Rostocker Mathematisches Kolloquium

Heft 50

S. G. Hristova S. I. Kostadinov M. B. Kapitanova	*Method of a small parameter for a class of nonautonomous systems of second order differential equations with impulses*	3
S. G. Hristova S. I. Kostadinov	*Integral inequalities for piecewice – continuous functions of many variables*	9
Lothar Berg Gerlind Plonka	*Refinement of Vectors of Bernstein Polynomials*	19
D. Lau	*Die maximalen Klassen von* $\widetilde{\boldsymbol{P}}_{\mathbf{3}}$	31
Tran-Ngoc Danh David E. Daykin	*Bezrukov-Gronau Order is Not Optimal*	45
Tran Ngoc Danh David E. Daykin	*Sets of 0,1 vectors with minimal sets of subvectors*	47
J. Synnatzschke	*Zur Erzeugung von Algebren durch Unteralgebren mit Quadrat Null*	53
Markus Kappert	*Eine Gradaussage für Posinome bester Approximation*	65
Dieter Schott	*Basic Properties of Fejer monotone Mappings*	71
Zeqing Liu	*Order completeness and stationary points*	85
Manfred Krüppel	*On the nearest point projection in Hilbert spaces with application to Nonlinear Ergodic Theory*	89

Universität Rostock

Fachbereich Mathematik

1997

Wissenschaftliche Leitung:	Prof. Dr. Strecker (Sprecher des Fachbereichs Mathematik) Dr. Werner Plischke
Redaktionelle Bearbeitung:	Dr. Werner Plischke
Herstellung der Druckvorlage:	S. Dittmer, W. Hartmann, H. Schubert

Universität Rostock
Fachbereich Mathematik
18051 Rostock
Bundesrepublik Deutschland

Redaktionsschluß: 2. August 1997

Das **Rostocker Mathematische Kolloquium** erscheint dreimal im Jahr und ist im Rahmen des Schriftentausches über die Universität Rostock, Universitätsbibliothek, Tauschstelle, 18051 Rostock, Bundesrepublik Deutschland, zu beziehen.

Außerdem bestehen Bezugsmöglichkeiten für Bestellungen aus Deutschland und dem Ausland über die Universität Rostock, Abteilung Wissenschaftspublizistik, 18051 Rostock, Bundesrepublik Deutschland

Zitat–Kurztitel: Rostock. Math. Kolloq. (1997) **50**

Universität Rostock
Abteilung Wissenschaftspublizistik
18051 Rostock, Bundesrepublik Deutschland
Telefon 4 40 55 20
Druck: Universitätsdruckerei
01000

Rostock. Math. Kolloq. **50**, 3–8 (1997) Subject Classification (AMS) **34A37 34E05**

S. G. Hristova
M. B. Kapitanova
S. I. Kostadinov

Method of a small parameter for a class of nonautonomous systems of second order differential equations with impulses

Systems with impulses find a wide application to mathematical modelling in ratio engineering, physics, chemistry, biology, etc. . The development of the qualitative theory of these systems is of particular interest since there solutions can be found in an explicit form only in exceptional cases.

An efficient method by means of which the periodic solutions of differential equations are investigated successfully is the method of a small parameter. In the present paper by means of this method a periodic solution of a nonautonomous system of second order impulsive differential equations is constructed. The case when the characteristic equation has some complex conjugate roots is studied.

Consider the system of impulsive differential equations

$$A\ddot{x} = Bx + f(t) + \sum_{k=1}^{\infty} \varepsilon^k f^{(k)}(t, x, \dot{x}, \ddot{x}) \tag{1}$$

$$\Delta x\big|_{t=\pi i} = Sx(\pi i) \tag{2}$$

$$\Delta \dot{x}\big|_{t=\pi i} = T\dot{x}(\pi i) \ , \tag{3}$$

where $x \in R^n$, A and B are $n \times n-$dimensional matrices, A is non-singular matrix, f and $f^{(k)}$ $(k = 1, 2, \dots)$ are $n-$dimensional vectors, $\Delta x|_{t=\pi i} = x(\pi i + 0) - x(\pi i - 0)$, $S = \operatorname{diag}(s_{11}, s_{22}, \dots, s_{nn})$, $T = \operatorname{diag}(t_{11}, t_{22}, \dots, t_{nn})$, $\varepsilon > 0$ is a small parameter, $i = 0, \pm 1, \pm 2, \dots$.

We will construct a $2\pi-$periodic solution of a system (1)–(3). We will consider the case when $t \in [0, 2\pi]$ and $i = 1$, i.e. the „jump“ condition (2) and (3) are at the piont $t = \pi$.

The equation

$$\det(P - \lambda^2 E) = 0 \ , \tag{4}$$

where $P = A^{-1}B = (p_{ij})_1^n$ and E is the $n-$dimensional identity matrix, will be called a characteristic equation of the system (1) – (3).

Let the equation (4) has l complex conjugate roots of the type $\underline{+}iq_j$, where $q_j(j = 1, 2, \ldots, l)$ are natural numbers.

We will call a generating system of the system (1) – (3) the following system

$$\ddot{x}^{(0)} = Px^{(0)} + G(t) \tag{5}$$

$$\Delta x^{(0)}\Big|_{t=\pi} = Sx^{(0)}(\pi) \tag{6}$$

$$\Delta \dot{x}^{(0)}\Big|_{t=\pi} = T\dot{x}^{(0)}(\pi) \ , \tag{7}$$

where $G(t) = A^{-1}f(t)$.

The general solution of the system (5) – (7) is of the following type

$$x^{(0)}(t) = \begin{cases} C^{(0)}\varphi(t) + \overline{x}^{(0)}(t) & \text{for } t \in [0, \pi] \\ \overline{C}^{(0)}\varphi(t) + \overline{x}^{(0)}(t) & \text{for } t \in (\pi, 2\pi] \end{cases} \ , \tag{8}$$

where $C^{(0)} = (C_1^{(0)}, c_2^{(0)}, \ldots, C_{2n}^{(0)})$, $\overline{C}^{(0)} = (\overline{C}_1^{(0)}, \ldots, \overline{C}_{2n}^{(0)})$. $\varphi(t)$ is $2n \times n-$dimensional matrix of fundamental solutions $\varphi_{j_k}(t)$ $(j = 1, \ldots, 2n)$ of the homogeneous system $\ddot{x} = Px$, and the function $\overline{x}^{(0)}(t)$ is a particular solution of (5) and ([1]). According to ([1]) we have

$$\overline{x}^{(0)}(t) = \int\limits_0^t G(t) \cdot \Xi(t - \lambda) d\lambda \ , \tag{9}$$

$\Xi(t)$ is a matrix of fundamental solutions $\xi_{s_j}(t)(s = \overline{1,n}; j = \overline{1,2n})$ of the homogeneous system, corresponding to (5), for which

$$\begin{array}{llll} \xi_{s_j}(0) = \delta_{s_j} \ , & \dot{\xi}_{s,n+j}(0) = \delta_{s_j} & \text{for} & s \neq j \\ \xi_{s_j}(0) = 0 \ , & \dot{\xi}_{s,n+j}(0) = 0 & \text{for} & s = j \ , \ (s, j = \overline{1,n}) \end{array} \tag{10}$$

and δ_{s_j} are the symbols of Kronecker.

In (8) $x^{(0)}(t) = (x_1^{(0)}(t), \ldots, x_n^{(0)}(t))$ and

$$x_s^{(0)}(t) = \begin{cases} C_1^{(0)}\,\varphi_{s1}(t) + \cdots + C_{2l}\,\varphi_{s,2l}(t) + \cdots + C_{2n}\,\varphi_{2n}^{(0)}(t) \text{ for } t \in [0, \pi] \\ \overline{C}_1^{(0)}\,\varphi_{s1}(t) + \cdots + \overline{C}_{2l}^{(0)}\,\varphi_{s,2l}(t) + \cdots + \overline{C}_{2n}^{(0)}\,\varphi_{2n}(t) \text{ for } t \in [\pi, 2\pi] \end{cases} \tag{11}$$

where $C_j^{(0)}$ and $\overline{C}_j^{(0)}$ $(j=\overline{1,n})$ are unknown constants and

$$\begin{aligned}
\varphi_{s1}(t) &= L_s^{(1)}\,\omega s\cos q_1 t\ ,\\
\varphi_{s2}(t) &= L_s^{(1)}\ \sin q_1 t\ ,\\
\varphi_{s3}(t) &= L_s^{(2)}\,\omega s\cos q_2 t\ ,\\
\varphi_{s4}(t) &= L_s^{(2)}\ \sin q_2 t\ ,\\
&\dots\dots\dots\dots\dots\dots\\
\varphi_{s,2l-1}(t) &= L_s^{(l)}\,\omega s\cos q_l t\ ,\\
\varphi_{s,2l}(t) &= L_s^{(l)}\ \sin q_l t\ ,\quad (s=\overline{1,n})\ .
\end{aligned} \tag{12}$$

In the notations (12) the constants $L_s^{(j)}$ $(j=\overline{1,l})$ are nonzero solutions of the systems

$$p_{s1}L_1^{(j)}+p_{s2}L_2^{(j)}+\dots+(p_{ss}+q_j)L_s^{(j)}+\dots+p_{sn}L_n^{(j)}=0\ ,\quad (s=\overline{1,n})\ .$$

The conditions for periodicity and the „jump“ conditions of the solution $x^{(0)}(t)$ of the system (5) – (7) are of the type

$$\begin{cases}
x_k^{(0)}(2\pi)-x_k^{(0)}(0)=0\\
\dot{x}_k^{(0)}(2\pi)-\dot{x}_k^{(0)}(0)=0\\
x_k^{(0)}(\pi+0)-x_k^{(0)}(\pi)=s_{kk}\cdot x^{(0)}(\pi)\\
\dot{x}_k^{(0)}(\pi+0)-\dot{x}_k^{(0)}(\pi)=t_{kk}\cdot \dot{x}^{(0)}(k)\ ,\quad (k=\overline{1,n})\ .
\end{cases} \tag{13}$$

From conditions (11) and (13) it follows that

$$\begin{cases}
L_k^{(1)}\left(\overline{C}_1^{(0)}-C_1^{(0)}\right)+L_k^{(2)}\left(\overline{C}_3^{(0)}-C_3^{(0)}\right)+\dots+\\
+L_k^{(l)}\left(\overline{C}_{2l-1}^{(0)}-C_{2l-1}^{(0)}\right)+\left(\overline{C}_{2l+1}^{(0)}-C_{2l+1}^{(0)}\right)\varphi_{k,2l+1}(0)+\\
+\dots+\left(\overline{C}_{2n}^{(0)}-C_{2n}^{(0)}\right)\varphi_{k,2n}(0)+\overline{x}_k^{(0)}(2\pi)=0 \qquad (k=\overline{1,n})
\end{cases} \tag{14}$$

$$\begin{cases}
p_1L_k^{(1)}\left(\overline{C}_2^{(0)}-C_2^{(0)}\right)+p_2L_k^{(2)}\left(\overline{C}_4^{(0)}-C_4^{(0)}\right)+\dots+\\
+p_lL_k^{(l)}\left(\overline{C}_{2l}^{(0)}-C_{2l}^{(0)}\right)+\left(\overline{C}_{2l+1}^{(0)}-C_{2l+1}^{(0)}\right)\cdot\dot{\varphi}_{k,2l+1}(0)+\\
+\dots+\left(\overline{C}_{2n}^{(0)}-C_{2n}^{(0)}\right)\dot{\varphi}_{k,2n}(0)+\dot{\overline{x}}^{(0)}(2\pi)=0 \qquad (k=\overline{1,n})
\end{cases} \tag{15}$$

$$\begin{cases}
(-1)^{p_1}\cdot L_k^{(1)}\overline{C}_1^{(0)}+(-1)^{p_2}\cdot L_k^{(2)}\cdot\overline{C}_3^{(0)}+\dots+(-1)^{p_l}L_k^{(2)}\overline{C}_{2l-1}^{(0)}-\\
-\overline{C}_{2l+1}^{(0)}\cdot\varphi_{k,2l+1}(\pi)-\dots-\overline{C}_{2n}^{(0)}\cdot\varphi_{k,2k}(\pi)=\\
=(s_{kk}+1)\left[(-1)^{p_1}\cdot L_k^{(1)}\cdot C_1^{(0)}+(-1)^{p_2}L_k^{(2)}\cdot C_3^{(0)}+\dots+\right.\\
\qquad\left.+(-1)^{p_l}\cdot L_k^{(l)}\cdot C_{2l-1}^{(0)}\cdot\varphi_{k,2l+1}(\pi)-\dots-C_{2n}^{(0)}\cdot\varphi_{k,2n}(\pi)\right]\ ,\qquad (k=\overline{1,n})
\end{cases} \tag{16}$$

$$\begin{cases} p_1\cdot(-1)^{p_1}L_k^{(1)}\overline{C}_2^{(0)}+p_2(-1)^{p_2}\cdot L_k^{(2)}\cdot\overline{C}_4^{(0)}+ \\ +p_l\cdot(-1)^{p_l}\cdot L_k^{(2)}\overline{C}_{2l}^{(0)}-\overline{C}_{2l+1}^{(0)}\cdot\dot{\varphi}_{k,2l+1}(\pi)- \\ -\overline{C}_{2n}^{(0)}\cdot\dot{\varphi}_{k,2k}(\pi) \\ =(t_{kk}+1)\Big[p_1\cdot(-1)^{p_1}L_k^{(1)}C_2^{(0)}+\dots+p_l(-1)^{p_l}L_k^{(l)}\cdot C_{2l}^{(0)}- \\ -C_{2l+1}^{(0)}\cdot\dot{\varphi}_{k,2l+1}(\pi)-\dots-C_{2n}^{(0)}\cdot\dot{\varphi}_{k,2\pi}(\pi)\Big]\ , \qquad (k=\overline{1,n}) \end{cases} \tag{17}$$

The determinant D of the system (14) - (17) with $4n$ linear equations and unknown $C_1^{(0)}$, $\dots, C_{2n}^{(0)}, \overline{C}_1^{(0)}, \dots, \overline{C}_{2n}^{(0)}$ is nonzero when S and T are nonzero matrices. Then this system has a unique solution $C_j^{(0)}, \overline{C}_j^{(0)} (j=\overline{1,2n})$. If we substitute the values of these constants in (11) then we find a 2π-periodic solution $x^{(0)}(t)=(x_1^{(0)}(t),\dots,x_n^{(0)}(t))$ in the generating system (5) - (7). The periodic solution of the system (1) - (3) can be developed in a power series and it coincides for $\varepsilon=0$ with the solution $x^{(0)}(t)=(x_1^{(0)}(t),x_2^{(0)}(t),\dots,x_n^{(0)}(t))$ of the generating system (5) - (7). So we seek the solution in the form

$$x(t)=x^{(0)}(t)+\varepsilon x^{(1)}(t)+\varepsilon^2x^{(2)}(t)+\dots, \tag{18}$$

where $x^{(s)}(t)=(x_1^{(s)}(t),x_2^{(s)}(t),\dots,x_n^{(s)}(t)),(s=\overline{1,n})$.
Let the functions $x^{(0)}(t),x^{(1)}(t),\dots,x^{(k-1)}(t)$ are found.
We substitute the expansion (18) into equations (1)-(3) and calculate the coefficients before ε of the power k. So we obtain the following system of differential equations for the functions $x^{(k)}(t)$:

$$\ddot{x}^{(k)}=P\cdot x^{(k)}+G(t) \tag{19}$$

$$\Delta x^{(k)}\big|_{t=\pi}=S\cdot x^{(k)}(t) \tag{20}$$

$$\Delta\dot{x}^{(k)}\big|_{t=\pi}=T\cdot\dot{x}^{(k)}(t)\ , \tag{21}$$

where $G^{(k)}=A^{-1}(\Theta^{(k)}+\overline{G}^{(k)})$, and

$$\Theta^{(k)}=\left(\frac{\partial f^{(1)}(\xi)}{\partial x}\right)\cdot x^{(k-1)}+\left(\frac{\partial f^{(1)}(\xi)}{\partial\dot{x}}\right)\cdot\dot{x}^{(k-1)}+\left(\frac{\partial f^{(1)}(\xi)}{\partial\ddot{x}}\right)\cdot\ddot{x}^{(k-1)}, \tag{22}$$

$\overline{G}(k)$ contains $x^{(0)}(t),\dots,x^{(k-2)}(t)$ and their first and second derivatives,

$$\xi=(t,x^{(0)}(t),\dot{x}^{(0)}(t),\ddot{x}^{(0)}(t)).$$

The general solution of the system (19) - (21) is of the type

$$x^{(k)}(t)=\begin{cases} C^{(k)}\varphi(t)+\overline{x}^{(k)}(t) & \text{for}\quad t\in[0,\pi] \\ \overline{C}^{(k)}\varphi(t)+\overline{x}^{(k)}(t) & \text{for}\quad t\in(\pi,2\pi] \end{cases}$$

where $\overline{x}^{(k)}(t)$ is a particular solution of (19) and

$$\overline{x}^{(k)}(t) = \int_0^t G^{(k)}(t) \cdot \Xi(t-\lambda)\partial\lambda,$$

$$C^{(k)} = (C_1^{(k)}, \dots, C_{2n}^{(k)}), \ \overline{C}^{(k)} = (\overline{C}_1^{(k)}, \dots ; \overline{C}_{2n}^{(k)}),$$

$C_j^{(k)}$ and $\overline{C}_j^{(k)} (j = \overline{1, 2n})$ are unknown constants.
The conditions for periodicity and „jump" conditions of the solution $x^{(k)}(t)$ of system (19)-(21) are of the type (13), where the superscript "0" is replaced by the superscript "k".
The constants $C_1^{(k)}, \dots, C_{2n}^{(k)}, \overline{C}_1^{(k)}, \dots, \overline{C}_{2n}^{(k)}$ are solution of a system of the type (14)-(17) where the superscript "0" is replaced by the superscript "k". The determinant of this system is equal to D. That's why the system has a unique solution $C_1^{(k)}, \dots, C_{2n}^{(k)}, \overline{C}_1^{(k)}, \dots, \overline{C}_{2n}^{(k)}$ if S and T are nonzero matrices.
In this way we can construct a 2π-periodic solution of the system (1)-(3) as a power series.

Theorem 1 *Let the following conditions be fulfilled:*

1. *The functions $f(t)$ and $f^{(k)}(t, x, \dot{x}, \ddot{x})(k = 1, 2, \dots)$ are continous and periodic functions with respect to t with a period 2π and they are analytic functions with respect to all arguments.*

2. *The matrices S and T are nonzero diagonal matrices.*

3. *The matrix A is non-singular.*

4. *The equation (4) has l complex conjugate roots of the type $\pm iq_j$ $(j = \overline{1, l})$ where q_j are natural numbres.*

Then the system with impulses (1)-(3) has a 2π-perodic solution which coincides with a 2π-periodic solution of the generating system when $\varepsilon = 0$.

Remark: The Theorem 1 is valid when all or part of the variables satisfy the „jump" conditions.

References

[1] **Paskalev, G. :** *Periodic solutions of non-autonomous linear systems of second order differential equations with constant coefficients.* Trav. sci. , Univ. de Plovdiv, vol. **12**, fasc. 1, 1974 (in Bulgarien)

received: November 24, 1993

Authors:

S. G. Hristova
University of Plovdiv
„Paissii Hilendarski“
Math. Fac.
Zar Assen 24
4000 Plovdiv
Bulgaria

M. B. Kapitanova
University of Plovdiv
„Paissii Hilendarski“
Math. Fac.
Zar Assen 24
4000 Plovdiv
Bulgaria

S. I. Kostadinov
University of Plovdiv
„Paissii Hilendarski“
Math. Fac.
Zar Assen 24
4000 Plovdiv
Bulgaria

Rostock. Math. Kolloq. **50**, 9–18 (1997)

Subject Classification (AMS)
34A40 45J05

S. G. Hristova
S. I. Kostadinov

Integral inequalities for piecewise – continuous functions of many variables

1 Introduction

There exist various generalizations of Bellman – Gronwall – Bihari inequalities [1] – [6]. These generalizations have been widely quoted by many mathematicians inproving uniqueness, boundedness, continuous dependence and stability.

In this paper we establish integral inequalities for piecewise – continuous functions of many variables. These generalizations are motivated by specific applications to the theory of impulsive partial differential equations.

2 Preliminaries

Let $x^{(k)} \in \mathbb{R}^n$, $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$, $k = 0, 1, \dots$ are fixed points so that for $i = \overline{1, n}$ the inequalities $x_i^{(k)} < x_i^{(k+1)}$, $k = 0, 1, \dots$ hold and $\lim\limits_{k \to \infty} x_i^{(k)} = +\infty$

Define the sets

$$D_k = \left\{ x = (x_1, \dots, x_n) : x_i \in \left[x_i^{(k)}, x_i^{(k+1)} \right], i = 1, \dots, n \right\}, k = 0, 1, 2, \dots, \ \Omega = \bigcup_{k=0}^{\infty} D_k.$$

Let $x, y \in \mathbb{R}^n$, $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$. We say that $x \geq (\leq y)$ if for $i = \overline{1, n}$ the inequalities $x_i \geq (\leq) y_i$ hold.

Let $x, y \in \mathbb{R}^n$, $x \leq y$. Denote

$$B(x,y) = [x_1, y_1] \times \cdots \times [x_n, y_n] \ ,$$
$$\int\limits_{x_1}^{y_1} \int\limits_{x_2}^{y_2} \ldots \int\limits_{x_n}^{y_n} ds_n ds_{n-1} \ldots ds_1 = \int\limits_{B(x,y)} \mathrm{ds} \ , \quad ds' = ds_n \ldots ds_{k+1} ds_{k-1} \ldots ds_1 \ ,$$
$$B_k(x,y) = [x_1, y_1] \times \cdots \times [x_{k-1}, y_{k-1}] \times [x_{k+1}, y_{k+1}] \times \cdots \times [x_n, y_n] \ ,$$

where $k : 1 \leq k \leq n$ is an arbitrary natural number, $(y_k, x') = (x_1, x_2, \ldots, x_{k-1}, y_k, x_{k+1}, \ldots, x_n)$.

We will use the following lemmas:

Lemma 1 */theorem* 1 [5]*/. Let for $s \in B(x_0, x)$ and x_0, $x \in R^n$, $x > x_0$ the following conditions be fulfilled:*

1. *The functions $u(x)$, $v(x) : R^n \to R$ and $\mu(x,s) : R^n \times R^n \to R$ are nonegative and continuous.*

2. *The inequality*

$$u(x) \leq a + \beta \int\limits_{B(x_0,x)} \left[v(y)u(y) + \int\limits_{B(x_0,y)} \mu(y,s)u(s)ds \right] dy$$

holds, where a and β are nonnegative constants. Then for $x \geq x_0$ the inequality

$$u(x) \leq a \exp\left\{ \beta \int\limits_{B(x_0,x)} \left[v(y) + \int\limits_{B(x_0,y)} \mu(y,s)ds \right] dy \right\}$$

holds.

Lemma 2 */theorem* 2 [6]*/. Let the functions $u(x) : R^n \to R^1$ and $f(x) : R^n \to R^1$ are continuous and nonnegative for $x_0 \leq x$, $x_0 \in R^n$. Let the inequality*

$$u(x) \leq c + \int\limits_{B(x_0,x)} f(s)\phi(u(s))ds$$

holds, where $c = \mathrm{const} > 0$,

$$F(t) = \int\limits_c^t \frac{d\tau}{\phi(\tau)} \ , \quad t \in (c, \sigma) \ , \qquad \int\limits_{B(x_0,x)} f(s)ds \subset \mathrm{Dom}(F^{-1})$$

and the function $\phi(\cdot) : R \to R$ is positive and increasing.

Then for $x \geq x_0$ the inequality

$$u(x) \leq F^{-1}\left(\int\limits_{B(x_0,x)} f(s)ds\right)$$

holds, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$.

3 Main results

Theorem 1 *Let the following conditions be fulfilled:*

1. *The function $u(x) : R^n \to R^1$ is defined, nonnegative, and piece – continuous for $x \geq x^{(0)}$ with points of interruption in the points $x^{(k)}$, $u(x^{(k)}) = u(x^{(k)} - 0)$ and for $k = 1, 2, \dots$ the inequalities $u(x^{(k)} + 0) - u(x^{(k)}) < \infty$ hold.*
2. *The functions $v(x) : R^n \to R$ and $w(x, y) : R^n \times R^n \to R$ are continuous and nonnegative for $x, y \geq x^{(0)}$ and the equalities $v(x) = 0$ for $x \in R^n|\Omega$ and $w(x, y) = 0$ for $x, y \overline{\in} \Omega$ hold.*
3. *The constants a, β and γ_k $(k = 1, 2, \dots)$ are nonnegative.*
4. *For $x \in \Omega$ the inequality*

$$u(x) \leq a + \beta \int\limits_{B(x^{(0)},x)} \left[v(s)u(s) + \int\limits_{B(x^{(0)},s)} w(s, \xi)u(\xi)d\xi\right] ds + \sum_{k:x^{(k)}<x} \gamma_k u(x^{(k)}) \tag{1}$$

holds.

Then for $x \in \Omega$ the inequality

$$u(x) \leq \prod_{k:x^{(k)}<x} (1 + \gamma_k) \exp\left\{\beta \int\limits_{B(x^{(0)},x)} \left[v(s) + \int\limits_{B(x^{(0)},s)} w(s, \xi)d\xi\right] ds\right\} \tag{2}$$

holds.

Proof. Let $x \in \mathcal{D}_0$. Then the inequality (1) can be written in the form

$$u(x) \leq a + \beta \int\limits_{B(x^{(0)},x)} \left[v(s)u(s) + \int\limits_{B(x^{(0)},s)} w(s, \xi)u(\xi)d\xi\right] ds\ . \tag{3}$$

According to Lemma 1 for $x \in \mathcal{D}_0$ the inequality

$$u(x) \leq a \exp\left\{\beta \int\limits_{B(x^{(0)},x)} \left[v(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)d\xi\right] ds\right\} \tag{4}$$

holds.

We assume the inequality (2) holds for $x \in \mathcal{D}_{k-1}$. We will prove the inequality (2) holds for $x \in \mathcal{D}_k$. According to the condition 2 of the theorem 1 for $x \in \mathcal{D}_k$ the inequality (1) can be written in the form

$$u(x) \leq \alpha + \beta \int\limits_{B(x^{(0)},x)} \left[v(s)u(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)u(\xi)d\xi\right] ds\ , \tag{5}$$

where

$$\alpha = a + \beta \int\limits_{\bigcup\limits_{i=0}^{k-1} \mathcal{D}_i} \left[v(s)u(s) + \int\limits_{B(x^{(i)},s)} w(s,\xi)u(\xi)d\xi\right] ds + \sum_{i=1}^{k} \gamma_i u(x^{(i)})\ . \tag{6}$$

By the assuming the inequality

$$\begin{aligned}
\alpha \leq{} & a + \beta \sum_{i=0}^{k-1} \int\limits_{\mathcal{D}_i} \left[v(s) a \prod_{j=1}^{i}(1+\gamma_j) \exp\left\{\beta \int\limits_{B(x^{(0)},s)} \left[v(y) + \int\limits_{B(x^{(0)},y)} w(y,\xi)d\xi\right] ds\right\} + \right. \\
& \left. + \int\limits_{B(x^{(i)},s)} w(s,\xi) a \prod_{j=1}^{i}(1+\gamma_j) \exp\left\{\beta \int\limits_{B(x^{(0)},s)} \left[v(y) + \int\limits_{B(x^{(0)},y)} w(y,\eta)d\eta\right] dy\right\} d\xi\right] ds + \\
& + \sum_{i=1}^{k} \gamma_i a \prod_{j=1}^{i-1}(1+\gamma_j) \exp\left\{\beta \int\limits_{B(x^{(0)},x^{(i)})} \left[v(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)d\xi\right] ds\right\} \leq \\
\leq{} & a + a\beta \sum_{i=0}^{k-1} \left\{\int\limits_{\mathcal{D}_i} \left[v(s) \prod_{j=1}^{i}(1+\gamma_j) \cdot \prod_{j=1}^{i-1} f_j \cdot g_i(s) + \right.\right. \\
& \left.\left. + \int\limits_{B(x^{(i)},s)} w(s,\xi) \prod_{j=1}^{i}(1+\gamma_j) \cdot \prod_{j=1}^{i-1} f_j \cdot g_i(\xi)d\xi\right] ds\right\} + \\
& + a \sum_{i=1}^{k} \gamma_i \prod_{j=1}^{i-1}(1+\gamma_j) \cdot \prod_{j=0}^{i-1} f_i\ ,
\end{aligned} \tag{7}$$

holds, where

$$\gamma_0 = 0 \ ,$$
$$f_j = \exp\left\{\beta \int\limits_{\mathcal{D}_j} \left[v(s) + \int\limits_{B(x^{(j)},s)} w(s,\xi)d\xi \right] ds \right\} \ ,$$
$$g_j(x) = \exp\left\{\beta \int\limits_{B(x^{(j)},x)} \left[v(s) + \int\limits_{B(x^{(j)},s)} w(s,\xi)d\xi \right] ds \right\} \ , \qquad x > x^{(j)} \ , \quad j = 0, 1, \ldots$$

and the symbols $\prod\limits_{i=1}^{0}(1+\gamma_i)$ and $\prod\limits_{i=1}^{-1} f_i$ denote the number 1.

From the inequality (7) and the monotonicity of the exponential function we obtain the inequality

$$\begin{aligned} \alpha \le a \Bigg\{ 1 + \sum_{i=0}^{k-1} \left[\prod_{j=1}^{i}(1+\gamma_j) \cdot \left(\prod_{j=1}^{i-1} f_j \right) \cdot (g_i(x^{(i+1)}) - 1) \right] + \\ + \sum_{i=1}^{k} \gamma_i \left(\prod_{j=1}^{i-1}(1+\gamma_j) \right) \cdot \left(\prod_{j=0}^{i-1} f_j \right) \Bigg\} = \\ = a \prod_{j=1}^{k}(1+\gamma_j) \cdot \left(\prod_{j=0}^{k-1} f_j \right) \ . \end{aligned} \tag{8}$$

From the inequalities (5), (8) and Lemma 1 we obtain the inequality

$$\begin{aligned} u(x) \le \alpha \exp\left\{\beta \int\limits_{B(x^{(k)},x)} \left[v(s) + \int\limits_{B(x^{(k)},s)} w(s,\xi)d\xi \right] ds \right\} \le \\ \le a \prod_{j=1}^{k}(1+\gamma_j) \cdot \exp\left\{\beta \int\limits_{B(x^{(0)},x)} \left[v(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)d\xi \right] ds \right\} \ . \end{aligned}$$

Remark 1. For $n = 1$ and $w(s,\xi) \equiv 0$ theorem 1 yields one result of [7].

Theorem 2 *Let the following conditions be fulfilled:*

1. *Condition* 1 *and* 2 *of the theorem* 1 *are satisfied.*

2. *The function* $g(x) : R^n \to R$ *is continuous, positive and non–decreasing for* $x \in \Omega$*.*

3. *For $x \in \Omega$ the inequality*

$$u(x) \leq g(x) + \beta \int\limits_{B(x^{(0)},x)} \left[v(s)u(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)u(\xi)d\xi \right] ds + \sum_{k:x^{(k)}<x} \gamma_k u(x^{(k)}) \ , \tag{9}$$

   holds, where the constants β and γ_i $(i = 1, 2, \dots)$ are nonnegative.

