STUDIA

UNIVERSITATIS BABEŞ-BOLYAI

Mathematica

CLUJ-NAPOCA 2006

Cluj University Press

4

Editorial Board of Studia Universitatis Babeş-Bolyai, Mathematica

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

ON A CLASS OF LINEAR POSITIVE BIVARIATE OPERATORS OF KING TYPE

OCTAVIAN AGRATINI

Dedicated to Professor Gheorghe Coman at his 70^{th} *anniversary*

Abstract. The concern of this note is to introduce a general class of linear positive operators of discrete type acting on the space of real valued functions defined on a plane domain. These operators preserve some test functions of Bohman-Korovkin theorem. Following our technique, as a particular class, a modified variant of the bivariate Bernstein-Chlodovsky operators is presented.

1. Introduction

Let $(L_n)_{n\geq 1}$ be a sequence of positive linear operators defined on the Banach space $C([a,b])$. A classical theorem of Bohman-Korovkin asserts: if $(L_n e_k)_{n\geq 1}$ converges to e_k uniformly on $[a,b]$, $k \in \{0,1,2\}$, for the test functions $e_0(x) = 1$, $e_1(x) = x$, $e_2(x) = x^2$, then $(L_n f)_{n\geq 1}$ converges to f uniformly on $[a,b]$, for each $f \in C([a,b])$.

J.P. King [8] has presented an example of linear and positive operators $V_n : C([0,1]) \to C([0,1])$, given as follows

$$(V_n f)(x) = \sum_{k=0}^{n} \binom{n}{k} (r_n^*(x))^k (1 - r_n^*(x))^{n-k} f\left(\frac{k}{n}\right), \ f \in C([0,1]), \ x \in [0,1], \quad (1)$$

Received by the editors: 01.04.2006.

2000 *Mathematics Subject Classification.* 41A36, 41A25.

Key words and phrases. linear positive operator, Bohman-Korovkin theorem, bivariate modulus of smoothness, Bernstein-Chlodovsky operator.

where $r_n^* : [0,1] \to [0,1]$,

$$r_n^*(x) = \begin{cases} x^2, & n = 1, \\ -\dfrac{1}{2(n-1)} + \sqrt{\dfrac{n}{n-1}x^2 + \dfrac{1}{4(n-1)^2}}, & n = 2, 3, \ldots \end{cases} \tag{2}$$

This sequence preserves two test functions e_0, e_2 and $(V_n e_1)(x) = r_n^*(x)$ holds. Based on Bohman-Korovkin criterion, we get $\lim\limits_{n\to\infty} (V_n f)(x) = f(x)$ for each f belonging to $C([0,1])$, $x \in [0,1]$.

Further results regarding V_n operator have been recently obtained by Gonska and Piţul [5]. Also, by using A-statistical convergence, an analog of King's result has been proved by O. Duman and C. Orhan [4].

In [1] we indicated a general technique to construct sequences of univariate operators of discrete type with the same property as in King's example, i.e., their degree of exactness is null, but they reproduce the third test function of the celebrated criterion.

The central issue of this paper is to present a sequence of bivariate operators with similar properties: to reproduce certain monomials of second degree and to form an approximation process.

2. **Preliminaries**

Following our announced aim, in this section we recall results regarding the univariate case. Also, basic results concerning the uniform approximation of functions by bivariate operators are delivered.

We set $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Following [1], we consider a sequence $(L_n)_{n\geq 1}$ of linear positive operators of discrete type acting on a subspace of $C(\mathbb{R}_+)$ and defined by

$$(L_n f)(x) = \sum_{k=0}^{\infty} u_{n,k}(x) f(x_{n,k}), \ x \geq 0, \ f \in \mathcal{F} \cap E_\alpha, \tag{3}$$

where $u_{n,k} : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous ($n \in \mathbb{N}, \ k \in \mathbb{N}_0$), $(x_{n,k})_{k\geq 0} := \Delta_n$ is a net on $\mathbb{R}_+$ and

$$\mathcal{F} := \{f : \mathbb{R}_+ \to \mathbb{R}_+ : \text{ the series in (3) is convergent}\},$$

$$E_\alpha := \{f \in C(\mathbb{R}_+) : \ (1+x^\alpha)^{-1} f(x) \text{ is convergent as } x \to \infty\},$$

$\alpha \geq 2$ being fixed. We mention that the right-hand side of (3) could be a finite sum. We assume that the following identities

$$(L_n e_0)(x) = 1, \quad (L_n e_1)(x) = x, \quad (L_n e_2)(x) = a_n x^2 + b_n x + c_n, \quad x \geq 0, \tag{4}$$

are fulfilled for each $n \in \mathbb{N}$. At this moment, $\{e_0, e_1, e_2\} \subset \mathcal{F} \cap E_\alpha$ holds. Moreover, we assume

$$a_n \neq 0, \ n \in \mathbb{N}, \quad \lim_{n\to\infty} a_n = 1, \quad \lim_{n\to\infty} b_n = \lim_{n\to\infty} c_n = 0.$$

Based on Bohman-Korovkin theorem these relations guarantee that $(L_n)_{n\geq 1}$ is a positive approximation process, more precisely $\lim_{n\to\infty} (L_n f)(x) = f(x)$ uniformly for every $f \in \mathcal{F} \cap E_\alpha$ and every x belonging to any compact $\mathcal{K} \subset \mathbb{R}_+$.

Since $(Le_1)^2 \leq (Le_0)(Le_2)$ is a common property of any linear positive operator L of summation type, we get

$$(a_n - 1)x^2 + b_n x + c_n \geq 0, \quad x \geq 0, \ n \in \mathbb{N}, \tag{5}$$

which implies

$$c_n \geq 0, \quad a_n \geq 1 \text{ for each } n \in \mathbb{N}, \tag{6}$$

and $\{n \in \mathbb{N} : \ a_n = 1\} \subset \{n \in \mathbb{N} : \ b_n \geq 0\}$. Further on, we are looking for the functions $v_n \in \mathbb{R}_+^{\mathbb{R}_+}$, $n \in \mathbb{N}$, such that $(L_n e_2)(v_n(x)) = x^2$ for each $x \geq 0$ and $n \in \mathbb{N}$, this means

$$a_n v_n^2(x) + b_n v_n(x) + c_n - x^2 = 0, \quad x \geq 0, \ n \in \mathbb{N}. \tag{7}$$

In what follows, throughout the paper, we take

$$c_n = 0, \quad n \in \mathbb{N}, \tag{8}$$

and

$$v_n(x) = \frac{1}{2a_n}(\sqrt{b_n^2 + 4a_n x^2} - b_n), \quad x \geq 0, \ n \in \mathbb{N}. \tag{9}$$

For each $n \in \mathbb{N}$, $v_n(x)$ is well defined and v_n is a continuous positive function. Also, relation (7) is verified.

Starting from (3) we define the univariate linear positive operators

$$(L_n^*f)(x)=\sum_{k=0}^{\infty}u_{n,k}(v_n(x))f(x_{n,k}),\quad x\geq 0,\ f\in\mathcal{F}\cap E_\alpha,\ n\in\mathbb{N},\tag{10}$$

where v_n is given by (9).

The following identities

$$L_n^*e_0=e_0,\quad L_n^*e_1=v_n,\quad L_n^*e_2=e_2\tag{11}$$

hold. Consequently, one has $\lim\limits_{n\to\infty}L_n^*f=f$ uniformly on compact intervals of $\mathbb{R}_+$ for every $f\in\mathcal{F}\cap E_\alpha$. This result follows from (11) and Korovkin criterion. For each n with the property $b_n\geq 0$ we get $v_n(0)=0$ and, consequently, one has $(L_n^*f)(0)=(L_nf)(0)$.

Setting $e_{i,j}(x,y)=x^iy^j$, $i\in\mathbb{N}_0$, $j\in\mathbb{N}_0$, $i+j\leq 2$, the test functions corresponding to the bidimensional case, we need a result due to Volkov [10].

Theorem 1. *Let I and J compact intervals of the real line. Let L_{m_1,m_2}, $(m_1,m_2)\in\mathbb{N}\times\mathbb{N}$, be linear positive operators applying the space $C(I\times J)$ into itself. If*

$$\lim_{m_1,m_2}L_{m_1,m_2}e_{i,j}=e_{i,j},\quad (i,j)\in\{(0,0),(1,0),(0,1)\},$$

$$\lim_{m_1,m_2}L_{m_1,m_2}(e_{2,0}+e_{0,2})=e_{2,0}+e_{0,2},$$

uniformly on $I\times J$, then the sequence $(L_{m_1,m_2}f)$ converges to f uniformly on $I\times J$ for any $f\in C(I\times J)$.

