for uniformly convergent sequences) of real functions.

13. Example. Let us take the functions $f_n, f\colon R \to R$ defined by $f(x) = 0$ for each $x \in R$ and

$$f_n(x) = \begin{cases} \frac{1}{n} & \text{for } x \in \bigcup_{k=0}^{\infty} [2k, 2k+1] \cup \bigcup_{k=1}^{\infty} [-2k, -2k+1] \\ 0 & \text{for } x \in \bigcup_{k=0}^{\infty} (2k-1, 2k) \cup \bigcup_{k=1}^{\infty} (-2k-1, -2k). \end{cases}$$

Then the f_n's are quasi-continuous functions, $f_n \to f$ uniformly and $C(f) \not\subset \bigcap_{n=1}^{\infty} \bigcup_{m \geqq n} C(f_m)$.

REFERENCES

[1] Engelking, R., *General Topology*, Monografie Matematyczne, Tom 60, PWN-Polish Scientific Publishers, Warszawa, 1977. *MR* **58** #18316b

[2] Ewert, J., On c-semicontinuous multivalued maps (to appear).

[3] Kelly, J. C., Bitopological spaces, *Proc. London Math. Soc.* (3) **13** (1963), 71–89. *MR* **26** #729

[4] Kempisty, S., Sur les fonctions quasicontinues, *Fund. Math.* **19** (1932), 184–197. *Zbl.* **5**, 198

[5] Kostyrko, P., On convergence of transfinite sequences, *Mat. Časopis Sloven. Akad. Vied* **21** (1971), 233–239. *MR* **46** #842

[6] Neubrunn, T., On quasicontinuity of multifunctions, *Math. Slovaca* **32** (1982), 147–154. *MR* **83h**: 54025

[7] Neubrunnová, A., A unified approach to the transfinite convergence and generalized continuity, *Acta Math. Univ. Comenianae* **44–45** (1984), 159–168. *MR* **86i**: 54010

[8] Popa, V., Asupra unei descompuneri quasicontinuităţii multifuncţiilor, *Stud. Cerc. Mat.* **27** (1975), 323–328. *MR* **54** #1162

[9] Sierpiński, W., Sur les suites transfinies convergentes de fonctions de Baire, *Fund. Math.* **1** (1920), 132–141. *Jb. Fortschritte Math.* **47**, 237

[10] Šalat, T., On transfinite sequences of B-measurable functions, *Fund. Math.* **78** (1973), 157–162. *MR* **48** #4991

(Received November 4, 1988)

WYŻSZA SZKOŁA PEDAGOGICZNA
WYDZIAŁ MATEMATYCZNO-PRZYRODNICZY
UL. ARCISZEWSKIEGO 22A
PL–76-200 SŁUPSK
POLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 313–330

ON THE LINE METHOD APPROXIMATIONS TO THE CAUCHY PROBLEM FOR PARABOLIC DIFFERENTIAL-FUNCTIONAL EQUATIONS

Z. KAMONT

Abstract

The initial value problem

$$\text{(i)}\quad \begin{aligned} &D_x z(x,y) = f(x,y,z(x,y),z,D_y z(x,y),D_{yy}z(x,y)),\ (x,y)\in[0,a]\times R^n,\\ &z(x,y)=\omega(x,y)\quad\text{for}\quad (x,y)\in[-\tau_0,0]\times R^n,\end{aligned}$$

is treated with the longitudinal method of lines. We consider a general class of differential-difference schemes corresponding to (i)

$$\text{(ii)}\quad \begin{aligned} &D_x z^{(m)}(x) = \phi_h(x,y^{(m)},z^{(m)}(x),z,\Delta z^{(m)}(x),\Delta^{(2)}z^{(m)}(x)),\\ &\qquad\qquad x\in[0,a],\ m\in Z,\\ &z^{(m)}(x)=\varphi_h^{(m)}(x)\quad\text{for}\quad x\in[-\tau_0,0],\ m\in Z,\end{aligned}$$

where Δ and $\Delta^{(2)}$ are difference operators. The aim of the paper is to give sufficient conditions for the convergence of the sequence $\{u_h\}$ where u_h is a solution of (ii), to a solution u of (i). We prove that if the method (ii) is stable and satisfies a consistency condition with respect to (i) then it is convergent. We assume that f and ϕ_h satisfy the Volterra condition. The proof of the convergence is based on theorems on countable systems of differential-functional inequalities.

1. Introduction

Let us denote by $C(X,Y)$ the class of all continuous functions with domain X and range in Y; X, Y being arbitrary metric spaces. Let $E= [0,a]\times R^n$, $E^{(0)}=[-\tau_0,0]\times R^n$ where $a>0$, $\tau_0\geqq 0$ and $\Omega=E\times R\times C(E^{(0)}\cup E,R)\times R^n\times R^n$. If $z\colon E^{(0)}\cup E\to R$ is a function of the variables (x,y), $y=(y_1,\dots,y_n)$, and there exist derivatives $D_{y_i}z$, $D_{y_iy_i}z$, $i=1,\dots,n$, then we denote $D_y z=(D_{y_1}z,\dots,D_{y_n}z)$ and $D_{yy}z=(D_{y_1y_1}z,D_{y_2y_2}z,\dots,D_{y_ny_n}z)$. Assume that $f\colon\Omega\to R$ and $\omega\colon E^{(0)}\to R$ are given functions. We consider

---

1980 *Mathematics Subject Classification* (1985 Revision). Primary 65M20; Secondary 35K15, 35R10.

Key words and phrases. Differential-functional inequalities, discretization with respect to spatial variables, discretization error, consistency condition, stable methods, Volterra condition.

Akadémiai Kiadó, Budapest

the differential-functional problem

$$(1) \qquad \begin{aligned} &D_x z(x,y) = f(x,y,z(x,y),z,D_y z(x,y),D_{yy}z(x,y)),\\ &z(x,y)=\omega(x,y) \quad \text{for} \quad (x,y)\in E^{(0)}. \end{aligned}$$

For $(x,y)\in E$ we define $H[x,y]=\{(t,s)=(t,s_1,\dots,s_n): t\in[-\tau_0,x],\ |s_i|\leqq |y_i| \text{ for } i=1,\dots,n\}$. Assume that the differential-functional problem (1) is the Volterra type, i.e. $f(x,y,p,z,q,r)=f(x,y,p,\overline{z},q,r)$ for $(x,y,p,q,r)\in E\times R\times R^n\times R^n$, $z,\overline{z}\in C(E^{(0)}\cup E,R)$ and $z\,|_{H[x,y]}=\overline{z}\,|_{H[x,y]}$. The problem (1) contains as particular cases, initial-value problems with a retarded argument. Differential-integral equations can be obtained from (1) by specializing the operator f.

The method of lines for partial differential equations consists in replacing spatial derivatives by difference expressions. Then the partial equation is transformed into a system of ordinary differential equations. The line method approximations for nonlinear differential problems of parabolic type were examined in [8]–[10], [12], [14]. The monograph [14] contains a large bibliography. The line method is also treated as a tool for proving of existence theorems for differential problems corresponding to parabolic equations [11], [13] or first order hyperbolic systems [4] (see also [15], [16]). A simple example of the line method for nonlinear differential-functional equations was considered in [3].

We introduce a general class of differential-difference scheme corresponding to (1). They are characterized by increment function ϕ_h and difference operators $\Delta=(\Delta_1,\dots,\Delta_n)$, $\Delta^{(2)}=(\Delta^{(2)}_{11},\dots,\Delta^{(2)}_{nn})$. The main problem in our investigations of a general class of line methods is to find such a differential-difference scheme which is stable and satisfies some consistency condition with respect to (1). We prove that if a nonlinear differential-difference method of the parabolic type is stable and satisfies a consistency condition with respect to (1) along solutions of (1) then it is convergent. The basic tool in our investigations are theorems on countable systems of differential-functional inequalities. In the first part of the paper we prove a comparison result for these systems. It will be a generalization of an adequate result from [8] where an initial value problem for a parabolic equation in two independent variables was considered. If we assume in our theorems that the right-hand sides of equations do not include the functional argument, then we obtain more general theorems than those results in [8], [12], [14]. Existence, uniqueness and convergence properties of the line method approximations are investigated under the classical assumption that the initial function $\omega\colon E^{(0)}\to R$ satisfies the condition $|\omega(x,y)|\leqq \text{const}\exp[B\|y\|^2]$, $(x,y)\in E^{(0)}$, where $\|\cdot\|$ is the Euclidean norm in R^n.

2. Differential-difference inequalities

The following assumption will be needed throughout the paper.

ASSUMPTION H_0. Suppose that the sequence $\varrho = \{\varrho_i\}_{i=0}^{\infty}$ satisfies the conditions:

1° $\varrho_i \in R_+$ for $i = 0, 1, 2, \ldots$, $R_+ = [0, +\infty)$, ϱ is strictly monotone increasing and $\lim\limits_{i\to\infty} \varrho_i = +\infty$,

2° there exists $b \in R_+$ such that for $i = 1, 2, \ldots$, we have

$$
\begin{aligned}
&\varrho_i(\varrho_{i+1} - \varrho_{i-1}) \leqq b,\\
&\varrho_i(\varrho_{i+1} - 2\varrho_i + \varrho_{i-1}) \leqq b.
\end{aligned}
\tag{2}
$$

We define a mesh on R^n in the following way. Suppose that Assumption H_0 is satisfied. Let $h > 0$ and

$$
\begin{aligned}
y_i^{(j)} &= h\varrho_j, \quad j = 0, 1, 2, \ldots, \quad i = 1, \ldots, n,\\
y_i^{(j)} &= -h\varrho_j, \quad j = -1, -2, \ldots, \quad i = 1, \ldots, n.
\end{aligned}
\tag{3}
$$

For $m = (m_1, \ldots, m_n)$ where m_i, $i = 1, \ldots, n$, are integers, we define $y^{(m)} = \left(y_1^{(m_1)}, \ldots, y_n^{(m_n)}\right)$ with $y_i^{(m_i)}$ given by (3). Let $Z = \{m = (m_1, \ldots, m_n)$: $m_j = 0, \pm 1, \pm 2, \ldots$ for $j = 1, \ldots, n\}$ and $E_h = \{(x, y^{(m)})\colon x \in [0, a],\ m \in Z\}$, $E_h^{(0)} = \{(x, y^{(m)})\colon x \in [-\tau_0, 0],\ m \in Z\}$. If $1 \leqq j \leqq n$ and $m \in Z$ then we denote

$$
\begin{aligned}
j(m) &= (m_1, \ldots, m_{j-1}, m_j + 1, m_{j+1}, \ldots, m_n),\\
-j(m) &= (m_1, \ldots, m_{j-1}, m_j - 1, m_{j+1}, \ldots, m_n).
\end{aligned}
$$

For a function $z\colon E_h^{(0)} \cup E_h \to R$ we write $z^{(m)}(x) = z(x, y^{(m)})$, $(x, y^{(m)}) \in E_h^{(0)} \cup E_h$. In order to replace spatial derivatives by difference expressions we introduce the operators $\Delta = (\Delta_1, \ldots, \Delta_n)$ and $\Delta^{(2)} = (\Delta_{11}^{(2)}, \Delta_{22}^{(2)}, \ldots, \Delta_{nn}^{(2)})$. If $z\colon E_h^{(0)} \cup E_h \to R$ then

$$
\Delta_i z^{(m)}(x) = \left(y_i^{(m_i+1)} = y_i^{(m_i-1)}\right)^{-1} \left(z^{(i(m))}(x) - z^{(-i(m))}(x)\right), \qquad i = 1, \ldots, n
\tag{4}
$$

and

$$
\Delta_{ii}^{(2)} z^{(m)}(x) = A_i^{(m)} z^{(i(m))}(x) + B_i^{(m)} z^{(m)}(x) + C_i^{(m)} z^{(-i(m))}(x), \qquad i = 1, \ldots, n,
\tag{5}
$$

where

$$
\begin{aligned}
A_i^{(m)} &= 2\left(y_i^{(m_i+1)} - y_i^{(m_i)}\right)^{-1}\left(y_i^{(m_i+1)} - y_i^{(m_i-1)}\right)^{-1},\\
B_i^{(m)} &= -2\left(y_i^{(m_i+1)} - y_i^{(m_i)}\right)^{-1}\left(y_i^{(m_i)} - y_i^{(m_i-1)}\right)^{-1},\\
C_i^{(m)} &= 2\left(y_i^{(m_i)} - y_i^{(m_i-1)}\right)^{-1}\left(y_i^{(m_i+1)} - y_i^{(m_i-1)}\right)^{-1}.
\end{aligned}
\tag{6}
$$

For $z\colon E_h^{(0)} \cup E_h \to R$ we write $\Delta z^{(m)}(x) = (\Delta_1 z^{(m)}(x), \dots, \Delta_n z^{(m)}(x))$, $\Delta^{(2)} z^{(m)}(x) = (\Delta_{11}^{(2)} z^{(m)}(x), \dots, \Delta_{nn}^{(2)} z^{(m)}(x))$, $(x, y^{(m)}) \in E_h^{(0)} \cup E_h$. Let $\mathcal{F}_c(E_h^{(0)}, R)$ denote the class of all functions z defined on $E_h^{(0)}$ taking values in R such that $z(\cdot, y^{(m)}) \in C([-\tau_0, 0], R)$ for $m \in Z$. We shall use vectorial inequalities, with the understanding that the same inequalities are satisfied between their corresponding components.

Assumption H_1. Suppose that

1° $\mathcal{C}_h \in \mathcal{F}_c(E_h^{(0)}, R)$ and there exist constants $A, B \in (0, +\infty)$ such that $|\mathcal{C}_h(x, y^{(m)})| \leqq A \exp[B\|y^{(m)}\|^2]$, $x \in [-\tau_0, 0]$, $m \in Z$,

2° the positive constants a, E, L, C, D satisfy the conditions

$$
D > C \quad \text{if} \quad n = 1, 2 \quad \text{and} \quad D > \frac{n}{2} C \quad \text{if} \quad n \geqq 3,
$$
$$
C > B, \quad E > D \quad \text{and} \quad a = (4EL)^{-1}.
$$

We define for $(x, y) \in [0, a] \times R^n$

$$
\begin{aligned}
g(x, y) &= (1 - 4DLx)^{-1} \exp\left[\frac{C\|y\|^2}{1 - 4DLx}\right],\\
w(x, y) &= g(x, y) \exp\left[-\frac{1}{2} P(y) + \gamma L x\right]
\end{aligned}
\tag{7}
$$

where

$$
\begin{aligned}
P(y) &= \left[\left(\frac{n}{4C}\right)^2 + n\|y\|^2\right]^{\frac{1}{2}}, \quad y \in R^n,\\
\gamma &= 1 + \frac{nE}{4(E - D)} + \frac{n^2}{4}.
\end{aligned}
\tag{8}
$$

We will use the symbol $\mathcal{F}_c(E_h^{(0)} \cup E_h, R)$ to denote the class of all functions $z\colon E_h^{(0)} \cup E_h \to R$ such that $z(\cdot, y^{(m)}) \in C([-\tau_0, a], R)$ for $m \in Z$. Let Γ_h denote the set of all $z \in \mathcal{F}_c(E_h^{(0)} \cup E_h, R)$ such that $|z(x, y^{(m)})| \leqq A_0 \exp[C\|y^{(m)}\|^2]$, $x \in [-\tau_0, 0]$, $m \in Z$, for a certain $A_0 \in R_+$ and

$$
\sup\left\{\frac{|z(x, y^{(m)})|}{g(x, y^{(m)})} : x \in [0, a],\ m \in Z\right\} < +\infty.
$$

We denote $\Omega_h = E_h \times R \times \Gamma_h \times R^n \times R^n$ and $I_0 = (0, b_0)$, $b_0 > 0$. Assume that for each $h \in I_0$ we have $\phi_h : \Omega_h \to R$. We consider the following line method for the problem (1)

$$\begin{aligned} D_x z^{(m)}(x) &= \phi_h(x, y^{(m)}, z^{(m)}(x), z, \Delta z^{(m)}(x), \Delta^{(2)} z^{(m)}(x)), \\ &\qquad x \in (0, a],\ m \in Z, \\ z^{(m)}(x) &= \mathcal{C}_h^{(m)}(x),\ x \in [-\tau_0, 0],\ m \in Z. \end{aligned} \tag{9}$$

If $x \in [0, a]$, $m \in Z$ then we denote $H_h[x, m] = \{(t, y^{(\bar{m})}) \colon t \in [-\tau_0, x],\ \bar{m} = (\bar{m}_1, \dots, \bar{m}_n),\ |\bar{m}_i| \leqq |m_i|,\ i = 1, \dots, n\}$. The function $\phi_h : \Omega_h \to R$ is said to satisfy the Volterra condition on Ω_h if for $x \in (0, a]$, $m \in Z$, $z, \bar{z} \in \Gamma_h$ such that $z\,|_{H_h[x,m]} = \bar{z}\,|_{H_h[x,m]}$ we have $\Phi_h(x, y^{(m)}, p, z, q, r) = \phi_h(x, y^{(m)}, p, \bar{z}, q, r)$ where $(z, y^{(m)}, p, q, r) \in E_h \times R \times R^n \times R^n$. For a function $z \in \Gamma_h$ we define

$$\|z\|_{x,m}^{(h)} = \max\left\{ |z(t, y^{(\bar{m})})| : (t, y^{(\bar{m})}) \in H_h[x, m] \right\}.$$

ASSUMPTION H_2. Suppose that the function $\phi_h : \Omega_h \to R$ of the variables (x, y, p, z, q, r), $q = (q_1, \dots, q_n)$, $r = (r_1, \dots, r_n)$, satisfies the conditions

1° for each $m \in Z$, $(p, q, r) \in R \times R^n \times R^n$, $h \in I_0$, we have $\phi_h(\cdot, y^{(m)}, p, z, q, r) \in C([0, a], R)$ where $z \in \Gamma_h$ and there exists a positive constant η such that

$$\phi_h(x, y, p, z, q, r) - \phi_h(x, y, p, z, q, \bar{r}) \geqq \eta \sum_{i=1}^{n} (r_i - \bar{r}_i) \quad \text{on } \Omega_h$$

where $r \geqq \bar{r}$,

2° ϕ_h is nondecreasing with respect to the functional argument and there exists $L_0 > 0$ such that

$$\begin{aligned} &|\phi_h(x, y, p, z, q, r) - \phi_h(x, y, \bar{p}, \bar{z}, \bar{q}, \bar{r})| \leqq \\ &\leqq L_0 \Big[|p - \bar{p}| + \|z - \bar{z}\|_{x,m}^{(h)} + \sum_{i=1}^{n} |q_i - \bar{q}_i| + \sum_{i=1}^{n} |r_i - \bar{r}_i| \Big] \end{aligned} \tag{10}$$

on Ω_h where $y = y^{(m)}$, assume that $L \geqq 2L_0$,

3° $|\phi_h(x, y^{(m)}, 0, 0, 0, 0)| \leqq A \exp[B\|y^{(m)}\|^2]$, $(x, y^{(m)}) \in E_h$, where A, B are the positive constants from Assumption H_1.

REMARK 1. It follows from (10) that ϕ_h satisfies the Volterra condition on Ω_h.

We define for $\alpha > 0$

$$\begin{aligned} v(x, y^{(m)}) &= \alpha(1 + x) w(x, y^{(m)}) \quad \text{for } (x, y^{(m)}) \in E_h, \\ v(x, y^{(m)}) &= \alpha w(0, y^{(m)}) \quad \text{for } (x, y^{(m)}) \in E_h^{(0)}. \end{aligned} \tag{11}$$

THEOREM 1. *If Assumptions H_0–H_2 are satisfied then there exists $\delta>0$, $\delta=\delta(C,D,E,\eta,L_0,\varrho)$ such that for $0<h<\delta$ we have*

(i) *the differential-difference problem (9) has the quasi-monotone property,*

(ii) *there exists $\alpha>0$ such that the function v given by (11) satisfies the inequalities*

$$D_x v^{(m)}(x) \geqq \phi_h\left(x, y^{(m)}, v^{(m)}(x), v, \Delta v^{(m)}(x), \Delta^{(2)} v^{(m)}(x)\right),$$
$$x\in(0,a], \quad m\in Z,$$

(iii) *if the functions $u,\tilde u\in\Gamma_h$ satisfy the initial inequality $u^{(m)}(x)\leqq\tilde u^{(m)}(x)$, $x\in[-\tau_0,0]$, $m\in Z$, and the differential inequalities*

$$(12)\qquad \begin{aligned} &D_x u^{(m)}(x) \leqq \phi_h\left(x, y^{(m)}, u^{(m)}(x), u, \Delta u^{(m)}(x), \Delta^{(2)} u^{(m)}(x)\right),\\ &D_x \tilde u^{(m)}(x) \geqq \phi_h\left(x, y^{(m)}, \tilde u^{(m)}(x), \tilde u, \Delta \tilde u^{(m)}(x), \Delta^{(2)} \tilde u^{(m)}(x)\right),\\ &x\in(0,a], \quad m\in Z,\end{aligned}$$

hold then $u^{(m)}(x)\leqq\tilde u^{(m)}(x)$ for $x\in[0,a]$, $m\in Z$.

PROOF. Suppose that $x\in(0,a]$, $m\in Z$, $z,\bar z\in\Gamma_h$, $z^{(m)}(x)=\bar z^{(m)}(x)$ and $z^{(\bar m)}(x)\leqq\bar z^{(\bar m)}(x)$ for $\bar m\in Z$. Let $K=\sup\{\varrho_{i+1}-\varrho_i : i=0,1,2,\dots\}$ and $\delta'=\frac{2\eta}{L_0K}$. Then for $0<h<\delta'$ we have

$$\begin{aligned}&\phi_h\left(x, y^{(m)}, \bar z^{(m)}(x), \bar z, \Delta \bar z^{(m)}(x), \Delta^{(2)} \bar z^{(m)}(x)\right) -\\ &-\phi_h\left(x, y^{(m)}, z^{(m)}(x), z, \Delta z^{(m)}(x), \Delta^{(2)} z^{(m)}(x)\right) \geqq\\ &\geqq \sum_{i=1}^{n}[\bar z^{(i(m))}(x)-z^{(i(m))}(x)+\bar z^{(-i(m))}(x)-z^{(-i(m))}(x)]\times\\ &\times\left(\frac{2\eta}{hK}-L_0\right)\left(y_i^{(m_i+1)}-y_i^{(m_i-1)}\right)^{-1}\geqq 0,\end{aligned}$$

which completes the proof of (i).

Now we prove that

$$\begin{aligned}&\left[L^{-1}D_x w(x,y)-\sum_{i=1}^{n} D_{y_iy_i} w(x,y)-\sum_{i=1}^{n}|D_{y_i} w(x,y)|-w(x,y)\right](w(x,y))^{-1}\geqq\\ (13)\qquad &\geqq\left[L^{-1}D_x g(x,y)-\sum_{i=1}^{n} D_{y_iy_i} g(x,y)\right](g(x,y))^{-1}\geqq\\ &\geqq 2(2D-nC)+4C\|y\|^2(D-C)\geqq 0,\end{aligned}$$

where $(x,y)\in(0,a]\times R^n$. It follows from (7), (8) and from the estimation

$$\frac{D_{y_i}w(x,y)}{w(x,y)}=\frac{2Cy_i}{1-4DLx}-\frac{1}{2}D_{y_i}P(y)\geqq$$
$$\geqq\frac{y_i}{2}\left[\frac{4C}{1-4DLx}-\frac{n}{P(y)}\right]\geqq 0\quad\text{for}\quad (x,y)\in(0,a]\times R^n_+,$$

that it suffices to prove (13) on $(0,a]\times R^n_+$. It follows from (7), (8) that

$$(w(x,y))^{-1}\Big[L^{-1}D_xw(x,y)-\sum_{i=1}^{n}D_{y_iy_i}w(x,y)-$$
$$-\sum_{i=1}^{n}|D_{y_i}w(x,y)|-w(x,y)\Big]=$$
$$=(g(x,y))^{-1}\Big[L^{-1}D_xg(x,y)-\sum_{i=1}^{n}D_{y_iy_i}g(x,y)\Big]+\tag{14}$$
$$+(g(x,y))^{-1}\sum_{i=1}^{n}D_{y_i}g(x,y)[D_{y_i}P(y)-1]-$$
$$-\frac{1}{4}\sum_{i=1}^{n}\Big\{[D_{y_i}P(y)-1]^2-2D_{y_iy_i}P(y)\Big\}+\gamma-1+\frac{n}{4},$$
$$(x,y)\in(0,a]\times R^n_+.$$

Since

$$t-t^2(\xi^2+t^2)^{-\frac{1}{2}}\leqq\frac{\xi}{2},\qquad \xi,t\in(0,+\infty),$$

then we obtain

$$\sum_{i=1}^{n}D_{y_i}g(x,y)[1-D_{y_i}P(y)](g(x,y))^{-1}\leqq\frac{n}{4}\frac{1}{1-4DLx},\tag{15}$$
$$(x,y)\in(0,a]\times R^n_+.$$

It follows from (8) that

$$[1-D_{y_i}P(y)]^2-2D_{y_iy_i}P(y)\leqq n+1,\qquad y\in R^n_+,\ i=1,\dots,n.\tag{16}$$

Using (14)-(16) we obtain (13).

We prove that to every $\varepsilon>0$ there exists a δ_1, $\delta_1=\delta_1(\varepsilon,C,D,E,\varrho)$, such that for $0<h<\delta_1$ we have

$$|\Delta_ig(x,y^{(m)})-D_{y_i}g(x,y^{(m)})|\leqq\left(1+|y_i^{(m_i)}|\right)g(x,y^{(m)}),\tag{17}$$
$$x\in(0,a],\quad m\in Z,\quad i=1,\dots,n,$$

and

$$\left|\Delta_{ii}^{(2)} g(x, y^{(m)}) - D_{y_i y_i} g(x, y^{(m)})\right| \leqq$$

(18)

$$\leqq \varepsilon \left[1 + (y_i^{(m_i)})^2\right] g(x, y^{(m)}), \qquad x \in (0, a],\ m \in Z,\ i = 1, \dots, n.$$

Let

$$c = \frac{CE}{E - D}, \quad \tilde{\eta}(x) = \frac{C}{1 - 4DLx}.$$

Suppose that $m_i > 0$. Then we have

$$\left[\Delta_i g(x, y^{(m)}) - D_{y_i} g(x, y^{(m)})\right] (g(x, y^{(m)}))^{-1} =$$

(19)

$$= \left(y_i^{(m_i+1)} - y_i^{(m_i-1)}\right)^{-1} \left\{\exp[\tilde{\eta}(x)[(y_i^{(m_i+1)})^2 - (y_i^{(m_i)})^2]] - \right.$$
$$\left. - \exp[\tilde{\eta}(x)[(y_i^{(m_i-1)})^2 - (y_i^{(m_i)})^2]]\right\} - 2\tilde{\eta}(x) y_i^{(m_i)} \leqq$$
$$\leqq (1 + \chi)\tilde{\eta}(x) \left[y_i^{(m_i+1)} - 2y_i^{(m_i)} + y_i^{(m_i-1)}\right] + 2\chi\tilde{\eta}(x) y_i^{(m_i)},$$
$$i = 1, \dots, n,$$

where $0 < \chi < 1$ and h satisfies the condition

(20) $$ch^2(\varrho_{i+1}^2 - \varrho_{i-1}^2) \leqq \log(1 + \chi), \quad i = 1, 2, \dots .$$

In a similar way we obtain

(21)

$$|\Delta_i g(x, y^{(m)}) - D_{y_i} g(x, y^{(m)})| (g(x, y^{(m)}))^{-1} \leqq$$
$$\leqq (1 + \chi)c \left|y_i^{(m_i+1)} - 2y_i^{(m_i)} + y_i^{(m_i-1)}\right| + 2c\chi |y_i^{(m_i)}|,$$
$$i = 1, \dots, n, \quad x \in (0, a], \quad m \in Z.$$

Now, conditions (2) and (21) imply (17) for sufficiently small h. We are going to show (18). It follows from (5)–(7) that

(22)

$$\left[\Delta_{ii}^{(2)} g(x, y^{(m)}) - D_{y_i y_i} g(x, y^{(m)})\right] (g(x, y^{(m)}))^{-1} =$$
$$= 2\left(y_i^{(m_i+1)} - y_i^{(m_i-1)}\right)^{-1} \left\{(y_i^{(m_i+1)} - y_i^{(m_i)})^{-1} \times \right.$$
$$\times \exp\left[\tilde{\eta}(x)((y_i^{(m_i+1)})^2 - (y_i^{(m_i)})^2)\right] -$$
$$-(y_i^{(m_i+1)} - y_i^{(m_i-1)})(y_i^{(m_i+1)} - y_i^{(m_i)})^{-1} (y^{(m_i)} - y_i^{(m_i-1)})^{-1} +$$
$$+(y_i^{(m_i)} - y_i^{(m_i-1)})^{-1} \exp\left[\tilde{\eta}(x)\left((y_i^{(m_i-1)})^2 - (y_i^{(m_i)})^2\right)\right] -$$
$$-2\tilde{\eta}(x) - 4(\tilde{\eta}(x))^2 (y_i^{(m_i)})^2.$$

Let $0 < \chi < 1$ and assume that h satisfies (20). Then we have

$$\left|\Delta_{ii}^{(2)} g(x, y^{(m)}) - D_{y_i y_i} g(x, y^{(m)})\right| (g(x, y^{(m)}))^{-1} \leqq$$
$$\leqq c^2 4 \Big[|y_i^{(m_i)}| |y_i^{(m_i+1)} - 2y_i^{(m_i)} + y_i^{(m_i-1)}| +$$
$$+ (y_i^{(m_i+1)} - y_i^{(m_i)})^2 + (y_i^{(m_i)} - y_i^{(m_i-1)})^2 +$$
$$+ \left|y_i^{(m_i+1)} - y_i^{(m_i)}\right| \left|y_i^{(m_i)} - y_i^{(m_i-1)}\right| +$$
$$+ 4\chi |y_i^{(m_i)}| |y_i^{(m_i+1)} - y_i^{(m_i)}| + \chi (y_i^{(m_i+1)} - y_i^{(m_i)})^2 \Big] +$$
$$+ 4\chi c^2 |y_i^{(m_i)}|^2, \quad i = 1, \dots, n, \; m \in Z, \; x \in (0, a].$$

Now, Assumption H_0 and the above estimations imply (18) for sufficiently small h.

It follows from (17), (18) that for any $\varepsilon > 0$ there exists $\delta_2 > 0$ such that for each h with $0 < h < \delta_2$ we have

$$[\Delta_i w(x, y^{(m)}) - D_{y_i} w(x, y^{(m)})] (w(x, y^{(m)}))^{-1} < \varepsilon (1 + y_i^{(m_i)}), \quad i = 1, \dots, n, \; x \in (0, a], \; m_i > 0, \tag{23}$$

and

$$\left[\Delta_{ii}^{(2)} w(x, y^{(m)}) - D_{y_i y_i} w(x, y^{(m)})\right] (w(x, y^{(m)}))^{-1} < \varepsilon (1 + y_i^{(m_i)})^2, \quad i = 1, \dots, n, \; x \in (0, a], \; m_i > 0. \tag{24}$$

Now we prove that there exists $\bar{\delta}$ such that for $0 < h < \bar{\delta}$ we have

$$D_x w^{(m)}(x) \geqq L\Big[w^{(m)}(x) + \sum_{i=1}^n |\Delta_i w^{(m)}(x)| + \sum_{i=1}^n \Delta_{ii}^{(2)} w^{(m)}(x)\Big], \quad m \in Z, \; x \in (0, a]. \tag{25}$$

We define $d = \min(2(2D - nC), 4C(D - C))$, $\varepsilon_1 = \frac{d}{2}$, $\varepsilon = \min(\varepsilon_0, \varepsilon_1)$ where $\varepsilon_0 > 0$ satisfies the inequality $\varepsilon_0^2 n + \varepsilon_0 (2n + 1) 4d - 4d^2 < 0$. Then there exists $\delta^* > 0$ such that estimations (23), (24) hold for $0 < h < \delta^*$ with the above defined ε. From (13), (23), (24) we have for $m_i > 0$, $i = 1, \dots, n$

$$\Big[\sum_{i=1}^n D_{y_i y_i} w^{(m)}(x) + \sum_{i=1}^n D_{y_i} w^{(m)}(x) + w^{(m)}(x)\Big] (w^{(m)}(x))^{-1} -$$
$$- (L w^{(m)}(x))^{-1} D_x w^{(m)}(x) + \sum_{i=1}^n [\Delta_{ii}^{(2)} w^{(m)}(x) - D_{y_i y_i} w^{(m)}(x)] (w^{(m)}(x))^{-1} +$$
$$+ \sum_{i=1}^n [\Delta_i w^{(m)}(x) - D_{y_i} w^{(m)}(x)] (w^{(m)}(x))^{-1} \leqq$$
$$\leqq \|y^{(m)}\|^2 (\varepsilon - d) + \|y^{(m)}\| 3\varepsilon \sqrt{n} + 2\varepsilon n - d < 0,$$

which completes the proof of (25) for $m_i > 0$, $i = 1, \dots, n$. In a similar way we prove (25) for $m \in Z$.

Now we prove (ii). It follows from (7) and from Assumption H_1 that there exists $\alpha > 0$ such that

$$w^{(m)}(0) \geqq A \exp\left[B\|y^{(m)}\|^2\right], \quad m \in Z. \tag{26}$$

Using (11), (25), (26) and Assumption H_2 we obtain

$$\begin{gathered}
D_x v^{(m)}(x) = \alpha(1+x) D_x w^{(m)}(x) + \alpha w^{(m)}(x) \geqq \\
\geqq L_0 v^{(m)}(x) + L_0 \alpha(1+x) \exp\left[-\frac{n}{8C} + \gamma L x\right] + \\
+ L \sum_{i=1}^{n} |\Delta_i v^{(m)}(x)| + L \sum_{i=1}^{n} |\Delta_{ii}^{(2)} v^{(m)}(x)| + \alpha w^{(m)}(0) \geqq \\
\geqq L_0 \left[v^{(m)}(x) + \|v\|_{x,m}^{(h)} + \sum_{i=1}^{n} |\Delta_i v^{(m)}(x)| + \sum_{i=1}^{n} |\Delta_{ii}^{(2)} v^{(m)}(x)| \right] + \\
+ A \exp\left[B\|y^{(m)}\|^2\right] \geqq \\
\geqq [\phi_h(x, y^{(m)}, v^{(m)}(x), v, \Delta v^{(m)}(x), \Delta^{(2)} v^{(m)}(x)) - \phi_h(x, y^{(m)}, 0, 0, 0, 0)] + \\
+ A \exp\left[B\|y^{(m)}\|^2\right] \geqq \phi_h(x, y^{(m)}, \\
v^{(m)}(x), v, \Delta v^{(m)}(x), \Delta^{(2)} v^{(m)}(x)), \quad x \in (0, a], \quad m \in Z,
\end{gathered}$$

which completes the proof of (ii).

We define

$$\overline{w}(x, y) = (1 - 4DLx)^{-1} \exp\left[\frac{(C+D)\|y\|^2}{2(1-4DLx)}\right] \exp\left[-\frac{1}{2} P(y) + \gamma L x\right],$$
$$(x, y) \in [0, a] \times R^n,$$

and

$$U(x, y^{(m)}) = \frac{u(x, y^{(m)})}{\overline{w}(x, y^{(m)})}, \quad \tilde{U}(x, y^{(m)}) = \frac{\tilde{u}(x, y^{(m)})}{\overline{w}(x, y^{(m)})}, \quad (x, y^{(m)}) \in E_h.$$

Let δ_0 be chosen such that for $0 < h < \delta_0$ we have

$$\begin{gathered}
D_x \overline{w}^{(m)}(x) \geqq L \left[\overline{w}^{(m)}(x) + \sum_{i=1}^{n} |\Delta_i \overline{w}^{(m)}(x)| + \sum_{i=1}^{n} \Delta_{ii}^{(2)} \overline{w}^{(m)}(x) \right], \\
m \in Z, \ x \in (0, a],
\end{gathered} \tag{27}$$

and

$$\frac{2\eta}{h(\varrho_{i+1}-\varrho_i)} - L - 2L\frac{\overline{w}^{(-i(m))}(x)}{\overline{w}^{(i(m))}(x)} > 0. \tag{28}$$

We define $V^{(m)}(x) = U^{(m)}(x) - \tilde{U}^{(m)}(x)$, $(x, y^{(m)}) \in E_h$, and $M_p = \max\{V^{(m)}(x)\colon\ x \in [0,a],\ m = (m_1, \dots, m_n),\ |m_i| \leqq p \text{ for } i = 1, \dots, n\}$ where p is a natural number. We prove that $M_p \leqq 0$ for $p = 1, 2, \dots$. Let us assume that there exists p_0 and $h \in (0, \delta_0)$ such that $M_{p_0} > 0$. Then we have $M_p \geqq M_{p_0} > 0$ for $p \geqq p_0$. We prove that for each $p \geqq p_0$ there exists $x_p \in (0, a]$ and j, $1 \leqq j \leqq n$, such that $M_p = V^{(m)}(x_p)$ and $m = (m_1, \dots, m_{j-1}, p, m_{j+1}, \dots, m_n)$ or $m = (m_1, \dots, m_{j-1}, -p, m_{j+1}, \dots, m_n)$. Suppose that there exist $p \geqq p_0$ and $\overline{m} = (\overline{m}_1, \dots, \overline{m}_n)$, $|\overline{m}_i| < p$ for $i = 1, \dots, n$, such that $0 < M_p = V^{(\overline{m})}(x_p)$. Then we have

$$D_x V^{(\overline{m})}(x_p) \geqq 0 \tag{29}$$

and

$$V^{(\overline{m})}(x_p) - V^{(j(\overline{m}))}(x_p) \geqq 0,\ V^{(\overline{m})}(x_p) - V^{(-j(\overline{m}))}(x_p) \geqq 0,\ j = 1, \dots, n,$$

where for each j, $1 \leqq j \leqq n$, equality holds in at most one place. Let $u^*(t, y^{(m)}) = \max\{u(t, y^{(m)}), \tilde{u}(t, y^{(m)})\}$ for $(t, y^{(m)}) \in H_h[x_p, \overline{m}]$ and $U^*(t, y^{(m)}) = u^*(t, y^{(m)})(\overline{w}(t, y^{(m)}))^{-1}$ where $t \in [0, x_p]$, $m = (m_1, \dots, m_n)$, $|m_i| \leqq p$ for $i = 1, \dots, n$. Since $\Delta^{(2)}\overline{w}^{(m)}(x) \geqq 0$ and

$$\begin{aligned}
\Delta^{(2)}_{jj} u^{(\overline{m})}(x_p) - \Delta^{(2)}_{jj}\tilde{u}^{(\overline{m})}(x_p) - V^{(\overline{m})}(x_p)\overline{w}^{(\overline{m})}(x_p) = \\
= A_j^{(\overline{m})}\overline{w}^{(j(\overline{m}))}(x_p)\left[V^{(j(\overline{m}))}(x_p) - V^{(\overline{m})}(x_p)\right] + \\
+ C_j^{(\overline{m})}\overline{w}^{(-j(\overline{m}))}(x_p)\left[V^{(-j(\overline{m}))}(x_p) - V^{(\overline{m})}(x_p)\right] < 0, \\
j = 1, \dots, n,
\end{aligned}$$

then

$$\begin{aligned}
D_x V^{(\overline{m})}(x_p)\overline{w}^{(\overline{m})}(x_p) + D_x\overline{w}^{(\overline{m})}(x_p)V^{(\overline{m})}(x_p) = \\
= D_x u^{(\overline{m})}(x_p) - D_x\tilde{u}^{(\overline{m})}(x_p) \leqq \\
\leqq [\phi_h(x_p, y^{(\overline{m})}, u^{(\overline{m})}(x_p), u^*, \Delta u^{(\overline{m})}(x_p), \Delta^{(2)}u^{(\overline{m})}(x_p)) - \\
- \phi_h(x_p, y^{(\overline{m})}, u^{(\overline{m})}(x_p), u^*, \Delta u^{(\overline{m})}(x_p), \Delta^{(2)}\tilde{u}^{(\overline{m})}(x_p) + V^{(\overline{m})}(x_p)\Delta^{(2)}\overline{w}^{(\overline{m})}(x_p))] + \\
+ [\phi_h(x_p, y^{(\overline{m})}, u^{(\overline{m})}(x_p), u^*, \Delta u^{(\overline{m})}(x_p), \Delta^{(2)}\tilde{u}^{(\overline{m})}(x_p) + \\
+ V^{(\overline{m})}(x_p)\Delta^{(2)}\overline{w}^{(\overline{m})}(x_p)) -
\end{aligned}$$

$$-\phi_h(x_p, y^{(\overline{m})}, \tilde{u}^{(\overline{m})}(x_p), \tilde{u}, \Delta\tilde{u}^{(\overline{m})}(x_p)\Delta^{(2)}\tilde{u}^{(\overline{m})}(x_p))] \leqq$$

$$\leqq -\eta \sum_{i=1}^{n} [A_i^{(\overline{m})} \overline{w}^{(i(\overline{m}))}(x_p)(V^{(\overline{m})}(x_p) - V^{(i(\overline{m}))}(x_p)) +$$

$$+C_i^{(\overline{m})} \overline{w}^{(-i(\overline{m}))}(x_p)(V^{(\overline{m})}(x_p) - V^{(-i(\overline{m}))}(x_p))] +$$

$$+L_o\Big[2V^{(\overline{m})}(x_p)\overline{w}^{(\overline{m})}(x_p) + \sum_{i=1}^{n} |V^{(i(\overline{m}))}(x_p)\Delta_i \overline{w}^{(\overline{m})}(x_p) +$$

$$+\overline{w}^{(-i(\overline{m}))}(x_p)\Delta_i V^{(\overline{m})}(x_p)| + \sum_{i=1}^{n} V^{(\overline{m})}(x_p)\Delta_{ii}^{(2)} \overline{w}^{(\overline{m})}(x_p)\Big].$$

The above estimation and (27) imply

$$D_x V^{(\overline{m})}(x_p)\overline{w}^{(\overline{m})}(x_p) \leqq -\eta \sum_{i=1}^{n} [A_i^{(\overline{m})} \overline{w}^{(i(\overline{m}))}(x_p)(V^{(\overline{m})}(x_p) - V^{(i(\overline{m}))}(x_p)) +$$

$$+C_i^{(\overline{m})} \overline{w}^{(-i(\overline{m}))}(x_p)(V^{(\overline{m})}(x_p) - V^{(-i(\overline{m}))}(x_p))] +$$

$$+L \sum_{i=1}^{n} (V^{(\overline{m})}(x_p) - V^{(i(\overline{m}))}(x_p))|\Delta_i \overline{w}^{(\overline{m})}(x_p)| +$$

$$+L \sum_{i=1}^{n} \overline{w}^{(-i(\overline{m}))}(x_p)[(V^{(\overline{m})}(x_p) - V^{(i(\overline{m}))}(x_p)) +$$

$$+(V^{(\overline{m})}(x_p) - V^{(-i(\overline{m}))}(x_p))](y_i^{(\overline{m}_i+1)} - y_i^{(\overline{m}_i-1)})^{-1}.$$

Then we have from (28)

$$D_x V^{(\overline{m})}(x_p) \leqq \sum_{i=1}^{n} \frac{\overline{w}^{(i(\overline{m}))}(x_p)}{\overline{w}^{(\overline{m})}(x_p)} (V^{(\overline{m})}(x_p) - V^{(i(\overline{m}))}(x_p)) \times$$

$$\times \left[L + \frac{2L\overline{w}^{(-i(\overline{m}))}(x_p)}{\overline{w}^{(i(\overline{m}))}(x_p)} - \frac{2\eta}{y_i^{(m_i+1)} - y_i^{(m_i)}} \right] (y_i^{(\overline{m}_i+1)} - y_i^{(\overline{m}_i-1)})^{-1} +$$

$$+ \sum_{i=1}^{n} \frac{\overline{w}^{(-i(\overline{m}))}(x_p)}{\overline{w}^{(\overline{m})}(x_p)} (V^{(\overline{m})}(x_p) - V^{(-i(\overline{m}))}(x_p)) \times$$

$$\times \left[L - \frac{2\eta}{y_i^{(\overline{m}_i)} - y_i^{(\overline{m}_i-1)}} \right] (y_i^{(\overline{m}_i+1)} - y_i^{(\overline{m}_i-1)})^{-1} < 0$$

which contradicts (29). Thus we see that for each $p \geqq p_0$ there exist $x_p \in (0,a]$ and j, $1 \leqq j \leqq n$, such that $M_p = V^{(m)}(x_p)$ and $m = (m_1, \dots, m_n)$, $|m_j| = p$. Since $u, \tilde{u} \in \Gamma_h$ then there is $\alpha_0 > 0$ such that

$$M_p = \frac{u^{(m)}(x_p) - \tilde{u}^{(m)}(x_p)}{\overline{w}^{(m)}(x_p)} \leqq \alpha_0 \frac{g(x_p, y^{(m)})}{\overline{w}(x_p, y^{(m)})}.$$

Because $\lim\limits_{p\to\infty} g(x_p, y^{(m)})(\overline{w}(x_p, y^{(m)}))^{-1} = 0$, we have $M_{p_0} = 0$ in contradiction to the assumption $M_{p_0} > 0$.

REMARK 2. If Assumptions H_0–H_2 are satisfied then there exist $\delta > 0$ ($\delta = \delta(C, D, E, \eta, L_0, \varrho)$) and $\alpha > 0$ such that for $0 < h < \delta$ the function

$$\bar{v}(x, y^{(m)}) = -\alpha w(0, y^{(m)}) \quad \text{for} \quad (x, y^{(m)}) \in E_h^{(0)},$$
$$\bar{v}(x, y^{(m)}) = -\alpha(1+x) w(x, y^{(m)}) \quad \text{for} \quad (x, y^{(m)}) \in E_h$$

satisfies the differential-difference inequalities

$$D_x \bar{v}^{(m)}(x) \leqq \phi_h\left(x, y^{(m)}, \bar{v}^{(m)}(x), \bar{v}, \Delta\bar{v}^{(m)}(x), \Delta^{(2)}\bar{v}^{(m)}(x)\right), \quad x \in (0,a], \quad m \in Z. \tag{30}$$

We omit the simple proof of (30).

3. Stability of the line method

THEOREM 2. *If Assumptions $H_0 - H_2$ are satisfied then there exists $h_0 > 0$ such that for $h \in I_0 = (0, h_0)$ we have*

(i) *the initial problem (9) has exactly one solution $u_h \in \Gamma_h$,*

(ii) *if $\overline{z} \in \Gamma_h$, $\gamma_h \geqq 0$ for $h \in I_0$ and*

$$\begin{aligned} &|C_h^{(m)}(x) - \overline{z}^{(m)}(x)| \leqq \gamma_h w^{(m)}(0) \quad \textit{for} \quad x \in [-\tau_0, 0],\ m \in Z, \\ &|D_x \overline{z}^{(m)}(x) - \phi_h(x, y^{(m)}, \overline{z}^{(m)}(x), \overline{z}, \Delta\overline{z}^{(m)}(x), \Delta^{(2)}\overline{z}^{(m)}(x))| \leqq \\ &\leqq \gamma_h w^{(m)}(x), \quad x \in (0,a], \quad m \in Z, \end{aligned} \tag{31}$$

then

$$|u_h^{(m)}(x) - \overline{z}^{(m)}(x)| \leqq \gamma_h (1+x) w^{(m)}(x), \qquad x \in [0,a],\ m \in Z. \tag{32}$$

PROOF. We define the sequence $\{v_i\}_{i=0}^{\infty}$, $v_i : E_h^{(0)} \cup E_h \to R$, as follows:

(a) $v_0(x, y^{(m)}) = v(x, y^{(m)})$ for $(x, y^{(m)}) \in E_h^{(0)} \cup E_h$ where v is given by (11) with α satisfying the assertion (ii) of Theorem 1,

(b) if v_i, $i \geqq 0$, is a known function then $v_{i+1}|_{E_h^{(0)}} = \mathcal{C}_h$ and v_{i+1} is a solution of the problem

$$(33)\qquad \begin{aligned} D_x z^{(m)}(x) = \phi_h\left(x, y^{(m)}, z^{(m)}(x), v_i, \Delta v_i^{(m)}(x), V[z, v_i]^{(m)}(x)\right) \\ x \in [0,a], \quad m \in Z, \\ z^{(m)}(0) = \mathcal{C}_h^{(m)}(0), \quad m \in Z, \end{aligned}$$

where $V[z, v_i] = (V_1[z, v_i], \dots , V_n[z, v_i])$ and

$$V_j[z, v_i]^{(m)}(x) = A_j v_i^{(j(m))}(x) + B_j z^{(m)}(x) + C_j v_i^{(-j(m))}(x), \qquad j = 1, \dots , n.$$

Since the right-hand side of (33) is Lipschitz continuous with respect to $z^{(m)}$, there exists a unique solution $v_{i+1}^{(m)}$ of (33) on $[0,a]$ for every $m \in Z$. Then we have $v_i \in \mathcal{F}_c(E_h^{(0)} \cup E_h, R)$ for $i = 0, 1, 2, \dots$. Using a simple theorem on differential inequalities ([2], [5]–[7], [14]) we get from the assertion (iii) of Theorem 1 (see also Remark 2) and from (33) that

$$(34)\qquad \tilde{v}^{(m)}(x) \leqq v_{i+1}^{(m)}(x) \leqq v_i^{(m)}(x) \leqq v^{(m)}(x), \quad x \in [0,a], \quad m \in Z.$$

Consider the sequence $\{v_i^{(m)}\}_{i=0}^{\infty}$ for a fixed $m \in Z$. Since the sequences $\{v_i^{(m)}\}_{i=0}^{\infty}$ are equicontinuous on $[0,a]$ for every $m \in Z$ it follows from (34) and from the Arzela–Ascoli theorem that there exists $u_h \in \Gamma_h$ such that

$$u_h|_{E_h^{(0)}} = \mathcal{C}_h \quad \text{and} \quad u_h^{(m)}(x) = \lim_{i \to \infty} v_i^{(m)}(x)$$

uniformly with respect to $x \in [0,a]$ where $m \in Z$. Considering the integral equations corresponding to (33) we get that u_h is a solution of (9). The uniqueness of a solution of (9) follows from the assertion (iii) of Theorem 1, and the proof of (i) is complete.

Now we prove that

$$(35)\qquad u_h^{(m)}(x) \leqq \bar{z}^{(m)}(x) + \gamma_h(1+x)w^{(m)}(x), \qquad x \in [0,a],\ m \in Z.$$

It follows from Assumption H_2 and from (31) that the function $\tilde{z}(x, y^{(m)}) = \bar{z}^{(m)}(x) + \gamma_h(1+x)w^{(m)}(x)$ for $(x, y^{(m)}) \in E_h$ and $\tilde{z}(x, y^{(m)}) = \bar{z}^{(m)}(x) + \gamma_h w^{(m)}(0)$ for $(x, y^{(m)}) \in E_h^{(0)}$ satisfies the conditions

$$D_x \tilde{z}^{(m)}(x) \geqq \phi_h\left(x, y^{(m)}, \tilde{z}^{(m)}(x), \tilde{z}, \Delta \tilde{z}^{(m)}(x), \Delta^{(2)} \tilde{z}^{(m)}(x)\right) +$$

$$+[\phi_h\left(x, y^{(m)}, \bar{z}^{(m)}(x), \bar{z}, \Delta \bar{z}^{(m)}(x), \Delta^{(2)} \bar{z}^{(m)}(x)\right) -$$

$$-\phi_h\left(x, y^{(m)}, \tilde{z}^{(m)}(x), \tilde{z}, \Delta\tilde{z}^{(m)}(x), \Delta^{(2)}\tilde{z}^{(m)}(x)\right)]+$$

$$+\gamma_h(1+x)D_x w^{(m)}(x) \geqq \phi_h\left(x, y^{(m)}, \tilde{z}^{(m)}(x), \tilde{z}, \Delta\tilde{z}^{(m)}(x), \Delta^{(2)}\tilde{z}^{(m)}(x)\right) -$$

$$-L_0\gamma_h(1+x)\left[2w^{(m)}(x) + \sum_{i=1}^{n}|\Delta_i w^{(m)}(x)| + \sum_{i=1}^{n}\Delta_{ii}^{(2)}w^{(m)}(x)\right]+$$

$$+\gamma_h(1+x)D_x w^{(m)}(x) \geqq$$

$$\geqq \phi_h\left(x, y^{(m)}, \tilde{z}^{(m)}(x), \tilde{z}, \Delta\tilde{z}^{(m)}(x), \Delta^{(2)}\tilde{z}^{(m)}(x)\right), \ x \in (0,a], \ m \in Z.$$

Since $u_h^{(m)}(x) \leqq \tilde{z}^{(m)}(x)$ for $x \in [-\tau_0, 0]$, $m \in Z$, and u_h is a solution of (9) then we have (35). In a similar way we prove

$$\overline{z}^{(m)}(x) - \gamma_h(1+x)w^{(m)}(x) \leqq u_h^{(m)}(x), \qquad x \in [0,a], \ m \in Z. \tag{36}$$

Estimations (35), (36) imply (32), which is our claim.

4. Convergence of the line method

In this part of the paper we prove that if the line method (9) is stable and satisfies a consistency condition then it is convergent. Let Γ denote the set of all $z \in C(E^{(0)} \cup E, R)$ such that $|z(x,y)| \leqq A_0 \exp[C\|y\|^2]$, $x \in [-\tau_0, 0]$, $y \in R^n$, for a certain $A_0 \in R_+$ and $\sup\{|z(x,y)|(g(x,y))^{-1} : (x,y) \in E\} < +\infty$.

THEOREM 3. *Suppose that*
1° *Assumptions $H_0 - H_2$ are satisfied,*
2° *$f \in C(\Omega, R)$ and there exists an $M > 0$ such that*

$$|f(x,y,p,z,q,r) - f(x,y,p,z,\overline{q},\overline{r})| \leqq M\left[\sum_{i=1}^{n}|q_i - \overline{q}_i| + \sum_{i=1}^{n}|r_i - \overline{r}_i|\right]$$

on Ω,
3° *$\overline{z}$ is a solution of* (1), *$\overline{z} \in \Gamma$ and there exists $\gamma_h^{(1)}$, $\gamma_h^{(2)}$, $h \in I_0$, such that*

$$\begin{aligned}&\left|D_{y_i}\overline{z}(x,y^{(m)}) - \Delta_i\overline{z}(x,y^{(m)})\right| \leqq \gamma_h^{(1)}(1-4DLx)^{-1}\exp\left[\frac{B\|y^{(m)}\|^2}{1-4DLx}\right]\\ &\left|D_{y_iy_i}\overline{z}(x,y^{(m)}) - \Delta_{ii}^{(2)}\overline{z}(x,y^{(m)})\right| \leqq \gamma_h^{(2)}(1-4DLx)^{-1}\exp\left[\frac{B\|y^{(m)}\|^2}{1-4DLx}\right],\end{aligned} \tag{37}$$

$$x \in [0,a], \quad m \in Z, \quad i = 1, \ldots, n,$$

and $\lim_{h\to 0} \gamma_h^{(i)} = 0$, $i = 1, 2$,

4° *the following consistency condition holds: for each* $h \in I_0$ *there exists* $\gamma_h^{(0)} \geqq 0$ *such that*

$$|\phi_h\left(x, y^{(m)}, \overline{z}_h^{(m)}(x), \overline{z}_h, \Delta\overline{z}_h^{(m)}(x), \Delta^{(2)}\overline{z}_h^{(m)}(x)\right) -$$

$$-f\left(x, y^{(m)}, \overline{z}^{(m)}(x), \overline{z}, \Delta\overline{z}^{(m)}(x), \Delta^{(2)}\overline{z}^{(m)}(x)\right)| \leqq \gamma_h^{(0)} w(x, y^{(m)})$$

$$x \in [0, a],\ m \in Z,$$

where $\overline{z}_h = \overline{z}\,|_{E_h^{(0)} \cup E_h}$ *and* $\lim_{h\to 0} \gamma_h^{(0)} = 0$,

5° *for each* $h \in I_0$ *there is* $\bar{\gamma}_h \geqq 0$ *such that* $\lim_{h\to 0} \bar{\gamma}_h = 0$ *and*

$$|\omega(x, y^{(m)}) - \mathcal{C}_h(x, y^{(m)})| \leqq \bar{\gamma}_h w(0, y^{(m)}), \qquad m \in Z,\ x \in [-\tau_0, 0],$$

6° u_h *is the solution of* (9) *on* $E_h^{(0)} \cup E_h$.

Under these assumptions there exists $\gamma_h \geqq 0$, $h \in I_0$, *such that* $\lim_{h\to 0} \gamma_h = 0$ *and*

$$|\overline{z}^{(m)}(x) - u_h^{(m)}(x)| \leqq \gamma_h(1 + x)w^{(m)}(x), \qquad x \in [0, a],\ m \in Z. \tag{38}$$

PROOF. We apply Theorem 2 for proving (38). Since

$$|D_x\overline{z}^{(m)}(x) - \phi_h\left(x, y^{(m)}, \overline{z}_h^{(m)}(x), \overline{z}_h, \Delta\overline{z}_h^{(m)}(x), \Delta^{(2)}\overline{z}_h^{(m)}(x)\right)| \leqq$$

$$\leqq |f\left(x, y^{(m)}, \overline{z}^{(m)}(x), \overline{z}, \Delta\overline{z}^{(m)}(x), \Delta^{(2)}\overline{z}^{(m)}(x)\right) -$$

$$-f\left(x, y^{(m)}, \overline{z}^{(m)}(x), \overline{z}, D_y\overline{z}^{(m)}(x), D_{yy}\overline{z}^{(m)}(x)\right)| +$$

$$+|f\left(x, y^{(m)}, \overline{z}^{(m)}(x), \overline{z}, \Delta\overline{z}^{(m)}(x), \Delta^{(2)}\overline{z}^{(m)}(x)\right) -$$

$$-\phi_h\left(x, y^{(m)}, \overline{z}_h^{(m)}(x), \overline{z}_h, \Delta\overline{z}^{(m)}(x), \Delta^{(2)}\overline{z}_h^{(m)}(x)\right)| \leqq$$

$$M\left[\sum_{i=1}^{n} |\Delta_i\overline{z}^{(m)}(x) - D_{y_i}\overline{z}^{(m)}(x)| + \sum_{i=1}^{n} |\Delta_{ii}^{(2)}\overline{z}^{(m)}(x) - D_{y_iy_i}\overline{z}^{(m)}(x)|\right] +$$

$$+w(x, y^{(m)})\gamma_h^{(0)}, \qquad x \in [0, a],\ m \in Z,$$

it follows from the conditions 3° and 5° that there exists $\gamma_h \geqq 0$, such that $\lim_{h\to 0} \gamma_h = 0$ and the estimations (31) hold. Now we obtain our assertion from Theorem 2.

REMARK 3. Let us assume that $\bar{z}\colon E^{(0)}\cup E\to R$ is a solution of (1) such that $D_{y_iy_iy_i}\bar{z}$, $i=1,\dots,n$, are continuous on E and there is $M_0\geqq 0$ such that

$$|D_{y_iy_iy_i}z(x,y)|\leqq M_0\exp\left[\frac{B\|y\|^2}{1-4DLx}\right],\qquad (x,y)\in E.$$

Then there are $\gamma_h^{(1)}$, $\gamma_h^{(2)}$, $h\in I_0$, such that $\lim_{h\to 0}\gamma_h^{(i)}=0$, $i=1,2$, and (37) holds.

REMARK 4. Let $S=\{s=(s_1,\dots,s_n)\colon s_i\in\{0,1\}$ for $i=1,\dots,n\}$. Suppose that $z\in\mathcal{F}_c(E_h^{(0)}\cup E_h,R)$. We define $W_hz\colon E^{(0)}\cup E\to R$ in the following way. Suppose that $(x,y)\in E^{(0)}\cup E$. Then there exists $m\in Z$ such that $y^{(m)}\leqq y<y^{(m+1)}$ where $m+1=(m_1+1,\dots,m_n+1)$. We define

$$(W_hz)(x,y)=\sum_{s\in S}z^{(m+s)}(x)\prod_{i=1}^{n}\left[\left(y_i-y_i^{(m_i)}\right)\left(y_i^{(m_i+1)}-y_i^{(m_i)}\right)^{-1}\right]^{s_i}\times$$
$$\times\left[1-\left(y_i-y_i^{(m_i)}\right)\left(y_i^{(m_i+1)}-y_i^{(m_i)}\right)^{-1}\right]^{1-s_i}.$$

Then we have ([1]) $W_hz\in C(E^{(0)}\cup E,R)$. Suppose that $f\colon\Omega\to R$. It is easy to formulate sufficient conditions for the convergence of the line method (9) if $\phi_h(x,y,p,z,q,r)=f(x,y,p,W_hz,q,r)$ on Ω_h.

