Studia Scientiarum Mathematicarum Hungarica

VOLUME 35
NUMBERS 1–2
1999

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 and French. It is published in yearly volumes of four issues (mostly double numbers published semiannually) by

AKADÉMIAI KIADÓ
H–1117 Budapest, Prielle Kornélia u. 4

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H–1364 Budapest

Tel.: +36 1 318 2875 Fax: +36 1 317 7166
e-mail: merza@math-inst.hu

Subscription information

Orders should be addressed to

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

For 1999 volume 35 is scheduled for publication. The subscription price is $ 164.00, air delivery plus $ 20.00.

Coden: SSMHAX May. 1999
Vol: 35 **Pages**: 1–260
Numbers: 1–2 **Whole**: 68

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 1–15

3-POLYTOPES WITH CONSTANT FACE WEIGHT

J. IVANČO and M. TRENKLER

Dedicated to Professor E. Jucovič on the occasion of his 70th birthday

Abstract

The weight of a face α in a 3-polytope is the sum of degrees of vertices which are incident with α. In the present paper we determine the number of different regular 3-polytopes for which the weight of each face is $w \geqq 9$. In the nonregular case we have similar results if $9 \leqq w \leqq 21$ and if $28 \leqq w$.

1. Introduction

Rosenfeld [5] and also Jendroľ and Jucovič [3] investigated 3-polytopes or maps with constant weight of edges. As an analogy, E. Jucovič suggested to study convex 3-polytopes with constant weight of faces and some basic properties of such 3-polytopes are studied in [1] by his student Bauer. The aim of the present paper is to contribute to the description of such 3-polytopes.

For a convex 3-polytope M, let $V(M)$, $E(M)$, $F(M)$ and $\Delta(M)$ (or only V, E, F and Δ) denote the vertex set of M, the edge set of M, the face set of M and the maximum degree of M, respectively. Let α be a k-gonal face of M, which is incident with vertices $A_1, \ldots, A_k$, where $\deg(A_1) \leqq \ldots \leqq \deg(A_k)$. The *type* of α is defined as the k-tuple of positive integers $(d_1, \ldots, d_k)$, where $d_i = \deg(A_i)$, for all $i = 1, \ldots, k$. The *charge* of α is $c(\alpha) := k - 6 + \sum_{i=1}^{k} \frac{2d_i - 6}{d_i}$ and the *weight* of α is $w(\alpha) := \sum_{i=1}^{k} d_i$.

Since Euler's formula for M can be rewritten as

$$\sum_{\alpha \in F} c(\alpha) = -12 \tag{1}$$

we get the following assertion (see also [4]).

LEMMA 1. *Every 3-polytope contains a face whose charge is negative.* □

If $\alpha \in F(M)$ and $X \subseteqq V(M)$, then the set of vertices of X which are incident with α is denoted by $X \cap \alpha$. The set $X \subseteqq V(M)$ is called the *strong*

1991 *Mathematics Subject Classification.* Primary 52B10; Secondary 05C10.
Key words and phrases. Polytope, constant weight of faces.

0081-6906/99/$ 5.00

set of M if all negative charge faces of M belong to F_X, where $F_X := \{\alpha \in F(M);\ X \cap \alpha \neq \emptyset\}$. The *relative charge* of a vertex A in a strong set X is

$$c_X(A) := \sum_{\alpha \in F_{\{A\}}} \frac{c(\alpha)}{|X \cap \alpha|}.$$

Since for every strong set X the formula (1) can be rewritten as

$$\sum_{A \in X} c_X(A) + \sum_{\alpha \in F - F_X} c(\alpha) = -12 \tag{2}$$

we get the following assertion.

LEMMA 2. *Every strong set of a 3-polytope contains a vertex whose relative charge is negative.* □

Since the constancy of weight of faces is a combinatorial property, it is useful to identify convex 3-polytopes with their graphs and next to consider polyhedral graphs, i.e. plane 3-connected graphs (see Grünbaum [2]). For a positive integer w, let $\mathfrak{M}(w)$ be the family of all polyhedral graphs whose all faces have the weight w. Similarly, by $\mathfrak{M}(w; k)$ we denote the family of all k-gonal graphs belonging to $\mathfrak{M}(w)$.

Let $V_k(M)$ denote the set of vertices of degree k in a polyhedral graph M and $v_k(M) = |V_k(M)|$. We can prove the following auxiliary result.

LEMMA 3. *If* $M \in \mathfrak{M}(w)$, *then*

$$\sum_{i \geq 3} \left(i^2 - \frac{w}{2} i + w \right) v_i(M) = 2w. \tag{3}$$

PROOF. Since a vertex A contributes to the weight of $\deg(A)$ faces we have

$$w|F| = \sum_{\alpha \in F} w(\alpha) = \sum_{A \in V} (\deg(A))^2 = \sum_{i \geq 3} i^2 v_i(M).$$

Hence $|F| = \frac{1}{w} \sum_{i \geq 3} i^2 v_i(M)$. Similarly, $|V| = \sum_{i \geq 3} v_i(M)$ and $|E| = \frac{1}{2} \sum_{i \geq 3} i v_i(M)$. Manipulations with these equalities and with Euler's formula yield the assertion. □

2. Regular cases

Evidently, $\mathfrak{M}(w; k) = \emptyset$ for $k \notin \{3, 4, 5\}$, because every 3-polytope contains a face incident with at most five vertices. Let us deal with the remaining cases.

THEOREM 1 ([1]).

$$|\mathfrak{M}(w;3)| = \begin{cases} 0 & \text{for } w \leqq 8 \text{ and } w = 10, \\ 1 & \text{for } w = 9,\ 11 \leqq w \leqq 14,\ 21 \leqq w \leqq 22 \text{ and } w \geqq 24, \\ 2 & \text{for } 16 \leqq w \leqq 20 \text{ and } w = 23, \\ 3 & \text{for } w = 15. \end{cases}$$

PROOF. It can easily be seen that all faces of $M \in \mathfrak{M}(w;3)$ are of the same type. Moreover, all neighbours of a vertex with odd degree must have the same degree as the other neighbours. Since M must contain a negative charge face, only the following types of faces can occur: $(3,3,3)$, $(3,2k,2k)$ for $2 \leqq k \leqq 5$, $(4,4,k)$ for $k \geqq 4$, $(4,6,2k)$ for $3 \leqq k \leqq 5$, $(5,5,5)$ and $(5,6,6)$.

Case 1. Faces of M are of type $(3,3,3)$ (of type $(5,5,5)$). Evidently, M is a graph of the tetrahedron (the regular icosahedron, respectively).

Case 2. Faces of M are of type $(3,2k,2k)$ where $k \in \{3,4,5\}$ (of type $(5,6,6)$). Then the graph $M_1 = M - V_3(M)$ ($M_1 = M - V_5(M)$) has faces of type (k,k,k) (of type $(3,3,3,3,3)$), i.e. M_1 is a graph of the regular $\frac{4k}{6-k}$-hedron (the regular dodecahedron). Therefore M is a graph of the Kleetope over the regular $\frac{4k}{6-k}$-hedron (the regular dodecahedron, respectively).

Case 3. Faces of M are of type $(4,4,k)$ for $k \geqq 3$ (of type $(3,4,4)$ for $k = 3$). Then the graph $M - V_k(M)$ is a circuit with k vertices. Thus M is a graph of the bipyramid with $2k$ faces.

Case 4. Faces of M are of type $(4,6,2k)$ for some $k \in \{3,4,5\}$. Then the graph $M - V_6(M)$ is homeomorphic to a graph of the regular $\frac{4k}{6-k}$-hedron (each edge of $\frac{4k}{6-k}$-hedron is replaced with a path of length 3). Therefore M is a dual graph of the Archimedean solid $(4,6,2k)$. □

THEOREM 2.

$$|\mathfrak{M}(w;4)| = \begin{cases} 0 & \text{for } w \leqq 11, \\ 1 & \text{for } 12 \leqq w \leqq 13 \text{ and } w \geqq 17, \\ 3 & \text{for } w = 14, \\ 4 & \text{for } w = 15, \\ \infty & \text{for } w = 16. \end{cases}$$

PROOF. Let M be a polyhedral graph belonging to $\mathfrak{M}(w;4)$. Negative charge faces are only of the following types: $(3,3,3,w-9)$ for $w \geqq 12$, $(3,3,4,w-10)$ for $14 \leqq w \leqq 21$, $(3,3,5,w-11)$ for $16 \leqq w \leqq 18$ and $(3,4,4,w-11)$ for $15 \leqq w \leqq 16$. Whence, $w - 11 \leqq \Delta(M) \leqq w - 9$.

Case 1. Suppose $\Delta(M) = w-9$. Then all faces incident with a maximum degree vertex A are of type $(3,3,3,w-9)$. Their neighbouring faces must be of the same type. Therefore there exists a maximum degree vertex B $(\neq A)$ which is incident with them. Thus M is a graph of the dual of antiprism.

Case 2. Suppose $\Delta(M) = w-10$. Then all faces incident with a maximum degree vertex A are of type $(3,3,4,w-10)$. The configurations of faces incident with A are illustrated in Figure 1. The first configuration

Fig. 1

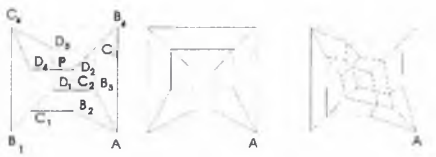


Fig. 2

is not possible in M because the face α is of type $(3,3,3,w-9)$ and so $\deg(B) = w-9 > \Delta$. The second configuration is possible only if $w = 14$ because $\deg(C) = w-10$, $\deg(D) \geqq w-11$ and so $w = w(\beta) \geqq 4+(w-10)+3+(w-11)$. The third configuration is possible for $14 \leqq w \leqq 15$ because $\deg(P) = \deg(Q) = w-11$ and so $w = w(\gamma) \geqq 4+2(w-11)+3$. Now, the faces of $F_{\{A\}}$ start a direct reconstruction of M. For example, consider the third configuration and $w = 14$ (see the first graph in Figure 2). Denote the neighbours of A successively by B_1, B_2, B_3, B_4. The fourth vertex of the face incident with B_i, A, B_{i+1} (subscripts being taken modulo 4) is denoted by C_i, $i = 1,\ldots,4$. The vertices C_i and C_{i+1} are distinct, otherwise there would be a 2-gonal face incident with B_i and C_i. Similarly, C_i and C_{i+2} are distinct because M is 3-connected. Let D_i, $i = 1,\ldots,4$, be the fourth vertex of the face incident with C_i, B_{i+1}, C_{i+1}. Since $\deg(B_i) = 3$, $\deg(C_i) = 4$ and $w = 14$, the vertex D_i is of degree 3. M is a polyhedral 4-gonal graph and so $D_i \notin \{A\} \cup \bigcup_{i=1}^{4} \{B_i, C_i\}$. D_i and D_{i+1} are distinct, otherwise there would be a 2-gonal face incident with D_i and C_{i+1}. As neighbours of D_i (D_{i+2}) are C_i, C_{i+1} (C_{i+2}, C_{i+3}) and $\deg(D_i) = 3$, D_i and D_{i+2} are distinct, too. Finally, denote by P the fourth vertex of the face α incident with D_1, C_2, D_2. Obviously, $\deg(P) = 4$. M is a polyhedral 4-gonal graph and so $P \notin \{A\} \cup \bigcup_{i=1}^{4} \{B_i, C_i, D_i\}$. Moreover, the neighbouring faces of α which belong to $F_{\{P\}}$ are incident with D_2, C_3, D_3 and D_1, C_1, D_4. Therefore, the neighbours of P must be D_1, D_2, D_3 and D_4, which complete the reconstruction. The same trivial ideas can be used in other cases. The details are therefore left to the reader. So, in this case, M is one of the graphs which are illustrated in Figure 2.

Case 3. Suppose $\Delta(M) = w - 11$ and $w \in \{17, 18\}$. Then M contains a face of type $(3, 3, 5, w - 11)$ and $V_5(M)$ is a strong set of M. Vertices of degree 5 and Δ are not adjacent because the contrary enforces either the first or the second of configurations in Figure 3 (encircled numbers are degrees of corresponding vertices). However, $w(\alpha) > w$ and $w(\beta) > w$, which is a contradiction. Moreover, a vertex of degree 5 is incident with at most one face of type $(3, 3, 5, w - 11)$ because two faces of this type enforce the third configuration in Figure 3 and $w(\gamma) > w$ (if two faces of type $(3, 3, 5, w-11)$ does not have a common edge then they have a common neighbouring face

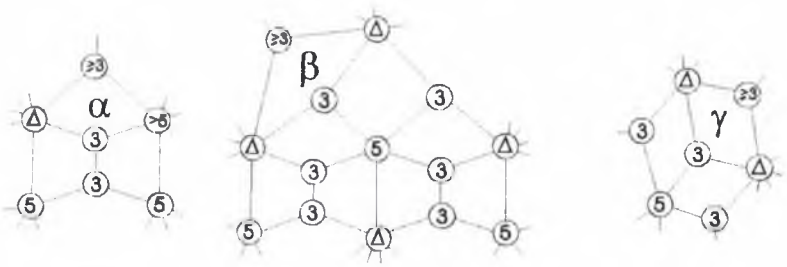


Fig. 3

which must be of the same type). If A is a vertex of $V_5(M)$ and a face $\alpha \in F_{\{A\}}$ is not of type $(3, 3, 5, w - 11)$, then only the following types of α can occur: $(3, 4, 5, 5)$, $(3, 4, 5, 6)$, $(4, 4, 5, 5)$, $(3, 5, 5, 5)$. The smallest possible contribution of α to the relative charge of A is either $\frac{1}{20}$ (α is of type $(3, 4, 5, 5)$ and $|V_5(M) \cap \alpha| = 2$), for $w = 17$, or $\frac{2}{15}$ (α is of type $(3, 5, 5, 5)$ and $|V_5(M) \cap \alpha| = 3$), for $w = 18$. Hence, either $c_{V_5(M)}(A) \geqq 4 \cdot \frac{1}{20} - \frac{1}{5} = 0$ or $c_{V_5(M)}(A) \geqq 4 \cdot \frac{2}{15} - \frac{2}{35} > 0$, in contradiction to Lemma 2. Therefore, no desired graph exists in this case.

Case 4. Suppose $w = 15$ and $\Delta(M) = 4$. Then all faces of M are of type $(3, 4, 4, 4)$. By Euler's formula and Lemma 3, $v_3(M)=8$, $v_4(M)=18$. First assume, that every vertex with degree 4 has a neighbour with degree 3. Therefore, $3v_3(M) - v_4(M) = 6$ vertices with degree 4 have two neighbours with degree 3. Hence, $M - V_3(M)$ is 6-gonal plane graph with 12 vertices of degree 3 and 6 vertices of degree 2. Moreover, each face of $M - V_3(M)$ is incident with at most one vertex of degree 2. Thus, $M - V_3(M)$ is homeomorphic to a 3-regular plane graph with twelve vertices and eight at least 5-gonal faces, in contradiction to Euler's formula. Therefore M contains a vertex $A \in V_4(M)$ with all neighbours also in $V_4(M)$. Similarly as in Case 1 (details are left to the reader) the faces of $F_{\{A\}}$ start a direct (besides one step) reconstruction of M. So, in this case M is one of two graphs which are

illustrated in Figure 4.

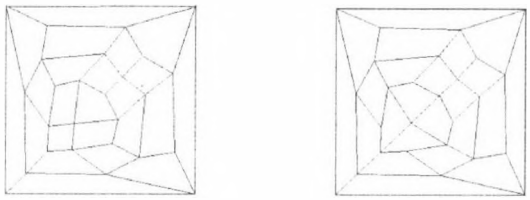


Fig. 4

Case 5. Suppose $w = 16$ and $\Delta(M) = 5$. In this case there are two infinite families of graphs, which are obtained from the two graphs in Figure 5 by inserting an arbitrary number of circuits into the belt of quadrangles (depicted by heavy lines) as it is illustrated in Figure 5. □

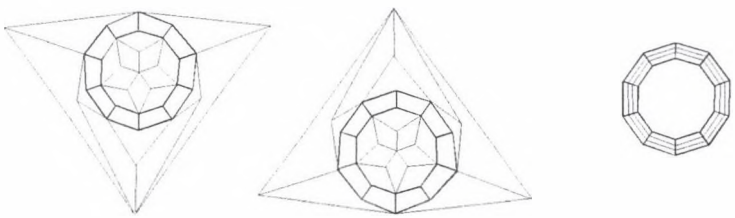


Fig. 5

REMARK. If the edge e of the 3-polytope M is incident with vertices A, B and faces α, β, the *complete weight* of e is defined as the number $a+b+m+n$, where a, b are degrees of A, B and m, n are numbers of edges of α, β. Let us denote by $\mathfrak{S}(w)$ the family of all 3-polytopes (polyhedral graphs) with every edge having complete weight w. For a polyhedral graph M, let the radial of M be the graph M^r such that there exists a bijective mapping $\psi\colon V(M) \cup F(M) \to V(M^r)$ satisfying: a face α is incident with a vertex A in M if and only if the vertices $\psi(\alpha)$ and $\psi(A)$ are adjacent in M^r. It can easily be seen that M belongs to $\mathfrak{S}(w)$ if and only if the radial of M belongs to $\mathfrak{M}(w;4)$. Therefore, one can describe all 3-polytopes belonging to $\mathfrak{S}(w)$ by the 3-polytopes of $\mathfrak{M}(w;4)$. Note that in the proof above there are described all 3-polytopes of $\mathfrak{M}(w;4)$ besides $w = 16$ and $\Delta = 5$. Starting from a vertex of degree 5 and verifying all possible neighbouring faces on every step, each graph of $\mathfrak{M}(16;4)$ with $\Delta = 5$ can be reconstructed (the reconstruction is very tedious). Using this method we get twenty graphs, illustrated in Figure 6, and two infinite families described in the proof. So, the description of $\mathfrak{M}(w;4)$ and $\mathfrak{S}(w)$ is complete.

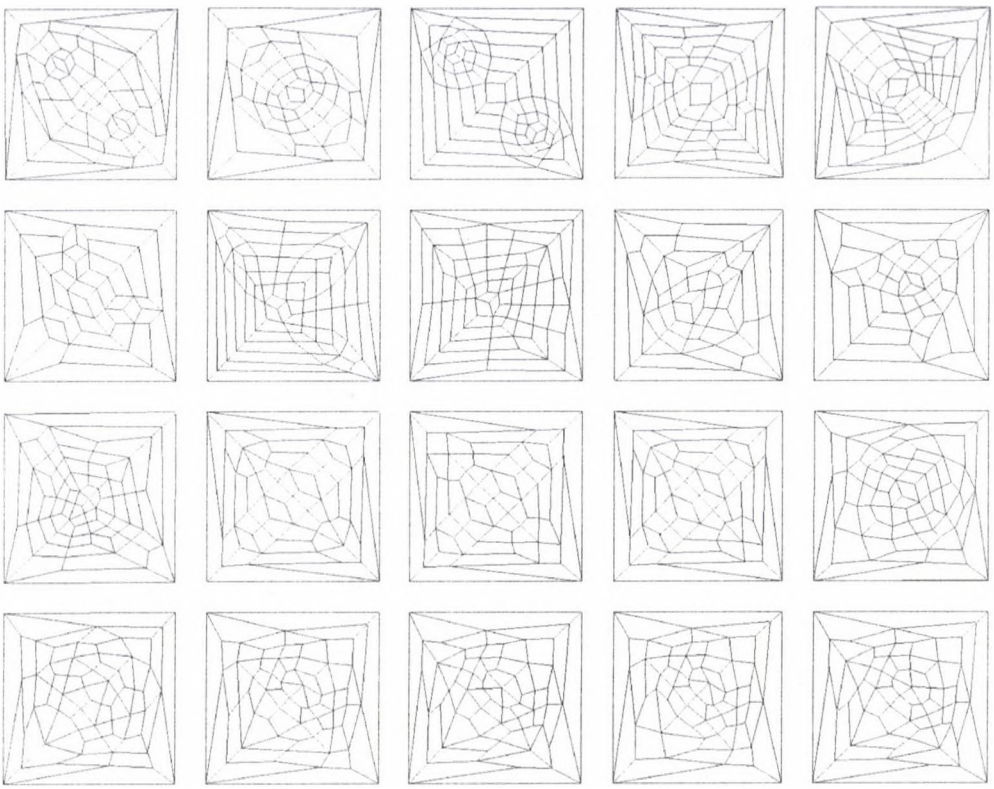


Fig. 6

THEOREM 3.

$$|\mathfrak{M}(w;5)| = \begin{cases} 0 & \text{for } w \leqq 14 \text{ and } w \geqq 18, \\ 1 & \text{for } w = 15 \text{ and } w = 16, \\ \infty & \text{for } w = 17. \end{cases}$$

PROOF. Let M be a polyhedral graph belonging to $\mathfrak{M}(w;5)$. Negative charge faces are only of the following types: $(3,3,3,3,3)$, $(3,3,3,3,4)$ and $(3,3,3,3,5)$. Whence, $15 \leqq w \leqq 17$.

Case 1. Suppose $w = 15$. Then all faces of M are of type $(3,3,3,3,3)$. Thus, M is a graph of the regular dodecahedron.

Case 2. Suppose $w = 16$. Then all faces of M are of type $(3,3,3,3,4)$. By Euler's formula and Lemma 3 , $v_3(M) = 32$, $v_4(M) = 6$. Moreover, each vertex $A \in V_4(M)$ enforce the configuration in Figure 7. $M - V_4(M)$ is 12-gonal plane graph with $v_3(M) - 4v_4(M) = 8$ vertices of degree 3 and $4v_4(M) = 24$ vertices of degree 2. Therefore, $M - V_4(M)$ is homeomorphic to a graph of

a cube (each edge of the cube is replaced with a path of length three - see heavy lines in Figure 7). Hence, M is a dual graph of the Archimedean solid $(3,3,3,3,4)$.

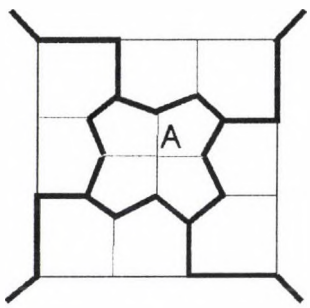


Fig. 7

Case 3. Suppose $w = 17$. Evidently, it is sufficient to find an infinite family of graphs belonging to $\mathfrak{M}(17;5)$. Such family can be obtained from arbitrary number of copies of the belt B and two copies of the cap C in Figure 8 by identifying their boundaries (depicted by heavy lines) as it is illustrated in the third configuration of Figure 8. □

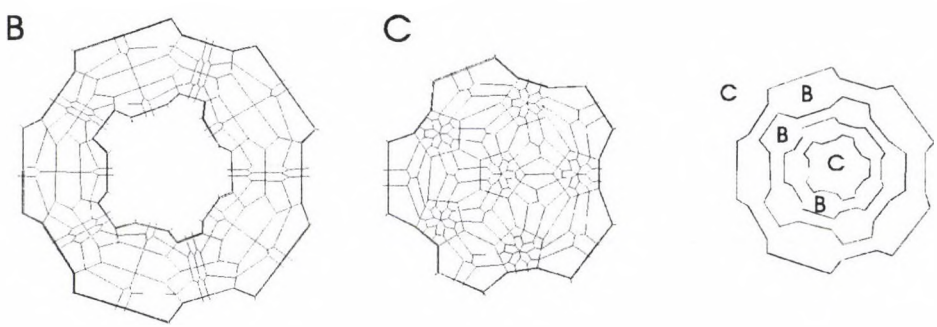


Fig. 8

REMARK. The description of all graphs belonging to $\mathfrak{M}(17;5)$ appears to be difficult. The problem of Eberdhard type (i.e. to determine all sequences $(v_3, v_4, \dots)$, such that there exists a graph $M \in \mathfrak{M}(17;5)$ having $v_k(M) = v_k$, for all $k \geqq 3$) seems to be more passable. Lemma 3 and Euler's formula imply $v_3 = 80 + 2v_4$, $v_5 = 12$ and $v_k = 0$ for $k \geqq 6$. Therefore, it is sufficient to determine v_4. Graphs described in the proof of Theorem 3 contain $120 + 75k$ $(k \geqq 1)$ vertices of degree 4. The authors have also found graphs with $365 + 75k$ $(k \geqq 0)$, 0, 60, 120, 300 and 450 vertices of degree 4 and proved that there exists no graph belonging to $\mathfrak{M}(17;5)$ with $0 < v_4 < 60$, but the problem is still open.

3. General cases

In what follows, we deal with $\mathfrak{M}(w)$. First, let us introduce some simple properties of $M \in \mathfrak{M}(w)$ which will be useful in the next:

(i) If $A \in V(M)$ is incident with a face of type (d_1, d_2, d_3), where $3 \leqq d_1 \leqq d_2 \leqq 5$, $d_3 = \deg(A)$, then all faces of $F_{\{A\}}$ are of the same type. (Both faces having a common edge with end vertices of degree d_3 and d_i ($i = 1$ or 2) are triangles of same type.)

(ii) If $A \in V(M)$ is incident with a face of type (d_1, d_2, d_3), where $3 \leqq d_1 < d_2 \leqq 5$, $d_3 = \deg(A)$, then $\deg(A) \equiv 0 \pmod 2$. (Degrees of neighbours of A are alternately d_1 and d_2.)

(iii) If M contains a face of type $(3, 3, d)$, then $d = 3$. (Only one face of M is not incident with the vertex of degree d and its weight is $3d$.)

(iv) If M contains a face of type $(3, 4, d)$, then either $d = 4$ or $d \geqq 8$. (By (ii), d is even. If $d > 4$, then the vertex of degree 3 is incident with k-gonal face α, where $k \geqq 4$. Hence, $3 + 4 + d = w(\alpha) \geqq 4 + 3 + 4 + 3$.)

(v) If M contains a face of type $(3, 5, d)$, then $d \geqq 8$. (By (ii), d is even. The vertex of degree 3 is incident with k-gonal face α, where $k \geqq 4$. Hence, $3 + 5 + d = w(\alpha) \geqq 5 + 3 + 5 + 3$.)

(vi) If M contains two faces of type $(4, 4, d)$ having two common vertices of degree 4, then M is a graph of the bipyramid with $2d$ faces. ($F(M) = F_{\{A\}} \cup F_{\{B\}}$, where $\deg(A) = \deg(B) = d$.)

(vii) If M contains two faces of type $(3, 3, 3, d)$, $d \geqq 7$, having two common vertices of degree 3, then M is a graph of the dual of antiprism with $2d$ faces. (Since A, B, where $\deg(A) = \deg(B) = d$, cannot lie on a common face, $F(M) = F_{\{A\}} \cup F_{\{B\}}$.)

(viii) If M contains two neighbouring faces of type $(3, d_1, d_2)$, where $3 \leqq d_1 < d_2$, then they have a common vertex of degree d_2. (Otherwise the third face incident with the vertex of degree 3 has weight at least $3 + d_2 + d_2$.)

(ix) If M contains two neighbouring face of types $(3, 4, d)$ and $(3, 4, 4, d-4)$, then $d = 8$. (If $\deg(A) = d$ and $\deg(B) = d - 4$, then all faces of $F_{\{A\}}$ ($F_{\{B\}}$) are of type $(3, 4, d)$ ($(3, 4, 4, d - 4)$, respectively). Moreover, $F(M) = F_{\{A\}} \cup F_{\{B\}}$.)

For $M \in \mathfrak{M}(w)$, let $V_*(M)$ denote the set $V_8(M) \cup V_9(M) \cup V_{\frac{w}{2}}(M)$ ($V_{\frac{w}{2}}(M) = \emptyset$, if w is odd). Now we prove the following

LEMMA 4 ([1]). *If $\mathfrak{M}(w)$ contains a graph M satisfying $|V_*(M)| \geqq 2$, then $|\mathfrak{M}(w)| = \infty$.*

PROOF. Suppose M_1, M_2 belong to $\mathfrak{M}(w)$ and $A_1 \in V_*(M_1)$, $A_2 \in V_*(M_2)$, where $\deg(A_1) = \deg(A_2) = d$. Let $B_{k,1}, \ldots, B_{k,d}$ ($k \in \{1, 2\}$) be neighbouring

vertices of A_k, where $B_{k,i}, A_k, B_{k,i+1}$ (where $B_{k,d+1} = B_{k,1}$) lie on a common face. The polyhedral graph $M_1 * M_2$ is defined as follows

(a) If $d = \frac{w}{2}$, then the graph $M_1 * M_2$ contains d independent edges $e_1, \ldots, e_d$ such that the graph $M_1 * M_2 - \{e_1, \ldots, e_d\}$ has two components, one is isomorphic to $M_1 - A_1$, the other to $M_2 - A_2$. Each edge e_i joins the two vertices corresponding to $B_{k,i}$ in the two components.

(b) If $d = 8$, then $M_1 * M_2$ contains a circuit with vertices $C_1, \ldots, C_8$ such that the graph $M_1 * M_2 - \{C_1, \ldots, C_8\}$ has two components, one is isomorphic to $M_1 - A_1$, the other to $M_2 - A_2$. Each vertex C_i is adjacent to the two vertices corresponding to $B_{k,i}$ in the two components.

(c) If $d = 9$, then $M_1 * M_2$ contains a circuit with vertices $C_1, \ldots, C_{18}$ such that the graph $M_1 * M_2 - \{C_1, \ldots, C_{18}\}$ has two components, one is isomorphic to $M_1 - A_1$, the other to $M_2 - A_2$. Each vertex C_{2i} (C_{2i-1}) is adjacent to the vertex corresponding to $B_{1,i}$ ($B_{2,i}$) in the first (the second, respectively) component.

Evidently, $M_1 * M_2 \in \mathfrak{M}(w)$. Moreover, if M_2 contains a vertex belonging to $V_*(M_2) - \{A_2\}$, then $M_1 * M_2$ contains a corresponding vertex of same degree. Therefore, $\mathfrak{M}(w)$ contains an infinite family of graphs which are constructed recursively as follows:

$$M_1 = M \text{ and } M_{k+1} = M_k * M. \qquad \square$$

Using this result, we are able to prove the following

THEOREM 4. *Let $w \leqq 21$ be a positive integer. Then*

$$|\mathfrak{M}(w)| = \begin{cases} 0 & \text{for } w \leqq 8 \text{ and } w = 10, \\ 1 & \text{for } w = 9 \text{ and } w = 11, \\ 2 & \text{for } w = 12 \text{ and } w = 13, \\ 4 & \text{for } w = 14, \\ 10 & \text{for } w = 15, \\ \infty & \text{for } 16 \leqq w \leqq 21. \end{cases}$$

PROOF. For $w \leqq 11$, $\mathfrak{M}(w) = \mathfrak{M}(w; 3)$, because every k-gonal face, where $k \geqq 4$, has a weight at least 12.

Faces of $M \in \mathfrak{M}(12)$ ($M \in \mathfrak{M}(13)$) are of type either $(4,4,4)$ or $(3,3,3,3)$ (either $(4,4,5)$ or $(3,3,3,4)$), because (ii), (iii) and (iv) eliminate $(3,4,5)$, $(3,5,5)$, $(3,3,6)$, $(3,3,7)$ and $(3,4,6)$. Since faces of distinct types cannot be neighbouring, $\mathfrak{M}(12) = \mathfrak{M}(12;3) \cup \mathfrak{M}(12;4)$ ($\mathfrak{M}(13) = \mathfrak{M}(13;3) \cup \mathfrak{M}(13;4)$, respectively).

Suppose $M \in \mathfrak{M}(14) - (\mathfrak{M}(14;3) \cup \mathfrak{M}(14;4))$. Then faces of M are of type $(4,4,6)$, $(3,3,3,5)$ or $(3,3,4,4)$, because (ii), (iii) and (v) eliminate $(3,4,7)$,

$(4,5,5)$, $(3,3,8)$ and $(3,5,6)$. By (i), all faces of $F_{\{A\}}$, where $\deg(A)=6$, are of type $(4,4,6)$ and by (vi), their other neighbouring faces are of type $(3,3,4,4)$. Thus, all vertices of degree 3 lie on a common 6-gonal face α. $w(\alpha)=18\neq 14$, a contradiction. Therefore, $\mathfrak{M}(14)=\mathfrak{M}(14;3)\cup\mathfrak{M}(14;4)$.

Now, let us assume that A is a maximum degree vertex of $M\in\mathfrak{M}(15)-\bigcup_{k=3}^{5}\mathfrak{M}(15;k)$. (iii) implies that $\Delta(M)\leqq 8$.

If $\Delta(M)=8$, then all faces of $F_{\{A\}}$ are of type $(3,4,8)$ and their neighbouring faces are of type $(3,4,4,4)$. Thus, M is the first of graphs in Figure 9.

If $\Delta(M)=7$, then by (i) and (ii), all faces of $F_{\{A\}}$ are of type $(4,4,7)$ and by (vi), their neighbouring faces are of type $(3,4,4,4)$. This implies that M contains a circuit of order 7 whose vertices have degrees alternately 3 and 4, a contradiction.

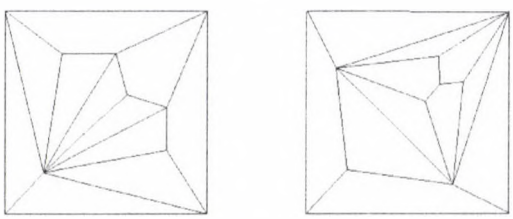


Fig. 9

If $\Delta(M)=6$, then only the following types of faces can occur: $(3,6,6)$, $(4,5,6)$, $(5,5,5)$, $(3,3,3,6)$, $(3,3,4,5)$, $(3,4,4,4)$, $(3,3,3,3,3)$. First, suppose that A is incident with a face of type $(4,5,6)$. Then by (i), all faces of $F_{\{A\}}$ are of the same type. Denote by B_i, $i=1,2,3$, the neighbours of A with degree 5. As $F_{\{B_i\}}$ contains two neighbouring faces of type $(4,5,6)$, the rest faces of $F_{\{B_i\}}$ are of type $(3,3,4,5)$. Moreover, the vertex C_i of degree 4 lying on the 4-gonal face with both neighbouring 4-gonal faces is not adjacent to B_i. Thus, the neighbouring faces of these quadrangles are of type either $(3,3,3,4)$ (if $B_i=B_{i+1}$) or $(3,3,3,4,4)$ (if $B_i\neq B_{i+1}$), in contradiction to $M\in\mathfrak{M}(15)$. Now, suppose that all faces of $F_{\{A\}}$ are of type $(3,6,6)$. Then all faces of M are incident with two vertices of degree 6. Therefore, they are of type $(3,6,6)$. So, $M\in\mathfrak{M}(15;3)$, a contradiction. Similarly, A is not incident with two neighbouring faces of type $(3,3,3,6)$, because either $M\in\mathfrak{M}(15;4)$ (if the neighbouring face of both considered faces is of type $(3,3,3,6)$), or $M\notin\mathfrak{M}(15)$ (if it is of type $(3,3,3,3,3)$), otherwise. Thus, A is incident with faces of types $(3,6,6)$, $(3,6,6)$, $(3,3,3,6)$, $(3,6,6)$, $(3,6,6)$, $(3,3,3,6)$. This implies that M is the second in Figure 9.

If $\Delta(M)\leqq 5$, then M contains no triangle, because any k-gonal face of M, where $k\geq 4$, is not incident with two vertices of degree 5. So, M contains neighbouring faces of types $(3,3,4,5)$ and $(3,3,3,3,3)$. This enforces a circuit of order 5 whose vertices have degree alternately 4 and 5, a contradiction.

The graph of the bipyramid with 16 (18) faces belongs to $\mathfrak{M}(16)$ ($\mathfrak{M}(17)$) and contains two vertices of degree 8 (9, respectively). The dual graph of Archimedean solid $(4,6,8)$ $((3,8,8), (4,6,10))$ belongs to $\mathfrak{M}(18)$ ($\mathfrak{M}(19)$, $\mathfrak{M}(20)$) and contains 6 (6, 12) vertices of degree 8 $(8, 10 = \frac{20}{2}$, respectively). Similarly, the graph illustrated in Figure 10 belongs to $\mathfrak{M}(21)$ and contains 14 vertices of degree 9. Therefore, for $16 \leqq w \leqq 21$, the assertion follows from Lemma 4. □

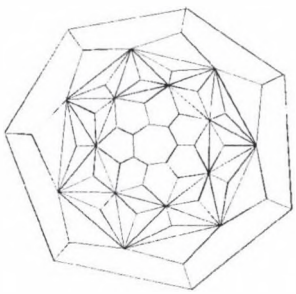


Fig. 10

REMARK. Moreover, the authors have proved that $\mathfrak{M}(16)$ contains only three graphs, which are illustrated in Figure 11, besides graphs described in the proofs of above theorems. The description of all graphs belonging to the rest infinite families $\mathfrak{M}(w)$ appears to be very difficult. The problem of Eberdhard type seems to be difficult, too. From (3), (ii) and (iii) we get some necessary conditions, but they are not sufficient.

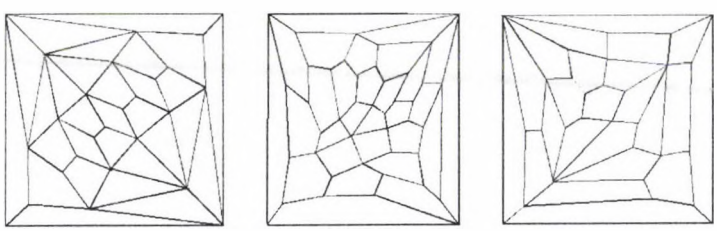


Fig. 11

For the edge e of a 3-polytope (polyhedral graph) with end vertices A, B, the *type* of e is defined as the couple (d_1, d_2), where $d_1 = \deg(A)$ and $d_2 = \deg(B)$. Let edges incident with a k-gonal face α be successively $e_1, \ldots, e_k$ and let the type of e_i be (d_i, d_{i+1}) for all $i = 1, \ldots, k$ (where $d_{k+1} = d_1$). Then

$$(4)\quad c(\alpha) = k - 6 + \sum_{i=1}^{k} \frac{2d_i - 6}{d_i} = \sum_{i=1}^{k} 3\left(1 - \left(\frac{2}{k} + \frac{1}{d_i} + \frac{1}{d_{i+1}}\right)\right) = \sum_{i=1}^{k} c(\alpha; e_i),$$

where $c(\alpha; e_i) := 3\left(1 - \left(\frac{2}{k} + \frac{1}{d_i} + \frac{1}{d_{i+1}}\right)\right)$ is the *contribution* of e_i to $c(\alpha)$. Note, if $k = 4$ and α contains no edge of type $(3, d)$, where $3 \leqq d \leqq 5$, or $k = 5$ and α contains no edge of type $(3, 3)$ or $k \geqq 6$, then $c(\alpha; e_i) \geqq 0$ for all $i = 1, \ldots, k$.

We conjecture $\mathfrak{M}(w) = \mathfrak{M}(w; 3) \cup \mathfrak{M}(w; 4)$, for all $w \geqq 22$. However, we are able to prove the following

THEOREM 5. *If* $w \geqq 28$, *then* $\mathfrak{M}(w) = \mathfrak{M}(w; 3) \cup \mathfrak{M}(w; 4)$.

PROOF. Assume indirectly that there exists a polyhedral graph $M \in \mathfrak{M}(w) - (\mathfrak{M}(w; 3) \cup \mathfrak{M}(w; 4))$. Its negative charge faces can be only of the following types: $(3, 4, w-7)$, $(3, 5, w-8)$, $(3, 6, w-9)$, $(3, 7, w-10)$ for $w \leqq 51$, $(3, 8, w-11)$ for $w \leqq 34$, $(3, 9, w-12)$ for $w \leqq 29$, $(4, 4, w-8)$ and $(3, 3, 3, w-9)$.

In the proof, an edge of M is called *weak* if it is incident with a negative charge face and the degrees of its end vertices are at most 9. Thus, every negative charge n-gonal face of M is incident with precisely $n-2$ weak edges which can be only of the following types: $(4, 4)$ and $(3, d)$, for $3 \leqq d \leqq 9$. Moreover, by (vi), (vii) and (viii), every weak edge of M is incident with precisely one negative charge face.

Now, we construct a function $b\colon F(M) \to \mathbb{R}$ such that $\sum_{\alpha \in F} b(\alpha) = \sum_{\alpha \in F} c(\alpha)$, according to the following rule. First, we put $b(\gamma) = c(\gamma)$ for every $\gamma \in F(M)$. Then for every weak edge e, which is incident with faces α and β, the amount $a(e) := \frac{|c(\alpha)|}{n-2}$ is subtracted from $b(\beta)$ and added to $b(\alpha)$ if α is a negative charge n-gonal face.

Let us verify that $b(\gamma) \geqq 0$ for every $\gamma \in F(M)$. Evidently, if γ is a negative charge face, then $b(\gamma) = 0$ and if γ is incident with no weak edge, then $b(\gamma) = c(\gamma) \geqq 0$. Therefore, let γ be a k-gonal face incident with a weak edge and $c(\gamma) \geqq 0$. Since every triangle incident with a weak edge is a negative charge face, $k \geqq 4$.

Let e be a weak edge incident with γ. If e is of type $(4, 4)$, then $c(\gamma; e) = \frac{3}{2} - \frac{6}{k}$ and $a(e) = \frac{6}{w-8}$. Thus, $c(\gamma; e) = \frac{3}{2} - \frac{6}{k} \geqq \frac{3}{2} - \frac{6}{5} = \frac{6}{28-8} \geqq \frac{6}{w-8} = a(e)$ for $k \geqq 5$. Similarly, if e is of type $(3, d)$ for $6 \leqq d \leqq 9$, then $c(\gamma; e) = 2 - \frac{3}{d} - \frac{6}{k}$ and $a(e) = \frac{6}{w-d-3} + \frac{6}{d} - 1$. So, $c(\gamma; e) \geqq a(e)$ for $k \geqq 6$ and $d = 6$ or $k \geqq 5$ and $d = 7$ or $k \geqq 4$ and $d \geqq 8$. If e is of type $(3, d)$ for $3 \leqq d \leqq 5$, then it can easily be seen that both of its neighbouring edges e_1, e_2 incident with γ cannot be weak. If e_1 is not weak, then we can counterbalance $a(e)$ by $c(\gamma; e) + c(\gamma; e_1)$ or only by $c(\gamma; e) + \frac{1}{2}c(\gamma; e_1)$ in case e_1 is adjacent to two weak edges of γ being of type $(3, d)$, $3 \leqq d \leqq 5$. As e_1 is of type (d, d_1),

degrees of vertices of γ	types of weak edges e_m	$c(\gamma)$	$\sum a(e_m)$
$(d,3,d,r)\ d\in\{5,6,7\}$	$2(3,d)$	$4-\frac{12}{d}-\frac{6}{r}$	$\frac{12}{d}+\frac{12}{w-3-d}-2$
$(7,3,n,r)\ n\in\{3,4\}$	$(3,7)$	$\frac{22}{7}-\frac{6}{n}-\frac{6}{r}$	$\frac{6}{w-10}-\frac{1}{7}$
$(8,3,n,r)\ n\in\{3,4,5\}$	$(3,8)$	$\frac{13}{4}-\frac{6}{n}-\frac{6}{r}$	$\frac{6}{w-11}-\frac{1}{4}$
$(9,3,n,r)\ n\in\{3,4,5\}$	$(3,9)$	$\frac{10}{3}-\frac{6}{n}-\frac{6}{r}$	$\frac{6}{w-12}-\frac{1}{3}$
$(4,4,i,r)$	$(4,4)$	$3-\frac{6}{i}-\frac{6}{r}$	$\frac{6}{w-8}$
$(6,3,3,r)$	$(3,6),(3,3)$	$1-\frac{6}{r}$	$\frac{9}{w-9}$
$(d,3,d,i,r)\ d\in\{4,5,6\}$	$2(3,d)$	$7-\frac{12}{d}-\frac{6}{i}-\frac{6}{r}$	$\frac{12}{d}+\frac{12}{w-3-d}-2$
$(6,3,3,i,r)$	$(3,6),(3,3)$	$4-\frac{6}{i}-\frac{6}{r}$	$\frac{9}{w-9}$
$(3,3,3,i,r)$	$2(3,3)$	$3-\frac{6}{i}-\frac{6}{r}$	$\frac{6}{w-9}$
$(4,4,3,3,r)$	$(4,4)$	$2-\frac{6}{r}$	$\frac{6}{w-8}$
$(d,3,3,i,r)\ d\in\{7,8,9\}$	$(3,d)$	$5-\frac{6}{d}-\frac{6}{i}-\frac{6}{r}$	$\frac{6}{d}+\frac{6}{w-3-d}-1$
$(4,3,4,i,j,r)$	$2(3,4)$	$7-\frac{6}{i}-\frac{6}{j}-\frac{6}{r}$	$\frac{12}{w-7}+1$
$(4,3,4,4,4,r)$	$2(3,4),(4,4)$	$4-\frac{6}{r}$	$\frac{6}{w-8}+\frac{12}{w-7}+1$
$(4,3,4,9,3,6)$	$2(3,4),(3,9)$	$\frac{10}{3}$	$\frac{2}{3}+\frac{12}{w-7}+\frac{6}{w-12}$
$(6,3,3,i,j,r)$	$(3,6),(3,3)$	$7-\frac{6}{i}-\frac{6}{j}-\frac{6}{r}$	$\frac{9}{w-9}$
$(6,3,3,4,4,r)$	$(3,6),(3,3),(4,4)$	$4-\frac{6}{r}$	$\frac{9}{w-9}+\frac{6}{w-8}$
$(6,3,3,d,3,d)\ d\in\{7,8,9\}$	$(3,6),(3,3),2(3,d)$	$5-\frac{12}{d}$	$\frac{9}{w-9}+\frac{12}{d}+\frac{12}{w-3-d}-2$
$(6,3,3,8,3,5)$	$(3,6),(3,3),(3,8)$	$\frac{61}{20}$	$\frac{9}{w-9}+\frac{6}{w-11}-\frac{1}{4}$
$(6,3,3,9,3,n)\ n\in\{4,5,6\}$	$(3,6),(3,3),(3,9)$	$\frac{13}{3}-\frac{6}{n}$	$\frac{9}{w-9}+\frac{6}{w-12}-\frac{1}{3}$
$(3,3,3,i,j,r)$	$2(3,3)$	$6-\frac{6}{i}-\frac{6}{j}-\frac{6}{r}$	$\frac{6}{w-9}$
$(3,3,3,4,4,r)$	$2(3,3),(4,4)$	$3-\frac{6}{r}$	$\frac{6}{w-9}+\frac{6}{w-8}$
$(3,3,3,d,3,d)\ d\in\{8,9\}$	$2(3,3),2(3,d)$	$4-\frac{12}{d}$	$\frac{6}{w-9}+\frac{12}{d}+\frac{12}{w-3-d}-2$
$(4,3,4,3,3,3,r)$	$2(3,4),2(3,3)$	$4-\frac{6}{r}$	$\frac{6}{w-9}+\frac{12}{w-7}+1$
$(4,3,4,3,3,6,r)$	$2(3,4),(3,3),(3,6)$	$5-\frac{6}{r}$	$\frac{9}{w-9}+\frac{12}{w-7}+1$
$(4,3,4,4,3,4,r)$	$4(3,4)$	$5-\frac{6}{r}$	$\frac{24}{w-7}+2$
$(4,3,4,4,3,4,i,r)$	$4(3,4)$	$8-\frac{6}{i}-\frac{6}{r}$	$\frac{24}{w-7}+2$

Table

$c(\gamma;e_1)=3-\frac{6}{k}-\frac{3}{d}-\frac{3}{d_1}\geqq 2-\frac{6}{k}-\frac{3}{d}=c(\gamma;e)$, $a(e)=\frac{6}{d}+\frac{6}{w-d-3}-1$ for $d>3$ and $a(e)=\frac{3}{w-9}$ for $d=3$. Therefore, $c(\gamma;e)+c(\gamma;e_1)\geqq a(e)$ for $k\geqq 7$ and $d=3,4$ or $k\geqq 6$ and $d=5$, and $c(\gamma;e)+\frac{1}{2}c(\gamma;e_1)\geqq a(e)$ for $k\geqq 6$ and $d=5$ or $k\geqq 9$ and $d=4$ or $k\geqq 7$ and $d=3$.

By (4), if the contribution of each edge of γ to $c(\gamma)$ is nonnegative and $a(e)\leqq c(\gamma;e)$ ($a(e)\leqq c(\gamma;e)+c(\gamma;e_1)$ or $a(e)\leqq c(\gamma;e)+\frac{1}{2}c(\gamma;e_1)$, respectively) for each weak edge e of γ, then $b(\gamma)\geqq 0$. For the rest of the possible cases

(i.e. γ contains an edge with negative contribution to $c(\gamma)$ or a weak edge e with $a(e) > c(\gamma; e)$ and $a(e) > c(\gamma; e) + c(\gamma; e_1)$ (or $a(e) > c(\gamma; e) + \frac{1}{2}c(\gamma; e_1)$), where e_1 is not weak edge adjacent to e) see Table, $c(\gamma)$ and the sum of $a(e_m)$ is determined for γ for all weak edges e_m incident with γ. Considering $w(\gamma) = w \geqq 28$, it can easily be verified that $c(\gamma) \geqq \sum a(e_m)$, which implies $b(\gamma) \geqq 0$.

This contradicts (1): $0 \leqq \sum_{\gamma \in F} b(\gamma) = \sum_{\gamma \in F} c(\gamma) = -12.$ □

REFERENCES

[1] Bauer, R., Weights of cells in complexes, Diploma-work, Košice, 1981 (in Slovakian).

[2] Grünbaum, B., *Convex polytopes*, Interscience Publishers John Wiley & Sons, Inc., New York, 1967. *MR* **37** #2085

[3] Jendrol', S. and Jucovič, E., On face-vectors of maps with constant weight of edges, *Studia Sci. Math. Hungar.* **17** (1982), 159–175. *MR* **85i**:05090

[4] Lebesgue, H., Quelques conséquences simples de la formule d'Euler, *J. Math. Pures Appl.* **19** (1940), 27–43. *MR* **1**, 316i

[5] Rosenfeld, M., Polytopes of constant weight, *Israel J. Math.* **21** (1975), 24–30. *MR* **52** #4132

(Received May 24, 1995)

PRÍRODOVEDECKÁ FAKULTA
UNIVERZITY P. J. ŠAFÁRIKA
JESENNÁ 5
SK–041 54 KOŠICE
SLOVAKIA

ivanco@duro.upjs.sk
trenkler@duro.upjs.sk

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 17–38

REGULAR COLOURED RANK 3 POLYHEDRA WITH TETRAGONAL VERTEX FIGURE

C. LEYTEM

1. Introduction

In 1977 B. Grünbaum [3] introduced a more general concept of regular polyhedron which allowed skew polygons as faces. Later A.Dress [1] gave a complete classification of the Grünbaum polyhedra using the concept of a Grünbaum system.

We are going to generalize the concept of a Grünbaum polyhedron, introducing polyhedra with two types of faces. In the classical case of the Whythoff construction, regular polyhedra are obtained by taking the fundamental vertex at the intersection of two non-orthogonal mirrors in the orthoscheme, whereas semi-regular polyhedra such as the cuboctahedron, icosidodecahedron can be obtained by taking the fundamental vertex on the intersection of the two orthogonal mirrors. We are going to proceed by analogy and take the fundamental vertex at the intersection of the fixed spaces of the two commuting involutions.

In Section 2 we recall the definitions of regular polyhedra and Grünbaum systems. In Section 3 we introduce the generalized Grünbaum systems and explain how they define coloured Grünbaum polyhedra. In Section 4 we present the classification of the rank 3 examples.

2. Regular Grünbaum polyhedra

2.1. Regular polygons. As the regular polygons are the building blocks of the Grünbaum polyhedra, we shortly recall the definition and notation.

A *polygon* $P = \{\dots, v_1, v_2, v_3, \dots\}$ in the Euclidean space $\mathbf{E} := \mathbf{E}^3$ is the figure formed by the distinct points *(vertices)* $\dots, v_1, v_2, v_3, \dots$ together with the segments *(edges)* $e_i = [v_i, v_{i+1}]$.

In case P is *finite* there is an additional edge $[v_{first}, v_{last}]$.

1991 *Mathematics Subject Classification.* Primary 52B15; Secondary 51M20.

Key words and phrases. Regular polyhedron, Grünbaum–Dress polyhedron, semi-regular polyhedron, flag-transitive symmetry group.

In case P is *infinite* there is an additional condition: Each compact subset of $\mathbf{E}$ meets only finitely many edges.

A *flag* of P is a pair consisting of a vertex v of P incident with an edge e of P.

A polygon is said to be *regular* if its group of *symmetries* acts transitively on the family of all flags of P. This group of symmetries can be generated by two fundamental isometries α_0 and α_1, α_0 fixing e_0 and α_1 fixing v_0.

In terms of the fundamental vertex v_0 and the fundamental isometries α_0 and α_1 the polygon is then given by

$$\{\dots,(\alpha_0\alpha_1)^i v_0,\dots,(\alpha_0\alpha_1)^2 v_0,\alpha_0\alpha_1 v_0,v_0,\alpha_1\alpha_0 v_0,(\alpha_1\alpha_0)^2 v_0,\dots,(\alpha_1\alpha_0)^i v_0,\dots\}.$$

A complete list is given in [1, 3]. It includes the plane, prismatic, antiprismatic and helical polygons.

2.2. Regular polyhedra. A *polyhedron* P in $\mathbf{E}$ is a family of polygons *(faces)* with the following properties [3]:

(i) each edge of a face is an edge of exactly one other face *(thinness)*;
(ii) the family of polygons is connected through edges;
(iii) each compact set meets only finitely many faces.

A *flag* of P is a triplet consisting of a vertex v, an edge e and a face f of P, all mutually incident.

A polyhedron is said to be *regular* if its group of *symmetries* acts transitively on the family of all flags of P. This group of symmetries can be generated by three fundamental isometries α_0 , α_1 and α_2, α_0 fixes e_0 and f_0, α_1 fixes v_0 and f_0 and α_2 fixes v_0 and e_0.

Illustrations, respectively a complete list can be found in [1, 3]. It includes the Platonic polyhedra, the plane tessellations and the Petrie–Coxeter polyhedra.

2.3. Discrete Grünbaum systems. To classify these regular polyhedra, Dress [1] introduced discrete Grünbaum systems:

A *discrete Grünbaum system* is a system $(v;\alpha_0,\alpha_1,\alpha_2)\in\mathbf{E}\times\mathrm{Iso}(\mathbf{E})^3$ ($\alpha_0,\alpha_1,\alpha_2$ isometries of $\mathbf{E}$) satisfying the conditions:

- $\alpha_0^2=\alpha_1^2=\alpha_2^2=(\alpha_0\alpha_2)^2=\mathrm{Id}_{\mathbf{E}}$;
- $\alpha_1 v=\alpha_2 v=v$;
- $\alpha_i|_{\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v}\neq\mathrm{Id}_{\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v}$, $i\in\{0,1,2\}$;
- $\langle\alpha_0,\alpha_1,\alpha_2\rangle$ is discrete as a group of isometries of $\mathbf{E}$.

As a discrete Grünbaum system univoquely defines a regular polyhedron with vertex set $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v$, edge set $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot[v,\alpha_0 v]$ and face set $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot\{\dots,v,\alpha_0 v,\alpha_1\alpha_0 v,\dots\}$, a classification of discrete Grünbaum polyhedra is equivalent to a classification of regular polyhedra (cf. [1]).

3. Regular coloured Grünbaum polyhedra

Motivated by the Wythoff construction for uniform polyhedra, we consider discrete Grünbaum systems which satisfy the relation $(\alpha_1\alpha_2)^2 = \mathrm{Id}_{\mathbf{E}}$ instead of $(\alpha_0\alpha_2)^2 = \mathrm{Id}_{\mathbf{E}}$. This modification will imply that the vertex figure is tetragonal.

3.1. Definition. A *discrete coloured Grünbaum system* is a system $(v; \alpha_0, \alpha_1, \alpha_2) \in \mathbf{E} \times \mathrm{Iso}(\mathbf{E})^3$ satisfying the conditions:

- $\alpha_0^2 = \alpha_1^2 = \alpha_2^2 = (\alpha_1\alpha_2)^2 = \mathrm{Id}_{\mathbf{E}}$;
- $\alpha_1 v = \alpha_2 v = v$;
- v, $\alpha_0 v$, $\alpha_1\alpha_0 v$, $\alpha_2\alpha_0 v$, $\alpha_1\alpha_2\alpha_0 v$ are all distinct;
- none of the segments $[v, \alpha_0 v]$, $[v, \alpha_1\alpha_0 v]$, $[v, \alpha_2\alpha_0 v]$, $[v, \alpha_1\alpha_2\alpha_0 v]$ is contained in one of the three other segments;
- $\langle \alpha_0, \alpha_1, \alpha_2 \rangle$ is discrete as a group of isometries of $\mathbf{E}$.

3.2. Construction. To a discrete coloured Grünbaum system one can associate a polyhedron with tetragonal vertex figure. First we have to describe the vertex set V, the coloured faces (black and white) and the flags associated to the Grünbaum system. This defines a coloured polyhedron. Note right away that the definition of a coloured Grünbaum system is symmetric in α_1 and α_2. It will be obvious that exchanging these two isometries will result in a switch in the colour of the faces.

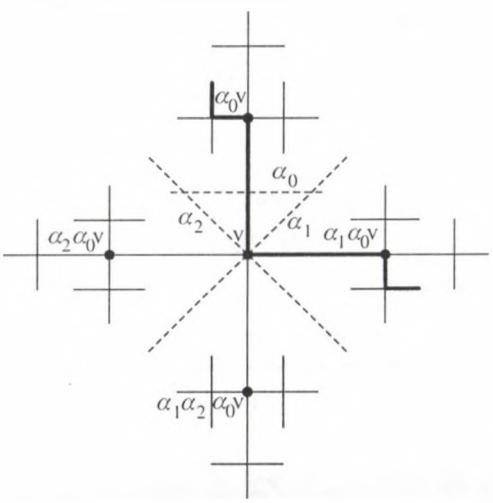


Fig. 1. The vertex figure of a coloured polyhedron

VERTEX SET. It is defined as $V = \langle \alpha_0, \alpha_1, \alpha_2 \rangle v$.

VERTEX FIGURE. The *fundamental vertex figure* (at v) is formed by the quadrangle $\{\alpha_2\alpha_0 v, \alpha_0 v, \alpha_1\alpha_0 v, \alpha_1\alpha_2\alpha_0 v\}$ (cf. Figure 1).

The *vertex figure (at an arbitrary vertex αv)* is obtained by conjugation.

First transform αv back to v by applying α^{-1}. Then apply successively α_0, $\alpha_1\alpha_0$, $\alpha_2\alpha_0$, $\alpha_1\alpha_2\alpha_0$ to v and transform back to αv. The vertex figure at αv is the quadrangle $\{\alpha\alpha_2\alpha_0 v, \alpha\alpha_0 v, \alpha\alpha_1\alpha_0 v, \alpha\alpha_1\alpha_2\alpha_0 v\}$.

FACES. As *first fundamental face (black)* we take the face defined as

$$b = \{\dots, \alpha_0 v, v, \alpha_1\alpha_0 v, \dots\}.$$

By 2.1 this is the face (represented as a bold black polygon in Figure 1)

$$\{\dots, (\alpha_0\alpha_1)^i v, \dots, (\alpha_0\alpha_1)^2 v, \alpha_0\alpha_1 v, v, \alpha_1\alpha_0 v, (\alpha_1\alpha_0)^2 v, \dots, (\alpha_1\alpha_0)^i v, \dots\}.$$

It is invariant under the transformation α_1:

$$\alpha_1 b = b,$$

and the transformation α_2 transforms it into

$$b' = \{\dots, \alpha_2\alpha_0 v, v, \alpha_1\alpha_2\alpha_0 v, \dots\}$$

which is also invariant under α_1:

$$\alpha_2 b = b',$$

$$\alpha_1 b' = b'.$$

As *second fundamental face (white)* we take the face defined as

$$w = \{\dots, \alpha_0 v, v, \alpha_2\alpha_0 v \dots\}.$$

By 2.1 this is the face

$$\{\dots, (\alpha_0\alpha_2)^i v, \dots$$
$$\dots, (\alpha_0\alpha_2)^2 v, \alpha_0\alpha_2 v, \alpha_0 v, v, \alpha_2\alpha_0 v, (\alpha_2\alpha_0)^2 v, \dots, (\alpha_2\alpha_0)^i v, \dots\}.$$

It is invariant under the transformation α_2:

$$\alpha_2 w = w,$$

and the transformation α_1 transforms it into

$$w' = \{\dots, \alpha_1\alpha_0 v, v, \alpha_1\alpha_2\alpha_0 v, \dots\}$$

which is also invariant under α_2:

$$\alpha_1 w = w',$$

$$\alpha_2 w' = w'.$$

At an arbitrary vertex αv, $\{\dots, \alpha\alpha_0 v, \alpha v, \alpha\alpha_1\alpha_0 v, \dots\}$ and $\{\dots, \alpha\alpha_2\alpha_0 v, \alpha v, \alpha\alpha_1\alpha_2\alpha_0 v, \dots\}$ are black faces invariant under $\alpha\alpha_1\alpha^{-1}$.

At an arbitrary vertex αv, $\{\dots, \alpha\alpha_0 v, \alpha v, \alpha\alpha_2\alpha_0 v, \dots\}$ and $\{\dots, \alpha\alpha_1\alpha_0 v, \alpha v, \alpha\alpha_1\alpha_2\alpha_0 v, \dots\}$ are white faces invariant under $\alpha\alpha_2\alpha^{-1}$.

FLAGS. Having thus defined a coloured polyhedron resulting from a discrete coloured Grünbaum system, we can try and describe the polyhedron as a regular coloured polyhedron by introducing coloured flags as follows:

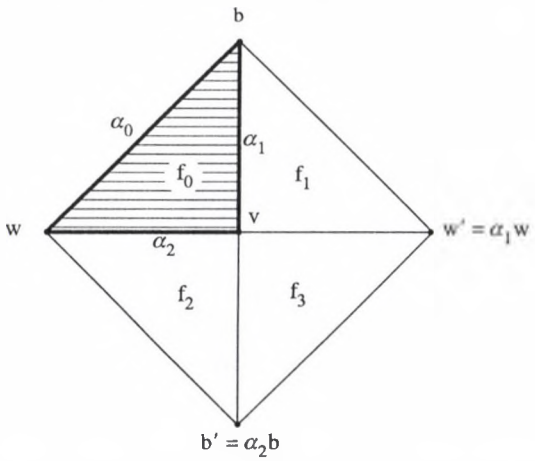


Fig. 2. A symbolic picture of the fundamental flag

As a *fundamental flag* we take

$$f_0 = [v, b, w].$$

Then, as in the case of a classical flag, α_0 fixes b and w, α_1 fixes b and v, α_2 fixes w and v.

The 4 flags surrounding v are f_0, $f_1 = \alpha_1 f_0 = [v, b, \alpha_1 w]$, $f_2 = \alpha_2 f_0 = [v, \alpha_2 b, w]$, $f_3 = \alpha_1\alpha_2 f_0 = [v, \alpha_2 b, \alpha_1 w]$. The situation is symbolically represented in Figure 2.

3.3. Definition. A *regular coloured polyhedron* P consists of a discrete set $V \subseteq \mathbf{E}^3$ of vertices denoted by v_i, and two sets of faces, the set B of black faces and the set W of white faces, a face being an ordered set of vertices $\{\dots, v_i, v_{i+1}, \dots\}$.

A flag is defined as a triplet $[v, b, w]$ with $b \cap w = \{v, v'\}$ ($v \neq v'$ and v' is consecutive to v) and vertices, faces and flags satisfy the following conditions:

- each vertex is contained in at least one face of each colour;
- each pair formed by a face and one of its vertices can be completed to exactly two different flags *(thin polyhedron)*;
- any two flags can be connected by a sequence of neighbouring flags, i.e., flags which differ by exactly one element *(flag-connected polyhedron)*;
- the group of colour-preserving automorphisms of the polyhedron acts transitively on the flags of P *(flag-transitive polyhedron)*;

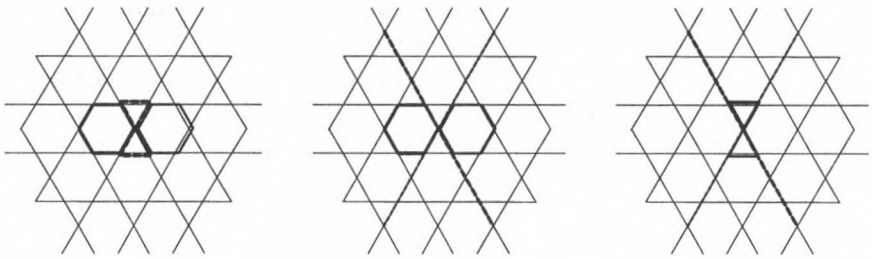


Fig. 3. Three planar coloured polyhedra related through Petrie operators

- each compact set meets only finitely many faces.

Three examples illustrating these concepts are given in Figure 3: the first one is a planar tiling with hexagons and triangles, the three fundamental symmetries α_i $(i = 1, 2, 3)$ are reflections in a line , the second one is related to this tiling and has as faces hexagons and apeirogons, the third one has triangles and apeirogons as faces (in these two examples the first two fundamental symmetries α_i $(i = 1, 2)$ are reflections in a line whereas the third symmetry α_2 is central).

3.4. Coloured Petrie operators. In analogy with the classical Petrie operator we give the following definition:

DEFINITION 3.4.1. The coloured Petrie operators associated to $(v; \alpha_0, \alpha_1, \alpha_2)$ are given by

$$P_1(v; \alpha_0, \alpha_1, \alpha_2) = (v; \alpha_0, \alpha_1\alpha_2, \alpha_2)$$

and

$$P_2(v; \alpha_0, \alpha_1, \alpha_2) = (v; \alpha_0, \alpha_1, \alpha_1\alpha_2).$$

PROPERTY 3.4.2. The coloured Petrie duals P_1 and P_2 transform a coloured Grünbaum system into a coloured Grünbaum system.

PROOF. This follows from the definition of a coloured Grünbaum system as α_1 and α_2 commute and α_i $(0 \leqq i \leqq 2)$ and $\alpha_1\alpha_2$ are involutions. □

We also have the following

PROPERTY 3.4.3.

$$P_i \circ P_i = \mathrm{Id} \quad (i = 1, 2),$$

$$P_i \circ P_j = P_i \quad (i = 1, 2;\ i \neq j)$$

followed by a switch in the colour of the faces in the second case.

Recall that the vertex figure of the fundamental vertex v is $\{\alpha_2\alpha_0 v, \alpha_0 v, \alpha_1\alpha_0 v, \alpha_1\alpha_2\alpha_0 v\}$ and that the fundamental black and white faces are $\{\ldots,$

$\alpha_0 v, v, \alpha_1\alpha_0 v, \dots\}$ and $\{\dots, \alpha_0 v, v, \alpha_2\alpha_0 v, \dots\}$. Applying a coloured Petrie operator changes the order of the vertices in the vertex figure of v.

- Applying P_1, the vertex figure becomes $\{\alpha_2\alpha_0 v, \alpha_0 v, \alpha_1\alpha_2\alpha_0 v, \alpha_1\alpha_0 v\}$ and the fundamental black and white faces will be $\{\dots, \alpha_0 v, v, \alpha_1\alpha_2\alpha_0 v, \dots\}$ and $\{\dots, \alpha_0 v, v, \alpha_2\alpha_0 v, \dots\}$.
- Applying P_2, the vertex figure becomes $\{\alpha_1\alpha_2\alpha_0 v, \alpha_0 v, \alpha_1\alpha_0 v, \alpha_2\alpha_0 v\}$ and the fundamental black and white faces will be $\{\dots, \alpha_0 v, v, \alpha_1\alpha_0 v, \dots\}$ and $\{\dots, \alpha_0 v, v, \alpha_1\alpha_2\alpha_0 v, \dots\}$.

4. Coloured rank 3 polyhedra

4.1. Introduction. As in Dress [1] we say that two coloured Grünbaum systems $(v; \alpha_0, \alpha_1, \alpha_2)$ and $(w; \beta_0, \beta_1, \beta_2)$ are *isometric* (resp. similar), if there exists an isometry (resp. similarity) $\gamma : \mathbf{E} \to \mathbf{E}$ with $\gamma v = w$ and $\gamma\alpha_i\gamma^{-1}|_{\langle\beta_0,\beta_1,\beta_2\rangle w} = \beta_i|_{\langle\beta_0,\beta_1,\beta_2\rangle w}$.

We want to classify the coloured rank 3 polyhedra up to isometry (resp. similarity).

As in Dress [1], we list polyhedra according to the dimensions and angles formed by the fixed point spaces of the involutions and we use his notations and terminology:

- We define the *rank* of a coloured polyhedron as the rank $\mathrm{rk}(G \cap T)$ of the intersection of $G = \langle\alpha_0, \alpha_1, \alpha_2\rangle$ with the translation group T of $\mathbf{E}$.
- We denote by $\dim\alpha$ the dimension of the fixspace $\mathbf{E}^\alpha$.
- We define $\widehat{(\alpha, \beta)}$ to be $0°$ unless $1 \leqq \dim\alpha, \dim\beta \leqq 2$ in which case we define $\widehat{(\alpha, \beta)}$ as the angle between $\mathbf{E}^\alpha$ and $\mathbf{E}^\beta$ or, in case $\mathbf{E}^\alpha \cap \mathbf{E}^\beta = \emptyset$, as the angle between properly intersecting parallel transforms of these two spaces. To avoid ambiguities, we will always assume $0° \leqq \widehat{(\alpha, \beta)} \leqq 90°$.
- Define two coloured Grünbaum systems $(v; \alpha_0, \alpha_1, \alpha_2)$ and $(w; \beta_0, \beta_1, \beta_2)$ to have the same *dimensionality* if $\dim\alpha_i = \dim\beta_i$ for $i = 0, 1, 2$ (implying $\dim\alpha_1\alpha_2 = \dim\beta_1\beta_2$).
- Define two coloured Grünbaum systems $(v; \alpha_0, \alpha_1, \alpha_2)$ and $(w; \beta_0, \beta_1, \beta_2)$ to have the same *angularity* if they have the same dimensionality and moreover $\widehat{(\alpha_0, \alpha_1)} = \widehat{(\beta_0, \beta_1)}$ and $\widehat{(\alpha_0, \alpha_2)} = \widehat{(\beta_0, \beta_2)}$.

A complete classification as proposed in [4] would now follow the lines of Dress [1], treating first the rank 0 case (among which the cuboctahedron and some other finite polyhedra discussed by Farris [2]), the rank 1 case (which exists in this context), the planar case, and finally the rank 2 and 3 cases for which we have

PROPERTY 4.1.1 (cf. Dress [1], Section 1). If $(v; \alpha_0, \alpha_1, \alpha_2)$ is a discrete Grünbaum system with group $G = \langle\alpha_0, \alpha_1, \alpha_2\rangle$ and if $\mathrm{rk}(G \cap T) \geqq 2$ then $\widehat{(\alpha_0, \alpha_1)} \in \{0°, 30°, 45°, 60°, 90°\}$ and $\widehat{(\alpha_0, \alpha_2)} \in \{0°, 30°, 45°, 60°, 90°\}$.

Note that, as α_1 and α_2 commute, one has necessarily $\mathbf{E}^{\alpha_1} \cap \mathbf{E}^{\alpha_2} \neq \emptyset$ and $\widehat{(\alpha_1, \alpha_2)} \in \{0°, 90°\}$.

4.2. The rank 3 classification. CONVENTION 4.2.1. To facilitate the classification we consider the triplet

$$(\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2))$$

in increasing order.

In fact, as $\langle \alpha_1, \alpha_2 \rangle$ is the Klein 4-group, we have three choices of unordered pairs of fundamental generators in it. Geometrically this change of generators corresponds to choosing different pairs of black and white fundamental faces in the vertex figure. A similar change of the fundamental flag can be obtained by applying the *coloured Petrie operators* to the coloured Grünbaum system $(v; \alpha_0, \alpha_1, \alpha_2)$ and will simplify the classification in view of the following property which results from 3.4.2:

PROPERTY 4.2.2. Two coloured Grünbaum systems $(v; \alpha_0, \alpha_1, \alpha_2)$ and $(w; \beta_0, \beta_1, \beta_2)$ are isometric (resp. similar) if and only if $P_i(v; \alpha_0, \alpha_1, \alpha_2)$ $(i = 1, 2)$ and $P_i(w; \beta_0, \beta_1, \beta_2)$ $(i = 1, 2)$ are isometric (resp. similar).

Now by the following property (α, β, γ isometries of $\mathbf{E}$):

PROPERTY 4.2.3 (Dress [1], Section 3, Corollary 1).

$$\{(\dim\alpha, \dim\beta, \dim\gamma) | \alpha, \beta, \gamma \neq \mathrm{Id}_{\mathbf{E}} \text{ and } \alpha^2 = \beta^2 = \gamma^2 = \alpha\beta\gamma = \mathrm{Id}_{\mathbf{E}}\}$$

$$= \{(a, b, c) \in \mathbf{N}^3 | 0 \leqq a, b, c \leqq 2 \text{ and } a + b + c \in \{3, 5\}\}$$

we necessarily have $\dim(\alpha_1) + \dim(\alpha_2) + \dim(\alpha_1\alpha_2) \in \{3, 5\}$.

Therefore $(\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2))$ can only be one of the following:

- $(0, 1, 2)$
- $(1, 1, 1)$
- $(1, 2, 2)$.

In the following lemmas we are going to exclude the dimensions of $\mathbf{E}^{\alpha_0}$ which correspond to systems of rank strictly smaller than 3 respectively have their vertices in a plane.

LEMMA 4.2.4. *If* $(\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2))$ *is* $(0, 1, 2)$, *then* $\dim(\alpha_0) = 1$ *and* $\mathbf{E}^{\alpha_0}$ *is neither parallel nor perpendicular to* $\mathbf{E}^{\alpha_2}$.

PROOF. If $\dim(\alpha_0) = 0$, then $\langle \alpha_0, \alpha_1, \alpha_2 \rangle \cdot v$ is contained in the plane $\langle \mathbf{E}^{\alpha_0}, \mathbf{E}^{\alpha_2} \rangle$ and therefore $\mathrm{rk}(G \cap T) \leqq 2$.

If $\dim(\alpha_0) = 1$ and $\mathbf{E}^{\alpha_0}$ is parallel (resp. perpendicular) to $\mathbf{E}^{\alpha_2}$ then the plane through $\mathbf{E}^{\alpha_2}$ and through (resp. perpendicular to) $\mathbf{E}^{\alpha_0}$ is invariant and therefore $\mathrm{rk}(G \cap T) \leqq 2$.

If $\dim(\alpha_0)=2$, denote $\mathbf{E}^{\alpha_0}_{\perp}$ the line perpendicular to $\mathbf{E}^{\alpha_0}$ and passing through v. Now $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v$ is contained in the plane $\langle\mathbf{E}^{\alpha_0}_{\perp},\mathbf{E}^{\alpha_2}\rangle$ and therefore $\mathrm{rk}(G\cap T)\leqq 2$. □

LEMMA 4.2.5. *If* $(\dim(\alpha_1),\ \dim(\alpha_2),\ \dim(\alpha_1\alpha_2))$ *is* $(1,1,1)$ *and* $\dim(\alpha_0)=1$ *then* $\mathbf{E}^{\alpha_0}$ *cannot be parallel to* $\mathbf{E}^{\alpha_1\alpha_2}$, *to* $\mathbf{E}^{\alpha_1}$ *or to* $\mathbf{E}^{\alpha_2}$; *if* $\dim(\alpha_0)=2$ *then* $\mathbf{E}^{\alpha_0}$ *cannot be parallel or perpendicular to* $\mathbf{E}^{\alpha_1\alpha_2}$, *to* $\mathbf{E}^{\alpha_1}$ *or to* $\mathbf{E}^{\alpha_2}$.

PROOF. If $\dim(\alpha_0)=1$, then in the first case e.g. $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v$ is contained in the plane $\langle\mathbf{E}^{\alpha_1},\mathbf{E}^{\alpha_2}\rangle$ and therefore $\mathrm{rk}(G\cap T)\leqq 2$.

If $\dim(\alpha_0)=2$, then in the first case e.g. $\langle\alpha_0,\alpha_1,\alpha_2\rangle\cdot v$ is contained in the plane $\langle\mathbf{E}^{\alpha_1},\mathbf{E}^{\alpha_2}\rangle$ and therefore $\mathrm{rk}(G\cap T)\leqq 2$. □

LEMMA 4.2.6. *If* $(\dim(\alpha_1),\dim(\alpha_2),\dim(\alpha_1\alpha_2))$ *is* $(1,2,2)$, *then* $\dim(\alpha_0)=1$. *Also* $\mathbf{E}^{\alpha_0}\cap\mathbf{E}^{\alpha_1}=\emptyset$ *and* $\mathbf{E}^{\alpha_0}$ *is not parallel to* $\mathbf{E}^{\alpha_2}$ *or* $\mathbf{E}^{\alpha_1\alpha_2}$ *and is not perpendicular to* $\mathbf{E}^{\alpha_1}$.

PROOF. If $\dim(\alpha_0)=0$, then $\mathrm{rk}(G\cap T)\leqq 2$ as the plane perpendicular to $\mathbf{E}^{\alpha_1}$ and containing $\mathbf{E}^{\alpha_0}$ is invariant.

If $\dim(\alpha_0)=1$ then $\mathbf{E}^{\alpha_0}\cap\mathbf{E}^{\alpha_1}=\emptyset$ as otherwise the system would be bounded. Also $\mathbf{E}^{\alpha_0}$ cannot be parallel to $\mathbf{E}^{\alpha_2}$ or $\mathbf{E}^{\alpha_1\alpha_2}$ as otherwise the system admits an invariant line. Finally $\mathbf{E}^{\alpha_0}$ cannot be perpendicular to $\mathbf{E}^{\alpha_1}$ as otherwise the system admits an invariant plane.

If $\dim(\alpha_0)=2$ and if $\mathbf{E}^{\alpha_0}$ is not parallel to $\mathbf{E}^{\alpha_1}$, then the system is finite as it admits an invariant point. If $\mathbf{E}^{\alpha_0}$ is parallel to $\mathbf{E}^{\alpha_1}$ then the system admits an invariant plane. □

Now, if $(w;\beta_0,\beta_1,\beta_2)$ is another Grünbaum system with the same angularity as $(v;\alpha_0,\alpha_1,\alpha_2)$, then in all three cases we can find an isometry γ with $\gamma v=w$, $\gamma\alpha_1\gamma^{-1}=\beta_1$ and $\gamma\alpha_2\gamma^{-1}=\beta_2$. Moreover we can compose γ with an isometry (resp. similarity) δ with $\delta w=w$, $\delta\beta_1\delta^{-1}=\beta_1$, $\delta\beta_2\delta^{-1}=\beta_2$ so that $\delta\gamma\alpha_0\gamma^{-1}\delta^{-1}=\beta_0$ if and only if the distances (resp. quotients of the distances) $d(\mathbf{E}^{\alpha_0},v)$, $d(\mathbf{E}^{\alpha_0},\mathbf{E}^{\alpha_1})$, $d(\mathbf{E}^{\alpha_0},\mathbf{E}^{\alpha_2})$ and $d(\mathbf{E}^{\alpha_0},\mathbf{E}^{\alpha_1\alpha_2})$ coincide with the corresponding distances (resp. quotients of the distances) in the system $(w;\beta_0,\beta_1,\beta_2)$. In more detail, we get the following

LEMMA 4.2.7. *Suppose* $(v;\alpha_0,\alpha_1,\alpha_2)$ *and* $(w;\beta_0,\beta_1,\beta_2)$ *are two Grünbaum systems with the same angularity. Then we have the following three cases:*

- $\dim(\alpha_0)=0$ *and* $(\dim(\alpha_1),\dim(\alpha_2),\dim(\alpha_1\alpha_2))$ *is* $(1,1,1)$*:*
  $(v;\alpha_0,\alpha_1,\alpha_2)$ *is isometric to* $(w;\beta_0,\beta_1,\beta_2)$ *if and only if the distances* $d=d(\mathbf{E}^{\alpha_0},v)$, $d_1=d(\mathbf{E}^{\alpha_0},\mathbf{E}^{\alpha_1})$ *and* $d_2=d(\mathbf{E}^{\alpha_0},\mathbf{E}^{\alpha_2})$ *coincide with the corresponding distances in the system* $(w;\beta_0,\beta_1,\beta_2)$. *Moreover,* $\max(d_1,d_2)<d<\sqrt{d_1^2+d_2^2}$.

  $(v;\alpha_0,\alpha_1,\alpha_2)$ *is similar to* $(w;\beta_0,\beta_1,\beta_2)$ *if and only if the quotients* $\dfrac{d}{d_1}>1$

and $\dfrac{d}{d_2} > 1$ *coincide with the corresponding quotients in the system* $(w; \beta_0, \beta_1, \beta_2)$.

- $\dim(\alpha_0) = 1$ *and* $(\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2))$ *is* $(0,1,2)$, $(1,1,1)$ *or* $(1,2,2)$*:*
  $(v; \alpha_0, \alpha_1, \alpha_2)$ *is isometric to* $(w; \beta_0, \beta_1, \beta_2)$ *if and only if the distances* $d = d(\mathbf{E}^{\alpha_0}, v)$ *and* $d' = d(\mathbf{E}^{\alpha_0}, \mathbf{E}^{\alpha_2})$ *(resp.* $d' = d(\mathbf{E}^{\alpha_0}, \mathbf{E}^{\alpha_1\alpha_2})$ *in the case* $(2,2,1)$*) coincide with the corresponding distances in the system* $(w; \beta_0, \beta_1, \beta_2)$. *Moreover,* $d' < d$.
  $(v; \alpha_0, \alpha_1, \alpha_2)$ *is similar to* $(w; \beta_0, \beta_1, \beta_2)$ *if and only if the quotient* $\dfrac{d}{d'} > 1$ *coincides with the corresponding quotient in the system* $(w; \beta_0, \beta_1, \beta_2)$.
- $\dim(\alpha_0) = 2$ *and* $(\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2))$ *is* $(1,1,1)$*:*
  $(v; \alpha_0, \alpha_1, \alpha_2)$ *is isometric to* $(w; \beta_0, \beta_1, \beta_2)$ *if and only if the distance* $d(\mathbf{E}^{\alpha_0}, v)$ *coincides with the corresponding distance in the system* $(w; \beta_0, \beta_1, \beta_2)$.
  $(v; \alpha_0, \alpha_1, \alpha_2)$ *is similar to* $(w; \beta_0, \beta_1, \beta_2)$.

Now consider a line $\mathbf{L}$ with a unitary directional vector $\vec{u}$ (resp. a plane $\mathbf{P}$ with a unitary normal vector $\vec{u}$), and an orthonormal basis $(\overrightarrow{e_0}, \overrightarrow{e_1}, \overrightarrow{e_2})$. Call ψ_i the angle formed by $\vec{u}$ and the basis vector $\overrightarrow{e_i}$ $(i = 0, 1, 2)$.

Decomposing $\vec{u}$ in the orthonormal basis:

$$\vec{u} = (\vec{u}|\overrightarrow{e_1})\overrightarrow{e_1} + (\vec{u}|\overrightarrow{e_2})\overrightarrow{e_2} + (\vec{u}|\overrightarrow{e_3})\overrightarrow{e_3}$$

we get:

$$\cos^2\psi_0 + \cos^2\psi_1 + \cos^2\psi_2 = 1.$$

Combining this remark with the preceeding lemmas and with Property 4.1.1. we obtain:

THEOREM 4.2.8. *If* $(v; \alpha_0, \alpha_1, \alpha_2)$ *is a non-planar and non-finite Grünbaum system without an invariant plane, then up to coloured Petrie duality and up to a switch in the colour of the faces, the 7-tupel*

$$(\dim(\alpha_0)|\dim(\alpha_1), \dim(\alpha_2), \dim(\alpha_1\alpha_2)|\widehat{(\alpha_0, \alpha_1)}, \widehat{(\alpha_0, \alpha_2)}, \widehat{(\alpha_0, \alpha_1\alpha_2)})$$

can assume the following values only:

(1) $(0|1,1,1|0°,0°,0°)$ *(2-parameter family of similarity classes),*

(2) $(1|0,1,2|0°,60°,30°)$, $(1|0,1,2|0°,45°,45°)$, $(1|0,1,2|0°,30°,60°)$ *(1-parameter families of similarity classes),*

(3) $(1|1,1,1|45°,60°,60°)$, $(1|1,1,1|60°,30°,90°)$, $(1|1,1,1|45°,45°,90°)$ *(1-parameter families of similarity classes),*

(4) $(1|1,2,2|45°,30°,30°)$, $(1|1,2,2|60°,45°,30°)$ *(1-parameter families of similarity classes),*

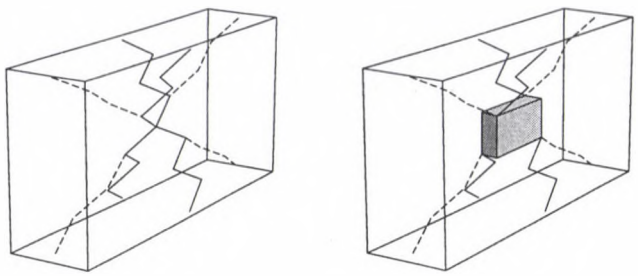


Fig. 4. The vertex figure of a coloured polyhedron $(0|111|00, 00, 00)$

(5) $(2|1, 1, 1|30°, 30°, 45°)$ *(1 similarity class).*

EXPLANATION 4.2.9.

1. This family (cf. Figure 4) contains the 1-parameter family of Grünbaum polyhedra denoted $\{\infty^{z(b)}, 4^{\alpha*(b)}/1\}$ [1, 3].
2. These polyhedra are related to the tilings of the plane by squares, respectively, by triangles and hexagons. Figure 5 represents three such polyhedra related to each other by the coloured Petrie operators.
3. The relation

$$\cos^2(\widehat{(\alpha_0, \alpha_1)}) + \cos^2(\widehat{(\alpha_0, \alpha_2)}) + \cos^2(\widehat{(\alpha_0, \alpha_1\alpha_2)}) = 1$$

restricts the number of possible polyhedra.

- We shall describe the class $(1|1, 1, 1|45°, 60°, 60°)$ in more detail below in Example 4.2.10.
- The other angularity classes correspond to polyhedra with vertical helical polygons based on the planar tilings with finite polygons, e.g. squares, respectively, triangles and hexagons. Figure 6 represents three such polyhedra related to each other by the coloured Petrie operators. In the class $(1|1, 1, 1|45°, 45°, 90°)$ we find the 1-parameter family of polyhedra $\{\infty^{\alpha(b),\pi/2}, 4^{\alpha*(b)}/1\}$ [1, 3].

4. The relation

$$\cos^2(\psi_0) + \cos^2(\psi_1) + \cos^2(\psi_2) = 1,$$

where ψ_i denotes the angle formed by $\mathbf{E}^{\alpha_0}$ and a basis vector $\overrightarrow{e_i}$ $(i = 0, 1, 2)$, restricts the number of possible polyhedra.

- In the first angularity class we find the regular polyhedron $\{6^{\pi/2}/1, 4\}$ having skew hexagons as faces (cf. Figure 7, $e = 0$). This polyhedron is related through a 1-parameter family of similarity classes (cf. Figure 7) to the polyhedron obtained by a truncation of $\{6, 6a\}$ and having plane and skew hexagons as faces (cf. Figure 7, $e = 0.5$ and Figure 8 where one skew hexagon bordered by six planar hexagons is shown).

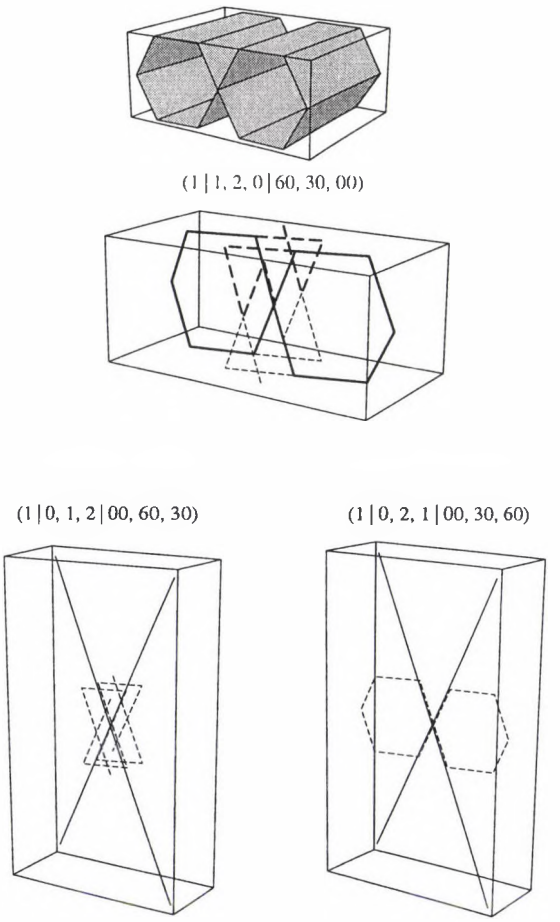


Fig. 5. Three coloured polyhedra based on a planar tiling with triangles and hexagons related through Petric operators

- In the second angularity class we mention two polyhedra, the first one related to a truncated $\{4, 6a\}$ and having plane squares and skew hexagons as faces (cf. Figure 9, where one skew hexagon bordered by six squares is shown), the second one related to a truncated $\{6, 4a\}$ and having plane hexagons and skew quadrangles as faces (cf. Figure 10, where four skew quadrangles, each one bordered by four hexagons, are shown).

5. This follows from the relation

$$\sin^2(\widehat{(\alpha_0, \alpha_1)}) + \sin^2(\widehat{(\alpha_0, \alpha_2)}) + \sin^2(\widehat{(\alpha_0, \alpha_1\alpha_2)}) = 1.$$

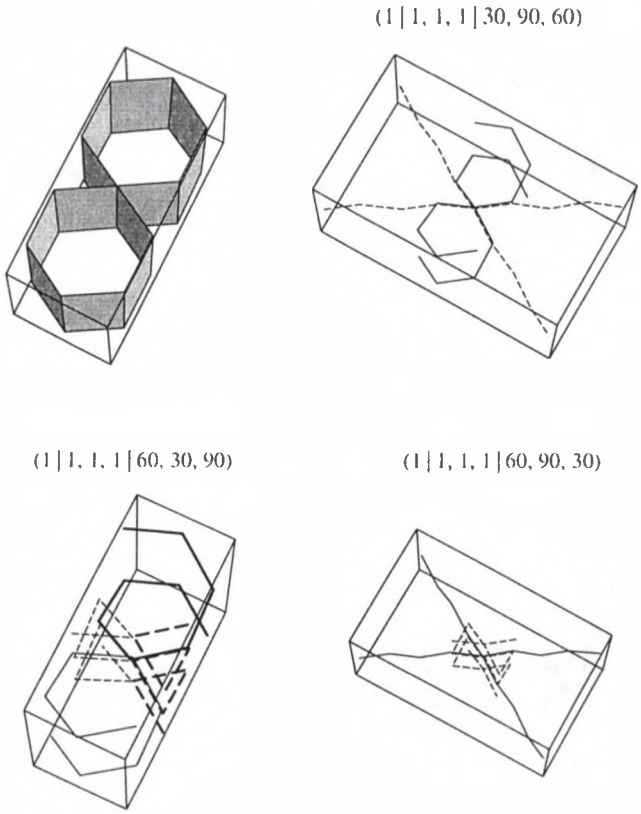


Fig. 6. Three coloured polyhedra based on a planar tiling with triangles and hexagons related through Petrie operators

In this class we find the Petrie–Coxeter polyhedron $\{6, 4^{48°\,12'}/1\}$ (cf. Figure 11). □

EXAMPLE 4.2.10. Consider the following analytical description of the three fundamental involutions of $(1|1,1,1|45°,60°,60°)$:

- $\mathbf{E}^{\alpha_0}$: $y = q$; $z = -x + p$ with $p, q \in \mathbf{R}$,
- $\mathbf{E}^{\alpha_1}$: $x = 0$; $y = 0$,
- $\mathbf{E}^{\alpha_2}$: $z = 0$; $x = y$,

and consider the point $P(p, q, 0) \in \mathbf{E}^{\alpha_0}$.

Then we get the following special cases:

- $p = 0$ $(1|1,1,1|45°,60°,60°)$ can obtained by applying a coloured Petrie operator to $(1|1,1,1|60°,60°,45°)$, which is the so-called 48-th Grünbaum polyhedron $\{\infty^{\pi/2,2\pi/3}, 4\}$ [1] (cf. Figure 12).

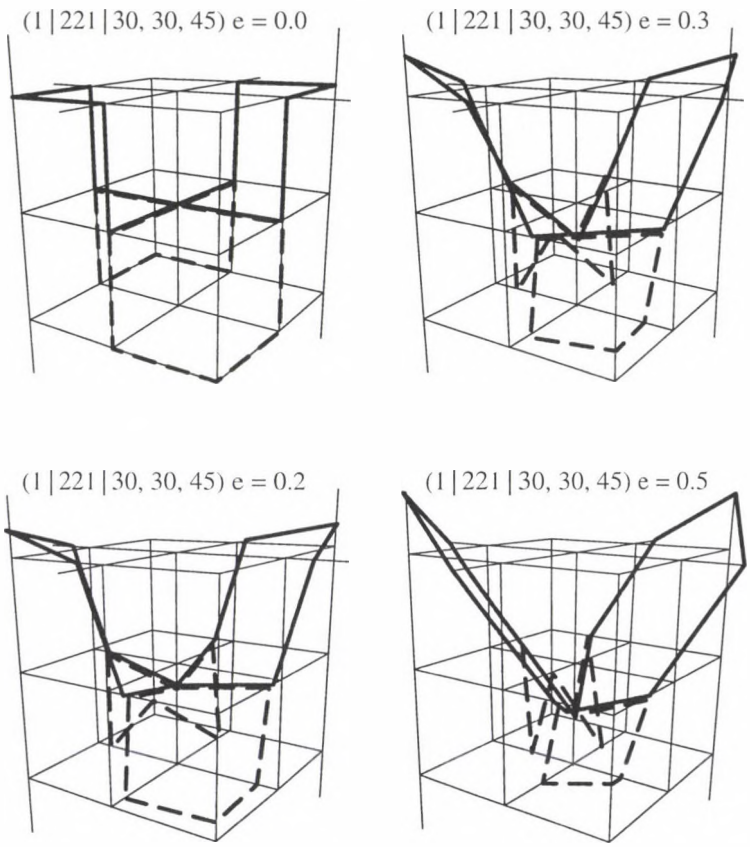
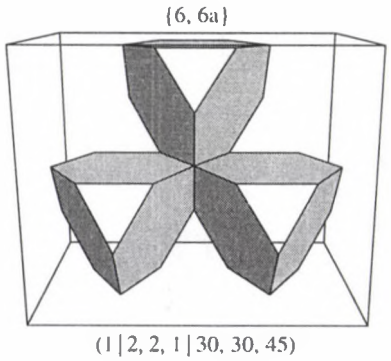


Fig. 7. The similarity class of $(1|221|30, 30, 45)$*: four examples*

- $q = 0$ This polyhedron is related to the Coxeter–Petrie polyhedron $\{\infty^{2\pi/3,2\pi/3}, 6^{\pi/3}/1\}$ [1, 3]. At each vertex it consists of two square polygons visible in $\{6, 4a\}$, and two Petrie polygons of $\{6, 4a\}$ (cf. Figure 13).

- $p = q$ The polyhedron can be thought of as obtained from helices with square base (cf. Figure 14).

Note that the 1-parameter similarity class of $(1, 1, 1, 1, 45°, 60°, 60°)$ can only be *discrete polyhedra* for well-chosen values of the ratio p/q. In fact, a polyhedron of this family can be thought of as obtained by fitting together square helical polygons with axis in direction of the orthonormal basis vectors $\overrightarrow{e_i}$ $(i = 0,1,2)$. Figure 15 represents the projection onto the (x, y)-plane of such a polyhedron.



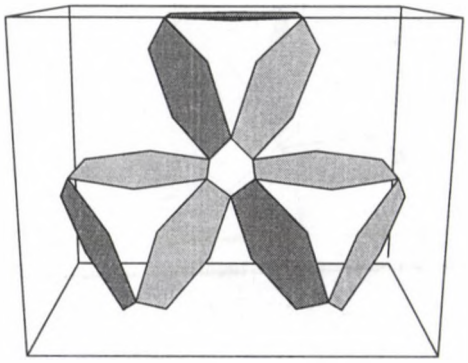
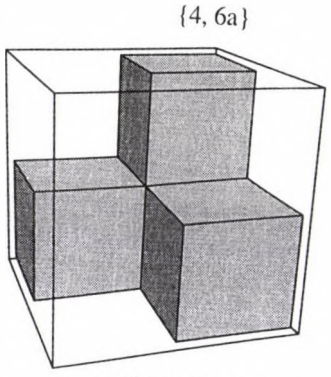


Fig. 8. $(1|221|30, 30, 45)$ *as a truncation of* $\{6, 6a\}$



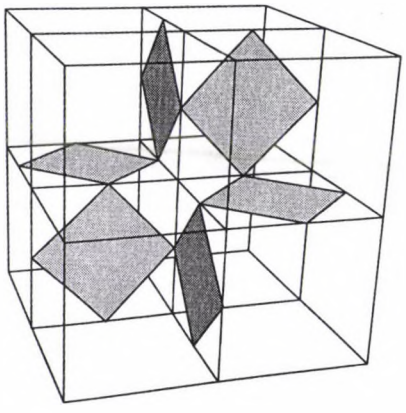


Fig. 9. (1|221|30, 45, 60) *as a truncation of* $\{4, 6a\}$

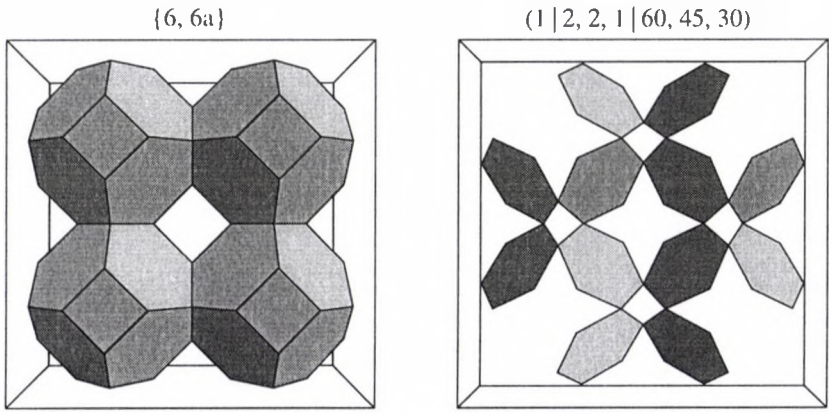


Fig. 10. $(1|221|60, 45, 30)$ *as a truncation of* $\{6, 4a\}$

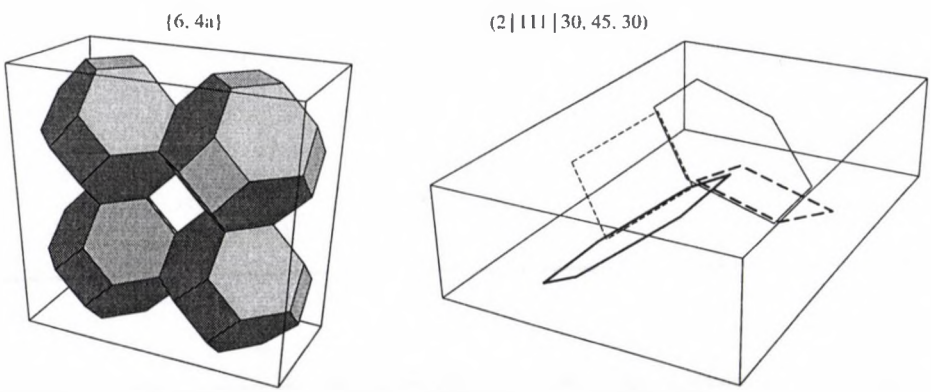


Fig. 11. The vertex figure of $(2|111|30, 45, 30)$ *related to* $\{6, 4a\}$

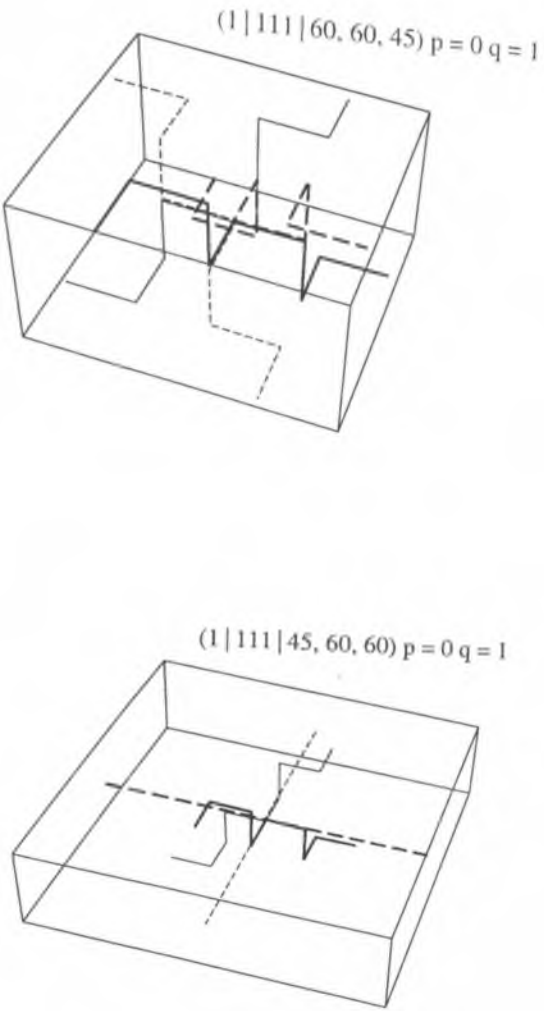
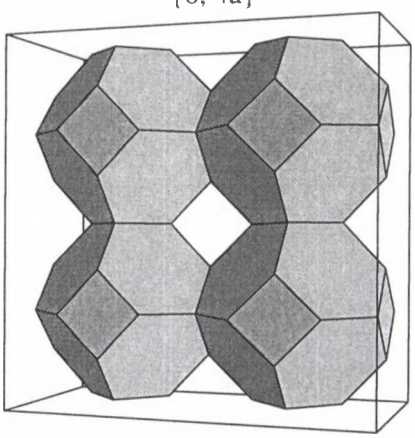


Fig. 12. $(1|111|45, 60, 60)$ *in relation to the* 48^{th} *Grünbaum polyhedron* $(1|111|60, 60, 45)$

{6, 4a}



(1 | 111 | 45, 60, 60) p = 1 q = 0

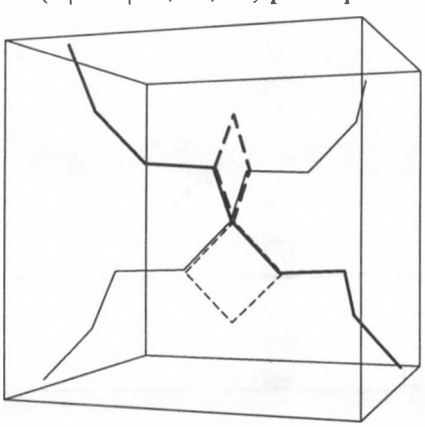


Fig. 13. (1|111|45, 60, 60) *related to* {6, 4a}

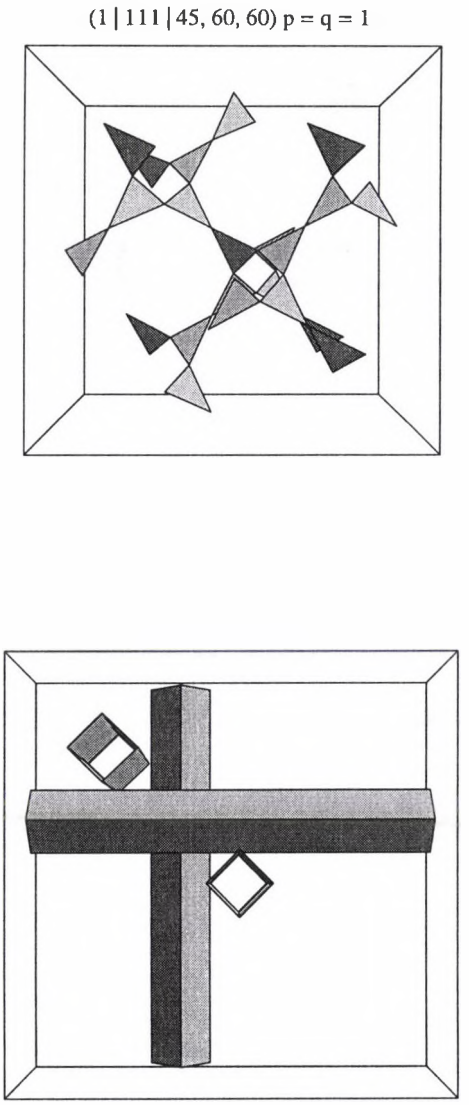


Fig. 14. (1|111|45, 60, 60) *related to four defining prisms*

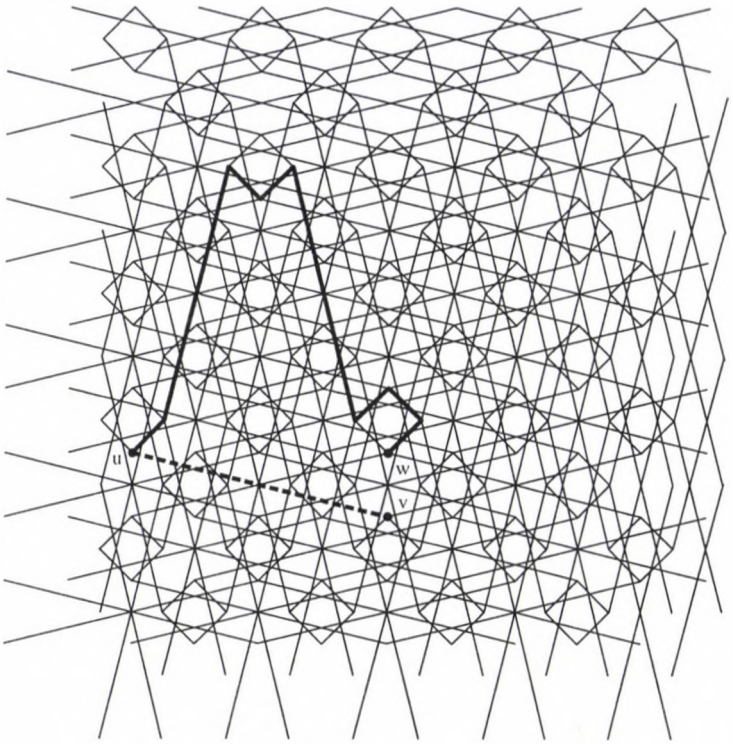


Fig. 15. Projection of $(1|111|45, 60, 60)$ *onto the* (x, y)*-plane*

Now consider helices of diameter 1 and denote $u'(d, 0, 0)$ a vertex which is a full turn along a helical polygon away from the preceeding vertex $u(0, 0, 0)$. Here d depends on p and q. Following the two helical paths indicated by dashed (resp. bold) segments on Figure 15, we first reach the vertices $v\left(\frac{d}{4}, y, 0\right)$ and $w(4, -y, 0)$. Repeating this movement with v and w as starting points, we reach after four iterations the two vertices $u'(d, 0, 0)$ and $u''(16, 0, 0)$ of the polyhedron.

Comparing coordinates, we now see that the polyhedron can only be discrete if d is rational.

4.3. Acknowledgements. We would like to express our gratitude to the organizers of the Budapest conference on Intuitive Geometry for their hospitality, and the participants together with the referee for their inspiring comments.

All figures have been realized in Mathematica (and redrawn in CorelDraw).

REFERENCES

[1] DRESS, A., A combinatorial theory of Grünbaum's new regular polyhedra. II. Complete enumeration, *Aequationes Math.* **29** (1985), 222–243. *MR* **87e**:51028

[2] FARRIS, S. L., Completely classifying all vertex-transitive and edge-transitive polyhedra. I. Necessary class conditions, *Geom. Dedicata* **26** (1988), 111–124. *MR* **89e**:52008

[3] GRÜNBAUM, B., Regular polyhedra – old and new, *Aequationes Math.* **16** (1977), 1–20. *MR* **57** #7353

[4] LEYTEM, C., Regular coloured polyhedra with tetragonal vertex figure (preprint).

(Received August 20, 1995)

LYCÉE CLASSIQUE DE DIEKIRCH
32, AVENUE DE LA GARE
L–9233 DIEKIRCH
GRAND-DUCHÉ DE LUXEMBOURG

cleytem@pop2.restena.lu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 39–51

POSITIVE SOLUTIONS OF SECOND ORDER QUASILINEAR ORDINARY DIFFERENTIAL EQUATIONS WITH GENERAL NONLINEARITIES

J. KIYOMURA, T. KUSANO and M. NAITO

1. Introduction

In this paper we are concerned with positive solutions of the quasilinear ordinary differential equation

$$(|x'|^{\alpha-1}x')' + p(t)f(x) = 0, \quad t \geqq t_0, \tag{1.1}$$

where the following conditions are always assumed:

$$\alpha > 0 \text{ is a constant;} \tag{1.2}$$

$$\begin{aligned} &p(t) \text{ is continuous on } [t_0, \infty), \ t_0 > 0, \text{ and} \\ &p(t) \geqq 0 \quad (t \geqq t_0); \end{aligned} \tag{1.3}$$

$$f(x) \text{ is continuous on } (0, \infty) \text{ and } f(x) > 0 \ \ (x > 0). \tag{1.4}$$

In the oscillation theory for differential equations one of the important problems is to find necessary and/or sufficient conditions for the equations under consideration to have positive solutions. For the equation (1.1), necessary and sufficient conditions can be established by restricting the nonlinearity of $f(x)$ to various classes of functions. As an example, consider the case $f(x) = x^\beta$, $\beta > 0$. In this case, equation (1.1) becomes

$$(|x'|^{\alpha-1}x')' + p(t)x^\beta = 0, \quad t \geqq t_0. \tag{1.5}$$

The existence and asymptotic behaviour of positive solutions of (1.5) have recently been studied by several authors. It is known (Elbert and Kusano [2]) that equation (1.5) has an unbounded positive solution $x(t)$ such that $\lim_{t\to\infty}[x(t)/t] = l_1$ for some $l_1 \in (0, \infty)$ if and only if

$$\int_{t_0}^{\infty} s^\beta p(s)ds < \infty; \tag{1.6}$$

1991 *Mathematics Subject Classification.* Primary 34C10; Secondary 34C11.
Key words and phrases. Positive solutions, quasilinear ordinary differential equations.

0081–6906/99/$ 5.00

and equation (1.5) has a bounded positive solution $x(t)$ such that $\lim_{t\to\infty} x(t)=l_2$ for some $l_2 \in (0,\infty)$ if and only if

$$\int_{t_0}^{\infty}\left(\int_{s}^{\infty} p(r)dr\right)^{1/\alpha} ds < \infty. \tag{1.7}$$

It is also known ([2]) that the superlinear equation (1.5) with $\beta > \alpha$ has a positive solution if and only if (1.7) holds, and the sublinear equation (1.5) with $0 < \beta < \alpha$ has a positive solution if and only if (1.6) holds. In the paper of Elbert and Kusano [2], a more general equation is considered. However, little is known about the case where f has a general nonlinearity.

The purpose of this paper is to investigate the existence and asymptotic behaviour of positive solutions of (1.1) for the case where f has a general nonlinearity. It is emphasized that, while condition (1.4) is assumed, no additional condition on $f(x)$ is hypothesized. We do not require any condition on $f(x)$ such as monotone conditions or growth conditions as $x \to \infty$. Instead, we need a kind of monotone condition on the coefficient $p(t)$.

In Section 2 we give necessary and sufficient conditions for (1.1) to have positive solutions $x(t)$ with special asymptotic properties as $t \to \infty$, and in Section 3 we provide a necessary and sufficient condition for (1.1) to have a positive solution $x(t)$. Here the terminology of "positive solution" of (1.1) is used in the eventual sense. That is, $x(t)$ is said to be a positive solution of (1.1) if and only if it is defined and is positive and satisfies (1.1) on some interval $[t_x, \infty)$, $t_x \geqq t_0$.

In the case $\alpha = 1$, the related oscillation theorems for equation (1.1) with general nonlinearity were obtained in the papers of Wong [4] and Burton and Grimmer [1]. Later, a systematic investigation was made in the paper of Naito [3], and the theorems in [4] and [1] were extended to a more general case. The results in this paper give a further extension of the corresponding results in [3].

2. Existence of maximal and minimal positive solutions

In this section we study the existence of positive solutions of (1.1) with special asymptotic properties as $t \to \infty$.

Let $x(t)$ be a positive solution of (1.1) on $[t_x, \infty)$, $t_x \geqq t_0$. We easily find that $x'(t)$ is nonnegative and nonincreasing for $t \geqq t_x$. Hence, $\lim_{t\to\infty} x'(t) = l_1$ exists and is nonnegative. We have $\lim_{t\to\infty} [x(t)/t] = l_1$ as an application of L'Hospital's rule. On the other hand, since $x(t)$ is positive and nondecreasing for $t \geqq t_x$, $\lim_{t\to\infty} x(t) = l_2$ exists in the extended real line $\overline{\mathbf{R}}$ and either $l_2 = \infty$ or l_2 is a positive finite value. It is clear that if $l_2 = \lim_{t\to\infty} x(t)$ is positive

and finite, then $l_1 = \lim_{t\to\infty}[x(t)/t] = 0$. It is also clear that if $l_1 = \lim_{t\to\infty}[x(t)/t]$ is positive and finite, then $l_2 = \lim_{t\to\infty} x(t) = \infty$. Then it is easy to see that, for a positive solution $x(t)$ of (1.1), one of the following three asymptotic conditions is satisfied:

$$\lim_{t\to\infty} \frac{x(t)}{t} \text{ exists as a finite positive value;} \tag{2.1}$$

$$\lim_{t\to\infty} \frac{x(t)}{t} = 0 \quad \text{and} \quad \lim_{t\to\infty} x(t) = \infty; \tag{2.2}$$

$$\lim_{t\to\infty} x(t) \text{ exists as a finite positive value.} \tag{2.3}$$

From these observations we see that, for a positive solution $x(t)$ of (1.1) on $[t_x, \infty)$, there are positive constants a_1 and a_2 such that

$$0 < a_1 \leqq x(t) \leqq a_2 t, \quad t \geqq t_x. \tag{2.4}$$

Thus, among all positive solutions of (1.1), those which are asymptotic to at $(a > 0)$ as $t \to \infty$ can be regarded as the maximal solutions, and those which are asymptotic to a $(a > 0)$ as $t \to \infty$ can be regarded as the minimal solutions. The purpose of this section is to present necessary and sufficient conditions for (1.1) to have positive solutions of these two special types.

THEOREM 2.1. *Suppose that $p(t)$ is decomposable in such a way that $p(t) = c(t)q(t)$, $c(t)$ is continuous and positive on $[t_0, \infty)$, $0 < \liminf_{t\to\infty} c(t) \leqq \limsup_{t\to\infty} c(t) < \infty$, and $t^\sigma q(t)$ is continuous and nondecreasing on $[t_0, \infty)$ for some real number σ.*

Then, equation (1.1) *has a positive solution $x(t)$ which satisfies* (2.1) *if and only if*

$$\int_{t_0}^{\infty} p(s)f(cs)ds < \infty \quad \textit{for some } c > 0. \tag{2.5}$$

PROOF. There is no loss of generality in assuming that $\sigma \geqq 0$. From the assumption on $c(t)$, we have

$$c_1 \leqq c(t) \leqq c_2, \quad t \geqq t_0, \tag{2.6}$$

for some positive constants c_1 and c_2.

We first prove the "only if" part. Let $x(t)$ be a positive solution of (1.1) satisfying (2.1). We suppose that $x(t)$ is defined, is positive and satisfies (1.1) on $[t_x, \infty)$, $t_x \geqq t_0$, and that

$$\lim_{t\to\infty} \frac{x(t)}{t} = l \in (0, \infty).$$

Then there exists $T \geqq 4t_x$ such that

$$\frac{l}{2}t \leqq x(t) \leqq 2lt, \quad t \geqq T. \tag{2.7}$$

It follows from (2.7) that

$$\frac{x(t)}{2l} \leqq t \leqq \frac{2x(t)}{l}, \quad t \geqq T. \tag{2.8}$$

Then, using (2.6) and (2.8) and the assumption on $p(t) = c(t)q(t)$, we get

$$p(t) \geqq kp\left(\frac{x(t)}{2l}\right), \quad t \geqq T, \tag{2.9}$$

where $k = c_1/[4^\sigma c_2]$. In fact, we compute as follows:

$$\begin{aligned} p(t) &= c(t)t^{-\sigma}t^\sigma q(t) \\ &\geqq c_1\left(\frac{2x(t)}{l}\right)^{-\sigma}\left(\frac{x(t)}{2l}\right)^\sigma q\left(\frac{x(t)}{2l}\right) \\ &\geqq \frac{c_1}{4^\sigma c_2}p\left(\frac{x(t)}{2l}\right), \quad t \geqq T. \end{aligned}$$

Now, integrating equation (1.1) from T to t, we obtain

$$|x'(t)|^{\alpha-1}x'(t) - |x'(T)|^{\alpha-1}x'(T) + \int_T^t p(s)f(x(s))ds = 0, \quad t \geqq T.$$

As mentioned above, we have $x'(t) \geqq 0$ for $t \geqq t_x$. Hence we find that

$$\int_T^\infty p(s)f(x(s))ds < \infty,$$

which together with (2.9) implies

$$\int_T^\infty p\left(\frac{x(s)}{2l}\right)f(x(s))ds < \infty. \tag{2.10}$$

Observe that $\lim_{t\to\infty} x'(t) = \lim_{t\to\infty}[x(t)/t] = l$. In particular, there is a positive constant d such that

$$\frac{d}{dt}\left(\frac{x(t)}{2l}\right) \leqq d, \quad t \geqq T.$$

Hence (2.10) gives

$$\int_T^\infty p\left(\frac{x(s)}{2l}\right) f(x(s)) \frac{d}{ds}\left(\frac{x(s)}{2l}\right) ds < \infty.$$

Then, letting $v = x(s)/2l$, we arrive at

$$\int_{x(T)/2l}^\infty p(v) f(2lv) dv < \infty.$$

Thus the integral condition (2.5) is satisfied.

We next prove the "if" part. We suppose that (2.5) holds. Put $K = 2^\sigma c_2/c_1$, where c_1 and c_2 are positive constants appearing in (2.6). Choose $T > 2t_0$ such that

$$\int_T^\infty p(s) f(cs) ds < \frac{1}{K}(2^\alpha - 1)c^\alpha. \tag{2.11}$$

Let $C^1[T, \infty)$ denote the Fréchet space of all C^1-functions on $[T, \infty)$ with the usual metric topology, and let X be the set of all functions $x \in C^1[T, \infty)$ such that $x(T) = cT$ and

$$ct \leqq x(t) \leqq 2ct \quad \text{and} \quad c \leqq x'(t) \leqq 2c \quad \text{for } t \geqq T. \tag{2.12}$$

Here $c > 0$ is a constant in (2.5). We define the mapping $M : X \to C^1[T, \infty)$ by

$$(Mx)(t) = cT + \int_T^t \left(c^\alpha + \int_s^\infty p(r) f(x(r)) dr\right)^{1/\alpha} ds, \quad t \geqq T.$$

We will show that the Schauder–Tychonoff theorem ensures the existence of a fixed point $x = Mx \in X$, and that this $x(t)$ is a positive solution of (1.1) satisfying (2.1).

(i) *M maps X into X.* Let $x \in X$. It is clear that $(Mx)(T) = cT$. Since $ct \leqq x(t) \leqq 2ct$ for $t \geqq T$, we have

$$\frac{x(t)}{2c} \leqq t \leqq \frac{x(t)}{c}, \quad t \geqq T.$$

As in the proof of the "only if" part, it can be verified that

$$p(t) \leqq Kp\left(\frac{x(t)}{c}\right), \quad t \geqq T. \tag{2.13}$$

From (2.11)–(2.13) it follows that

$$\begin{aligned} 0 \leqq \int_T^\infty p(r)f(x(r))dr &\leqq K \int_T^\infty p\left(\frac{x(r)}{c}\right) f(x(r)) \frac{d}{dr}\left(\frac{x(r)}{c}\right) dr \\ &\leqq K \int_T^\infty p(v)f(cv)dv \\ &\leqq (2^\alpha - 1)c^\alpha. \end{aligned}$$

Then we easily see that $ct \leqq (Mx)(t) \leqq 2ct$ and $c \leqq (Mx)'(t) \leqq 2c$ for $t \geqq T$.

(ii) *M is continuous on X.* Let x_i $(i = 1, 2, \dots)$ and x be functions in X such that $x_i(t) \to x(t)$, $x_i'(t) \to x'(t)$ as $i \to \infty$ uniformly on every compact subinterval of $[T, \infty)$. Note that all the inverse functions $x_i^{-1}(t)$ $(i = 1, 2, \dots)$ and $x^{-1}(t)$ are defined on the common interval $[cT, \infty)$. If $t \geqq T$, then we have

$$\begin{aligned} &\left| \int_t^\infty p(r)f(x_i(r))dr - \int_t^\infty p(r)f(x(r))dr \right| \\ &= \left| \int_{x_i(t)/c}^\infty p(x_i^{-1}(cs))f(cs) \frac{c}{x_i'(x_i^{-1}(cs))} ds \right. \\ &\quad \left. - \int_{x(t)/c}^\infty p(x^{-1}(cs))f(cs) \frac{c}{x'(x^{-1}(cs))} ds \right| \\ &\leqq \left| \int_{x_i(t)/c}^{x(t)/c} p(x_i^{-1}(cs))f(cs) \frac{c}{x_i'(x_i^{-1}(cs))} ds \right| \\ &\quad + c \int_{x(t)/c}^\infty \left| \frac{p(x_i^{-1}(cs))}{x_i'(x_i^{-1}(cs))} - \frac{p(x^{-1}(cs))}{x'(x^{-1}(cs))} \right| f(cs)ds \\ &\leqq K \left| \int_{x_i(t)/c}^{x(t)/c} p(s)f(cs)ds \right| \\ &\quad + c \int_T^\infty \left| \frac{p(x_i^{-1}(cs))}{x_i'(x_i^{-1}(cs))} - \frac{p(x^{-1}(cs))}{x'(x^{-1}(cs))} \right| f(cs)ds. \end{aligned}$$

Here we have used (2.13) with $x = x_i$ and (2.12) in the last step. Then, for any compact interval of the form $[T, S] \subset [T, \infty)$, we obtain

$$\begin{aligned}
&\sup_{T \leqq t \leqq S} \left| \int_t^\infty p(r) f(x_i(r)) dr - \int_t^\infty p(r) f(x(r)) dr \right| \\
&\leqq \frac{K}{c} \Big\{ \sup_{T \leqq s \leqq 2S} p(s) f(cs) \Big\} \Big\{ \sup_{T \leqq t \leqq S} |x_i(t) - x(t)| \Big\} \\
&\quad + c \int_T^\infty \left| \frac{p(x_i^{-1}(cs))}{x_i'(x_i^{-1}(cs))} - \frac{p(x^{-1}(cs))}{x'(x^{-1}(cs))} \right| f(cs) ds \\
&\to 0 \quad \text{as} \quad i \to \infty.
\end{aligned}$$

This fact shows that

$$\int_t^\infty p(r) f(x_i(r)) dr \to \int_t^\infty p(r) f(x(r)) dr \quad (i \to \infty)$$

uniformly on every compact subinterval of $[T, \infty)$. Then it is easy to see that $(Mx_i)'(t) \to (Mx)'(t)$ and $(Mx_i)(t) \to (Mx)(t)$ as $i \to \infty$ uniformly on every compact subinterval of $[T, \infty)$. Thus M is continuous on X.

(iii) $M(X)$ *is relatively compact.* Let $x \in X$. From (i) it follows that $ct \leqq (Mx)(t) \leqq 2ct$ and $c \leqq (Mx)'(t) \leqq 2c$ for $t \geqq T$. Put $C = c$ for $\alpha \geqq 1$ and $C = 2c$ for $0 < \alpha < 1$. If $t_1, t_2 \in [T, S]$, then

$$\begin{aligned}
&\alpha C^{\alpha-1} |(Mx)'(t_1) - (Mx)'(t_2)| \\
&\leqq \left| \{(Mx)'(t_1)\}^\alpha - \{(Mx)'(t_2)\}^\alpha \right| \\
&= \left| \int_{t_1}^{t_2} p(r) f(x(r)) dr \right| \\
&\leqq \Big\{ \sup_{T \leqq r \leqq S} p(r) \Big\} \Big\{ \sup_{cT \leqq u \leqq 2cS} f(u) \Big\} |t_1 - t_2|.
\end{aligned}$$

Thus $\{(Mx)(t)\}$ and $\{(Mx)'(t)\}$ are uniformly bounded and equicontinuous on every compact subinterval of $[T, \infty)$. Therefore, by the Ascoli–Arzelà theorem, $M(X)$ is relatively compact in the topology of $C^1[T, \infty)$.

From the above considerations we see that the Schauder–Tychonoff fixed point theorem can be applied to M, and so there exists an $x \in X$ such that $x(t) = (Mx)(t)$, $t \geqq T$. It is easily verified that $x(t)$ is a positive solution of (1.1) on $[T, \infty)$ and satisfies $\lim_{t \to \infty} [x(t)/t] = c > 0$. The proof of Theorem 2.1 is complete.

REMARK 2.1. Let the hypothesis on $p(t)$ in Theorem 2.1 be satisfied, and suppose that (2.5) holds. Then it is easy to see that

$$\int_{t_0}^{\infty} p(s)f(c's)ds < \infty \quad \text{for all } c' \geqq c,$$

where c is a positive number in (2.5). Thus the set of all $c>0$ satisfying (2.5) is an interval I which is contained in $\mathbf{R}_+ = (0, \infty)$. For example, if $p(t) = e^t$ and $f(x) = e^{-x}$, then $I = (1, \infty)$. By the proof of Theorem 2.1 we see that, if (2.5) is satisfied, then, for any c' with $c' \geqq c$, equation (1.1) has a positive solution $x(t)$ which satisfies $\lim_{t\to\infty}[x(t)/t] = c'$.

THEOREM 2.2. *Equation* (1.1) *has a positive solution* $x(t)$ *satisfying* (2.3) *if and only if*

$$\int_{t_0}^{\infty} \left(\int_{s}^{\infty} p(r)dr \right)^{1/\alpha} ds < \infty. \tag{2.14}$$

PROOF. Let $x(t)$ be a positive solution of (1.1) satisfying $\lim_{t\to\infty} x(t) = l$ with $l>0$. Since $x(t)$ is nondecreasing, there is a $T \geqq t_0$ such that

$$\frac{l}{2} \leqq x(t) \leqq l, \quad t \geqq T.$$

Integrating (1.1) from t to τ $(T \leqq t \leqq \tau)$, we get

$$|x'(\tau)|^{\alpha-1}x'(\tau) - |x'(t)|^{\alpha-1}x'(t) + \int_{t}^{\tau} p(s)f(x(s))ds = 0. \tag{2.15}$$

Since $x'(t)$ is nonnegative and nonincreasing, the existence of $\lim_{t\to\infty} x(t) = l$ implies $\lim_{t\to\infty} x'(t) = 0$. Then, let $\tau \to \infty$ in (2.15) to obtain

$$x'(t) = \left(\int_{t}^{\infty} p(s)f(x(s))ds \right)^{1/\alpha}, \quad t \geqq T.$$

Integrating the above equality from t to τ, and letting $\tau \to \infty$, we conclude that

$$x(t) = l - \int_{t}^{\infty} \left(\int_{s}^{\infty} p(r)f(x(r))dr \right)^{1/\alpha} ds, \quad t \geqq T.$$

From this it follows that

$$\int_T^\infty \left(\int_s^\infty p(r) f(x(r)) dr \right)^{1/\alpha} ds < \infty,$$

which implies

$$k_0^{1/\alpha} \int_T^\infty \left(\int_s^\infty p(r) dr \right)^{1/\alpha} ds < \infty$$

with $k_0 = \min\{f(u) : l/2 \leqq u \leqq l\}$. Thus (2.14) is satisfied.

Conversely, suppose that (2.14) is satisfied. Let c be an arbitrary positive number. Take $T > t_0$ such that

$$\int_T^\infty \left(\int_s^\infty p(r) dr \right)^{1/\alpha} ds < \frac{c}{2} K_0^{-1/\alpha}, \tag{2.16}$$

where $K_0 = \max\{f(u) : c/2 \leqq u \leqq c\}$, and define the subset X of $C[T, \infty)$ by

$$X = \left\{ x \in C[T, \infty) : \frac{c}{2} \leqq x(t) \leqq c, \ t \geqq T \right\}.$$

The space $C[T, \infty)$ is regarded as a Fréchet space with the topology of uniform convergence on every compact subinterval of $[T, \infty)$. We define the mapping $M : X \to C[T, \infty)$ by

$$(Mx)(t) = c - \int_t^\infty \left(\int_s^\infty p(r) f(x(r)) dr \right)^{1/\alpha} ds, \quad t \geqq T.$$

In view of (2.16), we see that M maps X into itself. It can be proved that M is continuous on X and that $M(X)$ is relatively compact in the topology of $C[T, \infty)$. Therefore, applying the Schauder–Tychonoff fixed point theorem, we find that there exists an $x \in X$ such that $x = Mx$. It is immediately verified that this fixed point $x(t)$ is a positive solution of (1.1) satisfying $\lim_{t \to \infty} x(t) = c > 0$. This sketches the proof of the "if" part. The details are left to the reader.

REMARK 2.2. By the proof of Theorem 2.2 we see that if (2.14) is satisfied, then, for any $c > 0$, equation (1.1) has a positive solution $x(t)$ such that $\lim_{t \to \infty} x(t) = c$.

3. Existence of positive solutions

The main purpose of this section is to establish a necessary and sufficient condition for the existence of a positive solution $x(t)$ of (1.1).

THEOREM 3.1. *Let $q \in C[t_0, \infty)$, $t_0 > 0$, be a function such that $p(t) \geqq q(t) \geqq 0$ $(t \geqq t_0)$ and $t^\beta q(t)$ is nondecreasing on $[t_0, \infty)$ for some $\beta < 1 + \alpha$. If equation* (1.1) *has a positive solution $x(t)$, then*

$$\int_{t_0}^{\infty} q(s) f(cs) ds < \infty \quad \text{for some } c > 0. \tag{3.1}$$

PROOF. It is enough to consider the case where $q(t) \not\equiv 0$ on $[t_0, \infty)$. We suppose that $x(t)$ is a positive solution of (1.1) on $[t_x, \infty)$, $t_x \geqq t_0$. Then $x'(t)$ is nonnegative and nonincreasing for $t \geqq t_x$, and $x(t)$ satisfies one of the asymptotic conditions (2.1)–(2.3). Assume that $x(t)$ satisfies (2.3). Then, by Theorem 2.2, we have (2.14). The assumption of the theorem implies $p(t) \geqq q(t) \geqq t_1^\beta q(t_1) t^{-\beta} > 0$ $(t \geqq t_1)$ for some $t_1 \geqq t_0$, and so we get

$$\int_{t_1}^{\infty} \left(\int_{s}^{\infty} r^{-\beta} dr \right)^{1/\alpha} ds < \infty.$$

But this is impossible under the assumption $\beta < 1 + \alpha$. Consequently, $x(t)$ does not satisfy (2.3), and hence we conclude that $\lim x(t) = \infty$ as $t \to \infty$.

Put $\gamma = \beta - 1$ $(< \alpha)$. Then we have

$$\begin{aligned} -\frac{d}{dt}(x'(t))^{\alpha-\gamma} &= -\frac{d}{dt}\left\{(x'(t))^\alpha\right\}^{(\alpha-\gamma)/\alpha} \\ &= \frac{\alpha-\gamma}{\alpha}(x'(t))^{-\gamma} p(t) f(x(t)) \end{aligned}$$

for $t \geqq t_x$. An integration of the above equality gives

$$-(x'(t))^{\alpha-\gamma} + (x'(t_x))^{\alpha-\gamma} = \frac{\alpha-\gamma}{\alpha} \int_{t_x}^{t} (x'(s))^{-\gamma} p(s) f(x(s)) ds.$$

Letting $t \to \infty$, we see that

$$\int_{t_x}^{\infty} (x'(s))^{-\gamma} p(s) f(x(s)) ds < \infty,$$

and hence

$$\int_{t_x}^{\infty} (x'(s))^{-\gamma} q(s) f(x(s)) ds < \infty. \tag{3.2}$$

Note that there are constants $a_1 > 0$ and $a_2 > 0$ such that (2.4) holds. Since $x'(t)$ is nonnegative and nonincreasing on $[t_x, \infty)$, we have

$$\begin{aligned} x(t) &= x(t_x) + \int_{t_x}^{t} x'(s) ds \\ &\geqq \int_{t/2}^{t} x'(s) ds \geqq \frac{1}{2} t x'(t) \end{aligned}$$

for $t \geqq 2t_x$. Since $x(t) \to \infty$ as $t \to \infty$, we can take T sufficiently large such that $x(t) \geqq a_2 t_0$ and $x(t) \geqq \frac{1}{2} t x'(t)$ for $t \geqq T$. Moreover, we may suppose without loss of generality that $\beta \geqq 0$. Then it follows from (2.4) and the nondecreasing property of $t^{\beta} q(t)$ that

$$\begin{aligned} q(t) &\geqq t^{-\beta} \left(\frac{x(t)}{a_2} \right)^{\beta} q\left(\frac{x(t)}{a_2} \right) \\ &\geqq \left(\frac{1}{2a_2} \right)^{\beta} (x'(t))^{\beta} q\left(\frac{x(t)}{a_2} \right) \end{aligned}$$

for $t \geqq T$. Therefore, by virtue of (3.2), we find that

$$\int_{T}^{\infty} x'(s) q\left(\frac{x(s)}{a_2} \right) f(x(s)) ds < \infty.$$

Then, letting $r = x(s)/a_2$ and noting that $s \to \infty$ as $r \to \infty$, we easily see that (3.1) with $c = a_2$ holds. The proof of Theorem 3.1 is complete.

REMARK 3.1. Suppose that $q(t)$ satisfies the hypothesis in Theorem 3.1. Then we easily see that if (3.1) holds, then

$$\int_{t_0}^{\infty} q(s) f(c's) ds < \infty \quad \text{for all } c' \text{ with } c' \geqq c,$$

where c is a positive number in (3.1).

COROLLARY 3.1. *Suppose that there exists a number* $\beta < 1+\alpha$ *such that* $\liminf_{t\to\infty} t^{\beta}p(t) > 0$. *If* (1.1) *has a positive solution* $x(t)$, *then*

$$\int_1^{\infty} x^{-\beta} f(x)dx < \infty. \tag{3.3}$$

PROOF. There are constants $k>0$ and $T \geqq t_0$ satisfying $p(t) \geqq kt^{-\beta}$ for $t \geqq T$. Apply Theorem 3.1 to the case $t_0 = T$ and $q(t) = kt^{-\beta}$ $(t \geqq T)$. We see that if (1.1) has a positive solution $x(t)$, then

$$\int_T^{\infty} s^{-\beta} f(cs)ds < \infty \quad \text{for some } c>0,$$

which is equivalent to (3.3).

It is to be noted that if $\limsup_{t\to\infty} t^{\beta}p(t) < \infty$ for some $\beta > 1+\alpha$, then equation (1.1) always has a positive solution $x(t)$ satisfying (2.3). This is easily derived from Theorem 2.2.

In Theorem 3.1 it is impossible to choose $\beta = 1+\alpha$. To see this, consider the equation

$$\left(|x'|^{\alpha-1}x'\right)' + \alpha(1-\lambda)\lambda^{\alpha}t^{-\alpha-1}x^{\alpha} = 0, \quad t \geqq 1, \tag{3.4}$$

where $\alpha > 0$ and $0 < \lambda < 1$. In this case we have $p(t) = \alpha(1-\lambda)\lambda^{\alpha}t^{-\alpha-1}$ and $f(x) = x^{\alpha}$. Take $q(t) \equiv p(t)$. Then, since $t^{1+\alpha}q(t)$ is a constant function, it is clearly nondecreasing. Equation (3.4) has a positive solution $x(t) = t^{\lambda}$ $(t \geqq 1)$. But Condition (3.1) is not satisfied. It is also impossible to choose $\beta = 1+\alpha$ in Corollary 3.1.

The next corollary gives a necessary and sufficient condition for (1.1) to have a positive solution.

COROLLARY 3.2. *Suppose that* $p(t)$ *is decomposable in such a way that* $p(t) = c(t)q(t)$, $c(t)$ *is continuous and positive on* $[t_0,\infty)$, $0 < \liminf_{t\to\infty} c(t) \leqq \limsup_{t\to\infty} c(t) < \infty$, *and* $t^{\beta}q(t)$ *is continuous and nondecreasing on* $[t_0,\infty)$ *for some* $\beta < 1+\alpha$.

Then equation (1.1) *has a positive solution* $x(t)$ *if and only if* (2.5) *holds.*

PROOF. The "if" part is contained in Theorem 2.1, and the "only if" part is easily derived from Theorem 3.1.

COROLLARY 3.3. *Suppose that* $p \in C^1[t_0,\infty)$, $p(t) > 0$ $(t \geqq t_0)$ *and that there exists a number* $\beta < 1+\alpha$ *such that*

$$\int_{t_0}^{\infty} \frac{(s^{\beta}p(s))'_-}{s^{\beta}p(s)}ds < \infty,$$

where $(t^\beta p(t))'_- = \max\{-(t^\beta p(t))', 0\}$.

Then equation (1.1) *has a positive solution* $x(t)$ *if and only if* (2.5) *holds.*

PROOF. We have only to apply Corollary 3.2 by taking

$$c(t) = \exp\left(-\int_{t_0}^{t} \frac{(s^\beta p(s))'_-}{s^\beta p(s)} ds\right)$$

and

$$q(t) = t_0^\beta p(t_0) t^{-\beta} \exp\left(\int_{t_0}^{t} \frac{(s^\beta p(s))'_+}{s^\beta p(s)} ds\right)$$

where $(t^\beta p(t))'_+ = \max\{(t^\beta p(t))', 0\}$.

As can be seen by equation (3.4), it is also impossible to choose $\beta = 1 + \alpha$ in Corollaries 3.2 and 3.3.

For the case $\alpha = 1$, the results of this paper were obtained by Naito [3].

REFERENCES

[1] BURTON, T. and GRIMMER, R., Oscillation, continuation, and uniqueness of solutions of retarded differential equations, *Trans. Amer. Math. Soc.* **179** (1973), 193–209. *MR* **48** #2523; see also *MR* **48** #6602

[2] ELBERT, Á. and KUSANO, T., Oscillation and non-oscillation theorems for a class of second order quasilinear differential equations, *Acta Math. Hungar.* **56** (1990), 325–336. *MR* **93b**:34039

[3] NAITO, M., Positive solutions of nonlinear differential inequalities, *Hiroshima Math. J.* **9** (1979), 769–785. *MR* **81a**:34011

[4] WONG, J. S. W., On second order nonlinear oscillation, *Funkcial. Ekvac.* **11** (1968), 207–234. *MR* **39** #7221

(Received December 1, 1995)

KUMAMOTO PREFECTURAL SEISEIKOU HIGH SCHOOL
KUMAMOTO 860–0862
JAPAN

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF SCIENCE
FUKUOKA UNIVERSITY
FUKUOKA 814–0180
JAPAN

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
EHIME UNIVERSITY
MATSUYAMA 790–8577
JAPAN

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 53–64

MATRIX TRANSFORMATIONS OF λ-BOUNDEDNESS FIELDS OF NORMAL MATRIX METHODS

A. AASMA[1]

Dedicated to Professor Károly Tandori on his 70th birthday

Abstract

In this paper we shall consider the boundedness and summability with speed. Let A be a normal matrix, B a triangular matrix, λ and μ monotonically increasing sequences (i.e. speeds). We shall prove a theorem that gives necessary and sufficient conditions for a matrix M to be transformation of the λ-boundedness field of A into the μ-boundedness field of B. For applications we shall consider the special case when A is a Riesz method and the λ-boundedness of Fourier expansions in Banach spaces by the method of Zygmund (Z, r).

Introduction

Let $\lambda = (\lambda_k)$ be a sequence with the property $0 < \lambda_k \uparrow$. A sequence $x = (x_k)$ is said to be *bounded with speed λ or λ-bounded* when the conditions

$$\lim_k x_k = \xi, \quad l_k = O(1)$$

are fulfilled, where

$$l_k = \lambda_k(x_k - \xi).$$

Let

$$m^\lambda = \{x = (x_k) \mid x \text{ is } \lambda\text{-bounded}\},$$

$$ms^\lambda = \left\{ x = (x_k) \,\middle|\, (X_n) \in m^\lambda, \text{ where } X_n = \sum_{k=0}^{n} x_k \right\}.$$

We note that the sequences $e = (1, 1, \dots, 1, \dots)$, $\lambda^{-1} = (\lambda_k^{-1})$ and $e^k = (0, \dots, 0, 1, 0, \dots)$ with 1 in kth position belong to m^λ. Also $m^\lambda \subset c$, where c is the space of convergent sequences.

1991 *Mathematics Subject Classification.* Primary 40D05; Secondary 41A25.

Key words and phrases. Matrix transformations, summability with speed, Fourier expansions, orders of approximation.

[1]The author was supported by ESF Grant No. 1685.

Let $A=(\alpha_{nk})$ be a matrix with $\alpha_{nk}\in\mathbf{C}$ $(n,k=0,1,\dots)$ and let x denote a sequence (x_k) or a series $\sum_k x_k$. Then x is said to be λ-*bounded by* A *or* A^λ-*bounded* if $Ax\in m^\lambda$, where $Ax=(A_nx)$ and

$$A_nx=\sum_k \alpha_{nk}x_k.$$

We denote the set of A^λ-bounded sequences (or series) by m_A^λ. Let $M=(m_{nk})$ with $m_{nk}\in\mathbf{C}$ $(n,k=0,1,\dots)$ be a matrix, $B=(\beta_{nk})$ with $\beta_{nk}\in\mathbf{C}$ $(n,k=0,1,\dots)$ a triangular matrix and $\mu=(\mu_k)$ a sequence with $0<\mu_k\uparrow$. We say that $M\in(m_A^\lambda,m_B^\mu)$ if the matrix transformation $y=Mx$ exists for each $x\in m_A^\lambda$ and $y\in m_B^\mu$. If Mx exists and

$$\lim_n B_n(Mx)=\lim_n A_nx$$

for each $x\in m_A^\lambda$, then we say that A and B are M-*consistent on* m_A^λ.

In this paper we shall find necessary and sufficient conditions for $M\in(m_A^\lambda,m_B^\mu)$ and for M-consistency of A and B *on* m_A^λ in the case when A is a normal matrix (i.e. $\alpha_{nn}\neq 0$ and $\alpha_{nk}=0$ if $k>n$) and B is a triangular one (Section 1). As an application we shall consider these conditions in the case $A=(R,p_n)$ (Section 2). Another application we get for Zygmund method $A=(Z,r)$. Namely, we shall consider the $(Z,r)^\lambda$-boundedness of Fourier expansions in Banach spaces with respect to a sequence of orthogonal projections (Section 3).

If λ and μ are bounded sequences, then $m_A^\lambda=c_A$ and $m_B^\mu=c_B$. The necessary and sufficient conditions for $M\in(c_A,c_B)$ were found in [1], and for special cases if A or A and B both are Cesàro methods, also in [2] and [9]. We note that in the case of diagonal matrix M this problem has been solved in [6, 7]. In particular, if $A=B=I$ (where I is identity method), the necessary and sufficient conditions for $M\in(m_A^\lambda,m_B^\mu)=(m^\lambda,m^\mu)$ can also be found in [6, 7].

1. Matrix transformations for the class (m_A^λ,m_B^μ)

A. During this paper we assume further that $\lambda=(\lambda_k)$ and $\mu=(\mu_k)$ are sequences with $\lambda_k,\mu_k\uparrow\infty$, A is a normal matrix with its inverse matrix $A^{-1}=(\eta_{kl})$ and B is a triangular matrix. First we notice that the matrix transformation $y=Mx$ exists for each $x\in m_A^\lambda$ if and only if the numbers m_{nk} are the convergence factors for m_A^λ for each $n=0,1,\dots$. Therefore, by the theorem 20.2 of [7], we have

LEMMA 1. *Let $M=(m_{nk})$ be an arbitrary matrix. Then the matrix transformation $y=Mx$ exists for each $x\in m_A^\lambda$ if and only if*

(i) *there exist finite limits* $\lim_r M_{nl}^r = M_{nl}$,

(ii) *there exist finite limits* $\lim_r \sum_{l=0}^r M_{nl}^r$,

(iii) $$\sum_l \frac{|M_{nl}|}{\lambda_l} = O_n(1),$$

(iv) $$\lim_r \sum_{l=0}^r \frac{|M_{nl}^r - M_{nl}|}{\lambda_l} = 0,$$

where

$$M_{nl}^r = \begin{cases} \sum_{k=l}^r m_{nk}\eta_{kl} & \text{if } l \leqq r, \\ 0 & \text{if } l > r. \end{cases}$$

Let $G=(g_{sk})=BM$ and

$$\gamma_{sl}^r = \begin{cases} \sum_{k=l}^r g_{sk}\eta_{kl} & \text{if } l \leqq r, \\ 0 & \text{if } l > r. \end{cases}$$

Now we can prove the main result of this paper.

THEOREM 1. *Let M be an arbitrary matrix. Then $M\in(m_A^\lambda, m_B^\mu)$ if and only if the conditions* (i)–(iv) *and the following conditions hold:*

(v) $(\varrho_s)\in m^\mu$,

(vi) *there exist the finite limits* $\lim_s \gamma_{sl} = \gamma_l$,

(vii) $$\sum_l \frac{|\gamma_l|}{\lambda_l} < \infty,$$

(viii) $$\mu_s \sum_l \frac{|\gamma_{sl} - \gamma_l|}{\lambda_l} = O(1),$$

where

$$\varrho_s = \lim_r \sum_{l=0}^r \gamma_{sl}^r, \quad \gamma_{sl} = \lim_r \gamma_{sl}^r.$$

PROOF. *Necessity.* Let $M\in(m_A^\lambda, m_B^\mu)$. Then the transformation $y=Mx$ exists for each $x\in m_A^\lambda$ and therefore the conditions (i)–(iv) are fulfilled. As B is triangular, we have

(1) $$\sum_{n=0}^s \beta_{sn} M_n x = G_s x$$

for each $x\in m_A^\lambda$. Now it follows from (1) that $G\in(m_A^\lambda, m^\mu)$. Moreover

(2) $$\sum_{k=0}^r g_{sk}x_k = \sum_{l=0}^r \gamma_{sl}^r t_l,$$

where $t_l = A_l x$, for each $x \in m_A^\lambda$. As the method A is normal there exists $x \in m_A^\lambda$ so that $(t_l) = (A_l x) = e$. Therefore condition (v) is fulfilled by (2).

Let, further

$$\lim_l t_l = \nu, \quad \beta_l = \lambda_l(t_l - \nu).$$

Then, by (2), the equalities

$$\sum_{k=0}^{r} g_{sk} x_k = \nu \sum_{l=0}^{r} \gamma_{sl}^r + \sum_{l=0}^{r} \frac{\gamma_{sl}^r}{\lambda_l} \beta_l \tag{3}$$

hold for each $x \in m_A^\lambda$. Now, the series $G_s x$ converge for each $x \in m_A^\lambda$ and the finite limits ϱ_s exist by condition (v). Hence we see from (3) that the matrix $(\gamma_{sl}^r/\lambda_l)$ transforms each bounded sequence (β_l) into convergent sequence for every s. Therefore (3) implies with the help of Theorem 2.1 of [3]

$$G_s x = \nu \varrho_s + \sum_l \frac{\gamma_{sl}}{\lambda_l} \beta_l. \tag{4}$$

Also the finite limit $\lim_s \varrho_s = \varrho$ exists by condition (v). Consequently, from (4) it follows that the matrix (γ_{sl}/λ_l) transforms each bounded sequence into convergent sequence. Thus conditions (vi) and (vii) are fulfilled,

$$\lim_s \sum_l \frac{|\gamma_{sl} - \gamma_l|}{\lambda_l} = 0, \tag{5}$$

and the equality

$$\lim_s G_s x = \nu \varrho + \sum_l \frac{\gamma_l}{\lambda_l} \beta_l \tag{6}$$

holds for each $x \in m_A^\lambda$ by Theorem 2.1 of [3]. Hence we have

$$\mu_s(G_s x - \lim_s G_s x) = \nu \mu_s(\varrho_s - \varrho) + \mu_s \sum_l \frac{\gamma_{sl} - \gamma_l}{\lambda_l} \beta_l \tag{7}$$

for each $x \in m_A^\lambda$. From (7) we see that the matrix $(\mu_s(\gamma_{sl} - \gamma_l)/\lambda_l)$ transforms each bounded sequence (β_l) into bounded sequence. Consequently, condition (viii) holds by Theorem 2.2 of [3].

Sufficiency. Let conditions (i)–(viii) be fulfilled. Then for each $x \in m_A^\lambda$ the transformation $y = Mx$ exists by Lemma 1 and equalities (1)–(3), where $t_l = A_l x$, hold. As

$$\lim_r \sum_{k=0}^{r} g_{sk} x_k = G_s x$$

for each $x \in m_A^\lambda$, then it follows from (3) that the equalities (4) hold for each $x \in m_A^\lambda$ by condition (v) and Theorem 2.1 of [3]. Moreover, condition (5) follows from (viii), because $\mu_s \neq O(1)$. Hence equality (6) holds for each $x \in m_A^\lambda$ by Theorem 2.1 of [3] and conditions (vi)–(viii). Thus equalities (7) are also valid for each $x \in m_A^\lambda$. Consequently, $M \in (m_A^\lambda, m_B^\mu)$ by (v) and Theorem 2.2 of [3].

REMARK 1. For a triangular matrix M conditions (i)–(iv) are redundant in Theorem 1.

We notice that in the special case if $\lambda_k = O(1)$ and $\mu_k = O(1)$, Theorem 1 of [1] for $M \in (c_A, c_B)$ immediately follows from Theorem 1. For a diagonal matrix M Theorem 1 reduces to Theorem 20.2 of [7] and for $A = B = I$ to Theorem 1 of [6].

If the method A preserves λ-boundedness, i.e. $m^\lambda \subset m_A^\lambda$, then from Theorem 1 we have

COROLLARY 1. *If* $m^\lambda \subset m_A^\lambda$ *and*

$$\gamma_{sl} = O(g_{sl}), \tag{8}$$

then in conditions (vii) *and* (viii) *of Theorem 1* $\Gamma = (\gamma_{sl})$ *can be replaced by* $G = (g_{sl})$.

PROOF. Let $M \in (m_M^\lambda, m_B^\mu)$. Now the relation $G \in (m_A^\lambda, m^\mu)$ follows from (2). As $m^\lambda \subset m_A^\lambda$, then $G \in (m^\lambda, m^\mu)$. Therefore the conditions

$$\sum_l \frac{|g_l|}{\lambda_l} < \infty, \tag{9}$$

$$\mu_s \sum_l \frac{|g_{sl} - g_l|}{\lambda_l} = O(1), \tag{10}$$

where

$$g_l = \lim_s g_{sl},$$

are fulfilled by Theorem 1 of [6]. This completes the proof because it follows from (8)–(10) that conditions (vii) and (viii) are fulfilled.

It is easy to see that in the case of normal method B condition (iii) follows from condition (vii). Therefore from Theorem 1 we get

COROLLARY 2. *If* B *is a normal method, then condition* (iii) *is redundant in Theorem 1.*

Let now a normal method $A = (\alpha_{nk})$ have the property $\alpha_{n0} = 1$. Then for its inverse matrix $A^{-1} = (\eta_{nk})$ the equalities

$$\sum_{i=0}^{k} \eta_{kl} = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k \geqq 1 \end{cases}$$

are valid by Theorem 9.2 of [3]. Therefore

$$\sum_{l=0}^{r} M_{nl}^{r} = m_{n0}, \qquad \sum_{l=0}^{r} \gamma_{sl}^{r} = g_{s0}.$$

Consequently, from Theorem 1, we have

COROLLARY 3. *If $A = (\alpha_{nk})$ has the property $\alpha_{n0} = 1$, then in Theorem 1 condition* (ii) *is redundant and condition* (v) *can be replaced by condition*

(ix) $\qquad e^0 \in m_G^{\mu}$.

COROLLARY 4. *Let M be a number matrix. Then A and B are M-consistent on m_A^{λ} if and only if conditions* (i)–(iv) *are fulfilled and*

(x) $\qquad \lim_s \varrho_s = 1,$

(xi) $\qquad \lim_s \sum_l \frac{|\gamma_{sl}|}{\lambda_l} = 0.$

PROOF. *Necessity.* Let A and B be M-consistent on m_A^{λ}. Then $M \in (m_A^{\lambda}, c_B)$. Therefore the conditions (i)–(iv) are fulfilled, equalities (4) hold for each $x \in m_A^{\lambda}$ and there exist the finite limits ϱ_s and the finite limit $\lim_s \varrho_s = \varrho$ (see the necessity part of the proof of Theorem 1). Let us show that $\varrho = 1$. Indeed, there exists $x \in m_A^{\lambda}$ so that $(t_l) = (A_l x) = e$ by the normality of A. As $\lim_s G_s x = \lim_l A_l x = 1$, then $\varrho = 1$ by (2), i.e. condition (x) is fulfilled. It follows from (4) that the matrix (γ_{sl}/λ_l) transforms the space of bounded sequences m into the space of 0-convergent sequences c_0. Therefore condition (xi) is fulfilled by Proposition 21 of [8].

Sufficiency. Let conditions (i)–(iv) and (x)–(xi) be fulfilled. Then the series ϱ_s are convergent and equalities (4) hold for each $x \in m_A^{\lambda}$ (see the proof of Theorem 1). As the matrix (γ_{sl}/λ_l) transforms m into c_0 by (xi), then A and B are M-consistent on m_A^{λ} by (4) and (x).

B. Let now X be a Banach space with norm $\| * \|$,

$$m^{\lambda}(X) = \{x = (x_k) \mid x_k \in X,\ \exists \lim x_k = \xi,\ \lambda_k \|x_k - \xi\| = O(1)\},$$

$$ms^{\lambda}(X) = \left\{x = (x_k) \,\middle|\, x_k \in X,\ (X_n) \in m^{\lambda}(X),\ \text{where } X_n = \sum_{k=0}^{n} x_k\right\},$$

$$m_A^{\lambda}(X) = \left\{x = (x_k) \mid x_k \in X,\ \exists \lim_n A_n x = \xi,\ \lambda_n \|A_n x - \xi\| = O(1)\right\}.$$

It is easy to see that $\xi e \in m^{\lambda}(X)$ for each $\xi \in X$ and the relations (3), (4), (6) hold for X-valued sequences (or series). Moreover, Theorems 2.1 and 2.2 of [3], Proposition 21 of [8] and Lemma 1 are also valid for X-valued convergent and bounded sequences (cf. [5], p. 115 and Remark 1 of [6]). Therefore we have

REMARK 2. All results of this paper are valid if m_A^λ and m_B^μ are replaced by $m_A^\lambda(X)$ and $m_B^\mu(X)$ in them.

2. Matrix transformations for the class (m_P^λ, m_B^μ)

Now we shall consider the case when A is a Riesz method. Let (p_n) be a sequence of nonzero complex numbers, $P_n = p_0 + \cdots + p_n \neq 0$, $P_{-1} = 0$ and let $P = (R, p_n) = (\alpha_{nk})$ be the series-to-sequence Riesz method generated by (p_n), i.e.,

$$\alpha_{nk} = \begin{cases} 1 - P_{k-1}/P_n & \text{if } k \leqq n, \\ 0 & \text{if } k > n. \end{cases}$$

It is well known that P is a normal method. Therefore P has the inverse matrix $P^{-1} = (\eta_{nk})$, where

$$\eta_{nk} = \begin{cases} P_k/p_k & \text{if } n = k, \\ -P_k(1/p_k + 1/p_{k+1}) & \text{if } n = k+1, \\ P_k/p_{k+1} & \text{if } n = k+2, \\ 0 & \text{if } n < k \text{ or } n > k+2 \end{cases} \tag{11}$$

(cf. [3], p. 116).

THEOREM 2. *Let $M = (m_{nk})$ be a matrix, P a Riesz method with properties $ms^\lambda \subset m_P^\lambda$,*

$$P_n = O(P_{n-1}), \tag{12}$$

$$\frac{P_n}{p_n} = O\left(\frac{P_{n+1}}{p_{n+1}}\right). \tag{13}$$

Then $M \in (m_P^\lambda, m_B^\mu)$ if and only if the condition (ix) *and the following conditions hold:*

(xii) $$\sum_l \frac{1}{\lambda_l} \left| P_l \Delta \frac{\Delta m_{nl}}{p_l} \right| = O_n(1),$$

(xiii) $$\lim_l \frac{P_l m_{nl}}{p_l \lambda_l} = 0,$$

(xiv) *there exist the finite limits* $\lim_s g_{sl} = g_l$,

(xv) $$\sum_l \frac{1}{\lambda_l} \left| P_l \Delta \frac{\Delta g_l}{p_l} \right| < \infty,$$

(xvi) $$\mu_s \sum_l \frac{1}{\lambda_l} \left| P_l \Delta \frac{\Delta(g_{sl} - g_l)}{p_l} \right| = O(1),$$

where

$$\Delta m_{sl} = \Delta_l m_{sl} = m_{sl} - m_{s,l+1}.$$

PROOF. *Necessity.* Let $M \in (m_P^\lambda, m_B^\mu)$. We shall show that conditions (ix) and (xii)–(xvi) are fulfilled. First we see by (11) that

$$M_{nl} = P_l \Delta \frac{\Delta m_{nl}}{p_l}, \tag{14}$$

$$M_{nl}^r = \begin{cases} M_{nl} & \text{if } l < r-1, \\ M_{n,r-1} - P_{r-1} m_{n,r+1}/p_r & \text{if } l = r-1, \\ P_r m_{nr}/p_r & \text{if } l = r, \\ 0 & \text{if } l > r. \end{cases} \tag{15}$$

Therefore condition (xii) is fulfilled by Theorem 1.

It is easy to see that

$$\sum_l \frac{|M_{nl}^r - M_{nl}|}{\lambda_l} = \left| \frac{P_{r-1} m_{n,r+1}}{p_r \lambda_{r-1}} \right| + \left| \frac{P_r m_{nr}}{p_r \lambda_r} - \frac{P_r}{\lambda_r} \Delta \frac{\Delta m_{nr}}{p_r} \right|. \tag{16}$$

Moreover, it follows from condition (xii) that

$$\lim_r \frac{P_r}{\lambda_r} \Delta \frac{\Delta m_{nr}}{p_r} = 0.$$

Hence condition (xiii) is fulfilled by Theorem 1.