In a more general frame, Volkov's theorem says: if X is a compact subset of the Euclidean space $\mathbb{R}^p$, then $\left\{\mathbf{1},pr_1,\dots,pr_p,\sum\limits_{j=1}^{p}pr_j^2\right\}$ is a Korovkin subset in $C(X)$. Here $\mathbf{1}$ stands for the constant function on X of constant value 1 and $pr_1,\dots,pr_p$ represent the canonical projections on X, this means $pr_j(x):=x_j$ for every $x=(x_i)_{1\leq i\leq p}\in X$, where $1\leq j\leq p$. For a thorough documentation the monograph of Altomare and Campiti [2; *page 245*] can be consulted.

3. **A class of bivariate operators**

Now we are going to present the tensor product extension of L_n^* to the bidimensional case.

Starting from the specified Δ_n net on $\mathbb{R}_+$, we consider $\Delta_{m_1} \times \Delta_{m_2}$, the corresponding net on $\mathbb{R}_+ \times \mathbb{R}_+$. Thus, $(x_{m_1,i}, x_{m_2,j})$, $(i,j) \in \mathbb{N}_0 \times \mathbb{N}_0$, are its knots.

Having in mind the notations of the previous section we introduce the bivariate linear positive operators acting on $\mathcal{D}$ and defined as follows

$$(L^*_{m_1,m_2} f)(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} u_{m_1,i}(v_{m_1}(x)) u_{m_2,j}(v_{m_2}(y)) f(x_{m_1,i}, x_{m_2,j}), \tag{12}$$

$(x,y) \in \mathbb{R}_+ \times \mathbb{R}_+$. For each index $m \in \mathbb{N}$, the functions $u_{m,k}$, $k \in \mathbb{N}_0$, enjoy the properties implied by (4) and v_m is given by (9). In the above $\mathcal{D}$ consists of all continuous functions $f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ with the properties: the series in (12) is convergent and $(1 + x^{\alpha_1})^{-1} f(x,y)$, $(1 + y^{\alpha_2})^{-1} f(x,y)$ are convergent as $x \to \infty$, $y \to \infty$ respectively, where $\alpha_1 \geq 2$, $\alpha_2 \geq 2$ are fixed. Clearly, $e_{i,j} \in \mathcal{D}$ for each $(i,j) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $i + j \leq 2$.

Theorem 2. *Let $L^*_{m_1,m_2}$ be defined by (12).*

(i) The following identities

$$L^*_{m_1,m_2} e_{0,0} = e_{0,0},\ L^*_{m_1,m_2} e_{2,0} = e_{2,0},\ L^*_{m_1,m_2} e_{0,2} = e_{0,2}, \tag{13}$$
$$(L^*_{m_1,m_2} e_{1,0})(x,y) = v_{m_1}(x),\ (L^*_{m_1,m_2} e_{0,1})(x,y) = v_{m_2}(y),\ (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+,$$

hold.

(ii) One has $\lim_{m_1,m_2} L^*_{m_1,m_2} f = f$ uniformly on compact subsets of $\mathbb{R}^2_+$ for every $f \in \mathcal{D}$.

Proof. (i) Taking into account (11) and (12), by a straightforward calculation the stated identities follow.

(ii) Based on (13), the result is implied by (9) and Theorem 1. □

We can explore the rate of convergence of $L^*_{m_1,m_2}$ operators in terms of the first order modulus of smoothness ω_f of the bivariate function f. It is known that for any real valued bounded function f, $f \in B(I \times J)$, where I and J are compact

intervals of the real line, the associated mapping ω_f is defined as follows:

$$\omega_f(\delta_1,\delta_2)=\sup\{|f(x_1,y_1)-f(x_2,y_2)|:\ (x_1,y_1),(x_2,y_2)\in I\times J,\\ |x_1-y_1|\le\delta_1,\ |x_2-y_2|\le\delta_2\},\ (\delta_1,\delta_2)\in\mathbb{R}_+\times\mathbb{R}_+. \tag{14}$$

Among the properties of ω_f investigated by A.F. Ipatov [7] we recall

$$\omega_f(\lambda_1\delta_1,\lambda_2\delta_2)\le(1+\lambda_1+\lambda_2)\omega_f(\delta_1,\delta_2),\quad \lambda_1>0,\ \lambda_2>0. \tag{15}$$

Let $\mathcal{K}\subset\mathbb{R}_+^2$ be a compact and let $\delta_1>0$, $\delta_2>0$ be fixed. Based on (15) and knowing that $L^*_{m_1,m_2}e_{0,0}=1$, for each $(x,y)\in\mathcal{K}$ we can write

$$|(L^*_{m_1,m_2}f)(x,y)-f(x,y)|$$

$$\le\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}u_{m_1,i}(v_{m_1}(x))u_{m_2,j}(v_{m_2}(y))|f(x_{m_1,i},x_{m_2,j})-f(x,y)|$$

$$\le\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}u_{m_1,i}(v_{m_1}(x))u_{m_2,j}(v_{m_2}(y))\omega_f\left(\frac{1}{\delta_1}|x_{m_1,i}-x|,\frac{1}{\delta_2}|x_{m_2,j}-y|\right)$$

$$\le\left(1+\frac{1}{\delta_1}\sum_{i=0}^{\infty}u_{m_1,i}(v_{m_1}(x))|x_{m_1,i}-x|+\frac{1}{\delta_2}\sum_{j=0}^{\infty}u_{m_2,j}(v_{m_2}(y))|x_{m_2,j}-y|\right)\omega_f(\delta_1,\delta_2).$$

On the other hand, Cauchy's inequality and the identities given by (13) imply

$$\sum_{i=0}^{\infty}u_{m_1,i}(v_{m_1}(x))|x_{m_1,i}-x|$$

$$\le\left(\sum_{i=0}^{\infty}u_{m_1,i}(v_{m_1}(x))\right)^{1/2}\left(\sum_{i=0}^{\infty}u_{m_1,i}(v_{m_1}(x))(x_{m_1,i}-x)^2\right)^{1/2}$$

$$=(2x^2-2xv_{m_1}(x))^{1/2},$$

and respectively

$$\sum_{j=0}^{\infty}u_{m_2,j}(v_{m_2}(y))|x_{m_2,j}-y|\le(2y^2-2yv_{m_2}(y))^{1/2}.$$

The above relations enable us to state the following estimate for the pointwise approximation.

Theorem 3. *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}_+^2$. The operators L_{m_1,m_2}, $(m_1,m_2)\in\mathbb{N}\times\mathbb{N}$, defined by (12) verify*

$$|(L_{m_1,m_2}f)(x,y)-f(x,y)|\leq\left(1+\frac{1}{\delta_1}\widetilde{v}_{m_1}(x)+\frac{1}{\delta_2}\widetilde{v}_{m_2}(y)\right)\omega_f(\delta_1,\delta_2), \tag{16}$$

for every $f\in\mathcal{D}$, $(x,y)\in\mathcal{K}$, $\delta_1>0$, $\delta_2>0$, where

$$\widetilde{v}_m(t)=\sqrt{2t^2-2tv_m(t)},\quad m\in\mathbb{N},\ t\geq 0, \tag{17}$$

and v_m is given at (9).

Remarks. 1° Based on Cauchy's inequality $(L^*e_1)^2\leq(L^*e_0)(L^*e_2)$ and relation (11) as well, we get $t\geq v_m(t)$, for each $t\geq 0$. Consequently, in (17) $\widetilde{v}_m$ is well defined.

2° Endowing $\mathbb{R}\times\mathbb{R}$ with the metric ρ, $\rho(z_1,z_2)=|x_1-x_2|+|y_1-y_2|$ for $z_k=(x_k,y_k)$, $k=1,2$, we could have estimated the rate of convergence using another type of modulus of smoothness given by

$$\omega_1(f;\delta)=\sup\{|f(z_1)-f(z_2)|:\ z_1\in\mathcal{K},\ z_2\in\mathcal{K},\ \rho(z_1,z_2)\leq\delta\},$$

for every $f\in B(\mathcal{K})$ and $\delta>0$. Clearly, (14) implies $\omega_f(\delta_1,\delta_2)\leq\omega_1(f;\delta_1+\delta_2)$. An overview on moduli of smoothness as well as some of their extensions can be found, e.g., in the monograph [2; *Section 5.1*].