REMARK 5. The results obtained in this paper can be extended to weakly coupled systems of parabolic differential-functional systems.

REFERENCES

[1] BRANDI, P., KAMONT, Z., SALVADORI, A., Differential and differential-difference inequalities related to mixed problems for first order partial differential-functional equations, *Atti Sem. Mat. Fis. Univ. Modena* **39** (1991), 255–276.

[2] CORDUNEANU, C., *Principles of Differential and Integral Equations*, Second edition, Chelsea, New York, 1977. *MR* **55** #12977

[3] KAMONT, Z. and ZACHAREK, S., The line method for parabolic differential-functional equations with initial boundary conditions of the Dirichlet type, *Atti Sem. Mat. Fis. Univ. Modena* **35** (1987), 249–262. *MR* **89g**: 65120

[4] ŁOJASIEWICZ, S., Sur le problème de Cauchy pour les systèmes d'équations aux dérivées partielles du premier ordre dans le cas hyperbolique de deux variables indépendantes, *Ann. Polon. Math.* **3** (1956), 87–117. *MR* **18**-485

[5] TURINICI, M., Differential inequalities via maximal element techniques, *Nonlinear Anal.* **5** (1981), 757–763. *MR* **82g**: 34014

[6] TURINICI, M., Time-dependent invariant polygonal domain for a class of functional-differential equations, *Rev. Roumaine Math. Pures Appl.* **26** (1981), 339–345. *MR* **82i**: 34079

[7] TURINICI, M., Function-differential inequalities via maximality procedures, *Nonlinear Anal.* **6** (1982), 1369–1384. *MR* **84e**: 34016

[8] VOIGT, A., Line method approximations to the Cauchy problem for nonlinear parabolic differential equations, *Numer. Math.* **23** (1974), 23–36. *MR* **50** #15370

[9] Voigt, A., The method of lines for nonlinear parabolic differential equations with mixed derivatives, *Numer. Math.* **32** (1979), 197–207. *MR* **80e**: 65087

[10] Walter, W., Die Linienmethode bei nichtlinearen parabolischen Differentialgleichungen, *Numer. Math.* **12** (1968), 307–321. *MR* **39** #3725

[11] Walter, W., Ein Existenzbeweis für nichtlineare parabolische Differentialgleichungen aufgrund der Linienmethode, *Math. Z.* **107** (1968), 173–188. *MR* **39** #3162

[12] Walter, W., Approximation für das Cauchy-Problem bei parabolischen Differentialgleichungen mit der Linienmethode, *Abstract Spaces and Approximation* (Proc. Conf., Oberwolfach, 1968), ISNM, vol. 10, Birkhäuser, Basel, 1969, 375–386. *MR* **41** #2954

[13] Walter, W., Existenzsätze in Großen für das Cauchy-Problem bei nichtlinearen parabolischen Differentialgleichungen mit der Linienmethode, *Math. Ann.* **183** (1969), 254–274. *MR* **40** #7625

[14] Walter, W., *Differential and Integral Inequalities*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 55, Springer-Verlag, Berlin–Heidelberg–New York, 1970. *MR* **42** #6391

[15] Walter, W. and Acker, A., On the global existence of solutions of parabolic differential equations with a singular nonlinear term, *Nonlinear Anal.* **2** (1978), 499–504. *MR* **80c**: 35054

[16] Walter, W., Schmitt, K. and Thompson, R. C., Existence of solutions of a nonlinear boundary value problem via the method of lines, *Nonlinear Anal.* **2** (1978), 519–535. *MR* **80i**: 65132

(Received December 10, 1988)

UNIWERSYTET GDAŃSKI
INSTYTUT MATEMATYKI
UL. WITA STWOSZA 57
PL–80-952 GDAŃSK
POLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 331–338

JOINS OF SOME VARIETIES OF SEMIGROUPS[1]

SHELLY L. WISMATH

1. Introduction and summary

The problem of describing the join of two or more varieties of semigroups is not in general an easy one. Viewing varieties as equationally defined classes, we would like when given equational descriptions of two varieties to produce a set of identities which will define their join. In this paper we investigate this problem for joins of various varieties V with the varieties of rectangular bands, and of k-nilpotent semigroups. The choice of which varieties to consider, and of defining sets of identities, has been greatly motivated by the author's work on hyperidentities for varieties of semigroups ([5], [6]). The results obtained, while not in themselves particularly deep, have proved useful in obtaining hyperidentity results for varieties of commutative semigroups.

Section 2 establishes the notation and terminology to be used, and describes the varieties of semigroups to be considered. It also outlines the two approaches, structural and syntactic, to be used in showing that a conjectured set of identities does indeed define the join of two specified varieties. The next two sections deal with joins of the form $V \vee W$, where W is either RB, the variety of rectangular bands, N_k, the variety of k-nilpotent semigroups, or a closely related variety.

2. Preliminaries

This section describes the notation and techniques to be used throughout. Our interest in varieties of semigroups will be in terms of the identities they satisfy; that is, we will view them as equational classes. We fix a countably infinite set of variables, including $x, y, z, w, x_1, x_2, \ldots, y_1, y_2, \ldots$, and use words from the free semigroup on this set. For any word u, $|u|$ denotes the length of u. An identity is then an equation $u = v$ where u and v are words. For any set I of identities, we use $V(I)$ to denote the class of all

1980 *Mathematics Subject Classification* (1985 Revision). Primary 20MXX.
Key words and phrases. Joins, varieties of semigroups, equational descriptions.

[1] The results described in this paper form part of the author's Ph.D. thesis, submitted to Simon Fraser University, Burnaby, Canada. The author is grateful for the help of her supervisor, Dr. N. R. Reilly, and for the financial support received from Simon Fraser University.

semigroups satisfying I; if I contains only one identity $u=v$, we simplify this to $V(u=v)$. A set I of identities is a basis for a variety V if $V(I)=V$, and hence all the identities satisfied by V are consequences of the identities in I. For convenience we list below the varieties of semigroups to be referred to in subsequent sections:

$A=V(xy=yx)$, the variety of abelian semigroups.

$M=V(xyzw=xzyw)$, the variety of medial semigroups.

$B=V(x^2=x)$, the variety of bands (idempotent semigroups).

$SL=V(xy=yx,\ x^2=x)=A\cap B$, the variety of semilattices.

$RB=V(xyz=xz)$, the variety of rectangular bands.

$NB=V(xyzw=xzyw,\ x^2=x)=M\cap B$, the variety of normal bands.

$N_k=V(x_1\dots x_k=y_1\dots y_k)$, the variety of k-nilpotent semigroups. Note that $N_2=Z$, the variety of zero semigroups.

$AN_k=V(xy=yx,\ x_1\dots x_k=y_1\dots y_k)$, the variety of abelian k-nilpotent semigroups.

$MN_k=V(xyzw=xzyw,\ x_1\dots x_k=y_1\dots y_k)$, the variety of medial k-nilpotent semigroups.

$A_m=V(xy=yx,\ xy^m=x)$, the variety of abelian groups of exponent m.

$A_{n,m}=V(xy=yx,\ x^n=x^{n+m})$, the variety of commutative semigroups satisfying $x^n=x^{n+m}$. Note that $A_{1,1}=SL$, and $A_{1,m}=SL\vee A_m$.

$B_{n,m}=V(x^n=x^{n+m})$, the variety of semigroups satisfying $x^n=x^{n+m}$.

$M_{n,m}=V(xyzw=xzyw,\ x^n=x^{n+m})$, the variety of medial semigroups satisfying $x^n=x^{n+m}$. Note that $M_{1,1}=NB$.

In the next two sections we produce sets of defining identities for various joins of these varieties. The choice of identities has been suggested in most cases by knowledge of what hyperidentity instances we have been able to obtain for the relevant varieties. To prove that a conjectured set of identities does indeed define the required join, we use two approaches, a structural one and a syntactic one. Let $V=V(I)$, $W=V(J)$, and $U=V(K)$, where I, J, and K are sets of identities, and $V\vee W\subseteqq U$. The syntactic approach is to consider the identities satisfied by $V\vee W$. If we can show that any non-trivial identity satisfied by $V\vee W$ is a consequence of the identities in K, then any such identity is also satisfied by U. From this it follows that $U\subseteqq V\vee W$, giving us $U=V\vee W$.

For the structural approach, we show that $U\subseteqq V\vee W$ by showing that every semigroup in U is a subdirect product of a semigroup in V and a semigroup in W. The following Lemma sets up the machinery to be used in proofs of this type.

LEMMA 2.1. *Let V and W be varieties of semigroups, and let C be any semigroup. Suppose that there is a map $\Theta: C\to C$ which satisfies the following conditions:*

1. *Θ is a homomorphism;*
2. *Θ is a retraction; that is, $\Theta^2=\Theta$;*

3. *the image* $\Theta(C)$ *is an ideal of* C*;*
4. $\Theta(C)$ *is in* V*;*
5. *the Rees quotient* $C/\Theta(C)$ *is in* W.

Then C *is a subdirect product of* $\Theta(C)$ *and* $C/\Theta(C)$*, so that* C *is in* $V \vee W$.

PROOF. Let ϱ be the canonical homomorphism from C to its Rees quotient $C/\Theta(C)$. The condition that Θ is a retraction ensures that the intersection of the kernels of ϱ and Θ is trivial. From this it follows (see [1]) that C is a subdirect product of the images $\Theta(C)$ of Θ and $C/\Theta(C)$ of ϱ. Conditions 4 and 5 then imply that C is in $V \vee W$.

3. Joins with RB

The first type of joins we consider are those of the form $V \vee RB$, for certain varieties V of semigroups. We will use for V the varieties A_m and $A_{n,m}$, and N_k, MN_k, and AN_k, for $n, m \geqq 1$ and $k \geqq 2$. A useful observation is that since $A_{1,m} = SL \vee A_m$, and $NB = RB \vee SL$, we have $A_{n,m} \vee RB$ equal to $A_{n,m} \vee NB$ for all n and $m \geqq 1$. Petrich has proved in [3] that for $m \geqq 2$, $A_{1,m} \vee RB = M_{1,m}$. For $A_{n,m} \vee RB$ when $n \geqq 2$ the structural approach used by Petrich does not work, and we turn instead to a syntactic method. Also unlike the $n = 1$ case, $A_{n,m} \vee RB$ is a proper subvariety of $M_{n,m}$ when $n \geqq 2$, as indicated by the identity $x^{n-1}yx = x^{n-1+m}yx$ used in the next result.

PROPOSITION 3.1. *Let* $n \geqq 2$ *and* $m \geqq 1$. *Then the variety* $A_{n,m} \vee RB$ *is defined by the identities*

$$xyzw = xzyw,\ x^n = x^{n+m},\ \text{and}\ x^{n-1}yx = x^{n-1+m}yx.$$

PROOF. Let the variety defined by the three given identities be called W. Clearly $A_{n,m} \vee RB \subseteq W$. For the opposite inclusion, suppose that $u = v$ is any non-trivial identity satisfied by $A_{n,m} \vee RB$. We show that W also satisfies $u = v$.

Since RB satisfies $u = v$, we know that u and v start with the same letter, x say, and end with the same letter, y say (with x and y possibly the same). Since $A_{n,m}$ satisfies $u = v$, u and v have the same content, and for each letter z in this content, either the number of occurrences of z in u is equal to the number of occurrences of z in v, or these two quantities are both $\geqq n$ and are congruent modulo m. Using this information we show how to deduce $u = v$ from the identities defining W.

We first transform u and v into a "standard form" $\overline{u}$ and $\overline{v}$ as follows. Write

$$\overline{u} = x^a z_1^{a_1} \dots z_p^{a_p} y^b \quad \text{and} \quad \overline{v} = x^c z_1^{c_1} \dots z_p^{c_p} y^e,$$

where $x, z_1, \dots, z_p, y$ are the distinct (except possibly $x = y$) letters appearing in u and v, and if $x = y$, then $b = e = 1$. The identities $u = \overline{u}$ and $v = \overline{v}$ hold in W, just by use of the medial identity.

If $x \neq y$, then from the above information we may deduce $\overline{u} = \overline{v}$ simply by using the identity $x^n = x^{n+m}$. Hence in this case, W satisfies $u = v$.

If $x = y$, we have $b = e = 1$ by construction. Again we may deal with the "interior" letters $z_1, \ldots, z_p$ using only $x^n = x^{n+m}$, so we may reduce this case to considering words $\overline{u}' = x^a wx$ and $\overline{v}' = x^c wx$, for some word w. From the comments above, either $a + 1 = c + 1$, or $a + 1$ and $c + 1$ are both $\geqq n$ and are congruent modulo m. If $a = c$, we are done; otherwise, both a and c are $\geqq n - 1$ and a and c are congruent modulo m, and we have two cases to consider.

If a and c are both $\geqq n$, with a and c congruent modulo m, then $x^a = x^c$ holds in W, and so does $\overline{u} = \overline{v}$. Finally, suppose that $a = n - 1$ and $c \geqq n$ (or dually). Then c is congruent to $n - 1$ modulo m, and c may be written as $km + n - 1$ for some $k \geqq 1$. But then $\overline{u}' = x^{n-1} wx$ and $\overline{v}' = x^{km+n-1} wx$, and $\overline{u}' = \overline{v}'$ holds in W by repeated use of the identity $x^{n-1} yx = x^{n-1+m} yx$. Hence in either case W satisfies $\overline{u} = \overline{v}$ and therefore also $u = v$.

For the remainder of this section we focus on the joins of some nilpotent varieties with RB.

PROPOSITION 3.2. *Let $k \geqq 3$. Then*
 i) $N_k \vee RB = V(x_1 \ldots x_k = x_1 y_2 \ldots y_{k-1} x_k)$, *and*
 ii) $MN_k \vee RB = V(xyzw = xzyw, x_1 \ldots x_k = x_1 y_2 \ldots y_{k-1} x_k)$.

PROOF. We will call the right-hand side variety from i) W_k. Clearly it contains $N_k \vee RB$. To prove the opposite inclusion, let C be any semigroup in W_k. Define a map $\Theta\colon C \to C$ by $\Theta(c) = c^k$, for all c in C. Then it is easily verified that Θ satisfies the five conditions of Lemma 2.1, giving the conclusion that C is in $N_k \vee RB$. The proof of ii) is very similar to the previous one, with the map $\Theta\colon C \to C$ as before. But now $\Theta(C)$ is in RB and $C/\Theta(C)$ is in MN_k, so that C is in $MN_k \vee RB$.

PROPOSITION 3.3. *Let $k \geqq 2$. Then $AN_k \vee RB = MN_k \vee RB$.*

PROOF. If $k = 2$, $AN_k = MN_k = N_k$, and the result is obvious, so we assume that $k \geqq 3$. Since $AN_k \vee RB \subseteq MN_k \vee RB$, it suffices to prove that every non-trivial identity satisfied by $AN_k \vee RB$ is also satisfied by $MN_k \vee RB$. So suppose that $AN_k \vee RB$ satisfies $u = v$. Since RB satisfies $u = v$, u and v have the same first letters and the same last letters. Thus if both $|u|$ and $|v|$ are $\geqq k$, then we are done: we use the identities defining $MN_k \vee RB$ (from the previous Proposition) to deduce $u = v$. Otherwise, consider the case where $|u| < k$ or $|v| < k$ (or both). Since AN_k satisfies $u = v$, there exists a chain $u = u_0 = u_1 = \ldots = u_\ell = v$, with each step $u_i = u_{i+1}$ a consequence of either $xy = yx$ or $x_1 \ldots x_k = y_1 \ldots y_k$. But steps which are consequences of the first of these identities do not change the length of words involved, while steps which are consequences of the second identity can only be used on words u_i of length $\geqq k$. Thus if $|u| < k$ or $|v| < k$, then $|u_i| < k$ for all $0 \leqq i \leqq \ell$, and in fact the abelian variety A satisfies $u = v$. Then $M = A \vee RB$ also satisfies $u = v$, so $MN_k \vee RB$ does, too.

4. Joins with N_k

Volkov [4] has proved that if V is a finitely based variety then so is $V \vee N_k$, for $k \geqq 2$; Clarke [2] has given a method for converting a basis of identities for a variety V into a basis of identities for $V \vee Z$. In this section we try to extend both these results. For certain varieties V we are able to produce identities which define $V \vee N_k$, for all $k \geqq 2$. We begin with the varieties $B_{1,m}$, including the variety $B = B_{1,1}$ of bands.

PROPOSITION 4.1. *Let $m \geqq 1$ and $k \geqq 2$. Let a be the first number $\geqq k$ which is congruent to 1 modulo m. Then the variety $B_{1,m} \vee N_k$ is defined by the identities*

$$x^a y = xy^a = (xy)^a, \quad x^a = x^{(m+1)a}, \ \text{and} \ x_1 \dots x_k = (x_1 \dots x_k)^a.$$

PROOF. Let W be the variety defined by the given identities. First, since a is congruent to 1 modulo m, $x^a = x$ holds in $B_{1,m}$, and so $B_{1,m} \vee N_k$ is contained in W. Conversely, let C be any semigroup in W. Define $\Theta : C \to C$ by $\Theta(c) = c^a$ for all c in C. Then

1. Θ is a homomorphism, since $x^a y^a = (xy)^{a^2} = (xy)^a$ holds in W because a^2 and a are congruent modulo m;
2. Θ is a retraction, since $x^{a^2} = x^a$ holds in W;
3. $\Theta(C)$ is an ideal of C, since for any c and d in C we have $c^a d = (cd)^a \in \Theta(C)$, and similarly dc^a is in $\Theta(C)$;
4. $\Theta(C)$ is in $B_{1,m}$, since $x^a = (x^a)^{m+1}$ holds in W;
5. $C/\Theta(C)$ is in N_k, since for any $c_1, \dots, c_k$ in C, $c_1 \dots c_k = (c_1 \dots c_k)^a \in \Theta(C)$;

Therefore by Lemma 2.1, C is in $B_{1,m} \vee N_k$.

PROPOSITION 4.2. *Let $W \subseteqq B_{1,m}$ for $m \geqq 1$. Let $k \geqq 2$, and let a be the first number $\geqq k$ which is congruent to 1 modulo m. Let Σ be a basis for W, with $x = x^{m+1}$ in Σ. Write $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$, where $\Sigma_1 = \{x = x^{m+1}\}$, $\Sigma_2 = \{u = v \in \Sigma : |u|, |v| \geqq k\}$, and $\Sigma_3 = \Sigma - (\Sigma_1 \cup \Sigma_2)$. Let $\Sigma_3^* = \{u^* = v^* : u = v \in \Sigma_3\}$, where u^* is obtained from u by replacing each letter x in u by x^a, each time it occurs. Then $W \vee N_k$ is defined by the identities in $\Sigma_2 \cup \Sigma_3^*$ plus the additional identities $x^a y = xy^a = (xy)^a$, $x_1 \dots x_k = (x_1 \dots x_k)^a$, and $x^a = x^{(m+1)a}$.*

PROOF. Note that since $x^a = x$ holds in W, $W \vee N_k$ satisfies all of these identities. The proof then follows exactly that of the previous proposition, using Θ, up to part 4. This time we have $\Theta(C)$ in W, since by construction $\Theta(C)$ satisfies all the identities in Σ. The conclusion follows.

COROLLARY 4.3. *Let $k \geqq 2$, and let W be a subvariety of $B_{1,m}$ for some $m \geqq 1$. Then any semigroup in $W \vee N_k$ is a subdirect product of a semigroup in W and a k-nilpotent semigroup.*

COROLLARY 4.4. *Let* $W = V(x = x^2, u = v)$ *be a variety of bands. For* $k \geqq 2$, $W \vee N_k$ *is defined by the identities* $x^k y = xy^k = (xy)^k$, $x_1 \ldots x_k = (x_1 \ldots x_k)^k$, $x^k = x^{2k}$, *and either* $u = v$, *if both* $|u|$ *and* $|v|$ *are* $\geqq k$, *or* $u^* = v^*$, *otherwise, where* u^* *and* v^* *are formed from* u *and* v, *respectively, by replacing each letter* x *by* x^k.

We now give an equational description for the varieties $M_{n,m} \vee MN_k$ for $n \geqq 1$. These varieties turn out to be significant in the investigation of hyperidentities for varieties of semigroups, and the equations used are motivated by hyperidentity equations. Note that when $n \geqq k$ we have $MN_k \subsetneqq M_{n,m}$, and hence $M_{n,m} \vee MN_k = M_{n,m}$, so we now consider the case where $n < k$.

NOTATION 4.5. Let $k \geqq 2$ and $n, m \geqq 1$, with $n < k$. Set $s = k - n + 1$. We use $\Sigma_{n,m,k}$ for the set of identities

$$x_1^n x_2 \ldots x_s = x_1^{n+m} x_2 \ldots x_s,$$
$$x_1 x_2^n x_3 \ldots x_s = x_1 x_2^{n+m} x_3 \ldots x_s,$$
$$\ldots,$$
$$x_1 \ldots x_{s-1} x_s^n = x_1 \ldots x_{s-1} x_s^{n+m},$$

and

$$xyzw = xzyw.$$

The special case $n = 1$ of the following proposition may be dealt with by a structural proof using Lemma 2.1, but for $n \geqq 1$ we must turn to a syntactic proof.

PROPOSITION 4.6. *Let* $m \geqq 1$, $k \geqq 2$, *and* $1 \leqq n < k$. *Then* $M_{n,m} \vee MN_k = V(\Sigma_{n,m,k})$.

PROOF. Since $M_{n,m} \vee MN_k \subsetneqq V(\Sigma_{n,m,k})$, it will suffice to prove that any identity satisfied by $M_{n,m} \vee MN_k$ is also satisfied by $V(\Sigma_{n,m,k})$.

Let $u = v$ be a non-trivial identity satisfied by $M_{n,m} \vee MN_k$, and hence by both $M_{n,m}$ and MN_k. Then either $u = v$ is a consequence of medial, and so is certainly satisfied by $V(\Sigma_{n,m,k})$, or both $|u|$ and $|v|$ are $\geqq k$. So we will now assume that $k \leqq |u| \leqq |v|$. We will prove that there is a sequence $u = u_0 = u_1 = \ldots = u_r = v$ such that each move $u_i = u_{i+1}$ is a consequence of the medial identity or the identity $x^n = x^{n+m}$, and such that $|u_i| \geqq k$ for all $0 \leqq i \leqq r$. From this it will follow that $V(\Sigma_{n,m,k})$ also satisfies $u = v$.

We now describe how to produce such a sequence. First, by repeated use of the medial identity, we may write any word w in a "standard form" $\overline{w}$ as follows. Rewrite any string $(w_1 \ldots w_1)^c$ in w as $w_1^c w_2^c \ldots w_1^c$. Then as in the proof of Proposition 3.3, express the rewritten string as

$$x^a y_1^{a_1} \ldots y_p^{a_p} y^b,$$

where $x, y_1, \dots y_p, y$ are the distinct (except possibly $x = y$) letters occurring in the word w; y_ℓ occurs a_ℓ times in w for $1 \leqq \ell \leqq p$; and x and y occur a and b times, respectively, in w, except that if $x = y$ then $b = 1$ and x occurs $a+1$ times in w.

Now by construction $M_{n,m}$ satisfies $u = \overline{u}$ and $v = \overline{v}$, and in fact there are deductions of these two identities involving only words of length $\geqq k$. Since $M_{n,m}$ satisfies $u = v$, it also satisfies $\overline{u} = \overline{v}$. Also, $|u| = |\overline{u}|$ and $|v| = |\overline{v}|$. Thus it will suffice to produce a deduction of $\overline{u} = \overline{v}$ in $M_{n,m}$ in which the length of any intermediate word is $\geqq k$.

Consider first the case where x and y are distinct letters. Then we can write

$$\overline{u} = x^a y_1^{a_1} \dots y_p^{a_p} y^b, \quad \text{and} \quad \overline{v} = x^c y_1^{c_1} \dots y_p^{c_p} y^d.$$

Since $M_{n,m}$ satisfies $\overline{u} = \overline{v}$, we must have $a_\ell = c_\ell$ or a_ℓ and c_ℓ both $\geqq n$ and congruent modulo m, for each $1 \leqq \ell \leqq p$, and similarly for a and c and b and d. For any variable z in $\overline{u}$, the net change in power on z as we go from $\overline{u}$ to $\overline{v}$ can then be accomplished as a series of moves of the form $z^e = z^{e+m}$ (an increase) or $z^{e+m} = z^e$ (a decrease), for some $e \geqq n$. It is clear that having grouped together all occurrences of each such variable z, the moves done to one variable are independent of those done to another, and such moves can be done in any order. Therefore we can arrange to move from $\overline{u}$ to $\overline{v}$ in such a way that all increases are done first, and then any decreases. Since $|u|$ and $|v|$ are $\geqq k$, this guarantees that any intermediate word in the sequence of moves also has length $\geqq k$, as required.

The case $x = y$ is handled in much the same way. This time we have

$$\overline{u} = x^a y_1^{a_1} \dots y_p^{a_p} x \quad \text{and} \quad \overline{v} = x^c y_1^{c_1} \dots y_p^{c_p} x,$$

which we will simplify a bit and write as $\overline{u} = x^a w x$ and $\overline{v} = x^c w' x$, where w and w' are words not involving the letter x. As in the $x \neq y$ case, we can always change w to w' using only $M_{n,m}$ identities by doing all the necessary increases first, then all the decreases, so that the greatest possible length is maintained. So we concentrate now on the letter x. If $a = c$, we are done. Otherwise, we must have a and c both $\geqq n$, and congruent modulo m. So again the net change in power on x is either an increase or a decrease, by a multiple of m. If this net change is an increase, we do it first, then change w to w' as previously described; if the net change is a decrease, we do it after the change from w to w' is made. In either case we move from $\overline{u} = x^a w x$ to $\overline{v} = x^c w' x$, maintaining at each stage a word-length $\geqq k$. This completes the proof of the proposition.

We will conclude this section with the join $A_{n,m} \vee RB \vee MN_k$. The syntactic proof given below combines the arguments used for $A_{n,m} \vee RB$ in Proposition 3.1 and for $M_{n,m} \vee MN_k$ in Proposition 4.6.

PROPOSITION 4.7. *Let $k \geqq 2$, $2 \leqq n < k$, and $m \geqq 1$. The variety $A_{n,m} \vee$*

$RB \vee MN_k$ is defined by the following identities:

$$x_1^n \ldots x_{k-n+1} = x_1^{n+m} \ldots x_{k-n+1},$$
$$x_1 x_2^n \ldots x_{k-n+1} = x_1 x_2^{n+m} \ldots x_{k-n+1},$$
$$\ldots,$$
$$x_1 \ldots x_{k-n+1}^n = x_1 \ldots x_{k-n+1}^{n+m},$$

and

$$x^{n-1} y_1 \ldots y_{k-n} x = x^{n-1+m} y_1 \ldots y_{k-n} x.$$

PROOF. Let U be the variety defined by the given identities. Certainly $A_{n,m} \vee RB \vee MN_k \subseteq U$. Conversely, we show that any non-trivial identity $u = v$ satisfied by $A_{n,m}$, RB and MN_k is also satisfied by U.

When MN_k satisfies $u = v$, either M and hence U satisfies $u = v$, and we are done, or $|u|$ and $|v|$ are both $\geqq k$. Since RB satisfies $u = v$, u and v have the same first letter, x say, and the same last letter, y say, with x and y possibly equal. Since $A_{n,m}$ satisfies $u = v$, u and v contain exactly the same letters, and for any letter z in u or v, either the number of occurrences of z in u is equal to the number of occurrences of z in v, or these two quantities are $\geqq n$ and are congruent modulo m. Therefore we will transform u and v into the standard from $\overline{u}$ and $\overline{v}$ of Propositions 3.1 and 4.6. As before, $A_{n,m}$ and $M_{n,m}$ still satisfy $u = \overline{u}$ and $v = \overline{v}$, and $|u| = |\overline{u}|$ and $|v| = |\overline{v}|$. The two cases, $x = y$ and $x \neq y$, are handled much as in the proof of Proposition 4.6.

REFERENCES

[1] BURRIS, S. and SANKAPPANAVAR, H. P., *A Course in Universal Algebra*, Graduate Texts in Mathematics, Vol. 78, Springer-Verlag, New York, 1981. *MR* **83k**: 08001

[2] CLARKE, G. T., Semigroup varieties of inflations of unions of groups, *Semigroup Forum* **23** (1981), 311–319. *MR* **83a**: 20083

[3] PETRICH, M., All subvarieties of a certain variety of semigroups, *Semigroup Forum* **7** (1974), 104–152. *MR* **52** #3395

[4] VOLKOV, M. V., On the join of varieties, *Simon Stevin* **58** (1984), 311–317. *MR* **86f**: 08007

[5] WISMATH, S. L., Hyperidentities for some varieties of semigroups, *Algebra Universalis* **27** (1990), 111–127. *MR* **90i**: 20062

[6] WISMATH, S. L., Hyperidentities for some varieties of commutative semigroups, *Algebra Universalis* **28** (1991), 245–273.

(Received September 12, 1988)

DEPARTMENT OF MATHEMATICS AND
COMPUTER SCIENCE
THE UNIVERSITY OF LETHBRIDGE
4401 UNIVERSITY DRIVE
LETHBRIDGE, ALBERTA
T1K 3M4
CANADA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 339–347

VARIETIES OF LOCALLY BOOLEAN ALGEBRAS

T. WESOLOWSKI

Abstract

Following J. Płonka (see [10]), an algebra $\mathfrak{A} = (A; \vee, \wedge, ')$ of type $\langle 2, 2, 1\rangle$ will be called a locally Boolean algebra if $(A\colon \vee, \wedge)$ is a distributive lattice and there exists a congruence $\sim$ of $\mathfrak{A}$ such that all congruence classes $[a]_\sim$, $a \in A$ are Boolean algebras with respect to the operations $\vee, \wedge$ and $'$. It was proved in [10] that the class $l(\mathbf{B})$ of all such algebras is a variety. In [14] all subdirectly irreducible members of $l(\mathbf{B})$ were described. In this paper using irreducibles we characterize the lattice of all subvarieties of $l(\mathbf{B})$.

0. Introduction

Our nomenclature is basically that of [5] and [6]. In [10] J. Płonka defined a locally Boolean algebra as an algebra $\mathfrak{A} = (A; \vee, \wedge, ')$ of type $\langle 2, 2, 1\rangle$ where $(A\colon \vee, \wedge)$ is a distributive lattice and there exists a congruence $\sim$ of $\mathfrak{A}$ such that all congruence classes $[a]_\sim$, $a \in A$ are Boolean algebras with respect to the operations $\vee, \wedge$ and $'$ restricted to $[a]_\sim$. Locally Boolean algebras have an interesting application in logic and were investigated from this point of view in [7]. In [13] a representation theorem for some algebras of this kind was given.

It was proved in [10] that the class $l(\mathbf{B})$ of all locally Boolean algebras is a variety determined by the following identities:

(1) identities in $\vee$ and $\wedge$ which define distributive lattices;
(2) $(x')' = x$;
(3) $(x \vee x')' = x \wedge x'$;
(4) $(x \vee y) \wedge (x \vee y)' = (x \wedge x') \vee (y \wedge y')$;
(5) $(x \wedge y) \vee (x \wedge y)' = (x \vee x') \wedge (y \vee y')$.

Further, there exists at most one congruence of an algebra $\mathfrak{A} = (A; \vee, \wedge, ')$ of type $\langle 2, 2, 1\rangle$ which decomposes $\mathfrak{A}$ into its Boolean subalgebras, namely

(i) $a \sim b$ iff $a \wedge a' = b \wedge b'$ for $a, b \in A$.

In [14] all subdirectly irreducible members of $l(\mathbf{B})$ were described. To construct these algebras we used disjunctive and codisjunctive distributive lattices.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 06DXX; Secondary 08BXX.

Key words and phrases. Congruence lattice, Boolean algebra, ordered sum of lattices, subdirectly irreducible algebra, variety, lattice of varieties of algebras.

The notion of a disjunctive lattice is an utilization of a notion of disjunctive poset for lattices. In accordance to the definition (see [3]) a poset $(L; \leqq)$ is disjunctive if for any $a, b \in L$, $b \nleqq a$ there exists $c \in L$ such that $c \leqq b$ and there does not exist in L a common lower bound of a and c. If a poset $(L; \leqq)$ has the least element 0_L, then $(L; \leqq)$ is called disjunctive if for any $a, b \in L$, $b \nleqq a$ there exists $c \in L \setminus \{0_L\}$ such that $c \leqq b$ and 0_L is the only common lower bound of a and c (see [12]). A disjunctive poset with the least element, which is a lattice will be called disjunctive lattice (see [14]). It is easy to see that a lattice $(L; \vee, \wedge)$ with zero 0_L is disjunctive iff for any $a, b \in L$ the following condition holds:

(ii) if $a < b$ then there exists $c \in L \setminus \{0_L\}$ such that $c \leqq b$ and $a \wedge c = 0_L$.

Dually, a lattice $(L; \vee, \wedge)$ with unit 1_L is called codisjunctive if for all $a, b \in L$ we have:

(iii) if $a < b$ then there exists $c \in L \setminus \{1_L\}$ such that $a \leqq c$ and $b \vee c = 1_L$.

Obviously each Boolean lattice is simultaneously disjunctive and codisjunctive and a finite disjunctive (codisjunctive) distributive lattice is Boolean. An infinite pseudo-Boolean lattice is an example of an infinite disjunctive distributive lattice which is not Boolean.

Denote by $\mathbf{DL}$ ($\mathbf{CL}$) the class of all disjunctive (codisjunctive) distributive lattices. For $\mathcal{L}_1 = (L_1; \vee, \wedge) \in \mathbf{CL}$ and $\mathcal{L}_2 = (L_2; \vee, \wedge) \in \mathbf{DL}$ with $L_1 \cap L_2 = \emptyset$ let $\mathcal{L}_1 \oplus \mathcal{L}_2 = (L_1 \cup L_2; \vee, \wedge, ')$ be an algebra of type $\langle 2, 2, 1 \rangle$ such that $(L_1 \cup \cup L_2; \vee, \wedge)$ is the order sum of $\mathcal{L}_1$ and $\mathcal{L}_2$ (see [1], p. 39) and the operation $'$ is defined as follows: $1'_{L_1} = 0_{L_2}$, $0'_{L_2} = 1_{L_1}$ and $a' = a$ otherwise. Denote by $\mathbf{2}$ the algebra $(\{0,1\}; \vee, \wedge, ')$ in which $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$ and $a' = a$ for all $a, b \in \{0, 1\}$. It was proved in [14] that algebras of the form

$$(*) \qquad \mathcal{L}_1 \oplus \mathcal{L}_2 \quad \text{where} \quad \mathcal{L}_1 \in \mathbf{CL} \quad \text{and} \quad \mathcal{L}_2 \in \mathbf{DL}$$

and algebras isomorphic to $\mathbf{2}$ are the only nondegenerated subdirectly irreducible members of $l(\mathbf{B})$. Note that for $\mathcal{L}_1, \mathcal{L}'_1 \in \mathbf{CL}$ and $\mathcal{L}_2, \mathcal{L}'_2 \in \mathbf{DL}$ we have:

(iv) if L'_1 is a sublattice of $\mathcal{L}_1$ such that $1_{L'_1} = 1_{L_1}$ then $\mathcal{L}'_1 \oplus \mathcal{L}_2$ is a subalgebra of $\mathcal{L}_1 \oplus \mathcal{L}_2$;

(v) if L'_2 is a sublattice of $\mathcal{L}_2$ such that $0_{L'_2} = 0_{L_2}$ then $\mathcal{L}_1 \oplus \mathcal{L}'_2$ is a subalgebra of $\mathcal{L}_1 \oplus \mathcal{L}_2$.

For a non-negative integer n denote by $\mathbf{2}^n = (B_n; \vee, \wedge)$ the Boolean lattice of all subsets of the set $\{0, 1, \dots, n-1\}$. Observe that each finitely generated subalgebra of an algebra of the form $(*)$ is finite, so using (iv) and (v) we obtain:

(vi) finitely generated subalgebra $\mathfrak{A}$ of an algebra of the form $(*)$ can be embedded into the algebra $\mathbf{2}^n \oplus \mathbf{2}^m$ for some $n, m \geqq 0$; if $\mathcal{L}_1 \cong \mathbf{2}^n$ for some $n \geqq 0$ then $\mathfrak{A}$ embeds into $\mathbf{2}^n \oplus \mathbf{2}^s$ for some $s \geqq 0$; if $\mathcal{L}_2 \cong \mathbf{2}^m$ for some $m \geqq 0$ then $\mathfrak{A}$ embeds into $\mathbf{2}^r \oplus \mathbf{2}^m$ for some $r \geqq 0$.

THEOREM 1. *Let $\mathcal{L} = (L; \vee, \wedge) \in \mathbf{DL}$. Then for each integer $n \geqq 1$ the following conditions are equivalent:*

(a) $|L| > 2^n$;
(b) *there exist elements* $d_0, \ldots, d_n \in L \setminus \{0_L\}$ *such that* $d_0 \wedge \ldots \wedge d_n = 0_L$ *and* $d_0 \wedge \ldots \wedge d_{i-1} \wedge d_{i+1} \wedge \ldots \wedge d_n \neq 0_L$ *for* $i = 0, 1, \ldots, n$.

PROOF. It follows immediately from the Kuroš–Ore theorem (see [2]).

THEOREM 2. *Let* $\mathcal{L} = (L; \vee, \wedge) \in \mathbf{CL}$. *Then for each integer* $n \geqq 1$ *the following conditions are equivalent:*
(a) $|L| > 2^n$;
(b) *there exist elements* $c_0, \ldots, c_n \in L \setminus \{1_L\}$ *such that* $c_0 \vee \ldots \vee c_n = 1_L$ *and* $c_0 \vee \ldots \vee c_{i-1} \vee c_{i+1} \vee \ldots \vee c_n \neq 1_L$ *for* $i = 0, 1, \ldots, n$.

COROLLARY 1. *Let* $\mathcal{L} \in \mathbf{DL}$ $(\mathcal{L} \in \mathbf{CL})$. *Then* $\mathcal{L}$ *is infinite iff for each* $n \geqq 0$ *the Boolean lattice* $(B_n; \vee, \wedge)$ *is (up to isomorphism) a subalgebra of* $\mathcal{L}$. *Moreover, if* $\mathcal{L} \in \mathbf{DL}$ *then* $0_L = 0_{B_n}$ *and if* $\mathcal{L} \in \mathbf{CL}$ *then* $1_L = 1_{B_n}$.

PROOF. If $\mathcal{L}$ is finite then it is isomorphic to $\mathbf{2}^n$ for some $n \geqq 0$. Hence $\mathbf{2}^{n+1}$ is not a subalgebra of $\mathcal{L}$. Let $\mathcal{L} \in \mathbf{DL}$ be infinite and $n \geqq 0$. Obviously $(B_0; \vee, \wedge)$ and $(B_1; \vee, \wedge)$ are subalgebras of $\mathcal{L}$. If $n - 1 \geqq 1$ then we can use Theorem 1 since $|L| > \mathbf{2}^{n-1}$. Hence there exist elements $d_0, \ldots, d_{n-1} \in \in L \setminus \{0_L\}$ such that $d_0 \wedge \ldots \wedge d_{n-1} = 0_L$ and all elements $e_i = d_0 \wedge \ldots \wedge d_{i-1} \wedge d_{i+1} \wedge \ldots \wedge d_{n-1}$, $i = 0, \ldots, n-1$ are different from 0_L. Obviously, $e_i \wedge e_j = = 0_L$ for $i \neq j$ $(i, j = 0, \ldots, n-1)$, so $e_0, \ldots, e_{n-1}$ are the all atoms of the sublattice L' of $\mathcal{L}$ generated by the set $\{e_0, \ldots, e_{n-1}\}$. Hence $\mathcal{L}' = (L'; \vee, \wedge)$ is isomorphic to $\mathbf{2}^n = (B_n; \vee, \wedge)$ and $0_L = 0_{L'} = 0_{B_n}$. The proof for $\mathcal{L} \in \mathbf{CL}$ is similar.

1. Equational characterization of some subvarieties of $l(\mathbf{B})$

For an ordinal $\alpha < \omega$ denote by $\mathbb{S}_\alpha^\omega$ the class of all algebras of the form $(*)$ where $|L_1| = \mathbf{2}^\alpha$. Symmetrically, if $\beta < \omega$ then $\mathbb{S}_\omega^\beta$ denotes the class of all algebras of the form $(*)$ with $|L_2| = \mathbf{2}^\beta$. If $\alpha, \beta < \omega$ then we denote by $\mathbb{S}_\alpha^\beta$ the intersection of $\mathbb{S}_\alpha^\omega$ and $\mathbb{S}_\omega^\beta$. Finally, let $\mathbb{S}_\omega^\omega$ be the class of all algebras of the form $(*)$. Let us put

(vii) $\mathbf{K}_\alpha^\beta = \mathrm{HSP}(\mathbb{S}_\alpha^\beta \cup \{\mathbf{2}\})$ for $\alpha, \beta < \omega + 1$.

Observe that $\mathrm{HSP}(\mathbb{S}_0^0) = \mathbf{B}$ where $\mathbf{B}$ is the variety of all Boolean algebras. Further, $\mathrm{HSP}(\{\mathbf{2}\}) = \mathbf{D}$ where $\mathbf{D}$ is the variety of all distributive lattices with additional unary operation $'$ satisfying $x' = x$. It is easy to see that if $\alpha \neq 0$ or $\beta \neq 0$ then $\mathbf{2} \in \mathrm{HSP}(\mathbb{S}_\alpha^\beta)$. In fact, if $\mathfrak{A} \in \mathbb{S}_\alpha^\beta$ for $\alpha + \beta > 0$ then using the congruence $\sim$ of $\mathfrak{A}$ defined by (i) we conclude that $\mathfrak{A}/\sim \in \mathbf{D}$ and $\mathfrak{A}/\sim$ has at least two elements. Therefore, $\mathbf{2}$ is isomorphic to a subalgebra of $\mathfrak{A}/\sim$. Hence we have

(viii) $\mathbf{K}_\alpha^\beta = \mathrm{HSP}(\mathbb{S}_\alpha^\beta)$ for $\alpha, \beta < \omega + 1$ and $\alpha + \beta > 0$.

REMARK. It was proved in [4] that $\mathbf{K}_0^0 = \mathbf{B} \vee \mathbf{D} = \mathbf{B} \times \mathbf{D}$, i.e. $\mathbf{K}_0^0$ consists of all direct products $\mathfrak{A}_1 \times \mathfrak{A}_2$ of algebras $\mathfrak{A}_1 \in \mathbf{B}$ and $\mathfrak{A}_2 \in \mathbf{D}$. Moreover, varieties $\mathbf{B}$ and $\mathbf{D}$ are the only proper and nontrivial subvarieties of $\mathbf{K}_0^0$.

For variables $x_0, x_1, \dots, x_\alpha$ where $0 < \alpha < \omega$, denote by $p(x_0, \dots, x_\alpha)$ the term $x_0 \wedge x_1 \wedge \dots \wedge x_\alpha$ and by $q(x_0, \dots, x_\alpha)$ the term $x_0 \vee x_1 \vee \dots \vee x_\alpha$. For $i = 0, 1, \dots, \alpha$ let $p_i(x_0, x_1, \dots, x_\alpha)$ be the term $x_0 \wedge \dots \wedge x_{i-1} \wedge x_{i+1} \wedge \dots \wedge x_\alpha$ and $q_i(x_0, x_1, \dots, x_\alpha)$ be the term $x_0 \vee \dots \vee x_{i-1} \vee x_{i+1} \vee \dots \vee x_\alpha$. If φ is a term of type $\langle 2,2,1 \rangle$ then we denote by $0(\varphi)$ the term $\varphi \wedge \varphi'$ and by $1(\varphi)$ the term $\varphi \vee \varphi'$. Observe that if $0(\varphi)^{\mathfrak{A}}$ and $1(\varphi)^{\mathfrak{A}}$ are the realizations of $0(\varphi)$ and $1(\varphi)$, respectively in an algebra $\mathfrak{A} = (A; \vee, \wedge, ') \in l(\mathbf{B})$ and φ is a term of variables $y_0, y_1, \dots, y_\alpha$, then for all $a_0, a_1, \dots, a_\alpha \in A$, $0(\varphi)^{\mathfrak{A}}(a_0, a_1, \dots, a_\alpha)$ is the zero and $1(\varphi)^{\mathfrak{A}}(a_0, a_1, \dots, a_\alpha)$ is the unit of the Boolean subalgebra $[\varphi^{\mathfrak{A}}(a_0, a_1, \dots, a_\alpha)]_\sim$ of $\mathfrak{A}$.

Let us consider the following identities:

$$(\mathrm{I}^0) \qquad 0(p_0(x_0, x_1)) \wedge 1(p_1(x_0, x_1)) = 0(p(x_0, x_1));$$

$$(\mathrm{I}^\alpha) \qquad 0(p_0(x_0, \dots, x_\alpha)) \wedge \dots \wedge 0(p_\alpha(x_0, \dots, x_\alpha)) = 0(p(x_0, \dots, x_\alpha)) \quad \text{for } 0 < \alpha < \omega;$$

$$(\mathrm{I}_0) \qquad 1(q_0(x_0, x_1)) \vee 0(q_1(x_0, x_1)) = 1(q(x_0, x_1));$$

$$(\mathrm{I}_\alpha) \qquad 1(q_0(x_0, \dots, x_\alpha)) \vee \dots \vee 1(q_\alpha(x_0, \dots, x_\alpha)) = 1(q(x_0, \dots, x_\alpha)) \quad \text{for } 0 < \alpha < \omega.$$

THEOREM 3. (a) *If $\alpha < \omega$ then the variety $\mathbf{K}_\omega^\alpha$ is determined by identities* (1)–(5) *and* (I^α)*;*

(b) *If $\alpha < \omega$ then the variety $\mathbf{K}_\alpha^\omega$ is determined by identities* (1)–(5) *and* (I_α)*;*

(c) *If $\alpha, \beta < \omega$ then the variety $\mathbf{K}_\alpha^\beta$ is determined by identities* (1)–(5), (I_α) *and* (I^β).

PROOF. (a) It is easy to see that each algebra $\mathcal{L}_1 \oplus \mathbf{2}^0 \in \mathbf{S}_\omega^0$ satisfies (I^0). On the other hand, if an algebra of the form $(*)$ satisfies (I^0) then $|L_2| = 1$. Indeed, otherwise there exists $a_1 \in L_2$ such that $a_0 = 0_{L_2} < a_1$ so $0(p_0(a_0, a_1)) \wedge 1(p_1(a_0, a_1)) = 0(a_1) \wedge 1(a_0) = a_1 \wedge a_0 \neq 1_{L_1} = 0(a_0) = 0(p(a_0, a_1))$.

Let $0 < \alpha < \omega$ and $\mathcal{L}_1 \oplus \mathcal{L}_2$ be an algebra of the form $(*)$ with $|L_2| = \mathbf{2}^\alpha$. Let us take $a_0, \dots, a_\alpha \in L_1 \cup L_2$ and put $0(p_0(a_0, \dots, a_\alpha)) \wedge \dots \wedge 0(p_\alpha(a_0, \dots, a_\alpha)) = b$. Since $p(a_0, \dots, a_\alpha) = a_i \wedge p_i(a_0, \dots, a_\alpha)) \leqq p_i(a_0, \dots, a_\alpha)$ so by (4) we have $0(p(a_0, \dots, a_\alpha)) \leqq 0(p_i(a_0, \dots, a_\alpha))$ for $i = 0, \dots, \alpha$. Hence $0(p(a_0, \dots, a_\alpha)) \leqq b$. We shall show that $b \leqq 0(p(a_0, \dots, a_\alpha))$.

If $a_i < 1_{L_1}$ for some $i \in \{0, \dots, \alpha\}$ then obviously $p(a_0, \dots, a_\alpha) < 1_{L_1}$ and $p_k(a_0, \dots, a_\alpha) < 1_{L_1}$ for $k = 0, \dots, i-1, i+1, \dots, \alpha$. Hence $b \leqq 0(p_0(a_0, \dots, a_\alpha)) \wedge \dots \wedge 0(p_{i-1}(a_0, \dots, a_\alpha)) \wedge 0(p_{i+1}(a_0, \dots, a_\alpha)) \wedge \dots \wedge 0(p_\alpha(a_0, \dots, a_\alpha)) = p_0(a_0, \dots, a_\alpha) \wedge \dots \wedge p_{i-1}(a_0, \dots, a_\alpha) \wedge p_{i+1}(a_0, \dots, a_\alpha) \wedge \dots \wedge p_\alpha(a_0, \dots, a_\alpha) = p(a_0, \dots, a_\alpha) = 0(p(a_0, \dots, a_\alpha))$.

If $a_0, \dots, a_\alpha \geqq 1_{L_1}$ and $a_i = 1_{L_1}$ for some $i \in \{0, \dots, \alpha\}$ then $p(a_0, \dots, a_\alpha) = 1_{L_1}$ and $p_k(a_0, \dots, a_\alpha) = 1_{L_1}$ for $k = 0, \dots, i-1, i+1, \dots, \alpha$. Hence $b \leqq$

$\leqq 0(p_0(a_0,\dots,a_\alpha)) \wedge \dots \wedge 0(p_{i-1}(a_0,\dots,a_\alpha)) \wedge 0(p_{i+1}(a_0,\dots,a_\alpha))\wedge$ $\wedge \dots \wedge 0(p_\alpha(a_0,\dots,a_\alpha)) = 1_{L_1} = 0(p(a_0,\dots,a_\alpha))$.

Let $a_0,\dots,a_\alpha \in L_2$. If $p_i(a_0,\dots,a_\alpha) = 0_{L_2}$ for some $i \in \{0,\dots,\alpha\}$ then also $p(a_0,\dots,a_\alpha) = 0_{L_2}$. Hence $b \leqq 0(p_i(a_0,\dots,a_\alpha)) = 1_{L_1} = 0(p(a_0,\dots,a_\alpha))$. If $p_i(a_0,\dots,a_\alpha) \neq 0_{L_2}$ for $i = 0,\dots,\alpha$ then $a_0,\dots,a_\alpha \in L_2 \setminus \{0_{L_2}\}$ and we can use Theorem 1 since $|L_2| \leqq \mathbf{2}^\alpha$. Hence $p(a_0,\dots,a_\alpha) \neq 0_{L_2}$ and therefore $b = p_0(a_0,\dots,a_\alpha) \wedge \dots \wedge p_\alpha(a_0,\dots,a_\alpha) = p(a_0,\dots,a_\alpha) = 0(p(a_0,\dots,a_\alpha))$.

We proved that each algebra from $\mathbb{S}^\alpha_\omega$ satisfies (I^α) so by (viii), each algebra from $\mathbf{K}^\alpha_\omega$ does.

Now let us assume that an algebra $\mathcal{L}_1 \oplus \mathcal{L}_2$ of the form ($*$) satisfies (I^α) where $0 < \alpha < \omega$. Then by Theorem 1, $|L_2| \leqq \mathbf{2}^\alpha$ so $\mathcal{L}_2$ is isomorphic to the Boolean lattice $\mathbf{2}^n = (B_n; \vee, \wedge)$ for some $n \leqq \alpha$. But B_n is (up to isomorphism) a subalgebra of $\mathbf{2}^\alpha = (B_\alpha; \vee, \wedge)$ and $0_{B_n} = 0_{B_\alpha}$. Thus by (v), $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\alpha_\omega$. Consequently, if $\mathfrak{A} \in l(\mathbf{B})$ satisfies (I^α) then each subdirectly irreducible factor of $\mathfrak{A}$ does, so $\mathfrak{A} \in \mathbf{K}^\alpha_\omega$.

(b) The proof is similar to that of (a) but we use Theorem 2.

(c) It follows from (a) and (b).

Corollary 2. (a) *If $\alpha < \omega$ then $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\alpha_\omega$ iff $|L_2| \leqq \mathbf{2}^\alpha$;*

(b) *If $\alpha < \omega$ then $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\omega_\alpha$ iff $|L_1| \leqq \mathbf{2}^\alpha$;*

(c) *If $\alpha, \beta < \omega$ then $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\beta_\alpha$ iff $|L_1| \leqq \mathbf{2}^\alpha$ and $|L_2| \leqq \mathbf{2}^\beta$.*

Proof. (a) If $|L_2| \leqq \mathbf{2}^\alpha$ then $|L_2| = \mathbf{2}^k$ for some $k \leqq \alpha$. Therefore $\mathcal{L}_1 \oplus \mathcal{L}_2$ is (up to isomorphism) a subalgebra of $\mathcal{L}_1 \oplus \mathbf{2}^\alpha \in \mathbf{K}^\alpha_\omega$ so $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\alpha_\omega$. If $|L_2| = 2 > \mathbf{2}^0$ and $L_2 = \{0_{L_2}, a\}$ where $0_{L_2} < a$ then

$$0\,(p_0(0_{L_2}, a)) \wedge 1\,(p_1(0_{L_2}, a)) = 0(a) \wedge 1(0_{L_2}) = a \wedge 0_{L_2} \neq 1_{L_1} = 0\,(p(0_{L_2}, a)),$$

so $\mathcal{L}_1 \oplus \mathcal{L}_2$ does not satisfy (I^0). Hence by Theorem 3(a), $\mathcal{L}_1 \oplus \mathcal{L}_2 \notin \mathbf{K}^\alpha_\omega$. If $|L_2| > \mathbf{2}^\alpha$ for $0 < \alpha < \omega$ then by Theorem 1, there exist elements $d_0,\dots,d_\alpha \in$ $\in L_2 \setminus \{0_{L_2}\}$ such that $p(d_0,\dots,d_\alpha) = 0_{L_2}$ and $0_{L_2} < p_i(d_0,\dots,d_\alpha)$ for $i =$ $= 0,\dots,\alpha$. Hence $0_{L_2} < 0(p_i(d_0,\dots,d_\alpha))$ for $i = 0,\dots,\alpha$ and therefore $0_{L_2} \leqq$ $\leqq 0(p_0(d_0,\dots,d_\alpha)) \wedge \dots \wedge 0(p_\alpha(d_0,\dots,d_\alpha))$. But $0(p(d_0,\dots,d_\alpha)) = 1_{L_1} <$ 0_{L_2} what proves that $\mathcal{L}_1 \oplus \mathcal{L}_2 \notin \mathbf{K}^\alpha_\omega$.

(b) The proof is similar to that of (a).

(c) It follows from (a) and (b).

Corollary 3. *Let $\alpha, \alpha', \beta, \beta' < \omega + 1$ be the ordinals. Then the following conditions hold:*

(a) $\quad \alpha < \alpha'$ *iff* $\mathbf{K}^\beta_\alpha \subsetneqq \mathbf{K}^\beta_{\alpha'}$;

(b) $\quad \beta < \beta'$ *iff* $\mathbf{K}^\beta_\alpha \subsetneqq \mathbf{K}^{\beta'}_\alpha$.

Proof. (a) Let $\alpha < \alpha'$. Hence $\alpha' \neq 0$ so by (vii) and (viii), $\mathbf{2} \in \mathbf{K}^\beta_{\alpha'}$. Let $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbb{S}^\beta_\alpha$. Then $|L_1| = 2^\alpha$ since $\alpha < \omega$. If $\alpha' = \omega$ then obviously $\mathcal{L}_1 \oplus \mathcal{L}_2 \in$ $\in \mathbf{K}^\beta_{\alpha'}$; if $\alpha' < \omega$ then $|L_1| = 2^\alpha < 2^{\alpha'}$ and by Corollary 2, also $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}^\beta_{\alpha'}$. Hence $\mathbf{K}^\beta_\alpha \subseteqq \mathbf{K}^\beta_{\alpha'}$. Since α is finite so $\alpha + 1$ is. Therefore, if $\beta < \omega$ then

$\mathbf{2}^{\alpha+1} \oplus \mathbf{2}^{\beta} \in \mathsf{K}^{\beta}_{\alpha+1} \subseteqq \mathsf{K}^{\beta}_{\alpha'}$, since $\alpha+1 \leqq \alpha'$. But $\mathbf{2}^{\alpha+1} \oplus \mathbf{2}^{\beta} \notin \mathsf{K}^{\beta}_{\alpha}$ by Corollary 2 (b). Hence $\mathsf{K}^{\beta}_{\alpha} \subsetneqq \mathsf{K}^{\beta}_{\alpha'}$. If $\beta=\omega$ and $\mathcal{L}_2 \in \mathsf{DL}$ then $\mathbf{2}^{\alpha+1} \oplus \mathcal{L}_2 \in \mathsf{K}^{\omega}_{\alpha+1} \subseteqq \mathsf{K}^{\omega}_{\alpha'}$, and $\mathbf{2}^{\alpha+1} \oplus \mathcal{L}_2 \notin \mathsf{K}^{\omega}_{\alpha}$. Therefore, $\mathsf{K}^{\beta}_{\alpha} \subsetneqq \mathsf{K}^{\beta}_{\alpha'}$ for each $\beta<\omega+1$. The proof that $\mathsf{K}^{\beta}_{\alpha} \subsetneqq \mathsf{K}^{\beta}_{\alpha'}$ implies $\alpha<\alpha'$ is obvious.

(b) The proof is similar to that of (a).

2. The lattice of all subvarieties of $l(\mathsf{B})$

Denote by $(\mathsf{V}(l(\mathsf{B})); \vee, \wedge)$ the lattice of all subvarieties of $l(\mathsf{B})$.

LEMMA 1. *If $\mathsf{K} \in \mathsf{V}(l(\mathsf{B}))$ and $\mathsf{K}^0_0 \nsubseteq \mathsf{K}$ then $\mathsf{K}=\mathsf{B}$ or $\mathsf{K}=\mathsf{D}$ or K is the trivial variety of type $\langle 2,2,1\rangle$.*

PROOF. It follows from (vii) that $\mathsf{K}^0_0 = \mathrm{HSP}(\{\mathbf{2}^0 \oplus \mathbf{2}^0, \mathbf{2}\})$. But $\mathbf{2}^0 \oplus \mathbf{2}^0$ is a subalgebra of each algebra of the form (*). Hence, if $\mathbf{2}^0 \oplus \mathbf{2}^0 \notin \mathsf{K}$ then $\mathsf{K}=\mathrm{HSP}(\{\mathbf{2}\})=\mathsf{D}$. If $\mathbf{2}^0 \oplus \mathbf{2}^0 \in \mathsf{K}$ then $\mathbf{2} \notin \mathsf{K}$ since $\mathsf{K}^0_0 \nsubseteq \mathsf{K}$. Therefore, if K is not trivial then by (vii) and (viii), $\mathbf{2}^0 \oplus \mathbf{2}^0$ is the only (up to isomorphism) non-degenerated subdirectly irreducible member of K. Thus $\mathsf{K}=\mathsf{B}$.

Let $\mathsf{V}_0(l(\mathsf{B}))$ be the dual ideal of $\mathsf{V}(l(\mathsf{B}))$ generated by the variety $\mathsf{K}^0_0 = \mathsf{B} \vee \mathsf{D}$. It follows from Lemma 1 and the Remark that $\mathsf{V}(l(\mathsf{B})) = \mathsf{V}_0(l(\mathsf{B})) \cup \{\mathsf{B}, \mathsf{D}, \mathsf{T}\}$ where $\mathsf{B} \wedge \mathsf{D} = \mathsf{T}$ and T is the trivial variety of type $\langle 2,2,1\rangle$.

THEOREM 4. *Varieties $\mathsf{K}^{\beta}_{\alpha}$ for $\alpha,\beta<\omega+1$ form a $\wedge$-subsemilattice of $\mathsf{V}_0(l(\mathsf{B}))$.*

PROOF. Let $\alpha=\min\{\alpha_1,\alpha_2\}$ and $\beta=\min\{\beta_1,\beta_2\}$ for α_1,α_2, $\beta_1,\beta_2 < \omega+1$. It follows from Corollary 2 and Corollary 3 that $\mathsf{K}^{\alpha_1}_{\beta_1} \wedge \mathsf{K}^{\alpha_2}_{\beta_2} = \mathsf{K}^{\beta}_{\alpha}$.