As the sequence $e^k \in ms^\lambda$ and $ms^\lambda \subset m_P^\lambda$ the condition (xiv) is fulfilled. Condition (ix) is satisfied by Corollary 3, since $\alpha_{n0} = 1$ for the method P. Further, we have by (11) that

$$\gamma_{sk} = P_k \Delta \frac{\Delta g_{sk}}{p_k}.$$

Consequently, conditions (xv) and (xvi) are fulfilled by Theorem 1.

Sufficiency. Let conditions (ix) and (xii)–(xvi) be fulfilled. We shall show that $M \in (m_P^\lambda, m_B^\mu)$. At first we see that conditions (i) and (iii) are fulfilled by (14), (15) and (xii). Also condition (iv) is fulfilled. Indeed, the condition

$$\frac{\lambda_n P_\nu}{\lambda_\nu P_n} = O(1) \quad (\nu \leqq n)$$

is satisfied by Lemma 1 of [6]. Hence

$$\frac{P_{r-1}}{\lambda_{r-1}} = O\left(\frac{P_{r+1}}{\lambda_{r+1}}\right).$$

Therefore we have

$$\frac{P_{r-1}m_{n,r+1}}{p_r\lambda_{r-1}} = O(1)\frac{P_{r+1}m_{n,r+1}}{\lambda_{r+1}p_r} = O(1)\frac{P_r m_{n,r+1}}{p_r\lambda_{r+1}}$$
$$= O(1)\frac{P_{r+1}m_{n,r+1}}{p_{r+1}\lambda_{r+1}} = o_n(1)$$

by (12), (13) and (xiii). Consequently, condition (iv) is fulfilled by (16), (xii) and (xiii).

From conditions (xiv)–(xvi) it follows that conditions (vi)–(viii) are fulfilled. Therefore $M \in (m_P^\lambda, m_B^\mu)$ by Theorem 1 and Corollary 3.

From Corollaries 3–4 and Theorem 2 immediately follows

COROLLARY 5. *Let M be a matrix, P be a Riesz method with properties* (12), (13) *and $ms^\lambda \subset m_P^\lambda$. Then P and B are M-consistent on m_P^λ if and only if conditions* (xii) *and* (xiii) *are fulfilled and*

(xvii) $$\lim_s g_{s0} = 1,$$

(xviii) $$\lim_s \sum_l \left| \frac{P_l}{\lambda l} \Delta \frac{\Delta g_{sl}}{p_l} \right| = 0.$$

In the next section we shall find that there exist special matrices P, M and B satisfying conditions of Theorem 2 and Corollary 5.

3. $(Z,r)^\lambda$-boundedness of Fourier expansions in Banach spaces

Here we shall apply the results of Section 2 to the summability of Fourier expansions in Banach spaces by Zygmund methods (Z,r). Namely, we shall characterize the relation between the speeds of approximation of Fourier expansions by different two methods of Zygmund.

Let X be a Banach space with norm $\|*\|$ and $\{T_k\}$ be a total sequence of mutually orthogonal continuous projections on X, i.e., T_k is a bounded linear operator of X into itself, $T_k f = 0$ for all k implies $f = 0$ and $T_j T_k = \delta_{jk} T_k$, where δ_{jk} is Kronecker's symbol. Then one may associate with each $f \in X$ its formal Fourier series expansion

$$f \sim \sum_k T_k f.$$

As we know, in the special case of $p_n = (n+1)^r - n^r$ $(r > 0)$ Riesz method P is called Zygmund method and denoted by (Z,r). Thus $(Z,r) = (\alpha_{nk})$, where

$$\alpha_{nk} = \begin{cases} 1 - (k/n+1)^r & \text{if } k \leqq n \\ 0 & \text{if } k > n. \end{cases}$$

Let, further, for every $f \in X$

$$Z_{rn}f = (Z,r)_n(T_kf) = \sum_{k=0}^{n}\left[1-\left(\frac{k}{n+1}\right)^r\right]T_kf.$$

We know that for the trigonometric system $\{T_k\}$ and for $0<\alpha<1$ the relation

$$(n+1)^\alpha\|Z_{1,n}f - f\| = O(1)$$

holds if and only if

$$f \in \operatorname{Lip}\alpha = \{f \in X \mid \|f(x+h)-f(x)\| = O_f(h^\alpha)\}$$

(cf. [4], p. 106). Now we shall show that the next result is valid.

COROLLARY 6. *Let X be a Banach space and $f \in X$. If $0<\beta\leqq\alpha$, $r>\alpha$, $t>\alpha$ and*

$$(n+1)^\alpha\|Z_{rn}f - f\| = O(1),$$

then

$$(n+1)^\beta\|Z_{tn}f - f\| = O(1). \tag{17}$$

PROOF. Let $\lambda_k = (k+1)^\alpha$, $\mu_k = (k+1)^\beta$ and $\lambda = (\lambda_k)$, $\mu = (\mu_k)$. Then $\lambda_k \neq O(1)$ and $\mu_k \neq O(1)$. Consequently, it is sufficient to show, by Remark 2, that the conditions of Theorem 2 and Corollary 5 are fulfilled for $P = (Z,r)$, $M = (Z,t)$ and $B = (\delta_{nk})$. First we notice that conditions (12), (13), (xii) and (xiii) are satisfied. As

$$\sum_{k=0}^{n}\Delta\alpha_{nk} = \sum_{k=0}^{n}\frac{(k+1)^r - k^r}{(n+1)^r} = 1,$$

then the inclusion $ms^\lambda \subset m^\lambda_{(Z,r)}$ is equivalent to the condition

$$S_n = (n+1)^\alpha\sum_{k=0}^{n}\frac{|\Delta\alpha_{nk}|}{(k+1)^\alpha} = O(1)$$

(cf. [6], p. 139). The last condition is fulfilled since

$$\begin{aligned} S_n &= (n+1)^{\alpha-r}\sum_{k=0}^{n}\frac{(k+1)^r - k^r}{(k+1)^\alpha} \\ &= O(1)(n+1)^{\alpha-r}\sum_{k=0}^{n}(k+1)^{r-1-\alpha} = O(1) \end{aligned}$$

by the mean-value theorem of Lagrange.

As now $g_{nk} = m_{nk}$, the conditions (xiv) and (xvii) are fulfilled, $g_l = 1$ and

$$L = \mu_s \sum_l \left| \frac{P_l}{\lambda_l} \Delta \frac{\Delta(g_{sl} - g_l)}{p_l} \right| = \mu_s \sum_{l=0}^{s} \left| \frac{P_l}{\lambda_l} \Delta \frac{\Delta g_{sl}}{p_l} \right| = L_1 + L_2 + L_3,$$

where

$$L_1 = \mu_s \sum_{l=0}^{s-2} \left| \frac{P_l}{\lambda_l} \Delta \frac{\Delta g_{sl}}{p_l} \right| = (s+1)^{\beta-t} \sum_{l=0}^{s-2} (l+1)^{r-\alpha} \Delta \frac{(l+1)^t - l^t}{(l+1)^r - l^r},$$

$$L_2 = \mu_s \frac{P_{s-1}}{\lambda_{s-1}} \Delta \frac{\Delta g_{sl}}{p_l}\bigg|_{l=s-1} = (s+1)^{\beta-t} s^{r-\alpha} \Delta \frac{s^t - (s-1)^t}{s^r - (s-1)^r},$$

$$L_3 = \mu_s \frac{P_s g_{ss}}{\lambda_s p_s} = (s+1)^{\beta+r-\alpha-t} \frac{(s+1)^r - s^t}{(s+1)^5 - s^r}.$$

Hence we have

$$\begin{aligned}
L_1 &= \frac{t}{r}(s+1)^{\beta-t} \sum_{l=0}^{s-2} (l+1)^{r-\alpha} \Delta (l+\theta_l^1)^{t-r} \\
&= \frac{t(t-r)}{r}(s+1)^{\beta-t} \sum_{l=0}^{s-2} (l+1)^{r-\alpha} (l+\theta_l^1+\theta_l^2)^{t-r-1} \\
&= O(1)(s+1)^{\beta-t} \sum_{l=0}^{s-2} (l+1)^{-\alpha+t-1} = O(1)(s+1)^{\beta-\alpha} = O(1), \\
L_2 &= \frac{t}{r}(s+1)^{\beta-t} s^{r-\alpha} \Delta (s-1+\theta_1)^{t-r} \\
&= \frac{t(t-r)}{r}(s+1)^{\beta-t} s^{r-\alpha} (s-1+\theta_1+\theta_2)^{t-r-1} \\
&= O(1)(s+1)^{\beta-t+r-\alpha+t-r-1} = O(1), \\
L_3 &= \frac{t}{r}(s+1)^{\beta+r-\alpha-t}(s+\theta_3)^{t-r} = O(1)(s+1)^{\beta-\alpha} = O(1)
\end{aligned}$$

$(0 < \theta_l^1, \theta_l^2, \theta_1, \theta_2, \theta_3 < 1)$ by the mean-value theorem of Cauchy. Therefore $L = O(1)$, i.e., condition (xvi) is fulfilled. Now it follows from (xvi) that also the condition (xviii) is satisfied. Thus the relation (17) holds by Theorem 2, Corollary 5 and Remark 2.

REFERENCES

[1] Aasma, A., Matrix transformations of summability fields, *Tartu Riikl. Ül. Toimetised* **770** (1987), 38–50 (in Russian). *MR* **89d**:40005

[2] Alpár, L., On the linear transformations of series summable in the sense of Cesàro, *Acta Math. Acad. Sci. Hungar.* **39** (1982), 233–243. *MR* **83f**:40002

[3] Baron, S., *Vvedenie v teoriyu summiruemosti ryadov* [Introduction to the theory of summability of series], Second edition, corrected and supplemented, Kirjastus „Valgus", Tallinn, 1977 (in Russian). *MR* **81j**:40007. See also *MR* **35** #4631

[4] Butzer, P. L. and Nessel, R. I., *Fourier analysis and approximation. I. One-dimensional theory*, Pure and Applied Mathematics, Vol. 40, Academic Press, New York – London, 1971. *MR* **58** # 23312

[5] Kangro, G., On matrix transformations of sequences in Banach spaces, *Izv. Akad. Nauk Eston. SSR Ser. Tehn. Fiz. Mat. Nauk* **2** (1956), 108–128 (in Russian). *MR* **20** #4121

[6] Kangro, G., Summability factors for series that are λ-bounded by the Riesz and Cesàro methods, *Tartu Riikl. Ül. Toimetised* **277** (1971), 136–154 (in Russian). *MR* **53** #3539

[7] Leiger, T., *Methods of functional analysis in the theory of summability*, Tartu Univ., Tartu, 1992.

[8] Stieglitz, M. and Tietz, H., Matrixtransformationen von Folgenräumen. Eine Ergebnisübersicht, *Math. Z.* **154** (1977), 1–16. *MR* **56** #5109

[9] Thorpe, B., Matrix transformations of Cesàro summable series, *Acta Math. Hungar.* **48** (1986), 255–265. *MR* **87m**:40014

(Received December 4, 1995)

TALLINNA PEDAGOOGIKAÜLIKOOL
MATEMAATIKA-INFORMAATIKAOSAKOND
NARVA MNT. 25
EE–10120 TALLINN
ESTONIA

antsa@tpu.ee

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 65–80

ON THE USE OF DIVIDED DIFFERENCES IN THE INVESTIGATION OF INTERPOLATORY POSITIVE LINEAR OPERATORS

D. D. STANCU

Dedicated to Professor E. W. Cheney on the occasion of his 70th birthday

1. Introduction

The divided differences (dd) have a great importance in the approximation theory of functions, since they constitute a basic mathematical tool for the representation of interpolatory type positive linear operators (iplo), for the investigation of monotonicity and convexity properties of sequences of such operators and for the evaluation of remainders in the corresponding approximation formulae. In particular, we mention that from the formulae discussed in this paper we can see that there is a closed connection between the shape of the approximants obtained by means of (iplo) and the second-order (dd).

We consider an (iplo) $L_m : C[a,b] \to C[a,b]$ defined by means of a formula of the following form

$$(L_m f)(x) = \sum_{k=0}^{m} q_{m,k}(x) f(x_{m,k}), \tag{1.1}$$

where the nodes $x_{m,k}$ are distinct points of the interval $I = [a,b]$, while $q_{m,k}$ are non-negative polynomials of degree m, for any $k \in \{0,1,\ldots,m\}$. We call L_m interpolatory because the values of $L_m f$ are expressed by means of a functional information consisting in knowing the values of the function f at the nodes $x_{m,k}$.

It is known that such operators can be constructed by using different methods: interpolatory, combinatorial, probabilistic or by using expansions of entire functions in series of polynomials.

In our paper [18] we have shown how the Bernstein-type polynomial

$$(B_m^{\langle\beta,\gamma\rangle} f)(x) = \sum_{k=0}^{m} p_{m,k}(x) f\left(\frac{k+\beta}{m+\gamma}\right), \tag{1.2}$$

1991 *Mathematics Subject Classification.* Primary 41A05, 41A10, 41A36, 41A80; Secondary 65D05.

Key words and phrases. Interpolation, divided differences, positive linear operators, representation of remainders.

0081–6906/99/$ 5.00

where $0 \leqq \beta \leqq \gamma$ and

$$p_{m,k}(x) = \binom{m}{k} x^k (1-x)^{m-k}, \tag{1.3}$$

can be constructed by an iterative two-points linear interpolation procedure.

By using the Gregory–Newton interpolation formula we have obtained [17], in particular, a Bernstein-type operator (depending on some parameters), defined by the following formula

$$(B_{m,p}^{\langle \alpha,\beta,\gamma \rangle} f)(x) = \sum_{k=0}^{m+p} \binom{m+p}{k} \frac{x^{[k,-\alpha]}(1-x)^{[m+p-k,-\alpha]}}{1^{[m+p,-\alpha]}} f\left(\frac{k+\beta}{m+\gamma}\right), \tag{1.4}$$

where $p \in \mathbb{N}_0$, $m \in \mathbb{N}$, $\alpha \geqq 0$, $0 \leqq \beta \leqq \gamma$ and f is a real-valued function defined on the interval $\left[0, 1 + \frac{p}{m}\right]$.

We denote by $y^{[n,h]}$ the factorial power of order n ($n \geqq 0$) and increment h of y, that is

$$y^{[n,h]} = y(y-h)\dots(y-(n-1)h), \quad y^{[0,h]} = 1.$$

This approximating operator was further investigated by H. H. Gonska and J. Meier [5], who called it "the Bernstein–Stancu operator".

If all the parameters are zero then one obtains the classical Bernstein operator, while if only p and α are zero, then one arrives at the operator given in (1.2)–(1.3), introduced first in our paper [16], which was included, under the name "Bernstein–Stancu operator", in the very important new book by F. Altomare and C. Campiti [1].

2. Representations by (dd)

In 1969 we published a memoir [15] dedicated to the use of probabilistic methods in the theory of uniform approximation of continuous functions. In that paper we gave a probabilistic method for obtaining representations for (iplo) in terms of finite differences of the function involved. Now we want to give an extension of this method.

We start from the Newton interpolation polynomial corresponding to the function f and the nodes considered at (1.2):

$$(N_m f)(t) = (N_m f)(t; x_{m,0}, x_{m,1}, \dots, x_{m,m}) =$$
$$= \sum_{j=0}^{n} u_{m,j}(t)[x_{m,0}x_{m,1}, \dots, x_{m,j}; f],$$

where

$$u_{m,0}(t)=1, \quad u_{m,j}(t)=(t-x_{m,0})(t-x_{m,1})\dots(t-x_{m,j-1}).$$

Here the brackets represent the symbol for (dd).

It is known that this polynomial satisfies the following interpolation properties

$$(N_m f)(x_{m,k})=f(x_{m,k}) \qquad (k=0(1)m). \tag{2.1}$$

Let Y_m be a real random variable having on the interval I the jump points $x_{m,k}$ and the corresponding jumps $p_{m,k}(x)$ $(k=0(1)m)$, where $x\in I$. Considering the compound random variable $Z_m=(N_m f)(Y_m)$ and calculating its expected value, we obtain

$$\begin{aligned} E(Z_m)=E(N_m f)(Y_m)&=\sum_{k=0}^{m}(N_m f)(x_{m,k})p_{m,k}(x)= \\ &=\sum_{j=0}^{m}\gamma_{m,j}(x)[x_{m,0},x_{m,1},\dots,x_{m,j};f], \end{aligned} \tag{2.2}$$

where

$$\gamma_{m,j}(x)=\sum_{k=0}^{m}p_{m,k}(x)u_{m,j}(x_{m,k}). \tag{2.3}$$

If we take the relations (2.1) into account, then we can write

$$\begin{aligned} (L_m f)(x)&=\sum_{k=0}^{m}p_{m,k}(x)f(x_{m,k})= \\ &=\sum_{j=0}^{m}\gamma_{m,j}(x)[x_{m,0},x_{m,1},\dots,x_{m,j};f]. \end{aligned} \tag{2.4}$$

Here we have a representation by (dd), on the nodes $x_{m,k}$, of the (iplo) L_m, applied to the function $f: I\to\mathbb{R}$.

Now let us consider the special case of the equally spaced nodes: $x_{m,k}=\alpha+kh_m$ $(k=0(1)m)$ $(h_m>0)$, where $a\leqq\alpha$, $x_m=a+mh_m\leqq b$. In this case we can write

$$[x_{m,0},x_{m,1},\dots,x_{m,j};f]=\frac{1}{j!h_m^j}(\Delta_{h_m}^j f)(\alpha)$$

and

$$u_{m,j}(t)=(t-\alpha)(t-\alpha-h_m)\dots(t-\alpha-(j-1)h_m)=(t-\alpha)^{[j,h_m]}.$$

Consequently, we have

$$\gamma_{m,j}(x) = \sum_{k=0}^{m} p_{m,k}(x)(x_{m,k} - \alpha)^{[j,h_m]}.$$

On the account of the equality

$$(x_{m,k} - \alpha)^{[j,h_m]} = (kh_m)^{[j,h_m]} = h_m^j k^{[j]},$$

where

$$k^{[j]} = k(k-1)\dots(k-j+1),$$

we find

$$\gamma_{m,j}(x) = h_m^j \mu_{m,[j]}(x),$$

with

$$\mu_{m,[j]}(x) = \sum_{k=0}^{m} k^{[j]} p_{m,k}(x). \tag{2.5}$$

Thus, in the case of equally spaced nodes we have

$$(L_m f)(x) = \sum_{j=0}^{m} \frac{1}{j!} \mu_{m,[j]}(x)(\Delta_{h_m}^j f)(\alpha). \tag{2.6}$$

For instance, assume that $a = 0$ and $b = 1$, while $\alpha = 0$ and $h_m = \frac{1}{m}$. In this case we have $x_{m,k} = \frac{k}{m}$ and formula (2.6) becomes

$$(L_m f)(x) = \sum_{j=0}^{m} \frac{1}{j!} \mu_{m,[j]}(x)(\Delta_{\frac{1}{m}}^j f)(0). \tag{2.7}$$

At (2.5) we have the factorial moment of order j of the random variable Y_m for which $P\left(Y_m = \frac{k}{m}\right) = p_{m,k}(x)$.

In order to calculate these moments it is helpful to use the corresponding factorial moment-generatic function, which is defined by $g_m(t) = E(t^{Y_m})$.

It is easy to see that we have

$$\mu_{m,[j]}(x) = g_m^{(j)}(1). \tag{2.8}$$

Formula (2.7) has been established first in our paper [15].

For illustration, we consider that Y_m is a discrete random variable having the jump points

$$x_{m,k} = \frac{k+\beta}{m+\gamma} \qquad (0 \leqq \beta \leqq \gamma;\ k = 0(1)m)$$

and jumps

$$p_{m,k}^{\langle\alpha\rangle}(x) = \binom{m}{k} \frac{x^{[k,-\alpha]}(1-x)^{[m-k,-\alpha]}}{1^{[m,-\alpha]}}, \tag{2.9}$$

where $0 \leqq x \leqq 1$ and α is a parameter which may depend on m, fulfilling the stipulation that

$$1^{[m,-\alpha]} = (1+\alpha)(1+2\alpha)\ldots(1+(m-1)\alpha) \neq 0,$$

then we arrive at the operator $S_m^{\langle\alpha,\beta,\gamma\rangle}$ defined by

$$(S_m^{\langle\alpha,\beta,\gamma\rangle} f)(x) = \sum_{k=0}^{m} p_{m,k}^{\langle\alpha\rangle}(x) f\left(\frac{k+\beta}{m+\gamma}\right), \tag{2.10}$$

which can be obtained from (1.4) for $p = 0$.

In this case the factorial moment-generatic function is

$$g_m^{\langle\alpha\rangle}(t) = \sum_{k=0}^{m} t^k p_{m,k}^{\langle\alpha\rangle}(x),$$

which corresponds to the Markov–Pólya probability distribution.

In the case $\alpha = 0$ it reduces to the moment-generatic function for the Bernoulli distribution

$$g_m(t) = (1-x+tx)^m$$

and we have

$$\mu_{m,[j]}(x) = g_m^{(j)}(1) = m^{[j]} x^{[j]}.$$

It follows that the Bernstein-type polynomial defined at (1.2)–(1.3) have the following representation in terms of finite differences:

$$\left(B_m^{\langle\beta,\gamma\rangle} f\right)(x) = \sum_{j=0}^{m} \binom{m}{j} \left(\Delta^j_{\frac{1}{m+\gamma}} f\right) \left(\frac{\beta}{m+\gamma}\right) x^j.$$

Assuming that $\alpha > 0$ and $0 < x < 1$, the polynomials (2.9) can be represented by means of the Euler beta function

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy \qquad (a,b>0),$$

namely

$$p_{m,k}^{\langle\alpha\rangle}(x)=\binom{m}{k}B\left(\frac{x}{\alpha}+k,\frac{1-x}{\alpha}+m-k\right)/B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right).$$

Therefore for the factorial moment-generatic function of the Markov–Pólya probability distribution we find

$$g_m^{\langle\alpha\rangle}(t)=\frac{1}{B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right)}\int_0^1 y^{\frac{x}{\alpha}-1}(1-y)^{\frac{1-x}{\alpha}-1}\left[\sum_{k=0}^{m}\binom{m}{k}(ty)^k(1-y)^{m-k}\right]dy.$$

Thus we have the following formula

$$g_m^{\langle\alpha\rangle}(t)=\frac{1}{B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right)}\int_0^1 y^{\frac{x}{\alpha}-1}(1-y)^{\frac{1-x}{\alpha}-1}(1-y+ty)^m dy.$$

In this case we obtain

$$\left(g_m^{\langle\alpha\rangle}(t)\right)^{(j)}=\frac{m^{[j]}}{B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right)}\int_0^1 y^{\frac{x}{\alpha}+j-1}(1-y)^{\frac{1-x}{\alpha}-1}(1-y+ty)^{m-j}dy.$$

It follows that

$$\mu_{m,[j]}(x)=\left(g_m^{\langle\alpha\rangle}(t)\right)_{t=1}^{(j)}=\left[B\left(\frac{x}{\alpha}+j,\frac{1-x}{\alpha}\right)/B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right)\right]m^{[j]}.$$

If we apply j times the known formula

$$B(a+1,b)=\frac{a}{a+b}B(a,b),$$

we find

$$B\left(\frac{x}{\alpha}+j,\frac{1-x}{\alpha}\right)=\frac{x^{[j,-\alpha]}}{1^{[j,-\alpha]}}B\left(\frac{x}{\alpha},\frac{1-x}{\alpha}\right).$$

Therefore we obtain the following representation of the Bernstein-type polynomial (2.10):

$$\left(S_m^{\langle\alpha,\beta,\gamma\rangle}f\right)(x)=\sum_{j=0}^{m}\binom{m}{j}\frac{x^{[j,-\alpha]}}{1^{[j,-\alpha]}}\left(\Delta^j_{\frac{1}{m+\gamma}}f\right)\left(\frac{\beta}{m+\gamma}\right),$$

or, in terms of divided differences:

$$\left(S_m^{\langle\alpha,\beta,\gamma\rangle}f\right)(x)=\sum_{j=0}^{m}\frac{m^{[j]}}{(m+\gamma)^j}\left[\frac{\beta}{m+\gamma},\frac{\beta+1}{m+\gamma},\cdots,\frac{\beta+j}{m+\gamma};f\right]\frac{x^{[j,-\alpha]}}{1^{[j,-\alpha]}}.$$

3. Monotonicity properties

For the investigation of the monotonicity properties of the sequence $(S_m^{\langle\alpha\rangle} f)$, obtained from (2.10) for $\beta = \gamma = 0$, one can use a formula which shows that there is a close connection between the shape of the polynomial $S_m^{\langle\alpha\rangle} f$ and the second-order (dd):

$$(S_{m+1}^{\langle\alpha\rangle} f)(x) - (S_m^{\langle\alpha\rangle} f)(x) = \tag{3.1}$$
$$= -\frac{x(1-x)}{m(m+1)(1+\alpha)} \sum_{k=0}^{m-1} p_{m-1,k}^{\langle\alpha\rangle}(x+\alpha, 1-x+\alpha) \left[\frac{k}{m}, \frac{k+1}{m+1}, \frac{k+1}{m}; f\right],$$

where we used the notation

$$p_{n,k}^{\langle\alpha\rangle}(u,v) = \binom{n}{k} u^{[k,-\alpha]} v^{[n-k,-\alpha]} / (u+v)^{[n,-\alpha]}. \tag{3.2}$$

By means of the relation (3.1) we are able to state the following result: if f is convex (concave) of first order on $[0,1]$, then the sequence $(S_m^{\langle\alpha\rangle} f)$ is decreasing (increasing) on $[0,1]$. In the case $\alpha = 0$ formula (3.1) was established in the paper [13].

The (dd) have a great importance also in the investigation of the monotonicity properties of the derivatives or prederivatives of the sequences of Bernstein-type polynomials.

In order to study the monotonicity of the derivative of order s $(0 \leqq s \leqq m)$ of the sequence of Bernstein polynomials, it is useful to consider a class of positive linear functionals.

Consider an integer r $(0 \leqq r \leqq m)$ and the following points of the interval $[a,b]$:

$$a_i = a + ih \quad (i = 0(1)m), \qquad b_j = a + jl \quad (j = 1(1)m),$$

where $0 < h \leqq (b-a)/m$, $0 < l < (b-a)/m$.

We associate to each function f defined on $[a,b]$, the linear functionals $T_k^{(\nu)}$ $(0 \leqq k \leqq m,\ 1 \leqq \nu \leqq r+1)$, defined recursively as follows

$$\begin{aligned} T_k^{(2)}(f) &= [a_k, a_{k+1}, b_{k+1}; f] && (0 \leqq k \leqq m-1) \\ T_k^{(\nu+1)}(f) &= T_{k+1}^{(\nu)}(f) - T_k^{(\nu)}(f) && (1 < \nu \leqq r,\ \ 0 \leqq k \leqq m-r). \end{aligned} \tag{3.3}$$

In [19] we have proved that $T_k^{(\nu+1)}$ $(\nu = 2.3, \ldots, r)$ can be represented as a linear combination, with positive coefficients, of ν (dd) of order $\nu+1$ of f.

On the other hand, we have obtained the following formula for the difference between the derivatives of order s $(0 \leqq s \leqq m)$ of two consecutive

Bernstein polynomials

$$
\begin{aligned}
&(B_{m+1}f)^{(s)}(x) - (B_m f)^{(s)}(x) = \\
&= -\frac{1}{m(m+1)}\left[(m-1)^{[s]}x(1-x)\sum_{k=0}^{m-s-1} p_{m-s-1,k}(x)T_{m,k}^{(s+2)}(f) + \right. \\
&\quad + s(m-1)^{[s-1]}(1-2x)\sum_{k=0}^{m-s} p_{m-s,k}(x)T_{m,k}^{(s+1)}(f) - \\
&\quad \left. - s(s-1)\sum_{k=0}^{m-s-1} p_{m-s+1,k}(x)T_{m,k}^{(s)}(f)\right].
\end{aligned} \tag{3.4}
$$

For illustrative purposes, let us consider some particular cases of this formula.

If $s=0$ we obtain

$$
(B_{m+1}f)(x) - (B_m f)(x) = -\frac{x(1-x)}{m(m+1)}\sum_{k=0}^{m-1} p_{m-1,k}(x)T_{m,k}^{(2)}(f), \tag{3.5}
$$

where

$$
T_{m,k}^{(2)}(f) = T_k^{(2)}(f) = \left[\frac{k}{m}, \frac{k+1}{m}, \frac{k+1}{m+1}; f\right].
$$

This formula, which was obtained first, under this form, in the paper [13], permits to state the classical result; if on $[0,1]$ the function f is convex (concave) of first order, then the sequence $(B_m f)$ is decreasing (increasing) on $[0,1]$. This result was established first by W. B. Temple [22] and later by O. Aramă [2] and B. Averbach (inserted in a paper by I. J. Schoenberg [11]). It should be mentioned that Temple and Averbach did not give a representation by means of (dd) for the difference of two consecutive terms of the sequence $(B_m f)$.

We mention that our short method for proving formula (3.5) was later used in the case of different generalized Bernstein operators by A. Jakimovski [7], V. M. Šalıverdiev [21] and O. Šabozov [20].

In the case $s=1$ formula (3.4) becomes

$$
\begin{aligned}
&(B_{m+1}f)'(x) - (B_m f)'(x) = \\
&= -\frac{1}{m(m+1)}\left[(m-1)x(1-x)\sum_{k=0}^{m-2} p_{m-2,k}(x)T_{m,k}^{(3)}(f) + \right. \\
&\quad \left. + (1-2x)\sum_{k=0}^{m-1} p_{m-1,k}(x)T_{m,k}^{(2)}(f)\right],
\end{aligned}
$$

where

$$T_{m,k}^{(3)}(f) =$$
$$= \frac{2}{m}\left[\frac{k}{m}, \frac{k+1}{m}, \frac{k+2}{m}, \frac{k+2}{m+1}; f\right] + \frac{1}{m+1}\left[\frac{k}{m}, \frac{k+1}{m}, \frac{k+1}{m+1}, \frac{k+2}{m+1}; f\right].$$

It is easy to see that if the function f is convex (concave) of first and second order on the interval $\left[0, \frac{1}{2}\right]$, then the sequence $(B_m f)'(x)$ is decreasing (increasing) on this interval, while if f is concave (convex) of first order and convex (concave) of second order on $\left[\frac{1}{2}, 1\right]$, then this sequence is decreasing (increasing) on this interval.

These results corresponding to the case $s = 1$ were obtained first in our paper [13].

By using formula (3.4) one can deduce similar results in the case $s \geqq 2$, [19].

We mention that G. Mastroianni and M. R. Occorsio [9] have studied the shape preserving properties of our operator $S_m^{(\alpha)}$, defined at (2.9)–(2.10), where we have to replace $\beta = \gamma = 0$, while B. Della Vecchia [3] has extended our results [19] concerning the monotonicity of the derivatives of the sequence of Bernstein polynomials to the operator $S_m^{(\alpha)}$, by replacing the differentiation operator by the prederivative operator D_α of Nörlund, with the increment α.

4. Representations of remainders by (dd)

We have discovered in 1962, while I was working at the University of Wisconsin – Madison, a representation by second-order (dd) of the remainder in Bernstein's approximation formula $f(x) = (B_m f)(x) + (R_m f)(x)$, namely

$$(4.1) \qquad (R_m f)(x) = -\frac{x(1-x)}{m} \sum_{k=0}^{m-1} p_{m-1,k}(x)\left[x, \frac{k}{m}, \frac{k+1}{m}; f\right].$$

This result was included in a paper [12] presented in 1963 at a Conference on Approximation Theory, organized by SIAM in Gatlinburg, Tennessee, which was published in the first issue of *SIAM J. Numer. Anal. Ser. B* **1** (1964), 137–162.

We mention that A. Jakimovski [7] has used our method for obtaining a similar formula with (4.1) for a generalized Bernstein operator.

In the case of the operator $S_m^{\langle\alpha\rangle}$, connected with the Markov–Pólya probability distribution, the remainder can be expressed by the following formula (4.2)

$$(R_m^{\langle\alpha\rangle} f)(x) = -\frac{x(1-x)}{m}\frac{1+\alpha m}{1+\alpha}\sum_{k=0}^{m-1} p_{m-1,k}^{\langle\alpha\rangle}(x+\alpha, 1-x+\alpha)\left[x, \frac{k}{m}, \frac{k+1}{m}; f\right],$$

where we used the notation (3.2).

It should be mentioned that D. Leviatan [8] has used the generalized (dd), in the sense of T. Popoviciu [10], for the representation of the remainder in the case of an operator representing a generalization of the Bernstein power-series.

Now we want to deduce a representation in terms of (dd) for the remainder of the approximation formula

$$f(x) = (S_{m,r} f)(x) + (R_{m,r} f)(x), \tag{4.3}$$

where $f : [0,1] \to \mathbb{R}$, $r \in \mathbb{N}_0$, $4r < m$, $S_{m,r}$ being an (iplo), defined by an expression of the following form

$$(S_{m,r} f)(x) = \sum_{k=0}^{m} w_{m,k,r}(x) f\left(\frac{k}{m}\right). \tag{4.4}$$

We assume that for

$$I_1 = \{k \mid 0 \leqq k < r\}, \qquad I_2 = \{k \mid r \leqq k < 2r\},$$
$$I_3 = \{k \mid 2r \leqq k < m-2r\}, \quad I_4 = \{k \mid m-2r \leqq k < m-r\},$$
$$I_5 = \{k \mid m-r \leqq k \leqq m\}$$

the fundamental polynomial $w_{m,k,r}(x)$ coincides, respectively, with

$$w_{m,k,r}^{(1)}(x) = \binom{m-2r}{k} x^k (1-x)^{m+2-2r-k},$$

$$w_{m,k,r}^{(2)}(x) = \binom{m-2r}{k} x^k (1-x)^{m+2-2r-k} + 2\binom{m-2r}{k-r} x^{k+1-r}(1-x)^{m+1-r-k},$$

$$w_{m,k,r}^{(3)}(x) = \binom{m-2r}{k} x^k (1-x)^{m+2-2r-k} + 2\binom{m-2r}{k-r} x^{k+1-r}(1-x)^{m+1-r-k} +$$
$$+ \binom{m-2r}{k-2r} x^{k+2-2r}(1-x)^{m-k},$$

$$w_{m,k,r}^{(4)}(x) = 2\binom{m-2r}{k-r} x^{k+1-r}(1-x)^{m+1-r-k} + \binom{m-2r}{k-2r} x^{k+2-2r}(1-x)^{m-k},$$

$$w_{m,k,r}^{(5)}(x) = \binom{m-2r}{k-2r} x^{k+2-2r}(1-x)^{m-k}.$$

It is easy to see that we can write

$$(S_{m,r}f)(x) = \sum_{k=0}^{m-2r} \binom{m-2r}{k} x^k (1-x)^{m+2-2r-k} f\left(\frac{k}{m}\right) +$$
$$+ 2\sum_{k=r}^{m-r} \binom{m-2r}{k-r} x^{k+1-r}(1-x)^{m+1-r-k} f\left(\frac{k}{m}\right) +$$
$$+ \sum_{k=2r}^{m} \binom{m-2r}{k-2r} x^{k+2-2r}(1-x)^{m-k} f\left(\frac{k}{m}\right).$$

By changing the index of summation $k-r=j$ in the second sum and $k-2r=i$ in the third one, and then denoting overall k as index of summation, we are able to obtain the following representation

$$(S_{m,r}f)(x) = (1-x)^2 \sum_{k=0}^{m-2r} p_{m-2r,k}(x) f\left(\frac{k}{m}\right) +$$
$$+ 2x(1-x) \sum_{k=0}^{m-2r} p_{m-2r,k}(x) f\left(\frac{k+r}{m}\right) + \tag{4.5}$$
$$+ x^2 \sum_{k=0}^{m-2r} p_{m-2r,k}(x) f\left(\frac{k+2r}{m}\right).$$

Now we can state and prove the following

THEOREM. *The remainder of the approximation formula* (4.3)–(4.4)–(4.5) *can be expressed by means of the second-order (dd) in the following form:*

$$(R_{m,r}f)(x) =$$
$$= -\frac{x(1-x)}{m^2}\left\{(m-2r)\sum_{k=0}^{m-2r-1} p_{m-2r-1,k}(x)\left((1-x)^2\left[x,\frac{k}{m},\frac{k+1}{m};f\right]+\right.\right.$$
$$+ 2x(1-x)\left[x,\frac{k+r}{m},\frac{k+r+1}{m};f\right] + x^2\left[x,\frac{k+2r}{m},\frac{k+2r+1}{m};f\right]\Big) + \tag{4.6}$$
$$\left. + 2r^2 \sum_{k=0}^{m-2r} p_{m-2r,k}(x)\left((1-x)\left[x\frac{k}{m},\frac{k+r}{m};f\right] + x\left[x,\frac{k+r}{m},\frac{k+2r}{m};f\right]\right)\right\}.$$

PROOF. Denoting by e_j the monomials $e_j(x) = x^j$ $(j \geqq 0)$, we can find at once the value of the operator $S_{m,r}$ for e_0, since

$$(S_{m,r}e_0)(x) = (x+(1-x))^{m+2-r} = 1.$$

Therefore we can write successively

$$(R_{m,r}f)(x) = f(x) - (S_{m,r}f)(x) =$$
$$= (1-x)^2 \sum_{k=0}^{m-2r} p_{m-2r,k}(x)\left[f(x) - f\left(\frac{k}{m}\right)\right] +$$
$$+ 2x(1-x)\sum_{k=0}^{m-2r} p_{m-2r,k}(x)\left[f(x) - f\left(\frac{k+r}{m}\right)\right] +$$
$$+ x^2 \sum_{k=0}^{m-2r} p_{m-2r,k}(x)\left[f(x) - f\left(\frac{k+2r}{m}\right)\right].$$

By using the first order (dd) we have

$$(R_{m,r}f)(x) = \frac{(1-x)^2}{m}\sum_{k=0}^{m-2r} p_{m-2r,k}(x)(mx-k)\left[x, \frac{k}{m}; f\right] +$$
$$+ \frac{2x(1-x)}{m}\sum_{k=0}^{m-2r} p_{m-2r,k}(x)(mx-r-k)\left[x, \frac{k+r}{m}; f\right] +$$
$$+ \frac{x^2}{m}\sum_{k=0}^{m-2r} p_{m-2r,k}(x)(mx-2r-k)\left[x, \frac{k+2r}{m}; f\right].$$

If we use the identities

$$mx - k = (m-2r-k)x - k(1-x) + 2rx,$$
$$mx - r - k = (m-2r-k)x - k(1-x) + 2rx - r,$$
$$mx - 2r - k = (m-2r-k)x - k(1-x) - 2r(1-x),$$

then, on account of the equalities

$$(m-2r-k)\binom{m-2r}{k} = (m-2r)\binom{m-2r-1}{k},$$
$$k\binom{m-2r}{k} = (m-2r)\binom{m-2r-1}{k-1},$$

we can write further

$$(R_{m,r}f)(x) =$$
$$= \frac{(1-x)^2}{m}\left\{(m-2r)x\sum_{k=0}^{m-2r-1}\binom{m-2r-1}{k}x^k(1-x)^{m-2r-k}\left[x, \frac{k}{m}; f\right] -\right.$$

$$
\begin{aligned}
&-(m-2r)(1-x)\sum_{k=1}^{m-2r}\binom{m-2r-1}{k-1}x^k(1-x)^{m-2r-k}\left[x,\frac{k}{m};f\right]+\\
&2rx\sum_{k=0}^{m-2r}p_{m-2r,k}(x)\left[x,\frac{k}{m};f\right]\Bigg\}+\\
&\frac{2x(1-x)}{m}\Bigg\{(m-2r)x\sum_{k=0}^{m-2r-1}\binom{m-2r-1}{k}x^k(1-x)^{m-2r-k}\left[x,\frac{k+r}{m};f\right]-\\
&-(m-2r)(1-x)\sum_{k=1}^{m-2r}\binom{m-2r-1}{k-1}x^k(1-x)^{m-2r-k}\left[x,\frac{k+r}{m};f\right]+\\
&+2rx\sum_{k=0}^{m-2r}p_{m-2r,k}(x)\left[x,\frac{k+r}{m};f\right]-\\
&-r\sum_{k=0}^{m-2r}p_{m-2r,k}(x)\left[x,\frac{k+r}{m};f\right]\Bigg\}+\\
&+\frac{x^2}{m}\Bigg\{(m-2r)x\sum_{k=0}^{m-2r-1}\binom{m-2r-1}{k}x^k(1-x)^{m-2r-k}\left[x,\frac{k+2r}{m};f\right]-\\
&-(m-2r)(1-x)\sum_{k=1}^{m-2r}\binom{m-2r-1}{k-1}x^k(1-x)^{m-2r-k}\left[x,\frac{k+2r}{m};f\right]\Bigg\}-\\
&-\frac{2rx^2(1-x)}{m}\sum_{k=0}^{m-2r}p_{m-2r,k}(x)\left[x,\frac{k+2r}{m};f\right].
\end{aligned}
$$

If in the second, fifth and eighth sums we change the index of summation $k=j+1$ and then denoting overall k as index of summation, by rearranging the terms we obtain

$$
\begin{aligned}
&(R_{m,r}f)(x)=\\
&=\frac{x(1-x)}{m}\Bigg\{(m-2r)(1-x)^2\sum_{k=0}^{m-2r-1}p_{m-2r-1,k}(x)\left(\left[x,\frac{k}{k};f\right]-\right.\\
&\left.-\left[x,\frac{k+1}{m};f\right]\right)\Bigg\}+\\
&+\frac{2x(1-x)}{m}\Bigg\{(m-2r)x(1-x)\sum_{k=0}^{m-2r-1}p_{m-2r-1,k}(x)\left(\left[x,\frac{k+r}{m};f\right]-\right.
\end{aligned}
\tag{4.7}
$$

$$- \left[x, \frac{k+r+1}{m}; f\right]\Big)\Big\} +$$

$$+ \frac{x(1-x)}{m}\left\{(m-2r)x^2 \sum_{k=0}^{m-2r-1} p_{m-2r-1,k}(x)\left(\left[x, \frac{k+2r}{m}; f\right] - \right.\right.$$

$$\left.\left. - \left[x, \frac{k+2r+1}{m}; f\right]\right)\right\} +$$

$$+ \frac{2rx^2(1-x)}{m} \sum_{k=0}^{m-2r} p_{m-2r,k}(x)\left(\left[x, \frac{k+r}{m}; f\right] - \left[x, \frac{k+2r}{m} f\right]\right).$$

According to the recurrence relation of (dd) we can write

$$\left[x, \frac{k}{m}; f\right] - \left[x, \frac{k+1}{m}; f\right] = -\frac{1}{m}\left[x, \frac{k}{m}, \frac{k+1}{m}; f\right],$$

$$\left[x, \frac{k+r}{m}; f\right] - \left[x, \frac{k+r+1}{m}; f\right] = -\frac{1}{m}\left[x, \frac{k+r}{m}, \frac{k+r+1}{m}; f\right],$$

$$\left[x, \frac{k+2r}{m}; f\right] - \left[x, \frac{k+2r+1}{m}; f\right] = -\frac{1}{m}\left[x, \frac{k+2r}{m}, \frac{k+2r+1}{m}; f\right],$$

$$\left[x, \frac{k}{m}; f\right] - \left[x, \frac{k+r}{m}; f\right] = -\frac{r}{m}\left[x, \frac{k}{m}, \frac{k+r}{m}; f\right],$$

$$\left[x, \frac{k+r}{m}; f\right] - \left[x, \frac{k+2r}{m}; f\right] = -\frac{r}{m}\left[x, \frac{k+r}{m}, \frac{k+2r}{m}; f\right].$$

By using these relations in (4.7) we arrive just at the representation (4.6) and so our theorem is proved

In the particular case $r = 0$ or $r = 1$ we have $S_{m,0} = S_{m,1} = B_m$ and (4.6) reduces to (4.1).

Formula (4.6) permits to see that

(i) $(R_{m,r}f)(0) = (R_{m,r}f)(1) = 0$, which shows that the approximating polynomial $S_{m,r}f$ is interpolatory at both ends of the interval $[0, 1]$;

(ii) The degree of exactness of formula (4.3) is one;

(iii) If f is a convex function of first-order on $[0, 1]$, without being linear, then we have $S_{m,r}f > f$ on $(0, 1)$, while if f is concave of first-order on $[0, 1]$, then $S_{m,r}f < f$ on $(0, 1)$;

(iv) If all the (dd) of second order of f are bounded on $[0, 1]$ then we have

$$|(R_{m,r}f)(x)| \leqq \left[1 + 2\frac{r(r-1)}{m}\right] \frac{x(1-x)}{m} M_2(f),$$

where $M_2(f)$ is the least upper bound of the absolute values of the second-order (dd) of f on $[0, 1]$;

(v) If we apply a criterion of T. Popoviciu [10] we can conclude that there exist three distinct points $t_{m,1}, t_{m,2}, t_{m,3}$ on $[0,1]$ such that

$$(R_{m,r}f)(x) = (R_{m,r}e_2)(x)[t_{m,1}, t_{m,2}, t_{m,3}; f].$$

If we replace $f = e_2$ in (4.6) then we find that

$$(R_m e_2)(x) = -\left[1 + 2\frac{r(r-1)}{m}\right]\frac{x(1-x)}{m};$$

(vi) The corresponding Voronovskaja theorem states that if the function f possesses a second derivative at a point x of $[0,1]$ then we have

$$(R_{m,r}f)(x) = -\left[1 + 2\frac{r(r-1)}{m}\right]\frac{x(1-x)}{2m}f''(x) + \frac{\varepsilon_m(x)}{m},$$

where $\varepsilon_m(x)$ tends to zero when m tends to infinity.

REFERENCES

[1] Altomare, F. and Campiti, M., *Korovkin-type approximation theory and its applications*, De Gruyter Studies in Mathematics, 17, Walter De Gruyter, Berlin, 1994. *MR* **95g**:41001

[2] Aramă, O., Properties concerning the monotonicity of the sequence of polynomials of interpolation of S. N. Bernstein and their applications to the study of approximation of functions, *Mathematica (Cluj)* **2** (**25**) (1960), 25–40 (in Russian). *MR* **23** #A1986

[3] Della Vecchia, B., On monotonicity of some linear positive operators, *Numerical Methods and Approximation Theory III* (Niš, 1987), Univ. Niš, Niš, 1988, 165–178. *MR* **90c**:41044

[4] Della Vecchia, B., On the approximation of function by means of the operators of D. D. Stancu, *Studia Univ. Babeş-Bolyai Math.* **37** (1992), 3–36. *MR* **95m**:41042

[5] Gonska, H. H. and Meier, J., Quantitative theorems on approximation by Bernstein-Stancu operators, *Calcolo* **21** (1984), 317–335. *MR* **87g**:41055

[6] Jakimovski, A. and Ramanujan, M. S., A uniform approximation theorem and its application to moment problems, *Math. Z.* **84** (1964), 143–153. *MR* **30** #2272

[7] Jakimovski, A., Generalized Bernstein polynomials for discontinuous and convex functions, *J. Analyse Math.* **23** (1970), 171–183. *MR* **42** #4926

[8] Leviatan, D., On the remainder in the approximation of functions by Bernstein-type operators, *J. Approximation Theory* **2** (1969), 400–409. *MR* **40** #6131

[9] Mastroianni, G. and Occorsio, M. R., Sulle derivate dei polinomi di Stancu, *Rend. Accad. Sci. Fis. Mat. Napoli* (4) **45** (1978), 273–281. *MR* **82a**:41019

[10] Popoviciu, T., Sur le reste dans certaines formules linéaires d'approximation de l'analyse, *Mathematica (Cluj)* **1** (**24**) (1959), 95–142. *MR* **23** #B2567

[11] Schoenberg, I. J., On variation diminishing approximation methods, *On numerical approximation* (Proc. Sympos., Madison, WI, 1958), ed. R. E. Langer, University of Wisconsin Press, Madison, 1959, 249–274. *MR* **21** #961

[12] Stancu, D. D., The remainder of certain linear approximation formulas in two variables, *SIAM J. Numer. Anal.* **1** (1964), 137–163. *MR* **31** #1503

[13] Stancu, D. D., On the monotonicity of the sequence formed by the first order derivatives of the Bernstein polynomials, *Math. Z.* **98** (1967), 46–51. *MR* **35** #3018

[14] Stancu, D. D., Approximation of functions by a new class of linear polynomial operators, *Rev. Roumaine Math. Pures Appl.* **13** (1968), 1173–1194. *MR* **38** #6278

[15] Stancu, D. D., Use of probabilistic methods in the theory of uniform approximation of continuous functions, *Rev. Roumaine Math. Pures Appl.* **14** (1969), 673–691. *MR* **40** #606

[16] Stancu, D. D., On a generalization of the Bernstein polynomials, *Studia Univ. Babeş–Bolyai Ser. Math. Phys.* **14** (1969) (no. 2), 31–45. *MR* **43** #775

[17] Stancu, D. D., Approximation of functions by means of some new classes of positive linear operators, *Numerische Methoden der Approximationstheorie*, Bd. 1 (Tagung, Math. Forschungsinstitut, Oberwolfach, 1971), Internat. Schriftenreihe Numer. Math., Band 16, eds. L. Collatz, G. Meinardus, Birkhäuser, Basel, 1972, 187–203. *MR* **52** #1107

[18] Stancu, D. D., The use of linear interpolation for the construction of a class of Bernstein polynomials, *Stud. Cerc. Mat.* **28** (1976), 369–379 (in Romanian). *MR* **54** #3226

[19] Stancu, D. D., Application of divided differences to the study of monotonicity of the derivatives of the sequence of Bernstein polynomials, *Calcolo* **16** (1979), 431–445. *MR* **82b**:41008

[20] Šabozov, O. S., Monotonicity conditions for a sequence of rational operators that generalize the Bernstein polynomials, *Dokl. Akad. Nauk Tadzik. SSR* **17** (1974), 12–15 (in Russian). *MR* **52** #3811

[21] Šahverdiev, V. M., Conditions for monotonicity of a certain form of generalized polynomials of S. N. Bernstein type and their derivatives, *Izv. Akad. Nauk Azerbaidzan SSR Ser. Fiz. Tehn. Mat. Nauk* **1968**, no. 5, 16–21 (in Russian). *MR* **40** #6125

[22] Temple, W. B., Stieltjes integral representation of convex functions, *Duke Math. J.* **21** (1954), 527–531. *MR* **16**, 22a

(Received February 14, 1996)

FACULTY OF MATHEMATICS AND INFORMATICS
UNIVERSITY BABEŞ–BOLYAI
STR. M. KOGĂLNICEANU NR. 1
RO–3400 CLUJ-NAPOCA
ROMANIA
ddstancu@math.ubbcluj.ro

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 81–97

EXTENDED INTERPOLATION WITH ADDITIONAL NODES IN SOME SOBOLEV-TYPE SPACES

M. R. CAPOBIANCO and M. G. RUSSO

Abstract

Convergence and boundedness of the extended Lagrange interpolating operator with additional nodes are studied in the space $L^p_{w,t}$ of Sobolev type.

1. Introduction

By considering two weight functions σ and τ in $[-1,1]$ and by denoting with $\{p_m(\sigma)\}_{m=0}^{\infty}$ and $\{p_n(\tau)\}_{n=0}^{\infty}$ the sequences of the corresponding systems of orthogonal polynomials, if the polynomial $Q_{m+n}=p_m(\sigma)p_n(\tau)$ has $m+n$ simple zeros in $(-1,1)$, it can be defined the extended interpolating polynomial, i.e., the Lagrange interpolating polynomial of degree $m+n-1$ which interpolates a given function f at the zeros of Q_{m+n}.

Extended interpolation and related matrices have many applications in numerical analysis and in approximation theory (cf. for instance [7], [8], [12] and [19]).

In this paper, letting

$$\sigma(x)\equiv w(x):=(1-x)^{\alpha}(1+x)^{\beta}\prod_{j=1}^{K}|\chi_j-x|^{\rho_j}, \tag{1.1}$$

i.e., with $w\in GSJ$, and

$$\tau(x)\equiv\bar{w}(x):=(1-x^2)w(x) \tag{1.2}$$

we will consider the extended interpolating polynomial $L_{2m+1,r,s}(w,\bar{w};f)$, which interpolates the function f at the $2m+1$ simple zeros $\{x_{m+1,i}(w)\}_{i=1}^{m+1}$ $\bigcup\{x_{m,i}(\bar{w})\}_{i=1}^{m}$ of $p_{m+1}(w)p_m(\bar{w})$ (we recall that $x_{m+1,i}(w)<x_{m,i}(\bar{w})<x_{m+1,i+1}(w)$ cf. [7]) and at the additional points $y_{m,j}$, $(j=1,2,\dots,s)$ and

1991 *Mathematics Subject Classification.* Primary 41A05; Secondary 65D05.

Key words and phrases. Extended Lagrange interpolation, orthogonal polynomials, Hilbert transform.

$z_{m,j}$, $(j = 1, 2, \ldots, r)$, chosen such that (cf. [1])

$$
\begin{aligned}
y_j &= -1 + j\frac{x_{m+1,1}(w) + 1}{s+1}, \quad j = 1, \ldots, s, \\
z_j &= 1 - j\frac{1 - x_{m+1,m+1}(w)}{r+1}, \quad j = 1, \ldots, r.
\end{aligned} \tag{1.3}
$$

With this choice, obviously, we have

$$
\begin{aligned}
x_{m+1,1}(w) - y_s \sim m^{-2} \sim z_1 - x_{m+1,m+1}(w), \\
1 + y_1 \sim m^{-2} \sim 1 - z_r.
\end{aligned} \tag{1.4}
$$

Operator $L_{2m+1,r,s}(w, \bar{w})$ was introduced in [7]; subsequently, in [6] and [12], it was considered as an operator mapping the space of continuous functions into the usual L^p-weighted space L^p_u (cf. definition in Section 2); in those papers necessary and sufficient conditions are stated for the boundedness of the operator and then estimates for the convergence to zero of the interpolation error in L^p_u, using the best uniform approximation, are found.

On the other hand it is well known that, even if the Lagrange operator is unbounded in L^p_u, it is very suitable, for many applications, to prove the boundedness of the operator in some subspaces of L^p_u. For this reason recently ordinary Lagrange interpolating operators (i.e., using the zeros of only one sequence of orthogonal polynomials) have been studied in [3] and in [5] in the Sobolev-type space $L^p_{u,t}$, defined by formula (2.11).

The introduction of additional points in these interpolatory processes becomes a necessity. In fact, for example in [4], the authors obtained a result on the uniform boundedness of operator $L_{2m+1}(w, \bar{w}) \equiv L_{2m+1,0,0}(w, \bar{w})$ in the space $L^p_{u,t}$, under conditions essentially mean the positivity of the exponents of u.

In this paper we find the boundedness of the operator $L_{2m+1,r,s}(w, \bar{w})$ in the space $L^p_{u,t}$, under wider conditions for u and w (not only regarding the positivity of the exponents of u) and show the crucial role of additional points implicitly. A crucial point for our purposes is the proof of an extended Marcinkiewicz inequality which, besides, it is a suitable result that can be used in many other applications (for example in estimating the error of extended quadrature formulas).

By using this inequality and then via one-sided approximation, we are able to give an estimate for the interpolation error in L^p-weighted spaces using the same norm of the weak derivatives of the function f. From this we can deduce a result of simultaneous convergence of the derivatives of the extended Lagrange interpolating polynomial to the corresponding derivatives of the function; this result is the tool that finally allows us to prove the boundedness of the operator $L_{2m+1,r,s}(w, \bar{w})$ in the space $L^p_{u,t}$.

2. Main results

At first we fix some notations.

A function u is called a "generalized Ditzian–Totik weight" ($u \in GDT$), if

$$u(x) = \prod_{j=0}^{M} |c_j - x|^{\Gamma_j} \log^{\gamma_j} \frac{e}{|c_j - x|}, \quad -1 \leqq x \leqq 1, \tag{2.1}$$

where $-1 = c_0 < c_1 < \cdots < c_{M-1} < c_M = 1$, $\Gamma_j > -1$ and $\gamma_j \geqq 0$, $j = 0, 1, \ldots, M$. When $\gamma_j = 0$, $j = 0, 1, \ldots, M$, u is a generalized smooth Jacobi weight (in short $u \in GSJ$) and when $\Gamma_j = \gamma_j = 0$, $j = 1, 2, \ldots, M-1$, u is a weight of Ditzian–Totik type (see for instance [5], [9]).

Let $u \in GDT$; we say that $f \in L^p_u([a,b])$, $-1 \leqq a \leqq b \leqq 1$, $1 \leqq p < \infty$, if and only if $\|f\|_{L^p_u([a,b])} := \|fu\|_{L^p([a,b])} < \infty$, where $\|g\|^p_{L^p([a,b])} := \int_a^b |g(x)|^p dx$. If $a = -1$ and $b = 1$, we write $f \in L^p_u$ and $\|f\|_{u,p} = \|fu\|_p < \infty$.

In the sequel $\mathbb{P}_m$ denotes the set of the algebraic polynomials of degree at most m. Then, if $u \in L_p$

$$E_m(f)_{u,p} := \inf_{P \in \mathbb{P}_m} \|f - P\|_{u,p} \tag{2.2}$$

is called the error of the best approximation by algebraic polynomials of the function f in the space L^p_u. Now let g be a bounded and measurable function, and let u be a weight function such that $u \in L_p$. We set

$$\tilde{E}_m(g)_{u,p} = \inf\{\|(q^+ - q^-)u\|_p, q^{\pm} \in \mathbb{P}_m, q^-(x) \leqq g(x) \leqq q^+(x), |x| \leqq 1\}. \tag{2.3}$$

$\tilde{E}_m(g)_{u,p}$ is called the error of the best one-sided approximation of the function g in L_p-space with weight u.

In the sequel we denote by "C" some positive constant that can be different in different formulas.

Now we fix some notations useful in the statements below. For sake of simplicity we denote $t_i \equiv t_{2m+r+s+1,i} = \cos \tau_i$, $i = 1, \ldots, 2m + r + s + 1$, the zeros of $A_s B_r p_{m+1}(w) p_m(\bar{w})$, where $A_s(x) = \prod_{j=1}^{s} (x - y_j)$, $B_r(x) = \prod_{j=1}^{r} (x - z_j)$, with y_j, z_j as in (1.4). Further we set $v^{(\gamma,\delta)}(x) = (1-x)^{\gamma}(1+x)^{\delta}$ with $\gamma, \delta > -1$, and $|x| \leqq 1$.

Now let $\{p_m(u^p)\}_{m=0}^{\infty}$ the sequence of orthonormal polynomials with positive leading coefficients related to the weight u^p; we denote

$$\lambda_m(u^p; x) = \frac{1}{\sum_{k=0}^{m-1} p_k^2(u^p; x)},$$

the m-th Christoffel function with respect to the weight u^p.

We state the following Marcinkiewicz-type inequality:

THEOREM 2.1. *Let w be a generalized smooth Jacobi weight and $u \in GDT$ with $u \in L_p$, $1 < p < \infty$. If for some non-negative integers r and s,*

$$\frac{u}{w} v^{(r-1,s-1)} \in L_p, \quad \frac{w}{u} v^{(1-r,1-s)} \in L_q, \qquad q^{-1} + p^{-1} = 1, \tag{2.4}$$

then, for every algebraic polynomial $P \in \mathbb{P}_{2m+r+s}$ the relations

$$C_1 \|Pu\|_p^p \leqq \sum_{i=1}^{2m+r+s+1} \lambda_m(u^p, t_i) |P(t_i)|^p \leqq C_2 \|Pu\|_p^p \tag{2.5}$$

holds, where C_1 and C_2 are some positive constants independent of P and m. Moreover if the inequality at the left-hand side in (2.5) holds true, then $\frac{u}{w} v^{(r-1,s-1)} \in L_p$ follows.

We observe that the second inequality in (2.5) can be obtained for more general u. In fact by using the same machinery as in [11], it is sufficient that only $u \in L_p$ and $a_1 \leqq m|\tau_i - \tau_{i-1}| \leqq a_2$, $i = 1, \ldots, 2m + r + s + 1$ with a_1, a_2 positive constants independent of m.

Inequality (2.5) can be used in different applications; in our context we use it to prove the theorem below.

THEOREM 2.2. *Under the assumptions of Theorem 2.1 for every bounded and measurable function $f : [-1, 1] \to \mathbb{R}$*

$$\| [f - L_{2m+1,r,s}(w, \bar{w}; f)] u \|_p \leqq C \tilde{E}_{2m+r+s}(f)_{u,p}, \tag{2.6}$$

where C is independent of f and m.

Under less restrictive conditions, [6] and [12] proved $\forall f \in C^0[-1, 1]$:

$$\|f - L_{2m+1,r,s}(w, \bar{w}; f)\|_{u,p} \leqq C E_m(f)_\infty, \tag{2.7}$$

where $E_m(f)_\infty$ is the error of best uniform approximation by algebraic polynomials. However, Theorem 2.2 gives a more refined rate of convergence, using

$$\tilde{E}_m(f)_{u,p} \leqq \tilde{E}_m(f)_\infty \|u\|_p = 2E_m(f)_\infty \|u\|_p.$$

Moreover, from Theorem 2.2 we can deduce convergence estimate also when f is unbounded at the endpoints of $[-1, 1]$, when (2.7) cannot be used; more precisely, denoting by AC_{LOC} the class of all absolutely continuous functions in any closed subset $[a, b] \subset (-1, 1)$ and letting $\varphi(x) = \sqrt{1 - x^2}$ and $I_m = [-1, 1] - [y_1, z_r]$, where $y_1 = -1 + \dfrac{1 + x_{m+1,1}}{s+1}$ and $z_r = 1 - \dfrac{1 - x_{m+1,m+1}}{r+1}$, we have the following

COROLLARY 2.1. *If $f \in AC_{LOC}$ and $f'\varphi^{\frac{2}{p}} u \in L_1$, then under the assumptions of Theorem 2.1*

$$\begin{aligned} &\| [f - L_{2m+1,r,s}(w, \bar{w}; f)]\, u\|_p \\ &\leqq \frac{C}{m} \|f'\varphi u\|_{L^p([y_1, z_r])} + \int_{I_m} |f'(t)| \varphi^{\frac{2}{p}}(t) u(t) dt \end{aligned} \tag{2.8}$$

with C independent of f and m. Moreover, if condition $f'\varphi^{\frac{2}{p}} u \in L_1$ is replaced by

$$f'\varphi u \in L_p,$$

then we have

$$\| [f - L_{2m+1,r,s}(w, \bar{w}; f)]\, u\|_p \leqq \frac{C}{m} E_{2m+r+s-1}(f')_{u\varphi,p}, \tag{2.9}$$

where C is some positive constant independent of f and m.

REMARK 1. First of all if we set $r = s = 0$ in (2.4) and (2.8) we get result obtained in [4] (cf. Theorem 2.1, Corollary 2.2, p. 4).

However, it is possible to apply the results given above in some cases the theorems in [4] cannot be used. For example, if we fix $u(x) = (1-x^2)^{-\frac{1}{3}}$, $p = 2$, then it does not exist any weight function of the type $w(x) = (1-x^2)^{\alpha}$, that verifies the assumption of Theorem 2.1 in [4]. On the other hand, Theorem 2.2 and Corollary 2.1 can be applied, in this case, with $\max\left\{r - \frac{11}{6}, s - \frac{11}{6}\right\} < \alpha < \min\left\{r - \frac{5}{6}, s - \frac{5}{6}\right\}$; for example if we fix $r = s = 1$ we find convergence results for all $\alpha \in \left(-\frac{5}{6}, \frac{1}{6}\right)$. Another example: let $w(x) = (1-x^2)^{\alpha}$ and $u(x) = (1-x^2)^{\gamma}$. In this case (2.4) means

$$\gamma > -\frac{1}{p},$$

$$\max\left\{\gamma + r - 2 + \frac{1}{p}, \gamma + s - 2 + \frac{1}{p}\right\} < \alpha < \min\left\{\gamma + r - 1 + \frac{1}{p}, \gamma + s - 1 + \frac{1}{p}\right\}.$$

Then, if $f(x) = \log(1+x)$, from (2.9) we obtain

$$\|[f - L_{2m+1,r,s}(w, \bar{w}; f)]u\|_p = O\left(m^{-2\gamma - \frac{2}{p}}\right).$$

Now we give a result on simultaneous convergence.

THEOREM 2.3. *Let k be a positive integer. If $f^{(k-1)} \in AC_{LOC}$ and $f^{(k)}\varphi^k \in L^p_u$, then under the assumptions of Theorem 2.1 with $\Gamma_k < 1 - \frac{1}{p}$, we have*

$$\left\| \left[f^{(i)} - L^{(i)}_{2m+1,r,s}(w, \bar{w}; f) \right] \varphi^i \right\|_{u,p} \leqq \frac{C}{m^{k-i}} \|f^{(k)}\varphi^k\|_{u,p} \tag{2.10}$$

for $i = 1, \ldots, k$ and with some constant C independent of f and m.

Let t be a non-negative integer; we define the space $L^p_{u,t}$ as the space of all functions $f \in L^p_u$ with

$$\|f\|_t := \left(\sum_{k=0}^{t} \|f^{(k)}\varphi^k\|_{u,p} \right)^{\frac{1}{p}} < \infty. \tag{2.11}$$

This space was introduced, for $p = 2$ and with and equivalent norm, by Sloan and Stephan [18]; subsequently, again for $p = 2$, in [2] the authors proved some suitable properties of this space, among others that $L^2_{u,t}$ is a Hilbert space. For $p \neq 2$ it is possible to prove, with the same technique used in [2] and by taking into account that L^p_u is a complete space, that $L^p_{u,t}$ is a Banach space. For this functional space of Sobolev type we have

THEOREM 2.4. *Let $t \geqq 1$ be a positive integer and u and w as in Theorem 2.3. Then*

$$\sup_m \|L_{2m+1,r,s}\|_{L^p_{u,t} \to L^p_{u,t}} < \infty. \tag{2.12}$$

Moreover if $f \in L^p_{u,t}$, then for $l \in \mathbb{N}$, with $0 \leqq l \leqq t$

$$\|f - L_{2m+1,r,s}(w, \bar{w}; f)\|_l \leqq \frac{C}{m^{t-l}} \|f\|_t \tag{2.13}$$

with some positive constant C independent of f and m.

REMARK 2. If we set

$$\sigma \equiv w_1(x) = (1-x)w(x), \quad \tau \equiv w_2(x) = (1+x)w(x)$$

it is well known [7] that the polynomial $p_m(w_1)p_m(w_2)$ has $2m$ simple zeros in $(-1, 1)$. So we can define the extended interpolating polynomial $L_{2m,r,s}(w_1, w_2; f)$ which interpolates f at the zeros of $p_m(w_1)p_m(w_2)$ and at the additional points of the matrix $\{y_j\}_{j=1}^s \cup \{z_j\}_{j=1}^r$. Then for an operator of this kind all the above theorems remain true under the same assumptions for f, u, and w.

3. Proofs of the main results

In the sequel, if A and B are two expressions depending on some variables, then we write $A \sim B$ if $|A/B|^{\pm 1} \leqq C$ uniformly for the variables under consideration.

Let $u \in GDT$ as in (2.1). We introduce the notation

$$u_m(x) = \prod_{j=1}^{M-1} \left(|c_j - x| + \frac{1}{m} \right)^{\Gamma_j} \log^{\gamma_j} \frac{e}{(|c_j - x| + m^{-1})} \left(\sqrt{1+x} + \frac{1}{m} \right)^{2\Gamma_0}$$

(3.1)

$$\times \log^{\gamma_0} \left(\frac{e}{1+x+m^{-2}} \right) \left(\sqrt{1-x} + \frac{1}{m} \right)^{2\Gamma_M} \log^{\gamma_M} \left(\frac{e}{1-x+m^{-2}} \right).$$

Now let $w \in GSJ$, as in (1.1). If $\{p_m(w;x)\}_{m=0}^{\infty}$ is the sequence of orthonormal polynomials with positive leading coefficients corresponding to the weight w, we have

$$|p_m(w;x)| < \frac{C}{\sqrt{w\varphi}}, \qquad |x| < 1, \tag{3.2}$$

$$|p_m(w;x)| \sim p_m(w;1) \sim m^{\alpha+\frac{1}{2}}, \quad 1 - m^{-2} \leqq x \leqq 1, \tag{3.3}$$

and

$$|p_m(w;x)| \sim (-1)^m p_m(w;-1) \sim m^{\beta+\frac{1}{2}}, \quad -1 \leqq x \leqq -1 + m^{-2} \tag{3.4}$$

uniformly for $m \in \mathbb{N}$ (cf. [16], Theorem 6.3.28, p. 120 and Theorem 9.33, p. 171).

Now we write explicitly the expression of the extended interpolating polynomial $L_{2m+1,r,s}(w, \bar{w}; f)$ at the zeros of $p_{m+1}(w)p_m(\bar{w})$ and at the additional points of the matrix $Y \cup Z \equiv \{y_j\}_{j=1}^{s} \cup \{z_j\}_{j=1}^{r}$. By

$$H_m(w; f; x) = \sum_{i=1}^{m} \frac{\lambda_{m,i}(w)}{x - x_{m,i}(w)} f(x_{m,i}(w)),$$

$$A_s(x) = \prod_{j=1}^{s} (x - y_j), \qquad B_r(x) = \prod_{j=1}^{r} (x - z_j),$$

where $\lambda_{m,i} \equiv \lambda_m(w; x)$ is the m-th Christoffel constant, we can write

$$L_{2m+1,r,s}(w, \bar{w}; f) =$$

(3.5)

$$= p_{m+1}(w)p_m(\bar{w}) \left\{ C_m^{-1} A_s B_r \left[H_{m+1}\left(w; v^{(1,1)} \frac{f}{A_s B_r} \right) - H_m\left(\bar{w}; \frac{f}{A_s B_r} \right) \right] \right.$$

$$\left. + A_s L_r \left(Z; \frac{f}{A_s p_{m+1}(w) p_m(\bar{w})} \right) + B_r L_s \left(Y; \frac{f}{B_r p_{m+1}(w) p_m(\bar{w})} \right) \right\},$$

where $\{C_m\} < \infty$ (cf. [7], formula (2.16), p. 203) and $L_r(Z; g)$ and $L_s(Y; g)$ are the Lagrange interpolating polynomials related to the matrices Z and Y, respectively. Moreover we can write

$$L_r\left(Z; \frac{f}{A_s p_{m+1}(w) p_m(\bar{w})}; x\right) = \sum_{k=1}^{r} \prod_{i=1 i\neq k}^{r} \frac{x - z_i}{z_k - z_i} \frac{f(z_k)}{A_s(z_k) p_{m+1}(w; z_k) p_m(\bar{w}; z_k)} \tag{3.6}$$

and

$$L_s\left(Y; \frac{f}{B_r p_{m+1}(w) p_m(\bar{w})}; x\right) = \sum_{k=1}^{s} \prod_{i=1 i\neq k}^{s} \frac{x - y_i}{y_k - y_i} \frac{f(y_k)}{B_r(y_k) p_{m+1}(w; y_k) p_m(\bar{w}; y_k)}. \tag{3.7}$$

To prove the theorems we need some auxiliary results. The first is a Remez-type inequality. Let $u \in GDT$, $Q \in \mathbb{P}_m$; there exists a $b_0 < m$ such that if $B \subset [-1, 1]$ with $\text{meas}(\cos^{-1} B) \leqq \dfrac{b}{m}$, $b \leqq b_0$, then, for $0 < p < \infty$, we have

$$\|Qu\|_p \leqq C \|Qu\|_{L^p([-1,1]\setminus B)}, \tag{3.8}$$

where "C" is independent of m and Q. This result, under more general assumption on u, can be found in [13] (Theorem 1, p. 2). The second result is the following. For $f \in L_1$ in $[-1, 1]$, the Hilbert transform $H(f)$ is defined by

$$H(f; t) = \lim_{\varepsilon \to 0} \int_{|x-t| \geqq \varepsilon} \frac{f(x)}{x - t} dx. \tag{3.9}$$

The operator H is bounded in L_p for $1 < p < \infty$ (M. Riesz). Further, if μ is a weight function, then H is a bounded operator in L^p_μ if and only if

$$\|\mu\|_{L_p(D)} \|\mu^{-1}\|_{L_q(D)} \leqq C(\text{meas}\, D), \quad \frac{1}{p} + \frac{1}{q} = 1, \tag{3.10}$$

for any interval $D \subseteqq (-1, 1)$ and with a positive constant independent of D (see, e.g., [15], [10]). Now, if $u \in GDT$ as in (2.1) and $-\dfrac{1}{p} < \Gamma_k < \dfrac{1}{q}$, $k = 0, \ldots, M$, then (3.10) is satisfied (cf. [5]) and so we can write

$$\|H(f)u\|_p \leqq C \|fu\|_p. \tag{3.11}$$

Property (3.11), together with condition (3.10), plays a crucial role in the proof of Theorem 2.1 (in particular, relations (2.4) come directly from condition (3.10)).

PROOF OF THEOREM 2.1. Firstly we set

$$A = \left[-1+am^{-2}, 1-am^{-2}\right] \setminus \left(\bigcup_{k=1}^{K}\left[\chi_k - am^{-1}, \chi_k + am^{-1}\right]\right)\left(\bigcup_{k=1}^{M-1}\left[c_k - am^{-1}, c_k + am^{-1}\right]\right), \tag{3.12}$$

where a is a fixed proper constant. Let $P \in \mathbb{P}_{2m+r+s}$; obviously $P(x) \equiv L_{2m+1,r,s}(w, \bar{w}; P; x)$. If we set $F(x) = |L_{2m+1,r,s}(w, \bar{w}; P; x)|$ and $G(x) = \operatorname{sgn}(L_{2m+1,r,s}(w, \bar{w}; P; x))$ we can write, (cf. (3.8))

$$\begin{aligned}\|Pu\|_p^p &\equiv \|L_{2m+1,r,s}(w, \bar{w}; P)u\|_p^p \leqq C\|L_{2m+1,r,s}(w, \bar{w}; P)u\|_{L_p(A)}^p \\ &= C\int_A |L_{2m+1,r,s}(w, \bar{w}; P; x)|^p u^p(x)dx \\ &= C\int_A L_{2m+1,r,s}(w, \bar{w}; P; x)F^{p-1}(x)u^p(x)G(x)dx.\end{aligned}$$

By recalling expressions (3.5)–(3.7), we have

$$\begin{aligned}\|Pu\|_p^p \leqq C\Bigg\{&\left|\sum_{i=1}^{m+1}\lambda_{m+1,i}(w)\frac{P(x_{m+1,i}(w))(1-x_{m+1,i}^2(w))}{A_s(x_{m+1,i}(w))B_r(x_{m+1,i}(w))}\right. \\ &\left.\times\int_A \frac{A_s(x)B_r(x)p_{m+1}(w;x)p_m(\bar{w};x)}{x-x_{m+1,i}(w)}F^{p-1}(x)u^p(x)G(x)dx\right| \\ &+\left|\sum_{i=1}^{m}\lambda_{m,i}(\bar{w})\frac{P(x_{m,i}(\bar{w}))}{A_s(x_{m,i}(\bar{w}))B_r(x_{m,i}(\bar{w}))}\right. \\ &\left.\times\int_A \frac{A_s(x)B_r(x)p_{m+1}(w;x)p_m(\bar{w};x)}{x-x_{m,i}(\bar{w})}F^{p-1}(x)u^p(x)G(x)dx\right| \\ &+\left|\sum_{k=1}^{r}\frac{P(z_k)}{A_s(z_k)p_{m+1}(w;z_k)p_m(\bar{w};z_k)}\right. \\ &\left.\times\prod_{i=1\,i\neq k}^{r}\frac{1}{z_k-z_i}\int_A (x-z_i)A_s(x)p_{m+1}(w;x)p_m(\bar{w};x)F^{p-1}(x)u^p(x)G(x)dx\right| \\ &+\left|\sum_{k=1}^{s}\frac{P(y_k)}{B_r(y_k)p_{m+1}(w;y_k)p_m(\bar{w};y_k)}\right.\end{aligned} \tag{3.13}$$

$$\times \prod_{i=1 i\neq k}^{s} \frac{1}{y_k - y_i} \int_A (x-y_i) B_r(x) p_{m+1}(w;x) p_m(\bar{w};x) F^{p-1}(x) u^p(x) G(x) dx \Bigg| \Bigg\}$$

$$\leqq C\,[J_1 + J_2 + J_3 + J_4]\,.$$

We begin to evaluate J_1. By recalling that

$$|A_s(x)B_r(x)| \sim v^{(r,s)}(x), \qquad |x| \leqq 1 - Cm^{-2} \tag{3.14}$$

we deduce

$$J_1 \leqq C \left| \sum_{i=1}^{m+1} \frac{\lambda_{m+1,i}(w) P(x_{m+1,i}(w)) \varphi^2(x_{m+1,i}(w))}{v^{(r,s)}(x_{m+1,i}(w))} \right.$$

$$\left. \times \int_A \frac{p_{m+1}(w;x) p_m(\bar{w};x) A_s(x) B_r(x)}{x - x_{m+1,i}(w)} F^{p-1}(x) u^p(x) G(x) dx \right| .$$

By Lemma 1, p. 6 in [13], we know that a polynomial $q_m(x)$ of degree $3m$ exists such that

$$q_m(x) \sim \frac{\bar{w}_m(x)}{v_m^{(r,s)}}, \quad |x| \leqq 1, \tag{3.15}$$

(for $\bar{w}_m$ and $v_m^{(r,s)}$, cf. notation (3.1)); then letting

$$Q(x) = p_{m+1}(w;x) p_m(\bar{w};x) A_s(x) B_r(x) q_m(x)$$

we can write

$$J_1 \leqq C \left| \sum_{i=1}^{m+1} \frac{\lambda_{m+1,i}(w) P(x_{m+1,i}(w)) \varphi^2(x_{m+1,i}(w))}{v^{(r,s)}(x_{m+1,i}(w))} \right.$$

$$\left. \times \int_A \frac{Q(x) - Q(x_{m+1,i}(w))}{x - x_{m+1,i}(w)} \frac{F^{p-1}(x) G(x) u^p(x)}{q_m(x)} dx \right| .$$

Now if

$$\pi(t) = \int_A \frac{Q(x) - Q(t)}{x - t} \frac{F^{p-1}(x) G(x) u^p(x)}{q_m(x)} dx \tag{3.16}$$

we observe that π is a polynomial, with respect to the variable t, of degree lm, with $l \in \mathbb{N}$. Then we have

(3.17)

$$J_1 \leqq C \sum_{i=1}^{m+1} \frac{\lambda_{m+1}(w; x_{m+1,i}(w))}{v^{(r,s)}(x_{m+1,i}(w))} \varphi^2(x_{m+1,i}(w)) |P(x_{m+1,i}(w)) \pi(x_{m+1,i}(w))|.$$

With similar arguments we can proceed to evaluate J_2, and so we have

$$J_2 \leqq C \sum_{i=1}^{m} \frac{\lambda_m(\bar{w}; x_{m,i}(\bar{w}))}{v^{(r,s)}(x_{m,i}(\bar{w}))} |P(x_{m,i}(\bar{w}))\pi(x_{m,i}(\bar{w}))|. \tag{3.18}$$

Now we evaluate J_3; by recalling expression (3.16) and that we set $B_r(x) = \prod_{k=1}^{r}(x - z_k)$ we have

$$J_3 = \left| \sum_{k=1}^{r} \prod_{j=1 j\neq k} \frac{1}{z_k - z_j} \frac{P(z_k)\pi(z_k)}{A_s(z_k)p_{m+1}(w; z_k)p_m(\bar{w}; z_k)} \right|.$$

Then by (3.3) we have

$$J_3 \leqq C \sum_{k=1}^{r} \frac{\varphi^2(z_k)}{v^{(r,s)}(z_k)} \frac{|P(z_k)\pi(z_k)|}{m^{2\alpha+2}}. \tag{3.19}$$

On the other hand we recall that if $\sigma \in GDT$ and $\lambda_m(\sigma, x)$ is the mth Christoffel function then (cf. [13], [16] Theorem 6.3.28, p. 120)

$$\lambda_m(\sigma, x) \sim \sigma_m(x) \left(\frac{\sqrt{1-x^2}}{m} + \frac{1}{m^2} \right), \quad |x| \leqq 1 \tag{3.20}$$

(for σ_m see notation (3.1)) and so, in particular, we have $\lambda_{m+1}(w; z_k) \sim m^{-2\alpha-2}$; therefore by (3.19) we obtain

$$J_3 \leqq C \sum_{k=1}^{r} \frac{\lambda_{m+1}(w; z_k)}{v^{(r,s)}(z_k)} \varphi^2(z_k)|P(z_k)\pi(z_k)|. \tag{3.21}$$

In a similar way we can evaluate J_4 and we find

$$J_4 \leqq C \sum_{k=1}^{s} \frac{\lambda_{m+1}(w; y_k)}{v^{(r,s)}(y_k)} \varphi^2(y_k)|P(y_k)\pi(y_k)|. \tag{3.22}$$

By $\{t_i\}_{i=1}^{2m+r+s+1} = \{y_j\}_{j=1}^{s} \cup \{x_{m+1,k}(w)\}_{k=1}^{m+1} \cup \{x_{m,k}(\bar{w})\}_{k=1}^{m} \cup \{z_j\}_{j=1}^{r}$ we can write

$$J_1 + J_2 + J_3 + J_4 \leqq C \sum_{i=1}^{2m+r+s+1} \frac{\lambda_m(u^p, t_i)}{u_m^p(t_i)v^{(r,s)}(t_i)} \bar{w}_m(t_i)|P(t_i)\pi(t_i)|. \tag{3.23}$$

Now, by applying the Hölder inequality to (3.23) we have

$$
\begin{aligned}
J_1 + J_2 + J_3 + J_4 \leqq C & \left[\sum_{i=1}^{2m+r+s+1} \lambda_m(u^p; t_i) |P(t_i)|^p \right]^{\frac{1}{p}} \\
& \times \left[\sum_{i=1}^{2m+r+s+1} \frac{\lambda_m(u^p; t_i) \bar{w}_m^q(t_i)}{u_m^{qp}(t_i) v^{(r,s)q}(t_i)} |\pi(t_i)|^q \right]^{\frac{1}{q}}.
\end{aligned} \tag{3.24}
$$

So from the Marcinkiewicz-type inequality in L_1 for GDT weight functions (see [13],Theorem 6, p. 5) we can deduce that

$$
\begin{aligned}
J_1 + J_2 + J_3 + J_4 \leqq C & \left[\sum_{i=1}^{2m+r+s+1} \lambda_m(u^p; t_i) |P(t_i)|^p \right]^{\frac{1}{p}} \\
& \times \left[\int_{-1}^{1} \left(\frac{\bar{w}(x)}{u^{(p-1)}(x) v^{(r,s)}(x)} \right)^q |\pi(x)|^q dx \right]^{\frac{1}{q}} \\
\leqq C & \left[\sum_{i=1}^{2m+r+s+1} \lambda_m(u^p; t_i) |P(t_i)|^p \right]^{\frac{1}{p}} \left\| \pi \frac{\bar{w}}{u v^{(r,s)}} \right\|_q .
\end{aligned} \tag{3.25}
$$

At this point we have to evaluate the quantity in L_q-norm at the right-hand side of (3.25).

Firstly, we observe that by (3.15) and the definition (3.9), we have

$$
|\pi(t)| \leqq \left| H\left(\frac{QF^{p-1}u^p}{q_m}; t \right) \right| + |Q(t)| \left| H\left((Fq_m)^{p-1} \left(\frac{uv^{(r-1,s-1)}}{w} \right)^p; t \right) \right| .
$$

Therefore, in reason of Remez inequality (3.8) and by the previous inequality, we obtain

$$
\begin{aligned}
\left\| \pi \frac{\bar{w}}{u v^{(r,s)}} \right\|_q & \leqq C \left\| \pi \frac{\bar{w}}{u v^{(r,s)}} \right\|_{L_q(A)} \\
& \leqq C \left\| H\left(\frac{QF^{p-1}u^p}{q_m} \right) \frac{\bar{w}}{u v^{(r,s)}} \right\|_{L_q(A)} \\
& \quad + C \left\| QH\left((Fq_m)^{p-1} \left(\frac{uv^{(r-1,s-1)}}{w} \right)^p \right) \frac{\bar{w}}{u v^{(r,s)}} \right\|_{L_q(A)} .
\end{aligned}
$$

Now by (3.2), (3.14) and (3.15), we have

$$
Q(x) \sim 1 \text{ on } A;
$$

and so, under the assumptions made for u and w, we can use inequality (3.11) and, again for (3.2), (3.14) and (3.15), we finally have

$$\left\| \pi \frac{\bar{w}}{uv^{(r,s)}} \right\|_q \leqq C \| F^{p-1} u^{p-1} \|_{L_q(A)} \leqq C \| F^{p-1} u^{p-1} \|_q = C \| Pu \|_p^{p-1}. \tag{3.26}$$

Hence the first inequality in (2.5) follows from (3.26), (3.25), and (3.13).

For the second inequality in (2.5) we can use the well-known Marcinkiewicz-type inequality due to Lubinsky–Máté–Nevai [11], after recalling again that a polynomial $q \in \mathbb{P}_{3m}$ such that $q \sim u_m^p$ exists [13]; so we have, also from (3.20)

$$\sum_{i=1}^{2m+r+s+1} \lambda_m(u^p, t_i) |P(t_i)|^p \leqq C \sum_{i=1}^{2m+r+s+1} |P(t_i)|^p u_m^p(t_i) \left(\frac{\sqrt{1-t_i^2}}{m} + \frac{1}{m^2} \right)$$

$$\leqq C \sum_{i=1}^{2m+r+s+1} |P(t_i)|^p q^p(t_i) \left(\frac{\sqrt{1-t_i^2}}{m} + \frac{1}{m^2} \right) \leqq C \int_{-1}^{1} |P(t) q(t)|^p dt$$

$$\leqq C \int_{-1}^{1} |P(t)|^p u_m(t)^p dt \leqq C \| P u_m \|_{L_p(A)}^p \leqq C \| Pu \|_p^p,$$

where we used again the Remez inequality (3.8).

Finally let f be a continuous function; we suppose that the first inequality in (2.5) is true. Then we apply it to the polynomial $L_{2m+1,r,s}(w, \bar{w}; f)$. So we have

$$\| L_{2m+1,r,s}(w, \bar{w}; f) u \|_p^p \leqq C \sum_{i=1}^{2m+r+s+1} \lambda_m(u^p; t_i) |f(t_i)|^p$$

$$\leqq C \| f \|_\infty \sum_{i=1}^{2m+r+s+1} \lambda_m(u^p; t_i) = C \| f \|_\infty \int_{-1}^{1} u^p(x) dx \leqq C \| f \|_\infty.$$

This chain of inequalities means that operator $L_{2m+1,r,s}(w, \bar{w})$ considered as a map of C^0 into L_u^p is bounded and so the first condition in (2.4) holds true (cf. [12], Corollary 1, p. 4). So the Theorem is completely proved. □

Proof of Theorem 2.2. Let f be a bounded and measurable function and $q^+, q^- \in \mathbb{P}_{2m+r+s}$, such that

$$q^-(x) \leqq f(x) \leqq q^+(x), \quad x \in [-1, 1].$$

Then we can write

$$\begin{aligned} &\|[f - L_{2m+1,r,s}(w,\bar{w};f)]u\|_p \\ &\leqq \|[f-q^-]u\|_p + \|L_{2m+1,r,s}(w,\bar{w};f-q^-)u\|_p \\ &\leqq \|[q^+ - q^-]\|_p + \|L_{2m+1,r,s}(w,\bar{w};f-q^-)u\|_p. \end{aligned} \tag{3.27}$$

Now from Theorem 2.1 we have

$$\begin{aligned} \|L_{2m+1,r,s}(w,\bar{w};f-q^-)u\|_p^p &\leqq C \sum_{i=1}^{2m+r+s+1} \lambda_m(u^p,t_i)|f-q^-|^p(t_i) \\ &\leqq C \sum_{i=1}^{2m+r+s+1} \lambda_m(u^p,t_i)|q^+-q^-|^p(t_i) \\ &\leqq C\|[q^+-q^-]u\|_p^p. \end{aligned} \tag{3.28}$$

Hence from (3.27) and (3.28) we can write

$$\|[f - L_{2m+1,r,s}(w,\bar{w};f)]u\|_p \leqq C\|[q^+-q^-]u\|_p \tag{3.29}$$

for some positive constant C and for every $q^+, q^- \in \mathbb{P}_{2m+r+s}$, such that $q^-(x) \leqq f(x) \leqq q^+(x)$, $x \in [-1,1]$. Then the Theorem follows making the infimum with respect to q^+, q^- in (3.29). □

Proof of Corollary 2.1. We set

$$f_m(x) = \begin{cases} f(y_1) & \text{if } x \in (-\infty, y_1], \\ f(x) & \text{if } x \in]y_1, z_r[, \\ f(z_r) & \text{if } x \in [z_r, \infty). \end{cases}$$

Then $L_{2m+1,r,s}(w,\bar{w};f) = L_{2m+1,r,s}(w,\bar{w};f_m)$ and

$$\begin{aligned} &\|[f - L_{2m+1,r,s}(w,\bar{w};f)]u\|_p \\ &\leqq \|[f-f_m]u\|_p + \|[f_m - L_{2m+1,r,s}(w,\bar{w};f_m)]u\|_p. \end{aligned} \tag{3.30}$$

The second term at the right-hand side of (3.30) can be evaluated by using Theorem 2.2 and so we have

$$\|[f - L_{2m+1,r,s}(w,\bar{w};f)]u\|_p \leqq \|[f-f_m]u\|_p + C\tilde{E}_{2m+r+s}(f_m)_{u,p}. \tag{3.31}$$

From Theorem 2.2 and Theorem 2.4 in [14]

$$\tilde{E}_{2m+r+s}(f_m)_{u,p} \leqq \frac{C}{m}\|f'\varphi u\|_{L_p([y_1,z_r])}, \tag{3.32}$$

$$\|[f-f_m]u\|_p \leqq C \int_{I_m} |f'(t)|\varphi^{\frac{2}{p}}(t)u(t)dt \tag{3.33}$$

with some positive constant C independent of f and m and where we set $I_m = [-1,1]\backslash[y_1, z_r]$. So the first part of the Corollary follows from (3.31)–(3.33).

Under the assumptions made for f, from [14] (cf. Corollary 2.6, p. 17) we have that

$$\|[f - f_m]u\|_p \leqq \frac{C}{m}\|f'\varphi u\|_p \tag{3.34}$$

and then

$$\|[f - L_{2m+1,r,s}(w, \bar{w}; f)]u\|_p \leqq \frac{C}{m}\|f'\varphi u\|_p. \tag{3.35}$$

So, if we apply (3.35) to the function $F(x) = f(x) - \int_{-1}^{x} P_{2m+r+s-1}(t)dt$, where $P_{2m+r+s-1} \in \mathbb{P}_{2m+r+s-1}$, it results

$$\|[f - L_{2m+1,r,s}(w, \bar{w}; f)]u\|_p \leqq \frac{C}{m}\|[f' - P_{2m+r+s-1}]\varphi u\|_p. \tag{3.36}$$

Then the last part of the Corollary follows by making the infimum with respect to $P_{2m+r+s-1}$ in (3.36). □

Now we need some auxiliary results. First of all it is known that if $u \in GDT$ and $1 \leqq p \leqq \infty$ then (cf. [13])

$$E_{n+1}(f)_{u,p} \leqq \frac{C}{n+1}E_n(f')_{u\varphi,p}. \tag{3.37}$$

Then we recall the following lemma (see [5]).

LEMMA 3.1. *Let $P \in \mathbb{P}_m$, k a positive integer and $u \in GDT$ such that $u \in L_p$ and $\Gamma_j < 1 - \frac{1}{p}$, for $j = 1, \ldots, M$. Then for every function f with $f^{(k-1)} \in AC_{LOC}$ and $f^{(k)}\varphi^k \in L^p_u$,*

$$\|(f - P)^{(k)}\varphi^k u\|_p \leqq C\left[E_{m-k}(f^{(k)})_{\varphi^k u,p} + m^k\|[f - P]u\|_p\right]. \tag{3.38}$$

PROOF OF THEOREM 2.3. First of all we observe that from Corollary 2.1 and by iterating (3.37) we have

$$\|[f - L_{2m+1,r,s}(w, \bar{w}; f)]u\|_p \leqq \frac{C}{m^i}E_{2m+r+s-i}(f^{(i)})_{\varphi^i u,p}.$$

So by Lemma 3.1 to $L_{2m+1,r,s}(w, \bar{w}; f)$ we can write

$$\|[f^{(i)} - L^{(i)}_{2m+1,r,s}(w, \bar{w}; f)]\varphi^i u\|_p \leqq CE_{2m+r+s-i}(f^{(i)})_{\varphi^i u,p}. \tag{3.39}$$

On the other hand, by iterating (3.37) once again, we obtain

$$
\begin{aligned}
E_{2m+r+s-i}(f^{(i)})_{\varphi^i u,p} &\leqq \frac{C}{m^{k-i}} E_{2m+r+s-k}(f^{(k)})_{\varphi^k u,p} \\
&\leqq \frac{C}{m^{k-i}} \|f^{(k)}\varphi^k u\|_p
\end{aligned} \tag{3.40}
$$

for all $i = 1, \dots, k$. So Theorem 2.3 follows immediately from (3.39) and (3.40). □

Proof of Theorem 2.4. The proof of (2.13) follows directly by Theorem 2.3; moreover we obtain (2.12) again by Theorem 2.3 (in the case $i = k$) observing that for every $k \leqq t$ it results

$$
\begin{aligned}
\left\| L^{(k)}_{2m+r+s+1}(w, \bar{w}; f)\varphi^k u \right\|_p &\leqq \left\| \left[f - L^{(k)}_{2m+r+s+1}(w, \bar{w}; f) \right] \varphi^k u \right\|_p \\
&\quad + \|f^{(k)}\varphi^k u\|_p.
\end{aligned}
$$

□

Acknowledgement. The authors are grateful to Prof. Giuseppe Mastroianni for his helpful suggestions. They also thank the referee for the accurate reading of the manuscript and the precious remarks.

REFERENCES

[1] Balázs, K. and Kilgore, T., Simultaneous approximation of derivatives by interpolation, *Approximation Theory VI*, Vol. 1 (College Station, TX, 1989), C. K. Chui, L. L. Schumaker, J. D. Ward, eds., Academic Press, Boston, MA, 1989, 25–27. *Zbl* **707**. 41025

[2] Berthold, D., Hoppe, W. and Silbermann, B., A fast algorithm for solving the generalized airfoil equation, *Orthogonal polynomials and numerical methods. J. Comput. Appl. Math.* **43** (1992), 185–219. *MR* **93k**:65106

[3] Capobianco, M. R. and Mastroianni, G., Uniform boundedness of Lagrange operator in some weighted Sobolev-type space, *Math. Nachr.* **187** (1997), 61–77. *MR* **98e**:41001

[4] Capobianco, M. R. and Russo, M. G., Extended interpolation in some Sobolev-type spaces, *Indian J. Math.* **37** (1995), 191–206. *MR* **98d**:41003

[5] Criscuolo, G. and Mastroianni, G., Fourier and Lagrange operators in some weighted Sobolev-type spaces, *Acta Sci. Math. (Szeged)* **60** (1995), 131–148. *MR* **96e**:41006

[6] Criscuolo, G., Mastroianni, G. and Nevai, P., Mean convergence of derivatives of extended Lagrange interpolation with additional nodes, *Math. Nachr.* **163** (1993), 73–92. *MR* **94g**:65011

[7] Criscuolo, G., Mastroianni, G. and Occorsio, D., Convergence of extended Lagrange interpolation, *Math. Comp.* **55** (1990), 197–212. *MR* **91c**:65008

[8] Criscuolo, G., Mastroianni, G. and Vértesi, P., Pointwise simultaneous convergence of extended Lagrange interpolation with additional knots, *Math. Comp.* **59** (1992), 515–531. *MR* **93a**:41003

[9] Ditzian, Z. and Totik, V., *Moduli of smoothness*, Springer Series in Computational Mathematics, 9, Springer-Verlag, New York – Berlin, 1987. *MR* **89h**:41002

[10] HUNT, R., MUCKENHOUPT, B. and WHEEDEN, R., Weighted norm inequalities for the conjugate function and Hilbert transform, *Trans. Amer. Math. Soc.* **176** (1973), 227–251. *MR* **47** #701

[11] LUBINSKY, D. S., MÁTÉ, A. and NEVAI, P., Quadrature sums involving pth powers of polynomials, *SIAM J. Math. Anal.* **18** (1987), 531–544. *MR* **89h**:41058

[12] MASTROIANNI, G., Approximation of functions by extended Lagrange interpolation, *Approximation and computation* (West Lafayette, IN, 1993), Internat. Ser. Numer. Math., 119, Birkhäuser Boston, Boston, MA, 1994, 409–420. *MR* **96c**:41007

[13] MASTROIANNI, G., Some weighted polynomial inequalities, Proceedings of the International Conference on Orthogonality, Moment Problems and Continued Fractions (Delft, 1994), *J. Comput. Appl. Math.* **65** (1995), 279–292. *MR* **96m**:41011

[14] MASTROIANNI, G. and VÉRTESI, P., Weighted L_p error of Lagrange interpolation, *J. Approx. Theory* **82** (1995), 321–339. *MR* **96f**:41004

[15] MIKHLIN, S. G. and PRÖSSDORF, S., *Singular integral operators*, Springer-Verlag, Berlin – Heidelberg – New York, 1986. *MR* **88e**:47097

[16] NEVAI, P. G., Orthogonal polynomials, *Mem. Amer. Math. Soc.* no. **213**, Amer. Math. Soc., Providence, RI, 1979. *MR* **80k**:42025

[17] NEVAI, P., Bernstein's inequality in L^p for $0 < p < 1$, *J. Approx. Theory* **27** (1979), 239–243. *MR* **80m**:41009

[18] SLOAN, I. H. and STEPHAN, E. P., Collocation with Chebyshev polynomials for Symm's integral equation on an interval, *J. Austral. Math. Soc. Ser. B* **34** (1992), 199–211. *MR* **93g**:45014

[19] SZABADOS, J., On the convergence of the derivatives of projection operators, *Analysis* **7** (1987), 349–357. *MR* **89h**:41010

(Received February 15, 1996)

ISTITUTO PER APPLICAZIONI DELLA MATEMATICA
C.N.R.
VIA PIETRO CASTELLINO 111
I–80131 NAPOLI
ITALY

capobian@iamna.iam.na.cnr.it

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DELLA BASILICATA
VIA NAZARIO SAURO 85
I–85100 POTENZA
ITALY

russo@unibas.it

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 99–111

SATURATION THEOREM FOR QUASI-PROJECTIONS

K. DZIEDZIUL

Abstract

We shall study properties of the box spline operators: quasi-interpolations and quasi-projections. We find relation between them. We prove the saturation theorem.

Introduction

In this paper we compare the properties of spline operators: quasi-interpolations and quasi-projections. Quasi-interpolations are examined in details in [B], see also [CD], [BHR]. The saturation theorem is proved in [DM1]. The Ciesielski-Dürmeyer operators are the simplest examples of quasi-projections, see [C1-2]. Z. Ciesielski applied them to density estimation. The definition of the quasi-projections appears in [BD2].

1. Spline operators

Let us review some standard facts on the box splines. Let $V = \{v_1, v_2, \ldots, v_n\}$ denote a set of not necessarily distinct vectors in $Z^d \setminus \{0\}$, such that

$$\operatorname{span}\{V\} = R^d.$$

We call such a set admissible. The box spline corresponding to V (denoted by $B(\cdot|V)$ or B_V) is defined by requiring that

$$\int_{R^d} f(x)B(x|V)\,dx = \int_{[0,1]^n} f(Vu)\,du$$

holds for any continuous function f on R^d.

We use standard convolution notation, see [B], [BR], [CD]. If $\{a(\alpha)\}$ and $\{b(\alpha)\}$, where $\alpha \in Z^d$, are two given sequences then the discrete convolution product is defined by

$$a * b(\alpha) = \sum_{\beta \in Z^d} a(\beta)b(\alpha - \beta).$$

1991 *Mathematics Subject Classification.* Primary 41A15, 41A63, 41A25.
Key words and phrases. Box splines, polynomials, saturation theorem.

Without danger of confusion, the semi-discrete convolution and the convolution are both denoted by $*$, i.e.,

$$\phi * a = \sum_{\alpha \in Z^d} a(\alpha)\phi(\cdot - \alpha),$$

$$\phi * \psi(x) = \int_{R^d} \phi(y)\psi(x-y)\,dy.$$

Moreover, we use the semi-discrete convolution operator

$$\phi *' f = \phi * f_{|Z^d} = \sum_{\alpha \in Z^d} f(\alpha)\,\phi(\cdot - \alpha).$$

We use the abbreviation

$$f_| = f_{|Z^d},$$

too. For $1 \leqq p < \infty$ and $k \in N$, W_p^k denotes the Sobolev space on R^d [S] and by W_∞^k we mean the closure of smooth functions with compact support in the norm

$$\|f\|_{\infty,k} = \sup_{x \in R^d} \sup_{|\alpha| \leqq k} |D^\alpha f(x)|,$$

where

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \quad \alpha = (\alpha_1, \ldots, \alpha_d), \quad |\alpha| = \alpha_1 + \cdots + \alpha_d.$$

By $\|\cdot\|_p$ we denote the standard L^p norm on R^d:

$$\|f\|_p = \left(\int_{R^d} |f(x)|^p\,dx \right)^{1/p}.$$

The inner product in $L^2(R^d)$ is denoted by

$$(f,g)_{R^d} = \int_{R^d} f(x)\overline{g(x)}\,dx.$$

By the Fourier transform we mean

$$\widehat{f}(x) = \int_{R^d} f(t)e^{-2\pi i x t}\,dt.$$

Now, let us introduce the notion of quasi-interpolation of Ch. Chui and H. Diamond [CD]. For given admissible V, let us introduce the function

$$N_V(x) = B(x + c_V|V), \tag{1.1}$$

which is symmetric with respect to the origin, and where

$$c_V = \frac{1}{2}\sum_{v\in V} v. \tag{1.2}$$

Consider the sequence $n_V = \{n_V(\alpha)\}$, where

$$n_V(\alpha) = \begin{cases} 1 - N_V(0) & \text{for } \alpha = 0 \\ -N_V(\alpha) & \text{for } \alpha \neq 0, \end{cases} \tag{1.3}$$

$\alpha \in Z^d$, and let

$$n_V^k = \underbrace{n_V * \cdots * n_V}_{k}. \tag{1.4}$$

Moreover, denote $\delta = \{\delta(\alpha)\}$, with

$$\delta(\alpha) = \begin{cases} 1 & \text{for } \alpha = 0 \\ 0 & \text{for } \alpha \neq 0. \end{cases}$$

DEFINITION 1.5. Let

$$m_{V,\varrho} = \delta + n_V + n_V^2 + \cdots + n_V^\varrho, \tag{1.6}$$

and f be a continuous function. The operator of quasi-interpolation $Q^{(V,\varrho)}$ is defined as follows:

$$\begin{aligned} Q^{(V,\varrho)} f &= N_V *' (f * m_{V,\varrho}) = \sum_{\alpha\in Z^d} (f * m_{V,\varrho})(\alpha) N_V(\cdot - \alpha), \\ Q_h^{(V,\varrho)} &= \sigma_h \circ Q^{V,V,\rho} \circ \sigma_{1/h}, \end{aligned} \tag{1.7}$$

where

$$\sigma_h f(\cdot) = f\left(\frac{\cdot}{h}\right). \tag{1.8}$$

For an admissible set V let

$$\varrho_V = \max\{\, r : \bigvee_{X\subset V} \#X = r,\ \operatorname{span}\{V \setminus X\} = R^d \,\}. \tag{1.9}$$

In [DM1] Dahmen and Micchelli considered a similar operator

$$Qf = \sum_{\alpha \in Z^d} f(\alpha + c_V) B_V(\cdot - \alpha)$$

and for $h > 0$

$$Q_h^{(V,\varrho)} = \sigma_h \circ Q^{V,V,\rho} \circ \sigma_{1/h}.$$

They proved the saturation theorem stating that if $f \in C^2(R^d)$ and $\varrho_V \geqq 2$ then

$$\lim_{h \to 0} h^{-2}(Q_h^{(V,0)} f - f) = \frac{1}{24} \sum_{v \in V} D_v^2 f, \tag{1.10}$$

where D_v is the directional derivative and with the uniform convergence on compact sets.

For the admissible sets V_1 and V_2, we define an admissible set

$$Y = V_1 \cup (-V_2),$$

where $-V_2$ is an admissible set consisting of vectors w such that

$$w \in -V_2 \Leftrightarrow -w \in V_2.$$

DEFINITION 1.11. The Ciesielski–Dürmeyer operator is defined as follows:

$$Q^{(V_1,V_2)} f = \sum_{\alpha \in Z^d} (f, B(\cdot - \alpha - c_Y | V_2))_{R^d} B(\cdot - \alpha | V_1), \tag{1.12}$$

where $c_Y = \frac{1}{2} \sum_{v \in Y} v$ and

$$Q_h^{(V_1,V_2)} = \sigma_h \circ Q^{(V_1,V_2)} \circ \sigma_{1/h}.$$

The saturation theorem for the tensor product B-spline operators is due to Ciesielski, see [C1-2]. We can formulate the theorem for general box-splines.

THEOREM 1.13. *Assume that V_1 and V_2 are admissible and $\varrho_{V_1} \geqq 2$, $\varrho_{V_2} \geqq 1$. Then for all functions f from Sobolev spaces $f \in W_p^2(R^d)$ and $1 \leqq p \leqq \infty$,*

$$\left\| \frac{Q_h^{(V_1,V_2)} f - f}{h^2} - \frac{1}{24} \sum_{v \in Y} D_v^2 f \right\|_p = o(1). \tag{1.14}$$

Note that the expressions (1.14) and (1.10) are similar. According to us the main reason is that for suitable chosen sets of directions V, X the corresponding quasi-projection and quasi-interpolation coincide on some polynomials.

DEFINITION 1.15. Let $V \subset Z^d \setminus \{0\}$ be admissible, and

$$Y = V_1 \cup (-V_2).$$

Moreover, put $n_X = \delta - B_{X|}$ and

$$m_{X,\varrho} = \delta + n_X + \cdots + n_X^{\varrho}. \tag{1.16}$$

The quasi-projections $Q_h^{(V,V,\varrho)}$ are defined as follows: for $f \in L^p(R^d)$, $1 \leqq p \leqq \infty$

$$Q^{(V,V,\varrho)}(f) = \sum_{\alpha \in Z^d} (f, B(\cdot - \alpha|V) * m_{X,\varrho})_{R^d} B(\cdot - \alpha|V), \tag{1.17}$$

$$Q_h^{(V,V,\varrho)} = \sigma_h \circ Q^{(V,V,\varrho)} \circ \sigma_{1/h}. \tag{1.18}$$

2. Saturation theorem

For a compactly supported function φ we define $S(\varphi)$ to be infinite span of the integer translates of φ:

$$S(\varphi) = \operatorname{span}\{\, \varphi(\cdot - \alpha) \,:\, \alpha \in Z^d \,\}. \tag{2.1}$$

As usual, let Π be the space of all d-variate polynomials, Π_k the subspace of Π of total degree at most k, and

$$\Pi(\varphi) = \Pi \cap S(\varphi).$$

It is known that [H]

$$\Pi_{\varrho V} \subset \Pi(B_V \cdot). \tag{2.2}$$

C. de Boor showed [B] that for all $f \in S(\varphi)$

$$\varphi *' f = f *' \varphi, \tag{2.3}$$

and

$$B_V \cdot *' \Pi(B_V \cdot) = \Pi(B_V \cdot). \tag{2.4}$$

Further, C. de Boor and A. Ron proved (see [BR], [RS]) the following

LEMMA 2.5. *For a polynomial* P

$$B_V *' P \in \Pi \Leftrightarrow B_V *' P = B_V * P. \tag{2.6}$$

Hence for all $P \in \Pi(B_V)$:

$$B_V *' P = B_V * P. \tag{2.7}$$

It is clear that (2.4)–(2.7) hold also for N_V.

THEOREM 2.8. *For all polynomials* $P \in \Pi_{\varrho_V}$ *and for all* $0 \leqq \varrho$,

$$Q^{(V,V,\varrho)}(P) = Q^{(X,\varrho)}(P). \tag{2.9}$$

Moreover, if $k \leqq \varrho_V$ *and* $k \leqq 2\varrho + 1$, *then*

$$Q^{(V,V,\varrho)} P = P \quad \text{for all} \quad P \in \Pi_k. \tag{2.10}$$

PROOF. Let

$$\check{f}(x) = f(-x)$$

and for a sequence $a = \{a(\alpha)\}$

$$\check{a}(\alpha) = a(-\alpha).$$

Then

$$(f, g(\cdot - \alpha))_{R^d} = f * \bar{\check{g}}(\alpha). \tag{2.11}$$

The function B_X is symmetric with respect to the origin. This implies that

$$B_X = N_X, \tag{2.12}$$

and the sequence n_X is symmetric with respect to the origin. Consequently, also $m_{X,\varrho}$ is symmetric with respect to the origin. So

$$(B_V * m_{X,\varrho})^{\check{}} = \check{B}_V * \check{m}_{X,\varrho} = \check{B}_V * m_{X,\varrho}. \tag{2.13}$$

From (2.11) and (2.13) it follows that

$$(P, B(\cdot - \alpha) * m_{X,\varrho})_{R^d} = P * (\check{B}_V * m_{X,\varrho})(\alpha).$$

Hence

$$Q^{(V,V,\varrho)}(P) = B_V *' (P * (\check{B}_V * m_{X,\varrho})) = B_V *' (\check{B}_V * (P * m_{X,\varrho})).$$

By the definition of the box splines, we get $B_{-V} = \check{B}_V$. Therefore

$$Q^{(V,V,\varrho)}(P) = B_V *' (B_{-V} * (P * m_{X,\varrho})).$$

Moreover, since $\varrho_{-V} = \varrho_V$, we have

$$\Pi_{\varrho_V} = \Pi_{\varrho_{-V}}. \tag{2.14}$$

From (2.4), (2.7), and (2.14) we conclude that for $P \in \Pi_{\varrho_V}$

$$B_{-V} * (P * m_{X,\varrho}) \in \Pi_{\varrho_V}.$$

Now from (2.7)

$$Q^{(V,V,\varrho)}(P) = B_V * B_{-V} * (P * m_{X,\varrho}).$$

We observe that

$$B_V * B_{-V} = B_X,$$

thus from (2.7) (2.12)

$$Q^{(V,V,\varrho)}(P) = B_X * (P * m_{X,\varrho}) = N_X *' (P * m_{X,\varrho}) = Q^{(X,\varrho)} P,$$

i.e., (2.9) is proved. By definition

$$\widehat{B_X}(\xi) = \prod_{v \in X} \frac{1 - e^{-2\pi i \xi v}}{2\pi i \xi v}.$$

Note that for β such that $|\beta| \leqq \varrho_V$ we have

$$D^\beta \widehat{B_V}(\alpha) = 0$$

for all $\alpha \in Z^d \setminus \{0\}$. Now, by Theorem 1, from [CD] we have

$$Q^{(X,\varrho)} P = P$$

for all $P \in \Pi_k$ with $k \leqq \varrho_V$, $k \leqq 2\varrho + 1$, which gives (2.10). □

The operator $Q^{(V,V,\varrho)}$ is local, i.e., for each $x \in R^d$, the value of $Q^{(V,V,\varrho)} f(x)$ depends only on the values of f on the bounded set $\Lambda_x \subset \{y \in R^d : \|y - x\| < s\}$. Let us estimate s. Denote by D the diameter of support of B_X. Then $s \leqq (\varrho + 2)D$.

This implies that there are constants s and C such that for any open set Ω

$$\|Q_h^{(V,V,\varrho)} f(x)\|_p(\Omega) \leqq C \|f(x)\|_p(\Omega_{hs}),$$

where

$$\Omega_{hs} = \{x \in R^d : \exists_{y \in \Omega} |x - y| < hs\}$$

and

$$\|f\|_p(\Omega) = \left(\int_\Omega |f(x)|^p \, dx \right)^p.$$

The general results on the order of approximation by local, polynomial-reproducing box-spline operators (see [K1-2], [BD], Proposition 3.4 [BHR]), and Theorem 2.8 imply the following result concerning the order of approximation by quasi-projections.

THEOREM 2.15. *Let $V \subset Z^d \setminus \{0\}$ be admissible and $\varrho \geqq 0$. Let $r = 2\varrho + 2$ in case $2\varrho + 1 \leqq \varrho_V$, and $r := \varrho_V + 1$ otherwise. Then for each p, $1 \leqq p \leqq \infty$ there is a constant C_p such that for all $f \in W_p^r$*

$$\|Q_h^{(V,V,\varrho)} f(x) - f(x)\|_p \leqq C_p h^r \sum_{|\beta|=r} \|D^\beta f\|_p. \tag{2.16}$$

In the sequel, the following result is needed.

LEMMA 2.17. *Let $2\varrho + 1 < \varrho_X$. Then for all polynomials P such that $\deg P \leqq 2\varrho + 2$*

$$Q^{(X,\varrho)} P = P + A_P, \tag{2.18}$$

where A_P is a constant depending on P.

PROOF. De Boor's formula

$$N_X *' P = N_X * P$$

implies that if $P \in \Pi(N_X)$ then $Q^{(X,\varrho)} P$ is a polynomial as well. Moreover, by the definition of quasi-interpolation

$$Q^{(X,\varrho)} P_| = (N_X *' (P * m_{X,\varrho}))_| =$$

$$= (N_X)_| * P_| * m_{X,\varrho} = (\delta - n_X) * (\delta + n_X + n_X^2 + \cdots + n_X^\varrho) * P_| =$$

$$(\delta - n_X) * (\delta + n_X + n_X^2 + \cdots + n_X^\varrho) * P_| = P_| - n_X^{\varrho+1} * P_|.$$

It is known that the operator

$$n_X : \Pi \to \Pi$$

is degree reducing, see [B], [CD], more precisely

$$\deg(P * n_X) + 2 \leqq \deg(P).$$

Thus if $\deg P \leqq 2\varrho + 2$ then

$$\deg(P * n_X^{\varrho+1}) = 0.$$

Hence

$$A_P = -P * n_X^{\varrho+1} = -P_| * n_X^{\varrho+1} \tag{2.19}$$

is a constant. This finishes the proof. □

Our next goal is to determine the constant in (2.19) for the monomials. Recall that for sufficiently regular functions the Poisson formula implies

$$\sum_{\alpha \in Z^d} f(\alpha) = \sum_{\alpha \in Z^d} \widehat{f}(\alpha), \tag{2.20}$$

see [SW].

Let us denote by A_β the constant (2.19) for the monomial

$$P = x^\beta = x^{\beta_1} \cdot \ldots \cdot x^{\beta_d}$$

with $|\beta| = 2\varrho + 2$, where $\beta = (\beta_1, \ldots, \beta_d)$. By Leibniz's formula

$$n_X^{\varrho+1} = (\delta - (N_X)_|)^{\varrho+1} = \delta + \sum_{k=1}^{\varrho+1} (-1)^k \binom{\varrho+1}{k} (N_X)_|^k.$$

From this and (2.19) we obtain

$$A_\beta = -P_| * n_X^{\varrho+1}(0) = -(P(0) + \sum_{k=1}^{\varrho+1} (-1)^k \binom{\varrho+1}{k} (N_X)_|^k * P_|(0)). \tag{2.21}$$

Let kX be an admissible set consisting of the vectors of X with multiplicity of k. Formula (2.7) implies now

$$(N_X)_| * P_| = (N_X *' P)_| = (N_X * P)_|,$$

hence

$$(N_X)|^2 * P_| = (N_X)_| * ((N_X)_| * P_|) = (N_X)_| * (N_X * P)_| = (N_X *' (N_X * P))_|.$$

Since $N_X * P \in \Pi_{\varrho X}$, we have

$$\begin{aligned}(N_X)|^2 * P_| &= (N_X * (N_X * P))_| = ((N_X * N_X) * P)_| \\ &= (N_{2X} * P)_| = (N_{2X} *' P)_| = (N_{2X})_| * P_|.\end{aligned}$$

Thus

$$(N_X)|^k * P_| = (N_{kX})_| * P_|.$$

Since $P(0) = 0$ we conclude that

$$A_\beta = -\sum_{k=1}^{\varrho+1} (-1)^k \binom{\varrho+1}{k} (N_{kX})_| * P_|(0).$$

The functions N_{kX} are symmetric. Therefore

$$(N_{kX})_| * P_|(0) = \sum_{\alpha \in Z^d} P(\alpha) N_{kX}(0 - \alpha) = \sum_{\alpha \in Z^d} P(\alpha) N_{kX}(\alpha). \tag{2.22}$$

Applying the Poisson formula for the functions $f = PN_{kX}$ we can rewrite (2.21) as

$$(N_{kX})_| * P_|(0) = \frac{1}{(2\pi i)^{2\varrho+2}} \sum_{\alpha \in Z^d} D^\beta \widehat{N}_{kX}(\alpha) = \frac{1}{(2\pi i)^{2\varrho+2}} D^\beta \widehat{N}_{kX}(0),$$

where

$$\widehat{N}_X(\xi) = \prod_{v \in X} \frac{\sin(\pi \xi v)}{\pi \xi v}.$$

Thus, we get

$$A_\beta = -\frac{1}{(2\pi i)^{2\varrho+2}} \sum_{k=1}^{\varrho+1} (-1)^k \binom{\varrho+1}{k} D^\beta \widehat{N}_{kX}(0)$$

$$= \frac{1}{(2\pi)^{2\varrho+2}} \sum_{k=1}^{\varrho+1} (-1)^{k+\varrho} \binom{\varrho+1}{k} D^\beta \widehat{N}_{kX}(0).$$

Denote $\beta! = \beta_1! \cdots \beta_d!$. Now, we are ready to state the saturation theorem for quasi-projections.

THEOREM 2.23. *Let $V \subset Z^d \setminus \{0\}$ be admissible, $2\varrho + 1 < \varrho_V$ and $1 \leqq p \leqq \infty$. Then for any $f \in W_p^{2\varrho+2}$*

$$\left\| \frac{Q_h^{(V,V,\varrho)} f - f}{h^{2\varrho+2}} - \sum_{|\beta|=2\varrho+2} \frac{1}{\beta!} A_\beta D^\beta f \right\|_p = o(1). \tag{2.24}$$

PROOF. We first prove (2.24) for $p = \infty$. Let f be any compactly supported function such that $f \in C^{2\varrho+2}$. Fix x. By Taylor's formula

$$f(y) = P_x(y) + R(x, y), \tag{2.25}$$

where P is the polynomial of the degree $2\varrho + 1$, and $R(x, y)$ is Taylor's remainder. Since

$$P_x(x) = f(x)$$

and from (2.10)

$$Q_h^{(V,V,\varrho)} P_x = P_x$$

we see that

$$Q_h^{(V,V,\varrho)} P_x(x) = f(x),$$

which gives

$$\frac{Q_h^{(V,V,\varrho)} f(x) - f(x)}{h^{2\varrho+2}} = \frac{Q^{(V,V,\varrho)} R(x, h\cdot)(x/h)}{h^{2\varrho+2}}.$$

Now, it suffices to show that

$$\sup_{x \in R^d} \left| \frac{Q^{(V,V,\varrho)} R_\beta(x, h\cdot)(x/h)}{h^{2\varrho+2}} - \frac{1}{\beta!} A_\beta D^\beta f(x) \right| \to 0 \quad \text{as } h \to 0,$$

where

$$R_\beta(x,y)=\frac{1}{\beta!}D^\beta f(\theta)(y-x)^\beta,\qquad |\beta|=2\varrho+2$$

and $\theta=\theta(x,y)$ is an intermediate point between x and y in Taylor's formula (2.25). For fixed x and h consider the polynomial

$$p_{h,x}(y)=(hy-x)^\beta=h^{2\varrho+2}y^\beta+q_{h,x}(y).$$

Note that $\deg q_{h,x}<2\varrho+2$ and it follows from the formulae (2.10) and (2.18) that $A_{q_{h,x}}=0$. This gives

$$A_{p_{h,x}}=h^{2\varrho+2}A_\beta$$

and

$$Q^{(V,V,\varrho)}(h\cdot-x)^\beta(x/h)=(h\frac{x}{h}-x)^\beta+A_\beta=h^{2\varrho+2}A_\beta.$$

Let us consider the function

$$T_\beta(x,y)=\frac{1}{\beta!}D^\beta f(x)(y-x)^\beta,\quad |\beta|=2\varrho+2.$$

Then

$$\frac{Q_h^{(V,V,\varrho)}T_\beta(x,\cdot)(x)}{h^{2\varrho+2}}=\frac{1}{\beta!}D^\beta f(x)\frac{Q^{(V,V,\varrho)}(h\cdot-x)^\beta(x/h)}{h^{2\varrho+2}}=\frac{1}{\beta!}A_\beta D^\beta f(x).$$

Note that, since the operator $Q^{(V,V,\varrho)}f$ is local, there are s and C independent of x and f such that

$$|Q^{(V,V,\varrho)}f(x)|\leqq C\sup_{|y-x|<s}|f(y)|.$$

Using these calculations we get

$$\left|\frac{Q^{(V,V,\varrho)}R_\beta(x,h\cdot)(x/h)}{h^{2\varrho+2}}-\frac{1}{\beta!}A_\beta D^\beta f(x)\right|=\left|\frac{Q^{(V,V,\varrho)}[R_\beta-T_\beta](x,h\cdot)(x/h)}{h^{2\varrho+2}}\right|$$

$$\leqq C\frac{\sup_{|y-x/h|<s}|R_\beta(x,hy)-T_\beta(x,hy)|}{h^{2\varrho+2}}$$

$$\leqq C\sup_{|z-x|<hs}|D^\beta f(z)-D^\beta f(x)|\frac{\sup_{|z-x|<hs}|(z-x)^\beta|}{h^{2\varrho+2}}$$

$$=Cs^{2\varrho+2}\sup_{|z-x|<hs}|D^\beta f(z)-D^\beta f(x)|.$$

Since the functions $D^\beta f$ are uniformly continuous, this gives (2.24) for $p=\infty$ and functions f of compact support. The result (2.24) for $1 \leqq p < \infty$ and compactly supported $f \in W_p^{2\varrho+2} \cap C^{2\varrho+2}$ follows from the just proved result for $p=\infty$ and the locality of quasi-projections. Moreover, Theorem 2.15 implies that the operators

$$K_h(f) = \frac{Q_h^{(V,V,\varrho)} f(x) - f(x)}{h^{2\varrho+2}} - \sum_{|\beta|=2\varrho+2} \frac{1}{\beta!} A_\beta D^\beta f(x)$$

are bounded on $W_p^{2\varrho+2}$, more precisely

$$\|K_h(f)\|_p \leqq C \sum_{|\beta|=2\varrho+2} \|D^\beta f\|_p.$$

The functions $f \in W_p^{2\varrho+2} \cap C^{2\varrho+2}$ with compact support are dense in $W_p^{2\varrho+2}$. By the triangle inequality

$$\|K_h(f)\|_p \leqq \|K_h(f-f_\epsilon)\|_p + \|K_h(f_\epsilon)\|_p.$$

Choosing appropriate functions f_ϵ we conclude that (2.24) holds for the functions from the Sobolev space $W_p^{2\varrho+2}$. □

REFERENCES

[BD1] Beška, M. and Dziedziul, K., Notes on Strang–Fix theorem, *Proc. Open Problems in Approximation Theory*, editor B. Bojanov, SCT Publishing, 1994, 16–24.

[BD2] Beška, M. and Dziedziul, K., Multiresolution and approximation and Hardy spaces, *J. Approx. Theory* **88** (1997), 154–167. *MR* **98f**:42027

[B] Boor, C. de, The polynomials in the linear span of integer translates of a compactly supported function, *Constr. Approx.* **3** (1987), 199–208. *MR* **88e**:41054

[BH] Boor, C. de and Höllig, K., *B*-splines from parallelepipeds, *J. Analyse Math.* **42** (1982/83), 99–115. *MR* **86d**:41008

[BHR] Boor, C. de, Höllig, K. and Riemenschneider, S., *Box splines*, Applied Mathematical Sciences, Vol. 98, Springer-Verlag, New York, 1993. *MR* **94k**:65004

[BR] Boor, C. de and Ron, A., The exponentials in the span of the multi-integer translates of a compactly supported function; quasi-interpolation and approximation order, *J. London Math. Soc.* (2) **45** (1992), 519–535. *MR* **94e**:41021

[CD] Chui, C. K. and Diamond, H., A natural formulation of quasi-interpolation by multivariate splines, *Proc. Amer. Math. Soc.* **99** (1987), 643–646. *MR* **88g**:41010

[CDR] Chui, C. K., Diamond, H. and Raphael, L. A., Interpolation by multivariate splines, *Math. Comp.* **51** (1987), 203–218. *MR* **89j**:41002

[C1] Ciesielski, Z., Nonparametric polynomial density estimation, *Probab. Math. Statist.* **9** (1988), 1–10. *MR* **89i**:62056

[C2] Ciesielski, Z., Asymptotic nonparametric spline density estimation in several variables, *Multivariate approximation and interpolation* (Duisburg, 1989), Internat. Ser. Numer. Math., Vol. 94, Birkhäuser-Verlag, Basel, 1990, 25–53. *MR* **92i**: 62067

[DM1] Dahmen, W. and Micchelli, C. A., Convexity of multivariate Bernstein polynomials and box spline surfaces, *Studia Sci. Math. Hungar.* **23** (1988), 265–287. *MR* **90g**:41005

[DM2] DAHMEN, W. and MICCHELLI, C. A., Translates of multivariate splines, *Linear Algebra Appl.* **52/53** (1983), 217–234. *MR* **85e**:41033

[H] HÖLLIG, K., Multivariate splines, *Approximation Theory* (New Orleans, LA, 1986), Proc. Sympos. Appl. Math., **36**, Amer. Math. Soc., Providence, RI, 1986, 103–127. *MR* **88c**:41020

[K1] KOWALSKI, J. K., Application of box splines to the approximation of Sobolev spaces, *J. Approx. Theory* **61** (1990), 53–73. *MR* **91b**:46033

[K2] KOWALSKI, J. K., A method of approximation of Besov spaces, *Studia Math.* **96** (1990), 183–193. *MR* **91h**:41014

[RS] RON, A. and SIVAKUMAR, N., The approximation order of box spline spaces, *Proc. Amer. Math. Soc.* **117** (1993), 473–482. *MR* **93d**:41010

[S] STEIN, E. M., *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30, Princeton University Press, Princeton, NJ, 1970. *MR* **44** #7280

[SW] STEIN, E. M. and WEISS, G., *Introduction to Fourier analysis on Euclidean spaces*, Princeton Mathematical Series, No. 32, Princeton University Press, Princeton, NJ, 1971. *MR* **46** #4102

(Received April 20, 1996)

WYDZIAŁ FIZYKI TECHNICZNEJ
I MATEMATYKI STOSOWANEJ
POLITECHNIKA GDAŃSKA
UL. G. NARUTOWICZA 11/12
PL–80-952 GDAŃSK–WRZESZCZ
POLAND

kdz@mif.pg.gda.pl

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 113–131

A PROBLEM OF ERDŐS–RÉVÉSZ ON ONE-DIMENSIONAL RANDOM WALKS

Z. SHI

Summary

The following problem is raised by Erdős and Révész [5]: let $\nu(n)$ be the time necessary for a simple symmetric random walk on the line at the n-th step to visit a new point, what can be said of the limsup behaviour of $\nu(n)$? It is established in [5] that $A \stackrel{\text{def}}{=} \limsup_{n\to\infty} \nu(n)/n(\log\log n)^2 \in [1/4\pi^2,\ 16/\pi^2]$. The exact value of A is obtained by Csáki [2]. In this paper, we present an integral test characterizing the upper functions of $\nu(n)$, and furthermore study the corresponding problem for a large class of real-valued random walks as well as for linear Wiener processes.

1. Introduction

Let $\{S_n\}_{n\geq 0}$ be a simple symmetric random walk in $\mathbb{Z}^d$ (but, very soon, $\mathbb{Z}$) with $S_0 = 0$, i.e. at each step, the random walk has probability $1/(2d)$ to visit each of its $(2d)$ neighbouring points. Erdős and Révész [5] raise the problem of investigating the "limsup" behaviour of

$$\nu(n) \stackrel{\text{def}}{=} \min\Big\{k \geq 1 : S_{n+k} \notin \{S_0, S_1, \ldots, S_n\}\Big\}. \tag{1.1}$$

In words, the question can be formulated as: how long does it take a simple symmetric random walk to visit a new site? This problem is very challenging, as is pointed out in [5]. Although not completely solved in [5], some interesting estimates are obtained by Erdős and Révész. Let us recall their result for dimension 1.

THEOREM 1.A (Erdős and Révész [5]). *Let $\nu(n)$ be as in* (1.1) *and let $d = 1$. Then*

$$\frac{1}{4\pi^2} \leq \limsup_{n\to\infty} \frac{\nu(n)}{n(\log\log n)^2} \leq \frac{16}{\pi^2} \quad a.s.. \tag{1.2}$$

The exact value of the limsup expression in (1.2) (in case $d = 1$) is determined by Csáki [2]:

$$\limsup_{n\to\infty} \frac{\nu(n)}{n(\log\log n)^2} = \frac{1}{\pi^2} \quad \text{a.s.}.$$

1991 *Mathematics Subject Classification.* Primary 60J65; Secondary 60J15.
Key words and phrases. Random walk, Wiener process, Lévy's class, integral test.

I am unable to solve the problem for higher dimensions, and shall limit myself to the study of random walks on the line. The aim of this paper is:

(a) to provide as much information as possible about the upper asymptotics of $\nu(n)$;
(b) to investigate the corresponding problem for a large class of one-dimensional random walks.

Via an integral criterion, we completely characterize the upper functions of $\nu(n)$ in case $d=1$.

THEOREM 1.1. *For $d=1$ and for any positive non-decreasing sequence $\{a_n\}_{n\geq 1}$, we have*

$$\mathbf{P}[\,\nu(n) > na_n;\ \text{i.o.}] = 0 \quad \text{or} \quad 1,$$

according as whether

$$\sum_n \exp(-\pi\sqrt{a_n}\,)$$

converges or diverges. Here and in the sequel, we adopt the usual symbol "i.o." *meaning "infinitely often" as the appropriate index tends to infinity.*

In Section 5, we shall study the corresponding problem for a general random walk in $\mathbb{Z}$ (of course, ν has to be interpreted as the time necessary for the random walk to exit from its range). Let us first look at the Brownian case.

Let $\{W(t);\, t\geq 0\}$ be a real-valued Wiener process starting from 0, and define, for each $t>0$,

$$\xi(t) \stackrel{\text{def}}{=} \inf\Big\{s>0\colon W(t+s)\notin [\inf_{0\leq u\leq t} W(u),\ \sup_{0\leq u\leq t} W(u)]\Big\}, \tag{1.3}$$

which is a continuous-time analogue to $\nu(n)$ introduced in (1.1). Not surprisingly, we have a similar version for the upper class of $\xi(t)$.

THEOREM 1.2. *Let $f>0$ be a non-decreasing function. Then*

$$\mathbf{P}[\,\xi(t) > tf(t);\ \text{i.o.}] = \begin{cases} 0 \\ 1 \end{cases} \Longleftrightarrow \int^{\infty} \frac{dt}{t} \exp(-\pi\sqrt{f(t)}\,) \begin{cases} < \\ = \end{cases} \infty.$$

In particular, we have

$$\limsup_{t\to\infty} \frac{\xi(t)}{t(\log\log t)^2} = \frac{1}{\pi^2} \quad a.s.. \tag{1.4}$$

REMARK. As is often the case for this kind of problem, Theorem 1.2 has a companion for small times (i.e. as t tends to 0), the statement and the proof of which are omitted.

The proof of Theorem 1.2, which is provided in Section 4, relies on some very accurate estimates (Lemmata 2.2, 3.2 and 3.3 below) on the first- and second-order distributional properties of ξ, developed in Sections 2 and 3, respectively. Section 5 is devoted to the study of the corresponding Erdős–Révész problem for general one-dimensional random walks, for which we obtain an integral test (the forthcoming Theorem 5.1). The latter yields Theorem 1.1 as a special case. We point out that, in order to be able to deal with the general random walk case, we shall actually prove in Section 4 a result slightly stronger than Theorem 1.2.

2. First-order distribution

Let $\{W(t); t \geqq 0\}$ be as before a standard linear Wiener process, and write, for $t \geqq 0$,

$$M_t \stackrel{\text{def}}{=} \sup_{0 \leqq u \leqq t} W(u), \qquad I_t \stackrel{\text{def}}{=} -\inf_{0 \leqq u \leqq t} W(u). \tag{2.1}$$

(Note that I_t is the *absolute value* of the infimum process.) The joint distribution of M_t and I_t is known for fixed $t > 0$ (cf. for example Itô and McKean [8, p. 31]):

$$\mathbf{P}(M_t < x,\ I_t < y) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \exp\left(-\frac{(2k+1)^2 \pi^2 t}{2(x+y)^2}\right) \sin \frac{(2k+1)\pi x}{x+y}, \tag{2.2}$$

for $x > 0$ and $y > 0$. The theta function, however, sometimes causes troubles for exact computations. Things look nicer if an independent random time is introduced.

FACT 2.1 (Yor [12], Imhof [7]). *Let T be an exponential random variable of mean 2, independent of W. For $t > 0$, we have*

$$\mathbf{P}(M_T < x,\ I_T < y) = 1 - \frac{\sinh x + \sinh y}{\sinh(x+y)}, \qquad x > 0,\ y > 0, \tag{2.3}$$

$$\mathbf{E} \exp\left(-\frac{\lambda^2}{2(M_t + I_t)^2}\right) = \frac{1}{\cosh^2(\lambda/2\sqrt{t})}, \qquad \lambda \in \mathbb{R}. \tag{2.4}$$

REMARK. Although (2.2) and (2.3) are mathematically equivalent, it is often a lot easier to use (2.3), thanks to its elegant form. Yor [12; Lecture 6] obtains (2.3) among many other related distributions (all of which are nice-looking, free of "troublesome" theta functions!) by virtue of the Gauss transform. From (2.3), it is possible to deduce (2.4), though the latter is first discovered by Imhof ([7]; see also Vallois [11] for an interpretation via

sample path decompositions) using direct computations, stated in [7] in a somewhat different form:

$$\mathbf{E}\exp\Big(-\frac{\lambda^2\theta_r}{2}\Big)=\frac{1}{\cosh^2(\lambda r/2)},\qquad r>0, \tag{2.5}$$

where $\theta_r \overset{\text{def}}{=} \inf\{t>0\colon M_t+I_t=r\}$. It is easily seen that (2.5) and (2.4) are equivalent. Indeed, by scaling, for any $t>0$,

$$\begin{aligned}\mathbf{P}(\theta_r<t)=\mathbf{P}(M_t+I_t>r)&=\mathbf{P}\Big(M_{1/r^2}+I_{1/r^2}>\frac{1}{\sqrt{t}}\Big)\\&=\mathbf{P}\Big(\frac{1}{(M_{1/r^2}+I_{1/r^2})^2}<t\Big),\end{aligned}$$

which means that θ_r has the same law as $1/(M_{1/r^2}+I_{1/r^2})^2$.

The main result of this section is the following accurate estimate of the first-order tail probability of $\xi(1)$.

LEMMA 2.2. *Let $\xi(t)$ be as in* (1.3). *There exists a finite absolute constant $C_1>1$ such that for $\lambda\geqq 1$,*

$$\frac{1}{C_1\sqrt{\lambda}}\exp(-\pi\sqrt{\lambda})\leqq\mathbf{P}\Big(\xi(1)>\lambda\Big)\leqq\frac{C_1}{\sqrt{\lambda}}\exp(-\pi\sqrt{\lambda}). \tag{2.6}$$

NOTATION. Throughout the paper, $C_k>1$ $(1\leqq k\leqq 45)$ denote unimportant finite constants. Their values either are universal or may depend only on the forthcoming parameter α.

PROOF OF LEMMA 2.2. Obviously we only have to deal with the situation when λ is very large. We have

$$\begin{aligned}\mathbf{P}\Big(\xi(1)>\lambda\Big)&=\mathbf{P}\Big(\sup_{0\leqq t\leqq\lambda}W(1+t)<M_1;\ -\inf_{0\leqq t\leqq\lambda}W(1+t)<I_1\Big)\\&=\mathbf{P}\Big(\sup_{0\leqq t\leqq\lambda}(W(1+t)-W(1))<M_1-W(1);\\&\qquad\qquad -\inf_{0\leqq t\leqq\lambda}(W(1+t)-W(1))<I_1+W(1)\Big).\end{aligned}$$

Since W has independent increments, and since $(M_1-W(1),I_1+W(1))$ has the same distribution as (M_1,I_1) (this is a straightforward consequence of the Brownian time inversion property), by writing $(\widetilde{M}_1,\widetilde{I}_1)$ for an independent copy of (M_1,I_1), we obtain

$$\mathbf{P}\Big(\xi(1)>\lambda\Big)=\mathbf{P}\Big(\sqrt{\lambda}\,\widetilde{M}_1<M_1;\ \sqrt{\lambda}\,\widetilde{I}_1<I_1\Big),$$

(which means that $\xi(1)$ is distributed as $\min((M_1/\widetilde{M}_1)^2, (I_1/\widetilde{I}_1)^2)$). Conditioning on the values of M_1 and I_1, it follows from (2.2) that

$$\mathbf{P}\Big(\xi(1) > \lambda\Big) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{2k+1}\mathbf{E}\Big[\exp(-\frac{(2k+1)^2\pi^2\lambda}{2(M_1+I_1)^2})\sin\frac{(2k+1)\pi M_1}{M_1+I_1}\Big]. \tag{2.7}$$

First, let us observe that by means of (2.4),

$$\begin{aligned}\Delta_1 &\overset{\text{def}}{=} \sum_{k=1}^{\infty}\frac{1}{2k+1}\mathbf{E}\exp(-\frac{(2k+1)^2\pi^2\lambda}{2(M_1+I_1)^2})\\ &= \sum_{k=1}^{\infty}\frac{1}{(2k+1)\,\cosh^2((2k+1)\pi\sqrt{\lambda}\,/2)}\\ &\leqq C_2\,\exp(-3\pi\sqrt{\lambda}).\end{aligned} \tag{2.8}$$

Now we try to establish the first inequality in (2.6). We shall make use of the independent exponential random variable T with $\mathbf{E}(T) = 2$. Since $(M_T, I_T) \overset{(\text{law})}{=} (\sqrt{T}\,M_1,\ \sqrt{T}\,I_1)$, we have

$$\begin{aligned}\Delta_2 &\overset{\text{def}}{=} \mathbf{E}\Big[\exp(-\frac{\pi^2\lambda}{2(M_1+I_1)^2})\sin\frac{\pi M_1}{M_1+I_1}\Big]\\ &= \mathbf{E}\Big[\mathbb{1}_{\{(M_1+I_1)\sqrt{T}>\pi\sqrt{\lambda}\}}\sin\frac{\pi M_1}{M_1+I_1}\Big]\\ &= \mathbf{E}\Big[\mathbb{1}_{\{M_T+I_T>\pi\sqrt{\lambda}\}}\sin\frac{\pi M_T}{M_T+I_T}\Big],\end{aligned} \tag{2.9}$$

where $\mathbb{1}$ denotes the indicator function. By means of the elementary estimate $\sin(\pi x) \geqq 2\min(x, 1-x)$ (for $0 \leqq x \leqq 1$), this yields

$$\begin{aligned}\Delta_2 &\geqq 2\mathbf{E}\Big[\frac{\min(M_T,\,I_T)}{M_T+I_T}\mathbb{1}_{\{M_T+I_T>\pi\sqrt{\lambda}\}}\Big]\\ &\geqq \frac{1}{\pi\sqrt{\lambda}}\mathbf{E}\Big[M_T\,\mathbb{1}_{\{0<M_T<\pi\sqrt{\lambda}/2;\,\pi\sqrt{\lambda}<M_T+I_T<2\pi\sqrt{\lambda}\}}\Big].\end{aligned}$$

According to (2.3),

$$\frac{d}{dx}\mathbf{P}\Big(M_T < x; I_T < y\Big) = \frac{\sinh y}{2\cosh^2((x+y)/2)}. \tag{2.10}$$

Therefore,

$$\Delta_2 \geqq \frac{1}{2\pi\sqrt{\lambda}} \int_0^{\pi\sqrt{\lambda}/2} x\Big(\frac{\sinh(2\pi\sqrt{\lambda}-x)}{\cosh^2(\pi\sqrt{\lambda})} - \frac{\sinh(\pi\sqrt{\lambda}-x)}{\cosh^2(\pi\sqrt{\lambda}/2)}\Big)\,dx$$

$$\geqq \frac{1}{C_3\sqrt{\lambda}} \int_0^{\pi\sqrt{\lambda}/2} x\,\exp(-x-\pi\sqrt{\lambda})\,dx \tag{2.11}$$

$$\geqq \frac{1}{C_4\sqrt{\lambda}}\exp(-\pi\sqrt{\lambda}).$$

By (2.7), $\mathbf{P}(\xi(1)>\lambda) \geqq (4/\pi)(\Delta_2-\Delta_1)$. Combining (2.8) with (2.11) immediately implies the first part of Lemma 2.2.

To verify its second inequality, let us go back to (2.9). Since $\sin(\pi x) \leqq \pi\min(x, 1-x)$ (for $0 \leqq x \leqq 1$), it follows that

$$\Delta_2 \leqq \pi\,\mathbf{E}\Big[\frac{\min(M_T,\, I_T)}{M_T+I_T}\,\mathbb{1}_{\{M_T+I_T>\pi\sqrt{\lambda}\}}\Big]$$
$$\leqq \frac{1}{\sqrt{\lambda}}\mathbf{E}\Big[\min(M_T,\, I_T)\,\mathbb{1}_{\{M_T+I_T>\pi\sqrt{\lambda}\}}\Big].$$

By symmetry and (2.10), we obtain

$$\Delta_2 \leqq \frac{2}{\sqrt{\lambda}}\mathbf{E}\Big[M_T\,\mathbb{1}_{\{I_T\geqq M_T;\, M_T+I_T>\pi\sqrt{\lambda}\}}\Big]$$
$$= \frac{2}{\sqrt{\lambda}}\mathbf{E}\Big[M_T\,\mathbb{1}_{\{M_T\leqq\pi\sqrt{\lambda}/2;\, I_T>\pi\sqrt{\lambda}-M_T\}}\Big]$$
$$+ \frac{2}{\sqrt{\lambda}}\mathbf{E}\Big[M_T\,\mathbb{1}_{\{M_T>\pi\sqrt{\lambda}/2;\, I_T>M_T\}}\Big]$$
$$= \frac{1}{\sqrt{\lambda}}\int_0^{\pi\sqrt{\lambda}/2} x\Big(\frac{2}{e^x} - \frac{\sinh(\pi\sqrt{\lambda}-x)}{\cosh^2(\pi\sqrt{\lambda}/2)}\Big)\,dx + \frac{1}{\sqrt{\lambda}}\int_{\pi\sqrt{\lambda}/2}^{\infty} x\Big(\frac{2}{e^x} - \frac{\sinh x}{\cosh^2 x}\Big)\,dx$$
$$\leqq \frac{C_5}{\sqrt{\lambda}}\int_0^{\pi\sqrt{\lambda}/2} x\,\exp(-x-\pi\sqrt{\lambda})\,dx + \frac{C_5}{\sqrt{\lambda}}\int_{\pi\sqrt{\lambda}/2}^{\infty} x\,\exp(-3x)\,dx$$
$$\leqq \frac{C_6}{\sqrt{\lambda}}\exp(-\pi\sqrt{\lambda}),$$

which, in view of (2.7) and (2.8), yields the second part of Lemma 2.2. □

3. Second-order distribution

Let $\xi(t)$ be as in (1.3). Define

$$\eta(t) \stackrel{\text{def}}{=} \xi(t) + t, \qquad t > 0, \tag{3.1}$$

which is the first exit time of the Wiener process W from $[-I_t, M_t]$ after t (M_t and I_t being respectively the supremum and the modulus of the infimum of W over $[0,t]$; cf. (2.1)). From (2.6), it is easily seen that

$$\frac{1}{C_7\sqrt{\lambda}} \exp(-\pi\sqrt{\lambda}) \leqq \mathbf{P}\Big(\eta(1) > \lambda\Big) \leqq \frac{C_7}{\sqrt{\lambda}} \exp(-\pi\sqrt{\lambda}), \tag{3.2}$$

for $\lambda \geqq 1$. An advantage of working with η is that

$$\text{the process } \eta \text{ is non-decreasing.} \tag{3.3}$$

In this section, we aim at estimating the probability $\mathbf{P}(x < \eta(s) < y, \eta(t) > z)$ for $1 < s < t < z/2$ and $2s < x < y < z$. First, let us treat a simple Laplace transform for a Gaussian distribution.

LEMMA 3.1. *Let $\mathcal{N}$ be an $\mathcal{N}(0,1)$ variable. Then for positive numbers a, b and λ,*

$$\mathbf{E}\exp(-\frac{\lambda^2}{(a+b\,|\mathcal{N}|)^2}) \leqq \exp(-\frac{\lambda^2}{4a^2}) + \exp(-\frac{\lambda}{\sqrt{2}\,b}). \tag{3.4}$$

PROOF. We have

$$\begin{aligned}
\mathbf{E}\exp(-\frac{\lambda^2}{(a+b\,|\mathcal{N}|)^2}) &\leqq \mathbf{E}\Big[\exp\Big(-\frac{\lambda^2}{(a+b\,|\mathcal{N}|)^2}\Big)\,\mathbb{1}_{\{|\mathcal{N}|\leqq a/b\}}\Big] \\
&\qquad + \mathbf{E}\Big[\exp\Big(-\frac{\lambda^2}{(a+b\,|\mathcal{N}|)^2}\Big)\,\mathbb{1}_{\{|\mathcal{N}|> a/b\}}\Big] \\
&\leqq \exp\Big(-\frac{\lambda^2}{4a^2}\Big) + \mathbf{E}\exp\Big(-\frac{\lambda^2}{4b^2N^2}\Big) \\
&= \exp\Big(-\frac{\lambda^2}{4a^2}\Big) + \exp\Big(-\frac{\lambda}{\sqrt{2}\,b}\Big),
\end{aligned}$$

as desired. □

Now let us estimate $\mathbf{P}(x < \eta(s) < y,\ \eta(t) > z)$. We begin with the easy case: $t \leqq y$.

LEMMA 3.2. *Let $1 < s < t < z/2$ and $2s < x < y < z$. If $t \leqq y$, then*

$$\mathbf{P}\Big(x < \eta(s) < y,\ \eta(t) > z\Big) \leqq \frac{C_7\sqrt{s}}{\sqrt{z}} \exp\Big(-\pi\frac{\sqrt{z}}{\sqrt{s}}\Big), \tag{3.5}$$

where C_7 is the constant introduced in (3.2).

PROOF. Since $t \leqq y$, by the definition of η (cf. (3.1)), on the event $\{x < \eta(s) < y,\ \eta(t) > z\}$, the Wiener process W stays in the tube $[-I_s, M_s]$ during $[s,t]$, which means that $\eta(s) = \eta(t)$. Therefore

$$\mathbf{P}\Big(x < \eta(s) < y,\ \eta(t) > z\Big) \leqq \mathbf{P}\Big(\eta(s) > z\Big) \leqq \frac{C_7\sqrt{s}}{\sqrt{z}} \exp\Big(-\pi\frac{\sqrt{z}}{\sqrt{s}}\Big),$$

the last inequality following from the scaling property and (3.2). □

The situation for $t > y$ becomes more delicate. For the applications we bear in mind, we need two estimates for the same probability term, the first being efficient when t is "relatively close" to s, the second when t is "extremely large".

LEMMA 3.3. *Let $1 < s < t < z/2$, $2s < x < y < z$ and $t > y$. Writing*

$$\Delta_3 \stackrel{\text{def}}{=} \frac{\sqrt{st}}{\sqrt{xz}} \exp\Big(-\frac{\pi\sqrt{x}}{\sqrt{s}} - \frac{\pi\sqrt{z}}{\sqrt{t}}\Big) + \exp\Big(-\frac{18\sqrt{z}}{\sqrt{t}}\Big),$$

we have

$$(3.6)\quad \mathbf{P}(x < \eta(s) < y,\ \eta(t) > z) \leqq C_8\,\Delta_3 + C_9\,\exp\Big(-\frac{\sqrt{z}}{C_{10}\sqrt{t}}\Big)\,\mathbf{P}(x < \eta(s) < y),$$

$$(3.7)\quad \mathbf{P}(x < \eta(s) < y,\ \eta(t) > z) \leqq C_{11}\,\Delta_3 + \frac{C_{12}\,s^{1/2}(z/t)^{1/4}}{\sqrt{t-y}}\,\mathbf{P}(x < \eta(s) < y).$$

PROOF. On $\{x < \eta(s) < y,\ \eta(t) > z\}$, we have $\eta(s) < t$. When W exits from $[-I_s, M_s]$ at time $\eta(s)$, it exits from either M_s or $-I_s$. Write

$$E = \Big\{x < \eta(s) < y;\ \eta(t) > z;\ W(\eta(s)) = -I_s\Big\}.$$

By symmetry,

$$\begin{aligned}\mathbf{P}\Big(x < \eta(s) < y,\ \eta(t) > z\Big) &= 2\mathbf{P}(E)\\ &= 2\mathbf{P}\Big(E;\ \sup_{s\leqq u\leqq t} W(u) > M_s\Big)\\ &\quad + 2\mathbf{P}\Big(E;\ \sup_{s\leqq u\leqq t} W(u) \leqq M_s\Big)\\ &\stackrel{\text{def}}{=} \Delta_4 + \Delta_5,\end{aligned} \tag{3.8}$$

with obvious notation. Let us estimate Δ_4 first. Since $\eta(s)$ is a stopping time,

$$\widehat{W}(u) \stackrel{\text{def}}{=} W(u + \eta(s)) - W(\eta(s)), \qquad u \geqq 0, \tag{3.9}$$

is Brownian motion, independent of $\mathcal{F}_{\eta(s)}$ ($\mathcal{F}$ being the natural completed filtration of W). And we can define the corresponding $\widehat{\xi}(u)$ and $\widehat{\eta}(u)$ exactly as in (1.3) and (3.1) respectively, taking $\widehat{W}$ in lieu of W. On the event $E \cap \{\sup_{s \leqq u \leqq t} W(u) > M_s\}$, $\xi(t)$ is nothing else but $\widehat{\xi}(t-\eta(s))$. In formulae, this yields

$$\begin{aligned}
\Delta_4 &\leqq 2\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y;\, W(\eta(s))=-I_s\}}\mathbf{P}\Big(\widehat{\xi}(t-\eta(s)) > z-t \,\Big|\, \mathcal{F}_{\eta(s)}\Big)\Big] \\
&= 2\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y;\, W(\eta(s))=-I_s\}}\mathbf{P}\Big(\widehat{\eta}(t-\eta(s)) > z-\eta(s)) \,\Big|\, \mathcal{F}_{\eta(s)}\Big)\Big] \\
&\leqq 2\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y;\, W(\eta(s))=-I_s\}}\mathbf{P}\Big(\widehat{\eta}(t) > z-t \,\Big|\, \mathcal{F}_{\eta(s)}\Big)\Big] \\
&= 2\mathbf{P}\Big(x<\eta(s)<y;\, W(\eta(s)) = -I_s\Big)\,\mathbf{P}\Big(\eta(t) > z-t\Big) \\
&\leqq 2\mathbf{P}\Big(\eta(s) > x\Big)\,\mathbf{P}\Big(\eta(t) > z-t\Big).
\end{aligned}$$

By scaling and (3.2), we obtain

$$\Delta_4 \leqq \frac{C_{13}\sqrt{st}}{\sqrt{xz}} \exp\Big(-\pi\frac{\sqrt{x}}{\sqrt{s}} - \pi\frac{\sqrt{z}}{\sqrt{t}}\Big). \tag{3.10}$$

To estimate Δ_5, again using the Wiener process $\widehat{W}$ (as well as the corresponding $\widehat{\xi}$ and $\widehat{\eta}$) introduced in (3.9) and we arrive at:

$$\Delta_5 \leqq 2\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y;\, W(\eta(s))=-I_s\}}\mathbf{P}\Big(F \,\Big|\, \mathcal{F}_{\eta(s)}\Big)\Big], \tag{3.11}$$

with

$$\begin{aligned}
F &\overset{\text{def}}{=} \Big\{\sup_{t-\eta(s)\leqq u\leqq z-\eta(s)} \widehat{W}(u) \leqq M_s + I_s;\, -\inf_{t-\eta(s)\leqq u\leqq z-\eta(s)} \widehat{W}(u) \leqq \widehat{I}_{t-\eta(s)}\Big\} \\
&\subseteq \Big\{\sup_{t-\eta(s)\leqq u\leqq z-\eta(s)} \widehat{W}(u) - \inf_{t-\eta(s)\leqq u\leqq z-\eta(s)} \widehat{W}(u) \leqq M_s + I_s + \widehat{I}_{t-\eta(s)}\Big\},
\end{aligned}$$

($-\widehat{I}$ being the infimum process associated with $\widehat{W}$, of course). Since the Wiener process has independent increments, it is easily seen using the strong Markov and scaling properties that

$$\mathbf{P}\Big(F \,\Big|\, \mathcal{F}_{\eta(s)}\Big) \leqq \mathbf{P}\Big(\sqrt{z-t}\,(\widetilde{M}_1 + \widetilde{I}_1) \leqq M_s + I_s + \widehat{I}_{t-\eta(s)} \,\Big|\, \mathcal{F}_{\eta(s)}\Big),$$

where $(\widetilde{M}_1, \widetilde{I}_1)$ denotes as before a random vector, distributed as (M_1, I_1), independent of all the other variables figuring in the above inequality. Since $\mathbf{P}(M_1 + I_1 < x) \leqq C_{14}\exp(-1/C_{15}\,x^2)$ for all $x > 0$ (this for example is a

straightforward consequence of the exact distribution of $M_1 + I_1$ evaluated by Feller [6]), we have

$$\begin{aligned}
\mathbf{P}\Big(F \,\Big|\, \mathcal{F}_{\eta(s)}\Big) &\leqq C_{14}\,\mathbf{E}\Big[\exp\Big(-\frac{z-t}{C_{15}\,(M_s+I_s+\widehat{I}_{t-\eta(s)})^2}\Big)\,\Big|\,\mathcal{F}_{\eta(s)}\Big] \\
&\leqq C_{14}\,\mathbf{E}\Big[\exp\Big(-\frac{z}{C_{16}\,(M_s+I_s+\widehat{I}_t)^2}\Big)\,\Big|\,\mathcal{F}_{\eta(s)}\Big] \\
&\leqq C_{14}\,\exp\Big(-\frac{z}{4C_{16}\,(M_s+I_s)^2}\Big) + C_{14}\,\exp\Big(-\frac{\sqrt{z}}{\sqrt{2C_{16}\,t}}\Big),
\end{aligned}$$

where in the last inequality we have applied (3.4) to $\lambda = \sqrt{z/C_{16}}$, $a = M_s + I_s$ and $b = \sqrt{t}$. Going back to (3.11), we obtain

$$\begin{aligned}
\Delta_5 &\leqq C_{14}\,\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y\}}\,\exp\Big(-\frac{z}{C_{17}\,(M_s+I_s)^2}\Big)\Big] \\
&\quad + C_{14}\,\exp\Big(-\frac{\sqrt{z}}{\sqrt{C_{18}\,t}}\Big)\,\mathbf{P}\Big(x<\eta(s)<y\Big) \\
&\leqq C_{14}\,\mathbf{E}\Big[\mathbb{1}_{\{x<\eta(s)<y\}}\,\exp\Big(-\frac{\sqrt{zt}}{144C_{17}\,s}\Big)\,\mathbb{1}_{\{M_s+I_s<12s^{1/2}(z/t)^{1/4}\}}\Big] \\
&\quad + C_{14}\,\mathbf{P}\Big(M_s+I_s>12s^{1/2}(z/t)^{1/4}\Big) \\
&\quad + C_{14}\,\exp\Big(-\frac{\sqrt{z}}{\sqrt{C_{18}\,t}}\Big)\mathbf{P}\Big(x<\eta(s)<y\Big) \\
&\leqq C_{19}\,\exp\Big(-\frac{\sqrt{z}}{C_{20}\,\sqrt{t}}\Big)\,\mathbf{P}\Big(x<\eta(s)<y\Big) + C_{21}\,\exp\Big(-\frac{18\sqrt{z}}{\sqrt{t}}\Big),
\end{aligned} \tag{3.12}$$

using the well-known Gaussian tail estimate $\mathbf{P}(|\mathcal{N}(0,1)| > \lambda) \leqq \exp(-\lambda^2/2)$ (for $\lambda > 0$). Assembling (3.8), (3.10) and (3.12) yields (3.6). To verify (3.7), we have to estimate Δ_5 in a different way. If $\widehat{W}$ is as defined in (3.9), we have

$$\begin{aligned}
\Delta_5 &\leqq \mathbf{P}\Big(x<\eta(s)<y;\ \sup_{0\leqq u\leqq t-\eta(s)}\widehat{W}\leqq M_s+I_s\Big) \\
&\leqq \mathbf{P}\Big(x<\eta(s)<y;\ \sup_{0\leqq u\leqq t-y}\widehat{W}\leqq M_s+I_s\Big) \\
&\leqq \mathbf{P}\Big(M_s+I_s>12s^{1/2}(z/t)^{1/4}\Big) \\
&\quad + \mathbf{P}\Big(x<\eta(s)<y\Big)\,\mathbf{P}\Big(\sup_{0\leqq u\leqq t-y}\widehat{W}\leqq 12s^{1/2}(z/t)^{1/4}\Big) \\
&\leqq 2\exp\Big(-\frac{18\sqrt{z}}{\sqrt{t}}\Big) + \frac{C_{22}\,s^{1/2}(z/t)^{1/4}}{\sqrt{t-y}}\,\mathbf{P}\Big(x<\eta(s)<y\Big),
\end{aligned} \tag{3.13}$$

using the fact that the density of $|\mathcal{N}(0,1)|$ is bounded (above) by 1. Combining (3.8), (3.10) and (3.13) now gives (3.7). Lemma 3.3 is proved. □

4. Proof of Theorem 1.2

As was pointed out in Section 1, in order to deal with the general random walk case, we shall prove something slightly stronger than Theorem 1.2.