3° Examining the construction of v_m we easily deduce $v_m(0)\leq v_m(x)\leq x$, for each $x\in\mathbb{R}_+$. Moreover, the mapping $x\mapsto x-v_m(x)$ is increasing one. For a compact $I=[\alpha,\beta]\subset\mathbb{R}_+$, we can write

$$\widetilde{v}_m(t)\leq\sqrt{2\beta}\left(\max_{t\in I}(t-v_m(t))\right)^{1/2}=\sqrt{2\beta}\sqrt{\beta-v_m(\beta)}.$$

Consequently, if $\mathcal{K}:=I\times J=[\alpha_1,\beta_1]\times[\alpha_2,\beta_2]\subset\mathbb{R}_+^2$ then, by choosing in (16) $\delta_j:=\sqrt{\beta_j-v_{m_j}(\beta_j)}$, $j\in\{1,2\}$, we obtain the following global estimate on the compact $\mathcal{K}$

$$\|L_{m_1,m_2}f-f\|_{C(\mathcal{K})}\leq(1+\sqrt{2\beta_1}+\sqrt{2\beta_2})\omega_f\left(\sqrt{\beta_1-v_{m_1}(\beta_1)},\sqrt{\beta_2-v_{m_2}(\beta_2)}\right).$$

Here $\|\cdot\|_{C(\mathcal{K})}$ stands for the usual sup-norm of the space $C(\mathcal{K})$.

4. **Example**

In order to obtain an approximation process of $L^*_{m_1,m_2}$-type, we focus our attention on Bernstein-Chlodovsky operators. Let $(h_n)_{n\geq 1}$ be a sequence of strictly positive real numbers verifying

$$\lim_{n\to\infty} h_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{h_n}{n} = 0.$$

The nth Bernstein-Chlodovsky operator [3], $L_n : C(\mathbb{R}_+) \to C(\mathbb{R}_+)$ is defined by

$$(L_n f)(x) = \begin{cases} \displaystyle\sum_{k=0}^{n} \binom{n}{k} \left(\frac{x}{h_n}\right)^k \left(1-\frac{x}{h_n}\right)^{n-k} f\left(\frac{h_n k}{n}\right), & \text{if} \quad 0 \leq x \leq h_n, \\ f(x), & \text{if} \quad x > h_n. \end{cases} \tag{18}$$

It is known that identities (4) are fulfilled and we get

$$\begin{cases} a_n = 1 - \dfrac{1}{n},\ b_n = \dfrac{h_n}{n},\ c_n = 0, & \text{if} \quad x \in [0, h_n], \\ a_n = 1,\ b_n = c_n = 0, & \text{if} \quad x > h_n. \end{cases} \tag{19}$$

Following (9) we obtain: for $n = 1$, $v_1(x) = x^2$, $x \geq 0$; for $n \geq 2$,

$$v_n(x) = \begin{cases} \dfrac{1}{2(n-1)} \left(\sqrt{h_n^2 + 4n(n-1)x^2} - h_n\right), & \text{if} \quad x \in [0, h_n], \\ x, & \text{if} \quad x > h_n. \end{cases} \tag{20}$$

Returning to (10) via (18), we obtain the modified univariate Bernstein-Chlodovsky operators L^*_n. Accordingly, based on (12), the bivariate extension for each $(x, y) \in [0, h_{m_1}] \times [0, h_{m_2}]$ and $f \in \mathcal{D}$ is defined by

$$(L^*_{m_1,m_2} f)(x,y) = \sum_{i=0}^{m_1} \sum_{j=0}^{m_2} c_{m_1,m_2}(i,j) v^i_{m_1}(x) v^j_{m_2}(y) (h_{m_1} - v_{m_1}(x))^i (h_{m_2} - v_{m_2}(y))^j$$

$$\times f\left(\frac{i}{m_1} h_{m_1}, \frac{j}{m_2} h_{m_2}\right), \tag{21}$$

where $c_{m_1,m_2}(i,j) = \binom{m_1}{i}\binom{m_2}{j} h_{m_1}^{-m_1} h_{m_2}^{-m_2}$ and v_m is described by (20).

We notice the following aspect. From (19) we get $a_1 = 0$ and this should be in contradiction with (6). In fact nothing is wrong because, this time, relation (5)

must hold only for $x \in [0, h_n]$, not for each $x \in \mathbb{R}_+$. Consequently, condition $a_n \geq 1$ in (6) is not necessary to take place.

Particular case. If we choose $h_n = 1$ in (18), then L_n becomes the classical nth Bernstein polynomial for each $n \in \mathbb{N}$. In this case relations (19) and (20) imply $v_n(x) = r_n^*(x)$, $x \in [0, 1]$, see (2). The King's operators (1) are reobtained.

Remarks. a) If we choose in (3) $u_{n,k}(x) := e^{-nx}\frac{(nx)^k}{k!}$ and $x_{k,n} := k/n$, the well-known Szász-Mirakyan-Favard operator is obtained. A variant of this operator in two dimensions was defined by Totik [9; *p.292*] as follows

$$(S_{n,m}f)(x,y) = e^{-nx-my}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(nx)^i}{i!}\frac{(my)^j}{j!}f\left(\frac{i}{n},\frac{j}{m}\right).$$

b) If we choose in (3) $u_{n,k}(x) := \binom{n-1+k}{k}x^k(1+x)^{-n-k}$ and $x_{k,n} := k/n$, the classical Baskakov operator is obtained. In [6] the authors have considered the Baskakov operator for functions of two variables given by

$$(A_{n,m}f)(x,y) = \frac{1}{(1+x)^n(1+y)^m}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}c_{i,j}(n,m)\left(\frac{x}{1+x}\right)^i\left(\frac{y}{1+y}\right)^j f\left(\frac{i}{n},\frac{j}{m}\right),$$

where $c_{i,j}(n,m) := \binom{n-1+i}{i}\binom{m-1+j}{j}$.

Following our technique, in the same manner we can obtain the modified variants given by relation (12) of the above two classes. By a short computation, relation (9) becomes

a) $v_n(x) = \dfrac{\sqrt{1+4n^2x^2}-1}{2n}$, b) $v_n(x) = \dfrac{\sqrt{1+4n(n+1)x^2}-1}{2(n+1)}$,

respectively.

References

[1] Agratini, O., *Linear operators that preserve some test functions*, International Journal of Mathematics and Mathematical Sciences, Volume 2006, Article ID94136, Pages 1-11, DOI 10.1155/IJJMS.

[2] Altomare, F. and Campiti, M., *Korovkin-type Approximation Theory and its Applications*, de Gruyter Studies in Mathematics, Vol.17, Walter de Gruyter, Berlin, 1994.

[3] Chlodovsky, I., *Sur le développment des fonctions définies dans un intervalle infini en séries de polynômes de M.S. Bernstein*, Compositio Math. **4**(1937), 380-393.

[4] Duman, O. and Orhan, C., *An abstract version of the Korovkin approximation theorem*, Publ. Math. Debrecen, **69**(2006), f.1-2, 33-46.

[5] Gonska, H. and Piţul, P., *Remarks on an article of J.P. King*, Schriftenreihe des Fachbereichs Mathematik, Nr.596(2005), Universität Duisburg-Essen, 1-8.

[6] Gurdek, M., Rempulska, L. and Skorupka, M., *The Baskakov operators for functions of two variables*, Collect. Math., **50**(1999), f.3, 289-302.

[7] Ipatov, A.F., *Estimate of the error and order of approximation of functions of two variables by S.N. Berstein's polynomials* (in Russian), Uč. Zap. Petrozavodsk Univ., **4**(1955), f.4, 31-48.

[8] King, J.P., *Positive linear operators which preserve* x^2, Acta Math. Hungar., **99**(2003), no.3, 203-208.

[9] Totik, V., *Uniform approximation by Szász-Mirakjan type operators*, Acta Math. Hung., **41**(1983), f.3-4, 291-307.

[10] Volkov, V.I., *On the convergence of sequences of linear positive operators in the space of continuous functions of two variables* (in Russian), Dokl. Akad. Nauk SSSR (N.S.), **115**(1957), 17-19.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: agratini@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

REPRESENTATION THEOREMS AND ALMOST UNIMODAL SEQUENCES

DORIN ANDRICA AND DANIEL VĂCĂREŢU

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. We define the almost unimodal sequences and we show that under some conditions the polynomial $P(X^k + n)$ is almost unimodal (Theorem 1.7). A nontrivial example of almost unimodality shows that the sequence $A_k^{(1)}(n)$, $k = -\frac{n(n+1)}{2}, \ldots, -1, 0, 1, \ldots, \frac{n(n+1)}{2}$ is symmetric and almost unimodal (Theorem 3.1). This result is connected to some representation properties of integers.

1. Almost unimodal sequences and polynomials

A finite sequence of real numbers $\{d_0, d_1, \ldots, d_m\}$ is said to be unimodal if there exists an index $0 \leq m^* \leq m$, called the mode of the sequence, such that d_j increases up to $j = m^*$ and decreases from then on, that is, $d_0 \leq d_1 \leq \cdots \leq d_{m^*}$ and $d_{m^*} \geq d_{m^*+1} \geq \cdots \geq d_m$. A polynomial is said to be unimodal if its sequence of coefficients is unimodal.