LEMMA 2. *Let $\mathsf{K} \in \mathsf{V}_0(l(\mathsf{B}))$ and $\alpha,\beta<\omega+1$. Then the following conditions hold:*

(a) *if $\mathsf{K}^m_{\alpha} \subseteqq \mathsf{K}$ for all $m<\omega$ then $\mathsf{K}^{\omega}_{\alpha} \subseteqq \mathsf{K}$;*

(b) *if $\mathsf{K}^{\beta}_{n} \subseteqq \mathsf{K}$ for all $n<\omega$ then $\mathsf{K}^{\beta}_{\omega} \subseteqq \mathsf{K}$;*

(a) *if $\mathsf{K}^m_n \subseteqq \mathsf{K}$ for all $n,m<\omega$ then $\mathsf{K}^{\omega}_{\omega} \subseteqq \mathsf{K}$.*

PROOF. (a) Let $\mathbf{2}^{\alpha} \oplus \mathcal{L}_2 \in \mathsf{K}^{\omega}_{\alpha}$, $\varphi=\psi$ be an identity from $\mathrm{Id}(\mathsf{K})$ on variables $x_0,\dots,x_r$ and let $a_0,\dots,a_r$ be the elements in $\mathbf{2}^{\alpha} \oplus \mathcal{L}_2$. By (vi), the subalgebra $\mathfrak{A}$ of $\mathbf{2}^{\alpha} \oplus \mathcal{L}_2$ generated by $a_0,\dots,a_r$ embeds into $\mathbf{2}^{\alpha} \oplus \mathbf{2}^m$ for some $m<\omega$ and therefore, $\mathfrak{A}$ satisfies $\varphi=\psi$. Consequently, $\mathbf{2}^{\alpha} \oplus \mathcal{L}_2$ satisfies $\varphi=\psi$, so $\mathbf{2}^{\alpha} \oplus \mathcal{L}_2 \in \mathsf{K}$. The proofs of (b) and (c) are similar to that of (a).

Now let us consider the Cartesian power $(\omega+1)^2$ of the chain of all ordinals $\alpha<\omega+1$. Obviously $\mathcal{L}_{\omega} = ((\omega+1)^2; \leqq)$ is a distributive lattice in which $\langle\alpha,\beta\rangle \leqq \langle\gamma,\delta\rangle$ iff $\alpha \leqq \gamma$ and $\beta \leqq \delta$. For a non-empty subset M of $(\omega+1)^2$ let $m(M)$ be the set of all maximal elements of M. Denote by $\mathcal{AC}$

the set of all anti-chains of $\mathcal{L}_\omega$. It is easy to see that each member of $\mathcal{AC}$ is finite. It is well-known that the relation $\prec$ defined on $\mathcal{AC}$ as follows

$A \prec B$ iff for each $x \in A$ there exists $y \in B$ such that $x \leqq y$

is a partial order.

LEMMA 3. *$(\mathcal{AC}; \prec)$ is a distributive lattice.*

PROOF. For $A, B \in \mathcal{AC}$ let us put $A \vee B = m(A \cup B)$ and $A \wedge B = m(\{x \wedge y\colon x \in A \text{ and } y \in B\})$. It is easy to see that $A \vee B$ is the supremum and $A \wedge B$ is the infimum of the set $\{A, B\}$. Hence $(\mathcal{AC}; \prec)$ is a lattice. We shall show that for $A, B, C \in \mathcal{AC}$ the following condition holds:

$$A \wedge (B \vee C) \prec (A \wedge B) \vee (A \wedge C).$$

Let $z = x_0 \wedge y_0 \in A \wedge (B \vee C)$. Then $x_0 \in A$, $y_0 \in m(B \cup C)$ and z is a maximal element of $\{x \wedge y\colon x \in A, y \in B \vee C\}$. If $y_0 \in B$ then z is maximal in the set $M_1 = \{x \wedge y\colon x \in A, y \in B\}$. Indeed, $z = x_0 \wedge y_0 \in M_1$ and if $x_0 \wedge y_0 \leqq x \wedge y$ for $x \in A$, $y \in B$ then $x_0 \wedge y_0 \leqq x \wedge t$ for some $t \in B \vee C$ since $B \prec B \vee C$. Hence $x_0 \wedge y_0 = x \wedge t \leqq x \wedge y \leqq x \wedge t$, so $x_0 \wedge y_0 = x \wedge y$. Similarly we prove that if $y_0 \in C$ then z is a maximal element of the set $M_2 = \{x \wedge y\colon x \in A, y \in C\}$. Consequently, $z \in A \wedge B$ or $z \in A \wedge C$ so $z \in A \wedge B \cup A \wedge C$. But each element from $A \wedge B \cup A \wedge C$ is contained in some element from $m(A \wedge B \cup A \wedge C)$.

Denote by $\mathcal{AC}_{\text{fin}}$ the set of all elements $\{\langle \alpha_1, \beta_1\rangle, \ldots, \langle \alpha_n, \beta_n\rangle\} \in \mathcal{AC}$ such that ordinals $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ are finite. Obviously we have

(ix) the set $\mathcal{AC}_{\text{fin}}$ is a sublattice of $(\mathcal{AC}; \prec)$.

THEOREM 5. *For each variety $\mathbf{K} \in \mathrm{V}_0(l(\mathbf{B}))$ there exists an anti-chain $\langle \alpha_1, \beta_1\rangle, \ldots, \langle \alpha_n, \beta_n\rangle$ of $\mathcal{L}_\omega$ such that*

$$(**) \qquad \mathbf{K} = \mathbf{K}_{\alpha_1}^{\beta_1} \vee \ldots \vee \mathbf{K}_{\alpha_n}^{\beta_n}.$$

Furthermore, if ordinals $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ are finite then the presentation $(**)$ is unique.

PROOF. If $\mathbf{K}_\alpha^\beta \subseteqq \mathbf{K}$ for all $\alpha, \beta < \omega$ then by Lemma 2(c), $\mathbf{K} = \mathbf{K}_\omega^\omega$. Let us assume that there exist $\alpha_0, \beta_0 < \omega$ such that $\mathbf{K}_{\alpha_0}^{\beta_0} \nsubseteq \mathbf{K}$. Hence $\mathbf{2}^{\alpha_0} \oplus \mathbf{2}^{\beta_0} \notin \mathbf{K}$ and $\mathbf{2}^\alpha \oplus \mathbf{2}^\beta \notin \mathbf{K}$ for $\langle \alpha_0, \beta_0\rangle \leqq \langle \alpha, \beta\rangle$. We define a subset $M(\mathbf{K})$ of $(\omega+1)^2$ in the following way. If α, β are finite then $\langle \alpha, \beta\rangle \in M(\mathbf{K})$ iff $\mathbf{2}^\alpha \oplus \mathbf{2}^\beta \in \mathbf{K}$. If $\langle \alpha, \beta\rangle = \langle \omega, \beta\rangle$ and β is finite then $\langle \alpha, \beta\rangle \in M(\mathbf{K})$ iff for all $n < \omega$, $\mathbf{2}^n \oplus \mathbf{2}^\beta \in \mathbf{K}$. Symmetrically, if $\langle \alpha, \beta\rangle = \langle \alpha, \omega\rangle$ and α is finite then $\langle \alpha, \beta\rangle \in M(\mathbf{K})$ iff for all $m < \omega$, $\mathbf{2}^\alpha \oplus \mathbf{2}^m \in \mathbf{K}$. Put $\langle \alpha_1, \beta_1\rangle, \ldots, \langle \alpha_n, \beta_n\rangle$ the all maximal elements of $M(\mathbf{K})$. It follows from Lemma 2 that $\mathbf{K}_\alpha^\beta \subseteqq \mathbf{K}$ for each $\langle \alpha, \beta\rangle \in M(\mathbf{K})$. Hence $\mathbf{K}_{\alpha_1}^{\beta_1} \vee \ldots \vee \mathbf{K}_{\alpha_n}^{\beta_n} \subseteqq \mathbf{K}$.

Let $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}$. If $\mathcal{L}_1, \mathcal{L}_2$ are finite then $|L_1| = \mathbf{2}^n$ and $|L_2| = \mathbf{2}^m$ for some $n, m < \omega$. Hence $\langle n, m\rangle \in M(\mathbf{K})$ and there exists $i \in \{1, \ldots, n\}$ such that $\langle n, m\rangle \leqq \langle \alpha_i, \beta_i\rangle \in m(M(\mathbf{K}))$. Therefore $\mathcal{L}_1 \oplus \mathcal{L}_2$ is a subalgebra of $\mathbf{2}^{\alpha_i} \oplus \mathbf{2}^{\beta_i} \in \mathbf{K}_{\alpha_i}^{\beta_i}$, so $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathbf{K}_{\alpha_i}^{\beta_i}$. If $\mathcal{L}_1$ is infinite and $\mathcal{L}_2$ is finite then $|L_2| = \mathbf{2}^{m_0}$ for some $m_0 < \omega$ and by Corollary 1, for each $n < \omega$, B_n is a sublattice of $\mathcal{L}_1$ such that $1_{B_n} = 1_{L_1}$. Hence by (iv), $\mathbf{2}^n \oplus \mathbf{2}^{m_0} \in \mathsf{K}$ for all $n < \omega$. Therefore $\langle \omega, m_0 \rangle \in \in M(\mathsf{K})$ and consequently $\langle \omega, m_0 \rangle \leqq \langle \alpha_i, \beta_i \rangle$ for some $\langle \alpha_i, \beta_i \rangle \in m(M(\mathsf{K}))$. Thus $\mathcal{L}_1 \oplus \mathcal{L}_2$ is a subalgebra of $\mathcal{L}_1 \oplus \mathbf{2}^{\beta_i} \in \mathsf{K}_{\alpha_i}^{\beta_i}$ and $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathsf{K}_{\alpha_i}^{\beta_i}$. Similarly, if $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathsf{K}$ where $\mathcal{L}_1$ is finite and $\mathcal{L}_2$ is not then $\mathcal{L}_1 \oplus \mathcal{L}_2 \in \mathsf{K}_{\alpha_i}^{\beta_i}$ for some $\langle \alpha_i, \beta_i \rangle \in m(M(\mathsf{K}))$. Thus $\mathsf{K} \subseteqq \mathsf{K}_{\alpha_1}^{\beta_1} \vee \ldots \vee \mathsf{K}_{\alpha_n}^{\beta_n}$ and consequently we get $(**)$. The second part of the theorem follows immediately from the Jónsson's Lemma (see [9]) since subvarieties of $l(\mathsf{B})$ are congruence distributive.

THEOREM 6. *The lattice* $\mathsf{V}_0(l(\mathsf{B}))$ *is generated by the set* $\{\mathsf{K}_\alpha^\beta \colon \alpha, \beta < < \omega + 1\}$. *A sublattice of* $\mathsf{V}_0(l(\mathsf{B}))$ *generated by the set* $\{\mathsf{K}_\alpha^\beta \colon \alpha, \beta < \omega\}$ *is isomorphic to the lattice* $(\mathcal{AC}_{\text{fin}}; \prec)$ *so it is distributive.*

PROOF. The first part of the theorem is obvious because of Theorem 5. Let $\mathsf{V}_{\text{fin}}(l(\mathsf{B}))$ be a sublattice of $\mathsf{V}_0(l(\mathsf{B}))$ generated by varieties K_α^β for $\alpha, \beta < < \omega$. Let us consider a mapping $\varphi \colon \mathcal{AC}_{\text{fin}} \to \mathsf{V}_{\text{fin}}(l(\mathsf{B}))$ defined as follows

$$\varphi(\{\langle \alpha_1, \beta_1 \rangle, \ldots, \langle \alpha_n, \beta_n \rangle\}) = \mathsf{K}_{\alpha_1}^{\beta_1} \vee \ldots \vee \mathsf{K}_{\alpha_n}^{\beta_n}.$$

It follows from Theorem 5 that φ is one-to-one and onto. Let $A = \{\langle \alpha_1, \beta_1 \rangle, \ldots, \langle \alpha_n, \beta_n \rangle\} \in \mathcal{AC}_{\text{fin}}$, $B = \{\langle \gamma_1, \delta_1 \rangle, \ldots, \langle \gamma_m, \delta_m \rangle\} \in \mathcal{AC}_{\text{fin}}$ and $\{\langle \xi_1, \mu_1 \rangle, \ldots, \langle \xi_r, \mu_r \rangle\} = A \vee B = m(A \cup B)$. It is easy to see that $\mathsf{K}_{\xi_1}^{\mu_1} \vee \ldots \vee \mathsf{K}_{\xi_r}^{\mu_r} = = \mathsf{K}_{\alpha_1}^{\beta_1} \vee \ldots \vee \mathsf{K}_{\alpha_n}^{\beta_n} \vee \mathsf{K}_{\gamma_1}^{\delta_1} \vee \ldots \vee \mathsf{K}_{\gamma_m}^{\delta_m}$ so $\varphi(A \vee B) = \varphi(A) \vee \varphi(B)$. Hence by Corollary 3 we have $A \prec B$ iff $\varphi(A) \subseteqq \varphi(B)$. We see that $\varphi(A \wedge B) \subseteqq \varphi(A) \wedge \wedge \varphi(B)$ and since φ is onto, $\varphi(A) \wedge \varphi(B) = \varphi(C)$ for some $C \in \mathcal{AC}_{\text{fin}}$. Hence $\varphi(C) \subseteqq \varphi(A)$ and $\varphi(C) \subseteqq \varphi(B)$ so $C \prec A$ and $C \prec B$. Thus $C \prec A \wedge B$, so $\varphi(A \wedge B) = \varphi(C)$.

REFERENCES

[1] BALBES, R. and DWINGER, P., *Distributive Lattices*, University of Missouri Press, Columbia, Mo., 1974. *MR* **51** #10185

[2] BIRKHOFF, G., *Lattice Theory*, 3rd ed., Amer. Math. Soc. Colloq. Publ., vol. 25, Amer. Math. Soc., Providence, R.I., 1967. *MR* **37** #2638

[3] BUCHI, J. R., Die Boole'sche Partialordnung und die Paarung von Gefuegen, *Portugaliae Math.* **7** (1948), 119–180. *MR* **11**-575.

[4] CHROMIK, W., On some axiomatics of the product of the class of distributive lattices and the class of Boolean algebras, *Demonstratio Math.* **13** (1980), 23–32. *MR* **81m**: 08011

[5] GRÄTZER, G., *General Lattice Theory*, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Math. Reihe, Bd. 52, Akademie-Verlag, Berlin, Birkhäuser-Verlag, Basel–Stuttgart, 1978. *MR* **80c**: 06001a.

[6] GRÄTZER, G., *Universal Algebra*, 2nd ed., Springer-Verlag, New York, 1979. *MR* **80g**: 08001.

[7] HAŁKOWSKA, K., On some generalizations of Boolean algebras, *Demonstratio Math.* **10** (1977), 731–749. *MR* **58** #5437.

[8] JECH, T. J., *Lectures in Set Theory, with Particular Emphasis on the Method of Forcing*, Lecture Notes in Math. Vol. 217, Springer-Verlag, Berlin–Heidelberg–New York, 1971. *MR* **48** #105

[9] JÓNSSON, B., Algebras whose congruence lattices are distributive, *Math. Scand.* **21** (1967), 110–121. *MR* **38** #5689

[10] PŁONKA, J., On bounding congruences in some algebras having the lattice structure, *Universal algebra and applications* (Warsaw, 1978), Banach Center Publications, Vol. 9, PWN – Polish Scientific Publishers, Warsaw, 1982, 203–207. *MR* **86f**: 06013

[11] SIKORSKI, R., On dense subsets of Boolean algebras, *Colloq. Math.* **10** (1963), 189–192. *MR* **27** #4773.

[12] TRACZYK, T., Wstęp do teorii algebr Boole'a [Introduction to the theory of Boolean algebras], *Biblioteka Matematyczna*, Tom 37, Państwowe Wydawnictwo Naukowe, Warszawa, 1970, p. 77 (in Polish). *MR* **49** #171.

[13] WESOŁOWSKI, T., On some locally pseudocomplemented distributive lattices, *Demonstratio Math.* **13** (1980), 907–918. *MR* **82e**: 06012.

[14] WESOŁOWSKI, T., Subdirectly irreducible locally Boolean algebras, *Studia Sci. Math. Hungar.* **23** (1988), 257–263. *MR* **89j**: 06012

(Received January 11, 1989)

WYŻSZA SZKOŁA PEDAGOGICZNA W OPOLU
SKRYTKA POCZTOWA NR. 313
PL–45–951 OPOLE
POLAND

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 349-354

A SHORT PROOF AND ANOTHER APPLICATION OF BROOKS–CHACON'S BITTING LEMMA

DINH QUANG LUU

1. Introduction

Let E be a Banach space and $L^1_E(\mathcal{A})$ denote the Banach space of all (equivalent classes of) $\mathcal{A}$-measurable E-valued Bochner integrable functions defined on a probability space $(\Omega, \mathcal{A}, P)$. The main aim of this note is to give a short new proof and another application of the following very important Brooks–Chacon's bitting lemma.

LEMMA (Brooks–Chacon). *Let (f_n) be an L^1-bounded sequence in $L^1_R(\mathcal{A})$. Then there exist a subsequence (f_{n_k}) of (f_n), a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k) = 1$ and a function $f_* \in L^1_R(\mathcal{A})$ such that for every p, the sequence $(f_{n_k} \mid A_p)$ converges to $f_* \mid A_p$ in the $\sigma(L^1_R(\mathcal{A}), L^\infty_R(\mathcal{A}))$-topology.*

2. A short proof of Brooks–Chacon's bitting lemma

Recently, Professors E. J. Balder [2] and C. Castaing and P. Clauzure [4] have noted that the above mentioned Brooks–Chacon bitting lemma would be very effectively applied to prove different existence results such as the existence theorem of V. L. Levin ([7], Theorem 1). Unfortunately, however, they could not find any proof of the above mentioned result essentially simpler than the first one given by J. K. Brooks and R. V. Chacon [3]. Thus here as an answer to their remark we would like to apply a very simple lemma of M. Slaby [10] to give a short proof of the following more general version of the Brooks–Chacon bitting lemma.

LEMMA. *Let E be a Banach space and (f_n) an L^1-bounded sequence in $L^1_E(\mathcal{A})$. Then there exist a subsequence (n_k) of N and a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k) = 1$ such that $(f_{n_k} \mid A_k)$ is uniformly integrable.*

1980 *Mathematics Subject Classifications.* Primary 41A65.

Key words and phrases. Chacon's bitting lemma, uniform integrability, best approximation.

PROOF. Let (f_n) be as in the lemma. Then the simple proof of M. Slaby ([10], Lemma 3.4) shows that there exist a subsequence (m_q) of N and a sequence (B_q) of $\mathcal{A}$ with $\lim_{q\to\infty} P(B_q)=1$ such that the sequence $(f_{m_q} \mid B_q)$ is uniformly integrable. Thus one can choose a subsequence (B_{q_k}) of (B_q) such that

$$P(B_{q_k}) \geqq 1 - 2^{-(k+1)}, \quad k \in N.$$

Therefore, if we put $A_k = \bigcap_{s\geqq k} B_{q_s}$, $k \in N$ then (A_k) is a nondecreasing sequence of $\mathcal{A}$ with $A_k \subset B_{q_k}$, $k \in N$ and $\lim_{k\to\infty} P(A_k)=1$. Furthermore, if we define $n_k = m_{q_k}$, $k \in N$ then Slaby's result yields that the sequence $(f_{m_{q_k}} \mid B_{q_k})$ is uniformly integrable, hence so is the sequence $(f_{n_k} \mid A_k)$. It completes the proof.

Having this general bitting lemma in hand we note that one can apply the Dunford–Pettis theorem ([5], IV, 2.1) to improve Brooks–Chacon's bitting lemma and its vector-valued versions recently given by E. J. Balder ([2], Lemma) and Ch. Castaing and P. Clauzure ([4], Theorem 3.2) as follows:

THEOREM 1. *Let E be a Banach space such that E and its topological dual E^* have the (RNP). Suppose that (f_n) is an L^1-bounded sequence in $L^1_E(\mathcal{A})$ such that for every $A \in \mathcal{A}$, the sequence $\left(\int_A f_n dP\right)$ is relatively weakly (r.w.) compact in E. Then there exist a subsequence (n_k) of N, a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k)=1$ and some $f_* \in L^1_E(\mathcal{A})$ such that the sequence $(I_{A_k} f_n)$ converges to f_* in the $\sigma(L^1_E(\mathcal{A}), L^\infty_{E^*}(\mathcal{A}))$-topology, where I_B denotes the characteristic function of B.*

PROOF. Let E and (f_n) be as in the theorem. By the Dunford–Pettis theorem ([5], IV, 2.1) it is enough to show that there exist a subsequence (n_k) of N and a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k)=1$ such that

(a) the sequence $(I_{A_k} f_{n_k})$ is uniformly integrable and

(b) for every $A \in \mathcal{A}$ the sequence $\left(\int_A I_{A_k} f_{n_k} dP = \int_{A\cap A_k} f_{n_k} dP\right)$ is r.w. compact in E.

For this purpose, let (n_k) and (A_k) be the sequences taken from the general bitting lemma. Then it is clear that (a) is satisfied. Thus we have to check only the property (b). But by the hypothesis, for every $a \in \mathcal{A}$, the sequence $\left(\int_A f_{n_k} dP\right)$ is r.w. compact in E, then by (a) and the property of (A_k), obviously so is the sequence $\left(\int_{A\cap A_k} f_{n_k} dP\right)$. More precisely, if (n_{k_s}) is a subsequence of (n_k) and the sequence $\left(\int_A f_{n_{k_s}} dP\right)$ weakly converges in E then the sequence $\left(\int_{A\cap A_{k_s}} f_{n_{k_s}} dP\right)$ weakly converges also and to the same limit. It proves (b) and the theorem.

Further, let K_c denote the set of all closed convex bounded and nonempty subsets of E. A multifunction $X : \Omega \to K_c$ is said to be weakly $\mathcal{A}$-measurable if for every open subset V of E the set $\{\omega \in \Omega : X(\omega) \cap V \neq \emptyset\} \in \mathcal{A}$. In addition, if the real-valued function $\omega \mapsto |X(\omega)| = \sup\{\|x\| : x \in X(\omega)\}$ is integrable then the multifunction X is called integrably bounded and written $X \in \mathcal{L}^1_C(E)$.

Now suppose that E is a separable Banach reflexive space and (X_n) is a uniformly integrable sequence in $L^1_C(E)$, i.e. the sequence $(|X_n(\omega)|)$ of real-valued functions is uniformly integrable. Then by Lemma 3.1 of D. Q. Luu [8], there exist a subsequence (n_k) of N and an $X \in \mathcal{L}^1_C(E)$ such that for all $A \in \mathcal{A}$ and $x \in E$ we have

$$\lim_{k\to\infty} \int_A \delta^*(x, X_{n_k}) dP = \int_A \delta^*(x, X) dP,$$

where given $x \in E$ and $K \in K_c$, $\delta(x, K) = \sup\{\langle x, y\rangle, y \in K\}$.

This result together with the above bitting lemma yields the following theorem which is a stronger version of Theorem 3.1 of C. Castaing and P. Clauzure [4].

THEOREM 2. *Let E be a separable Banach reflexive space and (X_n) an L^1-bounded sequence in $\mathcal{L}^1_C(E)$, i.e.*

$$\sup_{n\in N} \int_\Omega |X_n(\omega)| dP < \infty.$$

Then there exist a subsequence (n_k) of N, a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k) = 1$ and an $X \in \mathcal{L}^1_C(E)$ such that for all $A \in \mathcal{A}$ and $x \in E$ one has

$$\lim_{k\to\infty} \int_{A_k\cap A} \delta^*(x, X_{n_k}) dP = \int_A \delta^*(x, X) dP.$$

3. Applications

The main aim of this section is to apply the above result to establish the following existence theorem for best approximations in $L^1_E(\mathcal{A})$.

THEOREM 3. *Let E be as in Theorem 2. Suppose that H is a convex subset of $L^1_E(\mathcal{A})$, closed for convergence in probability and such that for every $A \in \mathcal{A}$, the set $\left\{\int_A h dP,\ h \in H\right\}$ is r.w. compact in E. Then for every $f \in L^1_E(\mathcal{A})$ there exists a function $h_* \in H$ which minimizes the L^1-distance of f from H, i.e.*

$$E(\|f - h_*\|) = \inf\{E(\|f - h\|),\ h \in H\}.$$

PROOF. Let E and H be as in the theorem (thus H is not necessarily $\mathcal{B}$-decomposable for any sub-σ-field $\mathcal{B}$ of $\mathcal{A}$). Let $f \in L^1_E(\mathcal{A})$ be given and (f_n) a sequence in H which minimizes the L^1-distance of f from H, i.e.

$$\lim_{n\to\infty} E(\|f - f_n\|) = d(f, H) := \inf\{E(\|f - h\|), h \in H\}.$$

Thus in particular, the sequence (f_n) satisfies all the hypothesis of Theorem 2. Therefore there exist a subsequence (n_k) of N, a nondecreasing sequence (A_k) of $\mathcal{A}$ with $\lim_{k\to\infty} P(A_k) = 1$ and some $h_* \in L^1_E(\mathcal{A})$ such that the sequence $(I_{A_k} \mid f_{n_k})$ converges to h_* in the σ-$(L^1_E(\mathcal{A}))$, $L^\infty_{E^*}(\mathcal{A})$-topology, equivalently, $(I_{A_k} \mid f_{n_k})$ converges weakly to h_* since E^* has the (RNP). Thus by Mazur's lemma, for every $p \in N$, $h_* \in \overline{\mathrm{co}}(I_{A_k} f_{n_k}, k \geqq p)$, where $\overline{\mathrm{co}}(\cdot)$ denotes the closed convex hull of $(\cdot)$. Therefore, there exists an increasing subsequence (k_s) of N and a sequence (φ_s) with

$$\varphi_s \in \mathrm{co}\left(I_{A_{k_s}} f_{n_{k_s}}, I_{A_{k_s+1}} f_{n_{k_s+1}}, \dots, I_{A_{k_{s+1}}} f_{n_{k_{s+1}}}\right)$$

such that (φ_s) converges in L^1-norm, hence in probability to h_*.

Now for every $s \in N$, let $\varphi_s = \sum_{j=0}^{k_{s+1}-k_s} a_j^s I_{A_{k_s+j}} f_{n_{k_s+j}}$ for a finite sequence $\{a_j^s\}_{j=0}^{k_{s+1}-k_s}$ with $a_j^s \geqq 0$ and $\sum_{j=0}^{k_{s+1}-k_s} a_j^s = 1$. Then it is easily seen that by the property of (A_k), every function

$$g_s := \sum_{j=0}^{k_{s+1}-k_s} a_j^s f_{n_{k_s+j}} \in \mathrm{co}\left(f_{n_{k_s}}, f_{n_{k_s+1}}, \dots, f_{n_{k_{s+1}}}\right) \subset H$$

and the sequence (g_s) converges also in probability to h_*, since $\{\|g_s - \varphi_s\| \neq 0\} \subset A^c_{k_s}$ with $P(A^c_{k_s}) \to 0$ as $s \to \infty$. Therefore by the property of H it follows that $h_* \in H$. We shall show that

$$E(\|f - h_*\|) = d(f, H).$$

To see this, let us define the functional $I\colon H \to R_+$ by

$$I(h) = E(\|f - h\|), \quad h \in H.$$

Then I is convex. On the other hand, the sequence $(\|f - g_s\|)$ contains a subsequence, say, $(\|f - g_{s_m}\|)$ converging a.s. to $\|f - h_*\|$. Then by the Fatou lemma we get

$$d(f, H) \leqq I(h_*) \leqq \liminf_m I(g_{s_m}) \leqq$$
$$\leqq \limsup_n I(f_n) = d(f, H).$$

It completes the proof.

Note that for the case, when H is a closed convex $\mathcal{B}$-*decomposable* subset of $L^1_E(\mathcal{A})$, where $\mathcal{B}$ is a sub-σ-field of $\mathcal{A}$, the above theorem was earlier proved by T. Shitani and T. Ando [1], N. Herrndorf [6], M. Valadier [11] and D. Q. Luu [9].

Finally, by using a proof similar to that of the above theorem, one can establish quickly the following existence result on best approximations in $L^1_E(\mathcal{A})$ with respect to the Pettis norm.

PROPOSITION 4. *Let E and H be as in Theorem 3. Suppose more that H is norm-bounded. Then for each $f \in L^1_E(\mathcal{A})$ there is a function $h_* \in H$ which minimizes the Pettis distance of f from H.*

Acknowledgement. The author wishes to express many thanks to Professor C. Castaing for useful discussions during his visit in the Hanoi Institute of Mathematics, November 1988 and to the referee for his invaluable comments on the first version of the paper.

REFERENCES

[1] ANDO, T. and SHINTANI, T., Best approximants in L^1-space, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **33** (1975/76), 33–39. *MR* **52** #1860

[2] BALDER, E. J., Short proof of an existence result of V. L. Levin, Univ. of Utrecht, Preprint, 1987; see also *Bull. Acad. Polon. Sci.* (to appear).

[3] BROOKS, J. K. and CHACON, R. V., Continuity and compactness of measures, *Adv. in Math.* **37** (1980), 16–26. *MR* **81m**: 28003

[4] CASTAING, C. and CLAUZURE, P., Version multivoque et vectorielle d'un résultat de Brooks–Chacon, *Séminaire d'Analyse Convexe*, Vol. 18 (Montpellier, 1988), Exp. No. 4, Univ. Sci. Tech. Languedoc, Montpellier, 1988. *MR* **91c**: 46049

[5] DIESTEL, J. and UHL, J. J. JR., *Vector Measures*, Mathematical Surveys, No 15, Amer. Math. Soc., Providence, R. I., 1977. *MR* **56** #12216

[6] HERRNDORF, N., Best ϕ and N_ϕ-approximation in Orlicz spaces of vector-valued functions, *Z. Wahrsch. Verw. Gebiete* **58** (1981), 309–329. *MR* **82m**: 46026

[7] LEVIN, V. L., Extremal problems with convex functionals that are lower semicontinuous with respect to convergence in measure, *Dokl. Akad. Nauk SSSR* **224** (1975), 1256–1259 (in Russian). *MR* **52** #11703

[8] DINH QUANG LU'U, Best approximations in the space of closed convex valued multifunctions, *Travaux Sém. Anal. Convexe* **12** (1982), Exp. 19, Montpellier. *MR* **84k**: 41028

[9] DINH QUANG LU'U, Best approximations in the space of Bochner integrable functions, *Math. Nachr.* **121** (1985), 287–293. *MR* **87b**: 41028

[10] SLABY, M., Strong convergence of vector-valued pramarts and subpramarts, *Probab. Math. Statist.* **5** (1985), 187–196. *MR* **87c**: 60042

[11] VALADIER, M., La multiaplication médianes conditionnelles, *Travaux Sém. Anal. Convexe* **12** (1982), Exp. 21. *MR* **85b**: 28011. See also, *Z. Wahrsch. Verw. Gebiete* **67** (1984), 279–282. *MR* **85k**: 41042

(Received January 16, 1989)

INSTITUTE OF MATHEMATICS
P.O.BOX 631, BO HO
HANOI
VIETNAM

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 355–360

ON THE GAME OF MISERY

J. PACH

In the spring of 1986 an "epidemic" raised its head at the Mathematical Sciences Research Institute, Berkeley, California. It was a very simple game that "infected" most visitors of the institute and temporarily diverted them from their researches.

The game, which will be referred to as the *Game of Misery*, can be played with finitely many piles of disks sitting at some integer points (positions) of a line. At each unit of time each pile independently is divided in half, with half the disks being moved one position to the left and the other half one position to the right. If there is an odd number of disks in a pile, then one disk stays at the same place and the remainder are divided evenly.

To make our description more "formal", let $a_{i,t}$ denote the number of disks sitting at point i at time t (i and t are integers, $-\infty < i < +\infty$, $t \geqq 0$). Then

$$a_{i,t+1} = \left\lfloor \frac{1}{2} a_{i-1,t} \right\rfloor + \left\lfloor \frac{1}{2} a_{i+1,t} \right\rfloor + \text{parity}(a_{i,t}),$$

where $\text{parity}(a_{i,t}) = 0$ or 1 according to whether $a_{i,t}$ is even or odd, respectively. For instance, if we start out with a single pile of 6 disks at position 0, then after 7 moves we reach a final configuration where each pile is either empty or contains exactly one disk. Moreover, the nonempty piles form two intervals of length 3 separated by one empty pile at position 0:

				6			
1st move			3	0	3		
2nd move		1	1	2	1	1	
3rd move		1	2	0	2	1	
4th move		2	0	2	0	2	
5th move	1	0	2	0	2	0	1
6th move	1	1	0	2	0	1	1
7th move	1	1	1	0	1	1	1

(Note that it is not even clear *a priori* that the Game of Misery will always end in finitely many steps! However, it is easily seen to be true.)

Some of the "victims of the Misery-epidemic" have recently published a paper in the American Mathematical Monthly [1] analyzing the game. They are mainly concerned with the case when the starting configuration is a single pile of $2n$ disks. According to their first result, in this case the final configuration consists of two intervals of n consecutive piles with one disk each, separated by an empty pile at position 0. This easily implies that, if we start out with a single pile of size $2n+1$, then at the final stage we get an interval of $2n+1$ piles of size 1, symmetric about 0.

In this note we will consider the general case, i.e., when we can start the game with any configuration. The following assertion (conjectured by a finite but nonempty subset of the authors [1]) is easily seen to be a generalization of the above facts.

THEOREM 1. *Suppose that in the initial configuration in the Game of Misery there are at most k nonempty piles. Then the final configuration consists of at most $2k$ intervals of consecutive piles of size* 1.

Moreover, equality holds here if and only if all piles in the initial configuration are even (i.e., $a_{i_j,0} > 0$ is even for $i_1 < i_2 < \ldots < i_k$) and

$$i_{j+1} - i_j \geqq \frac{a_{i_{j+1},0} + a_{i_j,0}}{2} + 2 \quad \textit{for} \quad 1 \leqq j < k.$$

PROOF. In a Game of Misery, let $\mathcal{S}$ and $\mathcal{F}$ denote the starting and the final configurations, respectively. Note that in $\mathcal{F}$ at each position there is at most one disk, otherwise we could continue the game.

A *maximal* interval I of consecutive positions is said to be a *block* of a configuration, if each position of I is occupied by at least one disk.

Assume that $\mathcal{F}$ has p blocks, i.e., it is the union of p intervals $I_1 \cup I_2 \cup \cup \ldots \cup I_p$ of consecutive piles of size 1. We are going to prove by induction that $\mathcal{S}$ contains at least $\lceil p/2 \rceil$ nonempty piles.

This clearly holds if $p \leqq 2$.

Let $p > 2$, and assume that the claim has already been proved for $1, 2, \ldots, p-1$. Starting with $\mathcal{F}$, let us play the following, so-called *Inverse Game*. In this new game a move is to pick some integer i such that neither of the piles at positions $i-1$ and $i+1$ is empty, and transfer one-one disk from each to position i. It is straightforward to see that, for a suitable sequence of the integers i, this game will terminate with the configuration $\mathcal{S}$. (We have to break each move of the original game into smaller steps and play the game "backwards".)

Fix any block I_j of $\mathcal{F}$, different from the leftmost and the rightmost ones $(1 < j < p)$. Consider the last configuration $\mathcal{C}$ occurring in the Inverse Game, where I_j still has an *intact* position, i.e., a position that has been occupied by exactly one disk all the time. Let I denote the block of $\mathcal{C}$ containing this position, and let $i_{\min}$ and $i_{\max}$ denote the smallest and the largest elements of I, respectively.

Observe that, at the time we reach configuration $\mathcal{C}$, *every* position of I is still intact, because we could not put a disk next to an intact position.

We distinguish two cases.

Case A: $\mathcal{C}$ is not the last configuration in the Inverse Game, i.e., $\mathcal{C} \neq \mathcal{S}$.

Then after the next move the last intact position of I disappears. Since we touch only 2 disks, $|I| \leqq 3$ must hold. In fact, one of the following two essentially different possibilities can occur (up to symmetry).

$$\begin{array}{rl} \mathcal{C}: & \ldots 0 \overbrace{\ 1\ }^{I} 0\ m \ldots \\ \text{next move:} & \ldots 0\ \ 0\ \ 2\ m-1 \ldots \\ & \text{Case } A_1 \end{array} \qquad \begin{array}{l} \ldots 0 \overbrace{1\,1\,1}^{I} 0 \ldots \\ \ldots 0\ 0\ 3\ 0\ 0 \ldots \\ \text{Case } A_2 \end{array}$$

During the whole Inverse Game no disk can jump from one side of $i_{\min} - \frac{1}{2}$ to the other. Obviously, this cannot happen before $i_{\min}$ seizes to be intact, and after that both positions $i_{\min} - 1$ and $i_{\min}$ are empty and remain empty, because we cannot put a disk on one of them unless the other one is not empty any more. Hence, $i_{\min} - \frac{1}{2}$ is a *cutpoint* of the game, i.e., the events occurring on one side of it are independent of whatever happens on the other side. Consequently, we can apply the induction hypothesis to the at least $j-1$ intervals of 1's on the left side, and to the at least $p-j+1$ intervals of 1's on the right side of $i_{\min} - \frac{1}{2}$ in $\mathcal{F}$, separately. We obtain that $\mathcal{S}$ has at least $\lceil (j-1)/2) \rceil + \lceil (p-j+1)/2 \rceil \geqq \lceil p/2 \rceil$ nonempty piles.

Case B: $\mathcal{C}$ is the last configuration in the Inverse Game, i.e., $\mathcal{C} = \mathcal{S}$.

Then $i_{\min} - \frac{1}{2}$ is again a cutpoint, for $i_{\min}$ remains intact during the whole game. Thus we can complete the argument in exactly the same way as in Case A.

This proves the claim, and hence the first part of Theorem 1.

It is clear that equality holds in the first statement of the theorem if and only if there is a sequence of empty positions $c_1 < c_2 < \ldots < c_{k-1}$ in $\mathcal{F}$ such that

(i) each c_j is a cutpoint;

(ii) $\mathcal{F}$ has exactly 2 blocks between c_j and c_{j+1} $(1 \leqq j < k-1)$, and two-two blocks on the left side of c_1 and on the right side of c_{k-1}, respectively;

(iii) each of these pairs of blocks becomes a single nonempty pile by the end of the Inverse Game.

Set $c_0 = -\infty$, $c_k = +\infty$, and for any j $(0 \leqq j < k)$ let I_{2j+1} and I_{2j+2} denote the blocks of $\mathcal{F}$ between c_j and c_{j+1}. I_{2j+1} and I_{2j+2} are separated by exactly one empty position i_{j+1}, otherwise we could find a cutpoint between these two blocks, contradicting condition (iii). On the other hand, if we start the Game of Misery with a single pile at position i, then the final configuration must be symmetric about i. Hence, (iii) implies that I_{2j+1} and I_{2j+2} are placed symmetrically about i_{j+1}. In particular, $|I_{2j+1}| = |I_{2j+2}|$ and $a_{i_{j+1},0} = |I_{2j+1}| + |I_{2j+2}|$ is even for every j. From this the second part of Theorem 1 follows immediately. □

Essentially the same argument allows us to establish the following stronger assertion.

THEOREM 2. *Suppose that the initial configuration in the Game of Misery is the union of at most k blocks of consecutive nonempty piles. Then the final configuration consists of at most $2k$ blocks of consecutive piles of size* 1. □

The interested reader is invited to analyze the cases of equality in this result.

In the sequel we shall investigate how long it takes to reach the final configuration in the Game of Misery. To this end, we first define another related game, the so-called *Free Game.* In this new game we also have finitely many piles of disks sitting at some integer positions of a line. An *elementary move* is to pick a pile with at least 2 disks and move one disk with one position to left, and the other with one position to the right. A *move* in the Free Game is the superposition of any nonempty set of nonconflicting elementary moves. The Free Game ends, if there are no more moveable disks left, i.e., each pile contains at most one disk.

It is clear that the Game of Misery is the special case of the Free Game, when we move all possible disks on every step. In this sense, the Game of Misery is the greediest way to play the Free Game. One of the most interesting observations made in [1] is that no matter how we play the Free Game, the final configuration is always the same. That is, the final configuration depends on the initial configuration only.

Given a Free Game F starting with a configuration $\mathcal{S}$, let $t_F(\mathcal{S})$ and $t(\mathcal{S})$ denote the number of moves it takes to reach the final configuration in F and in the corresponding Game of Misery, respectively.

We will make use of the following two simple facts.

LEMMA 1. $t(\mathcal{S}) \leqq t_F(\mathcal{S})$.

PROOF. Let k be the smallest number such that the k-th move in F is not identical with the k-th move in the Game of Misery. This means that at this step, in F we fail to move at least one pair of moveable disks (d_1, d_2) sitting at some position i. Since our disks are indistinguishable, and in the final configuration there is at most one disk at each position, we can assume that later on in F (say, at the k'-th step) we move d_1 and d_2 to the positions $i-1$ and $i+1$, respectively.

Let us modify F so that we move the pair (d_1, d_2) at the k-th step already, and then we do not touch them until after the k'-th move has been made. The movement of the rest of the disks remains unchanged. Thus we obtain another Free Game F_1 whose each move is legal, except that the k'-th move may become void. In any case, $t_{F_1}(\mathcal{S}) \leqq t_F(\mathcal{S})$. Meanwhile, we have reduced the sum $\sum je_j$ by at least 1, where e_j denotes the number of elementary moves comprising the j-th move of F. Continuing this process, we obtain a finite sequence of Free Games $F = F_0, F_1, F_2, \dots, F_m$ such that F_m is the Game of Misery, and $t_{F_{i+1}}(\mathcal{S}) \leqq t_{F_i}(\mathcal{S})$ for every $0 \leqq i < m$. □

LEMMA 2. *Suppose that the starting configuration $\mathcal{C}$ in the Game of*

Misery consists of $n+1$ *disks arranged in* n *consecutive nonempty piles, say at positions* 1 *through* n. *That is, for some* $1 \leqq k \leqq n$,

$$a_{i,0} = \begin{cases} 2 & \text{if } i = k, \\ 1 & \text{if } 1 \leqq i \leqq n,\ i \neq k, \\ 0 & \text{otherwise.} \end{cases}$$

Then $t(\mathcal{C}) = n$, *and in the final configuration all but one positions of the interval* $[0, n+1]$ *are occupied by a disk. More precisely,*

$$a_{i,n} = \begin{cases} 1 & \text{if } 0 \leqq i \leqq n+1,\ i \neq n+1-k, \\ 0 & \text{otherwise.} \quad \square \end{cases}$$

We leave the simple proof of this fact to the reader, in the hope that the following examples for $n = 8$ are sufficiently instructive.

21111111	12111111	11211111	11121111
102111111	20211111	12021111	11202111
110211111	102021111	20202111	12020211
111021111	110202111	102020211	20202021
111102111	111020211	110202021	102020202
111110211	111102021	111020202	1102020201
111111021	111110202	1112020201	1110202011
111111102	1111110201	1111102011	1111020111
1111111101	1111111011	1111110111	1111101111
11112111	11111211	11111121	11111112
11120211	11112021	11111202	111111201
11202021	11120202	111120201	111112011
12020202	112020201	111202011	111120111
202020201	120202011	112020111	111201111
1020202011	202020111	120201111	112011111
1102020111	1020201111	202011111	120111111
1110201111	1102011111	1020111111	201111111
1111011111	1110111111	1101111111	1011111111

THEOREM 3. *Let* $\mathcal{S}$ *be any configuration of* m *disks, and suppose that* $\mathcal{S}'$ *can be obtained from* $\mathcal{S}$ *by adding one more disk. Then* $t(\mathcal{S}') \leqq t(\mathcal{S}) + m$.

PROOF. Define a Free Game F whose starting configuration is $\mathcal{S}'$, as follows. The first $t(\mathcal{S})$ moves of F are identical to the moves of the Game of Misery played with $\mathcal{S}$, except that the extra disk of $\mathcal{S}'$ remains still. Then, either F ends, i.e., in the resulting configuration $\mathcal{C}'$ there is at most one disk at each position, or there is one pile of size 2. In the latter case, we continue F as a Game of Misery starting with $\mathcal{C}'$. According to Lemma 2, it takes at most $n \leqq m$ further steps to reach the final configuration, where n denotes the length of the block $\mathcal{C}$ in $\mathcal{C}'$ containing the pile of size 2. Hence, by Lemma 1, we get $t(\mathcal{S}') \leqq t_F(\mathcal{S}') \leqq t(\mathcal{S}) + n \leqq t(\mathcal{S}) + m$. $\square$

Strangely enough, it is perfectly conceivable that $t(\mathcal{S}')$ is substantially *smaller* than $t(\mathcal{S})$. Therefore, we are unable to decide whether the function $t(\mathcal{S})$ is continuous in the following sense. There exists an absolute constant c such that $|t(\mathcal{S}_1) - t(\mathcal{S}_2)| \leqq cd(\mathcal{S}_1, \mathcal{S}_2)(|\mathcal{S}_1| + |\mathcal{S}_2|)$ for any configurations $\mathcal{S}_1$ and $\mathcal{S}_2$, where $d(\mathcal{S}_1, \mathcal{S}_2)$ denotes the smallest number of disks that must be changed (added or deleted) to obtain $\mathcal{S}_2$ from $\mathcal{S}_1$.

In [1] it is shown that if the starting configuration is a single pile of n disks, then the number moves needed to finish the Game of Misery is $cn^2 + o(n^2)$ for some positive constant c. Our next result, which is an immediate consequence of Theorem 3, shows that, for any configuration $\mathcal{S}$ of n disks, $t(\mathcal{S})$ is at most quadratic in n.

COROLLARY. *Given any initial configuration of n disks in the Game of Misery, the final configuration is reached within $\binom{n}{2}$ moves.* □

REFERENCE

[1] ANDERSON, R. J., LOVÁSZ, L., SHOR, P., SPENCER, J., TARDOS, É. and WINOGRAD, S., Disks, balls and walls: Analysis of a combinatorial game, *Amer. Math Monthly* **96** (1989), 481–493. *MR* **90e**: 90169

(Received February 22, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 361–368

ON A NEW ITERATION FOR FINDING "ALMOST ALL" SOLUTIONS OF THE QUADRATIC EQUATION IN BANACH SPACE

I. K. ARGYROS

Abstract

We introduce a new iteration for solving quadratic equations in Banach space. Under certain assumptions the iterates are uniformly bounded below. In case of convergence we can obtain "almost" all solutions.

In this paper we introduce a new iteration

$$
\begin{aligned}
x_{n+1} &= B(x_n)^{-1}(L'(x_n) - y), \qquad n = 0, 1, 2, \dots \\
&= x_n - B(x_n)^{-1}\left(B(x_n, x_n) + y - L'(x_n)\right)
\end{aligned} \tag{1}
$$

for some $x_0 \in X$ in a Banach space X to prove existence and uniqueness of not necessarily "small" solutions of the quadratic equation

$$
x = y + L(x) + B(x, x), \qquad L' = I - L \tag{2}
$$

in a closed ball centered at a specific $z \in X$, where $y \in X$ is fixed, L is a bounded linear operator and B is a bounded symmetric bilinear operator on X. Equation (2) has been also studied under different assumptions in [1], [2], [3], [7], [9], [10].

Iteration (1) has the interesting property (under certain assumptions) that if $\|x_0\| \geqq d$ then $\|x_n\| \geqq d$ for all $n = 0, 1, 2, \dots$ and $d > 0$. In case of convergence (1) provides us with a solution x such that $\|x\| \geqq d$.

We also study the modified version of (1)

$$
x_{n+1} = x_n - B(z)^{-1}\left(B(x_n, x_n) + y - L'(x_n)\right), \quad n = 0, 1, 2, \dots. \tag{3}
$$

Finally, we provide two simple examples for (1) and (3).

DEFINITION 1. Denote by $L(X, X)$ the linear space over the field of real or complex numbers of all linear operators from X into X, then a linear operator B from X into $L(X, X)$ is called a *bilinear operator from X into X*.

The motivation for this definition is the observation that for any $x_1 \in X$, $L = B(x_1)$ is a linear operator from X into X, so that

$$
y = (B(x_1))(x_2) = B(x_1, x_2)
$$

1980 *Mathematics Subject Classifications.* Primary 46B15.
Key words and phrases. Bilinear-quadratic operator, Banach spaces.

Akadémiai Kiadó, Budapest

is an element of X for $x_2 \in X$.

DEFINITION 2. A linear operator L from X into X is said to be *bounded* if

$$\|L\| = \sup_{\|x\|=1} \|L(x)\| \tag{4}$$

is finite. The quantity $\|L\|$ is called the *norm* of L.

DEFINITION 3. A bilinear operator from X into X is said to be *bounded* if it is a bounded linear operator from X into $L(X,X)$. The *norm* $\|B\|$ of B is defined by (4), with B being considered to be an element of $L(X, L(X,X))$.

The inequality

$$\|B(x,y)\| \leqq \|B\| \cdot \|x\| \cdot \|y\|, \qquad x \in X, \; y \in X$$

is obvious from the above definition.

DEFINITION 4. A bilinear operator B from X into X is called *symmetric* if $B(x,y) = B(y,x)$ for all $x, y \in X$.

REMARK 1. The operator B in (2) is assumed to be symmetric without loss of generality since B can always be replaced by the *mean* $\overline{B}$ of B defined by

$$\overline{B}(x,y) = 1/2(B(x,y) + B(y,x)) \quad \text{for all} \quad x \in X, \; y \in X.$$

Note that $\overline{B}(x,x) = B(x,x)$ for all $x \in X$.

LEMMA 1. *Let L_1 and L_2 be bounded linear operators on X, where L_1 is invertible and $\|L_2\| \cdot \|L_1^{-1}\| < 1$. Then the operator $(L_1 + L_2)^{-1}$ is also invertible, and*

$$\|(L_1 + L_2)^{-1}\| \leqq \frac{\|L_1^{-1}\|}{1 - \|L_2\| \cdot \|L_1^{-1}\|}.$$

PROOF. The operator $(L_1 + L_2)^{-1}$ is invertible if the operator $I + L_1^{-1} L_2$ is invertible, since

$$(L_1 + L_2)^{-1} = (I + L_1^{-1} L_2)^{-1} L_1^{-1},$$

but

$$\|L_1^{-1} L_2\| \leqq \|L_1^{-1}\| \cdot \|L_2\| < 1,$$

so $I + L_1^{-1} L_2$ is invertible and

$$\|(L_1 + L_2)^{-1}\| = \left\|(I + L_1^{-1} L_2)^{-1} L_1^{-1}\right\| \leqq$$

$$\leqq \left\|(I + L_1^{-1} L_2)^{-1}\right\| \cdot \|L_1^{-1}\| \leqq \frac{\|L_1^{-1}\|}{1 - \|L_2\| \cdot \|L_1^{-1}\|}.$$

LEMMA 2. *Let $z \neq 0$ be fixed in X. Assume that the operator $B(z)$ is invertible, then $B(x)$ is also invertible for all*

$$x \in U(z,r) = \{x \in x / \|x - z\| < r\}$$

where r is such that $0 < r < [\|B\| \cdot \|B(z)^{-1}\|]^{-1}$.

PROOF. $B(x) = B(x-z) + B(z)$.
Now, it is enough to show

$$\|B(x-z)\| \cdot \|B(z)^{-1}\| \leqq \|B\| \|B(z)^{-1}\| \cdot \|x - z\| \leqq \|B\| \cdot \|B(z)^{-1}\| r < 1$$

true by hypothesis.

PROPOSITION 1. *Equation (2) has at most one solution $x \in X$ such that*

$$\|x\| < \frac{1 - \|L\|}{2\|B\|},$$

provided that $\|L\| < 1$.

PROOF. Let x_1, x_2 be distinct solutions of (2) with

$$\|x_1\| < \frac{1 - \|L\|}{2\|b\|}, \quad \|x_2\| < \frac{1 - \|L\|}{2\|B\|}.$$

By (2)

$$\|x_1 - x_2\| = \|(L + B(x_1 + x_2))(x_1 - x_2)\| \leqq$$
$$\leqq [\|L\| + \|B\|(\|x_1\| + \|x_2\|)]\|x_1 - x_2\| < \|x_1 - x_2\|$$

which is a contradiction.

PROPOSITION 2. *If (1) is well defined for all $n = 0, 1, 2, \ldots$ and $\|B(x_n)^{-1}\| \leqq c$ for some $c > 0$ such that $c\|L'\| < 1$. Then in case of convergence, the solution $x \in X$ satisfies the estimate*

$$\|x\| \leqq \frac{c\|y\|}{1 - \|L'\| \cdot c}.$$

PROOF. We have by (1)

$$\|x_{n+1}\| \leqq \|B(x_n)^{-1}\|(\|L'\| \cdot \|x_n\| + \|y\|)$$
$$\cdots\cdots\cdots\cdots$$
$$\leqq c^{n+1}\|L'\|^{n+1}\|x_0\| + c\|y\| \frac{1 - (c\|L'\|)^{n+1}}{1 - c\|L'\|}$$

since $c\|L'\| < 1$, by taking the limits in both sides of the above inequality as $n \to \infty$, we get

$$\|x\| \leqq \frac{c\|y\|}{1 - c\|L'\|}.$$

DEFINITION 5. Define the set M_1, M_2, by

$$M_1 = \{L \in L(X) / \|L\| \cdot \|x_n\| - \|y\| \leqq \| \|L(x_n)\| - \|y\| \|\}$$
$$M_2 = \{L \in L(X) / \|y\| - \|L\| \cdot \|x_n\| \leqq \| \|L(x_n)\| - \|y\| \|\}$$

where the x_n's, $n = 0, 1, 2, \ldots$ are given by (1) for some $x_0 \in X$. Note that $M_1, M_2 \neq 0$, since $I \in M_1$, $I \in M_2$.

PROPOSITION 3. *Let $L' \in M_1$ and $\|x_0\| \geqq d$ in (1) then $\|x_n\| \geqq d$, $n = 0, 1, 2, \ldots$ for some $d \in [d_1, d_2]$ where*

$$d_1 = \frac{\|L'\| - \sqrt{\|L'\|^2 - 4\|B\| \cdot \|y\|}}{2\|B\|},$$
$$d_2 = \frac{\|L'\| + \sqrt{\|L'\|^2 - 4\|B\| \cdot \|y\|}}{2\|B\|}$$

provided that $\|L'\|^2 - 4\|B\| \cdot \|y\| \geqq 0$.

PROOF. By (1)

$$\|B\| \cdot \|x_n\| \cdot \|x_{n+1}\| \geqq \|L'(x_n) - y\|$$

or

$$\|x_{n+1}\| \geqq \frac{\|L'(x_n) - y\|}{\|B\| \cdot \|x_n\|}, \quad \text{if} \quad \|x_n\| \geqq d, \quad k = 0, 1, 2, \ldots, n$$

then $\|x_{n+1}\| \geqq d$ if $\frac{\| \|L'(x_n)\| - \|y\| \|}{\|B\| \cdot \|x_n\|} \geqq d$ and since $L' \in M_1$ it is enough to show $\frac{\|L'\| \cdot \|x_n\| - \|y\|}{\|B\| \cdot \|x_n\|} \geqq d$ or $\|x_n\| \geqq \frac{\|y\|}{\|L'\| - d\|B\|}$ or $d \geqq \frac{\|y\|}{\|L'\| - d\|B\|}$ which is true for $d \in$ $\in [d_1, d_2]$.

The proof of the following proposition is omitted as similar to Proposition 3.

PROPOSITION 4. *Assume $L' \in M_2$ and the following are true:*
(i) *the hypotheses of Proposition 2,*
(ii) *there exists $d \geqq 0$ such that*

$$d \leqq \frac{c\|y\|}{1 - c\|L\|}, \quad d \leqq \frac{\|y\|}{d\|B\| + \|L'\|}$$

and

$$\frac{c\|y\|}{1 - c\|L'\|} \leqq \frac{\|y\|}{d\|B\| + \|L'\|}$$

then if $\|x_0\| \geqq d$ *in* (1) *then* $\|x_n\| \geqq d$. *For all* $n = 0, 1, 2, \ldots$.

DEFINITION 6. Let $z \neq 0$ be fixed in X. Assume that the linear operator $B(z)$ is invertible then by Lemma 2

$$\|B(x)^{-1}\| \leqq \frac{\|B(x)^{-1}\|}{1 - \|B(z)^{-1}\| \cdot \|B\| \cdot r}$$

where $x \in U(z, r) = \left\{x \in X \mid \|x - z\| < r < \left(\|B\| \cdot \|B(z)^{-1}\|\right)^{-1}\right\}$.

Define the operators P, T on $\overline{U}(z, r_0)$ by

$$P(x) = B(x, x) + y - L'(x), \quad T(x) = (B(x))^{-1}(L'(x) - y)$$

and the real polynomials $f(r)$, $g(r)$ on $\mathbf{R}$ by $f(r) = a'r^2 + b'r + c^1$, $g(r) = ar^2 + br + c$, where

$$\begin{aligned}
a' &= [\|B\|\|B(z)^{-1}\|]^2, \\
b' &= -2\|B\|\|B(z)^{-1}\|, \\
c' &= 1 - \|B(z)^{-1}\|\|L'\| - \|B\|\|B(z)^{-1}\|^2\|L'(z) - y\|, \\
a &= \|B\| \cdot \|B(z)^{-1}\|, \\
b &= \|B(z)^{-1}(L' - B(z)\| - 1, \\
c &= \|B(z)^{-1}P(z)\|.
\end{aligned}$$

Finally, note that for any operator T on $\overline{U}(z, r_0)$, $r_0 < r$, $\|T(w) - T(v)\| \leqq$ $\leqq q\|w - v\|$ for all $w, v \in \overline{U}(z, r)$, $r_0 < r$ where

$$q = \sup_{x \in \overline{U}(z, r_0)} \|T'(x)\|$$

and $T'(x)$ is the first Fréchet-derivative of T at $x \in X$.

THEOREM 1. *Let* $z \neq 0$ *be fixed in* X *such that* $(B(z))^{-1}$ *exists. Assume:*
(i) $b^2 - 4ac > 0$, $b < 0$;
(ii) $c' > 0$; *and*
(iii) *there exists* $r_0 > 0$ *such that* $f(r_0) > 0$ *and* $g(r_0) \leqq 0$.
Then the iteration $x_{n+1} = B(x_n)^{-1}(L'(x_n) - y)$, $n = 0, 1, 2, \ldots$ *is well defined and it converges to a unique solution* x *of the equation*

$$x = y + L(x) + B(x, x) \quad \textit{in} \quad \overline{U}(z, r_0)$$

for any $x_0 \in \overline{U}(z, r_0)$.

PROOF. *Claim 1.* T *is a well defined contraction on* $\overline{U}(z, r_0)$.

If $w, v \in \overline{U}(z, r_0)$ then

$$\|T(w) - T(v)\| \leqq q\|w - v\| \quad \text{for all } w, v \in \overline{U}(z, r_0)$$

where $q = \sup_{x \in \overline{U}(z, r_0)} \|T'(x)\|$. T is a contraction operator if $0 < q < 1$, but

$$\begin{aligned} \|T'(x)\| = \|B(x)^{-1}(-B(L'(x)), B(x)^{-1}\cdot) + \\ + B(x)^{-1}(B(y), B(x)^{-1}\cdot) + B(x)^{-1}L'\| \leqq \end{aligned}$$

$$\leqq \frac{\|B(z)^{-1}\|}{1 - \|B\| \cdot \|B(z)^{-1}\| r}\left[\|L'\| + \frac{\|B\| \cdot \|B(z)^{-1}\|}{1 - \|B\| \cdot \|B(z)^{-1}\| \cdot r}(\|L\|r + \|L'(z) - y\|)\right] < 1,$$

if $f(r_0) > 0$, which is true by hypothesis.

Claim 2. T maps $\overline{U}(z, r_0)$ into $\overline{U}(z, r_0)$.

If $w \in \overline{U}(z, r_0)$, then

$$T(w) - z = B(w)^{-1}[(L' - B(z))(x - z) - P(z)]$$

so

$$\|T(w) - z\| \leqq r_0$$

if

$$\frac{1}{1 - \|B\| \cdot \|B(z)^{-1}\| \cdot r_0}\left[\|B(z)^{-1}(L' - B(z))\|r_0 + \|B(z)^{-1}P(z)\|\right] \leqq r_0$$

or $g(r_0) \leqq 0$ which is true by hypothesis. The result follows from the contraction mapping principle.