Define for $t>0$ and $x>0$,

$$\eta_+(t,x) \stackrel{\text{def}}{=} \inf\Big\{ s \geqq t\colon W(s) \notin \Big[-I_t - x,\ M_t + x \Big] \Big\}, \tag{4.1}$$

$$\eta_-(t,x) \stackrel{\text{def}}{=} \begin{cases} \inf\{ s \geqq t\colon W(s) \notin [-(I_t - x),\ M_t - x] \} & \text{if } x \leqq (M_t + I_t)/2 \\ t & \text{otherwise,} \end{cases} \tag{4.2}$$

where, M and $-I$, introduced in (2.1), denote respectively the supremum and infimum processes associated with W. Since $\eta_+(t,x) > \eta(t) = \xi(t) + t$ and $\eta_-(t,x) < \xi(t) + t$, Theorem 1.2 is a straightforward consequence of the following result.

PROPOSITION 4.1. *Fix $0<\alpha<1/2$, and let $\eta_+(\cdot,\cdot)$ and $\eta_-(\cdot,\cdot)$ be as in* (4.1) *and* (4.2), *respectively. For any non-decreasing function $f>0$, define*

$$\mathcal{J}(f) \stackrel{\text{def}}{=} \int^{\infty} \frac{dt}{t} \exp(-\pi \sqrt{f(t)}\,). \tag{4.3}$$

We have

$$\mathcal{J}(f) < \infty \implies \mathbf{P}\Big[\eta_+(t, t^{1/2-\alpha}) > t f(t);\ \text{i.o.} \Big] = 0, \tag{4.4}$$

$$\mathcal{J}(f) = \infty \implies \mathbf{P}\Big[\eta_-(t, t^{1/2-\alpha}) > t f(t);\ \text{i.o.} \Big] = 1. \tag{4.5}$$

REMARK 4.2. As is well known (cf. Erdős [4] or Csáki [1]), we can limit ourselves to those functions f such that

$$\frac{(\log\log t)^2}{16} \leqq f(t) \leqq (\log\log t)^2. \tag{4.6}$$

PROOF of (4.4). Observe that $\eta(1)$ is a stopping time with respect to the natural completed filtration of W. Thus, it follows from the strong Markov and scaling properties that, for any fixed $x>0$,

$$\eta_+(1,x) = \eta(1) + x^2\sigma,$$

where σ is the first hitting time of 1 by a standard Wiener process, independent of $\eta(t)$. By means of (3.2), we have, for $\lambda \geqq 3$,

$$\begin{aligned} \mathbf{P}\Big(\eta_+(1,x) > \lambda\Big) &\leqq \mathbf{P}\Big(\eta(1) > \lambda - \sqrt{\lambda}\Big) + \mathbf{P}\Big(x^2\sigma > \sqrt{\lambda}\Big) \\ &\leqq \frac{C_7}{\sqrt{\lambda}} \exp\Big(-\pi\sqrt{\lambda - \sqrt{\lambda}}\,\Big) + \mathbf{P}\Big(|\mathcal{N}(0,1)| < \frac{x}{\lambda^{1/4}}\Big) \\ &\leqq \frac{C_{23}}{\sqrt{\lambda}} \exp\Big(-\pi\sqrt{\lambda}\Big) + \frac{x}{\lambda^{1/4}}, \end{aligned} \tag{4.7}$$

using the fact that the density function of $|\mathcal{N}(0,1)|$ is smaller than 1. Now let $f > 0$ be non-decreasing such that $\mathcal{J}(f) < \infty$. Obviously $f(t)$ goes to infinity. Taking $\lambda = f(t)$ and $x = t^{-\alpha}$ in (4.7), and by virtue of (3.2) and the scaling property, we obtain, for sufficiently large t,

$$\mathbf{P}\Big(\eta_+(t, t^{1/2-\alpha}) > t\, f(t)\Big) \leqq \frac{C_{24}}{\sqrt{f(t)}} \exp(-\pi\sqrt{f(t)}\,) + t^{-\alpha}. \tag{4.8}$$

Following Erdös [4], let us define $s_n \overset{\text{def}}{=} \exp(n/\log n)$ for sufficiently large n. From (4.8), it follows that

$$\begin{aligned} \mathbf{P}\Big(\eta_+(s_{n+1}, s_{n+1}^{1/2-\alpha}) > s_n f(s_n)\Big) &\leqq \frac{C_{24}}{\sqrt{s_n f(s_n)/s_{n+1}}} \exp\left(-\pi\sqrt{\frac{s_n\, f(s_n)}{s_{n+1}}}\right) + s_{n+1}^{-\alpha} \\ &\leqq \frac{C_{25}}{\sqrt{f(s_n)}} \exp(-\pi\sqrt{f(s_n)}\,) + s_{n+1}^{-\alpha}, \end{aligned}$$

which is summable for n, thanks to the convergence of $\mathcal{J}(f)$. According to the Borel-Cantelli lemma, (almost surely for large n), we have $\eta_+(s_{n+1}, s_{n+1}^{1/2-\alpha}) \leqq s_n f(s_n)$. Since $t \mapsto \eta_+(t, t^{1/2-\alpha})$ is non-decreasing, we have, for $t \in [s_n, s_{n+1}]$,

$$\eta_+(t, t^{1/2-\alpha}) \leqq \eta_+(s_{n+1}, s_{n+1}^{1/2-\alpha}) \leqq s_n f(s_n) \leqq t f(t),$$

as desired. □

The proof of (4.5) is more delicate. Assume (4.6) again, without loss of generality. Write

$$g(t) = f(t) + 1.$$

Let us fix a large initial index n_0, and define $t_n = \exp(C_{26}\, n/\log n)$ (for $n \geqq n_0$). Here $C_{26} > 1$ is an absolute constant so large that

$$C_7\, e^{-\pi C_{26}/5} \leqq \frac{1}{9C_7}, \tag{4.9}$$

C_7 being the constant figuring in (3.2). By the mean-value theorem, for $n_0 \leqq i < j$, we have

$$\frac{j-i}{2\log j} \leqq \frac{j}{\log j} - \frac{i}{\log i} \leqq \frac{j-i}{\log i}. \tag{4.10}$$

We need a preliminary result.

LEMMA 4.3. *Let $g > 0$ be a non-decreasing function satisfying* (4.6) *with $g = f + 1$, and let $i \geqq n_0$. We have*

$$\sum_{j \geqq i+(\log i)^3} \frac{g^{1/4}(t_j)}{\sqrt{t_j - t_{i+1}\, g(t_i)}} \leqq \frac{C_{27}}{\sqrt{t_i}}. \tag{4.11}$$

PROOF. Let $j \geqq k \overset{\text{def}}{=} [i + (\log i)^3]$ (i.e. the integer part of $i + (\log i)^3$). From (4.10) and (4.6) it follows that $t_j > 2t_{i+1}\, g(t_i)$. Therefore

$$\frac{g^{1/4}(t_j)}{\sqrt{t_j - t_{i+1}\, g(t_i)}} \leqq \frac{C_{28}\sqrt{\log\log t_j}}{\sqrt{t_j}} \leqq \int_{t_{j-1}}^{t_j} \frac{dt}{t_j - t_{j-1}} \frac{C_{28}\sqrt{\log\log t}}{\sqrt{t}}$$

$$\leqq C_{29} \int_{t_{j-1}}^{t_j} \frac{(\log\log t)^{3/2}}{t^{3/2}}\, dt,$$

which implies

$$\sum_{j \geqq k} \frac{g^{1/4}(t_j)}{\sqrt{t_j - t_{i+1}\, g(t_i)}} \leqq C_{29} \int_{t_{k-1}}^{\infty} \frac{(\log\log t)^{3/2}}{t^{3/2}}\, dt$$

$$\leqq C_{30} \frac{(\log\log t_{k-1})^{3/2}}{\sqrt{t_{k-1}}}$$

$$\leqq \frac{C_{31}}{\sqrt{t_i}},$$

the last inequality following from the fact that $t_j > i\, t_i$ for $j \geqq i + (\log i)^3$ (this is a straightforward consequence of (4.10)). □

The main difficulty in the proof of the divergent part of Theorem 1.2 is in applying the Borel–Cantelli lemma to a sequence of events which are not independent. One is tempted to choose some nice-looking stopping times, such as $\{\tau(r);\, r \geqq 0\}$, the inverse local times at 0 of W. After all, since $W(\tau(r)) = 0$, it looks as if it might take "a long time" for W to exit from $[-I_{\tau(r)}, M_{\tau(r)}]$ after $\tau(r)$. However, this turns out to be only utopic — it

is possible, with the aid of Theorem 3 of Pitman and Yor [10], to prove that $\limsup_{r\to\infty} \xi(\tau(r))/\tau(r)(\log\log\tau(r))^2 = 1/4\pi^2$ almost surely (which is of a relatively poor performance if compared with (1.4)). So we stick to our deterministic-time choice.

PROOF of (4.5). Take a function $f>0$ satisfying (4.6) such that $\mathcal{J}(f)=\infty$. Thus, for $g(t)=f(t)+1$, we have

$$\sum_n \frac{1}{\sqrt{g(t_n)}} \exp(-\pi\sqrt{g(t_n)}\,) = \infty, \tag{4.12}$$

(recall that $t_n = \exp(C_{26}\, n/\log n)$). Consider the events

$$E_n = \Big\{\eta_-(t_n, t_n^{1/2-\alpha}) \geqq t_n\, g(t_n)\Big\} \cap \Big\{\eta(t_n) < t_{n+1}\, g(t_n)\Big\} \cap \Big\{t_n^{1/2-\alpha} < \min(M_{t_n}, I_{t_n})\Big\},$$

for $n \geqq n_0 = n_0(\alpha)$. The extra condition $\eta_-(t_n, t_n^{1/2-\alpha}) < t_{n+1}\, g(t_n)$ does not considerably influence the probability of E_n, but makes life a lot easier (this is a trick I have learnt from Csáki [1]) as we shall see soon. Observe that on the event $\{x < \min(M_t, I_t)\}$,

$$\eta(t) \leqq \eta_-(t,x) + x^2\sigma, \tag{4.13}$$

where σ denotes again a variable distributed as the first hitting time of 1 by a linear Wiener process, independent of $\eta_-(t,x)$ (we shall not use the independence. however; (4.13) is not an identity, due to the fact that $W(t)$ may lay out of $[-I_t + x, M_t - x]$). Hence by scaling and (3.2), for $y \geqq t$, we have

$$\begin{aligned}
&\mathbf{P}\Big(\eta_-(t,x) > y;\ x < \min(M_t, I_t)\Big) \\
&\quad \geqq \mathbf{P}\Big(\eta(t) > y + \sqrt{yt}\,\Big) - \mathbf{P}\Big(x^2\sigma > \sqrt{yt}\,\Big) - \mathbf{P}\Big(\min(M_t, I_t) \leqq x\Big) \\
&\quad \geqq \frac{\sqrt{t}}{C_7\sqrt{y+\sqrt{yt}}} \exp\Big(-\pi\frac{\sqrt{y+\sqrt{yt}}}{\sqrt{t}}\Big) - \mathbf{P}\Big(|\mathcal{N}(0,1)| < \frac{x}{(yt)^{1/4}}\Big) \\
&\qquad\qquad - 2\mathbf{P}\Big(|\mathcal{N}(0,1)| < \frac{x}{\sqrt{t}}\Big) \\
&\quad \geqq \frac{\sqrt{t}}{7C_7\sqrt{y}} \exp\Big(-\pi\frac{\sqrt{y}}{\sqrt{t}}\Big) - \frac{x}{(yt)^{1/4}} - \frac{2x}{\sqrt{t}}.
\end{aligned}$$

(In the last inequality, we have used the relation $e^{-\pi/2}/\sqrt{2} > 1/7$ and the boundedness of Gaussian densities.) Consequently, using (3.2) and (4.6), we obtain

$$\mathbf{P}(E_n) \geqq \mathbf{P}\Big(\eta_-(t_n, t_n^{1/2-\alpha}) \geqq t_n g(t_n);\ t_n^{1/2-\alpha} < \min(M_{t_n}, I_{t_n})\Big)$$

$$- \mathbf{P}\Big(\eta(t_n) \geqq t_{n+1} g(t_n)\Big)$$

(4.14)

$$\begin{aligned}
&\geqq \frac{1}{7C_7\sqrt{g(t_n)}} \exp(-\pi\sqrt{g(t_n)}\,) - \frac{1}{t_n^{\alpha}\, g^{1/4}(t_n)} - \frac{2}{t_n^{\alpha}} \\
&\quad - \frac{C_7}{\sqrt{g(t_n)}} \exp\Big(-\pi \frac{\sqrt{t_{n+1}}}{\sqrt{t_n}} \sqrt{g(t_n)}\Big) \\
&\geqq \frac{1}{8C_7\sqrt{g(t_n)}} \exp(-\pi\sqrt{g(t_n)}\,) - \frac{C_7\, e^{-\pi C_{26}/5}}{\sqrt{g(t_n)}} \exp(-\pi\sqrt{g(t_n)}\,) \\
&\geqq \frac{1}{C_{32}\sqrt{g(t_n)}} \exp(-\pi\sqrt{g(t_n)}\,),
\end{aligned}$$

the last inequality following from (4.9). Thus by (4.12), we have

$$\sum_n \mathbf{P}(E_n) = \infty. \tag{4.15}$$

Another consequence of (4.14) and (3.2) is that

$$\mathbf{P}\Big(\eta(t_n) \geqq t_n g(t_n)\Big) \leqq C_{33}\, \mathbf{P}(E_n). \tag{4.16}$$

Now consider $n_0 \leqq i < j \leqq n$. Observe that

$$\mathbf{P}(E_i \cap E_j) \leqq \mathbf{P}\Big(t_i\, g(t_i) \leqq \eta(t_i) < t_{i+1}\, g(t_i);\ \eta(t_j) > t_j\, g(t_j)\Big). \tag{4.17}$$

There are two possible situations.

(i) First case: $j < j(i)$, where

$$j(i) \overset{\text{def}}{=} \inf\Big\{j \geqq n_0 : t_{j(i)} \geqq t_{i+1}\, g(t_i)\Big\}.$$

Using (4.17) and applying Lemma 3.2 to $s = t_i$, $t = t_j$, $x = t_i\, g(t_i)$, $y = t_{i+1}\, g(t_i)$ and $z = t_j\, g(t_j)$, we obtain

$$\mathbf{P}(E_i \cap E_j) \leqq \frac{C_7}{\sqrt{g(t_j)}} \exp\Big(-\pi \frac{\sqrt{t_j g(t_j)}}{\sqrt{t_i}}\Big).$$

According to (4.10) and (4.6), $\sqrt{t_j/t_i} \geqq 1 + (j-i)/4 \log j \geqq 1 + (j-i)/C_{34}\sqrt{g(t_j)}$, thus by virtue of (4.14),

$$\begin{aligned}
\mathbf{P}(E_i \cap E_j) &\leqq \frac{C_7\, e^{-\pi\sqrt{g(t_j)}}}{\sqrt{g(t_j)}} \exp\Big(-\frac{\pi}{C_{34}}(j-i)\Big) \\
&\leqq \frac{C_7\, e^{-\pi\sqrt{g(t_i)}}}{\sqrt{g(t_i)}} \exp\Big(-\frac{\pi}{C_{34}}(j-i)\Big) \\
&\leqq C_{35}\, e^{-(j-i)/C_{36}}\, \mathbf{P}(E_i),
\end{aligned}$$

which implies

$$\sum\sum_{n_0\leqq i<j\leqq n;\, j<j(i)} \mathbf{P}(E_i\cap E_j)\leqq C_{37}\sum_{i=n_0}^{n}\mathbf{P}(E_i). \tag{4.18}$$

(ii) Second case: $j\geqq j(i)$. We are entitled to apply Lemma 3.3. Indeed, in view of (4.10), (4.6) and (4.16), the inequalities in Lemma 3.3 readily yield

$$\mathbf{P}(E_i\cap E_j)\leqq C_{38}\,\mathbf{P}(E_i)\,\mathbf{P}(E_j)+C_{39}\,j^{-4}+C_{40}\,\mathbf{P}(E_i)\,j^{-1/C_{41}},$$

$$\mathbf{P}(E_i\cap E_j)\leqq C_{38}\,\mathbf{P}(E_i)\,\mathbf{P}(E_j)+C_{39}\,j^{-4}+C_{42}\,\frac{t_i^{1/2}g^{1/4}(t_j)}{\sqrt{t_j-t_{i+1}\,g(t_i)}}\mathbf{P}(E_i).$$

Since $t_{i+(\log i)^3}>t_{i+1}\,g(t_i)$, we have $j(i)\leqq i+(\log i)^3$. Accordingly,

$$\begin{aligned}\sum\sum_{n_0\leqq i<j\leqq n;\, j\geqq j(i)}\mathbf{P}(E_i\cap E_j)&\leqq C_{38}\Big(\sum_{i=n_0}^{n}\mathbf{P}(E_i)\Big)^2+C_{39}\sum_{i=n_0}^{n}\sum_{j=i}^{\infty}j^{-4}\\&\quad+C_{40}\sum_{i=n_0}^{n}\mathbf{P}(E_i)\sum_{j(i)\leqq j\leqq i+(\log i)^3}j^{-1/C_{41}}\\&\quad+C_{42}\sum_{i=n_0}^{n}\mathbf{P}(E_i)\sum_{j\geqq i+(\log i)^3}\frac{t_i^{1/2}g^{1/4}(t_j)}{\sqrt{t_j-t_{i+1}\,g(t_i)}}\\&\leqq C_{38}\Big(\sum_{i=n_0}^{n}\mathbf{P}(E_i)\Big)^2+C_{43}\sum_{i=n_0}^{n}i^{-3}+C_{44}\sum_{i=n_0}^{n}\mathbf{P}(E_i),\end{aligned}$$

by virtue of Lemma 4.3. This inequality, jointly considered with (4.18) and (4.15), yields

$$\liminf_{n\to\infty}\sum_{i=n_0}^{n}\sum_{j=n_0}^{n}\mathbf{P}(E_i\cap E_j)\Big/\Big(\sum_{i=n_0}^{n}\mathbf{P}(E_i)\Big)^2\leqq C_{45}\,.$$

According to Kochen and Stone's Borel–Cantelli lemma ([9]), we have $\mathbf{P}(E_n;\ \text{i.o.})\geqq 1/C_{45}$, which, by means of Kolmogorov's 0–1 law, yields (4.5). □

5. Random walks

Let $\{X_i\}_{i\geqq 1}$ be a sequence of real-valued independent and identically distributed random variables with

$$\mathbf{E}(X_1)=0,\qquad \mathbf{E}(X_1^2)=1\qquad\text{and}\qquad \mathbf{E}(|X_1|^{2+\delta})<\infty, \tag{5.1}$$

for some $\delta > 0$. Consider the random walk $S_n \overset{\text{def}}{=} \sum_{i=1}^{n} X_i$ (with $S_0 \overset{\text{def}}{=} 0$). We are interested in

$$\nu(n) \overset{\text{def}}{=} \min\Big\{k \geqq 1 : S_{n+k} \notin \Big[\min_{0 \leqq i \leqq n} S_i,\ \max_{0 \leqq i \leqq n} S_i\Big]\Big\}, \tag{5.2}$$

i.e. $\nu(n)$ stands for the step necessary for the random walk to exit from its range. In case of the simple random walk, $\nu(n)$ clearly corresponds to (1.1). We now present an integral test for $\nu(n)$ for general random walks on the line.

THEOREM 5.1. *Let $\{S_n\}_{n \geqq 0}$ be a real-valued random walk such that* (5.1) *holds. If $\nu(n)$ is defined as in* (5.2), *and if $\{a_n\}_{n \geqq 1}$ is a non-decreasing sequence, then*

$$\mathbf{P}\Big[\nu(n) > na_n;\ \text{i.o.}\Big] = \begin{cases} 0 \\ 1 \end{cases} \Longleftrightarrow \sum_n \exp(-\pi\sqrt{a_n}) \begin{cases} < \\ = \end{cases} \infty.$$

The proof of Theorem 5.1 relies on some known results recalled as follows.

FACT 5.2. *Assume that* (5.1) *holds. Possibly after redefinition on an enlarged probability space, there exists a standard Wiener process $\{W(t); t \geqq 0\}$ such that as n tends to infinity,*

$$|S_n - W(n)| = o(n^{1/(2+\delta)}) \quad a.s., \tag{5.3}$$

where δ is as in (5.1).

FACT 5.3. *We have*

$$\limsup_{n \to \infty} \frac{1}{\sqrt{2 \log n}} \max_{0 \leqq m \leqq n-1} \sup_{0 \leqq s \leqq 1} |W(m+s) - W(m)| = 1 \quad a.s.. \tag{5.4}$$

Fact 5.2 is a somewhat weaker version of the classical Komlós–Major–Tusnády strong approximation theorem, stated in the present form in Csörgő and Révész [3, p. 107]. Fact 5.3 is a particular case of the celebrated Csörgő–Révész large increments theorem, cf. [3, p. 30].

PROOF OF THEOREM 5.1. As usual, we assume without loss of generality that

$$\frac{(\log\log n)^2}{16} \leqq a_n \leqq (\log\log n)^2.$$

Let $\eta_+(\cdot,\cdot)$ be as in (4.1). First, suppose $\sum_n \exp(-\pi\sqrt{a_n}) < \infty$. Define $f(t) = na_n/(n+1)$ (for $n < t \leqq n+1$), which satisfies $\mathcal{J}(f) < \infty$ ($\mathcal{J}(f)$ being defined in (4.3)). According to (4.4), (almost surely for sufficiently large t), $\eta_+(t, t^{1/(2+\delta)}) \leqq tf(t)$. Thus $W(s) \notin [-I_t - t^{1/(2+\delta)},\ M_t + t^{1/(2+\delta)}]$ for some

$t < s \leqq tf(t)$. By (5.4), the fluctuations of the Wiener process between neighbouring integers are relatively negligible, it follows that

$$W([t]+l) \notin \Big[-I_t - \frac{t^{1/(2+\delta)}}{2},\ M_t + \frac{t^{1/(2+\delta)}}{2}\Big], \tag{5.5}$$

for some positive integer $l \leqq t\,(f(t)-1)$, with $[t]$ denoting the integer part of t. Let $t \in [n, n+1)$. It is confirmed by (5.3) and (5.4) that, when n is sufficiently large,

$$\Big[-I_t - \frac{t^{1/(2+\delta)}}{2},\ M_t + \frac{t^{1/(2+\delta)}}{2}\Big] \supseteq \Big[\min_{0\leqq i\leqq n} S_i - \frac{n^{1/(2+\delta)}}{3},\ \max_{0\leqq i\leqq n} S_i + \frac{n^{1/(2+\delta)}}{3}\Big].$$

In view of (5.5), we get

$$S_{n+l} \notin \Big[\min_{0\leqq i\leqq n} S_i,\ \max_{0\leqq i\leqq n} S_i\Big],$$

which means $\xi(n) \leqq (n+1)(f(n+1)-1) \leqq na_n$. This yields the desired convergent part of Theorem 5.1. The divergent part can be proved by a similar argument (using (4.5) instead of (4.4)), of which the details are omitted. □

ACKNOWLEDGEMENTS. I am grateful to Marc Yor for insightful comments on a first draft of the paper. Many thanks go to an anonymous referee for helpful suggestions.

REFERENCES

[1] CSÁKI, E., An integral test for the supremum of Wiener local time, *Probab. Theory Related Fields* **83** (1989), 207–217. *MR* **91a**:60206

[2] CSÁKI, E., A note on: "Three problems on the random walk in $\mathbb{Z}^{d}$" by P. Erdős and P. Révész, *Studia Sci. Math. Hungar.* **26** (1991), 201–205. *MR* **93k**:60172

[3] CSÖRGŐ, M. and RÉVÉSZ, P., *Strong approximations in probability and statistics*, Probability and Mathematical Statistics, Academic Press, New York, 1981. *MR* **84d**:60050

[4] ERDŐS, P., On the law of the iterated logarithm, *Ann. Math.* (2) **43** (1942), 419–436. *MR* **4**, 16j

[5] ERDŐS, P. and RÉVÉSZ, P., Three problems on the random walk in $\mathbb{Z}^{d}$, *Studia Sci. Math. Hungar.* **26** (1991), 309–320. *MR* **93k**:60171

[6] FELLER, W., The asymptotic distribution of the range of sums of independent random variables, *Ann. Math. Statist.* **22** (1951), 427–432. *MR* **13**, 140i

[7] IMHOF, J. P., On the range of Brownian motion and its inverse process, *Ann. Probab.* **13** (1985), 1011–1017. *MR* **86m**:60195

[8] ITÔ, K. and MCKEAN, H. P., *Diffusion processes and their sample paths*, Die Grundlehren der mathematischen Wissenschaften, Band 125, Springer, Berlin, 1965. *MR* **33** #8031

[9] KOCHEN, S. B. and STONE, C. J., A note on the Borel–Cantelli lemma, *Illinois J. Math.* **8** (1964), 248–251. *MR* **28** #4562

[10] PITMAN, J. W. and YOR, M., Dilatations d'espace–temps, réarrangements des trajectoires browniennes, et quelques extensions d'une identité de Knight, *C. R. Acad. Sci. Paris Ser. I Math.* **316** (1993), 723–726. *MR* **93k**:60208

[11] VALLOIS, P., Diffusion arrêtée au premier instant où l'amplitude atteint un niveau donné, *Stochastics Stochastics Rep.* **43** (1993), 93–115. *MR* **95f**:60092

[12] YOR, M., *Local times and excursions for Brownian motion: A concise introduction*, Lecciones en Matemáticas, Universidad Central de Venezuela, 1995.

(Received May 23, 1996)

LABORATOIRE DE STATISTIQUE THÉORIQUE
ET APPLIQUÉE
UNIVERSITÉ PARIS VI
TOUR 45–55 3e ÉTAGE
4 PLACE JUSSIEU
F–75252 PARIS Cedex 05
FRANCE
shi@ccr.jussieu.fr

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 133–138

HYPOCONTINUITY AND UNIFORM BOUNDEDNESS FOR BILINEAR MAPS

J. WU and R. LI

1. Introduction

For bilinear maps between topological vector spaces, Bourbaki has introduced the notion of hypocontinuity which lies between separate continuity and joint continuity for bilinear maps. By using the Basic Matrix Theorem, Antosik and Swartz studied the hypocontinuity for bilinear maps ([1] §6, [2]) and obtained sufficient conditions in the absence of completeness or barrelledness assumptions on the spaces involved. These results generalized the classical results of Bourbaki ([3]) and Mazur–Orlicz ([4]). However, we are more interested in the characterization of hypocontinuity for bilinear maps. In addition, Antosik and Swartz also studied the uniform boundedness and the equicontinuity for the family of bilinear maps and some sufficient conditions were obtained in ([1] §6, [2], [5]). In this paper, we present characterizations for these problems.

2. Hypocontinuity

Let E, F and G be topological vector spaces and $b: E \times F \to G$ be a bilinear map (i.e., the map $b(x,\cdot): F \to G$, $b(x,\cdot)(y) = b(x,y)$, and $b(\cdot,y): E \to G$, $b(\cdot,y)(x) = b(x,y)$, are linear maps for each x and y).

Let $\mathcal{N}$ be a family of bounded subsets of F, b is said to be $\mathcal{N}$-hypocontinuous if for each neighbourhood V of 0 in G and each $A \in \mathcal{N}$, there is a neighbourhood U of 0 in E such that $b(U, A) \subseteq V$. If for each $A \in \mathcal{N}$ when $x_i \to 0$ in E, $\lim_i b(x_i, y) = 0$ uniformly for $y \in A$, then b is said to be sequentially $\mathcal{N}$-hypocontinuous ([5] §4). If E is a paranormed space, then sequentially $\mathcal{N}$-hypocontinuous is equivalent to $\mathcal{N}$-hypocontinuous.

Recall that a sequence $\{x_k\}$ in E is $\mathcal{K}$-convergent if every subsequence of $\{x_k\}$ has a further subsequence $\{x_{n_k}\}$ such that the series $\sum x_{n_k}$ is convergent to an element $x \in E$ ([1] §3). A $\mathcal{K}$-convergent sequence obviously

1991 *Mathematics Subject Classification.* Primary 47A05.
Key words and phrases. Bilinear maps, hypocontinuity, uniform boundedness.

0081–6906/99/$ 5.00

converges to 0, but the converse is false in general although it does hold in complete metric linear spaces ([1] §3). A subset $B \subseteqq E$ is said to be $\mathcal{K}$-bounded if whenever $\{x_k\} \subseteqq B$ and $\{t_k\}$ is a scalar sequence converging to 0, the sequence $\{t_k x_k\}$ is $\mathcal{K}$-convergent ([1] §3). The families of all $\mathcal{K}$-convergent sequences ($\mathcal{K}$-bounded subsets) in E are denoted by $\mathcal{K}S(E)(\mathcal{K}B(E))$.

Let $(x_{ij})_{ij}$ be an infinite matrix in G. If $\{n_k\}$ is an increasing sequence of positive integers, then the matrix $(x_{n_i n_j})_{ij}$ is said to be a principle submatrix of $(x_{ij})_{ij}$.

LEMMA 1 (Basic Matrix Theorem). *Let G be an abelian topological group and $x_{ij} \in G$ for $i, j \in N$. Suppose*

(i) $\lim_i x_{ij} = x_j$ *exists for each j and*

(ii) *for each increasing sequence $\{m_j\}$ there is a subsequence $\{n_j\}$ of $\{m_j\}$ such that $\left\{\sum_j x_{in_j}\right\}_{i=1}^{\infty}$ is Cauchy.*

Then $\lim_i x_{ij} = x_j$ uniformly with respect to $j \in N$. In particular, $\lim_i x_{ii} = 0$ [6].

For hypocontinuity of bilinear maps, we now state our first result.

THEOREM 2. *Let E, F and G be paranormed spaces and $b: E \times F \to G$ be a separately continuous bilinear map, $\mathcal{N}$ be a family of bounded subsets of F, then b is $\mathcal{N}$-hypocontinuous if and only if for each $x_i \to 0$ in E and $A \in \mathcal{N}$, if $\{y_j\} \subseteqq A$, then for the infinite matrix $(\|b(x_i, y_j)\|)_{ij}$ there is a principle submatrix $(\|b(x_{n_i}, y_{n_j})\|)_{ij}$ such that $\lim_j \|b(x_{n_i}, y_{n_j})\|$ exists for each $i \in N$ and converges uniformly with respect to $i \in N$. Here $\|\cdot\|$ is the paranorm of G.*

PROOF OF NECESSITY. Let $x_i \to 0$ in E and $A \in \mathcal{N}$. If $\{y_j\} \subseteqq A$, we shall prove that for the infinite matrix $(\|b(x_i, y_j)\|)_{ij}$ there exists a principle submatrix satisfying the conditions of the theorem. Since b is $\mathcal{N}$-hypocontinuous, $\lim_i b(x_i, y_j) = 0$ uniformly with respect to $j \in N$. That is, for each $\varepsilon > 0$, there is $i_0 \in N$, whenever $i \geqq i_0$, $\|b(x_i, y_j)\| \leqq \varepsilon$ for all $j \in N$ holds. With no loss of generality, we may suppose for each $i \in N$, $\|b(x_i, y_j)\| < 1/2^i$ for all $j \in N$ holds. Take $i = 1$. Since $\|b(x_1, y_j)\| < \frac{1}{2}$ for all $j \in N$ holds, $\{\|b(x_1, y_j)\|\}$ is a bounded real number sequence. Hence, there is a subsequence $\{y_{j_k^{(1)}}\}$ of $\{y_j\}$ such that $\{\|b(x_1, y_{j_k^{(1)}})\|\}$ is a convergent sequence. Again, since $\{\|b(x_2, y_{j_k^{(1)}}\|\}$ is also a bounded real number sequence, there is a subsequence $\{y_{j_k^{(2)}}\}$ of $\{y_{j_k^{(1)}}\}$ such that $\{\|b(x_2, y_{j_k^{(2)}}\|\}$ is also a convergent sequence. Continuing this construction and by the diagonal method, we can obtain a subsequence $\{y_{n_j}\}$ of $\{y_j\}$ such that, for each $i \in N$, $\{\|b(x_i, y_{n_j})\|\}$ is a convergent sequence. Therefore, we obtain

a principle submatrix $(\|b(x_{n_i}, y_{n_j})\|)_{ij}$ of $(\|b(x_i, y_j)\|)_{ij}$ such that, for each $i \in N$, $\lim_j \|b(x_{n_i}, y_{n_j})\|$ exists and $\|b(x_{n_i}, y_{n_j})\| < 1/2^{n_i}$ for all $j \in N$ holds. Since $\sum_j 1/2^j < \infty$, the transpose matrix $(\|b(x_{n_i}, y_{n_j})\|)'_{ij}$ of $(\|b(x_{n_i}, y_{n_j})\|)_{ij}$ satisfies the conditions of Lemma 1. So for each $i \in N$, $\lim_j \|b(x_{n_i}, y_{n_j})\|$ exists and converges uniformly with respect to $i \in N$. The necessity holds.

PROOF OF SUFFICIENCY. If b is not $\mathcal{N}$-hypocontinuous, then there are $\varepsilon_0 > 0$, $x_i \to 0$ in E, $A \in \mathcal{N}$ and $\{y_j\} \subseteqq A$ such that

$$\|b(x_i, y_i)\| \geqq \varepsilon_0, \qquad i \in N. \tag{1}$$

For the infinite matrix $(\|b(x_i, y_j)\|)_{ij}$, by the conditions of the theorem, there exists a principle submatrix $(\|b(x_{n_i}, y_{n_j})\|)_{ij}$ such that for each $i \in N$, $\lim_j \|b(x_{n_i}, y_{n_j})\|$ exists and converges uniformly with respect to $i \in N$. Denote $a_i = \lim_j \|b(x_{n_i}, y_{n_j})\|$. We shall prove that $\lim_i a_i = 0$. In fact, since $\lim_j \|b(x_{n_i}, y_{n_j})\| = a_i$ converges uniformly with respect to $i \in N$, for any $\varepsilon > 0$, there exists $j_0 \in N$ such that for each $i \in N$,

$$\left| \|b(x_{n_i}, y_{n_{j_0}})\| - a_i \right| < \varepsilon/2.$$

Notice that b is separately continuous and $x_{n_i} \to 0$ in E, therefore there exists $i_0 \in N$ such that, whenever $i \geqq i_0$, $\|b(x_{n_i}, y_{n_{j_0}})\| < \frac{\varepsilon}{2}$. Hence, whenever $i \geqq i_0$, we have

$$|a_i| \leqq \left| \|b(x_{n_i}, y_{n_{j_0}})\| - a_i \right| + \|b(x_{n_i}, y_{n_{j_0}})\| < \varepsilon.$$

That is $\lim_i a_i = 0$. Thus we have $\lim_{i,j} \|b(x_{n_i}, y_{n_j})\| = 0$. In particular,

$$\lim_i \|b(x_{n_i}, y_{n_i})\| = 0.$$

This contradicts (1) and the theorem is proved.

If G is a topological vector space, then from ([7], P55) we know that the vector topology of G can be generated by a family of paranorms. So we have

THEOREM 3. *Let E, F and G be topological vector spaces and b: $E \times F \to G$ be a separately continuous bilinear map, $\mathcal{N}$ a family of bounded subsets of F. Then b is sequentially $\mathcal{N}$-hypocontinuous if and only if for each $x_i \to 0$ in E and $A \in \mathcal{N}$ and each continuous paranorm $\|\cdot\|$ of G, if $\{y_j\} \subseteqq A$, then for the infinite matrix $(\|b(x_i, y_j)\|)_{ij}$ there is a principle submatrix $(\|b(x_{n_i}, y_{n_j})\|)_{ij}$ such that $\lim_j \|b(x_{n_i}, y_{n_j})\|$ exists for each $i \in N$ and converges uniformly with respect to $i \in N$.*

By the proof of Theorem 2, we have the following result.

THEOREM 4. *Let E, F and G be topological vector spaces and $b: E\times F\to G$ be a separately continuous bilinear map, $\mathcal{N}$ be a family of bounded subsets of F. If for each $x_i\to 0$ in E, $A\in\mathcal{N}$ and $\{y_j\}\subseteq A$, there is a subsequence $\{y_{n_j}\}$ of $\{y_j\}$ such that $\lim_i b(x_i, y_{n_j})=0$ converges uniformly with respect to $j\in N$, then b must be sequentially $\mathcal{N}$-hypocontinuous.*

Using Theorem 4 and the Basic Matrix Theorem, we can present several hypocontinuity type Corollaries.

COROLLARY 5. *Let E, F and G be topological vector spaces and $b: E\times F\to G$ be a separately continuous bilinear map, then b is sequentially $\mathcal{KS}(F)$-hypocontinuous.*

Corollary 5 now yields the following very interesting generalization of a classical result of Mazur–Orlicz on joint continuity ([4]).

COROLLARY 6. *Let E, F be paranormed spaces and F be a $\mathcal{K}$-space ([1] §3), $b: E\times F\to G$ be a separately continuous bilinear map. Then b is continuous (i.e., jointly continuous).*

COROLLARY 7. *Let E, F and G be topological vector spaces and E be an M-space ([5], §4), $b: E\times F\to G$ be a separately continuous bilinear map. Then b is sequentially $\mathcal{KB}(F)$-hypocontinuous.*

Corollaries 5 and 7 generalized the classical result of Bourbaki on hypocontinuity ([3] §40.2).

3. Uniform boundedness

We next consider the uniform boundedness for a family of bilinear maps which is pointwise bounded on $E\times F$. By the proof methods of Theorems 2 and 3, we have the following uniform boundedness principle.

THEOREM 8. *Let E, F and G be topological vector spaces, τ be a family of bilinear maps of $E\times F\to G$ which is pointwise bounded on $E\times F$. Then τ is uniformly bounded on $A\times B\subseteq E\times F$ if and only if for each sequence $\{b_i\}\subseteq\tau$ and each sequence $\{(x_j, y_j)\}\subseteq A\times B$ and each continuous paranorm $\|\cdot\|$ of G, then for the infinite matrix $(\|b_i(x_j, y_j)/i\|)_{ij}$ there is a principle submatrix $(\|b_{n_i}(x_{n_j}, y_{n_j})/n_i\|)_{ij}$ such that for each $i\in N$, $\lim_j \|b_{n_i}(x_{n_j}, y_{n_j})/n_i\|$ exists and converges uniformly with respect to $i\in N$.*

By Theorem 8, we have

THEOREM 9. *Let E, F and G be topological vector spaces, τ be a family of bilinear maps of $E\times F\to G$ which is pointwise bounded on $E\times F$, $A\subseteq E$, $y\in F$. Then τ is uniformly bounded on $A\times\{y\}$ if and only if for each sequence $\{b_i\}\subseteq\tau$ and each sequence $\{(x_j, y)\}\subseteq A\times\{y\}$ and each continuous*

paranorm $\|\cdot\|$ *of* G, *for the infinite matrix* $(\|b_i(x_j, y)/i\|)_{ij}$ *there is a principle submatrix* $(\|b_{n_i}(x_{n_j}, y)/n_i\|)_{ij}$ *such that for each* $i \in N$, $\lim_j \|b_{n_i}(x_{n_j}, y)/n_i\|$ *exists and converges uniformly with respect to* $i \in N$.

Using Theorem 9 and the Basic Matrix Theorem, we have

THEOREM 10. *Let* E, F *and* G *be topological vector spaces,* τ *be a family of separately continuous bilinear maps of* $E \times F \to G$ *which is pointwise bounded on* $E \times F$. *Then* τ *is uniformly bounded on each product* $A \times \{y\} \subseteq E \times F$ *when*

(i) A *is a* $\mathcal{K}$*-convergent sequence in* E,

(ii) A *is a* $\mathcal{K}$*-bounded subset of* E.

By Theorem 10, we can formulate several Corollaries contained in ([1], [2], [5]).

COROLLARY 11. *Let* τ *be as in Theorem 10. Then* τ *is uniformly bounded on each product* $A \times B \subseteq E \times F$ *when*

(i) $A \times B \in \mathcal{K}S(E) \times \mathcal{K}S(F)$.

(ii) $A \times B \in \mathcal{K}B(E) \times \mathcal{K}S(F)$.

(iii) $A \times B \in \mathcal{K}B(E) \times \mathcal{K}B(F)$.

COROLLARY 12. *Let* τ *be as in Theorem 10. If* E *and* F *are* $\mathcal{A}$*-spaces* ([6]), *then* τ *is uniformly bounded on products of bounded subsets of* E *and* F.

4. Equihypocontinuity

Let E, F and G be topological vector spaces, τ be a family of separately continuous bilinear maps of $E \times F \to G$, $\mathcal{N}$ be a family of bounded subsets of F. We consider the following types of continuity ([5] §4):

(S1) τ is sequentially left equicontinuous, i.e., if $x_i \to 0$ in E, then $\lim b(x_i, y) = 0$ uniformly for $b \in \tau$ for each $y \in F$.

(S2) τ is sequentially $\mathcal{N}$-equihypocontinuous, i.e., if for each $A \in \mathcal{N}$ when $x_i \to 0$ in E, then $\lim b(x_i, y) = 0$ uniformly for $b \in \tau$, $y \in A$.

In [5], Antosik and Swartz showed that if $\cup\mathcal{N} = F$, then (S2) implies (S1). However, (S1) does not imply (S2) when $\mathcal{N} = B(F)$.

In this section we also have

THEOREM 13. *Let* E, F *and* G *be topological vector spaces,* τ *be a family of sequentially left equicontinuous bilinear maps of* $E \times F \to G$, $\mathcal{N}$ *be a family of bounded subsets of* F. *Then* τ *is sequentially* $\mathcal{N}$*-equihypocontinuous if and only if for each sequence* $x_i \to 0$ *in* E *and each sequence* $\{b_i\} \subseteq \tau$ *and each continuous paranorm* $\|\cdot\|$ *of* G, *if* $A \in \mathcal{N}$ *and* $\{y_j\} \subseteq A$, *then for the infinite matrix* $(\|b_i(x_i, y_j)\|)_{ij}$ *there is a principle submatrix* $(\|b_{n_i}(x_{n_i}, y_{n_j})\|)_{ij}$ *such*

that $\liminf_j \|b_{n_i}(x_{n_i}, y_{n_j})\|$ *exists for each* $i \in N$ *and converges uniformly with respect to* $i \in N$.

From Theorem 13 and the Basic Matrix Theorem, we also have

COROLLARY 14. *Let* τ *be as in Theorem 13. Then* τ *is sequentially* $\mathcal{KS}(F)$*-equihypocontinuous.*

COROLLARY 15. *Let* E *be an* M*-space and* τ *be as in Theorem 13. Then* τ *is* $\mathcal{KB}(F)$*-equihypocontinuous.*

Corollaries 14 and 15 generalize the Corollaries 5 and 7.

The authors would like to thank the referee for his useful remarks and suggestions.

REFERENCES

[1] ANTOSIK, P. and SWARTZ, C., *Matrix methods in analysis*, Lecture Notes in Mathematics, 1113, Springer-Verlag, Berlin – New York, 1985. *MR* **87b**:46079

[2] SWARTZ, C., Continuity and hypocontinuity for bilinear maps, *Math. Z.* **186** (1984), 321–329. *MR* **85h**:46006

[3] KÖTHE, G., *Topological vector spaces*. II, Grundlehren der mathematischen Wissenschaften, Bd. 237, Springer-Verlag, Berlin – New York, 1979. *MR* **81g**:46001

[4] MAZUR, S. and ORLICZ, W., Über Folgen linearer Operationen, *Studia Math.* **4** (1933), 152–157. *Zbl* **8**, 250

[5] ANTOSIK, P. and SWARTZ, C., Boundedness and continuity for bilinear operators, *Studia Sci. Math. Hungar.* **29** (1994), 387–395. *MR* **95m**:47001

[6] LI, R. and SWARTZ, C., Spaces for which the uniform boundedness principle holds, *Studia Sci. Math. Hungar.* **27** (1992), 379–384. *MR* **94h**:46015

[7] WILANSKY, A., *Modern methods in topological vector spaces*, McGraw-Hill, International Book Co., New York, 1978. *MR* **81d**:46001

(Received July 5, 1996)

DEPARTMENT OF MATHEMATICS
DAQING PETROLEUM INSTITUTE
ANDA 151400
PEOPLE'S REPUBLIC OF CHINA

DEPARTMENT OF MATHEMATICS
HARBIN INSTITUTE OF TECHNOLOGY
HARBIN 150006
PEOPLE'S REPUBLIC OF CHINA

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 139–145

ON THE PRIMITIVE ROOTS AND THE QUADRATIC RESIDUES MODULO p

W. ZHANG

Abstract

The main purpose of this paper is to prove the following conclusion: Let $p \geqq 3$ be a prime. Then for any quadratic residue a mod p, there exists a pair of primitive roots g_1 and g_2 mod p such that $a \equiv g_1 g_2 (\mathrm{mod} p)$. Let $N(a,p)$ denote the number of the solutions of this congruence. Then we have

$$N(a,p) = \frac{\phi^2(p-1)}{p-1} \sum_{d|u} \frac{\mu^2(d)d}{\phi^2(d)} \prod_{\substack{q|\frac{p-1}{d} \\ (d,q)=1}} \left(1 - \frac{1}{(q-1)^2}\right),$$

where $u = (\mathrm{ind}\, a, p-1)$, ind a denotes the index of a relative to some fixed primitive root mod p.

1. Introduction

Let n be a positive integer, $p \geqq 3$ be a prime, $m = p^n$. It is well known that there exists at least one primitive root mod m, and the number of all primitive roots mod m is equal to $\phi(\phi(m))$. The main purpose of this paper is to study the following two questions:

(a) For each integer a with $(a,m) = 1$, is there a pair of primitive roots g_1 and g_2 mod m such that the congruence

$$a^2 \equiv g_1 g_2 (\mathrm{mod}\, m)?$$

(b) If (a) is true, let $N(a^2, m)$ denote the number of the solutions of the congruence in (a). What can be said about the asymptotic properties of $N(a^2, m)$?

About these two problems, it seems that no one has studied them yet, at least I have not seen it before. The problems are interesting because they can help us to find some new relationship between the primitive roots and the quadratic residues mod p. In this paper, we use estimates for Gauss sums and the method of trigonometric sums to study the above two problems, and prove the following main conclusions:

1991 *Mathematics Subject Classification.* Primary 11A07.
Key words and phrases. Primitive roots, quadratic residues, representation of products.

THEOREM. *Let n be a positive integer, p be an odd prime, $m=p^n$. Then for any integer c with $(c,m)=1$ we have*

$$N(c^2,m)=\frac{\phi^2(\phi(m))}{\phi(m)}\sum_{d|u}\frac{\mu^2(d)d}{\phi^2(d)}\prod_{\substack{q|\frac{\phi(m)}{d}\\(d,q)=1}}\left(1-\frac{1}{(q-1)^2}\right),$$

where the product is over all distinct prime divisor q of $\dfrac{\phi(m)}{d}$ with $(q,d)=1$, $\phi(m)$ is the Euler function, and $u=(\operatorname{ind} c^2,\phi(m))$, $\operatorname{ind} c^2$ denotes the index of c^2 relative to some fixed primitive root of mod m.

From this theorem, we may immediately deduce the following three corollaries:

COROLLARY 1. *Let n be a positive integer, p be an odd prime, $m=p^n$. Then for the finite field F_m with any element $0\neq a\in F_m$, there exists a pair of generators g_1 and $g_2\in F_m$ such that*

$$a^2=g_1g_2.$$

COROLLARY 2. *Let p be an odd prime. Then for any quadratic residue a* mod p, *there exist two primitive roots g_1 and g_2* mod p *such that*

$$a\equiv g_1g_2(\operatorname{mod} p).$$

COROLLARY 3. *Let p be an odd prime. Then for any quadratic non-residue a* mod p, *there exist three primitive roots g_1, g_2 and g_3* mod p *such that*

$$a\equiv g_1g_2g_3(\operatorname{mod} p).$$

2. Proof of the theorem

To complete the proof of the theorem, we need the following two elementary lemmas.

LEMMA 1. *Suppose for the modulus $m\geq 3$ there exists a primitive root. Then for each integer n with $(m,n)=1$, we have the identity*

$$\sum_{k|\phi(m)}\frac{\mu(k)}{\phi(k)}\sum_{\substack{a=1\\(a,k)=1}}^{k}e\left(\frac{a\operatorname{ind} n}{k}\right)=\begin{cases}\frac{\phi(m)}{\phi(\phi(m))} & \text{if } n \text{ is a primitive root of } m;\\ 0 & \text{otherwise,}\end{cases}$$

where $\mu(m)$ is the Möbius function, $e(y)=e^{2\pi iy}$ and ind n *denotes the index of n relative to some fixed primitive root of m.*

PROOF (see Proposition 2.2 of reference [2]).

LEMMA 2. *Let n be a positive integer, p be an odd prime and $m = p^n$. Then for any integer u with $(u, m) = 1$ we have*

$$\sum_{\substack{\chi \bmod m \\ \text{order of } \chi = k}} \chi(u) = \sum_{d \mid (\text{ind } u, k)} \mu\left(\frac{k}{d}\right) d,$$

where k is a divisor of $\phi(m)$, $\text{ind } u$ denotes the index of u relative to some fixed primitive root of m.

PROOF. For each Dirichlet character $\chi \bmod m$ with order $= k$, we know that there exists one and only one integer $1 \leqq r \leqq k$ with $(r, k) = 1$ such that

$$\chi(u) = e\left(\frac{r \text{ind } u}{k}\right). \tag{1}$$

We also have the trigonometric identity

$$\sum_{a=1}^{m} e\left(\frac{au}{m}\right) = \begin{cases} m & \text{if } m \mid u; \\ 0 & \text{if } m \nmid u. \end{cases} \tag{2}$$

From (1) and (2) we get

$$\begin{aligned} \sum_{\substack{\chi \bmod m \\ \text{order of } \chi = k}} \chi(u) &= \sum_{\substack{r=1 \\ (r,k)=1}}^{k} e\left(\frac{r \text{ind } u}{k}\right) = \sum_{d \mid k} \mu(d) \sum_{r=1}^{k/d} e\left(\frac{r \text{ind } u}{k/d}\right) \\ &= \sum_{d \mid k} \mu\left(\frac{k}{d}\right) \sum_{r=1}^{d} e\left(\frac{r \text{ind } u}{d}\right) = \sum_{\substack{d \mid k \\ d \mid \text{ind } u}} \mu\left(\frac{k}{d}\right) d \\ &= \sum_{d \mid (\text{ind } u, k)} \mu\left(\frac{k}{d}\right) d. \end{aligned}$$

This proves Lemma 2.

PROOF OF THE THEOREM. First we claim the following two facts:

(c) Let $m = p^n$. Then for each integer x with $(x, m) = 1$, there exists exactly one $\overline{x}$ with $1 \leqq \overline{x} \leqq m - 1$ such that $x\overline{x} \equiv 1 (\bmod m)$.

(d) If x is a primitive root mod m and $x\overline{x} \equiv 1 (\bmod m)$, then $\overline{x}$ is also a primitive root mod m.

From the trigonometric identity (2) we have

$$N(c^2, m) = \frac{1}{m} \sum_{a=1}^{m}{}^{*} \sum_{b=1}^{m}{}^{*} \sum_{u=1}^{m} e\left(\frac{u(c^2 - ab)}{m}\right), \tag{3}$$

where $\sum^*$ denotes the summation over all primitive roots mod m.

Note that there are $\phi(\phi(m))$ primitive roots mod m in the interval $[1, m]$. Separating the summation $\sum_{u=1}^{m}$ in (3) into two parts and applying (c), (d) we get

$$
\begin{aligned}
N(c^2, m) &= \frac{1}{m}\sum_{u=1}^{m}\sum_{a=1}^{m}{}^{*}\sum_{b=1}^{m}{}^{*} e\left(\frac{u(c^2-ab)}{m}\right)\\
&= \frac{\phi^2(\phi(m))}{m} + \frac{1}{m}\sum_{u=1}^{m-1}\sum_{a=1}^{m}{}^{*}\sum_{b=1}^{m}{}^{*} e\left(\frac{u(c^2-ab)}{m}\right)\\
&= \frac{\phi^2(\phi(m))}{m} + \frac{1}{m}\sum_{u=1}^{m-1}\sum_{a=1}^{m}{}^{*}\sum_{b=1}^{m}{}^{*} e\left(\frac{uc^2\overline{a}-ub}{m}\right) \qquad (4)\\
&= \frac{\phi^2(\phi(m))}{m} + \frac{1}{m}\sum_{u=1}^{m-1}\left(\sum_{a=1}^{m}{}^{*} e\left(\frac{uc^2\overline{a}}{m}\right)\right)\left(\sum_{b=1}^{m}{}^{*} e\left(\frac{-ub}{m}\right)\right)\\
&= \frac{\phi^2(\phi(m))}{m} + \frac{1}{m}\sum_{u=1}^{m-1}\left(\sum_{a=1}^{m}{}^{*} e\left(\frac{uc^2a}{m}\right)\right)\left(\sum_{b=1}^{m}{}^{*} e\left(\frac{-ub}{m}\right)\right).
\end{aligned}
$$

The map which takes a with $(a, m) = 1$ to $e\left(\frac{r\operatorname{ind} a}{k}\right)$ is a Dirichlet character when $k|\phi(m)$. We shall denote this by $\chi(a; r, k)$. Applying Lemma 1 we can get

$$
\begin{aligned}
&\frac{\phi^2(\phi(m))}{m} + \frac{1}{m}\sum_{u=1}^{m-1}\left(\sum_{a=1}^{m}{}^{*} e\left(\frac{uc^2a}{m}\right)\right)\left(\sum_{b=1}^{m}{}^{*} e\left(\frac{-ub}{m}\right)\right)\\
&= \frac{1}{m}\sum_{u=1}^{m}\left(\frac{\phi(\phi(m))}{\phi(m)}\sum_{k|\phi(m)}\frac{\mu(k)}{\phi(k)}\sum_{\substack{r=1\\(r,k)=1}}^{k}\sum_{\substack{a=1\\(a,m)=1}}^{m} e\left(\frac{r\operatorname{ind} a}{k}\right) e\left(\frac{uc^2a}{m}\right)\right)\times\\
&\quad\times\left(\frac{\phi(\phi(m))}{\phi(m)}\sum_{h|\phi(m)}\frac{\mu(h)}{\phi(h)}\sum_{\substack{s=1\\(s,h)=1}}^{h}\sum_{\substack{b=1\\(b,m)=1}}^{m} e\left(\frac{s\operatorname{ind} b}{h}\right) e\left(\frac{-ub}{m}\right)\right)\\
&= \frac{\phi^2(\phi(m))}{m\phi^2(m)}\sum_{k|\phi(m)}\sum_{h|\phi(m)}\frac{\mu(k)\mu(h)}{\phi(k)\phi(h)}\sum_{\substack{r=1\\(r,k)=1}}^{k}\sum_{\substack{s=1\\(s,h)=1}}^{h}\sum_{u=1}^{m}\left(\sum_{a=1}^{m}\chi(a; r, k) e\left(\frac{uc^2a}{m}\right)\right)\times
\end{aligned}
$$

$$\times \left(\sum_{b=1}^{m} \chi(b;s,h) e\left(\frac{-ub}{m}\right)\right)$$

(5)

$$= \frac{\phi^2(\phi(m))}{m\phi^2(m)} \sum_{k|\phi(m)} \sum_{h|\phi(m)} \frac{\mu(k)\mu(h)}{\phi(k)\phi(h)} \sum_{\substack{r=1\\(r,k)=1}}^{k} \sum_{\substack{s=1\\(s,h)=1}}^{h} \sum_{a=1}^{m} \sum_{b=1}^{m} \chi(a;r,k)\chi(b;s,h)\times$$

$$\times \left(\sum_{u=1}^{m} e\left(\frac{u(c^2a-b)}{m}\right)\right)$$

$$= \frac{\phi^2(\phi(m))}{\phi^2(m)} \sum_{k|\phi(m)} \sum_{h|\phi(m)} \frac{\mu(k)\mu(h)}{\phi(k)\phi(h)} \sum_{\substack{r=1\\(r,k)=1}}^{k} \sum_{\substack{s=1\\(s,h)=1}}^{h} \sum_{\substack{a=1\\ac^2\equiv b(m)}}^{m} \sum_{b=1}^{m} \chi(a;r,k)\chi(b;s,h)$$

$$= \frac{\phi^2(\phi(m))}{\phi^2(m)} \sum_{k|\phi(m)} \sum_{h|\phi(m)} \frac{\mu(k)\mu(h)}{\phi(k)\phi(h)} \sum_{\substack{r=1\\(r,k)=1}}^{k} \sum_{\substack{s=1\\(s,h)=1}}^{h} \sum_{a=1}^{m} \chi(a;r,k)\chi(ac^2;s,h)$$

$$= \frac{\phi^2(\phi(m))}{\phi^2(m)} \sum_{k|\phi(m)} \sum_{h|\phi(m)} \frac{\mu(k)\mu(h)}{\phi(k)\phi(h)} \sum_{\substack{r=1\\(r,k)=1}}^{k} \sum_{\substack{s=1\\(s,h)=1}}^{h} \sum_{a=1}^{m} \chi(a;r,k)\chi(a;s,h)\chi(c^2;s,h).$$

Note the character sum identity

$$\sum_{a=1}^{m} \chi(a;r,k)\chi(a;s,h) = \begin{cases} \phi(m) & \text{if } k=h \text{ and } r+s=k; \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Let $v = (\text{ind } c^2, \phi(m))$. Then from (4), (5), (6) and Lemma 2 we have

$$N(c^2,m) = \frac{\phi^2(\phi(m))}{\phi(m)} \sum_{k|\phi(m)} \frac{\mu^2(k)}{\phi^2(k)} \sum_{\substack{r=1\\(r,k)=1}}^{k} \chi(c^2;r,k)$$

$$= \frac{\phi^2(\phi(m))}{\phi(m)} \sum_{k|\phi(m)} \frac{\mu^2(k)}{\phi^2(k)} \sum_{d|(v,k)} \mu\left(\frac{k}{d}\right)d \tag{7}$$

$$= \frac{\phi^2(\phi(m))}{\phi(m)} \sum_{d|v} d \sum_{k|\frac{\phi(m)}{d}} \frac{\mu^2(dk)\mu(k)}{\phi^2(dk)}.$$

If $(d,k)>1$, then $\dfrac{\mu^2(kd)}{\phi^2(kd)} = 0$. If $(d,k)=1$, then $\dfrac{\mu^2(kd)}{\phi^2(kd)} = \dfrac{\mu^2(d)\mu^2(k)}{\phi^2(d)\phi^2(k)}$. Thus

from (7) we have

$$N(c^2,m)=\frac{\phi^2(\phi(m))}{\phi(m)}\sum_{d|v}\frac{\mu^2(d)d}{\phi^2(d)}\sum_{\substack{k|\frac{\phi(m)}{d}\\(k,d)=1}}\frac{\mu(k)}{\phi^2(k)}$$

$$=\frac{\phi^2(\phi(m))}{\phi(m)}\sum_{d|v}\frac{\mu^2(d)d}{\phi^2(d)}\prod_{\substack{q|\frac{\phi(m)}{d}\\(q,d)=1}}\left(1-\frac{1}{(q-1)^2}\right), \tag{8}$$

where the product is over all distinct prime divisors q of $\frac{\phi(m)}{d}$ with $(q,d)=1$. This proves the Theorem.

Proof of Corollary 1. Let n be a positive integer, p be an odd prime and $m=p^n$. Note that $2|\phi(m)$ and

$$\sum_{\substack{k|\frac{\phi(m)}{d}\\(k,d)=1}}\frac{\mu(k)}{\phi^2(k)}=\begin{cases}\displaystyle\prod_{\substack{q|\frac{\phi(m)}{d}\\(q,d)=1}}\left(1-\frac{1}{(q-1)^2}\right)>0 & \text{if } 2|d;\\ 0 & \text{if } 2\nmid d.\end{cases} \tag{8}$$

From $2|v=(\operatorname{ind}a^2,\phi(m))=(2\operatorname{ind}a,\phi(m))$, (7) and (8) we obtain

$$N(a^2,m)=\frac{\phi^2(\phi(m))}{\phi(m)}\sum_{d|v}\frac{\mu^2(d)d}{\phi^2(d)}\prod_{\substack{q|\frac{\phi(m)}{d}\\(q,d)=1}}\left(1-\frac{1}{(q-1)^2}\right)$$

$$=\frac{\phi^2(\phi(m))}{\phi(m)}\sum_{2d|v}\frac{\mu^2(2d)2d}{\phi^2(2d)}\prod_{\substack{q|\frac{\phi(m)}{2d}\\(q,2d)=1}}\left(1-\frac{1}{(q-1)^2}\right)$$

$$=\frac{2\phi^2(\phi(m))}{\phi(m)}\sum_{\substack{d|\frac{v}{2}\\(2,d)=1}}\frac{\mu^2(d)d}{\phi^2(d)}\prod_{\substack{q|\frac{\phi(m)}{2d}\\(q,2d)=1}}\left(1-\frac{1}{(q-1)^2}\right)>0.$$

This completes the proof of Corollary 1. Similarly, we can also deduce Corollary 2 and Corollary 3.

Acknowledgements. The author expresses his gratitude to Professor Carl Pomerance for his helpful and detailed comments.

REFERENCES

[1] SCHMIDT, W. M., *Equations over finite fields. An elementary approach*, Lecture Notes in Mathematics, Vol. 536, Springer-Verlag, Berlin, 1976. *MR* **55** #2744

[2] NARKIEWICZ, W., *Classical problems in number theory*, PWN Polish Scientific Publishers, Warszawa, 1987, 79–80.

[3] APOSTOL, T. M., *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976. *MR* **55** #7892

(Received July 10, 1996)

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF GEORGIA
ATHENS, GA 30602
U.S.A.

Present address:

DEPARTMENT OF MATHEMATICS
NORTHWEST UNIVERSITY
XI'AN SHAANXI
PEOPLE'S REPUBLIC OF CHINA

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 147–150

FUNDAMENTAL REDUCIBILITY OF NORMAL OPERATORS ON KREIN SPACE

Ts. BAYASGALAN

Abstract

In the present paper we study fundamental reducibility of normal operators on Krein space. Fundamental reducibility of selfadjoint operators on Krein space, also fundamental reducibility of normal operators on Pontrjagin space has been studied in [1], [2]. For basic definitions and facts on Krein spaces and operators in these spaces we refer to [3].

LEMMA 1. *Let $\{T_i\}_{i=1}^n$ be a finite family of commuting fundamentally reducible bounded selfadjoint operators in Krein space H. Then there exists some fundamental decomposition of H, which reduces every operator from the family.*

PROOF. Assume that $\alpha \in \mathbb{C}$, $\alpha \neq \overline{\alpha}$ and

$$\|T_i\|_J < |\alpha| \quad (i = 1, 2, \ldots, n).$$

Then the operators $(T_i - \alpha I)^{-1}$ $(i = 1, 2, \ldots, n)$ are bounded. Setting

$$U_i = I + (\alpha - \overline{\alpha})(T_i - \alpha I)^{-1} \quad (i = 1, 2, \ldots, n),$$

we find that the family $\{U_i\}_{i=1}^n$ satisfies the conditions of Theorem 2.17 from [1]. Consequently, there exists some fundamental decomposition of H which reduces every operator from the family $\{U_i\}_{i=1}^n$, i.e.,

$$H = H^+[\dot{+}]H^-, \quad U_i H^{\pm} \subset H^{\pm} \quad (i = 1, 2, \ldots, n).$$

Thus from

$$T_i = \alpha I + (\alpha - \overline{\alpha})(U_i - I)^{-1}$$

it follows that $T_i H^{\pm} \subset H^{\pm}$ $(i = 1, 2, \ldots, n)$.

NOTE. For an infinite family of operators a similar statement is not true in general.

The proof of the following Lemma is clear.

1991 *Mathematics Subject Classification.* Primary 46C20.
Key words and phrases. Krein space, numerical ranges of operators.

The author would like to express his deep gratitude to doctor P. Jonas (Berlin, Germany) for his helpful comments, suggestions and discussions.

LEMMA 2. *Let T be a bounded normal operator in Krein space. Then T is fundamentally reducible if and only if the operators*

$$\operatorname{Re} T := \frac{T+T^+}{2}, \qquad \operatorname{Im} T := \frac{T-T^+}{2i}$$

are fundamentally reducible, where T^+ is the adjoint operator of T.

[1] and Lemma 2 imply the following

THEOREM 1. *Let T be a bounded normal operator in Krein space. The operator T is fundamentally reducible if and only if the conditions*

$$\sigma(\operatorname{Re} T) \subset \mathbb{R}, \qquad \sigma(\operatorname{Im} T) \subset \mathbb{R}, \tag{1}$$

$$\|(\operatorname{Re} T - i\gamma)^{-k}\|_J \leqq \frac{C_1}{|\gamma|^k},$$

$$\|(\operatorname{Im} T - i\gamma)^{-k}\|_J \leqq \frac{C_2}{|\gamma|^k}, \tag{2}$$

$$(k = 1, 2, \ldots,\ \gamma \neq 0,\ \gamma \in \mathbb{R})$$

are fulfilled, where $\|\cdot\|_J$ is the J-norm and C_1, C_2 are constants.

It is well known that a selfadjoint operator in Krein space is fundamentally reducible if and only if it is J-selfadjoint for some fundamental symmetry J. Also, it is fundamentally reducible if and only if it is similar to a J-selfadjoint operator. Analogous facts are true for unitary operators in Krein space.

But for normal operators we only have

LEMMA 3. *Let T be a bounded normal operator in Krein space. If T is fundamentally reducible, then T is J-normal.*

EXAMPLE. We set $H = \mathbb{C}^2$,

$$\begin{aligned} [(x_1, x_2), (y_1, y_2)] &= x_1\overline{y}_1 - x_2\overline{y}_2, \\ \langle (x_1, x_2), (y_1, y_2) \rangle &= x_1\overline{y}_1 + x_2\overline{y}_2, \end{aligned} \qquad ((x_1, x_2) \in \mathbb{C}^2,\ \ (y_1, y_2) \in \mathbb{C}^2),$$

$$T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then we have $T = T^+$, $TT^+ = T^+T$, but T is not fundamentally reducible.

COROLLARY. *If $T \neq 0$ is a bounded normal operator in Krein space such that $r(T) = 0$, where $r(T)$ is the spectral radius of T, then T is not fundamentally reducible. In particular, if $\dim H = 2$ and $T \neq 0$ is a bounded non-negative operator in H, then T is fundamentally reducible if and only if $T^2 \neq 0$.*

It is well known that if T is a bounded selfadjoint operator in H, λ is an eigenvalue of T and $\ker(T - \lambda I)$ is ortho-complemented, then λ is real and semi-simple.

Analogous fact is true for isometric operators. In case $\dim H < \infty$ these conditions are known to be sufficient, too, in order that $\ker(T - \lambda I)$ would be ortho-complemented.

LEMMA 4. *Let T be a bounded normal operator in Krein space and let $\ker(T-\lambda I)$ be ortho-complemented. Then λ is a semi-simple eigenvalue of T.*

PROOF. Let

$$H = \ker(T - \lambda I)[\dot{+}](\ker(T - \lambda I))^{\perp}.$$

Suppose that $(T - \lambda I)^r x = 0$. Then we have

$$(T - \lambda I)^{r-1} x_2 \in \ker(T - \lambda I),$$

where

$$x = x_1 + x_2, \quad x_1 \in \ker(T - \lambda I), \quad x_2 \in (\ker(T - \lambda I))^{\perp}.$$

On the other hand, because T is normal, we find

$$(T - \lambda I)(\ker(T - \lambda I))^{\perp} \subset (\ker(T - \lambda I))^{\perp}.$$

Thus we have

$$(T - \lambda I)^{r-1} x_2 \in (\ker(T - \lambda I))^{\perp}.$$

Consequently, $(T - \lambda I)^{r-1} x = 0$.

LEMMA 5. *Suppose $\dim H < \infty$, T is a bounded normal operator and λ is an eigenvalue of T. Then $\ker(T - \lambda I)$ is ortho-complemented if and only if the conditions*

(1) *λ is a semi-simple eigenvalue of T,*

(2) $\ker(T - \lambda I) = \ker(T^+ - \overline{\lambda} I)$

are fulfilled.

PROOF. If $\ker(T - \lambda I)$ is ortho-complemented, then λ is semi-simple by Lemma 4. Because $\ker(T - \lambda I)$ is ortho-complemented, hence $R(T^+ - \overline{\lambda} I)$ is non-degenerate. Let $x \in \ker(T - \lambda I)$. Then we have for every $y \in H$

$$\begin{aligned}\left[(T^+ - \overline{\lambda} I)x, (T^+ - \overline{\lambda} I)y\right] &= \left[(T^+ - \overline{\lambda} I)(T - \lambda I)x, y\right] \\ &= \left[(T - \lambda I)x, (T - \lambda I)y\right] = 0.\end{aligned}$$

Thus we obtain $\ker(T - \lambda I) \subset \ker(T^+ - \overline{\lambda} I)$.

Consequently, from

$$\dim \ker(T^+ - \overline{\lambda} I) = \dim \ker(T - \lambda I)$$

it follows that $\ker(T^+ - \overline{\lambda} I) = \ker(T - \lambda I)$.

Let, conversely, conditions (1)–(2) of Lemma 5 be fulfilled. If

$$\left[(T^+ - \overline{\lambda} I)x, (T^+ - \overline{\lambda} I)y\right] = 0 \quad \text{for every } y \in H,$$

then $(T-\lambda I)^2 x=0$, hence $(T-\lambda I)x=0$. Thus $R(T^+-\overline{\lambda}I)$ is non-degenerate.

EXAMPLE. Let $\dim H=2$. Assume that $(l_i)_{i=1}^2$ is a basis in H such that $[l_i,l_i]=0$ $(i=1,2)$. Define

$$Tl_1=il_1, \quad Tl_2=(-i)l_2.$$

Then $T=T^+$ and $\lambda=\pm i$ are semi-simple eigenvalues of T. But, we have $\ker(T-iI)\neq\ker(T+iI)$, and $\ker(T-\lambda I)$ is not ortho-complemented.

NOTE. If T is a fundamentally reducible operator in Krein space, then $\ker(T-\lambda I)$ is ortho-complemented for every $\lambda\in\mathbb{C}$. If H is a Pontrjagin space, then the bounded normal operator T is fundamentally reducible if and only if $\ker(T-\lambda I)$ is ortho-complemented for every $\lambda\in\mathbb{C}$ (see [1]).

THEOREM 2. *Let H be a Pontrjagin space and let T be a compact normal operator. Then T is fundamentally reducible if and only if the conditions*

(1) $\ker T$ *is ortho-complemented,*

(2) *every eigenvalue of T is semi-simple,*

(3) $\ker(T-\lambda I)=\ker(T^+-\overline{\lambda}I)$ $(\lambda\in\mathbb{C},\ \lambda\neq 0)$

are fulfilled.

PROOF. If T is fundamentally reducible, then by Lemma 3 T is a J-normal operator. Consequently, we have

$$\ker(T-\lambda I)=\ker(T^*-\overline{\lambda}I)=\ker(T^+-\overline{\lambda}I).$$

Let, conversely, the conditions of the theorem be fulfilled. The subspace $\ker(T-\lambda I)$ is ortho-complemented because $R(T^+-\overline{\lambda}I)$ is closed and $R(T^+-\overline{\lambda}I)$ is nondegenerate.

REFERENCES

[1] BAYASGALAN, TS., Fundamental reducibility of selfadjoint and unitary operators in spaces with an indefinite metric, *Studia Sci. Math. Hungar.* **20** (1985), 313–321 (in Russian). *MR* **88j**:47048

[2] MCENNIS, B. W., Fundamental reducibility of selfadjoint operators on Krein space, *J. Operator Theory* **8** (1982), 219–225. *MR* **84a**:47047

[3] BOGNÁR, J., *Indefinite inner product spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 78, Springer, Berlin, 1974. *MR* **57** #7125

(Received July 24, 1996)

DEPARTMENT OF MATHEMATICS
MONGOLIAN STATE UNIVERSITY
P.O. BOX 46/880
ULAAN–BAATAR
MONGOLIA

mnu_smcs@magicnet.mn

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 151–154

ON A PROBLEM OF DICKMEIS AND NESSEL CONCERNING THE APPROXIMATION BY BERNSTEIN POLYNOMIALS

L. IMHOF

Abstract

For $f \in C[0,1]$ let $B_n(f;x)$ denote the nth Bernstein polynomial and $\omega^*(f;t)$ the second modulus of smoothness. Continuing the investigations by W. Dickmeis and R. J. Nessel it is shown that for each abstract modulus of continuity ω there exists a counter-example $f_\omega \in C[0,1]$ such that on the one hand $\omega^*(f_\omega;t) = \mathcal{O}(\omega(t))$ and on the other hand $\limsup_{n\to\infty} |B_n(f_\omega;x) - f_\omega(x)|/\omega(x(1-x)/n) \geqq c > 0$ simultaneously for all $x \in (0,1)$. Furthermore, a pointwise lethargy assertion is established.

For $f \in C$, the space of continuous functions on [0,1], the Bernstein polynomials are defined by

$$B_n(f;x) := \sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right)\binom{n}{\nu} x^\nu (1-x)^{n-\nu}.$$

Let $\omega^*(f;t)$ denote the second modulus of smoothness, thus

$$\omega^*(f;t) := \sup_{0\leqq h\leqq t} \sup_{h\leqq x\leqq 1-h} |f(x-h) - 2f(x) + f(x+h)|, \qquad t \geqq 0,$$

and let ω be an abstract modulus of continuity, thus a continuous, non-decreasing, subadditive function on $[0,\infty)$ with $\omega(0) = 0$ and (additionally) $\lim_{t\to 0+} \omega(t)/t = \infty$. Then

$$\omega^*(f;t) = \mathcal{O}(\omega(t^2)) \quad \Longrightarrow \quad |B_n(f;x) - f(x)| \leqq M_f \omega(x(1-x)/n)$$

(cf. [1], [2, p. 308]). Against this background W. Dickmeis and R. J. Nessel posed the following problem (cf. [6]): Given ω, does there exist a counter-example $f_\omega \in C$ such that $\omega^*(f_\omega;t) = \mathcal{O}(\omega(t^2))$, but

$$\limsup_{n\to\infty} \frac{|B_n(f_\omega;x) - f_\omega(x)|}{\omega(x(1-x)/n)} \geqq c > 0 \tag{1}$$

simultaneously for all $x \in (0,1)$? In [5] a counter-example is constructed such that (1) is valid for all points x of a dense set of second category in $(0,1)$, while it is shown in [3] and [4] that there exists f_ω satisfying (1) almost everywhere, provided $\omega(t) = t^\alpha$, $0 < \alpha < 1$.

1991 *Mathematics Subject Classification.* Primary 41A25, 41A36.

Key words and phrases. Bernstein polynomials, pointwise error bound, interpolatory polynomials, quantitative resonance principle, sharpness everywhere.

THEOREM. *For each abstract modulus ω there exists a function $f_\omega \in C$ such that $\omega^*(f_\omega; t) = \mathcal{O}(\omega(t))$ and (1) holds simultaneously for all $x \in (0,1)$.*

The proof of this theorem is based on a convergence property of certain interpolation polynomials due to P. Erdős ([7, Theorem 3], see also [10, p. 53]) and on a quantitative extension of the classical resonance principle (cf. [3], [8]). To formulate the extended principle let $\|f\|$ denote the maximum (over [0,1]) norm of $f \in C$, and let C^* designate the set of all non-negative-valued, sublinear, bounded functionals F on C, thus

$$0 \leqq F(f+g) \leqq Ff + Fg, \quad F(af) = |a|Ff \qquad (f, g \in C,\ a \in \mathbb{R}),$$

$$\|F\|_{C^*} := \sup\{Ff : \|f\| \leqq 1\} < \infty.$$

In these terms one has the

RESONANCE PRINCIPLE. *For arbitrary index sets $\mathbf{T}$ and $\mathbf{R}_n$ $(n \in \mathbb{N})$ consider $U_t, V_{n,x} \in C^*$ $(t \in \mathbf{T},\ x \in \mathbf{R}_n,\ n \in \mathbb{N})$, a positive function σ on $\mathbf{T}$, and a null sequence $\tau_1 > \tau_2 > \cdots > 0$. Suppose there exist test elements $g_n \in C$ such that $(n \in \mathbb{N})$*

$$\|g_n\| \leqq c_1, \tag{2}$$

$$U_t g_n \leqq c_2 \min\{1, \sigma(t)/\tau_n\} \quad \text{for } t \in \mathbf{T}, \tag{3}$$

$$\|V_{n,x}\|_{C^*} \leqq c_3 \quad \text{for } x \in \mathbf{R}_n, \tag{4}$$

$$V_{n,x} g_k \leqq c_{4,k}\tau_n \quad \text{for } x \in \mathbf{R}_n,\ k < n, \tag{5}$$

$$V_{n,x} g_n \geqq c_5 > 0 \quad \text{for } x \in \mathbf{R}_n. \tag{6}$$

Then for each abstract modulus ω there exist naturals $n_1 < n_2 < \cdots$ and a counter-example $f_\omega \in C$ such that $U_t f_\omega \leqq 6c_2\omega(\sigma(t))$ for $t \in \mathbf{T}$ and

$$\limsup_{n\to\infty} \frac{V_{n,x} f_\omega}{\omega(\tau_n)} \geqq c_5 \quad \text{for } x \in \limsup_{k\to\infty} \mathbf{R}_{n_k}.$$

PROOF OF THE THEOREM. Consider the following quantities:

$$U_t f := \omega^*(f; t), \quad \sigma(t) := t, \quad t \in (0, 1/2] =: \mathbf{T}, \qquad \tau_n := 1/n,$$

$$V_{n,x} f := \max_{j=n,2n} |B_j(f; x) - f(x)|,$$

$$x \in \mathbf{R}_n := \left[\frac{1}{2} - \frac{1}{2\sqrt[2n]{2}}, \frac{1}{2} + \frac{1}{2\sqrt[2n]{2}}\right].$$

In view of the Erdős result mentioned there exist polynomials $g_1, g_2, \ldots$ having three properties: for every $n \in \mathbb{N}$, (i) the degree of g_n does not exceed $10n$; (ii) $g_n(k/2n) = 0$ or 1 according as k is even or odd $(k = 0, 1, \ldots, 2n)$; (iii) $\sup\{|g_n(x)| : -1 \leqq x \leqq 2\} \leqq c_1$ for some constant c_1. Obviously, the g_n

satisfy (2). Using elementary properties of the modulus (cf. [2, p. 44ff]) as well as Bernstein's inequality (cf. [9, p. 118f]) one obtains that

$$U_t g_n = \omega^*(g_n;t) \leqq \begin{cases} 4\|g_n\| \leqq 4c_1 \\ \\ 2t \sup\limits_{x\in[0,1]} |g_n'(x)| \leqq 20tn \sup\limits_{x\in[-1,2]} |g_n(x)| \leqq 20c_1\sigma(t)/\tau_n, \end{cases}$$

establishing (3) with $c_2 := 20c_1$. Since the Bernstein operators are positive and linear, one infers from $B_n 1 = 1$ that (4) holds for $c_3 := 2$. By Popoviciu's inequality (cf. [2, pp. 308, 330]), $\|g_k - B_n g_k\| \leqq \|g_k''\|/n$; thus (5) holds with $c_{4,k} := \|g_k''\|$. Finally, on account of the identities

$$1 = \left(\sum_{\nu=0,2,\ldots,2n} + \sum_{\nu=1,3,\ldots,2n-1} \right) \binom{2n}{\nu} x^\nu (1-x)^{2n-\nu} =: \Sigma_1(x) + \Sigma_2(x),$$

$$(1-2x)^{2n} = \Sigma_1(x) - \Sigma_2(x),$$

it follows that for $x \in \mathbf{R}_n$

$$B_{2n}(g_n;x) = \Sigma_2(x) = \frac{1-(1-2x)^{2n}}{2} \geqq \frac{1}{4},$$

and since $B_n g_n = 0$ one has $|B_j(g_n;x) - g_n(x)| \geqq 1/8 =: c_5$ for $j = n$ or $2n$ according as $|g_n(x)| \geqq 1/8$ or $\leqq 1/8$, respectively. This entails (6). Now an application of the resonance principle completes the proof. Note that $\limsup_{k\to\infty} \mathbf{R}_{n_k} = (0,1)$ whatever the sequence $n_1 < n_2 < \cdots$. □

In a similar way, employing [8, Corollary 2.2] instead of the resonance principle, one may obtain the following lethargy assertion.

COROLLARY. *For each null sequence $\varepsilon = (\varepsilon_n)_{n=1}^\infty \subset (0,\infty)$ there exists a counter-example $f_\varepsilon \in C$ such that*

$$\limsup_{n\to\infty} \varepsilon_n^{-1} |B_n(f_\varepsilon;x) - f_\varepsilon(x)| \geqq 1 \qquad \textit{for all } x \in (0,1).$$

ACKNOWLEDGEMENT. I would like to thank Professor R. J. Nessel and the referee for several helpful comments.

REFERENCES

[1] BERENS, H. and LORENTZ, G. G., Inverse theorems for Bernstein polynomials, *Indiana Univ. Math. J.* **21** (1972), 693–708. *MR* **45** #5638

[2] DEVORE, R. A. and LORENTZ, G. G., *Constructive approximation*, Grundlehren der mathematischen Wissenschaften, Bd. 303, Springer-Verlag, Berlin, 1993. *MR* **95f**:41001

[3] DICKMEIS, W., *Ein quantitatives Resonanzprinzip und Schärfe von punktweisen Fehlerabschätzungen fast überall*, Habilitationsschrift, RWTH Aachen, 1983.

[4] DICKMEIS, W., On quantitative condensation of singularities on sets of full measure, *Approx. Theory Appl.* **1** (1985), no. 5, 71–84. *MR* **87m**:41020

[5] DICKMEIS, W. and NESSEL, R. J., Condensation principles with rates, *Studia Math.* **75** (1982), 55–68. *MR* **84a**:41041

[6] DICKMEIS, W. and NESSEL, R. J., Problem 6: approximation by Bernstein polynomials, *Second Edmonton conference on approximation theory* (Edmonton, Alta., 1982), CMS Conf. Proc., **3**, Amer. Math. Soc., Providence, RI, 1983, 392–393. See: *MR* **84k**:41002

[7] ERDŐS, P., On some convergence properties of the interpolation polynomials, *Ann. of Math.* (2) **44** (1943), 330–337. *MR* **4**, 273e

[8] IMHOF, L. and NESSEL, R. J., A resonance principle with rates in connection with pointwise estimates for the approximation by interpolation processes, *Numer. Funct. Anal. Optim.* **16** (1995), 139–152. *MR* **95m**:41003

[9] NATANSON, I. P., *Konstruktive Funktionentheorie*, Akademie-Verlag, Berlin, 1955. *MR* **16**, 1100d

[10] SZABADOS, J. and VÉRTESI, P., *Interpolation of functions*, World Scientific Publishing Co., Inc., Teaneck, NJ, 1990. *MR* **92j**:41009

(Received July 26, 1996)

LEHRSTUHL A FÜR MATHEMATIK
RHEINISCH-WESTFÄLISCHE
TECHNISCHE HOCHSCHULE AACHEN
D–52056 AACHEN
GERMANY

imhof@stochastik.rwth-aachen.de

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 155–174

REMARKS ON THE MINIMAL RINGS OF CONVEX BODIES

MARIA MOSZYŃSKA

Abstract

Relationships between the minimal ring of direct sum and the minimal rings of summands are discussed. The minimal ring and its centre are proved to be continuous with respect to the Hausdorff metric.

The notion of minimal ring containing the boundary of a convex body A in $\mathbf{R}^n$ appeared in the literature already in 1924 for $n=2$ (see [2]). Its short history can be found in [1].

Bárány in [1] proved the existence and uniqueness of the minimal ring and its centre for arbitrary n.

Recently, it was proved that the centre of the minimal ring is selfdual with respect to polarity (see [6, Theorem 3.6]).

The purpose of the present paper is to study some properties of minimal rings. Section 1 concerns relationships between the minimal ring of the Euclidean direct sum $A_1 \oplus A_2$ in $\mathbf{R}^n$ for $A_i \subset E_i$ and the minimal rings of A_1 and A_2 in the linear subspaces E_1 and E_2, respectively. In Section 2 we prove that the minimal ring of a parallel body of A has the same centre and thickness as the minimal ring of A itself. In Section 3 we prove the continuity of minimal ring and its centre with respect to the Hausdorff metric.

The author is grateful to the referee of the first version of the paper for his/her valuable remarks and to Krzysztof Rudnik for his helpful assistance.

0. Preliminaries

We use the following terminology and notation.

Let ϱ be the Euclidean metric in $\mathbf{R}^n$.

Let $\mathcal{C}^n, \mathcal{K}^n$, and $\mathcal{K}_0^n$ be, respectively, the class of all non-empty compact subsets of $\mathbf{R}^n$, the class of all non-empty compact convex subsets of $\mathbf{R}^n$, and the class of convex bodies:

1991 *Mathematics Subject Classification.* Primary 52A20; Secondary 52A99.

Key words and phrases. Minimal ring of a convex body, centre of the minimal ring, Euclidean direct sum, parallel body.

0081–6906/99/$ 5.00

$$\mathcal{K}_0^n = \{A \in \mathcal{K}^n; \operatorname{int} A \neq \emptyset\}.$$ [1]

For any $x \in \mathbf{R}^n$ and $\alpha > 0$, let $B(x,\alpha)$ and $B_0(x,\alpha)$ be, respectively, the closed ball and the open ball with centre x and radius α. For convenience, singletons can be treated as degenerate balls (with radius 0).

For any $A \in \mathcal{K}_0^n$ and $x \in A$, let

$$R_A(x) := \inf\{\alpha; B(x,\alpha) \supset A\} \quad \text{and} \quad r_A(x) := \sup\{\alpha; B(x,\alpha) \subset A\}.$$

By Theorem 1 of Bárány [1], there exists a unique point of A at which $R_A - r_A$ attains its minimal value. As in [6], we denote this point by $c(A)$. It is the *centre of minimal (spherical) ring* containing $\operatorname{bd} A$,

$$\operatorname{ring} A := B(c(A), R(A)) \setminus B_0(c(A), r(A)),$$

where

$$R(A) := R_A(c(A)) \quad \text{and} \quad r(A) := r_A(c(A)).$$ [2]

We shall need the following theorem of Bárány.

THEOREM 0.1 ([1,Theorem 2]). *Let $A \in \mathcal{K}_0^n$ and $x_0 \in \mathbf{R}^n$. Then, $x_0 = c(A)$ if and only if there exist*

$$p_1, \ldots, p_k \in \operatorname{bd} A \cap \operatorname{bd} B(x_0, r(A, x_0))$$

and

$$q_1, \ldots, q_l \in \operatorname{bd} A \cap \operatorname{bd} B(x_0, R(A, x_0))$$

such that

$$\operatorname{conv}\left\{\frac{p_s - x_0}{r(A,x_0)}; s = 1, \ldots, k\right\} \cap \operatorname{conv}\left\{\frac{q_t - x_0}{R(A,x_0)}; t = 1, \ldots, l\right\} \neq \emptyset.$$

Let us note the following two simple facts.

PROPOSITION 0.2 (B. Zdrodowski). *For any triangle A in $\mathbf{R}^2$, $c(A)$ is the intersection of bisectrix of the longest side and bisectrix of the smallest angle of A.*

PROPOSITION 0.3. *If a convex body A in $\mathbf{R}^n$ is symmetric with respect to an affine subspace E of dimension $k \in \{0, \ldots, n-1\}$, then $c(A) \in E$.*

The first statement follows from Theorem 0.1. The second one is a direct consequence of the uniqueness of the centre of the minimal ring and its equivariance under isometries:

[1] R. Schneider in [8] refers to any compact convex set in $\mathbf{R}^n$ as a convex body.

[2] Bárány in [1] evidently assumes all the compact convex sets under consideration to have non-empty interiors, though he does not write it explicitly.

if $f : \mathbf{R}^n \to \mathbf{R}^n$ is an isometry, then

$$f(c(A)) = c(f(A)) \quad \text{for every} \quad A \in \mathcal{K}_0^n.$$

It is clear that for $n \geqq 2$ we can replace $\mathbf{R}^n$ by its arbitrary affine subspace E of dimension $k < n$. Let $\mathcal{K}(E)$ and $\mathcal{K}_0(E)$ be the counterparts of $\mathcal{K}^n$ and $\mathcal{K}_0^n$. Then, for the elements of $\mathcal{K}_0(E)$ we have the corresponding notions of minimal ring and its centre, ring_E and c_E.

We shall often use the notation $r(A, x)$ instead of $r_A(x)$, and $R(A, x)$ instead of $R_A(x)$. Consequently, for a convex body A in a subspace E, we write $r_E(A, x)$ and $R_E(A, x)$ for the radii of the suitable balls in E. We write $B_E(x, \alpha)$ for the ball with centre x and radius α, and bd_E for the boundary in E.

For any $\varepsilon > 0$ and $A \subset \mathbf{R}^n$, let A_ε be the *(outer) parallel body of A at distance ε* (compare, e.g.,[8]):

$$A_\varepsilon := \{x \in \mathbf{R}^n; \varrho(x, A) \leqq \varepsilon\}.$$

In particular, for $a \in \mathbf{R}^n$,

$$\{a\}_\varepsilon = B(a, \varepsilon).$$

It is well known and easy to check that for A closed

$$A_\varepsilon = A + \{0\}_\varepsilon.$$

Following [8], we use the symbol $A_{-\varepsilon}$ to denote the *inner parallel set of A at distance ε*:

$$A_{-\varepsilon} := \{x \in A; \{x\}_\varepsilon \subset A\}.$$

The symbols B^n and S^{n-1} are used for the unit ball and the unit sphere in $\mathbf{R}^n$:

$$B^n = \{0\}_1 \quad \text{and} \quad S^{n-1} = \mathrm{bd} B^n.$$

Let us recall that the Hausdorff metric ϱ_H in the class $\mathcal{C}^n$ (and, generally, in the class of compact subsets of any metric space) is defined by the formula

$$\varrho_H(A, B) := \max\{\sup_{x \in A} \varrho(x, B), \ \sup_{y \in B} \varrho(y, A)\},$$

or, equivalently,

$$\varrho_H(A, B) = \inf\{\varepsilon > 0; \ A \subset B_\varepsilon \ \& \ B \subset A_\varepsilon\}. \tag{0.1}$$

We use the symbol $\oplus$ for the direct sum of orthogonal linear subspaces of $\mathbf{R}^n$, and, generally, for subsets of such subspaces. More precisely, for $A_1, A_2 \subset \mathbf{R}^n$, let $A = A_1 \oplus A_2$ if and only if $A = A_1 + A_2$ and there exist orthogonal

linear subspaces E_1, E_2 of $\mathbf{R}^n$ such that $A_i \subset E_i$ for $i = 1, 2$ (compare [4] and [5]). We shall refer to $\oplus$ as the *Euclidean direct sum.*[3]

For any $A \in \mathcal{K}^n$ and $x \in \mathbf{R}^n$ there exists a unique point $\xi_A(x) \in A$ with

$$\varrho(x, \xi_A(x)) = \varrho(x, A).$$

The function $\xi_A : \mathbf{R}^n \to A$ is referred to as *metric projection.*

Following [8], we use the notation

$$\operatorname{pos} v := \{\lambda v; \lambda \geqq 0\}$$

for every non-zero vector $v \in \mathbf{R}^n$.

We denote by $H(A, v)$ the supporting hyperplane of A with outer normal vector v.

By $\operatorname{Nor}(A, p)$ we denote the normal cone of A at $p \in \operatorname{bd} A$ (i.e., the set of outer normal vectors to A at p).

The symbol $\Delta(a_1, \ldots, a_m)$ denotes the simplex with vertices $a_1, \ldots, a_m$; in particular, $\Delta(a_1, a_2)$ is the segment with endpoints a_1, a_2.

The usual scalar product of $x, y \in \mathbf{R}^n$ is denoted by $x \circ y$.

We shall need the following simple fact.

PROPOSITION 0.4. *Let $(f_k : \mathbf{R}^n \to \mathbf{R})_{k \in N}$ be a sequence of continuous functions uniformly convergent to f_0. If for every $k \in N \cup \{0\}$ there exists a unique point $x_k \in \mathbf{R}^n$ such that*

$$\forall x \quad f_k(x_k) \leqq f_k(x), \tag{0.2}$$

and the sequence $(x_k)_{k \in N}$ is bounded, then it is convergent and $\lim_k x_k = x_0$.

PROOF. If $(x_k)_{k \in N}$ is convergent, then evidently,

$$\lim_k f_k(x_k) = f_0(\lim_k x_k);$$

thus, passing to the limit in (0.2), we obtain

$$f_0(\lim_k x_k) \leqq f_0(x) \text{ for every } x,$$

whence $\lim_k x_k = x_0$ by the uniqueness of x_0. By the same argument we infer that any convergent subsequence of (x_k) has x_0 as the limit. Hence the sequence (x_k) is convergent, because it is bounded. □

[3] Using the notation of [5], we have $\oplus := \oplus_f$, where $f(t_1, t_2) = \sqrt{t_1^2 + t_2^2}$.

1. Minimal ring for Euclidean direct sum of convex bodies

We are interested in relationships between $\operatorname{ring}_{E_i} A_i$ for $i = 1, 2$ and $\operatorname{ring}(A_1 \oplus A_2)$.

Let us start with the following

PROPOSITION 1.1. *Let $A \in \mathcal{K}_0^n$, $A = A_1 \oplus A_2$, where $A_i \subset E_i$ for $i = 1, 2$, and let $a = a_1 + a_2$, $a_i \in A_i$. If, for $i = 1, 2$,*

$$\alpha_i = r_{E_i}(A_i, a_i), \quad \beta_i = R_{E_i}(A_i, a_i),$$

$$P_i = \{p \in \operatorname{bd}_{E_i}(A_i); \varrho(p, a_i) = \alpha_i\},$$

$$Q_i = \{q \in \operatorname{bd}_{E_i}(A_i); \varrho(q, a_i) = \beta_i\},$$

$$P = \{p \in \operatorname{bd} A; \varrho(p, a) = r(A, a)\} \text{ and } Q = \{q \in \operatorname{bd} A; \varrho(q, a) = R(A, a)\},$$

then

(i) $r(A, a) = \min\{\alpha_1, \alpha_2\}$,

(ii) $R(A, a) = \sqrt{\beta_1^2 + \beta_2^2}$,

(iii) $$P = \begin{cases} P_1 + a_2 & \text{if } \alpha_1 < \alpha_2 \\ (P_1 + a_2) \cup (a_1 + P_2) & \text{if } \alpha_1 = \alpha_2 \end{cases},$$

and

(iv) $Q = Q_1 + Q_2$.

PROOF. (i) By the assumptions,

$$B(a_i, \alpha_i) \subset A_i \quad \text{for } i = 1, 2 \tag{1.1}$$

and

$$P_1 \neq \emptyset \neq P_2.$$

We may assume that $\alpha_1 \leqq \alpha_2$. It suffices to prove

$$B(a, \alpha_1) \subset A \tag{1.2}$$

and

$$\varrho(p, a) = \alpha_1 \quad \text{for some point } p \in \operatorname{bd} A. \tag{1.3}$$

Let $\varrho(x, a) \leqq \alpha_1$ for some $x \in \mathbf{R}^n$ and let $x = x_1 + x_2$, where $x_i \in E_i$ for $i = 1, 2$. Then

$$\varrho(x_i, a_i) \leqq \sqrt{\varrho(x_1, a_1)^2 + \varrho(x_2, a_2)^2} = \varrho(x, a),$$

whence, by (1.1), $x_i \in A_i$ for $i = 1, 2$. This proves (1.2).

Let $p = p_1 + a_2$, where $p_1 \in P_1$.

Since

$$\operatorname{bd} A = ((\operatorname{bd}_{E_1} A_1) \oplus A_2) \cup (A_1 \oplus \operatorname{bd}_{E_2} A_2), \tag{1.4}$$

it follows that

$$p \in \operatorname{bd} A \quad \text{and} \quad \varrho(p, a) = \alpha_1 = \min\{\alpha_1, \alpha_2\}.$$

This proves (1.3).

(ii) By the assumptions,

$$B(a_i, \beta_i) \supset A_i \quad \text{for} \quad i = 1, 2 \tag{1.5}$$

and

$$Q_1 \neq \emptyset \neq Q_2.$$

It suffices to prove that

$$B\left(a, \sqrt{\beta_1^2 + \beta_2^2}\right) \supset A \tag{1.6}$$

and

$$q \in \operatorname{bd} A \quad \text{and} \quad \varrho(q, a) = \sqrt{\beta_1^2 + \beta_2^2} \quad \text{for some } q. \tag{1.7}$$

Let $x \in A$; then $x = x_1 + x_2$, where $x_i \in A_i$. Thus, by (1.5),

$$\varrho(x, a) = \sqrt{\varrho(x_1, a_1)^2 + \varrho(x_2, a_2)^2} \leqq \sqrt{\beta_1^2 + \beta_2^2},$$

which proves (1.6).

Let $q_i \in Q_i$ and let $q = q_1 + q_2$. Then q satisfies (1.7).

We have already proved the inclusion $\supset$ in (iii) and (iv). The easy proof of $\subset$ is left to the reader. □

Let us prove the following

THEOREM 1.2. *Let* $\mathbf{R}^n = E_1 \oplus E_2$, $\dim E_i \geqq 1$, *and let* A_i *be a convex body in* E_i *for* $i = 1, 2$, *with* $r_{E_1}(A_1) = r_{E_2}(A_2)$. *If at least one of the sets* A_1, A_2 *is centrally symmetric, then*

$$c(A_1 \oplus A_2) = c_{E_1}(A_1) + c_{E_2}(A_2).$$

PROOF. Let A_2 be centrally symmetric. Since c is equivariant under translations, we can assume that O is its centre of symmetry:

$$A_2 = -A_2. \tag{1.8}$$

Let $A := A_1 \oplus A_2$. Since c is equivariant under homotheties, we can assume that $r_{E_i}(A_i) = 1$. Let $\beta_i := R_{E_i}(A_i)$ and $\beta = R(A, 0)$. Then, by Proposition 1.1,

$$r(A) = 1 \text{ and } \beta = \sqrt{\beta_1^2 + \beta_2^2}.$$

By Theorem 0.1, for $i = 1, 2$ there exist finite sets

$$X_i \subset \operatorname{bd}_{E_i}(A_i) \cap \operatorname{bd} B_{E_i}(0, 1) \quad \text{and} \quad Y_i \subset \operatorname{bd}_{E_i}(A_i) \cap \operatorname{bd} B_{E_i}(0, \beta_i)$$

such that $\operatorname{conv} X_i \cap \operatorname{conv} \dfrac{Y_i}{\beta_i} \neq \emptyset$.

Let

$$a_i \in \operatorname{conv} X_i \cap \operatorname{conv} \frac{Y_i}{\beta_i} \quad \text{for} \quad i = 1, 2. \tag{1.9}$$

By (1.8), we can assume that $X_2 = -X_2$ and $Y_2 = -Y_2$, whence

$$0 \in \operatorname{conv} X_2 \cap \operatorname{conv} Y_2. \tag{1.10}$$

By (1.9) and (1.10),

$$\frac{\beta_1}{\beta} a_1 \in \operatorname{conv} \frac{Y_1 + Y_2}{\beta}. \tag{1.11}$$

Since $0 \in \operatorname{conv} X_2 \subset \operatorname{conv}(X_1 \cup X_2)$, it follows that

$$\frac{\beta_1}{\beta} \operatorname{conv}(X_1 \cup X_2) \subset \operatorname{conv}(X_1 \cup X_2);$$

thus, by (1.9),

$$\frac{\beta_1}{\beta} a_1 \in \operatorname{conv}(X_1 \cup X_2). \tag{1.12}$$

Since

$$X_1 \cup X_2 \subset S^{n-1} \cap \operatorname{bd} A \quad \text{and} \quad \frac{Y_1 \cup Y_2}{\beta} \subset S^{n-1} \cap \operatorname{bd} A,$$

in view of Theorem 0.1, conditions (1.11) and (1.12) imply the required assertion $c(A) = 0$. □

The following statement concerning cylinders is a direct consequence of Theorem 1.2.

COROLLARY 1.3. *Let* $A = A_1 \oplus A_2$, $A_i \subset E_i$ *for* $i = 1, 2$, *where at least one of* E_1, E_2 *is one-dimensional. If* $r_{E_1}(A_1) = r_{E_2}(A_2)$, *then*

$$c(A) = c_{E_1}(A_1) + c_{E_2}(A_2).$$

We shall now show that in Corollary 1.3 the assumption concerning small radii of the minimal rings is essential.

EXAMPLE 1.4. Let $n = 3$, and let e_1, e_2, e_3 be the unit vectors:

$$e_i = (\delta_i^1, \delta_i^2, \delta_i^3), \text{ where } \quad \delta_i^j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

Let, further,

$$E_1 = \operatorname{Lin}(e_1) \quad \text{and} \quad E_2 = \operatorname{Lin}(e_2, e_3).$$

We define A_1 and A_2 as follows: A_1 is the segment in E_1, with centre O and length equal to 2; A_2 is a non-isosceles acute angled triangle in E_2, whose circumscribed circle has centre O, and $r_{E_2}(A_2, O) > 1$ (Fig. 1). Then

$$c(A_1 \oplus A_2) = O. \tag{1.13}$$

Indeed, let $q_1, \ldots, q_6$ be the vertices of $A_1 \oplus A_2$; evidently,

$$q_i \in \operatorname{bd} B(O, R(A_1 \oplus A_2, O)) \cap \operatorname{bd}(A_1 \oplus A_2) \quad \text{for all } i,$$

whence

$$\operatorname{conv}\left\{\frac{q_i}{\|q_i\|}; i = 1, \ldots, 6\right\} = (R(A_1 \oplus A_2, O))^{-1}(A_1 \oplus A_2).$$

On the other hand, by Proposition 1.1, $r(A_1 \oplus A_2, O) = 1$ and thus

$$B(O, r(A_1 \oplus A_2, O)) \cap \operatorname{bd}(A_1 \oplus A_2) = \{p_1, p_2\}, \quad \text{where } p_i = ((-1)^i, 0, 0);$$

thus $\operatorname{conv}\{p_1, p_2\} = A_1$ and the two convex hulls have non-empty intersection. This proves (1.13).

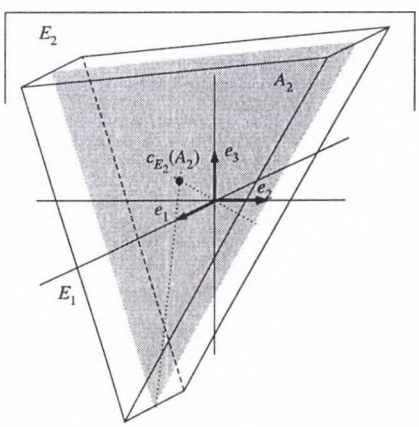


Fig. 1

Evidently, $c_{E_2}(A_2) \neq 0$, since, by Proposition 0.2, the point $c_{E_2}(A_2)$ differs from the centre of the circle circumscribed about A_2. □

Hence, also in Theorem 1.2 the assumption concerning small radii is essential. We are now going to prove that the assumption on A_1 or A_2 to be centrally symmetric also cannot be omitted. The counterexample (Example 1.7) is based on the following observation concerning the Euclidean direct sum of two isometric convex bodies.

PROPOSITION 1.5. *Let A_0 be a convex body in $\mathbf{R}^m$, $m \geq 2$, such that $c(A_0) = 0$, and let $\mathbf{R}^n = E_1 \oplus E_2$ where $\dim E_1 = m = \dim E_2$.*

If $A_i = f_i(A_0)$ for some linear isometry $f_i : \mathbf{R}^m \to E_i$, $i = 1, 2$, then the following two conditions are equivalent:

(i) $\quad c(A_1 \oplus A_2) = c_{E_1}(A_1) + c_{E_2}(A_2)$;

(ii) $\quad \operatorname{conv} \dfrac{P_0}{r(A_0)} \cap \sqrt{2} \operatorname{conv} \dfrac{Q_0}{R(A_0)} \neq \emptyset$,

where

$$P_0 = \{p \in \operatorname{bd} A_0; \|p\| = r(A_0)\}$$

and

$$Q_0 = \{q \in \operatorname{bd} A_0; \|q\| = R(A_0)\}.$$

PROOF. Since c is equivariant under isometries, without any loss of generality we can assume that

$$E_1 = \operatorname{Lin}(e_1, \ldots, e_m), \quad E_2 = \operatorname{Lin}(e_{m+1}, \ldots, e_n). \tag{1.14}$$

Then, we can treat $\mathbf{R}^n$ as $\mathbf{R}^m \times \mathbf{R}^m$, and thus we use the notation

$$(a, b) := (a_1, \ldots, a_m, b_1, \ldots, b_m)$$

for $a = (a_1, \ldots, a_m)$, $b = (b_1, \ldots, b_m) \in \mathbf{R}^m$.

Further, we can assume that for every $x \in \mathbf{R}^m$

$$f_1(x) = (x, 0), \quad f_2(x) = (0, x),$$

whence $f := f_2 f_1^{-1}$ is the reflection with respect to the linear subspace $E := \{(x, x); x \in \mathbf{R}^m\}$, i.e.,

$$f(x_1, \ldots, x_n) = (x_{m+1}, \ldots, x_n, x_1, \ldots, x_m) \quad \text{for every } (x_1, \ldots, x_n) \in \mathbf{R}^n. \tag{1.15}$$

Evidently,

$$c_{E_i}(A_i) = 0 \quad \text{for } i = 1, 2. \tag{1.16}$$

Let

$$P_i = f_i(P_0), \quad Q_i = f_i(Q_0) \quad \text{for } i = 1, 2,$$

and let

$$Z := \operatorname{conv} \frac{P_1 \cup P_2}{r(A_0)} \cap \operatorname{conv} \frac{Q_1 + Q_2}{\sqrt{2}R(A_0)}.$$

Then, by Theorem 0.1 combined with Proposition 1.1, Condition (1.16), and with the invariance of all the involved notions under isometries, (i) is equivalent to

$$Z \neq \emptyset. \tag{1.17}$$

In turn, since Z is convex and, by (1.15), it is symmetric with respect to E, it follows that (1.17) is equivalent to the condition

$$Z \cap E \neq \emptyset.$$

Hence, condition (i) holds if and only if there exist $p_1, \dots, p_k \in P_0$ and $\lambda_1^{(i)}, \dots, \lambda_k^{(i)} \geqq 0$ for $i = 1, 2$ such that $\Sigma_{j,i} \lambda_j^{(i)} = 1$ and

$$r(A_0)^{-1}(\Sigma_j \lambda_j^{(1)}(p_j, 0) + \Sigma_j \lambda_j^{(2)}(0, p_j)) \in \operatorname{conv} \frac{Q_1 + Q_2}{\sqrt{2}R(A_0)} \cap E.$$

It is easy to see that this holds if and only if there exist $p_j \in P_0$ and $\lambda_j \geqq 0$ for $j = 1, \dots, k$ such that

$$\Sigma_j \lambda_j = \frac{1}{2} \text{ and } a := \frac{\Sigma_j \lambda_j p_j}{r(A_0)} \in \operatorname{conv} \frac{Q_0}{\sqrt{2}R(A_0)}. \tag{1.18}$$

But $a \in \frac{1}{2} \operatorname{conv} \frac{P_0}{r(A_0)}$, whence (1.18) means that

$$\frac{1}{2} \operatorname{conv} \frac{P_0}{r(A_0)} \cap \operatorname{conv} \frac{Q_0}{\sqrt{2}R(A_0)} \neq \emptyset,$$

which is equivalent to (ii). □

We shall apply Proposition 1.5 for $n = 4$, i.e., for direct sum of two congruent plane convex bodies. We need the following elementary lemma.

LEMMA 1.6. *Let* $a_1, a_2, b_1, b_2 \in S^1$, $a_1 \neq a_2$, $b_1 \neq b_2$ *and*

$$\Delta(a_1, a_2) \cap \Delta(b_1, b_2) \neq \emptyset.$$

(i) *If* $b_1 \circ b_2 > 0$, *then*

$$\Delta(a_1, a_2) \cap \sqrt{2}\Delta(b_1, b_2) = \emptyset.$$

(ii) *If* $b_1 \circ b_2 < 0$ *and* $a_1 \circ b_1 = a_1 \circ b_2 > 0$, *then*

$$\Delta(a_1, a_2) \cap \sqrt{2}\Delta(b_1, b_2) \neq \emptyset.$$

PROOF. Evidently,

$$\sqrt{2}\Delta(b_1, b_2) \cap S^1 \neq \emptyset \quad \Leftrightarrow \quad \frac{\sqrt{2}}{2}(b_1 + b_2) \in B^2 \quad \Leftrightarrow \quad b_1 \circ b_2 \leqq 0.$$

This proves (i).

(ii) Let $x \in \Delta(a_1, a_2) \cap \Delta(b_1, b_2)$ and $b = \frac{1}{2}(b_1 + b_2)$. Since

$$\frac{\sqrt{2}}{2}(b_1 + b_2) \in \Delta(a_1, b) \cap \sqrt{2}\Delta(b_1, b_2),$$

by the Pasch Theorem, $\sqrt{2}\Delta(b_1, b_2) \cap \Delta(a_1, x) \neq \emptyset$, which proves (ii). □

EXAMPLE 1.7. Let $q_1, q_2 \in S^1$ with $0 < q_1 \circ q_2 < 1$, and let $p_1 = -p_2 = \frac{1}{2}(q_1 + q_2)$. Consider the ball

$$B = \frac{1}{2}(1 + \|p_1\|)B^2$$

and the strip

$$C = \{x \in \mathbf{R}^2; |\, x \circ p_1 \,| \leqq \|p_1\|^2\},$$

and let

$$A_0 := \operatorname{conv}(B \cap C \cup \{q_1, q_2\})$$

(Fig. 2). Then $c(A_0) = 0$. Thus, the points $a_i := \dfrac{p_i}{\|p_i\|}$ and $b_i := q_i$, $i = 1, 2$, satisfy the assumptions of Lemma 1.6 (i). Hence

$$\frac{\Delta(p_1, p_2)}{r(A_0)} \cap \sqrt{2}\frac{\Delta(q_1, q_2)}{R(A_0)} = \emptyset.$$

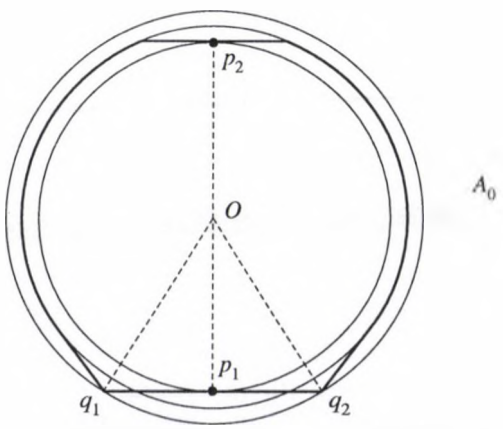


Fig. 2

Let now A_1 and A_2 be isometric copies of A_0 contained in two orthogonal linear subspaces E_1 and E_2 of $\mathbf{R}^4$. Then, by Proposition 1.5,

$$c(A_1 \oplus A_2) \neq c_{E_1}(A_1) + c_{E_2}(A_2).$$

□

Let us note that the idea of construction of the set A_0 in Example 1.7 is based on the idea of proof of Theorem 3 in [1].

Finally, we complete this section with a very simple positive result.

PROPOSITION 1.8. *If A_1 and A_2 are congruent triangles in orthogonal two-dimensional linear subspaces E_1 and E_2 of $\mathbf{R}^4$, then*

$$c(A_1 \oplus A_2) = c_{E_1}(A_1) + c_{E_2}(A_2).$$

PROOF. Let $A_i = f_i(A_0)$, where $A_0 = \Delta(a_1, a_2, a_3) \subset \mathbf{R}^2$. $f_i : \mathbf{R}^2 \to E_i$ is a linear isometry for $i = 1, 2$, and $c(A_0) = 0$.

Let, again,

$$P_0 = \{p \in \operatorname{bd} A_0; \|p\| = r(A_0)\}$$

and

$$Q_0 = \{q \in \operatorname{bd} A_0; \|q\| = R(A_0)\}.$$

By Proposition 1.5, it suffices to prove that

$$\operatorname{conv} \frac{P_0}{r(A_0)} \cap \sqrt{2} \operatorname{conv} \frac{Q_0}{R(A_0)} \neq \emptyset. \tag{1.19}$$

Case 1. $\varrho(a_1, a_2) \leqq \varrho(a_2, a_3) < \varrho(a_3, a_1)$. Then by Proposition 0.2,

$$Q_0 = \{a_1, a_3\} \quad \text{and} \quad P_0 = \{p_1, p_2\}, \quad \text{where } p_1 = \frac{1}{2}(a_1 + a_3).$$

Thus, by Theorem 0.1,

$$\frac{\Delta(p_1, p_2)}{\|p_i\|} \cap \frac{\Delta(a_1, a_3)}{\|a_1\|} \neq \emptyset. \tag{1.20}$$

Since, evidently, $a_1 \circ a_3 < 0$ and $p_1 \circ a_1 = p_1 \circ a_3 > 0$, by Lemma 1.6 (ii) it follows that

$$\frac{\Delta(p_1, p_2)}{r(A_0)} \cap \sqrt{2} \frac{\Delta(a_1, a_3)}{R(A_0)} \neq \emptyset, \tag{1.21}$$

i.e., (1.19) is satisfied.

Case 2. $\varrho(a_1, a_2) \leqq \varrho(a_2, a_3) = \varrho(a_3, a_1)$.

Then, by Proposition 0.2,

$$Q_0 = \{a_1, a_2, a_3\} \quad \text{and} \quad P_0 \supset \{p_1, p_2\},$$

where

$$p_1 = \frac{1}{2}(a_1 + a_3) \quad \text{and} \quad p_2 = \frac{1}{2}(a_2 + a_3).$$

Since $p_1 - p_2 \parallel a_1 - a_2$ and, by Theorem 0.1,

$$\frac{\Delta(p_1, p_2)}{r(A_0)} \cap \frac{\Delta(a_1, a_2, a_3)}{R(A_0)} \neq \emptyset,$$

it follows that (1.20) is satisfied. Since $a_1 \circ a_3 < 0$ and $p_1 \circ a_1 = p_1 \circ a_3 > 0$, as in Case 1, we obtain (1.21), which implies (1.19). □

2. Minimal ring of a parallel body

There is a simple and natural connection between ring A and ring A_ε: these two rings have common centre and the same thickness (Theorem 2.4). To prove this result, let us start with the following simple lemma, which is a direct consequence of additivity of the support function with respect to the Minkowski addition (see [8, Theorem 1.7.5(a)]).

LEMMA 2.1. *For every* $A, B \in \mathcal{K}^n$

$$A_\varepsilon \subset B_\varepsilon \quad \Leftrightarrow \quad A \subset B.$$

PROPOSITION 2.2. *Let* $A \in \mathcal{K}_0^n$. *For every* $x \in A$ *and* $\varepsilon > 0$

$$R_{A_\varepsilon}(x) = R_A(x) + \varepsilon \quad \text{and} \quad r_{A_\varepsilon}(x) = r_A(x) + \varepsilon.$$

PROOF. By Lemma 2.1,

$$\{\alpha > 0;\, A \subset B(x, \alpha)\} = \{\alpha > 0;\, A_\varepsilon \subset B(x, \alpha + \varepsilon)\},$$

whence

$$R_A(x) = R_{A_\varepsilon}(x) - \varepsilon.$$

Similarly,

$$\{\beta \geqq 0;\, B(x, \beta) \subset A\} = \{\beta \geqq 0;\, B(x, \beta + \varepsilon) \subset A_\varepsilon\},$$

whence

$$r_A(x) = r_{A_\varepsilon}(x) - \varepsilon.$$

□

PROPOSITION 2.3. *Let* $A \in \mathcal{K}_0^n$ *and* $\varepsilon > 0$. *For every* $x \in A_\varepsilon$ *there exists an* $a \in A$ *such that*

(i) $R_{A_\varepsilon}(a) \leqq R_{A_\varepsilon}(x)$

and

(ii) $r_{A_\varepsilon}(a) \geqq r_{A_\varepsilon}(x)$.

PROOF. Let $a := \xi_A(x)$. The assertions (i) and (ii) are trivial for $x \in A$. Assume that $x \in A_\varepsilon \setminus A$. Let x' be a point of A_ε most distant from x. Then

$$A_\varepsilon \subset B(x, \varrho(x', x)). \tag{2.2}$$

Let us notice that

$$\varrho(x', x) > \varepsilon. \tag{2.3}$$

Indeed, $x \in A_\varepsilon$ and the halfline $L = x + \mathrm{pos}(a - x)$ intersects $\mathrm{bd}\, A_\varepsilon$ in a point z satisfying

$$\varepsilon < \varrho(x, z) \leqq \varrho(x, x').$$

Thus (2.3) holds.

By Lemma 2.1, conditions (2.2) and (2.3) imply

$$A \subset B(x, \varrho(x', x) - \varepsilon). \tag{2.4}$$

Since $\sphericalangle(xap) \geqq \frac{\pi}{2}$, it follows that for every $p \in A$

$$\varrho(p, a) \leqq \varrho(p, x),$$

Thus, by (2.4),

$$A \subset B(a, \varrho(x', x) - \varepsilon),$$

whence

$$A_\varepsilon \subset B(a, \varrho(x', x)) = B(a, R_{A_\varepsilon}(x)).$$

This proves (i).

To verify (ii), notice first that

$$r_{A_\varepsilon}(a) \geqq \varepsilon$$

because $a \in A$. On the other hand,

$$r_{A_\varepsilon}(x) < \varepsilon;$$

indeed, let y be the point of intersection of $\mathrm{bd}\,(A_\varepsilon)$ with the halfline $L' := a + \mathrm{pos}(x - a)$; then

$$r_{A_\varepsilon}(x) \leqq \varrho(x, y) = \varrho(y, a) - \varrho(x, a) < \varrho(y, a),$$

where $\varrho(y, a) = \varrho(y, A) \leqq \varepsilon$, since $a = \xi_A(y)$ (cf. Theorem 4.8. (12) of [3]). □

As a direct consequence of Propositions 2.2 and 2.3, we obtain the following.

THEOREM 2.4. *For every* $A \in \mathcal{K}_0^n$ *and* $\varepsilon > 0$,
(i) $c(A_\varepsilon) = c(A)$,
(ii) $R(A_\varepsilon) = R(A) + \varepsilon$ *and* $r(A_\varepsilon) = r(A) + \varepsilon$.

For arbitrary $A \in \mathcal{K}_0^n$, the functions $R_A : A \to \mathbf{R}$ and $r_A : A \to \mathbf{R}$ can be extended over the whole $\mathbf{R}^n$ as follows (compare [1, p. 96]).

DEFINITION 2.5. For every $x \in \mathbf{R}^n$, let

$$F_A(x) := \{v \circ (p - x); p \in \operatorname{bd} A \text{ and } v \in S^{n-1} \cap \operatorname{Nor}(A, p)\},$$

$$\bar{R}_A(x) := \sup F_A(x), \quad \text{and} \quad \bar{r}_A(x) := \inf F_A(x).$$

As was noticed by Bárány (see the remark following Claim 3 of [1]), $\bar{R}_A$ is convex and $\bar{r}_A$ is concave.

We are now going to generalize Proposition 2.2 on the extended functions $\bar{R}_A$ and $\bar{r}_A$. Let us start with the following simple lemmas.

LEMMA 2.6. *For every* $A \in \mathcal{K}_0^n$ *and* $x \in \mathbf{R}^n$,

$$\bar{r}_A(x) = \delta(x)\varrho(x, \operatorname{bd} A),$$

where

$$\delta(x) = \begin{cases} 1 & \text{for } x \in A \\ -1 & \text{for } x \notin A. \end{cases}$$

PROOF. For $x \in A$ the assertion is obvious. Let $x \in \mathbf{R}^n \setminus A$. Then $\delta(x) = -1$ and $\varrho(x, \operatorname{bd} A) = \varrho(x, A)$, whence it suffices to prove that

$$-\varrho(x, A) = \inf\{v \circ (p - x); p \in \operatorname{bd} A \text{ and } v \in S^{n-1} \cap \operatorname{Nor}(A, p)\},$$

which is equivalent to the condition

$$\varrho(x, A) = \sup\{v \circ (x - p); p \in \operatorname{bd} A \text{ and } v \in S^{n-1} \cap \operatorname{Nor}(A, p)\}. \tag{2.5}$$

Evidently, for every $v \in S^{n-1}$,

$$v \circ (x - p) \leqq \varrho(x, p)$$

and the equality holds whenever

$$x - p = \varrho(x, p)v.$$

Thus, the right side of (2.5) is equal to $\varrho(x, p)$ if and only if $x - p \in \operatorname{Nor}(A, p)$, i.e., $p = \xi_A(x)$. This proves (2.5). □

LEMMA 2.7. *For every* $A \in \mathcal{K}_0^n$ *and* $x \in \mathbf{R}^n$,

$$\bar{R}_A(x) = \inf\{\alpha > 0; A \subset B(x, \alpha)\}.$$

PROOF. For $x \in A$ the formula is the definition of R_A.

Let $x \in \mathbf{R}^n \setminus A$. Evidently,

$$\sup F_A(x) = \varrho(x,p),$$

where $\varrho(x,p) = \sup\{\varrho(x,y); y \in \operatorname{bd} A\}$ and hence

$$\frac{p-x}{\varrho(p,x)} \in S^{n-1} \cap \operatorname{Nor}(A,p).$$

Then

$$A \subset B(x,\alpha_0) \quad \text{for} \quad \alpha_0 := \varrho(p,x)$$

and, moreover,

$$\alpha_0 = \inf\{\alpha > 0; A \subset B(x,\alpha)\}.$$

□

The following Proposition is a generalization of Proposition 2.2.

PROPOSITION 2.8. *Let $A \in \mathcal{K}_0^n$. For every $x \in \mathbf{R}^n$ and $\varepsilon > 0$,*

$$\bar{R}_{A_\varepsilon}(x) = \bar{R}_A(x) + \varepsilon \quad \textit{and} \quad \bar{r}_{A_\varepsilon}(x) = \bar{r}_A(x) + \varepsilon.$$

PROOF. By Lemma 2.7, the proof of the first formula is the same as for Proposition 2.2. It remains to prove the second one for $x \in \mathbf{R}^n \setminus A$.

It is easy to show that

$$\varrho(x, \operatorname{bd}(A_\varepsilon)) = \begin{cases} \varepsilon - \varrho(x, \operatorname{bd} A) & \text{if } x \in A_\varepsilon \setminus A, \\ \varrho(x, \operatorname{bd} A) - \varepsilon & \text{if } x \in \mathbf{R}^n \setminus A_\varepsilon \end{cases}$$

(cf. [3, Cor. 4.9]). Thus, by Lemma 2.6,

$$\bar{r}_{A_\varepsilon}(x) = \bar{r}_A(x) + \varepsilon \quad \text{for every} \quad x \in \mathbf{R}^n \setminus A.$$

□

3. Continuity of the minimal ring

Let us apply Proposition 2.8 to prove continuity of $\bar{R}$ and $\bar{r}$ with respect to the Hausdorff limit (Theorem 3.2). To this aim let us note a simple consequence of Lemmas 2.6 and 2.7.

LEMMA 3.1. *Let $A_1, A_2 \in \mathcal{K}_0^n$ and $A_1 \subset A_2$. Then, for every $x \in \mathbf{R}^n$,*

$$\bar{R}_{A_1}(x) \leqq \bar{R}_{A_2}(x) \quad \textit{and} \quad \bar{r}_{A_1}(x) \leqq \bar{r}_{A_2}(x).$$

THEOREM 3.2. *Let $A, A_k \in \mathcal{K}_0^n$ for every k. If $A = \lim_H A_k$, then $\bar{R}_{A_k}$ and $\bar{r}_{A_k}$ are uniformly convergent to $\bar{R}_A$ and $\bar{r}_A$, respectively.*

PROOF. Take an $\varepsilon > 0$. By the assumption, there exists $k_0 \in \mathbf{N}$ such that

$$A \subset (A_k)_\varepsilon \quad \text{and} \quad A_k \subset (A)_\varepsilon \quad \text{for every } k \geqq k_0.$$

Thus, by Lemma 3.1 and Proposition 2.8, for every $k \geqq k_0$ and $x \in \mathbf{R}^n$,

$$\bar{R}_A(x) \leqq \bar{R}_{A_k}(x) + \varepsilon \quad \text{and} \quad \bar{R}_{A_k}(x) \leqq \bar{R}_A(x) + \varepsilon$$

and, similarly,

$$\bar{r}_A(x) \leqq \bar{r}_{A_k}(x) + \varepsilon \quad \text{and} \quad \bar{r}_{A_k}(x) \leqq \bar{r}_A(x) + \varepsilon.$$

This completes the proof. □

We shall now prove that the centre of the minimal ring is continuous.

THEOREM 3.3. *The function $c : \mathcal{K}_0^n \to \mathbf{R}^n$ is continuous.*

PROOF. Let $A_k \in \mathcal{K}_0^n$ for $k \in \mathbf{N} \cup \{0\}$ and let $A = \lim_H A_k$. We consider the sequence $(f_k : \mathbf{R}^n \to \mathbf{R})_{k \in \mathbf{N}}$ and the function $f_0 : \mathbf{R}^n \to \mathbf{R}$ defined by the formulae:

$$f_k := \bar{R}_{A_k} - \bar{r}_{A_k} \quad \text{and} \quad f_0 = \bar{R}_A - \bar{r}_A.$$

Since, for every $k \in \mathbf{N} \cup \{0\}$, the function f_k is convex, it is continuous (compare [7, Cor.10.1.1]). By Theorem 3.2, f_k is uniformly convergent to f_0.

Let now $x_k = c(A_k)$for $k \in \mathbf{N}$ and let $x_0 = c(A)$. Then x_k is the unique point at which f_k attains its minimal value, for every $k \in \mathbf{N} \cup \{0\}$, and, since $c(A_k) \in A_k$, the set $\{x_k;\ k \in \mathbf{N}\}$ is bounded. Hence, by Proposition 0.4, $x_0 = \lim_k x_k$. □

COROLLARY 3.4. *The function* ring $: \mathcal{K}_0^n \to \mathcal{C}^n$ *is continuous.*

PROOF. Let $A, A_k \in \mathcal{K}_0^n$ and $A = \lim_H A_k$. Then,

$$R(A) = \lim_k R(A_k), \quad r(A) = \lim_k r(A_k), \quad \text{and} \quad c(A) = \lim_k c(A_k). \tag{3.1}$$

Indeed, let $x := c(A)$ and $x_k := c(A_k)$ for every k. Then, by Lemma 3.3, $x = \lim_k x_k$. Since $R(A) = R(A, x)$ and $R(A_k) = R(A_k, x_k)$, it follows that

$$|R(A_k) - R(A)| \leqq |\bar{R}(A_k, x_k) - \bar{R}(A_k, x)| + |\bar{R}(A_k, x) - \bar{R}(A, x)|.$$

Thus, by Lemma 3.2 and the continuity of R_A,

$$R(A) = \lim_k R(A_k).$$

The proof for r is analogous.

Now, condition (3.1) easily implies the statement. □

4. Final remarks

(A) To the author's best knowledge, it is an open question if there is an effective method of finding the centre of the minimal ring for arbitrary convex body. Some very special cases were mentioned in Preliminaries (Propositions 0.2 and 0.3).

It seems reasonable to believe that for convex polytopes the answer to the above question is positive. This optimistic point of view was a motivation for Section 3 of the present paper.

(B) The minimal ring of A is usually defined for $A \in \mathcal{K}_0^n$, i.e., for compact convex sets in $\mathbf{R}^n$ with non-empty interior (compare footnote 2). There are two natural ways to extend the definitions of $\operatorname{ring} A$ and $c(A)$ over the whole $\mathcal{K}^n$.

One possibility is to consider balls in the affine subspace spanned by A (compare Section 1). Then, of course,

$$\dim(\operatorname{ring}_E A) = \dim A$$

and the minimal ring of A depends only on A but not on the dimension of a Euclidean space containing A. As was noticed by K. Rudnik, $c: \mathcal{K}^n \to \mathbf{R}^n$ is not continuous:

EXAMPLE 4.1. There exists a sequence $(A_k)_{k\in\mathbf{N}}$ in $\mathcal{K}^3$, with $A_k \in \mathcal{K}_0^3$ for all k and $A = \lim_H A_k \in \mathcal{K}^3 \setminus \mathcal{K}_0^3$ such that

$$\lim_k c(A_k) \neq c_{\operatorname{aff} A}(A).$$

Indeed, let A be a non-equilateral triangle in $\operatorname{aff}(0, e_2, e_3)$ with O as the centre of circumscribed circle (like A_2 in Example 1.4) and let

$$a_k = (-k^{-1}, 0, 0), \quad b_k = (k^{-1}, 0, 0), \quad \text{and} \quad A_k := \triangle(a_k, b_k) \oplus A.$$

Then $A = \lim_H A_k$, while

$$c_{\operatorname{aff} A}(A) \neq 0 = \lim_k c(A_k).$$

□

The other possibility is to consider balls in $\mathbf{R}^n$ (instead of balls in $\operatorname{aff} A$) regardless of whether $\operatorname{int} A$ is empty or not. To avoid an ambiguity, let us denote this ring and its centre by "$\operatorname{ring}^{(n)}(A)$" and "$c^{(n)}(A)$", since they depend on the dimension of the ambient space $\mathbf{R}^n$. Of course, if $\operatorname{int} A = \emptyset$, then $r(A) = 0$ and $\operatorname{ring}^{(n)} A$ is a ball with radius $R(A)$.

Let us notice that in Sections 2 and 3 we did not make use of the assumption $\operatorname{int} A \neq \emptyset$. Thus, the results of these sections remain valid for $\operatorname{ring}^{(n)}$ and $c^{(n)}$. In particular, we obtain

COROLLARY 4.2. *The functions* $c^{(n)} : \mathcal{K}^n \to \mathbf{R}^n$ *and* $\mathrm{ring}^{(n)} : \mathcal{K}^n \to \mathcal{C}^n$ *are continuous.*

(C) Theorem 2.4 (i) states that $c\colon \mathcal{K}_0^n \to \mathbf{R}^n$ is, in some sense, "partially Minkowski additive":

if B is a ball, then

$$c(A+B) = c(A) + c(B).$$

However, as an application of Corollary 3.4, we obtain the following

COROLLARY 4.3. *The function* $c\colon \mathcal{K}_0^n \to \mathbf{R}^n$ *is not Minkowski additive.*

PROOF. Suppose that c is Minkowski additive. Since c is equivariant under isometries and, by Corollary 3.4, is continuous, from Theorem 3.4.2 [8] characterizing the Steiner point s, together with Remark 3.4.4 [8], it follows that

$$c(A) = s(A) \quad \text{for every} \quad A \in \mathcal{K}_0^n. \tag{4.1}$$

But condition (4.1) is evidently false. Indeed, let $A = \Delta(a_1, a_2, a_3)$ with $\sphericalangle(a_1 a_3 a_2) = \frac{\pi}{2}$ and $\varrho(a_1, a_3) = \varrho(a_2, a_3)$, and let $c(A) = O$. On the one hand, by Proposition 0.2, the point $c(A)$ is the centre of the circle inscribed in A, whence

$$\|a_3\| = \frac{\sqrt{2}}{2}\|a_1 + a_2\|.$$

On the other hand, if $s(A) = O$, then, by the formulae (5.4.11) and (5.4.12) in [8], $\frac{\pi}{4}(a_1 + a_2) + \frac{\pi}{2} a_3 = 0$, whence

$$\|a_3\| = \frac{1}{2}\|a_1 + a_2\|.$$

Thus $a_1 = -a_2$, which is impossible because $O = c(A) \in \operatorname{int} A$. 7 □

REFERENCES

[1] BÁRÁNY, I., On the minimal ring containing the boundary of a convex body, *Acta Sci. Math. (Szeged)* **52** (1988), 93–100. *MR* **89i**:52014

[2] BONNESEN, T., Über das isoperimetrische Defizit ebener Figuren, *Math. Ann.* **91** (1924), 252–268. *Jb. Fortschritte Math.* **50**, 487

[3] FEDERER, H., Curvature measures, *Trans. Amer. Math. Soc.* **93** (1959), 418–491. *MR* **22** #961

[4] GRUBER, P. M., Zur Charakterisierung konvexer Körper. Über einen Satz von Rogers und Shephard. II, *Math. Ann.* **184** (1970), 79–105. *MR* **41** #922

[5] MOSZYŃSKA, M., On the uniqueness problem for metric products, *Glas. Mat. Ser. III* **27** (**47**) (1992), 145–158. *MR* **94h**:54035

[6] MOSZYŃSKA, M. and ŻUKOWSKI, T., Duality of convex bodies, *Geom. Dedicata* **58** (1995), 161–173. *MR* **97g**:52009

[7] ROCKAFELLAR, T. R., *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, NJ, 1970. *MR* **43** #445

[8] SCHNEIDER, R., *Convex bodies: The Brunn–Minkowski theory*, Encyclopedia of Mathematics and its Applications, Vol. 44, Cambridge University Press, Cambridge, 1993. *MR* **94d**:52007

(Received August 6, 1996)

WYDZIAL MATEMATYKI, INFORMATYKI I MECHANIKI
INSTYTUT MATEMATYKI
UNIWERSYTET WARSZAWSKI
UL. BANACHA 2
PL–02-097 WARSZAWA
POLAND
mariamos@mimuw.edu.pl

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 175–184

COMPUTING MINIMUM AND BASIC SOLUTIONS OF LINEAR SYSTEMS USING THE HYPER-POWER METHOD

P. S. STANIMIROVIĆ

Abstract

An iterative method for computing the best approximate solution of a given system of linear equations is developed. The presented method is based on the hyper-power iterative process for computation of matrix products involving the Moore-Penrose inverse, introduced in [4], and have any high order $q \geqq 2$. Convergence properties of the method are studied. Also, we determine an optimal order q.

1. Introduction

The following notation will be used:

X, Y, Z are normed spaces with norms $\|\cdot\|$;

$\mathbb{C}^{m\times n}$, $\mathbb{C}^n$ are the set of $m \times n$ complex matrices and the set of n-dimensional complex vectors, respectively.

$B(X,Y)$ is the space of all bounded linear operators from X into Y;

$R(A)$, $N(A)$ are the range and the nullspace of A, respectively;

$A^\dagger$ is the Moore–Penrose inverse of A;

$\operatorname{rank}(A)$, $\operatorname{tr}(A)$ denote the rank and the trace of A, respectively.

$\mathbb{C}_r^{m\times n}$ is the set of $m \times n$ complex matrices whose rank is r;

$\rho(A)$ is the spectral radius of A;

$P_{R(A)}$ is the orthogonal projector on the range of A;

I_X, I_k, $\mathbb{O}$ denote the identity operator on X, the identity matrix of the order k, and the appropriate zero block, respectively.

Consider the following operator equation:

$$Ax = y, \qquad A \in B(X,Y), \qquad x \in X, \quad y \in Y. \tag{1}$$

If the solution exists, i.e., if for a given operator $A \in B(X,Y)$ and $y \in Y$ there exists an element (vector) $x \in X$ satisfying identically the equation (1), then the equation is said to be consistent (solvable), otherwise it is nonconsistent (unsolvable). It is obvious that $y \in R(A)$ is the necessary and sufficient condition of consistency.

1991 *Mathematics Subject Classification.* Primary 65F20; Secondary 15A09, 15A24.

Key words and phrases. Best approximate solution, hyper-power method, basic solutions, overdetermined system, Moore–Penrose inverse.

When the equation (1) is consistent and the operator A is invertible, then the unique solution is $x = A^{-1}y$. The determination of the general solution of (1) when A is non-invertible, i.e.,, when $N(A) \neq \{0\}$ and $b \in R(A)$, is possible through a generalization of the concept of inverse operators.

When the operator equation is nonconsistent, i.e., if $b \notin R(A)$, the functional of the error (residual vector) $\epsilon_x = Ax - y$ can be used as the measure of the nonconsistency. Most often the functional

$$\mu(A, y) = \inf_{x \in X} \|Ax - y\|$$

is used.

DEFINITION 1.1. A vector $x_0 \in X$ is said to be the *least-squares solution* (LSS solution) of (1) if

$$\|Ax_0 - y\| = \inf_{x \in X} \|Ax - y\|.$$

DEFINITION 1.2. Vector $x_0 \in X$ is said to be a *minimum-norm least-squares solution* (NLSS solution) of (1) if

$$\|Ax_0 - y\| = \inf_{x \in X} \|Ax - y\|$$

and

$$\|x_0\| < \|x\| \quad \text{for any} \quad x \neq x_0 \quad \text{for which} \quad \|Ax_0 - y\| = \|Ax - y\|.$$

NLSS solution is unique. This solution is also called the normal solution, or the best approximate solution.

PROPOSITION 1.1 ([6], [2]). *Let $A \in \mathbb{C}^{m \times n}$ and $b \in \mathbb{C}^m$. Then among the LSS solutions of $Ax = b$, $x = A^{\dagger}b$ is the one of the minimum norm. Conversely, if $X \in \mathbb{C}^{n \times m}$ has the property that, for all b, Xb is the NLSS solution of $Ax = b$, then $X = A^{\dagger}$.*

The paper is organized as follows. In the second section we briefly describe the hyper-power method and its modification for computing $A^{\dagger}B$, where A and B are matrices with identical number of rows [4]. In the third section we construct and investigate a method for computing NLSS solution $A^{\dagger}b$. The defined method is based on the modification of the hyper-power method, used for computing $A^{\dagger}B$. In this way, the constructed method is the first attempt to apply the hyper-power iterative method in computation of the best approximate solution of a linear system. The convergence rate is investigated, as well as the construction of the initial approximation and determination of the optimal order q of convergence. In the fourth section we develop corresponding iterative method which arises from the Neumann-type expansion. In the last section we describe the application of the presented method for computing the best approximate solution in computation of the *basic approximate solution* and give two illustrative examples.

2. Modification of the hyper-power method

The hyper-power iterative method was originally devised by Altman [1] for inverting a nonsingular bounded operator in a Banach space. This method is of any order $\geqq 2$. In [7] the convergence of the same method has been proved under a condition which is weaker than the one assumed in [1], and there have been derived better error estimates.

Zlobec in [13] defined two hyper-power iterative methods of an arbitrary high order $q \geqq 2$:

$$\begin{aligned} T_k &= I_X - Y_k A, \\ Y_{k+1} &= (I_X + T_k + \cdots + T_k^{q-1}) Y_k, \quad k = 0, 1, \ldots, \end{aligned} \tag{2}$$

$$\begin{aligned} T_k' &= I_Y - A Y_k', \\ Y_{k+1}' &= Y_k' (I_Y + T_k' + \cdots + T_k'^{q-1}), \qquad k = 0, 1, \ldots. \end{aligned} \tag{3}$$

It is well known [13] that if we take

$$Y_0 = Y_0' = \alpha A^*, \qquad 0 < \alpha \leqq \frac{2}{\operatorname{tr}(A^* A)},$$

then $Y_k \underset{k\to\infty}{\longrightarrow} A^\dagger$, $Y_k' \underset{k\to\infty}{\longrightarrow} A^\dagger$. In this way, the hyper-power iterative method is valid for generating the Moore–Penrose inverse.

The hyper-power iterative method of the order 2 is studied in [3], and also in [9], but in view of the singular value decomposition of a matrix.

In [4] the hyper-power iterative method is adapted for computing $A^\dagger B$, where $A \in \mathbb{C}^{m\times N}$ and $B \in \mathbb{C}^{m\times n}$ are arbitrary complex matrices with identical number of rows. The starting matrix Y_0 is chosen from the following conditions:

$$\begin{aligned} &Y_0 = A^* W A^*, \text{ for some } W \in \mathbb{C}^{m\times N} \text{ provided that} \\ &\rho\left(P_{R(A)} - A Y_0\right) < 1. \end{aligned} \tag{4}$$

Moreover, in [4] the following method is defined, ensuring the convergence of the sequence $\{Z_k\}$ to $A^\dagger B$, where $\{Z_k\}$ is defined by

$$\begin{aligned} &Y_0 \text{ is given by (4)}, \\ &Z_0 = Y_0 B, \\ &T_0 = I_N - Y_0 A, \\ &M_k = I_N + T_k + T_k^2 + \cdots + T_k^{q-1}, \\ &Z_{k+1} = M_k Z_k, \\ &T_{k+1} = T_k^q = I + M_k [T_k - I], \quad k = 0, 1, \ldots. \end{aligned} \tag{5}$$

Process (2) is superior to (3) (more efficient with respect to matrix multiplications) when $m > N$ [4]. Also, in [4] it is showed that (5) is an improvement (over using (2) to find $A^\dagger$ and then form $A^\dagger B$) only when $N > n$. In summary, the process (5) is recommended in the case $m > N > n$.

3. Computing $A^\dagger b$ by means of hyper-power method

Consider the overdetermined system $Ax = b$, where $A \in \mathbb{C}^{m\times N}$, and $b \in \mathbb{C}^m$. The process (5) is practical for computing the NLSS solution of an overdetermined system $Ax = b$, because of $m > N > 1$. Note that, in the case $m < N$, you had better to use process (3) in order to compute an approximation of $A^\dagger$ and follow this by a single multiplication to produce the NLSS solution $A^\dagger b$.

THEOREM 3.1. *Consider $b \in \mathbb{C}^m$ and $A \in \mathbb{C}^{m\times N}$ such that $m > N$ and $\operatorname{rank}(A) = r \geqq 2$. If $q \geqq 2$ is an integer, then the sequence $\{x_k \in \mathbb{C}^N\}$, defined by:*

$$
\begin{aligned}
&Y_0 \text{ is given by (4)},\\
&x_0 = Y_0 b,\\
&T_0 = I_N - Y_0 A,\\
&M_k = I_N + T_k + T_k^2 + \cdots + T_k^{q-1},\\
&x_{k+1} = M_k x_k,\\
&T_{k+1} = T_k^q = I + M_k[T_k - I], \quad k = 0, 1, \ldots
\end{aligned}
\tag{6}
$$

converges to the NLSS solution of the (overdetermined) system $Ax = b$, i.e., $x_k \to A^\dagger b$.

PROOF. Using the following known fact, proved in [4], valid for the process (5):

$$Z_k = Y_k B,$$

we immediately conclude

$$x_k = Y_k b.$$

Now, using $Y_k \to A^\dagger$ [13], we get $x_k \to A^\dagger b$. □

REMARK 3.1. The initial approximation Y_0, chosen in the general form (4) ensures convergence of the iterative process (2), i.e., the convergence $Y_k \underset{k\to\infty}{\longrightarrow} A^\dagger$. We can use more primitive conditions ensuring convergence of the method (2). Thus, the general initial approximation Y_0, given in (4) can be replaced by the following, more operative one ([3], [13]):

$$Y_0 = \alpha A^*, \qquad 0 < \alpha \leqq \frac{2}{\lambda_1(A^*A)}, \tag{7}$$

where $\dfrac{2}{\lambda_1(A^*A)}$ is the largest eigenvalue of AA^*.

Also, we can take [13]

$$Y_0 = \alpha A^*, \qquad 0 < \alpha \leqq \frac{2}{\operatorname{tr}(A^*A)}. \tag{8}$$

In [13] it is proved that the optimal value for α in (8) is $\dfrac{2}{\operatorname{tr}(A^*A)}$, and the optimal value for α in (7) is $\dfrac{2}{\lambda_1(A^*A)}$.

In the case of $\operatorname{rank}(A) = 1$ we need not to apply the iterative process (6), but the following result is applicable:

LEMMA 3.1. *In the case of* $\operatorname{rank}(A) = 1$, *the NLSS solution of the linear system* $Ax = b$ *is given by*

$$A^\dagger b = \frac{1}{\operatorname{tr}(A^*A)} A^* b.$$

PROOF. The proof immediately follows from the known result, valid in the case $\operatorname{rank}(A) = 1$ [13]:

$$A^\dagger = \frac{1}{\operatorname{tr}(A^*A)} A^*.$$

□

In the following lemma we investigate the convergence rate of the iterative process (6)

LEMMA 3.2. *Let given* $m \times n$ *matrix* A *of rank* $\operatorname{rank}(A) \geqq 2$. *Then the iterative process* (6) *for computing* $A^\dagger b$ *is of an arbitrary high order* $q \geqq 2$, *identical to the order of the corresponding hyper-power method* (2).

PROOF. Using the following result from Lemma 2.1:

$$x_k = Y_k b,$$

we get

$$\|x_k - A^\dagger b\| = \|Y_k b - A^\dagger b\| \leqq \|Y_k - A^\dagger\| \|b\|,$$

which implies

$$\frac{\|x_{k+1} - A^\dagger b\|}{\|x_k - A^\dagger b\|^q} \leqq \frac{\|Y_{k+1} - A^\dagger\|}{\|Y_k - A^\dagger\|^q} \|b\| \left\| \frac{Y_k - A^\dagger}{(Y_k - A^\dagger)b} \right\|^q.$$

In this way, the order of convergence of the process (6) is identical with the order of convergence of the process (2). □

The rate of convergence of the process (6) is an increasing function of q. However, the number of matrix multiplications required at each iteration is an increasing function of q. Now, we look for the optimal value of q, which minimizes the computation required to achieve a given degree of convergence.

THEOREM 3.2. *Consider a linear system $Ax=b$, where $A\in\mathbb{C}^{m\times N}$ and $b\in\mathbb{C}^m$. The optimal value of q in iterative process* (6) *for computing $A^\dagger b$ is $q=2$.*

PROOF. In [4] is shown that the optimal value of q for the process (5) for computing $A^\dagger B$, where $B\in\mathbb{C}^{m\times n}$, is that q which minimizes the following function $f(q)$:

$$f(q)=(n/N+q-1)/\ln q, \tag{9}$$

where $\ln q$ denotes the natural logarithm of q. According to (9), in our case, we should minimize the function

$$f_1(q)=(1/N+q-1)/\ln q, \quad N\geqq 2. \tag{10}$$

Consequently, we should obtain an integer solution of the equation

$$q\cdot\ln q-q-\frac{1}{N}+1=0, \qquad q\geqq 2. \tag{11}$$

From (11) we get $\ln q-1\leqq 0$, which implies that the integer solution of (11) is $q=2$. □

REMARK 3.2. The optimal value $q=2$ can be obtained using the following known result, valid for the function $f(q)$ [4]:

$$f(2)\leqq f(q) \text{ for all } q\neq 2 \text{ and } 0<n/N<0.71.$$

In our case is $n=1$, $N\geqq 2$, and consequently, $0<1/N\leqq 0.5$, which means $f(2)\leqq f(q)$ for all $q\neq 2$.

REMARK 3.3. Recall that the hyper-power method for computing generalized inverses is not self-correcting [11], [12]. We know Zielke's [11] iterative refinement process, which solves the self-correcting problem. Namely, the iterative refinement for computing the Moore–Penrose inverse of A has the following form:

$$\begin{aligned}\tilde{Y}_k&=A^*Y_k^*Y_kY_k^*A^*\\ \tilde{T}_k&=I-\tilde{Y}_kA\\ Y_{k+1}&=(I+\tilde{T}_k+\cdots+\tilde{T}_k^{q-1})\tilde{Y}_k.\end{aligned}$$

This modification is not necessary in each step. In our case, we can define the following refinement process, solving the self-correcting problem during the computation of $A^\dagger b$ by means of the hyper-power iterative method:

$$\begin{aligned}\tilde{x}_k&=A^*bx_k^*x_kx_k^*A^*b\\ M_k&=I+T_k+T_k^2+\cdots+T_k^{q-1},\\ \tilde{x}_{k+1}&=M_k\tilde{x}_k,\\ T_{k+1}&=T_k^q=I+M_k[T_k-I], \quad k=0,1,\ldots.\end{aligned}$$

In [10] it has been showed that the conjugate-gradient methods for computing the Moore–Penrose inverse can be used in refinement of the hyper-power method of the order 2 for computation of the Moore–Penrose inverse defined in [3]. The refinement of the hyper-power method can be done by replacing the hyper-power iterative method from [3] by the conjugate-gradient method from [10]. Consequently, in the same way, the iterative process (6) (of the order 2) can be refined by the conjugate-gradient method defined in [5].

4. Computing $A^{\dagger}b$ by means of infinite series

It is well known [13] that the q-th order hyper-power method generates the partial sums of the infinite series

$$\sum_{i=0}^{\infty}\left[(I-X_0A)^i X_0\right] \quad \text{or} \quad \sum_{i=0}^{\infty}\left[X_0(I-AX_0)^i\right].$$

More precisely,

$$Y_k=\sum_{i=0}^{q^k-1}\left[(I-X_0A)^i X_0\right] \quad \text{or} \quad Y_k'=\sum_{i=0}^{q^k-1}\left[X_0(I-AX_0)^i\right].$$

Zlobec [13] has shown that $A^{\dagger}$ can be computed by means of the infinite series if we choose $Y_0=\alpha A^*$, where $0<\alpha\leqq\dfrac{2}{\operatorname{tr}(A^*A)}$. Our strategy is to adapt the infinite series in order to compute $A^{\dagger}b$.

THEOREM 4.1. *Let the $m\times N$ matrix A of $\operatorname{rank}(A)=r\geqq 2$ be given with $q\geqq 2$. Then the sequence x_k, defined by the following iterative process*

$$Y_0=\alpha A^*, \qquad 0<\alpha\leqq\frac{2}{\operatorname{tr}(A^*A)},$$

$$x_k=\sum_{i=0}^{q^k-1}\left[(I-Y_0B)^i Y_0\right]b, \quad k=0,1,\ldots$$

converges to $A^{\dagger}b$.

5. Application and numerical results

The *basic approximate solution* to a linear system $Ax=b$, $A\in\mathbb{C}_r^{m\times n}$, is defined in [8] as a least squares solution with no more than r nonzero

components. Also, in [8] the following method is developed for construction of the matrix $A^{\ddagger}$, such that for every vector b, the basic approximate solution of the system $Ax=b$ is given by $x_b=A^{\ddagger}b$:

$$x_b=A^{\ddagger}b=\begin{bmatrix}B^{\dagger}\\ \mathbb{O}\end{bmatrix}b,$$

where the $m\times r$ matrix B contains r linearly independent columns of A.

Since $A^{\ddagger}b=\begin{bmatrix}B^{\dagger}b\\ \mathbb{O}\end{bmatrix}$, and $m\geqq r$, the method analogous to the process (6) can be applied in computation of the value $y_b=A^{\ddagger}b$:

$$\begin{aligned}
&Y_0=\begin{bmatrix}\alpha B^*\\ \mathbb{O}\end{bmatrix}, \qquad 0<\alpha\leqq\frac{2}{\operatorname{tr}(B^*B)},\\
&y_0=Y_0b=\begin{bmatrix}\alpha B^*b\\ \mathbb{O}\end{bmatrix},\\
&T_0=I-Y_0B,\\
&M_k=I+T_k+T_k^2+\cdots+T_k^{q-1},\\
&y_{k+1}=M_ky_k,\\
&T_{k+1}=T_k^q=I+M_k[T_k-I], \quad k=0,1,\ldots.
\end{aligned}\tag{12}$$

REMARK 5.1. Note that r linearly independent columns of A can be selected by means of the algorithm described in [8].

REMARK 5.2. If the rows $i_1,\ldots,i_r$ of A are linearly independent, then the basic approximate solution y_b can be generated from the best approximate solution $A^{\dagger}b$ in this way:

$$y_b=\begin{bmatrix}(A^{\dagger}b)_{i_1}\\ \cdots\\ (A^{\dagger}b)_{i_r}\\ \mathbb{O}\end{bmatrix}.$$

EXAMPLE 5.1. Consider a rank deficient matrix $A=\begin{pmatrix}2&0&2\\0&1&2\\1&1&3\\0&1&2\end{pmatrix}$ and the overdetermined linear system $Ax=b$, where $b=\begin{pmatrix}1\\2\\5/2\\2\end{pmatrix}$. Implementing the method (6) of the order $q=2$ by the initial value Y_0 selected from (8), in the package *MATHEMATICA*, we obtain the following numerical results:

$$\alpha = \frac{2}{\operatorname{tr}(A^*A)} = \frac{2}{31};$$

$$x_0 = Y_0 b = \alpha A^* b = \begin{pmatrix} \frac{9}{31} \\ \frac{13}{31} \\ \frac{35}{31} \end{pmatrix};$$

$$T_0 = I - Y_0 A, \qquad M_0 = I + T_0, \qquad x_1 = M_0 x_0 = \begin{pmatrix} -\frac{48}{961} \\ \frac{220}{961} \\ \frac{392}{961} \end{pmatrix};$$

$$T_1 = T_0^2, \qquad M_1 = I + T_1, \qquad x_2 = M_1 x_1 = \begin{pmatrix} -\frac{97920}{923521} \\ \frac{322072}{923521} \\ \frac{546224}{923521} \end{pmatrix};$$

$$T_2 = T_1^2, \qquad M_2 = I + T_2, \qquad x_3 = M_2 x_2 = \begin{pmatrix} -\frac{157929548160}{852891037441} \\ \frac{383307631024}{852891037441} \\ \frac{608685713888}{852891037441} \end{pmatrix}.$$

Continuing in a similar way, we get

$$x_5 \approx \begin{pmatrix} -0.249749 \\ 0.499853 \\ 0.749957 \end{pmatrix},$$

$$x_6 = \begin{pmatrix} -0.25 \\ 0.5 \\ 0.75 \end{pmatrix} = A^\dagger b.$$

EXAMPLE 5.2. In this example we compute the basic approximate solution of the system considered in Example 5.1. It can be seen that the first two columns of A are linearly independent. Applying the process (12) of the order 2, we conclude that the first two elements of the approximations y_i of $A^\dagger b$ are equal to the corresponding elements in x_i, $i = 1, 2, \ldots$, from Example 5.1 and the last element is the zero.

REFERENCES

[1] Altman, M., An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space, *Pacific J. Math.* **10** (1960), 1107–1113. *MR* **23** #A3461

[2] Ben-Israel, A. and Greville, T. N. E., *Generalized inverses: Theory and applications*, Wiley-Interscience, New York, 1974. *MR* **53** #469

[3] Ben-Israel, A. and Cohen, D., On iterative computation of generalized inverses and associated projections, *SIAM J. Numer. Anal.* **3** (1966), 410–419. *MR* **34** #3764

[4] Garnett, J., Ben-Israel, A. and Yau, S. S., A hyperpower iterative method for computing matrix products involving the generalized inverse, *SIAM J. Numer. Anal.* **8** (1971), 104–109. *MR* **43** #7048

[5] Kammerer, W. J. and Nashed, M. Z., On the convergence of the conjugate gradient method for singular linear operator equations, *SIAM J. Numer Anal.* **9** (1972), 165–181. *MR* **47** #7912

[6] Penrose, R., On best approximate solution of linear matrix equations, *Proc. Cambridge Philos. Soc.* **52** (1956), 17–19. *MR* **17**, 536d

[7] Petryshyn, W. V., On the inversion of matrices and linear operators, *Proc. Amer. Math. Soc.* **16** (1965), 893–901. *MR* **31** #6345

[8] Rosen, J. B., Minimum and basic solutions to singular linear systems, *J. Soc. Indust. Appl. Math.* **12** (1964), 156–162. *MR* **31** #3443

[9] Söderström, T. and Stewart, G. W., On the numerical properties of an iterative method for computing the Moore–Penrose generalized inverse, *SIAM J. Numer. Anal.* **11** (1974), 61–74. *MR* **49** #6589

[10] Tanabe, K., Conjugate-gradient method for computing the Moore–Penrose inverse and rank of a matrix, *J. Optimization Theory Appl.* **22** (1977), 1–23. *MR* **58** #24887

[11] Zielke, G., Iterative refinement of generalized matrix inverses now practicable, *ACM SIGNUM Newsletter* **13.4** (1978), 9–10.

[12] Zielke, G., A survey of generalized matrix inverses, *Computational Mathematics* (Warszawa, 1980), Banach Center Publ., PWN, Warsaw, 1984, 499–526. *MR* **87b**:65047

[13] Zlobec, S., On computing the generalized inverse of a linear operator, *Glasnik Mat. Ser. III* **2** (**22**) (1967), 265–271. *MR* **38** #3281

(Received September 10, 1996)

UNIVERSITY OF NIŠ
FACULTY OF PHILOSOPHY
DEPARTMENT OF MATHEMATICS
ĆIRILA I METODIJA 2
YU–18000 NIŠ
YUGOSLAVIA
pecko@archimed.filfak.ni.ac.yu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 185–206

GENERALIZED SOLUTIONS OF LOCAL INITIAL PROBLEMS FOR QUASI-LINEAR HYPERBOLIC FUNCTIONAL DIFFERENTIAL SYSTEMS

T. CZŁAPIŃSKI and Z. KAMONT

Abstract

Carathéodory solutions of quasi-linear hyperbolic systems in the second canonical form are investigated. Theorems on the existence, uniqueness and continuous dependence upon initial data are given. The method of bicharacteristics and integral inequalities are used. The local Cauchy problem is transformed into functional integral equations. The existence of solutions of this system is proved by using integral inequalities and the Banach fixed point principle.

1. Introduction

For any metric spaces X and Y let $C(X,Y)$ denote the class of all continuous functions from X into Y. We will denote by $M[k,n]$ the set of all $k \times n$ matrices with real elements. Suppose that $a>0$, $b=(b_1,\dots,b_n)$, $M=(M_1,\dots,M_n)\in R_+^n$, $R_+=[0,+\infty)$, and $b_i > M_i a$ for $1 \leqq i \leqq n$. Let E be the Haar pyramid

$$E=\{(x,y)\in R^{1+n}: x\in[0,a],\ y=(y_1,\dots,y_n),\ -b+Mx \leqq y \leqq b-Mx\}.$$

Here and subsequently the inequality between two vectors means that the same inequalities hold between their corresponding components. Write $E_0 = [-r_0,0]\times[-b,b]$ with $r_0\in R_+$ and

$$E_x=\{(t,s)=(t,s_1,\dots,s_n)\in E_0\cup E: t\leqq x\}.$$

Put $I[x,y]=\{t:\ (t,y)\in E_x\setminus E_0\}$, where $(x,y)\in[0,a]\times[-b,b]$. Write

$$S_x=[-b,b] \text{ for } x\in[-r_0,0] \text{ and } S_x=[-b+Mx,b-Mx] \text{ for } x\in[0,a].$$

Let $\Omega = E\times C(E_0\cup E,R^k)$ and assume that

$$A:\Omega\to M[k,k],\quad A=[A_{ij}]_{i,j=1,\dots,k},$$

$$\varrho:\Omega\to M[k,n],\ \varrho=[\varrho_{ij}]_{i=1,\dots,k,\,j=1,\dots,n},\ f:\Omega\to R^k,\ f=(f_1,\dots,f_k)$$

1991 *Mathematics Subject Classification.* Primary 35L45, 35D05, 35R10.