Unimodal polynomials arise often in combinatorics, geometry and algebra. The reader is referred to [BoMo] and [AlAmBoKaMoRo] for surveys of the diverse techniques employed to prove that specific families of polynomials are unimodal.

We recall few basic results concerning the unimodality.

Theorem 1.1. *If P is a polynomial with positive nondecreasing coefficients, then $P(X + 1)$ is unimodal.*

Received by the editors: 15.05.2006.

2000 *Mathematics Subject Classification.* Primary 11B99, 11K31, Secondary 40A99, 40C10.

Key words and phrases. Unimodal sequence, unimodal polynomial, almost unimodality, Erdös-Surányi sequence, complete sequence of integers.

Theorem 1.2. *Let $b_k > 0$ be a nondecreasing sequence. Then the sequence*

$$c_j = \sum_{k=j}^{m} b_k \binom{k}{j}, \quad 0 \leq j \leq m \tag{1.1}$$

is unimodal with mode $m^* = \left\lfloor \frac{m-1}{2} \right\rfloor$.

Theorem 1.3. *Let $0 \leq a_0 \leq a_1 \leq \cdots \leq a_m$ be a sequence of real numbers and $n \in \mathbb{N}$, and consider the polynomial*

$$P = a_0 + a_1 X + a_2 X^2 + \cdots + a_m X^m. \tag{1.2}$$

Then the polynomial $P(X+n)$ is unimodal with mode $m^* = \left\lfloor \frac{m}{n+1} \right\rfloor$.

We can reformulate Theorem 1.3 in terms of the coefficients of polynomial P.

Theorem 1.4. *Let $0 \leq a_0 \leq a_1 \leq \cdots \leq a_m$ be a sequence of real numbers and $n \in \mathbb{N}$. Then the sequence*

$$q_j = q_j(m,n) = \sum_{k=j}^{m} a_k \binom{k}{j} n^{k-j}, \quad 0 \leq j \leq m \tag{1.3}$$

is unimodal with mode $m^* = \left\lfloor \frac{m}{n+1} \right\rfloor$.

In order to introduce the almost unimodality of a sequence we need the following notion.

Definition 1.5. A finite sequence of real numbers $\{c_0, c_1, \ldots, c_n\}$ is called **almost nondecreasing** if it is nondecreasing excepting a subsequence which is zero.

It is clear that, if the sequence $\{c_0, c_1, \ldots, c_n\}$ is nondecreasing, then it is almost nondecreasing. The converse is not true, as we can see from the following example. The sequence $\{0, 1, 0, 2, 0, 3, \ldots, 0, m\}$ is almost nondecreasing but it is not nondecreasing.

Definition 1.6. A finite sequence of real numbers $\{d_0, d_1, \ldots, d_m\}$ is called **almost unimodal** if there exists an index $0 \leq m^* \leq m$, such that d_j almost increases up to $j = m^*$ and d_j almost decreases from then on.

As in the situation of unimodality, the index m^* is called the mode of the sequence. Also, a polynomial is said to be almost unimodal, if its sequence of coefficients is almost unimodal.

For instance, the polynomial

$$(X^k+1)^m = \binom{m}{0} + \binom{m}{1}X^k + \binom{m}{2}X^{2k} + \cdots + \binom{m}{m}X^{mk}$$

is almost unimodal for $k \geq 2$, but it is not unimodal.

The following result is useful in the study of almost unimodality.

Theorem 1.7. *Let $0 \leq a_0 \leq a_1 \leq \cdots \leq a_m$ be a sequence of real numbers, let n be a positive integer and consider the polynomial*

$$P = a_0 + a_1X + a_2X^2 + \cdots + a_mX^m.$$

Then for any integer $k \geq 2$, the polynomial $P(X^k+n)$ is almost unimodal.

Proof. We note that if Q is a unimodal polynomial, then for any $k \geq 2$ the polynomial $Q(X^k)$ is almost unimodal. Applying Theorem 1.3 we get that $P(X+n)$ is unimodal and now using the remark above it follows that $P(X^k+n)$ is almost unimodal with mode $m^* = k\left\lfloor \dfrac{m}{n+1} \right\rfloor$. □

Remark 1.8. If $n \geq m$, then $m^* = 0$, hence the sequence of coefficients of $P(X^k+n)$ is almost nonincreasing. For example, the sequence of coefficients of $(X^k+3)^3$ is

$$27, \underbrace{0, \ldots, 0}_{k-1}, 27, \underbrace{0, \ldots, 0}_{k-1}, 9, \underbrace{0, \ldots, 0}_{k-1}, 1.$$

2. **Some representation results for integers**

In 1960, P. Erdös and J. Surányi ([ErSu], Problem 5, pp.200) have proved the following result: Any integer k can be written in infinitely many ways in the form

$$k = \pm 1^2 \pm 2^2 \pm \cdots \pm n^2 \tag{2.1}$$

for some positive integer n and for some choices of signs $+$ and $-$.

In 1979, J. Mitek [Mi] has extended the above result as follows: For any fixed positive integer $s \geq 2$ the result in (2.1) holds in the form

$$k = \pm 1^s \pm 2^s \pm \cdots \pm n^s \tag{2.2}$$

The following notion has been introduced in [Dr] by M.O. Drimbe:

Definition 2.1. A sequence $(a_n)_{n\geq 1}$ of positive integers is an **Erdös-Surányi sequence** if any integer k can be represented in infinitely many ways in the form

$$k = \pm a_1 \pm a_2 \pm \cdots \pm a_n \tag{2.3}$$

for some positive integer n and for some choices of signs $+$ and $-$.

The main result in [Dr] is contained in

Theorem 2.2. *Any sequence $(a_n)_{n\geq 1}$ of positive integers satisfying:*

i) $a_1 = 1$,

ii) $a_{n+1} \leq 1 + a_1 + \cdots + a_n$, *for any positive integer* n,

iii) $(a_n)_{n\geq 1}$ *contains infinitely many odd integers,*

is an Erdös-Suranyi sequence.

As direct consequences of Theorem 2.1, in the paper [Dr], the following examples of Erdös-Suranyi sequences are pointed out:

1) The Fibonacci's sequence $(F_n)_{n\geq 0}$, where $F_0 = 1$, $F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$, for $n \geq 1$;

2) The sequence of primes $(p_n)_{n\geq 1}$.

We can see that the sequence $(n^s)_{n\geq 1}$ does not satisfy condition ii) in Theorem 2.2 but it is an Erdös Suranyi sequences, according to the result of J. Mitek [Mi] contained in (2.2). Following the paper [Ba] one can extend Theorem 2.2 in such way to include sequences $(n^s)_{n\geq 1}$. The following notion has been introduced in [Kl] by T. Klove:

Definition 2.3. A sequence $(a_n)_{n\geq 1}$ of positive integers is **complete** if any sufficiently great integer can be expressed as a sum of distinct terms of $(a_n)_{n\geq 1}$.

The above property is equivalent to the fact that for any sufficient great integer k there exists a positive integer $t = t(k)$ such that

$$k = u_1a_1 + u_2a_2 + \cdots + u_ta_t, \tag{2.4}$$

where $u_i \in \{0, 1\}$, $i = 1, 2, \ldots, t$.

The main result in [Ba] is contained in

Theorem 2.4. *Any complete sequence $(a_n)_{n\geq 1}$ of positive integers, containing infinitely many odd integers, is an Erdös-Surányi sequence.*

Proof. Let q can be represented as in (2.4). Let $S_n = a_1 + \cdots + a_n$, $n \geq 1$. The sequence $(S_n)_{n\geq 1}$ is increasing and it contains infinitely many odd integers but also infinitely many even integers. Let k be a fixed positive integer. One can find infinitely many integers S_p, having the same parity as k, such that $S_p > k + 2q$. Consider S_n a such integer and let $m = \dfrac{1}{2}(S_n - k)$. Because $q < m$, it follows that m can be represented as in (2.4). Taking into account that $m < S_n$, we have $m = u_1a_1 + \cdots + u_na_n$, where $u_i \in \{0, 1\}$, $i = 1, 2, \ldots, n$. Then, we have

$$k = S_n - 2m = (1 - 2u_1)a_1 + \cdots + (1 - 2u_n)a_n.$$

From $u_i \in \{0, 1\}$ we get $1 - 2u_i \in \{-1, 1\}$, $i = 1, 2, \ldots, n$. □

Remark 2.5. The result of J. Mitek [Mi] follows from Theorem 2.4 and from the property that the sequence $(n^s)_{n\geq 1}$ is complete, for any positive integer s. The completeness of $(n^s)_{n\geq 1}$ is a result of P. Erdös (see [Si], pp.395).