   Then for $x \in \Omega$ the inequality

$$u(x) \leq g(x) \prod_{k:x^{(k)}<x} (1+\gamma_k) \cdot \exp \left\{ \beta \int\limits_{B(x^{(0)},x)} \left[v(s) + \int\limits_{B(x^{(0)},s)} w(s,\xi)d\xi \right] ds \right\} \tag{10}$$

   holds.

Proof We devide the both sides of the inequality (9) by the function $g(x)$ and use the condition 2 of the theorem 2. We obtain

$$\frac{u(x)}{g(x)} \leq 1 + \beta \int\limits_{B(x^{(0)},x)} \left[v(s)\frac{u(s)}{g(s)} + \int\limits_{B(x^{(0)},s)} w(s,\xi)\frac{u(\xi)}{g(\xi)}d\xi \right] ds + \sum_{k:x^{(k)}<x} \gamma_k \frac{u(x^{(k)})}{g(x^{(k)})} \ . \tag{11}$$

From the inequality (11) and the theorem 1 it follows the inequality (10).

Theorem 3 *Let the following conditions be fulfilled:*

1. *The condition* 1 *of the theorem* 1 *is satisfied.*
2. *The function $v(x) : R^n \to R$ is continuous, non–negative, and $v(x) \equiv 0$ for $x \in R^n \backslash \Omega$.*
3. *The function $g(t) : R \to R$ is continuous, positive and non–decreasing for $t > 0$.*
4. *The functions $f_k(t) : R \to R$ $(k = 1, 2, \dots)$ are positive and non–decreasing for $t > 0$.*
5. *For every natural number k there exists a constant $M_k > 0$ so that for $t \geq s > 0$ the inequality*

$$G(\tau_k(t)) - G(\tau_k(s)) \leq M_k \left[G(t) - G(s) \right]$$

   holds, where

$$\tau_k(t) = t + f_k(t) \ , \quad G(t) = \int\limits_{t_0}^{t} \frac{d\xi}{g(\xi)} \ , \quad t > t_0 \ , \quad t_0 > 0 \ .$$

6. *For $x \geq x^{(0)}$ the inequality*

$$u(x) \leq a + \int\limits_{B(x^{(0)},x)} v(s)g(u(s))ds + \sum_{k:x^{(k)}<x} f_k(u(x^{(k)})) \ . \tag{12}$$

holds, where the constant a is positive.

Then for $x \in \Omega_1$ the inequality

$$u(x) \leq G^{-1}\left[G(R(x)) + \int\limits_{B(x^{(0)},x)} \left(\prod_{s<x^{(j)}<x} M_j\right) v(s)ds\right] \tag{13}$$

holds, where $R_0(x) \equiv a$ for $x \in \mathcal{D}_0$, $R(x) \equiv R_k = \tau_k(R_{k-1})$ for $x \in \mathcal{D}_k$,

$$\Omega_1 = \sup\left\{ x \geq x^0 : G(R(y)) + \right.$$

$$\left. + \int\limits_{B(x^{(0)},y)} \left(\prod_{s<x^{(j)}<y} M_j\right) v(s)ds \in \mathrm{Dom}\,(G^{-1}) \ , \quad y \in B(x^{(0)},x) \right\} \ .$$

Proof Let $x \in \mathcal{D}_0$. Then the inequality (12) can be written in the form

$$u(x) \leq a + \int\limits_{B(x^{(0)},x)} v(s)g(u(s))ds \ . \tag{14}$$

According to Lemma 2 for $x \in \mathcal{D}_0 \cap \Omega_1$ the inequality

$$u(x) \leq G^{-1}\left[G(a) + \int\limits_{B(x^{(0)},x)} v(s)ds\right] \ ,$$

holds.

We assume the inequality (13) holds for $x \in \mathcal{D}_{k-1} \cap \Omega_1$. We will prove that this inequality holds for $x \in \mathcal{D}_k \cap \Omega_1$. According to the condition 2 of theorem 3 the inequality (12) can be written in the form

$$u(x) \leq a + \sum_{i=0}^{k-1} \int\limits_{\mathcal{D}_i} v(s)g(u(s))ds + \int\limits_{B(x^{(k)},x)} v(s)g(u(s))ds + \sum_{i=1}^{k} f_i(u(x^{(i)})) \equiv w(x) \ .$$

The function $w(x)$ satisfies the inequality $u(x) \leq w(x)$. Then

$$w(x) \leq a + \sum_{i=0}^{k-1} \int\limits_{\mathcal{D}_i} v(s)g(w(s))ds + \int\limits_{B(x^{(k)},x)} v(s)g(w(s))ds + \sum_{i=1}^{k} f_i(w(x^{(i)})) \ . \tag{15}$$

We fix an arbitrary natural number $j : 1 \leq j \leq n$. Then the inequalities

$$\frac{\partial}{\partial x_j} w(x) \leq \int\limits_{B_j(x^{(k)},x)} v(x_j, s') g(w(x_j, s')) ds' \ , \tag{16}$$

$$\begin{aligned} w(x_j^{(k)} + 0, (x^{(k)})') &\leq w(x_j^{(k)} - 0, (x^{(k)})') + f_k(w(x_j^{(k)} - 0, (x^{(k)})')) = \\ &= w(x^{(k)}) - f_k(w(x^{(k)})) = \tau_k(w(x^{(k)})) \ , \end{aligned} \tag{17}$$

hold.

From the inequalities (16) and (17) and Bihari's lemma [1] it follows that the inequality

$$\begin{aligned} w(x) &\leq G^{-1} \left[G(w(x_j^{(k)} + 0, (x^{(k)})')) + \int\limits_{x_j^{(k)}}^{x_j} \int\limits_{B(x^{(k)},x)} v(x_j, s') ds' ds_j \right] = \\ &= G^{-1} \left[G(w(x_j^{(k)} + 0, (x^{(k)})')) + \int\limits_{B(x^{(k)},x)} v(s) ds \right] \end{aligned} \tag{18}$$

holds.

From the properties of the function $G(x)$ and the inequalities (17) and (18) it follows that the inequality

$$G(w(x)) \leq G(\tau_k(w(x^{(k)}))) + \int\limits_{B(x^{(k)},x)} v(s) ds \tag{19}$$

holds.

If the inequality $w(x^{(k)}) \leq R_{k-1}$ holds then from (18) it follows the inequality (13). Otherwise we obtain

$$\begin{aligned} G(\tau_k(w(x^{(k)}))) - G(R_k) &\leq M_k \left[G(w(x^{(k)})) - G(R_{k-1}) \right] \leq \\ &\leq M_k \int\limits_{B(x^{(0)},x)} \left(\prod_{s < x^{(j)} < x} M_j \right) v(s) ds \ . \end{aligned} \tag{20}$$

According to inequalities (19), (20) and the condition 2 of theorem 3 it follows that the inequality

$$\begin{aligned} G(w(x)) &\leq G(R_k) + \int\limits_{B(x^{(k)},x)} v(s) ds + M_k \int\limits_{B(x^{(0)},x)} \left(\prod_{s < x^{(j)} < x} M_j \right) v(s) ds = \\ &= G(R_k) + \int\limits_{B(x^{(0)},x)} \left(\prod_{s < x^{(j)} < x} M_j \right) v(s) ds \end{aligned}$$

holds.

Theorem 4 *Let the following conditions be fulfilled:*

1. *The conditions* 1, 2 *and* 5 *of theorem* 3 *are satisfied.*

2. *For* $x > x^{(0)}$ *the inequality*

$$u(x) \leq a + \int\limits_{B(x^{(0)},x)} v(s)u^m(s)ds + \sum_{i:x^{(i)}<x} \beta_i\, u(x^{(i)}) \ , \tag{21}$$

holds, where $m > 0$, $m \neq 1$ *and* $\beta_i \geq 0$ $(i = 1, 2, \dots)$ *are constants.*

Then for $x \in \Omega_2$ *the inequality*

$$u(x) \leq a \prod_{k:x^{(k)}<x} (1+\beta_k) \left\{ 1 + (1-m) \int\limits_{B(x^{(0)},x)} \left[\prod_{s<x^{(k)}<x} a(1+\beta_k)^{m-1} \right] v(s)ds \right\} \tag{22}$$

holds, where

$$\Omega_2 = \sup \left\{ x \geq x_0 : (m-1) \int\limits_{B(x^{(0)},y)} \left(\prod_{s<x^{(k)}<y} a(1+\beta_k)^{m-1} \right) v(s)ds < 1 \ , \right.$$
$$\left. y \in B(x^{(0)},x) \right\} \ .$$

Proof Define the functions $g(s) = s^m$, $\tau_k(s) = (1+\beta_k)s$, $R(x) = a \prod_{x^{(i)}<x} (1+\beta_i)$, $G(t) = \frac{t^{1-m}}{1-m} + \sigma$, $(\sigma > 0)$.

Then $M_i = (1+\beta_i)^{1-m}$. According to (21) and theorem 3 the inequality (22) holds.

References

[1] **Bihari, L. :** *A generalization of lemma of Bellman and its applications..* Acta Math. J. **10**, 643-647 (1943)

[2] **Beckenbach, E. F.**, and **Bellman, R. :** *Inequalities..* Springer – Verlag, Berlin / Heidelberg 1961

[3] **Lakshmikantham, V.**, and **Leela, S. :** *Differential and Integral inequalities..* Academic Press, New York/London 1969

[4] **Beesack, P. R. :** *Gronwall inequalities..* Carleton Math. Lecture Notes No. 11 (May, 1975)

[5] **Hristova, S. G.** , and **Bainov, D. D. :** *Some integral inequalities..* Glasnik Matematićki, vol. **13 (33)**, 249-253 (1978) (in russian)

[6] **Hristova, S. G.**, and **Bainov, D. D. :** *A generalization of Bihari's inequality..* Izv. AN Kaz. SSR, No. 5, 88-89 (1978) (in russian)

[7] **Samoilenko, A. M.**, and **Perestyuk, N. A. :** *Differential equations with impulsive effect..* Vista skola, Kiew (1987), pp. 284 (in russian)

received: January 21, 1994

Authors:

S. G. Hristova
University of Plovdiv
„Paissii Hilendarski“
Math. Fac.
Zar Assen 24
4000 Plovdiv
Bulgaria

S. I. Kostadinov
University of Plovdiv
„Paissii Hilendarski“
Math. Fac.
Zar Assen 24
4000 Plovdiv
Bulgaria

Rostock. Math. Kolloq. **50**, 19–30 (1997)

Subject Classification (AMS)
41A10 15A23 15A18

Lothar Berg and Gerlind Plonka

Refinement of Vectors of Bernstein Polynomials

ABSTRACT. For the case of Bernstein polynomials, the refinement mask is calculated recursively, and the refinement matrices are given explicitely. Moreover, the eigenvectors of the transposed refinement matrices are constructed, whereas the eigenvectors of the refinement matrices themselves can be determined by a theorem of Micchelli and Prautzsch.

1 INTRODUCTION

Let $n \in \mathbb{N}$ and let $\boldsymbol{b}^n(t) := (b_0^n(t), \ldots, b_n^n(t))^T$ be a vector of *uniformly refinable* real functions on $[0,1]$, i.e., there are $(n+1) \times (n+1)$ matrices $\boldsymbol{A}_0^n, \ldots, \boldsymbol{A}_{k-1}^n$ $(k \in \mathbb{N};\ k \geq 2)$ such that

$$\boldsymbol{b}^n\left(\frac{t+m}{k}\right) = \boldsymbol{A}_m^n\, \boldsymbol{b}^n(t) \tag{1.1}$$

is satisfied for $m = 0, \ldots, k-1$ and $t \in [0,1]$. These equations are called *refinement equations*, and the matrices $\boldsymbol{A}_m^n$ *refinement matrices* (cf. Micchelli and Prautzsch [5]). It is well–known that the refinement equations (1.1) are closely connected with corresponding subdivision algorithms which provide important techniques for the fast generation of curves (cf. [3, 5]). In [6] and [8], some applications of such equations in the theory of wavelets are discussed.

For polynomials $b_i^n(t)$ $(i = 0, \ldots, n)$ spanning the vector space of all polynomials of degree n, the matrices $\boldsymbol{A}_m^n$ in (1.1) always exist and are uniquely determined.

Here, we consider these matrices in the case of Bernstein polynomials

$$b_i^n(t) = \binom{n}{i} t^i\, (1-t)^{n-i}, \qquad i = 0, \ldots, n, \tag{1.2}$$

and study their spectral properties. In particular, we prove a recursion formula for the *refinement mask* of $\boldsymbol{b}^n$

$$\boldsymbol{A}^n(z) := \frac{1}{k} \sum_{m=0}^{k-1} \boldsymbol{A}_m^n \, z^m, \quad z \in \mathbb{C}. \tag{1.3}$$

Furthermore, we derive explicit formulas for the entries of the refinement matrices $\boldsymbol{A}_m^n$ as well as for their eigenvalues and corresponding eigenvectors. Note that for the special case $k = 2$, the corresponding subdivision algorithm is the de Casteljau algorithm (cf. [2]).

2 RECURSIVE COMPUTATION OF THE REFINEMENT MASK

First, we derive a simple recursion formula for the Fourier transform of the vector $\boldsymbol{b}^n(t)$ of Bernstein polynomials $b_i^n(t)$ ($t \in [0,1]$; $i = 0, \ldots, n$). For convenience, outside of $[0,1]$ the polynomials are defined by zero. Denoting the Fourier transform of a function $f \in L^2(\mathbb{R})$ by

$$\hat{f}(u) = \int_{\infty}^{\infty} f(t)\, e^{-iut} \, \mathrm{d}t, \quad u \in \mathbb{R},$$

we have:

Lemma 2.1 *For $n = 0$,*

$$\hat{\boldsymbol{b}}^0(u) = \hat{b}_0^0(u) = \frac{1 - e^{-iu}}{iu}. \tag{2.1}$$

For $n > 0$, the following recursion formula holds:

$$iu\, \hat{\boldsymbol{b}}^n(u) = \boldsymbol{C}_n(e^{-iu}) \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix}, \qquad u \in \mathbb{R}, \tag{2.2}$$

where the matrix $\boldsymbol{C}_n(z)$ with $z \in \mathbb{C}$ is an $(n+1) \times (n+1)$-matrix of the form

$$\boldsymbol{C}_n(z) := n \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & 1 & -1 \\ -z & 0 & \ldots & 0 & 1 \end{pmatrix}, \quad n \geq 1. \tag{2.3}$$

Proof: We put $b_{-1}^{n-1}(t) := \delta(t)/n$ and $b_n^{n-1}(t) := \delta(t-1)/n$ $(n \geq 1)$, where δ denotes the Dirac distribution. Then the known formula for the derivative of Bernstein polynomials $b_i^n(t)$ $(i = 1, \ldots, n-1)$

$$\mathrm{D}b_i^n(t) = n(b_{i-1}^{n-1}(t) - b_i^{n-1}(t))$$

can also be used for $i = 0$ and $i = n$, in view of the jumps of $b_0^n(t)$ at $t = 0$, and $b_n^n(t)$ at $t = 1$. Hence, we obtain for the vector $\boldsymbol{b}^n(t)$

$$\mathrm{D}\boldsymbol{b}^n(t) = n \begin{pmatrix} b_{-1}^{n-1} \\ \boldsymbol{b}^{n-1}(t) \end{pmatrix} - n \begin{pmatrix} \boldsymbol{b}^{n-1}(t) \\ b_n^{n-1} \end{pmatrix}.$$

Taking the Fourier transform, we infer

$$iu\,\hat{\boldsymbol{b}}^n(u) = n \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix} - n \begin{pmatrix} \hat{\boldsymbol{b}}^{n-1}(u) \\ e^{-iu}/n \end{pmatrix} = \boldsymbol{C}_n(e^{-iu}) \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix}. \quad \blacksquare$$

Remark 2.2 1. Note that

$$\det \boldsymbol{C}_n(z) = n^{n+1}\,(1-z).$$

2. The Bernstein polynomials $b_i^n(t)$ $(i = 0, \ldots, n)$ on $[0,1]$ can also be considered as B–splines defined by the multiple knots

$$\underbrace{0, \ldots, 0}_{n-i+1}, \underbrace{1, \ldots, 1}_{i+1}.$$

Generalizing this definition, we find that b_{-1}^n is determined by the knots $\underbrace{0, \ldots, 0}_{n+2}$ and analogously b_{n+1}^n by $\underbrace{1, \ldots, 1}_{n+2}$. Thus, the above definition of b_{-1}^n and b_{n+1}^n according to the distribution theory makes sense, also from this point of view (cf. [7]).

Example 2.3 For $n = 0$, we have (2.1). For $n = 1$ and $n = 2$, we find

$$\begin{aligned}
\hat{\boldsymbol{b}}^1(u) &= \frac{1}{(iu)^2} \begin{pmatrix} 1 & -1 \\ -e^{-iu} & 1 \end{pmatrix} \begin{pmatrix} iu \\ 1 - e^{-iu} \end{pmatrix} \\
&= \frac{1}{(iu)^2} \begin{pmatrix} iu - 1 + e^{-iu} \\ 1 - (1+iu)e^{-iu} \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
\hat{\boldsymbol{b}}^2(u) &= \frac{2}{(iu)^3} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -e^{-iu} & 0 & 1 \end{pmatrix} \begin{pmatrix} (iu)^2/2 \\ iu - 1 + e^{-iu} \\ 1 - (1+iu)e^{-iu} \end{pmatrix} \\
&= \frac{1}{(iu)^3} \begin{pmatrix} (iu)^2 - 2iu + 2 - 2e^{-iu} \\ 2(iu-2) + 2(2+iu)e^{-iu} \\ 2 - (2 + 2iu + (iu)^2)e^{-iu} \end{pmatrix}.
\end{aligned}$$

Now, we shall investigate the refinement equations (1.1) for the vector of Bernstein polynomials. Let $k \in \mathbb{N}$, $k \geq 2$ be given. After a substitution, we find that (1.1) is equivalent to the equation

$$\boldsymbol{b}^n(t) = \sum_{m=0}^{k-1} \boldsymbol{A}_m^n \, \boldsymbol{b}^n(kt - m), \quad t \in [0,1],$$

since at the right–hand side at most one term is different from zero. Fourier transform yields

$$\hat{\boldsymbol{b}}^n(u) = \boldsymbol{A}^n(e^{-iu/k}) \, \hat{\boldsymbol{b}}^n(u/k) \tag{2.4}$$

with $\boldsymbol{A}^n$ defined in (1.3). The refinement mask $\boldsymbol{A}^n$ can be characterized in the following way:

Theorem 2.4 *For $n = 0$, we have*

$$\boldsymbol{A}^0(z) = \frac{1}{k} \sum_{m=0}^{k-1} z^m = \frac{1 - z^k}{k\,(1-z)}. \tag{2.5}$$

For $n \geq 1$ and $z \neq 1$, the recursion formula

$$\boldsymbol{A}^n(z) = \frac{1}{k} \, \boldsymbol{C}_n(z^k) \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \boldsymbol{A}^{n-1}(z) \end{pmatrix} \boldsymbol{C}_n(z)^{-1} \tag{2.6}$$

is satisfied, where the matrices $\boldsymbol{C}_n$ are given in (2.3), and where $\mathbf{0}$ is a zero vector of suitable dimension.

Proof: For $n = 0$, we observe that

$$\boldsymbol{b}^0(t) = b_0^0(t) = 1, \quad t \in [0,1],$$

i.e., we have $\boldsymbol{A}_m^0 = 1$ $(m = 0, \ldots, k-1)$. Formula (1.3) implies that $\boldsymbol{A}^0(z) = \frac{1}{k} \sum_{m=0}^{k-1} z^m$, so that (2.5) is proved.
Now let $n > 0$. Then from (2.2) and (2.4) we obtain for $u \neq 0$

$$\begin{aligned} \frac{1}{iu} \boldsymbol{C}_n(e^{-iu}) \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix} &= \hat{\boldsymbol{b}}^n(u) = \boldsymbol{A}^n(e^{-iu/k}) \, \hat{\boldsymbol{b}}^n(u/k) \\ &= \boldsymbol{A}^n(e^{-iu/k}) \frac{k}{iu} \boldsymbol{C}_n(e^{-iu/k}) \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u/k) \end{pmatrix}. \end{aligned}$$

Hence, since $\boldsymbol{C}_n(z)$ is regular for $z \neq 1$,

$$\begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix} = k \, \boldsymbol{C}_n(e^{-iu})^{-1} \, \boldsymbol{A}^n(e^{-iu/k}) \, \boldsymbol{C}_n(e^{-iu/k}) \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u/k) \end{pmatrix}.$$

On the other hand, (2.4) with $n-1$ instead of n implies

$$\begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u) \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \boldsymbol{A}^{n-1}(e^{-iu/k}) \end{pmatrix} \begin{pmatrix} 1/n \\ \hat{\boldsymbol{b}}^{n-1}(u/k) \end{pmatrix}.$$

In these two equations, all entries are rational functions in u and $z = e^{-iu/k}$. Since z is a transcendent function in u, the both equations are identities too, if we consider u and z as independent variables. Moreover, the components of the vectors $(1/n,\, \hat{\boldsymbol{b}}^{n-1}(u))^T$ and $(1/n,\, \hat{\boldsymbol{b}}^{n-1}(u/k))^T$ are linearly independent in u, and the entries of the matrices can be considered as constants with respect to u. This implies, that the corresponding matrices are equal, i.e.,

$$\begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \boldsymbol{A}^{n-1}(e^{-iu/k}) \end{pmatrix} = k\, \boldsymbol{C}_n(e^{-iu})^{-1}\, \boldsymbol{A}^n(e^{-iu/k})\, \boldsymbol{C}_n(e^{-iu/k}),$$

so that also (2.6) is proved. ■

Remark 2.5 1. For $z = 1$, the refinement mask $\boldsymbol{A}^n(z)$ can be found by a limiting process $z \to 1$, since the elements of $\boldsymbol{A}^n(z)$ are continuous in z.
2. From the recursion formula in Theorem 2.4 and Remark 2.2, we can easily derive the determinant of $\boldsymbol{A}^n(z)$:

$$\det\, \boldsymbol{A}^n(z) = \left(\frac{1 - z^k}{k\,(1-z)} \right)^{n+1}.$$

3. For $z \neq 1$, the inverse matrix $\boldsymbol{C}_n(z)^{-1}$ is explicitly given by

$$\boldsymbol{C}_n(z)^{-1} = \frac{1}{n(1-z)} \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ z & 1 & \dots & 1 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ z & z & \ddots & 1 & 1 \\ z & z & \dots & z & 1 \end{pmatrix}.$$

4. For $n = 1$, (2.6) simplifies to

$$\boldsymbol{A}^1(z) = \frac{1}{k^2\,(1-z)^2} \begin{pmatrix} 1 & -1 \\ -z^k & 1 \end{pmatrix} \begin{pmatrix} k(1-z) & 0 \\ 0 & 1 - z^k \end{pmatrix} \begin{pmatrix} 1 & 1 \\ z & 1 \end{pmatrix}.$$

This formula can be generalized in the following way. By means of the direct sum of two quadratic matrices $\boldsymbol{A} \oplus \boldsymbol{B} := \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{B} \end{pmatrix}$ with a suitable zero matrix $\boldsymbol{0}$, and the ν–dimensional unit matrix $\boldsymbol{I}_\nu$, we define for $\nu = 0, \ldots, n$,

$$\boldsymbol{X}_\nu^n(z) := \boldsymbol{I}_{n-\nu} \oplus \frac{1}{\nu} \boldsymbol{C}_\nu(z^k), \qquad \boldsymbol{Y}_\nu^n(z) := \boldsymbol{I}_{n-\nu} \oplus \nu(1-z)\, \boldsymbol{C}_\nu(z)^{-1},$$

where $\boldsymbol{I}_0$ is dummy. Then (2.6) immediately implies

$$\boldsymbol{A}^n(z) = \frac{1}{k^{n+1}\,(1-z)^{n+1}}\, \boldsymbol{X}_n^n(z) \ldots \boldsymbol{X}_1^n(z)\, \boldsymbol{B}_n(z)\, \boldsymbol{Y}_1^n(z) \ldots \boldsymbol{Y}_n^n(z) \tag{2.7}$$

with

$$\boldsymbol{B}_n(z) := \operatorname{diag}\left(k^n(1-z)^n,\, k^{n-1}(1-z)^{n-1}, \ldots, k(1-z),\, 1-z^k\right).$$

In particular, we have

$$\begin{aligned} \boldsymbol{A}^2(z) &= \frac{1}{k^3(1-z)^3} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1+z^k & -2 \\ -z^k & -z^k & 1 \end{pmatrix} \begin{pmatrix} k^2(1-z)^2 & 0 & 0 \\ 0 & k(1-z) & 0 \\ 0 & 0 & 1-z^k \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & 1 & 1 \\ 2z & z+1 & 2 \\ z^2+z & 2z & z+1 \end{pmatrix}. \end{aligned}$$

For the matrix product $\boldsymbol{X}_n^n(z), \ldots, \boldsymbol{X}_1^n(z)$ occuring in (2.7) we easily find

$$\boldsymbol{X}_n^n(z) \ldots \boldsymbol{X}_1^n(z) = \left((-1)^{i+j} \binom{j}{i} - (-1)^{i+n} \binom{j}{n-i} \epsilon_{jn} z^k\right)_{i,j=0,\ldots,n},$$

where $\epsilon_{jn} := 1 - \delta_{jn}$ with the Kronecker symbol δ_{jn}. The product $\boldsymbol{Y}_1^n(z) \ldots \boldsymbol{Y}_n^n(z)$ has not such a simple explicit representation.

3 EXPLICIT REPRESENTATION OF REFINEMENT MATRICES

The determination of the refinement matrices $\boldsymbol{A}_m^n$ $(m = 0, \ldots, k-1)$ by means of the refinement mask (1.3) is not quite easy, so that we shall give an explicit representation of them in this section. By

$$\boldsymbol{A}_m^n = (a_{ij}), \quad i, j = 0, \ldots, n, \tag{3.1}$$

we introduce the entries a_{ij} of $\boldsymbol{A}_m^n$, which also depend on n, k and m.

Theorem 3.1 *The entries a_{ij} $(i, j = 0, \ldots, n)$ of $\boldsymbol{A}_m^n$ have the representation*

$$a_{ij} = \frac{1}{k^n}\,(k-m)^{n-i}\,m^i \sum_{\nu=0}^{j} \binom{j}{\nu}\binom{n-j}{\nu+i-j}\left(1-\frac{1}{k-m}\right)^{\nu}\left(1+\frac{1}{m}\right)^{j-\nu}. \tag{3.2}$$

Proof: The equations (1.1) and (1.2) imply

$$\binom{n}{i}\left(1-\frac{t+m}{k}\right)^{n-i}\left(\frac{t+m}{k}\right)^{i} = \sum_{j=0}^{n} a_{ij}\binom{n}{j}(1-t)^{n-j}\,t^j. \tag{3.3}$$

In view of

$$\begin{aligned} 1-\frac{t+m}{k} &= \frac{k-m}{k}\left(1-t+\left(1-\frac{1}{k-m}\right)t\right), \\ t+m &= m\left(1-t+\left(1+\frac{1}{m}\right)t\right), \end{aligned}$$

for $m \neq 0$, the left–hand side of (3.3) can be written as

$$\begin{aligned} &\binom{n}{i}\left(\frac{k-m}{k}\right)^{n-i}\left(\frac{m}{k}\right)^{i}\sum_{\nu=0}^{n-i}\binom{n-i}{\nu}(1-t)^{n-i-\nu}\left(1-\frac{1}{k-m}\right)^{\nu} t^{\nu} \\ &\times \sum_{\mu=0}^{i}\binom{i}{\mu}(1-t)^{i-\mu}\left(1+\frac{1}{m}\right)^{\mu} t^{\mu}. \end{aligned}$$

Putting $\nu + \mu = j$, we obtain from (3.3) by a comparison of coefficients

$$\binom{n}{j} a_{ij} = \binom{n}{i}\frac{1}{k^n}(k-m)^{n-i}\,m^i \sum_{\mu+\nu=j}\binom{n-i}{\nu}\binom{i}{\mu}\left(1-\frac{1}{k-m}\right)^{\nu}\left(1+\frac{1}{m}\right)^{\mu},$$

and in view of

$$\binom{n}{i}\binom{n-i}{\nu}\binom{i}{\mu} = \binom{n}{j}\binom{j}{\nu}\binom{n-j}{\nu+i-j},$$

finally (3.2).

According to $\binom{n-j}{\nu+i-j} = 0$ for $\nu < j - i$, formula (3.2) makes also sense for $m = 0$, where only the term with $\nu = j - i$ remains:

$$a_{ij} = \frac{1}{k^j}\binom{j}{i}(k-1)^{j-i} \quad \text{for} \quad m = 0. \qquad \blacksquare \tag{3.4}$$

Analogously, for $m = k - 1$, only the term with $\nu = 0$ remains:

$$a_{ij} = \frac{1}{k^{n-j}}\binom{n-j}{i-j}(k-1)^{i-j} \quad \text{for} \quad m = k-1. \tag{3.5}$$

The formulas (3.4) and (3.5) show that $\boldsymbol{A}_0^n$ is an upper and $\boldsymbol{A}_{k-1}^n$ a lower triangular matrix. Moreover, we only have one single term in (3.2) in the four cases
$i = 0$ with $\nu = j$,

$$a_{0j} = \frac{1}{k^n}\,(k-m)^{n-j}\,(k-m-1)^j,$$

$i = n$ with $\nu = 0$,

$$a_{nj} = \frac{1}{k^n}\,m^{n-j}\,(m+1)^j,$$

$j = 0$ with $\nu = 0$,

$$a_{i0} = \frac{1}{k^n}\binom{n}{i}\,(k-m)^{n-i}\,m^i, \tag{3.6}$$

and $j = n$ with $\nu = n - i$,

$$a_{in} = \frac{1}{k^n}\binom{n}{i}\,(k-m-1)^{n-i}\,(m+1)^i. \tag{3.7}$$

Remark 3.2 The equations (3.6) and (3.7) show that the last column of $\boldsymbol{A}_m^n$ equals to the first column of $\boldsymbol{A}_{m+1}^n$ $(m = 0, \dots, k-2)$. This follows also from the statement (a) of Theorem 5.1 in [5], if one uses the fact that $(1, 0, \dots, 0)^T$ and $(0, \dots, 0, 1)^T$ are eigenvectors of the matrices $\boldsymbol{A}_0^n$ and $\boldsymbol{A}_{k-1}^n$, respectively, corresponding to the eigenvalue one. The last fact can easily be seen from $a_{00} = 1$ for $m = 0$, $a_{nn} = 1$ for $m = k-1$, and the triangular structure of $\boldsymbol{A}_0^n$ and $\boldsymbol{A}_{k-1}^n$.

4 SPECTRAL PROPERTIES OF THE REFINEMENT MATRICES

The columns of $\boldsymbol{A}_m^n$ $(m = 0, \dots, k-1)$ possess simple generating functions.