We now state a theorem for (3) whose proof as similar to Theorem 1 is omitted. For simplicity we take $L = 0$ in (3).

THEOREM 2. *Let $z \neq 0$ be fixed in X. Assume:*

(i) *the linear operator $B(z)$ is invertible,*

(ii) $\|B(z)^{-1}(1 - B(z))\| < 1$ *and there exists r_0^0 such that $r_0^0 \in [r_1, r_2)$ where*

$$r_1 = \frac{1 - \|B(z)^{-1}(1 - B(z))\|}{2\|B\| \cdot \|B(z)^{-1}\|} -$$

$$- \frac{\left[(1 - \|B(z)^{-1}(I - B(z))\|)^2 - 4\|B\| \cdot \|B(z)^{-1}P(z)\|\|B(z)^{-1}\|\right]^{1/2}}{2\|B\| \cdot \|B(z)^{-1}\|}$$

$$r_2 = \frac{1 - \|B(z)^{-1}(I - B(z))\|}{2\|B\| \cdot \|B(z)^{-1}\|},$$

then (3) converges to a unique solution x of (2) (with $L = 0$) in $\overline{U}(z, r_0^0)$.

We now provide two simple examples for Theorems 1 and 2, respectively (with $L = 0$).

EXAMPLE 1. Let $X = 1R$ and consider the equation

$$x = .2x^2 - 1 \tag{5}$$

here $B(x,x) = .2x^2$, $y = -1$ and $1 - 4|b| \cdot |y| > 0$. Choose $z = 5$, then according to Definition 6 and Part 1 of Theorem 1 $c' = 1 - \|B\|^2 \|B(z)^{-1}\| \|y\|$ (sharper estimate if $L = 0$).

$$f(r) = .04r^2 - .4r + .96 \text{ with solutions } r_1' = 8.1622775, \quad r_2' = 1.8377225,$$
$$g(r) = .2r^2 - r + 1 = 0 \text{ with solutions } r_1, r_2 \text{ such that}$$
$$r_2 = 3.618033989 \text{ and } r_1 = 1.38196601.$$

Therefore, Theorem 1 can be applied if $r_2 \leqq r_0 < r_2'$ and then iteration (1) becomes

$$x_{n+1} = 5\left(1 + \frac{1}{x_n}\right) \qquad n = 0, 1, 2, \ldots$$

with $x_0 = z = 5$ we need 12 iterations to obtain the "large" solution of (5) which is $x = x_{12} = 5.8541010966$.

We know that x_{12} is the "large" solution of (5) since $\|x_n\| \geqq 5$, $n = 0, 1, 2, \ldots$ and by Proposition 1

$$\|x_n\| > \frac{1}{2|b|} = \frac{5}{2}, \quad n = 0, 1, 2, \ldots.$$

EXAMPLE 2. Let $x = R \times R$ equipped with the usual max-norm and consider the equation

$$\underline{x} = \underline{y} + \underline{x}^{tr} \underline{\underline{M}} \underline{x} \tag{6}$$

where, $\underline{y} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$, $y_1 = 1.55$, $y_2 = -.85$.

$$\underline{\underline{M}} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \; M_1 = \begin{bmatrix} -.45 & .01 \\ .9 & .02 \end{bmatrix}, \; M_2 = \begin{bmatrix} .01 & -.7 \\ .02 & .5 \end{bmatrix}$$

with $\underline{x}^{tr} \underline{\underline{M}} \underline{x} = \left[\begin{smallmatrix} \underline{x}^{tr} M_1 \underline{x} \\ \underline{x}^{tr} M_2 \underline{x} \end{smallmatrix}\right]$ then (6) can also be written as

$$\begin{aligned} x_1 &= -.45x_1^2 + .91x_1x_2 + .02x_2^2 + 1.55 \\ x_2 &= .01x_1^2 - .68x_1x_2 + .5x_2^2 - .85. \end{aligned}$$

Choose $z = \left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right]$, then

$$\begin{aligned} \|B\| &= 1.38 \\ \|B(z) - I\| &= .9 \\ \|B(z)^{-1}\| &= .555555 \\ \|P(z)\| &= .05. \end{aligned}$$

Theorem 2 can be applied in the ball $\overline{U}(z, r_0)$ for $r_0 \in [r_1, r_2)$ where $r_1 = = .061321367$ and $r_2 = .326086956$.

Choose $\underline{x}^0 = z$ and allow an error ε such that $\varepsilon \leqq 5.10^{-3}$ then we need five iterations

$$\underline{x}^{(1)} = \begin{bmatrix} -1.97222223 \\ .97368421 \end{bmatrix},$$

$$\underline{x}^{(2)} = \begin{bmatrix} -1.996301957 \\ .97283584 \end{bmatrix},$$

$$\underline{x}^{(3)} = \begin{bmatrix} -1.99715663 \\ .968165641 \end{bmatrix},$$

$$\underline{x}^{(4)} = \begin{bmatrix} -2.003524174 \\ .9654191 \end{bmatrix}, \quad \text{and}$$

$$\underline{x}^{(5)} = \begin{bmatrix} -2.00038145 \\ .96224933 \end{bmatrix}.$$

REFERENCES

[1] ANSELONE, P. M. and MOORE, R. H., An extension of the Newton–Kantorovič method for solving nonlinear equations, Tech. Rep. No. 520, U. S. Army Mathematics Research Center, University of Wisconsin, Madison, 1965. See *MR* **32** #8089

[2] ARGYROS, I. K., On a contraction theorem and applications, *Nonlinear functional analysis and its applications*, Part 1 (Berkeley, Calif., 1983), Proc. Sympos. Pure Math., Vol. 45, Part 1, Amer. Math. Soc., Providence, R. I., 1986, 51–53. *MR* **87h**: 65108

[3] ARGYROS, I. K., Quadratic equations in Banach space, perturbation techniques and applications to Chandrasekhar's and related equations, Ph. D. dissertation, University of Georgia, Athens, 1984. See *MR* **87d**: 47077

[4] BOWDEN, R. L. and ZWEIFEL, P. F., A Banach space analysis of the Chandrasekhar H-equation, *Astrophys. J.* **210** (1976), 219–232.

[5] CHANDRASEKHAR, S., *Radiative Transfer*, Dover Publ., New York, 1960. *MR* **22** #2446

[6] KELLEY, C. T., Solution of the Chandrasekhar H-equation by Newton's method, *J. Math. Phys.* **21** (1980), 1625–1628. *MR* **81j**: 85003

[7] KELLEY, C. T., Approximation of solutions of some quadratic integral equations in transport theory, *J. Integral Equations* **4** (1982), 221–237. *MR* **84j**: 85007

[8] MULLIKIN, T. W., Some probability distributions for neutron transport in a half-space, *J. Applied Probability* **5** (1968), 357–374. *MR* **38** #3929

[9] RALL, L. B., Quadratic equations in Banach spaces, *Rend. Circ. Mat. Palermo* (2) **10** (1961), 314–332. *MR* **26** #1731

[10] RALL, L. B., *Computational Solution of Nonlinear Operator Equations*, John Wiley Publ., New York, 1969. *MR* **39** #2289

[11] STIBBS, D. W. N. and WEIR, R. E., On the H-functions for isotropic scattering, *Monthly Not. Roy. Astr. Soc.* **119** (1959), 512–525. *MR* **22** #3576

(Received March 20, 1989)

DEPARTMENT OF MATHEMATICAL SCIENCES
CAMERON UNIVERSITY
LAWTON, OK 73505
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 369–378

A NEWTON-LIKE METHOD FOR SOLVING NONLINEAR EQUATIONS IN BANACH SPACE

I. K. ARGYROS

Abstract

In this paper we introduce iterations to solve nonlinear equations in a Banach space that are sometimes faster than the modified Newton's method, under the assumption that the nonlinear operators involved are once or twice Fréchet-differentiable.

Consider the equation

$$F(x)=0 \tag{1}$$

where F is twice Fréchet-differentiable at $z \in X$, nonlinear operator mapping a subset U of a Banach space X into a Banach space Y. We shall find it convenient to assume that U is a ball. Suppose that the approximation x_n has been found. To determine the next approximation x_{n+1} we replace (1) by the equation

$$F(x_n)+F'(x_n)(x-x_n)+\frac{1}{2}F''(x_n)(x-x_n,x-x_n)=0 \tag{2}$$

if the linear operators $\left[F'(z)-\frac{1}{2}F''(z)(x_n)\right]^{-1}$, $\left[F'(z)-\frac{1}{2}F''(z)(z)\right]^{-1}$ exist, then (2) suggests the iteration

$$x_{n+1}=x_n-\left[F'(z)-\frac{1}{2}F''(z)(x_n)\right]^{-1}F(x_n), \quad n=0,1,2,\ldots \tag{3}$$

or the modified version of (3)

$$x_{n+1}=x_n-\left[F'(z)-\frac{1}{2}F''(z)(z)\right]^{-1}F(x_n), \quad n=0,1,2,\ldots. \tag{4}$$

The above iterations converge to a solution x of (1) if the operators

$$T(x)=x-[F'(z)+B(x)]^{-1}F(x) \tag{5}$$

or the modified version of (5)

$$P(x)=x-[F'(z)+L]^{-1}F(x) \tag{6}$$

1980 *Mathematics Subject Classifications.* Primary 46B15.
Key words and phrases. Newton's method, bilinear-quadratic operator.

Akadémiai Kiadó, Budapest

have a fixed point in X, where B is a bounded symmetric bilinear operator on $X \times X$ and L is a bounded linear operator on X (usually, but not necessarily $B = -\frac{1}{2}F''(z)$, $L = -\frac{1}{2}F''(z)(z)$).

In this paper we give sufficient conditions for T and P to have unique fixed points in a closed ball centered at a specific $z \in X$ and then we compare (3) and (4) with the modified Newton's method [1], [2], [3], [4], [5], [6], [7]

$$x_{n+1} = x_n - (F'(z))^{-1}F(x_n), \quad n = 0, 1, 2, \ldots \tag{7}$$

using a simple scalar equation to show that (3) or (4) (most of the times) converge to a solution of (1) faster than (7).

DEFINITION 1. Denote by $L(X, Y)$ the linear space over the field of real or complex numbers of all linear operators from X into Y, then a linear operator B from X into $L(X, Y)$ is called a *bilinear operator from X into Y*.

The motivation for this definition is the observation that for any $x_1 \in X$, $L = B(x_1)$ is a linear operator from X into Y, so that

$$y = (B(x_1))(x_2) = B(x_1, x_2)$$

is an element of Y for $x_2 \in X$.

DEFINITION 2. A linear operator L from X into Y is said to be *bounded* if

$$\|L\| = \sup_{\|x\|=1} \|L(x)\| \tag{8}$$

is finite. The quantity $\|L\|$ is called the *norm* of L.

DEFINITION 3. A bilinear operator from X into Y is said to be *bounded* if it is a bounded linear operator from X into $L(X, Y)$. The *norm* $\|B\|$ of B is defined by (8), with B being considered to be an element of $L(X, L(X, Y))$.

The inequality

$$\|B(x, y)\| \leqq \|B\| \cdot \|x\| \cdot \|y\|, \quad x \in X, \quad y \in Y$$

is obvious from the above definitions.

DEFINITION 4. A bilinear operator B from X into X is called *symmetric* if

$$B(x, y) = B(y, x) \quad \text{for all} \quad x, y \in X.$$

REMARK 1. The operator B in (5) is assumed to be symmetric without loss of generality since B can always be replaced by the *mean* $\overline{B}$ of B defined by

$$\overline{B}(x, y) = \frac{1}{2}(B(x, y) + B(y, x)) \quad \text{for all} \quad x \in X, \ y \in Y.$$

Note that $\overline{B}(x,x) = B(x,x)$ for all $x \in X$.

DEFINITION 5. If F is an operator from X into Y, and for some $z \in X$ there exists a linear operator L from X into Y such that

$$\operatorname*{Lim}_{\|\Delta x\| \to 0} \frac{\|F(z+\Delta x) - F(z) - L\Delta x\|}{\|\Delta x\|} = 0,$$

then L is called the Fréchet derivative of F at z, denoted by $F'(z)$, and F is said to be differentiable once at z.

DEFINITION 6. If for some $r > 0$, F is differentiable once at all $x \in$ $\in U(z,r) = \{x \in X \mid \|x - z\| < r\}$, and a bilinear operator B from X to Y exists such that

$$\lim_{\|\Delta x\| \to 0} \frac{\|F'(z+\Delta x) - F'(z) - B\Delta x\|}{\|\Delta x\|} = 0,$$

B is called the second derivative of F at z, denoted by $F''(z)$, and F is said to be differentiable twice at z.

From now on we assume that $X = Y$, F is twice Fréchet differentiable at z and B is a bounded symmetric bilinear operator on X.

LEMMA 1. *Let L_1 and L_2 be bounded linear operators on X, where L_1 is invertible and $\|L_2\| \cdot \|L_1^{-1}\| < 1$. Then the operator $(L_1 + L_2)^{-1}$ is also invertible, and*

$$\|(L_1 + L_2)^{-1}\| \leqq \frac{\|L_1^{-1}\|}{1 - \|L_2\| \cdot \|L_1^{-1}\|}.$$

PROOF. The operator $(L_1 + L_2)^{-1}$ is invertible if the operator $I + L_1^{-1}L_2$ is invertible, since

$$(L_1 + L_2)^{-1} = (I + L_1^{-1}L_2)^{-1}L_1^{-1},$$

but

$$\|L_1^{-1}L_2\| \leqq \|L_1^{-1}\| \cdot \|L_2\| < 1,$$

so $I + L_1^{-1}L_2$ is invertible and

$$\|(L_1 + L_2)^{-1}\| = \|(I + L_1^{-1}L_2)^{-1}L_1^{-1}\| \leqq$$
$$\leqq \left\|(I + L_1^{-1}L_2)^{-1}\right\| \cdot \|L_1^{-1}\| \leqq \frac{\|L_1^{-1}\|}{1 - \|L_2\| \cdot \|L_1^{-1}\|}.$$

LEMMA 2. *If the linear operator $(F'(z) + B(z))^{-1}$ exists for some $z \in X$, then the linear operator*

$$[I + (F'(z))^{-1}B(x-z)]^{-1} \quad \textit{exists for every} \quad x \in U(z,r),$$

where r is such that

$$0 < r < \frac{1}{\|(F'(z)+B(z))^{-1}\|\|B\|}.$$

PROOF. By Lemma 1 it is enough to show

$$\|(F'(z)+B(z))^{-1}\| \cdot \|B(x-z)\| < 1,$$

or

$$\|(F'(z)+B(z))^{-1}\| \cdot \|B\| \cdot r < 1$$

which is true by hypothesis since $x \in U(z,r)$.

DEFINITION 7. Let z be fixed in X. Assume that the linear operator $\mathrm{C} = [F'(z)+B(z)]^{-1}$ exists and set $d = \|\mathrm{C}\|$, $e = \|B\|$. Define the linear operator A on $U(z,r) = \left\{x \in X \mid \|x-z\| < r < \frac{1}{de}\right\}$ by

$$A(x) = [I + (F'(z)+B(z))^{-1}B(x-z)]^{-1}.$$

Assume now that

$$\text{(c)} \qquad \begin{aligned} \|F'(x)-F'(y)\| &\leqq \ell_1\|x-y\|, \\ \|\mathrm{C}F(x)\| &\leqq n, \end{aligned}$$

where $x, y \in \overline{U}(z,r)$ and ℓ_1, n are nonnegative numbers. Define the numbers

$$\begin{aligned} h &= \|\mathrm{C}B\| \\ p &= \|\mathrm{C}B(z)\| \\ m &= \|\mathrm{C}F(z)\|. \end{aligned}$$

Note that $\|A(z)\| \leqq \frac{1}{1-der}$.

$$\|\mathrm{C}B(x)(x-z)\| \leqq \|\mathrm{C}B(x-z,x-z)\| + \|\mathrm{C}B(z,x-z)\| \leqq hr^2 + pr.$$

Define the real polynomials on $\mathbf{R}$ by

$$\begin{aligned} f(r) &= ar^2 + br + c, \\ g(r) &= a'r^2 + b'r + c' \end{aligned}$$

where

$$\begin{aligned} a &= de(de + d\ell_2 + h), \quad \ell_2 = 3\ell_1 \\ b &= dep - h - d\ell_2 - 2de, \quad \ell_3 = \frac{\ell_1}{2} \\ c &= 1 - p - hn \\ a' &= (\ell_3 + e)d + h \\ b' &= p - 1 \\ c' &= n. \end{aligned}$$

Finally, note that for any $x, y \in \overline{U}(z, r)$

$$T(x)-T(y)=[F'(z)+B(x)]^{-1}[F'(z)(x-y)-(F(x)-F(y))+B(x-z)(x-y)+ B(z)(x-y)+B(x-y)(F'(z)+B(y))^{-1}F(y)],$$

and

$$F'(z)(x-y)-(F(x)-F(y))=\int_0^1 (F'(z)-F'(x+t(y-x))(x-y)dt.$$

THEOREM 1. *Assume:*

(i) *The conditions* (c) *are satisfied for some* $z \in X$.

(ii) *There exists* r *such that* $f(r) > 0$ *and* $g(r) \leqq 0$.

Then (5) *has a unique fixed point in* $\overline{U}(z, r)$.

PROOF. T is well defined in $\overline{U}(z, r)$ by Lemma 2.

CLAIM 1. *T is a contraction operator on* $\overline{U}(z, r)$.

If $x, y \in \overline{U}(z, r)$, then according to Definition 7,

$$\|T(x)-T(y)\| \leqq \frac{1}{1-der}\left[d\ell_2 r+hr+p+\frac{hdn}{1-der}\right]\|x-y\|.$$

Now T is a contraction if

$$\frac{1}{1-der}\left[d\ell_2 r+hr+p+\frac{hdr}{1-der}\right] < 1$$

or $f(r) > 0$, which is true by (ii).

CLAIM 2. *T maps* $\overline{U}(z, r)$ *into* $\overline{U}(z, r)$.

If $x \in \overline{U}(z, r)$,

$$T(x)-z=A(x)\left[\mathrm{C}\int_0^1 (F'(z)-F'(z+t(x-z))(x-z)dt+\mathrm{C}B(x)(x-z)-\mathrm{C}F(z)\right],$$

then

$$\|T(x)-z\| \leqq r$$

if

$$\frac{1}{1-der}(d\ell_3 r^2+hr^2+pr+n) \leqq r$$

or

$$g(r) \leqq 0,$$

which is true by (ii).

The result now follows from the contraction mapping principle.

We now state a theorem for the modified equation (6). The proof as similar to the proof of Theorem 1 is omitted.

THEOREM 2. *Let z be as in Definition 7 and assume that there exists $r_0 \in [s,t)$ where*

$$s=\frac{1-\|(F'(z)+L)^{-1}L\|-\left[(1-\|(F'(z)+L)^{-1}L\|)^2-2\|(F'(z)+L)^{-1}\|\|(F'(z)+L)^{-1}F(z)\|\ell_1\right]^{1/2}}{\|(F'(z)+L)^{-1}\|\cdot\ell_1}$$

and

$$t=\frac{1-\|(F'(z)+L)^{-1}L\|}{\|(F'(z)+L)^{-1}\|\ell_1},$$

provided that the quantity under the radical is positive and

$$\|(F'(z)+L)^{-1}L\|<1.$$

Then (1) *has a unique solution x in $\overline{U}(z,r_0)$. Moreover, the rate of convergence $q(r_0)\in[u,1)$ where*

$$u=1-\left[(1-\|(F'(z)+L)^{-1}L\|)^2-2\|(F'(z)+L)^{-1}\|\|(F'(z)+L)^{-1}F(z)\|\ell_1\right]^{1/2}.$$

Assume that z is sufficient for the application of Newton's method and Theorem 2, then if q_N is the rate of convergence in Newton's method the iteration

$$x_{n+1}=x_n-[F'(z)+L]^{-1}F(x_n),\quad n=0,1,2,\dots \tag{9}$$

converges faster to a solution x of (1) if

$$q(r_0)<q_N=1-\left[1-2\ell_1\|F'(z)^{-1}\|\|F'(z)^{-1}F(z)\|\right]^{1/2}. \tag{10}$$

Denote by D, D_1 the quantities under the radicals in Theorem 2 and (10), respectively, then we have the following theorem.

THEOREM 3. *If the hypotheses in Newton's method* [4] *are satisfied then Theorem 2 can be applied also in* (1) *if*

$$\|L\|<\frac{1-[1-D_1]^{1/2}}{2\|F'(z)^{-1}\|}. \tag{11}$$

Moreover, if $q(r_0)<q_N$ then the iteration in (9) *converges faster to a solution x of* (1) *than the iteration in Newton's method* (7).

PROOF. By Lemma 1

$$\begin{aligned}\|(F'(z)+L)^{-1}L\|&=\left\|(I+F'(z)^{-1}L)^{-1}(F'(z))^{-1}L\right\|\leqq\\&\leqq\frac{\|F'(z)^{-1}\|\cdot\|L\|}{1-\|F'(z)^{-1}\|\cdot\|L\|}<1\qquad\text{by (11).}\end{aligned}\tag{12}$$

Now, using (11) and (12)

$$D \geqq \left(\frac{1-2\|F'(z)^{-1}\|\cdot\|L\|}{1-\|F'(z)^{-1}\|\cdot\|L\|}\right)^2 - \frac{2\ell_1\|F'(z)^{-1}\|\|F'(z)^{-1}F(z)\|}{(1-\|F'(z)^{-1}\|\cdot\|L\|)^2} =$$
$$= \frac{D_1 - 4\|F'(z)^{-1}\|\cdot\|L\|(1-\|F'(z)^{-1}\|\cdot\|L\|)}{(1-\|F'(z)^{-1}\|\cdot\|L\|)^2}$$

so, $D > 0$ if

$$D_1 > 4\|F'(z)^{-1}\|\cdot\|L\|(1-\|F'(z)^{-1}\|\cdot\|L\|)$$

which is true by (11). Therefore, Theorem 2 can be applied. The rest follows from the discussion after Theorem 2.

EXAMPLE. Let $X = \mathbf{R}\times\mathbf{R}$, be equipped with the max-norm. Define a bilinear operator B on X by the following calculation scheme:

$$B(x,y) = \left\{ (x_1,x_2)\begin{bmatrix} b_1^{11} & b_1^{12} \\ b_1^{21} & b_1^{22} \\ \hline b_2^{11} & b_2^{12} \\ b_2^{21} & b_2^{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} =$$

(13)
$$= \begin{bmatrix} b_1^{11}x_1 + b_1^{21}x_2 & b_1^{12}x_1 + b_1^{22}x_2 \\ b_2^{11}x_1 + b_2^{21}x_2 & b_2^{12}x_1 + b_2^{22}x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} =$$
$$= \begin{bmatrix} b_1^{11}x_1y_1 + b_1^{21}x_2y_1 + b_1^{12}x_1y_2 + b_1^{22}x_2y_2 \\ b_2^{11}x_1y_1 + b_2^{21}x_2y_1 + b_2^{12}x_1y_2 + b_2^{22}x_2y_2 \end{bmatrix},$$
$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in x.$$

It can easily be checked that B is a bilinear operator on X and as in [9] we can define the norm of B on X by

(14)
$$\|B\| = \sup_{\|x\|=1} \max_{(i)} \sum_{j=1}^{2}\left|\sum_{k=1}^{2} b_i^{jk}\xi_k\right|.$$

Define the linear operator $B(x)$ on X by

$$B(x)(y) = B(x,y)$$

where

$$B(x) = \begin{bmatrix} b_1^{11}x_1 + b_1^{21}x_2 & b_1^{12}x_1 + b_1^{22}x_2 \\ b_2^{11}x_1 + b_2^{21}x_2 & b_2^{12}x_1 + b_2^{22}x_2 \end{bmatrix}.$$

Let us now consider the quadratic system on X given by

(15)
$$F(x) = B(x,x) + L_1(x) + y = 0$$

where

$$B \sim \left[\begin{array}{cc} \frac{1}{2} & 1 \\ 1 & -\frac{3}{2} \\ \hline \frac{3}{2} & -1 \\ -1 & \frac{1}{2} \end{array}\right],$$

L_1 is a linear operator on X given by

$$L_1 = \frac{5}{2} I \quad (I, \text{ the identity operator on } X)$$

and

$$y = \begin{bmatrix} \frac{1}{16} \\ -\frac{1}{16} \end{bmatrix}.$$

Equation (15) can also be written using (13) as

$$\begin{aligned} \frac{1}{2}x_1^2 + 2x_1x_2 - \frac{3}{2}x_2^2 + \frac{5}{2}x_1 + \frac{1}{16} &= 0 \\ \frac{3}{2}x_1^2 - 2x_1x_2 + \frac{1}{2}x_2^2 + \frac{5}{2}x_2 - \frac{1}{16} &= 0. \end{aligned} \tag{16}$$

Let us choose

$$L = \frac{1}{10} I$$

and

$$z = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Then, obviously

$$\begin{aligned} \|L_1\| &= \frac{5}{2}, \\ \|y\| &= \frac{1}{16}, \\ \|L\| &= \frac{1}{10}, \end{aligned}$$

and using (14)

$$\|B\| = 4$$

and

$$\ell_1 = 2\|B\|.$$

Note that

$$F'(z) = 2B(z) + L_1 = L_1.$$

Then we can easily compute the quantities

$$\begin{aligned} D_1 &= .84 \\ q_N &= .0834849 \\ D &= .9230769 \\ u &= .0392311 \\ s &= .0025008 \\ t &= 3.125. \end{aligned}$$

Note that the hypotheses of Theorem 3 are satisfied for $r_0 \in [s, t)$ and by choosing $q(r_0) = u$ we observe that

$$q(r_0) < q_N$$

therefore, iteration (9) converges faster to a unique solution x of (15) in $U(0, r_0)$ than Newton's iteration.

Indeed, iterations (9) and (7) for solving (16) can now be written

$$x_{n+1} = x_n - \frac{10}{26} F(x_n) \tag{17}$$

and

$$\overline{x}_{n+1} = \overline{x}_n - \frac{10}{25} F(\overline{x}_n), \tag{18}$$

respectively, where

$$x_n = \begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix}, \quad \overline{x}_n = \begin{bmatrix} \overline{x}_{1,n} \\ \overline{x}_{2,n} \end{bmatrix}, \quad n = 0, 1, 2, \ldots$$

and

$$x_0 = \overline{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Let $\varepsilon = (.5)10^{-2}$ be the desired error tolerance that is

$$\|x - x_n\| \leqq \varepsilon \quad \text{for} \quad n \geqq N,$$

and

$$\|x - \overline{x}_n\| \leqq \varepsilon \quad \text{for} \quad n \geqq \overline{N}.$$

Then the true solution $x = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ is given by

$$\begin{aligned} x_1 &= -(24302916852540)10^{-2}, \\ x_2 &= (24062003442371)10^{-2}. \end{aligned}$$

Moreover, we have by (17) and (18)

$$x_{1,1} = -(24038461538462)10^{-2},$$
$$x_{2,1} = (23705087903960)10^{-2},$$
$$\overline{x}_{1,1} = -(25)10^{-2},$$
$$\overline{x}_{2,1} = (24625)10^{-2},$$
$$\overline{x}_{1,2} = -(24268665625)10^{-2},$$

and

$$\overline{x}_{2,2} = (24047248283457)10^{-2}.$$

We now observe that the number of steps N in (17) required to achieve the desired accuracy ε is

$$N = 1,$$

whereas the number of steps $\overline{N}$ in (18) required to achieve the same accuracy ε is

$$\overline{N} = 2.$$

REFERENCES

[1] BARTLE, R. G., Newton's method in Banach spaces, *Proc. Amer. Math. Soc.* **6** (1955), 827–831. *MR* **17**–176

[2] GOLDSTEIN, A. A., On Newton's method, *Numer. Math.* **7** (1965), 391–393. *MR* **33** #5033

[3] KRASNOSEL'SKIĬ, M. A., *Topological Methods in the Theory of Nonlinear Integral Equations*, Gostekhizdat, Moscow, 1956. *MR* **20** #3464

[4] KRASNOSEL'SKIĬ, M. A., *Approximate Solution of Operator Equations*, Wolters–Noordhoff Publ., Groningen, 1972. *MR* **52** #6515

[5] MIEL, G. J., An updated version of the Kantorovich theorem for Newton's method, Technical summary report, Mathematical research center, University of Wisconsin, Madison, 1980. See *MR* **83g**: 65062

[6] MOORE, R. H., Newton's method and variations, *Nonlinear Integral Equations* (Proc. Advanced Seminar Conducted by Math. Research Center, U.S. Army, Univ. Wisconsin, Madison, Wis., 1963), edited by P. M. Anselone, University of Wisconsin Press, Madison, Wis., 1964, 65–98. *MR* **29** #749

[7] POTRA, F. A. and PTAK, V., Sharp error bounds for Newton's process, *Numer. Math.* **34** (1980), 63–72. *MR* **81c**: 65027

[8] POTRA, F. A., An error analysis for the secant method, *Numer. Math.* **38** (1982), 427–445. *MR* **83h:** 65073

[9] RALL, L. B., Quadratic equations in Banach spaces, *Rend. Circ. Mat. Palermo* (2) **10** (1961), 314–332. *MR* **26** #1731

(Received March 20, 1989)

DEPARTMENT OF MATHEMATICAL SCIENCES
CAMERON UNIVERSITY
LAWTON, OK 73505
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 379–384

SPACES FOR WHICH THE UNIFORM BOUNDEDNESS PRINCIPLE HOLDS

R. LI and C. SWARTZ

1. Introduction

In [2], a theorem due to Antosik and Mikusinski concerning infinite matrices with entries in a metrizable topological group was systematically employed to treat a number of topics in functional analysis, classical analysis and measure theory. In particular, the matrix theorem was used to give a general form of the Uniform Boundedness Principle (UBP) that is valid with *no* completeness or barrelledness assumptions whatsoever on the domain space ([2] §4). Recently Li, Jin and Bu ([6], [7]) have extended the Antosik–Mikusinski matrix theorem to matrices with entries in an arbitrary topological group. This extended form of the matrix theorem now allows us to extend and improve some of the results of [2] for the UBP to arbitrary topological vector spaces (TVS) which are not metrizable. In this paper we discuss several such possible extensions and introduce a new class of spaces, called $\mathcal{A}$-spaces, which seem particularly natural for the UBP.

In Section 2 we discuss the basic matrix theorem. Since [6] is unpublished and [7] contains a very technical generalization of the Antosik–Mikusinski matrix theorem, for the sake of completeness, we give a straightforward proof of the matrix theorem for topological groups. In Section 3 we establish a general form of the UBP which is valid for arbitrary TVS. We next introduce a new class of spaces, called $\mathcal{A}$-spaces, for which a classical version of the UBP is valid. These spaces are more general than $\mathcal{K}$-spaces and enjoy more desirable properties than the $\mathcal{K}$-spaces. We give several examples of $\mathcal{A}$-spaces which are not $\mathcal{K}$-spaces. Finally, in Section 4, we discuss the relationship between uniform boundedness and equicontinuity. The results obtained yield equicontinuity versions of the UBP.

2. The basic matrix theorem

In this section we give a self-contained proof of the Antosik–Mikusinski Theorem ([2] 2.2) for group-valued matrices. In this section let G be an abelian topological group.

THEOREM 1. *Let $x_{ij} \in G$ for $i, j \in \mathbb{N}$. Suppose*

1980 *Mathematics Subject Classification* (1985 Revision). Primary 46A05

Key words and phrases. Locally convex space, uniform boundedness, equicontinuity, $\mathcal{K}$-spaces.

(I) $\lim_i x_{ij} = x_j$ *exists for each* j *and*

(II) *for each increasing sequence of positive integers* $\{m_j\}$ *there is a subsequence* $\{n_j\}$ *of* $\{m_j\}$ *such that* $\left\{\sum_{j=1}^{\infty} x_{in_j}\right\}_{i=1}^{\infty}$ *is Cauchy.*

Then $\lim_i x_{ij} = x_j$ *uniformly for* $j \in \mathbb{N}$. *In particular,*

$$\lim_i \lim_j x_{ij} = \lim_j \lim_i x_{ij} = 0 \quad \textit{and} \quad \lim_i x_{ii} = 0.$$

PROOF. If the conclusion fails, there are a closed, symmetric neighborhood U_0 of 0 and increasing sequences of positive integers $\{m_k\}$ and $\{n_k\}$ such that $x_{m_k n_k} - x_{n_k} \notin U_0$ for all k. Pick a closed, symmetric neighborhood U_1 of 0 such that $U_1 + U_1 \subseteqq U_0$ and set $i_1 = m_1$, $j_1 = n_1$. Since $x_{i_1 j_1} - x_{j_1} = (x_{i_1 j_1} - x_{i j_1}) + (x_{i j_1} - x_{j_1})$, there exists i_0 such that $x_{i_1 j_1} - x_{i j_1} \notin U_1$ for $i \geqq i_0$. Choose k_0 such that

$$m_{k_0} > \max\{i_1, i_0\},\ n_{k_0} > j_1 \quad \text{and set} \quad i_2 = m_{k_0},\ j_2 = n_{k_0}.$$

Then $x_{i_1 j_1} - x_{i_2 j_1} \notin U_1$ and $x_{i_2 j_2} - x_{j_2} \notin U_0$. Proceeding in this manner produces increasing sequences $\{i_k\}$ and $\{j_k\}$ such that $x_{i_k j_k} - x_{j_k} \notin U_0$ and $x_{i_k j_k} - x_{i_{k+1} j_k} \notin U_1$. For convenience, set $z_{k,\ell} = x_{i_k j_\ell} - x_{i_{k+1} j_\ell}$ so $z_{k,k} \notin U_1$.

Choose a sequence of closed, symmetric neighborhoods of 0, $\{U_n\}$, such that $U_n + U_n \subseteqq U_{n-1}$ for $n \geqq 1$. Note that

$$U_3 + U_4 + \ldots + U_m = \sum_{j=3}^{m} U_j \subseteqq U_2$$

for each $m \geqq 3$. By (I) and (II), $\lim_k z_{k\ell} = 0$ for each ℓ and $\lim_\ell z_{k\ell} = 0$ for each k so there is an increasing sequence of positive integers $\{p_k\}$ such that $z_{p_k p_\ell}, z_{p_\ell p_k} \in U_{k+2}$ for $k > \ell$. By (II) $\{p_k\}$ has a subsequence $\{q_k\}$ such that $\left\{\sum_{k=1}^{\infty} x_{i q_k}\right\}_{i=1}^{\infty}$ is Cauchy so $\lim_k \sum_{\ell=1}^{\infty} z_{q_k q_\ell} = 0$. Thus, there exists k_0 such that $\sum_{\ell=1}^{\infty} z_{q_{k_0} q_\ell} \in U_2$. Then for $m > k_0$

$$\sum_{\substack{\ell=1 \\ \ell \neq k_0}}^{m} z_{q_{k_0} q_\ell} = \sum_{\ell=1}^{k_0-1} z_{q_{k_0} q_\ell} + \sum_{\ell=k_0+1}^{m} z_{q_{k_0} q_\ell} \in \sum_{\ell=1}^{k_0-1} U_{k_0+2} + \sum_{\ell=k_0+1}^{m} U_{\ell+2} \subseteqq \sum_{\ell=3}^{m+2} U_\ell \subseteqq U_2$$

so $z_{k_0} = \sum_{\substack{\ell=1 \\ \ell \neq k_0}}^{\infty} z_{q_{k_0} q_\ell} \in U_2$. Thus,

$$z_{q_{k_0} q_{k_0}} = \sum_{\ell=1}^{\infty} z_{q_{k_0} q_\ell} - z_{k_0} \in U_2 + U_2 \subseteqq U_1.$$

This is a contradiction and establishes the result.

A matrix which satisfies conditions (I) and (II) is called a $\mathcal{K}$-matrix.

3. The UBP and $\mathcal{A}$-spaces

Let X and Y be TVS and let $L_s(X,Y)$ be the space of sequentially continuous linear operators from X into Y. If τ is the vector topology of X, a sequence $\{x_k\} \subseteqq X$ is said to be $\tau - \mathcal{K}$-convergent to 0 (or $\mathcal{K}$-convergent to 0 if the topology τ is understood) if every subsequence of $\{x_k\}$ has a subsequence $\{x_{n_k}\}$ such that the subseries $\sum_{k=1}^{\infty} x_{n_k}$ is τ-convergent to an element $x \in \in X$ ([2], 3.1). Following the analogue of the sequential characterization of boundedness in a TVS ([10]), Antosik introduced the notion of a $\mathcal{K}$-bounded set ([1]); a subset $A \subseteqq X$ is said to be $\tau - \mathcal{K}$-bounded (or $\mathcal{K}$-bounded if τ is understood) if for every sequence $\{x_k\} \subseteqq A$ and every scalar sequence $t_k \to 0$, the sequence $\{t_k x_k\}$ is $\tau - \mathcal{K}$-convergent to 0 ([1], [2] 3.4).

The classical version of the UBP for normed spaces asserts that if $\Gamma \subseteqq \subseteqq L_s(X,Y)$ is pointwise bounded on X and X is complete, then Γ is uniformly bounded on bounded subsets of X. This result fails to hold if X is not complete. The general versions of the UBP given in [2] are obtained by replacing the family of bounded subsets of X by either $\mathcal{K}$-convergent sequences or $\mathcal{K}$-bounded sets. We now give a form of the UBP which is valid for arbitrary TVS.

Throughout this section let $\Gamma \subseteqq L_s(X,Y)$. Let $\sigma(\Gamma)$ be the weakest topology on X such that all of the elements of Γ are continuous. The following UBP extends the UBP of [2] from the case of metric linear spaces to arbitrary TVS.

THEOREM 2. *If Γ is pointwise bounded on X, then Γ is*
(1) *uniformly bounded on $\sigma(\Gamma) - \mathcal{K}$-convergent sequences and*
(2) *uniformly bounded on $\sigma(\Gamma) - \mathcal{K}$-bounded subsets of X.*

PROOF. (1) Let $\{x_j\}$ be $\sigma(\Gamma) - \mathcal{K}$-convergent, $\{T_j\} \subseteqq \Gamma$ and $\{t_j\}$ a sequence of scalars which converges to 0. The matrix $M = [t_i T_i x_j]$ is a $\mathcal{K}$-matrix so by Theorem 1 $t_i T_i x_i \to 0$, and (1) is established.

(2) Let $B \subseteqq X$ be $\sigma(\Gamma) - \mathcal{K}$-bounded, $\{x_i\} \subseteqq B$, $\{T_i\} \subseteqq \Gamma$ and $\{t_j\}$ be a sequence of positive scalars which converges to 0. Then $M = \left[\sqrt{t_i} T_i (\sqrt{t_j} x_j)\right]$ is a $\mathcal{K}$-matrix so the result follows from Theorem 1.

We next introduce a class of spaces for which the analogue of one version of the classical UBP is valid. This class of spaces seems to be the appropriate class for which this form of the UBP holds.

DEFINITION 3. A TVS (X, τ) is said to be an $\mathcal{A}$-space if every τ-bounded subset of X is $\tau - \mathcal{K}$-bounded.

For such spaces, we obtain the following UBP from Theorem 2.

COROLLARY 4. *Let X be an $\mathcal{A}$-space. If Γ is pointwise bounded on X, then Γ is uniformly bounded on bounded subsets of X.*

The following propositions give a large number of examples of $\mathcal{A}$-spaces.

PROPOSITION 5. *If X is locally convex and sequentially complete, then X is an $\mathcal{A}$-space.*

PROOF. Let $A \subseteqq X$ be bounded, $\{x_j\} \subseteqq A$ and $t_j \to 0$. Given a subsequence of $\{t_j\}$, pick a further subsequence satisfying $\sum_{j=1}^{\infty} |t_{n_j}| < \infty$. Set $s_n = \sum_{j=1}^{n} t_{n_j} x_{n_j}$. If p is any continuous semi-norm on X, then $p(s_n - s_{n+p}) \leqq \sum_{j=n+1}^{n+p} |t_{n_j}| p(x_{n_j})$ so $\{s_n\}$ is a Cauchy sequence in X and, therefore, convergent.

Wilansky gives a list of sufficient conditions for a locally convex space to be sequentially complete in Table 30, p. 281 of [12]. All of these spaces are $\mathcal{A}$-spaces by Proposition 5.

From Proposition 5, we have the following examples of $\mathcal{A}$-spaces.

COROLLARY 6. *If X is semi-reflexive, then $(X, weak)$ is an $\mathcal{A}$-space.*

PROOF. A semi-reflexive space is quasi-complete ([8] 23.3.(2)).

For the same reason, we have

COROLLARY 7. *If X is a barrelled locally convex space, then $(X', \sigma(X', X))$ is an $\mathcal{A}$-space.*

The following simple observation can also be used to furnish examples of $\mathcal{A}$-spaces.

PROPOSITION 7. *If $T \in L_s(X, Y)$, then T carries $\mathcal{K}$-bounded sets to $\mathcal{K}$-bounded sets.*

COROLLARY 8. *Let (Z, Z') be a dual pair and let $\tau \subseteqq \sigma$ be two locally convex topologies which are compatible with this duality. If (Z, σ) is an $\mathcal{A}$-space, then (Z, τ) is an $\mathcal{A}$-space.*

PROOF. The identity from (Z, σ) to (Z, τ) is continuous, and, hence, every τ-bounded set is $\tau - \mathcal{K}$-bounded.

If X is a B-space, then by Corollary 8 (X, weak) is an $\mathcal{A}$-space which is not barrelled. Thus, the version of the UBP given in Corollary 4 is valid for spaces which are not necessarily barrelled (see the remark following [9] 39.3(2)).

Recall that a TVS X is said to be a $\mathcal{K}$-space if every sequence which converges to 0 is $\mathcal{K}$-convergent to 0 ([2] §3). Thus, a $\mathcal{K}$-space is obviously an $\mathcal{A}$-space. But, from Corollaries 6, 7 and 8, we see that (ℓ^p, weak), $1 < p < \infty$, and (ℓ^1, weak^*) are $\mathcal{A}$-spaces but are not $\mathcal{K}$-spaces [consider the unit vectors e_k which have a 1 in the k-th coordinate and 0 in the other coordinates]. These examples also show that a sequentially complete space need not be a $\mathcal{K}$-space, but by Proposition 5 are always $\mathcal{A}$-spaces. Klis' example ([5]) of a noncomplete normed $\mathcal{K}$-space shows that an $\mathcal{A}$-space need not be sequentially complete. The examples above also show that Corollary 7 is false for $\mathcal{K}$-spaces.

The examples above of $\mathcal{A}$-spaces which are not $\mathcal{K}$-spaces are not metrizable. Indeed, for metrizable spaces, we have

PROPOSITION 9. *If the metric linear space X is an $\mathcal{A}$-space, then X is a $\mathcal{K}$-space.*

PROOF. Let $x_k \to 0$. Then there exists a scalar sequence $t_k \uparrow \infty$ such that $t_k x_k \to 0$. Now $\{t_k x_k\}$ is bounded and, therefore, $\mathcal{K}$-bounded. Hence, $\{(1/t_k) t_k x_k\} = \{x_k\}$ is $\mathcal{K}$-convergent to 0.

The proof above obviously holds for what Khaleelulla calls braked spaces ([4]).

4. Equicontinuity and uniform boundedness

The conclusion of the UBP for operators defined on either F-spaces or barrelled spaces is often given in the form: if Γ is pointwise bounded on X, then Γ is equicontinuous ([2] 4.5, [9] 39.3.(2)). We next consider the relationship between equicontinuity and the conclusion given in Corollary 4. The results obtained in Proposition 11 and 13 can then be combined with Corollary 4 to obtain an analogue of this form of the UBP.

Consider the following:

(I) $\Gamma \subseteqq L(X, Y)$ is uniformly bounded on bounded subsets of X.

(II) Γ is equicontinuous.

It is routine to check that (II) always implies (I). That the converse implication does not hold in general follows from the example below.

EXAMPLE 10. Let X be c_0 with the weak topology. Then $\Gamma \subseteqq X' = \ell^1$ is uniformly bounded on weak bounded sets if and only if Γ is norm bounded in ℓ^1. The sequence $\{e_k\} \subseteqq \ell^1$ norm bounded but is not equicontinuous with respect to the weak topology of c_0 since $\{e_k\}$ is weakly convergent to 0 in c_0 but $\langle e_k, e_k \rangle = 1$.

We give sufficient conditions which guarantee that (I) implies (II).

PROPOSITION 11. *Let X and Y be locally convex with X infrabarrelled. Then (I) implies (II).*

PROOF. Let $V \subseteqq Y$ be an absolutely convex, closed neighborhood of 0. Put $U = \cap\{T^{-1}V : T \in \Gamma\}$. We show that U is a bornivore, and the result will then follow from the infrabarrelled assumption. That U is closed and absolutely convex is clear. Let $A \subseteqq X$ be bounded. Then $\Gamma(A)$ is bounded so there is $\lambda > 0$ such that $\Gamma(A) \subseteqq \lambda V$. Thus, $A \subseteqq \lambda U$, and U is a bornivore.

This result improves Proposition 7 of [11].

In a certain sense Proposition 11 is best possible for locally convex spaces. For, if X has the property that (I) and (II) are equivalent when Y is the scalar field, then strongly bounded subsets of X' are equicontinuous, and X is infrabarrelled ([8] 23.4.(4)).

From Proposition 11 and Corollary 4, we obtain

COROLLARY 12. *If X is an infrabarrelled $\mathcal{A}$-space, then X is barrelled. In particular, a locally convex, metrizable $\mathcal{A}$-space is barrelled.*

In [3] it is shown that a metric $\mathcal{K}$-space is a Baire space and, therefore, barrelled. Corollary 12 gives a locally convex generalization of this result.

In the non-locally convex case, we have

PROPOSITION 13. *If X is a metric linear space, then (I) implies (II).*

PROOF. Let $x_j \to 0$ in X. It suffices to show that $T_j x_j \to 0$ for every $\{T_j\} \subseteqq \Gamma$. Pick $t_j \uparrow \infty$ such that $t_j x_j \to 0$. Then $\{T_j(t_j x_j)\}$ is bounded by (I) so $(1/t_j)T_j(t_j x_j) = T_j x_j \to 0$.

In [2] §5 versions of the Banach–Steinhaus Theorem were developed using $\mathcal{K}$-convergent sequences which required no completeness or barrelledness assumptions. Using Theorem 1 these versions can now easily be extended to the case where both the domain and range spaces are arbitrary TVS ([2] 5.3, 5.4). We omit the precise statements and proofs since they are identical with those given in [2].

REFERENCES

[1] ANTOSIK, P., On uniform boundedness of families of mappings, *Proceedings of the Conference on Convergence Structures* (Szczyrk, 1979), Polsk. Akad. Nauk, Oddział Katowicach, Katowice, 1980, 1–16. *MR* **83i**: 46006a

[2] ANTOSIK, P. and SWARTZ, C., *Matrix Methods in Analysis*, Lecture Notes in Mathematics, Vol. 1113, Springer-Verlag, Berlin–New York, 1985. *MR* **87b**: 46079

[3] BURZYK, J., KLIS, C. and LIPECKI, Z., On metrizable abelian groups with a completeness-type property, *Colloq. Math.* **49** (1984), 33–39. *MR* **86c**: 22001

[4] KHALEELULLA, S. M., *Counterexamples in Topological Vector Spaces*, Lecture Notes in Mathematics, Vol. 936, Springer-Verlag, Berlin–Heidelberg, 1982. *MR* **84a**: 46002

[5] KLIS, C., An example of a noncomplete normed (K) space, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **26** (1978), 415–420. *MR* **58** #17795

[6] LI, R. and BU, Z., Basic matrix theorem (preprint).

[7] LI, R., JIN, C., and BU, Z., Mikusinski–Antosik basic matrix theorem is a fundamental principle of topological vector space theory.

[8] KÖTHE, G., *Topological Vector Spaces I*, Die Grundlehren der mathematischen Wissenschaften, Bd. 159, Springer-Verlag, Berlin, 1969. *MR* **40** #1750

[9] KÖTHE, G., *Topological Vector Spaces II*, Grundlehren der mathematischen Wissenschaften, Bd. 237, Springer-Verlag, Berlin, 1979. *MR* **81g**: 46001

[10] MAZUR, M. and ORLICZ, W., Über Folgen linearer Operationen, *Studia Math.* **4** (1933), 152–157. *Zbl.* **8**, 250

[11] SWARTZ, C., A generalisation of Mackey's theorem and the uniform boundedness principle, *Bull. Austral. Math. Soc.* **40** (1989), 123–128. *MR* **90k**: 46005

[12] WILANSKY, A., *Modern Methods in Topological Vector Spaces*, McGraw-Hill, New York, 1978. *MR* **81d**: 46001

(Received March 15, 1989)

DEPARTMENT OF MATHEMATICAL SCIENCES
COLLEGE OF ARTS AND SCIENCES
NEW MEXICO STATE UNIVERSITY
BOX 30001, DEPT. 3MB
LAS CRUCES, NM 88003
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 385–389

A PROBLEM IN GAME THEORY

M. HORVÁTH, I. JOÓ and Z. SZENTMIKLÓSSY

The minimax theorems form a central part of game theory. If we are given two sets X, Y and a function $f\colon X \times Y \to \mathbf{R}$, the equality

$$\sup_{x\in X}\inf_{y\in Y} f(x,y) = \inf_{y\in Y}\sup_{x\in X} f(x,y) \tag{1}$$

is called minimax equality. Several conditions ensuring (1) are known, we only mention some of the latest results [1–9] of Hungarian mathematicians. Denote by c^* the right of (1) and for some $c < c^*$ define

$$H^y = \{x : f(x,y) \geqq c\}, \qquad H_x = \{y : f(x,y) \geqq c\}.$$

In [5] is proved that (1) is equivalent to the assertion that

$$\bigcap_{y\in Y} H^y \neq \emptyset \quad \text{for all} \quad c < c^*. \tag{2}$$

If X is compact and the sets H^y are closed in X then (2) is ensured whenever the finite intersections of the sets H^y are nonempty.

An interval space on some topological space Y can be defined by a mapping

$$[\cdot,\cdot]\colon Y \times Y \to P(Y)$$

where $P(Y)$ means the set of subsets of Y and if the following conditions hold

$$y_1, y_2 \in [y_1, y_2] \tag{3}$$

$$[y_1, y_2] \subset Y \quad \text{is closed and connected} \tag{4}$$

for any $y_1, y_2 \in Y$. In an interval space we consider $[y_1, y_2]$ as an interval with endpoints y_1 and y_2. A set $K \subset Y$ is called convex if $y_1, y_2 \in K$ implies $[y_1, y_2] \subset K$.

In the following theorems we consider subsets $H_x \subset Y$, $H^y \subset X$ such that

$$x \in H^y \Leftrightarrow y \in H_x. \tag{5}$$

This means that there exists a set $Z \subset X \times Y$ such that

$$H_x = \{y : (x,y) \in Z\}, \quad H^y = \{x : (x,y) \in Z\}.$$

The second author of this paper raised the following

PROBLEM. Let X be a compact topological space, Y be an interval space. Define the set $H^y \subset X$, $H_x \subset Y$ satisfying (5) such that

a) $H^y \neq \emptyset$, $H^{y_1} \cap \ldots \cap H^{y_k}$ is closed and connected (may be, empty) for all $y_1, \ldots, y_k \in Y$.

b) $Y \setminus H_x$ is open and convex for all $x \in X$. Does it hold in this case that

$$H^{y_1} \cap \ldots \cap H^{y_k} \neq \emptyset \quad \text{for all} \quad y_1, \ldots, y_k \in Y?$$

We are not able to give the answer. In what follows we get positive answer using some further conditions.

THEOREM 1. *Let X be a compact M_1 space and $Y \in T_1$ be an interval space where the intervals are convex. Suppose the following property. For any convex set $K \subset Y$ contained in an interval and for any boundary point y of K there exists $y_1 \in K$ such that $(y, y_1] = [y, y_1] \setminus \{y\} \subset K$. Then* a) *and* b) *imply the finite intersection property*

$$H^{y_1} \cap \ldots \cap H^{y_k} \neq \emptyset \qquad y_1, \ldots, y_k \in Y. \tag{6}$$

PROOF. We establish first some consequences of a) and b). First,

$$y \in [y_1, y_2] \Rightarrow H^y \subset H^{y_1} \cup H^{y_2}. \tag{7}$$

Indeed, if $x \in H^y \setminus (H^{y_1} \cup H^{y_2})$ then $y_1, y_2 \in Y \setminus H_x$, $y \notin Y \setminus H_x$ which contradicts the convexity of $Y \setminus H_x$. Secondly

$$H^{y_1} \cap H^{y_2} = \emptyset, \quad y \in [y_1, y_2] \Rightarrow H^y \subset H^{y_1} \quad \text{or} \quad H^y \subset H^{y_2} \tag{8}$$

follows from a). Using this we remark first that it is enough to prove (6) for two sets, then the general case follows by induction. Indeed, if (6) holds for some k, then fix arbitrarily $y_1, \ldots, y_{k-1}$ and take $H^{y_1} \cap \ldots \cap H^{y_{k-1}}$ instead of X, $H^y \cap H^{y_1} \cap \ldots \cap H^{y_{k-1}}$ instead of H^y; H_x remains the same, and apply the statement for two members to obtain the case $k+1$. To see the case $k=2$ suppose indirectly that for some $a, b \in Y$ $H^a \cap H^b = \emptyset$. Define the sets

$$K_a = \{y \in [a, b] : H^y \subset H^a\}, \quad K_b = \{y \in [a, b] : H^y \subset H^b\}.$$

Then by (8) $[a, b] = K_a \cup^* K_b$, and the sets K_a, K_b are convex by (7). Since $[a, b]$ is connected, K_a and K_b cannot be simultaneously closed; e.g. there exists a boundary point y of K_a not belonging to K_a. We know that there exists $y_0 \in K_a$ such that $(y, y_0] \subset K_a$, $y \in K_b$. Fix a neighbourhood basis U_n of y. There exists a point $y_1 \in (y, y_0] \cap U_1$ (otherwise $\{y\}$ would be open and closed in $[y, y_0]$), there exists $y_2 \in (y, y_1] \cap U_2, \ldots, y_n \in (y, y_{n-1}] \cap \cap U_n, \ldots$. Since $H^y \subset H^b$ and $H^{y_n} \subset H^a$, (7) implies that $H^{y_{n+1}} \subset H^{y_n}$, and then $\bigcap_n H^{y_n} \neq \emptyset$ by the compactness. Take an $x \in \bigcap_n H^{y_n}$, then obviously

$x \notin H^y$. In other words, $y_n \in H_x$, $y \notin H_x$, in contradiction with b) since H_x is closed. Theorem 1 is proved.

REMARK. The compactness of X cannot be omitted above. Consider the following counterexample. Let $X = \mathbf{R}_+$, $Y = [0,1]$ and let

$$H_0 = \left[0, \frac{1}{2}\right] \cup \{1\},$$
$$H_x = \left[\frac{1}{2} + \frac{1}{2+x}, 1\right] \qquad x > 0$$

and define H^y by (5). It is easy to see that the other conditions of Theorem 1 hold, however

$$H^0 \cap H^{\frac{3}{4}} = \emptyset.$$

THEOREM 2. *Let X be a compact topological space, $Y \in T_1$ be an interval space where the intervals are convex and compact. Then* a) *and* b) *imply the finite intersection property* (6).

REMARK. With a sketched proof this statement is given in [10]. We give here the details.

PROOF. Again it is enough to prove (6) for two members. Suppose indirectly that $H^a \cap H^b = \emptyset$. We call an interval $[a', b']$ good if $H^{a'} \cap H^{b'} = \emptyset$. We establish a partial ordering between good subintervals of $[a, b]$, namely let

$$[a', b'] < [a'', b''] \quad \text{if} \quad [a', b'] \supset [a'', b''] \quad \text{and} \quad H^{a'} \supset H^{a''},\ H^{b'} \supset H^{b''}.$$

By the Kuratowski lemma we have a maximal ordered subset $\mathcal{Z}$. It is confinal with its well-ordered subset $[a_\xi, b_\xi]$, $\xi < \mathcal{K}$ where $\mathcal{K}$ is a regular well-ordering type. Let $\hat{a}$ resp. $\hat{b}$ be condensation points of the sequence a_ξ resp. b_ξ; they exist since $a_\xi, b_\xi \in [a, b]$ and $[a, b]$ is compact. Then $[\hat{a}, \hat{b}] \subset [a_\xi, b_\xi]$ for all ξ. Since $\xi < \xi'$ implies $H^{a_\xi} \supset H^{a_{\xi'}}$ hence by the compactness of X, $\bigcap_\xi H^{a_\xi} \neq \emptyset$. Let $x \in \bigcap_\xi H^{a_\xi}$, then $a_\xi \in H_x$ and consequently $\hat{a} \in H_x$, $x \in H^{\hat{a}} \subset H^{a_\xi}$. Analogously we get $H^{\hat{b}} \subset H^{b_\xi}$ for all ξ. So we see that $\hat{a} \neq \hat{b}$ and that $[\hat{a}, \hat{b}]$ is a maximal element of $\mathcal{Z}$. But this is impossible: by $Y \in T_1$ there exists $y \in (\hat{a}, \hat{b})$; now if $H^y \subset H^{\hat{a}}$ then $[y, \hat{b}] > [\hat{a}, \hat{b}]$ and if $H^y \subset H^{\hat{b}}$ then $[\hat{a}, y] > [\hat{a}, \hat{b}]$. The contradiction proves Theorem 2.

THEOREM 3. *Let X be a compact space, $Y \in T_1$ be an interval space. Suppose* a), b) *and*

c) $y_1, y_2 \in Y$, $y', y'' \in (y_1, y_2] \Rightarrow (y_1, y'] \cap (y_1, y''] \neq \emptyset$.

d) *If $y_1, y_2, y_\alpha \in Y$, y_α is a net converging to y_2 then* $[y_1, y_2) \subset \bigcup_\alpha [y_1, y_\alpha)$.

If a), b), c) d) *hold then* (6) *follows.*

PROOF. Suppose indirectly that $H^a \cap H^b = \emptyset$ and consider the sets K_a, K_b as in Theorem 1. Suppose for example that there exists a boundary point y of K_a with $y \notin K_a$. Take any point $y_1 \in K_a$; we shall prove that $(y, y_1] \subset K_a$. Indeed, let $y_\alpha \subset K_a$ be a net converging to y, then d) shows by the convexity of K_a that

$$(y, y_1] \subset \bigcup_\alpha (y_\alpha, y_1] \subset K_a.$$

Now c) shows that if $y', y'' \in (y, y_1]$ then there exists $\hat{y} \in (y, y'] \cap (y, y'']$, and hence $H^{y'} \cap H^{y''} \supset H^{\hat{y}} \neq \emptyset$. Since X is compact, there exists $x_0 \in \cap\{H^{y'} : y' \in \in (y, y_1]\}$. This means that $(y, y_1] \subset H_{x_0}$ and then $y \in H_{x_0}$, $x_0 \in H^y$. But this is impossible, because $H^y \subset H^b$ and $H^{y_1} \subset H^a$. This proves Theorem 3.

REMARK. The above results give three minimax theorems. Namely if $f \colon X \times Y \to \mathbf{R}$ is a function such that for all $c < c^*$ the sets $H^y = \{x : f(x, y) \geqq \geqq c\}$, $H_x = \{y : f(x, y) \geqq c\}$ satisfy a) and b) and if the spaces X, Y satisfy the conditions given in the statement of the above theorems then

$$\inf_y \sup_x f = \sup_x \inf_y f.$$

Now if the partial functions $x \mapsto f(x, y)$ for any fixed y are upper semicontinuous, then H^y is closed, and if the other partial functions are also upper semicontinuous, then H_x is also closed. We say that the functions $y \mapsto f(x, y)$ are quasiconvex if the sets $Y \setminus H_x$ are convex. The remaining part of a), namely that $H^{y_1} \cap \ldots \cap H^{y_k}$ be connected, can be ensured e.g. if we endow X with an interval structure and in this interval space the functions $x \mapsto f(x, y)$ are quasiconcave.