3. **Integral formulae and almost unimodality**

Consider an Erdös-Surányi sequence $(a_m)_{m\geq 1}$. If we fix n, then there are 2^n integers of the form $\pm a_1 \pm \cdots \pm a_n$. In this section we explore the number of ways to express an integer k in the form (2.3). Denote $A_k(n)$ to be this value. Using the method in [AnTo] let us consider the function

$$f_n(z) = \left(z^{a_1} + \frac{1}{z^{a_1}}\right)\left(z^{a_2} + \frac{1}{z^{a_2}}\right)\ldots\left(z^{a_n} + \frac{1}{z^{a_n}}\right) \tag{3.1}$$

It is clear that this is the generating function for the sequence $A_k(n)$, i.e. we may write

$$f_n(z) = \sum_{j=-S_n}^{S_n} A_j(n) z^j, \tag{3.2}$$

where $S_n = a_1 + \cdots + a_n$. It is interesting to note the symmetry of the coefficients in (3.2), i.e. $A_j(n) = A_{-j}(n)$. If we write $z = \cos t + i \sin t$, then by using DeMoivre's formula we may rewrite (3.1) as

$$f_n(z) = 2^n \cos a_1 t \cdot \cos a_2 t \ldots \cos a_n t \tag{3.3}$$

By noting that $A_k(n)$ is the constant term in the expansion $z^{-k} f_n(z)$, we obtain

$$\begin{aligned} z^{-k} f_n(z) &= 2^n (\cos kt - i \sin kt) \cos a_1 t \ldots \cos a_n t \\ &= A_k(n) + \sum_{j \neq k} A_j(n)(\cos(j-k)t + i \sin(j-k)t) \end{aligned} \tag{3.4}$$

Finally, making use of the fact that $\displaystyle\int_0^{2\pi} \cos mt dt = \int_0^{2\pi} \sin mt dt = 0$, we integrate (3.4) on the interval $[0, 2\pi]$ to find an elegant integral formula for $A_k(n)$:

$$A_k(n) = \frac{2^n}{2\pi} \int_0^{2\pi} \cos a_1 t \ldots \cos a_n t \cos kt dt \tag{3.5}$$

After integrating, we find that the imaginary part of $A_k(n)$ is 0, which implies the relation

$$\int_0^{2\pi} \cos a_1 t \ldots \cos a_n t \sin kt dt = 0 \tag{3.6}$$

for each k between $-S_n$ and S_n.

Applying formula (3.5) for Erdös-Surányi sequence $(m^s)_{m \geq 1}$, we get

$$A_k^{(s)}(n) = \frac{2^n}{2\pi} \int_0^{2\pi} \cos 1^s t \cos 2^s t \ldots \cos n^s t \cos kt dt,$$

where $A_k^{(s)}(n)$ denote the integer $A_k(n)$ for this sequence.

The following result gives a nontrivial example of almost unimodality.

Theorem 3.1. *The sequence* $A_k^{(1)}(n)$, $k = 0, 1, \ldots, \dfrac{n(n-1)}{2}$, *is almost nonincreasing and consequently, the sequence* $A_j^{(1)}(n)$, $j = -\dfrac{n(n+1)}{2}, \ldots, -1, 0, 1, \ldots, \dfrac{n(n+1)}{2}$ *is symmetric and almost unimodal.*

Proof. First of all we show that $A_k^{(1)}(n)$ is the number of representations of $\dfrac{1}{2}\left(\dfrac{n(n+1)}{2} - k\right)$ as $\sum\limits_{i=1}^{n} \varepsilon_i i$, where $\varepsilon_i \in \{0, 1\}$. Indeed, we note that if $\varepsilon \in \{0, 1\}$, then $1 - 2\varepsilon \in \{-1, 1\}$ and we have $\sum\limits_{i=1}^{n}(1 - 2\varepsilon_i)i = k$ if and only if

$$\frac{n(n+1)}{2} - 2\sum_{i=1}^{n} \varepsilon_i i = k,$$

hence

$$\sum_{i=1}^{n} \varepsilon_i i = \frac{1}{2}\left(\frac{n(n+1)}{2} - k\right). \tag{3.7}$$

Denote $B_k^{(1)}(n)$ the number of representations of $\dfrac{1}{2}\left(\dfrac{n(n+1)}{2} - k\right)$ in the form (3.7). It is clear that $B_k^{(1)}(n) = 0$ if and only if k and $\dfrac{n(n+1)}{2}$ have different parities. Also, we have $\dfrac{n(n+1)}{4} \leq j \leq \dfrac{n(n+1)}{2}$ for any integer j of the form $\dfrac{1}{2}\left(\dfrac{n(n+1)}{2} - k\right)$, $k = 0, 1, \ldots, \dfrac{n(n+1)}{2}$. Assume that we can write j as $\varepsilon_1 \cdot 1 + \varepsilon_2 \cdot 2 + \cdots + \varepsilon_n \cdot n$ and $\varepsilon_1 = 1$. Then, we have $j - 1 = \varepsilon_2 \cdot 2 + \cdots + \varepsilon_n \cdot n$, where $\varepsilon_2, \ldots, \varepsilon_n \in \{0, 1\}$. If we have in this sum three consecutive terms of the form $i - 1, 0, i + 1$, we can move 1 at the first position and obtain three consecutive terms of the form $i - 1, i, 0$. After another such step for other three consecutive terms $s - 1, 0, s + 1$, taking into account that a such map is injective it follows that $B_j^{(1)}(n) \leq B_{j-2}^{(1)}(n)$, hence $A_j^{(1)}(n) \leq A_{j-2}^{(1)}(n)$ if both $A_{j-2}^{(1)}(n)$ and $A_j^{(1)}(n)$ are not zero.

Remark 3.2. The conclusion of Theorem 3.1 is not generally true for $A_k^{(s)}(n)$, where $s \geq 2$ (see the values of $A_k^{(2)}(6)$ in the table below).

□

4. **Numerical results**

Numerical values for $A_k^{(1)}$ for n up to 9

$n=$	1
k	A_k
0	0
1	1

$n=$	2
k	A_k
0	0
1	1
2	0
3	1

$n=$	3
k	A_k
0	2
1	0
2	1
3	0
4	1
5	0
6	1

$n=$	4
k	A_k
0	2
1	0
2	2
3	0
4	2
5	0
6	1
7	0
8	1
9	0
10	1

$n=$	5
k	A_k
0	0
1	3
2	0
3	3
4	0
5	3
6	0
7	2
8	0
9	2
10	0
11	1
12	0
13	1
14	0
15	1

$n=$	6
k	A_k
0	0
1	5
2	0
3	5
4	0
5	4
6	0
7	4
8	0
9	4
10	0
11	3
12	0
13	2
14	0
15	2
16	0
17	1
18	0
19	1
20	0
21	1

$n=$	7
k	A_k
0	8
1	0
2	8
3	0
4	8
5	0
6	7
7	0
8	7
9	0
10	6
11	0
12	5
13	0
14	5
15	0
16	4
17	0
18	3
19	0
20	2
21	0
22	2
23	0
24	1
25	0
26	1
27	0
28	1

$n=$	8
k	A_k
0	14
1	0
2	13
3	0
4	13
5	0
6	13
7	0
8	12
9	0
10	11
11	0
12	10
13	0
14	9
15	0
16	8
17	0
18	7
19	0
20	6
21	0
22	5
23	0
24	4
25	0
26	3
27	0
28	2
29	0
30	2
31	0
32	1
33	0
34	1
35	0
36	1

$n=$	9
k	A_k
0	0
1	23
2	0
3	23
4	0
5	22
6	0
7	21
8	0
9	21
10	0
11	19
12	0
13	18
14	0
15	17
16	0
17	15
18	0
19	13
20	0
21	12
22	0
23	10
24	0
25	9
26	0
27	8
28	0
29	6
30	0
31	5
32	0
33	4
34	0
35	3
36	0
37	2
38	0
39	2
40	0
41	1
42	0
43	1
44	0
45	1

Numerical values for $A_k^{(2)}$ for n up to 6

$n=$	1
k	A_k
0	0
1	1

$n=$	2
k	A_k
1	0
2	0
3	1
4	0
5	1

$n=$	3
k	A_k
0	0
1	0
2	0
3	0
4	1
5	0
6	1
7	0
8	0
9	0
10	0
11	0
12	1
13	0
14	1

$n=$	4
k	A_k
0	0
1	0
2	1
3	0
4	1
5	0
6	0
7	0
8	0
9	0
10	1
11	0
12	1
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	1
21	0
22	1
23	0
24	0
25	0
26	0
27	0
28	1
29	0
30	1