Key words and phrases. Functional differential systems, second canonical form, generalized solutions, bicharacteristics.

are given functions of the variables (x, y, z), $z = (z_1, \ldots, z_k)$. Given an initial function $\phi = (\phi_1, \ldots, \phi_k) : E_0 \to R^k$, consider the Cauchy problem

$$(1) \quad \sum_{l=1}^{k} A_{il}(x,y,z)\left[D_x z_l(x,y) + \sum_{j=1}^{n} \varrho_{ij}(x,y,z) D_{y_j} z_l(x,y)\right] = f_i(x,y,z), \quad i = 1, \ldots, k,$$

$$(2) \quad z(x,y) = \phi(x,y) \text{ on } E_0.$$

We will consider existence and uniqueness for local generalized solutions of problem (1), (2) in the "almost everywhere" sense that is the solution $u : E_c \to R^k$, $0 < c \leqq a$, is continuous, possesses partial derivatives almost everywhere on $E_c \setminus E_0$ and satisfies (1) a. e. on $E_c \setminus E_0$.

Non-linear equations with first order partial derivatives have the following property: any classical solution of an initial problem exists locally with respect to x. This leads to generalized solutions in the sense almost everywhere or Carathéodory sense. Generalized solutions of quasi-linear equations are also investigated in the case that assumptions for the given functions are extended.

Numerous papers were published concerning Carathéodory solutions for hyperbolic problems. The main existence and uniqueness results for weakly coupled systems can be found in [3], [7], [9]. Integral differential equations with an initial condition and with unknown function of two variables was considered in [20]. The method of bicharacteristics is the main tool in these investigations. This method was adopted by L. Cesari [6]–[8] and P. Bassanini [1], [2] for quasi-linear hyperbolic systems in the second canonical form. The initial and boundary value problems were considered. These problems were global with respect to y. The paper [5] deals with the local initial value problem for semilinear hyperbolic systems without the functional dependence. Existence and uniqueness results in the Haar pyramid were presented.

Some non-linear hyperbolic systems can be reduced to quasi-linear systems in the second canonical form [10].

Recently numerous papers were published concerning functional differential equations. The existence and uniqueness of Carathéodory solutions is proved in [19] for systems with Volterra operators. A general class of functional differential problems was investigated in [11]–[14]. All these problems are global with respect to y. Note that the model of functional dependence introduced in [11] is not suitable for problems which are local with respect to y.

For further bibliography on hyperbolic functional differential problems see the survey paper [16].

Now we present relations between local and global (with respect to y) solutions of differential and functional differential systems.

Suppose that the function $\tilde{M} = (\tilde{M}_1, \ldots, \tilde{M}_n) \in C([0, a], R_+^n)$ is nondecreasing and $\tilde{M}(0) = 0$, $b - \tilde{M}(a) > 0$. Let

$$\tilde{E} = \{ (x, y) : \ x \in [0, a], \ \ -b + \tilde{M}(x) \leqq y \leqq b - \tilde{M}(x) \}.$$

Suppose that

$$\tilde{A} : \tilde{E} \times R^k \to M[k, k], \ \ \tilde{A} = \left[\tilde{A}_{ij} \right]_{i,j=1,\ldots,k},$$
$$\tilde{\varrho} : \tilde{E} \times R^k \to M[k, n], \ \ \tilde{\varrho} = [\, \tilde{\varrho}_{ij} \,]_{i=1,\ldots,k,\, j=1,\ldots,n},$$
$$\tilde{f} : \tilde{E} \times R^k \to R^k, \qquad \tilde{f} = (\tilde{f}_1, \ldots, \tilde{f}_k)$$

and $\tilde{\phi} : [-b, b] \to R^k$ are given functions.

Consider the quasilinear system without the functional dependence

$$\sum_{l=1}^{k} \tilde{A}_{il}(x, y, z(x, y)) \left[D_x z_l(x, y) + \sum_{j=1}^{n} \tilde{\varrho}_{ij}(x, y, z(x, y)) D_{y_j} z_l(x, y) \right] = \tilde{f}_i(x, y, z(x, y)), \ \ i = 1, \ldots, k, \tag{3}$$

with the initial condition

$$z(0, y) = \tilde{\phi}(y) \ \text{ on } [-b, b]. \tag{4}$$

For any interval $I = [a_0, a_1] \subset R$ let $L(I, R_+)$ be the set of all functions $\mu : I \to R_+$ such that

$$\int_{a_0}^{a_1} \mu(\tau)\, d\tau < +\infty.$$

We formulate now the following assumption on $\tilde{\varrho}$.

Assumption $\tilde{H}$. Suppose that the function $\tilde{\varrho}$ of the variables (x, y, p) is measurable in x for every (y, p) and it is continuous in (y, p) for almost all x. Assume that there exists a function $\gamma = (\gamma_1, \ldots, \gamma_n) \in L([0, a], R_+^n)$ such that for almost all $x \in [0, a]$

$$|\, \tilde{\varrho}_{ij}(x, y, p) \,| \leqq \gamma_j(x), \qquad 1 \leqq j \leqq n, \ 1 \leqq i \leqq k,$$

for y, p such that $(x, y, p) \in \tilde{E} \times R^k$, and

$$\tilde{M}(x) \geqq \int_0^x \gamma(\tau)\, d\tau \quad \text{for } x \in [0, a].$$

Initial problems for quasilinear systems without the functional dependence have the following property:

Existence and uniqueness results for the Cauchy problem (3), (4) can be deduced from known results for the global Cauchy problem [6]. More precisely, if

(i) Assumption $\tilde{H}$ is satisfied,

(ii) the functions $\tilde{A}$, $\tilde{\varrho}$, $\tilde{f}$ satisfy all the assumptions of Theorem 1 in [6] on the set $\tilde{E} \times R^k$ (instead of $[0,a] \times R^n \times [-d,d]$, $[-d,d] \subset R^k$),

(iii) the function $\tilde{\phi}$ satisfies all the assumptions of Theorem 1 in [6] on $[-b,b]$ (instead of R^n),

then there exists exactly one generalized solution $\tilde{u}$ of problem (3), (4). The solution $\tilde{u}$ is defined on the set $\tilde{E} \cap ([0,c] \times R^n)$ with $c \in (0,a]$ sufficiently small and it depends continuously on given functions.

This result can be proved by exactly the same methods as in [6], see also [1]–[3].

The situation is completely different for systems with the functional dependence. We discuss the problem.

Several authors introduced various hereditary structures for description different situations in partial differential equations. Let us recall some of the main settings. For simplicity, let $k=1$ and consider the nonlinear equation

$$D_x z(x,y) = F(x,y,T(z;x,y), D_y z(x,y)), \tag{5}$$

where $D_y z = (D_{y_1} z, \ldots, D_{y_n} z)$ and T is a delay operator. If T is given by $T(z;x,y) = z(x,y)$ then (5) reduces to a classical equation.

There are a lot of papers concerning equation (5) with T defined by

$$T(z;x,y) = z_{(x,y)}, \tag{6}$$

where the function $(x,y) \to z_{(x,y)}$ is a natural extension of the Hale operator for ordinary functional differential equations [15]. More precisely, let $B = [-r_0, 0] \times [-r, r]$, where $r_0 \in R_+$ and $r = (r_1, \ldots, r_n) \in R_+^n$. For a function $z : [-r_0, a] \times R^n$ and a point $(x,y) \in [0,a] \times R^n$ we put

$$z_{(x,y)}(t,s) = z(x+t, y+s), \ (t,s) \in B, \tag{7}$$

i. e. the function $z_{(x,y)}$ is the restriction of z to the set $[x-r_0, x] \times [y-r, y+r]$. Consider the equation (5), (6) with the initial condition

$$z(x,y) = \phi_0(x,y) \ \text{ on } [-r_0, 0] \times R^n. \tag{8}$$

This formulation is natural and suitable for initial problems which are global with respect to y. The paper [16] contains a survey of existence results for nonlinear equations and quasi-linear systems in the second canonical form.

It is evident from (7) that the formulation (5), (6) is not suitable for the local Cauchy problems considered on the Haar pyramid.

The second group of papers is connected with the initial problems for the equation

$$D_x z(x,y) = G(x,y,z,D_y z(x,y)), \tag{9}$$

where G is an operator of the Volterra type. If we assume that $G : \bar{E} \times C(E_0 \cup \bar{E}) \times R^n \to R$ then we can consider the initial problem consisting of equation (9) and the condition

$$z(x,y) = \phi_0(x,y) \text{ on } E_0. \tag{10}$$

Quasi-linear system (1) is generated by equation (9).

It follows from the above consideration that the results of the papers [11]–[14] are not applicable to problem (1), (2).

Until now there are not any results on the existence and uniqueness of generalized solutions of problem (1), (2).

An extension of the classic Hale operator to parabolic functional differential problems is presented in the monograph [21].

The aim of the paper is to prove a theorem on the existence and uniqueness of Carathéodory solutions of problem (1), (2). We use the method of bicharacteristics. Problem (1), (2) is transformed into a functional integral system. The existence and uniqueness of solutions of this system will be proved by using integral inequalities and by the Banach fixed point principle.

Hyperbolic systems with a deviated argument and integral differential systems can be obtained from (1) by specializing the operators A, ϱ and f. Our results in this paper are also motivated by applications of partial differential or functional differential equations considered in [4], [17], [18].

We will say that the function f satisfies the Volterra condition if for all $z, \bar{z} \in C(E_0 \cup E, R^k)$ and $(x,y) \in E$, such that $z(t,s) = \bar{z}(t,s)$ for $(t,s) \in E_x$ we have $f(x,y,z) = f(x,y,\bar{z})$. Throughout this paper we assume that f, $\varrho_i = (\varrho_{i1}, \ldots, \varrho_{in})$ and $A_i = (A_{i1}, \ldots, A_{ik})$, $1 \leqq i \leqq k$, satisfy the Volterra condition.

2. Bicharacteristics

For $y \in R^n$ and $\zeta = (\zeta_1, \ldots, \zeta_k) \in R^k$ we write

$$\|y\| = \sum_{i=1}^{n} |y_i|, \quad \|\zeta\| = \max\{|\zeta_i| : 1 \leqq i \leqq k\}.$$

(We use the same symbol $\|\cdot\|$ to denote the norms in R^n and R^k.) For

$$U \in M[k,n], \qquad U = [u_{ij}]_{i=1,\ldots,k,\, j=1,\ldots,n}$$

we define

$$\|U\| = \max\Big\{\sum_{j=1}^{n} |u_{ij}| : 1 \leqq i \leqq k\Big\} \text{ and } u_i = (u_{i1}, \ldots, u_{in}),\ 1 \leqq i \leqq k.$$

Now we define some function spaces.

Let $\|\cdot\|_x$ be the supremum norm in the space $C(E_x, R^k)$, where $0 \leqq x \leqq a$. We will use the symbol $C_L(E_x, R^k)$ to denote the space of all functions $z \in C(E_x, R^k)$ such that

$$\|z\|_{(x;L)} = \sup\left\{\frac{\|z(t,s) - z(\bar{t},\bar{s})\|}{|t-\bar{t}| + \|s-\bar{s}\|} : (t,s),\ (\bar{t},\bar{s}) \in E_x\right\} < +\infty$$

endowed with the norm

$$\|z\|_{(x;0,L)} = \|z\|_x + \|z\|_{(x;L)}.$$

Let

$$C(E_x, R^k; \kappa) = \{z \in C(E_x, R^k) : \|z\|_x \leqq \kappa\},$$
$$C_L(E_x, R^k; \kappa) = \{z \in C_L(E_x, R^k) : \|z\|_{(x;0,L)} \leqq \kappa\},$$

where $\kappa \in R_+$ and $0 \leqq x \leqq a$. Denote by $J[P]$, where $P = (P_0, P_1, P_2) \in R_+^3$, the set of all functions $\psi \in C(E_0, R^k)$ such that $\|\psi(x,y)\| \leqq P_0$ and

$$\|\psi(x,y) - \psi(\bar{x},\bar{y})\| \leqq P_1|x-\bar{x}| + P_2\|y-\bar{y}\|$$

on E_0. Suppose that $0 < c \leqq a$, $Q = (Q_0, Q_1, Q_2) \in R_+^3$ and $Q_i \geqq P_i$, $i = 1, 2, 3$. Let $K_{c,\phi}[Q]$ be the class of all functions $z \in C(E_c, R^k)$ such that

(i) $z(x,y) = \phi(x,y)$ on E_0,

(ii) $\|z(x,y)\| \leqq Q_0$ and

$$\|z(x,y) - z(\bar{x},\bar{y})\| \leqq Q_1|x-\bar{x}| + Q_2\|y-\bar{y}\| \text{ on } E_c.$$

Write $|Q| = Q_0 + Q_1 + Q_2$.

ASSUMPTION $H[\varrho]$. Suppose that

(1) the function $\varrho(\cdot, y, z) : I[a,y] \to M[k,n]$ is measurable for $(y,z) \in [-b,b] \times C(E_0 \cup E, R^k)$ and $\varrho(x,\cdot) : S_x \times C(E_x, R^k) \to M[k,n]$ is continuous for almost all $x \in [0,a]$,

(2) for $(y,z) \in S_x \times C(E_x, R^k)$ and for almost all $x \in [0,a]$ we have

$$|\varrho_{ij}(x,y,z)| \leqq M_j,\ \ 1 \leqq i \leqq k,\ \ 1 \leqq j \leqq n,$$

where $M = (M_1, \ldots, M_n)$ is the constant vector from the definition of the Haar pyramid,

(3) there is a nondecreasing function $\beta: R_+ \to R_+$ such that

$$\|\varrho(x,y,z) - \varrho(x,\bar{y},\bar{z})\| \leqq \beta(\kappa)\,[\,\|y-\bar{y}\| + \|z-\bar{z}\|_x\,]$$

for (y,z), $(\bar{y},\bar{z}) \in S_x \times C_L(E_x, R^k;\kappa)$ and for almost all $x \in [0,a]$.

Given $\phi \in J[P]$, $c \in (0,a]$ and $z \in K_{c.\phi}[Q]$ consider the Cauchy problem

$$\eta'(t) = \varrho_i(t,\eta(t),z), \quad \eta(x) = y, \tag{11}$$

where $1 \leqq i \leqq k$, $(x,y) \in E_c \setminus E_0$.

Suppose that Assumption $H[\varrho]$ is satisfied. Let $g_i[z](\cdot\,,x,y)$ be the solution of problem (11). Denoting by $[0, c_i(x,y)]$ the maximal interval on which this solution is defined we see that $(c_i(x,y), g_i[z](c_i(x,y),x,y)) \in \partial E_c$, where ∂E_c is the boundary of E_c. The function $g_i[z]$ is called the i-th bicharacteristic of system (1) corresponding to $z \in K_{c.\phi}[Q]$.

LEMMA 2.1. *Suppose that Assumption $H[\varrho]$ is satisfied and $c \in (0,a]$, $\phi, \bar{\phi} \in J[P]$, $z \in K_{c.\phi}[Q]$, $\bar{z} \in K_{c.\bar{\phi}}[Q]$. Then for each i, $1 \leqq i \leqq k$, the bicharacteristics $g_i[z](\cdot\,,x,y)$ and $g_i[\bar{z}](\cdot\,,x,y)$ defined on the intervals $[0, c_i(x,y)]$ and $[0, \bar{c}_i(x,y)]$ exist and moreover we have the estimates*

$$\|g_i[z](t,x,y) - g_i[z](t,\bar{x},\bar{y})\| \leqq [\,\|M\|\,|x-\bar{x}| + \|y-\bar{y}\|\,]\exp[\beta(|Q|)t]\,, \tag{12}$$

where (x,y), $(\bar{x},\bar{y}) \in E_c \setminus E_0$, $t \in [0, \min\{c_i(x,y),\ c_i(\bar{x},\bar{y})\}]$ and

$$\|g_i[z](t,x,y) - g_i[\bar{z}](t,x,y)\| \leqq t\,\|z-\bar{z}\|_t\,\beta(|Q|)\,\exp[\beta(|Q|)t]\,, \tag{13}$$

where $(x,y) \in E_c \setminus E_0$, $t \in [0, \min\{c_i(x,y), \bar{c}_i(x,y)\}]$.

PROOF. The existence and uniqueness of solutions (in the "almost everywhere" sense) of (11) follows from classical theorems since the right-hand side of the differential system fulfils the Carathéodory conditions and the Lipschitz estimate with respect the unknown function is satisfied. If we transform (11) into the integral equation

$$g_i[z](t,x,y) = y + \int_x^t \varrho_i\,(\tau, g_i[z](\tau,x,y), z)\,d\tau, \tag{14}$$

then by Assumption $H[\varrho]$ we get the estimate

$$\|g_i[z](t,x,y) - g_i[z](t,\bar{x},\bar{y})\,\| \leqq \|y-\bar{y}\| + \|M\|\,|x-\bar{x}|$$
$$+\left|\int_x^t \beta(|Q|)\,\|g_i[z](\tau,x,y) - g_i[z](\tau,\bar{x},\bar{y})\,\|d\tau\right|,$$

where (x,y), $(\bar{x},\bar{y}) \in E_c$, $t \in [0, \min\{c_i(x,y), c_i(\bar{x},\bar{y})\}]$. Hence (12) follows by the Gronwall inequality. The estimate

$$\|g_i[z](t,x,y) - g_i[\bar{z}](t,x,y)\|$$
$$\leqq \left|\beta(|Q|)\int_x^t \|z-\bar{z}\|d\tau\right| + \beta(|Q|)\left|\int_x^t \|g_i[z](\tau,x,y) - g_i[\bar{z}](\tau,x,y)\|d\tau\right|,$$

where $(x,y) \in E_c \setminus E_0$, $t \in [0, \min\{c_i(x,y), \bar{c}_i(x,y)\}]$, and the Gronwall inequality imply (13). This completes the proof of the lemma.

3. The integral operator and its properties

ASSUMPTION $H[f]$. Suppose that

(1) the function $f(\cdot,y,z): I[a,y] \to R^k$ is measurable for $(y,z) \in [-b,b] \times C(E_0 \cup E, R^k)$ and $f(x,\cdot): S_x \times C(E_x \times R^k)$ is continuous for almost all $x \in [0,a]$,

(2) there is a nondecreasing function $\alpha: R_+ \to R_+$ such that

$$\|f(x,y,z)\| \leqq \alpha(\kappa) \text{ for } (y,z) \in S_x \times C(E_x, R^k; \kappa)$$

for almost all $x \in [0,a]$, and

$$\|f(x,y,z) - f(x,\bar{y},\bar{z})\| \leqq \beta(\kappa)\,[\,\|y-\bar{y}\| + \|z-\bar{z}\|_x\,]$$

for (y,z), $(\bar{y},\bar{z}) \in S_x \times C_L(E_x, R^k; \kappa)$ almost everywhere on $[0,a]$.

ASSUMPTION $H[A]$. Suppose that

(1) $A \in C(\Omega, M[k,k])$ and there is $\nu > 0$ such that $\det A(x,y,z) \geqq \nu$ on Ω,

(2) the estimates

$$\|A(x,y,z)\| \leqq \alpha(\kappa) \text{ on } E \times C(E_x, R^k; \kappa)$$

and

$$\|A(x,y,z) - A(\bar{x},\bar{y},\bar{z})\| \leqq \beta(\kappa)\,[\,|x-\bar{x}| + \|y-\bar{y}\| + \|z-\bar{z}\|_{\tilde{x}}\,]$$

on $E \times C_L(E_{\tilde{x}}, R^k; \kappa)$, where $\tilde{x} = \max[x,\bar{x}]$ are satisfied.

REMARK 3.1. In the paper we prove that there exists a constant $c \in (0,a]$ such that problem (1), (2) has exactly one solution in the class $K_{c,\phi}[Q]$. We will need some estimates of the constant c. For simplicity of formulation of these estimates we assume that the functions f and A have the same estimate on $C(E_x, R^k; \kappa)$ and that they satisfy the local Lipschitz condition on $C_L(E_x, R^k; \kappa)$ with the coefficient α (see Condition (2) of Assumption $H[\varrho]$ and Condition (2) of Assumptions $H[f]$ and $H[A]$).

REMARK 3.2. Note that if Assumption $H[A]$ is satisfied then $A^{-1}(x,y,z)$ exists on Ω and $A^{-1} \in C(\Omega, M[k,k])$. Moreover, there are nondecreasing functions $\tilde{\alpha}$, $\tilde{\beta}: R_+ \to R_+$ such that

$$\|A^{-1}(x,y,z)\| \leqq \tilde{\alpha}(\kappa) \text{ on } E \times C(E_x, R^k; \kappa)$$

and

$$\|A^{-1}(x,y,z) - A^{-1}(\bar{x},\bar{y},\bar{z})\| \leqq \tilde{\beta}(\kappa)\,[|x-\bar{x}| + \|y-\bar{y}\| + \|z-\bar{z}\|_x]$$

on $E \times C_L(E_x, R^k; \kappa)$.

It is important in our considerations that we have assumed the Lipschitz condition for given functions on some special functions spaces. More precisely, we have assumed that

(i) the functions $\varrho(x,\cdot)$ and $f(x,\cdot)$ satisfy the Lipschitz condition on the space $S_x \times C_L(E_x, R^k; \kappa)$ for almost all $x \in [0,a]$,

(ii) the function A satisfies the Lipschitz condition on $E \times C_L(E_0 \cup E, R^k; \kappa)$.

The above conditions are local with respect to the functional variable.

Let us consider simplest assumptions on f, ϱ, A. Suppose that there is $\tilde{L} \in R_+$ such that for almost all $x \in [0,a]$

$$\|f(x,y,z) - f(x,\bar{y},\bar{z})\| \leqq \tilde{L}\,[\|y-\bar{y}\| + \|z-\bar{z}\|_x] \tag{15}$$

$$\|\varrho(x,y,z) - \varrho(x,\bar{y},\bar{z})\| \leqq \tilde{L}\,[\|y-\bar{y}\| + \|z-\bar{z}\|_x]\,, \tag{16}$$

where (y,z), $(\bar{y},\bar{z}) \in S_x \times C(E_0 \cup E, R^k)$ and

$$\|A(x,y,z) - A(\bar{x},\bar{y},\bar{z})\| \leqq \tilde{L}\,[|x-\bar{x}| + \|y-\bar{y}\| + \|z-\bar{z}\|_x] \tag{17}$$

on $E \times C(E_0 \cup E, R^k)$.

Of course, our results are true if we assume (15)–(17) instead of (i), (ii).

Now we show that the formulation (i), (ii) are important. More precisely, we show that there is a class of quasilinear systems satisfying (i), (ii) and do not satisfying (15)–(17).

Consider the system with a deviated argument

$$\sum_{l=1}^{k} \tilde{A}(x,y,z(\psi(x,y)))\left[D_x z_l(x,y) + \sum_{j=1}^{n} \tilde{\varrho}_{ij}(x,y,z(\psi(x,y)))\,D_{y_j} z_l(x,y)\right] = \tilde{f}(x,y,z(\psi(x,y))), \quad i = 1,\dots,k, \tag{18}$$

where $\tilde{A}$, $\tilde{\varrho}$, $\tilde{f}$ are given in Section 1 and

$$\psi = (\psi_0, \psi_1, \dots, \psi_n) \in C(E, E_0 \cup E).$$

We assume that $\psi_0(x,y) \leqq x$ for $(x,y) \in E$. We get system (18) by putting in (1)

$$f(x,y,z) = \tilde{f}(x,y,z(\psi(x,y))), \quad \varrho(x,y,z) = \tilde{\varrho}(x,y,z(\psi(x,y))),$$

$$A(x,y,z) = \tilde{A}(x,y,z(\psi(x,y))).$$

From now on we consider the function ϱ only.

Suppose that there are $\tilde{C}$, $C \in R_+$ such that

$$\|\tilde{\varrho}(x,y,p) - \tilde{\varrho}(x,\bar{y},\bar{p})\| \leqq \tilde{C}\,[\,\|y-\bar{y}\| + \|p-\bar{p}\|\,]$$

and

$$\|\psi(x,y) - \psi(x,\bar{y})\|_{n+1} \leqq C_0\|y-\bar{y}\|,$$

where $\|\cdot\|_{n+1}$ is the norm in the space R^{n+1}.

It is evident that for (y,z), $(\bar{y},\bar{z}) \in S_x \times C_L(E_x, R^k;\kappa))$ and for almost all $x \in [0,a]$ we get

$$\begin{aligned}\|\varrho(x,y,z) - \varrho(x,\bar{y},\bar{z})\| &= \|\tilde{\varrho}(x,y,z(\psi(x,y))) - \tilde{\varrho}(x,\bar{y},\bar{z}(\psi(x,\bar{y})))\| \\ &\leqq \tilde{C}\,(1+\kappa C_0)\,\|y-\bar{y}\| + \tilde{C}\|z-\bar{z}\|_x.\end{aligned}$$

Then Condition (3) of Assumption $H[\varrho]$ is satisfied.

We see at once that the function $\varrho(x,\cdot)$ does not satisfy the global Lipschitz condition (16) for (y,z), $(\bar{y},\bar{z}) \in S_x \times C(E_0 \cup E, R^k)$. Similar consideration apply to f and A.

Now we construct an integral operator corresponding to initial problem (1). (2). Suppose that $\phi \in J[P]$, $c \in (0,a]$, $z \in K_{c.\phi}[Q]$ and that

$$(g_1[z](\cdot\,,x,y), \ldots, g_k[z](\cdot\,,x,y)) = g[z](\cdot\,,x,y)$$

are bicharacteristics of (1) corresponding to z. Let

$$\begin{aligned}f[g,z](t,x,y) &= (f_1(t,g_1[z](t,x,y),z), \ldots, f_k(t,g_k[z](t,x,y),z))\,, \\ A[g,z](t,x,y) &= [A_{ij}(t,g_i[z](t,x,y),z)]_{i,j=1,\ldots,k}\,, \\ \Phi[g,z](t,x,y) &= [\phi_i(0,g_j[z](t,x,y))]_{i,j=1,\ldots,k}\,, \\ Z[g,z](t,x,y) &= [z_i(t,g_j[z](t,x,y))]_{i,j=1,\ldots,k}\,.\end{aligned}$$

For

$$U, V \in M[k,k], \qquad U = [u_{ij}]_{i,j=1,\ldots,k}, \qquad V = [v_{ij}]_{i,j=1,\ldots,k},$$

we denote by $U * V$ the vector $(d_1, \ldots, d_k)$, where

$$d_i = \sum_{j=1}^{k} u_{ij}v_{ji}, \qquad i = 1, \ldots, k.$$

Let us define the operator T_ϕ for all $z \in K_{c.\phi}[Q]$ by the formula

$$\begin{aligned}T_\phi z(x,y) &= A^{-1}(x,y,z)\,\{A[g,z](0,x,y) * \Phi[g,z](0,x,y)\} \\ &+ A^{-1}(x,y,z)\int_0^x \{D_t A[g,z](t,x,y) * Z[g,z](t,x,y) + f[g,z](t,x,y)\}\,dt\end{aligned} \tag{19}$$

for $(x,y) \in E_c \setminus E_0$ and

$$T_\phi z(x,y) = \phi(x,y) \text{ for } (x,y) \in E_0. \tag{20}$$

ASSUMPTION $H[Q]$. Suppose that the constants $(Q_0, Q_1, Q_2) \in R^3_+$ are such that

$$Q_0 > P_0, \qquad Q_1 > \max\left\{\tilde{\alpha}(Q_0)\alpha(Q_0)\left(1 + P_2\|M\|\right), P_1\right\},$$

$$Q_2 > P_2\left[1 + 2\tilde{\alpha}(Q_0)\alpha(Q_0)\right].$$

REMARK 3.3. The right-hand side of (19) is obtained in the following way.

Considering system (1) along bicharacteristics we get by (11) the relation

$$A[g,z](t,x,y) * D_t Z[g,z](t,x,y) = f[g,z](t,x,y).$$

Integrating it with respect to t from 0 to x, and making use (2) we get

$$A(x,y,z)z(x,y) = A[g,z](0,x,y) * \Phi[g,z](0,x,y)$$
$$+ \int_0^x \left\{D_t A[g,z](t,x,y) * Z[g,z](t,x,y) + f[g,z](t,x,y)\right\} dt.$$

If Assumptions $H[\varrho]$, $H[A]$ are satisfied then the derivative $D_t A[g,z](\cdot,x,y)$ exists almost everywhere and is integrable on $[0,c]$. Thus we may have used the integration by parts which yields the functional integral equation $z = T_\phi z$.

LEMMA 3.4. *If $\phi \in J[P]$ and Assumptions $H[\varrho]$, $H[f]$, $H[A]$, $H[Q]$ are satisfied then for sufficiently small $c \in (0,a]$ we have*

$$T_\phi : K_{c.\phi}[Q] \to K_{c.\phi}[Q].$$

PROOF. Suppose that $z \in K_{c.\phi}[Q]$. Using the relations

$$A^{-1}(x,y,z)\left\{A[g,z](x,x,y) * \Phi[g,z](x,x,y)\right\} = \phi(0,y),$$

$$A[g,z](x,x,y) - A[g,z](0,x,y) = \int_0^x D_t A[g,z](t,x,y)dt,$$

where $(x,y) \in E_c \setminus E_0$, we see that (19) is equivalent to

$$T_\phi z(x,y) = \phi(0,y) + A^{-1}(x,y,z)\left[\Delta_1(x,y) + \Delta_2(x,y) + \Delta_3(x,y)\right], \tag{21}$$

where

$$\Delta_1(x,y)=\int_0^x f[g,z](t,x,y)dt,$$

$$\Delta_2(x,y)=A[g,z](0,x,y)*\{\Phi[g,z](0,x,y)-\Phi[g,z](x,x,y)\}, \tag{22}$$

$$\Delta_3(x,y)=\int_0^x D_tA[g,z](t,x,y)*\{Z[g,z](t,x,y)-\Phi[g,z](x,x,y)\}\,dt.$$

By Assumptions $H[\varrho]$, $H[f]$, $H[A]$ we have

$$\begin{aligned}&\|\Delta_1(x,y)\|\leqq c\,\alpha(Q_0),\\ &\|\Delta_2(x,y)\|\leqq c\,\alpha(Q_0)P_2\|M\|,\\ &\|\Delta_3(x,y)\|\leqq c^2\,\tilde{\beta}(|Q|)\,[1+\|M\|]\,[Q_1+Q_2\|M\|\,].\end{aligned} \tag{23}$$

In the proof of the last inequality we have used the fact that the Lipschitz constant of the function $A[g,z](\cdot,x,y)$ is an upper bound of $\|D_t[g,z](t,x,y)\|$. The estimates (23) together with Remark 3.2 imply

$$\|T_\phi z(x,y)\|\leqq P_0+\tilde{\alpha}(Q_0)H_0(c),\qquad (x,y)\in E_c\setminus E_0,$$

where

$$H_0(c)=c\,\{\alpha(Q_0)\,(1+P_2\|M\|\,)+\beta(|Q|)\,[1+\|M\|]\,[Q_1+Q_2\|M\|]\,c\}\,.$$

Since

$$\lim_{c\to 0^+}H_0(c)=0\ \text{ and }\ Q_0>P_0$$

we may take $c\in(0,a]$ sufficiently small in order that

$$\|T_\phi z(x,y)\|\leqq Q_0\ \text{ for }(x,y)\in E_c\setminus E_0. \tag{24}$$

Now we establish the Lipschitz constants for $T_\phi z$. For (x,y), $(\bar{x},\bar{y})\in E_c\setminus E_0$ we have

$$T_\phi z(x,y)-T_\phi z(\bar{x},\bar{y})=\Gamma_0+\Gamma_1+\Gamma_2+\Gamma_3,$$

where

$$\begin{aligned}\Gamma_0&=\phi(0,y)-\phi(0,\bar{y})\\ &\quad+\left[A^{-1}(x,y,z)-A^{-1}(\bar{x},\bar{y},z)\right]\left[\Delta_1(x,y)+\Delta_2(x,y)+\Delta_3(x,y)\right],\end{aligned}$$

$$\Gamma_i=A^{-1}(\bar{x},\bar{y},z)\,[\Delta_i(x,y)-\Delta_i(\bar{x},\bar{y})]\,,\qquad i=1,2,3.$$

It follows from Assumptions $H[\varrho]$, $H[f]$, $H[A]$ and (23) that

$$\begin{aligned}&\|\Gamma_0\|\leqq P_2\|y-\bar{y}\|+\tilde{\beta}(|Q|)H_0(c)\,[\,|x-\bar{x}|+\|y-\bar{y}\|\,],\\ &\|\Gamma_1\|\leqq c\,\tilde{\alpha}(Q_0)\beta(|Q|)\exp\left[\beta(|Q|)c\right][\,\|M\|\,|x-\bar{x}|+\|y-\bar{y}\|\,],\end{aligned} \tag{25}$$

and

$$\begin{aligned}(26)\qquad \|\Gamma_3\| \leqq{}& c\,\tilde{\alpha}(Q_0)P_2\|M\|\alpha(Q_0)\beta(|Q|)\exp[\beta(|Q|)c]\,[\|M\|\,|x-\bar{x}|+\|y-\bar{y}\|]\\ &+\tilde{\alpha}(Q_0)\alpha(Q_0)\{P_2\exp[\beta(|Q|)c]\,[\|M\|\,|x-\bar{x}|+\|y-\bar{y}\|]+P_2\|y-\bar{y}\|\}.\end{aligned}$$

Integrating by parts we may write $\Delta_3(x,y)-\Delta_3(\bar{x},\bar{y})=\Lambda_1+\Lambda_2+\Lambda_3$, where

$$\begin{aligned}\Lambda_1 ={}& \{A[g,z](x,x,y)-A[g,z](x,\bar{x},\bar{y})\}*\{Z[g,z](x,x,y)-\Phi[g,z](x,x,y)\}\\ &-\{A[g,z](0,x,y)-A[g,z](0,\bar{x},\bar{y})\}*\{Z[g,z](0,x,y)-\Phi[g,z](x,x,y)\}\\ &-\int_0^x\{A[g,z](t,x,y)-A[g,z](t,\bar{x},\bar{y})\}*D_tZ[g,z](t,x,y)\,dt\end{aligned}$$

and

$$\begin{aligned}\Lambda_2 ={}& \int_0^x D_tA[g,z](t,\bar{x},\bar{y})*\{Z[g,z](t,x,y)-\Phi[g,z](x,x,y)\\ &-Z[g,z](t,\bar{x},\bar{y})+\Phi[g,z](\bar{x},\bar{x},\bar{y})\}\,dt,\\ \Lambda_3 ={}& \int_{\bar{x}}^x D_tA[g,z](t,\bar{x},\bar{y})*\{Z[g,z](t,\bar{x},\bar{y})-\Phi[g,z](\bar{x},\bar{x},\bar{y})\}\,dt.\end{aligned}$$

Applying Assumptions $H[\varrho]$, $H[A]$ we get the estimates

$$(27)\qquad \|\Lambda_i\| \leqq A_i(c)\,[\,|x-\bar{x}|+\|y-\bar{y}\|\,],\quad i=1,2,3,$$

with constants $A_i(c)>0$ such that

$$(28)\qquad \lim_{c\to 0^+}A_i(c)=0,\quad i=1,2,3.$$

It follows from (21), (24)–(28) that there exist $B_1(c)$, $B_2(c)>0$ such that

$$\begin{aligned}\|T_\phi z(x,y)-T_\phi z(\bar{x},\bar{y})\| \leqq{}& \{\tilde{\alpha}(Q_0)\,[\alpha(Q_0)+P_2\alpha(Q_0)\,\|M\|\,]+B_1(c)\}\;|x-\bar{x}|\\ &+\{P_2\,[1+2\tilde{\alpha}(Q_0)\alpha(Q_0)]+B_2(c)\}\;\|y-\bar{y}\|\end{aligned}$$

on $E_c\setminus E_0$ and

$$\lim_{c\to 0^+}B_1(c)=\lim_{c\to 0^+}B_2(c)=0.$$

Hence by Assumption $H[Q]$ there exists $c\in(0,a]$ sufficiently small in order that

$$(29)\qquad \|T_\phi z(x,y)-T_\phi z(\bar{x},\bar{y})\| \leqq Q_1|x-\bar{x}|+Q_2\|y-\bar{y}\|$$

on $E_c\setminus E_0$. Since $T_\phi z\in C(E_c,R^k)$ and estimates (24), (29) hold true also on E_0 we see that $T_\phi z\in K_{c,\phi}[Q]$, which completes the proof of the lemma.

4. Existence and uniqueness of solutions to the Cauchy problem

Set

$$\Gamma(x) = x\,\beta(|Q|)\exp\left[x\,\beta(|Q|)\right].$$

THEOREM 4.1. *Suppose that $\phi \in J[P]$ and Assumptions $H[\varrho]$, $H[A]$, $H[f]$, $H[Q]$ are satisfied. Then there exists $c \in (0, a]$ such that Problem (1), (2) has exactly one solution u in the class $K_{c.\phi}[Q]$. Moreover, if $\bar{\phi} \in J[P]$ and $\bar{u}$ is a solution of system (1) with the initial condition*

$$z(x, y) = \bar{\phi}(x, y) \quad \text{on } E_0, \tag{30}$$

then there exists $\Lambda_c \in R_+$ such that

$$\|u - \bar{u}\|_x \leqq \Lambda_c \|\phi - \bar{\phi}\|_0, \quad x \in [0, c]. \tag{31}$$

PROOF. Suppose that $c \in (0, a]$ is sufficiently small in order that $T_\phi : K_{c.\phi}[Q] \to K_{c.\phi}[Q]$, which can be done by force of Lemma 3.4.

If z, $\tilde{z} \in K_{c.\phi}[Q]$ then we have

$$T_\phi z(x, y) - T_\phi \tilde{z}(x, y) = B_0 + B_1 + B_2 + B_3,$$

where

$$B_0 = \left\{A^{-1}(x, y, z) - A^{-1}(x, y, \tilde{z})\right\}\left[\Delta_1(x, y) + \Delta_2(x, y) + \Delta_3(x, y)\right]$$

and

$$B_i = A^{-1}(x, y, \tilde{z})\left[\Delta_i(x, y) - \tilde{\Delta}_i(x, y)\right], \quad i = 1, 2, 3.$$

The functions Δ_i are given by (22) and $\tilde{\Delta}_i$ by the same relations that arise from (22) by replacing g, z with $\tilde{g}$, $\tilde{z}$.

It follows from Assumptions $H[\varrho]$, $H[f]$, $H[A]$ that

$$\|B_0\| \leqq \tilde{\beta}(|Q|)H_0(x)\|z - \tilde{z}\|_x,$$

$$\|\Delta_1(x, y) - \tilde{\Delta}_1(x, y)\| \leqq \beta(|Q|)\left[\Gamma(x) + 1\right]\|z - \tilde{z}\|_x,$$

$$\|\Delta_2(x, y) - \tilde{\Delta}_2(x, y)\| \leqq \left\{x\beta(|Q|)\left[\Gamma(x) + 1\right]P_2\|M\| + \alpha(Q_0)P_2\Gamma(x)\right\}\|z - \tilde{z}\|_x,$$

and

$$\|\Delta_3(x, y) - \tilde{\Delta}_3(x, y)\|$$
$$\leqq \left\{x\beta(Q_2)\left[\Gamma(x) + 1\right]\left[Q_1 + 2Q_2\|M\| + 1\right] + x\beta(|Q|)Q_1\right\}\|z - \tilde{z}\|_x.$$

In the last estimate we have used integration by parts analogously to the proof of Lemma 3.4. Hence there exists a function $\Lambda(x)$ such that

$$\|T_\phi z - T_\phi \tilde{z}\|_x \leqq \Lambda(x)\|z - \tilde{z}\|_x \tag{32}$$

and

$$\lim_{x\to 0^+} \Lambda(x) = 0.$$

If we choose $c \in (0, a]$ sufficiently small in order that $\Lambda(c) < 1$ then by the Banach fixed point theorem there is a unique solution $u \in K_{c,\phi}[Q]$ of the equation $z = T_\phi z$.

Now we prove that u is a solution of system (1). We have

$$u(x,y) = A^{-1}(x,y,u)\,\{A[g,u](0,x,y) * \Phi[g,u](0,x,y)\}$$

$$+A^{-1}(x,y,u)\int_0^x \{D_t A[g,u](t,x,y) * Z[g,u](t,x,y) + f[g,u](t,x,y)\}\,dt,$$

where $(x,y) \in E_c \setminus E_0$ and g is the solution of (11) with u instead of z. Multiplying the above relation by $A(x,y,u)$ and integrating its right-hand side by parts we obtain

$$\int_0^x \{-A[g,u](t,x,y) * D_t Z[g,u](t,x,y) + f[g,u](t,x,y)\}\,dt = 0,$$

and consequently we get our claim by the same considerations as in [6], [8].

Uniqueness of the solution of problem (1), (2) follows from the fact that any solution $\bar{z} \in K_{c,\phi}[Q]$ of the problem satisfies the equation $z = T_\phi z$ which has at most one solution.

Suppose that $u = T_\phi u$, $\bar{u} = T_{\bar{\phi}}\bar{u}$. Slightly modifying the estimates that we have used to get (32) we may analogously obtain

$$\|u - \bar{u}\|_x \leqq \Lambda(x)\|u - \bar{u}\|_x + \bar{\Lambda}(x)\|\phi - \bar{\phi}\|_0,$$

where

$$\bar{\Lambda}(x) = 1 + 2\tilde{\alpha}(Q_0)\alpha(Q_0) + x\tilde{\alpha}(Q_0)\beta(|Q|)\,[1 + \|M\|]\,,$$

and $\Lambda(x)$ is the same as in (32). Since we have assumed that $\Lambda(c) < 1$ the estimate (31) holds with

$$\Lambda_c = \bar{\Lambda}(c)\,(1 - \Lambda(c))^{-1}\,.$$

This completes the proof of the theorem.

5. Carathéodory solutions with generalized Lipschitz condition

Suppose that the function $M = (M_1, \ldots, M_n) \in C([0,a], R_+^n)$ is nondecreasing and $M(0) = 0$, $b_i - M_i(a) > 0$ for $1 \leqq i \leqq n$. Let E_0 be the set given in Section 1 and

$$E = \{(x,y) : x \in [0,a],\ -b + M(x) \leqq y \leqq b - M(x)\}.$$

Let E_x, $0 \leqq x \leqq a$, and $I[x,y]$, $(x,y) \in [0,a] \times [-b,b]$, be the sets defined in Section 1 with the above given E. Put

$$S_x = [-b,b] \text{ for } x \in [-r_0,0] \text{ and } S_x = [-b+M(x), b-M(x)] \text{ for } x \in [0,a].$$

Let $\Omega_0 = E \times R^k$ and assume that

$$A : \Omega_0 \to M[k,k], \quad \varrho : \Omega \to M[k,n], \quad f : \Omega \to R^k, \quad \phi : E_0 \to R^k$$

are given functions. Consider the quasilinear system

$$\sum_{l=1}^{k} A_{il}(x,y,z(x,y)) \left[D_x z_l(x,y) + \sum_{j=1}^{n} \varrho_{ij}(x,y,z) D_{y_j} z_l(x,y) \right] = f_i(x,y,z), \quad i = 1,\dots,k, \tag{33}$$

with initial Condition (2). The matrix A in system (33) depends on z in the classical sense.

In Section 4 we have considered Lipschitz continuous solutions of hyperbolic systems. Now we deal with more general class of solutions for system (33).

We will use the symbol $C_{0.L'}(E_x, R^k)$ to denote the set of all functions $z \in C(E_x, R^k)$ such that

$$\|z\|_{(x;L')} = \sup \left\{ \frac{\|z(t,s) - z(t,\bar{s})\|}{\|s - \bar{s}\|} : (t,s), (t,\bar{s}) \in E_x \right\} < \infty.$$

For $z \in C_{0.L'}(E_x, R^k)$ we write

$$\|z\|_{(x;0,L')} = \|z\|_x + \|z\|_{(x;L')}, \qquad 0 \leqq x \leqq a,$$

where $\|\cdot\|_x$ is the supremum norm in $C(E_x, R^k)$. Let $C(E_x, R^k; \kappa)$ be the set given in Section 2 and

$$C_{0.L'}(E_x, R^k; \kappa) = \left\{ z \in C_{0.L'}(E_x, R^k) : \|z\|_{(x;0,L')} \leqq \kappa \right\},$$

where $\kappa \in R_+$, $x \in [0,a]$.

Let $\omega_0 \in L([-r_0,0], R_+)$, $P = (P_0, P_1) \in R_+^2$. Denote by $J[\omega_0, P]$ the set of all functions $\psi \in C(E_0, R^k)$ such that $\|\psi(x,y)\| \leqq P_0$ and

$$\|\psi(x,y) - \psi(\bar{x},\bar{y})\| \leqq \left| \int_{\bar{x}}^{x} \omega_0(\tau) d\tau \right| + P_1 \|y - \bar{y}\| \text{ on } E_0.$$

Suppose that

$$c \in (0,a], \quad Q = (Q_0, Q_1) \in R_+, \quad Q_0 \geqq P_0, \quad Q_1 \geqq P_1$$

and $\omega \in L([-r_0, c]R_+)$, $\omega(t) \geqq \omega_0(t)$ for almost all $t \in [-r_0, 0]$. Suppose that $\phi \in J[\omega_0, P]$. Let $K_{c.\phi}[\omega, Q]$ be the set of all functions $z \in C(E_c, R^k)$ such that

(i) $z(x, y) = \phi(x, y)$ on E_0,

(ii) $\|z(x, y)\| \leqq Q_0$ and

$$\|z(x, y) - z(\bar{x}, \bar{y})\| \leqq \left| \int_{\bar{x}}^{x} \omega(\tau) d\tau \right| + Q_1 \|y - \bar{y}\| \text{ on } E_c.$$

Put $|Q| = Q_0 + Q_1$. Any function $z \in K_{c.\phi}[\omega, Q]$ satisfies (2). This function is a solution of (33) if it satisfies the system almost everywhere on $E_c \setminus E_0$.

Denote by Θ the class of all functions $\delta : [0, a] \times R_+ \to R_+$ such that $\delta(\cdot, t) \in L([0, a], R_+)$ for $t \in R_+$ and $\delta(t, \cdot)$ is nondecreasing on R_+ for almost all $t \in [0, a]$.

ASSUMPTION $\bar{H}[\varrho]$. Suppose that

(1) the function $\varrho(\cdot, y, z) : I[a, y] \to M[k, k]$ is measurable for $(y, z) \in [-b, b] \times C(E_0 \cup E, R^k)$ and $\varrho(x, \cdot) : S_x \times C(E_x, R^k) \to M[k, k]$ is continuous for almost all $x \in [0, a]$,

(2) there exists $\gamma = (\gamma_1, \ldots, \gamma_n) \in L([0, a], R_+^n)$ such that

$$|\varrho_{ij}(x, y, z)| \leqq \gamma_j(x), \quad 1 \leqq j \leqq n, \ 1 \leqq i \leqq k,$$

for $(y, z) \in S_x \times C(E_x, R^k)$ and for almost all $x \in [0, a]$,

(3) there exists $\beta_0 \in \Theta$ such that

$$\|\varrho(x, y, z) - \varrho(x, \bar{y}, \bar{z})\| \leqq \beta_0(x, \kappa) [\|y - \bar{y}\| + \|z - \bar{z}\|_x]$$

for $(y, z), (\bar{y}, \bar{z}) \in S_x \times C_{0.L'}(E_x, R^k; \kappa)$ almost everywhere on $[0, a]$,

(4) for $x \in [0, a]$ we have

$$M(x) \geqq \int_0^x \gamma(\tau) d\tau.$$

Given $\phi \in J[\omega_0, P], c \in (0, a]$ and $z \in K_{c.\phi}[\omega, Q]$, consider the Cauchy problem (11) and its solution $g_i[z](\cdot, x, y)$ with $(x, y) \in E_c \setminus E_0$.

LEMMA 5.1. *Suppose that Assumption $\bar{H}[\varrho]$ is satisfied and $c \in (0, a]$, $\phi, \bar{\phi} \in J[\omega, P]$, $z \in K_{c.\phi}[\omega, Q]$, $\bar{z} \in K_{c.\bar{\phi}}[\omega, Q]$.*

Then for each i, $1 \leqq i \leqq k$ the solutions $g_i[z](\cdot, x, y)$ and $g_i[\bar{z}](\cdot, x, y)$ are defined on such intervals $[0, c_i(x, y)]$ and $[0, \bar{c}_i(x, y)]$ that

$$(c_i(x, y), g_i[z](c_i(x, y), x, y)) \in \partial E_c \quad \text{and} \quad (\bar{c}_i(x, y), g_i[\bar{z}](\bar{c}_i(x, y), x, y)) \in \partial E_c.$$

Moreover, we have the estimates

$$\|g_i[z](t,x,y) - g_i[z](t,\bar{x},\bar{y})\|$$
$$\leqq \left[\|y-\bar{y}\| + \left|\int_{\bar{x}}^{x} \|\gamma(\tau)\|d\tau\right|\right] \exp\left[\left|\int_{x}^{t} \beta(\tau,|Q|)d\tau\right|\right],$$

where $(x,y),(\bar{x},\bar{y}) \in E_c \setminus E_0$, $t \in [0, \min\{c_i(x,y), c_i(\bar{x},\bar{y}\}]$, *and*

$$\|g_i[z](t,x,y) - g_i[\bar{z}](t,x,y)\|$$
$$\leqq \left|\int_{x}^{t} \beta(\tau,|Q|)\|z-\bar{z}\|_\tau d\tau\right| \exp\left[\left|\int_{0}^{t} \beta(\tau,|Q|)d\tau\right|\right],$$

where $(x,y) \in E_c \setminus E_0$, $t \in [0, \min\{c_i(x,y), \bar{c}_i(x,y)\}]$.

The existence and uniqueness of solutions of (11) follows from classical theorems. The proof of the estimates is based on the Gronwall inequality. Details are omitted.

ASSUMPTION $\bar{H}[f]$. Suppose that

(1) the function $f(\cdot,y,z): I[a,y] \to R^k$ is measurable for $(y,z) \in [-b,b] \times C(E_0 \cup E, R^k)$ and $f(x,\cdot): S_x \times C(E_x, R^k) \to R^k$ is continuous for almost all $x \in [0,a]$,

(2) there exist functions $\alpha, \beta \in \Theta$ such that

$$\|f(x,y,z)\| \leqq \alpha(x,\kappa) \text{ for } (y,z) \in S_x \times C(E_0 \cup E, R^k; \kappa)$$

almost everywhere on $[0,a]$ and

$$\|f(x,y,z) - f(x,\bar{y},\bar{z})\| \leqq \beta(x,\kappa)\,[\|y-\bar{y}\| + \|z-\bar{z}\|_x]$$

for $(y,z),(\bar{y},\bar{z}) \in S_x \times C_{0,L'}E(E_x, R^k; \kappa)$ almost everywhere on $[0,a]$.

ASSUMPTION $\bar{H}[A]$. Suppose that

(1) $A \in C(\Omega_0, M[k,k])$ and there is $\nu > 0$ such that $\det A(x,y,p) \geqq \nu$ on Ω_0,

(2) there are nondecreasing functions $\alpha, \beta: R_+ \to R_+$ and $\mu \in L([0,a], R_+)$ such that

$$\|A(x,y,p)\| \leqq \alpha(\kappa)$$

and

$$\|A(x,y,p) - A(\bar{x},\bar{y},\bar{p})\| \leqq \beta(\kappa)\,[\|y-\bar{y}\| + \|p-\bar{p}\|] + \left|\int_{\bar{x}}^{x} \mu(\tau)d\tau\right|,$$

where $(x, y, p,), (\bar{x}, \bar{y}, \bar{p}) \in \Omega_0$ and $\|p\|, \|\bar{p}\| \leqq \kappa$.

Now we formulate the integral functional problem corresponding to (33), (2). Suppose that $\phi \in J[\omega_0, P]$, $c \in (0, a]$, $z \in K_{c.\phi}[\omega, Q]$ and $g_i[z](\cdot, x, y)$, $1 \leqq i \leqq k$, are bicharacteristics of (33) corresponding to z. Let $f[g, z]$, $\Phi[g, z]$, $Z[g, z]$ be the functions defined in Section 2 and

$$A[g, z](t, x, y) = [A_{ij}(t, g_i[z](t, x, y), z(t, g_i[z](t, x, y)))]_{i,j=1,\ldots,k}.$$

Let us define the operator $\tilde{T}_\phi$ for all $z \in K_{c.\phi}[\omega, Q]$ by the formulas

$$\begin{aligned}\tilde{T}_\phi z(x, y) = {} & A^{-1}(x, y, z(x, y))\{A[g, z](0, x, y) * \Phi[g, z](0, x, y)\} \\ & + A^{-1}(x, y, z(x, y)) \int_0^x \Big[D_t A[g, z](t, x, y) * Z[g, z](t, x, y) \\ & + f[g, z](t, x, y)\Big] dt\end{aligned}$$

for $(x, y) \in E_c \setminus E_0$ and

$$\tilde{T} z(x, y) = \phi(x, y) \text{ on } E_0.$$

The main theorem in this section is the following

THEOREM 5.2. *Suppose that $\phi \in J[\omega_0, P]$ and Assumptions $\tilde{H}[\varrho]$, $\tilde{H}[f]$, $\tilde{H}[A]$ are satisfied.*

Then there are $c \in (0, a]$ and $\omega \in L([0, a], R_+)$, $Q \in R_+^2$ such that

(i) $\tilde{T}_\phi : K_{c.\phi}[\omega, Q] \to K_{c.\phi}[\omega, Q]$,

(ii) *the transformation $\tilde{T}_\phi$ has exactly one fixed point $u \in K_{c.\phi}[\omega, Q]$,*

(iii) *the function u is the Carathéodory solution of Problem* (33), (2),

(iv) *if $\bar{\phi} \in J[\omega_0, P]$ and $\bar{u} \in K_{c.\bar{\phi}}[\omega, Q]$ is a solution of* (33) *with the initial condition $z(x, y) = \bar{\phi}(x, y)$ on E_0 then there is $\tilde{\Lambda}_c \in R_+$ such that*

$$\|u - \bar{u}\|_x \leqq \tilde{\Lambda}_c \|\phi - \bar{\phi}\|_0, \quad x \in [0, c].$$

The proof of the theorem is similar to the proof of Theorem 4.1 and it is based on the Banach fixed point principle and on theorems on integral inequalities. Details are omitted.

REMARK 5.3. Suppose that the functions ϱ, f and φ are defined by

$$\varrho(x, y, z) = \bar{\varrho}(x, y, z(x, y)), \quad f(x, y, z) = \bar{f}(x, y, z(x, y)), \quad \varphi(x, y) = \bar{\varphi}(y),$$

where $\bar{\varrho}$, $\bar{f}$, $\bar{\varphi}$ are given in Section 1. Then problem (33), (2) reduces to the Cauchy problem without the functional dependence. Note that in this case our assumptions on given functions are identical with adequate conditions in

[6], where the systems without functional dependence were considered, see also [1]–[3].

REMARK 5.4. We have been working under the assumption that given functions satisfy the Lipschitz condition with respect to (y, z). The following examples show that this assumption is essential.

EXAMPLE 5.5. Let $k = n = 1$ and

$$F(p) = 0 \quad \text{for} \quad p < 0, \quad F(p) = \sqrt{p} \quad \text{for} \quad p \geqq 0,$$

$$\varphi(y) = 0 \quad \text{for} \quad y < 0, \quad \varphi(y) = y \quad \text{for} \quad y \geqq 0.$$

Consider the differential integral equation

(34)
$$D_x z(x,y) = \left[F(z(x,y)) + F\left(\int_0^y z(x,s)\,ds\right)\right] D_y z(x,y) - \sqrt{f(x,y)}\, D_y z(x,y)$$

with the initial condition

(35) $$z(0,y) = \varphi(y) \quad \text{for} \quad y \in R,$$

where

$$f(x,y) = \frac{1}{2}x^2 y + \frac{1}{2}y^2 + \frac{x}{12}\left[\sqrt{(x^2+4y)^3} - x^3\right] \quad \text{for } (x,y) \in [0,a] \times R_+,$$
$$f(x,y) = 0 \quad \text{for } (x,y) \in [0,a] \times R_-$$

and $R_- = (-\infty, 0]$.

It is easily seen that the functions

$$u_+(x,y) = \frac{1}{4}\left(x + \sqrt{x^2+4y}\right)^2, \quad (x,y) \in [0,a] \times (0,+\infty),$$

and

$$u_-(x,y) = 0, \quad (x,y) \in [0,a] \times (-\infty, 0)$$

are unique classical solutions of Problem (34), (35) on $[0,a] \times (0,+\infty)$ and $[0,a] \times (-\infty, 0)$, respectively.

Let

$$v(x,y) = u_+(x,y) \ \text{ on } [0,a] \times R_+, \quad v(x,y) = u_-(x,y) \ \text{ on } [0,a] \times (-\infty, 0).$$

The function v is not continuous at points $(x,0)$ for $x \in (0,a]$. It follows that Problem (34), (35) has not the generalized solution on the set $[0,\varepsilon] \times [-b,b]$ with $\varepsilon > 0$, $b > 0$.

EXAMPLE 5.6. Let $k = n = 1$ and

$$\varphi(y) = 0 \quad \text{for} \quad y < 0 \ \text{ and } \varphi(y) = \sqrt{y} \quad \text{for} \quad y \geqq 0.$$

Consider the Cauchy problem

(36)
$$D_x z(x,y) = \left[z(x,y) + \int_0^y z(x,s)ds\right] D_y z(x,y) - f(x,y) D_y z(x,y),$$

(37) $\qquad z(0,y) = \varphi(y) \quad \text{for} \quad y \in R,$

where

$$f(x,y) = \frac{xy}{2} + \frac{1}{2}\left[\sqrt{(x^2+4y)^3} - x^3\right] \quad \text{for} \quad (x,y) \in [0,a] \times R_+,$$
$$f(x,y) = 0 \quad \text{for} \quad (x,y) \in R_-.$$

It is easy to check that the functions

$$u_+(x,y) = \frac{1}{2}\left(x + \sqrt{x^2+4y}\right) \quad \text{for} \quad (x,y) \in [0,a] \times (0,+\infty),$$

and

$$u_-(x,y) = 0 \quad \text{for} \quad (x,y) \in [0,a] \times (-\infty,0)$$

are unique solutions of Problem (36), (37) on $[0,a] \times (0,+\infty)$ and $[0,a] \times (-\infty,0)$, respectively. Let

$$v(x,y) = u_+(x,y) \ \text{ on } [0,a] \times R_+, \quad v(x,y) = u_-(x,y) \ \text{ on } [0,a] \times (-\infty,0).$$

The function v is not continuous at points $(x,0)$ for $x \in (0,a]$. It follows that Problem (36), (37) has not the generalized solution on $[0,\varepsilon] \times [-b,b]$ with $\varepsilon > 0$, $b > 0$.

REMARK 5.7. Differential systems with a deviated argument and integral differential problems can be obtained from (33) by specializing ϱ and f. Functional differential systems considered in [19] are particular cases of (33) under suitable assumptions on Volterra operators.

REFERENCES

[1] BASSANINI, P., On a boundary value problem for a class of quasilinear hyperbolic systems in two independent variables, *Atti Sem. Mat. Fis. Univ. Modena* **24** (1975), 343–372. *MR* **55** #3548

[2] BASSANINI, P., On a recent proof concerning a boundary value problem for quasilinear hyperbolic systems in the Schauder canonic form, *Boll. Un. Mat. Ital. A* (5), **14** (1977), 325–332. *MR* **58** #11963

[3] BASSANINI, P., Iterative methods for quasilinear hyperbolic systems, *Boll. Un. Mat. Ital. B* (6) **1** (1982), 225–250. *MR* **83g**:35062

[4] BASSANINI, P. and SALVATORI, M. C., Un problema ai limiti per sistemi integrodifferenziali non lineari di tipo iperbolico, *Boll. Un. Mat. Ital. B* (5) **18** (1981), 785–798. *MR* **84b**:45015

[5] Cazzani-Nieri, M. G., Nuovi teoremi di esistenza per un problema di Cauchy nella classe delle funzioni assolutamente continue nell singole variabili, *Ann. Mat. Pura Appl.* (4) **167** (1994), 351–387. *MR* **96b**:35128

[6] Cesari, L., A boundary value problem for quasilinear hyperbolic systems in the Schauder canonic form, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **1** (1974), 311–358. *MR* **52** #1033

[7] Cesari, L., A boundary value problem for quasilinear hyperbolic systems, *Riv. Mat. Univ. Parma* (3) **3** (1974), 107–131. *MR* **55** #8574

[8] Cesari, L., Un problema ai limiti per sistemi di equazioni iperboliche quasi lineari nell forma canonica di Schauder, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* (8) **57** (1974), 303–307. *MR* **53** #13858

[9] Cinquini-Cibrario, M., Una classe di sistemi di equazioni a derivate parziali in piú variabili indipendenti, *Rend. Mat.* (7) **2** (1982), 499–522. *MR* **84g**:35034

[10] Cinquini-Cibrario, M., Sopra una classe di sistemi di equazioni non lineari a derivate parziali in piú variabili indipendenti, *Ann. Mat. Pura Appl.* (4) **140** (1985), 223–253. *MR* **87f**:35050

[11] Człapiński, T., A boundary value problem for quasilinear hyperbolic systems of partial differential-functional equations of the first order, *Boll. Un. Mat. Ital. B* (7) **5** (1991), 619–637. *MR* **92h**:35229

[12] Człapiński, T., On the Cauchy problem for quasilinear hyperbolic systems of partial differential-functional equations of the first order, *Z. Anal. Anwendungen* **10** (1991), 169–182. *MR* **93b**:35082

[13] Człapiński, T., Generalized solutions to boundary value problems for quasilinear hyperbolic systems of partial differential-functional equations, *Ann. Polon. Math.* **57** (1992), 177–191. *MR* **93j**:35175

[14] Człapiński, T. and Kamont, Z., Generalized solutions for quasi-linear hyperbolic systems of partial differential-functional equations, *J. Math. Anal. Appl.* **172** (1993), 353–370. *MR* **94b**:35288

[15] Hale, J. K. and Verduyn-Lunel, S. M., *Introduction to functional differential equations*, Springer, New York, 1993. *MR* **94m**:34169

[16] Kamont, Z., Initial value problems for hyperbolic differential-functional systems, *Boll. Un. Mat. Ital. B* (7) **8** (1994), 965–984. *MR* **95k**:35213

[17] Mattioli, N. M. and Salvatori, M. C., A theorem of existence and uniqueness in nonlinear dispersive optics, *Atti Sem. Mat. Fis. Univ. Modena* **28** (1979), 405–424 (in Italian). *MR* **82k**:78001

[18] Sinestrari, E. and Webb, G. F., Nonlinear hyperbolic systems with nonlocal boundary conditions, *J. Math. Anal. Appl.* **121** (1987), 449–464. *MR* **88d**:35124

[19] Turo, J., On some class of quasilinear hyperbolic systems of partial differential-functional equations of the first order, *Czechoslovak Math. J.* **36** (1986), 185–197. *MR* **88d**:35183

[20] Imanaliev, M. I. and Ved', Yu. A., A first-order partial differential equation with an integral coefficient, *Differentsial'nye Uravneniya* **25** (1989), 465–477 (in Russian). *MR* **90i**:35066

[21] Wu, J., *Theory and applications of partial functional-differential equations*, Applied Mathematical Sciences, 119, Springer, New York, 1996.

(Received September 25, 1996)

INSTYTUT MATEMATYKI
UNIWERSYTET GDAŃSKI
UL. WITA STWOSZA 57
PL–80-952 GDAŃSK
POLAND

czltsz@ksinet.univ.gda.pl
zkamont@ksinet.univ.gda.pl

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 207–215

APPROXIMATING SOLUTIONS OF OPERATOR EQUATIONS AND APPLICATIONS USING MODIFIED CONTRACTIONS

I. K. ARGYROS

Abstract

In this study we are concerned with the problem of approximating a locally unique solution of an operator equation, using inexact Newton-like iterations in a Banach space containing a nondifferentiable term. Earlier results guarantee the convergence of such iterations even in cases when the original Newton-like iteration also converges to the solution. We provide sufficient conditions even in cases when the original Newton-like method fails to converge. We achieve that by carefully choosing the operators involved as well as the residuals. Several applications are provided to show that our results apply where earlier results cannot. In particular we treat a nonlinear equation appearing especially in the harmonic motion of some radioactive particles. Related work can be found in [2], [3], [9].

1. Introduction

In this study we are concerned with the problem of approximating a locally unique solution x^* of the equation

$$F(x) + Q(x) = 0, \tag{1}$$

where F, Q are continuous nonlinear operators defined on some convex subset D of a Banach space E_1 with values in a Banach space E_2.

In a series of papers [2], [3] we introduced the inexact Newton-like method

$$y_n = x_n - A(x_n)^{-1}(F(x_n) + Q(x_n)) \tag{2}$$

$$x_{n+1} = y_n - z_n \quad (n \geqq 0) \tag{3}$$

for some fixed $x_0 \in D$ to approximate a locally unique solution x^* of equation (1). The points $z_n \in E_1$ $(n \geqq 0)$ are chosen so that the sequence $\{x_n\}$ $(n \geqq 0)$ converges to x^*. The linear operator $A(x) \in L(E_1, E_2)$ is usually chosen to be an approximation to the Fréchet-derivative $F'(x)$ of F for all $x \in D$. For $A(x) = F'(x)$, $Q(x) = 0$ $(x \in D)$ and $z_n = 0$ $(n \geqq 0)$, we obtain Newton's method. Sufficient conditions for the convergence of the inexact Newton-like method under various very general conditions has been given in [2], [3].

1991 *Mathematics Subject Classification.* Primary 65J15, 65B05, 47H17, 49D15.
Key words and phrases. Inexact Newton method, Banach space, Hilbert space.

In the above papers our results were compared favorably with earlier ones obtained by others and us for $Q(x)=0$ $(x\in D)$ and $z_n=0$ $(n\geqq 0)$ [7]–[11].

All earlier results guarantee the convergence of inexact Newton-like iteration (2)–(3) in cases when the original process

$$v_{n+1}=v_n-A(v_n)^{-1}(F(v_n)+Q(v_n)) \quad (n\geqq 0) \quad \text{with } v_0=x_0 \tag{4}$$

also converges to a solution x^* of equation (1) which is unique in $U(x^*,r^*)=\{x\in E_1 \mid \|x-x^*\|\leqq r^*\}$ for $r^*\geqq 0$ and provided that $U(x^*,r^*)\subseteq D$. In this study we will find convergence conditions even in cases when the original iteration (4) does not converge to x^*.

Besides the application of iteration (2)–(3) in solving fixed point problems we mention that the investigation of the global asymptotical stability of certain dynamic market systems also requires the convergence analysis of sequences generated by (2)–(3), where n denotes time and x_n is the state of the system at time period n. For such economic models see, for example [4], [5], [9]. Concerning approximation (3) we note that if, for example, an equation on the real line is solved, $F(x_n)+Q(x_n)\geqq 0$ $(n\geqq 0)$, and $A(x_n)$ overestimates the derivative, then y_n is always larger than the corresponding Newton iterate. In such cases, a positive z_n $(n\geqq 0)$ correction term is appropriate.

Finally several applications are provided to show that our results apply whereas earlier they do not. In particular we treat a nonlinear equation appearing in harmonic motion [4], [6].

2. Convergence analysis

We can now formulate our main result on the local convergence of iteration (2)–(3).

THEOREM 1. *Let $F,Q: D\subseteq E_1\to E_2$ be continuous operators. Assume:*

(i) *equation* (1) *has a unique solution $x^*\in D$;*

(ii) *for x_0 sufficiently close to x^*, $U(x^*,r^*)=\{x\in E_1 \mid \|x-x^*\|\leqq r^*\}\subseteq D$ for $r^*\geqq\|x_0-x^*\|$;*

(iii) *linear operator $A(x)$ is invertible for all $x\in U(x^*,r^*)$; and*

(iv) *there exist functions $z: U(x^*,r^*)\to D$, $a: U(x^*,r^*)\to[0,1)$ with*

$$a(x)\leqq b \quad \text{for some } b\in[0,1) \tag{5}$$

such that

$$\|x-A(x)^{-1}(F(x)+Q(x))-z(x)-x^*\|\leqq a(x)\|x-x^*\| \tag{6}$$

for all $x \in U(x^*, r^*)$.

Then the inexact Newton-like method $\{x_n\}$ $(n \geqq 0)$ *generated by* (2)–(3) *with* $z_n = z(x_n)$ $(n \geqq 0)$ *is well defined, remains in* $U(x^*, r^*)$ $(n \geqq 0)$ *and converges to the unique solution* x^* *of equation* $F(x) + Q(x) = 0$ *in* $U(x^*, r^*)$. *Moreover, the following estimate is true:*

$$\|x_{n+1} - x^*\| \leqq \prod_{0 \leqq k \leqq n} a(x_k)\|x_0 - x^*\| \leqq b^{n+1} r^* \quad (n \geqq 0). \tag{7}$$

PROOF. By Hypothesis (iii) the linear operator $A(x_0)$ is invertible since $x_0 \in U(x^*, r^*)$. Hence the iterate x_1 is well defined and by (2), (3), (5), (6) for $x = x_0$ we deduce that $x_1 \in U(x^*, r^*)$, and (7) is true for $n = 0$. Let us assume that $x_m \in U(x^*, r^*)$ and (7) is satisfied for $m = 0, 1, 2, \ldots, n$. Then the iterate x_{m+1} is well defined since the linear operator $A(x_m)$ is invertible. Moreover, using (2)–(3), (5), and (6) for $x = x_m$ we obtain in turn

$$\|x_{m+1} - x^*\| = \|x_m - A(x_m)^{-1}(F(x_m) + Q(x_m)) - z_m - x^*\| \leqq a(x_m)\|x_m - x^*\|$$

$$\leqq \prod_{0 \leqq k \leqq m} a(x_k)\|x_0 - x^*\| \leqq b^{m+1} r^* < r^*. \tag{8}$$

From estimate (8) we deduce that (7) is true, iteration $\{x_n\}$ $(n \geqq 0)$ remains in $U(x^*, r^*)$ for all $n \geqq 0$ and converges to x^* since $b \in [0, 1)$.

That completes the proof of the Theorem.

REMARK 1. Under the hypotheses of Theorem 1 we deduce that

$$\lim_{n \to \infty} z_n = -A(x^*)^{-1}(F(x^*) + Q(x^*)).$$

Hence x^* is a solution of equation (1) if and only if $\lim_{n \to \infty} z_n = 0$.

REMARK 2. We note that for $A(x) = F'(x)$, $Q(x) = 0$, $a(x) = b$ and $z(x) = 0$ $(x \in D)$ (6) reduces to a condition considered first by Kantorovich [8, Theorem XVIII.1.6], [4, Chapter 5].

We will now provide two applications where we show how to choose z, z_n $(n \geqq 0)$, a and b.

APPLICATION 1. Let $P, P_1 : D \subseteq E_1 \to E_2$ be operators such that P is Fréchet-differentiable, whereas P_1 is continuous on D. Choose

$$F(x) = x - P(x), \quad Q(x) = -P_1(x), \quad A(x) = F'(x) \quad (x \in D), \quad E_1 = E_2 \tag{9}$$

and

$$z_n = [(I - P'(w_n))^{-1} - (I - P'(x_n))^{-1}](x_n - (P(x_n) + P_1(x_n))) \quad (n \geqq 0), \tag{10}$$

where the sequence $\{w_n\}$ $(n \geqq 0)$ is in D.

With the above notation we can formulate the following local result:

THEOREM 2. *Assume:*

(i) *the first three conditions of Theorem 1 are satisfied;*

(ii) *there exist nonnegative constants* c_1, c_2, c_3 *such that for all* $x \in U(x^*, r^*)$

$$\|P'(x)\| \leqq c_1, \tag{11}$$

$$\|P_1(x) - P_1(x^*)\| \leqq c_2\|x - x^*\|, \tag{12}$$

and

$$c_1, c_3 \in [0, 1), \tag{13}$$

where

$$c_3 = \frac{4c_1 + c_2(1 + c_1)}{(1 - c_1)^2}. \tag{14}$$

Then the sequence $\{x_n\}$ $(n \geqq 0)$ *generated by* (2)–(3) *is well defined, remains in* $U(x^*, r^*)$ *for all* $n \geqq 0$ *and converges to the unique solution* x^* *of equation* (1) *in* $U(x^*, r^*)$ *provided that the sequence* $\{w_n\}$ $(n \geqq 0)$ *is in* $U(x^*, r^*)$.

Moreover, the following estimates are true for all $n \geqq 0$

$$\|y_n - x^*\| \leqq \frac{2c_1 + c_2}{1 - c_1}\|x_n - x^*\| \tag{15}$$

$$\|z_n\| \leqq \frac{2c_1(1 + c_1 + c_2)}{(1 - c_1)^2}\|x_n - x^*\| \tag{16}$$

and

$$\|x_{n+1} - x^*\| \leqq c_3\|x_n - x^*\|. \tag{17}$$

PROOF. The proof follows immediately from (9)–(14), (1)–(3) by using the following approximations for all $n \geqq 0$

$$\begin{aligned} y_n - x^* = (I - P'(x_n))^{-1}\{&(P(x_n) - P(x^*)) \\ &+ (P_1(x_n) - P_1(x^*)) - P'(x_n)(x_n - x^*)\}, \end{aligned} \tag{18}$$

$$\begin{aligned} z_n = (I - P'(w_n))^{-1}&(P'(w_n) - P'(x_n))(I - P'(x_n))^{-1}[(x_n - x^*) \\ &- (P(x^*) - P(x_n)) + (P_1(x^*) - P_1(x_n))] \end{aligned} \tag{19}$$

and

$$x_{n+1} - x^* = (y_n - x^*) - z_n. \tag{20}$$

REMARK 3. Under the hypotheses of Theorem 2, quantities z, a, b appearing in Theorem 1 can be defined as follows

$$z(x) = [(I - P'(w(x)))^{-1} - (I - P'(x))^{-1}](x - (P(x) + P_1(x))) \quad (x \in U(x^*, r^*))$$

for some function $w: U(x^*, r^*) \to U(x^*, r^*)$,

$$a(x) = c_3 = b \quad (x \in U(x^*, r^*)).$$

REMARK 4. Assume that there exists $c_4 \geqq 0$ such that

$$\|P'(w(x)) - P'(x)\| \leqq c_4 \|w(x) - x\| \tag{21}$$

for all $x \in U(x^*, r^*)$. Then under the hypotheses of Theorem 2, it can easily be seen from (19) that the right-hand side of (16) can be replaced by

$$\frac{c_3(1 + c_1 + c_2)}{(1 - c_1)^2} \|w_n - x_n\| \tag{22}$$

provided that $w(x_n) = w_n$ $(n \geqq 0)$.

We can now show an example for Theorem 2.

EXAMPLE 1. Consider the real function P defined by

$$P(x) = \begin{cases} -(3 - 2\sqrt{2})x & x \leqq (3 - 2\sqrt{2})^{-1}, \\ q(x) & (3 - 2\sqrt{2})^{-1} \leqq x \leqq 7, \\ (3 - 2\sqrt{2})(x - 8) & x \geqq 7, \end{cases}$$

where $q(x)$ joins the two linear portions of $P(x)$ smoothly with $|q'(x)| \leqq (3 - 2\sqrt{2})$. Set $P_1(x) = 0$ $(x \in R)$, $c_2 = 0$, $c_1 = 3 - 2\sqrt{2}$, $x_0 = (3 - 2\sqrt{2})^{-1}$, $x^* = 0$, $w_0 = F(x_0)$, $w_n = 0$ $(n \geqq 1)$. It can easily be seen that with the above choices the hypotheses of Theorem 2 are satisfied. Hence iteration (2)–(3) converges to the unique solution $x^* = 0$ of equation (1). Indeed we get $x_1 = 0 = x^*$. However, the original iteration (4) fails to converge.

APPLICATION 2. Set $E_1 = E_2 = \mathbf{R}^k$, with k a positive integer, in (1) $Q(x) = 0$, $F(x) = x - P(x)$ $(x \in D)$, where $P: D \to D$. Moreover, choose $A(x) = I$ $(x \in D)$, $z_n = -d_n F(x_n)$ $(n \geqq 0)$ for some $d_n \in [0, 1)$. Iteration (2)–(3) can now be written in the form

$$x_{n+1} = d_n x_n + (1 - d_n) P(x_n) \quad (n \geqq 0). \tag{23}$$

The convexity of D implies that the iteration sequence (23) exists for arbitrary $x_0 \in D$. Assume

(24) (C_1) $\|P(x)-x^*\| \leqq K(x)\|x-x^*\|$ for all $x \in D$,

where $x^* \in D$ is a fixed point of P, $\|\cdot\|$ is the Euclidean norm, and $K: D \to R$ is a real valued function;

(25) (C_2) $(x-x^*)^{\mathrm{tr}}(P(x)-x^*) \leqq L(x)\|x-x^*\|$ for all $x \in D$,

where $L: D \to R$ is a real valued function. By the Cauchy–Schwartz inequality we may assume that $L(x) \leqq K(x)$, otherwise we may replace $L(x)$ by $K(x)$. Set

$$K_n = K(x_n),\ L_n = L(x_n),$$
$$P(d_n) = d_n^2(1+K_n^2-2L_n) - 2d_n(K_n^2-L_n) + K_n^2. \tag{26}$$

With the above choices we showed in [5] the following result.

THEOREM 3. *Assume that for each $n \geqq 0$, either*

$$K_n \leqq 1-c_5, \quad L_n > K_n^2; \tag{27}$$

or

$$K_n \leqq 1,\ L_n \leqq K_n^2,\ L_n \leqq 1-c_5 \tag{28}$$

or

$$1 \leqq K_n \leqq c_6, \quad L_n \leqq 1-c_5, \tag{29}$$

where c_5, c_6 are fixed constants. Then with appropriate selection of d_n

$$\|x_{n+1}-x^*\| \leqq (1-\varepsilon)\|x_n-x^*\| \quad (n \geqq 0) \tag{30}$$

with some $\varepsilon > 0$, therefore iteration (23) converges to x^*.

REMARK 5. Let $K(x) = K$ and $L(x) = L$ for all $x \in D$. Assume that either $K < 1$, or $K \geqq 1$ and $L < 1$. Then iteration (23) converges to x^* with the appropriate selection of the coefficients d_n $(n \geqq 0)$.

REMARK 6. If the operator P is bounded and $-P$ is monotonic, then all conditions of Theorem 3 are satisfied. In this special case the best selection is

$$d_n = \frac{K^2}{1+K^2} \quad (n \geqq 0). \tag{31}$$

In this case $K(x) = K$ and $L(x) = 0$ for all $x \in D$.

REMARK 7. With the above notation a, b appearing in Theorem 1 can be chosen such that

$$a(x_n) = \sqrt{P(d_n)} \quad (n \geqq 0)$$

and

$$b = 1-\varepsilon.$$

EXAMPLE 2. Consider the nonlinear equation appearing frequently in harmonic motion ([5], [6], [9])

$$x = e_1 \cos x - e_2 x + e_3,$$

where e_1, e_2, e_3 are given positive constants such that $e_1 < e_2$. Choose $D = \mathbf{R}$ and set $P(x) = e_1 \cos x - e_2 x + e_3$. Since P decreases in x, we may select $L(x) = 0$ for all $x \in D$. If e_2 is sufficiently small, then P is a contraction and therefore iteration (4) (with the above choices of A, F, Q) converges. If e_2 is large enough, the P is not a contraction anymore, and iteration (4) diverges. However, all conditions of Remark 5 and the last case of Theorem 3 are satisfied, therefore iteration (23) converges with an appropriate selection of d_n $(n \geqq 0)$. Since K can be selected as $e_1 + e_2$, Remark 6 implies that the choice $d_n = \dfrac{(e_1 + e_2)^2}{1 + (e_1 + e_2)^2}$ is satisfactory.

In what follows we provide a semilocal convergence theorem for the inexact Newton-like iterations generated by (2)–(3).

THEOREM 4. *Let $F, Q: D \subseteq E_1 \to E_2$ be continuous operators. Assume:*

(i) *the linear operator $A(x)$ is invertible for all $x \in U(x_0, r_0)$ with $x_0 \in D$ and $r_0 \geqq 0$ such that $U(x_0, r_0) \subseteq D$;*

(ii) *there exist functions $z: U(x_0, r_0) \to D$, $p, q: U(x_0, r_0) \to [0, 1)$ and constants p_1, q_1 such that*

$$\|x - x_0 - A(x)^{-1}(F(x) + Q(x)) - z(x)\| \leqq p(x) r_0, \quad p(x) \leqq p_1 \leqq 1, \tag{32}$$

$$\|A(x)^{-1}(F(x) + Q(x)) + z(x)\| \leqq q(x) \|A(y)^{-1}(F(y) + Q(y)) + z(y)\|, \quad q(x) \leqq q_1 < 1 \tag{33}$$

for all $x, y \in U(x_0, r_0)$ with

$$x = y - A(y)^{-1}(F(y) + Q(y)) - z(y). \tag{34}$$

(iii) *The sequence $\{z_n\}$ $(n \geqq 0)$ with $z_n = z(x_n)$ is null.*

Then the inexact Newton-like method generated by (2)–(3) is well defined, remains in $U(x_0, r_0)$ for all $n \geqq 0$ and converges to a solution $x^* \in U(x_0, r_0)$ of equation (1). Moreover, the following estimates are true:

$$\|x_{n+1} - x_n\| \leqq q_1 \|x_n - x_{n-1}\| \quad (n \geqq 1) \tag{35}$$

and

$$\|x_n - x^*\| \leqq \frac{q_1^n}{1 - q_1} \|x_1 - x_0\| \quad (n \geqq 0). \tag{36}$$

PROOF. Using (i), (32) for $x = x_n$, (34) for $x = x_{n+1}$, $y = x_n$ and induction on $n \geqq 0$ we obtain that the iteration $\{x_n\}$ $(n \geqq 0)$ is well defined and remains in $U(x_0, r_0)$ for all $n \geqq 0$. Setting $x = x_n$ and $y = x_{n-1}$ in (33), we obtain by (34) that estimate (35) is true for all $n \geqq 1$. But (35) shows that the iteration $\{x_n\}$ $(n \geqq 0)$ is Cauchy in a Banach space E_1 and as such it converges to some $x^* \in U(x_0, r_0)$. By letting $n \to \infty$ in (2)–(3) and using hypothesis (iii) we deduce that $F(x^*) + Q(x^*) = 0$. Hence x^* is a solution of equation (1). From (35) we obtain in turn

$$\|x_{n+1} - x_n\| \leqq q_1 \|x_n - x_{n-1}\| \leqq q_1^2 \|x_{n-1} - x_{n-2}\| \leqq \cdots \leqq q_1^n \|x_1 - x_0\|.$$

Hence for all $m \geqq 0$ we get

$$\|x_{n+m} - x_n\| \leqq q_1^n \frac{1 - q_1^m}{1 - q_1} \|x_1 - x_0\|,$$

and by letting $m \to \infty$ we obtain (36).

That completes the proof of the Theorem.

EXAMPLE 3. Let $z(x) = 0$, $Q(x) = 0$ and $A(x) = F'(x_0)$ $(x \in D)$. Assume that the operator F is defined and Fréchet-differentiable in a ball $U(x_0, r)$, in which the Fréchet derivative $F'(x)$ satisfies a Lipschitz condition

$$\|F'(x_0)^{-1}(F'(x) - F'(y))\| \leqq l\|x - y\|.$$

Moreover, set $\|F'(x_0)^{-1}F(x_0)\| \leqq \eta_0$ and assume

$$h_0 = l\eta_0 < \frac{1}{2} \quad \text{and} \quad r_0 = \frac{1 - \sqrt{1 - 2h_0}}{l} \leqq r.$$

Then for all $x \in U(x_0, r_0)$ we get

$$\begin{aligned}
&\|x - x_0 - F'(x_0)^{-1}F(x)\| \\
&= \|F'(x_0)^{-1}(F'(x_0)(x - x_0) - (F(x) - F(x_0)) - F'(x_0)^{-1}F(x_0)\| \\
&\leqq \frac{l}{2}\|x - x_0\|^2 + \|F'(x_0)^{-1}F(x_0)\| \\
&\leqq \frac{l}{2} r_0^2 + \eta_0 = r_0.
\end{aligned}$$

Hence, we can choose $p(x) = p_1 = 1$ for all $x \in U(x_0, r_0)$ in (32). Similarly we can show that for all $x, y \in U(x_0, r_0)$ (33) is satisfied if we choose $q(x) = l\|x - x_0\|$ and $q_1 = 1 - \sqrt{1 - 2h_0}$.

The conclusions of Theorem 4 can now follow. We note that with the above choices our Theorem 4 reduces to Kantorovich's Theorem [8, Theorem XVIII.1.6].

Conclusion

In this study we examine the problem of approximating a locally unique solution of an operator equation, using inexact Newton-like iterations in a Banach space containing a nondifferentiable term. Earlier results guarantee the convergence of such iterations in cases when the original Newton-like iteration converges also. By choosing the operators involved as well as the residuals carefully, we showed convergence of the inexact Newton-like method to a solution of the operator equation even in cases when the original Newton-like iteration fails to converge. We provide an error analysis for our method. Several applications are also given to show that our results apply where earlier results cannot. In particular we treat a nonlinear equation appearing especially in the harmonic motion of some radioactive particles. Related work can be found especially in [2], [3], [5], [9].

REFERENCES

[1] Argyros, I. K., A unified approach for constructing fast two-step Newton-like methods, *Monatsh. Math.* **119** (1995), 1–22. *MR* **96a**:65093

[2] Argyros, I. K., A convergence theorem for Newton-like methods under generalized Chen–Yamamoto-type assumptions, *Appl. Math. Comput.* **61** (1994), 25–37. *MR* **95g**:65082

[3] Argyros, I. K., On the discretization of Newton-like methods, *Internat. J. Comput. Math.* **52** (1994), 161–170.

[4] Argyros, I. K. and Szidarovszky, F., *The theory and applications of iteration methods*, Systems Engineering Series, C.R.C. Press, Inc., Boca Raton, Florida, 1993. *MR* **95b**:65001

[5] Argyros, I. K. and Szidarovszky, F., On the convergence of modified contractions, *J. Comput. Appl. Math.* **55** (1994), 183–189. *MR* **96a**:65085

[6] Chandrasekhar, S., *Radiative transfer*, Dover Publications, Inc., New York, 1960. *MR* **22** #2446

[7] Chen, X. J. and Yamamoto, T., Convergence domains of certain iterative methods for solving nonlinear equations, *Numer. Funct. Anal. Optim.* **10** (1989), 37–48. *MR* **90a**:65137

[8] Kantorovich, L. V. and Akilov, G. P., *Functional analysis in normed spaces*, International Series of Monographs in Pure and Applied Mathematics, Vol. 46, The Macmillan Co., New York, 1964. *MR* **35** #4699

[9] Liu, D. and Szidarovszky, F., Global asymptotic stability of dynamic systems with modified contractions, *Appl. Math. Comput.* **43** (1991), 237–240. *MR CMP* 91 11

[10] Potra, F. A. and Ptak, V., Sharp error bounds for Newton's process, *Numer. Math.* **34** (1980), 63–72. *MR* **81c**:65027

[11] Zabrejko, P. P. and Nguen, D. F., The majorant method in the theory of Newton–Kantorovich approximations and the Ptak error estimates, *Numer. Funct. Anal. Optim.* **9** (1987), 671–684. *MR* **88h**:65129

(Received October 1, 1996)

DEPARTMENT OF MATHEMATICS
CAMERON UNIVERSITY
LAWTON, OK 73505
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 217–236

EXISTENCE AND CONSTRUCTION OF DEFINITE ESTIMATION FUNCTIONALS

K. DIETHELM

Abstract

Let G be a continuous linear functional on $C^s[a,b]$. An estimation functional for G is a functional of the form $E[f]=\sum_{j=1}^{n}\sum_{k=0}^{s}\alpha_{jk}f^{(k)}(x_j)$. For $r\geqq s$, the functional G admits r-positive definite estimation (r-negative definite estimation) if an estimation functional E for G exists such that, for every $f\in C^r[a,b]$ with $f^{(r)}\geqq 0$, there holds $G[f]-E[f]\geqq 0$ ($G[f]-E[f]\leqq 0$). In this paper, we state necessary and sufficient conditions on G for such estimation functionals to exist. In particular, we characterize the most interesting functionals, namely those that admit both r-positive definite estimation and r-negative definite estimation. We also solve this problem under the additional restriction that only estimation functionals of the form $E[f]=\sum_{j=1}^{n}\alpha_j f(x_j)$ are allowed. The proofs are constructive. Some examples are also included.

1. Introduction

Let G be a continuous linear functional on $C^s[a,b]$, where the (real) linear space $C^s[a,b]$ is endowed with the norm $\|f\|=\sum_{k=0}^{s-1}\left|f^{(k)}(a)\right|+\left\|f^{(s)}\right\|_\infty$. For the approximation of such a functional, one frequently uses (linear) point functionals, i.e. functionals of the form

$$E[f]=\sum_{j=1}^{n}\sum_{k=0}^{s}\alpha_{jk}f^{(k)}(x_j). \tag{1}$$

Such a functional is called an *estimation functional* for G. One of the most useful properties an estimation functional can have is definiteness: For $r\geqq s$, E is called r-positive definite (r-negative definite) with respect to G if, for every $f\in C^r[a,b]$ with $f^{(r)}\geqq 0$, there holds $G[f]-E[f]\geqq 0$ ($G[f]-E[f]\leqq 0$). (When there is no danger of confusion, we shall drop the reference "with respect to G" in the rest of this paper.) If an r-positive definite (r-negative

1991 *Mathematics Subject Classification.* Primary 41A80, 41A29.
Key words and phrases. Approximation of functionals, definiteness.

0081–6906/99/$ 5.00

definite) estimation functional E exists, we say that the functional G *admits r-positive definite estimation (r-negative definite estimation).* Classical examples are, for $r = 2$, the midpoint formula and the trapezoidal formula as estimations for the integral $G[f] = \int_a^b f(x)dx$ [1, pp. 60ff.]. A number of criteria are known that can be used to investigate whether a given estimation functional is definite or not, cf., e.g., Brass and Schmeisser [2] (where also some equivalent characterizations of definiteness are stated), Förster [6], [7], Köhler [8] or the author [4].

In this paper, we will discuss the conditions on G that are necessary and/or sufficient for definite estimation functionals to exist. The case that G admits both r-positive definite estimation and r-negative definite estimation is of particular interest since in this case we can give a guaranteed inclusion for the true value of $G[f]$ under the assumption that the r-th derivative of f does not change its sign: Let E^+ (E^-) be an r-positive definite (r-negative definite) estimation functional, then we have

$$E^+[f] \leqq G[f] \leqq E^-[f] \tag{2}$$

if, say, $f^{(r)} \geqq 0$. We shall see, however, that, in contrast to almost all previous observations, there exist functionals that admit r-positive definite estimation but not r-negative definite estimation (or vice versa) for some r. An example of such a functional has recently been described in [3] (see also § 4 below).

In the remainder of this section, for the convenience of the reader, we collect some well-known results which we shall use in the later sections. § 2 contains the results on the existence of definite estimation functionals. In § 3, we deal with the case that for the estimation of $G[f]$ only function values of f are available, but no information about derivatives. We thus have to investigate the problem under the restriction that only estimation functionals $\widehat{E}$ of the form

$$\widehat{E}[f] = \sum_{j=1}^{n} \alpha_j f(x_j) \tag{3}$$

are allowed instead of the more general form (1). We shall see that additional conditions must be imposed on G in order to ensure the existence of definite estimation functionals of this restricted form. Finally, § 4 will contain a number of examples.

The following results about the representation of continuous linear functionals can be found, e.g., in the book of Sard [9, Chapters 1 and 3].

First, we recall that a function $f : [a, b] \to \mathbb{R}$ is said to be a *normalized function of bounded variation* if f is of bounded variation on $[a, b]$, $f(a) = 0$, and $f(x + 0) = f(x)$ for $x > a$. The set of normalized functions of bounded variation will be denoted by V_0. Then we have the following representation theorem [9, p. 139]:

THEOREM 1.1. *Let G be a continuous linear functional on $C^s[a,b]$, and let $y \in [a,b]$. Then there exist uniquely determined real numbers $c_{0,y}$, $c_{1,y}$, $\ldots$, $c_{s-1,y}$, and a uniquely determined function $\mu_{s,y} \in V_0$ such that, for every $f \in C^s[a,b]$, there holds*

$$G[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s)}(x) d\mu_{s,y}(x). \tag{4}$$

The representation (4) for G will be called the *canonical representation* for G (with respect to the point y).

It will turn out that it depends on the properties of $\mu_{s,y}$ whether G admits s-definite estimation or not. Therefore, for the application of our results, we need some method to calculate $\mu_{s,y}$ when G is given. Such a method is given in [9, Chapter 3, equations (40) and (41)]:

THEOREM 1.2. *Under the assumptions of Theorem 1.1, we have that*

$$c_{j,y} = \frac{1}{j!} G[(\cdot - y)^j], \qquad 0 \leqq j \leqq s-1.$$

Furthermore, we have

$$\mu_{0,y}(x) = \begin{cases} 0 & \text{if } x = a, \\ \lim\limits_{n\to\infty} G[\theta_n(x,\cdot)] & \text{if } a < x \leqq b, \end{cases}$$

and, for $s \geqq 1$,

$$\mu_{s,y}(x) = \begin{cases} 0 & \text{if } x = a, \\ \dfrac{1}{(s-1)!} \lim\limits_{n\to\infty} G\left[\displaystyle\int_y (\cdot - t)^{s-1} \theta_n(x,t)\, dt\right] & \text{if } a < x \leqq b, \end{cases}$$

where

$$\theta_n(x,t) = \begin{cases} 1 & \text{if } t \leqq x, \\ 1 + n(x-t) & \text{if } x < t < x + 1/n, \\ 0 & \text{if } t \geqq x + 1/n. \end{cases}$$

We remark that a functional that is linear and continuous on $C^s[a,b]$ is also linear and continuous on all the sets $C^r[a,b]$ with $r \geqq s$. Thus, we have a canonical representation for such a functional for every $r \geqq s$. The following theorem establishes the relation between the measures $\mu_{s,y}$ and $\mu_{r,y}$.

THEOREM 1.3. *Let G be a continuous linear functional on $C^s[a,b]$. Consider the canonical representation of G with respect to y. We have*

$$c_{s,y} = \mu_{s,y}(b) \qquad \text{and}$$

$$\mu_{s+1,y}(x) = -\int_a^x \mu_{s,y}(t)dt + \mu_{s,y}(b)(x-y)_+.$$

Here, $(\cdot)_+$ denotes the truncated power function given by $t_+ = 0$ if $t < 0$ and $t_+ = t$ if $t \geq 0$.

PROOF. From Theorem 1.2 and Theorem 1.1, we have

$$s!c_{s,y} = G[(\cdot - y)^s] = \sum_{j=0}^{s-1} c_{j,y} \left[\frac{d^j}{dx^j}(x-y)^s\right]_{x=y} + \int_a^b \left[\frac{d^s}{dx^s}(x-y)^s\right] d\mu_{s,y}(x)$$

$$= s! \int_a^b d\mu_{s,y}(x) = s!\mu_{s,y}(b).$$

Furthermore, partial integration of (4) yields

$$G[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + f^{(s)}(b)\mu_{s,y}(b) - \int_a^b \mu_{s,y}(x)df^{(s)}(x)$$

$$= \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + f^{(s)}(b)\mu_{s,y}(b) + \int_a^b f^{(s+1)}(x)dM(x),$$

where $M(x) = -\int_a^x \mu_{s,y}(t)dt$. Now,

$$f^{(s)}(b)\mu_{s,y}(b) = \left(f^{(s)}(b) - f^{(s)}(y)\right)\mu_{s,y}(b) + f^{(s)}(y)\mu_{s,y}(b)$$

$$= f^{(s)}(y)c_{s,y} + \mu_{s,y}(b)\int_y^b f^{(s+1)}(x)dx,$$

and thus

$$G[f] = \sum_{j=0}^{s} c_{j,y} f^{(j)}(y) + \mu_{s,y}(b)\int_y^b f^{(s+1)}(x)dx + \int_a^b f^{(s+1)}(x)dM(x)$$

$$= \sum_{j=0}^{s} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s+1)}(x)d(M(x) + \mu_{s,y}(b)(x-y)_+).$$

From the uniqueness of the representation (4), the theorem follows. □

2. Approximation using general estimation functionals

In this section, we discuss the case of general estimation functionals. Since the proofs of the results are not essentially different from those concerning restricted estimation functionals (cf. §3), we only give explicit proofs for the latter. We remark here that, in the proofs of our results, we cannot only show that definite estimation functionals exist (if they exist), but we will actually construct such estimation functionals.

First of all, let G be a continuous linear functional on $C^s[a,b]$. Then, by obvious symmetry arguments, G admits s-negative definite estimation if and only if $-G$ admits s-positive definite estimation. Thus, we may restrict our attention on the one-sided case to functionals admitting s-positive definite estimation.

Our first result gives a characterization of the functionals having this property.

THEOREM 2.1. *Let G be a continuous linear functional on $C^s[a,b]$, given by its canonical representation* (4) *with respect to the point y. Then the following statements are equivalent:*

(a) *G admits s-positive definite estimation.*

(b) *For every $y \in [a,b]$, there exists a polynomial p_y of degree s and a step function τ_y with finitely many jumps such that the function $\mu_{s,y} - p_y - \tau_y$ is nondecreasing.*

(c) *There exist $y \in [a,b]$, a polynomial p_y of degree s and a step function τ_y with finitely many jumps such that the function $\mu_{s,y} - p_y - \tau_y$ is nondecreasing.*

The case $s = 0$ of Theorem 2.1 admits a particularly simple reformulation. Recall that, by Theorem 1.2, $\mu_{0,y}$ is actually independent of y.

COROLLARY 2.2. *Let G be a continuous linear functional on $C^0[a,b]$. Then G admits 0-positive definite estimation if and only if $\mu_{0,y}$ is decreasing at finitely many points only.*

We can also give a reformulation of Theorem 2.1 in the case $s \geqq 1$. For this purpose, we recall that a function $\phi \in V_0$ can be decomposed into an absolutely continuous function $\phi^{[\mathrm{ac}]}$ and a step function $\phi^{[\mathrm{s}]}$ with countably many jumps according to $\phi = \phi^{[\mathrm{ac}]} + \phi^{[\mathrm{s}]}$. The function $\phi^{[\mathrm{ac}]}$ will be called the *continuous part of ϕ*, and $\phi^{[\mathrm{s}]}$ will be called the *discontinuous part of ϕ*. Since $\phi^{[\mathrm{ac}]}$ is absolutely continuous, there exists a function $\phi^{[\mathrm{ac}]\prime} \in L_1[a,b]$ such that $\phi^{[\mathrm{ac}]}(x) - \phi^{[\mathrm{ac}]}(a) = \int_a^x \phi^{[\mathrm{ac}]\prime}(t)dt$ holds for $x \in [a,b]$. The function $\phi^{[\mathrm{ac}]\prime}$ will be called a *(generalized) derivative of $\phi^{[\mathrm{ac}]}$*. Then we have the following result.

COROLLARY 2.3. *Let $s \geqq 1$, and let G be a continuous linear functional on $C^s[a,b]$. Then the following statements are equivalent.*

(a) *G admits s-positive definite estimation.*
(b) *For every $y \in [a,b]$, the discontinuous part of $\mu_{s,y}$ has got finitely many jumps in negative direction, and, for the continuous part $\mu_{s,y}^{[\mathrm{ac}]}$ of $\mu_{s,y}$, there holds $\operatorname{ess\,inf}_{x\in[a,b]} \mu_{s,y}^{[\mathrm{ac}]\prime}(x) > -\infty$.*
(c) *There exists $y \in [a,b]$ such that $\mu_{s,y}$ fulfils the conditions of* (b).

As mentioned in the introduction, it is particularly useful if G admits both s-positive definite estimation and s-negative definite estimation. Based on Theorem 2.1, we can now characterize these functionals. We start with the case $s = 0$ and find that, among all continuous linear functionals on $C^0[a,b]$, only the trivial functionals admit 0-positive definite estimation and 0-negative definite estimation.

THEOREM 2.4. *Let G be a continuous linear functional on $C^0[a,b]$. Then G admits 0-positive definite estimation and 0-negative definite estimation if and only if there exists $n \in \mathbb{N}$ and numbers $x_1, x_2, \ldots, x_n \in [a,b]$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$ such that*

$$G[f] = \sum_{j=1}^{n} a_j f(x_j).$$

For functionals on $C^s[a,b]$, $s \geqq 1$, the situation is slightly more complex.

THEOREM 2.5. *Let $s \geqq 1$ and let G be a continuous linear functional on $C^s[a,b]$, given by its canonical representation* (4). *Then the following statements are equivalent:*

(a) *G admits s-positive definite estimation and s-negative definite estimation.*
(b) *For every $y \in [a,b]$, there exists a step function τ_y with finitely many steps such that $\mu_{s,y} - \tau_y$ fulfils a Lipschitz condition.*
(c) *There exist $y \in [a,b]$ and a step function τ_y fulfilling the condition of* (b).

As a consequence of the results above, we have the following very simple sufficient criterion.

COROLLARY 2.6. *Let $s > r \geqq 0$. Let G_1 be a continuous linear functional on $C^r[a,b]$, and let $G_2[f] := \sum_{j=1}^{n} \beta_j f^{(s)}(x_j)$ with $x_j \in [a,b]$ for $j = 1, 2, \ldots, n$. Then, $G := G_1 + G_2$ is a continuous linear functional on $C^s[a,b]$, and G admits s-positive definite estimation and s-negative definite estimation.*

REMARKS. 1. The case $G_2 \equiv 0$ is especially important. Explicitly, it reads: If G is a continuous linear functional on $C^r[a,b]$ and $s > r$, then G admits s-positive definite estimation and s-negative definite estimation.

2. We do not have equivalence in Corollary 2.6. To see this, consider a function $\phi \in C[a,b]$ with $\phi(a) = 0$ which is not of bounded variation, and

define $\Phi(x) := \int_a^x \phi(t)dt$. Then Φ fulfils a Lipschitz condition, and by Theorem 2.5, the functional G defined by $G[f] := \int_a^b f^{(s)}(x)d\Phi(x)$ admits s-positive definite estimation and s-negative definite estimation. Now, assume that G is a continuous linear functional on $C^{s-1}[a,b]$. Then, from the definition of G, a partial integration yields

$$G[f] = -\phi(b)f^{(s-1)}(b) - \int_a^b f^{(s-1)}(x)d\phi(x) = \int_a^b f^{(s-1)}(x)d\phi^*(x),$$

where $\phi^*(x) = -\phi(x)$ if $a \leqq x < b$ and $\phi^*(b) = -2\phi(b)$. Since ϕ is not of bounded variation, ϕ^* is not of bounded variation either. But by Theorem 1.1, there exists a function $\psi \in V_0$ such that

$$G[f] = \int_a^b f^{(s-1)}(x)d\psi(x).$$

Hence, for every $f \in C^{s-1}[a,b]$, we have that

$$\int_a^b f^{(s-1)}(x)d(\psi(x) - \phi^*(x)) = 0,$$

which yields $\phi^* = \psi$, a contradiction. Thus, G cannot be a continuous linear functional on $C^{s-1}[a,b]$. Another reasoning that can be applied to show this uses the fact that, since ϕ is not of bounded variation, a classical result [9] states that the integral $\int_a^b f^{(s-1)}(x)d\phi(x)$ in the definition of G does not exist for every $f \in C^{s-1}[a,b]$.

3. Approximation using restricted estimation functionals

In some situations, information about the derivatives of f is not available or can be obtained only under unreasonable difficulties. In this case, estimation functionals of the form (1) must be replaced by functionals of the form

$$E[f] = \sum_{j=1}^{n} \alpha_j f(x_j).$$

These functionals will be called *restricted estimation functionals.* Consequently, we say that a continuous linear functional G *admits restricted s-positive definite estimation (s-negative definite estimation)* if an s-positive definite (s-negative definite) restricted estimation functional E exists.

It is immediately clear that G admits s-positive definite estimation if it admits restricted s-positive definite estimation, and analoguously for negative definite estimation. The converse statement, however, is not true in general. In this section, we shall see how the additional conditions on G must be chosen. First, we consider the case $s = 0$. This is the only case where no additional conditions are necessary.

It is again obvious that we can restrict our attention on the one-sided case to the problem of restricted s-positive definite estimation.

THEOREM 3.1. *Let G be a continuous linear functional on $C^0[a,b]$. Then G admits restricted 0-positive definite estimation if and only if G admits 0-positive definite estimation.*

PROOF. The direction "⇒" is clear. To prove "⇐", we note that, if, say, G admits 0-positive definite estimation, the associated estimation functional is necessarily a restricted estimation functional. □

The principal result in the case $s = 1$ is the following theorem which states that the function $\mu_{1,y}$ in the canonical representation of G must have the properties described in § 2 and, additionally, it must not have jumps in the "wrong" direction.

THEOREM 3.2. *Let G be a continuous linear functional on $C^1[a,b]$. Then the following statements are equivalent.*

(a) *G admits restricted 1-positive definite estimation.*

(b) *For every $y \in [a,b]$, there exists a polynomial p_y of degree 1 such that the function $\mu_{1,y} - p_y$ is nondecreasing.*

(c) *There exist $y \in [a,b]$ and a polynomial p_y of degree 1 such that the function $\mu_{1,y} - p_y$ is nondecreasing.*

(d) *For every $y \in [a,b]$, the discontinuous part of $\mu_{1,y}$ has got no jumps in negative direction, and, for the continuous part $\mu_{1,y}^{[\mathrm{ac}]}$ of $\mu_{1,y}$, there holds $\operatorname{ess\,inf}_{x\in[a,b]} \mu_{1,y}^{[\mathrm{ac}]'}(x) > -\infty$.*

(e) *There exists $y \in [a,b]$ such that $\mu_{1,y}$ fulfils the conditions of* (d).

PROOF. The implications (b) ⇒ (c) and (d) ⇒ (e) are trivial.

For the proof of (c) ⇒ (a), define $E[f] := p_y'(a)(f(b) - f(a))$. Then a short calculation yields

$$G[f] - E[f] = \int_a^b f'(x)\,d(\mu_{1,y}(x) - p_y(x)) \geqq 0$$

if $f' \geqq 0$. Thus, E is a restricted 1-positive definite estimation functional for G.

To prove (a) $\Rightarrow$ (b), let $y \in [a, b]$. By assumption, there exists a restricted estimation functional E such that $G[f] - E[f] \geqq 0$ whenever $f' \geqq 0$. Note that this implies $G[f] = E[f]$ if f is a constant function. Thus, the canonical representation of $G - E$ is

$$G[f] - E[f] = \int_a^b f'(x) d\rho(x),$$

where ρ is nondecreasing and independent of y. Hence,

$$E[f] = c_{0,y} f(y) + \int_a^b f'(x) d(\mu_{1,y}(x) - \rho(x)).$$

Now, since E is a restricted estimation functional, we see that $\mu_{1,y} - \rho$ is a piecewise linear spline function and, in particular, it fulfils a Lipschitz condition, i.e. there exist real constants $\alpha_{1,y}, \alpha_{2,y}$ such that for every $x, \widehat{x} \in [a, b]$, we have

$$\alpha_{1,y} \leqq \frac{(\mu_{1,y}(x) - \rho(x)) - (\mu_{1,y}(\widehat{x}) - \rho(\widehat{x}))}{x - \widehat{x}} \leqq \alpha_{2,y}.$$

Setting $p_y(x) := \alpha_{1,y} x$, we obtain that $\mu_{1,y} - \rho - p_y$ is nondecreasing. Since ρ is also nondecreasing, statement (b) follows.

For the conclusion (b) $\Rightarrow$ (d), we note that, if either of the conditions of (d) would be wrong, then we would have a contradiction to statement (b) of the present theorem.

Finally, to prove (e) $\Rightarrow$ (c), choose $p_y(x) := x \operatorname*{ess\,inf}_{x \in [a,b]} \mu_{1,y}^{[\mathrm{ac}]'}(x)$. Then a simple calculation shows that $\mu_{1,y} - p_y$ is nondecreasing. □

In the case $s \geqq 2$, we must impose even more additional restrictions on the measure $\mu_{s,y}$: In addition to the conditions of § 2 and the condition on the jumps mentioned for the case $s = 1$, we must demand a certain "regular" behaviour near the end points of the interval $[a, b]$.

THEOREM 3.3. *Let $s \geqq 2$, and let G be a continuous linear functional on $C^s[a, b]$. Then the following statements are equivalent:*

(a) *G admits restricted s-positive definite estimation.*

(b) *For every $y \in (a, b)$, the following four conditions hold:*

(I_+) *$\mu_{s,y}$ does not have jumps in negative direction.*

(II_+) $\operatorname*{ess\,inf}_{x \in [a,b]} \mu_{s,y}^{[\mathrm{ac}]'}(x) > -\infty$.

(III_+) $\liminf_{x\to a}(x-a)^{1-s}\mu_{s,y}^{[\mathrm{ac}]'}(x) > -\infty.$

(IV_+) $\liminf_{x\to b}(b-x)^{1-s}\mu_{s,y}^{[\mathrm{ac}]'}(x) > -\infty.$

(c) *There exists* $y \in (a,b)$ *such that conditions* (I_+)–(IV_+) *of* (b) *hold.*

(d) *For* $y = a$, *conditions* (I_+), (II_+) *and* (IV_+) *of* (b) *hold, and there holds*

$$(\mathrm{III}_+^*)\quad \liminf_{x\to a}\left[(x-a)^{1-s}\mu_{s,a}^{[\mathrm{ac}]'}(x)+(-1)^s\sum_{j=0}^{s-1}c_{j,a}(a-x)^{-j}/(s-1-j)!\right] > -\infty.$$

(e) *For* $y = b$, *conditions* (I_+), (II_+) *and* (III_+) *of* (b) *hold, and there holds*

$$(\mathrm{IV}_+^*)\quad \liminf_{x\to b}\left[(b-x)^{1-s}\mu_{s,b}^{[\mathrm{ac}]'}(x)+\sum_{j=0}^{s-1}c_{j,b}(b-x)^{-j}/(s-1-j)!\right] > -\infty.$$

For the proof, we shall use

LEMMA 3.4. *Let* $s \geqq 0$, *and let* G *be a continuous linear functional on* $C^s[a,b]$. *For some arbitrary but fixed* $a < x_0 < x_1 < x_2 < \cdots < x_r < b$, *let* $\Pi_r[f]$ *be the interpolating polynomial for* f *with nodes* $x_0, x_1, \ldots, x_r$. *Then* G *admits restricted* s*-positive definite estimation if and only if* $G - G_r$ *admits restricted* s*-positive definite estimation, where* $G_r := G \circ \Pi_r$.

PROOF. For $0 \leqq j \leqq r$, let l_j denote the j-th Lagrange polynomial with respect to the nodes $x_0, x_1, \ldots, x_r$. Then

$$G_r[f] = G[\Pi_r[f]] = G\left[\sum_{j=0}^{r} f(x_j)l_j\right] = \sum_{j=0}^{r} f(x_j)G[l_j].$$

Therefore, we can see that G_r is a restricted estimation functional. Thus, if E_1 is an s-positive definite restricted estimation functional for G, then $E_1 - G_r$ is an s-positive definite restricted estimation functional for $G - G_r$. On the other hand, if E_2 is an s-positive definite restricted estimation functional for $G - G_r$, then $E_2 + G_r$ is an s-positive definite restricted estimation functional for G. □

PROOF OF THEOREM 3.3. (b) ⇒ (c) is obvious.

To prove (a) ⇒ (b), let $y \in (a,b)$, and let E be an s-positive definite restricted estimation functional for G. Then E has got the canonical representations

$$E[f] = \int_a^b f(x)d\epsilon_{0,y}(x) = \sum_{j=0}^{s-1} c_{j,y}f^{(j)}(y) + \int_a^b f^{(s)}(x)d\epsilon_{s,y}(x).$$

Here, $\epsilon_{0,y}$ is a step function. Therefore, by repeated application of Theorem 1.3, we can see that $\epsilon_{s,y}$ is a piecewise polynomial of degree s which has

got s-fold zeros at a and b. Furthermore, we can see from Theorem 1.3 that $\epsilon_{s,y}$ is Lipschitz continuous. Now, since E is an s-positive definite estimation functional for G, $\mu_{s,y} - \epsilon_{s,y}$ is nondecreasing. Therefore, $\mu_{s,y}$ cannot have jumps in negative direction, proving (I_+). Condition (II_+) follows in a similar way as in the proof of Theorem 3.2 (d). Since $\epsilon_{s,y}$ has got an s-fold zero at a, we have that, for some $\delta_a > 0$ and $x \in [a, a+\delta_a)$, $\epsilon_{s,y}(x) = \gamma_a(x-a)^s$. Since $\mu_{s,y}^{[\mathrm{ac}]} - \epsilon_{s,y}$ is nondecreasing, we have that, for $x \in [a, a+\delta_a)$,

$$\mu_{s,y}^{[\mathrm{ac}]\prime}(x) \geqq \epsilon'_{s,y}(x) = s\gamma_a(x-a)^{s-1},$$

which implies (III_+). Similarly, we obtain (IV_+) using the fact that $\epsilon_{s,y}$ has got an s-fold zero at b.

For the conclusion (c) $\Rightarrow$ (a), choose $a < x_0 < x_1 < \cdots < x_{s-1} < b$ and construct the functional G_{s-1} from Lemma 3.4. Then

$$G_{s-1}[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s)}(x) d\phi_{s,y}(x)$$

and

$$H[f] := G[f] - G_{s-1}[f] = \int_a^b f^{(s)}(x) d\psi_{s,y}(x),$$

where $\psi_{s,y} = \mu_{s,y} - \phi_{s,y}$. Since G_{s-1} is a restricted estimation functional, we have that

$$G_{s-1}[f] = \int_a^b f(x) d\phi_{0,y}(x),$$

where $\phi_{0,y}$ is a step function with a finite partition. Repeated application of Theorem 1.3 now yields that $\phi_{s,y}$ is continuous and a piecewise polynomial. Therefore, $\phi_{s,y}$ fulfils a Lipschitz condition. Thus, conditions (I_+) and (II_+) are equivalent to

($\mathrm{I'}_+$) $\psi_{s,y}$ does not have jumps in negative direction

and

(II'_+) $\operatorname{ess\,inf}_{x\in[a,b]} \psi_{s,y}^{[\mathrm{ac}]\prime}(x) > -\infty$,

respectively. It is another consequence of Theorem 1.3 that $\phi_{s,y}(x) = 0$ for $x \in [a, \min(x_0, y)]$ and $\phi_{s,y}(x) =$const for $x \in [\max(x_{s-1}, y), b]$. Thus, for $x \in [a, \min(x_0, y)] \cup [\max(x_{s-1}, y), b]$, there holds $\phi'_{s,y}(x) = 0$. Therefore, (III_+) is equivalent to

(III'_+) $\liminf_{x\to a} (x-a)^{1-s} \psi_{s,y}^{[\mathrm{ac}]\prime}(x) > -\infty$,

and (IV_+) is equivalent to

(IV'_+) $\liminf_{x\to b}(b-x)^{1-s}\psi_{s,y}^{[\mathrm{ac}]'}(x) > -\infty$.

Now, according to Lemma 3.4, the statement (a) will be proved if we show that H admits restricted s-positive definite estimation.

Consider first the case that $M := \operatorname{ess\,inf}_{x\in[a,b]} \psi_{s,y}^{[\mathrm{ac}]'}(x) \geqq 0$. Then $\psi_{s,y}$ is nondecreasing. Therefore, $H[f] \geqq 0$ if $f^{(s)} \geqq 0$. This yields that $E \equiv 0$ is a restricted s-positive definite estimation functional for H.

In the case $M < 0$, we construct the required estimation functional in the following way. We have

$$H[f] = \int_a^b f^{(s)}(x)\,d\psi_{s,y}^{[\mathrm{s}]}(x) + \int_a^b f^{(s)}(x)\psi_{s,y}^{[\mathrm{ac}]'}(x)\,dx.$$

Now, denote the divided difference of the function f with nodes $t_1, t_2, \ldots, t_k$ by $[t_1, t_2, \ldots, t_k]f$, and define $B_{s-1}(x) := [a, x_1, x_2, \ldots, x_{s-1}, b](\cdot - x)_+^{s-1}/(s-1)!$ to be the s-th Peano kernel of the divided difference. Then, following Schumaker [10, § 4.3], B_{s-1} is the basic spline of degree $s-1$. This spline has got $(s-1)$-fold zeros at a and b, and it is positive throughout the open interval (a, b). Therefore, by conditions (II'_+), (III'_+) and (IV'_+), there exists a constant α such that $\psi_{s,y}^{[\mathrm{ac}]'}(x) - \alpha B_{s-1}(x) \geqq 0$ for every $x \in [a, b]$. Now, since B_{s-1} is the Peano kernel of the divided difference, we have that for every $f \in C^s[a, b]$, there holds

$$[a, x_1, x_2, \ldots, x_{s-1}, b]f = \int_a^b f^{(s)}(x)B_{s-1}(x)\,dx.$$

Therefore, assuming $f^{(s)} \geqq 0$,

$$\begin{aligned} &H[f] - \alpha[a, x_1, x_2, \ldots, x_{s-1}, b]f \\ &= \int_a^b f^{(s)}(x)\,d\psi_{s,y}^{[\mathrm{s}]}(x) + \int_a^b f^{(s)}(x)\left(\psi_{s,y}^{[\mathrm{ac}]'}(x) - \alpha B_{s-1}(x)\right)dx \geqq 0. \end{aligned}$$

Since a divided difference is a restricted estimation functional, we have now found a restricted s-positive definite estimation functional for H completing the proof.

To see that (c) $\Leftrightarrow$ (d), we just have to note that, by Theorem 1.2,

$$\mu_{s,y}(x) - \mu_{s,a}(x) = \begin{cases} G[(\cdot - y)^s - (\cdot - a)^s]/s! & \text{if } x \geqq y, \\ G[(\cdot - x)^s - (\cdot - a)^s]/s! & \text{if } x < y. \end{cases} \tag{5}$$

Therefore, $\mu_{s,y} - \mu_{s,a}$ is a continuous piecewise polynomial, and hence Lipschitz continuous. Furthermore,

$$\mu_{s,y}^{[\mathrm{ac}]'}(x) - \mu_{s,a}^{[\mathrm{ac}]'}(x) = \begin{cases} 0 & \text{if } x > y, \\ -\sum_{j=0}^{s-1} c_{j,a}(a-x)^{s-1-j}/(s-1-j)! & \text{if } x < y. \end{cases}$$

From these relations, the equivalence can easily be seen.

Finally, for the proof of (c) $\Leftrightarrow$ (e), we proceed in a similar manner using the fact that

$$\mu_{s,y}(x) - \mu_{s,b}(x) = \begin{cases} G[(\cdot - y)^s - (\cdot - x)^s]/s! & \text{if } x \geqq y, \\ 0 & \text{if } x < y. \end{cases} \tag{6}$$

□

Now, we turn our attention once again towards the functionals admitting both restricted s-positive definite estimation and restricted s-negative definite estimation. As a consequence of the previous theorems, we obtain the following characterizations.

COROLLARY 3.5. (1) *Let G be a continuous linear functional on $C^0[a,b]$. Then G admits restricted 0-positive definite estimation and restricted 0-negative definite estimation if and only if there exist $n \in \mathbb{N}$ and numbers $x_1, x_2, \ldots, x_n \in [a,b]$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$ such that*

$$G[f] = \sum_{j=1}^{n} a_j f(x_j).$$

(2) *Let G be a continuous linear functional on $C^1[a,b]$. Then G admits restricted 1-positive definite estimation and restricted 1-negative definite estimation if and only if there exists $y \in [a,b]$ such that $\mu_{1,y}$ fulfils a Lipschitz condition. This is the case if and only if $\mu_{1,y}$ fulfils a Lipschitz condition for every $y \in [a,b]$.*

PROOF. The first part is an almost trivial consequence of Theorem 3.1 in connection with Theorem 2.4; the second part is an immediate consequence of Theorem 3.2. □

COROLLARY 3.6. *Let $s \geqq 2$, and let G be a continuous linear functional on $C^s[a,b]$. Then the following statements are equivalent:*

(a) *G admits restricted s-positive definite estimation and restricted s-negative definite estimation.*

(b) *For every $y \in (a,b)$, $\mu_{s,y}$ fulfils a Lipschitz condition and $\mu_{s,y}^{[\mathrm{ac}](j)}(a) = \mu_{s,y}^{[\mathrm{ac}](j)}(b) = 0$ for $j = 1, 2, \ldots, s-1$ and $\mu_{s,y}^{[\mathrm{ac}](s)}$ is bounded in a neighbourhood of a and in a neighbourhood of b.*

(c) *There exists $y \in (a,b)$ such that $\mu_{s,y}$ fulfils the conditions of* (b).

(d) $\mu_{s,a}$ *fulfils a Lipschitz condition and* $\mu_{s,a}^{[\mathrm{ac}](j)}(a) = (-1)^{j+1} c_{s-j,a}$ *and* $\mu_{s,a}^{[\mathrm{ac}](j)}(b) = 0$ *for* $j = 1, 2, \ldots, s-1$ *and* $\mu_{s,a}^{[\mathrm{ac}](s)}$ *is bounded in a neighbourhood of* a *and in a neighbourhood of* b.

(e) $\mu_{s,b}$ *fulfils a Lipschitz condition and* $\mu_{s,b}^{[\mathrm{ac}](j)}(a) = 0$ *and* $\mu_{s,b}^{[\mathrm{ac}](j)}(b) = -c_{s-j,b}$ *for* $j = 1, 2, \ldots, s-1$ *and* $\mu_{s,b}^{[\mathrm{ac}](s)}$ *is bounded in a neighbourhood of* a *and in a neighbourhood of* b.

PROOF. The implication (b)⇒(c) is obvious. To prove that (c)⇔(d) and (c)⇔(e), we proceed as in the respective parts of the proof of Theorem 3.3, using equations (5) and (6).

For the conclusion (a) ⇒ (b), we note that by conditions (I_+) and (II_+) of Theorem 3.3 and their counterparts, the function $\mu_{s,y}$ is continuous (and hence absolutely continuous). and that $\mu'_{s,y} = \mu_{s,y}^{[\mathrm{ac}]'}$ is bounded from both sides. Thus, $\mu_{s,y}$ fulfils a Lipschitz condition. Furthermore, from conditions (III_+) and its counterpart, we can see, for $j < s$,

$$\lim_{x\to a}(x-a)^{1-j}\mu'_{s,y}(x) = \lim_{x\to a}(x-a)^{s-j}\left[(x-a)^{1-s}\mu'_{s,y}(x)\right] = 0.$$

For $j = 1$, we obtain $\mu'_{s,y}(a) = 0$. By induction, using Taylor's formula, we obtain that $\mu_{s,y}^{[\mathrm{ac}](j)}(a) = 0$ for $j = 1, 2, \ldots, s-1$ and that $\mu_{s,y}^{[\mathrm{ac}](s)}$ is bounded in a neighbourhood of a. In the same way, using conditions (IV_+) and its counterpart, we prove the results on the behaviour of the derivatives of $\mu_{s,y}$ at b.

For the proof of (c) ⇒ (a), we recall that $\mu_{s,y}$ fulfils a Lipschitz condition. Hence it cannot have any jumps and its derivative is bounded. Consequently, conditions (I_+) and (II_+) of Theorem 3.3 and their counterparts hold. The condition on the derivatives ensures in particular that the derivatives exist. Thus,

$$(x-a)^{1-s}\mu'_{s,y}(x) = \sum_{j=0}^{s-2} \frac{\mu_{s,y}^{[\mathrm{ac}](j+1)}(a)}{j!}(x-a)^{j+1-s} + \frac{\mu_{s,y}^{[\mathrm{ac}](s)}(\xi)}{(s-1)!},$$

where, by assumption, the sum vanishes, and therefore the expression on the left-hand side remains bounded for $x \to a$ proving (III_+) and its counterpart. Conditions (IV_+) and its counterpart can be shown in a similar way. Thus, by Theorem 3.3, G admits restricted s-positive definite estimation and restricted s-negative definite estimation. □

REMARK. With respect to the construction of the restricted estimation functionals using B-Splines (i.e. Peano kernels of divided differences), we remark that, for the special case $G[f] = \int_a^b f(x)dx$, a similar process has been described by Ehrich and Förster [5].

We can now deduce some sufficient conditions for functionals to admit restricted definite estimation.

THEOREM 3.7. *Let G be a continuous linear functional on $C^s[a,b]$, and let $r > s$. If G admits restricted s-positive definite estimation and restricted s-negative definite estimation, then G admits restricted r-positive definite estimation and restricted r-negative definite estimation.*

PROOF. It is sufficient to give a proof for $r = s+1$. We investigate the canonical representations

$$G[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s)}(x) d\mu_{s,y}(x)$$

$$= \sum_{j=0}^{s} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s+1)}(x) d\mu_{s+1,y}(x),$$

where $y \in (a,b)$ is arbitrary but fixed. By Theorem 1.3, $\mu_{s+1,y}$ is Lipschitz continuous. On the interval $[a,y)$, there holds $\mu'_{s+1,y} = -\mu_{s,y}$. Thus, the boundary conditions which are required for $\mu_{s+1,y}$ at $x = a$ by Corollaries 3.5 and 3.6 are fulfilled because of the respective conditions fulfilled by $\mu_{s,y}$. Similarly, the conditions for $\mu_{s+1,y}$ at $x = b$ are also fulfilled. □

COROLLARY 3.8. *Let G be a continuous linear functional on $C^0[a,b]$, and let $s \geqq 1$. Then G admits restricted s-positive definite estimation and restricted s-negative definite estimation.*

PROOF. Choose an arbitrary $y \in (a,b)$, and consider the canonical representations

$$G[f] = \int_a^b f(x) d\mu_{0,y}(x) = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s)}(x) d\mu_{s,y}(x).$$

Since $s > 0$, $\mu_{s,y}$ is Lipschitz continuous by Theorem 1.3. Thus, Part 2 of Corollary 3.5 implies the desired result for $s = 1$. Now, the proof for arbitrary $s \geqq 1$ follows immediately from Theorem 3.7. □

Our next result deals with the case of a functional G that does not admit s-positive definite estimation or s-negative definite estimation. It states that, under certain assumptions, we can still get an inclusion for $G[f]$ in the sense of equation (2) using restricted estimation functionals if information on f outside the interval $[a,b]$ is available.

THEOREM 3.9. *Let $r > s \geqq 0$, let G be a continuous linear functional on $C^s[a,b]$, and let $a^* < a < b < b^*$. Then G is a continuous linear functional*

on $C^s[a^*, b^*]$, and, as such, it admits restricted r-positive definite estimation and restricted r-negative definite estimation. If additionally there exists $y \in [a, b]$ such that $\mu_{s,y}$ fulfils a Lipschitz condition on $[a, b]$, then G (interpreted as a continuous linear functional on $C^s[a^*, b^*]$) admits restricted s-positive definite estimation and restricted s-negative definite estimation.

REMARK. This result can be interpreted as an analogon to the case $G_2 \equiv 0$ of Corollary 2.6. Indeed, we may say that the latter result is the limit case $a^* \to a$, $b^* \to b$ of Theorem 3.9.

PROOF. The fact that G is a continuous linear functional on $C^s[a^*, b^*]$ is obvious.

Now, let G (on $C^s[a, b]$) have the canonical representation

$$G[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_a^b f^{(s)}(x) d\mu_{s,y}(x).$$

Then on $C^s[a^*, b^*]$, G has got the canonical representation

$$G[f] = \sum_{j=0}^{s-1} c_{j,y} f^{(j)}(y) + \int_{a^*}^{b^*} f^{(s)}(x) d\mu^*_{s,y}(x),$$

where

$$\mu^*_{s,y}(x) = \begin{cases} 0 & \text{for } a^* \leqq x < a, \\ \lim\limits_{z \to a+} \mu_{s,y}(z) & \text{for } x = a, \\ \mu_{s,y}(x) & \text{for } a < x \leqq b, \\ \mu_{s,y}(b) & \text{for } b < x \leqq b^*. \end{cases}$$

If $\mu_{s,y}$ fulfils a Lipschitz condition on $[a, b]$, then $\mu^*_{s,y}$ also fulfils a Lipschitz condition on $[a^*, b^*]$. Furthermore, ${\mu^*_{s,y}}^{(j)}(x) = 0$ for $x \in [a^*, a) \cup (b, b^*]$ and $j = 1, 2, \ldots, s$. Thus, by Corollaries 3.5 and 3.6, G admits restricted s-positive definite estimation and restricted s-negative definite estimation.

If $r > s$, we can see G as a continuous linear functional on $C^r[a^*, b^*]$ with the canonical representation

$$G[f] = \sum_{j=0}^{r-1} c_{j,y} f^{(j)}(y) + \int_{a^*}^{b^*} f^{(r)}(x) d\mu^*_{r,y}(x).$$

The usual reasoning with the help of Theorem 1.3 yields that $\mu^*_{r,y}$ is Lipschitz continuous. The fact that the boundary conditions of Corollaries 3.5 and 3.6 are fulfilled follows in the same way as above for $r = s$. Thus, G admits restricted r-positive definite estimation and restricted r-negative definite estimation. □

4. Examples

In this section, we apply the results of the previous sections to some of the most important examples of continuous linear functionals occurring in practice.

4.1. Derivatives. As a first example, we consider the case that we want to give an estimation for the functional

$$G[f] := f^{(s)}(\xi),$$

where $s \geqq 1$ and $\xi \in [a, b]$. Since G is already an estimation functional, it is clear that it admits r-positive definite estimation and r-negative definite estimation for every $r \geqq s$.

However, the question of the existence of restricted estimation functionals for G has got a more complex answer. The canonical representation of G (as a continuous linear functional on $C^s[a, b]$) is given by

$$G[f] = \int_a^b f^{(s)}(x) d\mu_{s,y}(x),$$

$$\mu_{s,y}(x) = \begin{cases} 0 & \text{if } x < \xi, \\ 1 & \text{if } x \geqq \xi, \end{cases} \text{ if } \xi > a \quad \text{and} \quad \mu_{s,y}(x) = \begin{cases} 0 & \text{if } x = \xi, \\ 1 & \text{if } x > \xi, \end{cases} \text{ if } \xi = a.$$

Thus $\mu_{s,y}^{[\mathrm{ac}]} \equiv 0$, and $\mu_{s,y}^{[\mathrm{s}]}$ has got a step in positive direction. Therefore, from Theorems 3.2 and 3.3, we obtain that G admits restricted s-positive definite estimation. An application of these theorems to the functional $-G$ shows that G does not admit restricted s-negative definite estimation.

For every $r > s$ and every $\xi \in (a, b)$, G admits restricted r-positive definite estimation and restricted r-negative definite estimation by Corollary 3.6 since in this case, $\mu_{r,\xi}(x) = 0$ in the canonical representation

$$G[f] = f^{(s)}(\xi) + \int_a^b f^{(r)}(x) d\mu_{r,\xi}(x).$$

If, however, $\xi = a$, then conditions (I_+), (II_+) and (IV_+) of Theorem 3.3 are fulfilled, but the limit condition (III_+^*) is fulfilled if and only if $r - s$ is even. Therefore, G admits restricted r-positive definite estimation if and only if $r - s$ is even. Analogously, we find that G admits restricted r-negative definite estimation if and only if $r - s$ is odd. For $\xi = b$, we see in a similar way that G admits restricted r-positive definite estimation for every $r > s$, and G never admits restricted r-negative definite estimation.

Let us give an interpretation of this situation in other words. Assume we can calculate function values of f, but no derivatives. Moreover, assume

we know that f is r times continuously differentiable, and that $f^{(r)} \geqq 0$. If we want to find an inclusion (in the sense of equation (2)) for $f^{(s)}(\xi)$ (where $r > s$), we can only succeed if information about f is available on both sides of ξ. If we have information on one side of ξ only (i.e. if $\xi = a$ or $\xi = b$), then one of the required restricted estimation functionals does not exist.

4.2. Weighted integrals. The classical problem in numerical integration is the estimation of the functional

$$G[f] := \int_a^b w(x) f(x) dx,$$

where the *weight function* $w \in L_1[a, b]$ is assumed to be fixed. Here, we set $W(x) := \int_a^x w(t) dt$ and see that the canonical representation of G is given by

$$G[f] = \int_a^b f(x) dW(x).$$

Obviously, W is absolutely continuous. Thus, G admits (restricted) 0-positive definite estimation if and only if W is nondecreasing. This is equivalent to $w(x) \geqq 0$ a.e.. Similarly we see that G admits (restricted) 0-negative definite estimation if and only if $w(x) \leqq 0$ a.e.. Hence G admits (restricted) 0-positive definite estimation and 0-negative definite estimation if and only if $w(x) = 0$ a.e., which is equivalent to $G \equiv 0$.

For every $s \geqq 1$, however, by Corollary 3.8, G admits restricted s-positive definite estimation and restricted s-negative definite estimation (and hence also unrestricted s-positive definite estimation and s-negative definite estimation) without any additional conditions.

4.3. Singular integrals. For the approximation of the Cauchy principal value integral

$$G[f] := \int_{-1}^{1} \frac{f(x)}{x - \lambda} dx$$

with $\lambda \in (-1, 1)$, the problem of finding definite estimation functionals has been considered in [3]. There, it has been stated that G admits restricted s-positive definite estimation and restricted s-negative definite estimation for every $s \geqq 2$. Moreover, G admits restricted 1-positive definite estimation but not restricted 1-negative definite estimation. The reason can be found by a look at the canonical representation of G:

$$G[f] = f(\lambda) \ln \frac{1 - \lambda}{1 + \lambda} + \int_{-1}^{1} f'(x) d\mu_{1,\lambda}(x),$$

$$\mu_{1,\lambda}(x) = 1 + x + (\lambda - x) \ln \frac{|\lambda - x|}{1 + \lambda \operatorname{sgn}(\lambda - x)}.$$

Here $\mu_{1,\lambda}$ is continuous and strictly increasing, but its derivative is not bounded from above. Hence $\mu_{1,\lambda}$ does not fulfil a Lipschitz condition, and by Corollary 2.3, G does not admit 1-negative definite estimation.

Similar results hold for integrals with singularities of higher order (interpreted in the finite-part sense).

5. Concluding remarks

We have discussed the problem of finding an inclusion for the value $G[f]$ in the sense of equation (2) under the assumption that some derivative of f has no change of sign. For this purpose it is most convenient to use restricted estimation functionals. In § 3, we have given criteria that can be used to show whether this is possible or not. Owing to the constructive nature of the proofs, we can really find the required functionals if they exist. If they do not exist, different things may happen: If the reason for the nonexistence is a problem with the boundary conditions of Corollary 3.6, we may use information on f from a larger interval (if such information exists and is available), cf. Theorem 3.9. In case of other problems like jumps in the "wrong" direction, or if such additional information is not available, we may still find an inclusion using general estimation functionals instead of their restricted relatives. In this case, conditions for the existence of such inclusions and a method for the construction are given in § 2. But, as the example of the Cauchy principal value integral in § 4.3 shows, there exist functionals which simply do not admit an inclusion of this type for some s.

The methods and results can also be used if only a one-sided estimation is sought and not an inclusion.

REFERENCES

[1] Brass, H., *Quadraturverfahren*, Studia Mathematica, Skript 3. Vandenhoeck & Ruprecht, Göttingen, 1977. *MR* **56** #1675

[2] Brass, H. and Schmeisser, G., Error estimates for interpolatory quadrature formulae, *Numer. Math.* **37** (1981), 371–386. *MR* **82j**:65012

[3] Diethelm, K., Definite quadrature formulae for Cauchy principal value integrals, *Approximation theory and function series* (Budapest, 1995), Bolyai Soc. Math. Stud. **5** (1996), 175–186. *MR* **97m**:41031

[4] Diethelm, K., A definiteness criterion for linear functionals and its application to Cauchy principal value quadrature, *J. Comput. Appl. Math.* **66** (1996), 167–176. *MR* **97c**:41033

[5] Ehrich, S. and Förster, K.-J., On exit criteria in quadrature using Peano kernel inclusions I: Introduction and basic results, *Z. Angew. Math. Mech.* **75** (1995), S625–S626.

[6] Förster, K.-J., A comparison theorem for linear functionals and its application in quadrature, *Numerical Integration* (Oberwolfach, 1981), ed. by G. Hämmerlin,

International Series of Numerical Mathematics, Vol. 57, Birkhäuser-Verlag, Basel, 1982, 66–76. *MR* **83k**:65003

[7] Förster, K.-J., Exit criteria and monotonicity in compound quadrature of Gaussian type, *Numer. Math.* **66** (1993), 321–327. *MR* **94i**:65030

[8] Köhler, P., A note on definiteness and monotonicity of quadrature formulae, *Z. Angew. Math. Mech.* **75** (1995), S645–S646.

[9] Sard, A., *Linear approximation*, American Mathematical Society, Providence, RI, 1963. *MR* **28** #1429. 2nd printing with corrections, 1982.

[10] Schumaker, L. L., *Spline functions: basic theory*, Pure and Applied Mathematics, J. Wiley & Sons, New York, 1981. *MR* **82j**:41001

(Received October 4, 1996)

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT HILDESHEIM
MARIENBURGER PLATZ 22
D–31141 HILDESHEIM
GERMANY

Present address:

INSTITUT FÜR ANGEWANDTE MATHEMATIK
TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG
POCKELSSTRASSE 14
D–38106 BRAUNSCHWEIG
GERMANY

k.diethelm@tu-bs.de

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 237–246

POINTWISE ESTIMATES FOR BERNSTEIN-TYPE OPERATORS

S. GUO and Q. QI

Abstract

For Bernstein-type operators and their combinations, Ditzian and Ivanov gave some equivalent relations. In this paper we extend these results in the pointwise case using the modulus of smoothness $\omega^{2r}_{\varphi^\lambda}(f,t)$ $(0 \leqq \lambda \leqq 1)$.

1. Introduction

The Bernstein-type integral operators discussed in this paper are given by

$$M_n f = M_n(f,x) = (n+1)\sum_{k=0}^{n} p_{n,k}(x) \int_0^1 p_{n,k}(t) f(t)\,dt, \tag{1.1}$$

where $p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$. Ditzian and Ivanov [1] constructed an operator $O_n f$ using linear combinations of $M_n f$ as follows:

$$O_n(f,x) = \sum_{i=0}^{2r-1} \alpha_i(n) M_{n_i}(f,x), \quad n_0 = n < n_1 < \cdots < n_{2r-1} \leqq An, \tag{1.2}$$

where A is independent of n. The operators $O_n f$ satisfy [1]

$$O_n(1,x) = 1, \quad O_n((\cdot - x)^m, x) = 0, \quad \text{for } m = 1, \ldots, 2r-1, \tag{1.3}$$

$$\sum_{i=0}^{2r-1} |\alpha_i(n)| \leqq B. \tag{1.4}$$

1991 *Mathematics Subject Classification.* Primary 41A25, 41A36.

Key words and phrases. Operators, linear combinations, moduli of smoothness, K-functionals.

Ditzian and Ivanov [1] showed for $r>\alpha$, $\varphi(x)=\sqrt{x(1-x)}$, $1\leqq p\leqq\infty$ that

$$\begin{aligned}(1.5)\qquad \|O_nf-f\|_p=O(n^{-\alpha}) &\Longleftrightarrow \omega_\varphi^{2r}(f,t)_p=O(t^{2\alpha})\\ &\Longleftrightarrow \|\varphi^{2r}M_n^{(2r)}f\|_p=O(n^{r-\alpha}).\end{aligned}$$

Recently, for Bernstein polynomials $B_n(f,x)=\sum_{k=0}^{n} f\left(\frac{k}{n}\right)p_{n,k}(x)$, Ditzian [2] gave an interesting pointwise estimate

$$(1.6)\qquad |B_n(f;x)-f(x)|\leqq C\omega_{\varphi^\lambda}^2\left(f,n^{-\frac{1}{2}}\varphi^{1-\lambda}(x)\right)\qquad (0\leqq\lambda\leqq 1).$$

However, [2] did not give the inverse estimate. Therefore we gave an inverse result for $0<\alpha<2$, published in [3],

$$(1.7)\qquad B_n(f,x)-f(x)=O\left((n^{-\frac{1}{2}}\varphi^{1-\lambda}(x))^\alpha\right)\Longleftrightarrow\omega_{\varphi^\lambda}^2(f,t)=O(t^\alpha).$$

In this paper, using $\omega_{\varphi^\lambda}^{2r}(f,t)$, we extend the results (1.5) for the case $p=\infty$ and show for $r>\alpha$, $\delta_n(x)=\varphi(x)+\frac{1}{\sqrt{n}}$, $0\leqq\lambda\leqq 1$ that

$$\begin{aligned}(1.8)\qquad O_n(f,x)-f(x)=O(\gamma_{n,\lambda}^{2\alpha}(x))&\Longleftrightarrow\omega_{\varphi^\lambda}^{2r}(f,t)=O(t^{2\alpha})\\ &\Longleftrightarrow\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)=O(\gamma_{n,\lambda}^{2(\alpha-r)}),\end{aligned}$$

where $\gamma_{n,\lambda}(x)=n^{-\frac{1}{2}}\delta_n^{1-\lambda}(x)$, $f\in C[0,1]$ and $\omega_{\varphi^\lambda}^{2r}(f,t)$ is defined by (1.9). Here we give some definitions (cf. [1], [2], [4]):

$$(1.9)\qquad \omega_{\varphi^\lambda}^{2r}(f,t)=\sup_{0<h\leqq t}\sup_{x\pm rh\varphi^\lambda(x)\in[0,1]}|\Delta_{h\varphi^\lambda(x)}^{2r}f(x)|,$$

$$(1.10)\qquad K_{\varphi^\lambda}(f,t^{2r})=\inf_{g^{(2r-1)}\in A.C_{loc}}\left(\|f-g\|_{C[0,1]}+t^{2r}\|\varphi^{2r\lambda}g^{(2r)}\|_{C[0,1]}\right),$$

$$(1.11)\qquad \overline{K}_{\varphi^\lambda}(f,t^r)=\inf_{g^{(2r-1)}\in A.C_{loc}}\left(\|f-g\|+t^{2r}\|\varphi^{2r\lambda}g^{(2r)}\|+t^{\frac{2r}{1-\frac{\lambda}{2}}}\|g^{(2r)}\|\right).$$

They are equivalent (cf. [4], Theorems 2.1.1 and 3.1.2). We write

$$(1.12)\qquad \omega_{\varphi^\lambda}^{2r}(f,t)\sim K_{\varphi^\lambda}(f,t^{2r})\sim\overline{K}_{\varphi^\lambda}(f,t^{2r}).$$

Throughout this paper C denotes a positive constant independent of n and x and not necessarily the same at each occurrence.

2. A direct theorem

In this section we show a direct estimate for $O_n f$.

THEOREM 1. *Let $f \in C[0,1]$, $r \in N$, then we have*

$$|O_n(f,x) - f(x)| \leqq C\omega^{2r}_{\varphi^\lambda}(f, \gamma_{n,\lambda}(x)). \tag{2.1}$$

PROOF. From (1.11) we may choose $g_n = g_{n,x,\lambda}$ for a fixed x and λ such that

$$\|f - g_n\| + \gamma^{2r}_{n,\lambda}(x)\|\varphi^{2r\lambda} g_n^{(2r)}\| + \gamma_{n,\lambda}^{\frac{2r}{1-\frac{\lambda}{2}}} \|g_n^{(2r)}\| \leqq C\omega^{2r}_{\varphi^\lambda}(f, \gamma_{n,\lambda}(x)). \tag{2.2}$$

We recall that [1]

$$O_n((\cdot - x)^k, x) = 0, \qquad k = 1, 2, \ldots, 2r-1. \tag{2.3}$$

For u between t and x we have (cf. [1] Lemma 5.3)

$$\frac{|t-u|^{2r-1}}{\varphi^{2r\lambda}(u)} \leqq \frac{|t-x|^{2r-1}}{\varphi^{2r\lambda}(x)}, \tag{2.4}$$

and

$$\frac{|t-u|^{2r-1}}{\delta_n^{2r\lambda}(u)} \leqq C\frac{|t-x|^{2r-1}}{\delta_n^{2r\lambda}(x)}. \tag{2.5}$$

Then, by (2.3) and (2.4), using [1] (5.4) we have

$$\begin{aligned} |O_n(g_n,x) - g_n(x)| &\leqq \left| O_n\left(\frac{1}{(2r-1)!} \int_x^t (t-u)^{2r-1} g_n^{(2r)}(u)du, x \right) \right| \\ &\leqq \sum_{i=0}^{2r-1} |\alpha_i(n)| M_{n_i}(|t-x|^{2r}, x)\|\varphi^{2r\lambda} g_n^{(2r)}\|\varphi^{-2r\lambda}(x) \\ &\leqq Cn^{-r}\delta_n^{2r}(x)\|\varphi^{2r\lambda} g_n^{(2r)}\|\varphi^{-2r\lambda}(x). \end{aligned} \tag{2.6}$$

Similarly, by (2.3) and (2.5), we have

$$\begin{aligned} |O_n(g_n,x) - g_n(x)| &\leqq Cn^{-r}\delta_n^{2r}(x)\|\delta_n^{2r\lambda} g_n^{(2r)}\|\delta_n^{-2r\lambda}(x) \\ &\leqq Cn^{-r}\delta_n^{2r(1-\lambda)}(x)\left(\|\varphi^{2r\lambda} g_n^{(2r)}\| + n^{-r\lambda}\|g^{(2r)}\| \right). \end{aligned} \tag{2.7}$$

Thus for $f \in C[0,1]$, $x \in E_n = \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$, when $\delta_n(x) \sim \varphi(x)$, by (2.2) and (2.6) we may deduce that

$$\begin{aligned} |O_n(f,x) - f(x)| &\leqq C\left(\|f - g_n\| + n^{-r}\delta_n^{2r(1-\lambda)}(x)\|\varphi^{2r\lambda} g_n^{(2r)}\| \right) \\ &\leqq C\omega^{2r}_{\varphi^\lambda}(f, \gamma_{n,\lambda}(x)). \end{aligned} \tag{2.8}$$

For $x \in E_n^c = \left[0, \frac{1}{n}\right) \cup \left(1 - \frac{1}{n}, 1\right]$, $\delta_n(x) \sim \frac{1}{\sqrt{n}}$, by (2.2) and (2.7) we have

$$\begin{aligned} |O_n(f,x) - f(x)| &\leqq C\Big(\|f - g_n\| + n^{-r}\delta_n^{2r(1-\lambda)}(x)\|\varphi^{2r\lambda}g^{(2r)}\| \\ &\quad + (n^{-\frac{1}{2}}\delta_n^{1-\lambda}(x))^{\frac{2r}{1-\frac{\lambda}{2}}}\|g^{(2r)}\|\Big) \\ &\leqq C\omega_{\varphi^\lambda}^{2r}(f, \gamma_{n,\lambda}(x)). \end{aligned} \tag{2.9}$$

From (2.8) and (2.9) we get (2.1).

3. A connection between the derivatives and the smoothness

In this section we will give an equivalence relation between the derivatives of $M_n f$ and the modulus of smoothness.

THEOREM 2. *Let $f \in C[0,1]$, $r \in N$, $0 \leqq \lambda \leqq 1$, then*

$$|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)| \leqq C\gamma_{n,\lambda}^{-2r}(x)\omega_{\varphi^\lambda}^{2r}(f, \gamma_{n,\lambda}(x)). \tag{3.1}$$

PROOF. To prove (3.1), in view of (1.12) it is sufficient to get

$$|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)| \leqq C\gamma_{n,\lambda}^{-2r}(x)\|f\|_\infty, \tag{3.2}$$

and

$$|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)| \leqq C\|\varphi^{-2r\lambda}f^{(2r)}\|_\infty. \tag{3.3}$$

First we prove (3.2). We discuss the two cases separately.

By [5] (3.9) we have

$$\begin{aligned} |M_n^{(2r)}(f,x)| &\leqq n^{2r}\left|\sum_{k=0}^{n-2r} p_{n-2r,k}(x)\sum_{j=0}^{2r}(-1)^{2r-j}\binom{2r}{j}(n+1)\int_0^1 p_{n,k+j}(t)f(t)dt\right| \\ &\leqq 2^{2r}n^{2r}\|f\|. \end{aligned} \tag{3.4}$$

If $x \in \left[0, \frac{1}{n}\right) \cup \left(1 - \frac{1}{n}, 1\right]$, then $\delta_n \sim \frac{1}{\sqrt{n}}$ and by (3.4) we have

$$\begin{aligned} |\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)| &\leqq \varphi^{2r\lambda}(x)|M_n^{(2r)}(f,x)| \\ &\leqq Cn^{-r(\lambda-2)}\|f\| \leqq C\gamma_{n,\lambda}^{-2r}(x)\|f\|. \end{aligned} \tag{3.5}$$

If $x \in \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$, then $\delta_n \sim \varphi(x)$ and by [1] (4.3) we have

$$|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)| \leqq C\varphi^{2r(\lambda-1)}(x)n^r\|f\| \leqq C\gamma_{n,\lambda}^{-2r}(x)\|f\|. \tag{3.6}$$

From (3.5) and (3.6) we obtain (3.2).

Next we prove (3.3). By [1] (3.6) and Hölder inequality we have (cf. [1] P75)

$$\begin{aligned}
&\left|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)\right| \\
&\leqq \frac{(n+1)!n!}{(n-2r)!(n+2r)!} \\
&\times\left(\sum_{k=0}^{n-2r} p_{n-2r,k}(x)\varphi^{2r\lambda}(x)\int_0^1 p_{n+2r,k+2r}(t)\varphi^{-2r\lambda}(t)dt\right)\|\varphi^{2r\lambda}f^{(2r)}\| \\
&\leqq C\left((n+1)\sum_{k=0}^{n-2r} p_{n-2r,k}(x)\varphi^{2r}(x)\int_0^1 p_{n+2r,k+2r}(t)\varphi^{-2r}(t)dt\right)^{\lambda}\|\varphi^{2r\lambda}f^{(2r)}\| \\
&\leqq C\|\varphi^{2r\lambda}f^{(2r)}\|.
\end{aligned}$$

This is (3.3). The proof of (3.1) is complete.

The inverse result is given as follows:

THEOREM 3. *Let* $f\in C[0,1]$, $r\in N$, $0<\alpha\leqq r$, $0\leqq\lambda\leqq 1$, *then*

$$|\varphi^{2r\lambda}(x)M_n^{(2r)}(f,x)|\leqq C\gamma_{n,\lambda}^{2(\alpha-r)}(x) \tag{3.7}$$

implies

$$\omega_{\varphi^\lambda}^{2r}(f,h)=O(h^{2\alpha}). \tag{3.8}$$

PROOF. We will use the commutative property of M_n:

$$M_n(M_mf)(x)=M_m(M_nf)(x)\quad\text{for } m,n\in N.$$

Let $0<t\leqq h<\frac{1}{16r}$, $rt\varphi^\lambda(x)<x<1-rt\varphi^\lambda(x)$. We note the fact

$$\max_{-r\leqq j\leqq r}\left\{\delta_n(x+jt\varphi^\lambda(x))\right\}\leqq 2\delta_n(x). \tag{3.9}$$

In fact, by symmetry, we need only consider the case $rt\varphi^\lambda(x)\leqq x\leqq\frac{1}{2}$. Then

$$0\leqq x-rt\varphi^\lambda(x)<x<x+rt\varphi^\lambda(x)\leqq 2x,$$

therefore, if $2x\leqq\frac{1}{2}$, then for $k=1,2,\ldots,r$

$$\begin{aligned}
\varphi(x-kt\varphi^\lambda(x))&\leqq\varphi(x)\leqq\varphi(x+kt\varphi^\lambda(x)) \\
&\leqq\varphi(2x)\leqq 2\varphi(x).
\end{aligned}$$

If $x \leqq \frac{1}{2} \leqq 2x$, then

$$\varphi(x \pm kt\varphi^{\lambda}(x)) \leqq \varphi\left(\frac{1}{2}\right) \leqq 2\varphi\left(\frac{1}{4}\right) \leqq 2\varphi(x).$$

By Theorem 1 and (3.9), we have for $m \in N$

$$\begin{aligned}
&|\Delta^{2r}_{t\varphi^{\lambda}(x)} M_m f(x)| \\
&\leqq \left| \sum_{j=0}^{2r} (-1)^{2r-j} \binom{2r}{j} \Big\{ O_n(M_m f, x + (j-r)t\varphi^{\lambda}(x)) \right. \\
&\quad \left. - M_m(f, x + (j-r)t\varphi^{\lambda}(x)) \Big\} \right| \\
&\quad + \left| \sum_{i=0}^{2r-1} \alpha_i(n) \Delta^{2r}_{t\varphi^{\lambda}(x)} M_{n_i}(M_m f)(x) \right| \qquad (3.10) \\
&\leqq 2^{2r} C \omega^{2r}_{\varphi^{\lambda}}(M_m f, \gamma_{n,\lambda}(x)) \\
&\quad + \sum_{i=0}^{2r-1} |\alpha_i(n)| \int \cdots \int_{-\frac{i}{2}\varphi^{\lambda}(x)}^{\frac{i}{2}\varphi^{\lambda}(x)} \left| M_m^{(2r)}\left(M_{n_i} f, x + \sum_{j=1}^{2r} u_j \right) \right| du_1 \cdots du_{2r}.
\end{aligned}$$

From (3.7) we can deduce that by $\delta_n(x) \sim \max\left\{\varphi(x), \frac{1}{\sqrt{n}}\right\}$

$$|\varphi^{2r\lambda}(x) M_n^{(2r)}(f, x)| \leqq C n^{-(\alpha-r)(2-\lambda)}, \qquad (3.11)$$

$$|\varphi^{2r\lambda}(x) M_n^{(2r)}(f, x)| \leqq C n^{-(\alpha-r)} \varphi^{2(1-\lambda)(\alpha-r)}(x). \qquad (3.12)$$

Using (3.3), (3.11) and (3.12), we have

$$\begin{aligned}
\left| M_m^{(2r)}\left(M_{n_i} f, x + \sum_{j=1}^{2r} u_j \right) \right| &\leqq \varphi^{-2r\lambda}\left(x + \sum_{j=1}^{2r} u_j \right) \| \varphi^{2r\lambda} M_{n_i}^{(2r)}(f) \| \\
&\leqq C \varphi^{-2r\lambda}\left(x + \sum_{j=1}^{2r} u_j \right) n^{-(\alpha-r)(2-\lambda)}, \qquad (3.13)
\end{aligned}$$

and

$$\begin{aligned}
&\left|M_m^{(2r)}\left(M_{n_i}f, x+\sum_{j=1}^{2r}u_j\right)\right| \\
&\leqq \varphi^{-2r\lambda+2(1-\lambda)(\alpha-r)}\left(x+\sum_{j=1}^{2r}u_j\right)\|\varphi^{2r\lambda+2(1-\lambda)(r-\alpha)}M_{n_i}^{(2r)}(f)\| \qquad (3.14)\\
&\leqq C\varphi^{-2r+2\alpha(1-\lambda)}\left(x+\sum_{j=1}^{2r}u_j\right)n^{-(\alpha-r)}.
\end{aligned}$$

By [6] (2.6) with Hölder inequality we can easily get for $0 \leqq \beta \leqq 2r$

$$\int\cdots\int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \varphi^{-\beta}\left(x+\sum_{j=1}^{r}u_j\right)du_1\cdots du_{2r} \leqq C\varphi^{-\beta}(x)t^{2r}\varphi^{2r\lambda}(x). \qquad (3.15)$$

Combining (3.10), (3.13)–(3.15) we obtain

$$\begin{aligned}
&\left|\Delta_{t\varphi^\lambda(x)}^{2r}M_mf(x)\right| \leqq C\omega_{\varphi^\lambda}^{2r}(M_mf,\gamma_{n,\lambda}(x)) \\
&+\sum_{i=0}^{2r-1}|\alpha_i(n)|C\min\{n^{-(\alpha-r)(2-\lambda)}t^{2r}, n^{-(\alpha-r)}t^{2r}\varphi^{2(1-\lambda)(\alpha-r)}(x)\} \qquad (3.16)\\
&\leqq C\left\{\omega_{\varphi^\lambda}^{2r}(M_mf,\gamma_{n,\lambda}(x))+t^r\gamma_{n,\lambda}^{2(\alpha-r)}(x)\right\}.
\end{aligned}$$

The following demonstration is very similar to [7] (cf. [7] (5.5)), we omit the details. From (3.16) we can deduce (3.8). The proof is complete.

4. An inverse theorem on O_nf

In this section, we will give an inverse result of Theorem 1.

THEOREM 4. *Let $f \in C[0,1]$, $r \in N$, $0<\alpha<r$, $0 \leqq \lambda \leqq 1$. Then we have*

$$|O_n(f,x)-f(x)| \leqq C\gamma_{n\lambda}^{2\alpha}(x) \qquad (4.1)$$

with a constant C independent of n and x if and only if

$$\omega_{\varphi^\lambda}^{2r}(f,t)=O(t^{2\alpha}). \qquad (4.2)$$

PROOF. We need only to prove that (4.1) $\Longrightarrow$ (4.2). By (4.1) we have

$$|\Delta^{2r}_{t\varphi^\lambda(x)}| \leqq |\Delta^{2r}_{t\varphi^\lambda}(O_n f(x) - f(x))| + |\Delta^{2r}_{t\varphi^\lambda} O_n f(x)|$$

$$\leqq C\gamma^{2\alpha}_{n,\lambda}(x) + \sum_{i=0}^{2r-1} |\alpha_i(n)| \int \cdots \int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left| M^{(2r)}_{n_i}\left(f, x + \sum_{j=1}^{2r} u_j\right)\right| du_1 \cdots du_{2r}$$

(4.3)

$$\leqq C\gamma^{2\alpha}_{n,\lambda}(x) + \sum_{i=0}^{2r-1} |\alpha_i(n)| \int \cdots \int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left(\left| M^{(2r)}_{n_i}\left(f - g, x + \sum_{j=1}^{2r} u_j\right)\right| \right.$$

$$\left. + \left| M^{(2r)}_{n_i}\left(g, x + \sum_{j=1}^{2r} u_j\right)\right| \right) du_1 \cdots du_{2r}.$$

By [1] (4.3), we can deduce that

(4.4)
$$\begin{aligned} |M^{(2r)}_n(f,x)| &= \varphi^{-2r}(x)|\varphi^{2r}(x) M^{(2r)}_n(f,x)| \\ &\leqq C\varphi^{-2r}(x) n^r \|f\|. \end{aligned}$$

Using this relation and (3.15) we get

(4.5)
$$\begin{aligned} &\int \cdots \int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left| M^{(2r)}_{n_i}\left(f - g, x + \sum_{j=1}^{2r} u_j\right)\right| du_1 \cdots du_{2r} \\ &\leqq Cn^r \|f - g\| \int \cdots \int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \varphi^{-2r}\left(x + \sum_{j=1}^{2r} u_j\right) du_1 \cdots du_{2r} \\ &\leqq Cn^r t^{2r} \varphi^{-2r(1-\lambda)}(x) \|f - g\|. \end{aligned}$$

Using (3.4), we have

(4.6)
$$\begin{aligned} &\int \cdots \int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left| M^{(2r)}_{n_i}\left(f - g, x + \sum_{j=1}^{2r} u_j\right)\right| du_1 \cdots du_{2r} \\ &\leqq Cn^{2r} t^{2r} \varphi^{2r\lambda}(x) \|f - g\|. \end{aligned}$$

Combining (4.5) and (4.6) we have

$$
\begin{aligned}
&\int\cdots\int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left|M_{n_i}^{(2r)}\left(f-g, x+\sum_{j=1}^{2r}u_j\right)\right| du_1\cdots du_{2r} \\
&\leqq Cn^r t^{2r}\varphi^{2r\lambda}(x)\min\{n^r, \varphi^{-2r}(x)\}\|f-g\| \\
&\leqq Cn^r t^{2r}\delta^{2r\lambda}(x)\delta_n^{-2r}(x)\|f-g\| \\
&= Ct^{2r}\gamma_{n,\lambda}^{-2r}(x)\|f-g\|.
\end{aligned} \tag{4.7}
$$

On the other hand, by (3.3) and (3.15), we have

$$
\begin{aligned}
&\int\cdots\int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \left|M_{n_i}^{(2r)}\left(g, x+\sum_{j=1}^{2r}u_j\right)\right| du_1\cdots du_{2r} \\
&\leqq C\int\cdots\int_{-\frac{t}{2}\varphi^\lambda(x)}^{\frac{t}{2}\varphi^\lambda(x)} \varphi^{-2r\lambda}\left(x+\sum_{j=1}^{2r}u_j\right) du_1\cdots du_{2r}\|\varphi^{2r\lambda}g^{(2r)}\| \\
&\leqq Ct^{2r}\|\varphi^{2r\lambda}g^{(2r)}\|.
\end{aligned} \tag{4.8}
$$

From (4.3), (4.7) and (4.8) we obtain

$$
\omega_{\varphi^\lambda}^{2r}(f,t) \leqq C\left(\gamma_{n,\lambda}^{2\alpha}(x) + \frac{t^{2r}}{\gamma_{n,\lambda}^{2r}(x)}\omega_{\varphi^\lambda}^{2r}(f, \gamma_{n,\lambda}(x))\right) \tag{4.9}
$$

and this implies, via the Berens–Lorentz Lemma [8], that

$$
\omega_{\varphi^\lambda}^{2r}(f,t) \leqq Ct^{2\alpha},
$$

which is the desired result.