Lemma 4.1 *For an arbitrary parameter λ we have for $j = 0, \dots, n$,*

$$\sum_{i=0}^{n} a_{ij}\,\lambda^i = \frac{1}{k^n}\Big(k + (\lambda-1)(m+1)\Big)^j\Big(k + m(\lambda-1)\Big)^{n-j}. \tag{4.1}$$

Proof: From (3.2) it follows that

$$\begin{aligned}\sum_{i=0}^{n} a_{ij}\,\lambda^i \;&=\; \frac{1}{k^n}\sum_{\nu=0}^{j}\binom{j}{\nu}\left(1-\frac{1}{k-m}\right)^{\nu}\left(1+\frac{1}{m}\right)^{j-\nu}\\ &\quad\times\sum_{i=0}^{n}\binom{n-j}{\nu+i-j}\,(k-m)^{n-i}\,(\lambda m)^i.\end{aligned}$$

Since $\binom{j}{\nu}\binom{n-j}{\nu+i-j} \neq 0$ only for $0 \le j-\nu \le i \le n-\nu \le n$, the sum over i equals to

$$\sum_{l=0}^{n-j} \binom{n-j}{l} (k-m)^{n-l-j+\nu} (\lambda m)^{l+j-\nu} = (k-m)^{\nu} (\lambda m)^{j-\nu} (k-m+\lambda m)^{n-j}$$

with $l = \nu + i - j$, and the assertion follows from

$$\sum_{\nu=0}^{j} \binom{j}{\nu} (k-m-1)^{\nu} (\lambda(m+1))^{j-\nu} = (k-m-1+\lambda(m+1))^{j}. \qquad \blacksquare$$

Corollary 4.2 *Equation (4.1) immediately implies that $\boldsymbol{A}_m^{n\,T}$ has the eigenvalue 1 with the eigenvector $(1,\dots,1)^T$, and the eigenvalue k^{-n} with the eigenvector $(1,\lambda,\dots,\lambda^n)^T$ and $\lambda = 1 + (1-k)/m$, so that $\boldsymbol{A}_m^n$ is a stochastic matrix with respect to the rows.*

By means of (4.1) it is also possible to construct eigenvectors of all eigenvalues k^{-j} ($j = 0,\dots,n$), but it is easier to derive them from the eigenvectors of $\boldsymbol{A}_m^n$, which were found in [5].

Theorem 4.3 *The matrix $\boldsymbol{A}_m^{n\,T}$ has the eigenvalues k^{-j} ($j = 0,\dots,n$) (where $k \ge 2$ is the dilation parameter in the refinement equation (1.1)) with the corresponding eigenvectors*

$$\left(\sum_{\nu=0}^{j} \binom{i}{j-\nu} \binom{n+\nu-j}{\nu} \left(\frac{m}{1-k}\right)^{\nu} \right), \tag{4.2}$$

where $i = 0,\dots,n$ denotes the row index.

Proof: Let $\boldsymbol{D}_n$ be the diagonal matrix of the eigenvalues

$$\boldsymbol{D}_n := \operatorname{diag}(1, k^{-1}, \dots, k^{-n}),$$

and $\boldsymbol{G}_n$, $\boldsymbol{U}_n$ matrices of the corresponding eigenvectors of $(\boldsymbol{A}_m^n)^T$ and $\boldsymbol{A}_m^n$, respectively. Then

$$(\boldsymbol{A}_m^n)^T \boldsymbol{G}_m = \boldsymbol{G}_m \boldsymbol{D}_n, \quad \boldsymbol{A}_m^n \boldsymbol{U}_m = \boldsymbol{U}_m \boldsymbol{D}_n$$

and therefore $(\boldsymbol{A}_m^n)^T \boldsymbol{U}_m^{-T} = \boldsymbol{U}_m^{-T} \boldsymbol{D}_n$ with $\boldsymbol{U}_m^{-T} := (\boldsymbol{U}_m^T)^{-1}$. Hence, we find

$$\boldsymbol{G}_m = \boldsymbol{U}_m^{-T} \boldsymbol{F}_m \tag{4.3}$$

with a diagonal matrix $\boldsymbol{F}_m$. According to Theorem 7.1 in [5] we have (after a correction of a misprint)

$$\boldsymbol{U}_m = \boldsymbol{U}_0 \boldsymbol{T}_n \left(\frac{m}{1-k}\right)$$

with

$$\boldsymbol{U}_0 = \left(\binom{n}{j} \binom{j}{i} (-1)^{i-j} \right), \qquad \boldsymbol{T}_n(a) = \left((-a)^{i-j} \binom{i}{j} \right).$$

Here and in the following matrices, we always denote the row index by i and the column index by j $(i, j = 0, \dots, n)$. It is easy to see that $\boldsymbol{T}_n^{-1}(a) = \boldsymbol{T}_n(-a)$ and

$$\boldsymbol{U}_0^{-1} = \left(\frac{\binom{j}{i}}{\binom{n}{i}} \right).$$

In order to obtain a simple result, we choose

$$\boldsymbol{F}_m := \operatorname{diag} \left(\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n} \right),$$

so that (4.3) and $\boldsymbol{U}_m^{-T} = \boldsymbol{U}_0^{-T} \boldsymbol{T}_n(-\frac{m}{1-k})^T$ imply

$$\boldsymbol{G}_m = \left(\frac{\binom{i}{j}}{\binom{n}{j}} \right) \cdot \left(\left(\frac{m}{1-k} \right)^{j-i} \binom{j}{i} \binom{n}{j} \right).$$

The entries of the matrix product on the right–hand side are

$$\sum_{l=0}^{n} \frac{\binom{i}{l}}{\binom{n}{l}} \left(\frac{m}{1-k} \right)^{j-l} \binom{j}{l} \binom{n}{j}$$

and in view of

$$\frac{\binom{j}{l}\binom{n}{j}}{\binom{n}{l}} = \binom{n-l}{j-l} = \binom{n+\nu-j}{\nu}$$

with $\nu = j - l$, these are exactly the entries of (4.2). ∎

Remark 4.4 1. For $j = n$ the components of the vector (4.2) are

$$\begin{aligned} \sum_{\nu=n-i}^{n} \binom{i}{\nu+i-n} \left(\frac{m}{1-k} \right)^{\nu} &= \left(1 + \frac{m}{1-k} \right)^{i} \left(\frac{m}{1-k} \right)^{n-i} \\ &= \left(\frac{m}{1-k} \right)^{n} \left(1 + \frac{1-k}{m} \right)^{i}. \end{aligned}$$

Thus, for the eigenvalue k^{-n} we get, up to a constant factor, indeed the same eigenvector as in Corollary 4.2.

2. Eigenvectors for matrices composed by binomial coefficients are also determined in [1] and [4].

ACKNOWLEDGEMENT

The second author was supported by the Deutsche Forschungsgemeinschaft.

References

[1] **L. Berg**, and **K. Engel :** *Spectral properties of matrices with products of binomial coefficients as entries.* In: J. M. Rassias (ed.), Functional Analysis, Approximation Theory and Numerical Analysis, World Sc., Singapore 1994, pp. 9–17

[2] **P. de Casteljau :** *Outillage méthodes calcul.* Andre Citroën Automobiles, SA, Paris, 1959

[3] **R. N. Goldman**, and **D. C. Heath :** *Linear subdivision is a strictly polynomial phenomenon.* Comp. Aided Geom. Design 1:269–278 (1984)

[4] **K. Engel**, and **C. Rommel :** *The Jordan normal form of matrices with products of binomial coefficients as entries.* Z. Angew. Math. Mech. **76** (5), 302-304 (1996)

[5] **C. Micchelli**, and **H. Prautzsch :** *Uniform refinement of curves.* Linear Algebra and Appl. 114/115:841–870 (1989)

[6] **C. A. Micchelli**, and **Y. Xu :** *Using the matrix refinement equation for the construction of wavelets on invariant sets.* Appl. Comput. Harmonic Anal. 1:391–401 (1994)

[7] **G. Plonka :** *Two–scale symbol and autocorrelation symbol for B–splines with multiple knots.* Advances in Comp. Math. 3: 1–22 (1995)

[8] **G. Plonka :** *Spline wavelets with higher defect.* In: P. J. Laurent, A. Le Méhaute, L. L. Schumaker (eds.), Wavelets, Images and Surface Fitting, A K Peters, Boston 1994, pp. 387–398

received: February 2, 1995

Authors:

Lothar Berg
Fachbereich Mathematik
Universität Rostock
D-18051 Rostock
Germany
e-mail: lothar.berg@mathematik.uni-rostock.de

Gerlind Plonka
Fachbereich Mathematik
Universität Rostock
D-18051 Rostock
Germany

Rostock. Math. Kolloq. **50**, 31–44 (1997)

Subject Classification (AMS)
08A55 08A40

Dietlinde Lau

Die maximalen Klassen von $\widetilde{P}_3$

Nachfolgend soll bewiesen werden, daß die Menge aller partiellen Funktionen $\widetilde{P}_3$ über einer 3-elementigen Menge genau 58 verschiedene (bez. Superposition) abgeschlossene maximale Teilmengen besitzt. Diese Aussage folgt zwar auch aus der von I. G. Rosenberg und L. Haddad in [2] [3] und [4] angegebenen Beschreibung sämtlicher maximalen abgeschlossenen Mengen aus partiellen Funktionen über beliebigen endlichen Mengen, jedoch soll hier gezeigt werden, wie man dies auch (ohne Kenntnis des allgemeinen Falls) elementar beweisen kann. Die hier vorgestellte Beweisvariante wurde bereits 1977 im Anhang der Dissertation der Autorin angegeben, jedoch später nicht publiziert, was nun (sehr verspätet) nachgeholt werden soll.

Wir beginnen mit einer Zusammenstellung der benötigten Grundbegriffe und Bezeichnungen. Seien

$$E_k := \{0, 1, ..., k-1\} \text{ und } \widetilde{E}_k := \{0, 1, ..., k-1, \infty\}.$$

$P_k^{(n)}$ bezeichne die Menge aller n-stelligen Funktionen $f^{(n)}$, die das n-fache kartesische Produkt E_k^n in E_k abbilden, und es sei

$$P_k := \sum_{n\geq 1} P_k^{(n)}.$$

In Verallgemeinerung dieser Mengen betrachten wir

$$\widetilde{P}_k^{(n)} := \{\, f^{(n)} \mid f^{(n)} : E_k^n \to \widetilde{E}_k \,\}$$

und

$$\widetilde{P}_k := \sum_{n\geq 1} \widetilde{P}_k^{(n)},$$

wobei $f^n(a_1,...,a_n) = \infty$ für „$f(a_1,a_2,...,a_n)$ ist nicht definiert" bzw. „$(a_1,...,a_n)$ gehört nicht zum Definitionsbereich von f" steht.
Falls die Stellenzahl von $f^{(n)} \in \widetilde{P}_k$ aus dem Zusammenhang ersichtlich ist, schreiben wir nur f. Den Wertebereich von f kürzen wir mit $W(f)$ ab. Die durch $e_i^{(n)}(x_1,...,x_n) := x_i$ bzw. $c_a^{(n)}(x_1,...,x_n) := a$ $(a \in \widetilde{E}_k)$ definierten Funktionen $e_i^{(n)}$ bzw. $c_a^{(n)}$ heißen *Projektionen* bzw. *Konstanten*. Insbesondere bezeichnet $c_\infty^{(n)}$ eine n-stellige Funktion, die für kein Tupel aus E_k^n definiert ist.
Bezeichnungen für die von den Konstanten verschiedenen einstelligen Funktionen aus P_3 sind in der Tabelle 1 zusammengefaßt.

x	$j_0(x)$	$j_1(x)$	$j_2(x)$	$j_3(x)$	$j_4(x)$	$j_5(x)$	$u_0(x)$	$u_1(x)$	$u_2(x)$	$u_3(x)$	$u_4(x)$	$u_5(x)$
0	1	0	0	1	1	0	2	0	0	2	2	0
1	0	1	0	1	0	1	0	2	0	2	0	2
2	0	0	1	0	1	1	0	0	2	0	2	2
x	$v_0(x)$	$v_1(x)$	$v_2(x)$	$v_3(x)$	$v_4(x)$	$v_5(x)$	$s_1(x)$	$s_2(x)$	$s_3(x)$	$s_4(x)$	$s_5(x)$	$s_6(x)$
0	2	1	1	2	2	1	0	0	1	1	2	2
1	1	2	1	2	1	2	1	2	0	2	0	1
2	1	1	2	1	2	2	2	1	2	0	1	0

Tabelle 1

Als *Operationen über* $\widetilde{P}_k$ betrachten wir die sogenannten *Mal'cev-Operationen* ζ, τ, Δ, ∇ und $*$, die für beliebige $f^{(n)}$ und $g^{(m)}$ aus $\widetilde{P}_k$ wie folgt definiert sind:

$$
\begin{aligned}
&(\zeta f)(x_1,x_2,...,x_n) := f(x_2,x_3,...,x_n,x_1),\\
&(\tau f)(x_1,x_2,...,x_n) := f(x_2,x_1,x_3,...,x_n),\\
&(\Delta f)(x_1,x_2,...,x_{n-1}) := f(x_1,x_1,x_2,...,x_{n-1}) \text{ für } n \geq 2,\\
&\zeta f = \tau f = \Delta f = f \text{ für } n = 1,\\
&(\nabla f)(x_1,x_2,...,x_{n+1}) := f(x_2,x_3,...,x_{n+1}) \text{ und}\\
&(f * g)(x_1,...,x_{m+n-1}) :=
\end{aligned}
$$

$$
\begin{cases} f(g(x_1,...,x_m),x_{m+1},....,x_{m+n-1}) & \text{für } g(x_1,...,x_m) \in E_k,\\ \infty & \text{sonst.} \end{cases}
$$

Die Menge aller Funktionen, die aus Funktionen einer Menge $A \subseteq \widetilde{P}_k$ mit Hilfe der oben definierten Mal'cev-Operationen in endlich vielen Schritten erhalten werden können, wird *Abschluß* von A genannt und mit $[A]$ bezeichnet. Ist $A = [A]$, so heißt A *abgeschlossene (Teil-)Menge* oder *(Teil-)Klasse von* $\widetilde{P}_k$. Eine echte Teilklasse A' der Klasse A nennt man *maximale Klasse von* A, wenn für alle $f \in A\backslash A'$ stets $[A' \cup \{f\}] = A$ gilt. Eine Menge B $(\subseteq A)$ heißt *vollständig in* A, wenn $[B] = A$ ist.

Zur Beschreibung von abgeschlossenen Teilmengen von $\widetilde{P}_k$ sind h-stellige (h-äre) Relationen geeignet, $h \geq 1$, d.h. Teilmengen von $\widetilde{E}_k^h$, deren Elemente $(a_1, ..., a_h)$ wir zumeist als Spalten

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_h \end{pmatrix}$$

aufschreiben werden und die Relation selbst in Form einer Matrix, deren Spalten die Elemente der Relation sind.
Wir sagen, eine *h-äre Relation ρ über $\widetilde{E}_k$ wird von einer Funktion $f \in \widetilde{P}_k$ bewahrt*, wenn für alle $\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n} \in \rho$ mit $\mathbf{r_i} := (r_{1i}, r_{2i}, ..., r_{hi})$, $i = 1, 2, ..., n$, stets

$$f(\mathbf{r_1}, ..., \mathbf{r_n}) := \begin{pmatrix} f(r_{11}, r_{12}, ..., r_{1n}) \\ f(r_{21}, r_{22}, ..., r_{2n}) \\ \vdots \\ f(r_{h1}, r_{h2}, ..., r_{hn}) \end{pmatrix} \in \rho$$

gilt, wobei für $\mathbf{a} \in \widetilde{E}_k^n \backslash E_k^h$ noch $f(\mathbf{a}) = \infty$ festgelegt sei.
Die Menge aller Funktionen aus P_k (bzw. $\widetilde{P}_k$), die die Relation $\rho \subseteq E_k^h$ (bzw. $\rho \subseteq \widetilde{E}_k^h$) bewahren, sei mit $Pol_k\rho$ (bzw. $Pol_{\widetilde{k}}\rho$) bezeichnet. Außerdem verwenden wir folgende Bezeichnung

$$POL_{\widetilde{k}}\rho := Pol_{\widetilde{k}}(\rho \cup (\widetilde{E}_k^h \backslash E_k^h)).$$

Die mit Hilfe einer gewissen Äquivalenzrelation ε auf E_h $(h \geq 1)$ definierbare h-äre Relation

$$\delta_{k,\varepsilon}^h := \{(a_0, a_1, ..., a_{h-1}) \in E_k^h \mid (i, j) \in \varepsilon \Longrightarrow a_i = a_j\}$$

wird *diagonale Relation* genannt. Zwecks einfacherer Beschreibung einer diagonalen Relation $\delta_{k,\varepsilon}^h$ geben wir diese Relation später oft in der Form

$$\delta_{k;\varepsilon_1,\varepsilon_2,...,\varepsilon_r}^h$$

an, wobei ε_1, ε_2, ..., ε_r gerade die aus mindestens zwei Elementen bestehende Äquivalenzklassen von ε sind.
In den nachfolgenden Sätzen wird öfter auch die folgende Relationenbezeichnung verwendet:

$$\iota_k^h := \{(a_0, a_1, ..., a_{h-1}) \in E_k^h \mid |\{a_0, a_1, ..., a_{h-1}\}| \leq h - 1 \}.$$

Als nächstes sollen einige Hilfsaussagen zusammengestellt werden.

Der folgende Satz wurde erstmalig von S. V. Jablonskij (siehe dazu [12] oder [14]) bewiesen:

Satz 1 *P_3 besitzt genau 18 maximale Klassen:*

- 3 *Klassen vom Typ* $\mathfrak{M}$ (*Klassen monotoner Funktionen*):

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 & 2 & 2 \end{pmatrix}, Pol_3\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 2 \\ 0 & 1 & 2 & 2 & 1 & 1 \end{pmatrix},$$

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 2 & 2 \end{pmatrix};$$

- 3 *Klassen vom Typ* $\mathfrak{U}$ (*durch nichttriviale Äquivalenzrelationen definierte Klassen*):

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{pmatrix}, Pol_3\begin{pmatrix} 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & 2 & 0 \end{pmatrix},$$

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 & 1 \end{pmatrix};$$

- 1 *Klasse vom Typ* $\mathfrak{S}$ (*autoduale Funktionen*):

$$Pol_3\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

- 9 *Klassen vom Typ* $\mathfrak{C}$ (*durch zentrale Relationen definierte Klassen*):
  $Pol_3\{0\}$, $Pol_3\{1\}$, $Pol_3\{2\}$, $Pol_3\{0,1\}$, $Pol_3\{0,2\}$, $Pol_3\{1,2\}$,

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 0 & 1 & 0 & 2 \\ 0 & 1 & 2 & 1 & 0 & 2 & 0 \end{pmatrix}, Pol_3\begin{pmatrix} 0 & 1 & 2 & 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 \end{pmatrix},$$

$$Pol_3\begin{pmatrix} 0 & 1 & 2 & 2 & 0 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 & 1 & 2 \end{pmatrix};$$

- 1 *Klasse vom Typ* $\mathfrak{L}$ (*lineare Funktionen*): $Pol_3\lambda_3$, *wobei*
  $\lambda_3 := \{\,(a,b,c,d) \in E_3^4 \,|\, a+b = c+d (mod\, 3)\}$;

- 1 *Klasse vom Typ* $\mathfrak{B}$: $Pol_3\iota_3^3$, *wobei*
  $\iota_3^3 := \{\,(a,b,c) \in E_3^3 \,|\, |\{a,b,c\}| \leq 2\,\}$.

Eine Teilmenge M von P_3 ist genau dann in P_3 vollständig, wenn M keine Teilmenge der oben angegebenen 18 maximalen Klassen von P_3 ist. □

Lemma 1 *Bezeichne f eine n-stellige Funktion aus P_3, die von mindestens 2 Variablen (o.B.d.A. seien dies x_1 und x_2) wesentlich abhängt, und sei $\{a, b, c\} = E_3$. Dann gilt:*
(a) $W(f) = E_3 \Rightarrow \exists\, \mathbf{r_1}, ..., \mathbf{r_n} \in \delta^3_{\{0,1\}} \cup \delta^3_{\{1,2\}} :\ f(\mathbf{r_1}, ..., \mathbf{r_n}) \in E_3^3 \backslash \iota_3^3$;
(b) $|W(f)| = 2 \Rightarrow \exists\, \mathbf{r_1}, ..., \mathbf{r_n} \in \delta^3_{\{a,b\}} \cup \delta^3_{\{b,c\}} :\ f(\mathbf{r_1}, ..., \mathbf{r_n}) \in \delta^3_{\{a,c\}} \backslash \delta^3_{\{0,1,2\}}$;
(c) $W(f) = E_3 \Rightarrow [\{f\} \cup \{\, g \in P_3^{(1)} \,|\, g(a) = g(b) \ \vee\ g(b) = g(c)\,\}] = P_3$.

Beweis: (a) ist ein Spezialfall des „Hauptlemmas von Jablonskij" und (b) eine leichte Folgerung aus diesem Lemma (siehe [10], S. 162).
(c): O.B.d.A. seien $a = 0$, $b = 1$ und $c = 2$. Offenbar gilt dann

$$\{\, g \in P_3^{(1)} \,|\, g(0) = g(1) \ \vee\ g(1) = g(2)\,\} = \{c_0, c_1, c_2, j_\alpha, u_\alpha, v_\alpha \,|\, \alpha \in \{0, 2, 3, 5\}\}.$$

Man prüft nun leicht nach, daß diese Menge einstelliger Funktionen keine Teilmenge von maximalen Klassen des Typs $\mathfrak{M}$, $\mathfrak{U}$, $\mathfrak{S}$, $\mathfrak{C}$ und $\mathfrak{L}$ ist. Da außerdem $W(f) = E_3$ und $f \in P_3 \backslash [P_3^{(1)}]$ vorausgesetzt ist, bewahrt f wegen (a) auch nicht die Relation ι_3^3, womit (c) aus Satz 1 folgt. $\square$.

Lemma 2 (a) *Für jedes $g \in \widetilde{P}_k \backslash (P_k \cup [\{c_\infty\}])$ gilt $[P_k \cup \{g\}] = \widetilde{P}_k$.*
(b) *$P_k \cup [\{c_\infty\}]$ ist die einzige maximale Klasse von $\widetilde{P}_k$, die P_k enthält.*

Beweis: (a): Sei $g^m \in \widetilde{P}_k \backslash (P_k \cup [\{c_\infty\}])$.
Da die Funktion $g' := e_2^2 * g$ $k + 1$ verschiedene Werte annimmt, können wir o.B.d.A. $W(g) = \widetilde{E}_k$ annehmen. Folglich gibt es $k + 1$ Tupel $\mathbf{a_i} := (a_{i1}, a_{i2}, ..., a_{im}) \in E_k^m$ mit $g(\mathbf{a_i}) = i$ ($i \in \widetilde{E}_k$). Bezeichne $f^{(n)}$ eine beliebige Funktion aus $\widetilde{P}_k$. In Abhängigkeit von f lassen sich folgende Funktionen f_j ($j = 1, 2, ..., m$) wie folgt definieren:

$$f_j(b_1, ..., b_n) \ =\ a_{ij} \iff f(b_1, ..., b_n) \ =\ i$$

($b_1, ..., b_n \in E_k;\ i \in \widetilde{E}_k$).
Damit haben wir $f(\mathbf{x}) = g(f_1(\mathbf{x}), ..., f_m(\mathbf{x}))$, womit $f \in [P_k \cup \{g\}]$.
(b) folgt unmittelbar aus (a). $\square$

Die nächsten 4 Lemmatas befassen sich mit der Maximalität gewisser Teilklassen von $\widetilde{P}_3$ (bzw. $\widetilde{P}_k$) in $\widetilde{P}_3$ (bzw. $\widetilde{P}_k$).

In [1] wurde die nachfolgende Aussage (a) und in [13] (bzw. in [8], [9]) wurde die Aussage (b) des folgenden Lemmas bewiesen:

Lemma 3 *Seien*

$$\rho_1 := \{(a, a, b, b), (a, b, a, b) \,|\, a, b \in E_k\,\},$$

$\rho_2 := \{(a,a,b,b),(a,b,a,b),(a,b,b,a)\,|\,a,b\in E_k\}$,
$\rho_i := \{(a_1,a_2,...,a_i)\in E_k^i\,|\,|\{a_1,...,a_i\}|\le i-1\}$ *($i = 3, ..., k$). Dann gilt*

(a) *Die Klassen $Pol_k\rho_i$ ($i=1,2,...,k$) sind die einzigen echten Teilklassen von P_k, die $P_k^{(1)}$ enthalten. Außerdem gilt*

$$[P_k^{(1)}] = Pol_k\rho_1 \subset Pol_k\rho_2 \subset ... \subset Pol_k\rho_{k-1} \subset Pol_k\rho_k \subset P_k.$$

(b) *Die Klassen $POL_{\widetilde{k}}\rho_i$ ($i = 1, 2, ..., k$) sind maximale Klassen von $\widetilde{P}_k$ und die einzigen maximalen Klassen von $\widetilde{P}_k$, die $P_k^{(1)}$ enthalten.* □

Lemma 4 *Sei $\varrho := \delta\cup\sigma$, wobei δ eine gewisse h-stellige diagonale Relation bezeichnet, die von E_k^h verschieden ist, und $\emptyset \neq \sigma \subseteq \{(a_1,...,a_h)\in E_k^h\,|\,|\{a_1,...,a_h\}| = h\}$ sei. Außerdem existiere zu jedem $\mathbf{a} := (a_1,...,a_h)\in\sigma$ eine gewisse Äquivalenzrelation ε_a auf E_k mit den folgenden zwei Eigenschaften:*
(1) *Für jedes $i\in\{1,...,h\}$ existiert genau eine Äquivalenzklasse von ε_a, die a_i enthält.*
(2) *Zu jedem $\mathbf{b}\in\varrho$ findet man in $POL_{\widetilde{k}}\varrho$ eine einstellige Funktion $g_{\mathbf{a},\mathbf{b}}$ mit $g_{\mathbf{a},\mathbf{b}}(\mathbf{a}) = \mathbf{b}$ und $g_{\mathbf{a},\mathbf{b}}(x) = g_{\mathbf{a},\mathbf{b}}(y)$ für alle $(x,y)\in\varepsilon_{\mathbf{a}}$.*
Dann ist $POL_{\widetilde{k}}\varrho$ eine maximale Klasse von $\widetilde{P}_k$.

Beweis: Offensichtlich ist $POL_{\widetilde{k}}\varrho \neq \widetilde{P}_k$. Sei $f\in\widetilde{P}_k\backslash POL_{\widetilde{k}}\varrho$. Da ϱ die Eigenschaften (1) und (2) besitzt, erhält man durch Einsetzen gewisser einstelliger Funktionen aus $POL_{\widetilde{k}}\varrho$ in f zu jedem $r\in\sigma$ eine gewisse einstellige Funktion h_r mit $h_r(r)\in E_k^h\backslash\varrho$. Sei $\sigma = \{r_1,...,r_m\}$. Zu $POL_{\widetilde{k}}\varrho$ gehören Funktionen, die auf Zeilen der Form

$$(x_1, x_2, g_{r_1}(x_1), g_{r_2}(x_1), ..., g_{r_m}(x_1), g_{r_1}(x_2), g_{r_2}(x_2), ..., g_{r_m}(x_2))$$

beliebige Werte aus $\widetilde{E}_k$ und sonst nur den Wert ∞ annehmen. Folglich sind beliebige Funktionen aus $P_k^{(2)}$ Superpositionen über $\{f\}\cup POL_{\widetilde{k}}\varrho$, woraus (nach bekannten Eigenschaften von P_k) $P_k \subseteq [\{f\}\cup POL_{\widetilde{k}}\varrho]$ folgt. Da zu $POL_{\widetilde{k}}\varrho$ offenbar Funktionen gehören, die genau $k+1$ verschiedene Werte annehmen, folgt aus Lemma 2, (a), daß $[\{f\}\cup POL_{\widetilde{k}}\varrho] = \widetilde{P}_k$ gilt. Also ist $POL_{\widetilde{k}}\varrho$ eine maximale Klasse von $\widetilde{P}_k$. □

i	τ_i	i	τ_i
1	$\{0\}$	2	$\{1\}$
3	$\{2\}$	4	$\{0,1\}$
5	$\{0,2\}$	6	$\{1,2\}$
7	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	8	$\begin{pmatrix} 0 \\ 2 \end{pmatrix}$
9	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$	10	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
11	$\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$	12	$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$
13	$\begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}$	14	$\begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}$
15	$\begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix}$	16	$\begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \end{pmatrix}$
17	$\begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \end{pmatrix}$	18	$\begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 0 & 2 & 1 \end{pmatrix}$
19	$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}$	20	$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 2 \end{pmatrix}$
21	$\begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix}$	22	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{pmatrix}$
23	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & 2 & 0 \end{pmatrix}$	24	$\begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 & 1 \end{pmatrix}$
25	$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \end{pmatrix}$	26	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & 1 & 1 \end{pmatrix}$
27	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 2 & 2 \end{pmatrix}$	28	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & 1 & 0 \end{pmatrix}$
29	$\begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 0 & 1 \end{pmatrix}$	30	$\begin{pmatrix} 0 & 1 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \end{pmatrix}$
31	$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 2 \\ 0 & 1 & 2 & 1 & 2 & 1 \end{pmatrix}$	32	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 & 2 \\ 0 & 1 & 2 & 1 & 1 & 0 \end{pmatrix}$
33	$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 & 2 & 2 \end{pmatrix}$	34	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 & 0 & 2 \\ 0 & 1 & 2 & 1 & 0 & 2 & 0 \end{pmatrix}$
35	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 1 & 0 & 2 & 1 \end{pmatrix}$	36	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 & 0 & 2 & 1 \end{pmatrix}$
37	$\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$	38	$\begin{pmatrix} 0 & 0 \\ 1 & 2 \\ 2 & 1 \end{pmatrix}$
39	$\begin{pmatrix} 0 & 2 \\ 1 & 1 \\ 2 & 0 \end{pmatrix}$	40	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 2 \end{pmatrix}$

Tabelle 2.1

i	τ_i	i	τ_i
41	$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}$	42	$\begin{pmatrix} 0 & 0 & 1 & 1 & 2 & 2 \\ 1 & 2 & 0 & 2 & 0 & 1 \\ 2 & 1 & 2 & 0 & 1 & 0 \end{pmatrix}$
43	$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix}$	44	$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 & 1 \end{pmatrix}$
45	$\begin{pmatrix} 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 & 0 \end{pmatrix}$	46	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 2 & 2 \end{pmatrix}$
47	$\begin{pmatrix} 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 2 & 1 & 2 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 \end{pmatrix}$	48	$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 1 & 2 & 1 & 2 & 0 & 2 & 0 & 1 \\ 0 & 1 & 2 & 2 & 1 & 2 & 0 & 1 & 0 \end{pmatrix}$

Tabelle 2.2

i	τ_i
49	$\{\,(0,1,2),(a,a,b)\,\vert\,a,b\in E_3\,\}$
50	$\{\,(0,1,2),(a,b,a)\,\vert\,a,b\in E_3\,\}$
51	$\{\,(0,1,2),(b,a,a)\,\vert\,a,b\in E_3\,\}$
52	$\{\,(0,1,2),(1,0,2),(a,a,b)\,\vert\,a,b\in E_3\,\}$
53	$\{\,(0,1,2),(2,1,0),(a,b,a)\,\vert\,a,b\in E_3\,\}$
54	$\{\,(0,1,2),(0,2,1),(b,a,a)\,\vert\,a,b\in E_3\,\}$
55	$\{\,(a,a,b,b),(a,b,a,b)\,\vert\,a,b\in E_3\,\}$
56	$\{\,(a,a,b,b),(a,b,a,b),(a,b,b,a)\,\vert\,a,b\in E_3\,\}$
57	$\{\,(a,b,c)\in E_3^3\,\vert\,\vert\{a,b,c\}\vert\ \le\ 2\,\}$
58	$E_3^2\cup\{(\infty,\infty)\}$