$n=$	5
k	A_k
0	0
1	0
2	0
3	2
4	0
5	2
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	1
14	0
15	1
16	0
17	0
18	0
19	0
20	0
21	1
22	0
23	1
24	0
25	0
26	0
27	1

$n=$	5
k	A_k
28	0
29	1
30	0
31	0
32	0
33	0
34	0
35	1
36	0
37	1
38	0
39	0
40	0
41	0
42	0
43	0
44	0
45	1
46	0
47	1
48	0
49	0
50	0
51	0
52	0
53	1
54	0
55	1

$n=$	6
k	A_k
0	0
1	2
2	0
3	0
4	0
5	0
6	0
7	1
8	0
9	2
10	0
11	1
12	0
13	1
14	0
15	1
16	0
17	1
18	0
19	1
20	0
21	1
22	0
23	1
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	2
32	0
33	2
34	0
35	0
36	0
37	0
38	0
39	2
40	0
41	2
42	0
43	0
44	0

$n=$	6
k	A_k
45	0
46	0
47	0
48	0
49	1
50	0
51	1
52	0
53	0
54	0
55	0
56	0
57	1
58	0
59	1
60	0
61	0
62	0
63	1
64	0
65	1
66	0
67	0
68	0
69	0
70	0
71	1
72	0
73	1
74	0
75	0
76	0
77	0
78	0
79	0
80	0
81	1
82	0
83	1
84	0
85	0
86	0
87	0
88	0
89	1
90	0
91	1

Numerical values for $A_0^{(1)}(n)$ and $A_0^{(2)}(n)$

(1)

n	A_0
1	0
2	0
3	2
4	2
5	0
6	0
7	8
8	14
9	0
10	0
11	70
12	124
13	0
14	0
15	722
16	1314
17	0
18	0
19	8220
20	15272
21	0
22	0
23	99820
24	187692
25	0
26	0
27	1265204
28	2399784
29	0
30	0
31	16547220
32	31592878
33	0
34	0
35	221653776
36	425363952
37	0
38	0
39	3025553180
40	5830034720
41	0
42	0
43	41931984034
44	81072032060
45	0
46	0
47	588431482334
48	1140994231458
49	0
50	0

n	A_0
51	8346638665718
52	16221323177468
53	0
54	0
55	119447839104366
56	232615054822964
57	0
58	0
59	1722663727780132
60	3360682669655028
61	0
62	0
63	25011714460877474
64	48870013251334676
65	0
66	0
67	365301750223042066
68	714733339229024336
69	0
70	0
71	53632882995852788800
72	105063310218141423340
73	0
74	0
75	79110709437891746598
76	155141342711178904962
77	0
78	0
79	1171806326862876802144
80	2300241216389780443900
81	0
82	0
83	17422684839627191647442
84	34230838910489146400266
85	0
86	0
87	259932234752908992679732
88	51110796628205911410542 4
89	0
90	0
91	38900805399055543953121 72
92	76547464704667766365081 50
93	0
94	0
95	58384150201994432824279356
96	11496359389815969968780 5154
97	0
98	0
99	87855297309635235880572 0000
100	17310240059487250166337 86324

(2)

n	A_0
1	0
2	0
3	0
4	0
5	0
6	0
7	2
8	2
9	0
10	0
11	2
12	10
13	0
14	0
15	86
16	114
17	0
18	0
19	478
20	860
21	0
22	0
23	5808
24	10838
25	0
26	0
27	55626
28	100426
29	0
30	0
31	696164
32	1298600
33	0
34	0
35	7826992
36	14574366
37	0
38	0
39	100061106
40	187392994
41	0
42	0
43	1223587084
44	2322159814
45	0
46	0
47	16019866270
48	30353305134
49	0
50	0

References

[AlAmBoKaMoRo] Alvarez, J., Amadis, M., Boros, G., Karp, D., Moll, V., Rosales, L., *An extension of a criterion for unimodality*, Elec. Jour. Comb. 8, #R10, 2001.

[AnTo] Andrica, D., Tomescu, I., *On an integer sequence related to a product of trigonometric functions and its combinatorial relevance*, Journal of Integer Sequences, Vol.5(2002).

[Ba] Badea, C., *On Erdös-Surányi sequences* (Romanian), R.M.T., Nr.1(1987), 10-13.

[BoMo] Boros, G., Moll, V., *A criterion for unimodality*, Elec. Jour. Comb. 6, #R10, 1999.

[BoMo] Boros, G., Moll, V., *An integral hidden in Gradshteyn and Rhyzik*, Jour. Comp. Appl. Math. 237, 272-287, 1999.

[Br] Brown, J.L., *Integer representations and complete sequences*, Mathematics Magazine, vol.49(1976), no.1, 30-32.

[Dr] Drimbe, M.O., *A problem of representation of integers* (Romanian), G.M.-B, 10-11(1983), 382-383.

[ErSu] Erdös, P., Surányi, J., *Selected chapters from number theory*, Tankönyvkiadó Vállalat, Budapest, 1960.

[Kl] Klove, T., *Sums of distinct elements from a fixed set*, Mathematics of Computation **29**(1975), No.132, 1144-1149.

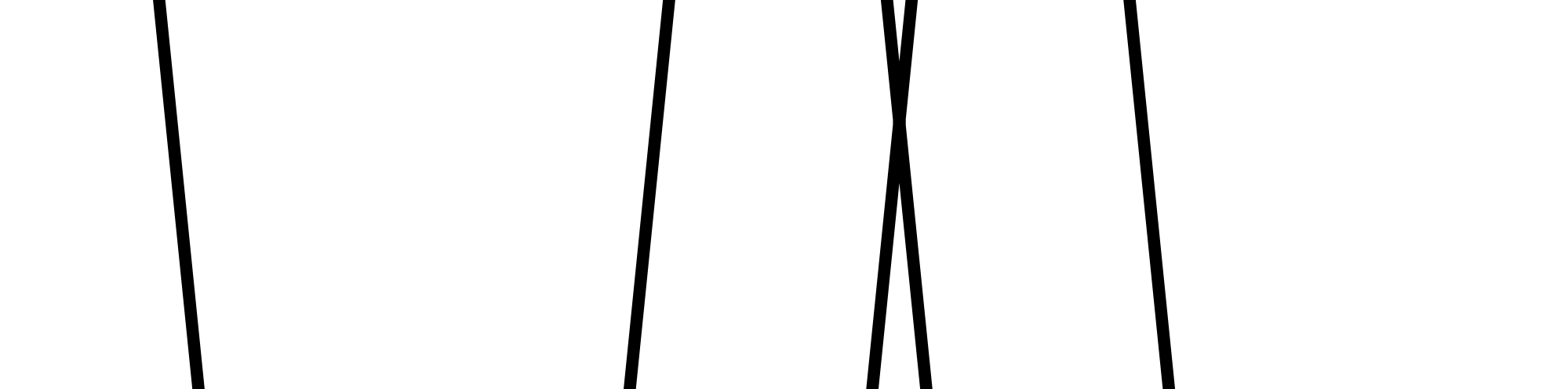
[Mi] Mitek, J., *Generalization of a theorem of Erdös and Surányi*, Annales Societatis Mathematical Polonae, Series I, Commentationes Mathematicae, XXI (1979).

[SaAn] Savchev, S., Andreescu, T., *Mathematical Miniatures*, The Mathematical Association of America, Anelli Lax Mathematical Library, Volume #43, 2003.

[Si] Sierpinski, W., *Elementary theory of numbers*, P.W.N., Warszawa, 1964.

[Vă] Văcăreţu, D., *Unimodal Polynomials: Methods and Techniques*, in "Recent Advances in Geometry and Topology", Proc. of The 6th International Workshop on Differential Geometry and Topology and The 3rd German-Romanian Seminar on Geometry (D. Andrica and P.A. Blaga, Eds.), Cluj University Press, 2004, pp.391-395.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: dorinandrica@yahoo.com



STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

SOME INFERENCES AND EXPERIMENTS ON FREE KNOTS SPLINE REGRESSION

PETRU P. BLAGA

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. Inferences and experiments on the simple spline regression with free knots are considered. For the first time an iterative procedure given in [2] to estimate the values of the free knots based on a multiple linear regression is recalled. Point estimators and confidence interval estimators on the spline regression coefficients and variance of the response, confidence interval estimators and (Scheffé [7]) simultaneous confidence interval estimators on the mean value response and prediction value are considered. Inferences are illustrated by some numerical experiments.