REMARK. Combining Theorems 1–4, we have proved the relation (1.8).

REFERENCES

[1] DITZIAN, Z. and IVANOV, K., Bernstein-type operators and their derivatives, *J. Approx. Theory* **56** (1989), 72–90. *MR* **90c**:41042

[2] DITZIAN, Z., Direct estimate for Bernstein polynomials, *J. Approx. Theory* **79** (1994), 165–166. *MR* **95h**:41018

[3] GUO, S., Inverse estimate for Bernstein polynomials (submitted).

[4] DITZIAN, Z. and TOTIK, V., *Moduli of smoothness*, Springer Series in Computational Mathematics, Vol. 9, Springer-Verlag, New York – Berlin, 1987. *MR* **89h**:41002

[5] HEILMANN, M., Direct and converse results for operators of Baskakov–Durrmeyer type, *Approx. Theory Appl.* **5** (1989), 105–127. *MR* **90k**:41025

[6] XIE, L., Uniform approximation by combinations of Bernstein polynomials, *Approx. Theory Appl.* **11** (1995), 36–51. *MR* **96m**:41006

[7] ZHOU, D., On a paper of Mazhar and Totik, *J. Approx. Theory* **72** (1993), 290–300. *MR* **94d**:41041

[8] BERENS, H. and LORENTZ, G., Inverse theorems for Bernstein polynomials, *Indiana Univ. Math. J.* **21** (1972), 693–708. *MR* **45** #5638

(Received October 30, 1996)

DEPARTMENT OF MATHEMATICS
HEBEI TEACHER'S UNIVERSITY
SHIJIAZHUANG 050016
PEOPLE'S REPUBLIC OF CHINA

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 247–258

ON FINITE AUTOMORPHISM GROUPS OF SIMPLE ARGUESIAN LATTICES

G. GRÄTZER and E. T. SCHMIDT

Abstract

Let $\mathfrak{G}$ be a finite group. In this paper we prove that there exists a *simple arguesian* lattice M whose automorphism group is isomorphic to $\mathfrak{G}$.

1. Introduction

G. Birkhoff [4] proved that every finite group can be represented as the automorphism group of a finite *distributive lattice*. R. Frucht [7] represented every finite group as the automorphism group of a finite *simple lattice* (of length three).

The only related result for modular lattices is due to E. Mendelsohn [12]: every group can be represented as the automorphism group of a *projective plane*. (See also L. Babai [1].) However, this projective plane cannot be coordinatized over a skewfield, or equivalently, it is not *arguesian* (does not satisfy the *arguesian identity*, see, e.g., [p. 199][8]). Indeed, the automorphism group of a projective plane over a field is very special; for instance, it must contain copies of the symmetric group on three elements.

The main result of this paper is the following

MAIN THEOREM. *Let $\mathfrak{G}$ be a finite group. Then there exists an interval finite, simple, arguesian lattice M such that the automorphism group of M is isomorphic to $\mathfrak{G}$.*

See Section 7 for a discussion of related problems.

NOTATION. For the basic concepts and notation, the reader is referred to [8]. $\mathfrak{M}_3$ is the five-element modular nondistributive lattice. Let A and B be lattices; let D be a dual ideal of A, let I be an ideal of B, and let φ be an isomorphism between D and I. We form the disjoint union of A and B

1991 *Mathematics Subject Classification.* Primary 06C05; Secondary 08A35.
Key words and phrases. Automorphism group, lattice, simple, modular, arguesian.

The research of the first author was supported by the NSERC of Canada.
The research of the second author was supported by the Hungarian National Foundation for Scientific Research, Grant No. T716432.

and identify each $d \in D$ with $d\varphi \in I$, obtaining the set C. We define $x \leqq y$ in C iff $x, y \in A$ and $x \leqq y$ in A, or $x, y \in B$ and $x \leqq y$ in B, or $x \in A$, $y \in B$ and there is a $z \in A \cap B \subseteq C$ with $x \leqq z$ in A and $z \leqq y$ in B. Then C is a lattice, which we call the *gluing* of A and B over D and I. Finally, a lattice L is called *interval finite*, if every interval of L is finite.

2. S-glued systems

The next two definitions and the theorem that follows are from Ch. Herrmann [11]; they are stated in a slightly more general form to facilitate their application in this paper.

DEFINITION 1. Let S be an interval finite lattice. The family of finite lattices

$$\mathfrak{L} = \{ L_s \mid s \in S \}$$

is an *S-glued system* iff the following conditions are satisfied for $s, t \in S$:

(1) If $s \leqq t$ and $L_s \cap L_t \neq \emptyset$, then $L_s \cap L_t$ is a dual ideal of L_s and an ideal of L_t.
(2) If $s \leqq t$ and $a, b \in L_s \cap L_t$, then the relation $a \leqq b$ holds in L_s iff it holds in L_t.
(3) If $s \prec t$, then $L_s \cap L_t \neq \emptyset$.
(4) $L_s \cap L_t \subseteq L_{s \wedge t} \cap L_{s \vee t}$.

DEFINITION 2. Let $\mathfrak{L} = \{ L_s \mid s \in S \}$ be an S-glued system. Let $L = \bigcup (L_s \mid s \in S)$ and define the partial order $\leqq$ on L as the transitive extension of the union of the partial orders on the L_s, $s \in S$. Then L is a lattice, the *S-glued sum* of $\mathfrak{L}$ and the L_s, $s \in S$, are the *blocks* of L.

THEOREM 1. *Let $\mathfrak{L} = \{ L_s \mid s \in S \}$ be an S-glued system, and let L be the S-glued sum of $\mathfrak{L}$. Then the following statements hold:*

(1) *A block of L is an interval.*
(2) *For $a, b \in L$, the relation $a \leqq b$ holds in L iff there exists a sequence $a = x_0, x_1, \ldots, x_n = b$ of elements of L and a sequence $s_1, \ldots, s_n$ of elements of S such that $s_i \leqq s_{i+1}$ in S for $i = 1, \ldots, n-1$, and $x_{i-1} \leqq x_i$ in L_{s_i} for $i = 1, \ldots, n$.*
(3) *If A and B are blocks indexed by comparable elements of S, then $A \cup B$ is a sublattice of L. If $A \cap B \neq \emptyset$, then this sublattice is a gluing of A and B over $A \cap B$.*
(4) *If $a \prec b$ in L, then $a \prec b$ in some block.*
(5) *If $s, t \in S$, $L_s = [a, b]$, and $L_t = [c, d]$, then $L_{s \vee t}$ is of the form $[a \vee c, e]$ for some $e \in L$, where $e \geqq b \vee d$.*
(6) *If each L_s, $s \in S$, is a complemented modular lattice, then L is a modular lattice, and the blocks can be recognized as maximal complemented intervals.*

(7) *Every modular lattice of finite length has a unique finest representation as an S-glued sum, namely, as the S-glued sum of its maximal complemented intervals.*

Now let $\mathfrak{L} = \{ L_s \mid s \in S \}$ be an S-glued system, and let L be the S-glued sum of $\mathfrak{L}$. Let $L_s = [a_s, b_s]$ in L. Let K be a lattice containing L as a sublattice, and define $\hat{L}_s = [a_s, b_s]_K$. It is easy to see that the $\hat{L}_s$, $s \in S$, also satisfies the conditions of Definition 1, so $\hat{\mathfrak{L}} = \{ \hat{L}_s \mid s \in S \}$ is an S-glued system. Therefore, we can form the S-glued sum of $\hat{\mathfrak{L}}$.

The following lemma is a crucial step in the proof of the Main Theorem.

LEMMA 1. *$\hat{\mathfrak{L}}$ is an S-glued system. The S-glued sum of $\hat{\mathfrak{L}}$ can be represented as the sublattice*

$$K_{\mathfrak{L}} = \bigcup ([a, b]_K \mid [a, b]_L \text{ is a block of } L)$$

of K.

PROOF. To show that $\hat{\mathfrak{L}}$ is an S-glued system, we have to verify Conditions 1–4 of Definition 1. To verify Condition 1, let $s, t \in S$, $s \leqq t$, and $x \in \hat{L}_s \cap \hat{L}_t$. Then $x \leqq b_s$ and $a_t \leqq x$, hence $a_t \leqq b_s$. It follows that $\hat{L}_s \cap \hat{L}_t = [a_t, b_s]_K$. Since $[a_t, b_s]_L$ is a dual ideal of L_s and an ideal of L_t, it follows that $[a_t, b_s]_K = \hat{L}_s \cap \hat{L}_t$ is a dual ideal of $\hat{L}_s$ and an ideal of $\hat{L}_t$, verifying Condition 1. Conditions 2 and 3 are trivial. To verify Condition 4, let $s, t \in S$; then $L_s \cap L_t \subseteq L_{s \wedge t} \cap L_{s \vee t}$, which implies that $a_{s \vee t} \leqq a_s \vee a_t$ and $b_{s \wedge t} \geqq b_s \wedge b_t$, which, in turn, imply Condition 4 for $\hat{\mathfrak{L}}$.

$K_{\mathfrak{L}}$ is a sublattice of K. Indeed, let $x_1, x_2 \in K_{\mathfrak{L}}$. Then there exist $[a_1, b_1]_L$ and $[a_2, b_2]_L$ that are blocks of L and satisfy $x_1 \in [a_1, b_1]_K$ and $x_2 \in [a_2, b_2]_K$. By Statement (5) of Theorem 1, there exists a block of L of the form $[a_1 \vee a_2, b_3]_L$. It follows that $x_1 \vee x_2 \in [a_1 \vee a_2, b_3]_K \subseteq K_{\mathfrak{L}}$. Dually, $x_1 \wedge x_2 \in K_{\mathfrak{L}}$.

It remains to show that the partial order on $K_{\mathfrak{L}}$ is the transitive extension of the union of the partial orders on the $\hat{L}_s$, $s \in S$ and so $K_{\mathfrak{L}}$ satisfies the condition in 2. This is really easy because of the following well-known statement:

Let A and B be lattices; let C be the gluing of A and B over the (isomorphic) dual ideal D of A and ideal I of B. Let E be any lattice containing A and B as sublattices such that $A \cap B = D = I$ in E. Then $A \cup B$ is a sublattice of E and it is isomorphic to C.

Now if $a, b \in K_{\mathfrak{L}}$ and $a \leqq b$ in $K_{\mathfrak{L}}$, then $a \in [u, v]_K$ and $b \in [w, z]_K$, where $u, v, w, z \in L$. By Statement (5) of Theorem 1, we can assume that $u \leqq w$. Let

$$u = u_0 \prec u_1 \prec \cdots \prec u_n = w$$

be a maximal chain in L. Then we can find blocks

$$[x_1, y_1]_L, \ [x_2, y_2]_L, \ldots, \ [x_n, y_n]_L$$

such that

$$u_0, u_1 \in [x_1, y_1],\ u_1, u_2 \in [x_2, y_2],\ \ldots,\ u_{n-1}, u_{n-1} \in [x_n, y_n].$$

By 1.3,

$$\bigcup ([x_i, y_i]_K \mid i = 1, \ldots, n)$$

can be obtained by repeated gluings, hence by the well-known statement we quoted, $a \leqq b$ in $K_{\mathfrak{L}}$ iff $a \leqq b$ in the S-glued sum of $\hat{\mathfrak{L}}$. □

3. Constructing a finite distributive lattice

We need the following theorem of G. Birkhoff [4]:

THEOREM 2. *Every finite group $\mathfrak{G}$ can be represented as the automorphism group of a finite distributive lattice D.*

The proof of the Main Theorem is based on a new proof of this result in G. Grätzer, H. Lakser and E. T. Schmidt [10]. We proceed now to outline this proof.

By R. Frucht [6] (see also G. Grätzer and H. Lakser [9], for an alternative proof), there exists an undirected finite graph $\langle V, E \rangle$ with no loops (that is, V is a set and E is a set of two-elements subsets of V) whose automorphism group is isomorphic to $\mathfrak{G}$. Since the automorphism group of $\langle V, E \rangle$ is the same as the automorphism group of the complement of $\langle V, E \rangle$, we can assume that $E \neq \emptyset$.

Let $V = \{v_1, v_2, \ldots, v_n\}$. Let F be the free distributive lattice over V with zero $0 = \bigwedge V$ and unit $1 = \bigvee V$. Define the element $\mathbf{o}$ of F as follows:

$$\mathbf{o} = \bigvee (x \wedge y \mid \{x, y\} \in E).$$

Then we can construct the finite distributive lattice D of Theorem 2 as follows:

$$D = [\mathbf{o}, 1].$$

The zero of D is $\mathbf{o}$. It is easily seen that $\{x, y\} \in E$ iff $x \wedge y \leqq \mathbf{o}$.

LEMMA 2. *The automorphism group of D is isomorphic to $\mathfrak{G}$.*

PROOF. Let α be an automorphism of $\langle V, E \rangle$. Then α has a natural extension to an automorphism $F(\alpha)$ of F, and $\mathbf{o}$ is a fixed point of $F(\alpha)$. So $F(\alpha)$ restricted to D yields an automorphism $D(\alpha)$ of D. The map $\alpha \to D(\alpha)$ is an isomorphism between the automorphism group of $\langle V, E \rangle$ and the automorphism group of D. □

For a more detailed exposition, see [10].

LEMMA 3. *Let* $w_1 = \mathbf{o} \vee v_1, \ldots, w_n = \mathbf{o} \vee v_n$, *and*

$$W = \mathbf{o} \vee V = \{w_1, \ldots, w_n\}.$$

Then D *is freely generated by* W, *subject to the relations*

$$w_i \wedge w_j = \mathbf{o}, \text{ for } \{v_i, v_j\} \in E.$$

PROOF. D is generated by W, and W satisfies the above relations, by the distributivity of F.

Now let L be a distributive lattice with 0, and let φ be a map of W into L subject to the condition that if $\{v_i, v_j\} \in E$, then $w_i\varphi \wedge w_j\varphi = 0$ in L. Let $\psi: V \to L$ be defined by $v_1\psi = w_1\varphi, \ldots, v_n\psi = w_n\varphi$. Since F is free, there is a homomorphism $\overline{\psi}: F \to L$ extending ψ. Let use define $\overline{\varphi}$ as the restriction of $\overline{\psi}$ to D. Then $\overline{\varphi}$ is a homomorphism of D into L, and $\overline{\varphi}$ extends φ; indeed,

$$w_i\overline{\varphi} = w_i\overline{\psi} = (\mathbf{o} \vee v_i)\overline{\psi} = \mathbf{o}\overline{\psi} \vee v_i\overline{\psi} = 0 \vee w_i\varphi = w_i\varphi,$$

as claimed. □

LEMMA 4. *Let* $P = J(D)$ *and* $Z = P - W$. *Then* W *is the set of maximal elements of* P, *every element of* P *is a meet of elements of* W, *and both* $P \cup \{\mathbf{o}\}$ *and* $Z \cup \{\mathbf{o}\}$ *are meet-subsemilattices of* D.

PROOF. Since V is a free generating set of F, we conclude, by Lemma 3, that every join-irreducible element of D is a meet of a finite nonempty subset of V. Let $a = \bigwedge W_1$, where W_1 is a nonempty finite subset of W, and let V_1 be the corresponding subset of V. Then $(\bigwedge V_1) \vee \mathbf{o} = \bigwedge W_1 = a$, so the interval $[\mathbf{o}, a]$ in D is isomorphic to the interval $[\mathbf{o} \wedge \bigwedge V_1, \bigwedge V_1]$ in F. Since $\bigwedge V_1$ is join-irreducible in F, it follows that a is join-irreducible in D. □

Observe that we can describe P as the poset of all nonempty subsets X of W excluding all "edges", that is, if $\{v_i, v_j\} \in E$, then $\{w_i, w_j\} \not\subseteq X$, and D is the lattice of hereditary subsets of P.

4. Constructing an infinite distributive lattice

We construct an infinite distributive lattice, $\overline{D}$, a stretched version of D, by replacing each w_i by an infinite chain C_i.

More formally, let

$$\overline{P} = Z \cup C,$$

where

$$C = C_1 \cup C_2 \cup \cdots \cup C_n,$$

and

$$C_i = \{w_{i,1}, w_{i,2}, \ldots\}, \quad w_{i,1} < w_{i,2} < \ldots.$$

We identify $w_{1,1}$ with w_1, ..., $w_{1,n}$ with w_n, respectively, and so we identify W with $\{w_{1,1}, \ldots, w_{n,1}\} \subseteq \overline{P}$. The partial order on $\overline{P}$ is the transitive closure of the union of the partial order on P and the orders on the C_i.

Let $\overline{D}$ be defined as an interval finite distributive lattice in which every element is a join of join-irreducible elements and satisfying $J(\overline{D}) = \overline{P}$. Equivalently, $\overline{D}$ is the lattice of finite hereditary subsets of $\overline{P}$: an element A of $\overline{D}$ is of the form

$$A = \bigcup (\{ x \mid x \leqq h \} \mid h \in H) \subseteq \overline{P},$$

for some finite subset H of $\overline{P}$. Every element of A is contained in a maximal element of A (which is in H); there are only finitely many maximal elements in A. If H is an antichain in $\overline{P}$, then the maximal elements in A form the set H.

Observe that, by Lemma 4, $P \subseteq \overline{P}$, and so D is a sublattice of $\overline{D}$, in fact, D is an ideal of $\overline{D}$. The zero of D, $\mathbf{o}$, is also the zero of $\overline{D}$. Now for $a \in \overline{P}$, we further identify a with $\{ x \mid x \leqq a \text{ and } x \in \overline{P} \}$, so $J(\overline{D}) = \overline{P}$.

LEMMA 5. *If $\{v_i, v_j\} \in E$, then $w_{i,k} \wedge w_{j,m} = \mathbf{o}$ in $\overline{D}$, for any $k, m \geqq 1$.*

PROOF. This is obvious, since the hereditary set corresponding to $w_{i,k}$ is

$$\{ z \mid z \in Z \text{ and } z \leqq w_i \} \cup \{w_{i,1}, \ldots, w_{i,k}\}.$$

So forming the intersection of the set corresponding to $w_{i,k}$ with the set corresponding to $w_{j,m}$ we obtain $\{ z \mid z \in Z,\ z \leqq w_i \text{ and } z \leqq w_j \} = \emptyset$ since $\{v_i, v_j\} \in E$. □

LEMMA 6. *$\overline{D}$ as a distributive lattice is freely generated by C subject to the partial ordering of C and the relations*

$$w_{i,k} \wedge w_{j,m} = w_{i,1} \wedge w_{j,1} \quad \text{for } i \neq j, \quad 1 \leqq i,\ j \leqq n, \quad k,\ m \geqq 1;$$
$$w_{i,k} \wedge w_{j,m} = \mathbf{o}, \quad \text{where } \{v_i, v_j\} \in E, \quad k,\ m \geqq 1.$$

PROOF. A routine computation shows that the poset of join-irreducible elements of the free distributive lattice described in this lemma is isomorphic to $\overline{P}$, hence the statement. □

The following concepts will help us characterize the elements of C in $\overline{D}$.

DEFINITION 3. Let L be a lattice. A *tight chain in L* is an infinite sequence of join-irreducible elements

$$x_0 \prec x_1 \prec \ldots$$

of L. A *tight element in L* is an element of a tight chain.

And here is the characterization:

LEMMA 7. *In $\overline{D}$, an element $c \in C$ can be characterized by the following two conditions:*

(1) *c is a tight element in $\overline{D}$.*

(2) *If a and b are distinct tight elements in $\overline{D}$ with $c < a$ and $c < b$, then a and b are comparable.*

PROOF. In $\overline{D}$, it is obvious that $w_{i,n} \prec w_{i,n+1} \prec \ldots$ is a tight chain; and conversely, every tight chain from some point on must be of this form.

Every $c \in C$ is in a tight chain, namely in some C_i. Moreover, if a and b are distinct tight elements in $\overline{D}$, then there are C_j and C_k and $\overline{a} \in C_j$, $\overline{b} \in C_k$ such that $a \leqq \overline{a} \in C_j$ and $b \leqq \overline{b} \in C_k$. Since $c < a \leqq \overline{a}$ and $c < b \leqq \overline{b}$, it follows that $i = j = k$; therefore $\overline{a}$, $\overline{b} \in C_i$ and so they are comparable.

Conversely, let c satisfy Conditions 1 and 2 of Lemma 7. If $c \notin C$, then $c \in Z$, so $c = \bigwedge W_1$ for some $W_1 \subseteq W$ with $1 < |W_1|$. So there are distinct a, $b \in W_1$ satisfying $c < a$, $c < b$, and a and b are not comparable. Since a and b are tight elements, this contradicts Condition 2 of Lemma 7. □

Let α be an automorphism of $\langle V, E \rangle$. The map $D(\alpha)$, defined in the proof of Lemma 2, is an automorphism of D, so it yields an automorphism of P. We extend $D(\alpha)$ to an automorphism $\overline{P}(\alpha)$ of $\overline{P}$ in the natural way: for $z \in Z$, let $z\overline{P}(\alpha) = zD(\alpha)$, and let $w_{i,k}\overline{P}(\alpha) = w_{j,k}$, if $w_i D(\alpha) = w_j$. In other words, $\overline{P}(\alpha)$ and $D(\alpha)$ act the same way on Z, and $\overline{P}(\alpha)$ acts on C_i as $D(\alpha)$ acts on w_i. The automorphism of $\overline{P}(\alpha)$ uniquely extends to an automorphism $\overline{D}(\alpha)$ of $\overline{D}$.

LEMMA 8. *The map $\alpha \to \overline{D}(\alpha)$ is an isomorphism between the automorphism group of $\langle V, E \rangle$ and the automorphism group of $\overline{D}$.*

PROOF. We already know that the map $\alpha \to \overline{D}(\alpha)$ embeds the automorphism group of $\langle V, E \rangle$ into the automorphism group of $\overline{D}$. To show that the map is onto, let β be an isomorphism of $\overline{D}$. By Lemma 7, we have an algebraic characterization of C, hence β must map C into itself. Thus β induces a map of W into itself, yielding a permutation α of V. It is easy to see that $\beta = \overline{D}(\alpha)$. □

5. Constructing a modular lattice

In this section, a *block* is a maximal complemented interval.

The next step is to embed $\overline{D}$ into a modular lattice M satisfying the conditions in the Main Theorem. This construction can be carried out for a large class of distributive lattices.

THEOREM 3. *Let L be a distributive lattice satisfying the following conditions:*

(1) *L is interval finite with zero.*

(2) *Every element in L is covered by finitely many elements.*

Then L can be embedded in an arguesian lattice M_L with the same two properties in which every block is an irreducible projective geometry; this embedding preserves covering. If, in addition, L satisfies the condition

(3) *Let a be an element of L that is not a bound of L; then there is an element $b \in L$ such that a and b are incomparable;*

then M_L is a simple lattice.

PROOF. We form the irreducible projective geometry G (identified with its subspace lattice) over the two-element field with $|J(L)|$ independent atoms, which we identify with $J(L)$, and embed L into G as follows:

Map $a \in L$ into $\bigvee((a] \cap J(L))$, where $(a] \cap J(L)$ is regarded as a set of independent atoms of G and the join is formed in G. This is a cover preserving embedding. We identify L with its image under this embedding into G.

Let S be the set of all blocks of L, partially ordered by $[a, b] \leqq [c, d]$ iff $a \leqq c$. It follows easily from Theorem 1 that S is an interval finite lattice. We consider L as the S-glued sum of the S-glued system $\mathfrak{L}$ formed by its blocks L_s, $s \in S$; obviously, $\mathfrak{L}$ satisfies the conditions of Definition 1. Every L_s is of the form $[a, b]_L$; define $\bar{L}_s = [a, b]_G$. We apply Lemma 1 to obtain that $\bar{\mathfrak{L}} = \{ \bar{L}_s \mid s \in S \}$ is an S-glued system, and the S-glued sum of $\bar{\mathfrak{L}}$ is

$$M_L = \bigcup ([a, b]_G \mid [a, b]_L \text{ is a block of } L).$$

By Statement (6) of Theorem 1, M_L is modular, and the maximal complemented intervals are of the form $[a, b]_G$, where $[a, b]_L$ is a block of L.

Now let us assume that, in addition, L satisfies Condition 3 of Theorem 3, and we prove that M_L is simple.

Every block of M_L is a simple lattice. M_L is interval finite; hence to prove it simple, it is sufficient to prove that if $[x, y]$ and $[y, z]$ are blocks of M_L and $[x, y]$ is collapsed by a congruence Θ, then so is $[y, z]$ (and dually).

Since $[x, y]$ and $[y, z]$ are blocks of M_L, it follows that x, y, $z \in L$; by Condition 3 of Theorem 3, there is an element $u \in L$ incomparable with y. So $y < y \vee u$, hence there is an element $p \in L$ satisfying $y \prec p \leqq y \vee u$. Similarly, there is an element $q \in L$ satisfying $y \wedge u \leqq q \prec y$. The same relations hold in M_L, and p is an atom in the block $[y, z]$, while q is a dual atom in the block $[x, y]$.

Obviously, the elements q, y, $(p \wedge u) \vee q$, and p form a covering Boolean interval in L, hence there is an element $m \in M_L$ such that the elements q, y, $(p \wedge u) \vee q$, p, and m form an $\mathfrak{M}_3$ in M_L.

Since Θ collapses $[x, y]$, therefore, $q \equiv y(\Theta)$. In $\mathfrak{M}_3$, we get that $y \equiv p(\Theta)$, hence Θ collapses $[y, z]$, as claimed. □

6. Proof of the Main Theorem

Let $M = M_{\overline{D}}$ as provided by Theorem 3. We show that M satisfies the requirements of the Main Theorem.

$\overline{D}$ obviously satisfies the first two conditions of Theorem 3.

Let $a \in \overline{D}$ and $a > 0$. Then in each C_i, for $1 \leqq i \leqq n$, there is a smallest a_i such that $a_i \leqq a$ fails. If no a_i is incomparable with a, then $a \leqq a_i$, for all $1 \leqq i \leqq n$, Since $a_1 \wedge \cdots \wedge a_n = 0$, it follows that $a = 0$, a contradiction. So Condition 3 of Theorem 3 is verified.

By Theorem 3, M is a simple modular lattice. It remains to show that the automorphism group of M is isomorphic to $\mathfrak{G}$. This is true because C is easy to recognize in M:

LEMMA 9. *In M, an element $c \in C$ can be characterized by the following two properties:*

(1) *c is a tight element in M.*

(2) *If a and b are distinct tight elements in M with $c < a$ and $c < b$, then a and b are comparable.*

PROOF. We only have to observe that if c is a tight element in M, then c is the unique lower cover of a join-irreducible element $d \in M$. Therefore, there exists in M a block of the form $[c, e]$, and so $c \in \overline{D}$ by Statement (6) of Theorem 1. So this lemma reduces to the statement of Lemma 7. □

In Section 4, for every automorphism α of $\langle V, E\rangle$, we have defined an automorphism $\overline{D}(\alpha)$ of $\overline{D}$. We claim that $\overline{D}(\alpha)$ defines a unique automorphism of M. Indeed, let $a \in M$. Then a is contained in a block $[c, d]$ of M. Since c, $d \in \overline{D}$, the block $[c\overline{\alpha}, d\overline{\alpha}]$ of M is well defined. The atoms of $[c, d]$ that are in $\overline{D}$ form a basis for $[c, d]$, and on those $\overline{D}(\alpha)$ is already defined. Therefore, $\overline{D}(\alpha)$ has a unique extension to an isomorphism between $[c, d]$ and $[c\overline{D}(\alpha), d\overline{D}(\alpha)]$. (This is obvious, but we would like to point out that in this step—and only in this step—we use the fact that we take a projective geometry over the *two-element* field. This step is equivalent to the following: if φ maps $\mathfrak{M}_3$ into itself and it maps two distinct atoms to two distinct atoms, then φ has a unique extension to an automorphism of $\mathfrak{M}_3$. This statement obviously fails for the subspace lattice of a projective line with more than three elements.) Since M is an S-glued sum of its blocks, $\overline{D}(\alpha)$ uniquely extends to an automorphism of M.

Therefore, the automorphism group of M is isomorphic to the automorphism group of $\overline{D}$; a reference to Lemma 8 completes the proof of the Main Theorem.

We presented the proof of this theorem in the spirit of its discovery: M is glued together from its parts, which are finite Boolean lattices completed to finite projective geometries; the Boolean lattices interface in $\overline{D}$, and this defines how the projective geometries face each other. This approach is very easy intuitively, and it can be easiest to formalize using S-glued sums.

Unfortunately, the arguesian identity is difficult to verify using this approach, so now all of M is built as a sublattice of a projective geometry making it possible not to use S-glued sums at all. Such an approach may be completely direct, or it may use the natural tolerance relations on modular lattices (see H.-J. Bandelt [3]). For yet another alternative approach, see A. Day and Ch. Herrmann [5].

7. Discussion

As customary, let $\mathcal{V}_2$ denote the quasivariety of all lattices that can be embedded into the subspace lattice of a projective geometry over the two-element field. We have proved the following stronger form of the Main Theorem:

MAIN THEOREM′. *Let $\mathfrak{G}$ be a finite group. Then there exists a lattice $M \in \mathcal{V}_2$ such that the automorphism group of M is isomorphic to $\mathfrak{G}$.*

Two problems suggest themselves.

PROBLEM 1. Let $\mathfrak{G}$ be a finite group. Does there exist a *finite simple arguesian lattice* M such that the automorphism group of M is isomorphic to $\mathfrak{G}$?

In this connection, one may mention a conjecture of L. Babai and D. Duffus [2]: for every natural number n, there is a group $\mathfrak{G}_n$ that cannot be represented as the automorphism group of a finite modular lattice of length at most n.

And here is a problem for infinite groups:

PROBLEM 2. Let $\mathfrak{G}$ be an arbitrary group. Does there exist a *simple arguesian lattice* M ($M \in \mathcal{V}_2$) such that the automorphism group of M is isomorphic to $\mathfrak{G}$?

8. New results

(Note added February 22, 1997)

The authors utilized the Main Theorem of this paper to prove the following:

INDEPENDENCE THEOREM. *Let D be a finite distributive lattice, and let $\mathfrak{G}$ be a finite group. Then there exists a modular lattice M such that the congruence lattice of M is isomorphic to D and the automorphism group of M is isomorphic to $\mathfrak{G}$.*

This result was presented in the Lattice Theory and Universal Algebra Seminar at the University of Manitoba on May 16, 1996; the paper containing

this result is about to be submitted for publication (*On the Independence Theorem of related structures for modular (arguesian) lattices*). We also solved Problem 2 (*On automorphism groups of simple arguesian lattices*, manuscript).

We sent the manuscript of this paper to Ch. Herrmann in June of 1996. In the fall, he sent us the manuscript *On automorphism groups of Arguesian lattices*, in which he solves both problems of Section 7. His paper is to appear in Acta Mathematica Hungarica.

REFERENCES

[1] Babai, L., Vector representable matroids of given rank with given automorphism group, *Discrete Math.* **24** (1978), 119–125. *MR* **80a**:05054

[2] Babai, L. and Duffus, D., Dimension and automorphism groups of lattices, *Algebra Universalis* **12** (1981), 279–289. *MR* **83a**:06006

[3] Bandelt, H.-J., Tolerance relations on lattices, *Bull. Austral. Math. Soc.* **23** (1981), 367–381. *MR* **83d**:06005

[4] Birkhoff, G., On groups of automorphisms, *Revista Unión Mat. Argentina* **11** (1946), 155–157 (in Spanish). *MR* **7**,411a

[5] Day, A. and Herrmann, Ch., Gluings of modular lattices, *Order* **5** (1988), 85–101. *MR* **89m**:06009

[6] Frucht, R., Herstellung von Graphen mit vorgegebener abstrakter Gruppe, *Compositio Math.* **6** (1938), 239–250. *Zbl* **20**, 078

[7] Frucht, R., Lattices with a given abstract group of automorphisms, *Canad. J. Math.* **2** (1950), 417–419. *MR* **12**,473d

[8] Grätzer, G., *General lattice theory*, Pure and Applied Mathematics, Vol. 75, Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers), New York–London; Lehrbücher und Monographien aus dem Gebiete der Exakten Wissenschaften, Mathematische Reihe, Band 52, Birkhäuser Verlag, Basel–Stuttgart; Akademie Verlag, Berlin, 1978. *MR* **80c**:06001a, 06001b

[9] Grätzer, G. and Lakser, H., Homomorphisms of distributive lattices as restrictions of congruences. II. Planarity and automorphisms, *Canad. J. Math.* **46** (1994), 3–54. *MR* **94m**:06008

[10] Grätzer, G., Lakser, H. and Schmidt, E. T., On a result of Birkhoff, *Period. Math. Hungar.* **30** (1995), 183–188. *MR* **96e**:06010

[11] Herrmann, Ch., S-verklebte Summen von Verbänden, *Math. Z.* **130** (1973), 255–274. *MR* **49** #7195

[12] Mendelsohn, E., Every group is the collineation group of some projective plane. *Foundations of geometry* (Proc. Conf., Univ. Toronto, Toronto, Ont., 1974), Univ. Toronto Press, Toronto, Ont., 1976, 175–182. *MR* **53** #9029

(Received December 16, 1996)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANITOBA
WINNIPEG, Manitoba
R3T 2N2
CANADA
gratzer@cc.umanitoba.ca

BUDAPESTI MŰSZAKI EGYETEM
MATEMATIKAI INTÉZETE
MŰEGYETEM RKP. 3
H–1521 BUDAPEST
HUNGARY
schmidt@euromath.vma.bme.hu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 259–259

BOOK REVIEW

Jenő Szép: **Vectorproducts and Applications**. First volume of the new series: *Pure and Applied Mathematics*, Akadémiai Kiadó, Budapest, 1998, 110 p.

The author who is well known in group theory, semigroup theory and ring theory, in this book (summarizing and completing nine research papers published during a time interval of twenty years (1975–1995)) exhibited his skill to invent a new algebraic structure.

His first idea to define setvector based on integer numbers, in a way a set of special vectors based on integer numbers form a setvector. Then a componentwise addition of integer numbers, defines the addition of the elements (vectors) of the setvector. The defined multiplication(s) produce a groupoid. Since in general a groupoid is not associative (neither commutative) parentheses must be used to avoid ambiguity. The structure of parentheses makes possible unique factorisation.

The book contains 10 chapters. It consists of two parts; the first part is a detailed discussion of the discrete case and the second part is a short description of the continuous case.

The author exhibits how his new algebraic structure (which he called Coded structure or Vectorproduct) can be applied to encipher and decipher plaintext. At present the most frequently used methods of enciphering (DES, RSA) based on groups and fields, the approach in this book seems to be new and hopefully very efficient. The author's idea can be more widely applied in cryptology. Access control, digital signature and digital fingerprint would be some further applications. Unfortunately the latter applications are not mentioned in the book. Some other practical applications such as simulation of processes in chemistry, nuclear physics and microbiology are briefly mentioned. Also the author observed the close connection between coded structures and Lindenmayer systems and Theory of automata. The reviewer thinks that this book has theoretical and practical values.

J. Dénes (Budapest)

0081-6906/99/$ 5.00

Typeset by TypoTEX Ltd., Budapest
PRINTED IN HUNGARY
Akadémiai Nyomda, Martonvásár

RECENTLY ACCEPTED PAPERS

DOSTER, W., Jeffreys' prior is the Hausdorff measure for the Hellinger and Kullback–Leibler distances

GAL, S. G. and SZABADOS, J., On the preservation of global smoothness by some interpolation operators

LANZINGER, H., A law of the single logarithm for moving averages of random variables under exponential moment conditions

ELSNER, C. and SANDER, J. W., On the distribution of residue classes of quadratic forms and integer-detecting sequences in number fields

SLEZÁK, B., A right inverse function theorem without assuming differentiability

LINDSTRÖM, B., A prime multiplier of the B_2-sets of Bose and Chowla

CARMONA, PH., PETIT, F., YOR, M. and PITMAN, J., On the laws of homogeneous functionals of the Brownian bridge

ZEMPLÉNI, A., Max-semigroups of bivariate random variables with Khinchine-type decomposition theorems

KÜNZI, H.-P. A., Remark on a result of Losonczi

PITMAN, J. and YOR, M., Path decompositions of a Brownian bridge related to the ratio of its maximum and amplitude

Manuscripts should be submitted in duplicate, typed in double spacing on only one side of the paper with wide margins. Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only upon receipt of the signed copy of the agreement.

Authors are encouraged to submit their papers electronically. All common dialects of TEX are welcome. The electronic file of an article should always be accompanied by a hardcopy, the printed version exactly corresponding to the electronic one.

Figures should be submitted on separate sheets, drawn either in India ink at their expected final size, or as printouts and matching files processed electronically, preferably as encapsulated PostScript (EPS) ones.

CONTENTS

HU ISSN 0081–6906

Studia Scientiarum Mathematicarum Hungarica

VOLUME 35
NUMBERS 3–4
1999

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 and French. It is published in yearly volumes of four issues (mostly double numbers published semiannually) by

AKADÉMIAI KIADÓ
H-1117 Budapest, Prielle Kornélia u. 4
http://www.akkrt.hu

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H-1364 Budapest

Tel.: +36 1 318 2875 Fax: +36 1 317 7166
E-mail: merza@math-inst.hu

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O.Box 245
H-1519 Budapest
Fax: +36 1 464 8221
E-mail: kiss.s@akkrt.hu

For 2000 volume 36 is scheduled for publication. The subscription price is $ 196.00, air delivery plus $ 20.00.

Coden: SSMHAX	December, 1999
Vol: 35	**Pages**: 261–478
Numbers: 3–4	**Whole**: 69

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 261–280

EIGENVALUE PROBLEMS FOR BESSEL'S EQUATION AND ZERO-PAIRS OF BESSEL FUNCTIONS

H. VOLKMER

Abstract

This paper studies an eigenvalue problem for Bessel's differential equation involving two complex parameters. The results are based on an investigation of zero-pairs of Bessel functions; these are pairs of complex numbers at which a Bessel function vanishes simultaneously. Properties of zero-pairs are derived from estimates satisfied by a quotient of Hankel functions.

1. Introduction

Eigenvalue problems for Bessel's differential equation belong to the best known and most intensively studied eigenvalue problems in applied mathematics. Let us consider the eigenvalue problem for a vibrating membrane occupying the region $0 < c \leqq r \leqq d$, $0 \leqq \phi \leqq \psi$ in polar coordinates r, ϕ ([3, V, §5]):

$$r(ru')' + (\lambda r^2 - \nu^2)u = 0, \quad u(c) = u(d) = 0. \tag{1.1}$$

The order ν is determined by ψ. The problem consists in determining those values of the spectral parameter λ for which (1.1) has a nontrivial solution. This is a regular Sturm–Liouville problem which has a monotonically increasing and positive sequence of eigenvalues $0 < \lambda_1(c,d) < \lambda_2(c,d) < \dots$ depending on c and d. We do not indicate the dependence of λ_n on the order ν because we consider ν as a fixed nonnegative number.

The eigenvalue problem (1.1) is closely related to the question of finding zero-pairs of Bessel functions of order ν; these are numbers a, b for which there exists a Bessel function of order ν which vanishes simultaneously at a and b. In fact, the solutions of the differential equation in (1.1) are of the form $C_\nu(\beta r)$, where $\beta^2 = \lambda$ and C_ν is an arbitrary solution of Bessel's differential equation

$$z(zC_\nu')' + (z^2 - \nu^2)C_\nu = 0 \tag{1.2}$$

1991 *Mathematics Subject Classification.* Primary 33C10; Secondary 34B30.

Key words and phrases. Bessel functions, Hankel functions, zeros of Bessel functions, eigenvalues of Bessel's equations.

0081–6906/99/$ 5.00

of order ν. Therefore, $\lambda > 0$ is an eigenvalue of (1.1) if and only if $a = \lambda^{1/2}c$, $b = \lambda^{1/2}d$ form a zero-pair of order ν.

In this paper we investigate the global behavior of the functions $\lambda_n(c,d)$ in their dependence on complex variables c and d. This includes a study of complex zero-pairs of Bessel functions. Let us first simplify the task by reducing $\lambda_n(c,d)$ to a function of one variable. This is possible because of the homogeneity relation

$$\lambda_n(tc,td) = t^{-2}\lambda_n(c,d), \quad t>0$$

which is easy to prove. Therefore, it will be sufficient to consider pairs c, d with $d-c=2$. The choice of the distance 2 between c and d is for convenience only. We write

$$c = \tau - 1, \quad d = \tau + 1, \tag{1.3}$$

where $\tau > 1$ is a new variable. Then, setting $y(x) = r^{1/2}u(r)$, $r = \tau - x$, our eigenvalue problem assumes the attractive form

$$y'' + \left(\lambda + \frac{\frac{1}{4} - \nu^2}{(x-\tau)^2}\right) y = 0, \tag{1.4}$$

$$y(-1) = y(1) = 0. \tag{1.5}$$

For given $\tau > 1$ (or $\tau < -1$), we again have a regular Sturm–Liouville problem with eigenvalues

$$0 < \lambda_1(\tau) < \lambda_2(\tau) < \dots \tag{1.6}$$

that agree with those of (1.1) under the substitution (1.3).

It is easy to show that the functions $\lambda_n(\tau)$ are analytic for $\tau > 1$ (see Section 2). It is therefore natural to ask for properties of the analytic continuation of $\lambda_n(\tau)$ into the complex τ-plane. Of course, the values of this analytic continuation will also be eigenvalues of (1.4), (1.5) in a sense to be specified later. In Section 2, we prove that $\lambda_n(\tau)$ is analytic at $\tau = \infty$ and can be continued analytically into a domain of the form $\operatorname{dist}(\tau, [-1,1]) > \epsilon_n > 0$ with $\epsilon_n \to 0$ as $n \to \infty$. Here $\operatorname{dist}(\tau, [-1,1])$ denotes the distance from τ to the line segment $[-1,1]$. It is to be expected that τ-values in the interval $(-1,1)$ are "critical" because then the regular singular point $x = \tau$ of (1.4) lies between the endpoints -1 and 1 appearing in the boundary conditions (1.5). The question now arises how the functions $\lambda_n(\tau)$ behave as τ approaches the segment $[-1,1]$. Computer calculations show branching between the functions $\lambda_n(\tau)$ in a neighborhood of $[-1,1]$. The existence of branch points is closely related to the phenomenon of level crossing of eigenvalues as described in Bender and Orszag [2, p. 350].

The location of branch points of the functions $\lambda_n(\tau)$ will depend on the value of ν. If $\nu = 1/2$, then $\lambda_n(\tau)$ is identically equal to $n^2\pi^2/4$. There are no branch points. In Section 5, as the main result of this paper, we prove that the functions $\lambda_n(\tau)$ do not have branch points in $\mathbb{C} \setminus [-1,1]$ if $\nu \in [1/3, 1/2]$. More precisely, we will prove the following theorem.

THEOREM 1.1. *If $\nu \in [1/3, 1/2]$, then the functions $\lambda_n(\tau)$, $n \in \mathbb{N}$, are analytic in the domain $\mathbb{C} \setminus [-1, 1]$ and at $\tau = \infty$. Moreover, these functions are also analytic on the segments $\tau + i0$ and $\tau - i0$ for $-1 < \tau < 1$, and they can be extended continuously into $\tau = 1$ and $\tau = -1$.*

Of course, the values of $\lambda_n(\tau + i0)$ will not match those of $\lambda_n(\tau - i0)$ if $\nu \neq 1/2$ because otherwise $\lambda_n(\tau)$ would be a bounded entire function and thus a constant function by Liouville's theorem.

The proof of Theorem 1.1 is based on a study of complex zero-pairs a, b of Bessel functions in Section 4. It should be noted that it is necessary to consider a and b as elements of the Riemann surface $\mathbb{C}_{\log}$ of the logarithm on which Bessel functions are analytic. The following observation will be crucial: a, b is a zero-pair of order ν if and only if there are complex constants A and B, not both zero, such that

$$AC_\nu(a) + BD_\nu(a) = 0, \quad AC_\nu(b) + BD_\nu(b) = 0,$$

where C_ν, D_ν form a fundamental set of solutions of Bessel's equation (1.2). It follows that a, b form a zero-pair of order ν if and only if

$$C_\nu(a)D_\nu(b) - C_\nu(b)D_\nu(a) = 0. \tag{1.7}$$

It will be convenient to choose $C_\nu = H_\nu^{(1)}$, $D_\nu = H_\nu^{(2)}$ because of the simple asymptotic behavior of the Hankel functions. Then (1.7) can be written in the form

$$\frac{H_\nu^{(2)}(a)}{H_\nu^{(1)}(a)} = \frac{H_\nu^{(2)}(b)}{H_\nu^{(1)}(b)}. \tag{1.8}$$

The quotient of Hankel functions will be investigated in Section 3.

In the final Section 6, we determine the behavior of the functions $\lambda_n(\tau)$ for τ in the critical interval $(-1, 1)$. This corresponds to a study of zero-pairs a, b of Bessel functions which have the property that 0 lies on the line segment connecting a and b.

Before we begin let us make some final remarks. Concerning the theory of Bessel functions, we refer to Watson's excellent treatise [13]. In particular, its Chapter 15 on the zeros of Bessel functions will be of interest to us. Recently, many new results on the zeros of Bessel functions have been discovered; we refer to [1, 4, 5, 6]. We do not intend to give a complete theory on the eigenvalue problem (1.4), (1.5). For instance, it would be of interest to investigate whether Theorem 1.1 (or variants thereof) remain valid for other ranges of ν. For example, if $\nu = 0$, then the functions $\lambda_n(\tau)$ are analytic and real-valued on the positive imaginary axis but there is branching between $\lambda_n(\tau)$ and $\lambda_{n+1}(\tau)$ exactly at $\tau = 0$ for all odd n. Thus Theorem 1.1 does not hold for $\nu = 0$ but we do not prove this here.

2. The eigenvalue problem

We consider the eigenvalue problem (1.4), (1.5) for a given order $\nu \geqq 0$. We say that (τ, λ) with $\tau \in \mathbb{C} \setminus [-1,1]$ and $\lambda \in \mathbb{C}$ is an *eigenpair* of order ν if there exists a nontrivial solution $y : [-1,1] \to \mathbb{C}$ of (1.4) and (1.5). This solution will be called an *eigenfunction* corresponding to (τ, λ). Clearly, if (τ, λ) is an eigenpair with eigenfunction $y(x)$, then $(-\tau, \lambda)$ is an eigenpair with eigenfunction $y(-x)$, and $(\overline{\tau}, \overline{\lambda})$ is an eigenpair with eigenfunction $\overline{y(x)}$.

If we allow $\tau = \infty$ in (1.4) (that means $(x-\tau)^{-2} = 0$), then (∞, λ) is an eigenpair if and only if there is $n \in \mathbb{N}$ such that $\lambda = n^2\pi^2/4$. The following lemma shows that all eigenpairs (τ, λ) are close to one of these eigenpairs provided that $|\tau|$ is large. The proof uses a standard method of perturbation theory; see [9, §1.5] or [7, Ch. 7].

THEOREM 2.1. *Let (τ, λ) be an eigenpair of order ν. Then there is $n \in \mathbb{N}$ such that*

$$\left|\lambda - \frac{1}{4}n^2\pi^2\right| \leqq \left|\frac{1}{4} - \nu^2\right| \operatorname{dist}(\tau, [-1,1])^{-2}.$$

PROOF. Let

$$w_n(x) = \sin\left(\frac{1}{2}n\pi(x+1)\right), \quad n \in \mathbb{N}$$

be the normalized eigenfunctions of

$$w'' + \lambda w = 0, \quad w(-1) = w(1) = 0.$$

The sequence w_n forms an orthonormal basis of $L^2(-1,1)$. Let (τ, λ) be an eigenpair with eigenfunction y. Then

$$w_n''y - w_n y'' = \left(\lambda - \frac{1}{4}n^2\pi^2\right) y w_n + \left(\frac{1}{4} - \nu^2\right)(x-\tau)^{-2} y w_n.$$

Thus

$$0 = w_n' y - w_n y'|_{-1}^{1} = \left(\lambda - \frac{1}{4}n^2\pi^2\right) \int_{-1}^{1} y w_n + \left(\frac{1}{4} - \nu^2\right) \int_{-1}^{1} (x-\tau)^{-2} y w_n.$$

Use Parseval's identity twice to obtain

$$\min_{n=1}^{\infty} \left|\lambda - \frac{1}{4}n^2\pi^2\right|^2 \int_{-1}^{1} |y|^2 \leqq \sum_{n=1}^{\infty} \left|\lambda - \frac{1}{4}n^2\pi^2\right|^2 \left|\int_{-1}^{1} y w_n\right|^2$$

$$= \left|\frac{1}{4} - \nu^2\right|^2 \sum_{n=1}^{\infty} \left| \int_{-1}^{1} (x-\tau)^{-2} y w_n \right|^2 = \left|\frac{1}{4} - \nu^2\right|^2 \sum_{n=1}^{\infty} \int_{-1}^{1} |x-\tau|^{-4} |y|^2$$

$$\leqq \left|\frac{1}{4} - \nu^2\right|^2 \operatorname{dist}(\tau, [-1,1])^{-4} \int_{-1}^{1} |y|^2.$$

Since $\int_{-1}^{1} |y|^2 > 0$, the desired statement follows. □

Let $Y(x;\tau,\lambda)$ be the solution of (1.4) which is uniquely determined by the initial values $y(-1)=0$ and $y'(-1)=1$. Set $D(\tau,\lambda) = Y(1;\tau,\lambda)$. The function $D(\tau,\lambda)$ is analytic for $\tau \in \mathbb{C} \setminus [-1,1]$ and $\lambda \in \mathbb{C}$. Its zeros are the eigenpairs. By introducing a new variable $1/\tau$, we see that $D(\tau,\lambda)$ is also analytic at $\tau = \infty$.

THEOREM 2.2. *Let $n \in \mathbb{N}$. Define $\epsilon_n \geqq 0$ by*

$$\epsilon_n^2 = \frac{8\left|\frac{1}{4} - \nu^2\right|}{(2n-1)\pi^2}$$

if $n \geqq 2$ and $\epsilon_1 := \epsilon_2$. Then there exists a uniquely determined analytic function $\lambda_n(\tau)$ for $\operatorname{dist}(\tau,[-1,1]) > \epsilon_n$ including $\tau = \infty$ such that $\lambda_n(\infty) = n^2\pi^2/4$ and such that $(\tau, \lambda_n(\tau))$ is an eigenpair for every τ.

PROOF. Let $n \in \mathbb{N}$. By Theorem 2.1, $D(\tau,\lambda) \neq 0$ for all λ on the circle $|\lambda - n^2\pi^2/4| = \pi^2(2n-1)/8$ and all τ with $\operatorname{dist}(\tau,[-1,1]) > \epsilon_n$. For $\tau = \infty$, there is exactly one zero $\lambda = n^2\pi^2/4$ within the circle (with regard to multiplicity). By Rouché's theorem, there is exactly one zero λ of $D(\tau,\lambda)$ within the circle for all τ with $\operatorname{dist}(\tau,[-1,1]) > \epsilon_n$. The induced function $\lambda_n(\tau)$ is analytic by the implicit function theorem. If $n=1$, then the proof has to be modified in an obvious way. □

The following observation follows from the proof of Theorem 2.2.

COROLLARY 2.3. *If $\delta \geqq 0$ is defined by*

$$\delta^2 = \frac{8}{3}\pi^{-2}\left|\frac{1}{4} - \nu^2\right|,$$

then all eigenpairs (τ,λ) with $\operatorname{dist}(\tau,[-1,1])) > \delta$ are given by $(\tau, \lambda_n(\tau))$, $n \in \mathbb{N}$.

The functions $\lambda_n(\tau)$ defined by Theorem 2.2 are even. If $\tau > 1$ or $\tau < -1$, then the functions $\lambda_n(\tau)$ agree with the sequence (1.6) of eigenvalues that we found by regular Sturm–Liouville theory. It should be noted that the

functions $\lambda_n(\tau)$ of (1.6) were defined for all $\tau > 1$ but this is not true for the functions $\lambda_n(\tau)$ of Theorem 2.2. It is clear that comparison of (1.4) with $w'' + \lambda w = 0$ cannot give optimal results for τ close to $[-1, 1]$.

It will be useful to allow τ to assume the values $\tau + i0$ and $\tau - i0$ if $-1 < \tau < 1$. A pair $(\tau \pm i0, \lambda)$ is called eigenpair if there is a solution $y(x)$ of (1.4), (1.5) that is analytic in $\operatorname{Im} x \leqq 0$, $x \neq \tau + i0$ or in $\operatorname{Im} x \geqq 0$, $x \neq \tau - i0$, respectively. We also allow $\tau = 1$ and $\tau = -1$. In this case the regular singular point $x = \tau$ of equation (1.4) coincides with one of the points ± 1 appearing in the boundary conditions (1.5). Note that the regular singular point has exponents $1/2 \pm \nu$. We call $(1, \lambda)$ an eigenpair if a solution of (1.4) belonging to the exponent $1/2 + \nu$ at $x = \tau$ (without the logarithmic term if ν is an integer) vanishes at $x = -1$. Using the Bessel function J_ν of the first kind, such a solution is given by

$$y(x) = (1 - x)^{1/2} J_\nu(\beta(1 - x)),$$

where $\beta^2 = \lambda \neq 0$. It follows that the eigenpairs $(1, \lambda)$ are given by $(1, j_{\nu,n}^2/4)$, $n \in \mathbb{N}$, where $0 < j_{\nu,1} < j_{\nu,2} < \dots$ denotes the monotonically increasing sequence of positive zeros of J_ν. A similar definition applies to eigenpairs of the form $(-1, \lambda)$. The λ-components of the eigenpairs $(-1, \lambda)$ agree with those of the eigenpairs $(1, \lambda)$.

Let $a, b \in \mathbb{C}_{\log}$. We say that a, b form a *zero-pair* (of Bessel functions) of order ν if there exists a nontrivial solution of (1.2) which vanishes at a and b.

We now indicate how eigenpairs are connected with zero-pairs.

LEMMA 2.4. *Let $\tau \in \mathbb{C} \setminus [-1, 1]$, $\beta \in \mathbb{C}$ with $\beta \neq 0$, and set*

$$a := \beta(\tau + 1), \quad b := \beta(\tau - 1).$$

Choose $\arg a, \arg b$ *such that* $|\arg a - \arg b| < \pi$. *Then (τ, β^2) is an eigenpair of order ν if and only if a, b is a zero-pair of order ν. The same equivalence holds for $\tau = \tau \pm i0$ with $\tau \in (-1, 1)$ if we choose* $\arg a$ *and* $\arg b$ *in such a way that* $\arg b - \arg a = \pm\pi$, *respectively.*

PROOF. This follows from the fact that the solutions of (1.4) with $\lambda = \beta^2$ are given by $y(x) = (\tau - x)^{1/2} C_\nu(\beta(\tau - x))$, where C_ν is an arbitrary solution of (1.2). If x runs from -1 to 1, $\beta(\tau - x)$ describes the line segment from a to b. If $\tau \in \mathbb{C} \setminus [-1, 1]$, then this line segment does not pass through 0. Then choosing $\arg a$ and $\arg b$ such that $|\arg a - \arg b| < \pi$, we see that the equivalence is true. If $\tau = \tau \pm i0$, then the line segment from a to b passes through zero. Choosing $\arg a$ and $\arg b$ as indicated, we obtain the desired equivalence. □

With $C_\nu(z)$ also $C_\nu(ze^{\pi i})$ and $\overline{C_\nu(\overline{z})}$ ($\arg \overline{z} := -\arg z$) solve (1.2). This implies the following lemma.

LEMMA 2.5. *If a, b is a zero-pair of order ν, then also $ae^{\pi i}, be^{\pi i}$ and $\overline{a}, \overline{b}$ are zero-pairs of order ν.*

3. Quotient of Hankel functions

We begin with the well known asymptotic formulas for the Hankel functions; see [13, p. 198]. For (small) $\delta > 0$, we have

$$H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)}(1 + \mathcal{O}(z^{-1}))$$

as $z \to \infty$ uniformly for $-\pi + \delta \leqq \arg z \leqq 2\pi - \delta$; and

$$H_\nu^{(2)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{-i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)}(1 + \mathcal{O}(z^{-1}))$$

as $z \to \infty$ uniformly for $-2\pi + \delta \leqq \arg z \leqq \pi - \delta$. This gives

$$\frac{H_\nu^{(2)}(z)}{H_\nu^{(1)}(z)} = e^{-2i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)}(1 + \mathcal{O}(z^{-1})) \tag{3.1}$$

as $z \to \infty$ uniformly for $-\pi + \delta \leqq \arg z \leqq \pi - \delta$. This asymptotic formula for the quotient of Hankel functions is not sufficient for our purposes because we also have to work with values of z close to 0. We therefore introduce the meromorphic function $Q_\nu(z)$ defined on $\mathbb{C}_{\log}$ by

$$Q_\nu(z) := e^{2i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)} \frac{H_\nu^{(2)}(z)}{H_\nu^{(1)}(z)}. \tag{3.2}$$

By (3.1), this function satisfies

$$Q_\nu(z) = 1 + \mathcal{O}(z^{-1}) \tag{3.3}$$

as $z \to \infty$ uniformly for $-\pi + \delta \leqq \arg z \leqq \pi - \delta$.

We note some further properties of Q_ν for $\nu \geqq 0$. Since

$$H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z), \quad H_\nu^{(2)}(z) = J_\nu(z) - iY_\nu(z), \tag{3.4}$$

and $J_\nu(z) = \mathcal{O}(1)$, $|Y_\nu(z)| \to \infty$ as $|z| \to 0$, we obtain

$$\frac{H_\nu^{(2)}(z)}{H_\nu^{(1)}(z)} \to -1 \quad \text{as} \quad |z| \to 0. \tag{3.5}$$

It follows that

$$Q_\nu(z) \to ie^{-\nu\pi i} \quad \text{as} \quad |z| \to 0. \tag{3.6}$$

By (3.4), $|H_\nu^{(1)}(z)| = |H_\nu^{(2)}(z)|$ for $\arg z = 0$ so that

$$|Q_\nu(z)| = 1 \quad \text{for} \quad \arg z = 0. \tag{3.7}$$

The Schwarz reflection principle yields

$$Q_\nu(\overline{z}) = \overline{Q_\nu(z)}^{-1}. \tag{3.8}$$

Thus the value of $Q_\nu(\overline{z})$ is the inversion of $Q_\nu(z)$ at the unit circle. The formulas (see [13, p. 75])

$$H_\nu^{(1)}(ze^{\pi i}) = -e^{-\nu\pi i}H_\nu^{(2)}(z),$$

$$H_\nu^{(2)}(ze^{\pi i}) = 2\cos(\nu\pi)H_\nu^{(2)}(z) + e^{\nu\pi i}H_\nu^{(1)}(z)$$

lead to

$$Q_\nu(ze^{\pi i}) = 2i\cos(\nu\pi)e^{-2iz} + Q_\nu(z)^{-1}. \tag{3.9}$$

Once we know $Q_\nu(z)$ for $0 \leqq \arg z \leqq \pi/2$, then (3.8) will determine $Q_\nu(z)$ for $-\pi/2 \leqq \arg z \leqq 0$, and (3.9) will determine $Q_\nu(z)$ for $\pi/2 \leqq \arg z \leqq \pi$. We will need to know the behavior of $Q_\nu(z)$ only in the sector $-\pi/2 \leqq \arg z \leqq \pi$.

A simple calculation using the definition of the modified Bessel functions I_ν and K_ν (see [13, p. 77]) gives

$$Q_\nu(ye^{i\frac{\pi}{2}}) = e^{-2y}\left(ie^{-\nu\pi i} + \pi\frac{I_\nu(y)}{K_\nu(y)}\right) \quad \text{for} \quad \arg y = 0. \tag{3.10}$$

Note that $I_\nu(y)$ and $K_\nu(y)$ are positive for $y > 0$. In particular, we find from (3.10) and

$$\sin(\nu\pi)K_\nu(y) = \frac{1}{2}\pi(I_{-\nu}(y) - I_\nu(y)), \quad \nu \notin \mathbb{Z}$$

that

$$\left|Q_\nu(ye^{i\frac{\pi}{2}})\right|^2 = e^{-4y}\left(1 + \pi^2\frac{I_\nu(y)I_{-\nu}(y)}{K_\nu(y)^2}\right) \quad \text{for} \quad \arg y = 0. \tag{3.11}$$

LEMMA 3.1. *Let $\nu \in [0, 1/2]$. Then the following estimates hold for $|Q_\nu(z)|$.*

(i) *If $0 \leqq \arg z \leqq \pi/2$, then $1 \leqq |Q_\nu(z)| \leqq 1 + 2\cos(\nu\pi)$;*
(ii) *if $-\pi/2 \leqq \arg z \leqq 0$, then $(1 + 2\cos(\nu\pi))^{-1} \leqq |Q_\nu(z)| \leqq 1$;*
(iii) *if $\pi/2 \leqq \arg z \leqq \pi$, then $1 - 2\cos(\nu\pi)e^{-2\operatorname{Im} z} \leqq |Q_\nu(z)| \leqq 1 + 2\cos(\nu\pi)$.*

PROOF. Let $\epsilon > 0$. We first show that

$$|Q_\nu(z)| \leqq 1 + 2\cos(\nu\pi) + \epsilon \quad \text{for} \quad 0 \leqq \arg z \leqq \pi. \tag{3.12}$$

This is true for $\arg z = 0$ and $\arg z = \pi$ by (3.7) and (3.9). By (3.3) it is true as $z \to \infty$ for $0 \leqq \arg z \leqq \pi/2$. By (3.3) and (3.9), it is true as $z \to \infty$ for $\pi/2 \leqq \arg z \leqq \pi$. The estimate (3.12) now follows from the maximum-modulus principle because $Q_\nu(z)$ is analytic for $0 \leqq \arg z \leqq \pi$ (and continuous at 0).

We used that $H_\nu^{(1)}(z)$ has no zeros in $0 \leqq \arg z \leqq \pi$; see [13, p. 511]. Letting $\epsilon \to 0$, we see that (3.12) holds with $\epsilon = 0$. This proves parts of (i) and (iii).

By [13, p. 441], we have

$$I_\nu(y)I_{-\nu}(y) = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} I_0(2y\cos\theta)\cos(2\nu\theta)\,d\theta.$$

Since $I_0(t) > 0$ for $t > 0$, this shows that $I_\nu(y)I_{-\nu}(y)$ is a monotonically decreasing function of $\nu \in [0,1]$ for every fixed $y > 0$. The formula [13, p. 181]

$$K_\nu(y) = \int_0^\infty e^{-y\cosh t}\cosh(\nu t)\,dt$$

shows that $K_\nu(y)$ is a monotonically increasing function of $\nu \geqq 0$ for every fixed $y > 0$; cf. [10, p. 251]. Hence, by (3.11), $|Q_\nu(ye^{\pi i/2})|$ is a monotonically decreasing function of $\nu \in [0,1/2]$ for every fixed $y > 0$. Since $Q_{1/2}(z) = 1$ for all z, we obtain $|Q_\nu(z)| \geqq 1$ for $\arg z = \pi/2$. We now use (3.3), (3.7) and the minimum-modulus principle to prove the remaining part of (i). The minimum-modulus principle is applicable because Q_ν is an analytic function without zeros in the sector $0 \leqq \arg z \leqq \pi/2$ if $\nu \in [0,1/2]$; see [13, p. 511].

Statement (ii) follows from (i) and (3.8). To complete the proof of (iii), note that (3.9) and (ii) imply, for $\pi/2 \leqq \arg z \leqq \pi$,

$$|Q_\nu(z)| \geqq |Q_\nu(ze^{-\pi i})|^{-1} - 2\cos(\nu\pi)e^{-2\operatorname{Im} z} \geqq 1 - 2\cos(\nu\pi)e^{-2\operatorname{Im} z}.$$

This completes the proof of the lemma. □

It should be noted that the lower bound for $|Q_\nu(z)|$ appearing in Lemma 3.1 (iii) can be negative. Of course, in such a case the bound is trivial. If $\nu \in (1/3,1/2]$, then $0 \leqq \cos(\nu\pi) < 1/2$ and the lower bound is positive. If $\nu \in [0,1/3]$ we lack a positive lower bound for $|Q_\nu(z)|$ in the sector $\pi/2 \leqq \arg z \leqq \pi$. In the borderline case $\nu = 1/3$, we still have a positive lower bound for $|Q_\nu(z)|$ in $\pi/2 \leqq \arg z < \pi$, $\operatorname{Im} z \geqq \epsilon > 0$ but none for $\arg z = \pi$. In fact, $Q_\nu(z)$ (or, equivalently, $H_\nu^{(2)}(z)$) has zeros in the sector $\pi/2 \leqq \arg z \leqq \pi$ if $\nu \in [0,1/3]$, the zeros lying on the ray $\arg z = \pi$ if $\nu = 1/3$.

LEMMA 3.2. *Let $\nu \in [0,1/2]$. Then the following estimates hold for $\arg Q_\nu(z)$.*

(i) *If $-\pi/2 \leqq \arg z \leqq \pi/2$, then $0 \leqq \arg Q_\nu(z) \leqq \left(\frac{1}{2} - \nu\right)\pi$;*

(ii) *if $\pi/2 \leqq \arg z \leqq \pi$ and $2\cos(\nu\pi)e^{-2\operatorname{Im} z} \leqq 1$, then*

$$-\phi - \left(\frac{1}{2} - \nu\right)\pi \leqq \arg Q_\nu(z) \leqq \phi,$$

where $\phi := \arcsin(2\cos(\nu\pi)e^{-2\operatorname{Im} z})$;

(iii) *if* $\nu \in [1/3, 1/2]$ *and* $\pi/2 \leqq \arg z \leqq \pi$, *then*

$$-4\left(\frac{1}{2}-\nu\right)\pi \leqq \arg Q_\nu(z) \leqq 3\left(\frac{1}{2}-\nu\right)\pi.$$

PROOF. (i) By (3.10), the stated inequality is true for $\arg z = \pi/2$. By (3.8), it is then also true for $\arg z = -\pi/2$. Now (3.3) and a variant of the maximum-modulus principle prove (i).

(ii) Let $\pi/2 \leqq \arg z \leqq \pi$ and $r := 2\cos(\nu\pi)e^{-2\operatorname{Im} z} \leqq 1$. Then, by (3.9), $Q_\nu(z)$ lies on the circle centered at $c = Q_\nu(ze^{-\pi i})^{-1}$ with radius r. By Lemma 3.1 (ii), the center c satisfies $|c| \geqq 1$. By part (i) of this lemma, $-(1/2-\nu)\pi \leqq \arg c \leqq 0$. The information on the radius and center of the circle implies that each point w on the circle satisfies $-\phi-(1/2-\nu)\pi \leqq \arg w \leqq \phi$. This proves (ii).

(iii) Since $2\sin x \leqq \sin(3x)$ for $0 \leqq x \leqq \pi/6$, we have

$$\phi \leqq \arcsin\left(2\sin\left(\left(\frac{1}{2}-\nu\right)\pi\right)\right) \leqq 3\left(\frac{1}{2}-\nu\right)\pi.$$

Now (iii) follows from (ii). □

Using Lemma 3.2 we find another positive lower bound for $|Q_\nu(z)|$.

LEMMA 3.3. *Let* $\nu \in [0, 1/2]$, $0 < \epsilon \leqq \pi/2$ *and* $n \in \mathbb{Z}$. *If* z *satisfies* $\pi/2 \leqq \arg z \leqq \pi$ *and*

$$(n-1)\pi + \frac{1}{4}\pi + \frac{1}{2}\epsilon \leqq \operatorname{Re} z \leqq n\pi + \frac{1}{2}\nu\pi - \frac{1}{2}\epsilon,$$

then $|Q_\nu(z)| \geqq \sin\epsilon$.

PROOF. By (3.9), $Q_\nu(z) = d - c$ with $d := -2i\cos(\nu\pi)e^{2iz}$ and $c := Q_\nu(ze^{-\pi i})^{-1}$. The assumptions on z show that d lies in the sector $\epsilon \leqq \arg d \leqq 2\pi-(1/2-\nu)\pi-\epsilon$. By Lemmas 3.1 and 3.2, c satisfies $-(1/2-\nu)\pi \leqq \arg c \leqq 0$ and $|c| \geqq 1$. It is easy to see that our estimates of c and d imply $|d-c| \geqq \sin\epsilon$. This yields the statement of the lemma. □

4. Estimates of zero-pairs

Let $a, b \in \mathbb{C}_{\log}$ be a zero-pair of Bessel functions of order ν. Then (1.8) and (3.2) imply $e^{-2ia}Q_\nu(a) = e^{-2ib}Q_\nu(b)$. If follows that

$$2\operatorname{Im}(a-b) = \log|Q_\nu(b)| - \log|Q_\nu(a)|, \tag{4.13}$$

$$2\operatorname{Re}(a-b) \equiv \arg Q_\nu(a) - \arg Q_\nu(b) \bmod 2\pi. \tag{4.14}$$

We now apply the results of the previous section to these formulas in order to obtain estimates for $a-b$ if a, b is a zero-pair.

THEOREM 4.1. *Let a, b be a zero-pair of order $\nu \in [0, 1/2]$. Then the following estimates hold for* $\operatorname{Im}(b-a)$.

(i) *If* $\arg a$ *and* $\arg b$ *lie both in* $[-\pi/2, \pi/2]$, *or both in* $[0, \pi]$, *then*

$$|\operatorname{Im}(b-a)| \leqq \frac{1}{2}\log(1+2\cos(\nu\pi));$$

(ii) *if* $\arg a \in [-\pi/2, 0]$, $\arg b \in [\pi/2, \pi]$ *and* $2\cos(\nu\pi)e^{-2\operatorname{Im} b} < 1$, *then*

$$0 \leqq \operatorname{Im}(b-a) \leqq -\frac{1}{2}\log\left(1-2\cos(\nu\pi)e^{-2\operatorname{Im} b}\right).$$

PROOF. We prove (i) by considering several cases.

(1) If $\arg a \in [-\pi/2, 0]$ and $\arg b \in (0, \pi/2]$, then (4.13) and Lemma 3.1 (i) (ii) imply $\operatorname{Im}(a-b) \geqq 0$. This is a contradiction which proves that this case is impossible if $\nu \in [0, 1/2]$.

(2) If $\arg a$, $\arg b \in [0, \pi/2]$, then (4.13) and Lemma 3.1 (i) give statement (i).

(3) If $\arg a \in [0, \pi/2]$, $\arg b \in [\pi/2, \pi]$ and $\operatorname{Im} b \leqq \operatorname{Im} a$, then (4.13) and Lemma 3.1 imply $0 \leqq \operatorname{Im}(a-b) \leqq \frac{1}{2}\log(1+2\cos(\nu\pi))$.

The remaining cases can be reduced to one of the three previous ones by using Lemma 2.5. For instance, if $\arg a \in [0, \pi/2]$, $\arg b \in [\pi/2, \pi]$ and $\operatorname{Im} a \leqq \operatorname{Im} b$, then we apply the result of the third case to $-\overline{b}$, $-\overline{a}$ in place of a and b, respectively. We obtain the desired statement. This completes the proof of (i).

(ii) follows immediately from (4.13) and Lemma 3.1. □

In the proof of Theorem 4.1 we saw that there is no zero-pair a, b of order $\nu \in [0, 1/2]$ with $\arg a \in [-\pi/2, 0]$ and $\arg b \in (0, \pi/2]$. As a corollary, we obtain the result that a Bessel function of order $\nu \in [0, 1/2]$ that is real-valued on $\arg z = 0$ (and thus has zeros in conjugate pairs) is zero-free in the union of the sectors $-\pi/2 \leqq \arg z < 0$ and $0 < \arg z \leqq \pi/2$. This result is due to Schafheitlin [12] (cf. [13, p. 482]) in the case of the Bessel functions Y_0 of the second kind.

Let us give an another application of Theorem 4.1. Consider a zero-pair a, b of order $\nu \in [0, 1/2]$ with $-\pi/2 \leqq \arg a \leqq 0$ and $\pi/2 \leqq \arg b \leqq \pi$. We claim that

$$\operatorname{Im} b \leqq \frac{1}{2}\log(1+2\cos(\nu\pi)).$$

In fact, if this were wrong, then Theorem 4.1 (ii) would imply that $\operatorname{Im}(b-a) < \operatorname{Im} b$ which contradicts $\operatorname{Im} a \leqq 0$. We conclude that a Bessel function of order $\nu \in [0, 1/2]$ that has a zero a in $-\pi/2 \leqq \arg a \leqq 0$ is zero-free in that part of the sector $\pi/2 \leqq \arg z \leqq \pi$ which lies above the line $\operatorname{Im} z = \frac{1}{2}\log(1+2\cos(\nu\pi))$. For example, a Bessel function that is real-valued

for $\arg z = 0$ has a zero a with $\arg a = 0$ (even infinitely many of them) so that this result is applicable. By using Lemma 3.3, we could find other zero-free regions but we do not go into the details here.

THEOREM 4.2. *Let a, b be a zero-pair of order $\nu \in [0, 1/2]$. Then the following estimates hold for* $\operatorname{Re}(a-b)$.

(i) *If* $\arg a, \arg b \in [-\pi/2, \pi/2]$, *then there is* $n \in \mathbb{Z}$ *such that*

$$|\operatorname{Re}(a-b) - n\pi| \leqq \frac{1}{2}\left(\frac{1}{2} - \nu\right)\pi;$$

(ii) *if* $\nu \geqq 1/3$, $\arg a \in [-\pi/2, \pi/2]$ *and* $\arg b \in [\pi/2, \pi]$, *then there is* $n \in \mathbb{Z}$ *such that*

$$-\frac{3}{2}\left(\frac{1}{2} - \nu\right)\pi \leqq \operatorname{Re}(a-b) - n\pi \leqq \frac{5}{2}\left(\frac{1}{2} - \nu\right)\pi.$$

PROOF. Both statements follow directly from (4.14) and Lemma 3.2. □

Let us give an application of Theorem 4.2 to the location of the positive zeros $j_{\nu,n}$ of the Bessel function J_ν. By (4.14), we have

$$j_{\nu,n} - j_{\nu,m} \equiv \frac{1}{2}\arg Q_\nu(j_{\nu,n}) - \frac{1}{2}\arg Q_\nu(j_{\nu,m}) \bmod \pi.$$

For fixed n, we let m go to infinity noting that ([13, p. 509])

$$j_{\nu,m} = m\pi - \frac{1}{4}\pi + \frac{1}{2}\nu\pi + o(1) \quad \text{as} \quad m \to \infty.$$

By (3.3) and Lemma 3.2, we find $k \in \mathbb{Z}$ such that

$$j_{\nu,n} - k\pi + \frac{1}{4}\pi - \frac{1}{2}\nu\pi \in \left[0, \frac{1}{2}\left(\frac{1}{2} - \nu\right)\pi\right] \tag{4.15}$$

if $\nu \in [0, 1/2]$. We know that $j_{\nu,n}$ is a monotonically increasing function of $\nu > -1$; see [13, p. 507]. Therefore,

$$n\pi - \frac{1}{2}\pi = j_{-1/2,n} \leqq j_{\nu,n} \leqq j_{1/2,n} = n\pi.$$

It follows that $k = n$ in (4.15). Hence we have proved that

$$n\pi - \frac{1}{4}\pi + \frac{1}{2}\nu\pi \leqq j_{\nu,n} \leqq n\pi \tag{4.16}$$

for $\nu \in [0, 1/2]$ and $n \in \mathbb{N}$. A related result was proved in a different way by Schafheitlin [11]; cf. [13, p. 490]. In a similar way, we can prove that

$$n\pi - \frac{1}{4}\pi - \frac{1}{2}\nu\pi \leqq j_{-\nu,n} \leqq n\pi - \nu\pi \tag{4.17}$$

for $\nu \in [0, 1/2]$ and $n \in \mathbb{N}$.

In the next section we will need the following estimates of $a - b$ if a, b is a zero-pair with $|\arg a - \arg b| \leqq \pi$.

THEOREM 4.3. *Let a, b be a zero-pair of order $\nu \in [1/3, 1/2]$ and $|\arg a - \arg b| \leqq \pi$. Then the following estimates hold.*

(i) *There is $n \in \mathbb{N}_0$ such that*

$$-\frac{3}{2}\left(\frac{1}{2}-\nu\right)\pi \leqq |\mathrm{Re}\,(a-b)| - n\pi \leqq \frac{5}{2}\left(\frac{1}{2}-\nu\right)\pi;$$

(ii) *if $\nu > 1/3$, then*

$$|\mathrm{Im}\,(a-b)| \leqq -\frac{1}{2}\log(1-2\cos(\nu\pi)).$$

PROOF. Let a, b be a zero-pair with $|\arg a - \arg b| \leqq \pi$. Using Lemma 2.5, it is easy to see that there is another zero-pair c, d with $|\mathrm{Re}\,(c-d)| = |\mathrm{Re}\,(a-b)|$ and $|\mathrm{Im}\,(c-d)| = |\mathrm{Im}\,(a-b)|$ satisfying one of following three statements: 1) $\arg c, \arg d \in [0, \pi/2]$; 2) $\arg c \in [0, \pi/2]$, $\arg d \in [\pi/2, \pi]$; 3) $\arg c \in [-\pi/2, 0]$, $\arg d \in [\pi/2, \pi]$ and $\arg d - \arg c \leqq \pi$. Now statement (i) follows from Theorem 4.2 (i) in case 1) and from part (ii) of the same theorem in the cases 2) and 3). Statement (ii) follows from Theorem 4.1 (i) in the cases 1) and 2) and from part (ii) of the same theorem in case 3). □

We will also need an estimate for $\mathrm{Im}\,(a-b)$ in the borderline case $\nu = 1/3$.

THEOREM 4.4. *Let a, b be a zero-pair of order $\nu = 1/3$ with $|\arg a - \arg b| \leqq \pi$. Then*

$$|\mathrm{Im}\,(a-b)| \leqq \frac{1}{2}\log\left(2 + \frac{2}{\pi}|\mathrm{Re}\,(a-b)|\right).$$

PROOF. As in the proof of Theorem 4.3 we have to consider three cases. In the first two cases, Theorem 4.1 (i) yields $|\mathrm{Im}\,(a-b)| \leqq \frac{1}{2}\log 2$. Therefore, it is sufficient to consider case 3): $\arg a \in [-\pi/2, 0]$, $\arg b \in [\pi/2, \pi]$ and $\arg b - \arg a \leqq \pi$. Then $\mathrm{Im}\,(b-a) \geqq 0$ so that (4.13) and Lemma 3.1 imply

$$0 \leqq \mathrm{Im}\,(b-a) \leqq -\frac{1}{2}\log|Q_{1/3}(b)|. \tag{4.18}$$

If $-\pi/2 \leqq \mathrm{Re}\,b \leqq 0$, then Lemma 3.3 with $n = 0$ and $\epsilon = \pi/3$ gives

$$|Q_{1/3}(b)| \geqq \sin\left(\frac{1}{3}\pi\right) = \frac{1}{2}\sqrt{3} \geqq 1/2.$$

Hence (4.18) shows that $0 \leqq \mathrm{Im}\,(b-a) \leqq \frac{1}{2}\log 2$. Therefore, it is sufficient to consider case 3) under the additional assumption that $\mathrm{Re}\,b \leqq -\pi/2$. Since $\arg b \leqq \arg a + \pi$, we obtain

$$\mathrm{Im}\,(b-a) = \mathrm{Im}\,b + |\mathrm{Im}\,a| \leqq \mathrm{Im}\,b + \frac{\mathrm{Im}\,b}{|\mathrm{Re}\,b|}\,\mathrm{Re}\,a \leqq \left(1 + \frac{2}{\pi}\mathrm{Re}\,a\right)\mathrm{Im}\,b. \tag{4.19}$$

If $\operatorname{Im} b > 0$, Theorem 4.1 yields a second estimate

$$0 \leqq \operatorname{Im}(b-a) \leqq -\frac{1}{2}\log(1-e^{-2\operatorname{Im} b}). \tag{4.20}$$

For abbreviation, let us set $s = 2\operatorname{Im} b > 0$ and $t = 1 + \frac{2}{\pi}\operatorname{Re} a \geqq 1$. Then (4.19) and (4.20) imply

$$0 \leqq \operatorname{Im}(b-a) \leqq \frac{1}{2}\min(st, -\log(1-e^{-s})). \tag{4.21}$$

We claim that, for all $s > 0$ and $t \geqq 1$,

$$\min(st, -\log(1-e^{-s})) \leqq \log(1+t). \tag{4.22}$$

In fact, the substitution $u = e^{-st}$ leads to the equivalent statement $\max(u, 1-u^{1/t}) \geqq (1+t)^{-1}$ for $0 < u < 1$ which is true because $1-u^{1/t} \geqq (1-u)/t$. Now (4.21) and (4.22) imply

$$0 \leqq \operatorname{Im}(b-a) \leqq \frac{1}{2}\log\left(2 + \frac{2}{\pi}\operatorname{Re} a\right).$$

Since $\operatorname{Re} a \leqq |\operatorname{Re}(a-b)|$ this completes the proof. □

5. Proof of the main theorem

We collect all permissible τ's into a set $\mathbb{C}^\star$. Thus $\mathbb{C}^\star$ consists of all $\tau \in \mathbb{C}\setminus[-1,1]$ together with $\tau = \infty$, the boundary points $\tau \pm i0$ for $\tau \in (-1,1)$ and $\tau = \pm 1$. This is a compact space.

THEOREM 5.1. *Let $\nu \in [1/3, 1/2]$. Let (τ, β^2) be an eigenpair of order ν with $\tau \in \mathbb{C}^\star$ and $\operatorname{Re}\beta \geqq 0$. Then there exists $n \in \mathbb{N}_0$ such that*

$$-\frac{3}{4}\left(\frac{1}{2}-\nu\right)\pi \leqq \operatorname{Re}\beta - \frac{1}{2}n\pi \leqq \frac{5}{4}\left(\frac{1}{2}-\nu\right)\pi \tag{5.1}$$

and

$$|\operatorname{Im}\beta| \leqq \begin{cases} -\dfrac{1}{4}\log(1-2\cos(\nu\pi)) & \text{if } \nu > 1/3, \\ \dfrac{1}{4}\log(3+2n) & \text{if } \nu = 1/3. \end{cases} \tag{5.2}$$

PROOF. Let (τ, β^2) be an eigenpair with $\tau \in \mathbb{C}\setminus[-1,1]$ or $\tau = \tau \pm i0$ with $\tau \in (-1,1)$ and $\operatorname{Re}\beta \geqq 0$. We can assume that $\beta \neq 0$. Define $a = \beta(\tau+1)$ and $b = \beta(\tau-1)$. By Lemma 2.4, a, b is a zero-pair of order ν with the indicated

choice of $\arg a$ and $\arg b$. Then $|\arg a - \arg b| \leqq \pi$. By Theorem 4.3 (i), there is $n \in \mathbb{N}_0$ such that

$$-\frac{3}{2}\left(\frac{1}{2}-\nu\right)\pi \leqq |\mathrm{Re}\,(a-b)| - n\pi \leqq \frac{5}{2}\left(\frac{1}{2}-\nu\right)\pi. \tag{5.3}$$

Since $\beta = (a-b)/2$, we obtain (5.1). Similarly, Theorem 4.3 (ii) implies (5.2) if $\nu > 1/3$. If $\nu = 1/3$, we use Theorem 4.4 in combination with (5.3) and obtain (5.2).

If $\tau = \infty$, then $\beta = n\pi/2$ for some $n \in \mathbb{N}$ and the statement is trivially true. If $\tau = \pm 1$, then $\beta = j_{\nu,n}/2$ for some $n \in \mathbb{N}$ and the statement follows from (4.16). □

It is important to note that (5.1), (5.2) define a collection of mutually disjoint rectangles R_n with R_n containing $n\pi/2$. Theorem 5.1 states that the union of the rectangles contains all β with $\mathrm{Re}\,\beta \geqq 0$ for which (τ, β^2) is an eigenpair of order ν for some $\tau \in \mathbb{C}^*$. This is all we need to prove the following main theorem. It is another (less important) task to make these rectangles as small as possible. In Section 6, we show a picture that gives an idea about the quality of the estimates of Theorem 5.1.

THEOREM 5.2. *Let $\nu \in [1/3, 1/2]$. For each $n \in \mathbb{N}$, there is a (uniquely determined) function $\beta_n(\tau)$ which is analytic in $\mathbb{C}\setminus[-1,1]$ and at $\tau = \infty$ with $\beta_n(\infty) = n\pi/2$ such that $(\tau, \beta_n(\tau)^2)$ is an eigenpair of order ν for all τ. This function is also analytic along $\tau \pm i0$ with $\tau \in (-1,1)$, and it can be extended to a continuous function on $\mathbb{C}^*$.*

PROOF. Let $n \in \mathbb{N}$. We recall that the function $D(\tau, \beta^2)$ is analytic for $\tau \in \mathbb{C}^*$, $\tau \neq \pm 1$ and $\beta \in \mathbb{C}$. Let R be a rectangle a little larger than R_n. Using Theorem 5.1, we see that $D(\tau, \beta^2) \neq 0$ along the boundary of the rectangle. If $\tau = \infty$, then $D(\tau, \beta^2)$ has exactly one (with regard to multiplicity) zero within the rectangle. By Rouché's theorem, we find that, for every τ, $D(\tau, \beta^2)$ has exactly one zero $\beta_n(\tau)$ within the rectangle. By the implicit function theorem, this function $\beta_n(\tau)$ is analytic.

We claim that $\beta_n(\tau)$ is continuous at $\tau = \pm 1$ if we set $\beta_n(\pm 1) = j_{\nu,n}/2$. It is sufficient to prove this for $\tau = 1$. We know that the range of β_n is bounded because it is contained in the rectangle R_n. Therefore, in order to prove that $\beta_n(\tau)$ is continuous at $\tau = 1$, it is is enough to show that $1 \neq \tau_k \to 1$, $\beta_n(\tau_k) \to \beta$ implies that $\beta = j_{\nu,n}/2$. Define $a_k = \beta_n(\tau_k)(\tau_k + 1)$ and $b_k = \beta_n(\tau_k)(\tau_k - 1)$. Then a_k, b_k is a zero-pair of order ν if we choose $\arg a_k$ and $\arg b_k$ according to Lemma 2.4. Using appropriate arguments we also have that $a_k \to 2\beta$ and $|b_k| \to 0$ as $k \to \infty$. We now use (1.8) for $a = a_k, b = b_k$ and (3.5). It follows that $H_\nu^{(2)}(2\beta)/H_\nu^{(1)}(2\beta) = -1$. Using (3.4) we conclude that $J_\nu(2\beta) = 0$. Since all zeros of J_ν are real ([13, p. 482]), we find that $2\beta = j_{\nu,m}$ for some $m \in \mathbb{N}$. Since β is in the rectangle R_n, m equals n. The proof is complete. □

The proof shows that the rectangle R_0 does not contain a β for which (τ, β^2) is an eigenpair for some $\tau \in C^\star$. The following two properties of the functions $\beta_n(\tau)$ follow easily.

COROLLARY 5.3. *Let (τ, λ) be an eigenpair of order $\nu \in [1/3, 1/2]$ with $\tau \in \mathbb{C}^\star$ and $\lambda \in \mathbb{C}$. Then there is $n \in \mathbb{N}$ such that $\beta_n(\tau)^2 = \lambda$.*

COROLLARY 5.4. *Let $\nu \in [1/3, 1/2]$ and $n \in \mathbb{N}$. Then, for every $\tau \in \mathbb{C}^\star$, $\beta := \beta_n(\tau)$ lies in the rectangle given by* (5.1) *and* (5.2).

6. The functions $\beta_n(\tau + i0)$

In this section we describe the functions $\beta_n(\tau + i0)$ of Theorem 5.2 for $\tau \in [-1, 1]$ and $\nu \in [1/3, 1/2]$. We know that these functions are analytic in $(-1, 1)$ and continuous in $[-1, 1]$.

We define $\alpha_0 = 0$ and

$$\alpha_k = \begin{cases} j_{-\nu,(k+1)/2} & \text{if } k \text{ is odd} \\ j_{\nu,k/2} & \text{if } k \text{ is even.} \end{cases}$$

The interlacing property of the zeros of J_ν and $J_{-\nu}$ shows that

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \dots . \tag{6.1}$$

For $n \in \mathbb{N}$ and $m = 0, \dots, 2n$, we define

$$\tau_{nm} = \frac{\alpha_m - \alpha_{2n-m}}{\alpha_m + \alpha_{2n-m}}. \tag{6.2}$$

It follows from (6.1) that, for every n,

$$-1 = \tau_{n0} < \tau_{n1} < \dots < \tau_{n,2n-1} < \tau_{n,2n} = 1 \tag{6.3}$$

so that the points τ_{nm}, $m = 0, \dots, 2n$, form a partition of the interval $[-1, 1]$. This partition is symmetric with respect to 0, that is, $\tau_{nm} = -\tau_{n,2n-m}$. In particular, we have $\tau_{nn} = 0$.

We also define

$$\beta_{nm} = \frac{1}{2}(\alpha_m + \alpha_{2n-m}).$$

LEMMA 6.1. *Let $\nu \in (0, 1/2)$. For every $n \in \mathbb{N}$ and $m = 0, 1, \dots, 2n$, $(\tau_{nm} + i0, \beta_{nm}^2)$ is an eigenpair of order ν. There are no other eigenpairs of the form $(\tau + i0, \lambda)$ with $\tau \in [-1, 1]$ and $\lambda > 0$.*

PROOF. If $m = 0$ or $m = 2n$, then $\beta_{nm} = j_{\nu,n}/2$ and $(\tau_{mn}, \beta_{mn}^2)$ is an eigenpair by definition. Let $m = 1, 2, \dots, 2n - 1$. Then

$$\beta_{nm}(\tau_{nm} + 1) = \alpha_m, \quad \beta_{nm}(\tau_{nm} - 1) = -\alpha_{2n-m}$$

which shows that $a := \beta_{nm}(\tau_{nm}+1)$ (with $\arg a = 0$), $b := \beta_{nm}(\tau_{nm}-1)$ (with $\arg b = \pi$) form a zero-pair of order ν. In fact, J_ν or $J_{-\nu}$ vanish at both a and b if m is even or odd, respectively. By Lemma 2.4, it follows that $(\tau_{mn}, \beta_{mn}^2)$ is an eigenpair.

Conversely, assume that $(\tau + i0, \beta^2)$ is an eigenpair with $\tau \in (-1,1)$ and $\beta > 0$. Then $a := \beta(\tau+1)$ (with $\arg a = 0$) and $b := \beta(\tau-1)$ (with $\arg b = \pi$) form a zero-pair. Thus there exists a nontrivial Bessel function C_ν of order ν that vanishes at a and b. We claim that C_ν is a multiple of either J_ν or $J_{-\nu}$. In fact, writing

$$C_\nu(z) = AJ_\nu(z) + BJ_{-\nu}(z),$$

we have

$$AJ_\nu(a) + BJ_{-\nu}(a) = 0, \quad AJ_\nu(b) + BJ_{-\nu}(b) = 0.$$

Since $J_{\pm\nu}(a)$ and $e^{\mp\pi\nu i}J_{\pm\nu}(b)$ are real numbers, we obtain that $B = 0$ or both A/B and $e^{2\pi\nu i}A/B$ are real numbers. Since $0 < \nu < 1/2$, either A or B must vanish. This establishes the claim. If C_ν is a multiple of J_ν, then there are $p, q \in \mathbb{N}$ such that $a = j_{\nu,p}$ and $b = -j_{\nu,q}$. It follows that $\tau = \tau_{nm}$, $\beta = \beta_{nm}$ with $m = 2p$ and $n = p + q$. Using a similar argument in the other case, we complete the proof. □

Let $\nu \in [1/3, 1/2]$. By Corollary 5.3 and Lemma 6.1, for every n, m, there exists $k \in \mathbb{N}$ such that $\beta_k(\tau_{nm}) = \beta_{nm}$. In order to determine k, we use the inequalities (4.16), (4.17) for the zeros $j_{\nu,n}$ and $j_{-\nu,n}$. We obtain

$$\begin{aligned} n\pi - \frac{1}{12}\pi &\leqq j_{\nu,n} \leqq n\pi \\ n\pi - \frac{1}{2}\pi &\leqq j_{-\nu,n} \leqq n\pi - \frac{1}{2}\pi + \frac{1}{6}\pi. \end{aligned}$$

This gives

$$\begin{aligned} \frac{1}{2}k\pi - \frac{1}{12}\pi \leqq \alpha_k \leqq \frac{1}{2}k\pi \qquad & \text{if } k \text{ is even} \\ \frac{1}{2}k\pi \leqq \alpha_k \leqq \frac{1}{2}k\pi + \frac{1}{6}\pi \qquad & \text{if } k \text{ is odd,} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{2}n\pi - \frac{1}{12}\pi \leqq \beta_{nm} \leqq \frac{1}{2}n\pi \qquad & \text{if } m \text{ is even} \\ \frac{1}{2}n\pi \leqq \beta_{nm} \leqq \frac{1}{2}n\pi + \frac{1}{6}\pi \qquad & \text{if } m \text{ is odd.} \end{aligned}$$

These estimates for β_{nm} and Corollary 5.4 imply the following theorem.

THEOREM 6.2. *Let* $\nu \in [1/3, 1/2]$. *Then*

$$\beta_n(\tau_{nm} + i0) = \beta_{nm}$$

for all $n \in \mathbb{N}$ *and* $m = 0, 1, \ldots, 2n$.

If $\nu \in [1/3, 1/2)$, then Lemma 6.1 and Theorem 6.2 show that $\operatorname{Im}\beta_n(\tau + i0)$ has exactly $2n+1$ zeros for $\tau \in [-1, 1]$, namely the numbers (6.3). We now wish to determine the sign of $\operatorname{Im}\beta_n(\tau + i0)$ between these zeros. We first calculate the derivative of $\beta_n(\tau)$.

LEMMA 6.3. *Let* $\nu \in [1/3, 1/2]$, $n \in \mathbb{N}$ *and* $\tau \in \mathbb{C} \setminus [-1, 1]$ *or* $\tau = \tau \pm i0$ *with* $\tau \in (-1, 1)$. *Set* $a := \beta_n(\tau)(\tau + 1)$, $b := \beta_n(\tau)(\tau - 1)$ *with* $\arg a$ *and* $\arg b$ *chosen as in Lemma 2.4. Let* C_ν *be a Bessel function of order* ν *that vanishes at* a *and* b, *and let* D_ν *be a Bessel function of order* ν *that is linearly independent from* C_ν. *Then*

$$\beta_n'(\tau) = \frac{1}{1-\tau^2} \frac{bD_\nu(b)^2 - aD_\nu(a)^2}{D_\nu(b)^2 - D_\nu(a)^2}.$$

PROOF. In the proof we denote the given value of τ by τ_0. Let τ denote a complex variable close to τ_0. Then, by (1.7),

$$C_\nu(\beta_n(\tau)(\tau+1))D_\nu(\beta_n(\tau)(\tau-1)) - C_\nu(\beta_n(\tau)(\tau-1))D_\nu(\beta_n(\tau)(\tau+1)) = 0.$$

We differentiate with respect to τ and set $\tau = \tau_0$. Since $C_\nu(a) = C_\nu(b) = 0$, this gives

$$\begin{aligned} &(\beta_n'(\tau_0)(\tau_0+1) + \beta_n(\tau_0))C_\nu'(a)D_\nu(b) \\ &= (\beta_n'(\tau_0)(\tau_0-1) + \beta_n(\tau_0))C_\nu'(b)D_\nu(a). \end{aligned} \tag{6.4}$$

Now the Wronskian relation

$$z(C_\nu(z)D_\nu'(z) - C_\nu'(z)D_\nu(z)) = \text{const}$$

yields

$$aC_\nu'(a)D_\nu(a) = bC_\nu'(b)D_\nu(b).$$

We use this in (6.4) to eliminate C_ν' and obtain the desired statement. □

We use Lemma 6.3 to calculate $\beta_n'(\tau_{nm})$ for odd m. We choose $D_\nu = J_\nu$ and obtain with $a = \beta_{nm}(\tau_{nm} + 1)$, $b = \beta_{nm}(\tau_{nm} - 1)$

$$\beta_n'(\tau_{nm} + i0) = \frac{1}{1-\tau_{nm}^2} \frac{bJ_\nu(b)^2 - aJ_\nu(a)^2}{J_\nu(b)^2 - J_\nu(a)^2}.$$

Setting $w = J_\nu(b)^2/J_\nu(a)^2$ we find

$$\operatorname{Im}\beta_n'(\tau_{nm} + i0) = (1-\tau_{nm}^2)^{-2}|1-w|^{-2}(a-b)\operatorname{Im} w.$$

Since $J_\nu(a)$ and $e^{-\nu\pi i}J_\nu(b)$ are real numbers, we see that $\operatorname{Im} w > 0$. This shows that $\operatorname{Im}\beta_n'(\tau_{nm} + i0) > 0$ for all odd m. This allows us to determine the sign of $\operatorname{Im}\beta_n(\tau + i0)$ for $\tau \in (-1, 1)$ as stated in the following theorem.

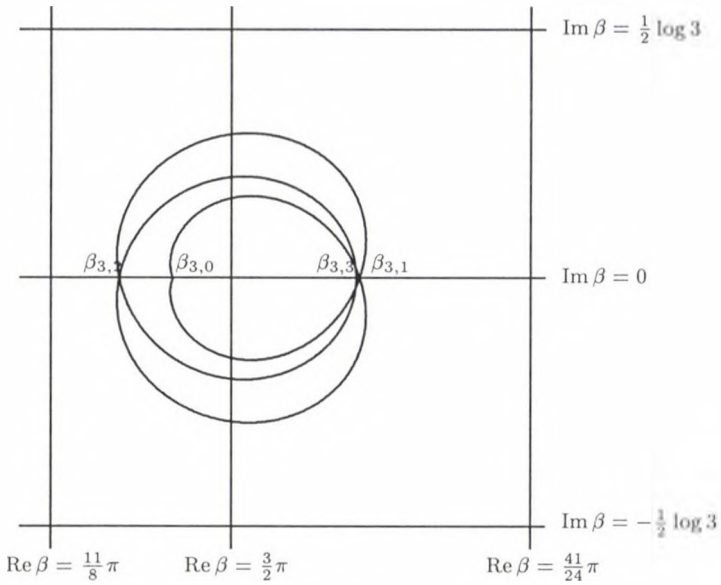


Fig. 1. The graph of $\beta_3(\tau + i0)$, $\tau \in (-1, 1)$, for $\nu = 1/3$

THEOREM 6.4 *Let* $\nu \in [1/3, 1/2)$. *If* m *is odd, then* $\operatorname{Im} \beta_n(\tau + i0) > 0$ *for* $\tau_{nm} < \tau < \tau_{n,m+1}$. *If* m *is even, then* $\operatorname{Im} \beta_n(\tau + i0) < 0$ *for* $\tau_{nm} < \tau < \tau_{n,m+1}$.

Figure 1 illustrates Theorem 6.4 and Corollary 5.4 for $\nu = 1/3$ and $n = 3$. The curve $\beta_3(\tau + i0)$, $\tau \in (-1, 1)$, starts at the indicated point $\beta_{3,0}$. The first part of the curve (for $\tau \in (\tau_{3,0}, \tau_{3,1})$) lies in the half-plane $\operatorname{Im} \beta < 0$. The large rectangle shows the bounds for $\beta_3(\tau)$ as stated in Corollary 5.4. The graph of $\beta_3(\tau + i0)$ was computed by solving the equation

$$J_\nu(a)J_{-\nu}(b) - J_\nu(b)J_{-\nu}(a) = 0, \quad a = \beta(\tau + 1), \quad b = \beta(\tau - 1)$$

for β by Newton's method. The Bessel functions $J_{\pm\nu}$ were computed by rational approximation; see [8].

REFERENCES

[1] BREEN, S., Uniform and lower bounds on the zeros of Bessel functions of the first kind, *J. Math. Anal. Appl.* **196** (1995), 1–17. *MR* **96k**:33004

[2] BENDER, C. and ORSZAG, S., *Advanced mathematical methods for scientists and engineers*, International Series in Pure and Applied Mathematics, McGraw-Hill, New York, 1978. *MR* **80d**:00030

[3] COURANT, R. and HILBERT, D., *Methods of mathematical physics*, vol. 1, Interscience Publishers, New York, 1953. *MR* **16**, 426a

[4] ELBERT, Á., Concavity of the zeros of Bessel functions, *Studia Sci. Math. Hungar.* **12** (1977), 81–88. *MR* **81d**:33004

[5] ELBERT, Á., An approximation for the zeros of Bessel functions, *Numer. Math.* **59** (1991), 647–657. *MR* **92h**:33008

[6] ELBERT, Á., LAFORGIA, A. and LORCH, L., Additional monotonicity properties of the zeros of Bessel functions, *Analysis* **11** (1991), 293–299. *MR* **93a**:33006

[7] KATO, T., *Perturbation theory for linear operators*, 2nd edition, Springer-Verlag, Berlin – New York, 1976. *MR* **53** #11389

[8] LUKE, Y., *The special functions and their approximations*, vols. 1–2, Academic Press, New York – London, 1969. *MR* **39** #3039, **40** #2909

[9] MEIXNER, J. and SCHÄFKE, F. W., *Mathieusche Funktionen und Sphäroidfunktionen mit Anwendungen auf physikalische und technische Probleme*, Die Grundlehren der mathematischen Wissenschaften, Band 71, Springer-Verlag, Berlin – Göttingen – Heidelberg, 1954. *MR* **16**, 586g

[10] OLVER, F. W. J., *Asymptotics and special functions*, Computer Science and Applied Mathematics, Academic Press, New York – London, 1974. *MR* **55** #8655

[11] SCHAFHEITLIN, P., Ueber die Gausssche und Besselsche Differentialgleichung und eine neue Integralform der letzteren, *J. für Math.* **114** (1894), 31–44. *Jb. Fortschritte Math.* **25**, 839

[12] SCHAFHEITLIN, P., Über die Nullstellen der Besselschen Funktionen zweiter Art, *Arch. Math. Phys.* **1** (1901), 133–137. *Jb. Fortschritte Math.* **32**, 466

[13] WATSON, G. N., *A treatise on the theory of Bessel functions*, Cambridge Univ. Press, Cambridge, 1944. *MR* **6**, 64a

(Received February 12, 1996)

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF WISCONSIN–MILWAUKEE
P.O. BOX 413
MILWAUKEE, WI 53201
U.S.A.

volkmer@csd.uwm.edu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 281–290

OPTIMAL PACKINGS OF UNIT SQUARES IN A SQUARE

S. EL MOUMNI

Abstract

Let $s(n)$ denote the edge length of the smallest square in which one can pack n unit squares whose interiors are pairwise disjoint. We prove that $s(7)=3$ and $s(15)=4$.

1. Introduction

In this note we determine, for $n=7$ and $n=15$, the edge length $s(n)$ of the smallest square in which one can pack n unit squares whose interiors are pairwise disjoint.

In 1975, P. Erdős and R. Graham [1] proved a remarkable theorem: If we denote by $m(z)$ the maximum number of unit squares that one can pack in a square of side z, and if $w(z)=z^2-m(z)$, then $w(z)=O(z^{7/11})$ $(7/11\cong 0.636)$. According to M. Gardner [2], H. Montgomery has improved this asymptotic result slightly, by proving that $w(z)=O(z^{\frac{(3-\sqrt{3})}{2}})$ $\left(\dfrac{3-\sqrt{3}}{2}\cong 0.633\right)$. In 1978, K. Roth and R. Vaughan [4] showed that $w(z)\geqq 10^{-100}(\|z\|z)^{1/2}$, where $\|z\|=\inf(|z-\lfloor z\rfloor|,|z-\lfloor z\rfloor-1|)$. F. Göbel remarked in [3] that, apart from the trivial result $s(k^2)=k$ for every $k\in\mathbb{N}_0$, the only values of n for which $s(n)$ is known are $n=2,3,5$ $(s(2)=2,\ s(3)=2,\ s(5)=2+\frac{\sqrt{2}}{2})$, and that E. Bajmoczy in Budapest established that $s(7)=3$, but the proof of this result has apparently never been published.

We are going to prove the following results:

THEOREM 1. $s(7)=3$.

THEOREM 2. $s(15)=4$.

1991 *Mathematics Subject Classification.* Primary 52C15.
Key words and phrases. Packing, squares.

2. Proof of Theorem 1

We first prove the following propositions:

PROPOSITION 1. *If we pack a unit square C_1 in a square C whose edge length is less than 2, then the center of C belongs necessarily to the interior of C_1.*

PROOF. Suppose, on the contrary, that we can pack a unit square C_1 in a square $C=(abcd)$ of side $2-2\epsilon$ $(\epsilon>0)$ in such a way that the center o of C is not in the interior $\overset{\circ}{C}_1$ of C_1. We denote by a', b', c', d', respectively, the midpoint of $[a,b]$, $[b,c]$, $[c,d]$, $[d,a]$. If $\overset{\circ}{C}_1$ intersects each of the open squares $(aa'od')$, $(a'bb'o)$, $(ob'cc')$, $(d'oc'd)$, then $o\in\overset{\circ}{C}_1$, a contradiction. Up to a symmetry of the square $(abcd)$, we may assume that $\overset{\circ}{C}_1\cap(\overline{ob'cc'})=\phi$. Denote by $C_2=(oefg)$ the unit square such that

$$[o,b']\subset[o,e] \quad \text{and} \quad [o,c']\subset[o,g].$$

Thus we have packed two unit squares C_1 and C_2 in a square of side $2-\epsilon$, contradicting the fact that $s(2)=2$.

PROPOSITION 2. *If we draw infinitely many parallel lines Δ_i $(i\in\mathbb{Z})$ in the euclidean plane E^2 in such a way that the distance between any two consecutive lines is a constant d satisfying $\frac{\sqrt{2}}{2}\leqq d<1$, then for any unit square C,*

$$\sum_{i\in\mathbb{Z}}|\overset{\circ}{C}\cap\Delta_i|\geqq\inf\left(2\left(\sqrt{2}-d\right),1\right).$$

PROOF. Since $d<1$, we have $\overset{\circ}{C}\cap(\bigcup_{i\in\mathbb{Z}}\Delta_i)\neq\phi$. We now distinguish two cases:

Case 1. There exists $j\in\mathbb{Z}$ such that Δ_j intersects $\overset{\circ}{C}$ in a segment $]p_j,q_j[$, where p_j and q_j belong to two opposite sides of C. Then $|\overset{\circ}{C}\cap\Delta_j|\geqq 1$, and so

$$\sum_{i\in\mathbb{Z}}|\overset{\circ}{C}\cap\Delta_j|\geqq\inf\left(2\left(\sqrt{2}-d\right),1\right).$$

Case 2. For each $i\in\mathbb{Z}$, if $\Delta_i\cap\overset{\circ}{C}\neq\phi$ and $\Delta_i\cap C=[p_i,q_i]$, then p_i and q_i belong to two consecutive sides of C. Consider the following two subcases:

Subcase 1. $\overset{\circ}{C}$ intersects only one Δ_j.

Subcase 2. $\overset{\circ}{C}$ intersects two consecutive lines Δ_j and Δ_{j+1} (it is impossible for $\overset{\circ}{C}$ to intersect three consecutive lines, because the diameter of C is $\sqrt{2}$ and $2d\geqq\sqrt{2}$).

Let s_1, s_2, s_3, s_4 be the vertices of C. We have $\Delta_j \cap C = [p_j, q_j]$, where p_j and q_j belong to two consecutive sides of C. Denote by s_3 the vertex common to these two consecutive sides, by s_1 the vertex opposite to s_3 (as in Figure 2), by α the angle between one of the sides of C containing s_1 and the line parallel to Δ_j passing through s_1 $(0 < \alpha < \pi/2)$.

Study of the 1st subcase. Let $d' = d(s_1, \Delta_j)$. We have $|p_j q_j| = \dfrac{\sin\alpha + \cos\alpha - d'}{\sin\alpha\cos\alpha}$, and so $|p_j q_j| \geqq 2\left(\sqrt{2} - d'\right) \geqq 2\left(\sqrt{2} - d\right)$.

Study of the 2nd subcase. We have

$$|p_j q_j| + |p_{j+1} q_{j+1}| = \frac{\sin\alpha + \cos\alpha - d}{\sin\alpha\cos\alpha}.$$

We conclude that in both subcases

$$\sum_{i\in\mathbb{Z}} |\overset{\circ}{C} \cap \Delta_i| \geqq 2(\sqrt{2} - d) \geqq \inf(1, 2(\sqrt{2} - d)).$$

PROOF OF THEOREM 1. Suppose that one can pack 7 unit squares in a square $(abef)$ of side $3-\alpha$ $(\alpha > 0)$. Since $s(5) = 2 + \frac{\sqrt{2}}{2}$, we have $3 - \alpha \geqq 2 + \frac{\sqrt{2}}{2}$, and so $\alpha \leqq 1 - \frac{\sqrt{2}}{2}$.

We dissect $(abef)$ into 9 little squares of side $1 - \epsilon$ (where $\alpha = 3\epsilon$), as shown in Figure 1.

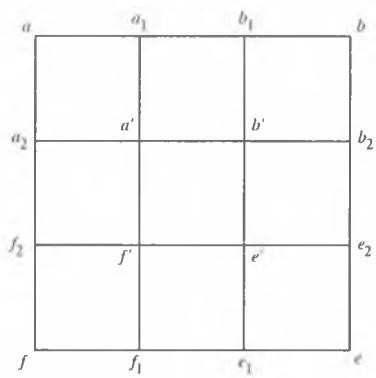


Fig. 1

Let $\mathcal{P}$ be a packing of 7 unit squares in the square $(abef)$. Note first that there are at least 3 unit squares whose interior does not contain any of the 4 points a', b', e', f', otherwise (by the pigeonhole principle) there would exist 2 unit squares whose interior contains one of the points a', b', e', f', contradicting the fact that the interiors of the 7 squares are pairwise disjoint.

Call C_1, C_2, C_3 these 3 unit squares. We will prove that the center o_i of C_i $(i=1,2,3)$ belongs necessarily to the union of the open rectangles $(a_1b_1e_1f_1)$ and $(a_2b_2e_2f_2)$. In order to do this, we will prove that, if it is not the case, then one of the points a', b', e', f' belongs necessarily to the interior of C_i. Indeed, up to a symmetry of the square, we may assume that o_i belongs to the closed square $(aa_1a'a_2)$. We have $C_i \subset (ab_1e'f_2)$ (of side $2-2\epsilon$) and, by Proposition 1, $a' \in \overset{\circ}{C}_i$.

Thus we have shown that the centers o_1, o_2, o_3 of the 3 squares C_1, C_2, C_3 belong to $(a_1b_1b'a') \cup (a'b'e'f') \cup (b'b_2e_2e') \cup (e'e_1f_1f) \cup (a_2a'f'f_2)$. However, two of the centers o_1, o_2, o_3 cannot belong simultaneously to one of these 5 squares of side $1-\epsilon$, otherwise there would be two centers o_i and o_j at distance $\geqq 1$ belonging to a square $(s_1s_2s_3s_4)$ of side $(1-\epsilon)$, where $s_1, s_2, s_3, s_4 \notin \overset{\circ}{C}_i \cup \overset{\circ}{C}_j$. Since the open disc $\overset{\circ}{D}(o_i, 1/2)$ of center o_i and radius $1/2$ is contained in $\overset{\circ}{C}_i$ and since $\overset{\circ}{D}(o_j, 1/2) \subset \overset{\circ}{C}_i$, we would have

$$\{o_i, o_j\} \cap \bigcup_{k=1}^{4} D(s_k, 1/2) = \phi,$$

and so o_i and o_j would belong to the shaded portion of Figure 2.

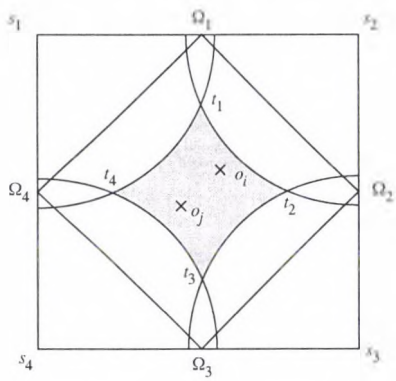


Fig. 2

Let r_1, r_2, r_3, r_4 be the midpoints of the sides of the square $(s_1s_2s_3s_4)$. We have $o_i \in (r_1r_2r_3r_4)$ and $o_j \in (r_1r_2r_3r_4)$. But the diameter of the square $(r_1r_2r_3r_4)$ is $1-\epsilon$. Therefore $|o_io_j| < 1$, and so $\overset{\circ}{C}_i \cap \overset{\circ}{C}_j \neq \phi$, a contradiction.

Let o_i be the center of the square C_i $(i=1,\ldots,7)$. It follows from the preceding arguments that any distribution of the centers of the 3 squares C_1, C_2, C_3 is equivalent (up to a symmetry of the large square) to one of the 3 cases represented in Figure 3.

Case 1. We first prove that $|\overset{\circ}{C}_1 \cap [a_2b_2]| = |\overset{\circ}{C}_1 \cap [a'b']|$. We have $D(o_1, 1/2) \subset C_1 \subset (abef)$, thus $d(o_1, [a_1, b_1]) \geqq 1/2$. On the other hand, $d([a_1, b_1], [a', b']) =$

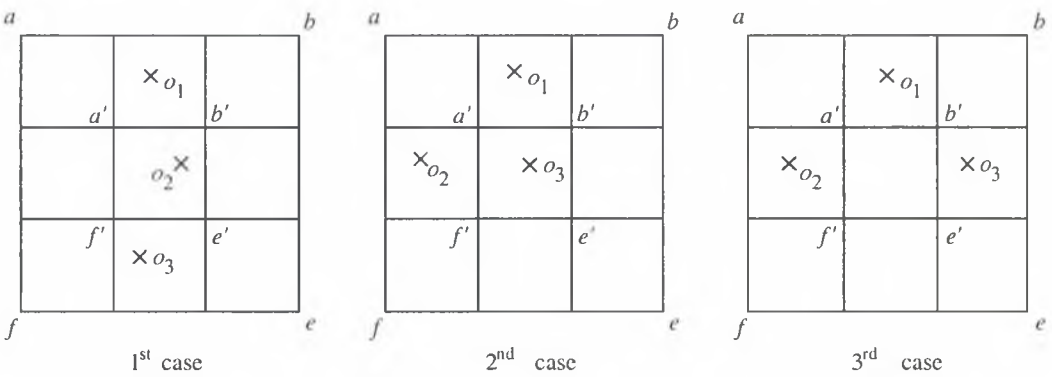


Fig. 3

$1-\epsilon<1$. It follows that $d(o_1,[a',b'])<1/2$. Therefore $\overset{\circ}{D}(o_1,1/2)\cap[a',b']\neq\phi$, and so $\overset{\circ}{C}_1\cap[a',b']\neq\phi$. The fact that $a'\notin\overset{\circ}{C}_1$ and $b'\notin\overset{\circ}{C}_1$, together with the convexity of $\overset{\circ}{C}_1$, imply that

$$\overset{\circ}{C}_1\cap[a_2,b_2]=\overset{\circ}{C}_1\cap[a',b'].$$

A similar argument shows that

$$\overset{\circ}{C}_3\cap[a_2,f_2]=\overset{\circ}{C}_3\cap[e',f'].$$

By Proposition 2,

$$|\overset{\circ}{C}_1\cap[a',b']|\geqq 2\sqrt{2}-2(1-\varepsilon)$$

and

$$|\overset{\circ}{C}_3\cap[e',f']|\geqq 2\sqrt{2}-2(1-\varepsilon).$$

If $\overset{\circ}{C}_2\cap[a',b']\neq\phi$, then $\overset{\circ}{C}_2\cap[a',b']$ is a segment whose endpoints belong to two neighbourly opposite sides of C_2 (otherwise $|a'b'|\geqq 1$, a contradiction). The same holds for $\overset{\circ}{C}_2\cap[b',e']$, $\overset{\circ}{C}_2\cap[e',f']$, $\overset{\circ}{C}_2\cap[f',a']$ and, by the convexity of $\overset{\circ}{C}_2$, we have

$$\overset{\circ}{C}_2\cap[a_2,b_2]=\overset{\circ}{C}_2\cap[a',b'] \quad\text{and}\quad \overset{\circ}{C}_2\cap[e_2,f_2]=\overset{\circ}{C}_2\cap[e',f'].$$

By Proposition 2,

$$|\overset{\circ}{C}_2\cap[a',b']|+|\overset{\circ}{C}_2\cap[e',f']|\geqq 2\sqrt{2}-2(1-\epsilon).$$

Therefore

$$\sum_{i=1}^{3} |\overset{\circ}{C}_i \cap [a', b']| + |\overset{\circ}{C}_i \cap [e', f']| \leqq |a'b'| + |e'f'|,$$

from which we deduce that $3\left(2\sqrt{2} - 2 + 2\epsilon\right) \leqq 2 - 2\epsilon$, that is $6\sqrt{2} < 8$, a contradiction. We conclude that this first case is impossible.

Cases 2 and 3. In each of these cases, the centers o_1 and o_2 of the squares C_1 and C_2 belong to the open squares $(ab_1b'a')$ and $(a_2a'f'f_2)$ of side $1 - \epsilon$ (the center o_i $(i = 1, 2)$ cannot belong to the segments $[a_1, f_1]$, $[b_1, e_1]$, $[a_2, b_2]$, $[f_2, e_2]$), otherwise one of the points a', b', e', f' would be in $\overset{\circ}{C}_i$, a contradiction.

In the same way as in the 1st case, we have

$$|\overset{\circ}{C}_1 \cap [a', b']| \geqq 2\sqrt{2} - 2(1 - \epsilon).$$

On the other hand, $\overset{\circ}{C}_1 \cap [a_1, a'] \neq \phi$, otherwise C_1 would be contained in the square (a_1be_2f') of side $2 - 2\epsilon$ and, by Proposition 1, $b' \in \overset{\circ}{C}_1$, contradicting the assumption that the interior of C_1 does not contain any of the 4 points a', b', e', f'.

Similarly, we have

$$|\overset{\circ}{C}_2 \cap [a', f']| \geqq 2\sqrt{2} - 2(1 - \epsilon) \quad \text{and} \quad \overset{\circ}{C}_2 \cap [a', a_2] \neq \phi.$$

Consider the points p, q, r, s, t, u such that

$$\begin{aligned} p, q \in [a', b'], \qquad & [pb'] = |a'q| = 2\sqrt{2} - 2 \\ r, s \in [a', f'], \qquad & |rf'| = |a's| = 2\sqrt{2} - 2 \\ t \in \overset{\circ}{C}_1 \cap [a_1, a'], \qquad & u \in \overset{\circ}{C}_2 \cap [a_2, a']. \end{aligned}$$

The points p, q, t belong to $\overset{\circ}{C}_1$. Similarly, r, s, u belong to $\overset{\circ}{C}_2$. The situation is summarized in Figure 4.

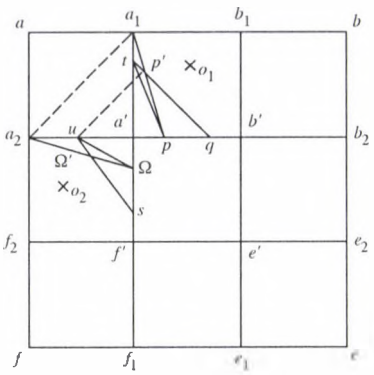


Fig. 4

By the convexity of $\overset{\circ}{C}_1$ and $\overset{\circ}{C}_2$, the triangles $(pqt) \subset \overset{\circ}{C}_1$ and $(rsu) \subset \overset{\circ}{C}_2$. Suppose that there exists a square C_4 such that $a' \in \overset{\circ}{C}_4$. Thus we have $\overset{\circ}{C}_4 \cap (pqt) = \phi$ and $\overset{\circ}{C}_4 \cap (rsu) = \phi$, and we deduce that C_4 is contained in the polygon $(aa_1pb_2ef_1ra_2)$. Consider the points p', r' such that

$$p' \in [a_1, p],\ r' \in [a_2, r],\ |p'r'| = 1 \text{ and } p'r' // a_1a_2.$$

We distinguish two cases:

Case 1. The center o_4 of C_4 belongs to the pentagon (aa_1pra_2). Now $d(o_4, [a_1, p]) \geqq \frac{1}{2}$, $d(o_4, [a_2, r]) \geqq \frac{1}{2}$, $|p'r'| = 1$, $|pr| = 2\sqrt{2} - 4 - \epsilon < 1$. We deduce that o_4 does not belong to the polygon $(p'prr')$, otherwise by drawing the line Δ parallel to $r'p'$ and passing through o_1, we would have $\Delta \cap [p', p] = p''$ and $\Delta \cap [r', r] = r''$, and so $|p''r''| \geqq 1$, contradicting the fact that $|pr| < |p''r''| < |p'r'| = 1$. Therefore $o_4 \in (aa_1p'r'a_2)$.

But two vertices of C_4 cannot belong to the open polygon $(r'p'b_2ef_1)$ because in this case the other two vertices would belong to $(aa_1p'r'a_2)$, therefore the intersection of C_4 and $[p', r']$ would be a segment whose endpoints belong to two opposite sides of C_4, and this implies $|p'r'| > 1$, a contradiction.

Hence there are three vertices of C_4 belonging to $(aa_1p'r'a_2)$, from which we deduce that there are two vertices of C_4 at distance $\sqrt{2}$ and belonging to the pentagon $(aa_1p'r'a_2)$.

We are going to prove that the diameter of the pentagon $(aa_1p'r'a_2)$ is less than $\sqrt{2}$, which yields a contradiction. For this it suffices to prove that

$$|ir'| = |ip'| < \frac{\sqrt{2}}{2}.$$

But

$$|ip'|^2 = \left(\left(\frac{2 - \sqrt{2} - \epsilon}{\sqrt{2} - 1} \right) \left(\frac{\sqrt{2}(1 - \epsilon) - 1}{2} \right) \right)^2 + \left(\frac{1}{2} \right)^2,$$

and so we have $|ir'| = |ip'| < \dfrac{\sqrt{2}}{2}$.

Case 2. The center o_4 of C_4 belongs to the pentagon $(rpb'e'f')$. There is a vertex s_1 of C_4 such that $s_1 \notin (rp'e'f')$ because $a' \in \overset{\circ}{C}_4$. The other vertices cannot belong to $(r'p'pr)$. We denote by s_2 and s_3 the vertices of C_4 adjacent to s_1. We have

$$\widehat{s_2 s_1 s_3} = \pi/2.$$

But

$$\widehat{s_2 a' s_3} > \widehat{s_2 s_1 s_3} = \pi/2 \quad \text{and} \quad \widehat{ra'p} \geqq \widehat{s_2 a' s_3} > \pi/2,$$

a contradiction.

Therefore a' is not in the interior of any of the other 4 squares. In conclusion:

If C_1 and C_2 are packed as in the 2nd case, then a' cannot belong to the interior of any of the other 4 squares packed in $(abef)$, therefore there is another square C_4 whose interior does not contain any of the points a', b', e', f', and so we have necessarily a packing of three unit squares equivalent to the 1st case, which is impossible.

If we are in 3rd case, the above arguments imply that a' and b' cannot belong to the interior of any of the other 4 unit squares packed in $(abef)$, thus there are two other squares C_4 and C_5 whose interior does not contain any of the points a', b', e', f', and so we have necessarily a packing equivalent to the 1st case, which is impossible.

We conclude that it is impossible to pack 7 unit squares in a square of side $3 - \alpha$ (with $\alpha > 0$).

On the other hand, one can clearly pack 7 unit squares in a square of side 3.

3. Proof of Theorem 2

We first prove the following proposition:

PROPOSITION 3. *If n is an integer $\geqq 3$ and if $s(n) \neq \lceil \sqrt{n} \rceil$, then*

$$s(n) \geqq \inf\left(\frac{n}{\lfloor \sqrt{n} \rfloor}, \frac{2\sqrt{2}n}{\dfrac{2n}{\lceil \sqrt{n} \rceil} + \lfloor \sqrt{n} \rfloor} \right).$$

PROOF. For any integer $n \geqq 3$, we have $\sqrt{n} \leqq s(n) \leqq \lceil \sqrt{n} \rceil$ (indeed, $n \leqq s(n)^2 \leqq \lceil \sqrt{n} \rceil^2$).

Let $\mathcal{P}$ be a packing of n unit squares C_i $(i=1,\dots,n)$ in a square $(abef)$ of side $s(n)$, and let a_j, b_j $(j=1,\dots,\lfloor\sqrt{n}\rfloor)$ be the points such that

$$|aa_1| = |a_{\lfloor\sqrt{n}\rfloor} f| = |a_j a_{j+1}| = \frac{s(n)}{\lceil\sqrt{n}\rceil},$$

and

$$|bb_1| = |b_{\lfloor\sqrt{n}\rfloor} e| = |b_j b_{j+1}| = \frac{s(n)}{\lceil\sqrt{n}\rceil} \quad (j=1,\dots,\lfloor\sqrt{n}\rfloor - 1).$$

If $s(n) < \lceil\sqrt{n}\rceil$, Proposition 2 shows that for every C_i

$$\sum_{j=1}^{\lfloor\sqrt{n}\rfloor} |C_i \cap [a_j, b_j]| \geqq \inf\left(2\left(\sqrt{2} - \frac{s(n)}{\lceil\sqrt{n}\rceil}\right), 1\right).$$

Moreover,

$$\sum_{i=1}^{n} \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} |C_i \cap [a_j, b_j]| \leqq \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} |a_j b_j| = \lfloor\sqrt{n}\rfloor(n).$$

Thus

$$n \inf\left(2\left(\sqrt{2} - \frac{s(n)}{\lceil\sqrt{n}\rceil}\right), 1\right) \leqq \lfloor\sqrt{n}\rfloor s(n),$$

and so

$$n \leqq \lfloor\sqrt{n}\rfloor s(n) \quad \text{or} \quad 2n\left(\sqrt{2} - \frac{s(n)}{\lceil\sqrt{n}\rceil}\right) \leqq \lfloor\sqrt{n}\rfloor s(n).$$

Therefore

$$s(n) \geqq \frac{n}{\lfloor\sqrt{n}\rfloor} \quad \text{or} \quad s(n) \geqq \frac{2\sqrt{2}n}{\dfrac{2n}{\lceil\sqrt{n}\rceil} + \lfloor\sqrt{n}\rfloor}.$$

We deduce that

$$s(n) \geqq \inf\left(\frac{n}{\lfloor\sqrt{n}\rfloor}, \frac{2\sqrt{2}n}{\dfrac{2n}{\lceil\sqrt{n}\rceil} + \lfloor\sqrt{n}\rfloor}\right).$$

PROOF OF THEOREM 2. Suppose that $s(15) \neq 4$. Then, by Proposition 3, $s(15) \geqq \inf\left(5, \dfrac{20\sqrt{2}}{7}\right) > 4$, contradicting the fact that $s(15) \leqq 4$. Therefore $s(15) = 4$.

In the same way, we can show that $s(8) = 3$ without using Theorem 1.

ACKNOWLEDGEMENT. I would like to thank Professor Jean Doyen for his encouragements throughout the preparation of this paper.

REFERENCES

[1] ERDŐS, P. and GRAHAM, R. L., On packing squares with equal squares, *J. Combinatorial Theory Ser. A* **19** (1975), 119–123. *MR* **51** #6595

[2] GARDNER, M., Some packing problems that cannot be solved by sitting on the suitcase, *Scientific American* **241** (1979), No. 4, 22–26.

[3] GÖBEL, F., Geometrical packing and covering problems, *Packing and covering in combinatorics*, A. Schrijver ed., Math. Centrum Tracts **106**, Mathematisch Centrum, Amsterdam, 1979, 179–199. *MR* **81b**:05001

[4] ROTH, K. F. and VAUGHAN, R. C., Inefficiency in packing squares with unit squares, *J. Combinatorial Theory Ser. A* **24** (1978), 170–186. *MR* **58** #7407

(Received January 24, 1996)

UNIVERSITÉ LIBRE DE BRUXELLES
DÉPARTEMENT DE MATHÉMATIQUE
CAMPUS PLAINE C.P. 216
BD DU TRIOMPHE
B–1050 BRUXELLES
BELGIUM

smoumni@cso.ulb.ac.be

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 291–330

ON THE EXCEPTIONAL SET FOR THE SUM OF A PRIME AND THE k-TH POWER OF A PRIME

C. BAUER

1. Introduction

It is well known from the work of Montgomery and Vaughan that the exceptional set $E(x)$ for the binary Goldbach conjecture, i.e. the set of even numbers not larger than a real number x which are not representable as the sum of two primes, can be estimated by $E(x) \ll x^{1-\delta}$ for a $\delta > 0$. Brünner, Perelli, Pintz [1] and later Zaccagnini [14] applied the method of Montgomery and Vaughan to the problem of the representation of a positive integer as the sum of a prime and the k-th power of a natural number. They obtained an estimate for the corresponding exceptional set comparable to the one of Montgomery and Vaughan. In this paper we improve, for even integers satisfying certain congruence conditions, upon their result by giving the following theorem:

THEOREM. *Let*

$$E_k(x) = |n : n \leqq x,\ 2|n,\ n \not\equiv 1 (\mathrm{mod}\, p)\ \forall p > 2 \text{ with } p-1|k,\ n \neq p_1 + p_2^k\ \forall p_1, p_2 \in P|, \tag{1.1}$$

where P denotes the set of primes. Then there exists an effectively computable constant $\Theta = \Theta(k)$ such that

$$E_k(x) \ll_k x^{1-\Theta}.$$

After this article had been written, the author became aware that in a still unpublished work Liu and Shung [7] have also proved the above theorem. Even though both works are based on the circle method, we feel that our work

1991 *Mathematics Subject Classification.* Primary 11L07; Secondary 11P32.
Key words and phrases. Additive prime number theory, Goldbach conjecture.

This article forms a part of the author's doctoral dissertation submitted to Professor Dr. D. Wolke from the Department of Mathematics at the University of Freiburg, Germany.

During the preparation of this article the author was holding a common scholarship by the Chinese State Education Commission and the German Academic Exchange Service (DAAD).

is still of interest because our method differs essentially from the method used in [7]. We basically apply the method of [1] and [14] to our problem, whereas Liu and Shung use a method developed in [8]. Where we appeal to the lemmas 4.6–4.9 in order to calculate the contribution of the intervals over the major arcs, Liu and Shung apply a completely different technique of the Lemmas 3.1 to 3.4 in [8]. Furthermore, in their Lemma 4.6 they make use of Jordan's theorem on Dirichlet's integral which makes it necessary to extend the integration over the major arcs to infinity. Here, instead, we proceed differently by calculating precisely the effect of the *P-excluded* zeros (defined below).

2. Notation

To a certain extent we follow the notation and the structure of the proof in [14]. We define: $e(x)=e^{2\pi ix}$; x is a sufficiently large real number, p denotes a prime number, $s=\sigma+it$ is a complex number, $\varrho=\beta+i\gamma$ denotes the generic zeros of the L-functions. By $\chi(=\chi_q), \chi^*(=\chi_q^*), \chi_0(=\chi_{0,q})$ we denote a character, a primitive character and a principal character (modulo q), respectively, whereas $\chi \bmod q \longleftrightarrow \chi^* \bmod q^*$ indicates that the character χ is induced by the primitive character χ^* with $q^*|q$; cond χ =conductor of χ. We denote the Möbius function by $\mu(n)$, the Euler function by $\phi(n)$, the number of prime divisors of n by $\omega(n)$, the divisor function by $\tau(n)$, the cardinality of a set A by $|A|$ and the greatest common divisor and the smallest common multiple of the integers a and b by (a,b) and $[a,b]$, respectively. P is the set of prime numbers and for any integer $l \geqq 1$ we define

$$S_l(\alpha)=\sum_{\frac{\sqrt[l]{x}}{2}\leqq p<\sqrt[l]{x}} \log p\, e(\alpha p^l), \qquad S_l(\chi,\alpha)=\sum_{\frac{\sqrt[l]{x}}{2}\leqq p<\sqrt[l]{x}} \chi(p)\log p\, e(\alpha p^l),$$

$$T_\varrho(\alpha)=\sum_{\frac{x}{2}\leqq m<x} m^{\varrho-1}e(m\alpha), \qquad T(\alpha)=T_1(\alpha),$$

and for a fixed $k \geqq 2$ we define

$$F_\varrho(\alpha)=\sum_{\frac{\sqrt[k]{x}}{2}\leqq m<\sqrt[k]{x}} m^{\varrho-1}e(m^k\alpha), \qquad F(\alpha)=F_1(\alpha).$$

$$\sum_{\substack{\chi \bmod q\\ \chi \text{ primitiv}}} = \sum_{\chi \bmod q}{}^{*}, \qquad \sum_{\substack{a=1\\(a,q)=1}}^{q} = \sum_{a=1}^{q}{}^{*}, \qquad \sum_{a\leqq n\leqq b} u_n = \sum_{a}^{b} u_n,$$

$$C_l(\chi,a)=\sum_{m=1}^{q}{}^{*}\chi(m)e\left(\frac{m^l a}{q}\right), \qquad C_1(\chi,1)=\tau(\chi),$$

for a character χ modulo q.

$$A(q,n,\chi_1,\chi_2)=\sum_{a=1}^{q}{}^{*}C_1(\chi_1,a)C_k(\chi_2,a)e\left(\frac{-an}{q}\right),$$

for characters χ_1 and χ_2 modulo q.

$$A(q,n,\chi_{0,q},\chi_{0,q})=A(q,n),\qquad r(x,n)=\sum_{\substack{p_1+p_2^k=n\\ \frac{x}{2}\leqq p_1<x\\ \frac{\sqrt[k]{x}}{2}\leqq p_2<\sqrt[k]{x}}}\log p_1\log p_2,$$

$$L_{\varrho,\varrho'}(x,n)=\sum_{\substack{m+l^k=n\\ \frac{x}{2}\leqq m<x\\ \frac{\sqrt[k]{x}}{2}\leqq l<\sqrt[k]{x}}}m^{\varrho-1}l^{\varrho'-1},\qquad L_{1,1}(x,n)=L(x,n),$$

$$\sigma(n,R,l)=\sum_{\substack{q\leqq R,\\ (q,l)=1}}\frac{A(q,n)}{\phi^2(q)},\qquad \sigma(n,R)=\sigma(n,R,1),$$

$$N(\sigma,T,\chi)=|\{\sigma:\ L(\sigma,\chi)=0\ ,\beta\geqq\sigma,\ 0\leqq|\gamma|\leqq T\}|,$$

$$N^{-}(\sigma,P,T)=\sum_{q\leqq P}\sum_{\chi \bmod q}{}^{*}N(\sigma,T,\chi),$$

where the possibly existing Siegel zero (relative to P) is excluded.

$$N(n,q)(=N(q))=\left|(m,l):m^k+l\equiv n(\mathrm{mod}\,q),m,l\in\{1,2,\dots,q\},(ml,q)=1\right|,$$

$$w(n,q)=\left|m:m^k\equiv n(\mathrm{mod}\,q),m\in\{1,2,\dots,q\},(m,q)=1\right|.$$

c_1, c_2, ... as well as the O- and $\ll$- constants are effectively computable positive constants which may depend on k.

3. Preliminary results

In the following we only argue for a fixed number k. We first quote:

LEMMA 3.1. *There exists a positive constant $c_1<1$ such that $L(s,\chi)\neq 0$ in the region*

$$\sigma\geqq 1-\frac{c_1}{\log T},\quad |t|\leqq T^{4k+7}$$

for all primitive characters $\chi \bmod q$, $q \leqq T$, $T \geqq 2$ *with the possible exception of at most one real primitive character* $\tilde{\chi} \bmod \tilde{r}$. *If it exists, the corresponding L-function has exactly one zero* $\tilde{\beta}$ *in the region given above, which is real, simple and satisfies*

$$\frac{c_2}{\tilde{r}^{1/2}\log^2\tilde{r}} \leqq 1-\tilde{\beta} \leqq \frac{c_1}{\log T}.$$

Furthermore, all the other zeros of the L-functions for primitive characters to moduls $q \leqq T$ *do not lie in the following region*

$$\sigma \geqq 1 - \frac{c_1}{\log T} \log\left(\frac{ec_1}{\delta(T)k(T)}\right), \quad |t| \leqq T^{4k+7},$$

where $\delta(T)$ *and* $k(T)$ *are defined by*

$$\delta(T) = \left\{ \begin{array}{ll} (1-\tilde{\beta})\log T & \text{if } \tilde{\beta} \text{ exists,} \\ 1 & \text{otherwise} \end{array} \right\}, \quad k(T) = \left\{ \begin{array}{ll} 1 & \text{if } \tilde{\beta} \text{ exists,} \\ c_1 & \text{otherwise} \end{array} \right\}.$$

PROOF. [2], chapter 14 and [3], paragraph 4.

Set $P_1 = x^{b_1}$, where b_1 is a sufficiently small constant specified later. Let us further choose $T = P_1$ in Lemma 3.1. With the notation of Lemma 3.1 let further

$$P_2 = x^{b_2} = \left\{ \begin{array}{l} P_1 \text{ if } \exists \tilde{r},\ \tilde{r} < P_1^{\lambda}, \\ P_1^{\lambda} \text{ otherwise} \end{array} \right\},$$

where λ, $0 < \lambda = \lambda(k) < \frac{1}{2}$ is a sufficiently small parameter specified later. Then Lemma 3.1 holds with $T = P_2, \lambda c_1$ instead of c_1 and $\tilde{r} \leqq P_2^{\lambda}$ (if $\tilde{\beta}$ exists). We define the $P_2, \lambda c_1$-$*$-*excluded zeros* as those zeros $s = \sigma + it$ of the $L(s,\chi)$-functions, where χ is a primitive character $\bmod q$, $q \leqq P_2$, in the region

$$\sigma \geqq 1 - \frac{16k^2 \log\log x}{\log x} \log\left(e\left(\frac{2}{\delta(P_2)}\right)^{\frac{1}{\log\log x}}\right), \quad |t| \leqq P_2^{4k+7},$$

excluding the Siegel zero (relative to P_2) and $\delta(P_2)$ is defined by Lemma 3.1 with $T = P_2$ and λc_1 instead of c_1. (Here e does not denote the exponential function, but the number e.) For any number P with $P = P_2^{\eta}$ for an $\eta \in]0,1]$ holds Lemma 3.1, obviously with $T = P$ and $\eta\lambda c_1$ instead of c_1. The $P, \lambda\eta c_1$-*excluded zeros* are defined as the zeros of $L(s,\chi)$-functions to a primitive character $\chi \bmod q$, $q \leqq P$, in the region

$$\sigma \geqq 1 - \frac{16k^2 \log\log x}{\log x} \log\left(e\left(\frac{2(4k+2)}{(4k+3)\delta(P)}\right)^{\frac{1}{\log\log x}}\right), \quad |t| \leqq P^{4k+7},$$

excluding the Siegel zero (relative to P) and $\delta(P)$ is defined by Lemma 3.1 with $T = P$ and the constant $\eta\lambda c_1$. We estimate the number of $P, \lambda\eta c_1$-*excluded zeros* by means of

LEMMA 3.2. *There exist constants c_3 and c_4 such that*

$$N^-(\alpha, T, T^{4k+7}) \leqq c_3\delta(T)T^{c_4(1-\alpha)},$$

where $\delta(T)$ is defined as in Lemma 3.1.

PROOF. See Zaccagnini [36], Lemma 3.2.

Applying this lemma we get for a sufficiently small b:

$$N^-\left(1 - \frac{16k^2 \log\log x}{\log x} \log\left(e\left(\frac{2(4k+2)}{(4k+3)\delta(P)}\right)^{\frac{1}{\log\log x}}\right), P_2, P_2^{4k+7}\right)$$

$$\leqq c_3\delta(P_2)\exp\left(16k^2 b_2 c_4 \log\log x - 16k^2 b_2 c_4 \log\frac{\delta(P_2)(4k+3)}{2(4k+2)}\right)$$

$$\leqq \delta^{5/6}(P_2)\log^{1/6} x.$$

So we find by $\delta(P) \leqq 1$ that there are not more than

$$\ll \log^{1/3} x \tag{3.1}$$

pairs of numbers (ϱ, ϱ'), where each of the two numbers is an $P, \lambda\eta c_1$-*excluded zero* or a Siegel zero (relative to P) or $= 1$. Now we prove that for every fixed P_2 we can find a P with $P = P_2^{\eta}$, $\eta \in \left[\frac{4k+2}{4k+3}, 1\right]$, for which further holds

$$\sigma \text{ is } P, \eta\lambda c_1 - \textit{excluded zero} \Rightarrow |\mathrm{Im}(\sigma)| \notin]P^{4k+3}, 16P^{4k+3}[. \tag{3.2}$$

First we have for a sufficiently large x and a fixed b_2:

$$16^{(\log x)^{1/6}} \leqq P_2^{1/4}. \tag{3.3}$$

Let $\{\gamma_1, \ldots, \gamma_m\}$ be the imaginary parts of the $P_2, \lambda c_1 - * - \textit{excluded zeros}$ with $|\gamma_i| \in [P_2^{4k+2}, P_2^{4k+3}]$ and $P_2^{4k+2} \leqq |\gamma_1| \leqq |\gamma_2| \leqq \ldots \leqq |\gamma_m| \leqq P_2^{4k+3}$. Estimating the $P_2, \lambda c_1 - * - \textit{excluded zeros}$ as in (3.1), we find by (3.3) that there holds at least one of the following three inequalities:

$$\exists t \in \{1, \ldots, m-1\} \text{ with } \frac{|\gamma_{t+1}|}{|\gamma_t|} > 16 \quad \text{or} \quad \frac{P_2^{4k+3}}{|\gamma_m|} \geqq P_2^{1/4} \quad \text{or} \quad \frac{|\gamma_1|}{P_2^{4k+2}} \geqq P_2^{1/4}.$$

Setting in the first case $|\gamma_t| = P^{4k+3}$, in the second case $|\gamma_m| = P^{4k+3}$ and in the third case $P_2^{4k+2} = P^{4k+3}$, we find a P with $P \in [P_2^{4k+2/4k+3}, P_2]$. (If there holds more than one of the three inequalities, then the definition of P^{4k+3} can be chosen arbitrarily among the possible choices.) But by the definition of a $P, \eta\lambda c_1 - \textit{excluded}$ and a $P_2, \lambda c_1 - * - \textit{excluded} - \textit{zero}$ every

$P, \eta\lambda c_1 - excluded\ zero$ is also an $P_2, \lambda c_1 - * - excluded\ zero$, because by the definition of $\delta(P)$ and $\delta(P_2)$ by Lemma 3.1 with the constant $c_1\lambda\eta$ and $c_1\lambda$, respectively and by $\delta(P_2) \leqq 1$ (by Lemma 1) holds:

$$\frac{4k+2}{4k+3}\frac{1}{\delta(P)} \leqq \frac{1}{\delta(P_2)}.$$

So every $P, \eta\lambda c_1 - excluded\ zero$, which does not satisfy the condition (3.2), would be a $P_2, \lambda c_1 - * - excluded\ zero$, which contradicts the choice of P. So P satisfies the condition (3.2). Then Lemma 3.1 holds with $T = P$, $c_1' = \eta\lambda c_1$ instead of c_1 and

$$\tilde{r} < P^{(4k+3/4k+2)\lambda} \tag{3.4}$$

(if the Siegel zero exists). In order to simplify the notation we write in the sequel $c_1' = c_1$ and the $P, \eta\lambda c_1 - excluded\ zeros$ will be denoted as the $P - excluded\ zeros$. Let the $P - excluded\ characters$ be the primitive characters $\chi(\mathrm{mod}\, q), q \leqq P$, for which $L(s, \chi) = 0$, where s is a $P - excluded\ zero$ and denote by the $P - excluded\ moduls$ the moduls belonging to the $P - excluded\ characters$. We will also use the following notation:

$$\theta = \{P - excluded\ characters\}, \quad \theta' = \{P - excluded\ zeros\},$$

$$\begin{aligned} P = x^b,\ \delta(P) = \delta,\ \tilde{\chi} &= \text{ exceptional character (to } P), \\ \tilde{\beta} &= \text{ Siegel zero (to } P). \end{aligned} \tag{3.5}$$

The unit interval $\left[\frac{1}{Q}, 1 + \frac{1}{Q}\right]$ is now divided into the disjunct major arcs M and the minor arcs m, which are defined by

$$M = \sum_{q \leqq P}{\sum_{a=1}^{q}}^{*} I(a,q), \quad I(a,q) = \left[\frac{a}{q} - \frac{1}{Q}, \frac{a}{q} + \frac{1}{Q}\right],$$

$$m = \left[\frac{1}{Q}, 1 + \frac{1}{Q}\right] \setminus M, \quad Q = xP^{-4k-3},$$

where P is defined by (3.2). We obtain

$$\begin{aligned} r(x,n) &= \int_{1/Q}^{1+(1/Q)} S(\alpha)S_k(\alpha)e(-n\alpha)d\alpha \\ &= \int_M S(\alpha)S_k(\alpha)e(-n\alpha) + \int_m S(\alpha)S_k(\alpha)e(-n\alpha) =: r_1(x,n) + r_2(x,n), \end{aligned} \tag{3.6}$$

where $r_1(x,n)$ and $r_2(x,n)$ are real, because the sets M and m are even mod 1.

4. Arithmetic and analytic lemmas

LEMMA 4.1. *Let* $q = q_1 q_2$ *and* $(q_1, q_2) = 1$.

(a) $N(q_1 q_2) = N(q_1) N(q_2)$.

(b) *For any prime number* p *and any natural number* $\alpha \geqq 2$ *holds:* $N(p^\alpha) = p^{\alpha-1} N(p)$.

(c) *For any natural number* r *holds:*

$$\frac{r}{\phi^2(r)} N(r) = \prod_{p|r} \frac{p}{(p-1)^2} N(p).$$

d) *Put* $s(p,n) := 1 + \dfrac{A(p,n)}{(p-1)^2}$. *Then we have*

$$s(p,n) = \frac{p}{(p-1)^2} N(p).$$

PROOF. (a) We note that every a with $1 \leqq a \leqq q$ can be written in a unique way as $a = a_1 q_2 + a_2 q_1$ with $1 \leqq a_i \leqq q_i$. We write

$$N(q) = \frac{1}{q} \sum_{a=1}^{q} \sum_{m=1}^{q} \cdot \sum_{l=1}^{q} \cdot e\left(\frac{m^k + l - n}{q} a\right),$$

split the summation over a in the two summations over a_1 and a_2 and after some arithmetical transformations get the lemma.

(b) By definition we have

$$N(p^\alpha) = \left| m : m^k \neq n (\mathrm{mod}\, p), m \in \{1, 2, \ldots, p^\alpha\}, (m,p) = 1 \right|.$$

For $\alpha \geqq 2$ we write for $(m,q) = 1$: $m = v + w p^{\alpha-1}$ with $1 \leqq v \leqq p^{\alpha-1}$, $(v,p) = 1$ and $0 \leqq w \leqq p-1$, from which we obtain

$$\begin{aligned} N(p^\alpha) &= \left| (v,w) : v^k \neq n (\mathrm{mod}\, p), 1 \leqq v \leqq p^{\alpha-1}, (v,p) = 1, 0 \leqq w \leqq p-1 \right| \\ &= p N(p^{\alpha-1}). \end{aligned}$$

Applying $(\alpha - 2)$-times this argument we get part (b).

(c) We get from (a) and (b)

$$\begin{aligned} \frac{r}{\phi^2(r)} N(r) &= \prod_{p^\alpha \| r} \frac{p^\alpha}{\phi^2(p^\alpha)} N(p^\alpha) = \prod_{p|r} \frac{p^\alpha p^{\alpha-1}}{(p-1)^2 (p^{\alpha-1})^2} N(p) \\ &= \prod_{p|r} \frac{p}{(p-1)^2} N(p). \end{aligned}$$

(d)

$$s(p,n) = 1 + \frac{\sum_{a=1}^{p}{}^{*} \sum_{m=1}^{p} \sum_{l=1}^{p}{}^{*} e\left(\frac{m^k+l-n}{p}a\right)}{(p-1)^2}$$

$$= \frac{\sum_{a=1}^{p} \sum_{m=1}^{p} \sum_{l=1}^{p}{}^{*} e\left(\frac{m^k+l-n}{p}a\right)}{(p-1)^2} = \frac{p}{(p-1)^2} N(p).$$

LEMMA 4.2. *For any natural number $k \geqq 1$, any primitive character χ modulo p^α, $\alpha \geqq 1$ and $(a,p) > 1$ holds:*

$$C_k(\chi, a) = 0.$$

PROOF. Writing $a' = a/p$ and $m = u + vp^{\alpha-1}$ we obtain for $\alpha \geqq 2$:

$$C_k(\chi, a) = \sum_{m=1}^{p^\alpha} \chi(m) e\left(\frac{a' m^k}{p^{\alpha-1}}\right) = \sum_{u=1}^{p^{\alpha-1}} e\left(\frac{a' u^k}{p^{\alpha-1}}\right) \sum_{v=1}^{p} \chi(u + vp^{\alpha-1}),$$

which is equal to zero because the last inner sum vanishes for primitive characters. For $\alpha = 1$ the lemma follows by the orthogonality relation of characters.

LEMMA 4.3. *For any natural number k, $q_1 q_2 = q$, $(q_2, q_1) = 1$, $\chi_a (\mathrm{mod}\, q) = \chi_{a_1} (\mathrm{mod}\, q_1) \chi_{a_2} (\mathrm{mod}\, q_2)$, $\chi_b (\mathrm{mod}\, q) = \chi_{b_1} (\mathrm{mod}\, q_1) \chi_{b_2} (\mathrm{mod}\, q_2)$, and $h = h_1 q_2 + h_2 q_1$*

(a) $C_k(\chi_a, h) = C_k(\chi_{a_1}, h_1) C_k(\chi_{a_2}, h_2)$.

(b) $A(q, n, \chi_a, \chi_b) = A(q_1, n, \chi_{a_1}, \chi_{b_1}) A(q_2, n, \chi_{a_2}, \chi_{b_2})$.

(c) *For any natural number $k \geqq 1$, any primitive character χ modulo q, $q > 1$ and $(a,q) > 1$:*

$$C_k(\chi, a) = 0.$$

PROOF. (a) is shown in the same way as Lemma 4.1 (a). Applying (a) we can show (b) in a similar way. (c) There exists a $p^\alpha \parallel q$, $\alpha \geqq 1$ with $(p,a) > 1$. Writing $a = a_2 p^\alpha + a_1 \frac{q}{p^\alpha}$, it is by part (a) enough to prove that $C_k(\chi_{p^\alpha}, a_1) = 0$. But this follows from Lemma 4.2 because of $(p, a_1) > 1$.

LEMMA 4.4. (a) *For any natural number n and prime number p*

$$A(p,n) = -(w(n,p) - 1)p - 1.$$

Let now be given any n which satisfies the congruence conditions in (1.1).

(b) *If at least one of the two characters* χ_1 *and* χ_2 *modulo* q, $q>1$ *is primitive, then*

$$|A(q,n,\chi_1,\chi_2)| \leqq \phi^2(q) \prod_{p|q} \left(1 - \frac{(w(n,p)-1)p+1}{(p-1)^2}\right).$$

(c) *For any characters* χ_1 *and* χ_2 *modulo* q:

$$|A(q,n,\chi_1,\chi_2)| \ll \phi^2(q) \log^{4k} q.$$

(d) *For any prime number* p *and* $s(p,n)$ *defined as in Lemma 4.1* (d)

$$s(p,n) > 0$$

holds true.

PROOF. (a) By the definition of $A(p,n)$ we have

$$A(p,n) = -\sum_{a=1}^{p-1}\sum_{m=1}^{p-1} e\left(\frac{m^k-n}{p}a\right) = -(w(n,p)-1)\,p-1.$$

(b) By Lemma 4.2 it holds:

$$\begin{aligned}\left|\phi^{-2}(q)A(q,n,\chi_1,\chi_2)\right| &= \left|\phi^{-2}(q)\sum_{a=1}^{q} C_1(\chi_1,a)C_k(\chi_2,a)e\left(\frac{-an}{q}\right)\right| \\ &= \left|\phi^{-2}(q)q \sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,q),\ 1\leqq l,m\leqq q,\\ (lm,q)=1}} \chi_1(l)\chi_2(m)\right| \\ &\leqq \phi^{-2}(q)qN(q) \leqq \prod_{p|q}\phi^{-2}(p)pN(p),\end{aligned}$$

where in the last step we have used Lemma 4.1 (c). Noting further that by the definition of $N(p)$ we have:

$$N(p) = \left|m:\ 1\leqq m\leqq p-1,\ m^k \not\equiv n(\mathrm{mod}\,p)\right| = p-1-w(n,p), \tag{4.1}$$

we see that the lemma holds by

$$\left|\phi^{-2}(q)A(q,n,\chi_1,\chi_2)\right| \leqq \prod_{p|q} p\left(\frac{p-1-w(n,p)}{(p-1)^2}\right)$$

$$= \prod_{p|q} \left(1 - \frac{(w(n,p)-1)p+1}{(p-1)^2}\right).$$

(c) The lemma is trivial for $q=1$. If the characters χ_1 and χ_2 satisfy the condition of part (b), then part (c) follows from part (b), $w(n,p) \leqq k$ and so

$$\prod_{p|q} \left(1 - \frac{(w(n,p)-1)p+1}{(p-1)^2}\right) \leqq \prod_{p \leqq q} \left(1 + \frac{4k}{p}\right) \ll \log^{4k} q.$$

In the other case we have

$$\chi_1 = \chi_1^* \chi_{0,l} \quad \text{or} \quad \chi_1 = \chi_{0,q},$$

where $q = q^* l$ and χ_1^* is a primitive character modulo q^*, $q^* > 1$. We quote Lemma 5.3 in [9], which states that for a character χ modulo $q \leftrightarrow \chi^*$ modulo q^* and $(a,q)=1$ it holds

$$C_1(\chi, a) = \overline{\chi(a)} \tau(\chi^*) \mu\left(\frac{q}{q^*}\right) \chi^*\left(\frac{q}{q^*}\right). \tag{4.2}$$

So if $\chi_1 = \chi_1^* \chi_{0,l}$, we can restrict ourselves to the q which satisfies:

$$\mu(l) \neq 0, \quad (l, q^*) = 1. \tag{4.3}$$

From this we get $\chi_2 = \chi_3 \chi_4$ with $\chi_3 = \chi_3 \bmod q^*$ and $\chi_4 = \chi_4 \bmod l$. So we obtain from Lemma 4.3 (c) and the first part of the proof:

$$|A(q, n, \chi_1 \chi_2)| \ll \phi^2(q^*) \log^{4k} q^* A(l, n, \chi_{0,l}, \chi_4). \tag{4.4}$$

Using further the estimate

$$C_k(\chi, a) \ll_\epsilon q^{1/2+\epsilon}, \tag{4.5}$$

which holds for $(a,q)=1$ and may be found in [13], note to Lemma 4, we obtain

$$A(l, n, \chi_{0,l}, \chi_4) = \sum_{a=1}^{l}{}^{*} C_1(\chi_{0,l}, a) C_k(\chi_4, a) e\left(\frac{-an}{l}\right) = \ll_\epsilon l^{3/2+\epsilon}. \tag{4.6}$$

So the lemma follows from (4.4) and (4.6). If $\chi_1 = \chi_{0,q}$ the lemma follows immediately by arguing like in (4.6).

(d) By Lemma 4.1 (a) it is enough to show that $N(p) > 0$. Because of (4.1) the lemma is proved if $w(n,p) = 0$. In the other case we know (see Ireland, Rosen [5], p. 45) that $w(n,p) = (k, p-1)$, so that by (4.1) the lemma is proved in the case $p-1 \nmid k$. By Fermat's little theorem we know for $p-1 | k$:

$$a^k \equiv 1 \pmod{p} \quad \forall a \text{ with } a \not\equiv 0 \pmod{p}.$$

So we obtain for $p-1|k$:

$$n \equiv 1 (\bmod\ p) \Longleftrightarrow w(n,p) = p-1 \Longleftrightarrow N(p) = 0,$$

which proves the lemma.