Tabelle 2.3

Lemma 5 *Die Klassen $POL_{\widetilde{3}}\tau_i$ ($i\in\{1,2,...,57\}$, siehe Tabelle 2.1 - 2.3) sind maximale Klassen von $\widetilde{P}_3$.*

Beweis: Für $i\in\{1,2,...,54\}$ beweist man das Lemma leicht mit Hilfe von Lemma 4. Falls $i\in\{55,56,57\}$, folgt die obige Behauptung aus Lemma 3. Die Maximalität von $POL_{\widetilde{3}}\tau_{58}$ ergibt sich aus $POL_{\widetilde{3}}\tau_{58}=\widetilde{P}_3\cup[\{c_\infty\}]$ und Lemma 2, (b). $\square$

Satz 2 (a) *$\widetilde{P}_3$ besitzt genau 58 maximale Klassen. Es sind dies die Mengen $POL_3\tau_i$ ($i =$ 1, 2, 3,..., 58), wobei die τ_i in den Tabellen 2.1 - 2.3 angegeben sind.*
(b) *Eine Menge $A \subseteq \widetilde{P}_3$ ist genau dann in $\widetilde{P}_3$ vollständig, wenn sie keine Teilmenge der angegebenen 58 maximalen Klassen ist.*

Beweis: Wegen Lemma 5 haben wir uns zum Beweis von (a) und (b) nur zu überlegen, daß eine in keine der Mengen $POL_{\widetilde{3}}\tau_i$ ($i \in \{1, 2, ..., 58\}$) enthaltene Teilmenge von $\widetilde{P}_3$ vollständig in $\widetilde{P}_3$ ist.
Sei $M \subseteq \widetilde{P}_3$ keine Teilmenge der im Satz genannten Klassen. Folglich gehören zu M Funktionen $f_i^{n_i} \notin POL\tau_i$ ($i = 1$, 2, ..., 58). Ist

$$\tau_i = (\sigma_1\ \sigma_2\ ...\ \sigma_{m_i})$$

eine h_i-stellige Relation, so können wir o.B.d.A. $n_i = m_i$ und

$$f_i(\sigma_1, \sigma_2, ..., \sigma_{m_i}) \in E_3^{h_i} \backslash \tau_i$$

vereinbaren.
Wie oben bereits erwähnt wurde, haben wir zu zeigen, daß $[M] = \widetilde{P}_3$ ist. Zuerst soll

$$\{c_0, c_1, c_2\} \subseteq [M] \tag{1}$$

nachgewiesen werden.
Die Funktion f_{37} ist einstellig und aus $P_3^{(1)} \backslash \{s_1\}$. Folgende 3 Fälle sind dann möglich:
Fall 1: $f_{37} = c_a$ ($a \in E_3$).
Offensichtlich haben wir in diesem Fall $\{c_0, c_1, c_2\} \subseteq [\{f_{37}, f_1, f_2, ..., f_6\}]$.
Fall 2: $|W(f_{37})| = 2$.
O.B.d.A. sei $W(f_{37}) = \{0, 1\}$, d.h., $f_{37} \in \{j_0, j_1, ..., j_5\}$. Wegen $j_2 * j_2 = c_0$, $j_0 * j_0 = j_5$, $j_3 * j_3 = c_1$, $j_4 * j_4 = j_1$ und Fall 1 genügt es, $f_{37} \in \{j_1, j_5\}$ weiter zu untersuchen.
2.1: $f_{37} = j_1$.
Wir bilden $f'_7 := f_7 * j_1 \in \{c_0, c_1, c_2, j_4, u_1, u_4, v_1, v_4\}$. Da $u_1 * u_1 = c_0$, $v_1 * v_1 = v_4$ und $u_4 * u_4 = c_2$, können wir uns wegen Fall 1 auf $f'_7 \in \{j_4, v_4\}$ beschränken.
2.1.1: $f'_7 = j_4$.
Einsetzen von f_{37} ($= j_1$) und f'_7 ($= j_4$) in f_{10} liefert eine einstellige Funktion $f'_{10}(x) := f_{10}(j_1(x), j_4(x))$ mit $f'_{10} \in \{c_0, c_1, c_2, u_1, u_4, v_1, v_4\}$. Wegen $u_1 * u_1 = c_0$, $u_4 * u_4 = c_2$ und $v_1 * v_1 = v_4$ bleibt noch $f'_{10} = v_4$ zu untersuchen. Es ist $v_4 * j_4 = v_1$. Bildet man $f'_{17}(x) := f_{17}(j_1(x), j_4(x), v_1(x), v_4(x))$, so gilt $f'_{17} \in \{c_0, c_1, c_2, u_1, u_4\}$ und $f'_{17} * f'_{17}$ ist eine Konstante, womit der Fall 2.1.1 auf den ersten Fall zurückgeführt worden ist.
2.1.2: $f'_7 = v_4$.

Die Funktion $f'_{14}(x) := f_{14}(j_1(x), v_4(x))$ gehört zu $\{c_0, c_1, c_2, j_4, u_1, u_4, v_1\}$. Wegen $u_1 * u_1 = c_0$, $u_4 * u_4 = c_2$, $j_1 * v_1 = j_4$ und Fall 2.1.1 ist auch der Fall 2.1.2 auf den ersten Fall zurückführbar.
2.2: $f_{37} = j_5$.
Dieser Fall läßt sich leicht wie der Fall 2.1 auf den ersten Fall zurückführen, indem man f_{16} anstelle von f_{17} und f_{13} anstelle von f_{14} verwendet.
Fall 3: $|(f_{37})| = 3$ und $f_{37} \neq s_1$.
Mit Hilfe der Funktionen f_{38}, ..., f_{42} ist auch dieser Fall zunächst auf die Fälle 1 oder 2 und damit auf Fall 1 zurückführbar.

Folglich gehören die Konstanten zu $[M]$**.**

Sei $f'_{43}(x) := f_{43}(c_0(x), c_1(x), c_2(x), x)$. Diese Funktion gehört zu $P_3^{(1)} \cap [M]$ und nimmt mindestens zwei verschiedene Werte an. Ist f'_{43} eine Permutation, so kann man mit Hilfe der Funktionen f_{44}, ..., f_{48} eine Funktion erhalten, die genau zwei verschiedene Werte aus E_3 annimmt. O.B.d.A. können wir also $W(f'_{43}) = E_2$ im folgenden voraussetzen, d.h., $f'_{43} \in \{j_0, j_1, ..., j_5\}$.
Sei

$$M_{a,b} := \{f \in P_3^{(1)} \mid f(a) = f(b)\}.$$

Als Nächstes soll

$$\exists\, a, b \in E_3 :\ a \neq b \wedge M_{a,b} \subset [M] \tag{2}$$

nachgewiesen werden. Wegen $j_0 * j_0 = j_5$, $j_4 * j_4 = j_1$ und Dualitätsgründen genügt es, die Fälle $f'_{43} \in \{j_1, j_2\}$ zu diskutieren.
Fall 1: $f'_{43} = j_1$.
Einsetzen der Konstanten und j_1 in f_{19} liefert $f'_{19} \in \{j_4, u_1, u_4, v_1, v_4\}$. Wegen $v_1 * v_1 = v_4$ und und $f_{22}(c_0, c_1, c_2, j_1, j_4) \in \{u_1, u_4, v_1, v_4\}$ können wir uns auf $f'_{19} \in \{u_1, u_4, v_4\}$ beschränken.
1.1: $f'_{19} = u_1$.
In diesem Fall läßt sich eine Funktion $f'_{25}(x) := f_{25}(c_0(x), c_1(x), c_2(x), j_1(x), u_1(x))$ mit $f'_{25} \in \{j_4, u_4, v_1, v_4\}$ konstruieren, von der wir o.B.d.A. $f'_{25} \in \{j_4, u_4, v_4\}$ annehmen können.
1.1.1: $f'_{25} = j_4$.
Es gilt $u_1 * j_4 = u_4$. Folglich ist die einstellige Funktion $f'_{34} := f_{34}(c_0, c_1, c_2, j_1, j_4, u_1, u_4) \in \{v_1, v_4\}$ eine Superposition über M. Wegen $v_1 * v_1 = v_4$ und $v_4 * j_4 = v_1$ gilt damit: $\{c_0, c_1, c_2, j_1, j_4, u_1, u_4, v_1, v_4\} = \{f \in P_3^{(1)} \mid f(0) = f(2)\} \subset [M]$.
1.1.2: $f'_{25} = u_4$.
Für die Funktion $f'_{23} := f_{23}(c_0, c_1, c_2, u_1, u_4)$ gilt $f'_{23} \in \{j_1, j_4, v_1, v_4\}$. Wegen $v_1 * v_1 = v_4$ und $j_4 * j_4 = j_1$ betrachten wir nur $f'_{23} \in \{j_1, v_4\}$.
1.1.2.1: $f'_{23} = j_1$.
Einsetzen von u_1 statt u_4 und u_4 statt u_1 in f_{23} ergibt anstelle von f'_{23} die Funktion $f''_{23} = j_4$.

Weiter wie in 1.1.1.

1.1.2.2: $f'_{23} = v_4$.

Einsetzen von u_1 statt u_4 und u_4 statt u_1 in f_{23} ergibt anstelle von f'_{23} die Funktion $f''_{23} = v_1$. Weiter ist $f'_{36} := f_{36}(c_0, c_1, c_2, u_1, u_4, v_1, v_4) \in \{j_1, j_4\}$. Falls $f'_{36} = j_1$ ist, setzen wir in f_{36} u_1 statt u_4, u_4 statt u_1, v_1 statt v_4 und v_4 statt v_1 in f_{36} ein und erhalten anstelle von f'_{36} die Funktion $f''_{36} = j_4$. Der Fall 1.1.2.2 ist damit ebenfalls auf den Fall 1.1.1 zurückführbar.

1.1.3: $f'_{13} = v_4$.

Eine Superposition über M ist dann $f'_{31} := f_{31}(c_0, c_1, c_2, j_1, u_1, v_4) \in \{j_4, u_4, v_1\}$. Wegen $j_1 * v_1 = j_4$ ist dieser Fall auf die Fälle 1.1.1 und 1.1.2 zurückführbar.

1.2: $f'_{19} = u_4$.

Wir können $f'_{28} := f_{28}(c_0, c_1, c_2, j_1, u_4) \in \{j_4, u_1, v_1, v_4\}$ bzw. o.B.d.A. (wegen $v_1 * v_1 = v_4$ und $u_4 * j_4 = u_1$) $f'_{28} \in \{u_1, v_4\}$ konstruieren.

1.2.1: $f'_{28} = u_1$.

Weiter wie im Fall 1.1.

1.2.2: $f'_{28} = v_4$.

Wir haben in diesem Fall $f'_{32} := f_{32}(c_0, c_1, c_2, j_1, v_4, u_4) \in \{j_4, u_1, v_1\}$. Wegen $v_4 * j_4 = v_1$ und $u_4 * v_1 = u_1$ können wir $f'_{32} = u_1$ annehmen, womit der Fall 1.2 auf den Fall 1.1 zurückführbar ist.

1.3: $f'_{19} = v_4$.

Bildet man man $f'_{26} := f_{26}(c_0, c_1, c_2, j_1, v_4)$, so ist $f'_{26} \in \{j_4, u_1, u_4, v_1\}$ bzw. o.B.d.A. (wegen $v_4 * j_4 = v_1$) $f'_{26} \in \{u_1, u_4, v_1\}$.

1.3.1: $f'_{26} \in \{u_1, u_4\}$.

Weiter wie unter 1.1 oder 1.2.

1.3.2: $f'_{26} = v_1$.

Es gilt $j_1 * v_1 = j_4$. Folglich läßt sich eine Funktion $f'_{35} := f_{35}(c_0, c_1, c_2, j_1, j_4, v_1, v_4) \in \{u_1, u_4\}$ konstruieren, d.h., 1.3.2 ist auf die Fälle 1.1 und 1.2 zurückführbar.

Im Fall 1 haben wir damit (2) gezeigt.

Fall 2: $f'_{43} = j_2$.

Dann gilt $f'_{19} := f_{19}(c_0, c_1, c_2, j_2) \in \{j_3, u_2, u_3, v_2, v_3\}$. Wegen $u_3 * u_3 = u_2$ und $v_3 * v_3 = v_2$ sei o.B.d.A. $f'_{19} \in \{j_3, u_2, v_2\}$.

2.1: $f'_{19} = j_3$.

Seien $f'_{22} := f_{22}(c_0, c_1, c_2, j_2, j_3)$ und $f''_{22} := f_{22}(c_0, c_1, c_2, j_3, j_2)$. Dann gilt $\{f'_{22}, f''_{22}\} \in \{\{u_2, u_3\}, \{v_2, v_3\}\}$. Als Superpositionen über $\{f'_{22}, f''_{22}, j_2, j_3, f_{34}, f_{35}\}$ erhält man folglich $\{u_2, u_3, v_2, v_3\}$, womit $M_{0,1} \subseteq [M]$ gezeigt wurde.

2.2: $f'_{19} = u_2$.

Wir haben in diesem Fall $f'_{25} := f_{25}(c_0, c_1, c_2, j_2, u_2) \in \{j_3, u_3, v_2, v_3\}$. Wegen $j_2 * u_3 = j_3$, $v_3 * v_3 = v_2$ und Fall 2.1 können wir $f'_{25} = v_2$ annehmen. Da $f_{33}(c_0, c_1, c_2, j_2, u_2, v_2) \in$

$\{j_3, u_3, v_3\}$ und $j_2 * u_3 = j_3$ sowie $j_2 * v_3 = j_3$, ist der Fall 2.2 auf den Fall 2.1 zurückführbar.
2.3: $f'_{19} = v_2$.
Analog zu 2.2.

Damit ist (2) bewiesen.

O.B.d.A. können wir folglich

$$M_{0,1} := \{\, f \in P_3^{(1)} \mid f(0) = f(1)\,\} = \{c_0, c_1, c_2, j_2, j_3, u_2, u_3, v_2, v_3\} \subset [M] \tag{3}$$

annehmen.
Einsetzen dieser Funktionen und Identifizieren von Variablen in f_{49} liefert eine Funktion $f'_{49} \in P_3^{(1)}$ mit $f'_{49}(0) \neq f'_{49}(1)$ und $f'_{49} \neq s_1$.
Fall 1: $|W(f'_{49})| = 2$.
In diesem Fall prüft man leicht nach, daß $M_{0,1} \cup M_{a,b} \subset [M]$ für ein gewisses $(a, b) \in \{(0, 2), (1, 2)\}$ gilt. O.B.d.A. sei $(a, b) = (1, 2)$. Außerdem kann man wegen (3) o.B.d.A. annehmen, daß für gewisse $\alpha, \beta, \gamma, \delta \in E_3$

$$f_{55} \begin{pmatrix} \alpha & \gamma \\ \alpha & \delta \\ \beta & \gamma \\ \beta & \delta \end{pmatrix} \in E_3^4 \backslash \tau_{55}$$

ist. Bildet man dann als Superposition über M die Funktion $f'_{55}(x, y) := f_{55}(g_1(x), g_2(y))$, wobei g_1 und g_2 gewisse Funktionen aus $M_{0,1} \cup M_{1,2}$ mit $g_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ und $g_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$ bezeichnen, so gilt $f'_{55} \in P_3 \backslash [P_3^{(1)}]$.
Falls $|W(f'_{55})| = 3$, folgt aus Lemma 1, (c), daß $P_3 \subseteq [M]$ ist, woraus sich mit Hilfe von Lemma 2, (a) und $f_{58} \in \widetilde{P}_3 \backslash [\{c_\infty\}]$ unmittelbar $[M] = \widetilde{P}_3$ ergibt, w.z.b.w.
Sei also nachfolgend

$$|W(f'_{55})| = 2. \tag{4}$$

Wegen Lemma 1, (b) existieren gewisse $a_1, a_2, a_3, b_1, b_2, b_3 \in E_3$ mit

$$f'_{55} \begin{pmatrix} a_1 & a_2 \\ b_1 & a_2 \\ a_1 & b_2 \end{pmatrix} = \begin{pmatrix} a_3 \\ b_3 \\ b_3 \end{pmatrix},$$

wobei $a_3 \neq b_3$ ist. Einsetzen gewisser Funktionen aus $M_{0,1} \cup M_{1,2}$ ($\subset [M]$) in f'_{55} liefert eine einstellige Funktion f''_{55} mit der Eigenschaft: $M_{0,2} \subseteq [\{f''_{55}\} \cup M_{0,1} \cup M_{1,2}]$. Mit Hilfe der

Funktion $f_{57} \in M$ und Lemma 1, (a) folgt hieraus $P_3^{(1)} \subset [M]$. Nach Lemma 3, (a) haben wir dann $Pol_3\tau_{56} \subseteq [M]$. Mit Hilfe der Funktionen f_{56} und f_{57} aus M erhält man hieraus leicht, daß sämtliche Funktionen aus P_3 Superpositionen über M sind. Wegen $f_{58} \in M$ und Lemma 2, (a) geht dies aber nur im Fall $[M] = \widetilde{P}_3$, w.z.b.w.

Fall 2: $|W(f'_{49})| = 3$.

f'_{49} ist in diesem Fall eine von s_1 verschiedene Permutation, mit deren Hilfe man im Fall $f_{49} \neq s_3$ wegen (3) eine gewisse einstellige Funktion g mit $g(0) \neq g(1)$ als Superposition über M bilden kann, womit wir weiter wie unter Fall 1 schließen können.

Ist $f'_{49} = s_3$, so ist es möglich, als Superposition über $M_{0,1} \cup \{f_{52}\}(\subset [M])$ eine einstellige Funktion h zu bilden, die entweder eine von s_1 und s_3 verschiedene Permutation ist oder genau 2 verschiedene Werte annimmt, jedoch nicht zu $M_{0,1}$ gehört. Folglich ist der Fall 2 vollständig auf den Fall 1 zurückführbar und unser Satz bewiesen. $\square$

Literatur

[1] **Burle, G. A. :** *Klassen der k-wertigen Logik, die alle Funktionen einer Veränderlichen enthalten (Russ.).* Diskret. Analiz **10**, 3-7 (1967)

[2] **Haddad, L.** und **Rosenberg, I. G. :** *Criterére général de complétude pour les algébras partielles finies.* C. R. cad. ci. Paris, Ser. I Math. **304**, 507-509 (1987)

[3] **Haddad, L. :** *Maximal partial clones determined by quasi-diagonal relations.* J. Inform. Process. Cybernet. EIK **24**, 7/8, 355-366 (1988)

[4] **Haddad, L.** und **Rosenberg, I. G. :** *Maximal partial clones determined by the areflexive relations.* Discrete Appl. Math. **24**, 1-3, 133-143 (1989)

[5] **Haddad, L.** und **Rosenberg, I. G.; Schweigert, D. :** *A maximal partial clone and a Słupecki-type criterion.* Acta WSci Math, (Szeged) **54**, 1 - 2, 89-98 (1990)

[6] **Haddad, L.** und **Rosenberg, I. G. :** *Partial Sheffer Functions.* Europ. Journal of Combinatorics. **12**, 9-21 (1991)

[7] **Haddad, L.** und **Rosenberg, I. G. :** *Completeness theory of finite partial algebras.* Algebra Universalis **29**, 378-401 (1991)

[8] **Lau, D. :** *Eigenschaften gewisser abgeschlossener Klassen in Postschen Algebren.* Dissertation (A). Universität Rostock 1977

[9] **Lau, D. :** *Über partielle Funktionenalgebren.* Rostock. Math. Kolloq. **33**, 23-48 (1988)

[10] **Lau, D. :** *Ein neuer Beweis für Rosenberg's Vollständigkeitskriterium.* EIK **28**, 149-195 (1992)

[11] **Lo Czukai :** *The completeness theory of partial many-valued logic functions.* Proc. XVII. Internat. Symp. on Multiple-valued Logic, pp. 118-121, Boston 1987

[12] **Pöschel, R.** und **Kalužnin, L. A. :** *Funktionen- und Relationenalgebren.* Berlin 1979

[13] **Romov, B. A. :** *Über maximale Unteralgebren der Algebra der partiellen Funktionen der mehrwertigen Logik.* (Russ.) Kibernetika **1**, 28-36 (1980)

[14] **Rosenberg, I. G. :** *Über die funktionale Vollständigkeit in den mehrwertigen Logiken.* Rozpravy Československe Akad. Ved. Řada Mat. Přirod. Věd **80**, 3-93 (1970)

[15] **Rosenberg, I. G. :** *Galois theory for partial algebras.* In: Freese, R. S., and Garcia, O. C.(Eds.): Universal algebra and lattice theory. Proc. 4. Intern. Conf., Puebla 1982, Lecture Notes in Math. **1004**, pp. 257 - 272, Berlin 1983

eingegangen: 10.04.1995

Autor:

Dietlinde Lau
Fachbereich Mathematik
Universität Rostock
Universitätsplatz 1
18051 Rostock

email: dietlinde.lau@mathematik.uni-rostock.de

Rostock. Math. Kolloq. **50**, 45–46 (1997)

Subject Classification (AMS)
05A99 06A07 68R15

Tran-Ngoc Danh; David E. Daykin

Bezrukov-Gronau Order is Not Optimal

Let N be the integers ≥ 0. For $n \in N$ let $V(n)$ be the vectors $\underset{\sim}{a} = (a_1, \dots, a_n)$ with $a_i \in N$, and $C(n)$ be the vectors with $a_i \in \{0,1\}$. For $1 \leq i \leq n$ let $\delta_i \underset{\sim}{a} \in V(n-1)$ be obtained from $\underset{\sim}{a}$ by deleting a_i. The shadow $\Delta \underset{\sim}{a}$ is $\{\delta_1 \underset{\sim}{a}, \dots, \delta_n \underset{\sim}{a}\} \subseteq V(n-1)$. For $A \subseteq V(n)$ we put $\Delta A = \bigcup\{\underset{\sim}{a} \in A\}\Delta \underset{\sim}{a}$.

Example 1: Let $E = \{000,\ 100,\ 010,\ 001,\ 110,\ 101,\ 011,\ 111,\ 200\}$,
$F = E \bigcup \{210,\ 201,\ 211,\}$, $G = E \bigcup \{020,\ 002,\ 210\}$, $H = \{00,\ 10,\ 01,\ 11,\ 20\}$ then $\Delta F = H \bigcup \{21\}$ but $\Delta G = H \bigcup \{02, 21\}$.

Now $V\underset{\sim}{a}$ is the i with $a_{i-1} > a_i \leq a_{i+1} \leq \dots \leq a_n$ (think of this as V shaped) and $\underset{\sim}{a}^* = \delta_{V\underset{\sim}{a}} \underset{\sim}{a}$. Given $\underset{\sim}{a},\ \underset{\sim}{b} \in V(n)$ let $\omega(\underset{\sim}{a},\ \underset{\sim}{b})$ be the last i with $a_i \neq b_i$. In [2] we made

Definition 1 *(V-order)* **We order $V(1),\ V(2),\ \dots$ in turn by $\underset{\sim}{a} < \underset{\sim}{b}$ if**

(i) $\underset{\sim}{a}^* < \underset{\sim}{b}^*$,

or

(ii) $\underset{\sim}{a}^* = \underset{\sim}{b}^*$ **and** $a_j < b_j$ **where** $j = \omega(\underset{\sim}{a},\ \underset{\sim}{b})$.

The first so many sets of an order is an **initial section** IS. Thus $E,\ F,\ H$ in Example 1 are IS of V-order. We say an order is **optimal** if for every $A \subseteq V(n)$ we get $|\Delta A| \geq |\Delta I|$, where I is the IS of the order of $V(n)$ with $|A| = |I|$. Strehl and Winkelmann defined SW-order and conjectured it was optimal [3]. We support that conjecture. In [2] we showed that V and SW orders are the same. Further we gave T-order for $C(n)$, and proved that both V-order and T-order are optimal over $C(n)$.

In [1] Bezrukov and Gronau gave BG-order and claimed it was optimal. The G of Example 1 is shown as an IS of BG-order so their claim is false. Sadly their proof is not correct even over $C(n)$ where BG-order is the same as V-order.

References

[1] **Bezrukov, S.L.,**, and **Gronau, H.-D.O.F. :** *A Kruskal-Katona type theorem.* Rostock Math. Kolloq. **46**, 71-80 (1992). (This may also found in the University of Bielefeld, Germany preprint series (1992) - 036.)

[2] **Danh T.-N.,**, and **Daykin D.E. :** *Orderring integer vectors for coordinate deletions.* (to appear)

[3] **Daykin D.E. :** *Ordered ranked posets, representations of integers and inequalities from extremal poset problems, Graphs and Order.* Proc. Conf. Banff., Ed.: I. Rival, 395 - 412, Canada (1984)

received: July 26, 1994

Authors:

Daykin D.E.
Universität of Readings
England
RG6 - 2AX

Dhan T.-N.
University of HoChiMinh City
227. Ngugen van Cu
HoChiMinh City
Vietnam

Rostock. Math. Kolloq. **50**, 47–52 (1997)

Subject Classification (AMS)
05A99 06A07 68R15

Tran-Ngoc Danh; David E. Daykin

Sets of 0,1 vectors with minimal sets of subvectors

ABSTRACT. Let $\mathcal{C}(n)$ be the 0,1 vectors $\underset{\sim}{a} = a_1 \dots a_n$. To get a subvector of $\underset{\sim}{a}$ delete any a_i. If $\mathcal{A} \subseteq \mathcal{C}(n)$ then $\Delta\mathcal{A}$ is the set of all subvectors of members of $\mathcal{A}$, so $\Delta\mathcal{A} \subseteq \mathcal{C}(n-1)$. Put $W\underset{\sim}{a} = a_1 + \dots + a_n$. We order $\mathcal{C}(n)$ by $\underset{\sim}{a} < \underset{\sim}{b}$ if (i) $W\underset{\sim}{a} < W\underset{\sim}{b}$ or (ii) $W\underset{\sim}{a} = W\underset{\sim}{b}$ and $1 = a_i > b_i = 0$ for the least i with $a_i \neq b_i$. We present a completely new proof of our <u>Theorem</u>. If $\mathcal{A} \subseteq \mathcal{C}(n)$ and $\mathcal{I}$ is the first $|\mathcal{A}|$ members in $\mathcal{C}(n)$ then $|\Delta\mathcal{I}| \leq |\Delta\mathcal{A}|$.

1 Introduction

Let $\mathcal{C}(n)$ be the set of all vectors $\underset{\sim}{a} = a_1 \dots a_n$ having $a_h = 0$ or $a_h = 1$ for $1 \leq h \leq n$. Further let $\delta_h\underset{\sim}{a}$ be the subvector in $\mathcal{C}(n-1)$ obtained from $\underset{\sim}{a}$ by deleting coordinate a_h. The shadow $\Delta\underset{\sim}{a}$ of $\underset{\sim}{a}$ is the set $\{\delta_1\underset{\sim}{a}, \dots, \delta_n\underset{\sim}{a}\}$. The shadow $\Delta\mathcal{A}$ of any $\mathcal{A} \subseteq \mathcal{C}(n)$ is $\cup\{\underset{\sim}{a} \in \mathcal{A}\}\Delta\underset{\sim}{a}$. We are concerned with this problem. Given $0 \leq k \leq 2^n$ find $\mathcal{A} \subseteq \mathcal{C}(n)$ with cardinality $|\mathcal{A}| = k$ and shadow size $|\Delta\mathcal{A}|$ minimal. We solved the problem in [1] 1993 by Theorem 1 below, where $\mathcal{C}(n)$ is V-ordered. If $\underset{\sim}{a}, \underset{\sim}{b} \in \mathcal{C}(n)$ and $\underset{\sim}{a} \neq \underset{\sim}{b}$ then $\alpha(\underset{\sim}{a}, \underset{\sim}{b})$ is the first i in $1 \leq i \leq n$ with $a_i \neq b_i$, also $W\underset{\sim}{a} = a_1 + \dots + a_n$.

Definition 1 *(V-order). Let $\underset{\sim}{a}, \underset{\sim}{b} \in \mathcal{C}(n)$ with $\underset{\sim}{a} \neq \underset{\sim}{b}$. Then $\underset{\sim}{a} < \underset{\sim}{b}$ if (i) $W\underset{\sim}{a} < W\underset{\sim}{b}$ or (ii) $W\underset{\sim}{a} = W\underset{\sim}{b}$ and $a_j > b_j$ with $j = \alpha(\underset{\sim}{a}, \underset{\sim}{b})$.*

By an initial section IS we mean the first so many members of $\mathcal{C}(n)$ in V-order. In particular $IS(\underset{\sim}{a}) = \{\underset{\sim}{x} \in \mathcal{C}(n) : \underset{\sim}{x} \leq \underset{\sim}{a}\}$.

Theorem 1 *If $\mathcal{I}, \mathcal{A} \subseteq \mathcal{C}(n)$ have $|\mathcal{I}| = |A|$ and $\mathcal{I}$ is an IS the $|\Delta\mathcal{I}| \leq |\Delta\mathcal{A}|$.*

In this note we present a completely new proof of Theorem 1. Any statement not proved in full detail is easy to verify. We use condensed notation. If $\underset{\sim}{u} = u_1 \dots u_r$ and $\underset{\sim}{v} = v_1 \dots v_s$ are vectors then $\underset{\sim}{u}\,\underset{\sim}{v}$ is the vector $u_1 \dots u_r v_1 \dots v_s$. If $\mathcal{A} \subseteq \mathcal{C}(n)$ and $\underset{\sim}{e} \in \mathcal{C}(1)$ then $\mathcal{A}\underset{\sim}{e} = \{\underset{\sim}{a}\underset{\sim}{e} :$

$\underset{\sim}{a} \in \mathcal{A}\}$. By $\mathcal{A}+\mathcal{B}$ we mean the union $\mathcal{A}\cup\mathcal{B}$ and are saying that $\mathcal{A}\cap\mathcal{B}=\emptyset$, the empty set. Each $\mathcal{A}\subseteq\mathcal{C}(n+1)$ has a partion $\mathcal{A}=A0+B1$ where $A,B\subseteq\mathcal{C}(n)$. When A,B are IS we say $\mathcal{A}$ is *part compressed* PC.

We tacitly use the fact that for $\underset{\sim}{x},\underset{\sim}{y}\in\mathcal{C}(n)$ and any vectors $\underset{\sim}{a},\underset{\sim}{b}$ we have $\underset{\sim}{a}\underset{\sim}{x}\underset{\sim}{b}<\underset{\sim}{a}\underset{\sim}{y}\underset{\sim}{b}$ iff $\underset{\sim}{x}<y$. Given $\underset{\sim}{a}\in\mathcal{C}(n)$, if $\underset{\sim}{a}=1\dots1$ put $V\underset{\sim}{a}=1$, ortherwise let $V\underset{\sim}{a}$ denote the last h in $1\le h\le n$ with $a_h=0$. Then $\underset{\sim}{a}^*=\delta_j\underset{\sim}{a}$ where $j=V\underset{\sim}{a}$. It is important that $\underset{\sim}{a}^*=\max\{\Delta\underset{\sim}{a}\}$. Of course $\mathcal{A}^*=\{\underset{\sim}{a}^*:\underset{\sim}{a}\in\mathcal{A}\}$.