1. Introduction

A multiple linear regression model with constant term is given by the functional relation

$$Y = \beta_0 + \sum_{k=1}^{r} \beta_k X_k + \varepsilon,$$

where Y is the response (dependent) variable, $X_1, \ldots, X_r$ are the regressor (independent) variables, and ε represents the error term (random noise).

The multiple linear regression analysis consists in the study of the influence of the variables $X_1, \ldots, X_r$ on the variable Y. This study is realized by the inferences on regression coefficients β_k, and error term ε. In this aim a sample of n data observations

Received by the editors: 01.08.2006.

2000 *Mathematics Subject Classification.* 65D10, 62J05, 62F10, 65C20.

Key words and phrases. Spline regression, multiple linear regression, confidence intervals.

are considered

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \quad \begin{pmatrix} 1 & x_{11} & \dots & x_{1r} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nr} \end{pmatrix} = (\mathbf{1}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_r) = \boldsymbol{X},$$

and the sample multiple linear regression can be written in the matrix form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \dots, \beta_r) \in \mathbb{R}^{r+1}$, $\boldsymbol{\varepsilon}^\top = (\varepsilon_1, \dots, \varepsilon_n) \in \mathbb{R}^n$. The classical multiple linear regression model supposes that the random vector $\boldsymbol{\varepsilon}$ follows the normal distribution $\mathcal{N}\left(\mathbf{0}; \sigma^2 \boldsymbol{I}_n\right)$, i.e. the components of $\boldsymbol{\varepsilon}$ are independent and identically distributed, each of them following the same normal distribution $\mathcal{N}\left(0; \sigma^2\right)$. A solution $(\boldsymbol{b}, \boldsymbol{e})$, with $\boldsymbol{b} \in \mathbb{R}^{r+1}$, $\boldsymbol{e} \in \mathbb{R}^n$, of the system of equations $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$ is called a fitted multiple linear regression, and the solution satisfying the least-squares criterion

$$\|\boldsymbol{e}\|^2 = \boldsymbol{e}^\top \boldsymbol{e} = \sum_{i=1}^{n} e_i^2 \longrightarrow \text{minim},$$

is called the fitted least-squares multiple linear regression.

It is well-known that the fitted least-squares coefficients are given by

$$\boldsymbol{b} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}, \tag{1}$$

and these are unbiased estimators of $\boldsymbol{\beta}$. Moreover, we have that

$$s^2 = \frac{1}{n-r-1} \sum_{k=1}^{n} e_k^2 \tag{2}$$

is an unbiased estimator for the parameter σ^2. We remark that the vector of fitted values $\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b}$ and the vector of residuals $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ can be expressed by the hat matrix

$$\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top,$$

namely $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, and $\boldsymbol{e} = (\boldsymbol{I}_n - \boldsymbol{H})\,\boldsymbol{y}$, respectively, where $\boldsymbol{I}_n$ denotes identity matrix of order n.

Also, we have the coefficient of determination given by

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}. \tag{3}$$

2. **Free knots simple spline regression**

The simple spline regression model with distinct free knots $\tau_1, \ldots, \tau_p$

$$Y = \sum_{k=0}^{m} \alpha_k X^k + \sum_{j=1}^{p} \beta_j (\tau_j - X)_+^m + \varepsilon \tag{4}$$

can be reduced to a multiple linear regression model with constant term, if one introduces the new $m+p$ regressor variables $X_k = X^k$, $k = \overline{1,m}$, and $X_{m+j} = (\tau_j - X)_+^m$, $j = \overline{1,p}$.

The spline technique became a very useful in regression analysis, see, for example, [4] and [9]. Some remarks on the number and positions of the knots τ_i are presented in [6] following the suggests given by Wold in [11]: (1) there should be as few knots as possible, with at least four or five data points per segment; (2) there should be no more than one extrem point and one point of inflexion per segment; (3) in so far as possible, the extrem points should be centred in the segment and the point of inflexion should be near the knots.

The transformation on the regressor variables given by Box and Tidwell [3], recalled in [6], was used in [2] to estimate positions of the knots $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$, $\tau_1 < \ldots < \tau_p$.

Let us consider a sample of n pairs of data (x_i, y_i), $i = \overline{1,n}$. The sample spline regression is reduced to a sample multiple linear regression

$$y_i = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ik} + \sum_{j=1}^{p} \beta_j x_{i,m+j} + \varepsilon_i, \quad i = \overline{1,n}, \tag{5}$$

where $x_{ik} = x_i^k$, $x_{i,m+j} = (\tau_j - x_i)_+^m$.

Using matrix notation for observations of response variable, coefficients of model, error terms

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_m \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

and design matrix of model

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} & x_{1,m+1} & \dots & x_{1,m+p} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nm} & x_{n,m+1} & \dots & x_{n,m+p} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & x_1 & \dots & x_1^m & (\tau_1 - x_1)_+^m & \dots & (\tau_p - x_1)_+^m \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^m & (\tau_1 - x_n)_+^m & \dots & (\tau_p - x_n)_+^m \end{pmatrix}$$

the regression model (5) has the matrix expression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}. \tag{6}$$

Taking into account that the knots of the spline regression are unknown, the following iterative procedure to obtain the knots τ_j is proposed.

For the first time an initial appropriate value $\boldsymbol{\tau}^{(0)} = \left(\tau_1^{(0)}, \dots, \tau_p^{(0)}\right)$ of $\boldsymbol{\tau}$ is considered. Thus, we have an initial spline regression model of type (4) with the attached multiple linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\delta} + \boldsymbol{\varepsilon}_0, \tag{7}$$

where

$$\boldsymbol{X}_0 = \begin{pmatrix} 1 & x_1 & \dots & x_1^m & \left(\tau_1^{(0)} - x_1\right)_+^m & \dots & \left(\tau_p^{(0)} - x_1\right)_+^m \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^m & \left(\tau_1^{(0)} - x_n\right)_+^m & \dots & \left(\tau_p^{(0)} - x_n\right)_+^m \end{pmatrix},$$

and corresponding vector of errors $\boldsymbol{\varepsilon}_0$. Based on (1), the least-squares estimators of the coefficients $\boldsymbol{\delta}$ of the initial model (7) are given by

$$\boldsymbol{d}_0 = (a_0, \dots, a_m; b_1, \dots, b_p)^\top = \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^\top \boldsymbol{y}.$$

For this multiple linear regression model, we have:

- the vector of fitted values (estimated values)

$$\hat{\boldsymbol{y}}^\top = \boldsymbol{X}_0 \boldsymbol{d}_0 = (\hat{y}_1, \dots, \hat{y}_n),$$

- the residual sum of squares

$$\|\boldsymbol{e}_0\|^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \tag{8}$$

- the residual mean squares (unbiased estimator of σ^2)

$$s_0^2 = \frac{1}{n-r-1} \|\boldsymbol{e}_0\|^2 \quad (\text{with } r = m + p), \tag{9}$$

- the coefficient of determination R_0^2 given by (3).

Then, the expanding of

$$h(X; \boldsymbol{\tau}) = h(X; \tau_1, \dots, \tau_p) = \sum_{j=1}^p \beta_j (\tau_j - X)_+^m$$

in Taylor series about the initial value $\boldsymbol{\tau}^{(0)}$ and ignoring terms of higher than first order, we obtain

$$h(X; \boldsymbol{\tau}) = h\left(X; \boldsymbol{\tau}^{(0)}\right) + \left(\boldsymbol{\tau} - \boldsymbol{\tau}^{(0)}\right)^\top \left.\frac{\partial h(X; \boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau} = \boldsymbol{\tau}^{(0)}} + \mathcal{O}\left(\eta^2\right)$$

where $\eta = \max_{j=\overline{1,p}} \left(\left| \tau_j - \tau_j^{(0)} \right| \right)$. Taking into account that

$$\frac{\partial h\left(X;\boldsymbol{\tau}\right)}{\partial \tau_j} = m\beta_j \left(\tau_j - X\right)_+^{m-1}, \quad j = \overline{1,p},$$

it results

$$h\left(X;\boldsymbol{\tau}\right) = h\left(X;\boldsymbol{\tau}^{(0)}\right) + \sum_{j=1}^{p} m\beta_j \left(\tau_j - \tau_j^{(0)}\right) \left(\tau_j^{(0)} - X\right)_+^{m-1} + \mathcal{O}\left(\eta^2\right).$$