LEMMA 4.5. *For two primitive characters* $\chi_1 \bmod q_1$ *and* $\chi_2 \bmod q_2$ *let* $q_3 = [q_1, q_2] \leqq P$. *If* n *satisfies the congruence conditions in* (1.1), *there holds:*

$$\sum_{\substack{q \leqq P \\ q \equiv 0 (\bmod q_3)}} \frac{|A(q, n, \chi_1 \chi_{0,q}, \chi_2 \chi_{0,q})|}{\phi^2(q)} \ll \log^{5k+1} P.$$

PROOF. For $q_1|q$ let $q_1 l = q$. Analogously to (4.3) we only have to treat those q that satisfy

$$\mu(l) \neq 0, \quad (l, q_1) = 1,$$

and for which, under the additional assumption $[q_1, q_2] = q_3$ and $q_3|q$,

$$\left(\frac{q}{q_3}, q_3\right) = 1 \tag{4.7}$$

holds. So we obtain

$$\chi_1 \chi_{0,q} = \chi_1 \chi_{0,\frac{q_3}{q_1}} \chi_{0,\frac{q}{q_3}}, \quad \chi_2 \chi_{0,q} = \chi_2 \chi_{0,\frac{q_3}{q_2}} \chi_{0,\frac{q}{q_3}},$$

and we further have by (4.2), Lemma 4.3 (b) and $w(p,n) \leqq k$:

$$|A(m,n)| \leqq \prod_{p|m} pk = mk^{\omega(m)}.$$

Using this, (4.7), Lemma 4.4 (c) and Lemma 4.5 in [14], we finally derive the lemma by

$$\begin{aligned}
&\sum_{\substack{q \leqq P \\ q \equiv 0 (\bmod q_3)}} \frac{|A(q, n, \chi_1 \chi_{0,q}, \chi_2 \chi_{0,q})|}{\phi^2(q)} \\
&= \frac{\left|A(q_3, n, \chi_1 \chi_{0,\frac{q_3}{q_1}}, \chi_2 \chi_{0,\frac{q_3}{q_2}})\right|}{\phi^2(q_3)} \sum_{\substack{m \leqq \frac{P}{q_3} \\ (m, q_3) = 1}} \frac{|A(m,n)|}{\phi^2(m)} \\
&\ll \log^{4k} P \sum_{m \leqq P} \frac{mk^{\omega(m)}}{\phi^2(m)} \ll \log^{4k+1} P \sum_{m \leqq P} \frac{k^{\omega(m)}}{m} \ll (\log P)^k.
\end{aligned}$$

LEMMA 4.6. *For all* ϱ *with* $0 \leqq \mathrm{Re}(\varrho) \leqq 1$ *and* $s \geqq c_5 k^2 \log k$ *it holds:*

$$\int_0^1 |F_\varrho(\alpha)|^{2s} d\alpha \ll x^{(2s/k)-1} .$$

PROOF. Considering the underlying diophantine equation this can be shown in the same way as Lemma 5.2 in [14].

LEMMA 4.7. (a) *Let* $2^k x^{-1} < \lambda < x^{\frac{1}{k}-1}$ *and* $0 \leqq \mathrm{Re}(\varrho) \leqq 1$. *There holds:*

$$\int_{-\lambda}^{\lambda} |F_\varrho(\alpha)|^2 d\alpha \ll x^{(2/k)-1}.$$

(b) *Let* $2x^{-1} < \lambda < 1$ *and* $0 \leqq \mathrm{Re}\,(\varrho) \leqq 1$. *There holds:*

$$\int_{-\lambda}^{\lambda} |T_\varrho(\alpha)|^2 d\alpha \ll x.$$

PROOF. (a) We define

$$u_n = \begin{cases} m^{\varrho-1} & \text{if } n = m^k \in [x/2^k, x[, \\ 0 & \text{otherwise.} \end{cases}$$

Then we get by Gallagher's lemma ([3], Lemma 1)

$$\int_{-\lambda}^{\lambda} |F_\varrho(\alpha)|^2 d\alpha \ll \int_{x/2^{k+1}}^{x} \left| \lambda \sum_{t}^{t+(2\lambda)^{-1}} u_n \right|^2 dt. \tag{4.8}$$

For the inner sum holds for a fixed $t \in [x/2^{k+1}, x]$

$$\sum_{t}^{t+(2\lambda)^{-1}} u_n \ll \sqrt[k]{(t+(2\lambda)^{-1}} - \sqrt[k]{t} \ll \lambda^{-1} x^{(1/k)-1}.$$

Substituting this in (4.8) we obtain the lemma. Part (b) is proved in the same way.

LEMMA 4.8. (a) *Let be given any* $\sigma = \beta + i\gamma$ *with* $0 \leqq \beta \leqq 1$ *and* $|\gamma| \leqq x/Q$. *Then for* $1/Q \leqq |\alpha| \leqq 1/2$:

$$T_\varrho(\alpha) \ll \frac{x^{\beta-1}}{|\alpha|}.$$

(b) *Let be given any* $\sigma = \beta + i\gamma$ *with* $0 \leqq \beta \leqq 1$ *and* $x^{1/2} \geqq |\gamma| > 16x/Q$. *Then for* $|\alpha| \leqq 1/Q$:

$$T_\varrho(\alpha) \ll \frac{x^\beta}{|\gamma|}.$$

PROOF. Part (a) is nearly identical to Lemma 12 in [1] and part (b) can be shown in the same way by appealing to Lemmas 4.2 and 4.8 in [11].

LEMMA 4.9. *If* $\varrho = \beta + i\gamma$ *with* $0 \leqq \beta \leqq 1$ *and* $|\gamma| \leqq P^{4k+3}$, *then for any* $s \geqq ck^2 \log k$ *and for all* ϱ' *with* $0 \leqq \operatorname{Re}(\varrho') \leqq 1$:

$$\int_{1/Q}^{1/2} |F_{\varrho'}(\alpha) T_\varrho(\alpha)| d\alpha \ll x^{(1/k)+\beta-1} P^{-\frac{2k+1}{s}}.$$

PROOF. Using Hölder's inequality, the Lemmata 4.6 and 4.8 (a) and the definition of Q this inequality can be shown in the same way as Lemma 5.8 in [14].

LEMMA 4.10. *If* $\varrho = \beta + i\gamma$ *with* $0 \leqq \beta \leqq 1$ *and* $16P^{4k+3} < |\gamma| \leqq P^{4k+7}$, *then there holds for all* ϱ' *with* $0 \leqq \operatorname{Re}(\varrho') \leqq 1$:

$$\int_{-\frac{1}{Q}}^{\frac{1}{Q}} |F_{\varrho'}(\alpha) T_\varrho(\alpha)| d\alpha \ll x^{1/k} P^{-2k-1}.$$

PROOF. Using the Lemmas 4.7 (a) and 4.8 (b) we get

$$\int_{-1/Q}^{1/Q} |F_{\varrho'}(\alpha) T_\varrho(\alpha)| d\alpha \ll \left(\int_{-1/Q}^{1/Q} |F_{\varrho'}(\alpha)|^2 d\alpha \right)^{1/2} \left(\int_{-1/Q}^{1/Q} |T_\varrho(\alpha)|^2 d\alpha \right)^{1/2}$$

$$\ll x^{1/k} P^{-2k-1}.$$

5. Lemmas for the singular series

LEMMA 5.1. (a) *For any character* χ *modulo* p^{α_1} *and* $\alpha_1 \geqq 0$

$$C_k(\chi\chi_0, a) = 0$$

holds if χ_0 *is the principal character to the modulus* p^α, $p \nmid a$ *and* $\alpha \geqq j + \max(j, \alpha_1)$, *where* $j = 1 + \operatorname{ord}_k(p)$ *and* $w = \operatorname{ord}_k(p) \Longleftrightarrow p^w \parallel k$.

(b) *For any primitive character* χ *modulo* p^α, $p \nmid a$, $w = \operatorname{ord}_k(p) \geqq 1$ *and* $\alpha \geqq 2w$ *it holds:*

$$C_k(\chi, a) = 0.$$

(c) *Let* χ *be any primitive character modulo* p^α *for any prime number* p *and a natural number* $\alpha \geqq 2$. *Then there holds for any integer* γ, $\alpha \geqq \gamma \geqq \alpha/2$:

$$\chi(1+p^\gamma) = e\left(\frac{c}{p^{\alpha-\gamma}}\right),$$

where $c = c(\gamma)$, $1 \leqq c \leqq p^{\alpha-\gamma}$ *is a natural number with* $p \nmid c$.

(d) *Let* χ *be any primitive character modulo* p^3 *for any prime number* $p < 2$. *Then it holds*

$$\chi(1+p) = e\left(\frac{c}{p^2}\right),$$

where $1 \leqq c \leqq p^2$, $p \nmid c$.

PROOF. (a) For $1 \leqq l \leqq p^\alpha$ we have $l = u + vp^{\alpha-j}$, $1 \leqq u \leqq p^{\alpha-j}$, $0 \leqq v \leqq p^j - 1$. By $\alpha \geqq j + \max(j, \alpha_1)$ is further $l^k \equiv u^k + vku^{k-1}p^{\alpha-j} \pmod{p^\alpha}$ and $l \equiv u \pmod{p^{\alpha_1}}$. So we get:

$$\begin{aligned} C_k(\chi\chi_0, a) &= \sum_{l=1}^{p^\alpha}{}^{*} \chi\chi_0(l) e\left(\frac{l^k a}{p^\alpha}\right) \\ &= \sum_{u=1}^{p^{\alpha-j}} \chi\chi_0(u) e\left(\frac{au^k}{p^\alpha}\right) \sum_{v=0}^{p^j-1} e\left(\frac{avku^{k-1}}{p^j}\right) = 0, \end{aligned}$$

because the inner sum vanishes for any p prime to u.

(b) We obtain in a similar way

$$C_k(\chi, a) = \sum_{l=1}^{p^\alpha}{}^{*} \chi(l) e\left(\frac{l^k a}{p^\alpha}\right) = \sum_{u=1}^{p^{\alpha-w}} e\left(\frac{au^k}{p^\alpha}\right) \sum_{v=0}^{p^w-1} \chi\left(u + vp^{\alpha-w}\right),$$

from which the lemma follows because the inner sum vanishes for a primitive character.

(c) It remains to show that $p \nmid c$. But if $p \mid c$, we obtain

$$\chi(1+ap^{\alpha-1}) = \chi^a(1+p^{\alpha-1}) = \chi^{ap^{\alpha-\gamma-1}}(1+p^\gamma) = e\left(\frac{acp^{\alpha-\gamma-1}}{p^{\alpha-\gamma}}\right) = 1,$$

which contradicts the primitivity of the character.

(d) Using $(1+p)^{p^2} \equiv 1 \pmod{p^3}$ for $p \neq 2$ (see, e.g., Ireland, Rosen [5]; S. 43) and $p \mid \binom{p}{2}$ for $p \neq 2$ the proof is analogous to the one of part (c).

LEMMA 5.2. *In the parts* (a)–(d) *let be given a natural number* $q = p^\alpha$, $\alpha \geqq 1$, *two characters* χ_1 *and* $\chi_2 \bmod q$ *and* $p \nmid k$, $p^\alpha \nmid n$.

(a) *For* $q = p$, χ_1 *primitive and* $\chi_2 = \chi_{0,q}$ *it holds:*

$$A(p, n, \chi_1, \chi_2) \leqq (k+1)p^{3/2}.$$

(b) *For* $q = p$, χ_1 *primitive and* $\chi_2 \neq \chi_{0,q}$*:*

$$A(p, n, \chi_1, \chi_2) \leqq kp^{3/2}.$$

(c) *For* $q = p^\alpha$, $\alpha \geqq 4$, χ_1, χ_2 *primitive and* $p^\beta \parallel n$, $\beta \leqq \left[\frac{\alpha}{4}\right]$*:*

$$A(p^\alpha, n, \chi_1, \chi_2) \leqq kp^{\alpha + \left[\frac{\alpha+1}{2}\right] + \left[\frac{\alpha}{4}\right]}.$$

(d) *For* $q = p^\alpha$, $\alpha \in \{2, 3\}$, χ_1, χ_2 *primitive and under the additional conditions* $p \neq 2$ *and* $p^\beta \parallel n$, $\beta \leqq 1$ *in the case* $\alpha = 3$ *holds:*

$$A(p^2, n, \chi_1, \chi_2) \ll_\epsilon kp^{(7/4)\alpha + \epsilon}.$$

(e) *Let be given the principal character* $\chi_{0,\alpha}$ *to the module* p^α *and a primitive character* χ_2 *to the module* p^{α_1} *with* $\alpha_1 < \alpha$. *Let* $p^\beta \parallel n$, $\beta \leqq \left[\frac{\alpha_1}{4}\right]$. *If with the notation of Lemma 5.1* (a) $\alpha_1 \geqq \max(\alpha - \operatorname{ord}_k(p), 6, \frac{2}{3}\alpha)$, *then there holds for any primitive character* χ_1 *modulo* p^α*:*

$$A(p^\alpha, n, \chi_1, \chi_{0,\alpha}\chi_2) \leqq k^2 p^{\alpha + \left[\frac{\alpha+1}{2}\right] + \left[\frac{\alpha}{4}\right] + 1}.$$

PROOF. We first transform $A(q, n, \chi_1, \chi_2)$ (and $A(q, n, \chi_1, \chi_{0,\alpha}\chi_2)$). Noting that in Parts (a)–(e) χ_1 is always primitive, that $|\tau(\chi_1)| = q^{1/2}$ and that (4.2) also holds for $(a, q) > 1$ for primitive characters, we see

$$\begin{aligned} A(q, n, \chi_1, \chi_2) &= \sideset{}{^*}\sum_{m,a=1}^{q} \chi_2(m) e\left(\frac{m^k - n}{q} a\right) \sideset{}{^*}\sum_{l=1}^{q} \chi_1(l) e\left(\frac{l}{q} a\right) \\ &= \tau(\chi_1) \sideset{}{^*}\sum_{m=1}^{q} \chi_2(m) \sideset{}{^*}\sum_{a=1}^{q} e\left(\frac{m^k - n}{q} a\right) \overline{\chi_1(a)} \qquad (5.1) \\ &= |\tau(\chi_1)|^2 \sideset{}{^*}\sum_{m=1}^{q} \chi_1(m^k - n)\chi_2(m) = qD(\chi_1, \chi_2), \end{aligned}$$

where $D(\chi_1, \chi_2) = \sideset{}{^*}\sum_{m=1}^{q} \chi_1(m^k - n)\chi_2(m)$.

(a) This case follows immediately from (13.3) in [14] and (5.1).

(b) For any integer n which is prime to n we can write any character χ modulo p as

$$\chi(n) = e\left(\frac{m \operatorname{ind}_g(n)}{p-1}\right),$$

where $m \in \{1, \dots, p-1\}$ and $\operatorname{ind}_g(n)$ denotes the index of n relative to a primitive root g of the reduced residue class system modulo p. Defining especially a character χ_s modulo p for $(n,p)=1$ by

$$\chi_s = e\left(\frac{\operatorname{ind}_g(n)}{p-1}\right)$$

(and $\chi_s(n)=0$, if $(n,p)>1$), we can write every character χ modulo p as $\chi = \chi_s^m$, $m \in \{1, \dots, p-1\}$, where $m = p-1 \Longleftrightarrow \chi = \chi_0$. We obtain:

$$D(\chi_1, \chi_2) = \sum_{m=1}^{p} \chi_s^{m_1}(m^k - n)\chi_s^{m_2}(m) = \sum_{m=1}^{p} \chi_s\left(\left(m^k - n\right)^{m_1} m^{m_2}\right),$$

where m_1, $m_2 \in \{1, \dots, p-2\}$. Let us denote F_p as the residue class system modulo p and $f(x) = (x^k - n)^{m_1} x^{m_2}$. With the notation of Theorem 2C′ in [10] (Weil's lemma) the character χ_s has the order $p-1$. If $f(x)$ is a $(p-1)$-th power in the sense of Theorem 2C′, every zero x_0 of $f(x) \in F_p[x]$ has the order $g_{x_0}(p-1)$, $g_{x_0} \in N$. Because of $p \nmid n$ and $m_2 \in \{1, \dots, p-2\}$ the order of the zero $x_0 = 0$ is $\neq g_{x_0}(p-1)$. $f(x)$ not having more than $(k+1)$-different zeros, the lemma now follows from Theorem 2C′ in [10].

(c) Let $\gamma = \left[\frac{\alpha+1}{2}\right]$. Writing every number a with $1 \leqq a \leqq p^\alpha$ as $a = u + vp^\gamma$, $1 \leqq u \leqq p^\gamma, 0 \leqq v \leqq p^{\alpha-\gamma} - 1$ and noting that for every integer a, $p \nmid a$ there exists a number $\overline{a}$ with $a\overline{a} \equiv 1 \pmod{p^\gamma}$, we get:

$$D(\chi_1, \chi_2) = \sum_{u=1}^{p^\gamma} \sum_{v=0}^{p^{\alpha-\gamma}-1} \chi_1(u^k - n + ku^{k-1}vp^\gamma)\chi_2(u + vp^\gamma)$$

$$= \sum_{u=1}^{p^\gamma} \chi_1(u^k - n)\chi_2(u) \sum_{v=0}^{p^{\alpha-\gamma}-1} \chi_1\left(1 + ku^{k-1}vp^\gamma \overline{(u^k - n)}\right)\chi_2(1 + \overline{u}vp^\gamma).$$

From this we obtain by Lemma 5.1 (c) and $(1+p^\gamma)^a \equiv 1 + ap^\gamma \pmod{p^\alpha}$ for two natural numbers c_1 and c_2, which are defined by

$$\chi_i(1+p^\gamma) = e\left(\frac{c_i}{p^{\alpha-\gamma}}\right), \qquad p \nmid c_i,\ i \in \{1,2\}: \tag{5.2}$$

(5.3)

$$D(\chi_1, \chi_2) = \sum_{u=1}^{p^\gamma} \chi_1\left(u^k - n\right)\chi_2(u) \sum_{v=0}^{p^{\alpha-\gamma}-1} e\left(\frac{c_1 k u^{k-1} v \overline{(u^k - n)}}{p^{\alpha-\gamma}}\right) e\left(\frac{c_2 \overline{u} v}{p^{\alpha-\gamma}}\right).$$

From (5.2) and (5.3) it is obvious that $(c_1c_2k\overline{(u^k-n)}u,p)=1$. Noting further that $a\overline{a}\equiv 1(\mathrm{mod}\,p^\gamma)\Longrightarrow a\overline{a}\equiv 1(\mathrm{mod}\,p^{\alpha-\gamma})$, we see that the inner sum in (5.3) $\neq 0$ if

(5.4)
$$c_1ku^{k-1}\overline{(u^k-n)}+c_2\overline{u}\equiv 0(\mathrm{mod}\ p^{\alpha-\gamma})\Longleftrightarrow u^k(c_1k+c_2)\equiv c_2n(\mathrm{mod}\ p^{\alpha-\gamma}).$$

If $p^\beta\parallel n$ and $p^\delta\parallel c_1k+c_2$, there holds (by the assumption of the lemma) $\beta\leqq[\frac{\alpha}{4}]<\alpha-\gamma$. So because of $(uc_2,p)=1$ a necessary condition for the solvability of the last congruence is $\beta=\delta$, in which case we can equivalently examine the congruence

$$u^k\frac{c_1k+c_2}{p^\beta}\equiv c_2\frac{n}{p^\beta}(\mathrm{mod}\ p^{\alpha-\gamma-\beta}),$$

which has mostly k solutions modulo $p^{\alpha-\gamma-\beta}$. So there are not more than $kp^{2\gamma-\alpha+\beta}$ numbers modulo p^γ for which the upper sum $\neq 0$. Together with (5.1) and (5.3) the lemma follows.

(d) We argue until (5.4) as in part (c). If p does not divide both n and c_1k+c_2, the congruence has not more than k solutions modulo $p^{\alpha-\gamma}$ and the result follows similarly to part (c). In the other case $p\parallel n$ and $p|c_1k+c_2$ we derive from (5.3) and (5.4):

$$D(\chi_1,\chi_2)=p\sum_{u=1}^{p^\gamma}\chi_1\left(u^k-n\right)\chi_2(u). \tag{5.5}$$

For any n prime to p we define

$$\chi_i(n)=e\left(\frac{m_i\,\mathrm{ind}_g(n)}{p^{\alpha-1}(p-1)}\right),$$

for $m_i\in\{1,\dots,p^{\alpha-1}(p-1)-1\}$ and $\mathrm{ind}_g(n)$ is the index of n relative to a primitive root g of the reduced residue system modulo p^α. Defining furthermore a character χ modulo p^α for $(n,p)=1$ by $\chi(n)=e\left(\frac{\mathrm{ind}_g(n)}{p^{\alpha-1}(p-1)}\right)$ (and $\chi(n)=0$, if $(n,p)>1$), we have

$$\chi_i=\chi^{m_i}. \tag{5.6}$$

χ is primitive by its definition, so we know by Lemma 5.1 (c) and (d) that $\chi(1+p)=e\left(\frac{c_3}{p^{\alpha-1}}\right)$ and $\chi_i(1+p)=e\left(\frac{c_i}{p^{\alpha-1}}\right)$, where $p\nmid c_i$, $i\in\{1,2,3\}$. By (5.6) it follows from this $c_i\equiv m_ic_3(\mathrm{mod}\,p^{\alpha-1})$ $(i\in\{1,2\})$ and so:

$$p|c_1k+c_2\Longrightarrow p|m_1k+m_2. \tag{5.7}$$

By (5.5) and (5.6) we know furthermore

$$D(\chi_1,\chi_2)=p\sum_{u=1}^{p^\gamma}\chi^{m_1}\left(u^k-n\right)\chi^{m_2}(u)=p\sum_{u=1}^{p^\gamma}\chi^{m_1k+m_2}(u)\,\chi^{m_1}\left(1-n\overline{u^k}\right),$$

where $\overline{u}$ is chosen such that $u^k\overline{u^k}\equiv 1(\mathrm{mod}\ p^{\alpha-1}=\mathrm{mod}\ p^\gamma)$, because so we get by $p\,\|\,n$: $nu^k\overline{u^k}\equiv n(\mathrm{mod}\ p^\alpha)$. Furthermore, we know from (5.7) $p|m_1k+m_2$, from which we derive by $\gamma=\alpha-1$ that

$$(h+p^\gamma)^{m_1k+m_2}\equiv h^{m_1k+m_2}(\mathrm{mod}\ p^\alpha)\ \forall h\in N.$$

So we get

$$\chi^{m_1k+m_2}(h+p^\gamma)=\chi\left((h+p^\gamma)^{m_1k+m_2}\right)=\chi\left(h^{m_1k+m_2}\right)=\chi^{m_1k+m_2}(h),$$

which shows that $\chi^{m_1k+m_2}$ is a character modulo p^γ. For $\alpha=2$ we get from the last identity for $D(\chi_1,\chi_2)$, $p\,\|\,n\Longleftrightarrow n=\tilde{n}p$, $(\tilde{n},p)=1$, (5.6) and $\chi_1(1+p)=e\left(\frac{c_1}{p}\right)$:

$$D(\chi_1,\chi_2)=p\sum_{u=1}^{p}\overline{\chi}^{m_1k+m_2}(u)e\left(\frac{-\tilde{n}c_1u^k}{p}\right)\ll_\epsilon p^{3/2+\epsilon},$$

where the last inequality is derived by applying (4.5) to $\chi^{m_1k+m_2}$. If $\alpha=2$ we can now derive the lemma by the last inequality and (5.1). If $\alpha=3$ we write any $u\in\{1,\dots,p^2\}$ as $u=v+wp$, $1\leqq v\leqq p$, $1\leqq w\leqq p-1$, getting so by (5.6) and the second last identity derived for $D(\chi_1,\chi_2)$:

$$\begin{aligned}D(\chi_1,\chi_2)&=p\sum_{v=1}^{p}\sum_{w=0}^{p-1}\overline{\chi}^{m_1k+m_2}(v+wp)\chi_1\left(1-\tilde{n}pv^k-\tilde{n}kv^{k-1}wp^2\right)\\&=p\sum_{v=1}^{p}\overline{\chi}^{m_1k+m_2}(v)\,\chi_1\left(1-\tilde{n}pv^k\right)\sum_{w=0}^{p-1}\overline{\chi}^{m_1k+m_2}\\&\quad\times(1+\overline{v}wp)\,\chi_1\left(1-\left(\overline{1-\tilde{n}pv^k}\right)\tilde{n}kv^{k-1}wp^2\right),\end{aligned}$$

where $a\overline{a}\equiv 1(\mathrm{mod}\ p)$, which implies $v\overline{v}wp\equiv wp(\mathrm{mod}\ p^2)$, which is sufficient, because $\overline{\chi}^{m_1k+m_2}$ has been shown to be a character modulo $p^{\alpha-1}=p^2$, and implies also $\left(1-\tilde{n}pv^k\right)\left(\overline{1-\tilde{n}pv^k}\right)\tilde{n}kv^{k-1}wp^2\equiv\tilde{n}kv^{k-1}wp^2(\mathrm{mod}\ p^3)$. We know by Lemma 5.1 (c) that

$$\chi_1(1+p^2)=e\left(\frac{c_4}{p}\right),\quad\overline{\chi}(1+p^2)=e\left(\frac{c_5}{p}\right),\quad p\nmid c_4c_5,$$

and, in general,

$$\chi_a\left(1+bp^2\right)=\chi_a^b\left(1+p^2\right), \quad \chi_a\in\{\chi_1,\overline{\chi}\}.$$

From (5.7) we know further that $m_1k+m_2=pc_6$ and so we get by $p|\binom{m_1k+m_2}{2}$ for $p>2$

$$(1+\overline{v}wp)^{m_1k+m_2}\equiv 1+\overline{v}wc_6p^2(\mathrm{mod}\ p^3),$$

from which we derive together with the last identity for $D(\chi_1,\chi_2)$:

$$D(\chi_1,\chi_2)=p\sum_{v=1}^{p}\overline{\chi}^{m_1k+m_2}(v)\,\chi_1\left(1-\tilde{n}pv^k\right)$$
$$\times\sum_{w=0}^{p-1}e\left(w\frac{c_5c_6\overline{v}-c_4\left(\overline{1-\tilde{n}pv^k}\right)\tilde{n}kv^{k-1}}{p}\right).$$

Similarly to part (c) we concentrate on the congruence

$$c_5c_6\overline{v}-c_4\left(\overline{1-\tilde{n}pv^k}\right)\tilde{n}kv^{k-1}\equiv 0(\mathrm{mod}\ p),$$

which for $p|c_6$ is not solvable because of $\left(c_4\left(\overline{1-\tilde{n}pv^k}\right)\tilde{n}kv^{k-1},p\right)=1$ and in the other case is equivalent to

$$\Longleftrightarrow v^k\left(-c_5c_6\tilde{n}p-c_4\tilde{n}k\right)+c_5c_6\equiv 0(\mathrm{mod}\ p).$$

By $(c_4c_5c_6\tilde{n}k,p)=1$ this congruence has at most k solutions modulo p, from which the lemma follows together with (5.1) for $\alpha=3$.

(e) Define $\lambda=\left[\frac{\alpha_1+1}{2}\right]+1$. We write a with $1\leqq a\leqq p^\alpha$ as $a=u+vp^\lambda$, $1\leqq u\leqq p^\lambda$, $0\leqq v\leqq p^{\alpha-\lambda}-1$. By the assumptions of the lemma we have $k=\tilde{k}p^{\alpha-\alpha_1+d}$, with $(\tilde{k},p)=1$, $d\geqq 0$ and for $b\geqq 3$

$$p^{2\lambda}\binom{k}{2}\equiv p^{b\lambda}\equiv 0\,(\mathrm{mod}\ p^\alpha).$$

Using this we get as in part (c)

$$D(\chi_1,\chi_{0,\alpha}\chi_2)=\sum_{u=1}^{p^\lambda}\chi_1\left(u^k-n\right)\chi_{0,\alpha}\chi_2(u)$$
$$\times\sum_{v=0}^{p^{\alpha-\lambda}-1}\chi_1\left(1+\tilde{k}p^{\alpha-\alpha_1+d}u^{k-1}vp^\lambda\overline{(u^k-n)}\right)\chi_{0,\alpha}\chi_2\left(1+\overline{u}vp^\lambda\right),$$

where $\overline{a}$ is chosen such that $a\overline{a} \equiv 1 (\mathrm{mod}\, p^{\lambda})$, in which way we get:

$$(u^k - n)\overline{(u^k - n)}\tilde{k}p^{\alpha-\alpha_1+d}u^{k-1}vp^{\lambda} \equiv \tilde{k}p^{\alpha-\alpha_1+d}u^{k-1}vp^{\lambda} (\mathrm{mod}\, p^{\alpha})$$

and $\overline{u}uvp^{\lambda} \equiv vp^{\lambda} (\mathrm{mod}\, p^{\alpha_1})$. By $\alpha - \alpha_1 + \lambda \geqq \frac{\alpha}{2}$ we get by Lemma 5.1 (c) and $\chi_{0,\alpha}\chi_2(m) = \chi_2(m)\, \forall m$ analogously to (5.2)

$$\chi_1(1 + p^{\alpha-\alpha_1+\lambda}) = e\left(\frac{c_1}{p^{\alpha_1-\lambda}}\right),$$

$$\chi_{0,\alpha}\chi_2(1 + p^{\lambda}) = \chi_2(1 + p^{\lambda}) = e\left(\frac{c_2}{p^{\alpha_1-\lambda}}\right),$$

where $p \nmid c_1 c_2$. We obtain as in (5.3)

$$D(\chi_1, \chi_{0,\alpha}\chi_2) = \sum_{u=1}^{p^{\lambda}} \chi_1\left(u^k - n\right)\chi_{0,\alpha}\chi_2(u)$$

$$\times \sum_{v=0}^{p^{\alpha-\lambda}-1} e\left(\frac{c_1\tilde{k}p^d u^{k-1}v\overline{(u^k - n)}}{p^{\alpha_1-\lambda}}\right) e\left(\frac{c_2\overline{u}v}{p^{\alpha_1-\lambda}}\right).$$

Arguing as before we see that because of $(c_2 u, p) = 1$ the inner sum can only be $\neq 0$ if $d = 0$, in which case we have to examine the congruence

$$u^k(c_1\tilde{k} + c_2) \equiv c_2 n (\mathrm{mod}\, p^{\alpha_1-\lambda}).$$

By $\beta < \alpha_1 - \lambda$ it is equivalent to the congruence

$$\frac{u^k(c_1\tilde{k} + c_2)}{p^{\beta}} \equiv \frac{c_2 n}{p^{\beta}} (\mathrm{mod}\, p^{\alpha_1-\lambda-\beta}),$$

that has at most k solutions modulo $p^{\alpha_1-\lambda-\beta}$, from which the lemma follows similarly to part (c).

LEMMA 5.3. *For any two primitive characters* $\chi_1 \bmod q_1$ *and* $\chi_2 \bmod q_2$ *with* $q_3 = [q_1, q_2] \leqq x^{\frac{1}{4}}$ *holds for all but* $\ll x q_3^{-1/16}$ *natural numbers* $n \in [(9/10)x, x[$:

$$A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3}) \ll q_3^{2-(1/32)}.$$

PROOF. The case $q_3 = 1$ is trivial. As in (4.3) we can concentrate on the case

$$q_3 = q_1 q_4,\ (q_1, q_4) = 1,\ \chi_1\chi_{0,q_3} = \chi_1\chi_{0,q_4},$$

$$\chi_2\chi_{0,q_3} = \chi_5\chi_6 \text{ with } \chi_5 \bmod q_1,\ \chi_6 \bmod q_4.$$

By applying Lemma 4.3 (b) and arguing as in (4.6) we obtain

$$A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3}) = A(q_1, n, \chi_1, \chi_5)A(q_4, n, \chi_{0,q_4}, \chi_6),$$
$$A(q_4, n, \chi_{0,q_4}, \chi_6) \ll q_4^{(3/2)+\epsilon}. \tag{5.8}$$

The lemma follows from (5.8), if χ_1 is the principal character to a module $q_1 \leqq q_3^{3/4}$, because in this case we get by (5.1) and (5.8):

$$|A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3})| \ll q_1^2\left(\frac{q_3}{q_1}\right)^{(3/2)+\epsilon}$$
$$\leqq q_1^{1/2}q_3^{(3/2)+\epsilon} \leqq q_3^{(15/8)+\epsilon} \leqq q_3^{2-(1/32)}.$$

So we assume in the following that χ_1 is a primitive character to a module $q_1 > q_3^{3/4}$. By Lemma 4.3 (b) we have

(5.9)
$$A(q_1, n, \chi_1, \chi_5) = \prod_{D\in\{A,B,C\}}\prod_{i=1}^{3}\prod_{\substack{p^\alpha \| q_1 \\ A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha})\in D_i}} A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}),$$

where $\chi_i = \prod_{p^\alpha \| q_1} \chi_{i,p^\alpha}$, $i \in \{1, 5\}$, $\chi_{1,p^\alpha} \bmod p^\alpha$, an empty product is equal to 1 and

$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_1 \Longleftrightarrow \alpha = 1,\ p|k,$
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_2 \Longleftrightarrow \alpha = 1,\ p\nmid kn,$
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_3 \Longleftrightarrow \alpha = 1,\ p\nmid k,\ p|n,$
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in B_1 \Longleftrightarrow \chi_{5,p_\alpha}$ primitive, $\alpha \geqq 2,\ p|k,$
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in B_2 \Longleftrightarrow \chi_{5,p_\alpha}$ primitive, $\alpha = 2,\ p\nmid k,\ p^\alpha \nmid n$ or χ_{5,p_α} primitive, $\alpha = 3,\ p\nmid k,\ p \neq 2,\ p^\beta \| n$ with $\beta \leqq 1$ or χ_{5,p_α} primitive, $\alpha \geqq 4,\ p\nmid k,\ p^\beta \| n$ with $\beta \leqq \left[\frac{\alpha}{4}\right]$,
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in B_3 \Longleftrightarrow \chi_{5,p_\alpha}$ primitive, $\alpha \geqq 2,$ $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \notin B_1 \cup B_2,$
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in C_1 \Longleftrightarrow \chi_{5,p_\alpha}$ not primitive, $\alpha \geqq 2,\ p^\beta \| n$ with $\beta > \left[\frac{\alpha}{6}\right]$,
$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in C_2 \Longleftrightarrow \chi_{5,p_\alpha}$ not primitive, $\alpha \geqq 2,\ p^\beta \| n$ with $\beta \leqq \left[\frac{\alpha}{6}\right]$, $\operatorname{cond}\chi_{5,p_\alpha} \geqq \max\left(\operatorname{ord}_k(p)+1, 6, \frac{2}{3}\alpha\right)$,

$$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in C_3 \Longleftrightarrow \chi_{5,p_\alpha} \text{ not primitive, } \alpha \geqq 2,$$
$$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \notin C_1 \cup C_2.$$

For $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_3 \cup B_3 \cup C_1$ we have by (5.1) trivially:

$$|A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha})| \leqq p^{2\alpha}. \tag{5.10}$$

In the following let cond $\chi_{5,p^\alpha} = \alpha_1$. For the estimation of $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in C_2$, by Lemma 5.1 (a) and by the relation $\operatorname{ord}_k(p) + 1 \leqq \alpha_1$, which holds by the definition of C_2, we can restrict our observations to the case $\alpha \leqq \operatorname{ord}_k(p) + \alpha_1$. By $\beta \leqq \left[\frac{\alpha}{6}\right] \leqq \left[\frac{\alpha_1}{4}\right]$ the conditions of Lemma 5.2 (e) are satisfied in this case. So we get by Lemma 5.2 (a)–(e) for $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_2 \cup B_2 \cup C_2$:

$$A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \leqq c_6 k^2 p^{(17/9)\alpha}. \tag{5.11}$$

For the estimation $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in C_3$, by Lemma 5.1 (a), we have only to look at the case $\alpha \leqq \operatorname{ord}_k(p) + \max(\operatorname{ord}_k(p) + 1, \alpha_1)$ and so $\operatorname{ord}_k(p) \geqq 1$. If the maximum on the right side is $\operatorname{ord}_k(p) + 1$, we have

$$\alpha \leqq 3 \operatorname{ord}_k(p).$$

In the other case it follows from the definition of C_3

$$\alpha_1 < \max\left(\operatorname{ord}_k(p) + 1, 6, \frac{2}{3}\alpha\right) \leqq \max\left(6 \operatorname{ord}_k(p), \frac{2}{3}\left(\operatorname{ord}_k(p) + \alpha_1\right)\right),$$

from which together with the equivalence

$$\alpha_1 < \frac{2}{3}\left(\operatorname{ord}_k(p) + \alpha_1\right) \Longleftrightarrow \alpha_1 < 2 \operatorname{ord}_k(p)$$

it follows that:

$$\alpha_1 < 6 \operatorname{ord}_k(p) \text{ and so } \alpha \leqq 6 \operatorname{ord}_k(p).$$

So we get in both cases

$$\alpha \leqq 6 \operatorname{ord}_k(p),$$

from which we get together with Lemma 5.1 (b) for $A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_1 \cup B_1 \cup C_3$:

$$|A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha})| \leqq p^{18 \operatorname{ord}_k(p)}. \tag{5.12}$$

We define now

$$f(q_1, n) = \prod_{\substack{p^\alpha \| q_1 \\ A(p^\alpha, n, \chi_{1,p^\alpha}, \chi_{5,p^\alpha}) \in A_1 \cup B_1 \cup C_3}} p^\alpha,$$

$$g(q_1,n)=\prod_{\substack{p^\alpha\|q_1\\ A(p^\alpha,n,\chi_{1,p^\alpha},\chi_{5,p^\alpha})\in A_3\cup B_3\cup C_1}} p^\alpha,$$

$$h(q_1,n)=\prod_{\substack{p^\alpha\|q_1\\ A(p^\alpha,n,\chi_{1,p^\alpha},\chi_{5,p^\alpha})\in A_2\cup B_2\cup C_2}} p^\alpha.$$

Then we have $f(q_1,n)g(q_1,n)h(q_1,n)=q_1$, $g(q_1,n)\leqq 8(q_1,n)^6$ and the three factors are pairwise prime. Defining characters $\chi_{q_i,d} \bmod d(q_1,n)$, $i\in\{1,5\}$, $d\in\{f,g,h\}$ with $\chi_i=\prod_{d\in\{f,g,h\}}\chi_{i,d}$, we get by Lemma 4.3 (b) and (5.9)–(5.12):

$$\begin{aligned}|A(q_1,n,\chi_1,\chi_5)| &= \prod_{d\in\{f,g,h\}} |A(d(q_1,n),n,\chi_{1,d},\chi_{2,d})|\\ &\leqq k^{18}(c_6k^2)^{\omega(q_1)}g^2(q_1,n)h^{17/9}(q_1,n)\\ &\ll (c_6k^2)^{\omega(q_1)}g^{1/9}(q_1,n)q_1^{17/9}\\ &\ll (c_6k^2)^{\omega(q_1)}(q_1,n)^{2/3}q_1^{17/9}.\end{aligned}\tag{5.13}$$

Let

$$A(x,q_1)=\left|n\in[(9/10)x,x[\,,\ (q_1,n)\geqq q_1^{1/10}\right|,$$

$$B(q_1)=\left|m \bmod q_1,\ (q_1,m)\geqq q_1^{1/10}\right|.$$

Then we have obviously

$$A(x,q_1)\ll\left(\frac{x}{q_1}+1\right)B(q_1),$$

and

$$B(q_1)\leqq\sum_{\substack{d|q_1\\ d\geqq q_1^{1/10}}}\frac{q_1}{d}\leqq\tau(q_1)q_1^{9/10},$$

from which we deduce

$$A(x,q_1)\ll xq_1^{-1/10}\tau(q_1)\leqq xq_3^{-3/40}\tau(q_3)\ll xq_3^{-1/16}.\tag{5.14}$$

The lemma follows now from (5.8), (5.13), (5.14) and $(c_6k^2)^{\omega(q_1)}\ll_\epsilon q_1^\epsilon$.

LEMMA 5.4. *For all n and all l holds:*

$$\sigma(n,R,l)\ll(\log R)^{k+1}.$$

PROOF. From Lemmas 4.3 (b), 4.4 (a), (4.2) and Lemma 4.5 in [14] it follows that

$$|\sigma(n,R,l)| \leqq \sum_{\substack{q\leqq R\\(q,l)=1}} \frac{\mu^2(q)|A(q,n)|}{\phi^2(q)} \leqq \sum_{q\leqq R} \frac{qk^{\omega(q)}}{\phi^2(q)} \ll \log R \sum_{q\leqq R} \frac{k^{\omega(q)}}{q} \ll (\log R)^{k+1}.$$

LEMMA 5.5. *Let $P = x^d$, where d is a positive constant $\leqq 1/10$. Let be given a set of natural numbers l_i, $1 \leqq i \leqq s \ll (\log x)^{1/3}$, with $\frac{P}{l_i} \geqq P^{\frac{4}{5}}$. Then for sufficiently small d there holds*

$$\sigma\left(n, \frac{P}{l_i}, l_i\right) = \prod_{\substack{p\leqq P\\(p,l_i)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) + O\left(P^{-\frac{1}{16}}\right)$$

for all but $\ll x^{1-\delta_1}$, $\delta_1 \geqq 0$ natural numbers $n \in [(9/10)x, x[$, which satisfy the congruence conditions in (1.1), and for all $i \in \{1,\dots,s\}$.

PROOF. The congruence conditions for n are required because of Lemma 4.5 (c). We first argue for a fixed $l \in \{l_1,\dots,l_s\}$ and set $\frac{P}{l} = R$. Defining $A(q,n,l) = \mu((q,l)^2)A(q,n)$ and noting Lemma 4.3 (b) and (4.2), we obviously have to estimate:

$$\begin{aligned} &\left|\sum_{q\leqq R} \frac{A(q,n)}{\phi^2(q)} - \prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\right| \\ &\leqq \left|\sum_{\substack{R<q<V\\ q\in\mathcal{D}}} \phi^{-2}(q)A(q,n,l)\right| + \left|\sum_{\substack{q\geqq V\\ q\in\mathcal{D}}} \phi^{-2}(q)A(q,n)\right| \\ &=: T_1(n,R) + T_2(n,R), \end{aligned} \tag{5.15}$$

where $V = \exp\left(\frac{\log P \log x}{\log\log x}\right)$ and

$$\mathcal{D} = \{q : q \in N,\ \mu(q) \neq 0,\ p|q \Rightarrow p \leqq P\}.$$

We first estimate $T_1(n,R)$. We have:

$$\phi^{-2}(q)A(q,n) = \phi^{-2}(q) \sum_{m|q} A_1(m,n)A_2(q/m,n), \tag{5.16}$$

where we define by Lemma 4.4 (a) and $w(n,p) = 0$ for $p|n$:

$$A_1(p,n) = \begin{cases} -\mu((p,l)^2)p(w(p,n)-1) & p\nmid n, \\ 0 & p|n, \end{cases}$$

$$A_2(p,n,l)=\begin{cases}-\mu((p,l)^2) & p\nmid n,\\ \mu((p,l)^2)(p-1) & p\mid n,\end{cases}$$
$$A_i(q,n)=\prod_{p\mid q}A_i(p,n),\ i\in\{1,2\},$$

and an empty product is equal to 1. For $p\nmid n$ it holds

$$w(n,p)=\left|m:m^k\equiv n(\mathrm{mod}\,p),m\in(1,2,\dots,p)\right|.$$

We obtain by Lemma 4.3 in [13], (4.2) and $|\tau(\chi)|\leqq p^{1/2}$ for $p\nmid n$:

$$\begin{aligned}w(n,p)&=1+\frac{1}{p}\sum_{a=1}^{p-1}e\left(\frac{-n}{p}a\right)\sum_{m=1}^{p}e\left(\frac{m^k}{p}a\right)\\&=1+\frac{1}{p}\sum_{\chi\in\mathcal{A}(p)}\tau(\chi)\sum_{a=1}^{p-1}e\left(\frac{-n}{p}a\right)\overline{\chi(a)}\\&=1+\frac{1}{p}\sum_{\chi\in\mathcal{A}(p)}|\tau(\chi)|^2\chi(-n)=1+\sum_{\chi\in\mathcal{A}(p)}\chi(-n),\end{aligned}$$

where $\mathcal{A}(p)$ denotes the set of non-principal characters χ modulo p, for which χ^k is the principal character and

$$|\mathcal{A}(p)|=(k,p-1)-1. \tag{5.17}$$

So we deduce for all p:

$$A_1(p,n)=-\mu((p,l)^2)p\sum_{\chi\in\mathcal{A}(p)}\chi(-n). \tag{5.18}$$

We obtain from (5.15) and (5.16)

$$\begin{aligned}T_1(n,R)\leqq&\sum_{\substack{R^{1/3}<m<V\\ m\in\mathcal{D}}}\phi^{-2}(m)|A_2(m,n)|\sum_{\substack{R/m<d<V/m,\,(d,m)=1\\ d\in\mathcal{D}}}\phi^{-2}(d)\,|A_1(d,n)|\\&+\sum_{\substack{m\leqq R^{1/3}\\ m\in\mathcal{D}}}\phi^{-2}(m)|A_2(m,n)|\left|\sum_{\substack{R/m<d<V/m\,(d,m)=1\\ d\in\mathcal{D}}}\phi^{-2}(d)A_1(d,n)\right|\\=:&F_1(n,R)+F_2(n,R).\end{aligned} \tag{5.19}$$

For $F_1(n,R)$ we get by $w(n,p)\leqq k$:

$$
\begin{aligned}
F_1(n,R) &\leqq R^{-1/3} \sum_{\substack{m<V\\ m\in\mathcal{D}}} \phi^{-2}(m) m |A_2(m,n)| \sum_{\substack{d<V\\ d\in\mathcal{D}}} \phi^{-2}(d)|A_1(d,n)| \\
&\leqq R^{-1/3} \prod_{\substack{p\leqq P\\ p\nmid n}} \left(1+\frac{p|A_2(p,n)|}{(p-1)^2}\right) R^{-1/3} \prod_{\substack{p\leqq P\\ p\mid n}} \left(1+\frac{p|A_2(p,n)|}{(p-1)^2}\right) \\
&\quad \times \prod_{p\leqq P} \left(1+\frac{|A_1(p,n)|}{(p-1)^2}\right) \qquad (5.20)\\
&\leqq R^{-1/3} \prod_{p\leqq P} \left(1+\frac{4}{p}\right) 3^{\omega(n)} \prod_{p\leqq P} \left(1+\frac{4(k-1)}{p}\right) \\
&\ll R^{-1/3} (\log P)^{4k} 3^{\omega(n)}.
\end{aligned}
$$

For the estimation of $F_2(n,R)$ we obtain by the definition of $A_1(d,n)$ and (5.18):

$$
\begin{aligned}
&\sum_{\substack{R/m<d<V/m,\ (d,m)=1\\ d\in\mathcal{D}}} \phi^{-2}(d) A_1(d,n) \\
&= \sum_{\substack{R/m<d<V/m,\ (d,m)=1\\ d\in\mathcal{D}}} \prod_{p\mid d} \left(-\frac{\mu((p,l)^2)}{(p-1)^2} p \sum_{\chi\in\mathcal{A}(p)} \chi(-n)\right) \qquad (5.21)\\
&= \sum_{\substack{R/m<d<V/m\\ d\in\mathcal{D}}} \sum_{\chi \bmod d}{}^{*} f(\chi)\chi(-n),
\end{aligned}
$$

where

$$
f(\chi) = \begin{cases} \prod_{p\mid d} \left(-\frac{p}{(p-1)^2}\right) & \text{if } \chi = \prod_{p\mid d} \chi_p \text{ with } \chi_p \in \mathcal{A}(p)\ \forall p\mid d,\ (ml,d)=1, \\ 0 & \text{otherwise.} \end{cases}
$$

By (5.17) we find for any positive number a and any $d\in\mathcal{D}$:

$$
\sum_{\chi \bmod d}{}^{*} |f(\chi)|^a \leqq (k-1)^{\omega(d)} \left(\prod_{p\mid d} \frac{p}{(p-1)^2}\right)^a \leqq \frac{(4^a(k-1))^{\omega(d)}}{d^a}. \qquad (5.22)
$$

Now we get from (5.21):

(5.23)

$$
\sum_{\substack{R/m<d<V/m,\ (d,m)=1\\ d\in\mathcal{D}}} \phi^{-2}(d) A_1(d,n,l) = \sum_{j=1}^{L} \sum_{\substack{Q_{j-1}<d\leqq Q_j,\\ d\in\mathcal{D}}} \sum_{\chi \bmod d}{}^{*} f(\chi)\chi(-n),
$$

where $Q_0 = R/m$, $Q_j = x^{j/2}$, $j = 1, \dots, L$, $L \leqq 2\dfrac{\log P}{\log\log x}$.

We have for a fixed j by (5.22), Lemma 6.5 in [6] and Lemma 4.5 in [14]:

$$\sum_{n\in[(9/10)x,x[}\left|\sum_{\substack{Q_{j-1}<d\leqq Q_j\\ d\in\mathcal{D}}}\sum_{\chi \bmod d}{}^{*} f(\chi)\chi(-n)\right|$$

$$\ll \left(x^{1/2}+Q_j^{1/j}\right)x^{1/2}\left(\log\left(x^j e\right)\right)^{(j^2-1)/2j}$$

(5.24)

$$\times\left(\sum_{Q_{j-1}<d\leqq Q_j}\sum_{\chi \bmod d}{}^{*}|f(\chi)|^{2j/(2j-1)}\right)^{(2j-1)/2j}$$

$$\ll x\left(\log\left(x^j e\right)\right)^{(j^2-1)/2j}\left(\frac{1}{Q_{j-1}^{1/(2j-1)}}\sum_{Q_{j-1}<d\leqq Q_j}\frac{(16k)^{\omega(d)}}{d}\right)^{(2j-1)/2j}$$

$$\ll \left(\frac{1}{Q_{j-1}^{1/(2j-1)}}(\log Q_j)^{16k}\right)^{(2j-1)/2j} x\left(\log\left(x^j e\right)\right)^{(j^2-1)/2j} Q_{j-1}^{-1/2j}(\log Q_j)^{16k}.$$

We deduce from (5.23) and (5.24)

$$\text{(5.25)}\qquad \sum_{n\in[(9/10)x,x]}\left|\sum_{\substack{R/m<d<V/m,\ (d,m)=1\\ d\in\mathcal{D}}}\phi^{-2}(d)A_1(d,n,l)\right|$$

$$\ll xQ_0^{-1/2}(\log x)^{16k}+x^{7/8}(\log x)^{32k}\sum_{j=2}^{L}\left(\log x^{j+1}\right)^{(j^2-1)/2j}.$$

For the sum in (5.25) we get for a sufficiently small d

$$\sum_{j=2}^{L}(\dots)\leqq\sum_{j=2}^{L}((j+1)\log x)^{j/2}\leqq 2\frac{\log P}{\log\log x}\left(3\frac{\log P}{\log\log x}\log x\right)^{\frac{\log P}{\log\log x}}\ll P^3.$$

From this and (5.25) it follows together with the definition of Q_0, $m\leqq R^{1/3}$ and a sufficiently small d:

$$\text{(5.26)}\qquad \sum_{n\in[(9/10)x,x[}\left|\sum_{\substack{R/m<d<V/m\ (d,m)=1\\ d\in\mathcal{D}}}\phi^{-2}(d)A_1(d,n)\right|$$

$$\ll x(\log x)^{32k}\left(P^{-\frac{1}{2}\frac{2}{3}\frac{4}{5}}+P^3x^{-1/8}\right)\ll xP^{-1/9}.$$

In order to finish the estimate of $F_2(n,R)$, we need the following result:

(5.27)
$$\sum_{\substack{m\leqq R^{1/3}\\ m\in\mathcal{D}}} \phi^{-2}(m)|A_2(m,n)| \leqq \prod_{\substack{p\leqq P\\ p\mid n}}\left(1+\frac{1}{p-1}\right)\prod_{\substack{p\leqq P\\ p\nmid n}}\left(1+\frac{1}{(p-1)^2}\right) \ll 2^{\omega(n)}.$$

Then (5.19), (5.26), (5.27) and $2^{\omega(n)} \leqq \tau(n) \ll_\epsilon n^\epsilon$ imply

$$\sum_{n\in[(9/10)x,x]}\sum_{i=1}^{s} F_2(n,P/l_i) \ll xP^{-1/10}. \tag{5.28}$$

So from the last expression, (5.19) and (5.20) we derive for all but $\ll x^{1-\delta_1}$ $n\in[(9/10)x,x[$, that satisfy the congruence conditions in (1.1):

$$T_1(n,P/l_i) \ll P^{-1/16} \tag{5.29}$$

for all l_i, $i\in\{1,\dots,s\}$. By Lemma 4.3 (b) we get for $T_2(n,R)$ and $v=\dfrac{\log\log x}{2\log P}$:

$$T_2(n,R) \leqq \sum_{q\in\mathcal{D}}\left(\frac{q}{V}\right)^v \phi^{-2}(q)\,|A(q,n,l)| \leqq V^{-v}\prod_{p\leqq P}\left(1+p^v\frac{|A(p,n)|}{(p-1)^2}\right).$$

By

$$V^{-v}=x^{-1/2}$$

and

$$p^v \leqq (\log x)^{1/2},$$

it follows for a sufficiently small d:

$$\begin{aligned} T_2(n,R) &\leqq x^{-1/2}\prod_{p\leqq P}\left(1+\frac{4k(\log x)^{1/2}}{p}\right)\\ &\ll x^{-1/2}(\log P)^{4k(\log x)^{1/2}} \ll x^{-1/3}. \end{aligned} \tag{5.30}$$

From (5.15), (5.29) and (5.30) the lemma follows.

LEMMA 5.6. (a) *For all n that satisfy the congruence conditions in* (1.1)

$$\prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right) \gg (\log P)^{-2k}.$$

(b) *For any two primitive characters* $\chi_1 \pmod{q_1}$ *and* $\chi_2 \pmod{q_2}$, $q_3 = [q_1 q_2] \leqq P$ *and all* n, *which satisfy the congruence condition in* (1.1)

$$\left|\frac{A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3})}{\phi^2(q_3)}\right| \prod_{\substack{p \leqq P \\ (p,q_3)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) \ll \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right).$$

holds true.

PROOF. (a) By $0 \leqq w(n,p) \leqq (k, p-1)$ and Lemma 4.4 (d):

$$\prod_{p \leqq P} \left(1 - \frac{(w(n,p)-1)p+1}{(p-1)^2}\right) \gg \prod_{2k < p \leqq P} \left(1 - \frac{2k}{p}\right) \gg (\log P)^{-2k},$$

from which the lemma can be deduced by Lemma 4.4 (a).

(b) If $q_3 = 1$, the lemma is obvious. For $q_3 > 1$ we distinguish the cases (i) $q_1 = q_3$ and (ii) $1 \leqq q_1 < q_3$. In the case (i) we immediately get the desired result from Lemma 4.4 (a) and (b) by

$$\left|\frac{A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3})}{\phi^{-2}(q_3)}\right| \prod_{\substack{p \leqq P \\ (p,q_3)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) \leqq \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right).$$

(ii) Analogously to (4.2) we have only to take into consideration such pairs q_3 and q_1, for which $\left(\frac{q_3}{q_1}, q_1\right) = 1$ and so, by Lemma 4.3 (b),

$$A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3}) = A(q_1, n, \chi_1, \chi_5) A\left(\frac{q_3}{q_1}, n, \chi_{0,\frac{q_3}{q_1}}, \chi_6\right),$$

for certain characters χ_5 and χ_6. Since in (4.6)

$$A\left(\frac{q_3}{q_1}, n, \chi_{0,\frac{q_3}{q_1}}, \chi_6\right) \ll \left(\frac{q_3}{q_1}\right)^{3/2+\epsilon},$$

furthermore, by Lemma 4.4 (a) and (d)

$$\prod_{p \mid \frac{q_3}{q_1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right)^{-1} \ll \prod_{p \mid \frac{q_3}{q_1}, p > 4k} \left(1 - \frac{2k}{p}\right)^{-1} \leqq 2^{\omega\left(\frac{q_3}{q_1}\right)} \ll_\epsilon \left(\frac{q_3}{q_1}\right)^{\epsilon}.$$

Using all this we get together with the result from (i)

$$\left|\frac{A(q_3, n, \chi_1\chi_{0,q_3}, \chi_2\chi_{0,q_3})}{\phi^2(q_3)}\right| \prod_{\substack{p \leqq P \\ (p,q_3)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) = \left|\frac{A(q_1, n, \chi_1, \chi_5)}{\phi^2(q_1)}\right|$$

$$\times \prod_{\substack{p\leqq P\\(p,q_1)=1}} \left(1+\frac{A(p,n)}{(p-1)^2}\right) \left| \frac{A\left(\frac{q_3}{q_1},n,\chi_{0,\frac{q_3}{q_1}},\chi_6\right)}{\phi^2\left(\frac{q_3}{q_1}\right)} \right| \prod_{p\mid\frac{q_3}{q_1}} \left(1+\frac{A(p,n)}{(p-1)^2}\right)^{-1}$$

$$\ll \prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\phi^{-2}\left(\frac{q_3}{q_1}\right)\left(\frac{q_3}{q_1}\right)^{3/2+\epsilon}\left(\frac{q_3}{q_1}\right)^{\epsilon} \ll \prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right).$$

6. The minor arcs

We obtain by Bessel's inequality and the prime number theorem

$$\sum_{(9/10)x\leqq n<x} r_2(x,n)^2 \leqq \int_m |S(\alpha)S_k(\alpha)|^2 d\alpha \ll x\log x \sup_{\alpha\in m}|S_k(\alpha)|^2.$$

By the definiton of the minor arcs and Theorem 1 in [4] we have

$$\sup_{\alpha\in m}|S_k(\alpha)| \ll x^{\frac{1+\epsilon}{k}}\left(\frac{1}{P}+\frac{1}{x^{1/2k}}+\frac{Q}{x}\right)^{1/4^{k-1}} \ll \frac{x^{\frac{1+\epsilon}{k}}}{P^{1/4^{k-1}}}.$$

Substituting this in the first estimate we obtain

$$\sum_{(9/10)x\leqq n<x} r_2(x,n)^2 \ll \frac{x^{1+(2/k)+\epsilon}}{P^{2/4^{k-1}}}. \tag{6.1}$$

7. The major arcs

Let us suppose in the following $l\in\{1,k\}$ and $S(\alpha)=S_1(\alpha)$. For $\alpha\in I(a,q)$ let $\alpha=\frac{a}{q}+\eta$. Because of $q\leqq P$ and $p>P$ for all p appearing in $S_l(\alpha)$ we get in a well-known way:

$$\begin{aligned} S_l(\alpha) &= \sum_{\frac{\sqrt{x}}{2}\leqq p<\sqrt{x}} \log p\, e\left(\frac{a}{q}p^l+\eta p^l\right) = \frac{1}{\phi(q)}\sum_{\chi \bmod q}\sum_{h=1}^{q}\overline{\chi}(h)e\left(\frac{ah^l}{q}\right) \\ &\times \sum_{\frac{\sqrt{x}}{l}\leqq p<\sqrt{x}} \chi(p)\log p\, e\left(\eta p^l\right) = \frac{1}{\phi(q)}\sum_{\chi \bmod q} C_l(\overline{\chi},a)S_l(\chi,\eta). \end{aligned} \tag{7.1}$$

Let $L=T$ if $l=1$ and $L=F$ if $l=k$. Now $W_l(\chi,\eta)$ is defined in the following way:

(i) For $\chi = \chi_{0,q}$ let

$$W_l(\chi,\eta) = S_l(\chi_{0,q},\eta) - L(\eta) + \sum_{\substack{\varrho\in\theta'\cup\tilde{\beta}\\ \zeta(\varrho)=0}} L_\varrho(\eta).$$

(ii) For $\chi = \chi_{0,q}\chi^*$ with $\chi^* \in \theta\cup\tilde{\chi}$, $\chi^* \neq \chi_{0,1}$ let

$$W_l(\chi,\eta) = S_l(\chi_{0,q}\chi^*,\eta) + \sum_{\substack{\varrho\in\theta'\cup\tilde{\beta}\\ L(\varrho,\chi^*)=0}} L_\varrho(\eta).$$

(iii) In all other cases let

$$W_l(\chi,\eta) = S_l(\chi,\eta).$$

We obtain

$$S_l\left(\frac{a}{q}+\eta\right) = \frac{1}{\phi(q)}C_l(\chi_0,a)L(\eta) + \frac{1}{\phi(q)}D_l(a,q,\eta) + \frac{1}{\phi(q)}E_l(a,q,\eta),$$

where

$$D_l(a,q,\eta) = \sum_{\chi \bmod q} C_l(\overline{\chi},a)W_l(\chi,\eta),$$

$$E_l(a,q,\eta) = -\sum_{\substack{\chi\in\theta\cup\tilde{\chi}\\ \text{cond }\chi\mid q}}\ \sum_{\substack{\varrho\in\theta'\cup\tilde{\beta}\\ L(\varrho,\chi)=0}} C_l(\chi_{0,q}\overline{\chi},a)L_\varrho(\eta).$$

Writing $W = W_1$, $E = E_1$ and $D = D_1$ we obtain from (3.6) and (7.1)

$$\begin{aligned} r_1(x,n) &= \sum_{q\leqq P}\sideset{}{^*}\sum_{a=1}^{q} e\left(\frac{-an}{q}\right)\int_{-1/Q}^{1/Q} S\left(\frac{a}{q}+\eta\right)S_k\left(\frac{a}{q}+\eta\right)e(-n\eta)d\eta \\ &= \sum_{q\leqq P}\frac{1}{\phi^2(q)}A(q,n)\int_{-1/Q}^{1/Q} T(\eta)F(\eta)e(-n\eta)d\eta \\ &\quad + \sum_{q\leqq P}\frac{1}{\phi^2(q)}\sideset{}{^*}\sum_{a=1}^{q} e\left(-\frac{an}{q}\right)C_1(\chi_0,a)\int_{-1/Q}^{1/Q} T(\eta)D_k(a,q,\eta)e(-n\eta)d\eta \\ &\quad + \sum_{q\leqq P}\frac{1}{\phi^2(q)}\sideset{}{^*}\sum_{a=1}^{q} e\left(-\frac{an}{q}\right)C_1(\chi_0,a)\int_{-1/Q}^{1/Q} T(\eta)E_k(a,q,\eta)e(-n\eta)d\eta \end{aligned}$$

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)C_k(\chi_0,a)\int_{-1/Q}^{1/Q}F(\eta)D(a,q,\eta)e(-n\eta)d\eta$$

(7.2)

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)C_k(\chi_0,a)\int_{-1/Q}^{1/Q}F(\eta)E(a,q,\eta)e(-n\eta)d\eta$$

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)\int_{-1/Q}^{1/Q}D_k(a,q,\eta)E(a,q,\eta)e(-n\eta)d\eta$$

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)\int_{-1/Q}^{1/Q}D_k(a,q,\eta)D(a,q,\eta)e(-n\eta)d\eta$$

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)\int_{-1/Q}^{1/Q}D(a,q,\eta)E_k(a,q,\eta)e(-n\eta)d\eta$$

$$+\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{a=1}^{q}{}^{*}e\left(-\frac{an}{q}\right)\int_{-1/Q}^{1/Q}E(a,q,\eta)E_k(a,q,\eta)e(-n\eta)d\eta$$

$$=: S_1+S_2+S_3+S_4+S_5+S_6+S_7+S_8+S_9.$$

In the following we only take into consideration such $n\in[(9/10)x,x[$, that satisfy the congruence conditions in (1.1).

8. The calculation of S_1–S_9

We first estimate S_4. Changing the summation over the characters according to the inducing primitive characters, we get by Lemma 4.7 (a) and Cauchy's inequality:

$$S_4=\sum_{q\leqq P}\frac{1}{\phi^2(q)}\sum_{\chi \bmod q}A(q,n,\overline{\chi},\chi_0)\int_{-1/Q}^{1/Q}F(\eta)W(\chi,\eta)e(-n\eta)d\eta$$

(8.1)

$$\ll x^{(1/k)-(1/2)}\sum_{r\leqq P}\sum_{\chi \bmod r}{}^{*}\sum_{\substack{q\leqq P\\ q\equiv 0(\bmod r)}}\frac{1}{\phi^2(q)}\,|A(q,n,\overline{\chi}\chi_{0,q},\chi_{0,q})|$$

$$\times \left(\int_{-1/Q}^{1/Q} |W(\chi_{0,q}\chi, \eta)|^2 d\eta \right)^{1/2}.$$

Because of $q \leqq P$ and $p > P$ we have $W(\chi_{0,q}\chi, \eta) = W(\chi, \eta)$, and so we get by (8.1) and Lemma 4.5

$$S_4 \ll x^{(1/k)-(1/2)} \sum_{r \leqq P} \sum_{\chi \bmod r}{}^{*} \left(\int_{-1/Q}^{1/Q} |W(\chi, \eta)|^2 d\eta \right)^{1/2}$$

$$\times \sum_{\substack{q \leqq P \\ q \equiv 0 (\bmod r)}} \frac{1}{\phi^2(q)} |A(q, n, \overline{\chi}\chi_{0,q}, \chi_{0,q})| \tag{8.2}$$

$$\ll \log^{5k+1} x \sum_{r \leqq P} \sum_{\chi \bmod r}{}^{*} \left(\int_{-1/Q}^{1/Q} |W(\chi, \eta)|^2 d\eta \right)^{1/2}.$$

We define now for an arbitrary primitive character $\chi \bmod r$:

$$\sum_{t}^{t+h}{}^{\#} \chi(p) \log p = \begin{cases} \sum_{t}^{t+h} \log p - \sum_{t}^{t+h} 1 & \text{if } r = 1, \\ \sum_{t}^{t+h} \chi(p) \log p & \text{if } r > 1. \end{cases}$$

Then we get by Lemma 1 in [3] and the definition of $W(\chi, \eta)$:

$$\int_{-1/qQ}^{1/qQ} |W(\chi, \eta)|^2 d\eta \ll \int_{\frac{x}{4}}^{x} \left| \frac{1}{Q} \sum_{\substack{t \leqq p \leqq t + \frac{Q}{2}, \\ \frac{x}{2} \leqq p < x}}{}^{\#} \chi(p) \log p \right|^2 dt,$$

from which we get by (8.2):

$$S_4 \ll x^{1/k} \log^{5k+1} x \sum_{r \leqq P} \sum_{\chi \bmod r}{}^{*} \max_{x/4 \leqq t \leqq x} \max_{h \leqq xP^{-4k-3}} (h + xP^{-4k-3})^{-1}$$

$$\times \left| \sum_{t}^{t+h}{}^{\#} \chi(p) \log p \right|. \tag{8.3}$$

Arguing exactly as in (19) in [1] we obtain for the last double sum

$$\sum\sum \ll \delta^{8k^2+1} \log^{-8k^2} x + P^{-1}. \tag{8.4}$$

If we combine (8.3) and (8.4) and argue in the same way for S_8, where we use the upper (3.1) for the number of the *P-exceptional zeros* over which is summed in S_8, we obtain

$$S_4 + S_8 \ll \frac{\delta^{8k^2+1} x^{1/k}}{\log^{8k^2-5k-1,5} x} + \frac{x^{1/k} \log^{5k+1,5} x}{P}. \tag{8.5}$$

Using Lemma 4.7 (b) we get in the same way for S_7:

$$\begin{aligned} S_7 &= \sum_{q \leqq P} \frac{1}{\phi^2(q)} \sum_{\chi \bmod q} \sum_{\chi_1 \bmod q} A(q, n, \overline{\chi_1}, \overline{\chi_2}) \int_{-1/Q}^{1/Q} W(\chi, \eta) W_k(\chi_1, \eta) e(-n\eta) d\eta \\ &\ll \left(\delta^{8k^2+1} x^{1/2} \log^{5k+1-8k^2} x + \frac{x^{1/2} \log^{5k+1} x}{P} \right) \\ &\quad \times \sum_{r_1 \leqq P} \sum_{\chi_1 \bmod r_1}{}^{*} \left(\int_{-1/Q}^{1/Q} |W_k(\chi_1, \eta)|^2 d\eta \right)^{1/2}. \end{aligned} \tag{8.6}$$

Arguing as in (8.3) and (8.4) we derive from this

$$S_7 \ll \left(\delta^{8k^2+1} x^{1/k} \log^{5k+1-8k^2} x + \frac{x^{1/k} \log^{5k+1} x}{P} \right) W_k, \tag{8.7}$$

where

$$\begin{aligned} W_k = \sum_{r_1 \leqq P} \sum_{\chi_1 \bmod r_1}{}^{*} & \max_{\sqrt[k]{x/(2^{k+1})} \leqq y \leqq \sqrt[k]{x}} \max_{h \ll x^{1/k} P^{-4k-3}} (h + \sqrt[k]{x} P^{-4k-3})^{-1} \\ & \times \left| \sum_{y}^{y+h}{}^{\#} \chi_1(p) \log p \right|, \end{aligned}$$

and

$$W_k \ll \delta^{8k+1} \log^{-8k} x + P^{-1}. \tag{8.8}$$

Combining (8.7) and (8.8) and arguing in the same way for S_2 and S_6 by using again (3.1) we obtain

$$S_2 + S_6 + S_7 \ll \frac{\delta^{8k+1} x^{1/k}}{\log^{3k-1,5} x} + \frac{x^{1/k} \log^{5k+1,5} x}{P}. \tag{8.9}$$

For S_1 we get by the Lemmas 4.5 and 4.9

$$
\begin{aligned}
S_1 &= \sum_{q \leqq P} \frac{1}{\phi^2(q)} A(q,n) \int_0^1 T(\eta) F(\eta) e(-n\eta) d\eta \\
&\quad + O\left(\sum_{q \leqq P} \frac{1}{\phi^2(q)} |A(q,n)| \int_{1/Q}^{1/2} |T(\eta) F(\eta)| d\eta \right) \\
&= \sigma(n,P) L(x,n) + O\left(x^{1/k} P^{\frac{-2k}{s}} \right).
\end{aligned} \tag{8.10}
$$

Noting that in the sum defining S_3 by (4.2) we only have to take into consideration such q with l cond $\chi = q$, for which $(l,\ \text{cond}\ \chi) = 1$ holds, we obtain in the same way as for S_1:

$$
\begin{aligned}
S_3 &= -\sum_{r \leqq P} \sum_{\substack{\chi \in \theta \cup \tilde{\chi} \\ \chi \bmod r}} \sum_{\substack{\varrho \in \theta' \cup \tilde{\beta} \\ L(\varrho,\chi)=0}} \frac{1}{\phi^2(r)} A(r,n,\chi_{0,r},\overline{\chi}) \sigma\left(n, \frac{P}{r}, r\right) L_{1,\varrho}(x,n) \\
&\quad + O\left(x^{1/k} P^{\frac{-2k}{s}} \right).
\end{aligned} \tag{8.11}
$$

For the calculation of the remaining terms we define

$$
\theta_1' = \left\{ \varrho \in \theta' \cup \tilde{\beta} :\ |\gamma| \leqq P^{4k+3} \right\}, \qquad \theta_2' = \theta' \cup \tilde{\beta} \setminus \theta_1',
$$

such that by (3.2):

$$
\varrho = \beta + i\gamma \in \theta_2' \Longrightarrow |\gamma| > 16 P^{4k+3}. \tag{8.12}
$$

So we obtain

(8.13)

$$
\begin{aligned}
S_5 &= -\sum_{\chi \in \theta \cup \tilde{\chi}} \sum_{\substack{\varrho \in \theta_1' \\ L(\varrho,\chi)=0}} \sum_{\substack{q \leqq P \\ \text{cond}\,\chi \mid q}} \frac{1}{\phi^2(q)} A(q,n,\overline{\chi}\chi_{0,q},\chi_{0,q}) \int_{-1/Q}^{1/Q} T_\varrho(\eta) F(\eta) e(-n\eta) d\eta \\
&\quad -\sum_{\chi \in \theta} \sum_{\substack{\varrho \in \theta_2' \\ L(\varrho,\chi)=0}} \sum_{\substack{q \leqq P \\ \text{cond}\,\chi \mid q}} \frac{1}{\phi^2(q)} A(q,n,\overline{\chi}\chi_{0,q},\chi_{0,q}) \int_{-1/Q}^{1/Q} T_\varrho(\eta) F(\eta) e(-n\eta) d\eta \\
&=: S_{5,1} + S_{5,2}.
\end{aligned}
$$

We first get from Lemma 4.5, Lemma 4.10 and (3.1)

$$S_{5,2} \leqq \sum_{\chi \in \theta} \sum_{\substack{\varrho \in \theta_2' \\ L(\varrho,\chi)=0}} \sum_{\substack{q \leqq P \\ \operatorname{cond} \chi | q}} \frac{1}{\phi^2(q)} |A(q, n, \overline{\chi}\chi_{0,q}, \chi_{0,q})|$$

$$\times \int_{-1/Q}^{1/Q} |T_\varrho(\eta) F(\eta)| d\eta \ll x^{1/k} P^{-2k}. \tag{8.14}$$

Arguing as for S_3, we get by appealing again to (3.1)

$$S_5 = -\sum_{r \leqq P} \sum_{\substack{\chi \in \theta \cup \tilde{\chi} \\ \chi \bmod r}} \sum_{\substack{\varrho \in \theta_1' \\ L(\varrho,\chi)=0}} \frac{1}{\phi^2(r)} A(r, n, \overline{\chi}, \chi_{0,r})$$

$$\times \sigma\left(n, \frac{P}{r}, r\right) L_{\varrho,1}(x, n) + O\left(x^{1/k} P^{\frac{-2k}{s}}\right). \tag{8.15}$$

For S_9 we get similarly to S_5

$$S_9 = \sum_{\chi \in \theta \cup \tilde{\chi}} \sum_{\substack{\varrho \in \theta_1' \\ L(\varrho,\chi)=0}} \sum_{\chi_1 \in \theta \cup \tilde{\chi}} \sum_{\substack{\varrho' \in \theta' \cup \tilde{\beta} \\ L(\varrho',\chi_1)=0}} \sum_{\substack{r \leqq P \\ [\operatorname{cond} \chi, \operatorname{cond} \chi_1]=r}} \frac{1}{\phi^2(r)} A(r, n, \overline{\chi}\chi_{0,r}, \overline{\chi_1}\chi_{0,r})$$

$$\times \sigma\left(n, \frac{P}{r}, r\right) L_{\varrho,\varrho'}(x, n) + O\left(x^{1/k} P^{\frac{-2k}{s}}\right). \tag{8.16}$$

9. Proof of the theorem

We first notice that obviously

$$|L_{\varrho,\varrho'}(X, n)| = \Big| \sum_{\substack{n - x < m^k \leqq n - (x/2) \\ \frac{\sqrt[k]{x}}{2} \leqq m < \sqrt[k]{x}}} (n - m^k)^{\varrho - 1} m^{\varrho' - 1} \Big| \ll x^{1/k} x^{\beta - 1} x^{\beta' - 1}. \tag{9.1}$$

Arguing in exactly the same way as in (35) in [1] or in Lemma 2.1 in [8], we obtain further that

$$\sum_{\varrho \in \theta'} x^{\beta - 1} + \sum_{\varrho \in \theta'} \sum_{\varrho' \in \theta'} x^{\beta - 1} x^{\beta' - 1} \leqq c_6 \exp\left(-\frac{c_1}{2b}\right) \delta^2 + x^{-1/2}, \tag{9.2}$$

where in the sequel we will neglect $x^{1/2}$, which in (9.10) will be shown to be permissible. We define further

$H = \{r = [r_1 r_2],\ r_i = P$-excluded module or exceptional module to P or $1\}$,

$G = \{r \in H,\ r \geqq P^{1/5}\}$.

Then we derive from Lemmas 5.3 and 5.4 and (9.1) that for any two characters $\chi_1 (\mathrm{mod}\ r_1), \chi_2 (\mathrm{mod}\ r_2) \in \{\theta \cup \tilde{\chi} \cup \chi_{0,1}\}$ with $r = [r_1, r_2] =:, r \in G$,

$$\frac{1}{\phi^2(r)} |A(r, n, \chi_1\chi_{0,r}, \chi_2\chi_{0,r})| |\sigma\Big(n, \frac{P}{r}, r\Big)| |L_{\varrho,\varrho'}(x,n)| \ll x^{1/k} P^{-1/240}, \tag{9.3}$$

holds for all but $\ll xP^{-1/80}$ $n \in [(9/10)x, x[$. If – in view of (3.1) – we apply Lemma 5.5 to all $r \in H \setminus G$ and note that $\bar{r} \notin G$ for a sufficiently small λ, then for all $n \in [(9/10)x, x[$ that satisfy the congruence conditions in (1.1) and $n \notin A(x)$ with $|A(x)| \ll xP^{-1/80} (\log x)^{1/3} + x^{1-\delta_1} \ll x^{1-\delta_2}$, $\delta_2 \geqq 0$, there holds by (7.2), (8.5), (8.9), (8.10), (8.11), (8.15) and (8.16):

$$\begin{aligned}
r_1(x,n) = {} & \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L(x,n) \\
& - \frac{1}{\phi^2(\tilde{r})} A(\tilde{r}, n, \tilde{\chi}, \chi_{0,\tilde{r}}) \prod_{\substack{p \leqq P \\ (p,\tilde{r})=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{\tilde{\beta},1}(x,n) \\
& - \frac{1}{\phi^2(\tilde{r})} A(\tilde{r}, n, \chi_{0,\tilde{r}}, \tilde{\chi}) \prod_{\substack{p \leqq P \\ (p,\tilde{r})=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{1,\tilde{\beta}}(x,n) \\
& + \frac{1}{\phi^2(\tilde{r})} A(\tilde{r}, n, \tilde{\chi}, \tilde{\chi}) \prod_{\substack{p \leqq P \\ (p,\tilde{r})=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{\tilde{\beta},\tilde{\beta}}(x,n) \\
& - \sum_{\substack{r \leqq P \\ r \notin G}} \sum_{\substack{\chi \in \theta \\ \chi \bmod r}} \sum_{\substack{\varrho \in \theta_1' \setminus \beta \\ L(\varrho,\chi)=0}} \frac{1}{\phi^2(r)} A(r, n, \overline{\chi}, \chi_{0,r}) \prod_{\substack{p \leqq P \\ (p,r)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{\varrho,1}(x,n) \\
& - \sum_{\substack{r \leqq P \\ r \notin G}} \sum_{\substack{\chi \in \theta \\ \chi \bmod r}} \sum_{\substack{\varrho \in \theta' \\ L(\varrho,\chi)=0}} \frac{1}{\phi^2(r)} A(r, n, \chi_{0,r}, \overline{\chi}) \prod_{\substack{p \leqq P \\ (p,r)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{1,\varrho}(x,n) \\
& \sum_{\chi \in \theta \cup \tilde{\chi}} \sum_{\substack{\varrho \in \theta_1' \\ L(\varrho,\chi)=0}} \sum_{\chi_1 \in \theta \cup \tilde{\chi}} \sum_{\substack{\varrho' \in \theta' \cup \tilde{\beta} \\ L(\varrho',\chi)=0 \\ (\varrho,\varrho') \neq (\beta,\tilde{\beta})}} \sum_{\substack{r \leqq P,\, r \notin G, \\ [\mathrm{cond}\,\chi,\, \mathrm{cond}\,\chi_1]=r}} \frac{1}{\phi^2(r)} A(r, n, \overline{\chi}\chi_{0,r}, \overline{\chi_1}\chi_{0,r}) \\
& \qquad \times \prod_{\substack{p \leqq P \\ (p,r)=1}} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) L_{\varrho,\varrho'}(x,n) + O(\dots) \\
& = B_1 + \dots + B_7 + O(x^{1/k} P^{\frac{-2k}{s}} + x^{1/k} \delta^2 \log^{1,5-3k} x),
\end{aligned} \tag{9.4}$$

where we have used (3.1) for the calculation of the error term. In the following s will be chosen fixed according to the preceding discussion. We first

get by (9.1), (9.2) and Lemma 5.6 (b)

$$B_5+\cdots+B_7 \ll x^{1/k}\prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\left(\sum_{\varrho\in\theta'}x^{\beta-1}+\sum_{\varrho\in\theta'}\sum_{\varrho'\in\theta'}x^{\beta-1}x^{\beta'-1}\right)$$

$$\leqq c_7\exp\left(\frac{-c_1}{2b}\right)\delta^2x^{1/k}\left|\prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\right|. \tag{9.5}$$

We further derive from Lemma 4.1 (c) and (d) that

$$\prod_{p\leqq P}\left(1+\frac{A(p,n)}{(p-1)^2}\right)=\prod_{\substack{p\leqq P\\(p,\tilde r)=1}}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\frac{\tilde r}{\phi^2(\tilde r)}\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}1. \tag{9.6}$$

In the same way as in the proof of Lemma 4.4 (b) we obtain for the characters $\chi_1,\chi_2\in\{\chi_{0,\tilde r},\tilde\chi\}$, which are not both equal to $\chi_{0,\tilde r}$:

$$A(\tilde r,n,\chi_1,\chi_2)=\tilde r\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}\chi_1(l)\chi_2(m). \tag{9.7}$$

So we get from (9.4), (9.6) and (9.7)

$$\begin{aligned}&B_1+B_2+B_3+B_4\\&=\prod_{\substack{p\leqq P\\(p,\tilde r)=1}}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\frac{\tilde r}{\phi^2(\tilde r)}\\&\times\Bigg((L(x,n))\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}1-L_{\tilde\beta,1}(x,n)\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}\tilde\chi(l)\\&-L_{1,\tilde\beta}(x,n)\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}\tilde\chi(m)+L_{\tilde\beta,\tilde\beta}(x,n)\sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}\tilde\chi(l)\tilde\chi(m)\Bigg)\\&=\prod_{\substack{p\leqq P\\(p,\tilde r)=1}}\left(1+\frac{A(p,n)}{(p-1)^2}\right)\frac{\tilde r}{\phi^2(\tilde r)}\\&\times\Bigg(\sum_{\substack{a+b^k=n\\\frac{x}{2}\leqq a<x\\\frac{\sqrt[k]{x}}{2}\leqq b<\sqrt[k]{x}}}\ \sum_{\substack{l+m^k\equiv n(\mathrm{mod}\,\tilde r)\\1\leqq l,m\leqq\tilde r,(lm,\tilde r)=1}}(1-\tilde\chi(l)a^{\tilde\beta-1})(1-\tilde\chi(m)b^{\tilde\beta-1}\Bigg)\end{aligned} \tag{9.8}$$

$$\geqq \left| \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) \right| \sum_{\substack{a+b^k=n \\ \frac{x}{2} \leqq a < x \\ \frac{\sqrt[k]{x}}{2} \leqq b < \sqrt[k]{x}}} (1-a^{\tilde{\beta}-1})(1-b^{\tilde{\beta}-1}),$$

where in the last step we have again argued as in (9.7). If the Siegel zero $\tilde{\beta}$ exists, we get

$$1 - P^{\tilde{\beta}-1} = (1-\tilde{\beta}) \log P \; P^{\gamma-1} \geqq c_{14}(1-\tilde{\beta}) \log P = c_{14}\delta.$$

Applying this to (9.8) we obtain

$$B_1 + B_2 + B_3 + B_4 \geqq \delta^2 x^{1/k} \left| \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) \right|,$$

which, by (9.8), obviously also holds if $\tilde{\beta}$ does not exist. So we get for a sufficiently small b from the last inequality, (9.5) and Lemma 5.6 (a)

$$(9.9) \quad |B_1 + \cdots + B_7| \gg \delta^2 x^{1/k} \left| \prod_{p \leqq P} \left(1 + \frac{A(p,n)}{(p-1)^2}\right) \right| \frac{1}{2} c_7 \gg \delta^2 x^{1/k} \log^{-2k} x.$$

If $\tilde{\beta}$ exists, we know by Lemma 3.1 and (3.4):

$$(9.10) \qquad \delta^2 = ((1-\tilde{\beta}) \log P)^2 \gg \frac{1}{P^{(4k+3)\lambda/(4k+2)} \log^2 x}.$$

Otherwise $\delta = 1$. We derive from (9.4), (9.9) and (9.10) that for $\lambda \leqq \min\left(\frac{1}{4^{k+1}}, \frac{k}{s}\right)$, $n \in [(9/10)x, x[\setminus A(x)$ and n satisfies the congruence conditions in (1.1):

$$r_1(x,n) \gg x^{1/k} \delta^2 \log^{-2k} x.$$

We further conclude from (6.1) that

$$r_2(x,n) \ll x^{1/k} P^{-1/4^k}$$

for all but $n \in [(9/10)x, x[\setminus B(x)$ with $|B(x)| \ll x^{1+\epsilon} P^{-6/4^k}$. So we get from (3.6) and the upper bound for λ

$$r(x,n) \gg x^{1/k} \delta^2 \log^{-2k} x$$

for all but $|A(x) \cup B(x)| \ll x^{1-\Theta}$, $\Theta > 0$ integers $n \in [(9/10)x, x[$, that satisfy the congruence conditions in (1.1). Splitting the interval $[1, x[$ into intervals of the type $[\frac{9}{10}t, t[$, we get the theorem.

REMARK. The author would like to thank Professor Dr. T. Zhan and Professor Dr. D. Wolke for their steady encouragement.

REFERENCES

[1] BRÜNNER, R., PERELLI, A. and PINTZ, J., The exceptional set for the sum of a prime and a square, *Acta Math. Hungar.* **53** (1989), 347–365. *MR* **91b**:11104

[2] DAVENPORT, H., *Multiplicative number theory*, Second edition, Revised by Hugh L. Montgomery, Graduate Texts in Mathematics, 74, Springer-Verlag, New York – Berlin, 1980. *MR* **82m**:10001

[3] GALLAGHER, P. X., A large sieve density estimate near $\sigma = 1$, *Invent. Math.* **11** (1970), 329–339. *MR* **43** #4775

[4] HARMAN, G., Trigonometric sums over primes, *Mathematika* **28** (1981), 249–254. *MR* **83j**:10045

[5] IRELAND, K. and ROSEN, M., *A classical introduction to modern number theory*, Graduate Texts in Mathematics, 84, Springer-Verlag, New York – Berlin, 1982. *MR* **83g**:12001

[6] LEUNG, M. C. and LIU M. C., On generalized quadratic equations in three prime variables, *Monatsh. Math.* **115** (1993), 133–167. *MR* **94g**:11085

[7] CHIU, S. F. and LIU, M. C., On exceptional sets for numbers representable by binary sums, *Rocky Mountain J. Math.* **26** (1996), 959–986. *MR* **98c**:11107

[8] LIU, M. C. and TSANG, K. M., Small prime solutions of some additive equations, *Monatsh. Math.* **111** (1991), 147–169. *MR* **92e**:11106

[9] MONTGOMERY, H. L. and VAUGHAN, R. C., The exceptional set in Goldbach's problem, *Acta Arith.* **27** (1975), 353–370. *MR* **51** #10263

[10] SCHMIDT, W. M., *Equations over finite fields. An elementary approach*, Lecture Notes in Mathematics, Vol. 536, Springer–Verlag, Berlin – New York, 1976. *MR* **55** #2744

[11] TITCHMARSH, E. C., *The theory of the Riemann zeta-function*, Second edition, Edited and with a preface by D. R. Heath-Brown, Oxford, Clarendon Press, 1986. *MR* **88c**:11049

[12] VAUGHAN, R. C., *The Hardy–Littlewood method*, Cambridge Tracts in Mathematics, 80, Cambridge University Press, Cambridge – New York, 1981. *MR* **84b**:10002

[13] VINOGRADOV, I. M., On the estimations of some simplest trigonometrical sums involving prime numbers, *Bull. Acad. Sci. URSS Sér. Math.* [*Izvestia Akad. Nauk SSSR*] **1939**, 371–396 and engl. summary 396–398 (in Russian). *Zbl* **24**, 293–294; *MR* **2**, 40

[14] ZACCAGNINI, A., On the exceptional set for the sum of a prime and a k-th power, *Mathematika* **39** (1992), 400–421. *MR* **94g**:11086

(Received June 8, 1996)

MATHEMATISCHES INSTITUT
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
ALBERTSTRASSE 23 B
D–79104 FREIBURG
GERMANY

Claus.Bauer@mn.oen.siemens.de

Current address:

IM ACKER 18
D–56332 OBERFELL
GERMANY

clausbauer@yahoo.com

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 331–338

ON CONVERGENCE IN PROBABILITY OF MARTINGALE-LIKE SEQUENCES

D. Q. LUU

Summary

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a complete probability space and $(\mathcal{A}_n)$ an increasing sequence of sub σ-fields of $\mathcal{A}$. A sequence (X_n) of Banach space-valued Bochner integrable functions is said to be a game fairer with time, if for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all q, $n \in \mathbf{N}$ with $p \leqq q \leqq n$ we have $\mathbf{P}(||X_q(n) - X_q|| > \varepsilon) < \varepsilon$, where $X_q(n)$ denotes the $\mathcal{A}_q$-conditional expectation of X_n. As a corollary of the main theorem we obtain that every game fairer with time (X_n), satisfying the condition: $\liminf_n \mathbf{E}(||X_n||) < \infty$ admits a unique decomposition: $X_n = M_n + P_n$, where (M_n) is a uniformly integrable martingale and (P_n) goes to zero in probability. In fact, we show that this result still holds for several classes of martingale-like sequences that considerably generalize the class of games fairer with time.

1. Notations and definitions

Throughout this note let $(\Omega, \mathcal{A}, \mathbf{P})$ be a complete probability space, $(\mathcal{A}_n)$ an increasing sequence of sub-σ-fields of $\mathcal{A}$ and $\mathbf{N}$ the set of all positive integers. Unless otherwise stated, we shall denote by V or U cofinal subsets of $\mathbf{N}$. Now given a separable Banach space F, let $L^1(F)$ stand for the Banach space of all (equivalence classes of) $\mathcal{A}$-measurable Bochner integrable functions $X : \Omega \to F$ with the norm: $\mathbf{E}(||X||) < \infty$. We shall consider only sequences (X_n) in $L^1(F)$ which are assumed to be adapted to $(\mathcal{A}_n)$, i.e. each X_n is $\mathcal{A}_n$-measurable. For more information on amarts, the reader is referred to the recent monograph of G. A. Edgar and L. Sucheston [7]. Here, we recall only the following

DEFINITION 1.1. A sequence (X_n) is said to be

(a) *a martingale*, if for all q, $n \in \mathbf{N}$ with $q \leqq n$ we have $X_q = X_q(n)$, where $X_q(n)$ denotes the $\mathcal{A}_q$-conditional expectation of X_n (cf. [14]);

(b) *a mil*, if for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all $n \in \mathbf{N}$ with $p \leqq n$ we have

$$\mathbf{P}\left(\sup_{p \leqq q \leqq n} ||X_q(n) - X_q|| > \varepsilon\right) < \varepsilon;$$

1991 *Mathematics Subject Classification.* Primary 60G48, 60B11.

Key words and phrases. Banach spaces, decomposition, convergence in probability, martingales, mils, sequential games fairer with time.

0081–6906/99/$ 5.00

(c) *a game which becomes fairer with time*, if for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all q, $n \in \mathbf{N}$ with $p \leqq q \leqq n$ we have

$$\mathbf{P}\left(||X_q(n) - X_q|| > \varepsilon\right) < \varepsilon.$$

Games fairer with time were first introduced by L. H. Blake [2] who proved that every real-valued game fairer with time which is a.s. bounded by an integrable function, converges in L^1. Later, this result was extended (independently) by A. G. Mucci [12] and S. Subramanian [15] to the uniformly integrable case. Recently, the amart theory has been extensively developed. Among many others, M. Talagrand [16] has introduced the class of mils as a generalization of martingales, uniform amarts [1], pramarts [11] and martingales in the limit [13]. The main structure Talagrand's theorem in [16] says that every F-valued mil (X_n) with $\liminf_n \mathbf{E}(||X_n||) < \infty$ admits a unique decomposition: $X_n = M_n + P_n$, $n \in \mathbf{N}$, where (M_n) is a uniformly integrable martingale and (P_n) goes to zero, a.s. Also it is known that every above-recalled class of martingale-like sequences is strictly contained in the next one. In the next section we shall apply the Talagrand's result to consider a decomposition and convergence of the following family of martingale-like sequences which considerably generalizes the class of all games fairer with time.

Definition 1.2. Let V be a cofinal subset of $\mathbf{N}$. We say that (X_n) is *a game which becomes fairer with the set V of times* (or briefly, *V-game*), if for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all $q \in \mathbf{N}$ and $v \in V$ with $p \leqq q \leqq v$ we have $\mathbf{P}\left(||X_q(v) - X_q|| > \varepsilon\right) < \varepsilon$.

In general, if (X_n) is a V-game for some cofinal subset V of $\mathbf{N}$ then (X_n) will be called *an $\mathbf{N}$-sequential game.*

Now, let N^c denote the set of all cofinal subsets of $\mathbf{N}$. Then N^c is a directed set filtering to the right with the order $(\leqq)$, given by: $V \leqq U$ iff $\mathrm{card}(U \setminus V)$ is finite. Clearly, by definition, in the space of all $\mathbf{N}$-sequential games, the classes of V-games, when V runs over N^c form an increasing family of classes of martingale-like sequences and the class of games fairer with time, i.e. the class of $\mathbf{N}$-games is the smallest one. Moreover, by Example 2.5 [10] the author has shown that the class of V-games coincides with that of U-games if and only if $\mathrm{card}(V \triangle U)$ is finite. Thus the class of $\mathbf{N}$-sequential games considerably generalizes that of games fairer with time. The main aim of the next section is to show the following

Theorem 1.3. *Let (X_n) be an E-valued V-game with*

$$\liminf_v \mathbf{E}(||X_v||) < \infty. \tag{1}$$

Then (X_n) admits a unique decomposition:

$$X_n = M_n + P_n, \quad n \in \mathbf{N}, \tag{2}$$

where (M_n) is a uniformly integrable martingale and (P_n) goes to zero in probability.

Consequently, if either the set $\{X_n(\omega)\}$ is relatively weakly compact a.s. or F has the Radon–Nikodym property (RNP) then (X_n) converges in probability to a Bochner integrable function.

It is worth noting that the theorem is independent of Theorem 2.2 [10], where instead of (1), we supposed that the subsequence (X_v) of (X_n) is *an L^1-amart*, i.e. for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all and v, $v' \in V$ with $p \leqq v \leqq v'$ we have

$$\mathbf{E}(||X_v(v') - X_v||) < \varepsilon, \quad \text{(cf. [8])}.$$

In this case we obtain a Riesz decomposition of the L^1-amart (X_v), hence of (X_n). So the proof of Theorem 2.2 [10] cannot be applied to prove the above main theorem. Finally, in the last section we shall discuss on some other generalizations of games for which the Riesz decomposition still holds.

2. Proof of the main Theorem 1.3

To show the theorem we need the following lemmas. The first is an extension of Lemma 2.2 of D. Q. Luu [9]. The second is a V-game version of a classical result of J. Neveu [14] for martingales and Theorem 6 of M. Talagrand [16] for mils which says that every L^1-bounded martingale (or mil) converges to zero, a.s. if it converges scalarly to zero, a.s. Let us begin with

LEMMA 2.1. *Every V-game contains a subsequence which is a mil.*

PROOF. Let (X_n) be a V-game. Then by definition there exists an increasing subsequence (v_n) of V such that for all q, $n \in \mathbf{N}$ with $q \leqq n$ we have

$$\mathbf{P}(||X_{v_q}(v_n) - X_{v_q}|| > 2^{-q}) < 2^{-q}.$$

Consequently, if $\varepsilon > 0$ is given then for all p, q, $n \in \mathbf{N}$ with $2^{-p+1} < \varepsilon$ and $p \leqq q \leqq n$ one get

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{p \leqq q \leqq n} ||X_{v_q}(v_n) - X_{v_q}|| > \varepsilon\Big) \\
&\leqq \sum_{q=p}^{n} \mathbf{P}(||X_{v_q}(v_n) - X_{v_q}|| > 2^{-q}) \\
&\leqq \sum_{q=p}^{\infty} \mathbf{P}(||X_{v_q}(v_n) - X_{v_q}|| > 2^{-q}) \\
&\leqq 2^{-p+1} < \varepsilon.
\end{aligned}$$

Then by definition, the subsequence (X_{v_n}) of (X_n) is a mil.

LEMMA 2.2. *Let (X_n) be a V-game satisfying* (1). *Suppose that (X_n) contains a subsequence, say $(X_u,\ u\in U)$, which converges to zero in probability. Then (X_n) converges itself to zero in probability.*

PROOF. Let (X_n), V and U be as supposed in the lemma. Assume on the contrary that (X_n) does not go to zero in probability. It means that there exists $a>0$ such that $\limsup_n \mathbf{P}(||X_n||>5a/4)>a$. We claim that:

For every $0<\varepsilon<a/4$ and $v_1\in V$ there exists $v_2\in V$ with $v_1\leqq v_2$ such that for each $A\in\mathcal{A}_{v_1}$ with $\mathbf{P}(A)<a/4$ and each $v\in V$ with $v_2\leqq v$ there exists a set $B\in\mathcal{A}_{v_2}$ with $B\cap A=\emptyset$ and $\mathbf{P}(B)<\varepsilon$ such that

$$\int_B ||X_v||dP\geqq a^2/4. \tag{3}$$

To prove the claim, let $0<\varepsilon<a/4$ and $v_1\in V$ be given. Then by definition, there exists $p\in\mathbf{N}$ with $v_1\leqq p$ so large that for all $v\in V$ with $p\leqq v$ we have

$$\mathbf{P}(||X_p(v)-X_p||>a/4)<\varepsilon/2 \tag{4}$$

and if we set $C=\{||X_p||>5a/4\}$ then $\mathbf{P}(C)>a$.

Consequently, there exists a finite sequence $\{x_j^*,\ j\leqq m\}$ of the unit ball of F^* such that if for every $j\leqq m$ we take

$$C_j^1=C\cap\Big\{\{(x_j^*,X_p)>5a/4\}\setminus\bigcup_{s<j}\{(x_j^*,X_s)>5a/4\}\Big\}$$

and $C^1=\bigcup_{j\leqq m}C_j^1$ then $\mathbf{P}(C^1)>7a/8$.

On the other hand, since $(X_u,\ u\in U)$ converges to zero in probability there is $u_1\in U$ with $p\leqq u_1$ such that if we put

$$D=\{||X_{u_1}||>a/2\}\quad\text{then}\quad\mathbf{P}(D)<\varepsilon/2.$$

Now let

$$v_2=\min\{v\in V,\ u_1\leqq v\},$$

$A\in\mathcal{A}_{v_1}$ with $\mathbf{P}(A)<a/4$ and $v\in V$ with $v_2\leqq v$. Then by (4), $\mathbf{P}(G)<\varepsilon/2$, where

$$G=\{||X_p(v)-X_p||>a/4\}.$$

Hence,

$$\mathbf{P}(C^2)>7a/8-3a/8=a/2,$$

where $C_j^2=C_j^1\setminus(G\cup A)$ and $C^2=\bigcup_{j\leqq m}C_j^2$. Similarly, let

$$H=\{||X_{u_1}(v)-X_{u_1}||>a/4\}.$$

Then by (4), $\mathbf{P}(H) < \varepsilon/2$. Further, let $D^1 = D \cup H$ then $\mathbf{P}(D^1) < \varepsilon$. Finally, for $j \leqq m$, let $B_j = C_j^2 \cap D^1$, $B = \bigcup_{j \leqq m} B_j$. Then $B \in \mathcal{A}_{u_1}$, $B \cap A = \emptyset$ and $\mathbf{P}(B) \leqq \mathbf{P}(D^2) < \varepsilon$. We shall show that constructed in such a way, the set B satisfies also (3), proving the claim. To see this, let $j \leqq m$ be any but fixed. Then we have

$$\int_{C_j^2} (x_j^*, X_v) d\mathbf{P} = \int_{C_j^2} (x_j^*, X_p(v) d\mathbf{P}) \geqq a\mathbf{P}(C_j^2)$$

since $C_j^2 \in \mathcal{A}_p$ and $(x_j^*, X_p(v)) \geqq (x_j^*, X_p) - a/4 \geqq 5a/4 - a/4 = a$ on C_j^2. Similarly, define $D_j^2 = (C_j^2 \setminus D^1)$ then $D_j^2 \in \mathcal{A}_{u_1}$ and on D_j^2 we have

$$(x_j^*, X_{u_1}(v)) \leqq (x_j^*, X_{u_1}) + a/4 = a/4 + a/4 = a/2,$$

hence, $\int_{D_j^2} (x_j^*, X_v) d\mathbf{P} = \int_{D_j^2} (x_j^*, X_{u_1}(v)) d\mathbf{P} \leqq a/2\mathbf{P}(C_j^2)$. But $B_j \cap D_j^2 = \emptyset$ and $C_j^2 = B_j \cup D_j^2$, $j \leqq m$ we get

$$\begin{aligned} \int_{B_j} ||X_v|| d\mathbf{P} &\geqq \int_{B_j} (x_j^*, X_v) d\mathbf{P} \\ &= \int_{C_j^2} (x_j^*, X_v) d\mathbf{P} - \int_{D_j^2} (x_j^*, X_v) d\mathbf{P} \\ &\geqq a\mathbf{P}(C_j^2) - a/2\mathbf{P}(C_j^2) = a/2\mathbf{P}(C_j^2). \end{aligned}$$

Thus by summation over $j \leqq m$ we have

$$\int_B ||X_v|| d\mathbf{P} \geqq a/2\mathbf{P}(C^2) \geqq a^2/4.$$

It proves (3) and the claim.

Now, returning to the proof of the lemma, we can construct by induction an increasing sequence (v_p) of V with the following property: whenever $A \in \mathcal{A}_{v_p}$ with $\mathbf{P}(A) < a/4$ and $v \in V$ with $v_{p+1} \leqq v$ there is a set $B \in \mathcal{A}_{v_{p+1}}$ with $B \cap A = \emptyset$, $\mathbf{P}(B) < a2^{-(p+1)}$ and $\int_B ||X_v|| d\mathbf{P} \geqq a^2/4$. Thus given $p \in \mathbf{N}$ and $v \geqq v_p$, one can construct by finite induction for $j \leqq p$ disjoint sets $B_j \in \mathcal{A}_{v_{p_j}}$ with $B_1 = \emptyset$, $\mathbf{P}(B_j) < a2^{-(j+1)}$ and $\int_{B_j} ||X_v|| d\mathbf{P} \geqq a^2/4$. Hence,

$$\int_B ||X_v|| d\mathbf{P} \geqq (p-1)a^2/4, \quad \text{where} \quad B = \bigcup_{j \leqq p} B_j.$$

This means that $\lim_v \mathbf{E}(||X_v||) = \infty$. It contradicts (1) and proves the lemma.

Finally, having these two lemmas in hand we can proceed to the proof of the main theorem. Indeed, let (X_n) and V be given as supposed in the theorem. It is clear that the subsequence, X_v, $(v \in V)$ is itself a game fairer with time w.r.t. $(\mathcal{A}_v)$. Consequently, by passing to an L^1-bounded subsequence, Lemma 2.1 implies that there exists a subsequence U of V such that the subsequence (X_u) is an L^1-bounded mil w.r.t. $(\mathcal{A}_u)$. Then by Theorem 8 [16], it follows that (X_n) admits a unique decomposition:

$$X_n = M_n + P_n, \quad n \in \mathbf{N}, \tag{5}$$

where (M_n) is a uniformly integrable martingale and the subsequence (P_u) is a mil w.r.t. $(\mathcal{A}_u,\ u \in U)$ that goes to zero, a.s. Clearly, (P_n) is still a V-game with $\liminf_n \mathbf{E}(||P_v||) < \infty$. This with the second lemma shows that (P_n) converges itself to zero in probability which completes the proof of Decomposition (2) and of the main part of the theorem. Here, (M_n) is uniformly integrable because according to the Talagrand's proof of Theorem 8 [16] we have $||M_n|| \leqq E_n(h)$, a.s. $n \in \mathbf{N}$, where the function $h = \liminf_u ||X_u||$ is integrable and $E_n(h)$ denotes the $\mathcal{A}_n$-conditional expectation of h. And only this uniform integrability of (M_n) guarantees the uniqueness of Decomposition (5), hence of Decomposition (2) required in the theorem. In addition, if the set $\{X_n(\omega)\}$ is relatively weakly compact, a.s. then, by Decomposition (2), so is the set $\{M_n(\omega)\}$. Consequently, either in the case or when F has the (RNP), (M_n) must converge a.s. and in L^1 to an F-valued Bochner integrable function X, according to a Chatterji's result in [5] or in [6], resp. Thus again by Decomposition (2) in the theorem, (X_n) converges itself in probability to X which completes the proof.

By the remark after Definition 1.2 and by the main theorem in the section we obtain the following corollary which contains all the results of L. A. Blake [2, 3], A. G. Mucci [12], S. Subramanian [15] and D. Q. Luu [9] and is new even in the real-valued case.

Corollary 2.3. *Let (X_n) be a game fairer with time or an $\mathbf{N}$-sequential game, resp. Suppose that (X_n) satisfies the following condition:* $\liminf_n \mathbf{E}(||X_n||) < \infty$ *or* $\limsup_n \mathbf{E}(||X_n||) < \infty$, *resp. Then (X_n) admits a unique decomposition: $X_n = M_n + P_n$, $n \in \mathbf{N}$, where (M_n) is a uniformly integrable martingale and (P_n) is a game fairer with time or an $\mathbf{N}$-sequential game, resp., that goes to zero in probability.*

3. Directed case

In this section, let $(D, \leqq)$ be a directed set filtering to the right with the nonempty set D^c of all increasing cofinal subsequences (t_n) of D and

$(\mathcal{A}_t,\ t \in D)$ an increasing family of sub-σ-fields of $\mathcal{A}$. Then for directed processes (X_t) in $L^1(E)$, adapted to $(\mathcal{A}_t)$ one can introduce the following

DEFINITION 3.1. A process (X_t) is called *a game which becomes fairer with the sequence (t_n) of times*, briefly $\{t_n\}$-game, if for every $\varepsilon > 0$ there exists $p \in \mathbf{N}$ such that for all $t \in D$ and $n \in \mathbf{N}$ with $t_p \leqq t \leqq t_n$ one has $\mathbf{P}(||X_t(t_n) - X_t|| > \varepsilon) < \varepsilon$. In general, if (X_t) is a $\{t_n\}$-game for some $(t_n) \in D^c$ then (X_t) is called *a D-sequential game.*

It is not hard to check that Theorem 1.3 can be extended to all $\{t_n\}$-games with $\liminf_n \mathbf{E}(||X_{t_n}||) < \infty$. Consequently, we get the following directed versions of the results of the previous section.

THEOREM 3.2. *Let (X_t) be a $\{t_n\}$-game or a D-sequential game, resp. Suppose that*

$$\liminf_n \mathbf{E}(||X_{t_n}||) < \infty \quad \textit{or} \quad \limsup_{t \in D} \mathbf{E}(||X_t||) < \infty, \quad \textit{resp.}$$

Then (X_t) admits a unique decomposition: $X_t = M_t + P_t$, $t \in D$, where (M_t) is a uniformly integrable martingale and (P_t) is a $\{t_n\}$-game or a D-sequential game, resp., that stochastically converges to zero.

ACKNOWLEDGEMENT. The author would like to express many thanks to Professors Cz. Ryll-Nardzewski and K. Musial for very nice and useful discussions during his stay at the Technical University of Wrocław, Poland, 1996–1997. Especially, the author is very much appreciated to the referee for his kindness and useful suggestions.

REFERENCES

[1] BELLOW, A., Uniform amarts: A class of asymptotic martingales for which strong almost sure convergence obtains, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **41** (1977/1978), 177–191. *MR* **57** #10806

[2] BLAKE, L. H., A generalization of martingales and two consequent convergence theorems, *Pacific J. Math.* **35** (1970), 279–283. *MR* **43** #1259

[3] BLAKE, L. H., A note concerning the L-convergence of a class of games which become fairer with time, *Glasgow Math. J.* **13** (1972), 39–41. *MR* **46** #4595

[4] CHACON, R. V. and SUCHESTON, L., On convergence of vector-valued asymptotic martingales, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **33** (1975), 55–59. *MR* **52** #15658

[5] CHATTERJI, S. D., Martingale convergence and the Radon–Nikodym theorem in Banach spaces, *Math. Scand.* **22** (1968), 21–41. *MR* **39** #7645

[6] CHATTERJI, S. D., Vector-valued martingales and their applications, *Probability in Banach spaces* (Proc. First Internat. Conf., Oberwolfach, 1975), Lecture Notes in Math., Vol. 526, Springer, Berlin, 1976, 33–51. *MR* **58** #24529

[7] EDGAR, G. A. and SUCHESTON, L., *Stopping times and directed processes*, Encyclopedia of Mathematics and its Applications, **47**, Cambridge Univ. Press, Cambridge, 1992. *MR* **94a**:60064

[8] LUU, D. Q., Application of set-valued Radon–Nikodym theorems to convergence of multi-valued L^1-amarts, *Math. Scand.* **54** (1984), 101–113. *MR* **86b**:60084

[9] Luu, D. Q., Decompositions and limits for martingale-like sequences in Banach spaces, *Acta Math. Vietnam* **13** (1988), 73–78. *MR* **91c**:60060
[10] Luu, D. Q., Convergence and lattice properties of a class of martingale-like sequences, *Acta. Math. Hungar.* **59** (1992), 273–281. *MR* **93j**:60003
[11] Millet, A. and Sucheston, L., Convergence of classes of amarts indexed by directed sets, *Canad. J. Math.* **32** (1980), 86–125. *MR* **81g**:60051
[12] Mucci, A., Limits for martingale-like sequence, *Pacific J. Math.* **48** (1973), 197–202. *MR* **50** #11425b
[13] Mucci, A., Another martingale convergence theorem, *Pacific J. Math.* **64** (1976), 539–541. *MR* **54** #8852
[14] Neveu, J., *Martingales à temps discret*, Masson, Paris, 1972. *MR* **53** #6728
[15] Subramanian, S., On a generalization of martingales due to Blake, *Pacific. J. Math.* **48** (1973), 275–278. *MR* **50** #11425a
[16] Talagrand, M., Some structure results for martingales in the limit and pramarts, *Ann. Probab.* **13** (1985), 1192–1203. *MR* **86m**:60021

(Received December 20, 1996)

HANOI INSTITUTE OF MATHEMATICS
P.O. BOX 631 BO-HO
HANOI
VIETNAM

Actual address:

INSTYTUT MATEMATYKI
STEFAN BANACH CENTER
UL. MOKOTOWSKA 25, SKR. POCZT. 137
PL–00-950 WARSZAWA
POLAND

minh16@easymail.it.com.pl

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 339–346

GENERALIZED COTANGENCY SETS IN PROJECTIVE SPACES

Gy. KISS

Abstract

The notion of *cotangency set* in the projective plane over any field was introduced by Bruen and Fisher [1]. They proved that a cotangency set never contains a quadrangle and deduced several theorems from this fact. In this paper a generalized definition of cotangency sets in the n-dimensional projective space is given. We prove some theorems about quadrics and Hermitian varieties which are consequences of the properties of cotangency sets.

1. Introduction

Let $\mathcal{P} = PG(n, F)$ be the n-dimensional projective space over the commutative field F. In the case $n = 2$ Bruen and Fisher [1] defined the cotangency set in the following way:

DEFINITION 1.1. Let $\mathcal{S}$ be a set of points in $\mathcal{P}$ and assume that there is an injective mapping f from $\mathcal{S}$ into the set of lines of $\mathcal{P}$ satisfying the two properties:

(a) if P is in $\mathcal{S}$ then $f(P)$ does not contain P;

(b) if P_1 and P_2 are distinct points of $\mathcal{S}$ then the points P_1, P_2 and the intersection of the lines $f(P_1)$ and $f(P_2)$ lie on a line.

They showed that a cotangency set never contains a quadrangle. The proof of this theorem is quite simple, but a number of consequences involving Hermitian arcs and conics follow quickly from the theorem by way of elementary arguments. In this paper we generalize the notion of cotangency set in higher dimensional spaces. The natural generalization is given in Definition 2.1. The main goal of the present paper is to prove the following generalization of the result of Bruen and Fisher: a proper n-cotangency set in $\mathcal{P}$

1991 *Mathematics Subject Classification.* Primary 51E20; Secondary 51A05.
Key words and phrases. Cotangency set, quadric, Hermitian variety.

This research was supported by the "Foundation for the Hungarian Sciences" of the Hungarian Credit Bank and by the Hungarian National Foundation for Scientific Research Grants No. F016302 and T017314.

never contains $n+2$ points in general position. We can also generalize the corresponding consequences involving non-singular quadrics and Hermitian varieties. Finally, in Section 3, we give some examples of 2-cotangency sets in $\mathcal{P}$ which contain $n+2$ points in general position. This means that the generalization of the theorem of Bruen and Fisher is not valid for 2-cotangency sets. But surprisingly, the generalized corollaries are true. These theorems were proved in [2].

2. n-dimensional cotangency sets

DEFINITION 2.1. Let $\mathcal{S}$ be a set of points in $\mathcal{P}$ and k be a natural number, $2 \leqq k \leqq n$. Assume that there is an injective mapping f from $\mathcal{S}$ into the set of hyperplanes of $\mathcal{P}$ satisfying the following two properties:

(a) if P is in $\mathcal{S}$ then $f(P)$ does not contain P;

(b) if $P_1, P_2, \ldots, P_k$ are distinct points of $\mathcal{S}$ then the subspace generated by the points $P_1, P_2, \ldots, P_k$ and the intersection of the k hyperplanes $f(P_1), f(P_2), \ldots, f(P_k)$ lie in a hyperplane.

Then $\mathcal{S}$ is called a k-cotangency set.

If the dimension of $\mathcal{P}$ is two then the only possible value for k is two, and our definition is the same as the original one. It is possible that a k-cotangency set contains m points $P_{i_1}, P_{i_2}, \ldots, P_{i_m}$ for some $m < k$ such that the subspace generated by the points $P_{i_1}, P_{i_2}, \ldots, P_{i_m}$ and the intersection of the m hyperplanes $f(P_{i_1}), f(P_{i_2}), \ldots, f(P_{i_m})$ lie in a hyperplane. If it happens, we call $\mathcal{S}$ an *m-degenerate* cotangency set. We would like to distinguish these sets from the "proper" cotangency sets.

DEFINITION 2.2. A k-cotangency set $\mathcal{S}$ is called proper k-cotangency set if there is no $m < k$ such that $\mathcal{S}$ is m-degenerate.

First we prove a simple lemma about proper cotangency sets.

LEMMA 2.3. *Let $\mathcal{S} = \{P_1, P_2, \ldots, P_l\}$ be a proper k-cotangency set in $\mathcal{P}$, $k > 2$. Let $\mathcal{H} = f(P_1)$. For $i > 1$ let $f'(P_i)$ be the intersection of $\mathcal{H}$ and $f(P_i)$, and let P_i' be the intersection of the line P_1P_i and $\mathcal{H}$. Then the points $P_2', P_3', \ldots, P_l'$ and the subspaces $f'(P_2), f'(P_3), \ldots, f'(P_l)$ form a proper $(k-1)$-cotangency set in $\mathcal{H}$.*

PROOF. Since f is an injection $f'(P_i)$ is an $(n-2)$-dimensional subspace in $\mathcal{P}$, thus $f'(P_i)$ is a hyperplane of $\mathcal{H}$. P_i' is well-defined because $P_1 \notin \mathcal{H}$. Let us denote the set of points $\{P_2', P_3', \ldots, P_l'\}$ by $\mathcal{S}_1'$.

First we prove that $P_i' \notin f'(P_i)$. If $P_i' \in f'(P_i)$ then the line P_1P_i meets $f(P_1) \cap f(P_i)$. Thus the two points P_1, P_i and the two hyperplanes $f(P_1)$, $f(P_i)$ in $\mathcal{S}$ form a 2-degenerate configuration which is a contradiction because $\mathcal{S}$ is a proper k-cotangency set.

Consider any $k-1$ points $P'_{i_1}, P'_{i_2}, \ldots, P'_{i_{k-1}}$ out of $\mathcal{S}'_1$, and the corresponding $k-1$ subspaces out of the set $\{f'(P_2), f'(P_3), \ldots, f'(P_l)\}$. If the subspace generated by the points $P'_{i_1}, P'_{i_2}, \ldots, P'_{i_{k-1}}$ and the intersection of the $k-1$ hyperplanes $f(P'_{i_1}), f(P'_{i_2}), \ldots, f(P'_{i_{k-1}})$ has dimension r in $\mathcal{P}$ then the subspace generated by the points $P_1, P_{i_1}, P_{i_2}, \ldots, P_{i_{k-1}}$ and the intersection of the k hyperplanes $f(P_1), f(P_{i_1}), f(P_{i_2}), \ldots, f(P_{i_{k-1}})$ has dimension $r+1$ in $\mathcal{P}$ because $P_1 \notin \mathcal{H}$ and $f(P'_{i_j}) \neq f(P_1)$. But $\mathcal{S}$ is a k-cotangency set hence $r+1 \leqq k-1$. Thus $r \leqq k-2$ which means that the points and hyperplanes of $\mathcal{H}$ satisfy property 2 of the definition of a cotangency set.

Finally we have to prove that $\mathcal{S}'_1$ is not m-degenerate. If the subspaces generated by any $m < k-1$ points $P'_{i_1}, P'_{i_2}, \ldots, P'_{i_m}$ and the intersection of the corresponding $(n-2)$-dimensional subspaces $f(P'_{i_1}), f(P'_{i_2}), \ldots, f(P'_{i_m})$ of $\mathcal{P}$ would be in a hyperplane of $\mathcal{H}$ – thus in an $(n-2)$-dimensional subspace of $\mathcal{P}$ – then the points $P_1, P_{i_1}, P_{i_2}, \ldots, P_{i_m}$ would form an $(m+1)$-degenerate configuration in $\mathcal{S}$. But this is a contradiction because $m+1 \leqq k$, and $\mathcal{S}$ is a proper k-cotangency set.

The lemma is proved.

Our main result is the following

THEOREM 2.4. *If $\mathcal{S}$ is a proper n-cotangency set in an n-dimensional space then $\mathcal{S}$ does not contain $n+2$ points in general position.*

PROOF. We prove by induction on n. If $n=2$, then this is the result of Bruen and Fisher [1]. Let $n>2$. Suppose that the theorem is true for $n-1$. Assume that $\mathcal{S}$ contains the points $P_1, P_2, \ldots, P_{n+2}$ which are in general position. Let $\mathcal{H} = f(P_1)$ and define $\mathcal{S}'_1$ in the same way as we have done in the proof of the previous lemma. $\mathcal{S}'_1$ is a proper $(n-1)$-cotangency set in $\mathcal{H}$ which is an $(n-1)$-dimensional projective space. $\mathcal{S}'_1$ contains the points $P'_2, P'_3, \ldots, P'_{n+2}$ which are in general position. This is a contradiction because the theorem is true for $(n-1)$-dimensional spaces. Thus $\mathcal{S}$ does not contain $n+2$ points in general position.

We now look at some applications.

COROLLARY 2.5. *Suppose that the characteristic of F is not equal to two. Let $\mathcal{S}$ be a set of points in $\mathcal{P}$ that is disjoint from a non-singular quadric $\mathcal{Q}$. Assume that any hyperplane generated by n linearly independent points of $\mathcal{S}$ has exactly one point in common with $\mathcal{Q}$, but no subspace generated by less than n points of $\mathcal{S}$ meets $\mathcal{Q}$. Then $\mathcal{S}$ consists of the vertices of a simplex circumscribed about $\mathcal{Q}$, or its points lie on a hyperplane which has one point in common with $\mathcal{Q}$.*

PROOF. Let $\mathcal{S} = \{P_1, P_2, \ldots, P_k\}$. For $P_i \in \mathcal{S}$ let $f(P_i)$ be its polar hyperplane with respect to the polarity defined by $\mathcal{Q}$. Since $P_i \notin \mathcal{Q}$, axiom (1) for a cotangency set is satisfied. Let $P_{i_1}, P_{i_2}, \ldots, P_{i_n}$ be n linearly independent points of $\mathcal{S}$ and let $Q_{i_1, i_2, \ldots, i_n}$ be the unique point in common of the

hyperplane generated by the points $P_{i_1}, P_{i_2}, \ldots, P_{i_n}$ and $\mathcal{H}$. From the properties of the polarity, $f(P_{i_j})$ passes through $Q_{i_1,i_2,\ldots,i_n}$ for each i_j. $Q_{i_1,i_2,\ldots,i_n}$ is the unique point in common of the n hyperplanes $f(P_{i_1}), f(P_{i_2}), \ldots, f(P_{i_n})$ because the points $P_{i_1}, P_{i_2}, \ldots, P_{i_n}$ are linearly independent. Thus axiom (2) for a cotangency set is satisfied. $\mathcal{S}$ is not m-degenerate for $m < n$ because no subspace generated by m points of $\mathcal{S}$ meets $\mathcal{Q}$. Hence $\mathcal{S}$ is a proper n-cotangency set. The statement follows from Theorem 2.4.

COROLLARY 2.6. *Let $\mathcal{S}$ be a set of points in $PG(n,p^{2r})$, p odd prime, that is disjoint from a Hermitian variety $\mathcal{H}$. Assume that any hyperplane generated by n points of $\mathcal{S}$ has exactly one point in common with $\mathcal{H}$, but no subspace generated by less than n points of $\mathcal{S}$ meets $\mathcal{H}$. Then $\mathcal{S}$ cannot contain $n+2$ points in general position.*

PROOF. It follows from Theorem 2.4 just as Corollary 2.5 did. We need to replace $\mathcal{Q}$ by $\mathcal{H}$ and interpret f as the polarity induced by $\mathcal{H}$.

3. Final remarks

The main theorem is not true if we replace the proper n-cotangency set by 2-cotangency set.

THEOREM 3.1. *If $n > 2$ and the commutative field F has at least 7 elements, then there is a 2-cotangency set $\mathcal{S}$ in $\mathcal{P} = PG(n,F)$ which contains $n+2$ points in general position.*

PROOF. We give an example. Let $\mathcal{S}$ be the set of points $\{P_0, P_1, ..., P_n, U\}$, where the points have the following coordinates:

$$P_0(0,0,\ldots,0,1);$$
$$P_i(0,0,\ldots,\underbrace{1}_{i},0,\ldots,0,1) \qquad \text{for } i = 1,2,\ldots,n;$$
$$U(1,1,\ldots,1,1).$$

Let the equations of the corresponding hyperplanes be

$$f(P_0): \quad x_{n+1} = 0;$$
$$f(P_i): \quad \sum_{j=1}^{n} a_{ij}x_j + x_{n+1} = 0;$$
$$f(U): \quad \sum_{j=1}^{n} A_j x_j + x_{n+1} = 0.$$

First we prove that $P_0, P_1, ..., P_n$ is a 2-cotangency set if $a_{ji} = \dfrac{1}{a_{ij}+1} - 1$ for $i > j$, and $a_{ii} = 0$.

P_i is not incident with $f(P_i)$ because $a_{ii}=0$ and $1\neq 0$. If both i and j are greater than 0 then the point

$$Q_{ij}(0,\dots,0,\underbrace{1}_{i},0,\dots,0,\underbrace{-1-a_{ji}}_{j},0,\dots,0,-a_{ji})$$

is incident with the line P_iP_j, and both of the hyperplanes $f(P_i)$ and $f(P_j)$ because $Q_{ij}=P_i-(1+a_{ji})P_j$ and $a_{ij}a_{ji}+a_{ij}+a_{ji}=0$. Hence the intersection of the two hyperplanes and the line joining the two points lie in a hyperplane. Consider the line P_0P_i. This meets $f(P_0)$ at the point $P_i-P_0=Q_i(0,\dots,0,\underbrace{1}_{i},0,\dots,0)$ and this point is incident with the hyperplane $f(P_i)$ because $a_{ii}=0$.

Now we determine A_i such that the set $U\cup\{P_0,P_1,\dots,P_n\}$ becomes a 2-cotangency set. The line P_0U meets $f(P_0)$ at the point $U-P_0$ which has coordinates $(1,1,\dots,1,0)$. This point is incident with $f(U)$ if and only if

$$\sum_{j=1}^{n}A_j=0. \tag{1}$$

The line P_iU for $i>0$ meets $f(P_i)$ at the point

$$R_i\Big(1,1,\dots,1,\underbrace{-\sum_{j=1}^{n}a_{ij}}_{i},1,\dots,1,-\sum_{j=1}^{n}a_{ij}\Big).$$

This point is incident with $f(U)$ if and only if

$$A_1+\cdots+A_{i-1}-A_i\Big(\sum_{j=1}^{n}a_{ij}\Big)+A_{i+1}+\cdots+A_n-\sum_{j=1}^{n}a_{ij}=0. \tag{2.i}$$

Let

$$A_i=\frac{1}{1+\sum_{j=1}^{n}a_{ij}}-1.$$

Then equation (2.i) becomes $\sum_{j=1}^{n}A_j=0$. We show that if $n>2$ then there exist a_{ij} such that $\sum_{i=1}^{n}A_i=0$. (If $n=2$ then the equation $A_1+A_2=0$ becomes $a_{1,2}^2/1+a_{1,2}=0$. Thus in this case there is no solution.)

First we define three matrices. Let

$$B_3=(b_{ij})=\begin{pmatrix} 0 & b & -\dfrac{4b(b+1)}{4b+1}\\ -\dfrac{b}{b+1} & 0 & \dfrac{b}{b+1}\\ \dfrac{4b(b+1)}{1-4b^2} & -\dfrac{b}{2b+1} & 0\end{pmatrix},$$

$$C_4 = (c_{ij}) = \begin{pmatrix} 0 & c & 0 & 0 \\ -\frac{c}{c+1} & 0 & \frac{c}{c+1} & 0 \\ 0 & -\frac{c}{2c+1} & 0 & \frac{c(3c+2)}{(c+1)(2c+1)} \\ 0 & 0 & -\frac{c(3c+2)}{5c^2+5c+2} & 0 \end{pmatrix},$$

and

$$D_r = (d_{ij})$$

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

The field F has at least seven elements, hence we can choose b, c and d such that

$$b(b+1)(1-4b^2)(4b+1) \neq 0,$$

$$c(2c+1)(3c+2)(5c^2+5c+2) \neq 0$$

and

$$d(d+1)(2d+1)(d^2+d-1)(d^2+2d+2) \neq 0.$$

Let us define

$$B_i = \frac{1}{1+\sum_{j=1}^{n} b_{ij}} - 1, \qquad C_i = \frac{1}{1+\sum_{j=1}^{n} c_{ij}} - 1$$

and

$$D_i = \frac{1}{1+\sum_{j=1}^{n} d_{ij}} - 1,$$

then elementary calculations give that

$$\sum_{i=1}^{n} B_i = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} D_i = 0.$$

Finally if $n = 3k + j$, where $j = 0, 1$ or 2, then let A be the $n \times n$ matrix

$$A = \begin{pmatrix} B_3 & & & & & \\ & B_3 & & & 0 & \\ & & \ddots & & & \\ & 0 & & B_3 & & \\ & & & & B_3 \text{ or } C_4 \text{ or } D_5 & \end{pmatrix}.$$

(The main diagonal of A contains $k-1$ times the submatrix B_3 and once the submatrix B_3, C_4 or D_5 according as $j = 0$, 1 or 2. All other entries of A are equal to 0.)

Define a_{ij} as the corresponding entry of the matrix A. Then $a_{ii} = 0$, because $b_{ii} = c_{ii} = d_{ii} = 0$; $a_{ij}a_{ji} + a_{ij} + a_{ji} = 0$, because $b_{ij}b_{ji} + b_{ij} + b_{ji} = 0$, $c_{ij}c_{ji} + c_{ij} + c_{ji} = 0$ and $d_{ij}d_{ji} + d_{ij} + d_{ji} = 0$. Finally

$$\sum_{i=1}^{n} A_i = 0$$

also holds, because

$$\sum_{i=1}^{n} A_i = (k-1) \sum_{i=1}^{3} B_i + \sum_{i=1}^{3+j} E_i = 0,$$

where $E_i = B_i$, C_i or D_i according as $j = 0$, 1 or 2.

Thus equation (1) has at least one nontrivial solution $\{A_1, A_2, \ldots, A_n\}$. Hence the points $U, P_1, P_2, \ldots, P_n$ form a 2-cotangency set.

Thus Theorem 2.2 is not true if we substitute the n-cotangency set by 2-cotangency set. But Corollaries 2.5 and 2.6 are valid in the following form:

THEOREM 3.2. *Suppose that the characteristic of F is not equal to two. Let $\mathcal{Q}$ be a non-singular quadric in $\mathcal{P}$ and let $\mathcal{S}$ be a set of points disjoint from $\mathcal{Q}$ such that the line joining any two points of $\mathcal{S}$ is a tangent to $\mathcal{Q}$. If the characteristic of K does not divide n then the cardinality of $\mathcal{S}$ is at most $n+1$, and hence $\mathcal{S}$ does not contain $n+2$ points in general position; while if the characteristic of K divides n, then the cardinality of $\mathcal{S}$ is at most $n+2$.*

THEOREM 3.3. *Let $\mathcal{H}$ be a Hermitian variety in $PG(3, p^{2r})$, p odd prime. Let $\mathcal{S}$ be a set of points disjoint from $\mathcal{H}$ such that the line joining any two points of $\mathcal{S}$ is a tangent to $\mathcal{H}$. If any four points of $\mathcal{S}$ form a tetrahedron then the cardinality of $\mathcal{S}$ is not greater than 10.*

These theorems were proved by Hirschfeld and Kiss [2].

We conclude with an open problem. Let $c(k)$ be the smallest number for which there is a proper $c(k)$-cotangency set in the k-dimensional projective space over F containing $k+2$ points in general position. Theorem 2.4 states that $2 \leqq c(k) < k$ for all k. It follows from Lemma 2.3 that if $c(k) = m$ then $c(k-1) \leqq m-1$. Determine $c(k)$ in general.

REFERENCES

[1] Bruen, A. A. and Fisher, J. C., An observation on certain point-line configurations in classical planes, *Discrete Math.* **106/107** (1992), 93–96. *MR* **93k**:51013

[2] Hirschfeld, J. W. P. and Kiss, Gy., Tangent sets in finite spaces, *Discrete Math.* **155** (1996), 107–119. *MR* **97d**:51010

[3] Hirschfeld, J. W. P. and Thas, J. A., *General Galois geometries*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1991. *MR* **96m**:51007

(Received January 23, 1997)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

kissgy@cs.elte.hu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 347–351

A NEW EXTENSION OF LUBELL'S INEQUALITY TO THE LATTICE OF DIVISORS

F. CHUDAK and J. GRIGGS

Abstract

P. L. Erdős and G. O. H. Katona gave an inequality involving binomial coefficients summed over an antichain in the product of two chains. Here we present the common generalization of this inequality and Lubell's famous inequality for the Boolean lattice to an arbitrary product of chains (lattice of divisors). We also describe the connection between this inequality and the LYM property.

1. Introduction

Let X be an n-set provided with a partition in M subsets X_i, called color classes, for $1 \leqq i \leqq M$. Let $n_i = |X_i|$ for all i. Associated with this coloring, we consider the poset $R(n_1, \ldots, n_m) = \{0 < \cdots < n_1\} \times \cdots \times \{0 < \cdots < n_M\}$, which consists of the product of M chains with ranks n_i. This poset is isomorphic to the lattice of divisors of $N = p_1^{n_1} \cdots p_M^{n_M}$, where the p_i's are distinct primes.

P. L. Erdős and G. O. H. Katona [3] discovered the following inequality for the product of just two chains in connection with their study of more-part Sperner families of subsets: For every antichain $I \subseteq R(n_1, n_2)$,

$$\sum_{(i_1,i_2)\in I} \frac{\binom{n_1}{i_1}\binom{n_2}{i_2}}{\binom{n_1+n_2}{i_1+i_2}} \leqq 1. \tag{1}$$

Their arguments were somehow lengthy, and a proof of a generalization for M colors was not apparent. We present such a generalization here along with some related observations.

1991 *Mathematics Subject Classification.* Primary 05A05; Secondary 06A07.
Key words and phrases. Poset, antichain, binomial inequalities.

The second author's research has been supported in part by NSA/MSP Grant MDA904-92H3053.

0081-6906/99/$ 5.00

THEOREM 1.1. *If $I \subseteq R(n_1, \ldots, n_m)$ is an antichain, then*

$$\sum_{(i_1,\ldots,i_M)\in I} \frac{\binom{n_1}{i_1}\cdots\binom{n_M}{i_M}}{\binom{n_1+\cdots+n_M}{i_1+\cdots+i_M}} \leq 1.$$

Notice that this extends Lubell's familiar inequality [7] for the Boolean lattice B_M of all subsets of an M-set, which is the case that all $n_i = 1$. In the next section we present two different proofs, both simpler than the original one in [3] for $M = 2$. The first is by counting chains, an argument that just extends Lubell's proof of Sperner's theorem [7]. We recently discovered the same proof, for $M = 2$ only, in a paper [1] of Ahlswede and Zhang.

It is also stated in [1] that (1) is just the LYM inequality for the poset (evidently, $R(n_1, n_2)$), which is not quite true.Let P be a ranked poset, with rank function $r : P \to \{0, 1, \ldots\}$. Let P_k denote the set of elements with rank k. Let $N_P(x)$ denote the number of elements of rank $r(x)$. We recall that P is said to be LYM provided that for every antichain $I \subseteq P$,

$$\sum_{x\in I} \frac{1}{N_P(x)} \leq 1.$$

It is well known that $R(n_1, \ldots, n_M)$ is LYM. (See [4] for a survey.) Note that the contribution of an element $x \in I$ to the sum in the LYM inequality depends only on its rank, which is not the case for inequality (1).

Our second proof of Theorem 1.1 shows that it is indeed the LYM inequality but for a weighted poset obtained naturally as a quotient of the Boolean lattice B_n of all subsets of X.

We must mention that (1) is in fact just a special case of an earlier inequality which lies at the heart of the proof of the product theorem for LYM posets, as presented in the survey by Greene and Kleitman ([4], p. 42). They show that for LYM, rank-log-concave posets P_1 and P_2 and maximum chains $C_1 \subseteq P_1$ and $C_2 \subseteq P_2$, every antichain $I \subseteq P_1 \times P_2$ satisfies

$$\sum_{(i_1,i_2)\in I\cap(C_1\times C_2)} \frac{N_{P_1}(i_1)N_{P_2}(i_2)}{N_{P_1\times P_2}(i_1, i_2)} \leq 1. \tag{2}$$

We obtain (1) when we take P_i to be the Boolean lattice B_{n_i} for $i = 1, 2$ in (2). Restricting the proof of Greene and Kleitman to this instance gives another proof of (1), although we cannot yet see how to extend it to prove Theorem 1.1 for general M. However, looking at (2) and Theorem 1.1 together, a common generalization is suggested, with (2) extended to general M and Theorem 1.1 extended to arbitrary LYM, rank-log-concave posets.

THEOREM 1.2. *If $P_1, \ldots, P_m$ are LYM and rank-log-concave posets, and $C_i \subseteq P_i$ are maximum chains $(i = 1, \ldots, m)$, then for any antichain $I \subseteq P_1 \times \cdots \times P_m$,*

$$\sum_{(i_1,\ldots,i_m)\in I\cap(C_1\times\cdots\times C_m)} \frac{N_{P_1}(i_1)\cdots N_{P_m}(i_m)}{N_{P_1\times\cdots\times P_m}(i_1,\ldots,i_m)} \leq 1.$$

We use the LYM Product Theorem of Harper, for weighted posets, to derive this result in Section 3. Note that it restricts to yet another proof of Theorem 1.1 when $P_i = B_{n_i}$.

2. Two proofs of Theorem 1.1

FIRST PROOF OF THEOREM 1.1. Suppose that I is an antichain as stated in Theorem 1.1. The total number of maximal chains in the product poset $\{0,\ldots,n_1\} \times \cdots \times \{0,\ldots,n_M\}$ is given by

$$\binom{n_1+\cdots+n_M}{n_1,\ldots,n_M} := \frac{(n_1+\cdots+n_M)!}{n_1!\cdots n_M!}.$$

For any vector $(i_1,\ldots,i_M)$, the number of maximal chains that pass through it is given by

$$\binom{i_1+\cdots+i_M}{i_1,\cdots,i_M}\binom{n_1-i_1+\cdots+n_M-i_M}{n_1-i_1,\ldots,n_M-i_M}.$$

Finally, since I is an antichain,

$$\sum_{(i_1,\ldots,i_M)\in I} \binom{i_1+\cdots+i_M}{i_1,\ldots,i_M}\binom{n_1-i_1+\cdots+n_M-iM}{n_1-i_1,\ldots,n_M-i_M} \leq \binom{n_1+\cdots+n_M}{n_1,\ldots,n_M},$$

and (1.1) follows after rewriting this last expression. □

SECOND PROOF OF THEOREM 1.1. We need to recall a well-known result derived from Lubell's proof of Sperner's Theorem.

THEOREM 2.1. *The Boolean lattice B_n of subsets of X has the LYM property.* □

A *weighted poset* is a pair (P, v), with P a finite ranked poset and v a function that assigns a positive real number to each element of P. A weighted poset (P, v) satisfies the *LYM inequality* if for any antichain $I \subseteq P$,

$$\sum_{x\in I} \frac{v(x)}{v(P_{r(x)})} \leq 1.$$

If P is a poset and G is a group of automorphisms of P, then the *quotient poset* P/G consists of the orbits of P under G ordered by $A \leq B$ in P/G whenever there exist $x \in A$ and $y \in B$ with $x \leq y$ in P.

We will use the following theorem due essentially to Harper (1974) [6]. (See [2] for a complete treatment.)

THEOREM 2.2. *A finite ranked poset P has the LYM property if and only if $(P/G, v)$ has the LYM property, where G is any subgroup of the group of automorphisms of P and $v(A)$ is the size $|A|$ of the class $A \in P/G$.* □

Now consider the subgroup G of permutations of X that are color preserving, that is, if $\sigma \in G$, $\sigma(X_i) \subseteq X_i$, for each $1 \leq i \leq M$. Clearly, G induces a subgroup of the group of automorphisms of 2^X, which we will still call G. It is immediate to check that the quotient poset $2^X/G$ with the canonical weight function as described in Theorem 2.2 is isomorphic to the weighted poset

$$P = (\{0, \ldots, n_1\} \times \cdots \times \{0, \ldots, n_M\}, v),$$

where $v((i_1, \ldots, i_M)) = \binom{n_1}{i_1} \cdots \binom{n_M}{i_M}$.

Now, since 2^X is LYM, by Theorem 2.2 (we are using the 'easy direction'), P is LYM. Hence, if $I \subseteq \{0, \ldots, n_1\} \times \cdots \times \{0, \ldots, n_M\}$ is an antichain, the LYM inequality ensures that

$$\sum_{(i_1, \ldots, i_M) \in I} \frac{v((i_1, \ldots, i_M))}{v(P_{r((i_1, \ldots, i_M))})} \leq 1.$$

Finally, the stated inequality follows from

$$\begin{aligned} v(P_{r((i_1, \ldots, i_M))}) &= \sum_{\substack{x_1 + \cdots + x_M = i_1 + \cdots + i_M \\ 0 \leq x_i \leq n_i}} \binom{n_1}{x_1} \cdots \binom{n_M}{x_M} \\ &= \binom{n_1 + \cdots + n_M}{i_1 + \cdots + i_M}. \end{aligned}$$

□

3. The proof of Theorem 1.2

A weighted poset (P, v) is said to be *weight-log-concave* if the sequence $\{v(P_k)\}$ is log-concave. We recall the following Product Theorem due to Harper [6].

THEOREM 3.1. *If (P_1, v_1) and (P_2, v_2) are weight-log-concave and satisfy the LYM inequality, then $(P_1 \times P_2, v_1 v_2)$ also satisfies the LYM inequality and is weight-log-concave.* □

By induction we obtain the following

COROLLARY 3.2. *If $(P_1, v_1), \ldots, (P_M, v_M)$ are weight-log-concave and satisfy the LYM inequality, then $(P_1 \times \cdots \times P_M, v_1 \cdots v_M)$ also satisfies the LYM inequality and is weight-log-concave.* □

To prove Theorem 1.2 we consider the weighted posets $(C_1, N_{P_1}), \ldots, (C_M, N_{P_M})$ and apply the corollary. The inequality in Theorem 1.2 is just the LYM inequality for $(C_1 \times \cdots \times C_M, N_{P_1} \cdots N_{P_M})$. □

REFERENCES

[1] AHLSWEDE, R. and ZHANG, Z., On cloud-antichains and related configurations, *Discrete Math.* **85** (1990), 225–245. *MR* **91i**:05003

[2] CHUDAK, F., On quotient posets and the LYM inequality and convex hulls of families of subsets, Master Thesis, Department of Mathematics, University of South Carolina, 1994.

[3] ERDŐS, P. L. and KATONA, G. O. H., Convex hulls of more-part Sperner families, *Graphs Combin.* **2** (1986), 123–134. *MR* **89d**:05002

[4] GREENE, C. and KLEITMAN, D. J., Proof techniques in the theory of finite sets, *Studies in Combinatorics* (ed. G. C. Rota), MAA Stud. Math. 17, Math. Assoc. America, Washington, DC, 1978, 22–79. *MR* **80a**:05006

[5] GRIGGS, J. R., Collections of subsets with the Sperner property, *Trans. Amer. Math. Soc.* **269** (1982), 575–591. *MR* **83d**:05003

[6] HARPER, L. H., The morphology of partially ordered sets, *J. Combinatorial Theory Ser. A* **17** (1974), 44–58. *MR* **51** #3008

[7] LUBELL, D., A short proof of Sperner's lemma, *J. Combinatorial Theory* **1** (1966), 299. *MR* **33** #2558

(Received February 7, 1997)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208
U.S.A.

Current address:

IBM
TJ WATSON RESEARCH CENTER
YORKTOWN HEIGHTS, NY 10598
U.S.A.

chudak@watson.ibm.com

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208
U.S.A

griggs@math.sc.edu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 353–358

AN ORLICZ–PETTIS THEOREM WITH APPLICATIONS TO $\mathcal{A}$-SPACES

J. WU and R. LI

1. Introduction

Let (E,τ) be a Hausdorff locally convex topological vector space (lcs) with continuous dual E'. Let E^s (E^b) be the space of all sequentially continuous (bounded) linear functionals defined on E. If E and F are a pair of vector spaces in duality, let $\sigma(E,F)(\tau(E,F),\beta(E,F))$ be the weak topology (Mackey topology, strong topology) on E from this duality. If X and Y are topological vector spaces, let $L(X,Y)(L_s(X,Y),B(X,Y))$ be the space of all continuous (sequentially continuous, bounded) linear operators from X into Y.

A series $\sum_j x_j$ in (E,τ) is said to be subseries convergent if for each nonempty $\Delta=\{j_1<j_2<\dots\}\subseteq\mathbf{N}$ there exists an $x_\Delta\in E$ such that $\sum_k x_{j_k}$ is τ-convergent to x_Δ. The classical Orlicz–Pettis theorem states that if the series $\sum_j x_j$ is subseries $\sigma(E,E')$-convergent, then $\sum_j x_j$ is also subseries $\tau(E,E')$-convergent ([3], [7]). In general, the series $\sum_j x_j$ is subseries $\sigma(E,E')$-convergent does not imply it must also be subseries $\beta(E,E')$-convergent ([3]).

A sequence $\{x_k\}$ in (E,τ) is said to be $\tau-\mathcal{K}$-convergent if each subsequence of $\{x_k\}$ has a subsequence $\{x_{n_k}\}$ such that the series $\sum_k x_{n_k}$ is τ-convergent to an element of E ([1], §3). A $\tau-\mathcal{K}$-convergent sequence is τ-convergent to 0, the converse does not hold, except in complete metric linear spaces ([1], §3). A subset B of (E,τ) is said to be $\tau-\mathcal{K}$-bounded if whenever $\{x_k\}\subseteq B$ and $\{t_k\}$ is a scalar sequence converging to 0, the sequence $\{t_kx_k\}$ is $\tau-\mathcal{K}$-convergent ([1], §3). A $\tau-\mathcal{K}$-bounded set is τ-bounded but, in general, the converse does not hold ([1], §3).

(E,τ) is said to be an $\mathcal{A}$-space if every τ-bounded subset of (E,τ) is $\tau-\mathcal{K}$-bounded ([4], Definition 3). Li Ronglu and Swartz in ([4], Proposition 5)

1991 *Mathematics Subject Classification.* Primary 46A03.
Key words and phrases. Orlicz–Pettis theorem, $\mathcal{A}$-space, full-invariant.

0081–6906/99/$ 5.00

proved that each sequentially complete locally convex space is an $\mathcal{A}$-space. But, in general, the converse does not hold. Locally convex spaces which are $\mathcal{A}$-spaces have been shown to enjoy many important properties, particular with respect to the Uniform Bounded Principle (UBP) and hypocontinuity for bilinear operators ([2], [4], [11]). $\mathcal{A}$-spaces seem to be a very natural class of spaces for which UBP holds. In this paper, at first, we prove an interesting result (Theorem 3) which can be viewed as an Orlicz–Pettis Theorem for l^p-multiplier convergent series $(0 < p \leqq 1)$. This result shows that an l^p-multiplier convergent series $(0 < p \leqq 1)$ is invariant with respect to all admissible topologies. From it we can show that $\mathcal{A}$-spaces under any admissible topology are still $\mathcal{A}$-spaces. That is, l^p-multiplier convergent series $(0 < p \leqq 1)$ and $\mathcal{A}$-spaces are invariant with respect to all admissible topologies. Note that these kinds of full-invariants in locally convex spaces theory are rare, except c_0-multiplier convergent series, l^p-multiplier convergent series $(1 \leqq p < +\infty)$ and some trivial facts [6]. Moreover, we also show that for each $E \in$ lcs, then $(E^s, \beta(E^s, E))$ is an $\mathcal{A}$-space. This implies that from each $E \in$ lcs we can obtain a large supply of $\mathcal{A}$-spaces.

2. An Orlicz–Pettis Theorem for l^p-mc series

Let $E \in$ lcs and $\sum_j x_j$ be a series in E. If for each $\{t_j\} \in l^p$ $(0 < p \leqq 1)$ the series $\sum_j t_j x_j$ converges in E, then the series $\sum_j x_j$ is said to be l^p-multiplier convergent (l^p-mc). Recently, Li and Swartz in [5] gave a few characterizations of Banach–Mackey spaces ([10] 10.4.3); this result is related to l^1-mc as below.

THEOREM 1. *For $E \in$ lcs, the following conditions are equivalent.*
(1) *E is Banach–Mackey space.*
(2) *If $\{x'_j\} \subseteqq E'$ is $\sigma(E', E)$-bounded and $\{t_j\} \in l^1$, then $\sum_j t_j x'_j \in E^s$.*
(3) *If $\{x'_j\} \subseteqq E'$ is $\sigma(E', E)$-bounded and $\{t_j\} \in l^1$, then $\sum_j t_j x'_j \in E^b$.*
(4) *If $\{x'_j\} \subseteqq E'$ is $\sigma(E', E)$-Cauchy, then $\lim_j x'_j \in E^b$.*

Now, we prove an Orlicz–Pettis Theorem for l^p-mc series $(0 < p \leqq 1)$ $\sum_j x_j$.

One of the tools used in the proofs below is the matrix theorem of Antosik and Mikusiński.

THEOREM 2 (Antosik–Mikusiński). *Let X be a topological vector space and $x_{ij} \in X$ for $i, j \in \mathbf{N}$. If*
(I) *$\lim_i x_{ij} = x_j$ exists for every j and*

(II) *for every increasing sequence of positive integers* $\{m_j\}$ *there is a subsequence* $\{n_j\}$ *of* $\{m_j\}$ *such that the sequence* $\left\{\sum_j x_{in_j}\right\}_i$ *converges, then* $\lim_i x_{ij} = x_j$ *uniformly for* $j \in \mathbf{N}$. *In particular,* $\lim_i x_{ii} = 0$.

A matrix $M = [x_{ij}]$ satisfying conditions (I) and (II) is called a $\mathcal{K}$-matrix. For proofs of more general forms of Theorem 2 see ([1], [4], [8]).

THEOREM 3. *Let* $E \in$ lcs *and* $\sum_j x_j$ *be a series in* E *and let* $0 < p \leqq 1$. *Then for every* $\{t_j\} \in l^p$ *the series* $\sum_j t_j x_j$ *is* $\sigma(E, E')$*-convergent if and only if for every* $\{t_j\} \in l^p$, *the series* $\sum_j t_j x_j$ *is* $\beta(E, E')$*-convergent.*

PROOF. $\Leftarrow$ is trivial.

$\Rightarrow$ At first, we show that $\{x_j\}$ is $\beta(E, E')$-bounded. Indeed, for every $\sigma(E', E)$-bounded subset B of E', it suffices to show that $\{\langle x_j, x'_j\rangle\}$ is bounded whenever $\{x'_j\} \subseteq B$. Let $t_j > 0$, $\lim t_j = 0$. Consider the matrix $M = \left[\langle \sqrt{t_j} x_j, \sqrt{t_i} x'_i\rangle\right]$. Since $\{x'_i\}$ is $\sigma(E', E)$-bounded, the columns of M converge to 0. If $\{m_j\}$ is an increasing sequence of positive integers, there is a subsequence $\{n_j\}$ of $\{m_j\}$ such that $\left\{\sqrt{t_{n_j}}\right\} \in l^p$ $(0 < p \leqq 1)$. It is not difficult to see that the series $\sum_j \sqrt{t_{n_j}} x_{n_j}$ is $\sigma(E, E')$-convergent to some $x \in E$. So, $\left\langle \sum_j \sqrt{t_{n_j}} x_{n_j}, \sqrt{t_i} x'_i \right\rangle = \sqrt{t_i}\langle x, x'_i\rangle \to 0$, and from Theorem 2, it follows that $t_j \langle x_j, x'_j\rangle \to 0$ so $\{\langle x_j, x'_j\rangle\}$ is bounded. This shows that $\{x_j\}$ is $\beta(E, E')$-bounded. Next, we show that for every $\{t_j\} \in l^p$ $(0 < p \leqq 1)$, the series $\sum_j t_j x_j$ is $\beta(E, E')$-convergent. Since $\{t_j\} \in l^p$ $(0 < p \leqq 1)$, we must have $\{t_j\} \in l^1$. Thus, we may suppose that $\sum_j |t_j| \leqq 1$ and $\sum_j t_j x_j$ is $\sigma(E, E')$-convergent to $x \in E$. Let $S_n = \sum_j^n t_j x_j$, then $\{S_n\}$ is $\sigma(E, E')$-convergent to x. We prove that $\{S_n\}$ is also $\beta(E, E')$-convergent to x. For every neighbourhood U of 0 in $(E, \beta(E, E'))$ there is a bounded subset A of $(E', \sigma(E', E))$ such that $A^0 \subseteq U$ and A^0 is $\sigma(E, E')$-closed and balanced. Here A^0 is the polar of A.

Since $\{x_j\}$ is $\beta(E, E')$-bounded and A is $\sigma(E', E)$-bounded, there is an $M > 0$ such that

$$\operatorname{Sup}\{|\langle x_j, y\rangle| : j \in \mathbf{N}, y \in A\} \leqq M.$$

Note that $\{t_j\} \in l^1$, there is $n_0 \in \mathbf{N}$ such that whenever $m, n \in \mathbf{N}$ and

$m > n \geqq n_0$ we have $\sum_{j=n}^{m} |t_j| M \leqq 1$. Thus, we have

$$\sup\left\{\left|\left\langle \sum_{j=n}^{m} t_j x_j, y \right\rangle\right| : y \in A\right\} \leqq \sum_{j=n}^{m} |t_j| M \leqq 1.$$

That is

$$S_m - S_{n-1} = \sum_{j=n}^{m} t_j x_j \in A^0 \subseteqq U.$$

Since A^0 is $\sigma(E, E')$-closed and balanced and $\{S_m\}$ is $\sigma(E, E')$-convergent to x, letting $m \to \infty$ we have $x - S_{n-1} \in A^0 \subseteqq U$. This shows that whenever $n \geqq n_0$ we have $S_{n-1} - x \in U$. It follows that $\sum_j t_j x_j$ is $\beta(E, E')$-convergent to x.

From Theorem 3 we can obtain a few important corollaries as follows.

COROLLARY 4. *Let (E, F) be a dual pair. Then $A \subseteqq E$ is $\sigma(E, F) - \mathcal{K}$-bounded set if and only if A is $\beta(E, F) - \mathcal{K}$-bounded set.*

This corollary improves Theorem 3.3.10 of [9].

COROLLARY 5. *Let (E, F) be a dual pair and $\sum_j x_j$ be a series in E. Then for every $\{t_j\} \in l^p$ $(0 < p \leqq 1)$ the series $\sum_j t_j x_j$ is $\sigma(E, F)$-convergent if and only if for any topology τ on E admissible with respect to (E, F), the series $\sum_j t_j x_j$ is τ-convergent for every $\{t_j\} \in l^p$ $(0 < p \leqq 1)$.*

COROLLARY 6. *Let (E, F) be a dual pair. Then all topologies on E admissible with respect to (E, F) have the same $\mathcal{K}$-bounded sets. In particular, if (E, τ) is an $\mathcal{A}$-space, then E is also an $\mathcal{A}$-space for any topology on E admissible with respect to (E, E').*

It follows from Corollary 6 that if (E, τ) is an $\mathcal{A}$-space, then $(E, \beta(E, E'))$ is also an $\mathcal{A}$-space. But, in general, the converse does not hold.

EXAMPLE 7. Let c_{00} be the space of all sequences which are eventually 0 and τ be the Sup-norm. Then $(c_{00}, \tau)' = l^1$. Consider the dual pair (l^1, c_{00}). $(l^1, \beta(l^1, c_{00})) = (l^1, \|\cdot\|_1)$ is a Banach space and, therefore, is an $\mathcal{A}$-space. But $(l^1, \sigma(l^1, c_{00}))$ is not an $\mathcal{A}$-space. In fact, if e_k is the sequence with a 1 in the kth coordinate and 0 elsewhere, then $\{ke_k\} \subseteqq l^1$ is $\sigma(l^1, c_{00})$-bounded, but $\{ke_k\}$ is not $\beta(l^1, c_{00})$-bounded, it follows from Corollary 6 that $\{ke_k\}$ is not $\sigma(l^1, c_{00}) - \mathcal{K}$-bounded.

3. Properties of $\mathcal{A}$-spaces

It follows from Corollary 6 that if (E,τ) is an $\mathcal{A}$-space, then (E,τ) is a Banach–Mackey space ([10], 10.4.3, [9], Theorem 3.3.12). In [5], Li and Swartz gave several characterizations of Banach–Mackey spaces; from them we can obtain a few basic properties of $\mathcal{A}$-spaces as follows.

THEOREM 1′. *If (E,τ) is an $\mathcal{A}$-space, then we have*

(2′) *If $\{x'_j\} \subseteq E'$ is $\sigma(E',E)$-bounded and $\{t_j\} \in l^p$ $(0 < p \leqq 1)$, then $\sum_j t_j x'_j \in E^s$.*

(3′) *If $\{x'_j\} \subseteq E'$ is $\sigma(E',E)$-bounded and $\{t_j\} \in l^p$ $(0 < p \leqq 1)$, then $\sum_j t_j x'_j \in E^b$.*

(4) *If $\{x'_j\} \subseteq E'$ is $\sigma(E',E)$-Cauchy, then $\lim_j x'_j \in E^b$.*

THEOREM 8. *If (E,τ) is an $\mathcal{A}$-space, then we have* (see [5], Theorems 21 and 22)

(5) *For every locally convex space F, $B((E,\beta(E,E')),F) \subseteq B(E,F)$.*

(6) *For every locally convex space F, $L((E,\beta(E,E')),F) \subseteq B(E,F)$.*

(7) $(E,\beta(E,E'))' \subseteq E^b$.

(8) $E'' \subseteq (E')^b$.

(9) *For every locally convex space F and every pointwise bounded family $\Gamma \subseteq L_s(E,F)$ is uniformly bounded on bounded subsets of E.*

(10) *For every locally convex space F if $\{T_k\} \subseteq L_s(E,F)$ and $\lim_k T_k x = Tx$ exists for every $x \in E$, then $T \in B(E,F)$.*

As the following theorem shows from every locally convex space we can obtain a large supply of $\mathcal{A}$-spaces.

THEOREM 9. *Let (E,τ) be a locally convex space, then $(E^s,\beta(E^s,E))$ is an $\mathcal{A}$-space.*

PROOF. Consider the dual pair (E,E^s). Then we have $(E,\sigma(E,E^s))' = E^s$ ([10], Theorem 8.2.12) and $(E,\sigma(E,E^s))^s = E^s$. In fact, let $f \in (E,\sigma(E,E^s))^s$ and $x_n \to 0$ in (E,τ). Then for every $g \in E^s$, $g(x_n) \to 0$. Thus $x_n \to 0$ in $(E,\sigma(E,E^s))$ and hence, $f(x_n) \to 0$. This shows that $f \in E^s$. That is, $(E,\sigma(E,E^s))^s = E^s$. Thus, $(E,\sigma(E,E^s))$ is a Mazur space ([10], 8.6.3). It follows from ([10], 8.6.6) that $(E^s,\beta(E^s,E))$ is complete. By ([4], Proposition 5), $(E^s,\beta(E^s,E))$ is an $\mathcal{A}$-space.

Let $(E^s,\beta(E^s,E))' = E^{s\prime}$. Consider the dual pair $E^s, E^{s\prime}$. Then from Corollary 6 and Theorem 9 we know that for any topology T on E^s admissible with respect to $(E^s,E^{s\prime})$, (E^s,T) is an $\mathcal{A}$-space. This shows that from every locally convex space (E,τ), we can obtain a large supply of $\mathcal{A}$-spaces. See also Theorem 3.4.8 of [9].

From ([9], Theorem 3.3.12) and ([10], Theorem 10.4.12) we have immediately

THEOREM 10. *If (E,τ) is a quasi-barrelled $\mathcal{A}$-space, then (E,τ) is barrelled. In particular, a bornological or metric $\mathcal{A}$-space is barrelled.*

REFERENCES

[1] ANTOSIK, P. and SWARTZ, C., *Matrix methods in analysis*, Lecture Notes in Mathematics, 1113, Springer-Verlag, Berlin – New York, 1985. *MR* **87b**:46079

[2] ANTOSIK, P. and SWARTZ, C., Boundedness and continuity for bilinear operators, *Studia Sci. Math. Hungar.* **29** (1994), 387–395. *MR* **95m**:47001

[3] DIEROLF, P., Theorems of the Orlicz–Pettis-type for locally convex spaces, *Manuscripta Math.* **20** (1977), 73–94. *MR* **55** #1018

[4] LI, R. and SWARTZ, C., Spaces for which the uniform boundedness principle holds, *Studia Sci. Math. Hungar.* **27** (1992), 379–384. *MR* **94h**:46015

[5] LI, R. and SWARTZ, C., Characterizations of Banach–Mackey spaces, *Chinese J. Math.* **24** (1996), 199–210. *MR* **97f**:46008

[6] LI, R. and CUI, C., An invariant with respect to all admissible (E,E')-polar topologies, *Chinese Ann. Math.* **19A** (1998), 289–294.

[7] MCARTHUR, C. W., On a theorem of Orlicz and Pettis, *Pacific J. Math.* **22** (1967), 297–302. *MR* **35** #4702

[8] SWARTZ, C., *An introduction to functional analysis*, Monographs and textbooks in pure and applied mathematics, 157, Marcel Dekker, New York, 1992. *MR* **93c**: 46002

[9] SWARTZ, C., *Infinite matrices and the gliding hump*, World Scientific Publ., Singapore, 1996. *MR* **98b**: 46002

[10] WILANSKY, A., *Modern methods in topological vector spaces*, McGraw-Hill, New York, 1978. *MR* **81d**:46001

[11] WU, J. and LI, R., Hypocontinuity and uniform boundedness for bilinear maps, *Studia Sci. Math. Hungar.* **35** (1999), 133–138.