2 Initial sections of V-order

Let $\mathcal{I}$ be an IS. Put $\mathcal{I}=G0+H1$ and $G=G_00+G_11$ and $H=H_00+H_11$.
Then

(1) G,H are IS,
(2) $H\subseteq G$ and $H_0\subseteq G_1$ so $H_1\subseteq H_0\subseteq G_1\subseteq G_0$,
(3) $\Delta\mathcal{I}=G=\mathcal{I}^*$,
(4) $\Delta(\mathcal{I}0)=\mathcal{I}$,
(5) $\Delta(\mathcal{I}1)=G0+G1$ not usually IS.

G_1	$G0$	$H1$	H_0
	0000		
	1000		
	0100		
00	0010		
		0001	00
	1100		
10	1010		
		1001	10
01	0110		
		0101	10
		0011	
11	1110		
		1101	11
		1011	
		0111	
		1111	

Figure 1. The V-order of $\mathcal{C}(4)$

In Figure 1 we see the V-orders (i) of $\mathcal{C}(2)$ in G_1, H_0, (ii) of $\mathcal{C}(3)$ in G, H and (iii) of $\mathcal{C}(4)$ in $G0 + H1$. In the figure $\mathcal{C}(4)$ is cut into slices by the last coordinates of its vectors. Imagine we have this figure for every $\mathcal{C}(n)$.

Consider $IS(\underset{\sim}{a})$ growing with $\underset{\sim}{a}$. We only get an increase in G_1 (resp. H_0) when $\underset{\sim}{a} = \underset{\sim}{x}10$(resp. $\underset{\sim}{a} = \underset{\sim}{y}01$) at the end of a 0-slice (resp. at the beginning of a 1-slice) and the increase is $\underset{\sim}{x}$ (resp. $\underset{\sim}{y}$). Moreover if the 0-slice is immediately before the 1-slice then $\underset{\sim}{x} = \underset{\sim}{y}$. All other vectors in a 0-slice (resp. 1-slice) end 00 (resp.11). Therefore (6) if $\mathcal{I} = IS(\underset{\sim}{x}10)$ then $G_1 = H_0 + \underset{\sim}{x}$ otherwide $G_1 = H_0$.

Let $n \geq 3$ and S be the 0-slice just before a 1-slice T. If $|T| \geq 2$ then $|S| = 1$, but if $|T| = 1$ then $|S| \geq 2$.

3 Part compressed families

Let D, E be any IS of $\mathcal{C}(n)$ and put $\mathcal{B} = D0+E1$ so $\mathcal{B}$ is PC. Next Γ is the set of $\underset{\sim}{a} \in \mathcal{C}(n+1)$ of the form $\underset{\sim}{a} = \underset{\sim}{u}\underset{\sim}{z}1$, where $\underset{\sim}{z} = 0 \ldots 0$ of $\dim \underset{\sim}{z} \geq 1$, and either $\underset{\sim}{u} = \emptyset$ or $\underset{\sim}{u}$ ends 1. Also $\gamma\underset{\sim}{a} = \underset{\sim}{u}1\underset{\sim}{z}$ if $\underset{\sim}{a} = \underset{\sim}{u}\underset{\sim}{z}1 \in \Gamma$. We can now define a map $\psi : \mathcal{B} \to \mathcal{C}(n+1)$ by

$$\psi\underset{\sim}{a} = \begin{cases} \gamma\underset{\sim}{a} & \text{if } \underset{\sim}{a} \in \Gamma \text{ and } \gamma\underset{\sim}{a} \notin \mathcal{B}, \\ \underset{\sim}{a} & \text{otherwise} \end{cases}$$

Theorem 2 *If $\mathcal{B}$ is PC then $\Delta\psi\mathcal{B} \subseteq \psi\Delta\mathcal{B}$.*

Proof: Observe that the two ψ are different. We can have $\mathcal{B} = D0 + E1$ as above. Let $\underset{\sim}{x} \in \Delta\psi\mathcal{B}$. There is a $\underset{\sim}{b} \in \psi\mathcal{B}$ with $\underset{\sim}{x} \in \Delta\underset{\sim}{b}$ and an $\underset{\sim}{a} \in \mathcal{B}$ with $\underset{\sim}{b} = \psi\underset{\sim}{a}$. We distinguish three possibilities. Case 1. $\underset{\sim}{a} \notin \Gamma$. Here $\underset{\sim}{b} = \underset{\sim}{a}$ and $\underset{\sim}{a}$ ends 0 or 11. Case 2. $\underset{\sim}{a} \in \Gamma$ and $\underset{\sim}{b} \neq \underset{\sim}{a}$. So $\underset{\sim}{b} = \underset{\sim}{u}1\underset{\sim}{z}$ and $\underset{\sim}{a} = \underset{\sim}{u}\underset{\sim}{z}1$. Case 3. $\underset{\sim}{a} \in \Gamma$ and $\underset{\sim}{b} = \underset{\sim}{a}$. Here $\underset{\sim}{a} = \underset{\sim}{u}\underset{\sim}{z}1$ and $\underset{\sim}{u}1\underset{\sim}{z}, \underset{\sim}{a}$ are both in $\mathcal{B}$.

Please look at the last line of Case 1 in Figure 2. There we are given $\underset{\sim}{a} = \underset{\sim}{w}11$, $\underset{\sim}{x} = \underset{\sim}{w}1$ and $\underset{\sim}{w}$ ends 0. So $\underset{\sim}{w} = \underset{\sim}{u}\underset{\sim}{z}$ where $\underset{\sim}{u} = \emptyset$ or $\underset{\sim}{u}$ ends 1. As shown $\underset{\sim}{u}1\underset{\sim}{z}1 \in \mathcal{B}$ yielding $\underset{\sim}{u}1\underset{\sim}{z}, \underset{\sim}{u}\underset{\sim}{z}1 \in \Delta\mathcal{A}$ and $\underset{\sim}{x} = \psi\underset{\sim}{x} \in \psi\Delta\mathcal{B}$ as required. In the last line of Case 2 we are given $\underset{\sim}{a}, \underset{\sim}{b}, \underset{\sim}{x}, \underset{\sim}{u}'$ as shown for some $\underset{\sim}{w}$. Now $\underset{\sim}{u}$ ends 1 so $\underset{\sim}{x} = \underset{\sim}{w}01\underset{\sim}{z} \in \Delta\mathcal{B}$ and $\underset{\sim}{x} = \psi\underset{\sim}{x} \in \psi\Delta\mathcal{B}$ again. On the final line we are in Case 3 with $\underset{\sim}{x} = \underset{\sim}{u}'\underset{\sim}{z}1$ and $\underset{\sim}{u}' = \underset{\sim}{w}z_1$ where $\underset{\sim}{w}$ ends 1 or $\underset{\sim}{w} = \emptyset$. Hence $\underset{\sim}{u} = \underset{\sim}{w}\underset{\sim}{z_1}1$ and $\underset{\sim}{c} = \underset{\sim}{w}\underset{\sim}{z_1}11\underset{\sim}{z} \in \mathcal{B}$ and $\underset{\sim}{c} > \underset{\sim}{d} = \underset{\sim}{w}1\underset{\sim}{z_1}1\underset{\sim}{z}$. Because D is IS this gives $\underset{\sim}{a}, \underset{\sim}{d} \in \mathcal{B}$ and $\underset{\sim}{w}1\underset{\sim}{z_1}\underset{\sim}{z}$, $\underset{\sim}{w}\underset{\sim}{z_1}\underset{\sim}{z}1 = \underset{\sim}{x} \in \Delta\mathcal{B}$ so $\underset{\sim}{x} \in \psi\Delta\mathcal{B}$. $\Diamond$

	$\utilde{x}$	$\psi\utilde{a}$ $\quad$ $\utilde{a}$	$\in \Delta\mathcal{B}$	Comment
Case 1.	$\utilde{w}'0$	$\utilde{w}0$	$\utilde{w}'0$	
	$\utilde{w} \notin \Gamma$	”	$\utilde{w}$	
	$\utilde{w} \in \Gamma$	”	$\utilde{u}1\utilde{z},\ \utilde{u}\utilde{z}1$	$\utilde{a} = \utilde{u}\utilde{z}10 > \utilde{u}1\utilde{z}0 \in \mathcal{B}$ as $D = IS$
	$\utilde{w}'11$	$\utilde{w}11$	$\utilde{w}'11$	
	$\utilde{w}1,\ w_n = 1$	”	$\utilde{w}1$	
	$\utilde{w}1,\ w_n = 0$	”	$\utilde{u}1\utilde{z},\ \utilde{u}\utilde{z}1$	$\utilde{a} = \utilde{u}\utilde{z}11 > \utilde{u}1\utilde{z}1 \in \mathcal{B}$ as $E = IS$
Case 2.	$\utilde{u}1\utilde{z}'$	$\utilde{u}1\utilde{z}$ $\quad$ $\utilde{u}\utilde{z}1$	$\utilde{u}\utilde{z}'1$	
	$\utilde{u}\utilde{z}$	” $\quad$ ”	$\utilde{u}\utilde{z}$	
	$\utilde{u}'1\utilde{z}$	” $\quad$ ”	$\utilde{u}'\utilde{z}1$	$\utilde{u}$ ends 1 or $\utilde{u}' = \emptyset$
	$\utilde{u}'1\utilde{z}$	” $\quad$ ”	$\utilde{w}01\utilde{z}$	$\utilde{u}' = \utilde{w}0, \utilde{u} = \utilde{w}01, \utilde{a} = \utilde{w}01\utilde{z}1$
Case 3	$\utilde{u}\utilde{z}$	$\utilde{u}\utilde{z}1$	$\utilde{u}\utilde{z}$	
	$\utilde{u}\utilde{z}'1$	”	$\utilde{u}1\utilde{z},\ \utilde{u}\utilde{z}'1$	
	$\utilde{u}'\utilde{z}1$	”	$\utilde{u}'1\utilde{z},\ \utilde{u}'\utilde{z}1$	$\utilde{u}'$ ends 1 or $\utilde{u}' = \emptyset$
	$\utilde{u}'\utilde{z}1$	”	” $\quad$ ”	$\utilde{u}'$ ends 0. See text.

Figure 2. Proof table for Theorem 2.

Theorem 3 [1] *Let D, E be IS of $C(n)$ with $E \subseteq D$. Put $D = D_0 0 + D_1 1$ and $E = E_0 0, + E_1 1$ and $\mathcal{B} = D0 + E1$. Then*

$$\Delta\mathcal{B} = \begin{cases} D & \text{if } E_0 \subseteq D_1, \\ D_0 0 + E_0 1 & \text{if } D_1 \subseteq E_0. \end{cases}$$

Proof: The result holds as $\Delta\mathcal{B} = D_0 0 \cup D_1 1 \cup E_0 1$ by (3).$\Diamond$

Example: To show that $D_1 = E_0$ does not imply $\mathcal{B} = IS$, let $D = IS(1110)$ and $E = IS(0010)$ then $D_1 = E_0 = IS(001)$ but $\mathcal{B} = IS(11100) \backslash 00011$ and $\mathcal{B} \neq IS$.

[1] Remark on Theorem 3. Let f be the final vector in $\mathcal{B}$. Let $\mathcal{I}$ be the largest IS in $\mathcal{B}$ and put $\mathcal{F} = \mathcal{B} \backslash \mathcal{I}$. If $D_1 \subseteq E_0$ then $\mathcal{B} \neq IS$ and $\utilde{f}$ ends 1 and $\Delta\mathcal{B} = \mathcal{I}^* + \mathcal{F}^*$. If $E_0 \subseteq D_1$ and $\mathcal{B} \neq IS$ then $\utilde{f}$ ends 0. We do not use these facts in this note.

4 Proof of Theorem 1.

The result holds trivially for $n = 1, 2$. To use induction we assume it holds for $n \leq m$. Let $\mathcal{A} \subseteq C(m+1)$ and put $\mathcal{A} = A0 + B1$ so

$$\Delta A = A \cup (\Delta A)0 \cup B \cup (\Delta B)1.$$

By exchanging 0 and 1 we may assume $|A| \geq |B|$. Let D, E be IS of $\mathcal{C}(m)$ with $|D| = |A|$ and $|E| = |B|$ so $E \subseteq D$. By induction $|\Delta D| \leq |\Delta A|$ and $|\Delta E| \leq |\Delta B|$. Put $D = D_0 0 + D_1 1$ and $E = E_0 0 + E_1 1$ and $\mathcal{B} = D0 + E1$ as usual so $\mathcal{B}$ is PC.

Case 1. $E_0 \subseteq D_1$. By Theorem 3 we have $|\Delta\mathcal{B}| = |D| = |A| \leq |\Delta\mathcal{A}|$. We forget $\mathcal{A}$ and study the slices of $\mathcal{B}$ in Figure 1. Let R, S, T, U be the consecutive slices with S the last 0-slice containing a vector of $\mathcal{B}$. Thus R, T are 1-slices and $U \cap \mathcal{B} = \emptyset$. Let $\underset{\sim}{x}10$ be the last vector of S, then $\underset{\sim}{x}01$ is the first vector of T.

Case 1.1 $\underset{\sim}{x}01 \in \mathcal{B}$. Here $R, S \subseteq \mathcal{B}$ and $\mathcal{B} = IS$.

Case 1.2 $\underset{\sim}{x}01 \notin \mathcal{B}$. Here $T \cap \mathcal{B} = \emptyset$. If $\mathcal{B} = IS$ we are done, so assume $\mathcal{B} \neq IS$. Let $\underset{\sim}{f}$ be the final vector of $\mathcal{B}$ so $\underset{\sim}{f} \in S$. There is a first $\underset{\sim}{e} \notin \mathcal{B}$ and $\underset{\sim}{e} < \underset{\sim}{f}$ and $\underset{\sim}{e}$ ends 1. We exchange $\mathcal{B}$ for $\mathcal{D} = (\mathcal{B} \backslash \underset{\sim}{f}) + \underset{\sim}{e}$. Then $\mathcal{D}$ is in Case 1. Also $|\Delta\mathcal{D}\| = |\Delta\mathcal{B}| - 1$ by Theorem 3, so we start on $\mathcal{D}$. Repetition proves Theorem 1 holds in Case 1.

Case 2. $D_1 \subseteq E_0$. By Theorem 3 and induction

$$|\Delta\mathcal{B}| = |D_0| + |E_0| = |\Delta D| + |\Delta E| \leq |\Delta A| + |\Delta\mathcal{B}| \leq |\Delta\mathcal{A}|.$$

So we deal with $\mathcal{B}$. Let T be the last 1-slice containing a vector of $\mathcal{B}$. Again let R, S, T, U be consecutive slices. Further let $\underset{\sim}{x}10, \underset{\sim}{x}01$ be as before. Then $\underset{\sim}{x}01 \in \mathcal{B}$ but by definition of this case $\underset{\sim}{x}10 \notin \mathcal{B}$. Let $\underset{\sim}{g}$ be the first vector of S.

Case 2.1 $\underset{\sim}{g} \in \mathcal{B}$. Here $E_0 = D_1 + \underset{\sim}{x}$ and $|\Delta\mathcal{B}| = |D_0| + |E_0| = |D| + 1$ by Theorem 3. We have $|S| \geq 2$ and this implies $|T| = 1$. If $\underset{\sim}{e}$ is the first vector not in $\mathcal{B}$, then $\underset{\sim}{e} \in S$, and $\mathcal{D} = (\mathcal{B} \backslash \underset{\sim}{x}01) + \underset{\sim}{e}$ is an IS. Again using Theorem 3 we get $|\Delta\mathcal{D}| = |D_0| = |D| + 1 = |\Delta\mathcal{B}|$. Theorem 1 holds in this Case 2.1.

Case 2.2 $\underset{\sim}{g} \notin \mathcal{B}$. Put $\mathcal{D} = \psi\mathcal{B}$ and note that $\mathcal{D} \neq \mathcal{B}$ because $\underset{\sim}{g} = \psi(\underset{\sim}{x}01)$, except in the trivial case $\underset{\sim}{g} = \underset{\sim}{z}$. But $|\mathcal{D}| = |\mathcal{B}|$ and Theorem 2 tells that $|\Delta\mathcal{D}| = |\Delta\mathcal{B}|$. If $\mathcal{D} = A0 + B1$ then $B \subset E \subseteq D \subset A$ so we can only enter this Case 2.2 a finite number of times. We forget $\mathcal{B}$ and work on $\mathcal{D}$. Since $\mathcal{D}$ may not be PC we must start at the very beginning $\Diamond$

References

[1] **Danh, T.-N.**, and **Daykin, D. E. :** *Ordering integer vectors for coordinate deletions.* (Submitted to J. London Math. Soc., 12. Feb 1994.)

[2] **Danh, T.-N.**, and **Daykin, D. E. :** *Set Casades and vector Valleys in Pascal's triangle.* (Submitted to Theta, 14 July 1994.)

[3] **Danh, T.-N.**, and **Daykin, D. E. :** *Bezrukov-Gronau Order is Not Optimal.* Rostock Math. Kolloq. **50**, 45-46 (1997)

received: April 19, 1995

Authors:

T.-N. Danh
University of HoCiMinh City
VietNam
227. Nguyen von Cu
HoCiMinh City

D. E. Daykin
University of Reading
England
RG6-2AX

Rostock. Math. Kolloq. **50**, 53–64 (1997)

Subject Classification (AMS)
16S15, 08A05, 08A30

J. Synnatzschke

Zur Erzeugung von Algebren durch Unteralgebren mit Quadrat Null

ABSTRACT. For a simple algebra $\mathcal{A}$ with unit, a condition depending on $n \in \mathbb{N}$ is given being equivalent to the possibility to generate $\mathcal{A}$ by n subalgebras having square zero and all except one of them having dimension 1. As a corollary under this condition, the algebra can be generated by two subalgebras with square zero such that one of them has dimension 1 or 2 in dependence on whether n is even or odd. In the case of the algebra $\mathcal{B}(E)$ of all continuous linear operators on a Banach space E, the condition is fulfilled if and only if E is the nth power $E = E_0^n$ of a Banach space E_0. This way by elementary considerations, not only a problem considered by W. Żelazko and afterwards by P. Šemrl is finished but also completely extended to arbitrary simple algebras with unit.

KEY WORDS. Generation of algebras by subalgebras, subalgebras with square zero, commutative subalgebras, operators in powers of Banach spaces

1 Einführung

Für eine Algebra $\mathcal{A}$ ist das Problem ihrer Erzeugung durch eine *minimale* Anzahl von Unteralgebren mit möglichst *einfacher* Struktur von auf der Hand liegendem grundsätzlichem Interesse. Wir werden in der vorliegenden Note Algebren betrachten, die letztendlich durch nur zwei Unteralgebren der Art erzeugt werden können, daß die Multiplikation in ihnen trivial und eine von ihnen nur ein- oder zwei-dimensional ist. Trotz der scheinenden Spezialisierung sind Algebren dieses Typs ganz und gar nicht „exotisch“, und die Klasse solcher Algebren ist noch hinreichend umfangreich, wie weiter unten ersichtlich sein wird.

Wir gehen davon aus, daß in der Situation der Algebra $\mathcal{B}(E)$ aller stetigen linearen Operatoren auf einem (nicht-trivialen) Banach-Raume E zwei nicht offensichtlich etwas gemeinsam habende Probleme, nämlich

- die Darstellbarkeit von E als k-te Potenz $E = E_0^k$ eines Banach-Raumes E_0, d.h. als direkte Summe $E = E_1 \oplus \ldots \oplus E_k$ paarweise zueinander isomorpher Banach-Räume $E_1, \ldots, E_k$ und
- die Erzeugung von $\mathcal{B}(E)$ durch l Unteralgebren $\mathcal{B}_i$ mit Quadrat Null, d.h. $\mathcal{B}_i^2 = \{xy : x, y \in \mathcal{B}_i\} = \{0\}$ $(i = 1, \ldots, l)$

einen gemeinsamen „Nenner" darin finden, daß folgende Bedingungen zueinander äquivalent sind (siehe [3, 6]):

(i) $E = E_0^2$

(ii) $\mathcal{B}(E)$ wird von einer Unteralgebra $\mathcal{B}_0$ und einem Operator V mit jeweils Quadrat Null erzeugt.

(iii) $\mathcal{B}(E)$ gestattet mit einer Unteralgebra $\mathcal{B}_0$ und einem Operator V mit jeweils Quadrat Null die Darstellung $\mathcal{B}(E) = \mathcal{B}_0 + V\mathcal{B}_0 + \mathcal{B}_0 V + V\mathcal{B}_0 V$.

Genauer, im Falle $E = E_0^n$ wird $\mathcal{B}(E)$ einerseits algebraisch durch n Unteralgebren mit Quadrat Null erzeugt, die bis auf eine alle ein-dimensional sind. Andererseits kann das System der $n-1$ ein-dimensionalen Unteralgebren durch eine einzige $(n-1)$-dimensionale Unteralgebra ersetzt werden. Gewissermaßen dazwischen liegt nun die sich als Folgerung ergebende „optimale" Aussage, daß $\mathcal{B}(E_0^n)$ durch nur zwei Unteralgebren mit Quadrat Null erzeugt wird, von denen die eine nur ein- oder zwei-dimensional in Abhängigkeit davon ist, ob n gerade oder ungerade ist.

Das inverse Problem, ob im Falle der Erzeugung von $\mathcal{B}(E)$ durch n Unteralgebren mit Quadrat Null nicht $E = E_0^n$ sei, war bereits von W. Żelasko aufgeworfen und für $n = 2$ positiv [6], im allgemeinen Fall von P. Šemrl [3] aber negativ beantwortet worden (siehe auch [7] und die dort zitierte Literatur). Hier setzen wir an, und zwar werden wir die Situation $E = E_0^n$ auf naheliegende Weise mit Hilfe von Operatoren charaktrisieren und dies auf eine beliebige einfache Algebra $\mathcal{A}$ mit Eins übertragen (Abschnitt 2), dann nach einigen Hilfssätzen (Abschnitt 3) als wichtigstes Resultat zueinander äquivalente Bedingungen dafür angeben, daß $\mathcal{A}$ durch n Unteralgebren mit Quadrat Null erzeugt wird, von denen bis auf eine alle ein-dimensional sind (Abschnitt 4), und abschließend als Folgerung eine Reihe weiterer Aussagen, insbesondere für den Fall n gerade bringen (Abschnitt 5).

Interessant und überraschend in diesem Zusammenhang war, daß im Falle der Erzeugung von $\mathcal{B}(E)$ durch zwei Unteralgebren mit Quadrat Null eine von ihnen als endlich-dimensional angenommen [3] und wiederum durch eine Anzahl ein-dimensionaler Unteralgebren ersetzt werden kann – ein Fakt, der im Rahmen unserer Situation einer einfachen Algebra $\mathcal{A}$ mit Eins, wie man leicht zeigen kann, erhalten bleibt.

Wir wollen noch bemerken, daß das Problem der *algebraischen* (d.h. diskret topologischen) Erzeugung einer Algebra durch Unteralgebren mit Quadrat Null von derjenigen der (nicht-diskret) *topologischen* Erzeugung sehr verschieden ist, denn es gibt z.B. Banach-Räume E derart, daß $\mathcal{B}(E)$ durch keine Anzahl von Unteralgebren mit Quadrat Null algebraisch erzeugt werden kann [3], während im Falle $\dim E > 1$ stets eine Erzeugung in der starken Topologie durch nur zwei solche Unteralgebren möglich ist [5]. Von Interesse beim Problem der algebraischen Erzeugung ist auch ein für eine Klasse allgemeiner Algebren $\mathcal{A}$ bewiesener Fakt, der im Falle $\mathcal{B}(E)$ besagt, daß das Ideal aller endlich-dimensionalen Operatoren durch zwei Unteralgebren ein-dimensionaler Operatoren mit Quadrat Null erzeugt werden kann [4].

Ferner, Unteralgebren mit Quadrat Null sind trivialerweise kommutativ. Interessiert allgemeiner die topologische Erzeugung einer Algebra durch *kommutative* Unteralgebren, so sei für den Fall eines separablen Hilbert-Raumes H noch angeführt, daß $\mathcal{B}(H)$ in der starken Topologie durch zwei Operatoren (folglich durch zwei kommutative Unteralgebren) erzeugt werden kann, die zudem als unitär [1] oder hermitesch [2] angenommen werden können.

2 Der Begriff der n-zyklischen Zerlegung der Eins

Wir vereinbaren, bei fixiertem $n \in \mathbb{N}$ $(n \geq 2)$ größere Werte natürlicher Indizes immer modulo n zu verstehen und erinnern daran, daß eine Algebra $\mathcal{A}$ *einfach* heißt, wenn $x\mathcal{A}y = \{0\}$ für gewisse Elemente $x, y \in \mathcal{A}$ nur im Falle $x = 0$ oder $y = 0$ möglich ist. Einen nicht-trivialen Vektorraum X werden wir *n-te Potenz* eines Vektorraumes X_0 nennen und $X = X_0^n$ schreiben, wenn $X = X_1 \oplus \ldots \oplus X_n$ mit paarweise zueinander isomorphen Vektorräumen $X_1, \ldots, X_n$ ist.

Bezeichnet $\mathcal{A} = \mathcal{L}(X)$ die Algebra aller linearen Operatoren in $X = X_0^n$, sind $F_i : X_i \to X_{i+1}$ Isomorphismen und P_i die Projektoren auf X_i, so werden durch $U_i = F_i^{-1}P_{i+1}$ und $V_i = F_iP_i$ zwei n-Tupel nicht-trivialer linearer Operatoren $(U_i)_{i=1}^n$ mit $U_iU_j = 0$ für $i \neq j-1$ und $(V_i)_{i=1}^n$ mit $V_iV_j = 0$ für $i \neq j+1$ definiert, die den beiden Bedingungen

(i) $V_iU_i = U_{i+1}V_{i+1}$

(ii) $U_iV_iU_i = U_i$ und $V_iU_iV_i = V_i$

genügen und für die $\sum_{i=1}^n U_iV_i = I_X$ (I_X der identische Operator in X) ist. Umgekehrt, sind in einer Algebra $\mathcal{A} = \mathcal{L}(X)$ zwei n-Tupel nicht-trivialer linearer Operatoren $(U_i)_{i=1}^n$

und $(V_i)_{i=1}^n$ mit den Eigenschaften (i) und (ii) gegeben, so werden durch $P_i = U_iV_i$ nicht-triviale Projektoren definiert und die durch sie bestimmten Unterräume X_i bilden eine direkte Zerlegung des Grundraumes X. Die Beziehung $P_i = (P_iU_iP_{i+1})(P_{i+1}V_iP_i)$ zeigt dabei, daß jeder Projektor P_i als Produkt von Operatoren $F_i = P_{i+1}V_iP_i : X_i \longrightarrow X_{i+1}$ und $G_i = P_iU_iP_{i+1} : X_{i+1} \longrightarrow X_i$ dargestellt werden kann, die sich mit Hilfe der obigen Eigenschaften als zueinander inverse Isomorphismen erweisen. Also ist X_i isomorph zu X_{i+1} und damit $X = X_0^n$. Diese Überlegungen geben nun Anlaß zu der folgenden

Definition 2.1 *Es sei $\mathcal{A}$ eine Algebra mit Eins e und $2 \leq n \in \mathbb{N}$. Unter einer n-zyklischen Zerlegung $(u, v)_n$ von e werden wir zwei n-Tupel nicht-trivialer Elemente $u = (u_i)_{i=1}^n$ mit $u_iu_j = 0$ für $i \neq j-1$ und $v = (v_i)_{i=1}^n$ mit $v_iv_j = 0$ für $i \neq j+1$ verstehen, die durch die beiden Bedingungen*

- **(i)** $v_iu_i = u_{i+1}v_{i+1}$
- **(ii)** $u_iv_iu_i = u_i$ *und* $v_iu_iv_i = v_i$

miteinander gekoppelt sind und für die zusätzlich $\sum_{i=1}^n u_iv_i = e$ gilt.

Beispiel 2.2 [1] Wir identifizieren die Algebra $\mathcal{A} = \mathcal{L}(\mathbb{R}^n)$ $(2 \leq n \in \mathbb{N})$ auf natürliche Weise mit der Algebra $\mathcal{M}_{n\times n}$ aller $(n \times n)$-Matrizen $A = (a_{\alpha\beta})_{\alpha,\beta=1}^n$ und definieren Matrizen $U_i = (u_{\alpha\beta}^{(i)})_{\alpha,\beta=1}^n$ $(1 \leq i \leq n-1)$ und $U_n = (u_{\alpha\beta}^{(n)})_{\alpha,\beta=1}^n$ gemäß

$$u_{\alpha\beta}^{(i)} = \begin{cases} 1 & \text{für } \alpha = i, \beta = i+1 \\ 0 & \text{sonst} \end{cases} \qquad \text{und} \qquad u_{\alpha\beta}^{(n)} = \begin{cases} 1 & \text{für } \alpha = \beta = n \\ 0 & \text{sonst.} \end{cases}$$

Sei $\{e_i\}_{i=1}^n$ die natürliche Basis in $\mathbb{R}^n$ und $P_i \in \mathcal{M}_{n\times n}$ der Projektor auf den von e_i erzeugten Unterraum. Dann ist $U_iV_i = P_i$ und $(U, V)_n$ mit $U = (U_i)_{i=1}^n$ und $V = (V_i)_{i=1}^n$ ist eine n-zyklische Zerlegung der Einheitsmatrix I in $\mathcal{M}_{n\times n}$.

Wir schließen für die Situation einer n-zyklischen Zerlegung der Eins einige Bemerkungen an. So ist aus $u_i = u_iu_{i+1}v_{i+1}$ und $v_i = v_iv_{i-1}u_{i-1}$ sofort klar, daß $u_iu_{i+1} \neq 0$ und $v_iv_{i-1} \neq 0$ sein muß (siehe auch Lemma 3.3). Wegen $u_iv_j = u_iu_{i+1}v_{i+1}v_j$ und $v_iu_j = v_iv_{i-1}u_{i-1}u_j$ ist auch sofort klar, daß $u_iv_j = 0$ und $v_iu_j = 0$ für $i \neq j$ sein muß und durch $p_i = u_iv_i$ paarweise zueinander orthogonale Projektoren definiert werden. Wegen unserer Vereinbarung, n übertreffende Werte von Indizes immer modulo n zu verstehen, kommen in der obigen Definition die Bedingungen $u_iu_j = 0$ für $i \neq j-1$ und $v_iv_j = 0$ für $i \neq j+1$ übrigens erst im Falle $n \geq 3$ voll zum Tragen, und man überlegt sich leicht, daß eine Algebra $\mathcal{A}$ genau dann eine 2-zyklische Zerlegung der Eins e hat, wenn es in ihr Elemente u und v mit Quadrat Null derart gibt, daß $uv + vu = e$ ist (siehe auch Theorem 5.5).

[1]Der Autor dankt dem Gutachter herzlich für dieses Beispiel.

3 Einige Hilfssätze

Im folgenden wird in einer Algebra $\mathcal{A}$ eine Konstruktion der Form

$$\mathcal{A}_1 = \mathcal{A}_0 + \sum_{i=2}^{n}(v_i \cdots v_2)\mathcal{A}_0 + \sum_{j=2}^{n}\mathcal{A}_0(v_n \cdots v_j) + \sum_{i,j=2}^{n}(v_i \cdots v_2)\mathcal{A}_0(v_n \cdots v_j) \tag{1}$$

betrachtet werden, wobei $\mathcal{A}_0$ eine Unteralgebra mit Quadrat Null ist und die Elemente $v_2, \dots, v_n$ insbesondere Teil einer n-zyklischen Zerlegung der Eins sein können.