REFERENCES

[1] JOÓ, I., A simple proof for von Neumann's minimax theorem, *Acta Sci. Math. (Szeged)* **42** (1980), 91–94. *MR* **81i**: 49008

[2] STACHÓ, L. L., Minimax theorems beyond topological vector spaces, *Acta Sci. Math. (Szeged)* **42** (1980), 157–164. *MR* **82a**: 49017

[3] JOÓ, I. and SÖVEGJÁRTÓ, A., A fixed point theorem, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **24** (1981), 9–11. *MR* **87m**: 47130

[4] JOÓ, I. and STACHÓ, L. L., A note on Ky Fan's minimax theorem, *Acta Math. Acad. Sci. Hungar.* **39** (1982), 401–407. *MR* **83i**: 90158

[5] JOÓ, I., Note on my paper: "A simple proof for von Neumann's minimax theorem", *Acta Math. Hungar.* **44** (1984), 363–365. *MR* **86f**: 49039

[6] KOMORNIK, V., Minimax theorems for upper semicontinuous functions, *Acta Math. Acad. Sci. Hungar.* **40** (1982), 159–163. *MR* **84h**: 49023

[7] HORVÁTH, M. and SÖVEGJÁRTÓ, A., On convex functions, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **29** (1986), 193–198. *MR* **89g**: 49015

[8] BOGMÉR, A., HORVÁTH, M., and JOÓ, I., Minimax theorems and convexity, *Mat. Lapok* **34** (1987), 149–170 (in Hungarian).

[9] HORVÁTH, M. and JOÓ, I., A Ky Fan-konvexitásról, *Mat. Lapok* **34** (1987), 137–140.
[10] JOÓ, I., On some convexities, *Acta Math. Hungar.* **54** (1989), 163–172. *MR* **90i**: 52003

(Received March 24, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 391–401

ON L^1-NORM CONVERGENCE OF HERMITE INTERPOLATION BASED ON THE ROOTS OF JACOBI, HERMITE, LAGUERRE AND SONIN–MARKOV POLYNOMIALS

S. SZABÓ

The norm convergence of interpolation processes is investigated by many authors. We mention first the classical result of Erdős and Turán [3] stating that for a general class of weights, the Lagrange interpolation converges in L^2-norm on finite intervals. For the case of infinite intervals the analogous theorem was proved by Balázs and Turán [4], [5]. In [6] the authors gave an estimate for the speed of convergence in L^2-norm of the Lagrange interpolation with Laguerre abscissas; it was based on a Jackson-type theorem corresponding to the Laguerre weight. For the Hermite interpolation on finite intervals a general result is given in [2] p. 419. The aim of this paper is to prove the L^1-norm convergence of Hermite interpolation formed by the Jacobi, Hermite, Laguerre or Sonin–Markov nodes.

Let $x_1 < \ldots < x_n$ be the nodes of the n-th interpolation. Let

$$w(x) = \prod_{i=1}^{n}(x - x_i)$$

and the polynomials

$$l_k(x) = l_{k,n}(x) = \frac{w(x)}{w'(x_k)(x - x_k)}$$

called the fundamental polynomials of the Lagrange interpolation. The Hermite interpolation polynomials of a function $f \in C^1$ are defined as

$$(1) \qquad H_n(f,x) = \sum_{k=1}^{n}\left[f(x_k)\Big(1 - \frac{w''(x_k)}{w'(x_k)}(x - x_k)\Big) + f'(x_k)(x - x_k)\right] l_k^2(x).$$

Consider first the Jacobi polynomials $p_n^{(\alpha,\beta)}(x)$, $\alpha, \beta > -1$ defined by

$$\int_{-1}^{1} p_n^{(\alpha,\beta)}(x) p_k^{(\alpha,\beta)}(x)(1 - x)^{\alpha}(1 + x)^{\beta} dx = \delta_{n,k}.$$

1980 *Mathematics Subject Classification* (1985 Revision). Primary 41A05; Secondary 41A10.

Key words and phrases. Hermite interpolation, L^1-norm convergence, orthogonal polynomials, Stone–Weierstrass theorem.

Akadémiai Kiadó, Budapest

Introduce the space

$$C(\gamma,\delta)=\{f\in C(-1,1): \lim_{|x|\to 1} f(x)(1-x)^{\gamma}(1+x)^{\delta}=0\}.$$

The Hermite interpolation polynomials for the Jacobi abscissas are defined by

$$(2)\quad 1-\frac{w''(x_k)}{w'(x_k)}(x-x_k)=\frac{1-x[\alpha-\beta+(\alpha+\beta+2)x_k]+(\alpha-\beta)x_k+(\alpha+\beta+1)x_k^2}{1-x_k^2}.$$

We shall prove first the L^1-norm boundedness of the operator sequence H_n.

LEMMA 1. *Let* $\alpha,\beta>-1$, $0<\varepsilon<\min(\alpha+1,\beta+1)$. *Then*

(3)

$$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|H_n(f,x)|dx\leqq c\max_{1\leqq k\leqq n}|f(x_k)|(1-x_k)^{\alpha+1-\varepsilon}(1+x_k)^{\beta+1-\varepsilon}+$$

$$+c\max_{1\leqq k\leqq n}|f'(x_k)|(1-x_k)^{\alpha+1-\varepsilon}(1+x_k)^{\beta+1-\varepsilon}.$$

PROOF. Since $-\alpha-1+\varepsilon<0$, $-\beta-1+\varepsilon<0$ hence

$$\left[(1-x)^{-\alpha-1+\varepsilon}(1+x)^{-\beta-1+\varepsilon}\right]^{(2n)}>0\qquad |x|<1$$

(see [9]). Denote $F(x)=(1-x)^{-\alpha-1+\varepsilon}(1+x)^{-\beta-1+\varepsilon}$. Then

$$(4)\qquad F(x)\geqq H_n(F,x)\qquad |x|<1.$$

Indeed, the function $F(x)-H_n(F,x)$ is positive near ± 1 and it has twofold zeros at x_k, $1\leqq k\leqq n$. If it admits also negative values then there must exist a $2n+1$-th zero. Using repeatedly the Rolle-theorem we get that the $2n$-th derivative $[F(x)-H_n(F,x)]^{(2n)}=[F(x)]^{(2n)}$ has a zero; this contradiction proves (4). Multiplying by $(1-x)^{\alpha}(1+x)^{\beta}$ and integrating (4) and using the orthogonality relation

$$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}(x-x_k)l_k^2(x)dx=0$$

we get

$$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}(1-x_k)^{-\alpha-1+\varepsilon}(1+x_k)^{-\beta-1+\varepsilon}l_k^2(x)dx\leqq\int_{-1}^{1}(1-x^2)^{\varepsilon-1}dx\leqq c.$$

Consequently

a) $$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\left|\sum_{k=1}^{n} f(x_k)l_k^2(x)\right|dx \leqq$$
$$\leqq \max_{1\leqq k\leqq n}|f(x_k)|(1-x_k)^{\alpha+1-\varepsilon}(1+x_k)^{\beta+1-\varepsilon}\cdot$$
$$\cdot\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}(1-x_k)^{-\alpha-1+\varepsilon}(1+x_k)^{-\beta-1+\varepsilon}l_k^2(x)dx \leqq$$
$$\leqq c \max_{1\leqq k\leqq n}|f(x_k)|(1-x_k)^{\alpha+1-\varepsilon}(1+x_k)^{\beta+1-\varepsilon} =: cA_n.$$

b) $$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\left|\sum_{k=1}^{n} f(x_k)\tfrac{1-x}{1-x_k}l_k^2(x)\right|dx \leqq$$
$$\leqq A_n\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}(1-x_k)^{-\alpha-1+\varepsilon}(1+x_k)^{-\beta-1+\varepsilon}\tfrac{1-x}{1-x_k}l_k^2(x)dx =$$
$$= A_n\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}(1-x_k)^{-\alpha-1+\varepsilon}(1+x_k)^{-\beta-1+\varepsilon}\tfrac{1-x_k}{1-x_k}l_k^2(x)dx \leqq cA_n.$$

Since in the Jacobi case we have

$$\left|1-\frac{w''(x_k)}{w'(x_k)}(x-x_k)\right| \leqq c\left(1+\frac{1-x}{1-x_k}\right), \qquad |x-x_k|\leqq 2$$

hence (3) follows from a) and b).

THEOREM 1. *Let* $\alpha,\beta>-1$, $0<\varepsilon<\min(\alpha+1,\beta+1)$ *and* $f,f'\in$ $\in C(\alpha+1-\varepsilon,\beta+1-\varepsilon)$. *Then*

(5) $$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|f(x)-H_n(f,x)|dx \to 0 \qquad (n\to\infty).$$

PROOF. By the Stone–Weierstrass theorem (see [8]) the polynomials are dense in $C(\alpha+1-\varepsilon,\beta+1-\varepsilon)$, i.e. there exist polynomials of order $\leqq n$ satisfying

(6) $$\sup_{|x|<1}|f'(x)-p_n'(x)|(1-x)^{\alpha+1-\varepsilon}(1+x)^{\beta+1-\varepsilon} \to 0 \qquad (n\to\infty).$$

We can also suppose $f(0)=p_n(0)$ and then

$$|f(x)-p_n(x)| = \left|\int_0^x (f'(t)-p_n'(t))\,dt\right| \leqq$$

$$\leqq c(1-x)^{-\alpha-1+\varepsilon}(1+x)^{-\beta-1+\varepsilon}\int_0^x |f'(t)-p_n'(t)|(1-t)^{\alpha+1-\varepsilon}(1+t)^{\beta+1-\varepsilon}dt$$

implies

$$\sup_{|x|<1} |f(x) - p_n(x)|(1-x)^{\alpha+1-\varepsilon}(1+x)^{\beta+1-\varepsilon} \to 0 \qquad (n \to \infty). \tag{7}$$

Consequently from Lemma 1 we obtain

$$\int_{-1}^{1}(1-x)^\alpha(1+x)^\beta|f(x) - H_n(f,x)|dx \leqq \int_{-1}^{1}(1-x)^\alpha(1+x)^\beta|f(x) - p_n(x)|dx +$$

$$+ \int_{-1}^{1}(1-x)^\alpha(1+x)^\beta|H_n(f-p_n,x)|dx \to 0 \qquad (n\to\infty)$$

by (6) and (7). Theorem 1 is proved.

Now introduce the normed Hermite polynomials

$$\int_{-\infty}^{\infty} h_n(x)h_k(x)e^{-x^2}dx = \delta_{n,k}.$$

Define the space

$$C(\lambda) := \{f \in C(\mathbf{R}) \colon \lim_{|x|\to\infty} f(x)e^{-\lambda x^2} = 0\}.$$

The Hermite interpolation polynomials are ([1])

$$H_n(f,x) = \sum_{k=1}^{n} f(x_k)(1 - 2x_k x + 2x_k^2)l_k^2(x) + \sum_{k=1}^{n} f'(x_k)(x - x_k)l_k^2(x).$$

LEMMA 2. *Let* $0 < \lambda < 1$. *Then*

$$\int_{-\infty}^{\infty} |H_n(f,x)|e^{-x^2}dx \leqq c \max_{1\leqq k\leqq n} |f(x_k)|e^{-\lambda x_k^2} + c \max_{1\leqq k\leqq n} |f'(x_k)|e^{-\lambda x_k^2}. \tag{8}$$

PROOF. From the trivial inequality $(e^{\lambda x^2})^{(2n)} > 0$, $x \in \mathbf{R}$ we get as in Lemma 1 that

$$e^{\lambda x^2} \geqq H_n(e^{\lambda t^2}, x) \qquad x \in \mathbf{R}.$$

Multiplying by e^{-x^2} and integrating we get

$$\int_{-\infty}^{\infty} e^{-x^2}\sum_{k=1}^{n} e^{\lambda x_k^2} l_k^2(x)dx \leqq \int_{-\infty}^{\infty} e^{-(1-\lambda)x^2}dx \leqq c$$

hence

a) $\int\limits_{-\infty}^{\infty} e^{-x^2}\left|\sum\limits_{k=1}^{n} f(x_k)l_k^2(x)\right|dx \leqq c \max\limits_{1\leqq k\leqq n} |f(x_k)|e^{-\lambda x_k^2} =: cB_n.$

b) $\int\limits_{-\infty}^{\infty} e^{-x^2}\left|\sum\limits_{k=1}^{n} f(x_k)x_k^2 l_k^2(x)\right|dx \leqq c \max\limits_{1\leqq k\leqq n} |f(x_k)|e^{-\lambda x_k^2} \int\limits_{-\infty}^{\infty} e^{-x^2}\sum\limits_{k=1}^{n} e^{\lambda' x_k^2} l_k^2(x)dx \leqq$

$\leqq cB_n$ if $\lambda, \lambda' < 1$.

c) $\int\limits_{-\infty}^{\infty} e^{-x^2}\left|\sum\limits_{k=1}^{n} f(x_k)x l_k^2(x)dx\right| \leqq cB_n \int\limits_{-\infty}^{\infty} |x|e^{-x^2} \sum\limits_{k=1}^{n} e^{\lambda x_k^2} l_k^2(x)dx \leqq$

$\leqq cB_n \int\limits_{-\infty}^{\infty} |x|e^{-(1-\lambda)x^2}dx \leqq cB_n.$

Lemma 2 follows from a), b), c).

THEOREM 2. *Let* $0<\lambda<1$ *and* $f, f' \in C(\lambda)$, *then*

$$\int_{-\infty}^{\infty} e^{-x^2}|f(x) - H_n(f,x)|dx \to 0 \qquad (n\to\infty). \tag{9}$$

PROOF. There exist polynomials p_n of order $\leqq n$ such that

$$\sup_{x\in\mathbf{R}} |f'(x) - p_n'(x)|e^{-\lambda x^2} \to 0 \qquad (n\to\infty). \tag{10}$$

We can further suppose that $f(0) = p_n(0)$ and then

$$|f(x) - p_n(x)| = \left|\int_0^x (f'(t) - p_n'(t))\,dt\right| \leqq o(1)\int_0^x e^{\lambda t^2}dt = o(1)e^{\lambda x^2}$$

i.e.

$$\sup_{x\in\mathbf{R}} |f(x) - p_n(x)|e^{-\lambda x^2} \to 0 \qquad (n\to\infty). \tag{11}$$

On the other hand

$$\int_{-\infty}^{\infty} e^{-x^2}|f(x) - p_n(x)|dx = o(1)\int_{-\infty}^{\infty} e^{-(1-\lambda)x^2}dx \to 0 \qquad (n\to\infty) \tag{12}$$

and then Lemma 2, (10), (11) and (12) imply

$$\int_{-\infty}^{\infty} e^{-x^2}|f(x) - H_n(f,x)|dx \leqq$$

$$\leqq \int_{-\infty}^{\infty} e^{-x^2}|f(x) - p_n(x)|dx + \int_{-\infty}^{\infty} e^{-x^2}|H_n(f-p_n,x)|dx \to 0 \quad (n\to\infty).$$

Theorem 2 is proved.

Now consider the Laguerre polynomials $l_n^{(\alpha)}(x)$, $\alpha > -1$ defined by

$$\int_0^\infty x^\alpha e^{-x} l_n^{(\alpha)}(x) l_k^{(\alpha)}(x)dx = \delta_{n,k}.$$

Define the space

$$C^\lambda(-\alpha+\varepsilon) := \left\{ f \in C(\mathbf{R}^+) : \lim_{x\to\infty} f(x)x^{\alpha-\varepsilon}e^{-\lambda x} = 0, \lim_{x\to 0+} f(x)x^{\alpha-\varepsilon}e^{-\lambda x} = 0 \right\}.$$

The Hermite interpolation polynomials are ([7])

$$H_n(f,x) = \sum_{k=1}^n f(x_k)\frac{x_k(x_k-\alpha)+x(\alpha+1-x_k)}{x_k} l_k^2(x) + \sum_{k=1}^n f'(x_k)(x-x_k)l_k^2(x).$$

LEMMA 3. *Let* $-1 < \varepsilon < \alpha$, $0 < \lambda < 1$. *Then*

(13)

$$\int_0^\infty x^\alpha e^{-x}|H_n(f,x)|dx \leqq c \max_{1\leqq k\leqq n} |f(x_k)|x_k^{\alpha-\varepsilon}e^{-\lambda x_k} + c \max_{1\leqq k \leqq n} |f'(x_k)|x_k^{\alpha-\varepsilon}e^{-\lambda x_k}.$$

PROOF. Since $\alpha - \varepsilon > 0$ hence

$$\left(\frac{e^{\lambda x}}{x^{\alpha-\varepsilon}}\right)^{(2n)} > 0 \qquad (x > 0)$$

(see [11]). Denote $F(x) = \frac{e^{\lambda x}}{x^{\alpha-\varepsilon}}$. Then we get as in Lemma 1 that

$$F(x) \geqq H_n(F,x) \qquad (x > 0).$$

Multiplying by $x^\alpha e^{-x}$ and integrating we get

$$\int_0^\infty x^\alpha e^{-x} \sum_{k=1}^n e^{\lambda x_k} x_k^{-\alpha+\varepsilon} l_k^2(x)dx \leqq \int_0^\infty x^\varepsilon e^{-(1-\lambda)x}dx \leqq c.$$

Hence

$$\int_0^\infty x^\alpha e^{-x}\left|\sum_{k=1}^n f(x_k)l_k^2(x)\right|dx \leqq$$

$$\leqq \max_{1\leqq k\leqq n} |f(x_k)|x_k^{\alpha-\varepsilon}e^{-\lambda x_k} \int_0^\infty x^\alpha e^{-x} \sum_{k=1}^n x_k^{-\alpha+\varepsilon} e^{\lambda x_k} l_k^2(x)dx \leqq$$

$$\leqq c \max_{1\leqq k\leqq n} |f(x_k)|x_k^{\alpha-\varepsilon}e^{-\lambda x_k}.$$

From this Lemma 3 follows with similar methods as in proof of Lemma 2.

THEOREM 3. *Let* $-1<\varepsilon<\alpha$, $0<\lambda<1$ *and* $f, f'\in C^\lambda(-\alpha+\varepsilon)$. *Then*

$$\int_0^\infty x^\alpha e^{-x}|f(x)-H_n(f,x)|dx\to 0 \qquad (n\to\infty). \tag{14}$$

PROOF. By the Stone–Weierstrass theorem there exist polynomials p_n of order $\leqq n$ such that

$$\sup_{x>0}|f'(x)-p_n'(x)|x^{\alpha-\varepsilon}e^{-\lambda x}\to 0 \qquad (n\to\infty). \tag{15}$$

We can suppose that $f(1)=p_n(1)$ and then

$$|f(x)-p_n(x)|=\left|\int_1^x (f'(t)-p_n'(t))\,dt\right|=o(1)\left|\int_1^x t^{\varepsilon-\alpha}e^{\lambda t}dt\right|=o(1)x^{\varepsilon-\alpha}e^{\lambda x},$$

i.e.

$$\sup_{x>0}|f(x)-p_n(x)|x^{\alpha-\varepsilon}e^{-\lambda x}\to 0 \qquad (n\to\infty). \tag{16}$$

On the other hand

$$\int_0^\infty x^\alpha e^{-x}|f(x)-p_n(x)|dx=o(1)\int_0^\infty x^\varepsilon e^{-(1-\lambda)x}dx\to 0 \qquad (n\to\infty) \tag{17}$$

and then Lemma 3, (15), (16) and (17) implies

$$\int_0^\infty x^\alpha e^{-x}|f(x)-H_n(f,x)|dx\leqq\int_0^\infty x^\alpha e^{-x}|f(x)-p_n(x)|dx+$$
$$+\int_0^\infty x^\alpha e^{-x}|H_n(f-p_n,x)|dx\to 0 \qquad (n\to\infty).$$

Theorem 3 is proved.

At last consider the case of the Sonin–Markov polynomials. The orthogonal polynomials $H_n^{(\beta)}(x)$ associated with the weight function $w(x)=e^{-x^2}|x|^\beta$ $(\beta>-1)$ are called Sonin–Markov polynomials and are generalizations of the Hermite polynomials $H_n(x)=H_n^{(0)}(x)$ ([1], [10]).

We investigate only the case $n=2m$. Let $\{x_i\}_{i=1}^{2m}$ be the zeros of the Sonin–Markov polynomial $H_{2m}^{(\beta)}(x)$. Then

$$\frac{w'(x_k)}{w'(x_k)}=\frac{2x_k^2-\beta}{x_k} \qquad (k=1,\dots,2m).$$

Define the space

$$C_{*}^{\lambda}(-\beta+\varepsilon)=\left\{f\in C(\mathbf{R}):\lim_{|x|\to\infty} f(x)|x|^{\beta-\varepsilon}e^{-\lambda x^2}=0\right\}.$$

LEMMA 4. *Let* $-1<\varepsilon<\beta$, $0<\lambda<1$. *Then*

$$\begin{aligned}(18)\qquad \int_{-\infty}^{\infty}|x|^{\beta}e^{-x^2}|H_{2m}(f,x)|dx \leqq c\max_{1\leqq k\leqq 2m}|f(x_k)||x_k|^{\beta-\varepsilon}e^{-\lambda x_k^2}+\\ +c\max_{1\leqq k\leqq 2m}|f'(x_k)||x_k|^{\beta-\varepsilon}e^{-\lambda x_k^2}.\end{aligned}$$

PROOF. It is known that

$$\frac{e^{\lambda x^2}}{|x|^{\beta-\varepsilon}}-\sum_{i=1}^{2m}\frac{e^{\lambda x_i^2}}{|x_i|^{\beta-\varepsilon}}l_i^2(x)\geqq 0,$$

see [11]. From this we obtain

$$\int_{-\infty}^{\infty}|x|^{\beta}e^{-x^2}\left|\sum_{k=1}^{2m}f(x_k)l_k^2(x)\right|dx\leqq$$

$$\max_{1\leqq k\leqq 2m}|f(x_k)||x_k|^{\beta-\varepsilon}e^{-\lambda x_k^2}\int_{-\infty}^{\infty}|x|^{\beta}e^{-x^2}\sum_{k=1}^{2m}|x_k|^{\varepsilon-\beta}e^{\lambda x_k^2}l_k^2(x)dx\leqq$$

$$\leqq c\max_{1\leqq k\leqq 2m}|f(x_k)||x_k|^{\beta-\varepsilon}e^{-\lambda x_k^2}\int_{-\infty}^{\infty}|x|^{\varepsilon}e^{-(1-\lambda)x^2}dx\leqq$$

$$\leqq c\max_{1\leqq k\leqq 2m}|f(x_k)||x_k|^{\beta-\varepsilon}e^{-\lambda x_k^2}.$$

From this Lemma 4 is following with similar methods as in proof of Lemma 2.

THEOREM 4. *Let* $-1<\varepsilon<\beta$, $0<\lambda<1$ *and* $f,f'\in C_{*}^{\lambda}(-\beta+\varepsilon)$. *Then*

$$(19)\qquad \int_{-\infty}^{\infty}|x|^{\beta}e^{-x^2}|f(x)-H_{2m}(f,x)|dx\to 0 \qquad (m\to\infty).$$

PROOF. By the Stone–Weierstrass theorem there exist polynomials of order $\leqq 2m$ satisfying

$$\sup_{x\in\mathbf{R}} |f'(x)-p'_{2m}(x)||x|^{\beta-\varepsilon}e^{-\lambda x^2}\to 0 \qquad (m\to\infty). \tag{20}$$

We can also suppose $f(0)=p_{2m}(0)$ and then

$$|f(x)-p_{2m}(x)|=\left|\int_0^x (f'(t)-p'_{2m}(t))\,dt\right|=o(1)\left|\int_0^x |t|^{\varepsilon-\beta}e^{\lambda t^2}dt\right|=o(1)|x|^{\varepsilon-\beta}e^{\lambda x^2},$$

i.e.

$$\sup_{x\in\mathbf{R}} |f(x)-p_{2m}(x)||x|^{\beta-\varepsilon}e^{-\lambda x^2}\to 0 \qquad (m\to\infty). \tag{21}$$

On the other hand
(22)

$$\int_{-\infty}^{\infty} |x|^\beta e^{-x^2}|f(x)-p_{2m}(x)|dx=o(1)\int_{-\infty}^{\infty}|x|^\varepsilon e^{-(1-\lambda)x^2}dx\to 0 \quad (m\to\infty)$$

and then Lemma 4, (20), (21) (22) implies

$$\int_{-\infty}^{\infty} |x|^\beta e^{-x^2}|f(x)-H_{2m}(f,x)|dx\leqq\int_{-\infty}^{\infty}|x|^\beta e^{-x^2}|f(x)-p_{2m}(x)|dx+$$

$$+\int_{-\infty}^{\infty}|x|^\beta e^{-x^2}|H_{2m}(f-p_{2m},x)|dx\to 0 \qquad (m\to\infty).$$

Theorem 4 is proved.

REMARK. Let

$$L_n(f,x)=\sum_{k=1}^n f(x_k)l_k(x)$$

be the Lagrange interpolation. Then we have for $f\in C\left(\frac{\alpha+1-\varepsilon}{2},\frac{\beta+1-\varepsilon}{2}\right)$, $\alpha,\beta>-1$ and $0<\varepsilon<\min(\alpha+1,\beta+1)$

$$\int_{-1}^{1}(1-x)^\alpha(1+x)^\beta|f(x)-L_n(f,x)|^2dx\to 0 \qquad (n\to\infty). \tag{23}$$

Indeed, take polynomials p_n such that

$$\sup_{|x|<1} |f(x)-p_n(x)|(1-x)^{\frac{\alpha+1-\varepsilon}{2}}(1-x)^{\frac{\beta+1-\varepsilon}{2}} \to 0 \qquad (n\to\infty).$$

Using the orthogonality relation

$$\int_{-1}^{1} l_k(x)l_j(x)(1-x)^{\alpha}(1+x)^{\beta}dx = 0 \qquad (k \neq j)$$

we obtain

$$\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|f(x)-L_n(f,x)|^2dx \leqq 2\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|f(x)-p_n(x)|^2dx+$$

$$+2\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|L_n(f-p_n,x)|^2dx \leqq$$

$$\leqq o(1)+2\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}|f(x_k)-p_n(x_k)|^2 l_k^2(x)dx =$$

$$= o(1)\left(1+\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\sum_{k=1}^{n}(1-x_k)^{-\alpha-1+\varepsilon}(1+x_k)^{-\beta-1+\varepsilon}l_k^2(x)dx\right) = o(1)$$

as we asserted. This result extends the function class for which the L^2-convergence holds. Namely Erdős and Turán [3] proved the convergence for $f \in C[-1,1]$; the class $C\left(\frac{\alpha+1-\varepsilon}{2}, \frac{\beta+1-\varepsilon}{2}\right)$ is larger, the function $f(x)$ may grow by the order $(1-x)^{\frac{-\alpha-1+\varepsilon}{2}}$ resp. $(1+x)^{\frac{-\beta-1+\varepsilon}{2}}$ if $x\to 1$ resp. $x \to -1$. Define

$$E_n(f) := \inf_{p_n \in \Pi_n} \sup_{|x|<1} |f(x)-p_n(x)|(1-x)^{\frac{\alpha+1-\varepsilon}{2}}(1+x)^{\frac{\beta+1-\varepsilon}{2}}$$

then the above proofs give the estimates

$$\text{(5')} \qquad \int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|f(x)-H_n(f,x)|dx \leqq cE_{2n-2}(f'),$$

$$\text{(23')} \qquad \int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}|f(x)-L_n(f,x)|^2dx \leqq c\left(E_{n-1}(f)\right)^2$$

and in the case of Hermite nodes, if

$$E_n^\lambda(f) := \inf_{p_n\in\Pi_n} \sup_{x\in\mathbf{R}} |f(x)-p_n(x)|e^{-\lambda x^2}$$

then

$$\int_{-\infty}^{\infty} e^{-x^2}|f(x)-H_n(f,x)|dx \leqq cE_{2n-2}^\lambda(f'). \tag{9'}$$

REFERENCES

[1] SZEGŐ, G., *Orthogonal Polynomials*, American Math. Soc. Colloq. Publ., Vol. 23, American Math. Soc., Providence, R. I., 1959. *MR* **21** #5029

[2] NATANSZON, I. P., *Konstruktív függvénytan* [Constructive theory of functions], Akadémiai Kiadó, Budapest, 1952 (in Hungarian). *MR* **15**-306

[3] ERDŐS, P. and TURÁN, P., On interpolation, I. Quadrature- and mean-convergence in the Lagrange-interpolation, *Ann. of Math.* (2) **38** (1937), 142–155. *Zbl.* **16**, 106

[4] BALÁZS, J. and TURÁN, P., Notes on interpolation, VIII. Mean convergence in infinite intervals, *Acta Math. Acad. Sci. Hungar.* **12** (1961), 469–474. *MR* **26** #4094

[5] BALÁZS, J. and TURÁN, P., Notes on interpolation, IX. Approximate representation of Fourier transform, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 215–220. *MR* **30** #5098

[6] JOÓ, I. and KY, N. X., Answer to a problem of Paul Turán, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **31** (1988), 229–241. *MR* **90h**: 41006

[7] JOÓ, I. and SZABADOS, J., On the weighted mean convergence of interpolating processes, *Acta Math. Hungar.* **57** (1991), 151–158.

[8] STONE, M. H., A generalized Weierstrass approximation theorem, *Studies in Math.* **1**, Prentice Hall, 1962. *Math. Mag.* **21** (1948), 167–184, 237–254. *MR* **10**-255

[9] JOÓ, I., An interpolation-theoretical characterization of the classical orthogonal polynomials, *Acta Math. Acad. Sci. Hungar.*, **26** (1975), 163–169. *MR* **51** # 1201

[10] KIS, O., Lagrange interpolation with nodes at the roots of Sonin–Markov polynomials, *Acta Math. Acad. Sci. Hungar.* **23** (1972), 389–417 (in Russian). *MR* **47** #680

[11] BALÁZS, C., On an extremal problem connected with the fundamental polynomials of interpolation, *Acta Math. Acad. Sci. Hungar.* **34** (1979), 307–315. *MR* **81c**: 41002

[12] BALÁZS, C. and JOÓ, I., On the speed of mean convergence of Lagrange and Hermite interpolation based on the roots of Laguerre polynomials, *Publ. Math.* (to appear).

(Received March 29, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 403–407

GRAPHS WITH MAXIMUM NUMBER OF STAR-FORESTS

Z. FÜREDI

Abstract

Let **H** denote the vertex disjoint union of stars of $a_1, \dots, a_t$ edges. Here it is proved that if $a_i > \log_2(t+1)$ for all $1 \leqq i \leqq t$ and e is sufficiently large $(e > e_0(a_1, \dots, a_t))$, then a star-forest of e edges and t components contains the largest number of (not necessarily induced) copies of **H**. A simple construction shows that the constraint $a_i = \Omega(\log t)$ cannot be omitted.

This (partly) settles a conjecture of Noga Alon.

1. Notations, preliminaries

Let **G** and **H** be simple graphs (i.e. undirected, finite, no loops and multiple edges) without isolated vertices. In this paper we investigate $N(\mathbf{G}, \mathbf{H})$, the number of subgraphs of **G** isomorphic to **H**. For simplicity, we suppose that the *edges* of the graphs are labelled, so, e.g., $N(\mathbf{K}^n, \mathbf{K}^m) = n(n-1)\dots(n-m+1)$. Let

$$N(e, \mathbf{H}) = \max\{N(\mathbf{G}, \mathbf{H}) : |E(\mathbf{G})| = e\},$$

the maximum number of ways as **H** can be embedded as a subgraph. **G** is called maximal with respect to **H** if $N(\mathbf{G}, \mathbf{H}) = N(|E(\mathbf{G})|, \mathbf{H})$.

A *star* $\mathbf{H}(a)$ is a graph of a edges, $a+1$ vertices with a degree a. The vertex disjoint union of $\mathbf{H}(a_1), \dots, \mathbf{H}(a_t)$ is denoted by $\mathbf{H}(a_1, \dots, a_t)$, and called a star-forest of type $(a_1, \dots, a_t)$. The vector $(a_1, \dots, a_t)$ is abbreviated as $\mathbf{a}$. In this paper we always suppose that $a_i \geqq 2$ for all i, and that $t \geqq 2$, except if otherwise stated.

Alon [1] determined the order of magnitude of $N(e, \mathbf{H})$ whenever **H** is an arbitrary given graph and $e \to \infty$.

CONJECTURE 1.1 (Alon [2]). *If* $\mathbf{H}(\mathbf{a})$ *is a star-forest and* **G** *is maximal with respect to* **H**, *then* **G** *is a star-forest, too.*

He proved the case $t \leqq 2$. The aim of this paper is to prove 1.1 for a large class of additional cases.

This research was supported in part by the Hungarian National Science Foundation under Grant No. 1812.

Akadémiai Kiadó, Budapest

Denote the polynomial

$$\sum x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_t}^{a_t}$$

by $p(a_1, \dots, a_t, x_1, \dots, x_n)$ or briefly by $p(\mathbf{a},\mathbf{x})$, where $i_1, \dots, i_t$ run over all the $n(n-1)\cdots(n-t+1)$ ordered t-tuples of $\{1, 2, \dots, n\}$. Let $p(\mathbf{a}, n)$ denote

$$\max\Big\{p(\mathbf{a},\mathbf{x}) : \mathbf{x} \geqq 0 \text{ and } \sum_{i=1}^{n} x_i = 1\Big\}.$$

Finally, let $p(\mathbf{a}) = \sup_{n \geqq t} p(\mathbf{a}, n)$.

During the proof $\varepsilon_1, \varepsilon_2, \dots$ and $c_1, c_2, \dots$ denote (explicitly computable) positive constants depending only on $\mathbf{a}$.

2. An asymptotic result

THEOREM 2.1. *Suppose that $a_i \geqq 2$ for all i and $\sum a_i = A$. Then $N(e, \mathbf{H}(\mathbf{a})) = p(\mathbf{a})e^A + O(e^{A-1})$, as e tends to infinity.*

PROOF. First we show that for some $n_0 = n_0(\mathbf{a})$ one has $p(\mathbf{a}, n_0) = p(\mathbf{a}, n)$ for all $n > n_0$, whenever all $a_i \geqq 2$. Suppose that $\mathbf{x}$ is a maximum point with $\mathbf{x} > 0$. Lagrange's multiplicator method gives that

$$\frac{\partial p(\mathbf{a},\mathbf{x})}{\partial x_j} = \lambda \tag{2.1}$$

for all $1 \leqq j \leqq n$. As every term in the polynomial $(\partial/\partial x_j)p(\mathbf{a},\mathbf{x})$ has degree $A-1$ and has a factor x_j (by $a_i \geqq 2$) we obtain

$$\frac{\lambda}{x_j} = \frac{1}{x_j}\frac{\partial}{\partial x_j} p(\mathbf{a},\mathbf{x}) \leqq (\max a_i - 1)\Big(\sum x_i\Big)^{A-2} = \max_i a_i - 1$$

implying

$$\lambda \leqq x_j(\max a_i - 1). \tag{2.2}$$

On the other hand, summing λx_j for all j (2.1) gives a lower bound for λ

$$\begin{aligned} \lambda = \lambda\Big(\sum x_j\Big) &= \sum x_j \frac{\partial}{\partial x_j} p(\mathbf{a},\mathbf{x}) = (A-t)p(\mathbf{a},\mathbf{x}) \geqq \\ &\geqq (A-t)p\Big(\mathbf{a}, \Big(\frac{1}{t}, \dots, \frac{1}{t}, 0, 0, \dots, 0\Big)\Big) = (A-t)\frac{t!}{t^A}. \end{aligned} \tag{2.3}$$

If $n \geqq t^A/t!$ and $x_j \leqq 1/n$, then (2.2) and (2.3) contradict each other.

Suppose that $p(\mathbf{a}) = p(\mathbf{a}, x_1, \dots, x_n)$ where $\mathbf{x} \geqq 0$, $\sum x_i = 1$. Then the graph $\mathbf{H}(\lfloor x_1 e \rfloor, \dots, \lfloor x_n e \rfloor)$ contains $p(\mathbf{a})e^A - O(e^{A-1})$ copies of $\mathbf{H}(\mathbf{a})$.

To prove the upper bound, consider an $\mathbf{H}(\mathbf{a})$-maximal graph $\mathbf{G}$ of e edges. First we claim that there is a set $C \subset V(\mathbf{G})$, $|C| \leqq c_1$ $(= c_1(\mathbf{a}))$, such that C intersects all edges of $\mathbf{G}$ and

$$\deg_G(v) \geqq \varepsilon_1 e \tag{2.4}$$

holds for all $v \in C$ for some $\varepsilon_1 = \varepsilon_1(\mathbf{a}) > 0$.

For an edge $E \in E(\mathbf{G})$ denote its *multiplicity* by $M(E)$, i.e. the number of occasions it appears in a subgraph of $\mathbf{G}$ isomorphic to $\mathbf{H}(\mathbf{a})$. Set $M_{\max} = \max\{M(E) : E \in E(\mathbf{G})\}$, and let $\{u, v\} \in E(\mathbf{G})$ be an edge with maximal multiplicity, $M(\{u, v\}) = M_{\max}$. As $p(\mathbf{a}) \geqq t^{-A}$ we have that

$$M_{\max} > \varepsilon_2 e^{A-1} \tag{2.5}$$

holds (for all $e \geqq A$).

Consider an arbitrary edge $\{p, q\} \in E(\mathbf{G})$ and suppose that $M(\{p, q\}) < \frac{1}{3} M_{\max}$. At least $\frac{2}{3} M_{\max}$ copies of $\mathbf{H}(\mathbf{a})$ contains $\{u, v\}$ but not $\{p, q\}$. At least half of these (i.e. $\geqq M/3$) has u as a center of a star. Then delete $\{p, q\}$ from $\mathbf{G}$ and add a new edge $\{u, w\}$ where $w \notin V(\mathbf{G})$. This operation increased $N(\mathbf{G}, \mathbf{H}(\mathbf{a}))$, a contradiction. We obtained that

$$M(E) > \varepsilon_3 e^{A-1} \tag{2.6}$$

holds for all edges $E \in E(\mathbf{G})$. Denote the degrees of the end points of the edge E by d_1, d_2, and let $d = \max\{d_1, d_2\}$. Then E is contained in at most

$$\sum_{\alpha=1,2} \sum_i \binom{d_\alpha}{a_i - 1} a^{A-a_i} < 2ta^{A-1} \left(\frac{d}{e}\right)^{\min a_i - 1} \leqq 2tde^{A-2}$$

star-forests of $\mathbf{G}$. Then (2.6) implies that at least one end point of E must have degree at least $(\varepsilon_3/2t)e$, yielding (2.4).

Finally, let $\mathbf{G}'$ be the bipartite graph obtained by deleting all edges inside $C, C = \{v_1, \dots, v_n\}$, $(n \leqq c_1)$. We get

$$N(\mathbf{G}, \mathbf{H}(\mathbf{a})) \leqq N(\mathbf{G}'(\mathbf{a}), \mathbf{H}) + \binom{c_1}{2}(A - t)e^{A-1}.$$

It is quite clear that for $x_i := \deg_{\mathbf{G}'}(v_i)/e$ one has

$$N(\mathbf{G}', \mathbf{H}(\mathbf{a})) \leqq p(\mathbf{a}, \mathbf{x})e^A + O(e^{A-1})$$

yielding the desired upper bound

$$N(\mathbf{G}, \mathbf{H}(\mathbf{a})) \leqq p(\mathbf{a}, \mathbf{x})e^A + O(e^{A-1}). \tag{2.7}$$

3. An exact statement

THEOREM 3.1. *Suppose that* $a_i > \log_2(t+1)$ *for all* $1 \leqq i \leqq t$ *and* $\mathbf{G}$ *is an* $\mathbf{H}(\mathbf{a})$*-maximal graph with* e *edges. If* $e > e_0(\mathbf{a})$, *then* $\mathbf{G}$ *is the union of* t *stars.*

This is not true in general. E.g., if $\mathbf{a} = (a, a, \dots, a)$, then

$$p(\mathbf{a}(x_1, \dots, x_t)) = \frac{t!}{t^{at}} < p(\mathbf{a}, (x_1, \dots, x_{t+1})) = \frac{(t+1)!}{(t+1)^{at}}$$

whenever $a \leqq \ln(t+1)$.

The main tool of the proof of 3.1 is the following technical lemma about $p(\mathbf{a},\mathbf{x})$. This lemma will be proved in the next section.

LEMMA 3.2. *Suppose that* $a_i > \log_2(t+1)$ *for all* $1 \leqq i \leqq t$, $A = \sum a_i$. *Suppose further that* $x_1, \dots, x_n \geqq \varepsilon$ *where* $n > t$. *Then*

$$p(\mathbf{a}) \geqq p(\mathbf{a},\mathbf{x}) + \frac{\varepsilon^A}{t}.$$

PROOF OF THEOREM 3.1. As we have seen in (2.4), there is a set $C = \{v_1, \dots, v_n\} \subset V(\mathbf{G})$ of large degrees ($\geqq \varepsilon_1 e$). Denote the degree sequence of C by $x_1 e, \dots, x_n e$. Then (2.7) implies that $|p(\mathbf{a},\mathbf{x}) - p(\mathbf{a})| = O(1/e)$. Then Lemma 3.2 gives that $n = t$.

There is no edge outside C, so each component of $\mathbf{H}(\mathbf{a})$ must intersect C. Hence each edge inside C has multiplicity 0, that is, C does not contain any edge by (2.6). Finally, it is clear that all vertices outside C must be of degree exactly one.

4. The proof of Lemma 3.2

Suppose that $x_1 \geqq x_2 \geqq \dots \geqq x_n \geqq \varepsilon$. Denote the sum of all terms of $p(\mathbf{a},\mathbf{x})$ containing x_i by p_i, and let $p_{n-1,n}$ denote the sum of terms containing both x_{n-1}, x_n. As x_n is the smallest of the x_i we have that $p_n \leqq (t/n)p(\mathbf{a},\mathbf{x})$. Similarly, as x_{n-1} is the second smallest of the x_i we obtain that

$$p_{n-1,n} \leqq \frac{t-1}{n-1} p_n \leqq \frac{t-1}{t} p_n. \tag{4.1}$$

Consider the ratio of the sum of distinct terms in p_n and p_{n-1} and use (4.1). We obtain

$$p_{n-1} - p_{n-1,n} \geqq (p_n - p_{n-1,n}) \left(\frac{x_{n-1}}{x_n} \right)^a \geqq \frac{p_n}{t} \left(\frac{x_{n-1}}{x_n} \right)^a, \tag{4.2}$$

where $a = \min a_i$. Now define

$$y_i = \begin{cases} x_i & \text{for } i = 1, 2, \dots, n-2 \\ x_{n-1} + x_n & \text{for } i = n-1 \\ 0 & \text{for } i = n. \end{cases}$$

Consider $p(\mathbf{a},\mathbf{y}) - p(\mathbf{a},\mathbf{x})$. We have that the increase of p is at least

$$(-p_n - p_{n-1} + p_{n-1,n}) + (p_{n-1} - p_{n-1,n})\left(\frac{x_n + x_{n-1}}{x_{n-1}}\right)^a. \tag{4.3}$$

Using (4.2) we have that the expression in (4.3) is at least

$$-p_n + \frac{p_n}{t}\left(\frac{x_{n-1}}{x_n}\right)^a\left(\left(\frac{x_n + x_{n-1}}{x_{n-1}}\right)^a - 1\right).$$

Here the coefficient of p_n/t is $(1+c)^a - c^a$ where $c = x_{n-1}/x_n \geqq 1$. So this coefficient is at least $2^a - 1 \geqq t + 1$. This implies that

$$p(\mathbf{a}) \geqq p(\mathbf{a}, n-1) \geqq p(\mathbf{a},\mathbf{y}) \geqq p(\mathbf{a},\mathbf{x}) + \frac{p_n}{t} \geqq p(\mathbf{a},\mathbf{x}) + \frac{\varepsilon^A}{t}.$$

Remarks, problems

It is probably not too difficult to give an asymptotic formula like in Theorem 2.1 for all $\mathbf{H}(\mathbf{a})$, when some $a_i = 1$ appear.

Another step to prove Conjecture 1.1 would be to get rid of the constraint $a_i > \log_2(t+1)$ in Theorem 3.1. It is easy to prove that if all $a_i \geqq 3$, then in a $\mathbf{H}(\mathbf{a})$-maximal $\mathbf{G}$ all the vertices outside C (see (2.4)) have degree 1.

It also seems to me a solvable question to investigate $N(\mathbf{G}, \mathbf{H})$ where now $\mathbf{G}$ and $\mathbf{H}$ are *multigraphs.*

REFERENCES

[1] ALON, N., On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math.* **38** (1981), 116–130. *MR* **82b**: 05078

[2] ALON, N., On the number of certain subgraphs contained in graphs with a given number of edges, *Israel J. Math.* **53** (1986), 97–120. *MR* **87j**: 05090

(Received February, 1989, in revised form September, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 409–432

ON THE ESTIMATE $(x_{\min} + x_{\max})/2$

I. JOÓ and S. SZABÓ

In practice often occurs the following problem. We have to measure some quantity and the results of n independent measurements are $x_1, \ldots, x_n$ real numbers. Give a "good" approximation for the quantity considered, using the values $x_1, \ldots, x_n$. Usually we consider $n^{-1}(x_1 + \ldots + x_n)$ as an approximation. Some critical remarks concerning this are given in [1]. Another possible approximation is e.g. $2^{-1}(x_{\min} + x_{\max})$ where $x_{\min} = \min(x_1, \ldots, x_n)$ and $x_{\max} = \max(x_1, \ldots, x_n)$. This approximation is very sensitive with respect to the errors, hence it is improbable that for $n \to \infty$ the exactness of the estimate increases. This estimate is investigated e.g. in [1], [2]. We measure the exactness of an estimate by the "interquantil halflength" — which is in usual notations $2^{-1}(Q_{3/4} - Q_{1/4})$. It is possible to prove that if ξ is a random variable with symmetric distribution, then $2^{-1}(Q_{3/4} - Q_{1/4}) \leqq \sqrt{2}D(\xi)$ if $D(\xi)$ exists (see e.g. [3], p. 309, formula (2)).

It is known that if ξ has uniform distribution, then the estimate $2^{-1}(x_{\min} + x_{\max})$ is better[1] than $n^{-1}(x_1 + \ldots + x_n)$. The aim of the present paper is to investigate the exactness of the estimate in the title and compare with that of the arithmetic mean. It is known ([5]) that the probability distribution of $(x_{\min}, x_{\max})$ is

$$f(x,y) = n(n-1)[F(y) - F(x)]^{n-2} f(x) f(y), \quad y \geqq x$$

where f resp. F denotes the density resp. distribution function of the random variable ξ. Hence one can obtain the distribution function $G(z)$ of $2^{-1}(x_{\min} + x_{\max})$ as follows:

$$G(z) = n \int_{-\infty}^{z} [F(2z - x) - F(x)]^{n-1} f(x)\,dx.$$

1980 *Mathematics Subject Classification* (1985 Revision). Primary 62F35; Secondary 62H10.

Key words and phrases. Distribution function, density function.

[1] I.e. $Q_{3/4} - Q_{1/4}$ tends to zero faster.

Akadémiai Kiadó, Budapest

We are looking for the solution — $y := Q(n)$ — of the equation

$$n \int_{-\infty}^{Q(n)} [F(2Q(n)-x) - F(x)]^{n-1} f(x)dx = 3/4 \tag{1}$$

where F resp. f denotes the distribution resp. density function of a symmetrical random variable, i.e. $F(-x) = 1 - F(x)$, but our method works also for non-symmetrical case. Important non-symmetrical distributions are e.g. the log-normal and beta-distribution, we return to this case in a next paper.

We investigate here the following types

1. $F(x) = 1 - e^{-w(x)}$ where
   a) $w(x)$ is polynomial (Theorem 3, Case I), e.g. $F(x) = 1 - 0.5e^{-x}$; $\alpha = \beta = c_5 = 1$, $a = \log 2$;
   b) $w(x)$ is logarithmical (Theorem 3, Case II), e.g. $F(x) = 1 - \frac{1}{1+x}$, $\gamma = \beta = c_6 = \delta = 1$;
   c) $w(x)$ has finite support (Theorem 3, Cases III, IV, V);

2. $f(x)$ has the form
   a) $f(x) = d_1 x^\delta e^{-\gamma x^\beta}$ (Theorem 3, Case VI), e.g. $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$, $d_1 = \frac{1}{\sqrt{2\pi}}$, $\delta = 0$, $\gamma = 1$, $\beta = 2$;
   b) $f(x) = d_2 \frac{x^\varepsilon}{(\alpha x^\beta + \gamma)^{\frac{\delta+\varepsilon+2}{\beta}}}$ (Theorem 3, Case VII), e.g. $f(x) = \frac{1}{\pi(1+x^2)}$ $d_2 = \frac{1}{\pi}$, $\varepsilon = 0$, $\alpha = \gamma = 1$, $\beta = 2$, $\delta = 0$.

We shall prove the following theorems.

THEOREM 1. *Let F be any distribution function of the form $F(x) = 1 - e^{-w(x)}$, $x \geqq 0$ such that $F(-x) = 1 - F(x)$, $(x \leqq 0)$, $w(0) = \ln 2$, $w'(x) > 0$, $(\ln f(x))'' \geqq 0$ $(x > 0)$, where $f(x) = F'(x)$. Let $\alpha(n)$, $\beta(n)$ be monotone increasing sequences, suppose $\alpha(n) = o(n)$ and let τ_1 and τ_2 be any real numbers such that $F(\tau_1) = 1 - \frac{\alpha(n)}{n}$ and $F(\tau_2) = 1 - \frac{1}{n\beta(n)}$. Then*

$$F'(\tau_1 + 2y) \leqq \frac{F'(\tau_1)}{3}\left[1 + O(e^{-\alpha(n)}) + O\left(\frac{1}{\beta(n)}\right)\right],$$

$$F'(\tau_2 + 2y) \geqq \frac{F'(\tau_2)}{3}\left[1 + O(e^{-\alpha(n)}) + O\left(\frac{1}{\beta(n)}\right)\right].$$

THEOREM 2. *If $(\ln f(x))'' \leqq 0$, $(x > 0)$ but the other assumptions of Theorem 1 are fulfilled, then we have*

$$F'(\tau_1 + 2y) \geqq \frac{F'(\tau_1)}{3}\left[1 + O(e^{-\alpha(n)}) + O\left(\frac{1}{\beta(n)}\right)\right],$$

$$F'(\tau_2 + 2y) \leqq \frac{F'(\tau_2)}{3}\left[1 + O(e^{-\alpha(n)}) + O\left(\frac{1}{\beta(n)}\right)\right].$$

REMARK. As we shall see in the proofs, in Theorems 1 and 2 it is enough to assume the desired properties of w only for $x \geqq x_1 > 0$.

From Theorems 1 and 2 we get the following Theorem 3 where the parameters are chosen so that in every case the resulting function F is a probability distribution function:

THEOREM 3. I. *If* $w(x) = c_5(x^\beta + a)^\alpha$ $(x \geqq x_1 > 0)$, $c_5, \alpha, \beta > 0$, $a \in \mathbf{R}$ *then*

$$y = \begin{cases} \frac{\log 3}{2\alpha\beta c_5^{1/\alpha\beta}}(\log 3n)^{\frac{1}{\alpha}-1}\left[(\log 3n)^{\frac{1}{\alpha}} - ac_5^{\frac{1}{\alpha}}\right]^{\frac{1}{\beta}-1} + O\left(\frac{\log\log n}{(\log n)^{2-\frac{1}{\alpha\beta}}}\right), & \alpha > 1 \\ \frac{\log 3}{2\alpha\beta c_5^{1/\alpha\beta}}(\log 3n)^{\frac{1}{\alpha\beta}-1} + O\left(\frac{\log\log n}{(\log n)^{2-\frac{1}{\alpha\beta}}}\right), & 0 < \alpha \leqq 1. \end{cases}$$

II. *If* $w(x) = c_6 \log^\gamma(x^\beta + \delta)$ $(x \geqq x_1 > 0)$, $c_6, \beta, \gamma > 0$, $\delta \in \mathbf{R}$ *then*

$$y = \begin{cases} \frac{1}{\beta c_6}\log n + O(1), & \gamma = 1 \\ \frac{1}{\beta}\left(\frac{1}{c_6}\log n\right)^{\frac{1}{\gamma}} + \left(\frac{1}{\gamma}-1\right)\log\log 3n + \log\left(\frac{\log 3}{2\beta\gamma c_6^{1/\gamma}}\right) + O\left(\frac{1}{(\log n)^{1-\frac{1}{\gamma}}}\right), & \gamma > 1 \\ \frac{1}{\beta}\left(\frac{1}{c_6}\log n\right)^{\frac{1}{\gamma}} + O\left((\log n)^{\frac{1}{\gamma}-1}\right), & 0 < \gamma < 1. \end{cases}$$

III. *If* $w(x) = c_7\left(\frac{1}{(h-x)^\beta} + f\right)^\alpha$ $(h > x \geqq x_1 > 0)$, $c_7, \alpha, \beta > 0$, $f \in \mathbf{R}$ *then*

$$y = \begin{cases} \frac{c_7^{\frac{1}{\alpha\beta}}}{2\alpha\beta}\log 3 \cdot (\log 3n)^{\frac{1}{\alpha}-1}\left[(\log 3n)^{\frac{1}{\alpha}} - fc_7^{\frac{1}{\alpha}}\right]^{-\frac{1}{\beta}-1} + O\left(\frac{\log\log n}{(\log n)^{2+\frac{1}{\alpha\beta}}}\right), & \alpha > 1 \\ \frac{c_7^{\frac{1}{\alpha\beta}}}{2\alpha\beta}\log 3 \cdot (\log 3n)^{-\frac{1}{\alpha\beta}-1} + O\left(\frac{\log\log n}{(\log n)^{2+\frac{1}{\alpha\beta}}}\right), & 0 < \alpha \leqq 1. \end{cases}$$

IV. *If* $w(x) = c_8\frac{1}{\log^\gamma((h-x)^\beta+1)}$ $(h > x \geqq x_1 > 0)$, $c_8, \beta, \gamma > 0$ *then*

$$y = \begin{cases} \frac{c_8^{\frac{1}{\gamma}}\log 3}{2\beta\gamma}e^{c_8^{\frac{1}{\gamma}}(\log n)^{-\frac{1}{\gamma}}}(\log 3n)^{-\frac{1}{\gamma}-1}\left[e^{c_8^{\frac{1}{\gamma}}(\log n)^{-\frac{1}{\gamma}}} - 1\right]^{\frac{1}{\beta}-1} + \\ + O\left(\frac{\log\log n}{(\log n)^{2+\frac{1}{\gamma\beta}}}\right), & \gamma > 1 \\ \frac{c_8^{\frac{1}{\gamma}}\log 3}{2\beta\gamma}(\log 3n)^{-\frac{1}{\gamma\beta}-1} + O\left(\frac{\log\log n}{(\log n)^{2+\frac{1}{\gamma\beta}}}\right), & 0 < \gamma \leqq 1. \end{cases}$$

V. *If* $w(x)=c_9\log^\gamma\left(\frac{1}{(h-x)^\beta}+\delta\right)$ $(h>x\geqq x_1,>0)$, $c_9,\beta,\gamma>0$, $\delta\in\mathbf{R}$ *then*

$$\log y=\begin{cases} -\frac{1}{c_9\beta}\log n+O(1), \ \ \gamma=1 \\ -\frac{1}{\beta}\left(\frac{1}{c_9}\log n\right)^{\frac{1}{\gamma}}+\left(\frac{1}{\gamma}-1\right)\log\log 3n+\log\left(\frac{\log 3}{2\beta\gamma c_9^{\frac{1}{\gamma}}}\right)+ \\ +O\left(\frac{\log\log n}{(\log n)^{1-\frac{1}{\gamma}}}\right), \ \ \gamma>1 \\ -\frac{1}{\beta}\left(\frac{1}{c_9}\log n\right)^{\frac{1}{\gamma}}+O\left((\log n)^{\frac{1}{\gamma}-1}\right), \ \ 0<\gamma<1. \end{cases}$$

VI. *If* $f(x)=d_1x^\delta e^{-\gamma x^\beta}$ $(x\geqq x_1>0)$, $d_1,\beta,\gamma>0$, $\delta\in\mathbf{R}$ *then*

$$y=\begin{cases} \frac{\log 3}{2\gamma}+O\left(\frac{1}{\log n}\right), \ \ \beta=1, \ \ \delta\neq 0 \\ \frac{\log 3}{2\gamma}+O(\exp(-\alpha(n))), \ \ \alpha(n)\nearrow+\infty, \ \ \alpha(n)=o(n), \ \ \beta=1, \ \ \delta=0 \\ \frac{\log 3}{2\beta\gamma^{\frac{1}{\beta}}}(\log n)^{\frac{1}{\beta}-1}+O\left(\frac{\log\log n}{(\log n)^{2-\frac{1}{\beta}}}\right), \ \ 0<\beta<1 \text{ or} \beta>1. \end{cases}$$

VII. *If* $f(x)=d_2\dfrac{x^\varepsilon}{(\alpha x^\beta+\gamma)^{\frac{\delta+\varepsilon+2}{\beta}}}$ $\qquad (x\geqq x_1>0)$, $d_2>0$,

$$\begin{gathered}(\alpha=0,\ \beta>0,\ \gamma>0,\ \delta\in\mathbf{R},\ \varepsilon<-1) \quad \textit{or} \\ (\alpha>0,\ \beta>0,\ \gamma\in\mathbf{R},\ \delta>-1,\ \varepsilon\in\mathbf{R}) \quad \textit{or} \\ (\alpha\in\mathbf{R},\ \beta<0,\ \gamma>0,\ \delta\in\mathbf{R},\ \varepsilon<-1)\end{gathered}$$

then

$$\log y=\begin{cases} \frac{1}{|\varepsilon+1|}\log n+O(1), \ \ \begin{cases}\alpha=0,\ \beta>0,\ \gamma>0,\ \delta\in\mathbf{R},\ \varepsilon<-1\\ \alpha\in\mathbf{R},\ \beta<0,\ \gamma>0,\ \delta\in\mathbf{R},\ \varepsilon<-1\end{cases} \\ \frac{1}{\delta+1}\log n+O(1), \ \ \ \alpha>0,\ \beta>0,\ \gamma\in\mathbf{R},\ \delta>-1,\ \varepsilon\in\mathbf{R}. \end{cases}$$

Applying the estimate $n^{-1}(x_1+\ldots+x_n)$ we obtain $y=c/\sqrt{n}$ (using the results of [6], p. 120). It follows that we can obtain better estimate than this one only in the case when f has finite support, e.g. in case V, if $\gamma\leqq 1$.

Remark that for the mentioned special cases we obtain:

1/a) $F(x)=1-0,5e^{-x}$ then

$$y=\frac{Q_{3/4}-Q_{1/4}}{2}=\frac{\log 3}{2}+O\left(\frac{\log\log n}{\log n}\right).$$

1/b) $F(x)=1-\frac{1}{1+x}$ then

$$y=\frac{Q_{3/4}-Q_{1/4}}{2}=\log n+O(1).$$

2/a) $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2}$ then

$$y=\frac{Q_{3/4}-Q_{1/4}}{2}=\frac{\log 3}{4}(\log n)^{-1/2}+O\Big(\frac{\log\log n}{(\log n)^{3/2}}\Big).$$

2/b) $f(x)=\frac{1}{\pi(1+x^2)}$ then

$$y=\frac{Q_{3/4}-Q_{1/4}}{2}=e^{\log n+O(1)}.$$

We prove only the case V; the proof of the other cases is similar. We need some lemmas.

We are looking for the solution of the equation (1). First we prove that this solution $Q(n)$ is positive. Suppose indirectly that $Q(n)\leqq 0$; then

$$3/4=n\int_{-\infty}^{Q(n)}[F(2Q(n)-x)-F(x)]^{n-1}f(x)dx\leqq$$

$$\leqq n\int_{-\infty}^{0}[F(-x)-F(x)]^{n-1}f(x)dx=1/2.$$

This contradiction proves that $Q(n)>0$, because we assumed that F is a symmetrical distribution with respect to 0, i.e.

$$F(-x)=1-F(x),\quad (x\leqq 0). \tag{2}$$

From (1) and (2) we obtain

$$n\int_{y}^{\infty}[F(x-2y)+F(x)-1]^{n-1}f(x)dx=1/4,\quad (n\geqq 1,\ y=Q(n)) \tag{3}$$

(see [2], formula (4)).