(Received February 26, 1997)

J. Wu
DEPARTMENT OF MATHEMATICS
DAQING PETROLEUM INSTITUTE
ANDA 151400
PEOPLE'S REPUBLIC OF CHINA

Present address:

DEPARTMENT OF MATHEMATICS
ZHE JIANG UNIVERSITY
HANG ZHOU 310 027
PEOPLE'S REPUBLIC OF CHINA

R. Li
DEPARTMENT OF MATHEMATICS
HARBIN INSTITUTE OF TECHNOLOGY
HARBIN 150006
PEOPLE'S REPUBLIC OF CHINA

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 359–373

ON MIRON'S GEOMETRY IN Osc^3M. II

IRENA ČOMIĆ

Abstract

R. Miron and Gh. Atanasiu in [15], [16], [17] studied the geometry of Osc^kM. Among many various problems which were solved, they introduced the adapted basis, the d-connection and gave its curvature theory. Different structures as almost product structure, metric structure were determined.

Here the attention on $E = Osc^3M$ will be restricted especially on variational problem and Zermello's conditions, but the transformation group is slightly different from that used in [15]. It will result in different theory.

1. Adapted basis in $T(Osc^3M)$ and $T^*(Osc^3M)$

Let $E = Osc^3M$ be a $4n$-dimensional C^∞-manifold. In some local chart (U, φ) some point $u \in E$ has coordinates

$$(x^a, y^{1a}, y^{2a}, y^{3a}) = (y^{0a}, y^{1a}, y^{2a}, y^{3a}) = (y^{\alpha a}),$$

where $x^a = y^{0a}$ and

$$a, b, c, d, e, \ldots = 1, 2, \ldots, n, \quad \alpha, \beta, \gamma, \delta, \kappa, \ldots = 0, 1, 2, 3.$$

If in some other chart (U', φ') the point $u \in E$ has coordinates $(x^{a'}, y^{1a'}, y^{2a'}, y^{3a'})$, then in $U \cap U'$ the allowed coordinate transformations are given by

$$\begin{aligned}
&\text{(a)} && x^{a'} = x^{a'}(x^1, x^2, \ldots, x^n)\\
&\text{(b)} && y^{1a'} = \frac{\partial x^{a'}}{\partial x^a} y^{1a} = \frac{\partial y^{0a'}}{\partial y^{0a}} y^{1a}\\
&\text{(c)} && y^{2a'} = \frac{\partial y^{1a'}}{\partial y^{0a}} y^{1a} + \frac{\partial y^{1a'}}{\partial y^{1a}} y^{2a}\\
&\text{(d)} && y^{3a'} = \frac{\partial y^{2a'}}{\partial y^{0a}} y^{1a} + \frac{\partial y^{2a'}}{\partial y^{1a}} y^{2a} + \frac{\partial y^{2a'}}{\partial y^{2a}} y^{3a}.
\end{aligned} \tag{1.1}$$

1991 *Mathematics Subject Classification.* Primary 53B25; Secondary 53B40.
Key words and phrases. Lagrange spaces of third order, variation problems.

Some nice examples of the space E can be obtained if the points $(x^a) \in M$ ($\dim M = n$) are considered as the points of the curve $x^a = x^a(t)$ and $y^{\alpha a}$, $\alpha = 1, 2, 3$ are defined by

$$y^{1a} = \frac{dx^a}{dt}, \quad y^{2a} = \frac{d^2x^a}{dt^2} = \frac{dy^{1a}}{dt}, \quad y^{3a} = \frac{d^3x^a}{dt^3} = \frac{dy^{2a}}{dt}.$$

M is the base manifold and $(x^a) \in M$ is the projection of $(x^a, y^{1a}, y^{2a}, y^{3a}) \in E$ on M. In [15], [16] $y^{\alpha a} = \frac{1}{\alpha!}\frac{d^\alpha x^a}{dt^\alpha}$, $\alpha = 1, \ldots, k$ and the transformations (1.1) have different form. If in $U \cap U'$ the equation

$$x^{a'} = x^{a'}(x^1(t), x^2(t), \ldots, x^n(t))$$

is valid, then it is easy to see that

$$\begin{aligned} y^{1a'} &= \frac{dx^{a'}}{dt} = y^{1a'}(x^a, y^{1a}), \\ y^{2a'} &= \frac{dy^{1a'}}{dt} = y^{2a'}(x^a, y^{1a}, y^{2a}), \\ y^{3a'} &= \frac{dy^{2a'}}{dt} = y^{3a'}(x^a, y^{1a}, y^{2a}, y^{3a}), \end{aligned} \tag{1.2}$$

satisfy (1.1b), (1.1c) and (1.1d), respectively, and the explicit form of (1.1) is the following:

$$\begin{aligned} x^{a'} &= x^{a'}(x^1, x^2, \ldots, x^n) \\ y^{1a'} &= \frac{\partial x^{a'}}{\partial x^a} y^{1a}, \\ y^{2a'} &= \frac{\partial^2 x^{a'}}{\partial x^a \partial x^b} y^{1a} y^{1b} + \frac{\partial x^{a'}}{\partial x^a} y^{2a}, \\ y^{3a'} &= \frac{\partial^3 x^{a'}}{\partial x^a \partial x^b \partial x^c} y^{1a} y^{1b} y^{1c} + 3\frac{\partial^2 x^{a'}}{\partial x^a \partial x^b} y^{1a} y^{2b} + \frac{\partial x^{a'}}{\partial x^a} y^{3a}. \end{aligned} \tag{1.3}$$

THEOREM 1.1. *The transformations determined by* (1.1) *form a group.*

With determination of the group of allowable coordinate transformations the first step to construction of some geometry is made. The second important step is the construction of the adapted basis in $T(E)$, which depends on the choice of the coefficients of the nonlinear connections, here denoted by N and M.

The following abbreviations

$$\partial_{\alpha a} = \frac{\partial}{\partial y^{\alpha a}}, \quad \alpha = 1, 2, 3, \quad \text{and } \partial_a = \partial_{0a} = \frac{\partial}{\partial x^a} = \frac{\partial}{\partial y^{0a}}$$

will be used. From (1.3) it follows

$$\partial_{0a}y^{0a'} = \partial_{1a}y^{1a'} = \partial_{2a}y^{2a'} = \partial_{3a}y^{3a'} = \frac{\partial x^{a'}}{\partial x^a} = A^{a'}_a,$$

$$\frac{dA^{a'}_a}{dt} = \partial_{0a}y^{1a'} = \frac{1}{2}\partial_{1a}y^{2a'} = \frac{1}{2}\frac{2}{3}\partial_{2a}y^{3a'} = \frac{\partial^2 x^{a'}}{\partial x^a \partial x^b}y^{1b} = B^{a'}_a, \tag{1.4}$$

$$\frac{dB^{a'}_a}{dt} = \partial_{0a}y^{2a'} = \frac{1}{3}\partial_{1a}y^{3a'} = \frac{\partial^3 x^{a'}}{\partial x^a \partial x^b \partial x^c}y^{1b}y^{1c} + \frac{\partial^2 x^{a'}}{\partial x^a \partial x^b}y^{2b} = C^{a'}_a,$$

$$\frac{dC^{a'}_a}{dt} = \partial_{0a}y^{3a'} = D^{a'}_a.$$

The natural basis $\bar{B}$ of $T(E)$ is

$$\bar{B} = \{\partial_{0a}, \partial_{1a}, \partial_{2a}, \partial_{3a}\} = \{\partial_{\alpha a}\}. \tag{1.5}$$

The elements of $\bar{B}$ with respect to (1.1) are not transformed as d-tensors. They satisfy the following relations:

$$\begin{aligned}
\partial_{0a} &= (\partial_{0a}y^{0a'})\partial_{0a'} + (\partial_{0a}y^{1a'})\partial_{1a'} + (\partial_{0a}y^{2a'})\partial_{2a'} + (\partial_{0a}y^{3a'})\partial_{3a'} \\
\partial_{1a} &= \qquad\qquad (\partial_{1a}y^{1a'})\partial_{1a'} + (\partial_{1a}y^{2a'})\partial_{2a'} + (\partial_{1a}y^{3a'})\partial_{3a'} \\
\partial_{2a} &= \qquad\qquad\qquad\qquad (\partial_{2a}y^{2a'})\partial_{2a'} + (\partial_{2a}y^{3a'})\partial_{3a'} \\
\partial_{3a} &= \qquad\qquad\qquad\qquad\qquad\qquad (\partial_{3a}y^{3a'})\partial_{3a'}.
\end{aligned} \tag{1.6}$$

The natural basis $\bar{B}^*$ of $T^*(E)$ is

$$\bar{B}^* = \{dx^a, dy^{1a}, dy^{2a}, dy^{3a}\} = \{dy^{\alpha a}\}. \tag{1.7}$$

The elements of $\bar{B}^*$ with respect to (1.1) are transformed in the following way (see (1.2)):

$$\begin{aligned}
dx^{a'} &= \frac{\partial x^{a'}}{\partial x^a}dx^a \Leftrightarrow dy^{0a'} = (\partial_{0a}y^{0a'})dy^{0a} \\
dy^{1a'} &= (\partial_{0a}y^{1a'})dy^{0a} + (\partial_{1a}y^{1a'})dy^{1a} \\
dy^{2a'} &= (\partial_{0a}y^{2a'})dy^{0a} + (\partial_{1a}y^{2a'})dy^{1a} + (\partial_{2a}y^{2a'})dy^{2a} \\
dy^{3a'} &= (\partial_{0a}y^{3a'})dy^{0a} + (\partial_{1a}y^{3a'})dy^{1a} + (\partial_{2a}y^{3a'})dy^{2a} + (\partial_{3a}y^{3a'})dy^{3a}.
\end{aligned} \tag{1.8}$$

The adapted basis B^* of $T^*(E)$ is given by

$$B^* = \{\delta y^{0a}, \delta y^{1a}, \delta y^{2a}, \delta y^{3a}\}, \tag{1.9}$$

where

$$\begin{aligned}
\delta y^{0a} &= dx^a = dy^{0a} \\
\delta y^{1a} &= dy^{1a} + M^{1a}_{0b}dy^{0b} \\
\delta y^{2a} &= dy^{2a} + M^{2a}_{1b}dy^{1b} + M^{2a}_{0b}dy^{0b} \\
\delta y^{3a} &= dy^{3a} + M^{3a}_{2b}dy^{2b} + M^{3a}_{1b}dy^{1b} + M^{3a}_{0b}dy^{0b}.
\end{aligned} \tag{1.10}$$

THEOREM 1.2. *The necessary and sufficient conditions that* $\delta y^{\alpha a}$ *are transformed as d-tensor field, i.e.*

$$\delta y^{\alpha a'} = \frac{\partial x^{a'}}{\partial x^a}\delta y^{\alpha a}, \quad \alpha = 0,1,2,3,$$

are the following equations:

$$\begin{aligned}
&\text{(a)} \quad M^{1a}_{0b}\partial_{1a}y^{1a'} = M^{1a'}_{0b'}\partial_{0b}y^{0b'} + \partial_{0b}y^{1a'}\\
&\text{(b)} \quad M^{2a}_{1b}\partial_{2a}y^{2a'} = M^{2a'}_{1c'}\partial_{1b}y^{1c'} + \partial_{1b}y^{2a'}\\
&\text{(c)} \quad M^{2a}_{0b}\partial_{2a}y^{2a'} = M^{2a'}_{0c'}\partial_{0b}y^{0c'} + M^{2a'}_{1c'}\partial_{0b}y^{1c'} + \partial_{0b}y^{2a'}\\
&\text{(d)} \quad M^{3a}_{2b}\partial_{3a}y^{3a'} = M^{3a'}_{2c'}\partial_{2b}y^{2c'} + \partial_{2b}y^{3a'}\\
&\text{(e)} \quad M^{3a}_{1b}\partial_{3a}y^{3a'} = M^{3a'}_{1c'}\partial_{1b}y^{1c'} + M^{3a'}_{2c'}\partial_{1b}y^{2c'} + \partial_{1b}y^{3a'}\\
&\text{(f)} \quad M^{3a}_{0b}\partial_{3a}y^{3a'} = M^{3a'}_{0c'}\partial_{0b}y^{0c'} + M^{3a'}_{1c'}\partial_{0b}y^{1c'} + M^{3a'}_{2c'}\partial_{0b}y^{2c'} + \partial_{0b}y^{3a'}.
\end{aligned} \tag{1.11}$$

From (1.11) and (1.4) it follows that (1.11) is a system in which equations of second, third and fourth order appeared, so there are infinitely many functions

$$\begin{aligned}
&M^{1a}_{0b} = M^{1a}_{0b}(x,y^1), \qquad M^{2a}_{1b} = M^{2a}_{1b}(x,y^1), \qquad M^{3a}_{2b} = M^{3a}_{2b}(x,y^1),\\
&M^{2a}_{0b} = M^{2a}_{0b}(x,y^1,y^2), \qquad M^{3a}_{1b} = M^{3a}_{1b}(x,y^1,y^2),\\
&M^{3a}_{0b} = M^{3a}_{0b}(x,y^1,y^2,y^3),
\end{aligned} \tag{1.12}$$

which are the solutions of (1.11). The adapted basis B^* ((1.9)) depends on the choice of M.

Let us denote the adapted basis of $T(E)$ by B, where

$$B = \{\delta_{0a}, \delta_{1a}, \delta_{2a}, \delta_{3a}\} = \{\delta_{\alpha a}\}, \tag{1.13}$$

and

$$\begin{aligned}
\delta_{0a} &= \partial_{0a} - N^{1b}_{0a}\partial_{1b} - N^{2b}_{0a}\partial_{2b} - N^{3b}_{0a}\partial_{3b},\\
\delta_{1a} &= \partial_{1a} - N^{2b}_{1a}\partial_{2b} - N^{3b}_{1a}\partial_{3b},\\
\delta_{2a} &= \partial_{2a} - N^{3b}_{2a}\partial_{3b},\\
\delta_{3a} &= \partial_{3a}.
\end{aligned} \tag{1.14}$$

THEOREM 1.3. *The necessary and sufficient conditions that* B ((1.13)) *be dual to* B^* ((1.9)) (*when* $\bar{B}$ ((1.5)) *is dual to* $\bar{B}^*$ ((1.7))), *i.e.*

$$\left\langle \delta_{\alpha a}, \delta y^{\beta b} \right\rangle = \delta^\beta_\alpha \delta^b_a,$$

are the following relations:

$$
\begin{aligned}
N^{1b}_{0a} &= M^{1b}_{0a}\\
N^{2b}_{0a} &= M^{2b}_{0a} - M^{2b}_{1c}N^{1c}_{0a}\\
N^{3b}_{0a} &= M^{3b}_{0a} - M^{3b}_{1c}N^{1c}_{0a} - M^{3b}_{2c}N^{2c}_{0a}\\
N^{2b}_{1a} &= M^{2b}_{1a}\\
N^{3b}_{1a} &= M^{3b}_{1a} - M^{3b}_{2c}N^{2c}_{1a}\\
N^{3b}_{2a} &= M^{3b}_{2a}.
\end{aligned}
\tag{1.15}
$$

THEOREM 1.4. *The necessary and sufficient conditions that* $\delta_{\alpha a}$ *with respect to* (1.1) *are transformed as d-tensors, i.e.*

$$
\delta_{\alpha a'} = \frac{\partial x^a}{\partial x^{a'}}\delta_{\alpha a}, \qquad \alpha = 0, 1, 2, 3,
\tag{1.16}
$$

are the following formulae:

$$
\begin{aligned}
N^{1b'}_{0a'}\partial_{0a}y^{0a'} &= N^{1c}_{0a}\partial_{1c}y^{1b'} - \partial_{0a}y^{1b'}\\
N^{2b'}_{0a'}\partial_{0a}y^{0a'} &= N^{2c}_{0a}\partial_{2c}y^{2b'} + N^{1c}_{0a}\partial_{1c}y^{2b'} - \partial_{0a}y^{2b'}\\
N^{3b'}_{0a'}\partial_{0a}y^{0a'} &= N^{3c}_{0a}\partial_{3c}y^{3b'} + N^{2c}_{0a}\partial_{2c}y^{3b'} + N^{1c}_{0a}\partial_{1c}y^{3b'} - \partial_{0a}y^{3b'}\\
N^{2b'}_{1a'}\partial_{1a}y^{1a'} &= N^{2c}_{1a}\partial_{2c}y^{2b'} - \partial_{1a}y^{2b'}\\
N^{3b'}_{1a'}\partial_{1a}y^{1a'} &= N^{3c}_{1a}\partial_{3c}y^{3b'} + N^{2c}_{1a}\partial_{2c}y^{3b'} - \partial_{1a}y^{3b'}\\
N^{3b'}_{2a'}\partial_{2a}y^{2a'} &= N^{3b}_{2a}\partial_{3b}y^{3b'} - \partial_{2a}y^{3b'}.
\end{aligned}
\tag{1.17}
$$

From (1.13) and (1.14) it follows

$$
\begin{aligned}
\partial_{3a} &= \delta_{3a}\\
\partial_{2a} &= \delta_{2a} + M^{3b}_{2a}\delta_{3b}\\
\partial_{1a} &= \delta_{1a} + M^{2b}_{1a}\delta_{2b} + M^{3b}_{1a}\delta_{3b}\\
\partial_{0a} &= \delta_{0a} + M^{1b}_{0a}\delta_{1b} + M^{2b}_{0a}\delta_{2b} + M^{3b}_{0a}\delta_{3b}.
\end{aligned}
\tag{1.18}
$$

THEOREM 1.5. *With respect to the coordinate transformation* (1.3) *the Liouville vector fields have the form*

$$
\begin{aligned}
\Gamma_{(1)} &= y^{1a}\partial_{3a},\\
\Gamma_{(2)} &= y^{1a}\partial_{2a} + 3y^{2a}\partial_{3a},\\
\Gamma_{(3)} &= y^{1a}\partial_{1a} + 2y^{2a}\partial_{2a} + 3y^{3a}\partial_{3a}.
\end{aligned}
\tag{1.19}
$$

In the geometry where Miron's transformation group is used ([15], [16], [17]) $\Gamma_{(1)}$ and $\Gamma_{(3)}$ are the same as here, but $\Gamma_{(2)} = y^{1a}\partial_{2a} + 2y^{2a}\partial_{3a}$.

The vector fields $\Gamma_{(\alpha)}$, $\alpha = 1,2,3$ given by (1.19) in the basis B has the form

$$
\begin{aligned}
&\Gamma_{(1)} = z_1^{3a}\delta_{3a},\\
&\Gamma_{(2)} = z_2^{2a}\delta_{2a} + z_2^{3a}\delta_{3a},\\
&\Gamma_{(3)} = z_3^{1a}\delta_{1a} + z_3^{2a}\delta_{2a} + z_3^{3a}\delta_{3a}.
\end{aligned} \tag{1.20}
$$

The relation between the components is given by:

$$
\begin{aligned}
&z_1^{3a} = y^{1a}, \quad z_2^{2a} = y^{1a}, \quad z_2^{3a} = 3y^{2a} + y^{1b}M_{2b}^{3a}\\
&z_3^{1a} = y^{1a}, \quad z_3^{2a} = 2y^{2a} + y^{1b}M_{1b}^{2a}\\
&z_3^{3a} = 3y^{3a} + 2y^{2b}M_{2b}^{3a} + y^{1b}M_{1b}^{3a}.
\end{aligned} \tag{1.21}
$$

The proof is obtained by (1.18). All z from (1.21) with respect to (1.3) are transformed as tensors of type $(1,0)$.

2. The adapted basis which is comprehensive with J

It is obvious that the introduced transformation group given by (1.1) instead of that introduced by R. Miron [16], [17] results a new adapted basis B ((1.13)) and B^* ((1.9)). These bases are dual to each other, their elements transform as d-vector (or covector) fields, but they are not convenient for the presentation of the almost tangent structure J, for which $J^4 = 0$ and $JT_H = T_{V_1}$, $JT_{V_1} = T_{V_2}$, $JT_{V_2} = T_{V_3}$, $JT_{V_3} = 0$. To obtain such a basis we take:

$$
\begin{aligned}
&\delta y^{0a} = dy^{0a} = dx^a\\
&\delta y^{1a} = dy^{1a} + M_{0b}^{1a}dy^{0b}\\
&\delta y^{2a} = \frac{1}{2}dy^{2a} + M_{1b}^{2a}dy^{1b} + M_{0b}^{2a}dy^{0b}\\
&\delta y^{3a} = \frac{1}{6}dy^{3a} + \frac{1}{2}M_{2b}^{3a}dy^{2b} + M_{1b}^{3a}dy^{1a} + M_{0b}^{3a}dy^{0b}.
\end{aligned} \tag{2.1}
$$

THEOREM 2.1. *The necessary and sufficient conditions that* $\delta y^{\alpha a}$ $(\alpha = 0,1,2,3)$ *given by* (2.1) *are transformed as d-tensor fields, are the following*

equations:

$$
\begin{aligned}
&M^{1a}_{0b}\partial_{0a}y^{0a'} = M^{1a'}_{0b'}\partial_{0b}y^{0b'} + \partial_{0b}y^{1a'}\\
&M^{2a}_{1b}\partial_{2a}y^{2a'} = M^{2a'}_{1b'}\partial_{1b}y^{1b'} + \frac{1}{2}\partial_{1b}y^{2a'}\\
&M^{3a}_{2b}\partial_{3a}y^{3a'} = M^{3a'}_{2b'}\partial_{2b}y^{2b'} + \frac{1}{3}\partial_{2b}y^{3a'}\\
&M^{2a}_{0b}\partial_{2a}y^{2a'} = M^{2a'}_{0b'}\partial_{0b}y^{0b'} + M^{1a'}_{1b'}\partial_{0b}y^{1b'} + \frac{1}{2}\partial_{0b}y^{2a'}\\
&M^{3a}_{1b}\partial_{3a}y^{3a'} = M^{3a'}_{1b'}\partial_{1b}y^{1b'} + \frac{1}{2}M^{3a'}_{2b'}\partial_{1b}y^{2b'} + \frac{1}{6}\partial_{1b}y^{3a'}\\
&M^{3a}_{0b}\partial_{3a}y^{3a'} = M^{3a'}_{0b'}\partial_{0b}y^{0b'} + M^{3a'}_{1b'}\partial_{0b}y^{1b'} + \frac{1}{2}M^{3a'}_{2b'}\partial_{0b}y^{2b'} + \frac{1}{6}\partial_{0b}y^{3a'}.
\end{aligned}
\tag{2.2}
$$

From (1.4) it follows that M^{1a}_{0b}, M^{2a}_{1b} and M^{3a}_{2b} have the same law of transformation, also M^{2a}_{0b} and M^{3a}_{1b} transform in the same way. This fact allows us to take

$$
M^{1a}_{0b} = M^{2a}_{1b} = M^{3a}_{2b}, \qquad M^{2a}_{0b} = M^{3a}_{1b}. \tag{2.3}
$$

If (2.3) is valid the adapted basis

$$
B'^{*} = \{\delta' y^{0a}, \delta' y^{1a}, \delta' y^{2a}, \delta' y^{3a}\} \tag{2.4}
$$

is given by

$$
\begin{aligned}
&\delta' y^{0a} = dx^a = dy^{0a}\\
&\delta' y^{1a} = dy^{1a} + M^{1a}_{0b}dy^{0b}\\
&\delta' y^{2a} = \frac{1}{2}dy^{2a} + M^{1a}_{0b}dy^{1b} + M^{2a}_{0b}dy^{0b}\\
&\delta' y^{3a} = \frac{1}{6}dy^{3a} + \frac{1}{2}M^{1a}_{0b}dy^{2b} + M^{2a}_{0b}dy^{1b} + M^{3a}_{0b}dy^{0b}.
\end{aligned}
\tag{2.5}
$$

THEOREM 2.2. *The structure J defined on $T^*(E)$ by*

$$
\begin{aligned}
&J(dy^{3a}) = 3dy^{2a}, \qquad J(dy^{2a}) = 2dy^{1a},\\
&J(dy^{1a}) = dy^{0a}, \qquad J(dy^{0a}) = 0
\end{aligned}
\tag{2.6}
$$

is a tensor field of type (1.1), *and satisfies the relation* $J^4 = 0$.

PROOF. From (2.6) and (2.5) it follows

$$
\begin{aligned}
&J(\delta' y^{3a}) = \delta' y^{2a}, \qquad J(\delta' y^{2a}) = \delta' y^{1a},\\
&J(\delta' y^{1a}) = \delta' y^{0a}, \qquad J(\delta' y^{0a}) = 0.
\end{aligned}
\tag{2.7}
$$

Let us denote by

$$B' = \{\delta'_{0a}, \delta'_{1a}, \delta'_{2a}, \delta'_{3a}\}$$

the adapted basis of $T(E)$ given by

$$\begin{aligned}
\delta'_{0a} &= \partial_{0a} - N^{1b}_{0a}\partial_{1b} - 2N^{2b}_{0a}\partial_{2b} - 6N^{3b}_{0a}\partial_{3b}\\
\delta'_{1a} &= \partial_{1a} - 2N^{1b}_{0a}\partial_{2b} - 6N^{2b}_{0a}\partial_{3b}\\
\delta'_{2a} &= 2\partial_{2a} - 6N^{1b}_{0a}\partial_{3b}\\
\delta'_{3a} &= 6\partial_{3a}.
\end{aligned} \tag{2.8}$$

THEOREM 2.3. *The adapted basis B' and B'^* are dual to each other if*

$$\begin{aligned}
N^{1b}_{0a} &= M^{1b}_{0a}\\
N^{2b}_{0a} &= M^{2b}_{0a} - M^{1b}_{0c}N^{1c}_{0a}\\
N^{3b}_{0a} &= M^{3b}_{0a} - M^{2b}_{0c}N^{1c}_{0a} - M^{1b}_{0c}N^{2c}_{0a}.
\end{aligned} \tag{2.9}$$

THEOREM 2.4. *The elements of basis B' given by* (2.9) *are transformed as d-tensor fields if*

$$\begin{aligned}
N^{1b'}_{0a'}\partial_{0a}y^{0a'} &= N^{1c}_{0a}\partial_{1c}y^{1b'} + \partial_{0a}y^{1b'}\\
N^{2b'}_{0a'}\partial_{0a}y^{0a'} &= N^{2c}_{0a}\partial_{2c}y^{2b'} + \frac{1}{2}N^{1c}_{0a}\partial_{1c}y^{2b'} - \frac{1}{2}\partial_{0a}y^{2b}\\
N^{3b'}_{0a'}\partial_{0a}y^{0a'} &= N^{3c}_{0a}\partial_{3c}y^{3b'} + \frac{1}{3}N^{2c}_{0a}\partial_{2c}y^{3b'} + \frac{1}{6}N^{1c}_{0a}\partial_{1c}y^{3b'} - \frac{1}{6}\partial_{0a}y^{3b'}.
\end{aligned} \tag{2.10}$$

From (2.6), (2.7) and (2.8) it follows:

$$J(\partial_{0a}) = \partial_{1a}, \quad J(\partial_{1a}) = 2\partial_{2a}, \quad J(\partial_{2a}) = 3\partial_{3a}, \quad J(\partial_{3a}) = 0, \tag{2.11}$$

$$J(\delta'_{0a}) = \delta'_{1a}, \quad J(\delta'_{1a}) = \delta'_{2a}, \quad J(\delta'_{2a}) = \delta'_{3a}, \quad J(\delta'_{3a}) = 0. \tag{2.12}$$

The tensor J in the basis $\bar{B}$ and $\bar{B}^*$ has the form:

$$J = dy^{0b}J^{1a}_{0b} \otimes \partial_{1a} + dy^{1b}J^{2a}_{1b} \otimes \partial_{2a} + dy^{2b}J^{3a}_{2b} \otimes \partial_{3a}, \tag{2.13}$$

where

$$J^{1a}_{0b} = \delta^a_b, \quad J^{2a}_{1b} = 2\delta^a_b, \quad J^{3a}_{2b} = 3\delta^a_b,$$

or in the matrix form

$$J = [dy^{0b}dy^{1b}dy^{2b}dy^{3b}] \begin{bmatrix} 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 3 & 0 \end{bmatrix} \otimes \begin{bmatrix} \partial_{0b}\\ \partial_{1b}\\ \partial_{2b}\\ \partial_{3b} \end{bmatrix}.$$

The tensor J in the basis B' and B'^* determined by (2.8) and (2.4) has the form

$$J = \delta' y^{0a} \otimes \delta'_{1a} + \delta' y^{1a} \otimes \delta'_{2a} + \delta' y^{2a} \otimes \delta'_{3a}. \tag{2.14}$$

It is easy to see that from (2.14) follow (2.6) and (2.12), and from (2.15) follow (2.7) and (2.13).

3. Variational problem of the Lagrangian of order three

DEFINITION 3.1. A differentiable Lagrangian of order three on a C^∞-manifold E is a function $L: E \to R$ differentiable on $\tilde{E}(\operatorname{rank}[y^{1a}] = 1)$ and continuous in the points of E, where y^{1a} are equal to zero.

From this definition it follows that

$$g_{ab}(x, y^1, \ldots, y^3) = \frac{1}{2}\partial_{3a}\partial_{3b}L^2 \tag{3.1}$$

is a symmetric d-tensor field of type $(0,2)$ on $\tilde{E}$. We say that the Lagrangian L is regular if $\operatorname{rank}[g_{ab}] = n$ on $\tilde{E}$.

DEFINITION 3.2. We call a Lagrange space of order three a pair $L^{(3)n} = (E, L)$, where L is a regular C^∞-Lagrangian of order 3 and the d-tensor field g_{ab} from (3.1) has a constant signature on $\tilde{E}$.

If the metric tensor G on $T(E)$ is defined by:

$$G = g_{ab}\delta y^{0a} \otimes \delta y^{0b} + g_{ab}\delta y^{1a} \otimes \delta y^{1b} + g_{ab}\delta y^{2a} \otimes \delta y^{2b} + g_{ab}\delta y^{3a} \otimes \delta y^{3b},$$

then T_H, T_{V_1}, T_{V_2}, T_{V_3} with respect to G are mutually orthogonal to each other.

Let $L: E \to R$ be a differentiable Lagrangian of order three and $c: t \in [0,1] \to (x^a(t))\partial_a \in M$ a smooth parametrized curve, such that $\operatorname{Im} c \subset U$, U being the domain of a local chart of the differentiable manifold M.

The extension c^* (of c) to $\tilde{E}$ is given by

$$c^*: t \in [0,1] \to x^a(t)\partial_a + d_t^1 x^a(t)\partial_{1a} + d_t^2 x^a(t)\partial_{2a} + d_t^3 x^a(t)\partial_{3a},$$

where the notations:

$$d_t^\alpha = \frac{d^\alpha}{dt^\alpha}, \quad y^{\alpha a} = d_t^\alpha x^a, \quad \alpha = 1, 2, 3$$

are used.

The integral of the action of the Lagrangian L along the curve c^* is given by

$$I_{(c^*)} = \int_0^1 L(x, d_t^1 x, d_t^2 x, d_t^3 x)dt = \int_0^1 L(x, y^1, y^2, y^3)dt. \tag{3.2}$$

We consider the curves c_ε^* on $\tilde{E}$:

$$c_\varepsilon^* : t \in [0,1] \to$$
$$\to (x^a(t) + \varepsilon v^a(t))\partial_{0a} + (y^{1a}(t) + \varepsilon v^{1a}(t))\partial_{1a} + (y^{2a}(t) + \varepsilon v^{2a}(t))\partial_{2a}$$
$$+ (y^{3a}(t) + \varepsilon v^{3a}(t))\partial_{3a},$$

where

$$v^a(t) = v^a(x^1(t), \dots, x^n(t)),$$
$$y^{\alpha a} = d_t^\alpha x^a, \quad v^{\alpha a} = d_t^\alpha v^a, \quad \alpha = 1,2,3,$$

$v^a(t)$ are C^∞-functions along c_ε^* and ε is a real number sufficiently small in absolute value, such that

$$x^a + \varepsilon v^a \in U \subset M.$$

We assume that

$$v^a(0) = v^a(1) = 0, \quad d_t^\alpha v^a(0) = d_t^\alpha v^a(1) = 0, \quad \alpha = 1,2. \tag{3.3}$$

The integral of action of Lagrangian L along c_ε^* is

$$I_{(c_\varepsilon^*)} = \int_0^1 L(x + \varepsilon v, d_t^1(x + \varepsilon v), d_t^2(x + \varepsilon v), d_t^3(x + \varepsilon v))dt. \tag{3.4}$$

A necessary condition that $I_{(c^*)}$ be an extremal value for $I_{(c_\varepsilon^*)}$ is

$$\left.\frac{dI_{(c_\varepsilon^*)}}{d\varepsilon}\right|_{\varepsilon=0} = 0. \tag{3.5}$$

Using the regularity, the operators $\frac{d}{d\varepsilon}$ and $\int$ can be permutated, i.e. we get

$$\begin{aligned}\frac{dI_{(c_\varepsilon^*)}}{d\varepsilon} &= \int_0^1 \frac{d}{d\varepsilon} L(x + \varepsilon v, d_t^1(x + \varepsilon v), d_t^2(x + \varepsilon v), d_t^3(x + \varepsilon v))dt \\ &= \int_0^1 [(\partial_{0a}L)v^a + (\partial_{1a}L)d_t^1 v^a + (\partial_{2a}L)d_t^2 v^a + (\partial_{3a}L)d_t^3 v^a]dt.\end{aligned} \tag{3.6}$$

As

$$(\partial_{1a}L)d_t^1 v^a = d_t^1((\partial_{1a}L)v^a) - (d_t^1\partial_{1a}L)v^a,$$
$$(\partial_{2a}L)d_t^2 v^a = d_t^1((\partial_{2a}L)d_t^1 v^a) - d_t^1((d_t^1\partial_{2a}L)v^a) + (d_t^2\partial_{2a}L)v^a$$

$$(\partial_{3a}L)d_t^3v^a = d_t^1((\partial_{3a}L)d_t^2v^a) - d_t^1((d_t^1\partial_{3a}L)d_t^1v^a) + d_t^1((d_t^2\partial_{3a}L)v^a) - (d_t^3\partial_{3a}L)v^a$$

the substitution of the above equations into (3.6) yields

$$\begin{aligned}\frac{dI_{(c_\varepsilon^*)}}{d\varepsilon} = \int_0^1 \Big\{ &(\partial_{0a}L - d_t^1\partial_{1a}L + d_t^2\partial_{2a}L - d_t^3\partial_{3a}L)v^a \\ &+ d_t^1[(\partial_{1a}L - d_t^1\partial_{2a}L + d_t^2\partial_{3a}L)v^a \\ &+ (\partial_{2a}L - d_t^1\partial_{3a}L)d_t^1v^a + \partial_{3a}Ld_t^2v^a]\Big\}dt.\end{aligned} \tag{3.7}$$

According to (3.3) the last part of (3.7) vanishes and we obtain

$$\frac{dI_{(c_\varepsilon^*)}}{d\varepsilon} = \int_0^1 (\partial_{0a}L - d_t^1\partial_{1a}L + d_t^2\partial_{2a}L - d_t^3\partial_{3a}L)v^a dt = 0.$$

As $v^a(t)$ are arbitrary functions we get

THEOREM 3.1. *In order that the integral of action $I(c^*)$ is an extremal value for the functionals $I(c_\varepsilon^*)$, it is necessary that the following Euler–Lagrange equations hold:*

$$E_a^0(L) = \partial_a L - d_t^1\partial_{1a}L + d_t^2\partial_{2a}L - d_t^3\partial_{3a}L = 0, \tag{3.8}$$

$$y^{1a} = \frac{dx^a}{dt}, \quad y^{2a} = \frac{d^2x^a}{dt^2}, \quad y^{3a} = \frac{d^3x^a}{dt^3}. \tag{3.9}$$

Using the relation (1.4) we get

$$E_a^0 = (\partial_a x^{a'})E_{a'}^0. \tag{3.10}$$

THEOREM 3.2. *Equation* (3.6) *is invariant with respect to the change of coordinates of type* (1.3) *if and only if the functions $v^a(x)$ are transformed as d-tensors, i.e. if $v^{a'} = (\partial_a x^{a'})v^a$.*

PROOF. If we introduce the notations:

$$\begin{aligned}&\delta_t^0v^a = v^a, \qquad \delta_t^1v^a = d_t^1v^a + M_{0b}^{1a}v^b, \\ &\delta_t^2v^a = d_t^2v^a + M_{1b}^{2a}d_t^1v^b + M_{0b}^{2a}v^b, \\ &\delta_t^3v^3 = d_t^3v^a + M_{2b}^{3a}d_t^2v^b + M_{1b}^{3a}d_t^1v^b + M_{0b}^{3a}v^b,\end{aligned} \tag{3.11}$$

and use (1.8), then the expression in (3.6) in the basis B has the form:

$$\begin{aligned}&[v^a\partial_{0a} + (d_t^1v^a)\partial_{1a} + (d_t^2v^a)\partial_{2a} + (d_t^3v^a)\partial_{3a}]L \\ &= [v^a\delta_{0a} + (\delta_t^1v^a)\delta_{1a} + (\delta_t^2v^a)\delta_{2a} + (\delta_t^3v^a)\delta_{3a}]L.\end{aligned} \tag{3.12}$$

If $v^a(t)$ under the coordinate transformation (1.3) transforms as d-vector field, i.e.

$$v^{a'}(t) = (\partial_a x^{a'}) v^a(t),$$

then $v^a(t)$, $d_t^1 v^a$, $d_t^2 v^a$ and $d_t^3 v^a$ transform as y^{1a}, y^{2a}, y^{3a} and $\dfrac{dy^{3a}}{dt}$, respectively. The comparison of (3.11) with (1.10) results that $\delta_t^0 v^a$, $\delta_t^1 v^a$, $\delta_t^2 v^a$, $\delta_t^3 v^a$ have the same transformation laws as $\dfrac{\delta y^{0a}}{dt}$, $\dfrac{\delta y^{1a}}{dt}$, $\dfrac{\delta y^{2a}}{dt}$ and $\dfrac{\delta y^{3a}}{dt}$, respectively, so they are d-vector fields.

4. Zermello's conditions in Osc^3M

The integral of action I_{c^*} does not depend on the parametrization of the curve c^* if

$$\int_0^1 L(x, y^1, y^2, y^3)dt = \int_0^1 L(x, y^{1'}, y^{2'}, y^{3'})ds, \tag{4.1}$$

for any change of parameter $s = s(t)$, where $s(t)$ is at least C^4-function, $s'(t) > 0$, $s(0) = 0$, $s(1) = 1$, and

$$y^{\alpha a'} = d_s^\alpha x^a = \frac{d^\alpha x^a}{ds^\alpha}, \quad \alpha = 1, 2, 3.$$

(4.1) will be satisfied if

$$L(x, y^1, y^2, y^3) = L(x, y^{1\prime}, y^{2\prime}, y^{3\prime})s', \tag{4.2}$$

where $s' = \dfrac{ds}{dt}$. We shall use the notation

$$s^{(\alpha)} = \frac{d^\alpha s}{dt^\alpha}, \quad \alpha = 1, 2, 3.$$

The equations which give the invariance of I_{c^*} from the parametrization of the curve c^* are called Zermello's conditions. By pure calculation we get:

$$\begin{aligned}
y^{1a} &= y^{1a'} s', \\
y^{2a} &= y^{2a'}(s')^2 + y^{1a'} s'', \\
y^{3a} &= y^{3a'}(s')^3 + y^{2a'} 3s's'' + y^{1a'} s''', \\
\frac{dy^{3a}}{dt} &= \frac{dy^{3a'}}{ds}(s')^4 + y^{3a'} 6s'^2 s'' + y^{2a'}(3(s'')^2 + 4s's''') + y^{1a'} s^{IV}.
\end{aligned} \tag{4.3}$$

Taking the partial derivatives of (4.2) with respect to s', s'', s''' and s^{IV} we get:

$$(\partial_{1a}L)y^{1a'} + (\partial_{2a}L)2s'y^{2a'} + (\partial_{3a}L)(3(s')^2y^{3a'} + 3s''y^{2a'}) = L(x, y^{1'}, y^{2'}, y^{3'}), \tag{4.4}$$

$$(\partial_{2a}L)y^{1a'} + (\partial_{3a}L)(3s'y^{2a'}) = 0 \tag{4.5}$$

$$(\partial_{3a}L)y^{1a'} = 0. \tag{4.6}$$

In (4.4)–(4.6) $L = L(x, y^1, y^2, y^3)$. If we multiply (4.4) with s', (4.5) with $2s''$, (4.6) with $3s'''$ and add all these equations we obtain:

$$\begin{aligned}&(\partial_{1a}L)y^{1a'}s' + 2(\partial_{2a}L)(y^{2a'}(s')^2 + y^{1a'}s'') \\ &\quad + 3(\partial_{3a}L)(y^{3a'}(s')^3 + 3y^{2a'}s's'' + y^{1a'}s''') \\ &= L(x, y^{1'}, y^{2'}, y^{3'})s'.\end{aligned}$$

The substitution of (4.3) and (4.2) into the above equations results in

$$(\partial_{1a}L)y^{1a} + 2(\partial_{2a}L)y^{2a} + 3(\partial_{3a}L)y^{3a} = L. \tag{4.7}$$

If we multiply (4.5) with s', (4.6) with $3s''$ and add all such obtained equations we get

$$(\partial_{2a}L)(y^{1a'}s') + 3(\partial_{3a}L)(y^{2a'}(s')^2 + y^{1a'}s'') = 0,$$

i.e.

$$(\partial_{2a}L)y^{1a} + 3(\partial_{3a}L)y^{2a} = 0. \tag{4.8}$$

From (4.6) it follows:

$$(\partial_{3a}L)y^{1a} = 0. \tag{4.9}$$

THEOREM 4.1. *Equations* (4.7)–(4.9) *are the Zermello's conditions in* Osc^3M.

The comparison of (4.7)–(4.9) with (1.21) yields:

THEOREM 4.2. *The Zermello's conditions in* Osc^3M *are:*

$$\Gamma_{(1)}L = 0, \quad \Gamma_{(2)}L = 0, \quad \Gamma_{(3)}L = L.$$

They are the necessary conditions for the invariance of I_{c^*} from the parametrization of the curve c^*.

REFERENCES

[1] *Lagrange and Finsler geometry. Applications to physics and biology*, Edited by P. L. Antonelli and R. Miron in cooperation with M. Anastasiei and Gh. Zet, Fundamental Theories of Physics, 76, Kluwer Academic Publishers Group, Dordrecht, 1996. *MR* **96j**: 53024

[2] Asanov, G. S., *Finsler geometry, relativity and gauge theories*, Fundamental Theories of Physics, D. Reidel Publ. Co., Dordrecht–Boston, MA, 1985. *MR* **87d**:53122

[3] Bejancu, A., Foundations of direction-dependent gauge theory, Seminarul de Mecanica, Univ. Timişoara **13** (1988), 1–60.

[4] Čomić, I., The curvature theory of strongly distinguished connection in the recurrent K-Hamilton space, *Indian J. Pure Appl. Math.* **23** (1992), 189–202. *MR* **93d**:53089

[5] Čomić, I., Curvature theory of recurrent Hamilton spaces with generalized Miron's d-connection, *An. Stiint. Univ. „Al. I. Cuza" Iasi Sect. Ia Mat.* **37** (1991), 467–476. *MR* **94i**:53015

[6] Čomić, I., Curvature theory of generalized second order gauge connections, *Publ. Math. Debrecen* **50** (1997), 97–106.

[7] Čomić, I., The curvature theory of generalized connection in Osc^2M, *Balkan J. Geom. Appl.* **1** (1996), 21–29. *MR* **98a**:53027

[8] Čomić, I. and Kawaguchi, H., The curvature theory of dual vector bundles and their subbundles, *Tensor (N.S.)* **55** (1994), 20–31. *MR* **95f**:53119

[9] Ikeda, S., On the theory of gravitational field in Finsler spaces, *Tensor (N.S.)* **50** (1991), 256–262. *MR* **93e**:53027

[10] Ikeda, S., Some generalized connection structures of the Finslerian gravitational field. II, *Tensor (N.S.)* **56** (1995), 318–324. *MR* **98a**:53034

[11] Kawaguchi, A., On the vectors of higher order and the extended affine connections, *Ann. Mat. Pura Appl.* (4) **55** (1961), 105–117. *MR* **24** #A1677

[12] Libermann, P. and Marle, C. M., *Symplectic geometry and analytical mechanics*, Mathematics and its Applications, 35, D. Reidel Publ. Co., Dordrecht – Boston, Mass., 1987. *MR* **88c**:58016

[13] Matsumoto, M., *Foundations of Finsler geometry and special Finsler spaces*, Kaiseisha Press, Shigaken, 1986. *MR* **88f**:53111

[14] Miron, R. and Anastasiei, M., *The geometry of Lagrange spaces: theory and applications*, Fundamental Theories of Physics, **59**, Kluwer Acad. Publ., Dordrecht, 1994. *MR* **95f**:53120

[15] Miron, R. and Atanasiu, Gh., Compendium sur les espaces Lagrange d'ordre supérieur, Seminarul de Mecanica 40, Universitatea din Timisoara, 1994.

[16] Miron, R. and Atanasiu, Gh., Differential geometry of the k-osculator bundle, *Rev. Roumaine Math. Pures Appl.* **41** (1996), 205–236. *MR* **97m**:53038

[17] Miron, R. and Atanasiu, Gh., Higher order Lagrange spaces, *Rev. Roumaine Math. Pures Appl.* **41** (1996), 251–262. *MR* **97i**:53024

[18] Miron, R. and Kawaguchi, T., Lagrangian geometrical theories and their applications to the physics and engineering dynamical systems, *Tensor Soc.* (to appear).

[19] Munteanu, Gh. and Atanasiu, Gh., On Miron-connections in Lagrange spaces of second order, *Tensor (N.S.)* **50** (1991), 241–247. *MR* **93k**:53023

[20] Munteanu, Gh., Metric almost tangent structures of the second order, *Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.)* **34** (**82**) (1990), 49–54. *MR* **92a**:53049

[21] Opris, D., Fibres vectoriels de Finsler et connexions associés, *The Proceedings of the National Seminar on Finsler Spaces* (Brasov, 1980), Univ. Timişoara, Timişoara, 1981. *MR* **83e**:53071

[22] Saczuk, J., On variational aspects of a generalized continuum, *Rend. Mat. Appl.* (7) **16** (1996), 315–327. *MR* **97m**:53119

[23] SARDANASHVILY, G. and ZAKHAROV, O., *Gauge gravitation theory*, World Scientific Publishing Co., Inc., River Edge, NJ, 1992. *MR* **93e**:83003

[24] TRAUTMAN, A., *Differential geometry for physicists*, Stony Brook lectures, Monographs and Textbooks in Physical Science, **2**, Bibliopolis, Naples, 1984. *MR* **86d**:53046

(Received February 27, 1997)

UNIVERSITY OF NOVI SAD
FACULTY OF TECHNICAL SCIENCES
YU–21000 NOVI SAD
YUGOSLAVIA

comirena@uns.ns.ac.yu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 375–380

EIN KONKRETES BEISPIEL ZU DEN SYMMETRISCHEN MÖBIUS-ZWANGLÄUFEN

J. TÖLKE

Dem Gedenken an J. Strommer gewidmet

Abstract

Because of the involved systems of differential equations there are some difficulties to construct examples for kinematic motions with more than one pair of centrodes. We give one for the symmetric Möbius-motions [2]. In the conformal model the centrodes are circles and the orbits in general quartics [1].

1. Standardschauplatz $\mathcal{S}$ der Möbiusgeometrie ist die Absolutquadrik Q_{41}^2 des hyperbolischen Raumes $P_1^3(\mathbb{R})$ vereinigt mit deren Außengebiet AQ_{41}^2 [3, S. 311]. Die Punkte von Q_{41}^2 heißen M-Punkte, die von AQ_{41}^2 M_A-Punkte. Bezeichnet $\langle\ ,\ \rangle$ die zugehörige symmetrische Bilinearform der Representation $V^4(\mathbb{R})$ von P_1^3, so ist die orthogonale Automorphismengruppe $\mathcal{M} \ni \varphi : V^4 \mapsto V^4$

$$\bigwedge_{x,y \in V^4} (\langle x, y\rangle = 0 \Longleftrightarrow \langle \varphi(x), \varphi(y)\rangle = 0), \qquad |\det(\varphi)| = 1$$

die *Möbiusgruppe*. Einparametrige Scharen $\varphi(t)$ $(t \in I \subset \mathbb{R})$ einer problemangepaßten Differentiationsordnung C^r heißen *Möbius-Zwangläufe* und $\varphi(t)x$ $(\langle x, x\rangle \geqq 0)$ die *Bahnkurve* von x. Sind $\varphi(t_0)x \in Q_{41}^2$ und $\dot{\varphi}(t_0)\varphi^{-1}(t_0)\{\varphi(t_0)x\}$ linear abhängig, so heißt $\varphi(t_0)x$ ein momentaner *Rastpol* $r(t_0)$ und $\varphi^{-1}(t_0)r(t_0) =: \bar{g}(t_0)$ ein momentaner *Gangpol* an der Stelle t_0.

Somit sind jene Eigenvektoren des bezüglich $\langle\ ,\ \rangle$ schiefen Endomorphismus $\psi(t_0) := \dot{\varphi}(t_0)\varphi^{-1}(t_0)$ von V^4, die M-Punkte bestimmen, die Rastpole. Die Sekulargleichung (es gilt $\det(\psi) \leqq 0$, $\operatorname{rang}(\psi)$ ist gerade für $t \in I$)

$$\det(\psi - \lambda \mathrm{id}) = \lambda^4 - \lambda^2/2 \operatorname{Spur}(\psi^2) - (-\det(\psi)) = 0,$$

in der die Vorzeichen der Koeffizienten gegenüber Parametertransformationen invariant sind, liefert eine momentane Klassifikation [5, S. 35] der Zwangläufe in 1. Differentiationsordnung. Es gibt vier Typen. Wie üblich betrachtet man nur solche Zwangläufe, die in I vom selben Typ sind. Die drei nicht parabolischen Zwanglauftypen (d.h. $\operatorname{Spur}(\psi^2(t)) \neq 0$ für $t \in I$) haben jeweils *zwei reelle* l.u. *Rastpole* $r_1(t), r_2(t)$. Nach H. Lehmann [5] lassen sie sich zu

---

1991 *Mathematics Subject Classification.* Primary 51B10; Secondary 53A17.

Key words and phrases. Möbius plane, rigid motion, symmetric Möbius motion, quartics.

einer lokalen *kanonischen* Basis $[r_1(t), r_2(t), q_1(t), q_2(t)] = V^4$ ergänzen. Diese ist durch die Eigenschaften $(i \neq j,\ i,j = 1,2)$

$$\psi(r_1) = \varrho r_1, \quad \psi(r_2) = -\varrho r_2, \quad \psi(q_1) = \sigma q_2, \quad \psi(q_2) = -\sigma q_1, \tag{1}$$

$$\langle r_i, r_i\rangle = \langle q_i, q_j\rangle = \langle r_i, q_j\rangle = 0, \quad \langle r_i, r_j\rangle = -\langle q_i, q_i\rangle = -1 \tag{2}$$

bis auf Transformationen der Gestalt

$$\begin{aligned} r_1^* &= \lambda r_1, & r_2^* &= \lambda^{-1} r_2, \\ q_1^* &= q_1 \cos\delta + q_2 \sin\delta, & q_2^* &= -q_1 \sin\delta + q_2 \cos\delta \end{aligned} \tag{3}$$

(mit Funktionen $\lambda, \delta \in C^r(I)$, $\lambda \neq 0$) bestimmt. Wegen (2) gelten Ableitungsgleichungen der Form

$$\begin{aligned} \dot r_1 &= \beta r_1 \qquad\quad + \gamma_1 q_1 + \gamma_2 q_2 \\ \dot r_2 &= \qquad -\beta r_2 + \gamma_3 q_1 + \gamma_4 q_2 \\ \dot q_1 &= \gamma_3 r_1 + \gamma_1 r_2 \qquad + \zeta q_2 \\ \dot q_2 &= \gamma_4 r_1 + \gamma_2 r_2 - \zeta q_1. \end{aligned} \tag{4}$$

Da die Funktionen γ_i nach (3) das Transformationsverhalten

$$\begin{aligned} \gamma_1^* &= \lambda(\gamma_1 \cos\delta + \gamma_2 \sin\delta), & \gamma_2^* &= -\lambda(\gamma_1 \sin\delta - \gamma_2 \cos\delta), \\ \gamma_3^* &= (\gamma_3 \cos\delta + \gamma_4 \sin\delta)/\lambda, & \gamma_4^* &= -(\gamma_3 \sin\delta - \gamma_4 \cos\delta)/\lambda \end{aligned}$$

besitzen, können wir o.B.d.A.

$$\gamma_2 = 0 \quad \text{für } t \in I \tag{5}$$

annehmen. Setzen wir

$$\bar g_i := \varphi^{-1} r_i, \quad \bar q_i := \varphi^{-1} q_i, \quad i = 1,2, \tag{6}$$

so gilt sinngemäß wieder (2) und damit Ableitungsgleichungen der Gestalt (4). Nehmen wir hierin die Ersetzungen $r_i \mapsto \bar g_i$, $q_i \mapsto \bar q_i$ vor und bezeichnen die Ableitungskoeffizienten mit demselben, aber gequerten Funktionszeichen, so gilt wegen (1)

$$\bar\beta = \beta - \varrho, \quad \bar\zeta = \zeta - \sigma, \quad \bar\gamma_i = \gamma_i, \qquad i = 1, \ldots, 4. \tag{7}$$

Machen wir für das Folgende die *Regularitätsforderung* $(\gamma_1^2 + \gamma_2^2)(\gamma_3^2 + \gamma_4^2) \neq 0$ für $t \in I$, so sind r_i bzw. $\bar g_i$ auf Q_{41}^2 Kurven, die *Rastpolbahn* bzw. *Gangpolbahn* $(i = 1,2)$. Die Kurven $r_i(t)$ und $\varphi(t_0)\bar g_i(t)$ berühren sich für jedes $t_0 \in I$ im Punkt $r_i(t_0) = \varphi(t_0)\bar g_i(t_0)$.

2. Mit W. Degen und S. Hartmann [2] heißt ein Möbius-Zwanglauf $\Sigma(t)$ *symmetrisch*, wenn ein die beiden Rastpole $r_i(t)$ festlassender Möbius-Zwanglauf $\bar{\varphi}(t)$ derart existiert, daß der Punkt $\bar{\varphi}x$ für jeden bezüglich Σ gangfesten M-Punkt x in Bezug auf Σ rastfest ist. Dies entspricht der *Queteletschen* Eigenschaft der euklidischen symmetrischen Rollungen [7]. Die Durchführung liefert für Σ mit (5) die kennzeichnenden Bedingungen

$$\text{(8)} \qquad \text{(a) } \varrho = 0, \quad \text{(b) } \gamma_4 = 0, \quad \text{(c) } \sigma - 2\zeta = 0$$

und für $\bar{\varphi}$ $(i = 1, 2)$

$$\text{(9)} \qquad \bar{\varphi}(r_i) = r_i, \quad \bar{\varphi}(q_1) = q_1, \quad \bar{\varphi}(q_2) = -q_2,$$

sodaß $\bar{\varphi}$ eine *Projektivspiegelung* an der Ebene des gemeinsamen *Tangentialkegelschnittes* der Polbahnen ist.

Setzt man Σ von der Differentiationsordnung C^ω voraus, so folgt [2]

$$\text{(10)} \qquad \bar{g}_i(t) = \Sigma^{-1}(t_0)\bar{\varphi}(t_0)r_i(t), \quad i = 1, 2.$$

3. Aus der Literatur [3, 6] ist uns kein Beispiel eines symmetrischen Möbius-Zwanglaufs bekannt; wir wollen ein solches für jenen Fall angeben, in dem die Polbahnen auf Q_{41}^2 Kegelschnitte sind. Sei

$$Q_{41}^2: \qquad x_1^2 + x_2^2 + x_3^2 - x_4^2 = 0$$

und

$$\text{(11)} \qquad \varphi(t) := \begin{matrix} 1 - 3s^2 & 3sc_1 & -\sqrt{6}sc_2 & 3sc_2 \\ 3sc & -2 - 3cc_1 & \sqrt{6}(1 + cc_2) & -3(1 + cc_2) \\ -\sqrt{6}s & 2\sqrt{6}(1 - c) & -5 + 6c & 3\sqrt{6}(1 - c) \\ -3s & 6(1 - c) & -3\sqrt{6}(1 - c) & 10 - 9c \end{matrix}$$

mit

$$s := \sin t, \quad c := \cos t, \quad c_1 := 1 - 2\cos t, \quad c_2 := 2 - 3\cos t.$$

Man überzeugt sich, daß $\varphi(t)$ ein Möbius-Zwanglauf mit $\det(\varphi(t)) = 1$ ist. Genau die M-Punkte

$$\text{(12)} \qquad \bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)^T, \quad x = \varphi\bar{x}, \quad 3x_3 - \sqrt{6}x_4 = 3\bar{x}_3 - \sqrt{6}\bar{x}_4 = 0$$

haben Kegelschnitte als Bahnkurven. Wir berechnen

$$\varphi^{-1}(t) = \begin{matrix} -2 + 3c^2 & 3sc & -\sqrt{6}s & 3s \\ 3sc_1 & -2 - 3cc_1 & 2\sqrt{6}(1 - c) & -6(1 - c) \\ -\sqrt{6}sc_2 & \sqrt{6}(1 + cc_2) & -5 + 6c & 3\sqrt{6}(1 - c) \\ -3sc_2 & 3(1 + cc_2) & -3\sqrt{6}(1 - c) & 10 - 9c. \end{matrix}$$

Damit bestimmt sich $\psi(t) := \dot{\varphi}(t)\varphi^{-1}(t)$

$$\psi(t) = \begin{matrix} 0 & -3 & \sqrt{6}c & -3c \\ 3 & 0 & \sqrt{6}s & -3s \\ -\sqrt{6}c & -\sqrt{6}s & 0 & 0 \\ -3c & -3s & 0 & 0. \end{matrix}$$

Also gilt $\det(\psi) = 0$, $\operatorname{Spur}(\psi^2) = -12$, $\varrho = 0$, $\sigma = -\sqrt{6}$ und damit für die Rastpolkurven $r_i(t)$

$$r_1 = (s/2, -c/2, 0, 1/2)^T, \quad r_2 = (s/2, -c/2, \sqrt{6}, 5/2)^T$$

und für die auf der reziproken Polare gelegen Punkte q_1, q_2

$$q_1 = (c, s, 0, 0)^T, \qquad q_2 = 1/\sqrt{6}(3s, -3c, \sqrt{6}, 3)^T.$$

Also folgt $\beta = \gamma_2 = \gamma_4 = 0$, $2\zeta - \sigma = 0$, $\gamma_1 = \gamma_3 = 1/2$, sodaß $\varphi(t)$ nach (8) *symmetrisch* ist.

Betrachten wir den bahnerzeugenden M-Punkt $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)^T$. Die Projektivspiegelung am Tangentialkegelschnitt $q_2(t = 0)$ liefert den M-Bildpunkt

$$(13) \quad x^* = (\bar{x}_1, -2\bar{x}_2 + \sqrt{6}\bar{x}_3 - 3\bar{x}_4, \sqrt{6}\bar{x}_2 - \bar{x}_3 + \sqrt{6}\bar{x}_4, 3\bar{x}_2 - \sqrt{6}\bar{x}_3 + 4\bar{x}_4)^T.$$

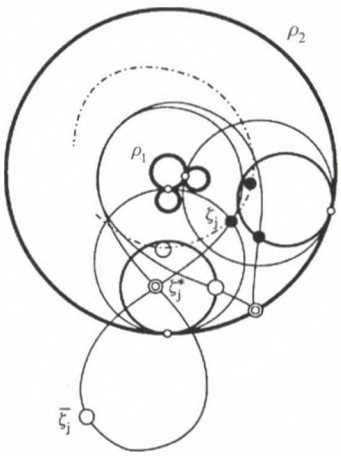
Damit folgt

$$(14) \qquad x = \varphi\bar{x} = x^* + \frac{\sqrt{2}}{\sqrt{3}} \left\{ \left(-\sqrt{6}\bar{x}_3 + 3\bar{x}_4 \right) c_2 - 3\bar{x}_1 s + 3\bar{x}_2 c_1 \right\} q_2.$$

D.h.: Die Bahnkurve $\varphi\bar{x}$ wird aus Q^2_{41} durch einen quadratischen Kegel κ mit der Kegelspitze x^* ausgeschnitten – sofern x^* nicht in der κ erzeugenden Kegelschnittsebene liegt. Andernfalls ist $\varphi\bar{x}$ der Kegelschnitt (12). Dabei gilt wegen (14): Der M_A-Punkt der Kegelschnittsebene vom $\varphi\bar{x}$ ist an *jeder* Parameterstelle zum M_A-Punkt q_2 konjugiert. Im konformen Möbius-Modell $\mathcal{M}$ sind die Bildkurven demzufolge i.a. *Quartiken* [1, S. 210f.] mit dem Bildpunkt von x^* als singulärem Punkt.

4. Wir wollen die Abbildung auf $\mathcal{M}$ auch noch analytisch verfolgen. Mit der Koordinatentransformation

$$\tilde{x} = Dx \quad \text{mit} \quad D := \begin{matrix} 0 & 0 & -v & v \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & v & v \end{matrix} \quad \text{mit} \quad v := 1/\sqrt{2}$$



Legende[1] Die beiden stärker ausgezogenen Kreise sind die Rastpolbahnen ϱ_1, ϱ_2. Die Gangpolbahnen $\gamma_1(t), \gamma_2(t)$ sind zu zwei Parameterstellen $t_0 = 0$, t_1 gezeichnet – jeweils mit den Tangentialkreisen $T(t_i)$. Die Bahnkurve $\zeta_j(t)$ von $\bar{\zeta}_j$ wird vom Punkt ζ_j^*, markiert mit einem Doppelnullkreis, durch Inversion an den Tangentialkreisen $T(t)$ erzeugt. Die mit einer Nullkreisscheibe versehenen Punkte gehören zur Bahnkurvenstelle t_i.

folgt

$$\hat{x} = D\varphi D^{-1}\hat{\bar{x}}, \tag{15}$$

womit der Anschluß an die übliche Darstellung [1, 3] gewonnen ist. Einem M_A-Punkt bzw. M-Punkt $\hat{p} = (p_0, p_1, p_2, p_3)^T$ ist in $\mathcal{M}$ der Kreis bzw. der Punkt

$$p_0(\xi^2 + \eta^2) - 2p_1\xi - 2p_2\eta + 2p_3 = 0 \quad \text{bzw.} \quad (p_1/p_0, p_2/p_0)$$

zugeordnet.

Die Polbahnen ϱ_1, ϱ_2 der momentanen *Rastpole* R_1 bzw. R_2

$$\xi_1 = \sqrt{2}\sin t, \qquad \eta_1 = -\sqrt{2}\cos t;$$
$$\xi_2 = \sqrt{2}/(5 - 2\sqrt{6})\sin t, \quad \eta_2 = -\sqrt{2}/(5 - 2\sqrt{6})\cos t$$

sind konzentrische Kreise. Das Bild des M_A-Punktes $\hat{q}_2$ ist der gemeinsame Tangentialkreis $T(t)$

$$(\xi - 3\sqrt{2}/(3-\sqrt{6})\sin t)^2 + (\eta + 3\sqrt{2}/(3-\sqrt{6})\cos t)^2 = r_T^2,\ r_T := 2\sqrt{3}/(3-\sqrt{6})$$

[1] Mein Dank gilt meinem Freund Dr. W. Schürrer für die Ausarbeitung der Figur.

der Polbahnen. Die Rastpolkreise gehen durch die Inversion $i(t)$ an $T(t)$ in die Gangpolkreise $\gamma_1(t), \gamma_2(t)$ über [2]. Nach dem zweiten Abschnitt gilt für die *Bahnkurve* des Punktes

$$\bar{\zeta} := (\bar{\xi}, \bar{\eta}): \quad \zeta(t) := (\xi(t), \eta(t)) = i^{-1}(t) i(0) \bar{\zeta}.$$

Sie durchsetzt an der Stelle t_0 den $\zeta(t_0)$ enthaltenden Kreis des zum hyperbolischen Büschel der beiden Momentanpole $R_i(t_0)$ orthogonalen Büschel senkrecht.

Der Punkt $\zeta^* := (\xi^*, \eta^*) := i(0)\bar{\zeta}$ ist der Bildpunkt von Dx^* in $\mathcal{M}$. Nach dem 3. Abschnitt besitzen die Punkte des zu *allen* Tangentialkreisen $T(t)$ orthogonalen Kreises K mit der Gleichung $\xi^2 + \eta^2 = \left(\sqrt{2}/\left(3 - \sqrt{6}\right)\right)^2$ ihn selbst als Bahnkurve. Liegt der Punkt ζ^* im Inneren des längs K geschlitzten Ringbereiches $\mathcal{R} := \bigcup_{t \in I} T(t)$, so inzidiert er mit genau zwei (sich schneidenden) Kreisen (die Bahnnormalen) $T(t_1), T(t_2)$. Also ist ζ^* ein *Knotenpunkt* der Bahnkurve $\zeta(t)$. Liegt ζ^* im Äußeren von $\mathcal{R}$ bzw. auf einem der beiden Rastpolkreise ϱ_1, ϱ_2, so sieht man analog, daß er ein *isolierter Doppelpunkt* bzw. eine *Spitze* der Bahnkurve ist.

Die Bahnkurvenverhältnisse zeigen eine gewisse Verwandtschaft mit jenen der *umgekehrten Ellipsenbewegung* [8], deren Bahnkurven sich bekanntlich auch durch symmetrische Kreisrollung erzeugen lassen.

LITERATURVERZEICHNIS

[1] Barner, M., Zur Möbius-Geometrie: Die Inversionsgeometrie ebener Kurven, *J. Reine Angew. Math.* **206** (1961), 192–220. *MR* **26** #5459

[2] Degen, W. und Hartmann, S., Zur Möbius-Kinematik, *Österreich. Akad. Wiss. Math. Naturwiss. Kl. S. B. II* **186** (1977), 179–192. *MR* **58** #2616

[3] Giering, O., *Vorlesungen über höhere Geometrie*, Friedr. Vieweg & Sohn, Braunschweig, 1982. *MR* **84c**:51002

[4] Jankovsky, Z., Möbiussche Bewegungen der Ebene mit mehrfach durchlaufenen Bahnkurven, *Apl. mat.* **30** (1985), 297–306. *MR* **87b**:53011

[5] Lehman, H., Zur Möbius-Kinematik, Dissertation, Univ. Freiburg, 1967.

[6] Pottmann, H., Kinematische Geometrie, *Geometrie und ihre Anwendungen*, edited by O. Giering and J. Hoschek, C. Hanser Verlag, München, 1994, 141–175. *MR* **95e**:51001

[7] Quetelet, L. A. J., Mémoire sur une nouvelle manière de considérer les caustiques, produites soit par reflexion, *Brux. Nouv. Mem.* **3** (1826), 89–140.

[8] Wunderlich, W., *Ebene Kinematik*, Bibliographisches Institut, Mannheim, 1970. *Zbl* **225**. 70002

(Eingegangen am 15. Mai 1997)

UNIVERSITÄT-GESAMTHOCHSCHULE SIEGEN
FACHBEREICH 6 – MATHEMATIK
HÖLDERLINSTRASSE 3
D–57076 SIEGEN
GERMANY

dekanat@math.uni-siegen.de

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 381–395

THE MEASURE OF NONCOMPACTNESS OF LINEAR OPERATORS BETWEEN SPACES OF M^{th}-ORDER DIFFERENCE SEQUENCES

E. MALKOWSKY and V. RAKOČEVIĆ

Abstract

In this paper we investigate linear operators between certain sequence spaces X and Y. Among other things, if X is any BK space and Y is a sequence space of bounded or convergent m^{th}-order differences, then we find necessary and sufficient conditions for infinite matrices A to map X into Y. Further the Hausdorff measure of noncompactness is applied to give necessary and sufficient conditions for A to be a compact operator.

1. Introduction and well-known results

We shall write ω for the set of all complex sequences $x = (x_k)_{k=0}^{\infty}$ and ϕ, l_∞, c and c_0 for the sets of all finite, bounded, convergent sequences and sequences convergent to naught, respectively, and finally, for $1 \leqq p < \infty$,

$$l_p = \left\{ x \in \omega : \sum_{k=0}^{\infty} |x_k|^p < \infty \right\}.$$

By e and $e^{(n)}$ $(n = 0, 1, \dots)$, we denote the sequences such that $e_k = 1$ for $k = 0, 1, \dots$, and $e_n^{(n)} = 1$ and $e_k^{(n)} = 0$ for $k \neq n$.

A *BK space* is a Banach sequence space with continuous coordinates.

A sequence $(b_n)_{n=0}^{\infty}$ in a linear metric space X is called a *(Schauder-) basis* if for each $x \in X$ there exists a unique sequence $(\lambda_n)_{n=0}^{\infty}$ of scalars such that $x = \sum_{n=0}^{\infty} \lambda_n b_n$.

A BK space $X \supset \phi$ is said to have AK if every sequence $x = (x_k)_{k=0}^{\infty} \in X$ has a unique representation $x = \sum_{n=0}^{\infty} x_n e^{(n)}$.

1991 *Mathematics Subject Classification.* Primary 40H05, 46A45; Secondary 47B07.

Key words and phrases. BK spaces, bases, matrix transformations, measure of noncompactness.

This joint research work was completed while the first author visited the University of Niš, Yugoslavia. He expresses his sincere gratitude to DAAD (German Academic Exchange Service), and the University of Niš for their financial support.

The work of the second author is supported by the Science Fund of Serbia, Grant number 04M03, through Matematički Institut.

Let $A=(a_{nk})_{n,k=0}^{\infty}$ be an infinite matrix of complex numbers and $x\in\omega$. Then we shall write

$$A_n(x)=\sum_{k=0}^{\infty}a_{nk}x_k,\quad (n=0,1,\dots)\quad\text{and}\quad A(x)=(A_n(x))_{n=0}^{\infty}.$$

For any subset X of ω, the set

$$X_A=\{x\in\omega: A(x)\in X\}$$

is called the *matrix domain of A in X*. For instance, if E is the matrix defined by $e_{nk}=1$ $(0\leqq k\leqq n)$ and $e_{nk}=0$ $(k>n)$ for all $n=0,1,\dots$, then $cs=c_E$ and $bs=(l_\infty)_E$ are the sets of convergent and bounded series.

2. Spaces of sequences of m^{th}-order differences and their β-duals

Let m be a positive integer throughout. We define the operators $\Delta^{(m)},\sum^{(m)}:\omega:\to\omega$ by

$$(\Delta^{(1)}x)_k=\Delta^{(1)}x_k=x_k-x_{k-1},\ (\textstyle\sum^{(1)}x)_k=\sum^{(1)}x_k=\sum_{j=0}^{k}x_j\quad(k=0,1,\dots),$$

$$\Delta^{(m)}=\Delta^{(1)}\circ\Delta^{(m-1)},\qquad \textstyle\sum^{(m)}=\sum^{(1)}\circ\sum^{(m-1)}\quad(m\geqq 2).$$

In the case where $m=1$, we shall write $\Delta=\Delta^{(1)}$ and $\sum=\sum^{(1)}$, for short.

For any subset X of ω, we define the set

$$X(\Delta^{(m)})=X_{\Delta^{(m)}}=\left\{x\in\omega:\Delta^{(m)}x\in X\right\}.$$

Here we shall be interested in the case where $X\in\{l_\infty,c,c_0\}$. The following results are well known (cf. [3]); they hold for all $m\geqq 1$ and $k=0,1,\dots$:

$$\left(\Delta^{(m)}x\right)_k=\sum_{j=0}^{m}(-1)^j\binom{m}{j}x_{k-j}=\sum_{j=\max\{0,k-m\}}^{k}(-1)^{k-j}\binom{m}{k-j}x_j,$$

$$\left(\textstyle\sum^{(m)}x\right)_k=\sum_{j=0}^{k}\binom{m+k-j-1}{k-j}x_j.$$

PROPOSITION 2.1 ([5, Proposition 1 and Theorem 1]). *The sets $l_\infty(\Delta^{(m)})$, $c(\Delta^{(m)})$ and $c_0(\Delta^{(m)})$ are BK spaces with respect to the norm $\|\cdot\|_{\Delta^{(m)}}$ defined by*

$$\|x\|_{\Delta^{(m)}}=\sup_k\left|\left(\Delta^{(m)}x\right)_k\right|=\sup_k\left|\sum_{j=0}^{m}(-1)^j\binom{m}{j}x_{k-j}\right|.$$

Further, $c_0(\Delta^{(m)})$ and $c(\Delta^{(m)})$ are closed subspaces of $l_\infty(\Delta^{(m)})$.

We define the sequences $b^k(m)$ by

$$b_n^{(-1)}(m) = \binom{m+n}{n} \quad (n = 0, 1, \dots),$$

and

$$b_n^{(k)}(m) = \begin{cases} 0 & (n \leqq k-1) \\ \binom{m+n-k-1}{n-k} & (n \geqq k) \end{cases} \qquad \text{for } k \geqq 1.$$

Then the sequence $\left(b^{(k)}(m)\right)_{k=0}^{\infty}$ is a basis of $c_0(\Delta^{(m)})$; every sequence $x = (x_k)_{k=0}^{\infty} \in c_0(\Delta^{(m)})$ has a unique representation

$$x = \sum_{k=0}^{\infty} \lambda_k(m) b^{(k)}(m), \quad \text{where } \lambda_k(m) = \left(\Delta^{(m)} x\right)_k \quad (k = 0, 1, \dots).$$

Then the sequence $\left(b^{(k)}(m)\right)_{k=-1}^{\infty}$ is a basis of $c(\Delta^{(m)})$; every sequence $x = (x_k)_{k=0}^{\infty} \in c(\Delta^{(m)})$ has a unique representation

$$x = l b^{(-1)}(m) + \sum_{k=0}^{\infty} (\lambda_k(m) - l) b^{(k)}(m), \qquad \text{where } l = \lim_{k \to \infty} \left(\Delta^{(m)} x\right)_k.$$

We shall use the following notations:

For any two sequences x and y, let $xy = (x_k y_k)_{k=0}^{\infty}$.

If X and Y are arbitrary subsets of ω and z any sequence, then we shall write

$$z^{-1} * X = \{x \in \omega : xz \in X\} \quad \text{and} \quad M(X, Y) = \bigcap_{x \in X} x^{-1} * Y.$$

In the special case, where $Y = cs$, the set

$$X^\beta = M(X, cs) = \left\{ a \in \omega : \sum_{k=0}^{\infty} a_k x_k \text{ converges for all } x \in X \right\}$$

is called the *β-dual of X*.

By $\mathcal{U}$, we denote the set of all sequences $u = (u_k)_{k=0}^{\infty}$ such that $u_k \neq 0$ for all $k = 0, 1, \dots$. If $u \in \mathcal{U}$, then we write

$$1/u = \left(\frac{1}{u_k}\right)_{k=0}^{\infty}.$$

Given any sequence a we define the sequence $R^{(m)}(a)$ by

$$R_k^{(1)}(a) = R_k(a) = \sum_{j=k}^{\infty} a_j \quad (k = 0, 1, \dots)$$

and

$$R^{(m)}(a) = R^{(1)}(R^{(m-1)}(a)) \quad (m \geqq 2)$$

provided the series converges. Further we write

$$X(R^{(m)}) = \left\{ x \in \omega : R^{(m)}(x) \in X \right\} \quad \text{for any } X \subset \omega.$$

The following well-known result gives the β-duals of the sets $l_\infty(\Delta^{(m)})$, $c(\Delta^{(m)})$ and $c_0(\Delta^{(m)})$.

PROPOSITION 2.2 ([5, Theorem 3]). *We write*

$$c_0^+ = \{x \in c_0 : x_k \geqq 0 \text{ for all } k\},$$

and put

$$M_\infty^\beta(m) = \left(((k^m))^{-1} * cs \right) \bigcap l_1(R^{(m)})$$
$$= \left\{ a \in \omega : \sum_{k=0}^{\infty} k^m a_k \text{ converges and } \sum_{k=0}^{\infty} |R_k^{(m)}(a)| < \infty \right\}$$

and

$$M_0^\beta(m) = \left(\bigcap_{v \in c_0^+} \left(\sum\nolimits^{(m)} v \right)^{-1} * cs \right) \bigcap l_1(R^{(m)})$$
$$= \left\{ a \in \omega : \sum_{k=0}^{\infty} a_k \sum_{j=0}^{k} \binom{m+k-j-1}{k-j} v_j \text{ converges for all } v \in c_0^+ \right\}$$
$$\bigcap \left\{ a \in \omega : \sum_{k=0}^{\infty} |R_k^{(m)}(a)| < \infty \right\}.$$

Then

$$\left(c(\Delta^{(m)}) \right)^\beta = \left(l_\infty(\Delta^{(m)}) \right)^\beta = M_\infty^\beta(m), \quad \left(c_0(\Delta^{(m)}) \right)^\beta = M_0^\beta(m),$$
$$\left(l_\infty(\Delta^{(m)}) \right)^\beta \neq \left(c_0(\Delta^{(m)}) \right)^\beta.$$

3. Matrix transformations

Let X and Y be two Banach spaces. By $B(X,Y)$, we denote the set of all continuous linear operators from X into Y, and we write

$$\|L\| = \sup\{\|L(x)\| : \|x\| = 1\}$$

for the operator norm of L. In the special case, where $Y = \mathbb{C}$, the complex numbers, we write $X^* = B(X,\mathbb{C})$ for the set of all continuous linear functionals on X, and

$$\|f\| = \sup\{|f(x)| : \|x\| = 1\} \quad (f \in X^*)$$

for the norm of the continuous linear functional f.

If X is a BK space and $a \in \omega$, then we put

$$\|a\|^* = \sup\left\{\left|\sum_{k=0}^{\infty} a_k x_k\right| : \|x\| = 1\right\}$$

provided the term on the right exists and is finite. This is the case whenever $a \in X^\beta$ (cf. [8, Theorem 7.2.9, p. 107]).

The following result is well known.

PROPOSITION 3.1 ([5, Lemma 4]). *On any of the spaces* $\left(c_0(\Delta^{(m)})\right)^\beta$, $\left(c(\Delta^{(m)})\right)^\beta$ *and* $\left(l_\infty(\Delta^{(m)})\right)^\beta$,

$$\|a\|^* = \|R^{(m)}(a)\|_1 = \sum_{k=0}^{\infty} |R_k^{(m)}(a)|.$$

If A is an infinite matrix of complex numbers, then we write A_n for the sequence in the n^{th} row of A. For any two subsets X and Y of ω, (X,Y) denotes the class of all infinite matrices that map X into Y. Thus $A \in (X,Y)$ if and only if $A_n \in X^\beta$ for all n, and $A(x) \in Y$ for all $x \in X$.

The following results are well known.

PROPOSITION 3.2 (cf. [6, Theorem 1]). *Let X and Y be BK spaces. Then $(X,Y) \subset B(X,Y)$, i.e. every $A \in (X,Y)$ defines an element $L_A \in B(X,Y)$, where*

$$L_A(x) = A(x) \quad (x \in X).$$

Further $A \in (X, l_\infty)$ if and only if

$$\|A\|^* = \sup_n \|A_n\|^* = \|L_A\| < \infty.$$

Finally, if $(b^{(k)})_{k=0}^{\infty}$ *is a basis of* X, Y *and* Y_1 *are BK spaces with* Y_1 *a closed subspace of* Y, *then* $A \in (X, Y_1)$ *if and only if* $A \in (X, Y)$ *and* $A(b^{(k)}) \in Y_1$ *for all* $k = 0, 1, \ldots$.

PROPOSITION 3.3 ([5, Theorem 4]).
(a) $A \in (l_\infty(\Delta^{(m)}), l_\infty)$ *if and only if*

$$A_n \in ((k^m))^{-1} * cs \quad \text{for all } n = 0, 1, \ldots \tag{3.1}$$

and

$$\sup_n \|A_n\|^* = \sup_n \left\| R^{(m)}(A_n) \right\|_1 < \infty. \tag{3.2}$$

Further, $(l_\infty(\Delta^{(m)}), l_\infty) = (c(\Delta^{(m)}), l_\infty)$.
(b) $A \in (c_0(\Delta^{(m)}), l_\infty)$ *if and only if condition* (3.2) *holds and*

$$A_n \in \bigcap_{v \in c_0^+} \left(\left(\sum\nolimits^{(m)} v \right)^{-1} * cs \right) \quad \text{for all } n = 0, 1, \ldots. \tag{3.3}$$

(c) $A \in (c_0(\Delta^{(m)}), c_0)$ *if and only if conditions* (3.2) *and* (3.3) *hold and*

$$\lim_{n\to\infty} \left(\sum_{j=k}^{\infty} \binom{m-1+j-k}{j-k} a_{nj} \right) = 0 \quad (k = 0, 1, \ldots). \tag{3.4}$$

(d) $A \in (c_0(\Delta^{(m)}), c)$ *if and only if conditions* (3.2) *and* (3.3) *hold and*

$$\lim_{n\to\infty} \left(\sum_{j=k}^{\infty} \binom{m-1+j-k}{j-k} a_{nj} \right) = l_k \quad (k = 0, 1, \ldots). \tag{3.5}$$

(e) $A \in (c(\Delta^{(m)}), c_0)$ *if and only if conditions* (3.2), (3.1), (3.4) *hold and*

$$\lim_{n\to\infty} \left(\sum_{j=0}^{\infty} \binom{m+j}{j} a_{nj} \right) = 0. \tag{3.6}$$

(f) $A \in (c(\Delta^{(m)}), c)$ *if and only if conditions* (3.2), (3.1), (3.5) *hold and*

$$\lim_{n\to\infty} \left(\sum_{j=0}^{\infty} \binom{m+j}{j} a_{nj} \right) = l_{-1}. \tag{3.7}$$

The following result reduces the characterization of the class (X, Y_T) to that of (X, Y) for triangles T, i.e. matrices T with $t_{nk} = 0$ $(k > n)$ and $t_{nn} \neq 0$ $(n = 0, 1, \ldots)$.

PROPOSITION 3.4. *Let T be a triangle.*

(a) *Then, for arbitrary subsets X and Y of ω, $A \in (X, Y_T)$ if and only if $B = TA \in (X, Y)$.*

(b) *Further, if X and Y are BK spaces and $A \in (X, Y_T)$, then*

$$\|L_A\| = \|L_B\|. \tag{3.8}$$

PROOF. (a) This is [4, Theorem 1].

(b) Let $A \in (X, Y_T)$. Since Y is a BK space and T a triangle, Y_T is a BK space with

$$\|y\|_{Y_T} = \|T(y)\|_Y \quad (y \in Y_T) \tag{3.9}$$

(cf. [8, Theorem 4.3.12, p. 63]). Thus A is continuous (cf. [8, Theorem 4.2.8, p. 56]), and consequently

$$\begin{aligned} \|L_A\| &= \sup\{\|L_A(x)\|_{Y_T} : \|x\| = 1\} \\ &= \sup\{\|A(x)\|_{Y_T} : \|x\| = 1\} < \infty. \end{aligned} \tag{3.10}$$

Further, since B is continuous,

$$\begin{aligned} \|L_B\| &= \sup\{\|L_B(x)\|_Y : \|x\| = 1\} \\ &= \sup\{\|B(x)\|_Y : \|x\| = 1\} < \infty. \end{aligned} \tag{3.11}$$

Let $x \in X$. Since $A_n \in X^\beta$ for all $n = 0, 1, \dots$, we have $x \in \omega_A$. Further $T_n \in \phi$ $(n = 0, 1, \dots)$, since T is a triangle. Thus

$$B(x) = (TA)(x) = T(A(x))$$

(cf. [8, Theorem 1.4.4, p. 8]), and (3.8) follows from (3.9), (3.10) and (3.11). □

As a corollary of Propositions 3.2 and 3.4, we have

COROLLARY 3.5. *Let X be a BK space.*

(a) *Then $A \in (X, l_\infty(\Delta^{(m)}))$ if and only if*

$$M(X, l_\infty(\Delta^{(m)})) = \sup_n \left\| \sum_{l=\max\{0,n-m\}}^{n} (-1)^{n-l} \binom{m}{n-l} A_l \right\|^* < \infty. \tag{3.12}$$

(b) *Further, if $(b^{(k)})_{k=0}^\infty$ is a basis of X, then $A \in (X, c_0(\Delta^{(m)}))$ if and only if condition (3.12) holds and*

$$\lim_{n\to\infty} \left(\sum_{l=\max\{0,n-m\}}^{n} (-1)^{n-l} \binom{m}{n-l} A_l(b^{(k)}) \right) = 0 \tag{3.13}$$

for each $k=0,1,\ldots$; $A\in(X,c(\Delta^{(m)}))$ *if and only if condition* (3.12) *holds and*

$$\lim_{n\to\infty}\left(\sum_{l=\max\{0,n-m\}}^{n}(-1)^{n-l}\binom{m}{n-l}A_l(b^{(k)})\right)=\alpha_k \tag{3.14}$$

for each $k=0,1,\ldots$.

REMARK 1. (a) If $X=l_p$ $(1\leqq p<\infty)$ and Y is any of the spaces $l_\infty(\Delta^{(m)})$, $c(\Delta^{(m)})$ and $c_0(\Delta^{(m)})$, then the conditions for $A\in(X,Y)$ follow from the respective ones in Corollary 3.5 by replacing the norm $\|\cdot\|^*$ in condition (3.12) by the natural norm on the β-dual of l_p, i.e. on l_q $(q=p/(p-1), 1<p<\infty; q=\infty, p=1)$ which is norm isomorphic to l_p^*. Hence we have

$$M(l_p,l_\infty(\Delta^{(m)}))=\begin{cases}\sup_n\left(\sum_{k=0}^{\infty}\left|\sum_{l=\max\{0,n-m\}}^{n}(-1)^{n-l}\binom{m}{n-l}a_{lk}\right|^q\right) & (1<p<\infty)\\ \sup_{n,k}\left|\sum_{l=\max\{0,n-m\}}^{n}(-1)^{n-l}\binom{m}{n-l}a_{lk}\right| & (p=1).\end{cases}$$

(b) Let s be a nonnegative integer.

If X is any of the spaces $l_\infty(\Delta^{(s)})$, $c(\Delta^{(s)})$ and $c_0(\Delta^{(s)})$, and Y is any of the spaces $l_\infty(\Delta^{(m)})$, $c(\Delta^{(m)})$ and $c_0(\Delta^{(m)})$, then the conditions for $A\in(X,Y)$ are obtained from the respective ones in Proposition 3.3 by replacing the entries of the matrix A by those of the matrix $B=TA$, for instance

$$\sup_n\|B_n\|^*=\sup_n\|R^{(s)}(B_n)\|_1<\infty,$$

where

$$B_n=\sum_{l=\max\{0,n-m\}}(-1)^{n-l}\binom{m}{n-l}A_l.$$

4. Measure of noncompactness and matrix transformations

If X and Y are metric spaces, then $f:X\to Y$ is a compact map if $f(Q)$ is relatively compact (i.e., if the closure of $f(Q)$ is compact subset of Y) subset of Y for each bounded subset Q of X. In this section, among other things, we investigate when, in some special cases (see Corollary 4.3), an operator L_A is compact. Our investigations use the measure of noncompactness. Recall that

if Q is a bounded subset of a metric space X, then the *Hausdorff measure of noncompactness* of Q is denoted by $\chi(Q)$, and

$$\chi(Q) = \inf\{\epsilon > 0 : Q \text{ has a finite } \epsilon\text{-net in } X\}.$$

The function χ is called the *Hausdorff measure of noncompactness*, and for its properties see ([1], [2] or [7]). Denote by $\overline{Q}$ the closure of Q. For the convenience of the reader, let us mention that: If Q, Q_1 and Q_2 are bounded subsets of a metric space (X, d), then

$$\begin{aligned}
\chi(Q) = 0 &\iff Q \text{ is a totally bounded set},\\
\chi(Q) &= \chi(\overline{Q}),\\
Q_1 \subset Q_2 &\iff \chi(Q_1) \leqq \chi(Q_2),\\
\chi(Q_1 \cup Q_2) &= \max\{\chi(Q_1), \chi(Q_2)\},\\
\chi(Q_1 \cap Q_2) &\leqq \min\{\chi(Q_1), \chi(Q_2)\}.
\end{aligned}$$

If our space X is a normed space, then the function $\chi(Q)$ has some additional properties connected with the linear structure. We have, e.g.

$$\begin{aligned}
\chi(Q_1 + Q_2) &\leqq \chi(Q_1) + \chi(Q_2),\\
\chi(\lambda Q) &= |\lambda|\chi(Q) \quad \text{for each} \quad \lambda \in \mathbb{C}.
\end{aligned}$$

If X and Y are normed spaces, then, for $A \in B(X, Y)$ the Hausdorff measure of noncompactness of A, denoted by $\|A\|_\chi$, is defined by $\|A\|_\chi = \chi(AK)$, where $K = \{x \in X : \|x\| \leqq 1\}$ is the unit ball in X. Further, A is compact if and only if $\|A\|_\chi = 0$, and $\|A\|_\chi \leqq \|A\|$.

Recall the following well known result (see e.g. [2, Theorem 6.1.1] or [1, 1.8.1]).

PROPOSITION 4.1. *Let X be a Banach space with a Schauder basis $\{e_0, e_1, \dots\}$, Q a bounded subset of X, and $P_n : X \to X$ the projector onto the linear span of $\{e_0, e_1, \dots, e_n\}$. Then*

$$\begin{aligned}
&\frac{1}{a} \limsup_{n\to\infty} \left(\sup_{x\in Q} \|(I - P_n)x\| \right) \leqq \chi(Q)\\
&\leqq \inf_n \sup_{x\in Q} \|(I - P_n)x\| \leqq \limsup_{n\to\infty} \left(\sup_{x\in Q} \|(I - P_n)x\| \right),
\end{aligned} \tag{4.1}$$

where $a = \limsup_{n\to\infty} \|I - P_n\|$.

Let us mention that concerning the number a in Proposition 4.1, if $X = c_0$, then $a = 1$, but if $X = c$, then $a = 2$ (see e.g. [2, p. 22]).

Concerning Proposition 3.3 and the measures of noncompactness we have

THEOREM 4.2. *Let A be as in Proposition 3.3, and for any integers m, n, r, $n > r$, set*

$$\|A\|^{(r)} = \sup_{n>r} \|R^{(m)}(A_n)\|_1. \tag{4.2}$$

Let X be either $c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, and let $A \in (X, c_0)$. Then we have:

$$\|L_A\|_\chi = \lim_{r\to\infty} \|A\|^{(r)}. \tag{4.3}$$

Let X be either $c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, and let $A \in (X, c)$. Then we have:

$$\frac{1}{2} \lim_{r\to\infty} \|A\|^{(r)} \leqq \|L_A\|_\chi \leqq \lim_{r\to\infty} \|A\|^{(r)}. \tag{4.4}$$

Let X be either $l_\infty(\Delta^{(m)})$, $c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, and let $A \in (X, l_\infty)$. Then we have:

$$0 \leqq \|L_A\|_\chi \leqq \lim_{r\to\infty} \|A\|^{(r)}. \tag{4.5}$$

PROOF. Let us remark that the limits in (4.3), (4.4) and (4.5) exist. Set $K = \{x \in X : \|x\| \leqq 1\}$. In the case $A \in (X, c_0)$ for $X = c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, by Proposition 4.1, we have:

$$\|L_A\|_\chi = \chi(AK) = \lim_{r\to\infty} \left[\sup_{x\in K} \|(I - P_r)Ax\|\right], \tag{4.6}$$

where $P_r : c_0 \to c_0$, $r = 0, 1, \ldots$, is the projector on the first $r+1$ coordinates, i.e., $P_r(x) = (x_0, x_1, \ldots, x_r, 0, 0, \ldots)$, $x = (x_k) \in c_0$ (let us remark that $\|I - P_r\| = 1, r = 0, 1, \ldots$). Further, by Proposition 3.3, we have

$$\|A\|^{(r)} = \sup_{x\in K} \|(I - P_r)Ax\|, \tag{4.7}$$

and by (4.6) we get (4.3).

To prove (4.4) let us remark that every sequence $x = (x_k)_{k=0}^\infty \in c$ has a unique representation

$$x = le + \sum_{k=0}^{\infty} (x_k - l)e^{(k)} \quad \text{where } l \in \mathbb{C} \text{ is such that } x - le \in c.$$

Let us define $P_r : c \to c$ by $P_r(x) = le + \sum_{k=0}^{r} (x_k - l)e^{(k)}$, $r = 0, 1, \ldots$. It is easy to prove that $\|I - P_r\| = 2$, $r = 0, 1, \ldots$. Now the proof of (4.4) is similar as in the case (4.3), and we omit it.

Let us prove (4.5). Now define $P_r : l_\infty \to l_\infty$, by $P_r(x) = (x_0, x_1, \ldots, x_r, 0, 0, \ldots)$, $x = (x_k) \in l_\infty$, $r = 0, 1, \ldots$. It is clear that

$$AK \subset P_r(AK) + (I - P_r)(AK).$$

Now, by the elementary properties of function χ we have

$$\begin{aligned} \chi(AK) &\leqq \chi(P_r(AK)) + \chi((I - P_r)(AK)) = \chi((I - P_r)(AK)) \\ &\leqq \sup_{x \in K} \|(I - P_r)Ax\|. \end{aligned} \tag{4.8}$$

Finally, by Proposition 3.3, we get (4.5). □

As a corollary of the above theorem, we have

COROLLARY 4.3. *Let A be as in Theorem 4.2. Then if $A \in (X, c_0)$ for $X = c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, or if $A \in (X, c)$ for $X = c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, then in all cases we have:*

$$L_A \text{ is compact if and only if } \lim_{r \to \infty} \|A\|^{(r)} = 0. \tag{4.9}$$

Further, if $A \in (X, l_\infty)$ for $X = l_\infty(\Delta^{(m)})$, $X = c_0(\Delta^{(m)})$ or $X = c(\Delta^{(m)})$, then we have:

$$L_A \text{ is compact if } \lim_{r \to \infty} \|A\|^{(r)} = 0. \tag{4.10}$$

The following example will show that it is possible for L_A in (4.10) to be compact in the case $\lim_{r \to \infty} \|A\|^{(r)} > 0$, and hence in general in (4.10) we have just "if".

EXAMPLE 4.4. Let the matrix A be defined by $A_n = e^{(0)}$ $(n = 0, 1, \ldots)$. Then, obviously, $R^{(m)}(A_n) = e^{(0)}$ for all n, and $A \in (l_\infty(\Delta^{(m)}), l_\infty)$. Further,

$$\|A\|^{(r)} = \sup_{n > r} \|R^{(m)}(A_n)\|_1 = \sup_{n > r} \|e^{(0)}\|_1 = 1 > 0 \quad \text{for all } r,$$

whence

$$\lim_{r \to \infty} \|A\|^{(r)} > 0.$$

Since $A(x) = x_0 e$ for all $x \in l_\infty(\Delta^{(m)})$, L_A is a compact operator.

Concerning Corollary 3.5 and the measures of noncompactness we have

THEOREM 4.5. *Let X be a BK space and let A be as in Corollary 3.5. Then for any integer m, n, r, $n > r$, set*

$$\|A\|_\Delta^{(r)} = \sup_{n > r} \left\| \sum_{j = \max\{0, n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} A_j \right\|^*. \tag{4.11}$$

Further, if X has a Schauder basis, and $A \in (X, c_0(\Delta^{(m)}))$, then we have:

$$\|L_A\|_\chi = \lim_{r\to\infty} \|A\|_\Delta^{(r)}. \tag{4.12}$$

If X has a Schauder basis, and $A \in (X, c(\Delta^{(m)}))$, then we have:

$$\frac{1}{2} \lim_{r\to\infty} \|A\|_\Delta^{(r)} \leqq \|L_A\|_\chi \leqq \lim_{r\to\infty} \|A\|_\Delta^{(r)}. \tag{4.13}$$

Finally, if $A \in (X, l_\infty(\Delta^{(m)}))$, then we have:

$$0 \leqq \|L_A\|_\chi \leqq \lim_{r\to\infty} \|A\|_\Delta^{(r)}. \tag{4.14}$$

PROOF. Let us remark that the limits in (4.12), (4.13) and (4.14) exist. Set $K = \{x \in X : \|x\| \leqq 1\}$. To prove (4.12), by Proposition 2.1 and Proposition 4.1, we have:

$$\|L_A\|_\chi = \chi(AK) = \lim_{r\to\infty} \left[\sup_{x\in K} \|(I - P_r)Ax\|\right], \tag{4.15}$$

where $P_r : c_0(\Delta^{(m)}) \to c_0(\Delta^{(m)})$, $r = 0, 1, \ldots$, is the projector defined by (see Proposition 2.1)

$$P_r(x) = \sum_{k=0}^{r} \lambda_k(m) b^{(k)}(m), \tag{4.16}$$

where $x = \sum_{k=0}^{\infty} \lambda_k(m) b^{(k)}(m) \in c_0(\Delta^{(m)})$ and $b^{(k)}(m)$ is a Schauder basis of $c_0(\Delta^{(m)})$. Let us remark that $\|I - P_r\| = 1$, $(r = 0, 1, \ldots)$. Further, by Corollary 3.5, we have

$$\|A\|_\Delta^{(r)} = \sup_{x\in K} \|(I - P_r)Ax\|. \tag{4.17}$$

To prove (4.13) let us remark (see Proposition 2.1) that $c(\Delta^{(m)})$ has the Schauder basis $b^{(k)}(m), k = -1, 0, 1, \ldots$, and every $x \in c(\Delta^{(m)})$ has a unique representation

$$x = l b^{(-1)}(m) + \sum_{k=0}^{\infty} (\lambda_k(m) - l) b^{(k)}(m), \quad \text{where} \quad l = \lim_{k\to\infty} (\Delta^{(m)} x)_k. \tag{4.18}$$

Now, let us define $P_r : c(\Delta^{(m)}) \to c(\Delta^{(m)})$, $r = 0, 1, \ldots$, by

$$P_r(x) = \sum_{k=0}^{r} (\lambda_k(m) - l) b^{(k)}(m). \tag{4.19}$$

It is easy to prove that $\|I-P_r\|=2$, $r=0,1,\dots$. Now the proof of (4.13) is similar as in the case (4.12), and we omit it.

Let us prove (4.14). Now define $P_r: l_\infty(\Delta^{(m)}) \to l_\infty(\Delta^{(m)})$, by $P_r(x) = (x_0, x_1, \dots, x_r, 0, 0, \dots)$, $x=(x_i)\in l_\infty(\Delta^{(m)})$, $r=0,1,\dots$. It is clear that

$$AK \subset P_r(AK) + (I-P_r)(AK).$$

Now, by the elementary properties of the function χ, we have

$$\begin{aligned}\chi(AK) &\leqq \chi(P_r(AK)) + \chi((I-P_r)(AK)) = \chi((I-P_r)(AK)) \\ &\leqq \sup_{x\in K} \|(I-P_r)Ax\|.\end{aligned} \tag{4.20}$$

Finally, by Proposition 3.4 and Corollary 3.5, we get (4.14). □

As a corollary of the above theorem, we have

COROLLARY 4.6. *Let X be a BK space and let A and $\|A\|_\Delta^{(r)}$ be as in Theorem 4.5. If X has a Schauder basis, and either $A\in(X, c_0(\Delta^{(m)}))$ or $A\in(X, c(\Delta^{(m)}))$, then*

$$L_A \text{ is compact if and only if } \lim_{r\to\infty} \|A\|_\Delta^{(r)} = 0. \tag{4.21}$$

Further, if $A\in(X, l_\infty(\Delta^{(m)}))$, then we have:

$$L_A \text{ is compact if } \lim_{r\to\infty} \|A\|_\Delta^{(r)} = 0. \tag{4.22}$$

Now, concerning Remark 1, we get several corollaries.

COROLLARY 4.7. *If either $A\in(l_p, c_0(\Delta^{(m)}))$ or $A\in(l_p, c(\Delta^{(m)}))$ $(1<p<\infty)$, then*

L_A *is compact if and only if*

$$\lim_{r\to\infty} \sup_{n>r} \left(\sum_{k=0}^{\infty} \left| \sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} a_{jk} \right|^q \right) = 0, \tag{4.23}$$

$$q = p/(p-1).$$

Further, if either $A\in(l_1, c_0(\Delta^{(m)}))$ or $A\in(l_1, c(\Delta^{(m)}))$, then

L_A *is compact if and only if* (4.24)

$$\lim_{r\to\infty} \sup_{n>r,k} \left| \sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} a_{jk} \right| = 0. \tag{4.25}$$

If $A \in (l_p, l_\infty(\Delta^{(m)}))$, $1 < p < \infty$, *then*

L_A *is compact if*

$$\lim_{r\to\infty} \sup_{n>r} \left(\sum_{k=0}^{\infty} \left| \sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} a_{jk} \right|^q \right) = 0, \tag{4.26}$$

$$q = p/(p-1).$$

Finally, if $A \in (l_1, l_\infty(\Delta^{(m)}))$, *then*

L_A *is compact if*

$$\lim_{r\to\infty} \sup_{n>r,k} \left| \sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} a_{jk} \right| = 0. \tag{4.27}$$

From Corollary 4.3, Proposition 3.3 and Remark 1 (b), we have

COROLLARY 4.8. *Let* s *and* m *be non negative integers. If* $A \in (X, c_0(\Delta^{(m)})$ *for* $X = c_0(\Delta^{(s)})$ *or* $X = c(\Delta^{(s)})$, *or if* $A \in (X, c(\Delta^{(m)})$ *for* $X = c_0(\Delta^{(s)})$ *or* $X = c(\Delta^{(s)})$, *then in all cases we have:*

L_A *is compact if and only if*

$$\lim_{r\to\infty} \sup_{n>r} \left\| R^{(s)} \left(\sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} A_j \right) \right\|_1 = 0. \tag{4.28}$$

Further, if $A \in (X, l_\infty(\Delta^{(m)})$ *for* $X = l_\infty(\Delta^{(s)})$, $X = c_0(\Delta^{(s)})$ *or* $X = c(\Delta^{(s)})$, *then we have:*

L_A *is compact if*

$$\lim_{r\to\infty} \sup_{n>r} \left\| R^{(s)} \left(\sum_{j=\max\{0,n-m\}}^{n} (-1)^{n-j} \binom{m}{n-j} A_j \right) \right\|_1 = 0. \tag{4.29}$$

REFERENCES

[1] AKHMEROV, R. R., KAMENSKII, M. I., POTAPOV, A. S., RODKINA, A. E. and SADOVSKII, B. N., *Measures of noncompactness and condensing operators*, Operator Theory: Advances and Applications, 55, Birkhäuser-Verlag, Basel, 1992. *MR* **92k**:47104

[2] BANÁS, J. and GOEBEL, K., *Measures of noncompactness in Banach spaces*, Lecture Notes in Pure and Applied Mathematics, 60, Marcel Dekker, New York and Basel, 1980. *MR* **82f**:47066

[3] HARDY, G. H., *Divergent series*, Oxford University Press, Oxford, 1973. First edition: Oxford, Clarendon Press, 1949. *MR* **11**, 25a

[4] MALKOWSKY, E., Linear operators in certain BK spaces, *Approximation theory and function series* (Budapest, 1995), Bolyai Soc. Math. Stud. **5**, János Bolyai Math. Soc., Budapest, 1996, 259–273. *CMP* 97, 08

[5] MALKOWSKY, E. and PARASHAR, S. D., Matrix transformations in spaces of bounded and convergent difference sequences of order m, *Analysis* **17** (1997), 87–97. *CMP* 97, 13

[6] MALKOWSKY, E. and RAKOČEVIĆ, V., The measure of noncompactness of linear operators between certain sequence spaces, *Acta Sci. Math. (Szeged)* **64** (1998), 151–170.

[7] RAKOČEVIĆ, V., *Funkcionalna analiza*, Naučna knjiga, Beograd, 1994.

[8] WILANSKY, A., *Summability through functional analysis*, North-Holland Mathematics Studies, No. 85, North-Holland Publ. Co., Amsterdam – New York, 1984. *MR* **85d**:40006

(Received July 24, 1997)

MATHEMATISCHES INSTITUT
UNIVERSITÄT GIESSEN
ARNDTSTRASSE 2
D–35392 GIESSEN
GERMANY

ema@math.uni-giessen.de

DEPARTMENT OF MATHEMATICS
FACULTY OF PHILOSOPHY
UNIVERSITY OF NIŠ
ĆIRILIA I METODIJA 2
YU–18000 NIŠ
YUGOSLAVIA

vrakoc@archimed.filfak.ni.ac.yu
vrakoc@baukerinter.net

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 397–414

ON THE PRESERVATION OF GLOBAL SMOOTHNESS BY SOME INTERPOLATION OPERATORS

S. G. GAL and J. SZABADOS*

1. Introduction

When one approximates an element f of a function space by a sequence of approximation operators $\{L_n(f)\}$, it is important to know, for example, the relation between the global smoothness properties of f and $L_n(f)$, i.e. if the following implication

$$f \in \mathrm{Lip}_M(\alpha;[a,b]) \Longrightarrow L_n(f) \in \mathrm{Lip}_{M'}(\alpha;[a,b]),\ n \in \mathbf{N}, \tag{1}$$

or the stronger condition

$$\omega(L_n(f);h) \leqq c\,\omega(f;h), \quad 0 \leqq h \leqq h_0,\ n \in \mathbf{N}, \tag{2}$$

are valid. Here

$$\mathrm{Lip}_M(\alpha;[a,b]) := \{f : [a,b] \to \mathbf{R};\ |f(x)-f(y)| \leqq M|x-y|^\alpha, \forall x,y \in [a,b]\},$$
$$0 < \alpha \leqq 1,$$

and ω represents the modulus of continuity.

The case when $L_n(f)$ are trigonometric polynomials is settled by the following well-known result of S. B. Stechkin.

THEOREM A ([21, pp. 229–230]). *If $f \in C_{2\pi}$ and $\{T_n\}_n$ is a sequence of trigonometric polynomials of order at most n satisfying*

$$\|f - T_n\| \leqq c_1 \omega_k\left(f;\frac{1}{n}\right), \quad \forall n \in \mathbf{N},$$

then there exists a constant $c(k) > 0$ (independent of f and n) such that for all $h > 0$ one has

$$\omega_k(T_n;h) \leqq c(k)\,\omega_k(f;h), \quad \forall n \in \mathbf{N},$$

1991 *Mathematics Subject Classification.* Primary 41A05; Secondary 41A17.
Key words and phrases. Interpolation operator, smoothness, modulus of continuity.

*Research of this author was supported by Hungarian National Science Foundation Grant No. T017425.

where ω_k is the well-known uniform modulus of smoothness of order k.

REMARK. In fact, in his paper Stechkin proved much more (see [21, Theorem 6, p. 330]), in the sense that in Theorem A, $\omega_k(f;h)$ can be replaced by a function $\varphi(h)$ having some suitable properties.

In the case when $\{L_n(f)\}_n$ are various linear (and sometimes positive) operators applied to a non-periodic function $f \in C[a,b]$, there exists an extensive literature realizing the above relations (1) and (2) (see e.g. [1–5], [8–9] and [11–18]). But when (1) does not hold for $\{L_n(f)\}_n$, a natural question arises: how much of the global smoothness of f is preserved by $\{L_n(f)\}_n$, which can be expressed, for example, in the following way: if $f \in \mathrm{Lip}_M(\alpha;[a,b])$ then there exist $\beta < \alpha$ and $M' > 0$ (independent of n) such that $L_n(f) \in \mathrm{Lip}_{M'}(\beta;[a,b])$ for all $n \in \mathbf{N}$.

REMARK. The following two examples show that the above property of partial preservation of the global smoothness does not depend on the approximation properties of the sequence $\{L_n(f)\}_n$.

Indeed, if $L_n(f) = S_n(f)$ represents the Fourier sum of order n of $f \in C_{2\pi}$, while it is well-known that $\{S_n(f)\}_n$ has no good approximation properties for all $f \in C_{2\pi}$, by [21, p. 231] we have

$$\omega(S_n(f);h) \leqq c\,\omega(f;h)\log\frac{1}{h},$$

which means that $\{S_n(f)\}$ has the property of partial preservation of the global smoothness of f.

As the second example, we choose $L_n(f) = P_n(f)$, the best approximation of $f \in C[-1,1]$ by algebraic polynomials of degree at most n. First let $x, y \in [-1,1]$, $|x-y| \leqq h$. By [25, Section 4.12, (20)] we get $\|P_n'(f)\| \leqq n^2\,\omega(P_n(f),1/n)$. Since by Jackson's theorem[1] $\|P_n(f)-f\| \leqq c\,\omega(f;1/n)$, we obtain

$$\begin{aligned}
|P_n(f,x)-P_n(f,y)| &= |P_n'(f)(\xi)|\,|x-y| \leqq \|P_n'(f)\|h \\
&\leqq hn^2\left[\omega\left(P_n(f)-f;\frac{1}{n}\right)+\omega\left(f;\frac{1}{n}\right)\right] \\
&\leqq hn^2\left[2\|P_n(f)-f\|+c\,\omega\left(f;\frac{1}{n}\right)\right] \\
&\leqq 3chn^2\,\omega\left(f;\frac{1}{n}\right) \leqq 6c\,\omega\left(f;\sqrt{h}\right), \quad n \in \mathbf{N}.
\end{aligned}$$

On the other hand, if $h > 1/n^2$, then

$$|P_n(f,x)-P_n(f,y)| \leqq |P_n(f,x)-f(x)|+|f(x)-f(y)|+|f(y)-P_n(f,y)|$$

[1] Throughout the paper, c, c_1, c_2, etc., will denote absolute positive constants, not necessarily the same at each occurrence.

$$\leqq 2c\,\omega\left(f;\frac{1}{n}\right)+\omega(f;h)\leqq(2c+1)\,\omega(f;\sqrt{h}),$$

which also means that $\{P_n(f)\}_n$ has the property of partial preservation of the global smoothness of f.

The purpose of the present paper is to consider the problem of partial preservation of the global smoothness for sequences $\{L_n(f)\}_n$ of interpolatory type. As it was proved for example in [4], if one takes as $L_n(f)$ the Hermite–Fejér operator based on the Chebyshev nodes of the first kind, then (1) does not hold for $\alpha=1$. In Section 2 we will prove some negative results about the (partial) preservation of global smoothness by interpolation polynomials, while in Section 3 we will obtain some positive results concerning the Hermite–Fejér polynomials based on the Chebyshev nodes of the first kind, the Lagrange interpolation polynomials based on the Chebyshev nodes of second kind and ± 1, as well as the Shepard operators.

2. Negative results

In [17], Kratz and Stadtmüller proved the estimate (2) (and then the constant c was improved in [4]) for discrete sequences $\{L_n(f)\}_n$ of the form

$$L_n(f,x)=\sum_{k=1}^{n} f(x_{k,n})p_{k,n}(x),\quad x\in[-1,1],\ f\in C[-1,1] \tag{3}$$

satisfying the conditions

$$\sum_{k=1}^{n} p_{k,n}(x)\equiv s_n \quad \text{is independent of} \quad x\in[-1,1], \tag{4}$$

$$\sum_{k=1}^{n} |p_{k,n}(x)|\leqq c_1,\quad x\in[-1,1], \tag{5}$$

$$p_{k,n}\in C^1[-1,1] \quad \text{and} \quad \sum_{k=1}^{n} |(x-x_{k,n})p'_{k,n}(x)|\leqq c_2,\quad x\in[-1,1] \tag{6}$$

for some constants $c_1, c_2>0$.

Since the Hermite–Fejér and Lagrange polynomials are of type (3), it is natural to ask if for these polynomials (4)–(6) hold. Unfortunately, this is not so for Lagrange interpolation because of the following

THEOREM 1. *Let $\mathcal{M}=\{\mathcal{M}_n\}_{n\in\mathbf{N}}$ ($\mathcal{M}_n=\{x_{k,n}\}_{k=1}^n$) be an arbitrary triangular matrix of interpolation nodes in $[-1,1]$ (i.e. $-1\leqq x_{n,n}<x_{n-1,n}<\cdots<x_{1,n}\leqq 1$, $n\in\mathbf{N}$) and $l_{k,n}(x)$ the fundamental polynomials of Lagrange interpolation on $\mathcal{M}_n$. Then for all $n\geqq 2$ we have*

$$1\leqq\liminf_{n\to\infty}\frac{\inf_{\mathcal{M}_n}\max_{|x|\leqq 1}\sum_{k=1}^n|x-x_{k,n}|\,|l'_{k,n}(x)|}{n}$$
$$\leqq\limsup_{n\to\infty}\frac{\inf_{\mathcal{M}_n}\max_{|x|\leqq 1}\sum_{k=1}^n|x-x_{k,n}|\,|l'_{k,n}(x)|}{n}\leqq 2. \tag{7}$$

PROOF. Let us denote $x_{k,n}=x_k$, $k=1,\dots,n$, and

$$A_n(x;\mathcal{M}_n):=\sum_{k=1}^n|x-x_{k,n}|\,|l'_{k,n}(x)|,$$

where

$$l_{k,n}(x):=\frac{\omega_n(x)}{\omega'_n(x_k)(x-x_k)},\quad k=1,\dots,n,$$
$$\omega_n(x):=\prod_{k=1}^n(x-x_k).$$

Consider an index $j\in\{1,2,\dots,n\}$ such that

$$|\omega'_n(x_j)|:=\max_{1\leqq k\leqq n}|\omega'_n(x_k)|.$$

Then we get

$$A_n(x_j;\mathcal{M}_n)=|\omega'_n(x_j)|\sum_{\substack{k=1\\k\neq j}}\frac{1}{|\omega'_n(x_k)|}\geqq n-1,$$

which proves the first inequality in (7).

To prove the second inequality, let us choose

$$\omega_n(x):=\frac{1}{2}[T_{n-2}(x)-T_n(x)]=\sin t\sin(n-1)t,\quad x=\cos t,$$

where $T_n(x):=\cos(n\arccos x)$ is the Chebyshev polynomial of degree n. Then $x_k=\cos t_k$, $t_k=\frac{(k-1)\pi}{n-1}$, $k=1,\dots,n$, and an easy calculation yields

$$|\omega'_n(x_k)|=\begin{cases}n-1 & \text{if } 2\leqq k\leqq n-2,\\ 2n-1 & \text{if } k=1,n,\end{cases}$$

and

$$\max_{|x|\leqq 1} |\omega_n(x)| \leqq 1, \quad \max_{|x|\leqq 1} |\omega_n'(x)| \leqq 2n-2.$$

Thus denoting an index j for which $|x-x_j| := \min_{1\leqq k\leqq n} |x-x_k|$ we obtain

$$A_n(x,\mathcal{M}_n) \leqq |\omega_n'(x)| \sum_{k=1}^{n} \frac{1}{|\omega_n'(x_k)|} + \sum_{k=1}^{n} \frac{|\omega_n(x)|}{|\omega_n'(x_k)||x-x_k|}$$

$$\leqq (2n-2)\left(\frac{n-2}{n-1}+\frac{2}{2n-2}\right)+2+\sum_{\substack{k=1\\k\neq j}}^{n} \frac{\sin t}{(n-1)|\cos t-\cos t_k|}$$

$$\leqq 2n+\frac{1}{n-1}\sum_{\substack{k=1\\k\neq j}}^{n}\left(\frac{1}{\sin\dfrac{t-t_k}{2}}+\frac{1}{\sin\dfrac{t+t_k}{2}}\right)$$

$$\leqq 2n+O\left(\sum_{\substack{k=1\\k\neq j}}^{n}\frac{1}{|j-k|}\right)=2n+O(\log n),$$

which completes the proof.

If $\{L_n(f)\}_n$ are the Hermite–Fejér polynomials based on the Chebyshev nodes of first kind, then obviously (4) and (5) hold with $s_n \equiv 1$ and $c_1 = 1$, but (6) cannot hold since then (2) would also hold which contradicts [4].

If $\{L_n(f)\}_n$ are the classical Lagrange polynomials for any system of nodes then it is known that (5) does not hold. As a conclusion, it seems that for classical interpolatory polynomials all of the three conditions (4) to (6) cannot be verified.

REMARKS. 1. This shortcoming can be removed as follows. Let $L_n(f,x)$ be of the form (3) with $p_{k,n}(x)$ satisfying (6). Then $f \in \operatorname{Lip}_M(1;[-1,1])$ implies $L_n(f) \in \operatorname{Lip}_{c_2 M}(1;[-1,1])$ for all $n \in \mathbf{N}$. Indeed, we have

$$|L_n(f,x)-L_n(f,y)| = \left|(x-y)\sum_{k=1}^{n} f(x_{k,n})p_{k,n}'(\xi_{x,y,n})\right|$$

$$= |x-y|\left|\sum_{k=1}^{n}[f(x_{k,n})-f(\xi_{x,y,n})]p_{k,n}'(\xi_{x,y,n})\right|$$

$$\leqq |x-y| M \sum_{k=1}^{n} |x_{k,n}-\xi_{x,y,n}|\,|p_{k,n}'(\xi_{x,y,n})| \leqq c_2 M|x-y|.$$

2. Let $x_{k,n} := -1 + \frac{2(k-1)}{n-1}$, $k = 1, \dots, n$, be the equidistant nodes in $[-1, 1]$ and $H_{2n-1}(f, x)$ the Hermite–Fejér interpolation polynomials on these nodes. D. L. Berman [7] proved that for $f(x) := x$, the sequence $\{H_{2n-1}(f, x)\}_n$ is unboundedly divergent for any $0 < |x| < 1$. Hence this sequence has no partial preservation of global smoothness. Indeed, if $f \in \mathrm{Lip}_1(1; [-1, 1])$ and if we suppose that there exist $0 < \alpha < 1$ and $M > 0$ such that $H_{2n-1}(f) \in \mathrm{Lip}_M(\alpha; [-1, 1])$ for all $n \in \mathbf{N}$, then

$$|H_{2n-1}(f, x) - H_{2n-1}(f, y)| \leqq M|x - y|^\alpha, \quad \forall x, y \in [-1, 1],\ n \in \mathbf{N}.$$

Taking $x = -1$, $0 < |y| < 1$ and letting $n \to \infty$ in the above inequality, we get a contradiction.

3. Positive results

First we consider two examples of trigonometric interpolation polynomials. For an $f \in C_{2\pi}$, let $I_n(f, x)$ be the trigonometric interpolation polynomial on the equidistant nodes in $[0, 2\pi)$. It is known (see e.g. [20]) that

$$\|f - I_n(f)\| \leqq c\,\omega\left(f; \frac{1}{n}\right) \log n, \quad \forall n \in \mathbf{N}.$$

Denoting $\varphi(h) := \omega(f; h) \log 1/h$, $0 < h < 1$, by [21, Theorem 6, p. 230] we get

$$\omega(I_n(f); h) \leqq c'\,\omega(f; h) \log \frac{1}{h}, \quad \forall 0 < h < 1.$$

The second example is the trigonometric Jackson interpolation polynomial $J_n(f, x)$ (see e.g. [22]). Concerning these polynomials, the second author proved [21] the estimate

$$\|f - J_n(f)\| \leqq c\left[\omega\left(f; \frac{1}{n}\right) + \omega\left(\tilde{f}; \frac{1}{n}\right)\right], \quad \forall n \in \mathbf{N},$$

where $\tilde{f}$ represents the trigonometric conjugate of $f \in C_{2\pi}$. Let $f \in \mathrm{Lip}_M \alpha$, $0 < \alpha \leqq 1$. If $\alpha < 1$ then it is known (see e.g. [6, p. 485]) that this is equivalent to $\tilde{f} \in \mathrm{Lip}_{\bar{M}} \alpha$, which, by the above estimate and by [23, Theorem 6, p. 230] give $\omega(J_n(f); h) \leqq ch^\alpha$, $h > 0$, $n \in \mathbf{N}$, i.e. we can say that in this case $\{J_n(f)\}_n$ completely preserves the global smoothness of f. Also, if $\alpha = 1$, then by e.g. [26, p. 157] it follows that $\omega(\tilde{f}; h) \leqq \bar{M} h \log 1/h$, which again together with the above estimate and with [21, Theorem 6, p. 230] yields

$$\omega(J_n(f); h) \leqq ch \log \frac{1}{h}, \quad 0 < h < 1,\ n \in \mathbf{N}.$$

REMARK. Let $x_k = \cos\frac{2k-1}{2n}\pi$, $k = 1, \dots, n$, be the roots of the Chebyshev polynomial $T_n(x)$, and the Hermite–Fejér polynomial of an $f \in C[-1,1]$ based on these roots,

$$H_n(f,x) = \sum_{k=1}^{n} f(x_k)(1 - xx_k)\frac{T_n^2(x)}{n^2(x-x_k)^2}.$$

The above considerations about the preservation properties of $J_n(f,x)$ induce a property of partial preservation for $H_n(f,x)$, too. Indeed, if we denote $g(t) = f(\cos t)$, then it is known that (see e.g. [22, p. 406])

$$H_n(f,x) \equiv J_{2n-1}(g,t), \quad t = \arccos x.$$

Now if $0 < \alpha < 1$ and $f \in \mathrm{Lip}_c\alpha$, then $J_{2n-1}(g)(t) \in \mathrm{Lip}\,\alpha$, which can be written as $(x = \cos u, y = \cos v)$

$$|H_n(f,x) - H_n(f,y)| = |J_{2n-1}(g)(u) - J_{2n-1}(g)(v)| \leqq c|u-v|^\alpha$$

$$= c\,|\arccos x - \arccos y|^\alpha \leqq \frac{c\pi}{\sqrt{2}}|x-y|^{\alpha/2}, \quad \forall x, y \in [-1,1]$$

(since by e.g. [10, p. 88, Problem 5], $\arccos x \in \mathrm{Lip}_{\frac{\pi}{\sqrt{2}}}(1/2;[-1,1])$, which means that $H_n(f) \in \mathrm{Lip}_{\frac{c\pi}{\sqrt{2}}}(\alpha/2;[-1,1])$ for all $n \in \mathbf{N}$).

If $\alpha = 1$, in the same way we get $\omega(H_n;h) \leqq c\left[h\left|\log\frac{1}{h}\right|\right]^{1/2}$. However, by a direct method we will improve the last consideration about $H_n(f)$.

THEOREM 2. *For any $f \in C[-1,1]$, $h > 0$ and $n \in \mathbf{N}$ we have*

$$\omega(H_n(f);h)$$

$$= \min\left[O\left(hn\sum_{k=1}^{n}\omega\left(f;\frac{1}{k^2}\right)\right), O\left(\frac{1}{n}\sum_{k=1}^{n}\omega\left(f;\frac{1}{k}\right) + \omega(f;h)\right)\right],$$

where the constants in "O" are independent of f, n and h.

PROOF. First we obtain an upper estimate for $|H_n'(f,x)|$. Let $x \in [-1,1]$ be fixed, the index j defined by $|x - x_j| := \min\limits_{1 \leqq k \leqq n} |x - x_k|$ and denote

$$A_k(x) := (1 - xx_k)\frac{T_n^2(x)}{n^2(x-x_k)^2}.$$

We have

$$A_k'(x) = -x_k\frac{T_n^2(x)}{(x-x_k)^2} + (1 - xx_k)\frac{2T_n(x)T_n'(x)}{(x-x_k)^2} - \frac{2(1-xx_k)T_n^2(x)}{(x-x_k)^3},$$

$k=1,\dots,n$, which immediately implies (by $1-xx_k=1-x^2+x(x-x_k)$)

$$|A_k'(x)| \leqq \frac{1}{(x-x_k)^2} + \frac{2(1-x^2)|T_n'(x)|}{(x-x_k)^2} + \frac{2|T_n'(x)|}{|x-x_k|} + \frac{2(1-x^2)}{|x-x_k|^3} + \frac{2}{(x-x_k)^2}, \quad k=1,\dots,n.$$

For simplicity, all the constants (independent of n and h) which appear will be denoted by c. Since

$$H_n'(f,x) = \sum_{k=1}^{n} f(x_k)A_k'(x) = \sum_{k=1}^{n} [f(x_k)-f(x)]A_k'(x),$$

we obtain

$$|H_n'(f,x)| \leqq \frac{c}{n^2} \sum_{\substack{k=1 \\ k\neq j}}^{n} \omega(f;|x-x_k|) \left[\frac{3}{(x-x_k)^2} + \frac{2(1-x^2)|T_n'(x)|}{(x-x_k)^2} + \frac{2(1-x^2)}{|x-x_k|^3} + \frac{2|T_n'(x)|}{|x-x_k|} \right] + c\,\omega(f;|x-x_j|)|A_j'(x)|. \tag{8}$$

The following known relation (see e.g. [24, p. 282]) will be frequently used:

$$|x-x_j| \leqq \frac{cj}{n^2}, \quad n\sqrt{1-x^2} \sim j, \quad |x-x_k| \sim \frac{|j^2-k^2|}{n^2}, \; k\neq j. \tag{9}$$

Now by (9) and by the combined Bernstein–Markov inequality we get

$$|A_j'(x)| \leqq \frac{n^2\|A_j\|}{n\sqrt{1-x^2}+1} \leqq \frac{cn^2}{j},$$

and

$$\omega(f;|x-x_j|)|A_j'(x)| \leqq c\,\omega\left(f;\frac{j}{n^2}\right)\frac{n^2}{j} \leqq c\,\omega\left(f;\frac{1}{n^2}\right)n^2.$$

Also, by (9) we obtain (using also the inequality $\omega(f;T)/T \leqq 2\,\omega(f;t)/t$ for $t \leqq T$)

$$\sum_{k\neq j} \frac{\omega(f;|x-x_k|)}{(x-x_k)^2} \leqq \frac{n^2}{j} \sum_{k\neq j} \frac{\omega(f;|x-x_k|)}{|x-x_k|}$$

$$\leqq \frac{cn^4}{j} \sum_{k\neq j} \frac{\omega\left(f;\dfrac{|j^2-k^2|}{n^2}\right)}{|j^2-k^2|} \leqq \frac{cn^4}{j} \sum_{k=1}^{n} \frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right),$$

$$\sum_{k\neq j}\frac{\omega(f;|x-x_k|)(1-x^2)|T_n'(x)|}{(x-x_k)^2}\leqq\sum_{k\neq j}\frac{\omega(f;|x-x_k|)n\sqrt{1-x^2}}{(x-x_k)^2}$$

$$\leqq cj\sum_{k\neq j}\frac{\omega(f;|x-x_k|)}{(x-x_k)^2}\leqq cn^4\sum_{k=1}^{n}\frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right),$$

$$\sum_{k\neq j}\frac{\omega(f;|x-x_k|)(1-x^2)}{|x-x_k|^3}\leqq c\sum_{k\neq j}\frac{\omega\left(f;\dfrac{|j^2-k^2|}{n^2}\right)\dfrac{j^2}{n^2}}{\dfrac{|j^2-k^2|^3}{n^6}}$$

$$\leqq cn^6\,\omega\left(f;\frac{1}{n^2}\right)\sum_{k\neq j}\frac{1}{|j^2-k^2|}\frac{j^2}{n^2}\leqq\frac{cn^6}{j^2}\,\omega\left(f;\frac{1}{n^2}\right)\frac{j^2}{n^2}\leqq cn^4\,\omega\left(f;\frac{1}{n^2}\right),$$

$$\sum_{k\neq j}\frac{\omega(f;|x-x_k|)|T_n'(x)|}{|x-x_k|}\leqq cn^2\sum_{k\neq j}\frac{\omega(f;|x-x_k|)}{|x-x_k|}$$

$$\leqq cn^4\sum_{k=1}^{n}\frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right).$$

Collecting now all of these estimates, by (8) we get

$$|H_n'(f,x)|\leqq cn^2\sum_{k=1}^{n}\frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right).$$

Since

$$n^2\sum_{k=1}^{n}\frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right)\sim n\int_{1/n}^{1}\frac{\omega(f;t^2)}{t^2}dt=n\int_{1}^{n}\omega\left(f;\frac{1}{u^2}\right)du$$

$$\sim n\sum_{k=1}^{n}\omega\left(f;\frac{1}{k^2}\right)$$

(we have used the equivalence between the Riemann integral sums and the integral itself, and made a substitution $t=1/u$), the above estimate becomes

$$|H_n'(f,x)|\leqq cn\sum_{k=1}^{n}\omega\left(f;\frac{1}{k^2}\right). \tag{10}$$

On the other hand, for $|x-y| \leqq h$ and by e.g. [24, Theorem 5.1, p. 168], we get

$$|H_n(f,x) - H_n(f,y)| \leqq |H_n(f,x) - f(x)| + |f(x) - f(y)| + |f(y) - H_n(f,y)|$$
$$\leqq 2\|H_n(f) - f\| + \omega(f;h) \leqq c \sum_{k=1}^{n} \frac{1}{k^2}\, \omega\left(f; \frac{k}{n}\right) + \omega(f;h).$$

But similarly to the above considerations,

$$\sum_{k=1}^{n} \frac{1}{k^2}\, \omega\left(f; \frac{k}{n}\right) \sim \frac{1}{n} \int_{1/n}^{1} \frac{\omega(f;t)}{t^2} dt = \frac{1}{n} \int_{1}^{n} \omega\left(f; \frac{1}{u}\right) du \sim \frac{1}{n} \sum_{k=1}^{n} \omega\left(f; \frac{1}{k}\right),$$

and therefore we get

$$\omega(H_n(f);h) = O\left[\frac{1}{n} \sum_{k=1}^{n} \omega\left(f; \frac{1}{k}\right)\right] + \omega(f;h). \tag{11}$$

Now, on the other hand, using (10) we obtain for $|x-y| \leqq h$,

$$|H_n(f,x) - H_n(f,y)| \leqq |H_n'(f,\xi)|h \leqq cnh \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right),$$

i.e.

$$\omega(H_n(f);h) = O\left[hn \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right)\right],$$

which together with (11) proves the theorem.

COROLLARY 3. *If $f \in \operatorname{Lip}\alpha$, $0 < \alpha \leqq 1$ then for all $n \in \mathbf{N}$ and $0 < h < 1$ we have*

$$\omega(H_n(f),h) = \begin{cases} O\left(h^{\frac{\alpha}{\max(2-\alpha,1+\alpha)}}\right) & \text{if } 0 < \alpha < 1/2 \text{ or } 1/2 < \alpha < 1, \\ O\left(\left[h \log \frac{1}{h}\right]^{\frac{2\alpha+1}{6}}\right) & \text{if } \alpha = 1/2 \text{ or } 1. \end{cases}$$

PROOF. The optimal point in Theorem 2 is when $h = v_n$, where

$$v_n = \frac{\sum_{k=1}^{n} \omega\left(f; \frac{1}{k}\right)}{n^2 \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right)}.$$

When $h < v_n$, the minimum in Theorem 2 is the first term, and when $h > v_n$, it is the second term. By simple calculations we have

$$v_n = \begin{cases} O(n^{\alpha-2}) & \text{if } 0 < \alpha < 1/2, \\ O\left(\dfrac{1}{n^{3/2}\log n}\right) & \text{if } \alpha = 1/2, \\ O(n^{-1-\alpha}) & \text{if } 1/2 < \alpha < 1, \\ O\left(\dfrac{\log n}{n^2}\right) & \text{if } \alpha = 1. \end{cases}$$

Hence, by using Theorem 2 we arrive at the statement of the corollary.

REMARKS. 1. Let $0 < \alpha < 1$. The obvious inequalities

$$\frac{\alpha}{2} < \frac{\alpha}{\max(2-\alpha, 1+\alpha)}, \quad h^{1/4} > \left[h \log \frac{1}{h}\right]^{1/3}$$

mean that the preservation property given by Corollary 3 is better than that given by the previous Remark.

2. It is an open question if the estimates of $\omega(H_n(f); h)$ in Theorem 2 and Corollary 3 are best possible. However, if we choose, for example, $f_0(x) := x \in \mathrm{Lip}_1(1; [-1, 1])$, then we can prove that $\omega(H_n(f_0); h) \sim \sqrt{h}$. Indeed, by

$$H_n(f_0, x) = x - \frac{T_n(x)T_{n-1}(x)}{n} = x - \frac{T_{2n-1}(x) + T_1(x)}{2n}$$

(see [4]), we get

$$|H_n'(f_0, x)| = \left|1 - \frac{T_{2n-1}'(x) + 1}{2n}\right| \geqq \frac{c}{\sqrt{1-x^2}} \geqq cn,$$

for all $x \in \left[\dfrac{1+x_1}{2}, 1\right]$ and

$$\omega\left(H_n(f_0); \frac{1-x_1}{2}\right) \geqq \left|H_n(f_0, 1) - H_n\left(f, \frac{1+x_1}{2}\right)\right|$$

$$= |H_n'(f_0, \xi)|\frac{1-x_1}{2} \geqq cn(1-x_1) = c\sqrt{\frac{1-x_1}{2}},$$

as claimed. Now we will prove that in fact $H_n(f_0) \in \mathrm{Lip}_M \frac{1}{2}$, for all $n \in \mathbf{N}$. Evidently, it suffices to prove that $\dfrac{T_{2n-1}(x)}{2n} \in \mathrm{Lip}_M \dfrac{1}{2}$ for all $n \in \mathbf{N}$. But by [10, Problem 5, p. 88]

$$\frac{|T_{2n-1}(x) - T_{2n-1}(y)|}{2n} =$$

$$= \frac{|\cos(2n-1)\arccos x - \cos(2n-1)\arccos y|}{2n}$$
$$\leqq \frac{(2n-1)|\sin\xi|\,|\arccos x - \arccos y|}{2n}$$
$$\leqq M|\arccos x - \arccos y| \leqq \frac{M\pi}{\sqrt{2}}|x-y|^{1/2},$$

which was to be proved.

Now consider the Lagrange interpolation polynomial $L_n(f)$ based on the Chebyshev nodes of second kind plus the endpoints ± 1. It is known [19] that

$$L_n(f,x) = \sum_{k=1}^{n} f(x_k) l_k(x),$$

where $x_k = \cos t_k$, $t_k = \dfrac{k-1}{n-1}\pi$, $k = 1, \ldots, n$, and

$$l_k(x) = \frac{(-1)^{k-1}\,\omega_n(x)}{(1+\delta_{k1}+\delta_{kn})(n-1)(x-x_k)}, \quad k = 1, \ldots, n$$

with $\omega_n(x) = \sin t \sin(n-1)t$, $x = \cos t$.

THEOREM 4. *For any $f \in C[-1,1]$, $h > 0$ and $n \in \mathbf{N}$ we have*

$$\omega(L_n(f);h) \leqq c \min\left[hn \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right), \omega\left(f; \frac{1}{n}\right)\log n + \omega(f;h)\right],$$

where $c > 0$ is independent of f, n and h.

PROOF. The method will follow the ideas in the proof of Theorem 2, taking also into account that the relations in (9) hold in this case, too. Therefore let $x \in [0,1]$ be fixed (the proof in case $x \in [-1,0]$ is similar), the index j defined by $\min|x - x_j| = \min\limits_{1 \leqq k \leqq n} |x - x_k|$, and let us denote $\omega_n(x) = U_n(x)(1-x^2)$, where $U_n(x) = \dfrac{\sin(n-1)t}{\sin t}$, $x = \cos t$. By simple calculations we get

$$l_k'(x) = \frac{(-1)^{k-1}}{(1+\delta_{k1}+\delta_{kn})(n-1)(x-x_k)} \times$$
$$\times \left[\frac{U_n'(x)(1-x^2)}{x-x_k} - \frac{2xU_n(x)}{x-x_k} - \frac{U_n(x)(1-x^2)}{(x-x_k)^2}\right],$$

which immediately implies (as in the proof of Theorem 2)

$$|L_n'(f,x)| \leqq \frac{1}{n-1}\sum_{k\neq j}\omega(f;|x-x_k|)\left[\frac{|U_n'(x)|(1-x^2)}{|x-x_k|}+\right.$$

$$\left.+\frac{2|U_n(x)|}{|x-x_k|}+\frac{|U_n(x)|(1-x^2)}{(x-x_k)^2}\right]+\omega(f;|x-x_j|)|l_j'(x)|.$$

Now the Bernstein–Markov inequality yields

$$|l_j'(x)| \leqq \frac{n^2}{n\sqrt{1-x^2}+1}\|l_j(x)\| \leqq \frac{cn^2}{j}\|l_j(x)\| \leqq \frac{cn^2}{j}.$$

Therefore

$$\omega(f;|x-x_j|)|l_j'(x)| \leqq \omega\left(f;\frac{j}{n^2}\right)\frac{cn^2}{j} \leqq n^2\,\omega\left(f;\frac{1}{n^2}\right).$$

Now we will use the obvious estimates

$$|U_n(x)|(1-x^2) \leqq \sqrt{1-x^2} \sim \frac{j}{n},$$

$$|U_n'(x)|(1-x^2) \leqq 2(n-1).$$

Thus we obtain

$$\frac{1}{n-1}\sum_{k\neq j}\omega(f;|x-x_k|)\frac{|U_n'(x)|(1-x^2)}{|x-x_k|}$$

$$\leqq 2\sum_{k\neq j}\frac{\omega(f;|x-x_k|)}{|x-x_k|} \leqq cn^2\sum_{k=1}^{n}\frac{1}{k^2}\,\omega\left(f;\frac{k^2}{n^2}\right)$$

$$\leqq cn\sum_{k=1}^{n}\omega\left(f;\frac{1}{k^2}\right),$$

$$\frac{1}{n-1}\sum_{k\neq j}\omega(f;|x-x_k|)\frac{|U_n(x)|}{|x-x_k|} \leqq \sum_{k\neq j}\frac{\omega(f;|x-x_k|)}{|x-x_k|} \leqq cn\sum_{k=1}^{n}\omega\left(f;\frac{1}{k^2}\right),$$

$$\frac{1}{n-1}\sum_{k\neq j}\omega(f;|x-x_k|)\frac{|U_n(x)|(1-x^2)}{|x-x_k|} \leqq \frac{cj}{n^2}\sum_{k\neq j}\frac{\omega(f;|x-x_k|)}{(x-x_k)^2}$$

$$\leqq \frac{cj}{n^2} \sum_{k \neq j} \frac{\omega(f; |x - x_k|)}{\frac{j}{n^2}|x - x_k|} \leqq c \sum_{k \neq j} \frac{\omega(f; |x - x_k|)}{|x - x_k|} \leqq cn \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right).$$

Collecting all of these estimates we obtain

$$|L_n'(f, x)| \leqq cn \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right),$$

which, by the same reasoning as in the proof of Theorem 2, yields

$$\omega(L_n(f); h) \leqq cnh \sum_{k=1}^{n} \omega\left(f; \frac{1}{k^2}\right). \tag{12}$$

On the other hand, for $|x - y| \leqq h$ we get

$$|L_n(f, x) - L_n(f, y)| \leqq 2\|L_n(f) - f\| + \omega(f; h),$$

which implies

$$\omega(L_n(f); h) \leqq 2\|L_n(f) - f\| + \omega(f; h).$$

Standard technique in interpolation theory (see [24]) gives

$$\|L_n(f) - f\| \leqq c\omega\left(f; \frac{1}{n}\right) \|\lambda_n\|,$$

where $\lambda_n(x) := \sum_{k=1}^{n} |l_k(x)|$, $x \in [-1, 1]$, is the Lebesgue function of interpolation. Here by (9)

$$\lambda_n(x) \leqq \sum_{k=1}^{n} \left| \frac{U_n(x)(1 - x^2)}{(n-1)(x - x_k)} \right| \leqq \sum_{k \neq j} \frac{|U_n(x)|(1 - x^2)}{(n-1)|x - x_k|} + |l_j(x)|$$

$$\leqq c \sum_{k \neq j} \frac{\frac{j}{n}}{(n-1)\frac{|j^2 - k^2|}{n^2}} + |l_j(x)|$$

$$\leqq c \sum_{k \neq j} \frac{k}{|k^2 - j^2|} + |l_j(x)| \leqq c \log n + |l_j(x)| \leqq c \log n.$$

Thus

$$\omega(L_n(f); h) \leqq c\omega\left(f; \frac{1}{n}\right) \log n + \omega(f; h), \tag{13}$$

which together with (12) proves the theorem.

COROLLARY 5. (i) *If* $f \in \operatorname{Lip}\alpha$, $0 < \alpha \leqq 1$, *then for all* $n \in \mathbf{N}$ *and* $h \in (0,1)$ *we have*

$$\omega(L_n(f);h) = \begin{cases} O\left[h^{\frac{\alpha}{2-\alpha}}\left(\log\frac{1}{h}\right)^{\frac{2-2\alpha}{2-\alpha}}\right] & \text{if } 0<\alpha<\frac{1}{2}, \\ O\left(h^{1/3}\log\frac{1}{h}\right) & \text{if } \alpha=\frac{1}{2}, \\ O\left[h^{\frac{\alpha}{1+\alpha}}\left(\log\frac{1}{h}\right)^{\frac{1}{1+\alpha}}\right] & \text{if } \frac{1}{2}<\alpha\leqq 1. \end{cases}$$

(ii) *If* $\omega(f;h) = O\left(\dfrac{1}{\log^{\beta}\frac{1}{h}}\right)$, $\beta > 1$ *then*

$$\omega(L_n(f);h) = O\left(\frac{1}{\log^{\beta-1}\frac{1}{h}}\right),$$

(All the constants in "O" are independent of n and h.)

PROOF. (i) Let $f \in \operatorname{Lip}\alpha$, $0 < \alpha \leqq 1$. Then (12)–(13) yield

$$\omega(L_n(f);h) = \begin{cases} O(n^{2-2\alpha}h) & \text{if } 0<\alpha<\frac{1}{2}, \\ O(nh\log n) & \text{if } \alpha=\frac{1}{2}, \\ O(nh) & \text{if } \frac{1}{2}<\alpha\leqq 1, \end{cases} \tag{14}$$

and

$$\omega(L_n(f);h) = O\left(\frac{\log n}{n^{\alpha}} + h^{\alpha}\right), \tag{15}$$

respectively. Now if n is smaller than

$$\left(\frac{1}{h}\log\frac{1}{h}\right)^{\frac{1}{2-\alpha}}, \quad h^{-2/3}, \quad \left(\frac{1}{h}\log\frac{1}{h}\right)^{\frac{1}{1+\alpha}},$$

in the cases $0 < \alpha < 1/2$, $\alpha = 1/2$, $1/2 < \alpha \leqq 1$, respectively, then we use the corresponding estimates in (14). Otherwise we use (15).

(ii) In this case we get from (12) and (13)

$$\omega(L_n(f);h) = O\left(\frac{n^2h}{\log n}\right)$$

and

$$\omega(L_n(f);h)=O\left(\frac{1}{\log^{\beta-1}n}+\frac{1}{\log^{\beta}\frac{1}{h}}\right).$$

As before, we use these estimates according as n is smaller or bigger than $\frac{1}{\sqrt{h}}\left(\log\frac{1}{h}\right)^{\frac{1}{2}-\beta}$.

Finally, let us consider the case of the Shepard interpolatory operator

$$S_{n,\lambda}(f,x)=\frac{\sum_{k=0}^{n} f(k/n)|x-k/n|^{-\lambda}}{\sum_{k=0}^{n}|x-k/n|^{-\lambda}},\qquad \lambda\geqq 1,\ n=1,2,\ldots,$$

defined for an arbitrary $f\in C[0,1]$. Since by [23, Theorem 1 and Lemma 2] we have estimates for $\|S_{n,\lambda}-f\|$ and $\|S'_{n,\lambda}\|$, by applying the above method we immediately get

$$\omega(S_{n,\lambda}(f);h)\leqq c\min\left[hn^{2-\lambda}\int_{1/n}^{1}\frac{\omega(f;t)}{t^{\lambda}}dt,\ n^{1-\lambda}\int_{1/n}^{1}\frac{\omega(f;t)}{t^{\lambda}}dt+\omega(f;h)\right]$$

$$\leqq ch^{\lambda-1}\int_{h}^{1}\frac{\omega(f;t)}{t^{\lambda}}dt,\qquad 0<h<1<\lambda,\ n\in\mathbf{N},$$

where the constant c is independent of n. In particular, for $1<\lambda\leqq 2$, $f\in \operatorname{Lip}\alpha$, $0<\alpha\leqq 1$ we get

$$\omega(S_{n,\lambda}(f);h)=\begin{cases} O(h^{\alpha}) & \text{if } \alpha<\lambda-1,\\ O\left(h^{\alpha}\log\frac{1}{h}\right) & \text{if } \alpha=\lambda-1,\\ O(h^{\lambda-1}) & \text{if } \lambda-1<\alpha.\end{cases}$$

Now if $\lambda>2$ then by

$$\int_{h}^{1}\frac{\omega(f;t)}{t^{\lambda}}dt\leqq\frac{2\omega(f;h)}{h}\int_{h}^{1}t^{1-\lambda}\,dt\leqq ch^{1-\lambda}\,\omega(f;h)$$

we get the simpler relation

$$\omega(S_{n,\lambda}(f);h)\leqq c\,\omega(f;h),\qquad 0<h<1,\ \lambda>2,\ n\in\mathbf{N}.$$

REMARK. The above results show that when $\lambda > 2$, or $1 < \lambda \leqq 2$ and $\alpha < \lambda - 1$, then the Shepard operators completely preserve the global smoothness of f. Also, the case $\lambda = 1$ remains unsolved, since in this case Lemma 2 in [23] does not give an estimate for $\|S'_{n,\lambda}\|$.

REFERENCES

[1] ADELL, J. A. and DE LA CAL, J., Preservation of moduli of continuity for Bernstein-type operators, *Approximation, Probability and Related Fields* (Santa Barbara, CA, 1993), G. A. Anastassiou and S. T. Rado, Eds., Plenum Press, New York, 1994, 1–18. *MR* **95k**:41034

[2] ADELL, J. A. and PÉREZ-PALOMARES, A., Global smoothness preservation properties for generalized Szász–Kantorovich operators, Preprint, 1996.

[3] ADELL, J. A. and PÉREZ-PALOMARES, A., Second modulus preservation inequalities for generalized Bernstein–Kantorovich operators, *Approximation and optimization*, Vol. 1 (Cluj-Napoca, 1996), Transilvania, Cluj-Napoca, 1997, 147–156. *MR* **98** (pending)

[4] ANASTASSIOU, G. A., COTTIN, C. and GONSKA, H. H., Global smoothness of approximating functions, *Analysis* **11** (1991), 43–57. *MR* **92h**:41047

[5] BLOOM, W. R. and ELLIOTT, D., The modulus of continuity of the remainder in the approximation of Lipschitz functions, *J. Approx. Theory* **31** (1981), 59–66. *MR* **84a**:41016

[6] BARI, N. K. and STECHKIN, S. B., Best approximations and differential properties of two conjugate functions, *Trudy Moskov. Mat. Obshch.* **5** (1956), 483–522 (in Russian). *MR* **18**, 303e

[7] BERMAN, D. L., Divergence of the Hermite–Fejér interpolation process, *Uspekhi Mat. Nauk* **13(80)** (1958), 143–148 (in Russian). *MR* **20** #4126

[8] BROWN, B. M., ELLIOTT, D. and PAGET, D. F., Lipschitz constants for the Bernstein polynomials of a Lipschitz continuous function, *J. Approx. Theory* **49** (1987), 196–199. *MR* **88d**:41023

[9] CISMASIU, C. S., Some properties of the linear positive operators of probabilistic type, *Seminar on Numerical and Statistical Calculus* (Cluj-Napoca, 1987), Preprint 87–9, Univ. "Babes–Bolyai", Cluj-Napoca, 1987, 129–133. *MR* **89j**:41036

[10] CHENEY, E. W., *Introduction to approximation theory*, McGraw-Hill, New York – Toronto – London, 1966. *MR* **36** #5568

[11] COTTIN, C. and GONSKA, H. H., Simultaneous approximation and global smoothness preservation, *Proceedings of the Second International Conference in Functional Analysis and Approximation Theory* (Acquafredda di Maratea, 1992), *Rend. Circ. Mat. Palermo* (2) *Suppl.* **33** (1993), 259–279. *MR* **95f**:41035

[12] DELLA VECCHIA, B., On the preservation of Lipschitz constants for some linear operators, *Boll. Un. Mat. Ital. B* (7) **3** (1989), 125–136. *MR* **90i**:41031

[13] GAVREA, I., Preservation of Lipschitz constants by linear transformations, *Itinerant Seminar on Functional Equations, Approximation and Convexity* (Cluj-Napoca, 1988), Preprint, 88–6, Univ. "Babes–Bolyai", Cluj–Napoca, 1988, 175–182. *MR* **91e**:41030

[14] GONSKA, H. H. and CAO, JIA-DING, Approximation by Boolean sums of positive linear operators. VI. Monotone approximation and global smoothness preservation, Preprint and talk given at the Seminarul itinerant "Tiberiu Popoviciu" de Ecuatii Functionale, Aproximare si Convexitate, Universitatea "Babes–Bolyai", Cluj–Napoca, May 1993.

[15] HAJEK, O., Uniform polynomial approximation, *Amer. Math. Monthly* **72** (1965), 681.

[16] KHAN, M. K. and PETERS, M. A., Lipschitz constants for some approximation operators of a Lipschitz continuous function, *J. Approx. Theory* **59** (1989), 307–315. *MR* **90k**:41031

[17] KRATZ, W. and STADTMÜLLER, U., On the uniform modulus of continuity of certain discrete approximation operators, *J. Approx. Theory* **54** (1988), 326–337. *MR* **90a**:41019

[18] LINDVALL, T., Bernstein polynomials and the law of large numbers, *Math. Sci.* **7** (1982), 127–139. *MR* **84d**:41009

[19] MASTROIANNI, G. and SZABADOS, J., Jackson order of approximation by Lagrange interpolation, *Proceedings of the Second International Conference in Functional Analysis and Approximation Theory* (Acquafredda di Maratea, 1992), *Rend. Circ. Mat. Palermo* (2) *Suppl.* **33** (1993), 375–386. *MR* **95g**:41021

[20] NATANSON, I. P., *Constructive function theory. Vol. 1. Uniform approximation*, Frederick Ungar Publishing Co., New York, 1964. – *Vol. 2. Approximation in mean*, 1965. – *Vol. 3. Interpolation and approximation quadratures*, 1965. *MR* **33** #4529

[21] STECKIN, S. B., On the order of the best approximations of continuous functions, *Izv. Akad. Nauk SSSR* **15** (1951), 219–242 (in Russian). *MR* **13**, 29d

[22] SZABADOS, J., On the convergence and saturation problem of the Jackson polynomials, *Acta Math. Acad. Sci. Hungar.* **24** (1973), 399–406. *MR* **49** #11124

[23] SZABADOS, J., Direct and converse approximation theorems for the Shepard operator, *Approx. Theory Appl.* **7** (1991), 63–76. *MR* **93b**:41029

[24] SZABADOS, J. and VÉRTESI, P., *Interpolation of functions,* World Scientific Publishing Co., Inc., Teaneck, NJ, 1990. *MR* **92j**:41009

[25] TIMAN, A. F., *Theory of approximation of functions of a real variable*, Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1960 (in Russian). *MR* **22** #8257

[26] ZYGMUND, A. *Trigonometrical series*, Dover Publications, New York, 1955. *MR* **17**, 361e

(Received November 3, 1997)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ORADEA
STR. ARMATEI ROMANE 5
RO–3700 ORADEA
ROMANIA

galso@math.uoradea.ro

MTA RÉNYI ALFRÉD MATEMATIKAI
KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

szabados@math-inst.hu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 415–426

REGULAR POLYHEDRA AND HAJÓS POLYHEDRA

KATALIN BOGNÁR MÁTHÉ and K. BÖRÖCZKY

Dedicated to the memory of György Hajós

1. Introduction

The regular polyhedra in the d-dimensional spaces of constant curvature are solutions of various extremal problems. Many characterizations are described in the book [8] of L. Fejes Tóth and in the survey article [9] of A. Florian. In these results, the inradius or circumradius, or the number of faces, or some quermassintegral of the polyhedron is prescribed.

On a seminar of G. Hajós in 1960, a method of L. Fejes Tóth in [7] (see III. 3) related to a packing of unit circles in the Euclidean plane was discussed. The idea is to cut off the corners of a Dirichlet–Voronoi cell by a circle of radius $\frac{2}{\sqrt{3}}$. The following problem was raised by the participants: find the polygon with minimal area among the convex polygons such that the polygons contain a circle of radius r_1 and vertices of the polygons lie on a concentric circle of radius r_2 $(r_1 < r_2)$. J. Molnár [10] proved that the solution is the so called *Hajós polygon* in any plane of constant curvature: each but possibly one side touches the inner circle. Observe that the Hajós polygon is a regular polygon for suitable r_1 and r_2.

The original problem can be rephrased in the following way: find the convex polygon P with minimal area such that for some $O \in P$, the distance of O from the edges is at least r_1 and from the vertices is at least r_2. We consider d-polyhedra under similar conditions. Let $0 < r_1 \leqq \cdots \leqq r_d$ where in the case of hyperbolic space, we allow $r_d = \infty$. We say that a polyhedron P satisfies the *distance condition* $(r_1, \ldots, r_d)$ with respect to some $O \in P$ if the distance of O and any k-flat containing some k-face of P is at least r_{d-k}. If $r_d = \infty$ in the hyperbolic space then the vertices of P are ideal points.

The distance condition originates from the proof of the so called simplex bound for the density of a packing of congruent balls by C. A. Rogers [11] in

1991 *Mathematics Subject Classification.* Primary 52C17, 52A15; Secondary 52B10.
Key words and phrases. Packing, polyhedra, spacious.

The research of the second named author is supported by OTKA T016131, OTKA T017314 and the Geometry Project of the Research Developments Foundation of the Hungarian Ministry of Culture and Education (FKFP 0152/1997).

the Euclidean, and by K. Böröczky [4] in any space of constant curvature. Note that the 2-dimensional case was settled earlier by L. Fejes Tóth (see [9]) in the Euclidean, and by J. Molnár [10] in any surface of constant curvature. We say that a packing of congruent balls satisfies the *spacious condition* $(r_1, \dots, r_d)$ if each Dirichlet–Voronoi cell satisfies the distance condition $(r_1, \dots, r_d)$. Note that there exist certain optimal $(r_1, \dots, r_d)$ depending on d and the curvature of the space such that any packing of balls of radius r_1 satisfies the spacious condition $(r_1, \dots, r_d)$. The papers mentioned above used the spacious condition in order to give a lower bound for the volume of any Dirichlet–Voronoi cell, which estimate in turn resulted in the simplex bound.

In the theorems below, certain regular polyhedron is given, and r_{d-k} is defined as the distance of a k-face and the center of this regular polyhedron.

THEOREM 1. *Let Q be a regular polyhedron in a 3-dimensional space of constant curvature, where ideal vertices are allowed in the case of the hyperbolic space. Denote by r_{3-k}, $k = 0, 1, 2$, the distance of the center and a k-face of Q. Then the volume of any polyhedron P satisfying the distance condition (r_1, r_2, r_3) is at least the volume of Q, with equality only if P and Q are congruent.*

In higher dimensional spaces, we need a technical assumption on P: P satisfies the *foot condition* with respect to $O \in P$ if the orthogonal projection of O onto the k-plane containing a k-face lies in the k-face. We denote the volume of Jordan measurable sets by $V(\cdot)$.

THEOREM 2. *Let Q be a regular polyhedron in a d-dimensional space of constant curvature, $d \geqq 4$, where ideal vertices are allowed in the hyperbolic space. Denote by r_{d-k}, $k = 0, \dots, d-1$, the distance of the center O of Q and a k-face of Q.*

(i) *In the Euclidean case, if a polyhedron P satisfies the distance condition $(r_1, \dots, r_d)$ then $V(P) \geqq V(Q)$ holds with equality only if P and Q are congruent.*

(ii) *In the hyperbolic or spherical case, if a polyhedron P satisfies with respect to O both the distance condition $(r_1, \dots, r_d)$ and the foot condition for k-faces, $k \geqq 3$, then $V(P) \geqq V(Q)$ holds with equality only if P and Q are congruent.*

Probably, in some cases the distance condition in Theorem 1 can be relaxed. For example, for some regular polyhedra with triangular faces, it might be enough to assume that the bounding planes and the vertices of P are not close, and no need for condition on the edges. J. Molnár raised the following problem in the three dimensional Euclidean space: determine the convex polyhedron of minimal volume if it contains a given ball and the vertices are taken from a concentric sphere.

An edge to edge spherical tiling in S^2 whose tiles are congruent to a given spherical Hajós polygon lying in an open hemisphere is called a spherical

Hajós tiling.

Call a 3-polyhedron in a space of constant curvature *Hajós polyhedron* if the faces are congruent Hajós polygons such that each face touches the inscribed ball in the circumcenter of the face, and hence the vertices of the polyhedron are contained in some sphere. Readily, each regular polyhedron is a Hajós polyhedron. Observe that Hajós tilings are exactly the radial projections of the faces of the corresponding Hajós polyhedra.

We also define the *asymptotic Hajós polyhedron* in the hyperbolic three space, as a polyhedron whose vertices are ideal points, the polyhedron has an inscribed ball and the radial projection of the faces onto S^2 is a Hajós tiling (assuming that S^2 is concentric with the inscribed ball). Note that if the center of S^2 is given then Hajós tilings and asymptotic Hajós polyhedra are in one to one correspondence.

The natural analogue of the problem of J. Molnár asks whether among polyhedra which contain the inscribed ball of the Hajós polyhedron and whose vertices lie on the circumsphere of a Hajós polyhedron, the Hajós polyhedron itself has the minimal volume. This problem has resisted all attemps so far.

B. Bollobás characterized Hajós tilings in the 60's but his result has never been published. We provide the list below, and verify in Section 3 that no other Hajós polyhedra exist.

THEOREM 3. *If a Hajós polyhedron is not regular then its faces are triangles whose two longer sides have equal length. In this case, denote the length of a shorter side by s and of a longer side by l. In the cases* (iii), ..., (vi), *there exists a unique polyhedron for any given s and n.*

The Hajós polyhedra are:

(i) *The regular polyhedra;*

(ii) *Tetrahedra whose four edges have length l, and the other two are opposite with length s;*

(iii) *Bipyramid over a regular n-gon with sidelength s, $n \geq 5$;*

(iv) *Consider a bipyramid as in* (iii) *for even $n \geq 6$. Cut the bipyramid into two by a plane containing the top and bottom vertex, and two opposite vertices of the n-gon, and hence the section is a square. Fix one part, and rotate the square onto itself together with the other part by $\pi/2$;*

(v) *The union of an antiprism over a regular n-gon, $n \geq 6$, and two pyramids (on top and bottom). The edges with length s are the sides of the two regular n-gons.*

(vi) *Consider the polyhedron as in* (v) *for odd $n \geq 7$. Cut the surface of the polyhedron into two by a spatial hexagon whose sides are edges of length l and whose vertices are the top vertex and the bottom vertex, and two opposite vertices of both regular n-gons. This hexagon has a rotational symmetry of degree $2\pi/3$ around the suitable axis. Fix*

one part of the surface, and rotate the other part by this $2\pi/3$ *degree rotation.*

REMARK. One type of asymptotic Hajós polyhedra are naturally the asymptotic regular polyhedra. On the other hand, while the faces of an asymptotic Hajós polyhedron corresponding to any of the cases (ii), ..., (vi) are regular asymptotic triangles, no face touches the inscribed ball of the polyhedron in the center of the incircle of the face.