Lemma 3.1 *Es sei $\mathcal{A}$ eine Algebra, $\mathcal{A}_0$ eine Unteralgebra mit Quadrat Null sowie der Eigenschaft $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$ und $v_i \in \mathcal{A}$ $(i = 2, \dots, n)$ beliebige Elemente. Dann wird durch* (1) *eine Unteralgebra definiert.*

Beweis: Offensichtlich ist $\mathcal{A}_1$ ein Vektorraum und die Eigenschaft $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$ sichert seine Abgeschlossenheit bezüglich der Multiplikation. □

Nicht jede Unteralgebra $\mathcal{A}_0$ mit Quadrat Null erfüllt die im vorigen Lemma benutzte Eigenschaft $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$. Wir werden als nächstes aber zeigen, daß dies stets für eine gewisse Erweiterung $\mathcal{B}_0 \supset \mathcal{A}_0$ der Fall ist (vgl. [6]).

Lemma 3.2 *Es sei $\mathcal{A}$ eine Algebra und $\mathcal{A}_0$ eine Unteralgebra mit Quadrat Null. Dann existiert eine Unteralgebra $\mathcal{B}_0 \supset \mathcal{A}_0$ mit Quadrat Null sowie $\mathcal{B}_0\mathcal{A}\mathcal{B}_0 \subset \mathcal{B}_0$.*

Beweis: Sowohl durch

$$\mathcal{B}_{0l} = \Big\{ b \in \mathcal{A} \Big| \; b\mathcal{A}_0 = \{0\}, \text{ und } b'\mathcal{A}_0 = \{0\} \text{ impliziert } b'b = 0 \Big\}$$

als auch

$$\mathcal{B}_{0r} = \Big\{ c \in \mathcal{A} \Big| \; \mathcal{A}_0 c = \{0\}, \text{ und } \mathcal{A}_0 c' = \{0\} \text{ impliziert } cc' = 0 \Big\}$$

wird eine Unteralgebra mit den angezeigten Eigenschaften definiert. Für die spätere Anwendung dieses Lemmas beim Beweis der Implikation (ii) $\Rightarrow$ (iii) im Theorem 4.2 setzen wir $\mathcal{B}_0 = \mathcal{B}_{0l} \cap \mathcal{B}_{0r}$. □

Als nächstes werden wir zeigen, daß die Konstruktion (1) im Falle einer n-zyklischen Zerlegung der Eins nicht trivial ist.

Lemma 3.3 *Es sei $\mathcal{A}$ eine Algebra mit n-zyklischer Zerlegung der Eins $\big((u_i)_{i=1}^n, (v_i)_{i=1}^n\big)$. Dann sind für $2 \le i < j \le n$ alle Produkte $v_i \cdots v_j$ mit jeweils lückenloser Indexfolge ungleich Null.*

Beweis: Wir starten mit der Beziehung $v_n = v_n u_n v_n = v_n v_{n-1} u_{n-1}$ und setzen in diese $v_{n-1} = v_{n-1} v_{n-2} u_{n-2}$ ein. Wiederholte Anwendung dieses Prinzips für fallenden Index führt zur Darstellung $v_n = (v_n \cdots v_2) v_1 (u_1 \cdots u_{n-1})$ und die Behauptung folgt □

Die Überlegungen zum Beweis der folgenden Aussage sind die gleichen wie bei Lemma 3.1, und wir unterlassen deshalb irgendwelche Ausführungen.

Lemma 3.4 *Es sei $\mathcal{A}$ eine Algebra, $\mathcal{A}_0$ eine Unteralgebra mit Quadrat Null sowie $\mathcal{A}_0 \mathcal{A} \mathcal{A}_0 \subset \mathcal{A}_0$ und $v_i \in \mathcal{A}$ $(i = 2, \ldots, n)$ Elemente mit $v_i v_j = 0$ $(i \neq j+1)$ sowie $v_{i+1} \mathcal{A}_0 v_j = \{0\}$ $(2 \leq i, j \leq n-1)$. Dann ist*

$$\mathcal{A}_2 = \sum_{2 \leq i \leq n} \mathbb{R} v_i + \sum_{\substack{i>j \\ 2 \leq i,j \leq n}} \mathbb{R}(v_i \cdots v_j) + \mathcal{A}_1 \tag{2}$$

mit $\mathcal{A}_1$ gemäß (1) gleich der von $\mathcal{A}_0$ und den Elementen v_i $(i = 2, \ldots, n)$ erzeugten Unteralgebra.

Schließlich formulieren wir das folgende

Lemma 3.5 *Es sei $\mathcal{A}$ eine Algebra mit Eins e, die von einer Unteralgebra $\mathcal{A}_0$ mit Quadrat Null sowie $\mathcal{A}_0 \mathcal{A} \mathcal{A}_0 \subset \mathcal{A}_0$ und Elementen $v_i \in \mathcal{A}$ $(i = 2, \ldots, n)$ mit $v_i v_j = 0$ $(i \neq j+1)$ sowie $v_{i+1} \mathcal{A}_0 v_j = \{0\}$ $(2 \leq i, j \leq n-1)$ erzeugt wird. Dann gestattet $\mathcal{A}$ die Darstellung (1).*

Beweis: Nach Lemma 3.4 hat $\mathcal{A}$ die Form (2). Wir betrachten die ihr entsprechende Darstellung von e und bezeichnen in dieser mit e_{ij} die Elemente aus den Summanden $(v_i \cdots v_2) e_{ij} (v_n \cdots v_j)$ von (1). Multipliziert man die Darstellung von links mit v_i und von rechts mit v_{i-1} durch, so wird sich $v_i v_{i-1} = (v_i \cdots v_2) e_{(i-1)i} (v_n \cdots v_{i-1})$ $(3 \leq i \leq n)$ ergeben. Das zeigt, daß alle Summanden unter dem zweiten Summenzeichen von (2) bereits in $\mathcal{A}_1$ enthalten sind. Auf gleiche Weise können wir uns davon überzeugen, daß auch alle Summanden unter dem ersten Summenzeichen von (2) bereits in $\mathcal{A}_1$ enthalten sind und damit $\mathcal{A} = \mathcal{A}_1$ ist. □

4 Das Hauptresultat

Gestattet eine Algebra $\mathcal{A}$ eine Darstellung der Form (1), so kann jedes ihrer Elemente in der Form

$$a = a_0 + \sum_{i=1}^{n} (v_i \cdots v_2) a_i' + \sum_{i=1}^{n} a_j'' (v_n \cdots v_j) + \sum_{i,j=1}^{n} (v_i \cdots v_2) a_{i,j} (v_n \cdots v_j) \tag{3}$$

mit gewissen Elementen $a_0, a_i', a_j'', a_{ij} \in \mathcal{A}_0$ dargestellt werden und es steht nun noch die Frage nach deren Eindeutigkeit.

Lemma 4.1 *Es sei $\mathcal{A}$ eine einfache Algebra mit Eins, $v_i \in \mathcal{A}$ $(i = 2, \dots, n)$ Elemente mit $v_i v_j = 0$ $(i \neq j+1)$ und $\mathcal{A}_0$ eine Unteralgebra mit Quadrat Null sowie $v_{i+1}\mathcal{A}_0 = \mathcal{A}_0 v_j = \{0\}$ $(2 \leq i, j \leq n-1)$. Gestattet dann $\mathcal{A}$ die Darstellung (1), so sind in der entsprechenden Darstellung (3) für $a \in \mathcal{A}$ die Elemente a_0, a_i', a_j'' und a_{ij} eindeutig bestimmt.*

Beweis: Es kann $a = 0$ angenommen und muß dann $a_0 = a_i' = a_j'' = a_{kl} = 0$ gezeigt werden. Dazu schreiben wir (3) in der Form

$$\sum_{i=2}^{n}(v_i \cdots v_2)\left(a_i' + \sum_{j=2}^{n} a_{ij}(v_n \cdots v_j)\right) + \left(a_0 + \sum_{j=2}^{n} a_j''(v_n \cdots v_j)\right) = 0 \qquad (4)$$

und schränken uns auf einen Hinweis dessen ein, wie die Aussage für die Elemente a_0 und a_j'' im zweiten Summanden bewiesen werden kann (der Beweis für die Elemente im ersten Summanden würde analog verlaufen). Dazu bezeichnen wir den zweiten Summanden von (4) mit b und die Summe in ihm mit s und weisen nacheinander mit der gleichen Argumentation $b = 0, a_0 = 0, a_j''(v_n \cdots v_j) = 0, \ \dots \ , a_j'' v_n = 0$ und schließlich $a_j'' = 0$ nach. So ist z.B. wegen $\mathcal{A}_0^2 = \{0\}$ offensichtlich $\mathcal{A}_0 b = \{0\}$ und wegen $v_{i+1}\mathcal{A}_0 = \{0\}$ ist $v_{i+1} b = 0$ für $2 \leq i \leq n$. Multipliziert man (4) von links mit v_2, so folgt auch $v_2 b = 0$. Damit ist $\mathcal{A} b = \{0\}$ und deshalb notwendig $b = 0$. □

Theorem 4.2 *Es sei $\mathcal{A}$ eine einfache Algebra mit Eins e und $v_i \in \mathcal{A}$ $(i = 2, \dots, n)$ Elemente mit $v_i v_j = 0$ $(i \neq j+1)$. Dann sind die folgenden Bedingungen zueinander äquivalent:*

- **(i)** *Die Elemente v_i $(i = 2, \dots, n)$ können zu einer n-zyklischen Zerlegung der Eins $\big((u_i)_{i=1}^n, (v_i)_{i=1}^n\big)$ ergänzt werden.*
- **(ii)** *Es existiert eine Unteralgebra $\mathcal{A}_0$ mit Quadrat Null und $v_{i+1}\mathcal{A}_0 = \mathcal{A}_0 v_j = \{0\}$ $(2 \leq i, j \leq n-1)$ derart, daß $\mathcal{A}$ von $\mathcal{A}_0$ und $v_2, \dots, v_n$ erzeugt wird.*
- **(iii)** *Es existiert eine Unteralgebra $\mathcal{A}_0$ mit Quadrat Null und $v_{i+1}\mathcal{A}_0 = \mathcal{A}_0 v_j = \{0\}$ $(2 \leq i, j \leq n-1)$ derart, daß sich $\mathcal{A}$ in der Form (1) darstellen läßt.*

Beweis: *Die Implikation* (i) ⇒ (ii): Sei $\big((u_i)_{i=1}^n, (v_i)_{i=1}^n\big)$ eine n-zyklische Zerlegung der Eins e. Durch $\mathcal{A}_0 = v_1 \mathcal{A} v_1$ wird dann offensichtlich eine Unteralgebra mit $\mathcal{A}_0^2 = \{0\}$ und $v_{i+1}\mathcal{A}_0 = \mathcal{A}_0 v_j = \{0\}$ definiert. Ferner, mit $u_{i+1}v_{i+1} = v_i u_i = v_i u_i v_i u_i = (v_i v_{i-1}) u_{i-1}(u_i)$ beginnend und hierin nacheinander $u_k = v_{k-1}u_{k-1}u_k$ $(k = i-1, \dots, 2)$ einsetzend gelangt man zur Darstellung $u_{i+1}v_{i+1} = (v_i \cdots v_2) v_1 (u_1 \cdots u_i)$. Analog, mit $u_j v_j = u_j v_j u_j v_j = (u_j u_{j+1}) v_{j+1}(v_j)$ beginnend und hierin nacheinander $v_l = u_{l+1}v_{l+1}v_l$ $(l = j+1, \dots, n)$ einsetzend gelangt man zur Darstellung $u_j v_j = (u_j \cdots u_n u_1) v_1 (v_n \cdots v_j)$. In der trivialen Identität $a = eae$ für ein beliebiges Element $a \in \mathcal{A}$ ersetzen wir nun e links und rechts entsprechend durch die beiden Darstellungen

$$e = v_1 u_1 + \sum_{i=2}^{n} u_{i+1} v_{i+1} \qquad \text{und} \qquad e = u_1 v_1 + \sum_{j=2}^{n} u_j v_j$$

und die Summanden $u_{i+1}v_{i+1}$ und $u_j v_j$ durch deren oben erhaltene Darstellungen. Hieraus ergibt sich sofort, daß $\mathcal{A}$ von $\mathcal{A}_0$ und $v_2, \dots, v_n$ erzeugt wird.

Die Implikation (ii) $\Rightarrow$ (iii): Die Eigenschaft $v_{i+1}\mathcal{A}_0 = \mathcal{A}_0 v_j = \{0\}$ $(2 \le i, j \le n-1)$ überträgt sich offensichtlich auf die im Beweis von Lemma 3.2 konstruierte Unteralgebra $\mathcal{B}_0$, so daß von Anfang an $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$ angenommen werden kann und Lemma 3.5 die Aussage (iii) sichert.

Die Implikation (iii) $\Rightarrow$ (i): Wie im vorigen Schritt kann $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$ und (1) angenommen werden. Sei

$$e = e_0 + \sum_{i=2}^{n} (v_i \cdots v_2) e_i' + \sum_{j=2}^{n} e_j''(v_n \cdots v_j) + \sum_{i,j=2}^{n} (v_i \cdots v_n) e_{ij} (v_n \cdots v_j) \tag{5}$$

die entsprechende Darstellung der Eins. Wir multiplizieren diese von links bzw. rechts mit v_i $(2 \le i \le n)$ durch. Gleichsetzung beider Resultate führt mit Hilfe von Lemma 4.1 zu

$$e = (v_n \cdots v_2) a + a (v_n \cdots v_2) + \sum_{i=2}^{n-1} (v_i \cdots v_2) a (v_n \cdots v_{i+1}) \tag{6}$$

mit einem gewissen nicht-trivialen Element $a \in \mathcal{A}_0$. Wird diese Darstellung mit v_i durchmultipliziert (gleich ob von links oder rechts), so erhält man sofort die Beziehung $v_i = (v_i \cdots v_2) a (v_n \cdots v_i)$ $(2 \le i \le n)$. Setzt man ferner die Darstellung (6) in $e^2 = e$ ein, so kann mit Hilfe von Lemma 4.1 auf $a(v_n \cdots v_2)a = a$ geschlossen werden. Durch

$$\begin{aligned} u_2 &= a(v_n \cdots v_3) \\ u_i &= (v_{i-1} \cdots v_2) a (v_n \cdots v_{i+1}) \qquad (3 \le i \le n-1) \\ u_n &= (v_{n-1} \cdots v_2) a \end{aligned}$$

werden dann Elemente $u_i \in \mathcal{A}$ $(2 \le i \le n)$ definiert, die zusammen mit $u_1 = v_n$ und $v_1 = u_n$ eine n-zyklische Zerlegung der Eins $\big((u_i)_{i=1}^n, (v_i)_{i=1}^n\big)$ bilden. □

5 Folgerungen

Es sei $\mathcal{A}$ eine einfache Algebra mit n-zyklischer Zerlegung der Eins. Aus Theorem 4.2 ergibt sich dann einerseits auf offensichtliche Weise, daß $\mathcal{A}$ durch n Unteralgebren mit Quadrat Null erzeugt wird, die bis auf eine alle ein-dimensional sind. Andererseits kann deren Gesamtheit durch eine einzige $(n-1)$-dimensionale Unteralgebra ersetzt werden. Als in gewissem Sinne dazwischen liegendes „optimales“ Resultat wird sich im folgenden noch ergeben, daß $\mathcal{A}$ durch zwei Unteralgebren $\mathcal{A}_0$ und $\mathcal{A}_0'$ mit jeweils Quadrat Null und $\dim \mathcal{A}_0' = 1$ oder $\dim \mathcal{A}_0' = 2$ in Abhängigkeit davon erzeugt wird, ob n gerade oder ungerade ist.

Definition 5.1 *Wir werden eine n-zyklische Zerlegung der Eins einer Algebra* gerade *oder* ungerade *nennen, je nachdem, ob n gerade oder ungerade ist.*

Lemma 5.2 *Es sei $\mathcal{A}$ eine einfache Algebra mit n-zyklischer Zerlegung der Eins. Ist $n = r \cdot s$ mit $2 \leq r, s \in \mathbb{N}$, so hat $\mathcal{A}$ auch eine r-zyklische Zerlegung der Eins.*

Beweis: Eine r-zyklische Zerlegung der Eins $\left((u_i')_{i=1}^r, (v_i')_{i=1}^r\right)$ kann sofort durch $u_i' = \sum_{l=1}^s u_{(l-1)r+i}$ und $v_i' = \sum_{l=1}^s v_{(l-1)r+i}$ $(i = 1, \ldots, r)$ angegeben werden. □

Theorem 5.3 *Eine einfache Algebra $\mathcal{A}$ mit n-zyklischer Zerlegung der Eins kann durch zwei Unteralgebren $\mathcal{A}_0$ und $\mathcal{A}_0'$ mit Quadrat Null und $\dim \mathcal{A}_0' = 1$ oder $\dim \mathcal{A}_0' = 2$ in Abhängigkeit davon erzeugt werden, ob n gerade oder ungerade ist.*

Beweis: Ist n gerade, so kann nach Lemma 5.2 $e = uv + vu$ mit $u^2 = v^2 = 0$ angenommen und $\mathcal{A}_0 = u\mathcal{A}u$ sowie $\mathcal{A}_0' = \mathbb{R}v$ gesetzt werden.

Ist dagegen n ungerade und $\left((u_i)_{i=1}^{2l+1}, (v_i)_{i=1}^{2l+1}\right)$ eine $n = (2l+1)$-zyklische Zerlegung der Eins, so werden durch $u = \sum_{i=1}^l v_{2i}$ und $v = \sum_{i=1}^l u_{2i}$ Elemente mit Quadrat Null sowie $uu_1 = u_1u = vv_1 = v_1v = 0$ und $e = uv + vu + u_1v_1$ definiert, und es kann $\mathcal{A}_0 = u\mathcal{A}u + u\mathcal{A}u_1 + u_1\mathcal{A}u + u_1\mathcal{A}u_1$ und $\mathcal{A}_0' = \mathrm{Lin}\,\{v, v_1\}$ gesetzt werden. □

Beispiel 5.4 Ausgehend vom Beispiel 2.2 für $n = 4$ liefert Lemma 5.2 eine 2-zyklische Zerlegung $UV + VU = I$ der Eins in der Algebra $\mathcal{A} = \mathcal{M}_{4\times 4}$. Diese wird folglich durch die Unteralgebra

$$\mathcal{A}_0 = \begin{pmatrix} 0 & a & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & c & 0 & d \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (a, b, c, d \in \mathbb{R})$$

und die Matrix $V_0 = U_1 + U_3$ erzeugt. Dabei ist $\mathcal{A}_0^2 = \{0\}$ und $V_0^2 = 0$.

Für $n = 5$ erhalten wir $UV + VU + U_1V_1 = I$ mit $U = V_2 + V_4$ und $V = U_2 + U_4$. Die Algebra $\mathcal{A} = \mathcal{M}_{5\times 5}$ wird durch die Unteralgebra

$$\mathcal{A}_0 = \begin{pmatrix} 0 & a & 0 & b & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & b & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & d & 0 \end{pmatrix} \qquad (a, b, c, d, \in \mathbb{R})$$

und die beiden Matrizen V und V_1 erzeugt. Dabei ist $\mathcal{A}_0^2 = \{0\}$ und $V^2 = 0$ sowie $V_1^2 = 0$.

Wir werden nun über Theorem 4.2 für $n = 2$ hinaus noch mehrere zueinander äquivalente Bedingungen dafür angeben, daß eine einfache Algebra eine *gerade* zyklische Zerlegung der Eins gestattet. Den mehr oder weniger auf der Hand liegenden Beweis unterlassen wir.

Theorem 5.5 *Es sei $\mathcal{A}$ eine einfache Algebra mit Eins e. Dann sind die folgenden Bedingungen zueinander äquivalent.*

(i) *Es existiert eine gerade zyklische Zerlegung der Eins.*

(ii) *Es existiert eine 2-zyklische Zerlegung der Eins.*

(iii) *Es existieren Elemente u und v mit Quadrat Null derart, daß $uv + vu = e$ ist.*

(iv) *Es existiert eine Unteralgebra $\mathcal{A}_0$ und ein Element v mit jeweils Quadrat Null derart, daß $\mathcal{A}$ von $\mathcal{A}_0$ und v erzeugt wird.*

(v) *Es existiert eine Unteralgebra $\mathcal{A}_0$ und ein Element v mit jeweils Quadrat Null derart, daß $\mathcal{A} = \mathcal{A}_0 + v\mathcal{A}_0 + \mathcal{A}_0 v + v\mathcal{A}_0 v$ ist.*

Beispiel 5.6 Beispiel 2.2 liefert im Falle $n = 2$ eine Matrix $V_0 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ und durch $\mathcal{B}_0 = \mathbb{R}B_0$ mit $B_0 = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ eine Unteralgebra mit jeweils Quadrat Null. Wegen

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = cB_0 + V(aB_0) + (dB_0)V + V(bB_0)V$$

gilt dabei für die Algebra $\mathcal{A}$ die Zerlegung $\mathcal{A} = \mathcal{B}_0 + V\mathcal{B}_0 + \mathcal{B}_0 + V\mathcal{B}_0 V$.

Ein Theorem analog zu Theorem 5.5 für Algebren $\mathcal{A}$ mit *ungerader* zyklischer Zerlegung der Eins kann nicht formuliert werden. Wir wollen aber in diesem Zusammenhang bemerken, daß $\mathcal{A}$ nach Theorem 5.3 von zwei Unteralgebren $\mathcal{A}_0$ und $\mathcal{A}_0' = \mathrm{Lin}\,\{v, v_1\}$ mit Quadrat Null sowie $\mathcal{A}_0\mathcal{A}\mathcal{A}_0 \subset \mathcal{A}_0$ erzeugt wird und $e = e_1v + ve_2 + e_3v_1$ mit $e_1, e_2, e_3 \in \mathcal{A}_0$ ist. Im allgemeinen Fall der Erzeugung einer beliebigen Algebra mit Eins durch zwei solche Unteralgebren ist dagegen $e = e_1v + ve_2 + e_3v_1 + v_1e_4$, also in dieser Darstellung der Eins ein zusätzlicher Summand v_1e_4 vorhanden.

Literatur

[1] **Davis, C. :** *Generators of the ring of bounded operators.* Proc. Amer. Math. Soc. **6**, 970 – 972 (1955)

[2] **Nordgren, E. A., Radjabalipour, M., Radjavi, H.** und **Rosenthal, P. :** *Quadratic operators and invariant subspaces.* Studia Math. **88**, 263 – 268 (1988)

[3] **Šemrl, P. :** *On algebraic generation of $B(X)$ by two subalgebras with square zero.* Studia Math. **97**, 139 – 142 (1990)

[4] **Synnatzschke, J. :** *Zur Erzeugung des Ideals der endlichdimensionalen Elemente einer Algebra durch zwei Unteralgebren eindimensionaler Elemente mit trivialem Quadrat.* Rostock. Math. Kolloq. **45**, 21 – 24 (1991)

[5] **Żelazko, W. :** *$B(X)$ is generated in strong operator topology by two subalgebras with square zero.* Proc. Royal Irish Acad. (Sec. A) **88**, 19 – 21 (1988)

[6] **Żelazko, W. :** *Algebraic generation of $\mathcal{B}(X)$ by two subalgebras with square zero.* Studia Math. **90**, 205 – 212 (1988)

[7] **Żelazko, W. :** *Generation of $B(X)$ by two commutative subalgebras – results and open problems.* Banach Center Publ. **30**, 363 – 367 (1994)

eingegangen: 12.05.1995 **revidierte Fassung:** 1.3.1996

Autor:

J. Synnatzschke
Institut für Mathematik der Universität Leipzig
Augustuspl. 10
D – 04109 Leipzig
Germany

Rostock. Math. Kolloq. **50**, 65–70 (1997)

Subject Classification (AMS)
41A10, 41A29, 49K30, 65K10

Markus Kappert

Eine Gradaussage für Posinome bester Approximation

ABSTRACT. In this paper we consider the least-squares approximation by posinoms. We prove a statement about the degree of the best-approximation posinom in a special case. The proof is based on the Kuhn-Tucker-Theorem.

KEY WORDS. least - squares approximation, posinoms, degree

1 Einleitung

In verschiedenen Arbeiten haben sich L. Berg [1], [2], [3], [4] und B. Thielcke [8] mit der Interpolation einer Menge

$$S_n := \{(t_i, f_i)|i = 0, \dots, n\} \text{ mit } 0 = t_0 < t_1 < \dots < t_n \ (n \in \mathbb{N}_0)$$

durch Posinome [7], d.h. durch Polynome mit nichtnegativen Koeffizienten beschäftigt. Es liegt nahe, die Fragestellung zu erweitern und die Approximation im euklidischen Sinne durch Posinome zu betrachen. Gesucht ist dann ein Posinom P, für das

$$\sum_{i=0}^{n} (P(t_i) - f_i)^2 = min! \tag{1.1}$$

gilt. Hierbei ist auch $t_0 > 0$ zugelassen. Existiert ein Posinom P, das (1.1) erfüllt, so gibt es keine anderen Posinome, die (1.1) erfüllen (vgl. [3], S. 60). Wir nennen P das Posinom bester Approximation. Weitere Ergebnisse finden wir bei L. Berg in [3], [4], [5].
In Folgendem erkennen wir den praktischen Nutzen der Approximation durch Posinome: Sind bei der Approximation einer Menge S_n Oszillationen der Funktion und ihrer Ableitungen unerwünscht, so liegt es nahe, mit Posinomen zu approximieren. Insbesondere, wenn die

Wahl des Polynomgrads noch offen ist, bietet die Theorie der Posinome eine eindeutige Entscheidungsmöglichkeit (vgl. [1], S. 503). Genauso verhält es sich in Fällen, in denen man bemüht ist, diffizile nichtlineare Approximationen zu finden, ohne festzustellen, daß eine Gerade, die möglicherweise den Anstieg Null hat, die bessere Lösung ist.

2 Angabe und Beweis der neuen Aussage

Wir wollen hier eine Aussage über den Grad des Posinoms bester Approximation in einem Spezialfall beweisen:

Satz 1 *Es sei*

$$S_k := \{(t_i, f_i)|\ i = 0, \dots, k\} \text{ mit } 0 \leq t_0 < t_1 < \dots < t_k \quad (k \in \mathbb{N}_0).$$

Es sei $n \in \mathbb{N}_0$. Das Posinom bester Approximation zu S_n existiere und habe den Grad Null. Dann hat auch das Posinom bester Approximation zu S_{n+1}, wenn es existiert und S_{n+1} nicht interpoliert, den Grad Null.

Beweis:
Es sei $m \in \mathbb{N}_0$ beliebig, und es sei A die folgende $(n+1) \times (m+1)$ - Matrix:

$$A = \begin{pmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^m \\ 1 & t_1 & t_1^2 & \cdots & t_1^m \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & t_n & t_n^2 & \cdots & t_n^m \end{pmatrix}.$$

Es seien ferner

$$\vec{x}^T = (x_0, \dots, x_m) \in \mathbb{R}^{m+1} \text{ und } \vec{f}^T = (f_0, \dots, f_n) \in \mathbb{R}^{n+1}.$$

Dann ist die Bestimmung des Posinoms bester Approximation gleichbedeutend mit der Lösung des quadratischen Optimierungsproblems

$$\vec{x}^T\ A^T A\vec{x} - 2\vec{f}^T A\vec{x} = min!$$

unter der Nebenbedingung

$$\vec{x} \geq \vec{0}.$$

Dieses Problem ist äquivalent zu den Kuhn-Tucker-Bedingungen (vgl. [6])

$$\vec{v} = A^T A\vec{x} - A^T\vec{f} \geq \vec{0}, \quad \vec{x}^T\vec{v} = 0, \quad \vec{x} \geq 0 \tag{2.2}$$

mit $\vec{v} \in \mathbb{R}^{m+1}$.
Es sei $P_n(t) = x_0^{(n)}$ das Posinom bester Approximation zu S_n.
Ist P_n das Interpolationsposinom zu S_n und $f_{n+1} \geq x_0^{(n)}$, so existiert ein Interpolationsposinom P_{n+1} zu S_{n+1}. Ist P_n das Posinom bester Approximation zu S_n, aber nicht das Interpolationsposinom zu S_n, und gilt $f_{n+1} > x_0^{(n)}$, so existiert kein Posinom bester Approximation zu S_{n+1} (vgl. [3], Theorem 6). Wir können also voraussetzen, daß $f_{n+1} \leq x_0^{(n)}$ gilt.
Es sind die folgenden Fälle zu betrachten:

1. Fall:
Es ist

$$P_n(t) = x_0^{(n)} = \frac{1}{n+1} \sum_{i=0}^{n} f_i > 0.$$

Es sei $f_{n+1} \geq -\sum_{i=0}^{n} f_i$.
Ist $m = 0$, so ist $x_0^{(n+1)} := \frac{1}{n+2} \sum_{i=0}^{n+1} f_i \geq 0$. Dann gilt $v_0^{(n+1)} = 0$ in (2.2).
Es sei $m \in \mathbb{N}$ beliebig.
Wegen $v_0^{(n)} = 0$ und $x_j^{(n)} = 0$ für $j = 1, \ldots, m$ in (2.2) gilt für $j = 1, \ldots, m$:

$$v_j^{(n)} = \sum_{i=0}^{n} t_i^j \cdot x_0^{(n)} - \sum_{i=0}^{n} t_i^j \cdot f_i \geq 0.$$

Wiederum ist $x_0^{(n+1)} := \frac{1}{n+2} \sum_{i=0}^{n+1} f_i \geq 0$. Ferner sei $x_j^{(n+1)} = 0$ für $j = 1, \ldots, m$.

Dann gilt $v_0^{(n+1)} = 0$ in (2.2), und wir müssen zeigen, daß

$$v_j^{(n+1)} = \sum_{i=0}^{n+1} t_i^j \cdot x_0^{(n+1)} - \sum_{i=0}^{n+1} t_i^j \cdot f_i \geq 0 \qquad (j = 1, \ldots, m)$$

ist. Betrachten wir für $j = 1, \ldots, m$ nun $v_j^{(n+1)}$ als Funktion von f_{n+1}. Dann gilt für $j = 1, \ldots, m$:

$$\begin{aligned} \left(v_j^{(n+1)}\right)'(f_{n+1}) &= \sum_{i=0}^{n+1} t_i^j \cdot \frac{1}{n+2} - t_{n+1}^j \\ &< t_{n+1}^j - t_{n+1}^j = 0. \end{aligned}$$

Also ist $v_j^{(n+1)}$ für $j = 1, \ldots, m$ streng monoton fallend in f_{n+1}. Wir können daher wegen $f_{n+1} \leq x_0^{(n)}$ wie folgt abschätzen:

$$\begin{aligned} v_j^{(n+1)}(f_{n+1}) &\geq v_j^{(n+1)}(x_0^{(n)}) \\ &= \left(\sum_{i=0}^{n+1} t_i^j\right) \cdot \frac{\sum_{i=0}^{n} f_i + x_0^{(n)}}{n+2} - \sum_{i=0}^{n} t_i^j \cdot f_i - t_{n+1}^j \cdot x_0^{(n)} \\ &= \sum_{i=0}^{n+1} t_i^j \cdot x_0^{(n)} - \sum_{i=0}^{n} t_i^j \cdot f_i - t_{n+1}^j \cdot x_0^{(n)} \\ &\geq \sum_{i=0}^{n+1} t_i^j \cdot x_0^{(n)} - \sum_{i=0}^{n} t_i^j \cdot x_0^{(n)} - t_{n+1}^j \cdot x_0^{(n)} = 0, \end{aligned}$$

wobei $j = 1, \ldots, m$.
Mithin ist nach den Kuhn-Tucker-Bedingungen $P_{n+1}(t) = x_0^{(n+1)}$ das gesuchte Posinom bester Approximation.
Es seien $f_{n+1} < -\sum_{i=0}^{n} f_i$ und $m \in \mathbb{N}_0$ beliebig. Dann setzen wir $x_j^{(n+1)} = 0$ für $j = 0, \ldots, m$. Aufgrund von $x_j^{(n+1)} = 0$ für $j = 0, \ldots, m$ in (2.2) ist für diese j

$$\begin{aligned} v_j^{(n+1)} &= -\sum_{i=0}^{n+1} t_i^j \cdot f_i \\ &\geq -\sum_{i=0}^{n} t_i^j \cdot f_i + t_{n+1}^j \cdot \sum_{i=0}^{n} f_i \\ &\geq -\sum_{i=0}^{n} t_i^j \cdot x_0^{(n)} + t_{n+1}^j \cdot \sum_{i=0}^{n} f_i \\ &\geq -\left(\sum_{i=0}^{n} t_i^j - (n+1) \cdot t_{n+1}^j\right) \cdot x_0^{(n)} \geq 0 \end{aligned}$$

wegen $x_0^{(n)} > 0$ äquivalent zu

$$\sum_{i=0}^{n} t_i^j \leq (n+1) \cdot t_{n+1}^j.$$

Die letzte Zeile ist aber wegen

$$\sum_{i=0}^{n} t_i^j \leq \sum_{i=0}^{n} t_{n+1}^j \quad (j = 0, \ldots, m)$$

wahr, so daß auch hier nach den Kuhn-Tucker-Bedingungen $P_{n+1}(t) = x_0^{(n+1)}$ das gesuchte Posinom bester Approximation ist.