Using $\int\limits_{y}^{\infty}=\int\limits_{y}^{2y}+\int\limits_{2y}^{\infty}$ and taking into account that for $x\in[y,2y]$ $-1/2\leqq F(x-2y)+F(x)-1\leqq 1/2$ holds, we obtain

$$\Big|\int_{y}^{2y}\Big|\leqq\int_{y}^{2y}(1/2)^{n-1}f(x)dx\leqq\Big(\frac{1}{2}\Big)^{n}. \tag{4}$$

Denote $T_n(y) := \int_{2y}^{\infty} [F(x-2y)+F(x)-1]^{n-1} f(x)dx$. From (3) and (4) we obtain

$$(4n)^{-1} - 2^{-n} \leqq T_n(y) \leqq (4n)^{-1} + 2^{-n}, \quad (n \geqq 4). \tag{5}$$

First we give a simple estimate for $T_n(y)$. To this we need the following assumption on F:

$$F(x) = 1 - e^{-w(x)}, \;\; w(0) = \ln 2, \;\; w(x) \nearrow \infty \;\; (x \to \infty), \tag{6}$$

further suppose that there exist such function $\varphi_1(t)$ or $\varphi_2(t)$ (at least one of them) such that $\varphi_1, \varphi_2 \geqq 0$ and

$$w(x) - \varphi_1(t_1) \leqq w(x-t_1) \leqq w(x) - \varphi_2(t_1) \quad \text{if} \quad x \geqq t_1 \geqq 0. \tag{7}$$

According to $w(x-t) \geqq w(0)$ it is reasonable to assume that $w(0) \leqq w(x) - \varphi_1(t_1)$; hence, for $x = t_1$ we get $w(0) \leqq w(t_1) - \varphi_1(t_1) \leqq w(0)$, i.e. $\varphi_1(t_1) = w(t_1) - w(0)$. If our assumptions on F are fulfilled, then

$$1 - e^{-w(x)}\left(1 + e^{\varphi_1(2y)}\right) \leqq F(x-2y) + F(x) - 1 \leqq 1 - e^{-w(x)}\left(1 + e^{\varphi_2(2y)}\right)$$

$(x \geqq 2y)$. Here we have

$$e^{-w(x)}\left(1 + e^{\varphi_1(2y)}\right) \leqq e^{-w(2y)}(1 + e^{w(2y)-w(0)}) = e^{-w(2y)} + 0.5 \leqq 1,$$

i.e.

$$T_n(y) \begin{cases} \leqq \int_{2y}^{\infty} \left[F(x)(1+e^{\varphi_2(2y)}) - e^{\varphi_2(2y)}\right]^{n-1} f(x)dx \\ \geqq \int_{2y}^{\infty} \left[F(x)(1+e^{\varphi_1(2y)}) - e^{\varphi_1(2y)}\right]^{n-1} f(x)dx, \end{cases}$$

$$\int_{2y}^{\infty} \left[F(x)\left(1 + e^{\varphi_i(2y)}\right) - e^{\varphi_i(2y)}\right]^{n-1} f(x)dx =$$

$$= \int_{2y}^{\infty} \frac{1}{n(1+e^{\varphi_i(2y)})} \left[\left(F(x)(1+e^{\varphi_i(2y)}) - e^{\varphi_i(2y)}\right)^n\right]' dx =$$

$$= \frac{1}{n(1+e^{\varphi_i(2y)})}\left[1 - \left(F(2y)(1+e^{\varphi_i(2y)}) - e^{\varphi_i(2y)}\right)^n\right]$$

and so

$$\begin{aligned} &T_n(y) \leqq \left[n(1+e^{\varphi_2(2y)})\right]^{-1}, \\ &T_n(y) \geqq \left[n(1+e^{\varphi_1(2y)})\right]^{-1}\left[1 - \left(F(2y)(1+e^{\varphi_1(2y)}) - e^{\varphi_1(2y)}\right)^n\right]. \end{aligned} \tag{8}$$

Here we have

$$F(2y)(1+e^{\varphi_1(2y)})-e^{\varphi_1(2y)}=$$
$$=1-e^{-w(2y)}(1+e^{\varphi_1(2y)})\begin{cases}\leqq F(2y-2y)+F(2y)-1\leqq F(0)=1/2\\ \geqq 0\end{cases}$$

which means that

$$T_n(y)\geqq\left[n(1+e^{\varphi_1(2y)})\right]^{-1}(1-2^{-n}). \tag{9}$$

From (5), (8) and (9) we get for y:

$$\frac{1}{n(1+e^{\varphi_1(2y)})}\leqq(4n)^{-1}+O(2^{-n})\leqq\frac{1}{n(1+e^{\varphi_2(2y)})},\qquad(n\geqq 4). \tag{10}$$

In the special case $\varphi_1\equiv\varphi_2$ according to (7) we have $w(x)-w(t_1)+w(0)\leqq\leqq w(x-t_1)\leqq w(x)-w(t_1)+w(0)$, i.e. $w(x-t_1)=w(x)-w(t_1)+w(0)$, which is the Cauchy function equation, the solution of which (according to our assumption) is $w(x)=c_1x+w(0)$, $(x\geqq 0)$, i.e. $\varphi_1(t_1)=\varphi_2(t_1)=c_1t_1$ $(t_1\geqq 0)$. Substituting this into (10) we get

$$y=\frac{\log 3}{2c_1}+\frac{1}{c_1}O\left(\frac{n}{2^n}\right)\quad(n\geqq n_0). \tag{11}$$

Here the implicit constant is absolute.

REMARKS. 1. If only one of φ_1 or φ_2 exists, then we have in (8) and in (9) the corresponding inequality.

2. Now we give a simple sufficient condition for the existence of φ_1 resp. φ_2.

On the existence of φ_1: We saw that if φ_1 exists, then $\varphi_1(t_1)=w(t_1)--w(0)$. Using (7) it follows

$$w(x)-w(x-t_1)\leqq w(t_1)-w(0),\quad x\geqq t_1\geqq 0. \tag{12}$$

Here at $x=t_1$ we have equality. If the left-hand side decreases for $x\geqq\geqq t_1$ then (12) is fulfilled, hence it is enough to assume $w'(x)-w'(x-t_1)\leqq\leqq 0$. According to the mean value theorem we have $0\geqq w'(x)-w'(x-t_1)==t_1\cdot w''(\xi)$, $\xi\in(x,x-t_1)$ if this theorem is applicable. We obtained: for the existence of φ_1 the assumption $w''(x)\leqq 0$, $x>0$ is sufficient.

On the existence of φ_2: From (7) we get $\varphi_2(t_1)\leqq w(x)-w(x-t_1)$, $x\geqq\geqq t_1\geqq 0$, which gives at $x=t_1$ the inequality $\varphi_2(t_1)\leqq w(t_1)-w(0)$. Because of (8) and (10) we have to choose φ_2 as large as possible. Investigate: What is the largest possible φ_2, and what is a simple sufficient condition for this? For the largest possible φ_2 we have $\varphi_2(t_1)=w(t_1)-w(0)$, and in this case, according to (7) we have

$$w(t_1)-w(0)\leqq w(x)-w(x-t_1),\quad x\geqq t_1\geqq 0. \tag{13}$$

This is just the opposite of (12), and a similar argument as above shows that $w''(x) \geqq 0$, $x > 0$ is sufficient for this.

3. Suppose that $F(x) = 1$ for $x > x_0 > 0$, $x_0 = \inf\{t\colon F(t) = 1\}$. Then according to (1): $y < x_0$ and for w given in (6) we have $w(x) \nearrow \infty$ as $x \to x_0$ instead of $w(x) \nearrow \infty$, $x \to \infty$ (given in (6)). According to (7) we have $y < x_0/2$ and hence we can obtain (8) and (9).

Next assume (2) and (6) but substitute (7) with a new one, in order to obtain more exact estimate for y. We have

$$\int_{2y}^{\infty} (F(x-2y) + F(x) - 1)^{n-1}(F'(x-2y) + F'(x))dx =$$
$$= n^{-1}\left[(F(x-2y) + F(x) - 1)^n\right]_{2y}^{\infty} = n^{-1} + O(n^{-1}2^{-n}).$$

Using (7) we get

$$\tag{14} \begin{aligned} \frac{3}{4n} + O(2^{-n}) &= \int_{2y}^{\infty} (F(x-2y) + F(x) - 1)^{n-1} F'(x-2y)dx = \\ &= \int_{0}^{\infty} (F(x) + F(x+2y) - 1)^{n-1} F'(x)dx \end{aligned}$$

and we can write (7) in the following form

$$\tag{15} (4n)^{-1} + O(2^{-n}) = \int_{0}^{\infty} (F(x) + F(x+2y) - 1)^{n-1} F'(x+2y)dx.$$

Hence, it is reasonable to compare $F'(x)$ and $F'(x+2y)$. First investigate a special case $F'(x) = F'(x+2y)g(y)$, i.e. $f(x) = f(x+2y)g(y)$ where according to (6) $f(x) > 0$ $(x \geqq 0)$. Hence $f(0)/g(y) = f(2y)$, i.e. $g\left(\frac{x}{2} + y\right) = g\left(\frac{x}{2}\right)g(y)$, $g > 0$. This is the "multiplicative" Cauchy equation; its solution, according to our assumption, is $g(x) = e^{c_2 x}$, $c_2 > 0$, $x \geqq 0$, i.e. $f(x) = c_3 e^{-c_2/2\,x} = c_3 e^{-c_4 x}$; $c_3, c_4 > 0$, $x \geqq 0$. Then, from (14) and (15) we obtain

$$(4n)^{-1} + O(2^{-n}) = \int_{0}^{\infty} (F(x) + F(x+2y) - 1)^{n-1} f(x+2y)dx =$$
$$= e^{-2c_4 y} \int_{0}^{\infty} (F(x) + F(x+2y) - 1)^{n-1} f(x)dx = e^{-2c_4 y}\left(\frac{3}{4n} + O(2^{-n})\right).$$

Hence,

$$y = \frac{\log 3}{2c_4} + \frac{1}{c_4} O(n2^{-n}) \tag{16}$$

where the implicit constant is absolute.

Now return to our general problem. Let $h(x) := w'(x) - \frac{w''}{w'}(x)$, $w'(x) > 0$, $x > 0$. We investigate the following two cases separately.

Case 1. $h(x)$ is monotone decreasing (this is fulfilled if $w''(x) \leqq 0$ and $w''(x)/w'(x)$ is monotone increasing).

Case 2. $h(x)$ is monotone increasing (this is fulfilled if $w''(x) \geqq 0$ and $w''(x)/w'(x)$ is monotone decreasing).

Substitute the assumption (7) by one of these cases. We investigate only Case 1, because Case 2 is similar.

LEMMA 1. *Under the assumptions of Case 1 the function $F'(x)/F'(x+2y)$ is monotone decreasing in x ($0 < y$ is fixed).*

PROOF. The statement follows immediately from the following identities:

$$\frac{F'(x)}{F'(x+2y)} = e^{w(x+2y)-w(x)} \frac{w'(x)}{w'(x+2y)},$$

$$\frac{d}{dx}\left[\frac{F'(x)}{F'(x+2y)}\right] =$$

$$= e^{w(x+2y)-w(x)} \frac{w'(x)}{w'(x+2y)} \left[w'(x+2y) - w'(x) + \frac{w''(x)}{w'(x)} - \frac{w''(x+2y)}{w'(x+2y)}\right] =$$

$$= e^{w(x+2y)-w(x)} \frac{w'(x)}{w'(x+2y)} [h(x+2y) - h(x)]. \quad \square$$

It is easy to see that

$$w(x) = -\log(1 - F(x)) \quad \text{and hence} \quad h(x) = -\frac{f'}{f}(x). \tag{17}$$

We shall substitute the integrals in (14) and (15) by that of $\int\limits_{\tau_1}^{\tau_2}$ where we choose τ_1 and τ_2 so that $\int\limits_0^{\tau_1}, \int\limits_{\tau_2}^{\infty} = o(n^{-1})$ be fulfilled. To this let $\alpha(n)$ and $\beta(n)$ be positive increasing sequences, suppose $\alpha(n) = o(n)$ and choose τ_1 and τ_2 so that $F(\tau_1) = 1 - \frac{\alpha(n)}{n}$ and $F(\tau_2) = 1 - \frac{1}{n\beta(n)}$ be fulfilled.

LEMMA 2. *We have the estimates*

$$\int_0^{\tau_1}(F(x)+F(x+2y)-1)^{n-1}F'(x)dx = O(n^{-1}e^{-\alpha(n)}), \tag{18}$$

$$\int_{\tau_2}^{\infty}(F(x)+F(x+2y)-1)^{n-1}F'(x)dx = O\Big(\frac{1}{n\beta(n)}\Big), \tag{19}$$

(20)
$$\int_{\tau_1}^{\tau_2}(F(x)+F(x+2y)-1)^{n-1}F'(x)dx = \Big[1+O\Big(\frac{1}{\beta(n)}\Big)+O(e^{-\alpha(n)})\Big]\frac{3}{4n},$$

(21)
$$\int_{\tau_1}^{\tau_2}(F(x)+F(x+2y)-1)^{n-1}F'(x+2y)dx = \Big[1+O\Big(\frac{1}{\beta(n)}\Big)+O(e^{-\alpha(n)})\Big]\frac{1}{4n}.$$

PROOF. The estimate (18) follows immediately from the sequence of estimates:

$$\int_0^{\tau_1}(F(x)+F(x+2y)-1)^{n-1}F'(x)dx \leqq \int_0^{\tau_1}(F(x))^{n-1}F'(x)dx = n^{-1}[F^n(x)]_0^{\tau_1} =$$
$$= n^{-1}F^n(\tau_1) = n^{-1}\exp\Big[-n\log\frac{1}{1-n^{-1}\alpha(n)}\Big] \leqq n^{-1}\exp\Big[-n\frac{\alpha(n)}{n}\Big].$$

The proof of (19) follows from

$$\int_{\tau_2}^{\infty}(F(x)+F(x+2y)-1)^{n-1}F'(x)dx \leqq \int_{\tau_2}^{\infty}(F(x))^{n-1}F'(x)dx =$$
$$= n^{-1}[1-F^n(\tau_2)] = n^{-1}\Big[1-\Big(1-\frac{1}{n\beta(n)}\Big)^n\Big] =$$
$$= n^{-1}\Big[1-\exp\Big(-n\log\frac{1}{1-\frac{1}{n\beta(n)}}\Big)\Big] \leqq n^{-1}\Big[1-\exp\Big(-\frac{2}{\beta(n)}\Big)\Big] \leqq n^{-1}\frac{2}{\beta(n)}.$$

$\Big($Here we have used that $\log\frac{1}{1-\frac{1}{n\beta(n)}} \leq \frac{1}{n\beta(n)}+\frac{1}{n^2\beta^2(n)} \leqq \frac{2}{n\beta(n)}$.$\Big)$ For the proof of (20) take into consideration (14), (18) and (19). We obtain:

$$\int_{\tau_1}^{\tau_2} (F(x)+F(x+2y)-1)^{n-1}F'(x+2y)dx =$$

$$= -\left[1+O\Big(\frac{1}{\beta(n)}\Big)+O(e^{-\alpha(n)})\right]\frac{3}{4n}+$$

$$+\int_{\tau_1}^{\tau_2}(F(x)+F(x+2y)-1)^{n-1}[F'(x)+F'(x+2y)]dx =$$

$$= -\left[1+O\Big(\frac{1}{\beta(n)}+e^{-\alpha(n)}\Big)\right]\frac{3}{4n}+n^{-1}\left[(F(x)+F(x+2y)-1)^n\right]_{\tau_1}^{\tau_2}.$$

Here we have

$$(F(\tau_2)+F(\tau_2+2y)-1)^n \leqq F^n(\tau_2) = 1-(1-F^n(\tau_2)) = 1-O\Big(\frac{1}{\beta(n)}\Big)$$

further

$$(F(\tau_2)+F(\tau_2+2y)-1)^n \geqq (2F(\tau_2)-1)^n = \Big(1-\frac{2}{n\beta(n)}\Big)^n \geqq 1-O\Big(\frac{1}{\beta(n)}\Big)$$

i.e.

$$(F(\tau_2)+F(\tau_2+2y)-1)^n = 1-O\Big(\frac{1}{\beta(n)}\Big),$$

further

$$(F(\tau_1)+F(\tau+2y)-1)^n \leqq F^n(\tau_1) \leqq \exp(-\alpha(n))$$

and we have proved also (21). □

From Lemma 1 we get

$$\frac{F'(\tau_2)}{F'(\tau_2+2y)} \leqq \frac{F'(x)}{F'(x+2y)} \leqq \frac{F'(\tau_1)}{F'(\tau_1+2y)}$$

and hence, taking into consideration (20) and (21),

$$\left[1+O(e^{-\alpha(n)})+O\Big(\frac{1}{\beta(n)}\Big)\right]\frac{3}{4n} \leqq \frac{F'(\tau_1)}{F'(\tau_1+2y)}\left[1+O(e^{-\alpha(n)})+O\Big(\frac{1}{\beta(n)}\Big)\right]\frac{1}{4n}$$

and

$$\left[1+O(e^{-\alpha(n)})+O\Big(\frac{1}{\beta(n)}\Big)\right]\frac{3}{4n} \geqq \frac{F'(\tau_2)}{F'(\tau_2+2y)}\left[1+O(e^{-\alpha(n)})+O\Big(\frac{1}{\beta(n)}\Big)\right]\frac{1}{4n}$$

follow. Theorems 1 and 2 are proved.

REMARKS. 1. The assumptions on w are used only on the interval $[\tau_1,\infty)$, hence it is enough to assume the desired properties of w only for $x \geqq x_1 > 0$. According to the conditions we have $\tau_1 < \tau_2$ $(n \geqq n_0)$.

2. If $F(x)=1$, for $x \geqq x_0 > 0$ $(x_0 < +\infty)$, then the above calculations are meant in the interval $[x_1, x_0]$, $(x_0 > x_1 > 0)$. In this case $\tau_1 < \tau_2 < x_0$, $\tau_1 \to x_0$ $(n \to \infty)$ according to our assumptions further we obtain from (21): $\tau_1 + 2y < x_0$, i.e. $y < \frac{1}{2}(x_0 - \tau_1)$ and hence $y \to 0$ $(n \to \infty)$. We can say more than this if we suppose that there exists an interval I of the form $I = (x_0 - \varepsilon, x_0)$ $(\varepsilon > 0)$, where F is differentiable and $\inf\limits_I F' \geqq p > 0$ with an absolute constant p. Then by the mean value theorem we obtain: $\frac{F(x_0)-F(\tau_1)}{x_0-\tau_1} = F'(\xi)$, $\xi \in (\tau_1, x_0)$. Hence $\alpha(n)/n \cdot f(\xi) = x_0 - \tau_1$, and we can choose $\alpha(n)$ so that $y = O(1/n)$. In the case when f is monotone increasing in the interval (x_2, x_0) where $0 \leqq \leqq x_2 < x_0$ (obviously $x_2 < \tau_1$ if $n \geqq n_0$) then $\tau_2 + 2y \geqq x_0$ is fulfilled. Indeed, in the opposite case, because f is monotone increasing, hence (21) contradicts to (20). If we suppose that f is monotone increasing and F is differentiable in an interval of the form $I = (x_0 - \varepsilon, x_0)$, $(\varepsilon > 0)$, further $\sup\limits_I F' \leqq q$ $(q > 0$ is an absolute constant), then by a similar calculation as above we get $y \geqq k_1/n$ where k_1 is a constant, independent of n. (Here we have to choose $\beta(n)$ to be a constant independent of n.)

We shall often apply

LEMMA 3. *Let $x > 0$ be sufficiently large, $s > 0$ be any fixed number further a be a real number such that $a = O(x)$ if $a \geqq 0$ and $|a| \leqq qx$ if $a < 0$ where $0 < q < 1$. Then*

$$(26)\quad (a+x)^s - x^s = sa(a+x)^{s-1} + O\left(sa^2(a+x)^{s-2}\right) + O\left(s^2a^2(a+x)^{s-2}\right).$$

The implicit constants are independent of a, x, s, they depend only on the implicit constant in the condition $a = O(x)$.

PROOF. The statement follows immediately from the following estimates

$$(a+x)^s - x^s = (a+x)^s\left(1 - \left(\frac{x}{a+x}\right)^s\right),$$

$$\left(\frac{x}{x+a}\right)^s = e^{s\log x/(a+x)} = 1 + s\log\frac{x}{a+x} + s^2 O\left(\log^2\left(\frac{x}{a+x}\right)\right),$$

$$\log\frac{x}{a+x} = \log\left(1 - \frac{a}{a+x}\right) = -\frac{a}{a+x} + O\left(\left(\frac{a}{a+x}\right)^2\right). \quad \square$$

LEMMA 4. *Let $y \geqq q > 0$ where q is an absolute constant. Then*

$$(27)\quad \int\limits_y^\infty s^a e^{-s}\,ds = y^a e^{-y}(1 + O(1/y))$$

where the implicite constant depends only on a.

PROOF. Integrating by part

$$\int_y^\infty s^a e^{-s} ds = y^a e^{-y} + a \int_y^\infty s^{a-1} e^{-s} ds,$$

hence (27) follows for the case $a=0$. In the case of $a>0$ by repeated integration by part ensure that the exponent of s in the integral be $\leqq 0$ and then we have

$$\int_y^\infty s^{a-k} e^{-s} ds \leqq y^{a-k} \int_y^\infty e^{-s} ds = y^{a-k} e^{-y}.$$

Here $k>0$ is an integer such that $a-k \leqq 0$ and $a-k+1>0$.

In the case of $a<0$ we have

$$\int_y^\infty s^{a-1} e^{-s} ds \leqq y^{a-1} \int_y^\infty e^{-s} ds = y^{a-1} e^{-y}$$

and Lemma 4 is proved. □

PROOF OF THE THEOREM. We prove only the case V, because the other cases are similar.

V: $w(x) = c_9 \log^\gamma\Big(\frac{1}{(h-x)^\beta}+\delta\Big)$, $(h > x \geqq x_1 > 0)$, $c_9, \beta, \gamma > 0$, $\delta \in \mathbf{R}$. Consider the cases 1) $\gamma=1$, 2) $\gamma>1$, 3) $0<\gamma<1$ separately. For each case we have

$$\frac{1}{(h-\tau_1)^\beta} = \exp\left(\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)-\delta, \quad \frac{1}{(h-\tau_2)^\beta} = \exp\left(\left(\frac{1}{c_9}\log n\beta(n)\right)^{\frac{1}{\gamma}}\right)-\delta.$$

$$w'(x) = c_9\beta\gamma \log^{\gamma-1}\Big(\frac{1}{(h-x)^\beta}+\delta\Big)\frac{1}{(h-x)(1+\delta(h-x)^\beta)},$$

$$w'(x) > 0 \quad (h > x \geqq x_1 > 0).$$

$$w''(x) = c_9\beta\gamma\log^{\gamma-2}\Big(\frac{1}{(h-x)^\beta}+\delta\Big)\frac{1}{(h-x)^2(1+\delta(h-x)^\beta)^2}\times$$

$$\times\left[\beta(\gamma-1)+\log\Big(\frac{1}{(h-x)^\beta}+\delta\Big)\Big(\delta(\beta+1)(h-x)^\beta+1\Big)\right].$$

$$\frac{w''}{w'}(x) = \log^{-1}\Big(\frac{1}{(h-x)^\beta}+\delta\Big)\frac{1}{(h-x)(1+\delta(h-x)^\beta)}\times$$

$$\times\left[\beta(\gamma-1)+\log\Big(\frac{1}{(h-x)^\beta}+\delta\Big)(\delta(\beta+1)(h-x)^\beta+1)\right].$$

$$\left(\frac{w''}{w'}\right)'(x)=\log^{-2}\left(\frac{1}{(h-x)^{\beta}}+\delta\right)\frac{1}{(h-x)^{2}(1+\delta(h-x)^{\beta})^{2}}\times$$
$$\times\Big\{-\beta^{2}(\gamma-1)+[\beta(\gamma-1)+(\gamma-1)\beta\delta(\beta+1)(h-x)^{\beta}]\times$$
$$\times\log\left(\frac{1}{(h-x)^{\beta}}+\delta\right)+\left[1+\delta(\beta+1)(\beta+2)(h-x)^{\beta}+\delta^{2}(\beta+1)(2\beta+1)(h-x)^{2\beta}\right]\times$$
$$\times\log^{2}\left(\frac{1}{(h-x)^{\beta}}+\delta\right)\Big\}.$$

Case 1. In this case

$$h'(x)=\frac{1}{(h-x)^{2}[1+\delta(h-x)^{\beta}]^{2}}\times$$
$$\times\left(c_{9}\beta-1+\delta(\beta+1)(c_{9}\beta-\beta-2)(h-x)^{\beta}-\delta^{2}(\beta+1)(2\beta+1)(h-x)^{2\beta}\right).$$

Consider the cases i) $c_9\beta-1>0$, ii) $c_9\beta-1=0$, iii) $c_9\beta-1<0$.

i) In this case $h'(x)>0$ ($h>x\geqq x_1>0$), hence the conditions of Theorem 2 are fulfilled in the interval $[x_1,h]$. We have seen earlier that $2y+\tau_1<h$, and $2y+\tau_1>x_1$ if $n\geqq n_0$, so Theorem 2 is applicable. We obtain

$$3\left(1+O\left(\frac{1}{\beta(n)}\right)+O(e^{-\alpha(n)})\right)\geqq\frac{F'(\tau_1)}{F'(\tau_1+2y)}=\frac{\alpha(n)}{n}\times$$
$$\tag{28}\times\frac{(h-\tau_1-2y)^{\beta+1}\left(\frac{1}{(h-\tau_1-2y)^{\beta}}+\delta\right)}{(h-\tau_1)^{\beta+1}\left(\frac{1}{(h-\tau_1)^{\beta}}+\delta\right)}\exp\left(c_9\log\left(\frac{1}{(h-\tau_1-2y)^{\beta}}+\delta\right)\right)$$

Here

$$\log\left(\frac{1}{(h-\tau_1)^{\beta}}+\delta\right)=\beta\log\frac{1}{h-\tau_1}+O\left((h-\tau_1)^{\beta}\right),$$
$$\log\left(\frac{1}{(h-\tau_1-2y)^{\beta}}+\delta\right)=\beta\log\frac{1}{h-\tau_1-2y}+O\left((h-\tau_1)^{\beta}\right),$$
$$\log\log\left(\frac{1}{(h-\tau_1)^{\beta}}+\delta\right)=\log\beta+\log\log\frac{1}{h-\tau_1}+O\left((h-\tau_1)^{\beta}\right),$$
$$\log\log\left(\frac{1}{(h-\tau_1-2y)^{\beta}}+\delta\right)=\log\beta+\log\log\frac{1}{h-\tau_1-2y}+O\left((h-\tau_1)^{\beta}\right).$$

Using these, we can write (28) in the following form:

$$\log\frac{3n}{\alpha(n)}+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)+O\left((h-\tau_1)^{\beta}\right)-\log\frac{1}{h-\tau_1}\geqq$$
$$\geqq(c_9\beta-1)\log\frac{1}{h-\tau_1-2y}\left(1+O\left(\exp\left(-\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)\left(\log\frac{n}{\alpha(n)}\right)^{-1}\right)\right).$$

Hence

$$\left(\frac{3n(h-\tau_1)}{\alpha(n)}\right)^{\frac{1}{c_9\beta-1}}\exp\left(O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)+O\left(\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9}}\right)\right)\geqq\frac{1}{h-\tau_1-2y}.$$

Here

$$h-\tau_1=\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9\beta}}\left(1+O\left(\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9}}\right)\right).$$

Hence

$$2y\leqq h-\tau_1-3^{-\frac{1}{c_9\beta-1}}\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9\beta}}\left(1+O\left(\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9}}\right)\right)\times$$
$$\times\exp\left(O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)+O\left(\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9}}\right)\right).$$

Using $e^t=1+O(t)$, $0\leqq t\leqq c_0$ we can write the right-hand side as follows:
(29)

$$2y\leqq\left(1-3^{-\frac{1}{c_9\beta-1}}\right)\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9\beta}}\left(1+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)+O\left(\left(\frac{n}{\alpha(n)}\right)^{-\frac{1}{c_9}}\right)\right).$$

Now we give lower estimate for y.

Distinguish two cases: a) $y\geqq(h-\tau_2)/2$, b) $y<(h-\tau_2)/2$. In the case a) we have $\log y\geqq-\frac{1}{c_9\beta}\log n+O(\log\beta(n))$. In the case b) we can apply Theorem 2 and we obtain:

$$3\left(1+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)\right)\leqq\frac{F'(\tau_2)}{F'(\tau_2+2y)}$$

and hence, by a similar calculation as above we get

$$2y\geqq\left(1-3^{-1/(c_9\beta-1)}\right)(n\beta(n))^{-1/(c_9\beta)}\left(1+O\left(e^{-\alpha(n)}+\frac{1}{\beta(n)}\right)+O((n\beta(n))^{-1/c_9}\right).$$

According to a) and b) we obtain the following lower estimate

$$\log y\geqq-\frac{1}{c_9\beta}\log n+O(\log\beta(n)). \tag{30}$$

From (29) and (30) choosing $\alpha(n)$ and $\beta(n)$ to be appropriate constants independent on n we get

$$\log y=-(c_9\beta)^{-1}\log n+O(1). \tag{31}$$

ii) In this case

$$h'(x)=\frac{1}{(h-x)^2[1+\delta(h-x)^\beta]^2}(-\delta)(\beta+1)(h-x)^\beta\left(\beta+1-\delta(2\beta+1)(h-x)^\beta\right).$$

Consider the cases $*)$ $\delta<0$, $**)$ $\delta=0$, $***)$ $\delta>0$ separately.

$*)$ Now $h'(x)>0$ $(h>x\geqq x_1>0)$, hence the conditions of Theorem 2 are fulfilled in the interval $[x_1,h]$. We have seen earlier that $2y+\tau_1<h$ and hence we get the following upper estimate for y:

$$y<(h-\tau_1)/2. \tag{32}$$

Now we give lower estimate for y. Consider the cases a) $y\geqq(h-\tau_2)/2$, and b) $y<(h-\tau_2)/2$ separately. In the case a) we have $\log y\geqq-\log n+O(\log\beta(n))$. In case b) we can apply Theorem 2, we obtain

$$3\left(1+O(e^{-\alpha(n)})+O\Big(\frac{1}{\beta(n)}\Big)\right)\leqq\frac{F'(\tau_2)}{F'(\tau_2+2y)}=\frac{1}{n\beta(n)}\times$$
$$\times\frac{(h-\tau_2-2y)^{\beta+1}}{(h-\tau_2)^{\beta+1}}\frac{\left(\frac{1}{(h-\tau_2-2y)^\beta}+\delta\right)}{\left(\frac{1}{(h-\tau_2)^\beta}+\delta\right)}\left(\frac{1}{(h-\tau_2-2y)^\beta}+\delta\right)^{c_9}.$$

Hence, taking into consideration $c_9\beta=1$ we get

$$3n\beta(n)(h-\tau_2)\left[1+\delta(h-\tau_2)^\beta\right]\left(1+O\Big(e^{-\alpha(n)}+\frac{1}{\beta(n)}\Big)\right)\leqq$$
$$\leqq\Big(1+\delta(h-\tau_2-2y)^\beta\Big)^{c_9+1}$$

where $h-\tau_2=(n\beta(n))^{-1}\left(1+O\left((n\beta(n))^{-1/c_9}\right)\right)$. Taking into account $\log(1+t)=t+O(t^2)$, $0\leqq|t|\leqq c_0<1$ we obtain

$$\frac{\log 3}{c_9+1}+\frac{\delta(h-\tau_2)^\beta}{c_9+1}+O\Big(e^{-\alpha(n)}+\frac{1}{\beta(n)}\Big)+O\left((h-\tau_2)^{2\beta}\right)\leqq\delta(h-\tau_2-2y)^\beta,$$

i.e.

$$\frac{(h-\tau_2)^\beta}{c_9+1}+O\left((h-\tau_2)^{2\beta}\right)+\frac{\log 3}{(c_9+1)\delta}+O\Big(e^{-\alpha(n)}+\frac{1}{\beta(n)}\Big)\geqq(h-\tau_2-2y)^\beta.$$

Let $\alpha(n)$, $\beta(n)\geqq K$ where K is a sufficiently large constant, (independent of n). Then we have

$$\frac{\log 3}{(c_9+1)\delta}+O\Big(e^{-\alpha(n)}+\frac{1}{\beta(n)}\Big)<0$$

and hence from the estimate above we get

$$2y\geqq h-\tau_2-(h-\tau_2)\left[(c_9+1)^{-1}+O\left((h-\tau_2)^\beta\right)\right]^{1/\beta}.$$

For $n \geqq n_0$ we have $\frac{1}{c_9+1} + O((h-\tau_2)^\beta) \leqq L < 1$ (L is an absolute constant — independent of n). From a) and b) we obtain:

$$\log y \geqq -\log n + O(\log \beta(n)). \tag{33}$$

If $\alpha(n)$ and $\beta(n)$ are appropriately chosen constants (independent of n) then from (32) and (33) we get:

$$\log y = -\log n + O(1). \tag{34}$$

In the case (**) and (***) we have proved before that $2y+\tau_1 < h$ and hence

$$y < (h-\tau_1)/2. \tag{35}$$

Now give lower estimate for y. Obviously $f'(x) = e^{-w(x)}\left(w''(x) - [w'(x)]^2\right)$, hence

$$f'(x) = e^{-w(x)}\left(1 - c_9\beta + \delta(\beta+1)(h-x)^\beta\right)\frac{c_9\beta}{(h-x)^2\left(1+\delta(h-x)^\beta\right)^2}.$$

According to the assumptions $f'(x) \geqq 0$, $x_2 < x < h$, i.e. f is monotone increasing in the interval (x_2, h), and hence, as we have proved above $\tau_2 + 2y \geqq h$ follows, i.e.

$$y \geqq (h-\tau_2)/2. \tag{36}$$

From (35) and (36) we obtain

$$\log y = -\log n + O(1), \tag{37}$$

if $\alpha(n)$ and $\beta(n)$ are appropriate constants (independent of n).

(iii) We have proved before that $2y + \tau_1 < h$, hence

$$y < (h-\tau_1)/2. \tag{38}$$

Now we give lower estimate for y. We have in this case

$$f'(x) = e^{-w(x)}\left(1 - c_9\beta + \delta(\beta+1)(h-x)^\beta\right)\frac{c_9\beta}{(h-x)^2\left(1+\delta(h-x)^\beta\right)^2}.$$

Now, according to the assumptions $f'(x) \geqq 0$ $(x_2 < x < h)$, i.e. f is monotone increasing in the interval (x_2, h), and hence, as we have proved, $\tau_2 + 2y \geqq h$ follows, i.e.

$$y \geqq (h-\tau_2)/2. \tag{39}$$

From (38) and (39) we obtain

$$\log y = -\frac{1}{c_9\beta}\log n + O(1)$$

if $\alpha(n)$ and $\beta(n)$ is chosen appropriately (independently of n).

Case 2. In this case $h'(x) > 0$ $(h > x \geqq x_1 > 0)$, hence the conditions of Theorem 2 are fulfilled on the interval $[x_1, h]$. We have seen above that $2y + \tau_1 < h$ further $2y + \tau_1 > x_1$ if $n \geqq n_0$. We can apply Theorem 2, we obtain (41)

$$3\Big(1 + O\Big(e^{-\alpha(n)} + \frac{1}{\beta(n)}\Big)\Big) \leqq \frac{F'(\tau_1)}{F'(\tau_1 + 2y)} = \frac{\alpha(n)}{n}\frac{\log^{\gamma-1}\Big(\frac{1}{(h-\tau_1)^\beta} + \delta\Big)}{\log^{\gamma-1}\Big(\frac{1}{(h-\tau_1-2y)^\beta} + \delta\Big)}\times$$

$$\times\frac{(h-\tau_1-2y)^{\beta+1}\Big(\frac{1}{(h-\tau_1-2y)^\beta} + \delta\Big)}{(h-\tau_1)^{\beta+1}\Big(\frac{1}{(h-\tau_1)^\beta} + \delta\Big)}\exp\Big(c_9\log^\gamma\Big(\frac{1}{(h-\tau_1-2y)^\beta} + \delta\Big)\Big).$$

a) First we prove that $y/(h-\tau_1) = o(1)$.

$$\log\Big(\frac{1}{(h-\tau_1)^\beta} + \delta\Big) = \beta\log\frac{1}{h-\tau_1} + O\left((h-\tau_1)^\beta\right),$$

$$\log\Big(\frac{1}{(h-\tau_1-2y)^\beta} + \delta\Big) = \beta\log\frac{1}{h-\tau_1-2y} + O\left((h-\tau_1)^\beta\right),$$

$$\log\log\Big(\frac{1}{(h-\tau_1)^\beta} + \delta\Big) = \log\beta + \log\log\frac{1}{h-\tau_1} + O\left((h-\tau_1)^\beta\right),$$

$$\log\log\Big(\frac{1}{(h-\tau_1-2y)^\beta} + \delta\Big) = \log\beta + \log\log\frac{1}{h-\tau_1-2y} + O\left((h-\tau_1)^\beta\right),$$

$$h-\tau_1 = \exp\Big(-\frac{1}{\beta}\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Big)\Big(1+O\Big(\exp\Big(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Big)\Big)\Big).$$

Using these we can write (41) in the following form

$$\log\frac{n}{\alpha(n)} + \log 3 + O\Big(e^{-\alpha(n)} + \frac{1}{\beta(n)}\Big) +$$

$$+O\Big((h-\tau_1)^\beta\Big) - (\gamma-1)\log\log\frac{1}{h-\tau_1} - \log\frac{1}{h-\tau_1}\Big) \geqq$$

$$\geqq c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1-2y}\Big(1 + O\Big(\frac{(h-\tau_1-2y)^\beta}{|\log(h-\tau_1-2y)|}\Big)\Big) -$$

$$-(\gamma-1)\log\log\frac{1}{h-\tau_1-2y} - \log\frac{1}{h-\tau_1-2y}.$$

Here

$$\frac{(h-\tau_1-2y)^\beta}{|\log(h-\tau_1-2y)|}=O\left(\frac{(h-\tau_1)^\beta}{|\log(h-\tau_1)|}\right)=$$
$$=O\left(\exp\left(-\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)\left(\log\frac{n}{\alpha(n)}\right)^{-\frac{1}{\gamma}}\right),$$

$$\frac{\log\log\frac{1}{h-\tau_1-2y}}{\log^\gamma\frac{1}{h-\tau_1-2y}}\leqq\frac{\log\log\frac{1}{h-\tau_1}}{\log^\gamma\frac{1}{h-\tau_1}},$$

$$\log^{1-\gamma}\frac{1}{h-\tau_1-2y}\leqq\log^{1-\gamma}\frac{1}{h-\tau_1},$$

hence we get

$$\log\frac{n}{\alpha(n)}+\log 3+O\left(e^{-\alpha(n)}+\frac{1}{\beta(n)}\right)+ \tag{42}$$
$$+O\left((h-\tau_1)^\beta\right)-(\gamma-1)\log\log\frac{1}{h-\tau_1}-\log\frac{1}{h-\tau_1}\geqq$$
$$\geqq c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1-2y}\left(1+O\left(\exp\left(-\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)\left(\log\frac{n}{\alpha(n)}\right)^{-\frac{1}{\gamma}}\right)-\right.$$
$$\left.-\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1}}-\frac{1}{c_9\beta^\gamma}\log^{1-\gamma}\frac{1}{h-\tau_1}\right).$$

Let

$$x:=O\left(\exp\left(-\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)\left(\log\frac{n}{\alpha(n)}\right)^{-\frac{1}{\gamma}}\right)+$$
$$+\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1}}+\frac{1}{c_9\beta^\gamma}\log^{1-\gamma}\frac{1}{h-\tau_1}.$$

Obviously, $0<x<1/2$, $n\geqq n_1$, $x\to 0$ $(n\to\infty)$, hence from (42) we get

$$\log\frac{n}{\alpha(n)}\left(1+\frac{\log 3}{\log\frac{n}{\alpha(n)}}+O\left(\frac{\exp(-\alpha(n))}{\log\frac{n}{\alpha(n)}}+\frac{1}{\beta(n)\log\frac{n}{\alpha(n)}}+\frac{(h-\tau_1)^\beta}{\log\frac{n}{\alpha(n)}}\right)-\right.$$
$$\left.-\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{\log\frac{n}{\alpha(n)}}-\frac{\log\frac{1}{h-\tau_1}}{\log\frac{n}{\alpha(n)}}\right)\frac{1}{1-x}\geqq c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1-2y}$$

and from this, taking into account $\frac{1}{1-x}=1+\frac{x}{1-x}$ we obtain

$$\frac{1}{\beta}\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg(1+\frac{\log 3}{(1-x)\log\frac{n}{\alpha(n)}}+O\Bigg(\frac{\exp(-\alpha(n))}{\log\frac{n}{\alpha(n)}}\Bigg)+$$
$$+O\Bigg(\frac{1}{\beta(n)\log\frac{n}{\alpha(n)}}\Bigg)+O\Bigg(\frac{(h-\tau_1)^{\beta}}{\log\frac{n}{\alpha(n)}}\Bigg)+$$
$$+O\Bigg(\exp\Bigg(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)\Big(\log\frac{n}{\alpha(n)}\Big)^{-\frac{1}{\gamma}}\Bigg)+$$
$$+\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{(1-x)c_9\beta^{\gamma}\log^{\gamma}\frac{1}{h-\tau_1}}-\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{(1-x)\log\frac{n}{\alpha(n)}}+\frac{1}{(1-x)c_9\beta^{\gamma}}\log^{1-\gamma}\frac{1}{h-\tau_1}-$$
$$-\frac{\log\frac{1}{h-\tau_1}}{(1-x)\log\frac{n}{\alpha(n)}}\Bigg)\geqq\log\frac{1}{h-\tau_1-2y}.$$

Here

$$\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{(1-x)c_9\beta^{\gamma}\log^{\gamma}\frac{1}{h-\tau_1}}-\frac{(\gamma-1)\log\log\frac{1}{h-\tau_1}}{(1-x)\log\frac{n}{\alpha(n)}}=$$
$$=O\Bigg(\exp\Bigg(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)\log\log\frac{n}{\alpha(n)}\Big(\log\frac{n}{\alpha(n)}\Big)^{-1-\frac{1}{\gamma}}\Bigg),$$

$$\frac{\log^{1-\gamma}\frac{1}{h-\tau_1}}{(1-x)c_9\beta^{\gamma}}-\frac{\log\frac{1}{h-\tau_1}}{(1-x)\log\frac{n}{\alpha(n)}}=O\Bigg(\exp\Bigg(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)\Big(\log\frac{n}{\alpha(n)}\Big)^{-1}\Bigg),$$

and hence

$$2y\leqq h-\tau_1-\exp\Bigg(-\frac{1}{\beta}\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg(1+\frac{\log 3}{(1-x)\log\frac{n}{\alpha(n)}}+$$
$$+O\Bigg(\frac{\exp(-\alpha(n))}{\log\frac{n}{\alpha(n)}}+\frac{1}{\beta(n)\log\frac{n}{\alpha(n)}}\Bigg)+$$
$$+O\Bigg(\exp\Bigg(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)\Big(\log\frac{n}{\alpha(n)}\Big)^{-\frac{1}{\gamma}}\Bigg)\Bigg).$$

Hence

$$\frac{2y}{h-\tau_1} \leqq 1-\exp\left(-\frac{\log 3}{(1-x)\beta c_9^\gamma\left(\log\frac{n}{\alpha(n)}\right)^{1-\frac{1}{\gamma}}}+O\left(\frac{\exp(-\alpha(n))}{\left(\log\frac{n}{\alpha(n)}\right)^{1-\frac{1}{\gamma}}}\right)+\right.$$

$$\left.+O\left(\frac{1}{\beta(n)\left(\log\frac{n}{\alpha(n)}\right)^{1-\frac{1}{\gamma}}}\right)+O\left(\exp\left(-\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)\right)\right)\times$$

$$\times\left(1+O\left(\exp\left(-\left(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\right)\right)\right).$$

Let $\alpha(n),\ \beta(n)\geqq K$ where K is a sufficiently large constant independent of n. Then we have $y/(h-\tau_1)=O\left((\log n/\alpha(n))^{\frac{1}{\gamma}-1}\right)$, i.e. we have proved: $y/(h-\tau_1)=o(1)$.

b) Using the estimates above, we can give more exact estimates for y. From (41) we get

$$\log\frac{3n}{\alpha(n)}+O\left(e^{-\alpha(n)}+\frac{1}{\beta(n)}\right)+O\left((h-\tau_1)^\beta\right)\geqq c_9\beta^\gamma\log^\gamma\frac{1}{h-\tau_1-2y}\times$$

$$\times\left(1+O\left(\frac{(h-\tau_1)^\beta}{|\log(h-\tau_1)|}\right)\right)+\log(h-\tau_1-2y)-\log(h-\tau_1)+(\gamma-1)\log\log\frac{1}{h-\tau_1}-$$

$$-(\gamma-1)\log\log\frac{1}{h-\tau_1-2y}.$$

Here

$$\log(h-\tau_1-2y)-\log(h-\tau_1)=\log\left(1-\frac{2y}{h-\tau_1}\right)=O\left(\frac{y}{h-\tau_1}\right)=$$

$$=O\left(\left(\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}-1}\right),$$

$$\log\log\frac{1}{h-\tau_1-2y}=\log\log\frac{\frac{1}{h-\tau_1}}{1-\frac{2y}{h-\tau_1}}=\log\log\frac{1}{h-\tau_1}+\log\left(1+\frac{\log\left(1-\frac{2y}{h-\tau_1}\right)}{\log(h-\tau_1)}\right)=$$

$$=\log\log\frac{1}{h-\tau_1}+O\left(\frac{y}{h-\tau_1}\frac{1}{|\log(h-\tau_1)|}\right).$$

Using these we obtain

$$\frac{1}{\beta}\left(\frac{1}{c_9}\log\frac{3n}{\alpha(n)}\right)^{\frac{1}{\gamma}}\left(1+O\left(\frac{\exp(-\alpha(n))}{\log\frac{n}{\alpha(n)}}\right)+O\left(\frac{1}{\beta(n)\log\frac{n}{\alpha(n)}}\right)+\right.$$

$$\left.+O\left(\left(\log\frac{n}{\alpha(n)}\right)^{\frac{1}{\gamma}-2}\right)+O\left(\frac{(h-\tau_1)^\beta}{|\log(h-\tau_1)|}\right)\geqq\log\frac{1}{h-\tau_1-2y}.\right.$$

Hence

$$2y \leqq (h-\tau_1)\Bigg(1-\exp\Bigg(-\frac{1}{\beta}\Big(\frac{1}{c_9}\log\frac{3n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg[1+O\Bigg(\frac{\exp(-\alpha(n))}{\log\frac{n}{\alpha(n)}}\Bigg)+$$
$$+O\Bigg(\frac{1}{\beta(n)}\frac{1}{\log\frac{n}{\alpha(n)}}\Bigg)+O\Bigg(\Big(\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}-2}\Bigg)+O\Bigg(\frac{(h-\tau_1)^{\beta}}{|\log(h-\tau_1)|}\Bigg)\Bigg]+$$
$$+\frac{1}{\beta}\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}+O\Bigg(\exp\Bigg(-\Big(\frac{1}{c_9}\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)\Bigg)\Bigg)\Bigg).$$

According to Lemma 3

$$\Big(\log\frac{3n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}-\Big(\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}=\frac{1}{\gamma}\log 3(\log 3n)^{\frac{1}{\gamma}-1}+O\Big(\frac{\log\alpha(n)}{(\log n)^{2-\frac{1}{\gamma}}}\Big),$$
$$\exp\Bigg(-\Big(\log\frac{n}{\alpha(n)}\Big)^{\frac{1}{\gamma}}\Bigg)=\exp\Big(-(\log n)^{\frac{1}{\gamma}}\Big)\Bigg(1+O\Big(\frac{\log\alpha(n)}{(\log n)^{1-\frac{1}{\gamma}}}\Big)\Bigg),$$

so we get

$$2y\leqq\exp\Bigg(-\frac{1}{\beta}\Big(\frac{1}{c_9}\log n\Big)^{\frac{1}{\gamma}}\Bigg)\Bigg(1+O\Big(\frac{\log\alpha(n)}{(\log n)^{1-\frac{1}{\gamma}}}\Big)\Bigg)\times$$
$$\times\Bigg(1-\exp\Bigg[-\frac{\log 3}{\beta\gamma c_9^{1/\gamma}}(\log 3n)^{\frac{1}{\gamma}-1}+O\Bigg(\frac{\log\alpha(n)}{(\log n)^{2-\frac{1}{\gamma}}}\Bigg)+O\Bigg(\frac{\exp(-\alpha(n))}{\Big(\log\frac{n}{\alpha(n)}\Big)^{1-\frac{1}{\gamma}}}\Bigg)+$$
$$+O\Bigg(\frac{1}{\beta(n)\Big(\log\frac{n}{\alpha(n)}\Big)^{1-\frac{1}{\gamma}}}\Bigg)+O\Bigg(\Big(\log\frac{n}{\alpha(n)}\Big)^{\frac{2}{\gamma}-2}\Bigg)\Bigg]\Bigg).$$

Using the estimate $e^t=1+t+O(t^2)$ $|t|\leqq t_0$ we obtain the following upper estimate for y:

(43)
$$y\leqq\exp\Bigg(-\frac{1}{\beta}\Big(\frac{1}{c_9}\log n\Big)^{\frac{1}{\gamma}}\Bigg)\Bigg(\frac{\log 3}{2\beta\gamma c_9^{1/\gamma}}(\log 3n)^{\frac{1}{\gamma}-1}+O\Bigg(\frac{\exp(-\alpha(n))}{\Big(\log\frac{n}{\alpha(n)}\Big)^{1-\frac{1}{\gamma}}}\Bigg)+$$
$$+O\Bigg(\frac{1}{\beta(n)\Big(\log\frac{n}{\alpha(n)}\Big)^{1-\frac{1}{\gamma}}}\Bigg)+O\Bigg(\frac{\log\alpha(n)}{(\log n)^{2-\frac{2}{\gamma}}}\Bigg)\Bigg).$$

Now we give lower estimate for y. It follows from (43) that for $n\geqq n_1$ we have $2y+\tau_2<h$, hence we can apply Theorem 2, we get

$$3\left(1+O\left(e^{-\alpha(n)}+\frac{1}{\beta(n)}\right)\right)\leqq\frac{F'(\tau_2)}{F'(\tau_2+2y)}=\frac{1}{n\beta(n)}\times$$

$$\times\frac{\log^{\gamma-1}\left(\frac{1}{(h-\tau_2)^\beta}+\delta\right)}{\log^{\gamma-1}\left(\frac{1}{(h-\tau_2-2y)^\beta}+\delta\right)}\times\frac{(h-\tau_2-2y)^{\beta+1}\left(\frac{1}{(h-\tau_2-2y)^\beta}+\delta\right)}{(h-\tau_2)^{\beta+1}\left(\frac{1}{(h-\tau_2)^\beta}+\delta\right)}\times \tag{44}$$

$$\times\exp\left(c_9\log^\gamma\left(\frac{1}{(h-\tau_2-2y)^\beta}+\delta\right)\right).$$

Here

$$\frac{y}{h-\tau_2}=\frac{y}{h-\tau_1}\frac{h-\tau_1}{h-\tau_2}=O\left((\log n/\alpha(n))^{\frac{1}{\gamma}-1}\right).$$

Using this, a similar calculation as above gives:

$$\begin{aligned}y\geqq\exp\left(-\frac{1}{\beta}\left(\frac{1}{c_9}\log n\right)^{\frac{1}{\gamma}}\right)\Bigg(&\frac{\log 3}{2\beta\gamma c_9^{1/\gamma}}(\log 3n)^{\frac{1}{\gamma}-1}+\\ &+O\left(\frac{\exp(-\alpha(n))}{(\log n\beta(n))^{1-\frac{1}{\gamma}}}\right)+O\left(\frac{(\log n\beta(n))^{\frac{1}{\gamma}-1}}{\beta(n)}\right)+O\left(\frac{\log\beta(n)}{(\log n)^{2-\frac{2}{\gamma}}}\right)\Bigg).\end{aligned} \tag{45}$$

From (43) and (45) we get with $\alpha(n)=\beta(n)=\log n$:

$$y=\exp\left(-\frac{1}{\beta}\left(\frac{1}{c_9}\log n\right)^{\frac{1}{\gamma}}\right)\left(\frac{\log 3}{2\beta\gamma c_9^{1/\gamma}}(\log 3n)^{\frac{1}{\gamma}-1}+O\left(\frac{\log\log n}{(\log n)^{2-\frac{2}{\gamma}}}\right)\right). \tag{46}$$

Case 3. We have proved earlier that $2y+\tau_1<h$. Hence we get

$$y<(h-\tau_1)/2. \tag{47}$$

Now we give lower estimate for y. Because

$$f'(x)=e^{-w(x)}(w''(x)-(w'(x))^2)$$

where

$$k_3\log^{\gamma-1}\frac{1}{h-x}\frac{1}{(h-x)^2}\leqq w''(x)\leqq k_4\log^{\gamma-1}\frac{1}{h-x}\frac{1}{(h-x)^2},$$

$$k_5\log^{\gamma-1}\frac{1}{h-x}\frac{1}{h-x}\leqq w''(x)\leqq k_6\log^{\gamma-1}\frac{1}{h-x}\frac{1}{h-x}$$

according to the assumptions we obtain $f'(x)\geqq 0$, $x_2<x<h$, i.e. f is monotone increasing in the interval (x_2,h). Hence, as we have seen before $\tau_2+2y\geqq$ $\geqq h$ follows, i.e.

$$y\geqq(h-\tau_2)/2. \tag{48}$$

From (47) and (48) we obtain

$$\log y = -\frac{1}{\beta}\Big(\frac{1}{c_9}\log n\Big)^{\frac{1}{\gamma}} + O\Big((\log n)^{\frac{1}{\gamma}-1}\Big)$$

if $\alpha(n)$ and $\beta(n)$ are appropriately chosen constants (independent of n). Thus case V of the Theorem is proved. □

In a next paper we shall investigate the cases

$$f(x) = \frac{1}{\sqrt{2\pi}x} e^{-\ln^2 x/2} \quad \text{"lognormal" ([4], p. 228),}$$

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \ (p,q>0,\ \ 0<x<1) \quad \text{"beta" ([4], p. 240)}$$

$$f(x) = \lambda e^{-\lambda x}, \ \ x>0, \ \ \lambda>0 \quad \text{"exponential" ([4], p. 188).}$$

REFERENCES

[1] STEINER, F., *Geofizikai adatfeldolgozás, robusztus becslések*, Tankönyvkiadó, Budapest, 1985 (in Hungarian).

[2] CSERNYÁK, L., Flanken-Diagnostik, Publications of the Technical University for Heavy Industry, Series A, Miskolc (in Hungarian).

[3] RÉNYI, A., *Valószínűségszámítás* [The calculus of probabilities], Tankönyvkiadó, Budapest, 1954 (in Hungarian). *MR* **17**–46

[4] PRÉKOPA, A., *Valószínűségelmélet*, Műszaki Könyvkiadó, Budapest, 1962 (in Hungarian).

[5] HALÁSZ, G., MÁRIALIGETI, J. and ZOBORY, I., Statisztikus módszerek a műszaki gyakorlatban, Kézirat, BME Továbbképző Intézete, Budapest, 1976 (in Hungarian).

[6] POSTNIKOV, A. G., *Introduction to analytic number theory*, Nauka, Moscow, 1971 (in Russian). *MR* **55** #7895

(Received May 23, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 433–447

THE MASS CENTRE AND THE GRAVITY CENTRE ON THE HYPERBOLIC PLANE

TEODÓRA WIEGAND

Abstract

In this article the position of the gravity point of the hyperbolic triangle is determined by the help of the triangle sides. It is well known that in the Euclidean plane the gravity centre and the mass centre (it is the unique point around which the rigid particle system with equal masses given by the vertices of the triangle has a free rotation) of all the triangles coincide. We point out that in the hyperbolic plane only the equiangular triangle has this property. Finally we point out what kind of connection exists in general case between the gravity centre and the mass centre in the hyperbolic plane.

We denote in the following by H^2 and E^2 the hyperbolic and Euclidean planes; the points of H^2 are $A, B, C\dots$; the points of E^2 are $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$. Let $\overline{XY}$ $(\overline{\mathcal{X}\mathcal{Y}})$ be the natural distance of the points $X, Y \in H^2$ $(\mathcal{X}, \mathcal{Y} \in E^2)$.

We call the lines connecting the vertices with the midpoints of the opposite sides the gravity lines of a triangle also in H^2 as well in E^2. It is known that the gravity lines intersect in one point also in the hyperbolic plane (cf. [3]), which point G is called the gravity centre of the triangle.

THEOREM 1. *For the position of the gravity centre G of the triangle ABC in the hyperbolic plane we have:*

$$\operatorname{sh}\overline{AG} = \sqrt{\frac{\operatorname{sh}^2\overline{AC} + \operatorname{sh}^2\overline{AB} + 2\cdot\operatorname{ch}\overline{AC}\cdot\operatorname{ch}\overline{AB} - 2\cdot\operatorname{ch}\overline{BC}}{2(\operatorname{ch}\overline{BC} + \operatorname{ch}\overline{AC} + \operatorname{ch}\overline{AB}) + 3}}$$

$$\cos GAB\angle = \frac{\operatorname{sh}^2\overline{AB} + \operatorname{ch}\overline{AC}\cdot\operatorname{ch}\overline{AB} - \operatorname{ch}\overline{BC}}{\operatorname{sh}\overline{AB}\cdot\sqrt{\operatorname{sh}^2\overline{AC} + \operatorname{sh}^2\overline{AB} + 2\cdot\operatorname{ch}\overline{AC}\cdot\operatorname{ch}\overline{AB} - 2\cdot\operatorname{ch}\overline{BC}}}.$$

PROOF. First we define a mapping (model) of the hyperbolic plane H^2 onto the Euclidean plane E^2 (cf. [4]), and we use the rectangular coordinates of E^2 for the description of some transformations of the hyperbolic plane.

The model: Let us take two oriented mutually perpendicular lines OX and OY in H^2, let the positive sense of rotation be chosen according to

1980 *Mathematics Subject Classification* (1985 Revision). Primary 51M10; Secondary 51N30.

Key words and phrases. Hyperbolic plane, gravity centre.

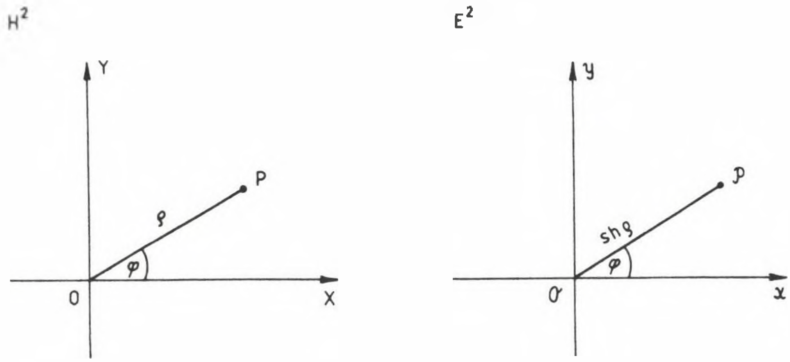


Fig. 1

$XOY\angle = \frac{\pi}{2}$, and let us take a rectangular coordinate system $\mathcal{OXY}$ in E^2 (Fig. 1).

Let the point O of H^2 be represented by the Euclidean origin, and if for any point P different from O

$$XOP\angle = \varphi, \quad \overline{OP} = \varrho$$

then P mapped in E^2 by the point $\mathcal{P}$ for which

$$\mathcal{XOP}\angle = \varphi, \quad \overline{\mathcal{OP}} = \operatorname{sh} \varrho.$$

In this way we have made a one-to-one mapping between the points of the hyperbolic plane and those of the whole Euclidean plane.

The properties of the model:

— the images of the lines passing through the origin and the equidistant curves belonging to them are straight lines;

— the images of the lines non-passing through the origin are hyperbolas, their asymptotes pass through the Euclidean origin; and the images of the equidistant curves belonging to them are also hyperbolas, their asymptotes are parallel to the asymptotes belonging to the image of the base line;

— if d is the distance between the equidistant curve and the base line then

- the distance of image straight lines is $\operatorname{sh} d$ when the base line passes through the origin,
- the distance between the asymptotes of the images is $\operatorname{sh} d$ when the base line does not pass through the origin.