2. Extremality of the regular polyhedra

In this section, we work in a d-dimensional space of constant curvature, and by S we always denote some ball. The density of S in a Jordan measurable set T is defined as

$$d(T,S)=\frac{V(S\cap T)}{V(T)}.$$

An orthoscheme $A_0\dots A_d$ is a d-simplex such that for $i=1,\dots,d-1$, the subspaces determined by $A_0,\dots,A_i$ and $A_i,\dots,A_d$ are totally orthogonal, namely, any line containing A_i and lying in the first subspace is orthogonal to any line containing A_i and lying in the second subspace.

The following three lemmas are related to certain statements and methods in Böröczky [4] and Böröczky and Florian [6]. Note that in the special case $d=3$, the results of Böröczky [4] were already proved in Böröczky and Florian [6].

LEMMA 1. *Consider a ball* S *with center* A_0 *and orthoschemes* $T=A_0A_1\dots A_d$ *and* $\widetilde{T}=A_0\widetilde{A}_1\dots\widetilde{A}_d$ *where* A_d *and* $\widetilde{A}_d$ *are allowed to be ideal points in the case of the hyperbolic space. Assume that* $A_0A_i\geqq A_0\widetilde{A}_i$ *for* $i=1,\dots,d$, *and* S *does not intersect the hyperplanes* $A_1\dots A_d$ *and* $\widetilde{A}_1\dots\widetilde{A}_d$. *Then* $d(T,S)\leqq d(\widetilde{T},S)$, *and equality holds only if* T *and* $\widetilde{T}$ *are congruent.*

SKETCH OF THE PROOF. Lemma 1 basically coincides with Lemma 10 in Böröczky [4], p. 256 (note that the dimension of the space is denoted by n in that paper). The difference is that now we allow $\widetilde{A}_d$ (and hence A_d) to be an ideal point. If this is the case, the proof in [4] still carries through word by word (see Sections 2 and 3, p. 244–256), with the following change: instead of $BA_d<CA_d$, write $\angle CBA_d>\angle BCA_d$.

If A_d is ideal and we are not interested in the case of equality, the finite case yields the statement by limiting arguments. □

The limit density at a lower dimensional orthoscheme $T=A_0A_1\dots A_k$, $1\leqq k<d$, is also needed in the arguments for the proof of Theorem 2. It is defined as

$$d(T,A_k,S)=\lim_{\substack{A_i\to A_k\\ i=k+1,\dots,d}} d(A_0\dots A_kA_{k+1}\dots A_d,S),$$

where $A_0 \dots A_k A_{k+1} \dots A_d$ is assumed to be an orthoscheme.

Denote by $S(O,r)$ the ball with center O and radius r (if the space is hyperbolic and $r=\infty$ then $S(O,r)$ is naturally the whole space). Set $r = A_0A_k$ and denote by Π the $(d-k+1)$-plane containing A_{k-1} and totally orthogonal to $A_0 \dots A_{k-1}$. If σ is any $(d-k)$-dimensional Jordan measurable subset of the boundary of $S(O,r) \cap \Pi$ then

$$d(T, A_k, S) = d(\operatorname{conv}\{A_0, \dots, A_{k-1}, \sigma\}, S).$$

Lemma 2 below is Lemma 11 in Böröczky [4]. We repeat the proof for sake of completeness.

LEMMA 2. *Consider a ball S with center A_0, a k-dimensional orthoscheme $T = A_0A_1 \dots A_k$ $(k < d)$, and a d-dimensional orthoscheme $\tilde{T} = A_0\tilde{A}_1 \dots \tilde{A}_d$. Assume that $A_0A_i \geqq A_0\tilde{A}_i$ for $i = 1, \dots, k-1$, $A_0A_k \geqq A_0\tilde{A}_d$ and S does not intersect the planes $A_1 \dots A_k$ and $\tilde{A}_1 \dots \tilde{A}_d$. Then $d(T, A_k, S) < d(\tilde{T}, S)$.*

PROOF. Let

$$T_0 = A_0A_1 \dots A_{k-1}A'_k \dots A'_d$$

be an orthoscheme such that $A_0A'_i > A_0\tilde{A}_i$ for $i = k, \dots, d-1$ and $A_0A'_d = A_0A_k$. Then it follows by the definition of $d(T, A_k, S)$ and by Lemma 1 that

$$d(T, A_k, S) \leqq d(T_0, S) < d(\tilde{T}, S). \qquad \square$$

The last lemma is a rather technical one. The three-dimensional version with finite vertices was proved in Böröczky and Florian [6].

LEMMA 3. *Consider $T = \operatorname{conv}\{A_0, A_1, \dots, A_{d-3}, p\}$ with the following properties: T is a full dimensional and p is a two dimensional convex polyhedron, $A_0, A_1, \dots, A_{d-3}$ are vertices of T and if $d \geqq 4$ then the subspaces determined by $A_0, A_1, \dots, A_k$ and $A_k, \dots, A_{d-3}, p$ are totally orthogonal, $k = 1, \dots, d-3$. In addition, assume that the orthogonal projection B of A_0 onto the 2-plane of p lies outside of p, and the vertices of p have the same distance from A_0 (or each is an ideal point in the hyperbolic space). Observe that p has a unique side DE separating B from p, and denote the orthoscheme $\operatorname{conv}\{A_0, A_1, \dots, A_{d-3}, B, C, D\}$ by T_0 where C is the closest point of DE to A_0. With these assumptions, if S is any ball with center A_0 that does not intersect the $(d-1)$-plane determined by $A_1, \dots, A_{d-3}, p$ then $d(T, S) < d(T_0, S)$.*

PROOF. Note that

$$(1) \qquad T = \operatorname{conv}\{A_0, A_1, \dots, A_{d-3}, B, p\} \setminus \operatorname{conv}\{A_0, A_1, \dots, A_{d-3}, B, D, E\}.$$

For any side s_i of p different from DE, denote the closest point of s_i to A_0 by U_i. Let V_i be one of the endpoints of s_i (U_i is the midpoint of s_i

unless V_i is an ideal point). Then $\text{conv}\{A_0, A_1, \ldots, A_{d-3}, B, p\}$ can be dissected into orthoschemes which are congruent to one of $T_i = \text{conv}\{A_0, A_1, \ldots, A_{d-3}, B, U_i, V_i\}$, and $\text{conv}\{A_0, A_1, \ldots, A_{d-3}, B, D, E\}$ can be dissected into orthoschemes congruent to T_0. We deduce by $A_0C < A_0U_i$ and Lemma 1 that $d(T_i, S) < d(T_0, S)$ for $i = 1, \ldots, m$, which in turn yields the lemma by (1). □

The proof of Theorem 1 uses ideas in Böröczky and Florian [6], p. 240, where the case of special (finite) values for r_1, r_2 and r_3 was considered.

PROOF OF THEOREM 1. Based on the center O of Q, divide Q into congruent orthoschemes. Let $\widetilde{T} = O\widetilde{A}_1\widetilde{A}_2\widetilde{A}_3$ be one of these orthoschemes, where A_i sits in a face of dimension $3-i$, and hence $r_i = O\widetilde{A}_i$.

Assume that P is situated so that O has distance at least r_i from any $(3-i)$-plane spanned by some $(3-i)$-face of P. Approximating closely the part of the boundary of $S(O, r_3)$ contained in $\text{int}P$, a polyhedron $P' \subset P$ can be constructed whose vertices are on the boundary of $S(O, r_3)$ and the polyhedron still satisfies the distance condition (r_1, r_2, r_3) with respect to O. If $r_3 = \infty$ then simply take $P' = P$.

We define a tiling of P' by skew pyramids and orthoschemes, always with apex O in the following way: Let F be a face of P' and let B be the orthogonal projection of O onto the plane of F. If $B \notin F$ then we simply take the skew pyramid $\text{conv}\{O, F\}$. If $B \in F$ then dissect $\text{conv}\{O, F\}$ by orthoschemes $OA_1A_2A_3$ where $A_1 = B$, A_2 is the closest point of an edge of F to O, and A_3 is an endpoint of this edge. Observe that if r_3 is finite then A_2 is the midpoint of the corresponding edge.

Let T be a tile defined above. It follows by Lemma 1 if T is an orthoscheme, and by Lemma 3 if T is a skew pyramid that $d(T, S(O, r_1)) \leqq d(\widetilde{T}, S(O, r_1))$ with equality only if T is congruent with $\widetilde{T}$. Therefore $V(Q) \leqq V(P') \leqq V(P)$, with equality only if Q and P are congruent. □

REMARK. Even if $0 < r_1 < r_2 < r_3$ do not originate from a regular polyhedron, one can still define the orthoscheme $\widetilde{T}$ corresponding to r_1, r_2, r_3. Let P be a polyhedron satisfying the distance condition (r_1, r_2, r_3) with respect to O. The same argument as above yields that $d(P, S(O, r_1)) \leqq d(\widetilde{T}, S(O, r_1))$, which in turn results in a lower bound for the volume of P. If (r_1, r_2, r_3) is the optimal distance condition for a Dirichlet–Voronoi cell in a packing of congruent balls and P is a Dirichlet–Voronoi cell then these arguments yield the simplex bound in Böröczky and Florian [6].

In the proof of Theorem 1, we did not need the foot condition. The point is that after intersecting with $S(O, r_3)$, the foot condition holds for the edges. On the other hand, if the foot condition does not hold for a 2-face then Lemma 3 solves the problem.

Unfortunately, no analogue of Lemma 3 is known for higher dimensional faces, and so we had to assume the foot condition in Theorem 2 for k-faces with $k \geqq 3$.

The proof of Theorem 2 is based on ideas in Böröczky [4]. That paper uses special values of r_i derived from conditions on a Dirichlet–Voronoi cell P in a packing of congruent balls. In this case, a suitable analogue of the foot condition automatically holds for $P \cap S(O, r_d)$ (O is the corresponding center) because P is a Dirichlet–Voronoi cell.

PROOF OF THEOREM 2. Dissect Q into congruent orthoschemes, and denote one of them by $\widetilde{T}$. Set $S = S(O, r_1)$.

First consider the Euclidean case, and assume that O is the base point for the distance condition. For any flag $F_0 \subset \cdots \subset F_{d-1}$ of the faces of P, $\dim F_i = i$, define the *skew orthoscheme* $T = \mathrm{conv}\{O, A_1, \ldots, A_d\}$ where A_i is the closest point of F_{d-i} to O. One can dissect P into non-degenerate skew orthoschemes of the above type.

Then according to Theorem III.6 in [3], the inequality

$$d(T, S) \leqq d(\widetilde{T}, S)$$

holds for any non-degenerate skew orthoscheme T with equality if and only if T and $\widetilde{T}$ are congruent. In turn, we conclude Theorem 2 (i).

Now assume that P is a polyhedron in the hyperbolic or spherical d-space satisfying with respect to O both the distance condition $(r_1, \ldots, r_d)$ and the foot condition for k-faces, $k \geqq 3$. The idea is to dissect $P \cap S(O, r_d)$ into orthoschemes. If a face F of P intersects $\mathrm{int} S(O, r_d)$ and $\dim F \neq 2$ then the foot condition does hold for F, and hence two things cause problems: we do not have foot condition for 2-faces, and also some spherical regions bound $P \cap S(O, r_d)$. Anyway, one can still dissect $P \cap S(O, r_d)$ into a tiling where each tile is of the form $T = \mathrm{conv}\{O, A_1, \ldots, A_k, \sigma\}$. Here A_i is the closest point of the $(d-i)$-plane determined by a $(d-i)$-face F_{d-i} of P, $F_{d-k} \subset \cdots \subset F_{d-1}$, $A_i \in F_{d-i}$ and σ is a $(d-k-1)$-dimensional set lying on the relative boundary of $F_{d-k} \cap S(O, r_d)$. There are three types of tiles according to the type of σ.

(i) If σ is a point A_d then $T = OA_1 \ldots A_d$ is an orthoscheme.

(ii) σ is a convex domain in a Euclidean 2-plane, and the orthogonal projection B of A_{d-3} onto the 2-plane of σ lies outside of σ.

(iii) σ is a $(d-k-1)$-dimensional spherical Jordan measurable set lying in the relative boundary of $F_{d-k} \cap S(O, r_d)$. Observe that $T = \mathrm{conv}\{O, \sigma\}$ can occur, where σ is a $(d-1)$-dimensional spherical Jordan measurable set on the boundary of $S(O, r_d)$, and then we set $k = 0$.

Note that if $r_d = \infty$ then only cases (i) and (ii) can occur.

We claim for any tile T in $P \cap S(O, r_d)$ that

$$d(T, S) \leqq d(\widetilde{T}, S), \tag{2}$$

with equality only if T and $\widetilde{T}$ are congruent.

If T is an orthoscheme then the claim follows by Lemma 1.

Assume that T is of type (iii) (and hence r_d is finite), and choose an $A_{k+1} \in \sigma$. Then $d(T,S) = d(T, A_{k+1}, S)$, and Lemma 2 yields the claim.

Finally, assume that T is of type (ii). Since B is not in σ, there exists a side DE of σ separating B from σ. Set

$$T_0 = \text{conv}\{O, A_1, \ldots, A_{d-3}, B, C, D\},$$

where C is the point of DE closest to O. Approximating σ by polygons, Lemma 3 yields that $d(T,S) \leqq d(T_0,S)$. On the other hand, we have $OC > r_{d-1}$, and hence $d(T_0,S) < d(T,S)$ holds by Lemma 1. Therefore we conclude (2), which in turn yields the Theorem. □

K. Bognár Máthé considered polyhedra in the Euclidean 3-space satisfying certain special distance conditions. In the papers [1] and [2], she showed that the dual Archimedean semi-regular polyhedra $(3,4,3,4)$ and $(3,5,3,5)$ have minimal volumes in suitable classes of polyhedra.

3. Hajós polyhedra

The goal of the section is to prove Theorem 3.

The polyhedra described in Theorem 3 (i), ...,(v) readily exist. The existence of the last type of Hajós polyhedron (more precisely, the existence of the corresponding Hajós tiling) is established in II. 2.2.1 below.

Consider a Hajós tiling whose tiles are congruent to the spherical Hajós polygon P.

If P is a regular polygon then Euler's theorem yields that the Hajós polyhedron is a regular polyhedron. So assume that P is not a regular polygon.

Denote the length of the longer sides of P by a, the shorter side by b, the larger angles by β and the smaller angle by α. Finally, let O be the circumcenter of P, and let B_1 and B_2 be the end points of the side of length b. Since P is contained in some open hemisphere, we have

$$OB_1 = OB_2 < \frac{\pi}{2}. \tag{3}$$

The reflected image of a tile through the short side is also a tile. Call the union of the tile and the reflected image a *twin*. Since the twins also tile S^2, the number f of tiles in the tiling is even. We call the tiling determined by the twins as *twin tiling*.

Euler's theorem yields that P has at most five sides.

I *P is not a pentagon or a quadrilateral.*

Assume that P is a k-gon, $k=4,5$. Then $2\alpha+2\beta>2\pi$ holds by $\alpha<\beta$, and vertices where angle β occurs show that

$$\alpha+2\beta=2\pi, \quad \text{and hence} \quad \alpha<\frac{2\pi}{3}. \tag{4}$$

Since each side of length a of P can be seen by an angle less than $\frac{2\pi}{k-1}$ from O, we have

$$\beta>\alpha>\frac{(k-3)\pi}{k-1}. \tag{5}$$

If $k=5$ then (5) yields that each vertex of the Hajós tiling has degree three. We deduce that $f=12$, and hence

$$3\alpha+2\beta=\frac{4\pi}{12}+3\pi=\frac{10\pi}{3},$$

which in turn contradicts (4).

So assume that $k=4$. Among the angles in the tiling, we have the same number of α's and β's. We deduce by (4) that there exists a vertex, where only angles of size α lie. Now $\frac{\pi}{3}<\alpha<\frac{2\pi}{3}$ yields that either $\alpha=\frac{2\pi}{5}$ or $\alpha=\frac{\pi}{2}$. Then the angle of the triangle OB_1B_2 at B_1 is either $\pi/2$ or $3\pi/5$, respectively. We deduce that $OB_1\geqq\frac{\pi}{2}$ holds in both cases, which contradicts (3).

II *P is a triangle.*

Then P is an isosceles triangle whose two longer sides are equal.

II 1 $f=4$.

In this case, the Hajós polyhedron is some tetrahedron, which is readily the one described in Theorem 3 (ii).

II 2 $f\geqq 6$.

Instead of the Hajós tiling, from this point we consider only the twin tiling. Now the twin is a spherical rhombus with angles α and 2β, which may degenerate to a 2-gon (if $\beta=\pi/2$). Denote by $f_t=f/2$ the number of twins. Counting the area of a twin yields that

$$\alpha+2\beta=\frac{f_t+2}{f_t}\pi. \tag{6}$$

We deduce by (6) and $f_t\geqq 3$ that at any vertex only at most two 2β's meet. Call a vertex of type zero, one or two according to the number of 2β's meeting at the vertex.

The number of α's and 2β's (as the angles of twins) is the same. Thus there exists a vertex, where the number of α's is at least (or at most) the number of 2β's. Since $\pi < \alpha + 2\beta < 2\pi$, there exists a vertex of type either zero or one (where the number of α's is greater), and there exists a vertex of type two. In particular, at a vertex of type two, we have either

$$\alpha + 2 \cdot 2\beta = 2\pi \quad \text{or} \quad 2 \cdot 2\beta = 2\pi.$$

Call the vertex of a twin *sharp* (*flat*) if the angle at the vertex is α (2β).

II 2.1 $\beta = \pi/2$.

Then the rhombus is a 2-gon with angle α. Since $\alpha < \beta = \pi/2$, we deduce that $f_t \geqq 5$.

If there exists a vertex of type zero then the corresponding polyhedron is the one described in Theorem 3 (iii) (and $n = f_t$). Otherwise, there exists a vertex of type one. At this vertex, the flat vertex of a twin meets the sharp vertex of $\frac{1}{2}f_t$ other twins, which case is described in Theorem 3 (iv). Then $n = f_t$ is even, and hence $n \geqq 6$.

II 2.2 $\alpha + 4\beta = 2\pi$.

In this case, (6) yields that

$$\alpha = \frac{4}{f_t}\pi \quad \text{and} \quad \beta = \frac{f_t - 2}{2f_t}\pi.$$

Then $\alpha < \beta$ yields that $f_t > 10$.

II 2.2.1 There exists a vertex with type zero.

Let A be the vertex of type zero where $\frac{1}{2}f_t$ twins meet. The $\frac{1}{2}f_t$ neighbouring vertices of A are of type two, which are the sharp vertices of $\frac{1}{2}f_t$ additional twins. This way all twins have been enumerated, and the additional twins meet at a common vertex D. The whole arrangement has spherical rotational symmetry of angle α around A, and hence A and D are opposite points. Therefore the corresponding polyhedron is the one described in Theorem 3 (v), and $f_t = 2n$.

Assume that $\frac{1}{2}f_t = n$ is odd. Then

$$2\beta = \frac{\frac{1}{2}f_t - 1}{2}\alpha,$$

and hence the construction described in Theorem 3 (vi) can be actually performed.

II 2.2.2 There exists no vertex with type zero.

Then there exists a vertex A of type one. Since

$$\frac{\frac{1}{2}f_t+1}{2}\alpha+2\beta=2\pi,$$

we have that $\frac{1}{2}f_t=2m+1$ for some integer $m\geqq 3$, $2\beta=m\alpha$. The vertices of type two have degree three, and any vertex of type one is the sharp vertex of $m+1$ twins.

Denote by T the twin which has a flat vertex at A, and by $T_1,\dots,T_{m+1}$ the twins which have a sharp vertex at A. Let $C_1,\dots,C_m$ be the neighbours of A not contained in T. Then $C_1,\dots,C_m$ are sharp vertices of m twins, whose other sharp vertex is a common point D. The union H of these last m twins and $T_1,\dots,T_{m+1}$ is a spherical region bounded by a hexagonal line whose sides have length a and the angles are alternately either 2β or $2\pi-2\beta$.

Since the twin is determined by its angles, all the twins in H are also contained in the twin tiling constructed in II. 2.2.1, and hence A and D are opposite points on S^2.

Observe that D has type one. Thus there exists an $(m+1)^{\text{st}}$ twin which has a sharp vertex at D, and hence the other sharp vertex of this twin is a common vertex $\bar{A}$ with T. One of the twins sharing an edge with T has a sharp vertex at A and a flat vertex at $\bar{A}$. Denote this twin by $\bar{T}$. Then the whole construction based on A and T can be repeated word by word for $\bar{A}$ and $\bar{T}$, resulting in a set $\bar{H}$. The twins in $\bar{H}$ can be obtained from the twins in H by spherical reflection through the midpoint of the edge $A\bar{A}$. The hexagonal line bounding H is also the boundary of $\bar{H}$, and hence each twin is contained either in H or in $\bar{H}$. We conclude that the tiling is the one described in Theorem 3 (vi), and the spherical hexagonal line bounding H corresponds to the spatial hexagon mentioned in Theorem 3 (vi).

ACKNOWLEDGEMENT. We would like to thank Endre Makai for helpful discussions.

REFERENCES

[1] BOGNÁR MÁTHÉ, K., Über Kugelsysteme unter Geräumigkeitsbedingungen, *Studia Sci. Math. Hungar.* **28** (1993), 431–445. *MR* **95b**:52032

[2] BOGNÁR MÁTHÉ, K., Über ein halbreguläres Polyeder, *Studia Sci. Math. Hungar.* (to appear).

[3] BÖRÖCZKY, K., Gömbkitöltések állandó görbületű terekben [Sphere packing in spaces of constant curvature, in Hungarian] II, *Mat. Lapok* **26** (1975), 67–90. *MR* **58** #24015

[4] BÖRÖCZKY, K., Packing of spheres in spaces of constant curvature, *Acta Math. Acad. Sci. Hungar.* **32** (1978), 243–261. *MR* **80h**:52014

[5] BÖRÖCZKY, K., Closest packing and loosest covering, *Diskrete Geometrie* 3. Kolloq., Salzburg, 1985, 329–334. *Zbl* **563**.00013

[6] BÖRÖCZKY, K. and FLORIAN, A., Über die dichteste Kugelpackung in hyperbolischen Raum, *Acta Math. Acad. Sci. Hungar.* **15** (1964), 237–245. *MR* **28** #3369

[7] Fejes Tóth, L., *Lagerungen in der Ebene, auf der Kugel und im Raum*, Die Grundlehren der mathematischen Wissenschaften, Band 65, Springer, Berlin, 1953. *MR* **15**, 248b. Second edition in 1972. *MR* **50** #5603

[8] Fejes Tóth, L., *Regular figures*, Pergamon Press, Oxford, 1964. *MR* **29** #2705

[9] Florian, A., Extremum problems for convex discs and polyhedra, *Handbook of convex geometry*, edited by P. M. Gruber and J. M. Wills, North-Holland, Amsterdam, 1993, 177–221. *MR* **94e**:52001

[10] Molnár, J., Körelhelyezések állandó görbületű felületeken [Circle packings on surfaces of constant curvature, in Hungarian], *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **12** (1962), 223–263. *MR* **28** #1535

[11] Rogers, C. A., *Packing and covering*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 54, Cambridge Univ. Press, Cambridge, 1964. *MR* **30** #2405

(Received November 12, 1997)

YBL MIKLÓS MŰSZAKI FŐISKOLA
MATEMATIKA ÉS ÁBRÁZOLÓ GEOMETRIA TANSZÉK
THÖKÖLY ÚT 74
H–1146 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

boroc@ludens.elte.hu
boroczky@cs.elte.hu

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 427–444

ASYMPTOTIC BEHAVIOR OF POSITIVE SOLUTIONS TO NONLINEAR SINGULAR DIFFERENTIAL EQUATIONS OF SECOND ORDER

T. TANIGAWA

1. Introduction

We are concerned with positive solutions to second order differential equations with singular nonlinear terms of the type

$$(A) \qquad (p(t)|y'|^{\alpha-1}y')' = q(t)y^{-\beta}, \quad t \geqq a,$$

for which the following conditions are always assumed to hold:

(a) α and β are positive constants;

(b) $p(t)$ and $q(t)$ are positive continuous functions on $[a,\infty)$, $a \geqq 0$;

(c) $p(t)$ satisfies

$$(1.1) \qquad \int_a^\infty (p(t))^{-\frac{1}{\alpha}}\,dt = \infty.$$

By a positive solution of (A) on $J \subset [a,\infty)$ we mean a function $y : J \to (0,\infty)$ which is continuously differentiable on J together with $p|y'|^{\alpha-1}y'$ and satisfies the equation at every point of J. Our attention will be paid exclusively to the case where J is a positive half-line of the form $[t_0,\infty)$, $t_0 \geqq a$. A solution of (A) is said to be *proper* if it can be continued to ∞ and *singular* otherwise. Clearly, a singular solution must vanish at the right endpoint of its maximal interval of existence which is bounded.

There may or may not exist singular solutions of (A). It is shown (Section 2) that (A) is essentially free from singular solutions if $\beta > 1$ and that (A) does possess singular solutions if $\beta < 1$ and $\beta < \alpha$. As regards proper solutions, the equation (A) is always shown to have such solutions which may exhibit a variety of asymptotic behavior as $t \to \infty$. We classify the totality of positive proper solutions into several types according to their asymptotic behavior at infinity (Section 3), and establish conditions guaranteeing the

1991 *Mathematics Subject Classification.* Primary 34C11.

Key words and phrases. Quasilinear differential, singular nonlinearity, positive solution, asymptotic behavior.

existence of proper solutions of each of the classified types (Sections 4 and 5). As a result we are able to get fairly precise information about the structure of the solution set of (A). It is of interest to observe (Section 6) that the results for (A) can be applied to the qualitative study of spherically symmetric positive solutions to singular partial differential equations involving the m-Laplacian of the form

$$\text{(B)} \qquad \operatorname{div}(|Du|^{m-2}Du) = c(|x|)u^{-n}, \quad x \in E_a,$$

where $m > 1$ and $n > 0$ are constants, Du denotes the gradient of u in $\mathbf{R}^N$, $N \geqq 2$, $E_a \subset \mathbf{R}^N$ is the exterior of the ball centered at the origin and with radius $a > 0$, and $c : [a, \infty) \to (0, \infty)$ is a continuous function.

Differential equations with singular nonlinearities such as (A) and (B) are encountered in natural and physical sciences and there has been an increasing interest in their theoretical investigations: see, for example, the papers [1, 3–11]. A particular mention should be made of a paper by Motai and Usami [9] which is devoted to the asymptotic analysis of positive proper solutions of a special case of the equation (A) in which $p(t) \equiv 1$. The present work is designed to extend their main results to a more general case of (A) with $p(t) \not\equiv 1$ subject to (1.1) and to append a couple of propositions regarding singular solutions of (A) which are not considered in [9].

We remark that the qualitative study of differential equations involving operators of the type $(p(t)|y'|^{\alpha-1}y')'$ goes back to a pioneering paper by Elbert [2], which has been so influential as to have motivated a number of mathematicians to develop further the theory of [2] in many directions. The present paper could be considered as a work in this development.

2. Existence and nonexistence of singular solutions

We begin by showing that there is a large class of equations of the form (A) which admits no singular solutions.

THEOREM 1.1. *There is no singular solution of* (A) *if* $\beta > 1$ *and if* $p(t)$ *and* $q(t)$ *are locally of bounded variation on* $[a, \infty)$.

PROOF. It suffices to prove this theorem for the case where $p(t)$ and $q(t)$ are of class C^1 on $[a, \infty)$. Let $y(t)$ be any solution of (A) defined in some right neighborhood of $t = a$. Associated with $y(t)$ we define the function

$$V[y](t) = \frac{\alpha}{\alpha+1}(p(t))^{1+\frac{1}{\alpha}}|y'(t)|^{\alpha+1} + (p(t))^{\frac{1}{\alpha}}q(t)\frac{(y(t))^{1-\beta}}{\beta-1}.$$

A simple calculation gives

$$\frac{d}{dt}V[y](t) = \left[(p(t))^{\frac{1}{\alpha}}q(t)\right]'\frac{(y(t))^{1-\beta}}{\beta-1} \leqq \frac{\left[(p(t))^{\frac{1}{\alpha}}q(t)\right]'_+}{(p(t))^{\frac{1}{\alpha}}q(t)}V[y](t),$$

where $[f'(t)]_+ = \max\{f'(t), 0\}$, from which we have

$$V[y](t) \leqq V[y](a) \exp\left(\int_a^t \frac{[(p(s))^{\frac{1}{\alpha}} q(s)]'_+}{(p(s))^{\frac{1}{\alpha}} q(s)} ds \right) \tag{2.1}$$

in an interval of existence of $y(t)$. It follows that the function $(p(t))^{\frac{1}{\alpha}} q(t)(y(t))^{1-\beta}/(e-1)$ is bounded from above by the right-hand side of (2.1). Since $\beta > 1$ this shows that $y(t)$ cannot be zero at any finite point to the right of a, that is, $y(t)$ can be continued to $t = \infty$. This completes the proof.

A question naturally arises: Does the equation (A) with $\beta < 1$ possess a singular solution? A partial answer to this question follows.

THEOREM 1.2. *Suppose that $\beta < 1$ and $\beta < \alpha$. Then, for any $T > a$, the differential equation* (A) *possesses a singular solution having $[a, T)$ as its maximal interval of existence.*

PROOF. Let $k = \dfrac{\alpha + 1}{\alpha + \beta}$ and define the constants K_1 and K_2 by

$$K_1 = \frac{1}{k} \left(\frac{q_*/p^*}{1 - k\beta} \right)^{\frac{1}{\alpha}}, \quad K_2 = \frac{1}{k} \left(\frac{q^*/p_*}{1 - k\beta} \right)^{\frac{1}{\alpha}},$$

where we have used the notation

$$f^* = \max_{[a,T]} f(t), \quad f_* = \min_{[a,T]} f(t).$$

Defining

$$c_1 = K_1^{\frac{\alpha^2}{\alpha^2 - \beta^2}} K_2^{-\frac{\alpha\beta}{\alpha^2 - \beta^2}}, \quad c_2 = K_1^{-\frac{\alpha\beta}{\alpha^2 - \beta^2}} K_2^{\frac{\alpha^2}{\alpha^2 - \beta^2}}$$

and noting that $K_1 \leqq K_2$ implies $c_1 \leqq c_2$, we consider the set $Y \subset C[a, T]$ and the integral operator $\mathcal{F} : Y \to C[a, T]$ given by

$$Y = \{y \in C[a, T] : c_1 (T - t)^k \leqq y(t) \leqq c_2 (T - t)^k, t \in [a, T]\}$$

and

$$(\mathcal{F}y)(t) = \int_t^T \left[(p(s))^{-1} \int_s^T q(r)(y(r))^{-\beta} dr \right]^{\frac{1}{\alpha}} ds, \quad t \in [a, T].$$

It is not difficult to show that $\mathcal{F}$ is a continuous mapping which sends Y into a compact subset of Y. Thus, Schauder's fixed point theorem applies and there exists a function $y \in Y$ such that $y = \mathcal{F}y$, that is,

$$y(t) = \int_t^T \left[(p(s))^{-1} \int_s^T q(r)(y(r))^{-\beta} dr \right]^{\frac{1}{\alpha}} ds, \quad t \in [a, T].$$

Differentiation of this integral equation shows that $y(t)$ is a positive solution of (A) on $[a, T)$ and decreases to zero as $t \to T-0$. This completes the proof.

We note that the condition (1.1) with regard to $p(t)$ is not needed in the above theorems.

3. Classification of the positive proper solutions

A. *Classification.* We start with a remark that the equation (A) always has positive proper solutions. In fact, because of the presence of a negative exponent $-\beta$ in (A), it is easily seen that a solution $y(t)$ having any prescribed initial values $y(t_0) > 0$ and $y'(t_0) \geqq 0$, $t_0 \geqq a$, is increasing and can be continued to infinity.

We are interested in the variety of asymptotic behavior of positive proper solutions of (A) as $t \to \infty$. Here and in what follows extensive use will be made of the function

$$P(t) = \int_a^t (p(s))^{-\frac{1}{\alpha}} ds, \quad t \geqq a. \tag{3.1}$$

It is clear that $P(t) \to \infty$ as $t \to \infty$ because of (1.1). Let $y(t)$ be a positive solution (A) on $[t_0, \infty)$, $t_0 \geqq a$. Since (A) implies that $p(t)|y'(t)|^{\alpha-1}y'(t)$ is increasing, there are two possibilities either

$$p(t)|y'(t)|^{\alpha-1}y'(t) < 0 \quad \text{for} \quad t \geqq t_0 \tag{3.2}$$

or there is $t_1 \geqq t_0$ such that

$$p(t)|y'(t)|^{\alpha-1}y'(t) > 0 \quad \text{for} \quad t \geqq t_1. \tag{3.3}$$

Suppose that (3.2) holds. Then, $y'(t) < 0$ for $t \geqq t_0$, and $p(t)|y'(t)|^{\alpha-1}y'(t) = -p(t)(-y'(t))^{\alpha}$ increases to a nonpositive limit as $t \to \infty$. We claim that this limit is zero. If this is not the case, then there would be a constant $k > 0$ such that

$$-p(t)(-y'(t))^{\alpha} \leqq -k^{\alpha} \quad \text{or} \quad p(t)(-y'(t))^{\alpha} \geqq k^{\alpha}, \quad t \geqq t_1.$$

Then, $-y'(t) \geqq k(p(t))^{-\frac{1}{\alpha}}$, $t \geqq t_1$, and integrating this inequality, we obtain

$$y(t_1) - y(t) \geqq k \int_{t_1}^t (p(s))^{-\frac{1}{\alpha}} ds, \quad t \geqq t_1.$$

This is impossible, because the left-hand side is bounded, while the right-hand side grows to ∞ as $t \to \infty$ by (1.1). Therefore, we must have

$$\lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) = 0. \tag{3.4}$$

Concerning the limit of $y(t)$ as $t \to \infty$ there are two possibilities: either $\lim_{t\to\infty} y(t) = \text{const} > 0$ or $\lim_{t\to\infty} y(t) = 0$.

Suppose next that (3.3) holds. Then $y'(t) > 0$ for $t \geqq t_1$ and $p(t)|y'(t)|^{\alpha-1}y'(t) = p(t)(y'(t))^{\alpha}$ tends to a positive constant or grows to ∞ as $t \to \infty$, or equivalently

$$\lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) = \text{const} > 0 \tag{3.5}$$

or

$$\lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) = \infty. \tag{3.6}$$

L'Hospital's rule shows that

$$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t),$$

where $P(t)$ is given by (3.1), and so (3.5) is equivalent to

$$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \text{const} > 0, \tag{3.7}$$

and (3.6) is equivalent to

$$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \infty. \tag{3.8}$$

The above observations suggest that the following four cases are possible for the asymptotic behavior of positive proper solutions $y(t)$ of (A):

(I) $$\lim_{t\to\infty} y(t) = 0;$$

(II) $$\lim_{t\to\infty} y(t) = \text{const} > 0;$$

(III) $$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \text{const} > 0;$$

(IV) $$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \infty.$$

Solutions of types (I) and (II) are called *decaying solutions* and *uniformly positive solutions*, respectively. Solutions of types (III) and (IV) are collectively termed *growing solutions*, and those of type (IV) are referred to as *strongly growing solutions*.

B. *Integral equations.* The problem of existence of positive proper solutions of (A) belonging to the types (I)–(IV) will be discussed in detail in the next sections, where a crucial role is played by the integral equations characterizing the solutions of all of the corresponding types. Our purpose here is to derive them by direct integrations from the equation (A).

Let $y(t)$ be a proper solution of type (I) or (II) defined on $[t_0, \infty)$, $t_0 \geqq a$. Rewrite the equation (A) as

$$(-p(t)(-y'(t))^{\alpha})' = q(t)(y(t))^{-\beta}$$

and integrate it from t to ∞. Using (3.4), we obtain

$$-y'(t) = \left[(p(t))^{-1} \int_t^{\infty} q(s)(y(s))^{-\beta} ds\right]^{\frac{1}{\alpha}}, \quad t \geqq t_0, \tag{3.9}$$

which, upon one more integration over $[t, \infty)$, yields

$$y(t) = c + \int_t^{\infty} \left[(p(s))^{-1} \int_s^{\infty} q(r)(y(r))^{-\beta} dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_0, \tag{3.10}$$

where $c = \lim_{t\to\infty} y(t)$; $c = 0$ if $y(t)$ is of type (I) and $c > 0$ if $y(t)$ is of type (II). The derivation of (3.9) and (3.10) ensures that $q(t)(y(t))^{-\beta}$ and $\left[(p(t))^{-1} \int_t^{\infty} q(s)(y(s))^{-\beta} ds\right]^{\frac{1}{\alpha}}$ are integrable on $[t_0, \infty)$.

Let $y(t)$ be a proper solution of type (III) defined on $[t_1, \infty)$, $t \geqq a$. An integration of (A) rewritten as

$$(p(t)(y'(t))^{\alpha})' = q(t)(y(t))^{-\beta} \tag{3.11}$$

over $[t, \infty)$ shows that $q(t)(y(t))^{-\beta}$ is integrable in $[t_1, \infty)$ and

$$y'(t) = \left[(p(t))^{-1} \left(\omega^{\alpha} - \int_t^{\infty} q(s)(y(s))^{-\beta} ds\right)\right]^{\frac{1}{\alpha}}, \quad t \geqq t_1, \tag{3.12}$$

where

$$\omega = \lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) > 0,$$

from which we have

$$y(t) = y(t_1) + \int_{t_1}^{t} \left[(p(s))^{-1} \left(\omega^{\alpha} - \int_s^{\infty} q(r)(y(r))^{-\beta} dr\right)\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1. \tag{3.13}$$

If $y(t)$ is a type (IV)-solution of (A) on $[t_1, \infty)$, then integrating (3.11) twice from t_1 to t, we obtain

$$y'(t) = \left[(p(t))^{-1}\left(p(t_1)(y'(t_1))^{\alpha} + \int_{t_1}^{t} q(s)(y(s))^{-\beta} ds\right)\right]^{\frac{1}{\alpha}}, \quad t \geqq t_1, \tag{3.14}$$

and

$$y(t) = y(t_1) + \int_{t_1}^{t} \left[(p(s))^{-1}\left(p(t_1)(y'(t_1))^{\alpha} + \int_{t_1}^{s} q(r)(y(r))^{-\beta} dr\right)\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1, \tag{3.15}$$

of which (3.15) is the desired integral equation for $y(t)$. From (3.14) and (3.6) we see that $q(t)(y(t))^{-\beta}$ is not integrable on $[t_1, \infty)$.

4. Existence of increasing proper solutions

This section concerns the question of existence of (eventually) increasing positive proper solutions of types (III) and (IV). Our aim is to establish sharp criteria for (A) to possess positive solutions of these two types.

THEOREM 4.1. *There exists a positive proper solution $y(t)$ of* (A) *such that* $\lim_{t\to\infty} y(t)/P(t) = \infty$ *if and only if*

$$\int_{b}^{\infty} q(t)(P(t))^{-\beta} dt = \infty \quad \textit{for any } b > a. \tag{4.1}$$

THEOREM 4.2. *There exists a positive proper solution $y(t)$ of* (A) *such that* $\lim_{t\to\infty} y(t)/P(t) = \text{const} > 0$ *if and only if*

$$\int_{b}^{\infty} q(t)(P(t))^{-\beta} dt < \infty \quad \textit{for any } b > a. \tag{4.2}$$

PROOF OF THEOREM 4.1. (The "if" part.) Let $y(t)$ be a positive solution of (A) defined on $[t_1, \infty)$, $t_1 > a$, and satisfying $\lim_{t\to\infty} y(t)/P(t) = \infty$. There is a constant $k > 0$ such that $y(t) \geqq kP(t)$ for $t \geqq t_1$. Combining this inequality with the fact that $q(t)(y(t))^{-\beta}$ is not integrable on $[t_1, \infty)$ (see the remark at the end of the preceding section), we see that

$$\infty = \int_{t_1}^{\infty} q(t)(y(t))^{-\beta} dt \leqq k^{-\beta} \int_{t_1}^{\infty} q(t)(P(t))^{-\beta} dt,$$

which implies the truth of (4.1).

(The "only if" part.) Suppose that (4.1) holds. Let $t_1 > a$ be fixed and consider the solution $y(t)$ of (A) determined by the initial conditions $y(t_1) = y_1 > 0$ and $(p(t_1))^{\frac{1}{\alpha}} y'(t_1) = y_1' \geqq 0$. As remarked at the beginning of Section 3, for any such y_1 and y_1', $y(t)$ can be continued to ∞ as a growing positive solution. We claim that

$$\lim_{t\to\infty} y(t)/P(t) = \lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) = \infty. \tag{4.3}$$

Suppose the contrary. Then there is a constant $l > 0$ such that $y(t) \leqq lP(t)$ for $t \geqq t_1$, and we have from (3.14)

$$(p(t))^{\frac{1}{\alpha}} y'(t) \geqq \left[\int_{t_1}^{t} q(s)(y(s))^{-\beta} ds\right]^{\frac{1}{\alpha}} \geqq l^{-\frac{\beta}{\alpha}} \left[\int_{t_1}^{t} q(s)(P(s))^{-\beta} ds\right]^{\frac{1}{\alpha}}, \quad t \geqq t_1,$$

which shows that $q(t)(P(t))^{-\beta}$ is integrable on $[t_1, \infty)$, contradicting (4.1). Therefore, $y(t)$ must satisfy (4.3), and so it must be a strongly growing solution of (A). This completes the proof.

REMARK 4.1. It is natural to ask how fast strongly growing solutions (of type (IV)) of (A) grow as $t \to \infty$. Let $y(t)$ be such a solution on $[t_1, \infty)$, $t_1 > a$. Then in view of (3.15) we obtain

$$y(t) \geqq \int_{t_1}^{t} \left[(p(s))^{-1} \int_{t_1}^{s} q(r)(y(r))^{-\beta} dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1.$$

Since $y(t)$ is increasing, we have

$$y(t) \geqq (y(t))^{-\frac{\beta}{\alpha}} \int_{t_1}^{t} \left[(p(s))^{-1} \int_{t_1}^{s} q(r) dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1,$$

from which it follows that

$$y(t) \geqq \left\{\int_{t_1}^{t} \left[(p(s))^{-1} \int_{t_1}^{s} q(r) dr\right]^{\frac{1}{\alpha}} ds\right\}^{\frac{\alpha}{\alpha+\beta}}, \quad t \geqq t_1. \tag{4.4}$$

Let us introduce the notation for any $\tau \geqq a$

$$R(t;\tau) = \int_{\tau}^{t} \left[(p(s))^{-1} \int_{\tau}^{s} q(r) dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq \tau. \tag{4.5}$$

Then (4.4) says that $y(t)$ grows at least as fast as $[R(t;a)]^{\frac{\alpha}{\alpha+\beta}}$ as $t \to \infty$. Noting from (3.15) and (4.4) that

$$y(t) \geqq S(t) := \max\left\{\eta, [R(t;t_1)]^{\frac{\alpha}{\alpha+\beta}}\right\}, \quad t \geqq t_1,$$

where $\eta = y(t_1) > 0$, and using this inequality in (3.15), we obtain

$$y(t) \leqq \eta + \int_{t_1}^{t} \left[(p(s))^{-1}\left(\zeta^{\alpha} + \int_{t_1}^{s} q(r)(S(r))^{-\beta} dr\right)\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1, \tag{4.6}$$

where $\zeta = (p(t_1))^{\frac{1}{\alpha}} y'(t_1) > 0$, which shows that $y(t)$ grows faster as $t \to \infty$ than any constant multiple of the function

$$\int_{a}^{t} \left[(p(s))^{-1} \int_{a}^{s} q(r)(1 + R(r;a))^{-\frac{\alpha\beta}{\alpha+\beta}} dr\right]^{\frac{1}{\alpha}} ds. \tag{4.7}$$

Information about the growth of strongly growing solutions of (A) drawn above from the integral equation (3.15) is fairly sharp in that in some cases the functions (4.5) and (4.7) have the same order of growth as $t \to \infty$ (see the examples given in Section 6).

PROOF OF THEOREM 4.2. (The "only if" part.) Let $y(t)$ be a positive solution of type (III) of (A) defined on $[t_1, \infty)$, $t_1 > a$. There exists a constant $k > 0$ such that $y(t) \leqq kP(t)$ for $t \geqq t_1$. Combining this inequality with the known fact that $q(t)(y(t))^{-\beta}$ is integrable on $[t_1, \infty)$, we have

$$k^{-\beta} \int_{t_1}^{\infty} q(t)(P(t))^{-\beta} dt \leqq \int_{t_1}^{\infty} q(t)(y(t))^{-\beta} dt < \infty,$$

which implies (4.2).

(The "if" part.) Suppose that (4.2) holds. Let $t_1 > a$ be fixed and choose $k > 0$ so large that

$$\int_{t_1}^{\infty} q(t)(P(t))^{-\beta} dt \leqq 2^{-\alpha-\beta}(2^{\alpha} - 1)k^{\alpha+\beta}. \tag{4.8}$$

Consider the set $Y \subset C[t_1, \infty)$:

$$Y = \left\{y \in C[t_1, \infty) : \frac{1}{2}kP(t) \leqq y(t) \leqq kP(t),\ t \geqq t_1\right\} \tag{4.9}$$

and the mapping $\mathcal{F}: Y \to C[t_1, \infty)$

(4.10)

$$(\mathcal{F}y)(t) = \frac{1}{2}kP(t_1) + \int_{t_1}^{t} \left[(p(s))^{-1} \left(k^{\alpha} - \int_{s}^{\infty} q(r)(y(r))^{-\beta} dr \right) \right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1.$$

Clearly, Y is a closed convex subset of the Fréchet space $C[t_1, \infty)$ with the topology of uniform convergence on compact subintervals of $[t_1, \infty)$.

From (4.10) and (4.8) we see that, for any $y \in Y$ and $t \geqq t_1$,

$$\begin{aligned}(\mathcal{F}y)(t) &\leqq \frac{1}{2}kP(t_1) + k \int_{t_1}^{t} (p(s))^{-\frac{1}{\alpha}} ds \\ &\leqq k \int_{a}^{t_1} (p(s))^{-\frac{1}{\alpha}} ds + k \int_{t_1}^{t} (p(s))^{-\frac{1}{\alpha}} ds \\ &= k \int_{a}^{t} (p(s))^{-\frac{1}{\alpha}} ds = kP(t),\end{aligned}$$

and

$$\begin{aligned}(\mathcal{F}y)(t) &\geqq \frac{1}{2}kP(t_1) + \int_{t_1}^{t} \left[(p(s))^{-1} \left(k^{\alpha} - \left(\frac{k}{2} \right)^{-\beta} \int_{t_1}^{\infty} q(r)(P(r))^{-\beta} dr \right) \right]^{\frac{1}{\alpha}} ds \\ &\geqq \frac{1}{2}kP(t_1) + \frac{1}{2}k \int_{t_1}^{t} (p(s))^{-\frac{1}{\alpha}} ds = \frac{1}{2}kP(t),\end{aligned}$$

which implies that $\mathcal{F}y \in Y$, and hence $\mathcal{F}$ maps Y into itself. If $\{y_\nu\}$ is a sequence in Y converging to $y \in Y$ in $C[t_1, \infty)$, then using the Lebesgue dominated convergence theorem it can be shown without difficulty that the sequence $\{(\mathcal{F}y_\nu)(t)\}$ converges to $(\mathcal{F}y)(t)$ uniformly on any compact subinterval of $[t_1, \infty)$. This shows that $\mathcal{F}$ is a continuous mapping. Furthermore, since

$$\frac{1}{2}kP(t) \leqq (\mathcal{F}y)(t) \leqq kP(t) \quad \text{and} \quad 0 \leqq (\mathcal{F}y)'(t) = k(p(t))^{-\frac{1}{\alpha}}$$

for all $y \in Y$ and $t \geqq t_1$, we know that the set $\mathcal{F}(Y) = \{Fy : y \in Y\}$ is a relatively compact subset of $C[t_1, \infty)$. Thus, all the hypotheses of the Schauder–Tychonoff fixed point theorem are satisfied for the operator $\mathcal{F}$ acting on Y,

and so there exists an element $y \in Y$ such that $y \in \mathcal{F}y$, that is,
(4.11)

$$y(t) = \frac{1}{2}kP(t_1) + \int_{t_1}^{t} \left[(p(s))^{-1} \left(k^{\alpha} - \int_{s}^{\infty} q(r)(y(r))^{-\beta} dr \right) \right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_1.$$

Then differentiation of (4.11) leads to the conclusion that $y(t)$ is a positive solution of (A) on $[t_1, \infty)$ with the required asymptotic property $\lim_{t\to\infty} (p(t))^{\frac{1}{\alpha}} y'(t) = \lim_{t\to\infty} y(t)/P(t) = k$. This completes the proof of Theorem 4.2.

5. Decreasing proper solutions

Now we turn our attention to the set of decreasing positive proper solutions of (A) which, as mentioned in Section 3, can be partitioned into two subclasses (I) and (II).

It should be noticed that any positive decaying solution of (A) defined near ∞ can be continued as a solution over the entire interval $[a, \infty)$. Indeed, let $y(t)$ be such a solution defined on $[t_1, \infty)$, $t_1 > a$. Continue it as a solution of (A) to the left of t_1 and let J be the maximal interval of existence of the continuation, again denoted by $y(t)$. From (A) the function $(p(t))^{\frac{1}{\alpha}} y'(t)$ is increasing on J and it is negative on $[t_1, \infty)$ by hypothesis, we see that $y'(t) < 0$ on J, so that $y(t)$ is decreasing there. We claim that $J = [a, \infty)$. If $J \neq [a, \infty)$, then there must exist $t_0 > a$, the left end point of J, such that $y(t) \to \infty$ as $t \to t_0 + 0$. But this is impossible, since letting $t \to t_0 + 0$ in the equation

$$y(t) = y(t_1) + \int_{t}^{t_1} \left[(p(s))^{-1} \left(p(t_1)(-y'(t_1))^{\alpha} + \int_{s}^{t_1} q(r)(y(r))^{-\beta} dr \right) \right]^{\frac{1}{\alpha}} ds,$$

$$t_0 < t \leqq t_1,$$

which follows from (A) by direct integrations, we find that the limit $\lim_{t\to t_0+0} y(t)$ is finite. The contradiction obtained shows that J must coincide with $[a, \infty)$ as claimed, so that $y(t)$ can be continued up to a.

We are interested in getting criteria for (A) to possess positive decaying solutions of types (I) and (II). If $y(t)$ is one such solution defined on $[t_0, \infty)$, $t_0 \geqq a$, then, as pointed out in Section 3, both $q(t)(y(t))^{-\beta}$ and $\left[(p(t))^{-1} \int_{t}^{\infty} q(s)(y(s))^{-\beta} ds \right]^{\frac{1}{\alpha}}$ are integrable on $[t_0, \infty)$. This fact combined

with the inequality $y(t) \leqq y(t_0)$, $t \geqq t_0$, implies that

$$\int_a^\infty q(t)dt < \infty \quad \text{and} \quad \int_a^\infty \left[(p(t))^{-1} \int_t^\infty q(s)ds\right]^{\frac{1}{\alpha}} dt < \infty. \tag{5.1}$$

It can be proved that the condition (5.1) is also a sufficient condition for the existence of a type (II)-solution of (A).

THEOREM 5.1. *Let $c > 0$ be a given positive constant. Then there exists a positive proper solution $y(t)$ of* (A) *such that* $\lim_{t\to\infty} y(t) = \text{const} > 0$ *if and only if* (5.1) *is satisfied.*

PROOF. We need only to prove the "if" part of the theorem. Let $c > 0$ be fixed arbitrarily and choose $t_0 \geqq a$ so large that

$$\int_{t_0}^\infty \left[(p(t))^{-1} \int_t^\infty q(s)ds\right]^{\frac{1}{\alpha}} dt \leqq c^{1+\frac{\beta}{\alpha}}.$$

Define the set Y by

$$Y = \{y \in C[t_0, \infty) : c \leqq y(t) \leqq 2c,\ t \geqq t_0\}$$

and the mapping $\mathcal{F}$ by

$$(\mathcal{F}y)(t) = c + \int_t^\infty \left[(p(s))^{-1} \int_s^\infty q(r)(y(r))^{-\beta} dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_0.$$

It is easy to check that $\mathcal{F}$ is well defined on Y and sends Y into itself. Proceeding as in the proof of Theorem 4.2, we can show that $\mathcal{F}$ is a continuous mapping and that the set $\mathcal{F}(Y)$ is relatively compact in $C[t_0, \infty)$. Therefore, by the Schauder–Tychonoff fixed point theorem, there exists $y \in Y$ such that $y = \mathcal{F}y$, i.e.,

$$y(t) = c + \int_t^\infty \left[(p(s))^{-1} \int_s^\infty q(r)(y(r))^{-\beta} dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq t_0. \tag{5.3}$$

By (5.3) it is clear that $y(t)$ is a positive solution of (A) on $[t_0, \infty)$ satisfying $\lim_{t\to\infty} y(t) = c$, which completes the proof of Theorem 5.1.

As regards the positive decaying solutions (of type (I)) of (A), we have been unable to prove or disprove that (5.1) is also sufficient for their existence. A more stringent condition than (5.1) is needed for us in order to construct a desired solution as a solution of the integral equation

$$y(t) = \int_t^\infty \left[(p(s))^{-1} \int_s^\infty q(r)(y(r))^{-\beta} dr\right]^{\frac{1}{\alpha}} ds, \quad t \geqq a, \tag{5.4}$$

which is the integral equation (5.3) with $c = 0$.

THEOREM 5.2. *There exists a positive proper solution* $y(t)$ *of* (A) *such that* $\lim_{t\to\infty} y(t)=0$ *if, in addition to* (5.1), *the inequality*

$$\int_a^\infty \left[(p(s))^{-1}\int_t^\infty q(s)(Q(s))^{-\frac{\alpha\beta}{\alpha+\beta}}ds\right]^{\frac{1}{\alpha}}dt<\infty, \tag{5.5}$$

holds, where

$$Q(t)=\int_t^\infty \left[(p(s))^{-1}\int_s^\infty q(r)dr\right]^{\frac{1}{\alpha}}ds. \tag{5.6}$$

PROOF. For $n\in\mathbf{N}$ let $y_n(t)$ be a positive proper solution of (A) defined on $[a,\infty)$ and satisfying $\lim_{n\to\infty} y_n(t)=1/n$. The existence of $y_n(t)$ is ensured by Theorem 5.1 and Remark 5.1, and $y_n(t)$ satisfies

$$y_n(t)=\frac{1}{n}+\int_t^\infty \left[(p(s))^{-1}\int_s^\infty q(r)(y_n(r))^{-\beta}dr\right]^{\frac{1}{\alpha}}ds, \quad t\geqq a. \tag{5.7}$$

Noting that $y_n(t)$ is decreasing, we have from (5.7) $y_n(t)\geqq (y_n(t))^{-\frac{\beta}{\alpha}}Q(t)$, $t\geqq a$, or

$$y_n(t)\geqq (Q(t))^{\frac{\alpha}{\alpha+\beta}}, \quad t\geqq a. \tag{5.8}$$

We use (5.8) in (5.7) to obtain

$$y_n(t)\leqq\frac{1}{n}+\int_t^\infty \left[(p(s))^{-1}\int_s^\infty q(r)(Q(r))^{-\frac{\alpha}{\alpha+\beta}}dr\right]^{\frac{1}{\alpha}}ds, \quad t\geqq a, \tag{5.9}$$

and

$$|y_n'(t)|\leqq\left[(p(t))^{-1}\int_t^\infty q(s)(Q(s))^{-\frac{\alpha}{\alpha+\beta}}ds\right]^{\frac{1}{\alpha}}, \quad t\geqq a. \tag{5.10}$$

The above inequalities show that the sequence $\{y_n(t)\}$ is uniformly bounded and locally equicontinuous on $[a,\infty)$, so that there exists a subsequence $\{y_{n_k}(t)\}$ of $\{y_n(t)\}$ which converges uniformly on any compact subinterval of $[a,\infty)$. We now let $n=n_k$ in (5.7) and pass to the limit as $k\to\infty$. By means of the Lebesgue convergence theorem we conclude that the limit function $y(t)=\lim_{t\to\infty} y_{n_k}(t)$ satisfies the integral equation (5.4) for $t\geqq a$. That

$y(t) > 0$ for $t \geqq a$ is an immediate consequence of (5.8). Therefore $y(t)$ is a positive decaying solution of (A) on $[a, \infty)$. This completes the proof.

REMARK 5.1. From (5.8) and (5.9) it follows that the solution $y(t)$ constructed in Theorem 5.2 is subject to the estimates

$$(5.11) \qquad (Q(t))^{\frac{\alpha}{\alpha+\beta}} \leqq y(t) \leqq \int_t^\infty \left[(p(s))^{-1} \int_s^\infty q(r)(Q(r))^{-\frac{\alpha\beta}{\alpha+\beta}} dr \right]^{\frac{1}{\alpha}} ds, \quad t \geqq a.$$

Since (5.11) can be derived directly from (5.4) under the assumption that $y(t)$ is positive everywhere, it is a property that is possessed by all possible positive decaying solutions of (A). The accuracy of the decay estimates (5.11) will be tested by means of examples given in Section 6.

REMARK 5.2. We combine the four theorems proven above to derive useful information about the structure of the set of all positive proper solutions of (A).

(i) Suppose that $q(t)(P(t))^{-\beta}$ is not integrable on $[b, \infty)$, $b > a$. Then all positive proper solutions $y(t)$ of (A) are strongly growing: $\lim_{t\to\infty} y(t)/P(t) = \infty$.

(ii) Suppose that $q(t)$ is not integrable on $[a, \infty)$ but $q(t)(P(t))^{-\beta}$ is integrable on $[b, \infty)$, $b > a$. Then, all positive proper solutions $y(t)$ of (A) grow exactly as fast as constant multiples of $P(t)$ as $t \to \infty$: $\lim_{t\to\infty} y(t)/P(t) = \text{const} > 0$.

(iii) Suppose that $q(t)$ is integrable on $[a, \infty)$ but $\left[(p(t))^{-1} \int_t^\infty q(s)ds \right]^{\frac{1}{\alpha}}$ is not integrable on $[b, \infty)$, $b > a$. Noting that $q(t)(P(t))^{-\beta}$ is integrable on $[b, \infty)$, $b > a$, we have the same conclusion as in (ii).

(iv) Suppose that $q(t)$ is integrable on $[a, \infty)$ and $\left[(p(t))^{-1} \int_t^\infty q(s)ds \right]^{\frac{1}{\alpha}}$ is integrable on $[b, \infty)$, $b > a$. Then (A) possesses both growing and decaying positive proper solutions. Increasing solutions $y(t)$ have the property that $\lim_{t\to\infty} y(t)/P(t) = \text{const} > 0$. There always exist positive solutions of (A) which decrease to positive constants as $t \to \infty$. An additional condition is required to ensure the existence of a positive solution of (A) decaying to zero as $t \to \infty$.

6. Examples

A. The main results obtained above will be illustrated by the example

$$(6.1) \qquad (p(t)|y'|^{\alpha-1}y')' = \lambda(p(t))^{-\frac{1}{\alpha}}(P(t))^{\gamma}y^{-\beta},$$

where $\alpha > 0$, $\beta > 0$, γ and $\lambda > 0$ are constants, $p(t)$ is a positive continuous function on $[a, \infty)$ satisfying (1.1) and $P(t)$ is defined by (3.1). We consider this equation which is a special case of (A) with $q(t) = \lambda(p(t))^{-\frac{1}{\alpha}}(P(t))^{\gamma}$, in $[a', \infty)$, $a' > a$. Simple calculations show that:

$$\int_{a'}^{\infty} q(t)dt < \infty \Longleftrightarrow \gamma < -1; \tag{6.2}$$

$$\int_{a'}^{\infty} \left[(p(t))^{-1} \int_{t}^{\infty} q(s)ds \right]^{\frac{1}{\alpha}} dt < \infty \Longleftrightarrow \gamma < -\alpha - 1; \tag{6.3}$$

$$\int_{a'}^{\infty} q(t)(P(t))^{-\beta} dt < \infty \Longleftrightarrow \gamma < \beta - 1; \tag{6.4}$$

$$\int_{a'}^{\infty} \left[(p(t))^{-1} \int_{t}^{\infty} q(s)(Q(s))^{-\frac{\alpha\beta}{\alpha+\beta}} ds \right]^{\frac{1}{\alpha}} dt < \infty \Longleftrightarrow \gamma < -\alpha - 1. \tag{6.5}$$

Here $Q(t)$ stands for the function defined by (5.6).

In view of Remark 5.3 combined with the above results we have the following statements:

(i) If $\gamma \geqq \beta - 1$, then all positive proper solution $y(t)$ of (6.1) have the property that $\lim_{t\to\infty} y(t)/P(t) = \infty$. Applying Remark 4.1 to (6.1), we see that they satisfy for all sufficiently large t

$$c_1 P(t) \leqq y(t) \leqq c_2 P(t)(\log P(t))^{\frac{1}{\alpha}} \quad \text{if} \quad \gamma = \beta - 1$$

and

$$c_1 (P(t))^{\frac{\alpha+\gamma+1}{\alpha+\beta}} \leqq y(t) \leqq c_2 (P(t))^{\frac{\alpha+\gamma+1}{\alpha+\beta}} \quad \text{if} \quad \gamma > \beta - 1,$$

for some positive constants c_1 and c_2.

(ii) If $-\alpha - 1 \leqq \gamma < \beta - 1$, then all positive proper solutions $y(t)$ of (6.1) have the property that $\lim_{t\to\infty} y(t)/P(t) = \text{const} > 0$.

(iii) If $\gamma < -\alpha - 1$, then the set of positive proper solutions of (6.1) is composed of three types of solutions $y(t)$, $z(t)$, $w(t)$ with the properties that

$$\lim_{t\to\infty} \frac{y(t)}{P(t)} = \text{const} > 0, \quad \lim_{t\to\infty} z(t) = \text{const} > 0, \quad \lim_{t\to\infty} w(t) = 0,$$

respectively. Remark 5.2 specialized to (6.1) shows that there exist positive constants k_1 and k_2 such that

$$k_1 (P(t))^{\frac{\alpha+\gamma+1}{\alpha+\beta}} \leqq w(t) \leqq k_2 (P(t))^{\frac{\alpha+\gamma+1}{\alpha+\beta}}$$

for all sufficiently large t.

As a more concrete example of (6.1) we give the equation

$$(e^{-\alpha t}|y'|^{\alpha-1}y')' = \lambda e^{\mu t}y^{-\beta}, \quad t \geqq 0, \tag{6.6}$$

where μ is a constant. All positive proper solutions $y(t)$ of (6.6) satisfy $\lim_{t\to\infty} e^{-t}y(t) = \infty$ if $\mu \geqq \beta$ and $\lim_{t\to\infty} e^{-t}y(t) = \text{const} > 0$ if $-\alpha \leqq \mu < \beta$. If $\mu < -\alpha$, then (6.6) possesses three types of positive solutions $y(t)$, $z(t)$ and $w(t)$ such that $\lim_{t\to\infty} e^{-t}y(t) = \text{const} > 0$, $\lim_{t\to\infty} z(t) = \text{const} > 0$ and $\lim_{t\to\infty} w(t) = 0$. The decaying solution $w(t)$ is bounded from above and below for all large t by constant multiples of $e^{\frac{\alpha+\mu}{\alpha+\beta}t}$.

B. It can be shown that the results for (A) can be applied to the qualitative study of spherically symmetric positive solutions to partial differential equations involving the m-Laplace operator of the form

$$\operatorname{div}(|Du|^{m-2}Du) = c(|x|)u^{-n}, \quad x \in E_a, \tag{6.7}$$

where $m > 1$ and $n > 0$ are constants, $x = (x_1, \ldots, x_N) \in \mathbf{R}^N$, $N \geqq 2$, $Du = (\partial u/\partial x_1, \ldots, \partial u/\partial x_N)$, $|\cdot|$ denotes the Euclidean length of an N-vector, $E_a = \{x \in \mathbf{R}^N : |x| > a\}$, $a > 0$, and $c(t)$ is a positive continuous function on $[a, \infty)$. A spherically symmetric function $u(x) = y(|x|)$ is a solution of (6.7) if and only if $y(t)$ satisfies the ordinary differential equation

$$(t^{N-1}|y'|^{m-2}y')' = t^{N-1}c(t)y^{-n}, \quad t \geqq a, \tag{6.8}$$

which is a special case of (A) with $\alpha = m-1$, $\beta = n$, $p(t) = t^{N-1}$ and $q(t) = t^{N-1}c(t)$. The condition (1.1) is satisfied for (6.8) if and only if $m \geqq N$, in which case the function $P(t)$ given by (3.1) can be taken to be

$$P(t) = \log\frac{t}{a} \text{ if } m = N, \quad P(t) = \frac{m-1}{m-N}t^{\frac{m-N}{m-1}} \text{ if } m > N. \tag{6.9}$$

Assuming, in particular, that $m > N$, we consider the following special case of (6.7)

$$\operatorname{div}(|Du|^{m-2}Du) = |x|^l u^{-n}, \quad x \in E_a, \tag{6.10}$$

where l is a constant, the one-dimensional version of which, corresponding to (6.8), is

$$(t^{N-1}|y'|^{m-2}y')' = t^{N+l-1}y^{-n}, \quad t \geqq a. \tag{6.11}$$

As it is easily seen, (6.11) is a special case of (6.1) in which $p(t) = t^{N-1}$, $\alpha = m-1$, $\beta = n$,

$$\gamma = \frac{m(N-1)+l(m-1)}{m-N} \quad \text{and} \quad \lambda = \Big(\frac{m-N}{m-1}\Big)^{\gamma}.$$

The conditions (6.2), (6.3), (6.4) written for (6.11) then become

$$l < -N,\ l < -m \ \text{ and } \ l < -N + \frac{n(m-N)}{m-1},$$

respectively. Using this fact and applying the known results for (6.1) to (6.11), we can deduce nontrivial information about the asymptotic behavior of spherically symmetric positive solutions of the partial differential equation (6.10).

(i) If $l \geqq -N + \dfrac{n(m-N)}{m-1}$ then all symmetric positive proper solutions $u(x)$ of (6.10) have the property that

$$\lim_{|x|\to\infty} |x|^{-\frac{m-N}{m-1}} u(x) = \infty.$$

(ii) If $-m \leqq l < -N + \dfrac{n(m-N)}{m-1}$ then all symmetric positive proper solutions $u(x)$ of (6.10) have the property that

$$\lim_{|x|\to\infty} |x|^{-\frac{m-N}{m-1}} u(x) = \text{const} > 0.$$

(iii) If $l < -m$ then (6.10) has three types of spherically symmetric positive solutions $u(x)$, $v(x)$, $w(x)$ with the properties that

$$\lim_{|x|\to\infty} |x|^{-\frac{m-N}{m-1}} u(x) = \text{const} > 0, \quad \lim_{|x|\to\infty} v(x) = \text{const} > 0, \quad \lim_{|x|\to\infty} w(x) = 0,$$

respectively. The decaying solution $w(x)$ satisfies

$$k_1 |x|^{\frac{m+l}{m+n-1}} \leqq w(x) \leqq k_2 |x|^{\frac{m+l}{m+n-1}}$$

for some positive constants k_1 and k_2 and for all sufficiently large $|x|$.

REFERENCES

[1] Dalmasso, R., Solutions d'équations elliptiques semi-linéaires singulières, *Ann. Mat. Pura Appl.* (4) **153** (1988), 191–201. *MR* **90g**:35049

[2] Elbert, Á., A half-linear second order differential equation, *Qualitative Theory of Differential Equations* (Szeged, 1979), Colloq. Math. Soc. János Bolyai, **30**, North-Holland, Amsterdam – New York, 1981, 153–180. *MR* **84g**:34008

[3] Evtukhov, V. M., On the asymptotic behavior of monotone solutions of nonlinear differential equations of Emden–Fowler type, *Differentsial'nye Uravneniya* **28** (1992), 1076–1078 (in Russian). *MR* **93j**:34042

[4] Furusho, Y., Kusano, T. and Ogata, A., Symmetric positive entire solutions of second-order quasilinear degenerate elliptic equations, *Arch. Rational Mech. Anal.* **127** (1994), 231–254. *MR* **95f**:35073

[5] KIGURADZE, I. T. and SHEKHTER, B. L., Singular boundary value problems for second-order ordinary differential equations, *Current problems in mathematics. Newest results*, Vol. 30, VINITI, Moscow, 1987, 105–201 (in Russian). Translated in *J. Soviet Math.* **43** (1988), no. 2, 2340–2417. *MR* **89f**:34022

[6] KUROKIBA, M., KUSANO, T. and WANG, J., Positive solutions of second order quasilinear differential equations with singular nonlinearities, *Differentsial'nye Uravneniya* **32** (1996), 1630–1637 (in Russian). *CMP* 98, 08. Translation: *Differential Equations* **32** (1996), 1623–1629

[7] KUSANO, T. and SWANSON, C. A., Entire positive solutions of singular semilinear elliptic equations, *Japan J. Math. (N.S.)* **11** (1985), 145–155. *MR* **88b**:35070

[8] KVINIKADZE, G. G., A singular boundary value problem for nonlinear ordinary differential equations, *Ninth international conference on nonlinear oscillations*, Vol. 1 (Kiev, 1981), Naukova Dumka, Kiev, 1984, 166–168 (in Russian). *CMP* 17, 17

[9] MOTAI, M. and USAMI, H., On positive decaying solutions of singular quasilinear ordinary differential equations (preprint).

[10] TALIAFERRO, S., On the positive solutions of $y'' + \phi(t)y^{-\lambda} = 0$, *Nonlinear Anal.* **2** (1978), 437–446. *MR* **80d**:34041

[11] USAMI, H., On positive decaying solutions of singular Emden–Fowler-type equations, *Nonlinear Anal.* **16** (1991), 795–803. *MR* **92e**:34015

(Received January 7, 1998)

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF SCIENCE
FUKUOKA UNIVERSITY
8–19–1 NANAKUMA
JONAN-KU
FUKUOKA 814–0180
JAPAN

tanigawa@sf.sm.fukuoka-u.ac.jp

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 445–455

ON THE LAWS OF HOMOGENEOUS FUNCTIONALS OF THE BROWNIAN BRIDGE

P. CARMONA, F. PETIT, J. PITMAN and M. YOR

Abstract

In this note, we give a general and elementary method, which allows to compute the distributions of a large number of interesting functionals of the standard Brownian bridge.

1. Introduction

Let $(B_t; t \geq 0)$ be a Brownian motion starting from 0, and $f : \mathbb{R} \to \mathbb{R}$ a locally bounded Borel function. It is well known that the computation of the law of $\int_0^S f(B_u)du$ is more involved when S is a fixed time, t say, than when S is equal to either $T_a = \inf\{t; B_t = a\}$, or $\tau_l = \inf\{t; L_t > l\}$, where $(L_t; t \geq 0)$ denotes the local time of B at 0. An obvious "reason" for this is that the value of Brownian motion B at time t is not fixed, whereas $B_{T_a} = a$ and $B_{\tau_l} = 0$.

A classical manner to overcome the difficulty for time t is to replace t by S_λ, an independent exponential time of parameter λ, and use Feynman–Kac formula, which allows to compute:

$$\mathbb{E}\left[\exp\left(-\mu \int_0^{S_\lambda} f(B_u)du\right)\right] = \lambda \int_0^\infty e^{-\lambda t}\mathbb{E}\left[\exp\left(-\mu \int_0^t f(B_u)du\right)\right]dt$$

in terms of the solutions of a Sturm–Liouville equation (see, e.g. Jeanblanc–Pitman–Yor ([10]) for a discussion of the Feynman–Kac formula in relation with certain decompositions of Brownian paths, and/or Brownian excursion theory; see also, in the same vein, [20], exercise 4.20, chapter XII). In the course of such computations, one obtains in fact the joint Laplace transform of $\int_0^{g_{S_\lambda}} f(B_u)du$ and $\int_{g_{S_\lambda}}^{S_\lambda} f(B_u)du$, where $g_t = \sup\{u < t; B_u = 0\}$. Recalling that the standard Brownian bridge $(b_u; 0 \leq u \leq 1)$ may be represented as

1991 *Mathematics Subject Classification.* Primary 60J55, 60J60, 60J65.

Key words and phrases. Reflecting Brownian motion, Bessel processes, Ray–Knight theorems, generalized arc-sine laws.

$\left(\frac{1}{\sqrt{g_t}} B_{ug_t}; 0 \leqq u \leqq 1\right)$ whereas the standard meander may be represented as $\left(\frac{1}{\sqrt{t-g_t}} |B_{g_t+u(t-g_t)}|; 0 \leqq u \leqq 1\right)$, these computations actually yield quite some information about functionals of the Brownian bridge and Brownian meander. See, e.g., Pitman–Yor ([15], [16], [17]) for a number of results on the Brownian bridge and Brownian meander derived in this manner; see also Revuz–Yor ([20], exercise 3.8, chapter XII).

In this note, we present a different elementary method, which allows to compute the distributions of a large number of interesting functionals of the standard Brownian bridge. Again, we essentially rely upon the representation of the Brownian bridge $(b_u; 0 \leqq u \leqq 1)$ as $\left(\frac{1}{\sqrt{g}} B_{ug}; 0 \leqq u \leqq 1\right)$ which, moreover, is independent of $g = \sup\{t \leqq 1; B_t = 0\}$.

NOTATIONS. In the following, we denote by
N: a standard Gaussian variable;
Z_a: a gamma variable of parameter a;
$Z_{a,b}$: a beta variable of parameters a and b.

2. A basic identity in law

We consider $(A_t; t \geqq 0)$ an increasing process, adapted to the filtration of B and which scales jointly with B. Precisely, we assume that there exists a process $F_t(\cdot)$ on the canonical space $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ such that

$$A_t(\omega) = F_t(B(\omega)) \quad \text{and} \quad F_{ct}(w) = cF_t\left(\frac{1}{\sqrt{c}} w(c\cdot)\right),$$

for every $c > 0$, and $t \geqq 0$.

Note that, in particular, for every $c > 0$,

$$(B_{ct}, A_{ct}; t \geqq 0) \overset{(\text{law})}{=} (\sqrt{c}\, B_t, c\, A_t; t \geqq 0),$$

but our hypothesis is stronger. Let us introduce $\alpha_t \overset{\text{def}}{=} \inf\{s; A_s > t\}$, the inverse of A. All the results in this note shall be derived from the next

PROPOSITION 1.

$$A_g \overset{(\text{law})}{=} \frac{1}{\alpha_1 + B^2_{\alpha_1} \hat{T}}, \tag{1}$$

where $\hat{T}$ is a stable variable of parameter $\frac{1}{2}$, independent of the standard Brownian motion B.

PROOF. For $a \geqq 0$, we consider the set:

$$\{A_g \leqq a\} = \{g \leqq \alpha_a\} = \{1 \leqq d_{\alpha_a}\},$$

where $d_t \overset{\text{def}}{=} \inf\{u \geqq t; B_u = 0\}$.

From the scaling property of A, we deduce:

$$\{A_g \leqq a\} \overset{(\text{law})}{=} \{1 \leqq a d_{\alpha_1}\}.$$

Hence, we have: $A_g \overset{(\text{law})}{=} \dfrac{1}{d_{\alpha_1}}$, and the identity (1) follows from the strong Markov property for B, applied at time α_1. □

COROLLARY 2. *Let S be an independent exponential time of parameter $\dfrac{1}{2}$. Then, denoting $A^{(b)} = F_1(b(\cdot))$, where b is a Brownian bridge, we have*

$$(2) \qquad P\left(|N|\sqrt{A^{(b)}} \geqq x\right) = P(SA_g \geqq x^2) = \mathbb{E}\left[\exp\left(-\frac{x^2}{2}\alpha_1 - x|B_{\alpha_1}|\right)\right],$$

where N denotes a standard Gaussian variable, independent of b.

Following the previous corollary, we introduce the notion of Gauss transform $\mathcal{G}\mu$, of μ, the law of X, an $\mathbb{R}_+$-valued random variable: $\mathcal{G}\mu$ is the law of $|N|X$, where N is a standard Gaussian variable, independent of X.

The next (easy) lemma describes some important relationship between the laws of μ and $\mathcal{G}\mu$.

LEMMA 3. *The law of $|N|X$ has a density $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ which is characterized by*

$$\mathbb{E}\left[\exp\left(-\frac{\lambda^2}{2}X^2\right)\right] = \frac{1}{2}\int_{-\infty}^{+\infty} \exp(i\lambda x)\phi(|x|)dx.$$

3. Applications

EXAMPLE 1. The supremum and infimum of the Brownian bridge.

We consider $A_t \overset{\text{def}}{=} \sup_{0 \leqq s \leqq t} B_s^2$. Then, $\alpha_1 = T_1^* = \inf\{t; |B_t| = 1\}$, hence, equation (1) becomes:

$$\sup_{0 \leqq s \leqq g} B_s^2 \overset{(\text{law})}{=} \frac{1}{T_1^* + \hat{T}}$$

and Corollary 2 yields:

$$P\left(|N| \sup_{0 \leqq s \leqq 1} |b_s| \geqq x\right) = \mathbb{E}\left[\exp\left(-\frac{x^2}{2}(T_1^* + \hat{T})\right)\right] = \frac{e^{-x}}{\operatorname{ch} x}.$$

Hence,

$$P\Big(|N| \sup_{0\leq s\leq 1} |b_s| \leq x\Big) = \operatorname{th} x. \tag{3}$$

More generally, the same elementary method also yields:

$$P(|N|\sigma^+ \leq x, |N|\sigma^- \leq y) = \frac{2}{\coth x + \coth y}, \tag{4}$$

where $\sigma^{\pm} \overset{\text{def}}{=} \sup_{0\leq s\leq 1} b_s^{\pm}$. Another way to obtain (4) is presented in [20], Exercise (4.24), Chapter XII, and is based upon excursion theory arguments.

EXAMPLE 2. Lévy's uniform law.

We consider $A_t = A_t^+ \overset{\text{def}}{=} \int_0^t 1_{(B_s>0)} ds$, and we want to check Lévy's result, i.e. that $A^+(b) \overset{\text{def}}{=} \int_0^1 1_{(b_s\geq 0)} ds$ is uniform on $[0,1]$. First, recall the following relation between beta variables:

$$Z_{a,b} Z_{a+b,c} \overset{(\text{law})}{=} Z_{a,b+c},$$

where, on the left-hand side, the beta variables are independent. Since we have $A_g^+ \overset{(\text{law})}{=} gA^+(b)$ and we know that $g \overset{(\text{law})}{=} Z_{\frac{1}{2},\frac{1}{2}}$, it suffices to prove

$$A_g^+ \overset{(\text{law})}{=} Z_{\frac{1}{2},\frac{3}{2}} \tag{5}$$

to recover Lévy's uniform law. In this case, we have

$$\alpha_t = \alpha_t^+ \overset{\text{def}}{=} \inf\{s; A_s^+ > t\}.$$

Writing $t = A_t^+ + A_t^-$, we have:

$$\alpha_1^+ = 1 + A^-_{\alpha_1^+} = 1 + A^-_{\tau(2(\frac{1}{2} l_{\alpha_1^+}))} \overset{(\text{law})}{=} 1 + \Big(\frac{1}{2} l_{\alpha_1^+}\Big)^2 A^-_{\tau_2}$$

because $l_{\alpha_1^+}$ is independent of the local time process $(l_{\tau_2}^{-x}; x \geq 0)$. Then, if $\tilde{\beta}$ is a reflecting Brownian motion whose local time at 0 is denoted by $\tilde{l}$, and if $\tilde{\tau}_1$ is a stable variable of index $\frac{1}{2}$ independent of $\tilde{\beta}$, we may write:

$$(\alpha_1^+, B_{\alpha_1^+}) = (1 + \tilde{l}_1^2 \tilde{\tau}_1, \tilde{\beta}_1).$$

Then, formula (1) becomes:

$$A_g^+ \overset{(\text{law})}{=} (1+\tilde{l}_1^2\hat{\tau}_1+\tilde{\beta}_1^2\hat{T})^{-1} \overset{(\text{law})}{=} (1+(\tilde{l}_1+\tilde{\beta}_1)^2\tau^*)^{-1},$$

thanks to the additivity properties of stable variables of parameter $\frac{1}{2}$, denoting by τ^* another stable $\left(\frac{1}{2}\right)$ variable independent of the pair $(\tilde{l}_1, \tilde{\beta}_1)$. Then, noticing that:

$$(\tilde{l}_1+\tilde{\beta}_1)^2 \overset{(\text{law})}{=} 2Z_{\frac{3}{2}} \quad \text{and} \quad \tau^* \overset{(\text{law})}{=} \frac{1}{N^2} \overset{(\text{law})}{=} \frac{1}{2Z_{\frac{1}{2}}},$$

we find:

$$A_g^+ \overset{(\text{law})}{=} \frac{Z_{\frac{1}{2}}}{Z_{\frac{1}{2}}+Z_{\frac{3}{2}}} \overset{(\text{law})}{=} Z_{\frac{1}{2},\frac{3}{2}}.$$

EXAMPLE 3. Extensions to perturbed Brownian motions.

Here, we want to recover, thanks to the identity (1), the following result due to the second author (see [13] and [21], page 102, formula (8.6)):

$$A_g^{-,\mu} \overset{(\text{law})}{=} Z_{\frac{1}{2},\frac{1}{2}+\frac{1}{2\mu}},$$

where, denoting by L the local time at 0 of the Brownian motion B, we have written $A_t^{-,\mu} \overset{\text{def}}{=} \int_0^t 1_{(|B_s|\leq \mu L_s)} ds$.

By equation (1), we have:

$$A_g^{-,\mu} \overset{(\text{law})}{=} (\alpha_1^{-,\mu}+(B_{\alpha_1^{-,\mu}})^2\hat{T})^{-1}$$

and, if we denote

$$\alpha_1^{-,\mu} = 1 + A^{+,\mu}_{\alpha_1^{-,\mu}},$$

we obtain

$$A_g^{-,\mu} \overset{(\text{law})}{=} \left(1+\left(\frac{1}{2}l^{\mu}_{\alpha_1^{-,\mu}}\right)^2\tilde{T}+(B_{\alpha_1^{-,\mu}})^2\hat{T}\right)^{-1},$$

and thanks to the same arguments as in the second example, it is equivalent to prove that

$$\left(\frac{1}{2}l^{\mu}_{\alpha_1^{-,\mu}}+|B_{\alpha_1^{-,\mu}}|\right)^2 \overset{(\text{law})}{=} 2Z_{\frac{1}{2}+\frac{1}{2\mu}},$$

which is shown in Theorem 1 of [5].

EXAMPLE 4. The supremum of Brownian local times.

Here, we consider $A_t \overset{\text{def}}{=} \sup_{a\in\mathbb{R}}(l_t^a)^2$.

By Corollary 2 and [8] we have

$$P\left(|N|\sqrt{A^{(b)}} \geqq x\right) = P(SA_g \geqq x^2)$$

$$= \mathbb{E}\left[\exp\left(-\frac{x^2}{2}\alpha_1 - x|B_{\alpha_1}|\right)\right] = \left(\frac{\frac{x}{2}}{\operatorname{sh}\left(\frac{x}{2}\right)}\right)^2 \frac{\phi(2;3;x)}{\phi(1;1;x)},$$

where $\phi(a;b;z)$ is the Kummer function (also denoted $M(a,b,z)$; see [12] and [1] who use respectively the first and the second notations):

$$\phi(a;b;z) = \sum_{k\geqq 0} \frac{(a)_k}{(b)_k k!} z^k,$$

denoting $(a)_k = a(a+1)\dots(a+k-1)$. Then, we have:

$$P\left(|N| \sup_{a\in\mathbb{R}} l_1^a(b) \geqq x\right) = \frac{x - 1 + e^{-x}}{2\operatorname{sh}^2(\frac{x}{2})}.$$

REMARK 4. On the contrary, knowing $P(A^{(b)} \geqq y)$, we may recover $P(SA_g \geqq x^2)$. In the particular case where $A_t \overset{\text{def}}{=} \sup_{a\geqq 0}(l_t^a(|B|))^2$, we obtain, thanks to Theorem 8.1 of [7]:

$$P\left[|N| \sup_{a\geqq 0} l_1^a(|b|) \leqq x\right] = P\left[\sqrt{2Z_1} \sup_{a\geqq 0} l_g^a(|B|) \leqq x\right] = \operatorname{th}\left(\frac{x}{4}\right).$$

EXAMPLE 5. Successive heights for the Brownian bridge.

For $t > 0$, let us consider the sequence

$$M_1(t) \overset{\text{def}}{=} \sup_{0\leqq s\leqq t} |B_s| > M_2(t) > \dots > M_k(t) > \dots$$

of ranked heights of excursions of the absolute value of Brownian motion B up to time t. Let us define $(M_k^{(b)})_{k\geqq 1}$ the analogous quantity for the standard Brownian bridge. We consider $A_t \overset{\text{def}}{=} (M_n(t))^2$ for some $n \geqq 1$. Then denoting

$$\alpha_1^{(n)} \overset{\text{def}}{=} \inf\{s; (M_n(s))^2 > 1\} = \inf\{s; M_n(s) > 1\},$$

we have, thanks to Corollary (2):

$$P[|N|M_n^{(b)} \geqq x] = P(SA_g \geqq x^2)$$

$$= \mathbb{E}\left[\exp\left(-\frac{x^2}{2}\alpha_1^{(n)} - x|B_{\alpha_1^{(n)}}|\right)\right] = e^{-x}\mathbb{E}\left[\exp\left(-\frac{x^2}{2}\alpha_1^{(n)}\right)\right].$$

PROPOSITION 5. *If X is a random variable, we denote $X_{(k)}$ the sum of k independent variables distributed as X. We have:*

$$\alpha_1^{(n)} \overset{(\text{law})}{=} T^*_{(n)} + \tilde{T}_{(n-1)}, \tag{6}$$

where:

both variables on the right-hand side are assumed to be independent;

$T^* \overset{(\text{law})}{=} \inf\{s>0; |B_s|=1\}$;

$\tilde{T} \overset{(\text{law})}{=} \inf\{s>0; B_s=1\}$.

As a consequence of identity (6), we have:

$$\begin{aligned} P[|N|M_n^{(b)} \geqq x] &= e^{-x}\mathbb{E}\left[\exp\left(-\frac{x^2}{2}T^*\right)\right]^n \mathbb{E}\left[\exp\left(-\frac{x^2}{2}\tilde{T}\right)\right]^{n-1} \\ &= e^{-x}\left(\frac{1}{\operatorname{ch} x}\right)^n (e^{-x})^{n-1} = (1-\operatorname{th} x)^n. \end{aligned} \tag{7}$$

Thus, we recover the one-dimensional marginal results given in [18] about the Markovian sequence $(|N|M_k^{(b)})_{k\geqq 1}$. Unfortunately, our method does not extend easily to yield multi-dimensional distributions.

PROOF OF PROPOSITION 5. We introduce the two following sequences of stopping times:

$$T_1^{(0)} = T_0^{(0)} = 0,$$

and for any integer $k \geqq 0$,

$$T_1^{(k+1)} = \inf\{s > T_0^{(k)}; |B_s| = 1\} \quad \text{and} \quad T_0^{(k+1)} = \inf\{s > T_1^{(k+1)}; B_s = 0\}.$$

Then clearly:

$$\alpha_1^{(n)} = T_1^{(n)} = \sum_{k=0}^{n-1}(T_1^{(k+1)} - T_0^{(k)}) + \sum_{k=1}^{n-1}(T_0^{(k)} - T_1^{(k)}).$$

Identity (6) immediately follows, thanks to the strong Markov property of Brownian motion. □

Similarly, we may define the sequence $(M_k^{(b)+})_{k\geqq 1}$ associated with the successive positive heights of excursions. Formula (7) now becomes:

$$P[|N|M_n^{(b)+} \geqq x] = \exp(-2nx). \tag{8}$$

Further results in the same vein may be deduced from [18], where excursion theory arguments are used.

4. Extension to δ-dimensional Bessel process

The same work may be done, replacing the standard Brownian motion $(B_t)_{t\geq 0}$ by a δ-dimensional Bessel process $(R_t^{(\delta)})_{t\geq 0}$, with $\delta < 2$. First we recall the following lemma, which follows from the well-known time reversal result between $R^{(\delta)}$ and $R^{(4-\delta)}$ (for the identity in law (ii), see Getoor ([9])):

LEMMA 6. *Let us denote $\hat{T}^{(\delta)}$ the first time a δ-dimensional Bessel process starting from 1 reaches 0. Then, we have*

$$\hat{T}^{(\delta)} \underset{\text{(i)}}{\overset{\text{(law)}}{=}} \Lambda^{(4-\delta)} \overset{\text{def}}{=} \sup\{t; R_t^{(4-\delta)} = 1\} \underset{\text{(ii)}}{\overset{\text{(law)}}{=}} \frac{1}{2\, Z_{1-\frac{\delta}{2}}}.$$

Then, the same arguments as in Proposition 1 give, with obvious notations:

$$A_g \overset{\text{(law)}}{=} \left(\alpha_1 + (R_{\alpha_1}^{(\delta)})^2 \hat{T}^{(\delta)}\right)^{-1}. \tag{9}$$

Let us denote $(r_t^{(\delta)})_{0\leq t\leq 1}$ the δ-dimensional Bessel bridge, Z_μ an independent Gamma variable of parameter $\mu = 1 - \dfrac{\delta}{2}$. Since

$$g \overset{\text{def}}{=} \sup\{t \leq 1; R_t^{(\delta)} = 0\} \overset{\text{(law)}}{=} Z_{\mu,1-\mu}$$

we have, denoting by S an independent exponential time of parameter $\dfrac{1}{2}$:

$$\begin{aligned} P\left(\sqrt{2Z_\mu}\,\sqrt{A^{(r^{(\delta)})}} \geq x\right) &= P(SA_g \geq x^2) = \mathbb{E}\left[\exp\left(-\frac{x^2}{2A_g}\right)\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{x^2}{2}(\alpha_1 + R_{\alpha_1}^2 \hat{T}^{(\delta)})\right)\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{x^2}{2}\left(\alpha_1 + R_{\alpha_1}^2 \frac{1}{2Z_\mu}\right)\right)\right]. \end{aligned} \tag{10}$$

In the special case where $A_t \overset{\text{def}}{=} \sup\limits_{0\leq s\leq t} (R_s^{(\delta)})^2$, we obtain:

$$P\left(\sqrt{2Z_\mu} \sup_{0\leq s\leq 1} r_s^{(\delta)} \geq x\right) = \mathbb{E}\left[\exp\left(-\frac{x^2}{2} T_1^{(\delta)}\right)\right] \mathbb{E}\left[\exp\left(-\frac{x^2}{4Z_\mu}\right)\right], \tag{11}$$

where $T_1^{(\delta)}$ is the first hitting time of 1 by a δ-dimensional Bessel process starting from 0, so that (see [15] for example):

$$\begin{aligned} P\left(\sqrt{2Z_\mu} \sup_{0\leq s\leq 1} r_s^{(\delta)} \geq x\right) &= \frac{x^{-1+\frac{\delta}{2}}}{2^{-1+\frac{\delta}{2}}\Gamma(\frac{\delta}{2}) I_{-1+\frac{\delta}{2}}(x)} \frac{x^{1-\frac{\delta}{2}} K_{1-\frac{\delta}{2}}(x)}{2^{-\frac{\delta}{2}}\Gamma(1-\frac{\delta}{2})} \\ &= \frac{2K_\mu(x)}{\Gamma(\mu)\Gamma(1-\mu) I_{-\mu}(x)}. \end{aligned} \tag{12}$$

Thus, we recover the following result (see [15], [17] and [16]):

$$P\Big(\sqrt{2\,Z_\mu}\sup_{0\leqq s\leqq 1} r_s^{(\delta)} \leqq x\Big) = \frac{I_\mu(x)}{I_{-\mu}(x)}, \tag{13}$$

a formula which is closely related to Kiefer's series expansions for the law of $\Big(\sup_{0\leqq s\leqq 1} r_s^{(\delta)}\Big)$ (see Kiefer [11], and again [15] and [16]).

REMARK 7. It is noteworthy that the ratio on the right-hand side of (13) also occurs in the following:

$$P_x[T_0^{(\delta)} > t / R_t^{(\delta)} = y] = \frac{I_\mu(z)}{I_{-\mu}(z)}, \quad \text{where } z = \frac{xy}{t}$$

and $T_0^{(\delta)} = \inf\{t; R_t^{(\delta)} = 0\}$.

REMARK 8. In the case $\delta = 1$, we recover (3) in Example 1 above:

$$P\Big(\sqrt{2Z_{\frac{1}{2}}}\sup_{0\leqq s\leqq 1} |b_s| \leqq x\Big) = \operatorname{th} x,$$

i.e. $\sup_{a\geqq 0} l_1^a(|b|) \overset{(\text{law})}{=} 4 \sup_{0\leqq s\leqq 1} |b_s|$, thanks to Remark 4. Thus, from equation (8.1) of [7], we recover the well-known identity:

$$\frac{1}{2}\sup_{0\leqq s\leqq 1} m_s \overset{(\text{law})}{=} \sup_{0\leqq s\leqq 1} |\dot{b}_s|,$$

where $(m_s; 0 \leqq s \leqq 1)$ is a Brownian meander.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A., *Handbook of mathematical functions, with formulas, graphs and mathematical tables*, Fifth printing, with corrections, National Bureau of Standards Applied Mathematics Series, Vol. 55, National Bureau of Standards, Washington, DC, 1966. *MR* **34** #8607. Dover Publications, Inc., New York, 1966. *MR* **34** #8606; see also *MR* **31** #1400

[2] BIANE, PH. and YOR, M., Sur la loi des temps locaux browniens pris en un temps exponentiel, *Séminaire de Probabilités, XXII*, Lecture Notes in Math., 1321, Springer-Verlag, Berlin – New York, 1988, 454–466. *MR* **90e**:60103

[3] BIANE, PH. and YOR, M., Valeurs principales associées aux temps locaux browniens, *Bull. Sci. Math.* (2) **111** (1987), 23–101. *MR* **88g**:60188

[4] CARMONA, PH., PETIT, F. and YOR, M., Some extensions of the arcsine law as partial consequences of the scaling property of Brownian motion, *Probab. Theory Related Fields* **100** (1994), 1–29. *MR* **95f**:60089

[5] CARMONA, PH., PETIT, F. and YOR, M., An identity in law involving reflecting Brownian motion, derived from generalized arc-sine laws for perturbed Brownian motions, *Stochastic Process. Appl.* (1999), 323–334.

[6] CHAUMONT, L. and DONEY, R. A., Pathwise uniqueness for perturbed versions of Brownian motion and reflected Brownian motion, *Probab. Theory Related Fields* **113** (1999), 519–534.

[7] CSÖRGŐ, M., SHI, Z. and YOR, M., Some asymptotic properties of the local time of the uniform empirical process, *Bernoulli* **6** (1999) (to appear).

[8] EISENBAUM, N., Un théorème de Ray–Knight lié au supremum des temps locaux browniens, *Probab. Theory Related Fields* **87** (1990), 79–95. *MR* **92a**:60168

[9] GETOOR, R. K., The Brownian escape process, *Ann. Probab.* **7** (1979), 864–867. *MR* **80h**:60102

[10] JEANBLANC, M., PITMAN, J. and YOR, M., The Feynman–Kac formula and decomposition of Brownian paths, *Mat. Appl. Comput.* **16** (1997), 27–52. *MR* **98f**:60156

[11] KIEFER, J., K-sample analogues of the Kolmogorov–Smirnov and Cramér–v. Mises tests, *Ann. Math. Statist.* **30** (1959), 420–447. *MR* **21** #1668

[12] LEBEDEV, N. N., *Special functions and their applications*, Dover Publications, Inc., New York, 1972. *MR* **50** #2568

[13] PETIT, F., Sur le temps passé par le mouvement brownien au dessus d'un multiple de son supremum, et quelques extensions de la loi de l'arcsinus, Ph. D. Thesis, Université Paris VII, 1992.

[14] PETIT, F., Quelques extensions de la loi de l'arcsinus, *C. R. Acad. Sci. Paris Sér. I Math.* **315** (1992), 855–858. *MR* **93g**:60176

[15] PITMAN, J. W. and YOR, M., Random Brownian scaling and splicing of Bessel processes, *Ann. Probab.* **26** (1998), 1683–1702.

[16] PITMAN, J. W. and YOR, M., Homogeneous functionals of Brownian motion, 1999 (in preparation).

[17] PITMAN, J. W. and YOR, M., Quelques identités en loi pour les processus de Bessel, Hommage à P. A. Meyer et J. Neveu, *Astérisque* (1996), No. 236, 249–276. *MR* **98c**:60106

[18] PITMAN, J. W. and YOR, M., Ranked functionals of Brownian excursions, *C. R. Acad. Sci. Paris Sér. I Math.* **326** (1998), 93–97.

[19] PITMAN, J. W. and YOR, M., Laplace transforms related to excursions of a one-dimensional diffusion, *Bernoulli* **5** (1999), 249–255.

[20] REVUZ, D. and YOR, M., *Continuous martingales and Brownian motion*, 2nd edition, Grundlehren der mathematischen Wissenschaften, Band 293, Springer-Verlag, Berlin, 1994. *MR* **95h**:60072

[21] YOR, M., *Some aspects of Brownian motion. Part I. Some special functionals*, Lectures in Mathematics ETH Zürich, Birkhäuser-Verlag, Basel, 1992. *MR* **93i**:60155

[22] YOR, M., *Some aspects of Brownian motion. Part II. Some recent martingale problems*, Lectures in Mathematics ETH Zürich, Birkhäuser-Verlag, Basel, 1997. *MR* **98e**:60140

[23] YOR, M., Some remarks about the joint law of Brownian motion and its supremum, *Séminaire de Probabilités XXXI*, Lecture Notes in Math., 1655, Springer, Berlin, 1997, 306–314.

(Received July 15, 1998)

LABORATOIRE DE STATISTIQUE ET PROBABILITÉS
UNIVERSITÉ PAUL SABATIER
118 ROUTE DE NARBONNE
F–31062 TOULOUSE Cedex 4
FRANCE

carmona@cict.fr

LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES
UNIVERSITÉ PARIS VI
TOUR 56, 3° ÉTAGE
4 PLACE JUSSIEU
F–75252 PARIS Cedex 05
FRANCE

fpe@ccr.jussieu.fr

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
367 EVANS HALL
BERKELEY, CA 94720–3860
U.S.A.

pitman@stat.berkeley.edu

LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES
UNIVERSITÉ PARIS VI
TOUR 56, 3° ÉTAGE
4 PLACE JUSSIEU
F–75252 PARIS Cedex 05
FRANCE

secret@proba.jussieu.fr

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 457–474

PATH DECOMPOSITIONS OF A BROWNIAN BRIDGE RELATED TO THE RATIO OF ITS MAXIMUM AND AMPLITUDE

J. PITMAN and M. YOR

Abstract

We give two new proofs of Csáki's formula for the law of the ratio $1-Q$ of the maximum relative to the amplitude (i.e. the maximum minus minimum) for a standard Brownian bridge. The second of these proofs is based on an absolute continuity relation between the law of the Brownian bridge restricted to the event $(Q \leqq v)$ and the law of a process obtained by a Brownian scaling operation after back-to-back joining of two independent three-dimensional Bessel processes, each started at v and run until it first hits 1. Variants of this construction and some properties of the joint law of Q and the amplitude are described.

1. Introduction

In his study of asymptotic distributions arising from empirical processes in non-parametric statistics, Smirnov [25] showed that the formula

$$(1)\quad \mathbf{P}(I \leqq a, M \leqq b) = \sum_{k=-\infty}^{\infty} \exp(-2k^2(a+b)^2) - \sum_{k=-\infty}^{\infty} \exp(-2[b+k(a+b)]^2)$$

for $a, b \geqq 0$ defines the joint distribution of a pair of non-negative random variables (I, M). Doob [11] showed that (I, M) may be constructed as

$$I := -\inf_{0 \leqq u \leqq 1} b_u \quad \text{and} \quad M := \sup_{0 \leqq u \leqq 1} b_u,$$

where $(b_u, 0 \leqq u \leqq 1)$ is a standard Brownian bridge. Besides the many applications of this law of (I, M) in the theory of empirical processes (for which see Shorack and Wellner [24, §2.2]), this law is of interest on account of some of its remarkable properties which can be found scattered in the probabilistic

1991 *Mathematics Subject Classification.* Primary 60J65.

Key words and phrases. Williams' decomposition, range, three-dimensional Bessel process, Brownian scaling.

*Research supported in part by N.S.F. Grant DMS 97-03961.

0081-6906/99/$ 5.00

literature. To quickly recall some of these properties, the asymptotic distribution of the Kolmogorov–Smirnov statistic is that of the absolute maximum $I \vee M = \sup_{0 \leq u \leq 1} |b_u|$, which can be read from (1) as

$$\mathbf{P}(I \vee M \leq b) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 b^2). \tag{2}$$

As explained by Vervaat's [27] construction of a Brownian excursion from Brownian bridge, the law of the maximum of a standard Brownian excursion found by Kennedy [16] and Chung [9] is identical to the law of $I + M$, known as the *amplitude* or *range* of the bridge, whose distribution is given by the formula [12]

$$\mathbf{P}(I + M > b) = 2 \sum_{k=1}^{\infty} (4k^2 b^2 - 1) \exp(-2k^2 b^2) \tag{3}$$

for $b \geq 0$. See also [2] for a survey of transformations related to Vervaat's construction. As observed by Chung [9], the distribution of $I \vee M$ is characterized by the Laplace transform

$$\mathbf{E} \exp\Big(-\frac{1}{2}\lambda^2 (I \vee M)^2\Big) = \frac{\frac{\pi}{2}\lambda}{\sinh(\frac{\pi}{2}\lambda)}, \tag{4}$$

while that of $I + M$ is characterized by the companion formula

$$\mathbf{E} \exp\Big(-\frac{1}{2}\lambda^2 (I + M)^2\Big) = \left(\frac{\frac{\pi}{2}\lambda}{\sinh(\frac{\pi}{2}\lambda)}\right)^2. \tag{5}$$

Consequently, the law of $(I + M)^2$ equals the law of the sum of two independent copies of $(I \vee M)^2$. For $x \geq 0$, $y > 0$ let $T^{(3)}_{x,y}$ denote the first hitting time of y by a $\mathrm{BES}^{(3)}_x$ process $(R^{(3)}_{x,t}, t \geq 0)$, that is a three-dimensional Bessel process started at x, which may be constructed as $R^{(3)}_{x,t} := \sqrt{(x + B_{1,t})^2 + B^2_{2,t} + B^2_{3,t}}$ where the $(B_{i,t}, t \geq 0)$ for $i = 1, 2, 3$ are three independent standard Brownian motions started at 0. It is well known that for $y > 0$

$$\mathbf{E} \exp\Big(-\frac{1}{2}\lambda^2 T^{(3)}_{0,y}\Big) = \frac{y\lambda}{\sinh(y\lambda)} \tag{6}$$

so the identities (4) and (5) amount to the equalities in distribution

$$(I \vee M)^2 \overset{d}{=} T^{(3)}_{0,\pi/2} \quad \text{and} \quad (I + M)^2 \overset{d}{=} T^{(3)}_{0,\pi/2} + \widehat{T}^{(3)}_{0,\pi/2}, \tag{7}$$

where $\widehat{T}^{(3)}_{0,\pi/2}$ is an independent copy of $T^{(3)}_{0,\pi/2}$. As far as we know there is still no satisfying explanation in terms of Brownian paths for these remarkable identities found by Chung. For further discussion of these results, their relation to the functional equations satisfied by the Jacobi theta and Riemann theta functions, and various applications, see [5, 4, 30].

Let $Q := I/(I+M)$. Csáki [10, Theorem 2] deduced from (1) a fairly complicated expression for $\mathbf{P}(I+M<u, Q<v)$, from which he obtained by letting $u \to \infty$ the remarkable formula [10, (2.12)]

$$\mathbf{P}(Q \leqq v) = 2v^2(1-v)\sum_{n=1}^{\infty} \frac{1}{n^2 - v^2} = (1-v)(1-\pi v \cot(\pi v)) \tag{8}$$

for $0 < v < 1$. Section 2 of this paper presents a novel approach to Csáki's formula (8) via the alternative expression

$$\mathbf{P}(Q \leqq v) = 2v^2(1-v)\int_0^{\infty} d\lambda \left(\frac{\sinh(v\lambda)}{v\sinh(\lambda)}\right)^2. \tag{9}$$

By (6), for $T^{(3)}_{v,1}$ the hitting time of 1 by a $\mathrm{BES}^{(3)}_v$ process, there is the standard formula

$$\mathbf{E}\exp\left(-\frac{1}{2}\lambda^2 T^{(3)}_{v,1}\right) = \frac{\sinh(v\lambda)}{v\sinh(\lambda)} \tag{10}$$

so if we let $\widehat{T}^{(3)}_{v,1}$ denote an independent copy of $T^{(3)}_{v,1}$, and set

$$T^*_v := T^{(3)}_{v,1} + \widehat{T}^{(3)}_{v,1}$$

then

$$\mathbf{E}\exp\left(-\frac{1}{2}\lambda^2 T^*_v\right) = \left(\frac{\sinh(v\lambda)}{v\sinh(\lambda)}\right)^2. \tag{11}$$

In Section 3, the appearance of this quantity as the integrand in (9) is explained in terms of the path decomposition at the maximum for the Brownian bridge, deduced as in Pitman–Yor [20] from Williams' [28] path decomposition at the maximum for a one-dimensional diffusion. The path decomposition of the bridge at its maximum allows the law of the bridge restricted to the event $(Q \leqq v)$ to be constructed by a random Brownian scaling operation from a back-to-back joining of the paths of two independent $\mathrm{BES}^{(3)}_v$ processes run until their first hits of 1. In Section 4 we deduce some corollaries of this result involving the joint law of M and Q. Section 5 presents a more refined result, which gives an explicit description of the law of the

bridge conditioned on $Q = q$ for an arbitrary $q \in [0,1]$. We note in particular that in the limiting case $q = 0$ this conditional distribution on $C[0,1]$ is absolutely continuous with respect to the law of a standard Brownian excursion, with a density factor at $\omega \in C[0,1]$ that is proportional to $(\sup_{0 \leq u \leq 1} \omega_u)^2$. In Section 6 we present some further identities involving the local time of the bridge at 0 up to time 1. Finally, Section 7 records some basic properties of the distribution of Q determined by Csáki's formula (8).

2. A derivation of Csáki's formula

Let $|N|$ denote the absolute value of a standard Gaussian variable N, so

$$\mathbf{P}(|N| \leq x) = \int_0^x \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}y^2} dy$$

and assume that N is independent of the bridge $(b_t, 0 \leq t \leq 1)$. Our starting point is the formula

$$\mathbf{P}(|N| I \leq x, |N| M \leq y) = \frac{2}{\coth x + \coth y}, \tag{12}$$

which we have discussed already in [23, Ex. (4.24) of Chapter XII]. See also [8, 22]. As shown by Perman and Wellner [18], the Smirnov–Doob formula (1) can be deduced from (12) by inversion of Laplace transforms. But since

$$Q := \frac{I}{I+M} = \frac{|N| I}{|N| I + |N| M} \tag{13}$$

we can proceed directly from (12) to the distribution of Q, without consideration of Laplace transforms. Easily from (12), for $x, y \geq 0$

$$\mathbf{P}(|N| I \leq x, |N| M \in dy) = \frac{2 \sinh^2(x) dy}{\sinh^2(x+y)} \tag{14}$$

which combined with (13) gives

$$\mathbf{P}(Q \leq v) = \mathbf{P}\left(|N| I \leq \frac{v}{(1-v)} |N| M\right) \tag{15}$$

$$= \int_0^\infty dy \frac{2 \sinh^2(\frac{vy}{1-v})}{\sinh^2(\frac{y}{1-v})} \tag{16}$$

$$= (1-v)\int_0^\infty d\lambda \frac{2\sinh^2(v\lambda)}{\sinh^2(\lambda)}, \tag{17}$$

so we have arrived at formula (9). To complete the proof of Csáki's formula (8), it only remains to check the identity

$$\int_0^\infty d\lambda \frac{2\sinh^2(v\lambda)}{\sinh^2(\lambda)} = 1 - \pi v \cot(\pi v). \tag{18}$$

But after expanding

$$2\sinh^2(v\lambda) = \cosh(2v\lambda) - 1 = \sum_{n=1}^\infty \frac{(2v)^{2n}}{(2n)!}\lambda^{2n}$$

the identity (18) follows easily from the classical identities [13, 3.523.2]

$$\int_0^\infty d\lambda \frac{\lambda^{2n}}{\sinh^2(\lambda)} = \pi^{2n}|B_{2n}| \quad (n = 1, 2, \dots), \tag{19}$$

where B_m is the mth Bernoulli number, and [13, 1.411.7]

$$\sum_{n=1}^\infty \frac{2^{2n}|B_{2n}|}{(2n)!}x^{2n} = 1 - x\cot x \quad (|x| < \pi). \tag{20}$$

3. Path decomposition at the maximum

We start by formulating the path decomposition of the Brownian bridge at its maximum in terms of the following construction, which we adapt from [28, 29, 19, 5, 20]. See also [21] for variations of this construction and [14, 15, 26] for other decompositions of the Brownian path involving the range process and $\mathrm{BES}^{(3)}$ pieces.

CONSTRUCTION 1. Given two continuous path processes with random finite lifetimes, each with initial value 0 and final value z, say $R := (R(t), 0 \leqq t \leqq \eta)$ and $(\widehat{R} := (\widehat{R}(t), 0 \leqq t \leqq \widehat{\eta})$ with $R(\eta) = \widehat{R}(\widehat{\eta}) = z$, construct a random element r of $C[0,1]$, say

$$r := (r(u), 0 \leqq u \leqq 1) := \mathrm{BRIDGE}\left[(R(t), 0 \leqq t \leqq \eta); (\widehat{R}(t), 0 \leqq t \leqq \widehat{\eta})\right]$$

with $r(0) = r(1) = 0$ by first pasting R and $\widehat{R}$ back to back and then transforming the resulting path by Brownian scaling to have lifetime 1; that is

$$r(u) := \begin{cases} \zeta^{-1/2} R(u\zeta) & \text{if } 0 \leqq u \leqq V \\ \zeta^{-1/2} \widehat{R}((1-u)\zeta) & \text{if } V \leqq u \leqq 1, \end{cases} \tag{21}$$

where $\zeta := \eta + \widehat{\eta}$ and $V := \eta/\zeta$.

In the following applications, η and $\widehat{\eta}$ will be the first hitting times of some level $z > 0$ by the processes R and $\widehat{R}$, respectively. Then V is evidently the a.s. unique time at which r attains its maximum level, so V is a measurable function of r with

$$\sup_{0 \leqq u \leqq 1} r(u) = r(V) = z\zeta^{-1/2}$$

and R and $\widehat{R}$ can then be recovered from r via the formulae

$$\begin{aligned} \zeta &= z^2/r^2(V) \\ (R(t), 0 \leqq t \leqq \eta) &= (zr(t/\zeta)/r(V), 0 \leqq t \leqq V\zeta) \\ (\widehat{R}(t), 0 \leqq t \leqq \widehat{\eta}) &= (zr(1 - t/\zeta)/r(V), 0 \leqq t \leqq (1-V)\zeta). \end{aligned}$$

So the joint distribution of $(R, \widehat{R})$ determines the distribution of $r := \mathrm{BRIDGE}[R; \widehat{R}]$, and vice versa.

THEOREM 2. *Let $(b_u, 0 \leqq u \leqq 1)$ be a standard Brownian bridge, and let*

$$(b^*_u, 0 \leqq u \leqq 1) := \mathrm{BRIDGE}\left[(B_t, 0 \leqq t \leqq \sigma_1); (\widehat{B}_t, 0 \leqq t \leqq \widehat{\sigma}_1)\right], \tag{22}$$

where $(B_t, 0 \leqq t \leqq \sigma_1)$ and $(\widehat{B}_t, 0 \leqq t \leqq \widehat{\sigma}_1)$ are two independent copies of a standard Brownian motion started at 0 and run until its first hitting time of 1. Then for every non-negative measurable function F defined on the path space $C[0,1]$ there is the identity

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1)] = \sqrt{2\pi}\, \mathbf{E}[F(b^*_u, 0 \leqq u \leqq 1) M^*], \tag{23}$$

where

$$M^* := \sup_{0 \leqq u \leqq 1} b^*_u = 1/\sqrt{\sigma_1 + \widehat{\sigma}_1}. \tag{24}$$

PROOF. Copy the proof of [20, Theorem 3.1] in dimension 1, with the one-dimensional Bessel process $(|B_t|, t \geqq 0)$ replaced by $(B_t, t \geqq 0)$. □

COROLLARY 3. *Fix $0 < v < 1$ and let*

$$(\widetilde{b}_{v,u}, 0 \leqq u \leqq 1) := \mathrm{BRIDGE}\left[(R^{(3)}_{v,t} - v, 0 \leqq t \leqq T^{(3)}_{v,1}); (\widehat{R}^{(3)}_{v,t} - v, 0 \leqq t \leqq \widehat{T}^{(3)}_{v,1})\right], \tag{25}$$

where $(R^{(3)}_{v,t}, 0 \leqq t \leqq T^{(3)}_{v,1})$ *and* $(\widehat{R}^{(3)}_{v,t}, 0 \leqq t \leqq \widehat{T}^{(3)}_{v,1})$ *are two independent copies of a* $BES^{(3)}_v$ *process run until its first hitting time of* 1. *Then for every non-negative measurable function* F *defined on the path space* $C[0,1]$ *there is the identity*

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1)1(Q \leqq v)] = \sqrt{2\pi}\, v^2 \mathbf{E}\left[F(\widetilde{b}_{v,u}, 0 \leqq u \leqq 1)\widetilde{M}_v\right], \tag{26}$$

where

$$\widetilde{M}_v := \sup_{0 \leqq u \leqq 1} \widetilde{b}_{v,u} = \frac{(1-v)}{\sqrt{T^*_v}} \quad \textit{with} \quad T^*_v := T^{(3)}_{v,1} + \widehat{T}^{(3)}_{v,1}. \tag{27}$$

PROOF. In (23) replace $F(\cdots)$ by

$$F(\cdots)1(Q \leqq v) = F(\cdots)1(I/M \leqq a) \text{ where } v = a/(a+1),\ a = v/(1-v)$$

to see that

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1)1(Q \leqq v)] = \sqrt{2\pi}\, \mathbf{E}[F(b^*_u, 0 \leqq u \leqq 1)M^* 1(G_a)] \tag{28}$$

for G_a the event

$$G_a := (I_{\sigma_1} > -a) \cap (\widehat{I}_{\widehat{\sigma}_1} > -a),$$

where $I_t := \inf_{0 \leqq u \leqq t} B_u$ and hats indicate corresponding variables defined in terms of the other independent Brownian motion. Since $\mathbf{P}(G_a) = (a/(a+1))^2 = v^2$, formula (28) can be recast as

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1)1(Q \leqq v)] = \sqrt{2\pi}\, v^2 \mathbf{E}[F(b^*_u, 0 \leqq u \leqq 1)M^* \mid G_a]. \tag{29}$$

But conditionally on G_a the processes $(B_t, 0 \leqq t \leqq \sigma_1)$ and $(\widehat{B}_t, 0 \leqq t \leqq \widehat{\sigma}_1)$ are two independent copies of Brownian motion started at 0 and run until its hitting time of 1, with conditioning to hit 1 before $-a$. By mapping the interval $[-a, 1]$ linearly to $[0,1]$, and scaling time by a factor of $(a+1)^2 = 1/(1-v)^2$, these two processes can be constructed from two independent copies of Brownian motion started at v and run until its hitting time of 1, with conditioning to hit 1 before 0. As shown by Williams [28], such a conditioned Brownian motion is a copy of $(R^{(3)}_{v,t}, 0 \leqq t \leqq T^{(3)}_{v,1})$. Thus the processes $(B_t, 0 \leqq t \leqq \sigma_1)$ and $(\widehat{B}_t, 0 \leqq t \leqq \widehat{\sigma}_1)$ given G_a are distributed like two independent copies of the process

$$\left(\frac{R^{(3)}_{v,t(1-v)^2} - v}{1-v}, 0 \leqq t \leqq \frac{T^{(3)}_{v,1}}{(1-v)^2} \right). \tag{30}$$

Thus (29) holds with the process $(b_t^*, 0 \leqq t \leqq 1)$ conditioned on G_a replaced by $(\widetilde{b}_t, 0 \leqq t \leqq 1)$ defined as in (25), and with the density factor M^* in (29) replaced by the corresponding quantity defined in terms of the $\mathrm{BES}_v^{(3)}$ processes, that is

$$\widetilde{M}_v := \frac{1}{\sqrt{(T_v^*)/(1-v)^2}} = \frac{(1-v)}{\sqrt{T_v^*}},$$

and these substitutions in (29) yield (26). □

As a check on formula (26), we note that the previous formula (23) is recovered from (26) in the limit as $v \uparrow 1$. To see this, observe that as $v \uparrow 1$ the distribution of the process in (30) converges to that of $(B_t, 0 \leqq t \leqq \sigma_1)$, and hence the distribution of the process $(\widetilde{b}_{v,u}, 0 \leqq u \leqq 1)$ converges to that of $(b_u^*, 0 \leqq u \leqq 1)$. For a discussion of the limiting case of (26) as $v \downarrow 0$, see the end of Section 5.

4. Some consequences of the path decomposition

If in (26) we take $F(b_u, 0 \leqq u \leqq 1) = M^{-1} f(M)$ with $M := \sup_{0 \leqq u \leqq 1} b_u$ as before, and f an arbitrary non-negative Borel function, then we deduce from (26) that

$$\mathbf{E}(M^{-1} f(M) 1(Q \leqq v)) = \sqrt{2\pi}\, v^2 \mathbf{E}\left[f\left(\frac{1-v}{\sqrt{T_v^*}} \right) \right], \tag{31}$$

where the distribution of T_v^* is determined by the Laplace transform (11). In particular, for arbitrary real r

$$\mathbf{E}(M^r 1(Q \leqq v)) = \sqrt{2\pi}\, v^2 (1-v)^{r+1} \mathbf{E}((T_v^*)^{-(r+1)/2}). \tag{32}$$

For any non-negative random variable X there is the formula

$$\mathbf{E}(X^{-p}) = \frac{2^{1-p}}{\Gamma(p)} \int_0^\infty d\lambda \lambda^{2p-1} \mathbf{E} \exp\left(-\frac{1}{2}\lambda^2 X \right) \quad (p > 0) \tag{33}$$

obtained by application of Fubini's theorem. So (32) combined with (11) yields

(34)

$$\mathbf{E}(M^r 1(Q \leqq v)) = \sqrt{2\pi}(1-v)^{r+1} \frac{2^{\frac{1-r}{2}}}{\Gamma(\frac{r+1}{2})} \int_0^\infty d\lambda \lambda^r \left(\frac{\sinh(v\lambda)}{\sinh(\lambda)} \right)^2 \quad (r > -1).$$

This formula determines the distribution of M restricted to the event $(Q \leqq v)$ by a Mellin transform. In the special case $r = 0$ we recover from (34) the alternative form (9) of Csáki's formula (8).

By another application of formulae (31) and (11), we deduce the following characterization of the law of M restricted to the event $(Q \leqq v)$: for all real ξ and $0 < v < 1$

$$\text{(35)} \quad \mathbf{E}\left[\frac{1}{M}\exp\left(-\frac{\xi^2}{2M^2}\right)1(Q \leqq v)\right] = \sqrt{2\pi}\,\frac{\sinh^2(\xi v/\bar{v})}{\sinh^2(\xi/\bar{v})} \quad \text{where } \bar{v} := (1-v).$$

5. Conditioning the bridge on Q

Formulae for various conditional expectations given $Q = v$ are obtained by differentiating formulae of the previous section with respect to v. For instance, in the special case $r = -1$ formula (32) simplifies to give for $0 \leqq v \leqq 1$

$$\text{(36)} \qquad \mathbf{E}[M^{-1}1(Q \leqq v)] = \sqrt{2\pi}\,v^2$$

and hence by differentiation

$$\text{(37)} \qquad \mathbf{E}(M^{-1} \mid Q = v) = 2\sqrt{2\pi}\,v/f_Q(v),$$

where, by application of Csáki's formula (8),

$$\text{(38)} \qquad f_Q(v) := \mathbf{P}(Q \in dv)/dv = \frac{d}{dv}(1-v)(1-\pi v\cot(\pi v)).$$

See Section 8 for further discussion of this density. By differentiation of formula (35) we obtain for $0 < v < 1$, with $\bar{v} := 1 - v$,

$$\text{(39)} \qquad \mathbf{E}\left[\frac{1}{M}\exp\left(-\frac{\xi^2}{2M^2}\right)1(Q \in dv)\right] = \frac{2\sqrt{2\pi}\,\xi}{\bar{v}^2}\,\frac{\sinh(\xi v/\bar{v})\sinh(\xi)}{\sinh^3(\xi/\bar{v})}dv.$$

If we apply this formula with $\lambda := \xi/\bar{v}$ and v replaced by q then in terms of the amplitude

$$A := I + M = M/(1-Q)$$

we deduce the simpler formula

$$\text{(40)} \qquad \mathbf{E}\left[\frac{1}{A}\exp\left(-\frac{\lambda^2}{2A^2}\right) \;\middle|\; Q = q\right] = \frac{2\sqrt{2\pi}\,\lambda\sinh(\lambda q)\sinh(\lambda(1-q))}{f_Q(q)\sinh^3(\lambda)}$$

for $0 < q < 1$. In view of (6) and (10) this can be interpreted as follows. Let

$$T_q := T^{(3)}_{0,1} + T^{(3)}_{q,1} + T^{(3)}_{1-q,1},$$

where $T^{(3)}_{x,y}$ is as before the first hitting time of y by a $\mathrm{BES}^{(3)}_x$ process, and we now assume that the three random times $T^{(3)}_{0,1}$, $T^{(3)}_{q,1}$, and $T^{(3)}_{1-q,1}$ are independent. Then from (6) and (10) we have

$$\mathbf{E}\exp\left(-\frac{1}{2}\lambda^2 T_q\right) = \frac{\lambda\sinh(\lambda q)\sinh(\lambda(1-q))}{q(1-q)\sinh^3(\lambda)}. \tag{41}$$

Let

$$A_q := 1/\sqrt{T_q}. \tag{42}$$

Then (41) allows (40) to be rewritten

$$\mathbf{E}\left[\frac{1}{A}\exp\left(-\frac{\lambda^2}{2A^2}\right)\,\middle|\, Q=q\right] = \frac{2\sqrt{2\pi}\,q(1-q)}{f_Q(q)}\mathbf{E}\left[\exp\left(-\frac{\lambda^2}{2A_q^2}\right)\right]. \tag{43}$$

It now follows by uniqueness of Laplace transforms that for an arbitrary non-negative Borel function g and $0<q<1$ there is the identity

$$\mathbf{E}[g(A)\,|\,Q=q] = \frac{2\sqrt{2\pi}\,q(1-q)}{f_Q(q)}\mathbf{E}(A_q g(A_q)). \tag{44}$$

That is to say, the conditional density of A at a given $Q=q$ is identical to $af_{A_q}(a)/\mathbf{E}(A_q)$, where f_{A_q} is the density of $A_q := 1/\sqrt{T_q}$. In particular, by (44) for $g=1$,

$$\mathbf{E}(A_q) = \frac{1}{2\sqrt{2\pi}}\frac{f_Q(q)}{q(1-q)}. \tag{45}$$

Formula (61) gives bounds which imply that $\mathbf{E}(A_q)$ lies in the interval (0.19, 0.22) for all $q\in(0,1)$. In view of (41), (42) and (33) for $p=1/2$, we see that (45) amounts to the identity

$$\int_0^\infty d\lambda \frac{\lambda\sinh(\lambda q)\sinh(\lambda(1-q))}{\sinh^3(\lambda)} = \frac{f_Q(q)}{4}. \tag{46}$$

This identity can also be deduced by integration of (40) with respect to $d\lambda$. Since

$$4\int_0^v dq\,\lambda\sinh(\lambda q)\sinh(\lambda(1-q)) = 2\lambda v\cosh(\lambda)-\sinh(\lambda)+\sinh(\lambda-2v\lambda) \tag{47}$$

the identity (46) is in turn equivalent to

$$\int_0^\infty d\lambda \frac{(2\lambda v\cosh(\lambda)-\sinh(\lambda)+\sinh(\lambda-2v\lambda))}{\sinh^3(\lambda)} = P(Q\leqq v) \tag{48}$$

as given by Csáki's formula (8). We were able to confirm this by symbolic integration using *Mathematica*.

The above discussion invites an interpretation in terms of a path decomposition of the bridge conditioned on $Q = q$. Such an interpretation is provided by the following corollary of Theorem 2, which extends the previous formula (44) from an identity of one-dimensional distributions to an identity of distributions on the path space $C[0,1]$.

Fix $q \in (0,1)$. Take three independent $\mathrm{BES}^{(3)}$ processes, with starting levels 0, q and $1-q$, say R_0, R_q and $\widehat{R}_{1-q}$, whose hitting times of 1 are $T^{(3)}_{0,1}$, $T^{(3)}_{q,1}$ and $\widehat{T}^{(3)}_{1-q,1}$. Define a continuous path $S := (S(w), 0 \leqq w \leqq T_q)$, starting at q at time 0, and ending at q at time $T_q := T^{(3)}_{0,1} + T^{(3)}_{q,1} + T^{(3)}_{1-q,1}$, by concatenation of the three paths

$$(R_q(t), 0 \leqq t \leqq T^{(3)}_{q,1}),$$

$$\left(R_0(T^{(3)}_{0,1} - u), 0 \leqq u \leqq T^{(3)}_{0,1}\right),$$

$$\left(1 - R_{1-q}(T^{(3)}_{1-q,1} - v), 0 \leqq v \leqq T^{(3)}_{1-q,1}\right).$$

Let $(b^{\dagger}_{q,u}, 0 \leqq u \leqq 1)$ be the process derived from S by the Brownian scaling operation $b^{\dagger}_{q,u} := (S(uT_q) - q)/\sqrt{T_q}$. So by construction, $(b^{\dagger}_{q,u}, 0 \leqq u \leqq 1)$ is a process starting and ending at 0 whose amplitude is $A_q := 1/\sqrt{T_q}$ as above, with the feature that the process attains its maximum value before its minimum.

COROLLARY 4. *Let $\rho_{\min}$ denote the a.s. unique time that the Brownian bridge $(b_u, 0 \leqq u \leqq 1)$ attains its minimum on $[0,1]$, and $\rho_{\max}$ the corresponding time for the maximum. Then for every non-negative measurable function F defined on the path space $C[0,1]$ there is the identity*

$$\begin{aligned} &\mathbf{E}[F(b_u, 0 \leqq u \leqq 1) \mid \rho_{\max} < \rho_{\min}, Q = q] \\ &= \frac{2\sqrt{2\pi}\, q(1-q)}{f_Q(q)} \mathbf{E}\left[F(b^{\dagger}_{q,u}, 0 \leqq u \leqq 1) A_q\right], \end{aligned} \tag{49}$$

where $A_q := 1/\sqrt{T_q}$ is the amplitude of $(b^{\dagger}_{q,u}, 0 \leqq u \leqq 1)$.

PROOF. By application of (23), and the definition of conditional expectations, we deduce that for $0 < q < 1$

$$\begin{aligned} &\mathbf{E}[F(b_u, 0 \leqq u \leqq 1) \mid \rho_{\max} < \rho_{\min}, Q = q] \\ &= \frac{\mathbf{E}[F(b^*_u, 0 \leqq u \leqq 1) M^* \mid \rho^*_{\max} < \rho^*_{\min}, Q^* = q]}{\mathbf{E}[M^* \mid \rho^*_{\max} < \rho^*_{\min}, Q^* = q]}, \end{aligned}$$

where M^*, $\rho^*_{\max}$, $\rho^*_{\min}$ and Q^* are M, $\rho_{\max}$, $\rho_{\min}$ and Q evaluated for $(b^*_u, 0 \leqq u \leqq 1)$ instead of $(b_u, 0 \leqq u \leqq 1)$. In particular, by construction $M^* = 1/\sqrt{\sigma_1 + \widehat{\sigma}_1}$. Now, from the construction of $(b^*_u, 0 \leqq u \leqq 1)$, we see that the event $(\rho^*_{\max} < \rho^*_{\min}, Q^* = q)$ is identical to the event $(I_{\sigma_1} > c, \widehat{I}_{\widehat{\sigma}_1} = c)$ where $c/(c+1) = q$, $c = q/(1-q)$. With this conditioning, the process $(B_u, 0 \leqq u \leqq \sigma_1)$ becomes a Brownian motion run until it first reaches 1, conditioned to reach 1 before reaching $-c$, while the process $(\widehat{B}_u, 0 \leqq u \leqq \widehat{\sigma}_1)$ is a Brownian motion run until it first reaches 1 and conditioned on $\inf_{0 \leqq u \leqq \widehat{\sigma}_1} \widehat{B}_u = -c$. According to Williams' path decomposition at the minimum [28], the latter process can be constructed by concatenation of two $\mathrm{BES}^{(3)}$ pieces. After rescaling as in the proof of Corollary 3 these two fragments are represented by the second two paths in the concatenation of three paths which defines the process S, and the argument is completed similarly to the proof of Corollary 3. □

Note that due to the invariance of the bridge under time reversal, the event $(\rho_{\max} < \rho_{\min})$ appearing above is an event of probability $1/2$ that is independent of the pair (Q, A). Corollary 4 combined with this remark provides an explicit description of the unique family of conditional distributions for $(b_u, 0 \leqq u \leqq 1)$ given $Q = q$ that is weakly continuous in q for $q \in [0, 1]$. In particular, the law of $(b_u, 0 \leqq u \leqq 1)$ given $Q = 0$ is obtained either by letting $q \downarrow 0$ in Corollary 4, or by conditioning on $(Q \leqq v)$ and letting $v \downarrow 0$ in Corollary 3. (See formulae (58) and (59) for the required asymptotics of $f_Q(v)$ and $\mathbf{P}(Q \leqq v)$ as $v \downarrow 0$.) Let $(\widetilde{b}_{0,u}, 0 \leqq u \leqq 1)$ be the process defined by formula (25) for $v = 0$. That is, $(\widetilde{b}_{0,u}, 0 \leqq u \leqq 1)$ is constructed by putting back-to-back two independent copies of a $\mathrm{BES}^{(3)}_0$ run until its first hit of 1, then Brownian scaling to obtain lifetime 1. Then formula (26) implies that

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1) \mid Q = 0] = \frac{3\sqrt{2\pi}}{\pi^2} \mathbf{E}\left[F(\widetilde{b}_{0,u}, 0 \leqq u \leqq 1)\widetilde{M}_0\right], \tag{50}$$

where

$$\widetilde{M}_0 := \sup_{0 \leqq u \leqq 1} \widetilde{b}_{0,u} = \frac{1}{\sqrt{T^*_0}} \quad \text{with} \quad T^*_0 := T^{(3)}_{0,1} + \widehat{T}^{(3)}_{0,1}. \tag{51}$$

It is known [7] that the law of $(b_u, 0 \leqq u \leqq 1)$ given $I = 0$, defined similarly as a weak limit, is the law of a standard Brownian excursion (or $\mathrm{BES}^{(3)}$ bridge), as determined by [20, Theorem 3.1 with $\delta = 3$],

$$\mathbf{E}[F(b_u, 0 \leqq u \leqq 1) \mid I = 0] = \sqrt{\frac{\pi}{2}} \mathbf{E}\left[F(\widetilde{b}_{0,u}, 0 \leqq u \leqq 1)(\widetilde{M}_0)^{-1}\right]. \tag{52}$$

Thus the limit in distribution as $v \downarrow 0$ of $(b_u, 0 \leqq u \leqq 1)$ given $(Q \leqq v)$, as determined by (50), is not the same as the limit in distribution as $v \downarrow 0$ of

$(b_u, 0 \leqq u \leqq 1)$ given $(I \leqq v)$, as determined by (52), despite the identity of the events $(Q = 0)$ and $(I = 0)$. See Billingsley [6, p. 441] for similar variations of the classical Borel paradox. As a check on the constants of integration, it is known [5] that the mean squared maximum of a Brownian excursion is $\pi^2/6$. Thus (52) for $F(\cdots) = M^2$ gives

$$\frac{\pi^2}{6} = \mathbf{E}(M^2 \mid I = 0) = \sqrt{\frac{\pi}{2}}\, \mathbf{E}(\widetilde{M}_0)$$

in agreement with (50) for $F(\cdots) = 1$.

6. Some further identities

The formula (12) which we used as our starting point was derived in [23] as a consequence of the following trivariate identity, which characterizes the joint law of (I, M, L), where

$$L := \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^1 dt\, 1(|b_t| \leqq \varepsilon)$$

is the *local time at* 0 *of the bridge up to time* 1:

$$\mathbf{P}(|N| I \leqq x, |N| M \leqq y, |N| L \in dl) = \exp\left(-\frac{l}{2}(\coth x + \coth y)\right) dl. \tag{53}$$

We note that Corollary 3 could be applied to give another characterization of the law of (I, M, L).

Let $(\tau_l, l \geqq 0)$ denote the usual local time process at zero of a standard Brownian motion $(B_t, t \geqq 0)$. As shown in [3], the law of the *pseudo-bridge* $(b_t^{\#}, 0 \leqq t \leqq 1)$ defined by

$$b_t^{\#} := B_{t\tau_1} / \sqrt{\tau_1}$$

is absolutely continuous with respect to that of the bridge, with density $(\sqrt{\pi/2}\, L)^{-1}$. Equivalently, for every non-negative measurable function F defined on the path space $C[0, 1]$ there is the identity

$$\mathbf{E}[F(b_t, 0 \leqq t \leqq 1)] = \sqrt{\frac{\pi}{2}}\, \mathbf{E}\left[F(b_t^{\#}, 0 \leqq t \leqq 1) L^{\#}\right], \tag{54}$$

where $L^{\#} = 1/\sqrt{\tau_1}$ is the local time at 0 of $(b_t^{\#}, 0 \leqq t \leqq 1)$ up to time 1. In terms of the Brownian motion (B_t), define

$$I_t := -\inf_{0 \leqq u \leqq t} B_u; \qquad M_t := \sup_{0 \leqq u \leqq t} B_u; \qquad A_t := I_t + M_t; \qquad Q_t := \frac{I_t}{A_t}.$$

It was shown in [21] that Q_{τ_1} has uniform distribution on $(0,1)$. In view of (54) and Csáki's formula (8), this implies

$$\mathbf{E}[L^{-1}g(Q)] = \sqrt{\frac{\pi}{2}} \int_0^1 dv\, g(v)$$

for all non-negative Borel functions g. This formula can also be obtained quite easily from Theorem 2. From this formula we deduce that

$$\mathbf{E}[L^{-1} \mid Q = v] = \sqrt{\frac{\pi}{2}} \frac{1}{f_Q(v)}. \tag{55}$$

Compared with (37), this gives the curious formula

$$\mathbf{E}[M^{-1} \mid Q = v] = 4v\, \mathbf{E}[L^{-1} \mid Q = v]. \tag{56}$$

As shown by Lévy [17], the random variables M and $2L$ have identical Rayleigh distributions, with $\mathbf{P}(M > x) = \mathbf{P}(2L > x) = \exp(-2x^2)$ for $x > 0$. The conditional distribution of M given $Q = v$, which is determined by (35), could also be described by a series density derived from (1). It does not seem easy to describe the conditional law of L given $Q = q$ so explicitly, though the density of $|N|L$ on the event $(Q \leqq v)$ can be read from (53), and this could be used to give integral expressions for conditional moments of L given $Q \leqq v$ or $Q = v$.

7. The distribution of Q

We record in this section some properties of the distribution of Q which follow from Csáki's formula (8) for $\mathbf{P}(Q \leqq v)$. By differentiation of (8), the density at $q \in (0,1)$ is

$$f_Q(q) = \frac{\pi^2 q(1-q)}{\sin^2 \pi q} + (2q-1)\pi \cot \pi q - 1. \tag{57}$$

It is easily checked using (57) that

$$f_Q(q) = f_Q(1-q) \sim \frac{2\pi^2}{3} q \quad \text{as } q \downarrow 0, \tag{58}$$

where the first equality is obvious from the symmetry of Brownian bridge with respect to a sign change. Easily from (58)

$$\mathbf{P}(Q \leqq q) = \mathbf{P}(Q \geqq 1-q) \sim \frac{\pi^2}{3} q^2 \quad \text{as } q \downarrow 0. \tag{59}$$

This distribution of Q is close in most respects to the beta(2,2) distribution with density $6q(1-q)$. Both densities are concave and symmetric about $1/2$. The beta(2,2) distribution is slightly more peaked, with modal density $3/2 = 1.5$ at $q = 1/2$, whereas

$$f_Q(1/2) = \frac{\pi^2}{4} - 1 = 1.4674\ldots. \tag{60}$$

The density of the law of Q relative to the beta(2,2) law is subject to the bounds

$$0.978 \approx \frac{\pi^2 - 4}{6} \leqq \frac{f_Q(q)}{6q(1-q)} \leqq \frac{\pi^2}{9} \approx 1.097, \tag{61}$$

where the lower bound is attained at $1/2$ and the upper bound is sharp at $0+$ and $1-$. The total variation distance between these two densities was found by numerical integration using *Mathematica* to be

$$\int_0^1 dq |f_Q(q) - 6q(1-q)| \approx 0.019. \tag{62}$$

For $n > 0$ the nth moment $E(Q^n) = \int_0^1 dq\, q^n f_Q(q)$ can be evaluated by integration by parts as follows:

$$\begin{aligned} E(Q^n) &= \int_0^1 dv (1 - P(Q \leqq v)) n v^{n-1} \\ &= \frac{n}{n+1} + n\pi \int_0^1 dv v^n (1-v) \cot(\pi v). \end{aligned} \tag{63}$$

For $m = 1, 2, \ldots$ there is the classical identity [1, 23.2.17]

$$\int_0^1 dv B_{2m+1}(v) \cot(\pi v) = 2(2m+1)!(-1)^{m+1} \frac{\zeta(2m+1)}{(2\pi)^{2m+1}}, \tag{64}$$

where $B_n(v)$ is the nth Bernoulli polynomial, which is of degree n with rational coefficients, and $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ is the Riemann zeta function. Also, by symmetry,

$$E[(Q - 1/2)^{2m-1}] = 0. \tag{65}$$

It follows that for $n = 1, 2, \ldots$

$$E(Q^n) = \frac{n}{n+1} + \sum_{m=1}^{\lfloor n/2 \rfloor} a_{n,m} \frac{\zeta(2m+1)}{\pi^{2m}} \tag{66}$$

for some rational coefficients $a_{n,m}$ determined by (63), (64) and (65). For instance

$$E(Q) = \frac{1}{2}; \quad E(Q^2) = \frac{2}{3} - \frac{3\zeta(3)}{\pi^2}; \quad E(Q^3) = \frac{3}{4} - \frac{9\zeta(3)}{2\pi^2}; \tag{67}$$

$$E(Q^4) = \frac{4}{5} - \frac{8\zeta(3)}{\pi^2} + \frac{30\zeta(5)}{\pi^4}; \quad E(Q^5) = \frac{5}{6} - \frac{25\zeta(3)}{2\pi^2} + \frac{75\zeta(5)}{\pi^4}. \tag{68}$$

REFERENCES

[1] Abramowitz, M. and Stegun, I. A. (editors), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications Inc., New York, 1966. *MR* **34** #8606

[2] Bertoin, J. and Pitman, J., Path transformations connecting Brownian bridge, excursion and meander, *Bull. Sci. Math.* (2) **118** (1994), 147–166. *MR* **95b**:60097

[3] Biane, Ph., Le Gall, J. F. and Yor, M., Un processus qui ressemble au pont brownien, *Séminaire de Probabilités XXI*, Lecture Notes in Math., 1247, Springer, Berlin–New York, 1987, 270–275. *MR* **89d**:60145

[4] Biane, Ph., Pitman, J. and Yor, M., Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions, Technical Report No. 569, Department of Statistics, University of California, Berkeley, 1999. Available via www.stat.berkeley.edu/users/pitman

[5] Biane, Ph. and Yor, M., Valeurs principales associées aux temps locaux Browniens, *Bull. Sci. Math.* (2) **111** (1987), 23–101. *MR* **88g**:60188

[6] Billingsley, P., *Probability and measure*, 3rd edition, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1995. *MR* **95k**:60001

[7] Blumenthal, R. M., Weak convergence to Brownian excursion, *Ann. Probab.* **11** (1983), 798–800. *MR* **85e**:60083

[8] Carmona, Ph., Petit, F., Pitman, J. and Yor, M., On the laws of homogeneous functionals of the Brownian bridge, *Studia Sci. Math. Hungar.* **35** (1999), 445–455.

[9] Chung, K. L., Excursions in Brownian motion, *Ark. Mat.* **14** (1976), 155–177. *MR* **57** #7791

[10] Csáki, E., On some distributions concerning maximum and minimum of a Wiener process, *Analytic function methods in probability theory* (Proc. Colloq. Methods of Complex Anal. in the Theory of Probab. and Statist., Debrecen, 1977), ed. by B. Gyires, Colloq. Math. Soc. János Bolyai, **21**, North-Holland, Amsterdam – New York, 1979, 43–52. *MR* **81b**:60079

[11] Doob, J., Heuristic approach to the Kolmogorov–Smirnov theorems, *Ann. Math. Statistics* **20** (1949), 393–403. *MR* **11**, 43a

[12] Gnedenko, B. V., Kriterien für die Unveränderlichkeit der Wahrscheinlichkeitsverteilung von zwei unabhängigen Stichprobenreihen, *Math. Nachr.* **12** (1954), 29–66 (in Russian). *MR* **16**, 498c

[13] GRADSHTEYN, I. S. and RYZHIK, I. M., *Table of integrals, series, and products*, Corrected and enlarged edition, edited by Alan Jeffrey, Academic Press, New York, 1980. *MR* **81g**:33001

[14] HSU, P. and MARCH, P., Brownian excursions from extremes, *Séminaire de Probabilités XXII*, edited by J. Azéma, P.-A. Meyer and M. Yor, Lecture Notes in Mathematics, 1321, Springer-Verlag, Berlin – New York, 1988, 502–507. *MR* **89c**:60005

[15] IMHOF, J. P., A construction of the Brownian path from BES^3 pieces, *Stochastic Process. Appl* **43** (1992), 345–353. *MR* **94c**:60133

[16] KENNEDY, D. P., The distribution of the maximum Brownian excursion, *J. Appl. Probability* **13** (1976), 371–376. *MR* **53** #6769

[17] LÉVY, P., Sur certains processus stochastiques homogènes, *Compositio Math.* **7** (1939), 283–339. *Zbl* **22**.059

[18] PERMAN, M. and WELLNER, J., An excursion approach to the Kolmogorov–Smirnov statistic (in preparation).

[19] PITMAN, J. and YOR, M., A decomposition of Bessel bridges, *Z. Wahrsch. Verw. Gebiete* **59** (1982), 425–457. *MR* **84a**:60091

[20] PITMAN, J. and YOR, M., Decomposition at the maximum for excursions and bridges of one-dimensional diffusions, *Itô's stochastic calculus and probability theory*, ed. by N. Ikeda, S. Watanabe, M. Fukushima and H. Kunita, Springer-Verlag, Tokyo, 1996, 293–310.

[21] PITMAN, J. and YOR, M., Random Brownian scaling identities and splicing of Bessel bridges, *Ann. Probab.* **26** (1998), 1683–1702.

[22] PITMAN, J. and YOR, M., Laws of homogeneous functionals of Brownian motion (in preparation).

[23] REVUZ, D. and YOR, M., *Continuous martingales and Brownian motion*, 2nd edition, Springer-Verlag, Berlin – Heidelberg, 1994. *MR* **95h**:60072

[24] SHORACK, G. R. and WELLNER, J. A., *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, 1986. *MR* **88e**:60002

[25] SMIRNOV, N. V., On the estimation of the discrepancy between empirical curves of distribution for two independent samples, *Bull. MGU* **2** (1939), 3–14 (in Russian). *Bull. Math. Univ. Moscou, Sér. internat.* **2** (1939), fasc. 2, 1–16. *Zbl* **23**.249

[26] VALLOIS, P., Decomposing the Brownian path via the range process, *Stochastic Process. Appl.* **55** (1995), 211–226. *MR* **96a**:60067

[27] VERVAAT, W., A relation between Brownian bridge and Brownian excursion, *Ann. Probab.* **7** (1979), 143–149. *MR* **80b**:60107

[28] WILLIAMS, D., Path decomposition and continuity of local time for one-dimensional diffusions. I, *Proc. London Math. Soc.* (3) **28** (1974), 738–768. *MR* **50** #3373

[29] WILLIAMS, D., *Diffusions, Markov processes, and martingales, Vol. I. Foundations*, Probability and Mathematical Statistics, Wiley, Chichester, 1979. *MR* **80i**: 60100

[30] WILLIAMS, D., Brownian motion and the Riemann zeta-function, *Disorder in physical systems*, ed. by G. R. Grimmett and D. J. A. Welsh, Clarendon Press, Oxford; Oxford Univ. Press, New York, 1990, 361–372. *MR* **91h**:60094

(Received October 1, 1998)

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
367 EVANS HALL
BERKELEY, CA 94720-3860
U.S.A.

pitman@stat.berkeley.edu

LABORATOIRE DE PROBABILITÉS
UNIVERSITÉ PARIS VI
TOUR 56 – 3° ÉTAGE
4 PLACE JUSSIEU
F–75252 PARIS Cedex 05
FRANCE

secret@proba.jussieu.fr

Studia Scientiarum Mathematicarum Hungarica 35 (1999), 475–478

BOOK REVIEWS

Asymptotic Methods in Probability and Statistics, *A Volume in Honour of Miklós Csörgő*, Ed. by B. Szyszkowicz, Elsevier Science B.V., Amsterdam, 1998, xxxiii, 889 pp. ISBN 0 444 50083 9.

ICAMPS'97, an International Conference on Asymptotic Methods in Probability and Statistics was organized and held in honour of Professor Miklós Csörgő at Carleton University, Ottawa, Canada 8-13 July, 1997.

The present volume is the Proceedings of this Conference, containing 55 papers, mainly on the subjects Miklós Csörgő was a main contributor. The volume consists of 17 Parts and reflects the developments in the last few years in the fields of invariance principles, local times and other additive functionals, iterated processes, change-point and other nonparametric problems, including empirical and quantile processes, etc.. The volume starts with a Preface by B. Szyszkowicz, a detailed review of the scientific work of M. Csörgő.

Part 1: Limit theorems for variously mixing and quasi-associated random variables. (Contributors: S. Csörgő, P. Kowalski–Z. Rychlik, T. M. Lewis, M. Peligrad.)

Part 2: Central limit theorems for logarithmic averages. (I. Berkes, E. Csáki–A. Földes.)

Part 3: Strong approximations, weighted approximations. (A. R. Dabrowski–H. Dehling, P. Deheuvels, P. W. Glynn, G. R. Shorack.)

Part 4: Empirical distributions and processes. (K. Ghoudi–B. Remillard, P. Massart–E. Rio, L. Takács.)

Part 5: Iterated random walks. (K. Grill, P. Révész.)

Part 6: Fine analytic path behavior of the oscillations of stochastic processes. (S. Keprta, W. V. Li, Z. Y. Lin–Y. C. Qin, C. R. Lu–H. Yu, J. Steinebach, Y. Xiao.)

Part 7: Multiparameter stochastic processes. (B. Chen, B. G. Ivanoff–N. C. Weber.)

Part 8: Results related to studies of local time and hitting times of Bessel processes. A cautionary note on limiting sigma-algebras. (E. Csáki–A. Földes, D. L. Hanson, Y. Hu–M. Yor.)

Part 9: Large deviations, small ball problems, self normalization. (D. A. Dawson–J. Gärtner, S. Feng, D. Khoshnevisan–Z. Shi, Q.-M. Shao.)

Part 10: Stochastic bifurcations (K. Burdzy–D. M. Frankel–A. Pauzner)

Part 11: Change-point analysis, U-statistics, non-smooth functions, comparison distributions. (E.-E. A. A. Aly, J. A. Correa, B. Freidlin–J. L. Gastwirth, E. Gombay–L. Horváth, M. Hušková, F. Lombard, H.-G. Müller, E. Parzen.)

Part 12: Empirical reliability, survival analysis. (L. Rejtő–G. Tusnády, M. D. Rothmann–R. P. Russo, H. Yu, R. Zitikis.)

Part 13: Gaussian bootstrap, Monte Carlo simulation. (M. D. Burke, B. J. Eastwood–V.R. Eastwood.)

Part 14: Autoregressive and moving average schemes. (G. Haiman, M. Rosenblatt.)

Part 15: Nonparametric curve estimation, regression diagnostics. (R. J. Kulperger, P. Major–L. Rejtő, S. Portnoy.)

Part 16: Testing statistical hypotheses. (M. Alvo–P. Cabilio, J. Babb–A. Rogatko–S. Zacks, T. Inglot–W. C. M. Kallenberg–T. Ledwina.)

0081–6906/99/$ 5.00

Part 17: Tail index estimation, order statistics of order statistics. (S. Csörgő–L. Viharos, R. J. Tomkins.)

The papers in this volume are valuable contributions to asymptotic methods in probability and statistics and is recommended to researchers in this field.

E. Csáki (Budapest)

The Art and Craft of Problem Solving, by Paul Zeitz, John Wiley & Sons, Inc., New York – Chichester, 1999, xvii + 334 pp. ISBN 0 471 13571 2.

How to become a famous mathematician? What distinguishes the contest winner young talent from the average teenager around the corner? Is there any way to improve your, or your kid's, problem solving capacity? The book of Paul Zeitz has a lot to say about these questions.

The author was a member of the very first US team to participate in the International Mathematical Olympiad, and almost twenty years later he has been coaching several recent teams. The intellectual challenge of solving mathematical problems captured him while in high school, and rooted deeply. "As a missionary for the problem solving culture" he writes in the Preface, "*The Art and Craft of Problem Solving* is a first approximation of my attempt to spread the gospel." He compares problem solving to hiking. As the hiker is rewarded by the scenery both en route and at the destination, similarly "[t]he problem solver climbs to the top of mountains, sees hitherto undreamed vistas. The problem solver arrives at places of amazing beauty, and experiences ecstasy which is amplified by the effort expended to get there." The book tries to teach those techniques, methods, tricks and know-hows, tactics and strategic thinking, which makes the reader capable of making longer tours on higher, and, gradually, more dangerous mountains of mathematical thinking.

Part I is an excellent introduction to basic, psychological and non-psychological, techniques of problem solving. A successful problem solver must have several qualities, which can be developed through hard work, such as confidence, concentration, creativity, and open-mindedness. "Just because a problem seems impossible does not mean that it is impossible. Never admit defeat after a cursory glance." There are stories about how someone solved a long-standing open problem just because had no idea of the enormous (unsuccessful) effort put previously into the problem. Also "[n]ice guys may or may not finish last, but *good, obedient boys and girls solve fewer problems than naughty and mischievous ones.*" Moral: break, or at least bend the rules if they do not lead to the solution. Each rule is spelled out explicitly, and is richly illustrated with either "folklore" problems, or problems from different math competitions. Certain problems are recurring at different sections, offering new insight, or new ways of attack.

Once over Part I, and having the basic skills, we could enter the more messy Part II, entitled "Specifics." Here the problems are grouped not by the method which could help to solve them, rather by the topic they belong: Algebra, Combinatorics, Number Theory, and Calculus. As in Part I, all sections end with at least a dozen problems and exercises illustrating the section's main ideas. Approaching the crux of the book, more and more problems are solved, more and more tricks and methods are introduced. After we have learnt how to make the first steps, how to arrange our initial ideas, we have to recognize that the essence of mathematical problem solving cannot be captured by a few well chosen recipes, there is always a place for new ideas, for unexpected connections. And here is the only point Paul Zeitz's excellent book missed. He convinces the reader that (math) problems can be solved, that this is an intellectual challenge, and gives satisfaction and the feeling of a well-done work; this is not a privilege of a few but can be learnt. What he does not say is essentially two important facts. First, there is a significant difference

between problems set up for solution in a contest (or in a newspaper) and problems arising in the everyday life (of mathematics, say). While the former ones do have a simple solution (otherwise they could not be posed), the latter ones do not necessarily have any. There are hopeless mathematical problems, and they must be avoided during the first years of study. Secondly, and this is more important, while there are certain techniques which must be mastered, this does not mean that for all problems discussed in the book the most elegant solutions should pop out from the student's head. For several problems it took years, sometimes decades, for a simple proof to appear. It is unjust even to suggest that "this is the way it should be solved, and if you cannot figure it out by yourself, you won't be a good mathematician." Several deep and important mathematical facts, theorems, even problems have only proofs and solutions which are ugly, lengthy, full of sweat and struggle. The most rewarding experience is to find a truly nice, illuminating, simple solution – a proof from The Book. But finding any, even the ugliest one, is equally satisfactory, and the fame goes with the first proof.

I recommend "The Art and Craft of Problem Solving" to those, who want to learn problem solving techniques; to those who teach that kind of people; and to those who simply want to amuse themselves by the myriad of wonderful brain-twisters and mathematical puzzles. The required mathematics never goes beyond the first undergraduate level. Finally, let me quote three simple problems from the book, one from the beginning, one from about midway, and one from the end. You may try your own claws on them.

Problem 2.1.27, (b) Of all the books at a certain library, if you select one at random, then there is a 90% chance that it has illustration. Of all the illustrations in all the books, if you select one at random, then there is 90% chance that it is in color. If the library has 10,000 books, then what is the minimum number of books that must contain colored illustrations?

Problem 4.1.22 The 20 members of a local tennis club have scheduled exactly 14 two-person games among themselves, with each member playing at least one game. Prove that within this schedule there must be a set of 6 games with 12 distinct players.

Problem 8.4.25 Let $P = \{4, 8, 9, 16, \dots\}$ be the set of perfect powers, i.e. the set of positive integers of the form a^b where a and b are integers greater than 1. Prove that

$$\sum_{j \in P} \frac{1}{j-1} = 1.$$

László Csirmaz (Budapest)

Methods in Ring Theory, Proceedings of the Trento Conference, Eds. V. Drensky, A. Giambruno, S. Sehgal, Lecture Notes in Pure and Applied Mathematics, Vol. 198, Marcel Dekker, Inc., New York–Basel–Hong Kong, 1998.

From the Preface: "This volume contains the proceedings of the Methods in Ring Theory conference held in the small town of Levico Terme in the Italian Dolomites. The aim of the conference was to present and give an update on interesting techniques and methods in two important branches of modern ring theory: group algebras and algebras with polynomial identities. ... The papers of most of the principal speakers and some other contributions are included."

Here is the table of contents; expository papers are marked with an asterisk.

L. Carini and A. Regev: Young derivation of the trace cocharacters of the 2×2 matrices

S. P. Coelho and C. Polcino Milies: Torsion subgroups of units in artinian rings

M. Domokos: Polynomial ideals and identities of matrices

* V. Drensky: Gelfand–Kirillov dimension of PI-algebras

A. Giambruno and M. V. Zaicev: PI-algebras and codimension growth

K. Hoechsmann: Unit bases in small cyclic group rings

* E. Jespers: Units in integral group rings: A survey

A. Kemer: On the multilinear components of the regular prime varieties

Z. S. Marciniak and S. K. Sehgal: Units in group rings and geometry

* D. S. Passman: The semiprimitivity of group algebras

* V. M. Petrogradsky: Scale for codimension growth of Lie algebras

A. Pianzola and J. Valencia: The Hopf shuttle algebra and Lie series

A. Popov: Graded identities and cocharacters of products of T-ideals

* C. Procesi: Deformations of representations

* D. M. Riley: The Engel condition, semigroup identities, and collapsibility in associative rings

J. Szigeti: Idempotent ideals in Lie nilpotent rings

A. Valenti: Units and group identities in rings

M. V. Zaicev: Varieties and identities of affine Kac–Moody algebras

The volume is a precious contribution to ring theory and especially researchers in group rings and polynomial identities shall not miss it. Several papers from it (e.g. Procesi's excellent presentation) address readers coming from a much broader field, and this makes the book a good acquisition for the library of any Department with research interest in ring theory.

L. Márki (Budapest)

Studia Scientiarum Mathematicarum Hungarica

Volume 35

Akadémiai Kiadó, Budapest

1999

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Volume 35

CONTENTS

RECENTLY ACCEPTED PAPERS

Toma, V. and Zuzčák, I., Bases and homeomorphisms in polytopological spaces

Kovács, K., On a characterization of cyclic groups by sums and differences

Bala, P., On range cyclic operator algebras. II

Pouzet, M. and Rosenberg, I. G., Embedding and absolute retracts of relational systems

Ament, P. and Blind, G., Packing 34 circles in a square

Blahota, I. and Gát, G., Pointwise convergence of double Vilenkin–Fejér means

Saran, J., Supplement to a result of G. P. Steck and G. J. Simmons on the distribution of (D_{mn}^{+}, R_{mn}^{+})

Dumitrescu, A., Planar sets with few empty convex polygons

Slezák, B., On the primitive of the differential one-forms

Jüttler, B., Arbitrarily weak linear convexity conditions for multivariate polynomials

Khan, L. A., Abujabal, H. A. S. and Alghamdi, M. A., On the Riesz representation theorem in topological vector spaces

Bazzanella, D. and Languasco, A., On the asymptotic formula for Goldbach numbers in short intervals

Doster, W., Jeffrey's prior is the Hausdorff measure for the Hellinger and Kullback–Leibler distances

Major, P., Almost sure functional limit theorems. Part II. The case of independent random variables

Menke, B., On longest cycles in grid graphs

Kupitz, Y. S. and Martini, H., On the weak circular intersection property

Berthet, P. and Shi, Z., Small ball estimates for Brownian motion under a weighted sup-norm

Romaguera, S. and Salbany, S., Dieudonné complete bispaces

Crvenkovic, S., Dolinka, I. and Vinčić, M., Equational bases for some 0-direct unions of semigroups

Manuscripts should be submitted in duplicate, typed in double spacing on only one side of the paper with wide margins. Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only upon receipt of the signed copy of the agreement.

Authors are encouraged to submit their papers electronically. All common dialects of TEX are welcome. The electronic file of an article should always be accompanied by a hardcopy, the printed version exactly corresponding to the electronic one.

Figures should be submitted on separate sheets, drawn either in India ink at their expected final size, or as printouts and matching files processed electronically, preferably as encapsulated PostScript (EPS) ones.

CONTENTS

HU ISSN 0081-6906