2. Fall:
Es ist

$$P_n(t) = x_0^{(n)} = 0.$$

Es sei $m \in \mathbb{N}_0$ beliebig. Wegen $x_j^{(n)} = 0$ für $j = 0, \ldots, m$ in (2.2) ist für diese j:

$$v_j^{(n)} = -\sum_{i=0}^{n} t_i^j \cdot f_i \geq 0.$$

Weiter gilt $f_{n+1} \leq x_0^{(n)} = 0$.
Wir setzen $x_0^{(n+1)} = 0$ und $x_j^{(n+1)} = 0$ für $j = 1, \ldots, m$, sofern $m \in \mathbb{N}$. Dann gilt für $j = 0, \ldots, m$:

$$\begin{aligned} v_j^{(n+1)} &= -\sum_{i=0}^{n+1} t_i^j \cdot f_i \\ &= -\sum_{i=0}^{n} t_i^j \cdot f_i - t_{n+1}^j \cdot f_{n+1} \geq 0. \end{aligned}$$

Somit folgt, wie oben, daß auch in diesem Fall die Behauptung bewiesen ist. Mithin ist alles gezeigt. ■

Abschließend möchte ich noch Herrn Prof. Dr. L. Berg danken, der mich auf die Fragestellung aufmerksam gemacht hat und mich freundlich beraten hat.

Literatur

[1] **Berg, L. :** *Zur Identifikation mehrfach monotoner Funktionen.* Z. Angew. Math. Mech. **65**, 497 - 507 (1985)

[2] **Berg, L. :** *Über vollmonotone Polynome durch vier gegebene Punkte.* Optimization **16**, 265 - 272 (1985)

[3] **Berg, L. :** *Existence, characterization and application of interpolating minimal polynomials.* Optimization **18**, 55 - 63 (1987)

[4] **Berg, L. :** *Interpolation and approximation by completely monotonous polynomials.* In: Greenspan, D., und Rózsa, P. (Ed.): Numerical Methods. Colloq. Math. Soc. János Bolyai **50**, pp. 589 - 604, Amsterdam 1991

[5] **Berg, L. :** *On the least-squares approximation of four points in the plain by posinoms.* Z. Angew. Math. Mech.**76** (4), 241-244 (1996)

[6] **Collatz, L.** und **Wetterling, W. :** *Optimierungsaufgaben.* Berlin 1971

[7] **Marusciac, I. :** *On a geometric programming problem with linear constraints.* Math., Rev. Anal. Numer. Theor. Approximation, Anal. Numer. Theor. Approximation **9**, 209 - 219 (1980)

[8] **Thielcke, B. :** *Zwei Algorithmen zur Berechnung interpolierender Minimalpolynome.* Rostock. Math. Kolloq. **30**, 56 - 62 (1986)

eingegangen: 22. Juni 1995

Autor:

Markus Kappert
Fachbereich Mathematik
Gerhard – Mercator – Universität
Gesamthochschule Duisburg
D-47048 Duisburg
Deutschland

Rostock. Math. Kolloq. **50**, 71–84 (1997)

Subject Classification (AMS)
65J05, 49D20

Dieter Schott

Basic Properties of Fejer monotone Mappings

Summary. *We consider certain classes of Fejer monotone mappings and their basic properties. These properties are the starting point of a convergence theory for corresponding iterative methods which are widely used to solve convex problems.*

1 Introduction

The present paper continues the paper [6] about basic properties of Fejer monotone sequences. In the applications such a sequence is often generated by a single Fejer monotone mapping g or by a sequence (g_k) of such mappings which can naturally be derived from the considered problem with the solution set M. More precisely we investigate set-valued mappings $g: Q \to \mathbb{P}(Q)$ with domain Q in a Hilbert space H and range in the power set of Q which reduce the distance between an element of Q and each element of the problem subset M. Then iterative methods

$$x_{k+1} \in g_k(x_k)\,, \quad x_0 \in Q$$

arise converging under slight additional assumptions to any element x^* in M (see e.g. [5] and the references cited there). These methods represent Fejer monotone sequences. In [5] a theory of strong convergent Fejer methods was developed using some basic properties of Fejer monotone mappings and sequences given here, in the preceding paper [6] and in the succeeding paper [7].

2 Fejer monotone mappings

Let H be a (real) Hilbert space. We consider a nonempty, convex and closed subset Q of H (set of feasible elements) and set-valued mappings (multioperators) $g : Q \to \mathbb{P}(Q)$, where $\mathbb{P}(Q)$ contains all nonempty subsets of Q. For g we introduce fixed point sets in the weak and in the strong sense, namely

$$F_{-}(g) = \{x \in Q : x \in g(x)\} \ , \ F_{+}(g) = \{x \in Q : \{x\} = g(x)\} \ ,$$

where $F_{+}(g) \subseteq F_{-}(g)$. As usual mappings (operators) $g : Q \to Q$ are integrated as imbeddings. Here both kinds of fixed point sets coincide with $F(g) = \{x \in Q : x = g(x)\}$. Further we introduce some relations and operations between set-valued mappings. Let $g \, , \, g_1 \, , \, g_2 \, : \, Q \to \mathbb{P}(Q)$. Then we define for all $y \in Q$:

$$(co\, g)(y) := co\; g(y) \quad , \quad g_2 \subseteq g_1 \ : \ g_2(y) \subseteq g_1(y) \ ,$$

$$(g_1 \cap g_2)(y) := g_1(y) \cap g_2(y) \quad , \quad (g_1 \cup g_2)(y) := g_1(y) \cup g_2(y) \ ,$$

$$(g_1 \circ g_2)(y) := \{x \in Q : x \in g_1(z) \, , \, z \in g_2(y)\} \ .$$

Now the central concepts are explained.

Definition 2.1 *Let M be a nonempty subset of Q. Then g is said to be M-*Fejer monotone *($g \in \mathbb{F}(M)$) iff*

$$\|z - x\| \le \|y - x\| \quad \forall x \in M \ , \ \forall y \in Q \ , \ \forall z \in g(y) \, .$$

It is said to be regularly M-Fejer monotone *($g \in \mathbb{F}_r(M)$) iff additionally*

$$y \notin g(y) \quad \forall y \in Q \setminus M \quad (or \quad F_{-}(g) \subseteq M) \ .$$

It is said to be strictly M-Fejer monotone *($g \in \mathbb{F}_{<}(M)$) iff additionally*

$$\|z - x\| < \|y - x\| \quad \forall x \in M \ , \ \forall y \in Q \setminus M \ , \ \forall z \in g(y) \, .$$

It is called (regularly, strictly) Fejer monotone *($g \in \mathbb{F} \, , \, g \in \mathbb{F}_r \, , \, g \in \mathbb{F}_{<}$) iff it is (regularly, strictly) M-Fejer monotone for any M.*

The set

$$C(g) = \{x \in Q \ : \ \|z - x\| \le \|y - x\| \quad \forall y \in Q \ , \ \forall z \in g(y)\}$$

is called the Fejer carrier *of g.*

Remarks 2.2 The above definitions are slight modifications of concepts given in [2] and [1] for special situations. In [2] a strictly Fejer monotone mapping g on $\mathbb{R}^n$ is said to be Fejer monotone. In [1] a strictly Fejer monotone continuous operator g on $\mathbb{R}^n$ is called *paracontractive*. But in contrast to our definition $C(g)$ is allowed to be empty.

Remarks 2.3

1. Obviously strictly Fejer monotone mappings g are regularly Fejer monotone and regularly Fejer monotone mappings are Fejer monotone (relative to the same set M), that is

$$\mathbb{F}_<(M) \subseteq \mathbb{F}_r(M) \subseteq \mathbb{F}(M) \quad \forall M \quad , \quad \mathbb{F}_< \subseteq \mathbb{F}_r \subseteq \mathbb{F} \; .$$

2. It is evident that

$$g \in \mathbb{F}_r(M) \iff g \in \mathbb{F}(M) \; , \; z \neq y \; : \; \forall y \in Q \setminus M \; , \; \forall z \in g(y) \; .$$

In the following $\mathbb{F}_\star$ stands either for $\mathbb{F}_r$ or for $\mathbb{F}_<$. If $\mathbb{F}_{(\star)}$ occurs, then the statement holds choicewise with and without $\star$.

Lemma 2.4 *Let (g_k) be a sequence of (regularly, strictly) Fejer monotone mappings all acting on Q . Then the iterative method*

$$x_0 \in Q \quad , \quad x_{k+1} \in g_k(x_k) \quad , \quad k \in \mathbb{N}$$

generates a (regularly, strictly) Fejer monotone sequence in the sense given in [6], respectively.

Proof: The definitions of Fejer monotone sequences in [6] and of Fejer monotone mappings are analogous. The latter are transformed in the first if y and z are replaced by x_k and x_{k+1}, respectively. ■

Theorem 2.5 *The following properties hold:*

a) $g \in \mathbb{F} \quad \iff \quad C(g) \neq \emptyset \quad \iff \quad g \in \mathbb{F}(C(g))$,

b) $g \in \mathbb{F}(M) \quad \Longrightarrow \quad g \in \mathbb{F}(C(g))$,

c) $g \in \mathbb{F}(M) \; , \; \emptyset \neq N \subseteq M \quad \Longrightarrow \quad g \in \mathbb{F}(N)$,

d) $g \in \mathbb{F}(M) \quad \Longrightarrow \quad M \subseteq C(g) \subseteq F_+(g)$,

e) $g \in \mathbb{F}_r(M) \quad \Longrightarrow \quad M = C(g) = F_+(g) = F_-(g)$.

Proof: Let be $g \in \mathbb{F}$. Then g is M-Fejer monotone for a certain nonempty set M. Hence $C(g)$ is nonempty, too. Further $C(g)$ supplies the maximal set for which Fejer monotony can be reached. This means $M \subseteq C(g)$ and $g \in \mathbb{F}(C(g))$. But this statement results again in $g \in \mathbb{F}$. Consequently, the assertions a) and b) hold. Assertion c) is obvious by definition. Let be $g \in \mathbb{F}(M)$ such that the estimate $\|z-x\| \leq \|y-x\|$ is fulfilled for $x \in C(g)$, $y \in Q$ and $z \in g(y)$. Then we can replace there y by x which leads to $z = x$ for all $z \in g(x)$, that is $x \in F_+(g)$. Therefore assertion d) is true. Finally we assume $g \in \mathbb{F}_r(M)$ and $x \in F_-(g)$. The second assumption implies $x \in g(x)$ and the first $x \notin g(x)$ for all $x \in Q \backslash M$. So we conclude $x \in M$. Hence, we get the relation chain $F_+(g) \subseteq F_-(g) \subseteq M$, which supplies in connection with d) also assertion e). ∎

Now we introduce critical points of Fejer monotone mappings which play an outstanding part in addition to the fixed points. Let $B(x,r)$ denote the closed ball with midpoint x and radius r.

Definition 2.6 *The mapping* $G_M : Q \to \mathbb{P}(Q)$ *with*

$$G_M(y) = \bigcap_{x \in M} B(x, \|y-x\|)$$

is said to be the Fejer zone *of the nonempty set* M*. The derived mapping* ∂G_M *, where* $\partial G_M(y)$ *is the boundary of* $G_M(y)$*, is called the* Fejer boundary *of* M*. Finally*

$$P_g(y) = (g(y) \cap \partial G_{C(g)}(y)) \backslash \{y\}$$

defines a mapping P_g *induced by* $g \in \mathbb{F}$ *whose domain* $D(P_g)$ *contains the so-called* critical points *of* g *and whose range* $R(P_g)$ *contains the so-called* exposed points *of* g*.*

Remarks 2.7 1. Because of $\mathbb{F}_<(M) \subseteq \mathbb{F}_r(M)$ and Theorem 2.5 e) the reference set M is uniquely determined for regularly and all the more for strictly Fejer monotone mappings g. Hence it is not necessary to state M in these cases. Besides we have

$$g \in \mathbb{F}(M)\ ,\ g \notin \mathbb{F}_\star(M) \ \Rightarrow\ \nexists N \subset M \ :\ g \in \mathbb{F}_\star(N)\ .$$

2. Without doubt $G_M \in \mathbb{F}(M)$ is satisfied. Moreover, any mapping $g \in \mathbb{F}(M)$ is characterized by the relation $g \subseteq G_M$. In so far the Fejer zone G_M is the maximal

M-Fejer monotone mapping. Additionally we obtain in view of this fact, Definition 2.6 and Theorem 2.5 b)

$$g \subseteq G_{C(g)} \subseteq G_M \ .$$

Obviously the images $G_M(y)$ represent nonempty, bounded and closed convex sets for all y since they arise from ball intersections and contain y. The Fejer boundary ∂G_M has the images

$$\partial G_M(y) = \{z \in Q : \exists x \in M \ : \ \|z - x\| = \|y - x\|\} \ \cap \ G_M(y)$$

and supplies beside its fixed points the exposed points of the Fejer zone G_M. So the critical points of g are the arguments which possess image elements on the Fejer boundary of its Fejer carrier $C(g)$ being no fixed points. These image elements are just the exposed points of g.

3. Theorem 2.5 d) shows that Fejer monotone mappings g have fixed points (in the strong sense). If g is regularly Fejer monotone, all the fixed points ly in $M = C(g)$ by Theorem 2.5 e). But g can possess critical points (outside of M). If g is strictly Fejer monotone, then there are no critical points.

Theorem 2.8 *The Fejer carrier $C(g)$ of g is convex and closed.*

Proof: The verification runs along the same lines as in [6, Theorem 2.6], where the Fejer carrier $C(x_k)$ of a Fejer monotone sequence (x_k) is proven to be convex and closed. The starting point is the reformulation of $C(g)$ as

$$C(g) = \{x \in Q : (y - z, y + z - 2x) \geq 0 \quad \forall y \in Q \ , \ \forall z \in g(y)\} \ . \blacksquare$$

If we define the sets

$$Q(y, z) = \{x \in Q \ : \ \|z - x\| \leq \|y - x\|\}$$

which correspond to halfspaces with normals $y - z$ and elements $\frac{1}{2}\,(y + z)$ on the boundary, then the Fejer carrier of g can obviously be described in the form

$$C(g) = \bigcap_{y \in Q, z \in g(y)} Q(y, z) \ .$$

The geometrical characterization of $Q(y, z)$ follows immediately from the scalar product representation of the squared defining inequality (compare also with the above reformulation of $C(g)$).

Theorem 2.9 *The following properties hold:*

a) $g \in \mathbb{F}(M) \quad \Longrightarrow \quad \overline{co}\, g \in \mathbb{F}(M)$,

b) $g \in \mathbb{F}_\star(M) \quad \Longrightarrow \quad co\, g \in \mathbb{F}_\star(M)$,

c) $g_1 \in \mathbb{F}_{(\star)}(M)\, ,\, g_2 \subseteq g_1 \quad \Longrightarrow \quad g_2 \in \mathbb{F}_{(\star)}(M)$,

d) $g_1 \in \mathbb{F}_{(\star)}(M)\, ,\, g_2 \in \mathbb{F}_{(\star)}(M) \quad \Longrightarrow \quad g_1 \cup g_2 \in \mathbb{F}_{(\star)}(M)$.

e) $g_1 \in \mathbb{F}(M)\, ,\, g_2 \in \mathbb{F}_{(<)}(M) \quad \Longrightarrow \quad g_1 \circ g_2 \in \mathbb{F}_{(<)}(M)$.

Proof: a) Let be $x \in M$, $y \in Q$, $z \in g(y)$, $z' \in g(y)$ and $\lambda \in (0,1)$. Then $g \in \mathbb{F}(M)$ implies $\|z-x\| \le \|y-x\|$ and $\|z'-x\| \le \|y-x\|$. Consequently for $z_\lambda = \lambda\, z + (1-\lambda)\, z'$ the estimate

$$\begin{aligned} \|z_\lambda - x\| &= \|\lambda\,(z-x) + (1-\lambda)\,(z'-x)\| \\ &\le \lambda\,\|z-x\| + (1-\lambda)\,\|z'-x\| \\ &\le \lambda\,\|y-x\| + (1-\lambda)\,\|y-x\| \;=\; \|y-x\| \end{aligned}$$

holds. Hence $co\, g \in \mathbb{F}(M)$. Now we consider any $z \in (\overline{co}\, g)(y)$. Then there is a sequence (z_n) with $z_n \in (co\, g)(y)$ for all n and $\lim_{n\to\infty} z_n = z$. Because of $co\, g \in \mathbb{F}(M)$ we have $\|z_n - x\| \le \|y-x\|$ for all n . So the continuity of the norm supplies

$$\|z-x\| = \lim_{n\to\infty} \|z_n - x\| \le \|y-x\| \;,$$

that is $\overline{co}\, g \in \mathbb{F}(M)$.

b) Let be $x \in M$, $y \in Q\backslash M$, $z \in g(y)$, $z' \in g(y)$ and $\lambda \in (0,1)$. Then $g \in \mathbb{F}_<(M)$ implies $\|z-x\| < \|y-x\|$ and $\|z'-x\| < \|y-x\|$. Thus we get

$$\begin{aligned} \|z_\lambda - x\| &\le \lambda\,\|z-x\| + (1-\lambda)\,\|z'-x\| \\ &< \lambda\,\|y-x\| + (1-\lambda)\,\|y-x\| = \|y-x\| \;. \end{aligned}$$

Hence also $co\, g \in \mathbb{F}_<(M)$.

Now we assume $g \in \mathbb{F}_r(M)$. Observing the chain of estimates for $\|z_\lambda - x\|$ in part a) of the proof the equation $\|z_\lambda - x\| = \|y-x\|$ has the consequence

$$\begin{aligned} \|\lambda\,(z-x) + (1-\lambda)\,(z'-x)\| &= \lambda\,\|z-x\| + (1-\lambda)\,\|z'-x\| \;, \\ &= \|y-x\| \;. \end{aligned}$$

The assumptions $\lambda \in (0,1)$, $\|z-x\| \le \|y-x\|$ and $\|z'-x\| \le \|y-x\|$ show that

$$\|z-x\| = \|z'-x\| = \|y-x\| \;.$$

Because of $y \in Q\backslash M$ and $x \in M$ we have $y - x \neq 0$ and by the second equation also $z - x \neq 0$ and $z' - x \neq 0$. Since the Hilbert space H is strictly convex, which means that

$$\|u + v\| = \|u\| + \|v\| \implies \exists\, c > 0 \,:\, v = cu$$

holds for elements $u \neq 0$, $v \neq 0$ of H, the first equation gives $z' - x = d\,(z - x)$ for a certain $d > 0$ and therefore $\|z' - x\| = d\,\|z - x\|$. Paying again attention to the second equation we obtain $d = 1$, $z' = z$ and finally $z_\lambda = z$. As $g \in \mathbb{F}_r(M)$ and $y \in Q\backslash M$ leads to $z \neq y$, we have for such y also $z_\lambda \neq y$. But this means $co\, g \in \mathbb{F}_r(M)$.

c) , d) Here the assertions follow immediately from the definition of the inclosed concepts.

e) First we suppose $g_1 \in \mathbb{F}(M)$ and $g_2 \in \mathbb{F}(M)$. Further let be $x \in M$ and $y \in Q$. Then for $u \in (g_1 \circ g_2)(y)$ there is an element $z \in g_2(y)$ with $u \in g_1(z)$ such that the estimates

$$\|u - x\| \leq \|z - x\| \leq \|y - x\|$$

are satisfied. So $g_1 \circ g_2 \in \mathbb{F}(M)$ follows. If g_2 lies even in $F_<(M)$, then for $y \in Q\backslash M$ arises $\|z - x\| < \|y - x\|$ and consequently $\|u - x\| < \|y - x\|$. Hence in this case $g_1 \circ g_2 \in \mathbb{F}_<(M)$. ■

Remarks 2.10 The results in Theorem 2.9 can be combined to get new results. For instance a) and d) supply

$$g_1 \in \mathbb{F}(M)\,,\, g_2 \in \mathbb{F}(M) \implies \overline{co}\,(g_1 \cup g_2) \in \mathbb{F}(M)\;.$$

Further c) leads to

$$g_1 \in \mathbb{F}_{(\star)}(M) \quad \text{or} \quad g_2 \in \mathbb{F}_{(\star)}(M) \implies g_1 \cap g_2 \in \mathbb{F}_{(\star)}(M)\;.$$

Besides e) contains the special case

$$g_1 \in \mathbb{F}_<(M)\,,\, g_2 \in \mathbb{F}_<(M) \implies g_1 \circ g_2 \in \mathbb{F}_<(M)\;.$$

By the way, e) induces the question, wether the strict case can be reached also if only the first term g_1 is strict. But $g_1 \circ g_2 \in \mathbb{F}_<(M)$ is satisfied for $g_1 \in \mathbb{F}_<(M)$ and $g_2 \in \mathbb{F}(M)$ only under additional assumptions. For instance, $g_2(y) \cap M = \emptyset$ implies $\|u - x\| < \|z - x\|$ and hence $\|u - x\| < \|y - x\|$ for $u \in (g_1 \circ g_2)(y)$ since $z \in g_2(y)$ is then in $Q\backslash M$. Besides the question arises what happens in the regular case. Here it does not suffice for $g_1 \circ g_2 \in \mathbb{F}_r(M)$ that both g_1 and g_2 are in $\mathbb{F}_r(M)$. Namely, $\|u - x\| = \|z - x\| = \|y - x\|$, $z \neq y$, $u \neq z$ can lead to $u = y$.

Corollary 2.11 *If* $g \in \mathbb{F}(M)$ *and* $\lambda \in (0,1)$ *, then*

$$g_\lambda = (1-\lambda)I + \lambda\, g \in \mathbb{F}(M)\ , \tag{2.1}$$

where I *represents the identity mapping.*

Proof: By assumption it is $g \in \mathbb{F}(M)$. Obviously we have also $I \in \mathbb{F}(M)$. Using Theorem 2.9 a) and d) we get $\overline{co}(I \cup g) \in \mathbb{F}(M)$ (see also Remark 2.10). Since g_λ belongs to the convex hull of I and g, Theorem 2.9 c) shows the assertion. ■

Remark 2.12 The above result has an important modification. If $g \in \mathbb{F}_r(M)$ holds, then even $g_\lambda \in \mathbb{F}_<(M)$ is true. More precisely, g_λ turns out to be in the subclass $\mathbb{F}^\alpha(M)$ of $\mathbb{F}_<(M)$ with $\alpha = \frac{1-\lambda}{\lambda}$ (for the definition see Example 3.3). This is proven in [7] by a more detailed investigation.

3 Examples

3.1 General examples of Fejer monotone mappings

Example: 3.1 We start with a well-known class of operators $g : Q \to Q$, namely the *nonexpansive* operators ($g \in \mathbb{L}$) defined by

$$\|g(y) - g(x)\| \le \|y - x\| \quad \forall y \in Q\ ,\ \forall x \in Q\ . \tag{3.2}$$

Here we have $F(g) = F_+(g) = F_-(g)$. We call g *regularly nonexpansive* ($g \in \mathbb{L}_r$) if g has additionally fixed points ($F(g) \neq \emptyset$). In this case

$$\|g(y) - x\| \le \|y - x\|$$

arises for $x \in F(g)$. Hence such an operator g is Fejer monotone relative to $M = F(g)$. Observing Theorem 2.5 d) we get also $C(g) = F(g)$. So g is even regularly Fejer monotone ($\mathbb{L}_r \subseteq \mathbb{F}_r$). Under the following conditions a nonexpansive operator g is for instance regular:

a) Q is bounded (Browder, Göhde, Kirk, e.g. [8, p. 115]) and therefore weakly compact,

b) $Q = H$ and g is linear ($0 \in F(g)$),

c) g satisfies $\|g(y) - g(x)\| < \|y - x\|$ for all $x, y \in Q$ with $x \neq y$ and g is compact (e.g. [3, p. 46-48], [4, p. 510-512]),

d) g is contractive (Lipschitz-continuous with constant $L < 1$, Banach).

In the linear case b) nonexpansivity means

$$\|g(y)\| \le \|y\| \quad \forall y \in H$$

because of

$$\|g(y) - g(x)\| = \|g(y-x)\| \le \|y - x\| \ .$$

This corresponds to Fejer monotony relative to 0 and automatically also relative to $F(g)$, since

$$\|g(y) - x\| = \|g(y) - g(x)\|$$

holds for all $x \in F(g)$. Hence for linear operators all classes $\mathbb{L}$, $\mathbb{L}_r$, $\mathbb{F}$ and $\mathbb{F}_r$ coincide.

Example: 3.2 A subclass of the regularly nonexpansive operators form the *strictly nonexpansive operators* g ($g \in \mathbb{L}_<$) given by $F(g) \neq \emptyset$ and

$$\|g(y) - g(x)\| < \|y - x\| \quad \forall x\,,\, y \in Q \ : \ g(y) - g(x) \neq y - x \ . \tag{3.3}$$

This definition differs from that in [1, p. 306] essentially by the additional demand that $g \in \mathbb{L}_<$ has to possess fixed points. For $x \in F(g)$ and $y \in Q \backslash F(g)$ the condition $g(y) - g(x) \neq y - x$ is fulfilled. So we get

$$\begin{aligned} \|g(y) - x\| &\le \|y - x\| \quad \forall x \in F(g)\,,\ \forall y \in Q\,, \\ \|g(y) - x\| &< \|y - x\| \quad \forall x \in F(g)\,,\ \forall y \in Q \backslash F(g)\,. \end{aligned}$$

Hence g is strictly Fejer monotone ($\mathbb{L}_< \subseteq \mathbb{F}_<$). The mapping g satisfying (3.3) has for instance fixed points and is therefore strictly nonexpansive under the conditions a) and b) stated in Example 3.1. The assumptions about g cited there in c) and d) lead directly to strictly nonexpansive operators since (3.3) is then fulfilled automatically.

One simple and important example for a strictly Fejer monotone operator is given in [1]. Let M be a nonempty, convex and closed subset of H and let P denote the metric projector onto M. Then the relaxed projector

$$P_\lambda = (1-\lambda)I + \lambda\, P \quad , \quad \lambda \in (0, 2) \tag{3.4}$$

is strictly nonexpansive ([1, p. 307]). In [7] this result will be sharpened (see also the remark in Example 3.3 below). Observe that the parameter λ can vary here in a larger interval than in Corollary 2.11 for the relaxation (2.1).

If g is strictly nonexpansive and linear (condition b) above), then the defining relations reduce to

$$\|g(y)\| < \|y\| \quad \forall y \notin F(g) \ .$$

Hence for linear operators strict nonexpansivity means strict Fejer monotony ($\mathbb{L}_< = \mathbb{F}_<$ in the linear case).

Example: 3.3 Let M be a nonempty subset of Q and α a positive number. Then g is said to be *α-strongly M-Fejer monotone* ($g \in \mathbb{F}^\alpha(M)$) iff it is

$$\|y-x\|^2 - \|z-x\|^2 \geq \alpha\|y-z\|^2 \quad \forall x \in M \ , \ \forall y \in Q \ , \ \forall z \in g(y) \tag{3.5}$$

and

$$y \notin g(y) \quad \forall y \in Q \backslash M \ . \tag{3.6}$$

It is called *strongly Fejer monotone* ($g \in \mathbb{F}_s$) iff it is α-strongly M-Fejer monotone for any M and any α.
The condition (3.5) in the above definition ensures that the mapping g is M-Fejer monotone. The condition (3.6) supplies the regularity of g. But g is also strictly M-Fejer monotone because of $z \neq y$ for $y \in Q\backslash M$ and $z \in g(y)$. Namely, such y, z generate a positive right-hand side in the inequality (3.5). Hence we have again $M = C(g)$.
The *basic examples* of Fejer monotone mappings are strongly Fejer monotone. For instance the relaxed projector P_λ in (3.4) turns out to be α-strongly M-Fejer monotone with $\alpha = \frac{2-\lambda}{\lambda}$ (see [7]). Consequently P itself is 1-strongly M-Fejer monotone. A further important example is given by the relaxed operator

$$T_\lambda(y) \quad = \quad \begin{cases} y - \lambda \frac{b(y)-c}{\|b'(y)\|^2}\, b'(y) & \text{if} \quad b(y) > c \\ y & \text{if} \quad b(y) \leq c \end{cases} \quad , \quad \lambda \in (0,2) \quad ,$$

where b is a convex and continuously differentiable functional on Q. This operator is proven in [1, p. 308] to be strictly Fejer monotone. But we show in [7] that a generalization of T_λ is even α-strongly Fejer monotone with $\alpha = \frac{2-\lambda}{\lambda}$. By the way, in this case we get $C(T_\lambda) = \{x \in Q : b(x) \leq c\}$ for the Fejer carrier.
The subclass $\mathbb{F}_s$ is described in detail in the succeding paper [7].