While proving we admit first that in the hyperbolic plane there exists a point which having been chosen as the coordinate system origin of the hyperbolic plane, the gravity centre of the triangle determined by the points $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ obtained in the Euclidean representation is the Euclidean origin. (The triangle $\mathcal{ABC}$ does not coincide with the image of the triangle ABC; this image is namely surrounded at least by one arc of hyperbola.) We admit

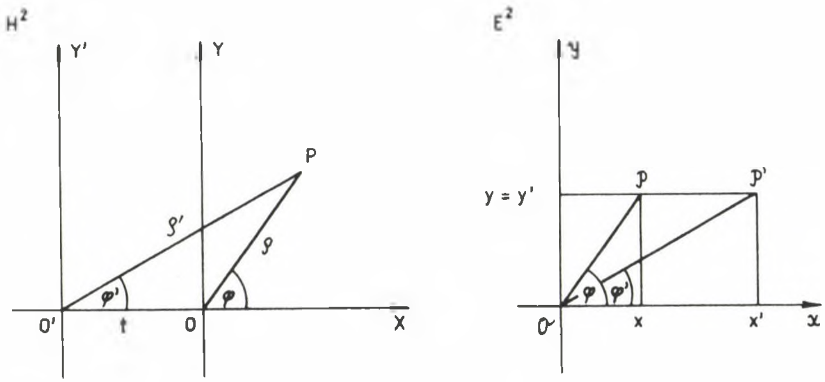


Fig. 2

that in the hyperbolic plane there exists only one such point and it is the gravity centre of the triangle ABC; finally we determine the position of the gravity centre.

LEMMA 1. *Let be given in H^2 the translation of the coordinate system OXY along the axis OX to the oriented distance $\overline{O'O} = t$ into the coordinate system $O'XY'$. Let $\mathcal{P}$ and $\mathcal{P}'$ be the images of the point $P \in H^2$ in the Euclidean model corresponding to the coordinate systems OXY and $O'XY'$, respectively. The Euclidean coordinates of $\mathcal{P}$ and $\mathcal{P}'$ satisfy*

$$\begin{aligned} x' &= x \cdot \operatorname{ch} t + \sqrt{1 + x^2 + y^2} \cdot \operatorname{sh} t \\ y' &= y \end{aligned}$$

(Note: $\mathcal{P}$ and $\mathcal{P}'$ image points are described in the same Euclidean coordinate system.)

PROOF. For the triangle $OO'P$ in Fig. 2 (cf. [1])

$$\frac{\sin(180^\circ - \varphi)}{\sin \varphi'} = \frac{\operatorname{sh} \varrho'}{\operatorname{sh} \varrho} \qquad \text{(sine rule)}$$

$$\operatorname{sh} t \cdot \operatorname{cth} \varrho = \sin(180^\circ - \varphi) - \operatorname{ctg} \varphi' + \cos(180^\circ - \varphi) \cdot \operatorname{ch} t$$
$$\text{(cotangent rule)}$$

from which it follows:

$$\begin{aligned} y' &= \operatorname{sh} \varrho' \cdot \sin \varphi' = \operatorname{sh} \varrho \cdot \sin \varphi = y \\ x' &= \operatorname{sh} \varrho' \cdot \cos \varphi' = \operatorname{sh} \varrho' \cdot \sin \varphi' \cdot \operatorname{ctg} \varphi' = \\ &= \operatorname{sh} r \cdot \cos \varphi \cdot \operatorname{ch} t + \operatorname{ch} \varrho \cdot \operatorname{sh} t = \\ &= x \cdot \operatorname{ch} t + \sqrt{1 + x^2 + y^2} \cdot \operatorname{sh} t. \end{aligned}$$

LEMMA 2. *Let ABC be a triangle in H^2. We denote by $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ the images of the points $A, B, C \in H^2$ in the Euclidean model corresponding to any OXY coordinate system in H^2. There exists such a point $T \in H^2$ that in the case of $O = T$ the gravity centre of the triangle defined by the points $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ is $\mathcal{O}$.*

PROOF. First let OXY be any arbitrary coordinate system and in the representation corresponding to it let the gravity centre of the triangle $\mathcal{ABC}$ be denoted by $\mathcal{G}$. Because of the rotation invariance of the model mapping we can suppose that $\mathcal{G}$ lies on the axis $\mathcal{OX}$. When the polar coordinates of the points A, B, C are denoted by (p,α), (q,β), (r,γ), respectively, then the polar coordinates of the points $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ are $(\operatorname{sh} p,\alpha)$, $(\operatorname{sh} q,\beta)$, $(\operatorname{sh} r,\gamma)$, respectively. For the coordinates x_g, y_g of the point $\mathcal{G}$ we have:

$$x_g = \frac{x_a + x_b + x_c}{3} = \frac{\operatorname{sh} p \cdot \cos\alpha + \operatorname{sh} q \cdot \cos\beta + \operatorname{sh} r \cdot \cos\gamma}{3}$$
$$y_g = \frac{y_a + y_b + y_c}{3} = 0.$$

We are looking for the coordinate transformation corresponding to Lemma 1 in which the gravity centre of the triangle determined by the image points $\mathcal{A}'$, $\mathcal{B}'$, $\mathcal{C}'$ corresponding to the coordinate system $O'XY'$ is $\mathcal{O}$. For the coordinates of the points $\mathcal{A}'$, $\mathcal{B}'$, $\mathcal{C}'$ we have

$$\begin{aligned} x_{a'} &= x_a \cdot \operatorname{ch} t + \sqrt{1 + x_a^2 + y_a^2} \cdot \operatorname{sh} t = \\ &= x_a \cdot \operatorname{ch} t + \operatorname{ch} p \cdot \operatorname{sh} t \\ y_{a'} &= y_a \end{aligned}$$

similarly:

$$\begin{aligned} x_{b'} &= x_b \cdot \operatorname{ch} t + \operatorname{ch} q \cdot \operatorname{sh} t \\ y_{b'} &= y_b \\ x_{c'} &= x_c \cdot \operatorname{ch} t + \operatorname{ch} r \cdot \operatorname{sh} t \\ y_{c'} &= y_c \end{aligned}$$

When the gravity centre of the triangle $\mathcal{A}'\mathcal{B}'\mathcal{C}'$ is $\mathcal{O}$ then

$$\begin{aligned} 0 &= x_{a'} + x_{b'} + x_{c'}, \\ 0 &= y_{a'} + y_{b'} + y_{c'}. \end{aligned}$$

The second equality is satisfied for any arbitrary t, however, the first one

$$\begin{aligned} 0 = (\operatorname{sh} p \cdot \cos\alpha + \operatorname{sh} q \cdot \cos\beta + \operatorname{sh} r \cdot \cos\gamma) \cdot \operatorname{ch} t + \\ + (\operatorname{ch} p + \operatorname{ch} q + \operatorname{ch} r) \cdot \operatorname{sh} t \end{aligned}$$

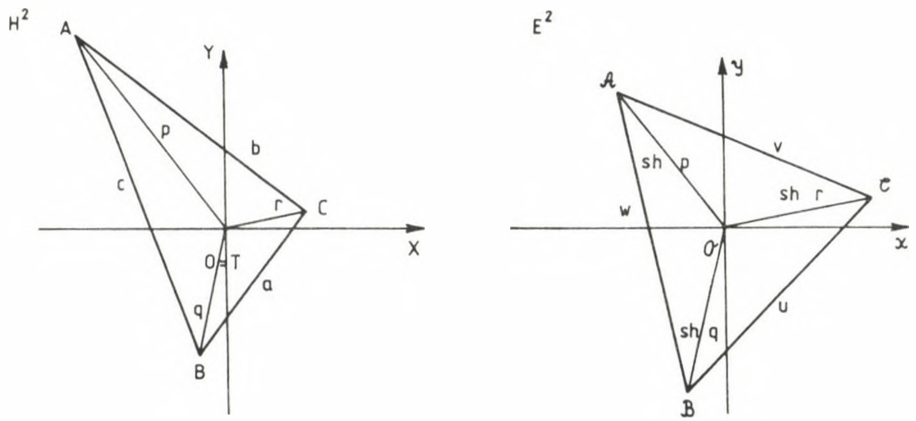


Fig. 3

from where

$$\mathrm{th}\, t = -\frac{\mathrm{sh}\, p \cdot \cos\alpha + \mathrm{sh}\, q \cdot \cos\beta + \mathrm{sh}\, r \cdot \cos\gamma}{\mathrm{ch}\, p + \mathrm{ch}\, q + \mathrm{ch}\, r}$$

i.e. t is uniquely determined. The point $O' = T$, being at a distance with sign $\overline{TO} = t$ from O, meets the requirement that in the case of a representation belonging to any arbitrary coordinate system with origin T the gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$.

LEMMA 3. *Let ABC be a triangle in H^2 and T a point which has the properties mentioned above and let OXY be any arbitrary coordinate system. Let the appropriate Euclidean image points be denoted by $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{T}$. If $\mathcal{G}$ is the gravity centre of the triangle $\mathcal{ABC}$ then $\mathcal{O}$, $\mathcal{T}$, $\mathcal{G}$ are collinear, $\mathcal{T}$ separates $\mathcal{O}$ and $\mathcal{G}$ and their distance is:*

$$\overline{\mathcal{OG}} = \frac{1}{3}\sqrt{2(\mathrm{ch}\,\overline{BC} + \mathrm{ch}\,\overline{AC} + \mathrm{ch}\,\overline{AB}) + 3} \cdot \mathrm{sh}\,\overline{OT}.$$

PROOF. Let us take first in H^2 a coordinate system with origin T. In this case the gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$. The cosine rules for the triangles ABO, BCO, CAO are (Fig. 3, cf. [1]):

$$\begin{aligned}
\mathrm{ch}\, a &= \mathrm{ch}\, q \cdot \mathrm{ch}\, r - \mathrm{sh}\, q \cdot \mathrm{sh}\, r \cdot \cos(\gamma - \beta)\\
\mathrm{ch}\, b &= \mathrm{ch}\, p \cdot \mathrm{ch}\, r - \mathrm{sh}\, p \cdot \mathrm{sh}\, r \cdot \cos(\alpha - \gamma)\\
\mathrm{ch}\, c &= \mathrm{ch}\, p \cdot \mathrm{ch}\, q - \mathrm{sh}\, p \cdot \mathrm{sh}\, q \cdot \cos(\beta - \alpha).
\end{aligned}$$

Adding these

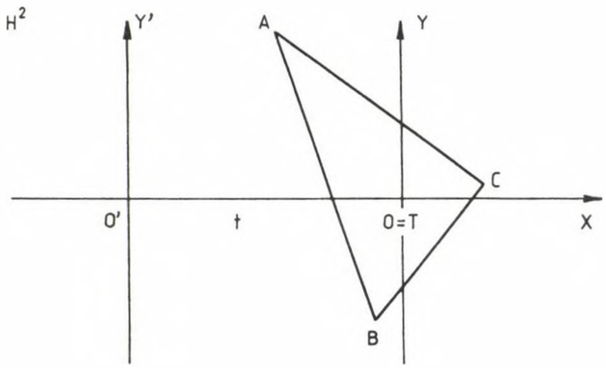


Fig. 4

$$\text{(1)}\qquad \begin{aligned}\operatorname{ch} a+\operatorname{ch} b+\operatorname{ch} &= \operatorname{ch} q\cdot\operatorname{ch} r+\operatorname{ch} p\cdot\operatorname{ch} r+\operatorname{ch} p\cdot\operatorname{ch} q-\\ &-(\operatorname{sh} q\cdot\operatorname{sh} r\cdot\cos(\gamma-\beta)+\operatorname{sh} p\cdot\operatorname{sh} r\cdot\cos(\alpha-\gamma)+\operatorname{sh} p\cdot\operatorname{sh} q\cdot\cos(\beta-\alpha)).\end{aligned}$$

Since the gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$, therefore

$$\begin{aligned}0 &= \operatorname{sh} p\cdot\cos\alpha+\operatorname{sh} q\cdot\cos\beta+\operatorname{sh} r\cdot\cos\gamma\\ 0 &= \operatorname{sh} p\cdot\sin\alpha+\operatorname{sh} q\cdot\sin\beta+\operatorname{sh} r\cdot\sin\gamma.\end{aligned}$$

Squaring and adding the two equalities we have

$$\text{(2)}\qquad \begin{aligned}0=\operatorname{sh}^2 p+\operatorname{sh}^2 q+\operatorname{sh}^2 r+2(\operatorname{sh} q\cdot\operatorname{sh} r\cdot\cos(\gamma-\beta)+\\ +\operatorname{sh} p\cdot\operatorname{sh} r\cdot\cos(\alpha-\gamma)+\operatorname{sh} p\cdot\operatorname{sh} q\cdot\cos(\beta-\alpha)).\end{aligned}$$

Adding the double of (1) to (2) we have

$$2(\operatorname{ch} a+\operatorname{ch} b+\operatorname{ch} c)=(\operatorname{ch} p+\operatorname{ch} q+\operatorname{ch} r)^2-3$$

from where

$$\text{(3)}\qquad \operatorname{ch} p+\operatorname{ch} q+\operatorname{ch} r=\sqrt{2(\operatorname{ch} a+\operatorname{ch} b+\operatorname{ch} c)+3}.$$

Let us make a coordinate transformation in H^2 according to Lemma 1 (Fig. 4).

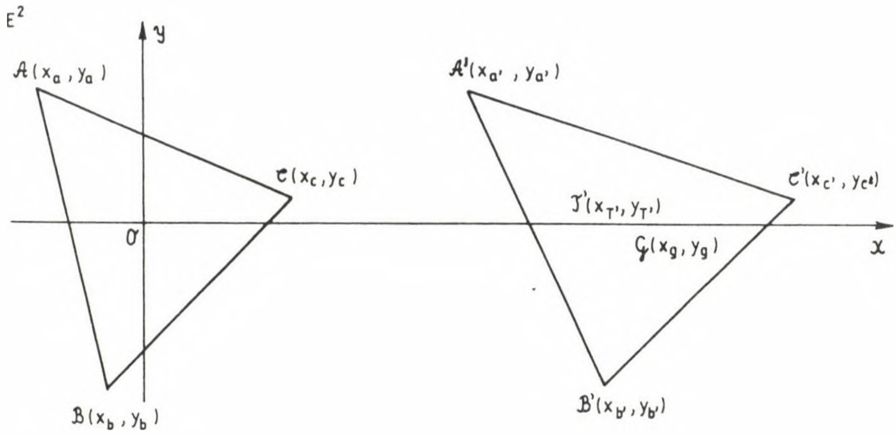


Fig. 5

For the coordinates of the points $\mathcal{A}'$, $\mathcal{B}'$, $\mathcal{C}'$, $\mathcal{T}'$ which are the images of the points A, B, C, T corresponding to the system $O'XY'$ (Fig. 5):

$$\begin{aligned}
x_{a'} &= x_a \cdot \operatorname{ch} t + \operatorname{ch} p \cdot \operatorname{sh} t \\
y_{a'} &= y_a \\
x_{b'} &= x_b \cdot \operatorname{ch} t + \operatorname{ch} q \cdot \operatorname{sh} t \\
y_{b'} &= y_b \\
x_{c'} &= x_c \cdot \operatorname{ch} t + \operatorname{ch} r \cdot \operatorname{sh} t \\
y_{c'} &= y_c \\
x_{T'} &= \operatorname{sh} t \\
y_{T'} &= 0.
\end{aligned}$$

Since the gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$, therefore

$$\begin{aligned}
x_a + x_b + x_c &= 0 \\
y_a + y_b + y_c &= 0
\end{aligned}$$

and so for the coordinates of the gravity centre $\mathcal{G}$ of the triangle $\mathcal{A'B'C'}$

$$\begin{aligned}
x_g &= \frac{1}{3}(x_{a'} + x_{b'} + x_{c'}) = \frac{1}{3}(\operatorname{ch} p + \operatorname{ch} q + \operatorname{ch} r) \cdot \operatorname{sh} t \\
y_g &= \frac{1}{3}(y_{a'} + y_{b'} + y_{c'}) = 0
\end{aligned}$$

replacing (3)

$$x_g = \frac{1}{3}\sqrt{2(\operatorname{ch} a + \operatorname{ch} b + \operatorname{ch} c) + 3} \cdot \operatorname{sh} t \qquad (5)$$
$$y_g = 0.$$

Since $x_g > x_{T'}$ therefore $\mathcal{T}'$ separates $\mathcal{O}$ and $\mathcal{G}$.

On the basis of the above results if we take any rectangular coordinate system OXY in H^2 the following is true: Let us map the points A, B, C, T in E^2 corresponding to the coordinate system OXY and draw the gravity centre $\mathcal{G}$ of the triangle $\mathcal{ABC}$. Rotating both coordinate systems with $\mathcal{GOX}\angle = \vartheta$ (Fig. 6), on the basis of (4) and (5) $\mathcal{G}$ and $\mathcal{T}$ are on the axis $\mathcal{OX}'$, $\mathcal{T}$ separates $\mathcal{O}$ and $\mathcal{G}$ and their distance is

$$\overline{\mathcal{OG}} = \frac{1}{3}\sqrt{2(\operatorname{ch} \overline{BC} + \operatorname{ch} \overline{AC} + \operatorname{ch} \overline{AB}) + 3} \cdot \operatorname{sh} \overline{OT} \qquad (6)$$

because of the rotation invariance of the mapping.

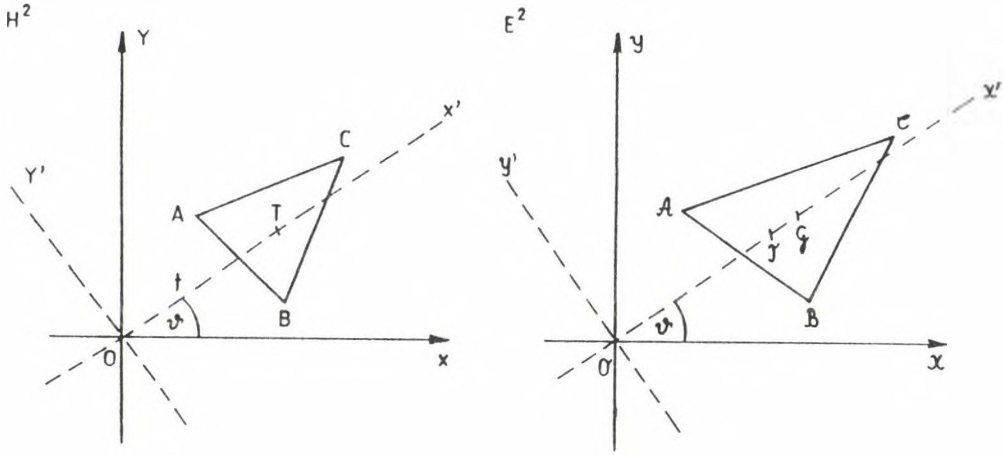


Fig. 6

LEMMA 4. *T is the gravity centre of the triangle ABC.*

PROOF. Let us apply Lemma 3 to the following case:

— $O = A$;

— the axis OX is the line of the sides AB and its positive half-line contains B.

Then to the triangle ABC a triangle $\mathcal{ABC}$ is given for which

$$\overline{\mathcal{AB}} = \operatorname{sh} c, \quad \overline{\mathcal{AC}} = \operatorname{sh} b, \quad \mathcal{CAB}\angle = CAB\angle$$

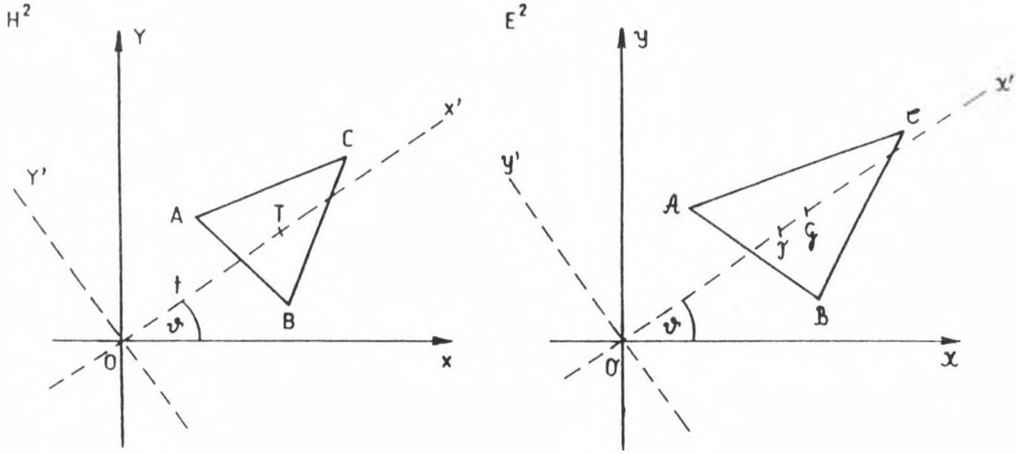


Fig. 7

(further on the coordinate systems can be left out). Then (Fig. 7) according to the congruence of the triangle $\mathcal{F}_a\mathcal{C}\mathcal{T}_1$ and $\mathcal{F}_a\mathcal{B}\mathcal{T}_2$

$$\operatorname{sh} b \cdot \sin \eta = \operatorname{sh} c \cdot \sin \vartheta,$$

in the rectangular triangles AD_1C and ABD_2 of H^2 (cf. [1])

$$\operatorname{sh} \overline{D_1C} = \operatorname{sh} b \cdot \sin \eta$$
$$\operatorname{sh} \overline{BD_2} = \operatorname{sh} c \cdot \sin \vartheta$$

that is

$$\overline{D_1C} = \overline{BD_2}.$$

In the rectangular triangles CD_1N and D_2NB

$$\operatorname{sh} \overline{CN} = \frac{\operatorname{sh} \overline{CD_1}}{\sin D_1NC\angle}, \qquad \operatorname{sh} \overline{BN} = \frac{\operatorname{sh} \overline{D_2B}}{\sin D_2NB\angle}$$

from where

$$\overline{CN} = \overline{BN}$$

because of $D_1NC\angle = D_2NB\angle$. Therefore T lies on the gravity line passing through the vertex A and the midpoint of the opposite side. The proof is similar for the two gravity lines. In this way it has been proved that there exists only one such point T and it is the gravity centre of the triangle.

The numerical determination of G location: in the triangle $\mathcal{ABC}$ (Fig. 7)

$$s = \frac{\sqrt{2\operatorname{sh}^2 b + 2\operatorname{sh}^2 c - \overline{\mathcal{BC}}^2}}{3}$$

where $\overline{BC}^2 = \mathrm{sh}^2 b + \mathrm{sh}^2 c - 2\mathrm{sh}\, b \cdot \mathrm{sh}\, c \cdot \cos\alpha$. $\cos\alpha$ can be expressed by the hyperbolic cosine rule belonging to the triangle ABC (cf. [1])

$$\cos\alpha = \frac{\mathrm{ch}\, b \cdot \mathrm{ch}\, c - \mathrm{ch}\, a}{\mathrm{sh}\, b \cdot \mathrm{sh}\, c}$$

therefore

$$s = \frac{\sqrt{\mathrm{sh}^2 b + \mathrm{sh}^2 c + 2\mathrm{ch}\, b \cdot \mathrm{ch}\, c - 2\mathrm{ch}\, a}}{3}$$

and with the help of (6)

$$\mathrm{sh}\, t = \sqrt{\frac{\mathrm{sh}^2 b + \mathrm{sh}^2 c + 2\mathrm{ch}\, b \cdot \mathrm{ch}\, c - 2\mathrm{ch}\, a}{2(\mathrm{ch}\, a + \mathrm{ch}\, b + \mathrm{ch}\, c) + 3}}$$

$$\mathrm{sh}\, \overline{AG} = \sqrt{\frac{\mathrm{sh}^2 \overline{AC} + \mathrm{sh}^2 \overline{AB} + 2 \cdot \mathrm{ch}\, \overline{AC} \cdot \mathrm{ch}\, \overline{AB} - 2 \cdot \mathrm{ch}\, \overline{BC}}{2(\mathrm{ch}\, \overline{BC} + \mathrm{ch}\, \overline{AC} + \mathrm{ch}\, \overline{AB}) + 3}}.$$

For the triangle $\mathcal{A}\mathcal{F}_c\mathcal{G}$

$$\overline{\mathcal{G}\mathcal{F}_c}^2 = s^2 + \frac{\mathrm{sh}^2 c}{4} - s \cdot \mathrm{sh}\, c \cdot \cos\vartheta,$$

where

$$\overline{\mathcal{G}\mathcal{F}_c}^2 = \frac{1}{9}\overline{\mathcal{C}\mathcal{F}_c}^2 = \frac{1}{9} \cdot \frac{2\mathrm{sh}^2 b + 2\overline{BC}^2 - \mathrm{sh}^2 c}{4}$$

replacing it and expressing $\cos\vartheta$

$$\cos\vartheta = \frac{\mathrm{sh}^2 c + \mathrm{ch}\, b \cdot \mathrm{ch}\, c - \mathrm{ch}\, a}{\mathrm{sh}\, c \cdot \sqrt{\mathrm{sh}^2 b + \mathrm{sh}^2 c + 2\mathrm{ch}\, b \cdot \mathrm{ch}\, c - 2\mathrm{ch}\, a}}$$

$$\cos SAB\angle = \frac{\mathrm{sh}^2 \overline{AB} + \mathrm{ch}\, \overline{AC} \cdot \mathrm{ch}\, \overline{AB} - \mathrm{ch}\, \overline{BC}}{\mathrm{sh}\, \overline{AB} \cdot \sqrt{\mathrm{sh}^2 \overline{AC} + \mathrm{sh}^2 \overline{AB} + 2 \cdot \mathrm{ch}\, \overline{AC} \cdot \mathrm{ch}\, \overline{AB} - 2 \cdot \mathrm{ch}\, \overline{BC}}}.$$

In this way we have proved the theorem.

The mass centre M of a triangle ABC in H^2 is defined by the inequality

$$\mathrm{sh}^2 \overline{AM} + \mathrm{sh}^2 \overline{BM} + \mathrm{sh}^2 \overline{CM} \leq \mathrm{sh}^2 \overline{AP} + \mathrm{sh}^2 \overline{BP} + \mathrm{sh}^2 \overline{CP} \quad (\forall P \in H^2). \tag{7}$$

This point is uniquely determined and characterized by the property that the rigid particle system with equal masses given by the vertices of the triangle has a unique free rotation around this point (cf. [2]).

THEOREM 2. *Let ABC be any triangle with the mass centre M in the hyperbolic plane and $A'B'C'$ an other triangle for which we have*

$$\overline{MA'} = 2\overline{MA}, \quad \overline{MB'} = 2\overline{MB}, \quad \overline{MC'} = 2\overline{MC}$$

and the points M, A, A' and M, B, B' and M, C, C' are collinear. The point M is the gravity centre of this triangle $A'B'C'$. The gravity centre and the mass centre of the triangle coincide if and only if when the sides of the triangle are equal.

LEMMA 5. *"Magnifying" the hyperbolic triangle ABC from its mass centre M in an arbitrary ratio, the mass centre of the obtained triangle $A'B'C'$ is also M if and only if when the triangle ABC is an equiangular triangle.*

("Magnifying" means: $\overline{MA'} = \varepsilon \cdot \overline{MA}$, $\overline{MB'} = \varepsilon \cdot \overline{MB}$, $\overline{MC'} = \varepsilon \cdot \overline{MC}$, and the points M, A, A' and M, B, B' and M, C, C' are collinear; where ε is any arbitrary positiv number.)

PROOF. Let (p, α), (q, β), (r, γ) and (ϱ, φ) be the polar coordinates of the points A, B, C and any point P of the hyperbolic plane, respectively, in any coordinate system OXY (Fig. 8).

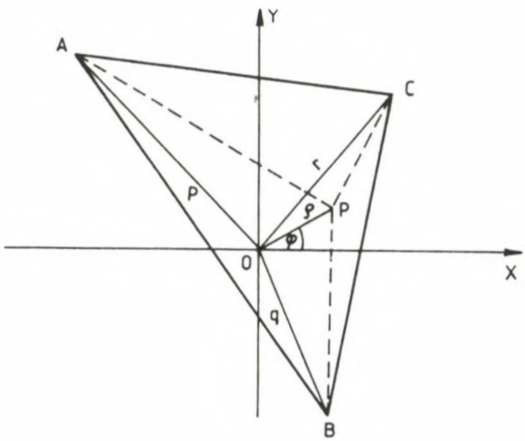


Fig. 8

Let us establish the expression of the right side of (7) as a function of ϱ and φ: let us apply the hyperbolic cosine rule for the triangles AOP, BOP, COP (cf. [1]):

$$\operatorname{ch} \overline{AP} = \operatorname{ch} p \cdot \operatorname{ch} \varrho - \operatorname{sh} p \cdot \operatorname{sh} \varrho \cdot \cos(\alpha - \varphi)$$
$$\operatorname{ch} \overline{BP} = \operatorname{ch} q \cdot \operatorname{ch} \varrho - \operatorname{sh} q \cdot \operatorname{sh} \varrho \cdot \cos(\beta - \varphi)$$
$$\operatorname{ch} \overline{CP} = \operatorname{ch} r \cdot \operatorname{ch} \varrho - \operatorname{sh} r \cdot \operatorname{sh} \varrho \cdot \cos(\gamma - \varphi)$$

$$f(\varrho,\varphi)=\operatorname{sh}^2\overline{AP}+\operatorname{sh}^2\overline{BP}+\operatorname{sh}^2\overline{CP}=\operatorname{ch}^2\overline{AP}+\operatorname{ch}^2\overline{BP}+\operatorname{ch}^2\overline{CP}-3=$$
$$=(\operatorname{ch}^2p+\operatorname{ch}^2q+\operatorname{ch}^2r)\cdot\operatorname{ch}^2\varrho-2[\operatorname{ch}p\cdot\operatorname{sh}p\cdot\cos(\alpha-\varphi)+$$
$$+\operatorname{ch}q\cdot\operatorname{sh}q\cdot\cos(\beta-\varphi)+\operatorname{ch}r\cdot\operatorname{sh}r\cdot\cos(\gamma-\varphi)]\cdot\operatorname{ch}\varrho\cdot\operatorname{sh}\varrho+$$
$$+[\operatorname{sh}^2p\cdot\cos^2(\alpha-\varphi)+\operatorname{sh}^2q\cdot\cos^2(\beta-\varphi)+\operatorname{sh}^2r\cdot\cos^2(\gamma-\varphi)]\cdot\operatorname{sh}^2\varrho-3.$$

This function has its minimum in a unique point M of the hyperbolic plane (this has been proved by P. Nagy (cf. [2])). This point M is the origin O of the coordinate system if and only if when the value of the function $f'_\varrho(\varrho,\varphi)$ and $f'_\varphi(\varrho,\varphi)$ is zero in the case $\varrho=0$ and φ is arbitrary. That is the parameters $p,q,r,\alpha,\beta,\gamma$ determining the triangle ABC have to satisfy the following equality for all φ in the case of $M=O$:

$$\operatorname{sh}2p\cdot\cos(\alpha-\varphi)+\operatorname{sh}2q\cdot\cos(\beta-\varphi)+\operatorname{sh}2r\cdot\cos(\gamma-\varphi)=0. \tag{8}$$

This condition is equivalent to the assumption that the mass centre of the triangle ABC is O.

Let us assume now that for the triangle ABC $M=O$, that is (8) is satisfied. Let us "magnify" the triangle ABC from O in ε ratio. O is the mass centre of the obtained triangle $A'B'C'$ if and only if when

$$\begin{aligned}\operatorname{sh}2(\varepsilon p)\cdot\cos(\alpha-\varphi)+\operatorname{sh}2(\varepsilon q)\cdot\cos(\beta-\varphi)+\\+\operatorname{sh}2(\varepsilon r)\cdot\cos(\gamma-\varphi)=0\end{aligned} \tag{9}$$

is satisfied for all φ.

By the notations $\cos(\alpha-\varphi)=x$, $\cos(\beta-\varphi)=y$, $\cos(\gamma-\varphi)=z$ (8) and (9) can be interpreted as equations of planes lying in the Euclidean space. The equations are satisfied by the same non-collinear points (x,y,z) if the mass centre of the triangles ABC and $A'B'C'$ is O. It means that the two planes coincide, that is the normals

$$\underline{n}(\operatorname{sh}2p,\operatorname{sh}2q,\operatorname{sh}2r)\quad\text{and}\quad\underline{n}'(\operatorname{sh}2\varepsilon p,\operatorname{sh}2\varepsilon q,\operatorname{sh}2\varepsilon r)$$

are parallel. That is satisfied only if

$$p=q=r.$$

In this case from the conditions (8) and (9) remains:

$$\cos(\alpha-\varphi)+\cos(\beta-\varphi)+\cos(\gamma-\varphi)=0\qquad(\forall\varphi)$$
$$(\cos\alpha+\cos\beta+\cos\gamma)\cdot\cos\varphi+(\sin\alpha+\sin\beta+\sin\gamma)\cdot\sin\varphi=0\qquad(\forall\varphi)$$

which is satisfied only if when

$$\begin{aligned}\cos\alpha+\cos\beta+\cos\gamma&=0\quad\text{and}\\ \sin\alpha+\sin\beta+\sin\gamma&=0.\end{aligned} \tag{10}$$

Let us we assume first that $\gamma = 0$. In this case from (10)

$$\begin{aligned}\cos\alpha + \cos\beta &= -1\\ \sin\alpha + \sin\beta &= 0,\end{aligned}$$

its solution (apart from the multiple of 2π):

$$\alpha = \frac{2\pi}{3}, \quad \beta = \frac{4\pi}{3} \quad \text{(or inversely)}.$$

The arbitrary rotated values of these angles are also solutions of (10) that is

$$\begin{aligned}\cos\left(\frac{2\pi}{3} + \delta\right) + \cos\left(\frac{4\pi}{3} + \delta\right) &= -\cos\delta\\ \sin\left(\frac{2\pi}{3} + \delta\right) + \sin\left(\frac{4\pi}{3} + \delta\right) &= -\sin\delta.\end{aligned}$$

Other solution does not exist because transforming the equations of (10) into the forms

$$\begin{aligned}\cos\alpha + \cos\beta &= -\cos\gamma\\ \sin\alpha + \sin\beta &= -\sin\gamma,\end{aligned}$$

squaring and adding them we get the equation:

$$\cos(\alpha - \beta) = -\frac{1}{2}.$$

Its solutions are $\alpha - \beta = \frac{2\pi}{3}$ and $\alpha - \beta = \frac{4\pi}{3}$. Accordingly, the triangle of the hyperbolic plane can be "magnified" from the mass centre M if and only if in a way that the mass centre of the obtained triangle is also M when

$$\overline{AM} = \overline{BM} = \overline{CM} \quad \text{and} \quad AMB\angle = BMC\angle = CMA\angle = \frac{2\pi}{3}$$

that is the triangle ABC is an equiangular triangle.

Proof of Theorem 2. Let again the position of the coordinate system OXY be arbitrary in H^2 and let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be the notations of the images of the points A, B, C in the Euclidean representation corresponding to OXY. It is known that in E^2 the gravity centre $\mathcal{G}$ of the triangle $\mathcal{ABC}$ is determined by the following inequality:

$$\text{(11)} \qquad \overline{\mathcal{AG}}^2 + \overline{\mathcal{BG}}^2 + \overline{\mathcal{CG}}^2 \leqq \overline{\mathcal{AP}}^2 + \overline{\mathcal{BP}}^2 + \overline{\mathcal{CP}}^2 \quad (\forall \mathcal{P} \in E^2).$$

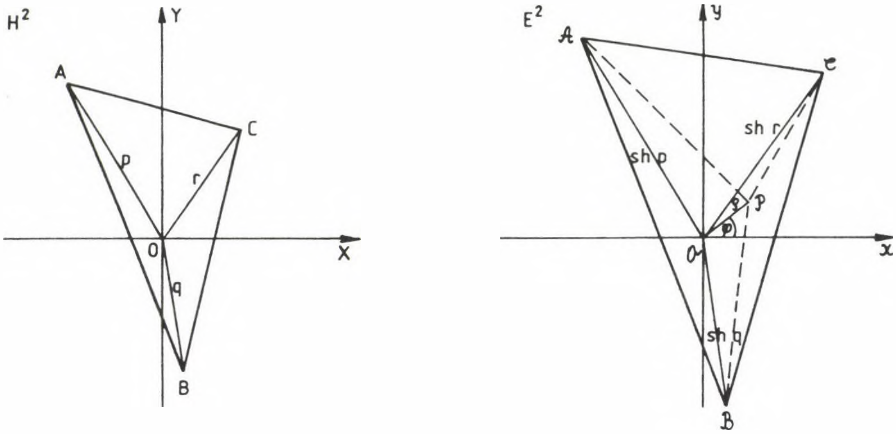


Fig. 9

Let us establish the expression of the right side of (11) as a function of the polar coordinates (ϱ, φ) of the point $\mathcal{P}$: applying the cosine rule for the triangles $\mathcal{AOP}$, $\mathcal{BOP}$, $\mathcal{COP}$ (Fig. 9) we have:

$$g(\varrho,\varphi) = \overline{\mathcal{AP}}^2 + \overline{\mathcal{BP}}^2 + \overline{\mathcal{CP}}^2 = \operatorname{sh}^2 p + \operatorname{sh}^2 q + \operatorname{sh}^2 r + 3\varrho^2 -$$
$$-2(\operatorname{sh} p \cdot \cos(\alpha - \varphi) + \operatorname{sh} q \cdot \cos(\beta - \varphi) + \operatorname{sh} r \cdot \cos(\gamma - \varphi)) \cdot \varrho.$$

In the Euclidean geometry the function $g(\varrho, \varphi)$ has its minimum in one point of the plane (in the Euclidean gravity centre).

The gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$ if and only if when the value of the functions $g'_\varrho(\varrho, \varphi)$ and $g'_\varphi(\varrho, \varphi)$ is zero in the case $\varrho = 0$ and φ is arbitrary, that is

$$\text{(12)} \qquad \operatorname{sh} p \cdot \cos(\alpha - \varphi) + \operatorname{sh} q \cdot \cos(\beta - \varphi) + \operatorname{sh} r \cdot \cos(\gamma - \varphi) = 0 \quad (\forall \varphi).$$

Let us now choose in H^2 the mass centre of the triangle ABC again as the origin of the coordinate system. Then (8) is satisfied. However, satisfying (8) in the Euclidean plane means that according to (12) the gravity centre of the triangle $\mathcal{A'B'C'}$ is $\mathcal{O}$, where the triangle $\mathcal{A'B'C'}$ corresponds in E^2 to the triangle $A'B'C'$ which is "magnified" in double ratio from $M = O$. But according to Lemmas 2 and 4 the point $M = O$ is the gravity centre of the triangle $A'B'C'$.

If the gravity centre and the mass centre of the hyperbolic triangle ABC coincide, then in the case of $M = O$ in the Euclidean plane the gravity centre of the triangle $\mathcal{ABC}$ is $\mathcal{O}$ according to Lemmas 2 and 4. That means that (8) and (12) are simultaneously satisfied, i.e., according to Lemma 5, the sides of the triangle ABC are equal.

The theorem is proved.

REFERENCES

[1] Szász, P., *Bevezetés a Bolyai–Lobacsevszkij-féle geometriába* [Introduction to Bolyai–Lobačevskian geometry], Disquisitiones Mathematicae Hungaricae, No. 5, Akadémiai Kiadó, Budapest, 1973. *MR* **54** #1072

[2] Nagy, P. T., Dynamical invariants of rigid motions on the hyperbolic plane, *Geometriae Dedicata* (to appear).

[3] Gruber, F., Neuer Beweis des Transversalensatzes in der absoluten Geometrie, *Sitzungsberichte der Heidelberger Akad. Wiss. Math. naturwiss. Kl.*, 1929, Abh. 18. *Jb. Fortschritte Math.* **55**, 955

[4] Wiegand, T., A "polar coordinate model" of the hyperbolic plane, *Publ. Math. Debrecen* (to appear).

(Received May 2, 1989)

JANUS PANNONIUS TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
MATEMATIKA TANSZÉK
IFJÚSÁG ÚTJA 6
H–7624 PÉCS
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 449–454

ON THE LEBESGUE FUNCTION OF $(0,1,2)$ INTERPOLATION

P. VÉRTESI

1. Introduction and preliminary results

If $X = \{x_{kn}\} \subset [-1,1]$, $1 \leqq k \leqq n$, $n = 1,2,\dots$, is an arbitrary interpolatory matrix, i.e.

(1.1) $$-1 \leqq x_{nn} < x_{n-1,n} < \dots < x_{1n} \leqq 1, \quad n = 1,2,\dots,$$

the unique Hermite–Fejér interpolatory polynomial $H_{nm}(f,X,x)$ of order $\leqq$ $\leqq nm-1$ ($m \geqq 1$, fixed integer) for a continuous function $f(x)$ in $[-1,1]$ ($f \in C$, shortly) is defined by

(1.2) $$H_{nm}^{(t)}(f,X,x_{kn}) = \delta_{0t} f(x_{kn}), \quad k = 1,2,\dots,n, \quad t = 0,1,\dots,m-1.$$

By (1.2), H_{nm}, using some obvious short notations, can be written as follows:

(1.3) $$H_{nm}(f,X,x) = \sum_{k=1}^{n} f(x_k) h_{knm}(X,x)$$

where for the polynomials h_k of degree exactly $mn-1$,

$$h_k^{(t)}(x_j) = \delta_{0t}\delta_{kj}, \quad 1 \leqq k,j \leqq n, \quad 0 \leqq t \leqq m-1.$$

The corresponding Lebesgue functions and Lebesgue constants

(1.4) $$\lambda_{nm}(X,x) := \sum_{k=1}^{n} |h_{knm}(X,x)|, \qquad \Lambda_{nm}(X) := \|\lambda_{nm}(X,x)\|$$

($\|f(x)\| = \max_{-1 \leqq x \leqq 1} |f(x)|$) play a decisive role in the convergence and divergence behaviour of H_{nm} interpolation. For $m = 1$ (Lagrange interpolation) G. Faber [1] proved that $\Lambda_{n1}(X) > c \log n$ for *any* system of nodes i.e. for arbitrary fixed X there exists an $f \in C$ such that $\overline{\lim}_{n\to\infty} |H_{n1}(f,X,x)| = \infty$.

Considering the pointwise behaviour of H_{n1} we refer to our paper P. Erdős and P. Vértesi [2] where we proved that *for any fixed X and $\varepsilon > 0$*

Akadémiai Kiadó, Budapest

there exist sets $E_n = E_n(X, \varepsilon)$ *with* $|E_n| \leqq \varepsilon$ *and* $\mu = \mu(\varepsilon) > 0$ *such that with* $R = (-\infty, \infty)$

$$\lambda_{n1}(X, x) > \mu \log n \quad \text{if} \quad x \in R \setminus E_n. \tag{1.5}$$

The above statement implies that

$$\overline{\lim_{n\to\infty}} \lambda_{n1}(X, x) = \infty \quad \textit{a.e. on the real line} \tag{1.6}$$

i.e. for almost every fixed $x_0 \in R$, $\overline{\lim\limits_{n\to\infty}} |H_{n1}(f, X, x_0)| = \infty$ with a proper $f = f_{x_0} \in C$. (Relation (1.6) can be obtained as follows. If (1.6) were not true, then $\overline{\lim\limits_{n\to\infty}} \lambda_{n1}(x) < \infty$ on a proper set G with $|G| = \delta > 0$. Let $\varepsilon = \delta/2$ be fixed. By (1.5), $\lambda_{n1}(x) > \mu \log n$ if $x \in S_n$ where $|R \setminus S_n| \leqq \varepsilon$. Let $S = \bigcap\limits_{k=1}^{\infty} \left(\bigcup\limits_{i=k}^{\infty} S_i \right) := \bigcap\limits_{k=1}^{\infty} Q_k$. Then $R \setminus S = \bigcup\limits_{k=1}^{\infty} (R \setminus Q_k) := \bigcup\limits_{k=1}^{\infty} P_k$ where by $Q_1 \supset Q_2 \supset \ldots$ we get $P_1 \subset P_2 \subset \ldots$ whence $|R \setminus S| = \lim\limits_{k\to\infty} |P_k| \leqq \varepsilon$. By construction $\overline{\lim\limits_{n\to\infty}} \lambda_{n1}(x) = \infty$ if $x \in S$ whence $G \subset R \setminus S$ i.e. $|G| \leqq |R \setminus S|$, a contradiction.)

The process H_{n3} for $X^{(\alpha,\beta)}$ (Jacobi nodes) was investigated in R. Sakai [3] and P. Vértesi [4]. In [4]

$$\Lambda_{n3}(X^{(\alpha,\beta)}) \sim \max \left(\log n, n^{3\alpha+3/2}, n^{3\beta+3/2} \right) \tag{1.7}$$

was proved. *For arbitrary interpolatory matrix* X, J. Szabados and A. K. Varma [5] proved the Faber type result

$$\Lambda_{n3}(X) \geqq c \log n$$

using the important observation

(1.9)

$$|h_{kn3}(x)| \geqq \frac{3}{2} \left(\sum_{i\neq k} \frac{1}{(x_k - x_i)^2} \right) (x - x_k)^2 |l_k^3(x)| \geqq \frac{3}{2} \frac{(x - x_k)^2}{(x_k - x_{k\pm1})^2} |l_k^3(x)|, \ 1 \leqq k \leqq n,$$

being valid for *arbitrary* X. Here with

$$\omega(x) = \omega_n(X, x) = c_n \prod_{k=1}^{n} (x - x_k), \quad n = 1, 2, \ldots, \tag{1.10}$$

$$\ell_k(x) = \ell_{kn}(X, x) = \frac{\omega(x)}{\omega'(x_k)(x - x_k)}, \quad 1 \leqq k \leqq n, \tag{1.11}$$

are the fundamental polynomials of Lagrange interpolation, and $k \pm 1$ means that the relation holds true with any of the signs (if it has meaning).

2. Results

2.1. The main goal of this paper is to get an estimation for λ_{n3} similar to (1.5). Namely we prove

THEOREM 2.1. *Let $\varepsilon > 0$ be any given number. Then for any interpolatory matrix $X \subset [-1,1]$ there exist sets $E_n = E_n(X,\varepsilon)$ with $|E_n| \leqq \varepsilon$ and $\eta = \eta(\varepsilon)$ such that*

$$\lambda_{n3}(X,x) > \eta \log n \quad \textit{whenever} \quad x \in R \setminus E_n. \tag{2.1}$$

COROLLARY 2.2. *If $\{S_n\}$ are arbitrary measurable sets then for any $X \subset$ $\subset [-1,1]$*

$$\int_{S_n} \lambda_{n3}(X,x)dx > (|S_n| - \varepsilon)\eta \log n \tag{2.2}$$

(ε and η are defined in Theorem 2.1).

2.2 REMARKS. 1. Considering the Chebyshev matrix, we can see that the order of (2.1) and (2.2) is the best possible.

2. If for a fixed matrix X and *odd m*

$$|h_{knm}(X,x)| \geqq c_m \frac{(x-x_k)^{m-1}}{(x_k - x_{k\pm 1})^{m-1}} |\ell_k^m(x)|, \quad 1 \leqq k \leqq n, \tag{2.3}$$

then with the previous notations one can prove the relation

$$\lambda_{nm}(X,x) > \eta \log n \quad \textit{if} \quad x \in R \setminus E_n \tag{2.4}$$

(see formula (3.9)). In many cases it is enough to suppose (2.3) for "many" values of k to prove (2.4). We omit the details. As an example, by R. Sakai [7] and R. Sakai, P. Vértesi [8] *relation* (2.3) *holds true* if $X = X^{(\alpha,\beta)}$, $\alpha, \beta > -1$, *m is arbitrary odd* (cf. 3.5 for recent results).

3. If m is even, $\Lambda_{nm}(X^{(\alpha,\beta)}) = O(1)$ if $-1/2 - 2/m \leqq \alpha, \beta \leqq -1/2 + 1/m$ and $|\alpha - \beta| \leqq 2/m$ (cf. P. Vértesi [4]) whence (2.1) cannot hold for arbitrary X.

3. Proof

3.1. The proof is based on P. Erdős, P. Vértesi [2], P. Vértesi [6] and relation (1.9). In what follows let $J_k = J_{kn} = [x_{k+1,n}, x_{kn}]$, $k = 0, 1, \dots, n$, $n = 1, 2, \dots, x_{0n} \equiv 1$, $x_{n+1,n} \equiv -1$. Further with $0 < q_k = q(J_{kn}) \leqq \frac{1}{2}$ let

$$J_k(q_k) = [x_{k+1} + q_k|J_k|, x_k - q_k|J_k|]. \tag{3.1}$$

Let $z_k = z_k(q_k)$ be defined by

$$|\omega_n(z_k)| = \min_{x \in J_k(q_k)} |\omega_n(x)|, \quad 0 \leqq k \leqq n, \tag{3.2}$$

finally let

$$|J_i, J_k| = \max(|x_{i+1} - x_k|, |x_{k+1} - x_i|), \quad 0 \leqq i, k \leqq n, \tag{3.3}$$

$$\varrho(J_i, J_k) = \min(|x_{i+1} - x_k|, |x_{k+1} - x_i|), \quad 0 \leqq i, k \leqq n. \tag{3.4}$$

3.2. For the "long" intervals we quote [6, Lemma 3.1].

LEMMA 3.1. *Let $|J_{kn}| > \delta_n := n^{-1/6}$ (k is fixed, $0 \leqq k \leqq n$). Then for any $\{s_n\}$ with $(\log n)^{-2} \leqq s_n \leqq 1/4$, we can define the index $t = t(k,n)$ and the set $e_{kn} \subset J_{kn}$ so that $|e_{kn}| \leqq 4s_n|J_{kn}|$, moreover*

$$|\ell_{tn}(x)| \geqq 3^{\sqrt{n}} \quad \text{if} \quad x \in J_{kn} \setminus e_{kn} \quad \text{and} \quad n \geqq n_1 \tag{3.5}$$

(n_1 is an absolute positive constant, large enough).

Let $A_n := \bigcup\limits_{|J_k| > \delta_n} J_k$ and $B_n := \bigcup\limits_{|J_k| > \delta_n} \left(J_k\left(\frac{1}{n}\right) \setminus e_k\right)$. Obviously $B_n \subset A_n$ further if $s_n = (\log n)^{-2}$, $|A_n \setminus B_n| \leqq \sum\limits_{k=0}^{n} \left(\frac{2}{n} + \frac{4}{\log^2 n}\right) |J_k| < \frac{1}{\log n}$, say $(n \geqq n_1)$. Now let $x \in B_n$. Then by (1.9) and (3.5)

$$\lambda_{n3}(x) \geqq |h_t(x)| \geqq \frac{3}{2} \frac{1}{2^2 n \cdot n^{1/6}} 3^{3\sqrt{n}} > \log^2 n, \quad \text{say} \quad (n \geqq n_1)$$

whence we obtain (2.1) for A_n apart from a set of measure $\leqq (\log n)^{-1}$ if $n \geqq n_1$.

3.3. Now we consider the "short" intervals, i.e. when $|J_k| \leqq \delta_n (= n^{-1/6})$. We prove (cf. [6, Lemma 3.2])

LEMMA 3.2. *If $1 \leqq k, r \leqq n-1$ then*

$$|h_k(x)| + |h_{k+1}(x)| \geqq \frac{q_k^3}{72} \left|\frac{\omega(z_r)}{\omega(z_k)}\right|^3 \frac{|J_k|}{|J_r, J_k|}, \quad n \geqq 6, \tag{3.6}$$

provided $x \in J_r(q_r)$, $\varrho(J_r, J_k) \geqq 2\delta_n$ and $|J_r| \leqq \delta_n$.

PROOF OF THE LEMMA. First we verify

$$\frac{2}{3} \leqq \left|\frac{z_r - x_s}{x - x_s}\right| \leqq 2 \quad \text{if} \quad x \in J_r(q_r), \quad s = k, k+1. \tag{3.7}$$

Indeed,

$$\left|\frac{z_r - x_s}{x - x_s}\right| \geqq \frac{|z_r - x_s| + \delta_n - \delta_n}{|z_r - x_s| + \delta_n} \geqq 1 - \frac{\delta_n}{3\delta_n} = \frac{2}{3},$$
$$\left|\frac{z_r - x_s}{x - x_s}\right| \leqq \frac{|z_r - x_s| - \delta_n + \delta_n}{|z_r - x_s| - \delta_n} \leqq 1 + \frac{\delta_n}{\delta_n} = 2$$

whence comes (3.7). By (3.2) and (3.7)

$$(3.8)\quad |\ell_s(x)| = \left|\frac{\omega(x)}{\omega(z_r)}\frac{z_r - x_s}{x - x_s}\right| |\ell_s(z_r)| \geqq \frac{2}{3}|\ell_s(z_r)|, \qquad s = k, k+1 \quad x \in J_r(q_r).$$

Using (1.9), (3.8), the definition of q_k and (3.7) we get (by $\ell_k(z_k) > 0$, and $\ell_{k+1}(z_k) > 0$)

$$|h_k(x)| + |h_{k+1}(x)| \geqq$$
$$\geqq \frac{3}{2}\left(\frac{2}{3}\right)^3 \left\{|\ell_k^3(z_r)|\frac{(x - x_k)^2}{(x_k - x_{k+1})^2} + |\ell_{k+1}^3(z_r)|\frac{(x - x_{k+1})^2}{(x_k - x_{k+1})^2}\right\} =$$
$$= \frac{4}{9}\left|\frac{\omega^3(z_r)}{\omega^3(z_k)}\right| \left\{\ell_k^3(z_k)\frac{|x_k - z_k|^3}{|z_r - x_k|}\frac{1}{(x_k - x_{k+1})^2}\frac{(x - x_k)^2}{(z_r - x_k)^2} + \right.$$
$$\left. + \ell_{k+1}^3(z_k)\frac{|x_{k+1} - x_k|^3}{|z_r - x_{k+1}|}\frac{1}{(x_k - x_{k+1})^2}\frac{(x - x_{k+1})^2}{(z_r - x_{k+1})^2}\right\} \geqq$$
$$\geqq \frac{4}{9}\left|\frac{\omega^3(z_r)}{\omega_3(z_k)}\right| \frac{q_k^3|J_k|^3}{|J_r, J_k|}\frac{1}{|J_k|^2}\left(\frac{1}{2}\right)^2 \left\{\ell_k^3(z_k) + \ell_{k+1}^3(z_k)\right\}, \qquad x \in J_r(q_r)$$

which is (3.6) by $\ell_k(z_k) + \ell_{k+1}(z_k) \geqq 1$ $(n \geqq 6)$, whence $\{\dots\} \geqq 1/8$. □

REMARK. Replacing formula (1.9) by (2.3), the same argument gives (3.9)

$$|h_{knm}(x)| + |h_{k+1,nm}| \geqq c(m)q_k\left|\frac{\omega(z_r)}{\omega(z_k)}\right|^m \frac{|J_k|}{|J_r, J_k|}, \quad n \geqq 6,\ 1 \leqq k,\ r \leqq n-1.$$

A reformulation of [6, Lemma 3.3] is

LEMMA 3.3. *Let $I_k = (a_k, b_k)$, $1 \leqq k \leqq t$, $t \geqq 2$ be any t disjoint intervals in $[-1, 1]$ with $|I_k| \leqq \delta$, $q_k = q$ $(1 \leqq k \leqq t)$ and $\sum_{k=1}^{t} |I_k| = \nu$. Let $\xi \geqq \delta$ be fixed. Supposing that for a certain integer $R \geqq 4$, we have $\nu \geqq 2^R\xi/q$, there exists an index s $(1 \leqq s \leqq t)$ such that*

$$\sum_{\substack{k=1\\ \varrho(I_s, I_k) \geqq \xi}}^{t} \frac{|I_k|}{|I_s, I_k|} \geqq \frac{R}{8}\nu - \frac{3}{4q}.$$

Now applying the argument of [2, 3.5–3.7], we can complete the proof for the interval [1, 1].

3.4. If $|x| \geqq 1+\varrho$, by $x^{n-1} = \sum_{k=1}^{n} x_k^{n-1} \ell_k(x)$ we get $(1+\varrho)^{n-1} \leqq \sum_{k=1}^{n} |\ell_k(x)|$ whence, using an argument similar to that at Part 3.2, we get $\lambda_{n3}(x) > \log^2 n$ (say) if $|x| \geqq 1+\varrho$. This completes the proof for the whole real line. □

3.5. Remark (December 3, 1991). Very recently we have proved Theorem 2.1 for arbitrary $\lambda_{nm}(X, x)$, m odd (cf. [9]).

REFERENCES

[1] Faber, G., Über die interpolatorische Darstellung stetiger Funktionen, *Jahresbericht der Deutschen Math. Ver.* **23** (1914), 192–210. *Jb. Fortschritte Math.* **45**, 381

[2] Erdős, P. and Vértesi, P., On the Lebesgue function of interpolation, *Functional Analysis and Approximation* (Oberwolfach, 1980), Internat. Ser. Numer. Math., Vol. 60, Birkhäuser, Basel–Boston, Mass., 1981, 299–309. *MR* **83k**: 41002

[3] Sakai, R., Hermite–Fejér interpolation prescribing higher order derivatives, *J. Approx. Theory* (to appear).

[4] Vértesi, P., Hermite–Fejér interpolations of higher order I, *Acta Math. Hungar.* **54** (1989), 135–152. *MR* **90k**: 41008

[5] Szabados, J. and Varma, A. K., On (0, 1, 2) interpolation in uniform metric, *Proc. Amer. Math. Soc.* **109** (1990), 975–979. *MR* **90k**: 41007

[6] Vértesi, P., New estimation for the Lebesgue function of Lagrange interpolation, *Acta Math. Acad. Sci. Hungar.* **40** (1982), 21–27. *MR* **84e**: 41005

[7] Sakai, R., Certain unbounded Hermite-Fejér interpolatory polynomial operators, *Acta Math. Hungar.* (to appear).

[8] Sakai, R. and Vértesi, P., Hermite-Fejér type interpolations of higher order III (to appear).

[9] Vértesi, P., Lebesgue function type sums of Hermite interpolation, *Acta Math. Hungar.* (to appear).

(Received May 31, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H-1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 455–462

DIE REGELFLÄCHEN DES E_n, DIE EINE AUS EBENEN KURVEN BESTEHENDE KONGRUENZSCHAR TRAGEN. II

H. WRESNIK

1. In [8] ist es dem Verfasser gelungen, alle EK-Regelflächen Φ des n-dimensionalen euklidischen Raumes E_n zu bestimmen und Parameterdarstellungen derselben anzugeben. Dabei zeigte sich, daß Φ in einem höchstens 5-dimensionalen Teilraum des E_n liegt. Desweiteren wurde die von den Trägerebenen der Scharkurven gebildete 3-Regelfläche Ψ betrachtet.

Im zweiten Teil dieser Arbeit wollen wir nun die EK-Regelflächen näher untersuchen. Wir geben die in [8] gefundenen Parameterdarstellungen nochmals an.

$$\vec{x}(u,v)=\begin{pmatrix} a_1(u)\\ 0\\ 0\\ B_2\sin u\end{pmatrix}+v\begin{pmatrix}\cos u\\ \sin u\\ k\\ 0\end{pmatrix}, \tag{1}$$

mit $a_1 \in C^3$, k und B_2 von Null verschiedene reelle Zahlen.

$$\vec{x}(u,v)=\begin{pmatrix} \cos\beta(A_1\cos u+B_1\sin u)\\ 0\\ \sin\beta(A_1\cos u+B_1\sin u)\\ A_2\cos u+B_2\sin u\end{pmatrix}+v\begin{pmatrix}\cos u\\ \sin u\\ k\\ 0\end{pmatrix}, \tag{2}$$

wobei $(A_2,B_2)\neq(0,0)$ und für $k=0$ noch zusätzlich $\beta\neq 0$ und $(A_1,B_1)\neq$ $\neq(0,0)$ gilt und

$$\vec{x}(u,v)=\begin{pmatrix} 0\\ 0\\ \cos\beta(A_1\cos u+B_1\sin u)\\ \sin\beta(A_1\cos u+B_1\sin u)\\ A_2\cos u+B_2\sin u\end{pmatrix}+v\begin{pmatrix}\cos u\\ \sin u\\ k\\ 0\\ 0\end{pmatrix}, \tag{3}$$

mit $(A_i,B_i)\neq(0,0)$, $i=1,2$, $\beta\neq 0$, $A_2B_1-A_1B_2\neq 0$ und im Fall $k=0$ gilt zusätzlich $\beta\neq\frac{\pi}{2}$.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 51N20; Secondary 53A05.

Key words and phrases. Ruled surfaces, plane family of congruences

Den Darstellungen (2) und (3) entnimmt man, daß die Scharkurven der EK-Schar Ellipsen sind, während dies für die EK-Regelflächen mit der Parameterdarstellung (1) nicht gilt, da ja die Scharkurven von der beliebigen Funktion $a_1(u)$ abhängen.[1]

Wie die Untersuchung der von den Scharkurventrägerebenen gebildeten 3-Regelfläche Ψ in [8] gezeigt hat, besitzen alle Trägerebenen genau für die EK-Regelflächen mit der Parameterdarstellung (1) einen gemeinsamen Fernpunkt. Nach den in [8] erwähnten Ergebnissen von H. Vogler sind also in den Fällen (2) und (3) die Scharkurven durch den Bewegungsparameter affin aufeinander bezogen.

Wir wollen nun die Normalprojektionen der EK-Regelflächen mit den Parameterdarstellungen (1) und (2) aus den Fernpunkten der z- bzw. t-Achse auf die dazu orthogonalen 3-Räume betrachten. Werden sie mit Φ_z und Φ_t bezeichnet, so erhält man im Fall (2)

$$\vec{x}(u,v) = \begin{pmatrix} \cos\beta(A_1 \sin u + B_1 \cos u) \\ 0 \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ 0 \end{pmatrix} \tag{4}$$

bzw.

$$\vec{x}(u,v) = \begin{pmatrix} \cos\beta(A_1 \cos u + B_1 \sin u) \\ 0 \\ \sin\beta(A_1 \cos u + B_1 \sin u) \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \end{pmatrix}. \tag{5}$$

Wie man nun unmittelbar erkennt, handelt es sich dabei um EK-Regelflächen in einem 3-dimensionalen euklidischen Raum. Die Ebenenkoordinaten der Scharkurventrägerebenen berechnen sich zu

$$(0 : A_2 v : (A_1 B_2 - A_2 B_1)\cos\beta + B_2 v : -(A_1 \cos\beta + v)v) \tag{6}$$

bzw.

$$(kv(A_1 v \cos\beta + v) : A_1 \sin\beta : B_1 \sin\beta : -(A_1 \cos\beta + v)). \tag{7}$$

In [2] bzw. [7] wurde gezeigt, daß die EK-Regelflächen des euklidischen 3-Raums nach der von den Trägerebenen der Scharkurven umhüllten Torse Λ in fünf Klassen unterteilt werden können. Ist nämlich Φ konoidal, so gibt es den

Fall $\mathcal{K}$: Λ ist ein irreduzibler Kegel 2. Klasse und

Fall $\mathcal{B}$: Λ ist ein Ebenenbüschel mit eigentlicher Trägergeraden.

Ist Φ hingegen eine Böschungsregelfläche, so gibt es den

Fall $\mathcal{T}$: Λ ist eine irreduzible Torse 3. Klasse,

Fall $\mathcal{Z}$: Λ ist ein irreduzibler parabolischer Zylinder 2. Klasse und

---

[1] Nur wenn $a_1(u)$ der Differentialgleichung $\dot{a}_1 + \dddot{a}_1 = 0$ genügt, so sind auch in (1) die Scharkurven Ellipsen.

Fall $\mathcal{P}$: Λ ist ein Parallelebenenbüschel.
Daher folgt aus (6), daß Φ_z, im allgemeinen vom Typ $\mathcal{K}$ ist; nur für $A_2=0$ oder $B_2=0$ und $\beta=\frac{\pi}{2}$ oder $B_1=B_2=0$ ist Φ_z vom Typ $\mathcal{B}$.

Φ_t hingegen ist, wie man (7) entnimmt, für $k=0$ vom Typ $\mathcal{B}$ und für $k\neq 0$ im allgemeinen vom Typ $\mathcal{Z}$, nur für $A_1=B_1=0$ oder $\beta=0$ vom Typ $\mathcal{P}$.

Für die EK-Regelflächen mit der Parameterdarstellung (1) erhält man für die Normalprojektionen Φ_z und Φ_t

$$\vec{x}(u,v)=\begin{pmatrix} a_1(u) \\ 0 \\ B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ 0 \end{pmatrix} \tag{8}$$

bzw.

$$\vec{x}(u,v)=\begin{pmatrix} a_1(u) \\ 0 \\ 0 \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \end{pmatrix}. \tag{9}$$

Da im konoidalen Fall nur die EK-Regelflächen vom Typ $\mathcal{B}$ beliebige Scharkurven besitzen können, ist (8) stets von diesem Typ, während (9) stets vom Typ $\mathcal{P}$ ist.

Bemerkung 1. Die Scharkurven der EK-Regelflächen Φ des E_n werden auf die Scharkurven der Bilder Φ_z und Φ_t abgebildet.

Bemerkung 2. Nur in (8) tritt jede (vom Typ her mögliche) EK-Regelfläche des E_3 als Bild auf.

2. Nun sollen die Zentral- bzw. die Kehlräume der Erzeugenden einer EK-Regelfläche Φ betrachtet werden. Nach [6] kann eine Erzeugende e von Φ zylindrisch sein, einen Kehlpunkt K oder einen Zentralpunkt Z besitzen. Dabei wird Φ im ersten Fall längs e von ein- und derselben Tangentialebene berührt, im zweiten Fall stimmen die Tangentialebenen in allen von K verschiedenen Punkten von e überein, während K singulär ist — die Erzeugende e heißt in diesem Fall torsal — und im letzten Fall besitzen verschiedene Punkte von e auch verschiedene Tangentialebenen an Φ. Der Zentralpunkt Z ist dadurch gekennzeichnet, daß seine Tangentialebene auf die asymptotische Ebene (das ist die Tangentialebene im Fernpunkt von e an Φ normal steht. e heißt in diesem Fall regulär.

Ist nun Φ durch

$$\vec{x}(u,v)=\vec{\mathrm{l}}(u)+v\,\vec{\mathrm{e}}(u) \tag{10}$$

mit $u\in I$, $v\in R$ und $\vec{\mathrm{e}}$, $\vec{\mathrm{l}}\in C^1$ gegeben, so gehört der Zentralpunkt bzw. der Kehlpunkt einer Erzeugenden e von Φ nach [1] zum Parameterwert $v=$ $=-\frac{\dot{\vec{\mathrm{l}}}\dot{\vec{\mathrm{e}}}}{\dot{\vec{\mathrm{e}}}^2}$.

Für die Striktionslinie s einer EK-Regelfläche Φ (das ist die Menge der Zentralpunkte bzw. Kehlpunkte), erhält man somit

$$\vec{s}(u) = \begin{pmatrix} a_1(u) + \dot{a}_1(u)\sin u\cos u \\ \dot{a}_1(u)\sin^2 u \\ k\dot{a}_1(u)\sin u \\ B_2 \sin u \end{pmatrix} \tag{11}$$

$$\vec{s}(u)= \begin{pmatrix} \cos\beta(A_1\cos^3 u + B_1 \sin u(1+\cos^2 u)) \\ \cos\beta(B_1\cos u - A_1 \sin u)\sin^2 u \\ A_1(\sin\beta\cos u - k\cos\beta\sin^2 u) + B_1(\sin\beta\sin u + k\cos\beta\sin u\cos u) \\ A_2\cos u + B_2\sin u \end{pmatrix} \tag{12}$$

und

$$\vec{s}(u) = \begin{pmatrix} 0 \\ 0 \\ \cos\beta(A_1\cos u + B_1\sin u) \\ \sin\beta(A_1\cos u + B_1\cos u) \\ A_2\cos u + B_2\sin u \end{pmatrix}. \tag{13}$$

Bemerkung 3. Die Striktionslinien der Normalprojektionen Φ_z und Φ_t von Φ sind die Projektionen der Striktionslinie s von Φ.

Bemerkung 4. Die Striktionslinie (12) der EK-Regelfläche (2) ist im allgemeinen eine rationale Kurve 6. Ordnung, jene der EK-Regelflächen (3) gehört stets der EK-Schar an und ist somit eine Ellipse.

Mit obigen Überlegungen lassen sich nun auch jene EK-Regelflächen mit den Parameterdarstellungen (1) und (2) angeben, deren Striktionslinien der EK-Schar angehören. Diese Regelflächen sind dann Regelflächen mit ebenen Kurven konstanten Striktionsabstandes.[2]

Da nach den Bemerkungen 1 und 3 die Scharkurven der EK-Schar und die Striktionslinie von Φ bei den obigen Normalprojektionen ihre Bedeutung beibehalten, kann man die Fragestellung auf die EK-Regelflächen des E_3 zurückspielen.

Betrachten wir zunächst die EK-Regelflächen mit der Parameterdarstellung (1), so findet man in [4] oder [9], daß die Striktionslinie der Regelfläche (8) genau dann auch Scharkurve ist, wenn

$$a_1(u) = A\ln\left|\tan\frac{u}{2}\right| \tag{14}$$

[2] Für die EK-Regelflächen in einem E_3 wurde diese Frage von H. Sachs [4] beantwortet.

mit $A \in R$ gilt. Da mit (13) auch die Striktionslinie der EK-Regelfläche (9) der EK-Schar angehört, erhält man als Parameterdarstellung der Lösungsflächen

$$\vec{x}(u,v) = \begin{pmatrix} A \ln \left|\tan \frac{u}{2}\right| \\ 0 \\ 0 \\ B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix}, \tag{15}$$

mit $A \in R$.

Nun betrachten wir die EK-Regelflächen mit der Parameterdarstellung (2). Mit analogen Überlegungen erkennt man, daß die Striktionslinie der durch (5) gegebenen EK-Regelfläche Φ_t genau dann der EK-Schar angehört, wenn für $k=0$ die Bedingung $\beta = \frac{\pi}{2}$ und für $k \neq 0$ die Bedingungen $\beta = \frac{\pi}{2}$ oder $A_1 = B_1 = 0$ erfüllt sind.

Mit diesen Werten besitzt auch die Normalprojektion Φ_z von Φ die gewünschte Eigenschaft, weshalb man als Parameterdarstellung der Lösungsflächen erhält

$$\vec{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ A_1 \cos u + B_1 \sin u \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{16}$$

mit $(A_2, B_2) \neq (0,0)$, wobei im Falle $k=0$ noch zusätzlich $(A_1, B_1) \neq (0,0)$ gelten muß.

3. In Verallgemeinerung des Dralls einer Regelfläche des E_3 definierte R. Koch in [1] für eine durch (10) gegebene Regelfläche des E_n eine Drallfunktion $d(u)$ durch

$$d(u) = \frac{1}{\dot{\vec{e}}^{\,2}} \sqrt{\dot{\vec{e}}^{\,2} \left(\dot{\vec{l}}^{\,2} - (\dot{\vec{l}}\,\vec{e})^2 \right) - (\dot{\vec{l}}\,\dot{\vec{e}})^2}\,. \tag{17}$$

Er ziegte weiters, daß eine torsale Erzeugende $e(u)$ durch $d(u) = 0$ und abwickelbare Regelflächen durch auf I identisch verschwindenden Drall gekennzeichnet sind.

Aus (18) erhält man somit für den Drall der EK-Regelflächen (1), (2) und (3)

$$d(u) = B_2 \cos u \tag{18}$$

$$d(u) = \sqrt{(A_2 \sin u - B_2 \cos u)^2 - \alpha (A_1 \sin u - B_1 \cos u)^2} \tag{19}$$

mit $\alpha := \sin\beta(2k\cos\beta\cos u + (k^2-1)\sin\beta)$ und

$$d(u) = \sqrt{(1 - k^2\cos^2\beta)(A_1 \sin u - B_1 \cos u)^2 + (A_2 \sin u - B_2 \cos u)^2}. \tag{20}$$

Damit erkennt man nun, daß im Intervall $I = [0, 2\pi]$ die EK-Regelflächen (1) genau zwei reelle torsale Erzeugenden besitzen, die zu den Parameterwerten $u = \frac{\pi}{2}$ und $u = \frac{3\pi}{2}$ gehören, während die EK-Regelflächen (2) und (3) zwei (im Sinne der algebraischen Wurzelzählung) torsale Erzeugenden besitzen, die für

$$\sin\beta(2k\cos\beta + (k^2-1)\sin\beta) < 0$$

bzw.

$$k^2\cos^2\beta - 1 < 0$$

nicht reell sind.

Die oben angegebenen Drallfunktionen ermöglichen es auch, die abwikkelbaren EK-Regelflächen anzugeben.

Da es im Fall (1) wegen (18) offensichtlich keine solchen Regelflächen gibt, wenden wir uns zunächst den EK-Regelflächen mit der Parameterdarstellung (3) zu. Aus (20) erhält man mit der Forderung $d(u) \equiv 0$ auf I die drei Bedingungen

$$\alpha A_1^2 + A_2^2 = 0, \quad \alpha B_1^2 + B_2^2 = 0, \quad \alpha A_1 B_1 + A_2 B_2 = 0$$

mit $\alpha = 1 - k^2\cos^2\beta$.

Daraus folgt $A_2 = \pm\sqrt{-\alpha}A_1$ und $A_2 = \pm\sqrt{-\alpha}B_1$, was aber wegen

$$A_1B_2 - A_2B_1 = \pm(-\alpha A_1B_1 + \alpha A_1B_1) = 0$$

keine Lösungsflächen liefert.

Es bleibt somit der Fall der EK-Regelflächen mit der Parameterdarstellung (2) zu diskutieren. Nach einigen Umformungen der Drallfunktion $d(u)$ erhält man aus der Forderung $d(u) \equiv 0$ die Bedingungen

$$\begin{gathered} A_2^2 - \mu A_1^2 + B_2^2 - \mu B_1^2 = 0 \\ -A_2^2 + \mu A_1^2 + B_2^2 - \mu B_1^2 = 0 \\ \mu A_1B_1 - A_2B_2 = 0 \\ \lambda A_1 B_1 = 0 \\ \lambda(A_1^2 + 3B_1^2) = 0 \\ \lambda(A_1^2 - B_1^2) = 0 \end{gathered} \tag{21}$$

mit $\lambda = 2k\sin\beta\cos\beta$ und $\mu = (k^2-1)\sin^2\beta$.

Aus $(21_{1,2,3})$ folgt $A_2 = \pm\sqrt{\mu}A_1$ und $B_2 = \pm\sqrt{\mu}B_1$. Ist nun $\lambda \neq 0$, so ergibt sich aus $(21_{5,6})$ $A_1 = B_1 = 0$ und damit $A_2 = B_2 = 0$, was aber keine EK-Regelfläche (vom Typ (2)) liefert.

Also muß $k\sin\beta\cos\beta = 0$ gelten. Ist $k = 0$, dann ist $\mu < 0$ und es gibt keine reellen Lösungsflächen. Ist $\beta = 0$, so ist $\mu = 0$ und damit gilt auch $A_2 = B_2 = 0$, was keine Lösung liefert.

Also muß $\beta = \frac{\pi}{2}$ gelten und damit gibt es genau für $\mu = k^2 - 1 > 0$ reelle Lösungsflächen mit der Parameterdarstellung

$$\overrightarrow{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ A_1 \cos u + B_1 \sin u \\ \sqrt{k^2-1}A_1 \cos u + \sqrt{k_2 - 1}B_1 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{22}$$

mit $k^2 > 1$ und $(A_1, B_1) \neq (0,0)$. Nach (16) handelt es sich bei diesen Regelflächen um Torsen mit ebenen Kurven konstanten Kehlkurvenabstandes. Die Kehlkurve stimmt mit der Leitkurve in (22) überein.

4. Zum Abschluß der Untersuchungen wollen wir noch jene EK-Regelflächen Φ bestimmen, deren Scharkurven Orthogonaltrajektorien der Erzeugenden von Φ sind.[3] Dazu muß wegen eines bekannten Ergebnisses von C. F. Gauss [5] nur geprüft werden, unter welchen Bedingungen die Leitkurve l in (10) Orthogonaltrajektorie der Erzeugenden von Φ ist.

Für die EK-Regelflächen mit der Parameterdarstellung (1) folgt aus dieser Bedingung $a_1(u) \cos u = 0$ und damit $a_1(u) \equiv c$, $c \in R$. Somit ergibt sich als Normalform der Lösungsflächen

$$\overrightarrow{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{23}$$

mit $k, B_2 \neq 0$.

Für die EK-Regelflächen vom Typ (2) ergeben sich aus der Forderung, daß l eine Orthogonaltrajektorie der Erzeugenden sein soll, die Bedingungen

$$kA_1 \sin\beta = kB_1 \sin\beta = A_1 \cos\beta = B_1 \cos\beta = 0$$

und damit erhält man für $k = 0$ und $\beta = \frac{\pi}{2}$ die Lösungsflächen

$$\overrightarrow{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ A_1 \cos u + B_1 \sin u \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ 0 \\ 0 \end{pmatrix} \tag{24}$$

mit $(A_i, B_i) \neq (0,0)$, $i = 1,2$, während $k = 0$ und $\beta \neq \frac{\pi}{2}$ wegen $A_1 = B_1 = 0$ keine Lösung liefert.

Ist hingegen $k \neq 0$, so ergibt sich $A_1 = B_1 = 0$ und damit erhält man als Parameterdarstellung der Lösungsflächen

$$\overrightarrow{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \end{pmatrix} \tag{25}$$

[3] Die analoge Fragestellung für Regelflächen des E_3 wurde von H. Sachs in [3] behandelt.

mit $k \neq 0$ und $(A_2, B_2) \neq (0,0)$.

Für die EK-Regelflächen mit der Parameterdarstellung (3) führt die Forderung für l auf die Bedingungen $kA_1 \cos\beta = kB_1 \cos\beta = 0$, was für $k = 0$ als Parameterdarstellung der Lösungsflächen

$$\vec{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ \cos\beta(A_1 \cos u + B_1 \sin u) \\ \sin\beta(A_1 \cos u + B_1 \sin u) \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{26}$$

ergibt mit $\beta \neq 0, \frac{\pi}{2}$ und $(A_i, B_i) \neq (0,0)$ für $i = 1,2$ und für $k \neq 0$ besitzen die Lösungsflächen die Parameterdarstellung

$$\vec{x}(u,v) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ A_1 \cos u + B_1 \sin u \\ A_2 \cos u + B_2 \sin u \end{pmatrix} + v \begin{pmatrix} \cos u \\ \sin u \\ k \\ 0 \\ 0 \end{pmatrix} \tag{27}$$

mit $k \neq 0$, $(A_i, B_i) \neq (0,0)$, $i = 1,2$.

LITERATURVERZEICHNIS

[1] KOCH, R., Striktionseigenschaften von Regelscharen im E^n, *Monatsh. Math.* **83** (1977), 133–146. *MR* **57** # 13734

[2] MEIRER, K., Über windschiefe Flächen mit ausgezeichneten Scharen ebener Kurven, *First Geometry Colloquium* (Graz, 1980), Ber. No. 164, Bericht, 163–167, Forschungszentrum Graz, Graz, 1981. *MR* **82j**: 53006

[3] SACHS, H., Die Strahlflächen mit durchwegs ebenen Orthogonaltrajektorien der Erzeugenden, *Arch. Math. (Basel)* **21** (1970), 437–445. *MR* **43** # 1057

[4] SACHS, H., Über die Kurven konstanten Striktionsabstandes auf windschiefen Flächen, *Monatsh. Math.* **74** (1970), 445–461. *MR* **43** # 3929

[5] BLASCHKE, W. und LEICHTWEISS, K., *Elementare Differentialgeometrie*, Fünfte vollständig neubearbeitete Auflage von K. Leichtweiss, Die Grundlehren der mathematischen Wissenschaften, Band 1, Springer-Verlag, Berlin–New York, 1973. *MR* **50** # 3122

[6] FRANK, H. und GIERING, O., Verallgemeinerte Regelflächen, *Math. Z.* **150** (1976), 261–271. *MR* **54** # 5984

[7] VOGLER, H., Räumliche Zwangläufe mit ebenen Bahnkurven, *Second Zagreb–Graz Meeting of Mathematicians* (Plitvice, 1980), Ber. No. 162, Bericht, 154–162, Forschungszentrum Graz, Graz, 1981. *MR* **82j**: 53008

[8] WRESNIK, H., Die Regelflächen des E_n, die eine aus ebenen Kurven bestehende Kongruenzschar tragen I, *Studia Sci. Math. Hungar.* **27** (1992), 133–142.

[9] WRESNIK, H., *Regelflächen mit ebenen Kongruenzscharen*, Dissertation, TU – Graz, 1988.

(Eingegangen am 22. Juni 1988)

INSTITUT FÜR GEOMETRIE
TECHNISCHE UNIVERSITÄT GRAZ
KOPERNIKUSGASSE 24
A-8010 GRAZ
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 463–470

LARGE PRIME FACTORS OF SUMS

I. Z. RUZSA

1. Introduction

Let $A \subset [1, N]$, $B \subset [1, N]$ be sets of integers, $|A| = k$, $|B| = l$. Let P denote the largest prime factor of the product

$$S = \prod_{a \in A,\, b \in B} (a + b). \tag{1.1}$$

We shall find estimates for P in terms of k, l and N, with particular attention on the case $k = l$.

Balog and Sárközy [1] proved

$$P > \frac{(kl)^{1/2}}{16 \log N}$$

assuming that $kl \gg N(\log N)^2$. For $kl \gg N^{2-c/\log\log N}$ Sárközy and Stewart [3] improved this to

$$P \gg \frac{(kl)^{1/2}}{\log R \log\log R}, \qquad R = \frac{N}{(kl)^{1/2}}.$$

For $k = l \gg N$ this shows $P \gg k$, and nothing better can be expected since A and B may consist of the numbers $1, 2, \dots, k$.

The proofs in the above papers were based on the large sieve and on Hardy and Littlewood's method of exponential sums. These methods are extremely flexible and can be used to find small prime divisors or to estimate the number of prime divisors in a given range as well, or to find a prime p that appears in some sum $a + b$, $a \in A$, $b \in B$ with a high exponent.

Our aim is to present an elementary method that gives sharper estimates of P for $k = l$ and works for a much wider range. On the other hand, it is quite rigid and seemingly it cannot be applied to anything else.

The principal results are the following.

THEOREM 1. *Let A, B be sets of integers, $|A| = |B| = k$. Let $N = \max\{|a+b|: a \in A, b \in B\}$. Let S be the product defined in* (1.1). *Assume that $S \neq 0$ and let P be the largest prime divisor of S. We have*

$$P \geqq c\frac{k \log k}{\log N} \log \frac{\log N}{\log k} \tag{1.2}$$

with an absolute constant $c > 0$.

Moreover, for $k > C(\varepsilon)\sqrt{N}$ we have

$$P > (2/e - \varepsilon)k. \tag{1.3}$$

THEOREM 2. *Let A, B be disjoint sets of integers, $|A| = |B| = k$. Let $N = \max\{|a-b|: a \in A, b \in B\}$. Put*

$$D = \prod_{a \in A,\, b \in B} |a - b| \tag{1.4}$$

and let P be the largest prime divisor of D. The estimate (1.2), *and for $k > C(\varepsilon)\sqrt{N}$ the estimate* (1.3) *holds.*

REMARKS. 1) Since we did not restrict the sign of elements of A and B, Theorems 1 and 2 are equivalent (we replace B by $-B$). In the proofs it will be more convenient to work with differences.

2) (1.2) guarantees $P \to \infty$ if $k(\log\log N)/(\log N) \to \infty$. It guarantees $P \gg k$ for $k \gg N^\alpha$ with any fixed $\alpha > 0$.

3) If $A = B = \{1, 2, \dots, k\}$, then the largest prime divisor of S is the largest prime below $2k$. One is tempted to conjecture that $P > (2-\varepsilon)k$ at least for $k \gg N$ and summands of equal signs. If different signs are allowed, then $A = \{-k/2, \dots, k/2\}$, $B = \{-k, -k+1, \dots, -k/2+1, k/2+1, \dots, k\}$ yields $P \sim 3k/2$.

2. An inequality

We prove the following inequality.

LEMMA 2.1. *Let $a_1 < a_2 < \ldots < a_k$ and $b_1 < b_2 < \ldots < b_l$ be integers, $k \leqq l$. Put*

$$U = \prod_{i<j} (a_j - a_i),$$

$$V = \prod_{i<j} (b_j - b_i),$$

$$D = \prod_{\substack{i,j \\ a_i \neq b_j}} |a_i - b_j|\,.$$

We have

$$|D| \geqq U^{l/k} V^{k/l} (4/9)^k. \tag{2.1}$$

We start with some preparation.

LEMMA 2.2. *Let f be a real function defined for $x \geqq 0$, increasing and concave. Let $a_1, \dots, a_k$, $b_1, \dots, b_l$ be real numbers. We have*

$$\begin{aligned} &2 \sum_{i=1}^{k} \sum_{j=1}^{l} f(|a_i - b_j|) \geqq \\ &\geqq \frac{l}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} f(|a_i - a_j|) + \frac{k}{l} \sum_{i=1}^{l} \sum_{j=1}^{l} f(|b_i - b_j|). \end{aligned} \tag{2.2}$$

PROOF. We may assume that f is bounded; otherwise, we can replace it by $f^*(x) = f(\min(x, M))$, where $M = \max(|a_i| + |b_j|)$. A bounded f can be written in the form

$$f(x) = c - dg(x), \tag{2.3}$$

with real c, $d \geqq 0$ and a function g satisfying $g(0) = 1$, $g(x) \geqq 0$, decreasing and convex.

Extending this g to negative numbers by $g(-x) = g(x)$ we obtain a positive definite function. (These functions are known as Pólya's characteristic functions; see Pólya [2].) This means that the inequality

$$\sum_{i=1}^{n} \sum_{j=1}^{n} g(x_i - x_j) z_i z_j \geqq 0 \tag{2.4}$$

holds for arbitrary real x_i and complex z_i. We apply (2.4) with $n = k + l$, $x_i = a_i$ for $i = 1, \dots, k$, $x_{k+j} = b_j$, $j = 1, \dots, l$. Put $z_i = \sqrt{(l/k)}$ for $i \leqq k$ and $z_i = -\sqrt{(k/l)}$ for $i > k$. We obtain

$$2 \sum \sum g(a_i - b_j) \leqq \frac{l}{k} \sum \sum g(a_i - a_j) + \frac{l}{k} \sum \sum g(b_i - b_j).$$

We substitute (2.3) to obtain (2.2). □

LEMMA 2.3. *Let $a_1, \dots, a_k$, $b_1, \dots, b_l$ be integers, f an increasing real-valued function defined on nonnegative integers and satisfying*

$$f(n+1) - 2f(n) + f(n-1) \geqq 0 \quad (n \geqq 1). \tag{2.5}$$

Then we have (2.2).

PROOF. We extend f to nonnegative real numbers by linear interpolation between consecutive integers, and apply Lemma 2.2. □

PROOF of Lemma 2.1. We apply Lemma 2.3 for the function

$$f(n)=\begin{cases}\log n & \text{if } n\geqq 3,\\ \frac{\log 3}{3}n & \text{if } n\leqq 2.\end{cases}$$

This modified logarithm satisfies (2.5), moreover $f(n)\geqq \log n$ for $n\geqq 1$ and $f(0)=0$. We obtain

$$\sum_{\substack{i,j\\ a_i\neq b_j}} f\left(|a_i-b_j|\right)\geqq \frac{l}{k}\sum_{i<j}\log(a_j-a_i)+\frac{k}{l}\sum_{i<j}\log(b_j-b_i). \tag{2.6}$$

We omitted the factor 2 because from the pairs (a,a') and (a',a) we include only one.

Now consider the sum

$$\sum_{\substack{i,j\\ a_i\neq b_j}} f\left(|a_i-b_j|\right)-\log|a_i-b_j|\,. \tag{2.7}$$

For a fixed $a_i=a$ the only values of b_j that can give a nonzero summand in (2.7) are $a+1$, $a+2$, and altogether these give at most

$$2\left(\frac{\log 3}{3}+\frac{2}{3}\log 3-\log 2\right)=\log\frac{9}{4}.$$

Thus the sum in (2.7) is $\leqq k\log(9/4)$. Combining this inequality with (2.6) we obtain (2.1) in logarithmic form. □

REMARK. The factor $(4/9)^k$ cannot be omitted, as the example $A=\{2,4,6,8\}$, $B=\{1,3,7,9\}$ shows. I have no conjecture about its optimal value.

3. Inequalities for the exponents of primes

Let k, l, a_i, b_j, U, V and D be the same as in the previous section. For any integer $m\geqq 2$ let $u(m)$ denote the number of pairs (i,j), $1\leqq i<j\leqq k$ such that $m\mid a_i-a_j$, and define $v(m)$ similarly for the numbers b_i. Finally, let $w(m)$ be the number of those pairs (i,j), $1\leqq i\leqq k$, $1\leqq j\leqq l$ for which $m\mid a_i-b_j$.

LEMMA 3.1. *We have*

$$\frac{l}{k}u(m)+\frac{k}{l}v(m)\geqq w(m)-\frac{k+l}{2}. \tag{3.1}$$

If $w(m) = 0$, *then we have*

$$\frac{l}{k}u(m) + \frac{k}{l}v(m) \geqq \frac{2kl}{m} - \frac{k+l}{2}. \tag{3.2}$$

PROOF. Let x_t $(t = 1, \ldots, m)$ be the number of integers a_i satisfying $a_i \equiv t \pmod m$, and define y_t similarly for the b_i. We have

$$u(m) = \sum \binom{x_t}{2}, \quad v(m) = \sum \binom{y_t}{2}, \quad w(m) = \sum x_t y_t. \tag{3.3}$$

Now we apply the inequality

$$xy \leqq \frac{1}{2}\left(\frac{l}{k}x^2 + \frac{k}{l}y^2\right) = \frac{l}{k}\binom{x}{2} + \frac{k}{l}\binom{y}{2} + \frac{lx + ky}{2}$$

for the numbers x_t, y_t. Summing up and taking into account $\sum x_t = k$, $\sum y_t = l$, in view of (3.3) we obtain (3.1).

Assume now $w(m) = 0$. Let K be the number of subscripts t with $x_t \neq 0$. By the square-mean inequality we have

$$\sum x_t^2 \geqq \frac{1}{K}\left(\sum x_t\right)^2 = k^2/K.$$

Since we must have $x_t y_t = 0$ for all t, the number of nonzero values of x_t is at most $m - K$. Now similarly we find

$$\sum y_t^2 \geqq \frac{1}{m-K}\left(\sum y_t\right)^2 = k^2/(m-K).$$

Adding these equations we get

$$\frac{l}{k}\sum x_t^2 + \frac{k}{l}\sum y_t^2 \geqq \frac{klm}{K(m-K)} \geqq \frac{4kl}{m}. \tag{3.4}$$

Applying (3.3) we can deduce (3.2). $\square$

For a prime p, let α_p, β_p and δ_p denote the exponent of p in U, V and D, respectively.

LEMMA 3.2. *We have*

$$\frac{l}{k}\alpha_p + \frac{k}{l}\beta_p \geqq \delta_p - \frac{k+l}{2}\left[\frac{\log N}{\log p}\right]. \tag{3.5}$$

If $\delta_p = 0$, *then we have*

$$\frac{l}{k}\alpha_p + \frac{k}{l}\beta_p \geqq \frac{2kl}{p} - \frac{k+l}{2}. \tag{3.6}$$

PROOF. We have

$$\alpha_p = \sum u(p^j), \quad \beta_p = \sum v(p^j), \quad \delta(p) = \sum w(p^j).$$

Since $|a_i - b_j| \leqq N$, the largest exponent of p that can occur in an $a_i - b_j$ is at most $\left[\frac{\log N}{\log p}\right]$. Summing up the inequalities (3.1) for this range we obtain (3.5).

To deduce (3.6) we use (3.2), keeping only the term $m = p$. $\square$

4. Proof of the theorems

We prove Theorem 2. We retain the notations of the previous sections, but we fix $k=l$. We also assume $P<k$, since otherwise (1.2) and (1.3) are obvious.

We start from Lemma 2.1. We decompose U, V and D into primes and take the logarithm to find

$$\sum_p (\alpha_p+\beta_p-\delta_p)\log p \leqq k \log 9/4 < k.$$

We use Lemma 3.2 to estimate the coefficient of $\log p$. We use (3.5) for $p\leqq P$, (3.6) for $P<p\leqq 2k$, and drop the terms with $p>2k$ (they are nonnegative by the assumption $P<k$). This gives the inequality

$$\sum_{P<p<2k}\left(\frac{2k}{p}-1\right)\log p < 1+\sum_{p\leqq P}\left[\frac{\log N}{\log p}\right]\log p. \tag{4.1}$$

To show (1.2) we estimate the right side by

$$\leqq 1+\pi(P)\log N \leqq c\frac{P\log N}{\log P}.$$

On the left side we use $2k/p-1>k/p$ for $p\leqq k$ and drop the terms with $k<p<2k$. To estimate the sum we use

$$\sum_{P<p<k}\frac{\log p}{p} > \log(k/P)-c.$$

In this way we obtain

$$k\log(k/p) \leqq c\left(\frac{P\log N}{\log P}+k\right).$$

By introducing the variable $u=e^{-c}k/p$ this can be written as

$$u\log u < c'\frac{\log N}{\log P}.$$

To solve this inequality we use the following observation: if $x\log x \leqq$ $\leqq y$, $x>0$, $y>1$, then $x\leqq 2y/(1+\log y)$. (Hint: add x times the inequality $\log(y/x)\leqq (y/x)-1$.) This yields

$$u=e^{-c}k/P \leqq c''\frac{\log N}{\log P}\Big/\left(1+\log\left(\frac{\log N}{\log P}\right)\right).$$

Rearranging and using the assumption $P < k$ we get

$$\frac{P}{\log P} \gg \frac{k}{\log N}\left(1 + \log\left(\frac{\log N}{\log k}\right)\right),$$

from which (1.2) immediately follows.

To prove the sharp estimate (1.3) for large k we must be more careful. We assume $k \gg C\sqrt{N}$; from (1.2) we already know $P \gg k$, hence $P > \sqrt{N}$ if C is large enough.

In the right side of (4.1) we distinguish the cases $p > \sqrt{N}$, for which $\left[\frac{\log N}{\log p}\right] = 1$ and the rest; this yields the upper estimate

$$1 + \vartheta(P) - \vartheta(\sqrt{N}) + \pi(\sqrt{N}) \log N,$$

where $\vartheta(x) = \sum_{p \leqq x} \log p$. Now (4.1) yields

$$2k \sum_{P<p<2k} \frac{\log p}{p} < 1 + \vartheta(2k) - \vartheta(\sqrt{N}) + \pi(\sqrt{N}) \log N.$$

Using the usual estimates for primes we find

$$2k \log \frac{2k}{P} < 2k + O(\sqrt{N}).$$

For $u = \frac{2k}{eP}$ this yields

$$u \log u \ll \sqrt{N}/k.$$

Now if $u \leqq 1$, the proof is finished. Otherwise we use $u \log u \geqq \log u$ to obtain

$$u < e^{c\sqrt{N}/k} < 1 + \varepsilon$$

for $k > C(\varepsilon)\sqrt{N}$. $\square$

As we have already observed, Theorem 1 follows from Theorem 2 via replacing B by $-B$.

5. Remarks on the case $k \neq l$

Assume $k < l$. An obvious estimate for P can be obtained by omitting some elements of B, retaining only k of them. In the considerations of Sections 2 and 3 there was nothing specific for $k = l$. On the other hand, observe that (3.6) yields a nontrivial estimate only for $p < \frac{4kl}{k+l} < 4k$.

It is easy to improve (3.6) for large p. Assume $p > 2k$. Then in (3.4) (with $m = p$) we have $K \leqq k < p/2$, thus $K(p-K) \leqq k(p-k)$, which gives the estimate

$$\frac{l}{k}\alpha_p + \frac{k}{l}\beta_p \geqq \frac{lp}{2(p-k)} - \frac{k+l}{2} = \frac{k(p-k-l)}{2(p-k)},$$

nontrivial up to $p < k + l$. This indeed leads to an improvement of the bound given by Theorems 1–2, by about a factor $\log(l/k)$, rather than l/k that can be conjectured.

The problem lies more in inequality (3.5). It is sharp, with equality occurring if, in the notation of Section 3, x_t is proportional to y_t. For $m > k$ this means that the l elements of B lie in $k < m$ residue classes mod m, which cannot typically happen. To exclude this case one may resort to the large sieve. This works, however, only if $kl \gg N$, and it yields only a minor improvement over Balog and Sárközy's result quoted in the Introduction (the logarithm can be omitted from the denominator). I conjecture that the real order of P is l unless k is very small, say, that we have $P > c(\varepsilon)l$ with $c(\varepsilon) > 0$ under the assumption $l \geqq k > N^{\varepsilon}$.

REFERENCES

[1] Balog, A. and Sárközy, A., On sums of sequences of integers II, *Acta Math. Hungar.* **44** (1984), 169–179. *MR* **86g**: 11056b

[2] Pólya, G., Remarks on characteristic functions, *Proc. Berkeley Sympos. Math. Stat. Probability* 1945, 1946 (ed. J. Neyman), Univ. California Press, Berkeley, 1949, 115–123. *MR* **10**–463

[3] Sárközy, A. and Stewart, C. L., On divisors of sums of integers II, *J. Reine Angew. Math.* **365** (1986), 171–191. *MR* **88f**: 11088

(Received May 2, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 27 (1992), 471–477

SEPARATORY SUBLATTICES AND SUBSEMILATTICES

L. LIBKIN* and I. MUCHNIK

Abstract

A subalgebra of an algebra is called *separatory* if its complement is a subalgebra as well. In this paper we study separatory sublattices and subsemilattices.

1. Introduction

Let $\langle A, \Omega \rangle$ be an algebra with a carrier A and a signature Ω. We say that a subalgebra $\langle B, \Omega \rangle$ of $\langle A, \Omega \rangle$ is *separatory* if its complement $A - B$ is a subalgebra again, i.e. if it is closed under all operations from Ω. This paper studies the properties of separatory sublattices and subsemilattices.

Let us introduce some terminology. Throughout the paper S will denote a semilattice and the semilattice operation will be denoted by $\circ$. We use the customary letter L for lattices and $\wedge$ and $\vee$ for the infimum and supremum operations. We assume that the semilattices are ordered by letting $x \leqq y$ iff $x \circ y = y$, i.e. S is always a join-semilattice. $J(S)$ stands for the set of *irreducible* elements of S, i.e. $x \in J(S)$ iff $x = y \circ z$ implies $x = y$ or $x = z$. $J(L)$ is the set of *join-irreducible* elements of L and $M(L)$ is the set of *meet-irreducible* elements where $J(L) = J(\langle L, \vee \rangle)$ and $M(L) = J(\langle L, \wedge \rangle)$. The lattice of subalgebras of $\langle A, \Omega \rangle$ will be denoted by $Sub\langle A, \Omega \rangle$. An algebra $\langle A, \Omega \rangle$ is called *idempotent* if all operations in Ω are idempotent. The signatures will be omitted if they are understood.

Incomparability of x and y will be denoted by $x \| y$. M_3 and N_5 stand for the five-element non-distributive and non-modular lattices (diamond and pentagon). An interval is a set $[x, y] = \{z \mid x \leqq z \text{ and } z \leqq y\}$. The lattice of intervals of a lattice L is denoted by $Int\, L$.

The notion of a separatory subalgebra is similar to that of a half-space and this observation motivated our study of the separation properties of separatory subalgebras. An analogue of the classical separation theorem from convex analysis holds for arbitrary semilattices which is shown in Section 2. In the lattice case, however, separatory sublattices even may not exist, and

*While working on the final version, this author was supported by NSF Grants IRI-86-10617 and CCR-90-57570 and ONR Grant NOOO14-88-K0634.

Akadémiai Kiadó, Budapest

the separation theorem holds only in a very restrictive subclass of distributive lattices. But a slight relaxation of the statement of the separation theorem characterizes the variety of distributive lattices. Finally, in Section 4, we use interval representations of sub(semi)lattices of finite distributive (semi)lattices of [1, 5, 7] to describe the separatory sub(semi)lattices. The conditions saying that tops (or tops and bottoms in the lattice case) of the intervals used in the representations form a chain (chains in the lattice case) are necessary and sufficient for a sub(semi)lattice to be separatory.

For arbitrary idempotent algebras the notion of being separatory can be characterized through pseudocomplements. In fact,

LEMMA 1. *Let $\langle A, \Omega\rangle$ be an idempotent algebra. Then a subalgebra $\langle B, \Omega\rangle$ is separatory iff it has a pseudocomplement in $Sub\langle A, \Omega\rangle$.*

PROOF. Let B have a pseudocomplement C in $Sub\,A$. Assume B is not separatory, i.e. $\omega(x_1, \dots, x_n) \in B$ for some $\omega \in \Omega$ and $x_1, \dots, x_n \notin B$. Clearly, at least one $x_i \notin C$ and the subalgebra $\{x_i\}$ is not a subset of C although $\{x_i\} \cap B = \emptyset$ which contradicts the definition of a pseudocomplement. The other direction is immediate. □

2. Separatory subsemilattices

In this section we state and prove the separation theorem for semilattices. Based on that theorem, we list some properties of the separatory subsemilattices.

THEOREM 1 (Separation Theorem). *Let S be an arbitrary semilattice and S_1, S_2 its disjoint subsemilattices. Then S_1 and S_2 can be separated by a separatory subsemilattice, i.e. there exists a separatory subsemilattice $S' \subseteqq S$ such that $S_1 \subseteqq S'$, $S_2 \subseteqq S - S'$.*

PROOF. Let $S_0(x)$ denote the minimal subsemilattice of S containing a subsemilattice S_0 and $x \in S$, i.e. $S_0(x) = S_0 \cup \{x\} \cup \{x \circ s \mid s \in S_0\}$. We claim that for every $x \notin S_1 \cup S_2$ either $S_1(x) \cap S_2 = \emptyset$ or $S_2(x) \cap S_1 = \emptyset$. Assume that both intersections are nonempty, i.e. there exist $x_1 \in S_1(x) \cap S_2$ and $x_2 \in S_2(x) \cap S_1$. Then $x_1 = x \circ s_1$, $x_2 = x \circ s_2$ where $s_1 \in S_1$, $s_2 \in S_2$. Then $S_2 \ni (x \circ s_1) \circ s_2 = (x \circ s_2) \circ s_1 \in S_1$. This contradiction proves our claim. To finish the proof, consider a family $\{(S_1', S_2') \mid S_1', S_2' \in Sub\,S,\ S_1', S_2'$ disjoint, $S_1 \subseteqq S_1',\ S_2 \subseteqq S_2'\}$. By Zorn's lemma, this family has a maximal element (S', S''). Moreover, $S' \cup S'' = S$ follows from the claim proved above. Thus, S' separates S_1 from S_2.

This result immediately implies that any semilattice with two or more elements has a proper separatory subsemilattice. Since a single element is a subsemilattice, for any $S_0 \in Sub\,S$ and any $x \notin S_0$ there is a separatory subsemilattice S_x' separating S_0 from x. Obviously, $S_0 = \cap\{S_x' \mid x \notin S_0\}$. Therefore, COROLLARY 1. *Every subsemilattice of an arbitrary semilattice is an intersection of separatory subsemilattices.* □

COROLLARY 2. *If S is finite, every element of $M(Sub\,S)$ is a separatory subsemilattice.* □

REMARK. Matrix representations of separatory subsemilattices of finite semilattices were studied in [3]. Those representations were based on the result for the Boolean case, or, to be more precise, for the semilattices of form $\langle \mathbf{2}^U - \{\emptyset\}, \cup \rangle$, where U is a finite set. Separatory subsemilattices of such semilattice are in 1-1 correspondence with $|U| \times |U|$ symmetric 0-1 absolutely determined matrices where "absolutely determined" means that every submatrix has a saddle-point. This correspondence is established by $S \mapsto A = \|a_{ij}\|_{i,j=1,\dots,n}$ such that $a_{ij} = 1$ iff $\{x_i, x_j\} \in S$, where $U = \{x_1, \dots, x_n\}$.

3. Separatory sublattices

In contrast to the semilattice case, separatory sublattices may not always exist. Any prime ideal is a separatory sublattice; hence, the existence of separatory sublattices is guaranteed for distributive lattices. But one can easily verify that the lattice *Part*(4) of equivalence relations on a four-element set does not have proper separatory sublattices. And neither does any finite partition lattice *Part*(n). Therefore, one could not expect the statement of the separation theorem for lattices to hold in a large class of lattices, and, indeed, only a very restrictive class admits the separation theorem. However, intervals can be separated in any distributive lattice.

We say that a lattice L satisfies the *separation condition*, if for any disjoint sublattices $L_1, L_2 \subseteq L$ there exists a separatory sublattice L' such that $L_1 \subseteq L'$, $L_2 \subseteq L - L'$.

A lattice is called *series-parallel*, if it does not contain a subposet whose diagram looks like the letter **N** [6]. This is equivalent to saying that L is series-parallel if $a \vee c = b$, $b \wedge d = c$ hold for no four distinct elements $a, b, c, d \in$ $\in L$. The following result appeared in [4]:

THEOREM 2. *A lattice L satisfies the separation condition iff it is distributive and series-parallel.* □

In the finite case a distributive series-parallel lattice is an ordinal sum of chains and four-element Boolean lattices, its width being one or two. This is indeed a very small class, and our intention is to enlarge it by weakening the separation condition.

We say that a lattice L satisfies the *interval separation condition* if for any two disjoint intervals $I_1, I_2 \subseteq L$ there exists a separatory sublattice $L' \subseteq L$ such that $I_1 \subseteq L'$, $I_2 \subseteq L - L'$.

THEOREM 3. *An arbitrary lattice satisfies the interval separation condition iff it is distributive.*

PROOF. The interval separation condition is inherited by sublattices. Since five-element non-modular and non-distributive lattices do not satisfy it (which can be checked directly as there are only a finite number of intervals), any lattice with the interval separation condition is distributive.

Conversely, let L be distributive and $[x_1, y_1] \cap [x_2, y_2] = \emptyset$, $x_1 \leqq y_1$, $x_2 \leqq y_2$. We claim that there are an ideal $\mathcal{I}$ and coideal $\mathcal{D}$ such that $\mathcal{I} \cap \mathcal{D} = \emptyset$, one interval is contained in $\mathcal{I}$ and the other in $\mathcal{D}$. In fact, if $x_1 \| y_2$, then $\mathcal{D} = [x_1)$ and $\mathcal{I} = (y_2]$. If $x_2 \| y_1$ or $y_1 \leqq x_2$, then $\mathcal{D} = [x_2)$ and $\mathcal{I} = (y_1]$. Finally, if $y_1 \geqq x_2$ and x_1 and y_2 are comparable, then y_2 cannot be greater than x_1 (otherwise $y_1 \wedge y_2$ would belong to both intervals). Hence, $y_2 \leqq x_1$ and $\mathcal{D} = [x_1)$, $\mathcal{I} = (y_2]$. Having proved the existence of $\mathcal{I}$ and $\mathcal{D}$, note that there is a prime ideal $\mathcal{P}$ such that $\mathcal{I} \subseteqq \mathcal{P}$, $\mathcal{D} \cap \mathcal{P} = \emptyset$, see [2]. Hence, $\mathcal{P}$ is a separatory sublattice separating $[x_1, y_1]$ from $[x_2, y_2]$.

REMARK. It was proved in [8] that the separation condition for the lattice of intervals of a *complete* lattice is equivalent to distributivity.

It was shown above that meet-irreducible subsemilattices of finite semilattices are separatory. Again, in the lattice case the situation is much more complicated. For example,

PROPOSITION 1. *Let L be a finite lattice. Suppose that $M(Sub\,L)$ contains only separatory sublattices. Then L does not have a sublattice isomorphic to M_3.*

PROOF. Assume that L does have a sublattice $\{0, a, b, c, e\}$ isomorphic to M_3, where $0 < a, b, c < e$. Let L' be a maximal sublattice of L which contains e and c but does not contain 0, a and b. If $L' = L_1 \cap L_2$ and $L' \neq L_1, L_2$, then both L_1 and L_2 contain an element from $\{0, a, b\}$ and $0 \in L'$ since $a \wedge c = b \wedge c = 0$. This contradiction shows that $L' \in M(Sub\,L)$, but L' is not a separatory sublattice ($a, b \notin L'$, but $a \vee b = e \in L'$).

COROLLARY 3. *If L is a finite modular lattice and all meet-irreducible sublattices are separatory, then L is distributive.* □

The requirement that all meet-irreducible elements be separatory sublattices turns out to be equivalent to distributivity, if we consider $Int\,L$ instead of $Sub\,L$. Furthermore, those meet-irreducibles can be described explicitly.

PROPOSITION 2. *Let L be a finite lattice. The following are equivalent:*
1) *Every element of $M(Int\,L)$ is a separatory sublattice;*
2) *$M(Int\,L)$ consists of all prime ideals and coideals of L;*
3) *L is distributive.*

PROOF. $1 \Rightarrow 3$. Suppose that 1 holds and L is not distributive. Then L contains either N_5 or M_3 as a sublattice. Therefore each interval of N_5 (or

M_3) must be represented as an intersection of separatory sublattices of N_5 (or M_3) which are intervals. But one can easily check that this is not the case.

$3 \Rightarrow 2$. It follows from the proof of Theorem 3 that each interval of a finite distributive lattice is an intersection of prime ideals and coideals. Hence, it is enough to prove that any prime ideal (coideal) is meet-irreducible in $Int\, L$ for a finite distributive lattice L. Assume that $\mathcal{P} \subsetneqq L$ is a prime ideal and $\mathcal{P} = [x, y] \cap [u, w]$. Then $\mathcal{P} = (a]$ where $a \in M(L)$. Hence $x \wedge u = 0$ (the least element of L) and $y \wedge w = a$, i.e. either $\mathcal{P} = [x, y]$ or $\mathcal{P} = [u, w]$. Thus, $\mathcal{P} \in M(Int\, L)$. Similarly, any prime coideal is meet irreducible in $Int\, L$.

$2 \Rightarrow 1$ is immediate. The proof is now complete.

4. Separatory sublattices and subsemilattices in distributive case

Recall that a semilattice S is called *distributive*, if $x \leqq y \circ z$ implies the existence of $y' \leqq y$ and $z' \leqq z$ such that $x = y' \circ z'$ [2]. It is known that the finite distributive semilattices and only they arise as distributive lattices considered as join-semilattices, i.e. they are $\langle L, \vee\rangle$ where $\langle L, \vee, \wedge\rangle$ is a finite distributive lattice.

The next two theorems from [1, 5, 7] describe the structure of sub(semi)-lattices of finite distributive (semi)lattices.

THEOREM 4 ([1, 5]). *Let S be a finite distributive semilattice. Then its subsemilattices and only they can be represented as*

$$S_0 = S - \bigcup_{i \in I} [a_i, x_i], \quad \forall i \in I\colon a_i \in J(S),\ x_i \in S. \tag{1}$$

THEOREM 5 ([7]). *Let L be a finite distributive lattice. Then its sublattices and only they can be represented as*

$$L_0 = L - \bigcup_{i \in I} [a_i, x_i], \quad \forall i \in I\colon a_i \in J(L),\ x_i \in M(L). \tag{2}$$

Surprisingly little has to be added to the above representations in order to characterize the separatory sub(semi)lattices.

THEOREM 6. *Let S be a finite distributive semilattice. Then its separatory subsemilattices and only they can be represented as in* (1) *where $\{x_i \mid i \in I\}$ is a chain in S.*

THEOREM 7. *Let L be a finite distributive lattice. Then its separatory sublattices and only they can be represented as in* (2) *where $\{a_i \mid i \in I\}$ and $\{x_i \mid i \in I\}$ are chains in L.*

Before we prove these theorems, we need an auxiliary combinatorial lemma.

LEMMA 2. *Let $B = \{b_1, \ldots, b_n\}$ be a subset of a finite semilattice S. Suppose that for any pair $b', b'' \in S$ there exists an element $b \geqq b' \circ b''$ such that we are allowed to exchange either b' or b'' for b. Then, using only exchanges allowed, B can be transformed into a chain whose minimal element belongs to B in a finite number of steps.*

PROOF. We claim that B can be transformed into a set B_0 such that $b_i \in B_0$ for an index $i \in 1, \ldots, n$ and $b \geqq b_i$ for any $b \in B_0$. Prove this claim by induction on n. If B has a unique minimal element, no exchanges are needed and we are done. Suppose there are at least two minimal elements, b_i and b_j, in B. By induction hypothesis, using a finite number of exchanges, $B - \{b_j\}$ can be transformed into $B' = \{b'_k \mid k \in K\}$ such that b_l is the minimal element in B' for some index $l \in 1, \ldots, n$, $l \neq j$. Notice that $b_l \not\leqq b_j$. If $b_l \geqq \geqq b_i$, then $B_0 = B' \cup \{b_i\}$ and we are done. If $b_l \| b_i$, consider all pairs (b'_k, b_j), $k \in K$. If for some k an exchange of b_j for $b \geqq b_j \circ b'_k$ is allowed, $B_0 = B' \cup \cup \{b\}$. If for any pair the only allowed exchange is b'_k for $b''_k \geqq b'_k \circ b_j$, then $B_0 = \{b''_k \mid k \in K\} \cup \{b_j\}$. The claim is proved.

Now the lemma can be proved by induction on n. If n is 1 or 2, we are done. For $n > 2$ transform B into $B' = \{b'_k \mid k \in K\}$ such that $b'_l = b_i$ and $b'_k \geqq b'_l$ for all $k \in K$. By induction hypothesis, $B' - \{b'_l\}$ can be transformed into a chain B'' whose minimal element is in B' and therefore not less than b_i. Then $B'' \cup \{b_i\}$ is the desired chain whose minimal element b_i is in B. $\square$

PROOF OF THEOREM 6. If $\{x_i \mid i \in I\}$ is a chain, then $\bigcup\limits_{i \in I} [a_i, x_i]$ is a subsemilattice of S and S_0 (1) is separatory. Conversely, if S_0 (1) is separatory, consider two intervals $[a_i, x_i]$ and $[a_j, x_j]$ from the representation (1). Let $x = x_i \circ x_j$. It is enough to show that either $[a_i, x] \cap S_0 = \emptyset$ or $[a_j, x] \cap S_0 = \emptyset$. Then applying Lemma 2 will yield the desired result immediately.

Suppose that there exist $z_i \in [a_i, x] \cap S_0$ and $z_j \in [a_j, x] \cap S_0$. Let $z = z_i \circ z_j$. Since $z \leqq x = x_i \circ x_j$ and S is distributive, there exist $x'_i \leqq x_i$ and $x'_j \leqq x_j$ such that $z = x'_i \circ x'_j$. Now, $z \geqq a_i$ and $z \geqq a_j$ imply $z = (x'_i \circ a_i) \circ (x'_j \circ a_j)$. Notice that $x'_i \circ a_i \in [a_i, x_i]$ and $x'_j \circ a_j \in [a_j, x_j]$. Therefore, $z \notin S_0$ because S_0 is separatory. But on the other hand, $z = z_i \circ z_j \in S_0$. This contradiction proves our claim and Theorem 6. $\square$

PROOF OF THEOREM 7. Let both $\{a_i \mid i \in I\}$ and $\{x_i \mid i \in I\}$ be chains. Then it can be easily shown that $\bigcup\limits_{i \in I} [a_i, x_i]$ is a sublattice of L. Therefore, L_0 (2) is separatory.

Conversely, let L_0 (2) be separatory. Let $[a_i, x_i]$ and $[a_j, x_j]$ be two distinct intervals in the representation (2). We are going to show that there is such $x \in M(L)$ that $x \geqq x_i \vee x_j$ and either $[a_i, x] \cap L_0 = \emptyset$ or $[a_j, x] \cap L_0 = = \emptyset$. Reason by contradiction. Assume that for any $x \in M(L)$, $x \geqq x_i \vee x_j$, there exist $a_i(x) \in [a_i, x] \cap L_0$ and $a_j(x) \in [a_j, x] \cap L_0$. Let $a^k = \bigwedge(a_k(x) \mid x \geqq x_i \vee x_j, x \in M(L))$, $k = i, j$. Then $a^k \in [a_k, x_i \vee x_j] \cap L_0$, $k = i, j$. Clearly, $a^i, a^j \in L_0$.

Let $p = a^i \vee a^j \in L_0$. Then $p \wedge x_i \in [a_i, x_i]$ and $p \wedge x_j \in [a_j, x_j]$. Hence, L_0 being separatory, $(p \wedge x_i) \vee (p \wedge x_j) \notin L_0$. By distributivity, $(p \wedge x_i) \vee (p \wedge x_j) = p \wedge (x_i \vee x_j) = p \in L_0$. This contradiction proves our claim.

The existence of such $x \in M(L)$ and Lemma 2 show that we can turn the tops of the intervals in (2) into a chain without changing the bottoms, i.e. $L_0 = L - \cup([a_i, c_i] \mid i \in I)$, where $\{c_i \mid i \in I\}$ is a chain. Since the dual lattice L^* is distributive, the same is true of the bottoms of the intervals, which gives us a desired representation. □

Acknowledgements. The authors would like to thank Professor E. T. Schmidt and an anonymous referee for several helpful comments.

REFERENCES

[1] Demetrovics, J., Libkin, L. and Muchnik, I., Functional dependencies and the semilattice of closed classes, *MFDBS* 89 (Visegrád, 1989), *Lecture Notes in Computer Science*, Vol. 364, Springer-Verlag, Berlin–New York, 1989, 136–147. *MR* **90h**: 68010

[2] Grätzer, G., *General Lattice Theory*, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Math. Reihe, Bd. 52, Birkhäuser Verlag, Basel–Stuttgart, Springer Verlag, 1978. *MR* **80c**: 06001a

[3] Gurvich, V. and Libkin, L., Absolutely determined matrices, *Math. Soc. Sci.* **20** (1990), 1–18.

[4] Libkin, L., Separation theorems for lattices, *MTA SZTAKI Közlemények* **39** (1988), 93–100. *MR* **90c**: 06010

[5] Libkin, L., and Muchnik, I., Separatory subsemilattices and their properties, *MTA SZTAKI Közlemények* **39** (1988), 83–92. *MR* **90c**: 06009

[6] Lonc, Z. and Rival, I., Chains, antichains and fibres, *J. Combin. Theory Ser. A* **44** (1987), 207–228. *MR* **88c**: 06006

[7] Rival, I., Maximal sublattices of finite distributive lattices. II, *Proc. Amer. Math. Soc.* **44** (1974), 263–268. *MR* **50** #12843

[8] Van de Vel, M., Binary convexities and distributive lattices, *Proc. London Math. Soc.* (3) **48** (1984), 1–33. *MR* **85c**: 52004a

(Received July 1, 1989)

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF PENNSYLVANIA
200 SOUTH 33RD STREET
PHILADELPHIA, PA 19104–6389
U.S.A.
Email: libkin @ saul.cis.upenn.edu

BIO-MOLECULAR CENTER
BOSTON UNIVERSITY
BOSTON, MA 02215
U.S.A.

Typeset by TYPOTEX Ltd, Budapest
PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda Vállalat, Budapest

Studia Scientiarum Mathematicarum Hungarica

Volume 27

Akadémiai Kiadó, Budapest

1992

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Volume 27

CONTENTS

RECENTLY ACCEPTED PAPERS

YADAV, S. P., Saturation order of some approximation processes in certain Banach spaces

LASSAK, M. and VÁSÁRHELYI, É., Covering a plane convex body with negative homothetical copies

KY, N. X. and HOA, N. Q., Uljanov-type embedding theorems in the theory of weighted polynomial approximation

NICULESCU, S. P. and OMEY, E., Asymptotic behaviour of the distribution function of multivariate nonhomogeneous renewal processes

ZHOU, S. P., On the density of rational combinations of $\{x^{\lambda_n}\}$ for some complex sequence λ_n

SLEZÁK, B., Implicit function theorem in uniform spaces

MASTROIANNI, G. and VÉRTESI, P., Simultaneous pointwise approximation of Lagrange interpolation

MEENAKSHI, A. and INDIRA, R., On Schur complements in a conjugate EP matrix

FÉNYES, T., On the Fourier transform of the Bessel function with respect to the order

DEÁK, J., Spaces from pieces I

ATANASSOV, K. T., On the Sperner's lemma

MILICI, S. and TUZA, ZS., Coverable graphs

CONTENTS

HU ISSN 0081-6906

Index: 26.748