3.2 Fejer monotone mappings on $\mathbb{R}$

We turn to Fejer monotone mappings with $H = \mathbb{R}$. Let $g : Q \to \mathbb{P}(Q)$ be such a mapping on real intervals $Q = [a, b]$. The reference set M is then a subinterval $[c, d]$. The norm inequality for Fejer monotony specializes to $|z - x| \leq |y - x|$ which is at first considered for $Q = \mathbb{R}$ and fixed $x \in \mathbb{R}$. Then the inequality describes the left and right located domain $\mathbb{G}_x$ between the straight lines with ascents $+1$ and -1 through the point (x, x) in the $y-z-$plane. If x is allowed to vary in the interval M, then the solution set of the inequality reduces to the graph $\mathbb{G}_M(\mathbb{R}) = \bigcap_{x \in M} \mathbb{G}_x$ of G_M (see Definition 2.6) consisting of the two quarter planes (apart from the vertices)

$$\mathbb{G}_1 = \{(y,z) : \ y < c \, , \ y \leq z \leq 2c - y\}$$

$$\mathbb{G}_2 = \{(y,z) : \ y > d \, , \ 2d - y \leq z \leq y\}$$

with the half-line boundaries B_1^+ , B_1^- and B_2^- , B_2^+ , respectively, and the connecting straight line segment

$$\mathbb{G}_3 = \{(y,z) : \ c \leq y \leq d \, , \ z = y\} \, .$$

Hence $g : \mathbb{R} \to \mathbb{P}(\mathbb{R})$ is M-Fejer monotone iff the graph of g is contained in $\mathbb{G}_M(\mathbb{R})$. Especially, the graph has to coincide with $\mathbb{G}_3$ over M. The mapping g is even regularly M-Fejer monotone iff additionally the graph does not meet $B_1^+ \cup B_2^+$, and strictly M-Fejer monotone iff the graph meets no boundary segment of $\mathbb{G}_1$ and $\mathbb{G}_2$. Finally, the graph of α-strongly M-Fejer monotone mappings is additionally restricted (see Example 3.3). Namely, the graph has to be located outside of $\mathbb{G}_3$ in angle domains of $\mathbb{G}_1$ and $\mathbb{G}_2$ arising if the legs with ascent -1 are replaced by legs with ascent $\frac{\alpha-1}{\alpha+1}$. This can easily be derived if you start with the defining inequality

$$(y-x)^2 - (z-x)^2 \geq \alpha \, (y-z)^2 \, ,$$

substitute $u = y - x$, $v = z - x$ and use the third binomial formula to simplify the expression for $y \neq z$ and $u \neq v$, respectively.
If $Q \subset \mathbb{R}$ holds, then $\mathbb{G}_M(\mathbb{R})$ is reduced to $\mathbb{G}_M(Q) := \{(y,z) \in \mathbb{G}_M(\mathbb{R}) \ : \ y \in Q\}$. Obviously the left and right lying subdomains $\mathbb{G}_1$, $\mathbb{G}_2$ are restricted then in the same way.

Now we present for illustration some simple examples which show the relationship between the mentioned classes of mappings. All examples apart from the last relate to the special case $g : \mathbb{R} \to \mathbb{R}$ and $M = C(g) = \{0\}$. Since M is fixed, we will omit M in the following notations. It is easy to modify the mappings in such a way that also the cases $Q \subset \mathbb{R}$ and $\{0\} \subset C(g)$ are reflected. According to Theorem 2.9 it is also no problem to produce examples with $g : \mathbb{R} \to \mathbb{P}(\mathbb{R})$. Starting with $g_1 \, , \, g_2 : \mathbb{R} \to \mathbb{R}$ and $g_1 \, , \, g_2 \in \mathbb{F}_{(\star)}$ the derived mapping $co\,(g_1 \cup g_2)$ is also in $\mathbb{F}_{(\star)}$.

Examples 3.4 1. First we consider the smooth and unbounded function

$$g(y) = y \, \sin \, y \, .$$

The estimate $|g(y)| \leq |y|$ shows $g \in \mathbb{F}$. But in view of

$$F(g) = \{y \in \mathbb{R} : \sin\, y = 1\} \cup \{0\}$$

we have $g \notin \mathbb{F}_r$. By the way, the set

$$D(P_g) = \{y \in \mathbb{R} : \sin\, y = -1\}$$

of critical points (see Definition 2.6) is infinite, too. Since g is not regularly Fejer monotone, g can not be nonexpansive. Naturally, this shows also the derivative $g'(y) = y\, \cos\, y + \sin\, y$, which increases absolutely over all limits if $|y|$ tends to infinity in such a way that the zeros of $\cos\, y$ are avoided.

2. By a slight modification we get the continuous, but not everywhere differentiable function

$$g(y) = -y\,|\sin\, y|\ .$$

Again we have $|g(y)| \leq |y|$ such that $g \in \mathbb{F}$. But because of $F(g) = \{0\}$ it is even $g \in \mathbb{F}_r$, while $g \notin \mathbb{F}_<$ in view of

$$D(P_g) = \{y \in \mathbb{R} : |\sin\, y| = 1\} \neq \emptyset\ .$$

Moreover we have $|g'(y)| = |y\, \cos\, y + \sin\, y|$ for all y with $\sin\, y \neq 0$. Consequently g' is unbounded and g is not nonexpansive.

3. The first function with a reducing scalar factor supplies

$$g(y) = \gamma\, y\, \sin\, y \quad , \quad 0 < |\gamma| < 1\ .$$

Here we get $|g(y)| \leq |\gamma|\,|y| < |y|$ for $y \neq 0$. Hence $g \in \mathbb{F}_<$. In this case $|y| - |g(y)|$ increases over all limits if $|y|$ tends to infinity. This follows from

$$|y| - |g(y)| \geq |y| - |\gamma\, y| \geq (1 - |\gamma|)|y|\ .$$

Nevertheless g is unbounded. It can be shown that g is strongly Fejer monotone ($g \in \mathbb{F}^\alpha$ with $\alpha = \frac{1-|\gamma|}{1+|\gamma|}$, see [7]). But g is again not nonexpansive (compare with the first example above).

Examples 3.5 1. First we define g for nonnegative arguments by

$$g(y) \;=\; \begin{cases} \frac{y}{2} & \text{if} \quad 0 \leq y \leq 2 \\ \frac{y}{2} + \frac{1}{2}\sqrt{y^2 - 4} & \text{if} \quad y > 2 \end{cases}\ .$$

Then we extend g by $g(-y) = -g(y)$ to the negative domain. Here we obtain $|g(y)| < |\frac{y}{2} + \frac{1}{2}\sqrt{y^2}| = |y|$ for $y > 2$ and therefore $|g(y)| < |y|$ for all y such that $g \in \mathbb{F}_<$ holds. By the way, g is even 3-strongly Fejer monotone (see [7]). Further we have $\lim_{|y|\to\infty} (\, y - g(y)\,) = 0$. Because of

$$\lim_{y\to 2+0} g'(y) = \lim_{y\to 2+0} \frac{1}{2}(1 + \frac{y}{\sqrt{y^2-4}}) = +\infty$$

the function g is far away from being nonexpansive. We get similar examples by choosing $-g$ or by extending g as an even function to the negative domain. But observe that the modified functions remain strictly Fejer monotone while they lose the property of strong Fejer monotony.

2. The smooth and bounded function

$$g(y) = \sin\, y^2$$

given already in [1, p. 309] satisfies again $|g(y)| < |y|$ for $y \neq 0$. Thus $g \in \mathbb{F}_<$ holds. Moreover, we have even $g \in \mathbb{F}_s$ (see [7]). On the other hand the derivative $g'(y) = 2y \cos y^2$ is unbounded. Hence g is not nonexpansive.

3. The simple example $g(y) = -y$ shows that a regularly nonexpansive function has not to be strictly Fejer monotone. Finally it is easy to check that the continuous and piecewise linear function

$$g(y) \quad = \quad \begin{cases} n & \text{if} \quad y \in [2n, 2n+1) \\ y - n - 1 & \text{if} \quad y \in [2n+1, 2n+2) \end{cases} \quad , \quad n \in \mathbb{N} \quad ,$$

is strictly nonexpansive and consequently strictly Fejer monotone if we use the extension $g(-y) = -g(y)$ to the negative domain. Moreover, g is obviously 1-strongly Fejer monotone (see [7]). But observe that the opposite function $-g$ is not strictly nonexpansive, although it is regularly nonexpansive and α-strongly Fejer monotone with a certain $\alpha < 1$ (see also [7]).

Example 3.6 We study the function

$$g(y) \quad = \quad \begin{cases} y & \text{if} \quad -1 \le y \le 1 \\ -1 & \text{if} \quad y < -1 \\ 1 & \text{if} \quad y > 1 \end{cases} \quad .$$

On the one hand g is $\{0\}$-Fejer monotone. On the other hand we have $C(g) = F(g) = [-1, 1]$. It is easy to check that g is even 1-strongly $[-1, 1]$-Fejer monotone (see [7]). By Remark 2.7.2 we obtain $g \subseteq G_{[-1,1]} \subseteq G_{\{0\}}$. Moreover we can state

$$g(y) \subset G_{[-1,1]}(y) \quad \text{for} \quad |y| > 1 \quad , \quad G_{[-1,1]}(y) \subset G_{\{0\}}(y) \quad \text{for} \quad y \neq 0$$

in this case.

References

[1] **L. Elsner, I. Koltracht**, and **M. Neumann :** *Convergence of sequential and asynchronous nonlinear paracontractions.* Numer. Math., **62**, 305-319 (1992)

[2] **I.I. Eremin**, and **V.D. Mazurov :** *Nestacionarnye Processy Programmirovanija.* Nauka, Moskva 1979

[3] **H. Jeggle :** *Nichtlineare Funktionalanalysis.* Stuttgart 1979

[4] **L.W. Kantorowitsch**, and **G.P. Akilow :** *Funktionalanalysis in normierten Räumen.* Berlin 1964

[5] **D. Schott :** *Iterative solution of convex problems by Fejer monotone methods.* Numer. Funct. Anal. Optimiz. 1323-1357 (1995)

[6] **D. Schott :** *Basic properties of Fejer monotone sequences.* Rostock. Math. Kolloq. **49**, 57-74 (1995)

[7] **D. Schott :** *Basic properties of uniformly Fejer monotone sequences.* (submitted to J. Comput. Appl. Math.)

[8] **E. Zeidler :** *Vorlesungen über nichtlineare Funktionalanalysis I.* Leipzig 1976

received: March 8, 1996

Author:

Dieter Schott
Neubrandenburger Str. 49a
18055 Rostock
Germany

Rostock. Math. Kolloq. **50**, 85–88 (1997)

Subject Classification (AMS)
54H25

Zeqing Liu

Order completeness and stationary points

ABSTRACT. We give characterizations of order completeness by using stationary point theorems.

KEY WORDS. Order completeness, stationary point, ordered metric space.

1 Introduction

The notion of order complete metric spaces was introduced in [2]. It is worth emphasizing that the notion of order complete metric spaces is a non trivial extension of the notion of complete metric spaces. Conserva and Rizzo [1] characterized a class of order complete metric spaces as those ones in which every map of a suitable family has at least one fixed point. Turinici [2] obtained a necessary condition of order completeness.

In this paper we prove that the converse of Turinici′s result is true and obtain several criteria of order completeness.

Throughout this paper, let (X, d) be a metric space, $\leq$ an order (that is, a reflexive, antisymmetric and transitive relation) on X. N denotes the set of all positive integers. A sequence $\{x_n\}_{n\in N}$ in X is said to be $\leq$-monotone if and only if $x_n \leq x_m$ for all n, m in N with $n \leq m$. (X, d) is said to $\leq$-complete if and only if every $\leq$-monotone Cauchy sequence in$X$ is convergent. For $A \subset X$, let $\delta(A)$ and $\overline{A}$ denote the diameter and closure of A, respectively; $\alpha(A) = \inf\{\varepsilon > 0$: there exists a finite convering of A with sets having a diameter less than $\varepsilon\}$ if $\delta(A) < \infty$ and $\alpha(A) = \infty$ if $\delta(A) = \infty$. $CL(X)$ denotes the family of all nonempty closed subsets of X; $BC(X)$ denotes the family of all nonempty bounded closed subsets of X; 2^X denotes the family of all nonempty subsets of X. Define an order $>$ on 2^X with the aid of the above order $\leq$ on X as follows: for A, B in 2^X, we write $A > B$ if and only if $A \supset B$ and, for every a in A there exits a point b in B with $a \leq b$.

2 Characterizations of order completeness

Our results are as follows:

Theorem 1 *For any metric space (X,d) and any order $\leq$ on X the following statements are equivalent:*

(1) *(X,d) is $\leq$-complete;*

(2) *If $\{F_n\}_{n\in N}$ is a $>$-monotone sequence in $BC(X)$ satisfying $\alpha(F_n)\to 0$ as $n\to\infty$, then $\bigcap_{n\in N} F_n \neq \Phi$;*

(3) *If $\{F_n\}_{n\in N}$ is a $>$-monotone sequence in $BC(X)$ satisfying $\delta(F_n)\to 0$ as $n\to\infty$, then $\bigcap_{n\in N} F_n \neq \Phi$.*

Proof: (1) $\Rightarrow$ (2) is given in [2].

(2) $\Rightarrow$ (3) Note that $\alpha(F_n)\leq\delta(F_n)$. Then (3) follows immediately from (2).

(3) $\Rightarrow$ (1) Assume that $\{x_n\}_{n\in N}$ is a $\leq$-monotone Cauchy sequence in X. Define

$A_n=\overline{\{x_k : k\geq n\}}$ for all n in N. Then $\{A_n\}_{n\in N}$ is a $>$-monotone sequence and A_n is in $BC(X)$ for each n in N. It is evident that $\delta(A_n)=\delta\left(\{x_k : k\geq n\}\right)\to 0$ as $n\to\infty$. It follows from (3) that $\bigcap_{n\in N} A_n\neq\Phi$. Let w be in $\bigcap_{n\in N} A_n$. Then $d(x_n,w)\leq\delta(A_n)\to 0$ as $n\to\infty$; i.e., (X,d) is $\leq$-complete.

This completes the proof.

Theorem 2 *For any metric space (X,d) and any order $\leq$ on X, (1) is equivalent to each of the following:*

(4) *If $\{x_n\}_{n\in N}$ is a $\leq$-monotone sequence in X satisfying $\alpha\left(\overline{\{x_k : k\geq n\}}\right)\to 0$ as $n\to\infty$, then $\bigcap_{n\in N}\overline{\{x_k : k\geq n\}}\neq\Phi$;*

(5) *If $\{F_n\}_{n\in N}$ is a sequence in $CL(X)$ such that $\alpha(F_n)\to 0$ as $n\to\infty$, and that $F_n\supset F_{n+1}$ for each n in N and that there exists a $\leq$-monotone sequence $\{x_n\}_{n\in N}$ in X with $x_n\in F_n$ for all n in N, then $\bigcap_{n\in N} F_n\neq\Phi$;*

(6) *If $\{F_n\}_{n\in N}$ is a sequence in $CL(X)$ such that $\delta(F_n)\to 0$ as $n\to\infty$, and that $F_n\supset F_{n+1}$ for each n in N and that there exists a $\leq$-monotone sequence $\{x_n\}_{n\in N}$ in X with $x_n\in F_n$ for all n in N, then $\bigcap_{n\in N} F_n\neq\Phi$.*

(7) *If f is a map of X into $CL(X)$ satisfying*

(i) *$fy\subset fx$ for all x in X and all y in fx;*

(ii) *there exists a $\leq$-monotone sequence $\{x_n\}_{n\in N}$ in X such that x_{n+1} is in fx_n for each n in N and that $\delta(fx_n) \to 0$ as $n \to \infty$,*

then f has a stationary point w in X; i.e., $fw = \{w\}$;

(8) *If f is a map of X into $BC(X)$ satisfying* (i) *and* (ii), *then f has a stationary point.*

Proof: (1)$\Rightarrow$(4) From the proof of Theorem 4.1 in [2] it is easy to see that (1) implies (4).

(4)$\Rightarrow$(5) Since $\overline{\{x_k : k \geq n\}} \subset F_n$ for each n in N, it follows that $\alpha\left(\overline{\{x_k : k \geq n\}}\right) \leq \alpha(F_n) \to 0$ as $n \to \infty$.

By (4) we have $\Phi \neq \bigcap_{n\in N} \overline{\{x_k : k \geq n\}} \subset \bigcap_{n\in N} F_n$.

(5)$\Rightarrow$(6) Since $\alpha(F_n) \leq \delta(F_n)$, it follows that (5) implies (6).

(6)$\Rightarrow$(7) Take $F_n = fx_n$ for all n in N. By (6) we conclude that $A = \bigcap_{n\in N} F_n \neq \Phi$. Note that $\delta(A) \leq \delta(F_n) \to 0$ as $n \to \infty$. Consequently $\delta(A) = 0$; i.e., $A = \{w\}$ for some w in X. (i) ensures that $fw \subset fx_n$ for all n in N. Hence $fw \subset \bigcap_{n\in N} fx_n = A = \{w\}$. It is clear that $fw = \{w\}$.

(7)$\Rightarrow$(8) Note that $BC(X) \subset CL(X)$. Hence (8) follows from (7).

(8)$\Rightarrow$(1) Suppose that (X, d) is not $\leq$-complete. Then there exists a $\leq$-monotone Cauchy sequence $\{x_n\}_{n\in N}$ in X such that $\{x_n\}_{n\in N}$ is not convergent. Without loss of generality we may assume that $x_n \neq x_m$ for all distinct n, m in N. Take $A_0 = X$ and $A_n = \{x_k : k \geq n\}$ for each n in N. It is easy to verify that A_n is in $BC(X)$ for all n in N. Define a function $i : X \to N \bigcup \{0\}$ by $i(x) = n$ if $x \in A_n$ and $x \overline{\in} A_{n+1}$. Now we construct a map f of X into $BC(X)$ by $fx = A_{i(x)+1}$ for each x in X. It is clear that f has no stationary point. For any x, y in X with $y \in fx = A_{i(x)+1}$ we have $i(y) \geq i(x) + 1$ by the definition of i. This means that $fy = A_{i(y)+1} \subset A_{i(x)+1} = fx$; i.e., f satisfies (i). Note that $x_n \in A_n$ and $x_n \overline{\in} A_{n+1}$ for every n in N. It follows that $i(x_n) = n$ for all n in N. Consequently $x_{n+1} \in A_{n+1} = A_{i(x_n)+1} = fx_n$ for each n in N and $\delta(fx_n) = \delta(A_{n+1}) \to 0$ as $n \to \infty$; i.e., f satisfies (ii). By (8) f has a stationary point. This is a contradiction.

This completes the proof.

References

[1] **Conserva, V.**, and **Rizzo, S.** : *Fixed points and completeness.* Math. Japonica **38**, 901 - 903 (1993)

[2] **Turnici, M. :** *Maximal elements in a class of order complete metric spaces.* Math. Japonica **25**, 511 - 517 (1980)

received: March 19, 1996

Author:

Zeqing Liu
Department of Mathematics
Liaoning Normal University
Dalian
Liaoning 116029 P. R. China

Rostock. Math. Kolloq. **50**, 89–94 (1997)

Subject Classification (AMS)
47H09

Manfred Krüppel

On the nearest point projection in Hilbert spaces with application to Nonlinear Ergodic Theory

ABSTRACT. Let P be the nearest point projection of a Hilbert space H onto a closed and convex subset K and $\Sigma\lambda_i x_i$ any convex combination of the points x_i in H. We give an estimate for $||P(\Sigma\lambda_i x_i) - \Sigma\lambda_i P x_i||$. A consequence of this estimate is the following: Let T be a nonexpansive selfmapping of a bounded, closed and convex subset C of H, let $F(T)$ denote the fixed point set of T which is nonempty, closed and convex, and for $x \in C$ let be $S_n x = (x + Tx + \ldots + T^{n-1}x)/n$. Then the sequence $\{\mathrm{Proj}_{F(T)} S_n x\}$ converges to a fixed point p, which is the weak limit of the sequence $\{S_n x\}$ too.

Keywords and phrases: Nearest point projection, nonexpansive mapping, inequalities, Nonlinear Ergodic Theorem, Hilbert space

Let K be a nonempty, closed and convex subset of a (real) Hilbert space H. Then it is well known that for any x in H the nearest point projection $p = \mathrm{Proj}_K x$ is well defined as the unique point of K which is nearest to x, i.e. $||x - p|| \le ||x - z||$ for all z in K (cf. [4]). Further it is known that the nearest point projection in Hilbert space is nonexpansive. In this note we shall apply the identity

$$||\lambda x + (1-\lambda)y||^2 = \lambda||x||^2 + (1-\lambda)||y||^2 - \lambda(1-\lambda)||x-y||^2 \tag{1}$$

for all x, y in H and $0 \le \lambda \le 1$. For any n points $x_1, x_2, \ldots, x_n$ in H the following generalized identity holds

$$\left|\left|\sum_{i=1}^{n} \lambda_i x_i\right|\right|^2 = \sum_{i=1}^{n} \lambda_i ||x_i||^2 - \sum_{i<j} \lambda_i \lambda_j ||x_i - x_j||^2 \tag{2}$$

where $\lambda_i \ge 0$ and $\Sigma\lambda_i = 1$.

Lemma 1 *If P is the nearest point projection of H onto the closed convex subset K, then holds it for all z in K*

$$||x - Px||^2 + ||z - Px||^2 \le ||x - z||^2.$$

Proof: Let z be any point in K. Since K is convex the point $[\lambda z + (1-\lambda)Px]$ lies in K too. By (1), we have

$$||x - [\lambda z + (1-\lambda)Px]||^2 = \lambda||x-z||^2 + (1-\lambda)||x-Px||^2 - \lambda(1-\lambda)||z-Px||^2.$$

Because of $||x - Px|| \leq ||x - [\lambda z + (1-\lambda)Px]||$ it follows

$$\lambda||x-Px||^2 \leq \lambda||x-z||^2 - \lambda(1-\lambda)||z-Px||^2.$$

Dividing by λ and let tends $\lambda \to 0$ the lemma is proved.

Theorem 2 *If P is the nearest point projection of the Hilbert space H onto the closed convex subset K, then it holds for any convex combination of points $x_1, x_2, \ldots, x_n$ in H*

$$\left|\left|\sum_{i=1}^n \lambda_i Px_i - P\left(\sum_{i=1}^n \lambda_i x_i\right)\right|\right|^2 \leq \sum_{i<j} \lambda_i\lambda_j \langle Px_i - Px_j,\, (I-P)x_i - (I-P)x_j\rangle.$$

Proof: Let be $x = \Sigma\lambda_i x_i$. By lemma 1, we have

$$||x-Px||^2 + ||\sum \lambda_i Px_i - Px||^2 \leq ||x - \sum \lambda_i Px_i||^2 = ||\sum \lambda_i(x_i - Px_i)||^2.$$

From (2) it follows

$$||\sum \lambda_i(x_i - Px_i)||^2 = \sum \lambda_i||x_i - Px_i||^2 - \sum_{i<j} \lambda_i\lambda_j||(x_i - Px_i) - (x_j - Px_j)||^2.$$

Again by lemma 1 it is

$$\sum \lambda_i||x_i - Px_i||^2 \leq \sum \lambda_i||x_i - Px||^2 - \sum \lambda_i||Px_i - Px||^2.$$

Because of the identity (2) we have

$$||x-Px||^2 = ||\sum \lambda_i(x_i - Px)||^2 = \sum \lambda_i||x_i - Px||^2 - \sum_{i<j} \lambda_i\lambda_j||x_i - x_j||^2$$

and

$$||\sum \lambda_i Px_i - Px||^2 = ||\sum \lambda_i(Px_i - Px)||^2 = \sum \lambda_i||Px_i - Px||^2 - \sum_{i<j} \lambda_i\lambda_j||Px_i - Px_j||^2.$$

Altogether, we obtain

$$2||\sum \lambda_i Px_i - Px||^2 \leq \sum_{i<j} \lambda_i\lambda_j(||x_i - x_j||^2 - ||Px_i - Px_j||^2 - ||x_i - x_j - (Px_i - Px_j)||^2).$$

By definition of the inner product the theorem is proved.

Corollary 3 *If P is a nearest point projection in a Hilbert space onto a closed and convex subset K, then for any two points x and y in H it holds*

$$\langle Px - Py, (x - y) - (Px - Py)\rangle \geq 0.$$

From this one can conclude several properties of a nearest point projection P in a Hilbert space, e.g. P is nonexpansive (cf. [5], [6]). Since P is nonexpansive, it is by the inequality of Schwarz

$$\begin{aligned}\langle Px - Py, (x - y) - (Px - Py)\rangle &= \langle Px - Py, x - y\rangle - ||Px - Py||^2 \\ &\leq ||Px - Py||\,||x - y|| - ||Px - Py||^2 \\ &\leq ||x - y||^2 - ||Px - Py||^2.\end{aligned}$$

So from theorem 2 the inequality of Zarantonello

$$\left|\left|\sum_{i=1}^{n} \lambda_i Px_i - P\left(\sum_{i=1}^{n} \lambda_i x_i\right)\right|\right|^2 \leq \sum_{i<j} \lambda_i\lambda_j(||x_i - x_j||^2 - ||Px_i - Px_j||^2)$$

follows which is valid for each nonexpansive mapping in Hilbert space.

Corollary 4 *Let be P the nearest point projection in a Hilbert space onto a closed and convex subset K, then for any convex combination of points $x_1, x_2, \dots, x_n$ in H and any p in K*

$$||P(\sum \lambda_i x_i) - \sum \lambda_i Px_i||^2 \leq d \sum_{i=1}^{n} \lambda_i ||Px_i - p||$$

where

$$d = \sup_{i,j}(||x_i - x_j|| - ||Px_i - Px_j||).$$

Proof: By theorem 2 we get

$$\begin{aligned}||P(\sum \lambda_i x_i) - \sum \lambda_i Px_i||^2 &\leq \sum_{i<j} \lambda_i\lambda_j(\langle Px_i - Px_j, x_i - x_j\rangle - ||Px_i - Px_j||^2) \\ &\leq \sum_{i<j} \lambda_i\lambda_j ||Px_i - Px_j||\,(||x_i - x_j|| - ||Px_i - Px_j||) \\ &\leq d \sum_{i<j} \lambda_i\lambda_j \|Px_i - Px_j\| \\ &\leq d \sum_{i<j} \lambda_i\lambda_j (||Px_i - p|| + ||Px_j - p||) \\ &\leq d \sum_{i<j} \lambda_i ||Px_i - p||.\end{aligned}$$

Corollary 5 *The nearest point projection P in a Hilbert space onto a closed and convex subset K is demiclosed,i.e. if $x_n \rightharpoonup x$ (weakly) and $Px_n \to p$ then $Px = p$.*

Proof: To given $\varepsilon > 0$ there exists an integer n_0 with $||Px_n - p|| \leq \varepsilon$ for $n \geq n_0$. Because of a theorem of Mazur there exists a finite convex combination of the points $x_n (n = n_0, \dots, m)$ so that

$$\left|\left|\sum_{i=n_0}^{m} \lambda_i x_i - x\right|\right| < \varepsilon.$$

Since P is nonexpansive, we obtain by corollary 4

$$\begin{aligned} ||Px - p|| &\leq \left|\left|Px - P\left(\sum_{i=n_0}^{m} \lambda_i x_i\right)\right|\right| + \left|\left|P\left(\sum_{i=n_0}^{m} \lambda_i x_i\right) - \sum_{i=n_0}^{m} \lambda_i Px_i\right|\right| + \left|\left|\sum_{i=n_0}^{m} \lambda_i Px_i - p\right|\right| \\ &\leq \left|\left|x - \sum_{i=n_0}^{m} \lambda_i x_i\right|\right| + \sqrt{d \sum_{i=n_0}^{m} \lambda_i ||Px_i - p||} + \sum_{i=n_0}^{m} \lambda_i ||Px_i - p|| \\ &\leq 2\varepsilon + \sqrt{d\varepsilon}. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the corollary is proved.

Corollary 6 *Let P be the nearest point projection in a Hilbert space onto a closed and convex subset K, let (x_n) be a bounded sequence with $Px_n \to p$. For any regular matrix (a_{nk}), i.e.*

$$a_{n,k} \geq 0, \qquad \sum_{k=0}^{\infty} a_{n,k} = 1, \qquad \lim_{n\to\infty} a_{n,k} = 0 \tag{3}$$

let be

$$y_n = \sum_{k=0}^{\infty} a_{n,k} x_k$$

Then $Py_n \to p$.

Proof: By corollary 4, we have

$$\begin{aligned} ||Py_n - p|| &\leq ||P(\sum a_{n,k} x_k) - \sum a_{n,k} Px_k|| + ||\sum a_{n,k} Px_k - p|| \\ &\leq \sqrt{d \sum a_{n,k} ||Px_k - p||} + \sum a_{n,k} ||Px_k - p||. \end{aligned}$$

Now the conclusion follows by a limit theorem due to Toeplitz (cf. [6]).

Remark: Theorem 2 is not valid in an arbitrary Banach space X, as the following example shows: Let be $X = l_3^p, x_1 = (1,0,0), x_2 = (0,1,0)$ and $K = \{x = t(1,1,1) : |t| \leq 1\}$.

Let P be the nearest point projection of X onto K, then $Px_1 = Px_2 = y$ where $y = (2^{1/p-1}+1)^{-1}(1,1,1)$ but $P((x_1+x_2)/2) = 2^{-1+1/(p-1)}y = y$ if and only if $p = 2$.

Finally we give an application for the ergodic theory of nonexpansive mappings in Hilbert space. Let C be a nonempty bounded closed and convex subset of a Hilbert space H and $T : C \to C$ a nonexpansive mapping, i.e. $||Tx - Ty|| \leq ||x - y||$ for all x, y in C. In 1975 Baillon [1] has proved the first ergodic theorem for nonexpansive mappings, namely that for each x in C the sequence of CESARO-means

$$S_n x = \frac{x + Tx + \ldots + T^{n-1}x}{n}$$

converges weakly for $n \to \infty$ to a fixed point of T. Brezis and Browder [2] generalized Baillons result to the more general summation method

$$A_n x = \sum_{k=0}^{\infty} a_{n,k} T^k x,$$

where $(a_{n,k})$ is any strongly regular matrix, i.e. it holds (3) and additionally

$$s_n = \sum_{k=0}^{\infty} |a_{n,k+1} - a_{n,k}| \to 0 \qquad \text{for } n \to \infty.$$

Theorem 7 *Let C be a bounded, closed, convex subset of a Hilbert space H, let $T : C \to C$ be a nonexpansive mapping and for $x \in C$ let $A_n x$ be a summation method with any strongly regular matrix. Then the fixed point set $F(T)$ is nonempty, closed, convex and there exists a fixed point p in $F(T)$ such that holds:*

(i) $Proj_{F(T)} T^n x$ converges to p.

(ii) $Proj_{F(T)} A_n x$ converges to p.

(iii) $A_n x$ converges weakly to p.

Proof: 1. That the fixed point set $F(T)$ is nonempty, closed and convex is firstly proved by Browder [3].
2. Point (i) is proved by Baillon [1] and Pazy [7].
3. By point (i) and corollary 6 it follows (ii).
4. Point (iii) is the mean ergodic theorem in the form of Brezis and Browder [2].

References

[1] **Baillon, J. B. :** *Un théorème de type ergodic pour les contractions nonlinéares dans un espace de Hilbert.* C.R.Acad.Sci. Paris **280**, 1511–1514 (1975)

[2] **Brezis, H.**, and **Browder, F. E. :** *Nonlinear Ergodic Theorems.* Amer. Math. Soc. vol. **82**, no. 6, 959–961 (1976)

[3] **Browder, F. E. :** *Fixed point theorems for noncompact mappings in Hilbert space.* Proc. Nat. Acad. Sci. U.S.A. **53**, 1272–1276 (1966)

[4] **Goebel, K.**, and **Kirk, W. A. :** *Topics in metric fixed point theory.* Cambridge University press (1990)

[5] **Goebel, K.**, and **Reich, S. :** *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings.* New York 1984

[6] **Fichtenholz G. M. :** *Differential- und Integralrechnung II.* p.337, Berlin 1964

[7] **Pazy, A. :** *On the asymptotic behaviour of iterates of nonexpansive mappings in Hilbert space.* Israel J.Math. **426**, no. 2, 197–204 (1977)

received: June 17, 1996

Author:

Manfred Krüppel
Universität Rostock
Fachbereich Mathematik
Universitätsplatz 1
18051 Rostock
Germany