Thus, an extended spline regression model is obtained:

$$Y = \sum_{k=0}^{m} \alpha_k X^k + \sum_{j=1}^{p} \beta_j \left(\tau_j^{(0)} - X\right)_+^m + \sum_{j=1}^{p} m\beta_j \left(\tau_j - \tau_j^{(0)}\right) \left(\tau_j^{(0)} - X\right)_+^{m-1} + \tilde{\varepsilon},$$

with a corresponding extended multiple linear regression

$$Y = \alpha_0 + \sum_{k=1}^{m} \alpha_k X_k + \sum_{j=1}^{p} \beta_j X_{m+j} + \sum_{j=1}^{p} \gamma_j X_{m+p+j} + \tilde{\varepsilon},$$

where $\gamma_j = m\beta_j \left(\tau_j - \tau_j^{(0)}\right)$, and the additional regressor variables are given by $X_{m+p+j} = \left(\tau_j^{(0)} - X\right)_+^{m-1}$, $j = \overline{1,p}$. In this way, we have the sample extended multiple linear regression

$$y_i = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ik} + \sum_{j=1}^{p} \beta_j x_{i,m+j} + \sum_{j=1}^{p} \gamma_j x_{i,m+p+j} + \tilde{\varepsilon}_i,\ i = \overline{1,n}.$$

We denote by

$$\tilde{\delta}^\top = \left(\alpha_0, \ldots, \alpha_m; \beta_1, \ldots, \beta_p; \gamma_1, \ldots, \gamma_p\right), \quad \tilde{\boldsymbol{\varepsilon}}^\top = \left(\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_p\right),$$

and

$$\tilde{\boldsymbol{X}} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1,m+2p} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{n,m+2p} \end{pmatrix},$$

the vector of coefficients, vector of error terms, and design matrix of the sample extended multiple linear regression. Here, for each $i = \overline{1,n}$, we have

$$x_{ik} = x_i^k,\ k = \overline{1,m};$$
$$x_{i,m+j} = \left(\tau_j^{(0)} - x_i\right)_+^m,\ x_{i,m+p+j} = \left(\tau_j^{(0)} - x_i\right)_+^{m-1},\ j = \overline{1,p}.$$

Thus, the sample extended multiple linear regression has the matrix form

$$\boldsymbol{y} = \tilde{\boldsymbol{X}}\,\tilde{\boldsymbol{\delta}} + \tilde{\varepsilon}. \tag{10}$$

Because $\tilde{\varepsilon} = \varepsilon + \mathcal{O}\left(\eta^2\right)$, it results that $E\left(\tilde{\varepsilon}\right) = E\left(\varepsilon\right) + \mathcal{O}\left(\eta^2\right)\boldsymbol{I}_n \approx \boldsymbol{0}$, and $Var\left(\tilde{\varepsilon}\right) = Var\left(\varepsilon + \mathcal{O}\left(\eta^2\right)\boldsymbol{I}_n\right) = Var\left(\varepsilon\right) = \sigma^2\boldsymbol{I}_n$.

The least-squares estimators of the coefficients $\tilde{\boldsymbol{\delta}}$ are given by

$$\tilde{\boldsymbol{d}} = \left(\tilde{a}_0, \ldots, \tilde{a}_m; \tilde{b}_1, \ldots, \tilde{b}_p; \tilde{c}_1, \ldots, \tilde{c}_p\right)^\top = \left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}^\top\boldsymbol{y}.$$

Refering to $\gamma_j = m\beta_j\left(\tau_j - \tau_j^{(0)}\right)$, $j = \overline{1,p}$, we obtain

$$\tau_j = \tau_j^{(0)} + \frac{\gamma_j}{m\beta_j}, \quad j = \overline{1,p},$$

and new estimations of coefficients of the linear model (5) can be calculated, considering the new positions of the knots

$$\tau_j^{(0)} := \tau_j^{(0)} + \frac{\tilde{c}_j}{mb_j},\ j = \overline{1,p}.$$

Note that the estimations b_j, $j = \overline{1,p}$, of the coefficients β_j, $j = \overline{1,p}$, obtained on the linear model (6), generally differ from the estimations $\tilde{b}_j$, $j = \overline{1,p}$, of the coefficients β_j, $j = \overline{1,p}$, obtained on the linear model (10).

It is remarked in [6] that the procedure of Box and Tidwell [3] converges quite rapidly, but the round-off error is potentially a problem and successive values of $\boldsymbol{\tau}$ may oscillate widly unless enough decimal places are carried. Convergence problems may be encountered in cases where the error standard deviation of response variable Y is large or when the range of the regressor variable X is very small compared to its expectation.

Table 1 contains the data generated by using the function ([9], p. 45)

$$f(x) = 4.26\left(\mathrm{e}^{-x} - 4\mathrm{e}^{-2x} + 3\mathrm{e}^{-3x}\right), \quad x \in [0, 3.3]. \tag{11}$$

The values of the dependent variable Y are give by

$$y_i = f(x_i) + \varepsilon_i, \quad i = \overline{1,n},$$

where $x_i = (i-1)/30$, $i = \overline{1,100}$, and ε_i, $i = \overline{1,100}$, are independent random numbers following the normal distribution $\mathcal{N}(0; 0.02)$, i.e. the random vector $\boldsymbol{\varepsilon}^\top = (\varepsilon_1, \ldots, \varepsilon_n)$ has multivariate normal distribution with the mean value $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and martrix of covariance is $Var(\boldsymbol{\varepsilon}) = 0.04\boldsymbol{I}_n$.

Table 2 contains the knots τ_i, estimated coefficients a_i and b_i of fitted spline regressions: linear ($m = 1$) with $p = 1, 2, 3$ knots, quadratic ($m = 2$) with $p = 1, 2$ knots, and cubic ($m = 3$) with one knot. The corresponding sum of residual squares (8), residual mean squares (9), and coefficient of determination (3) for each of the fitted spline regresion are given in the same table.

The procedure to obtain the free knots ends if two successive iterations of knots differ less than $\frac{1}{2}10^{-2}$, else the maximum number of iterations is 500. In the second case, the free knots correspond to the minumum norm difference of two successive iterations of the knots of the spline regression no more than 500 iterations.

Figures 1–6 correspond to the six spline regressions and contain for each of them: plot of fitted spline regression (by continuous line), scatter diagram (by circles), positions of knots (by squares), and plot of generator function (11) (by dashed line).

3. **Confidence intervals**

We are interested in giving confidence intervals on the coefficients $\delta_i = \alpha_i$ or β_i of the multiple linear regression (7). If one assumes that the error term $\boldsymbol{\varepsilon}_0$ is normally distributed $\mathcal{N}\left(\mathbf{0}, \sigma^2\boldsymbol{I}_n\right)$, i.e. (7) is a classical multiple linear model, then

i	y_i	i	y_i	i	y_i	i	y_i	i	y_i
1	−0.087	21	−0.516	41	−0.148	61	0.296	81	0.343
2	−0.590	22	−0.789	42	0.242	62	0.231	82	0.373
3	−0.439	23	−0.326	43	−0.005	63	0.511	83	0.397
4	−0.571	24	−0.092	44	0.503	64	−0.085	84	0.005
5	−0.986	25	−0.505	45	0.072	65	0.372	85	0.241
6	−0.614	26	−0.146	46	0.350	66	0.463	86	0.242
7	−0.683	27	−0.020	47	0.298	67	0.426	87	−0.012
8	−0.973	28	−0.545	48	0.079	68	0.392	88	0.037
9	−0.926	29	−0.471	49	−0.163	69	0.281	89	0.397
10	−0.965	30	−0.027	50	0.265	70	0.404	90	0.150
11	−1.032	31	−0.183	51	0.081	71	0.378	91	0.249
12	−0.833	32	0.072	52	0.410	72	0.209	92	0.185
13	−1.069	33	0.132	53	0.393	73	0.180	93	0.036
14	−0.481	34	0.144	54	0.633	74	0.192	94	0.047
15	−0.905	35	0.290	55	0.415	75	−0.048	95	0.243
16	−0.810	36	0.194	56	0.169	76	0.195	96	−0.040
17	−0.572	37	0.325	57	0.375	77	0.261	97	0.302
18	−0.722	38	−0.130	58	0.097	78	0.295	98	0.256
19	−0.701	39	0.129	59	0.294	79	0.516	99	−0.026
20	−0.795	40	0.123	60	0.288	80	0.153	100	0.081

TABLE 1. Values of the dependent variable Y

each of the statistics

$$t_i = \frac{a_i - \alpha_i}{s_i}, \quad i = \overline{0, m},$$
$$t_{m+i} = \frac{b_i - \beta_i}{s_{m+i}}, \quad i = \overline{1, p},$$

	m=1			m=2		m=3
	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=1$
τ_1	1.715	0.061	0.061	0.061	0.433	0.649
τ_2		1.506	1.360	1.749		
τ_3			2.030			
a_0	0.605	0.433	0.691	−1.119	−0.216	−1.951
a_1	−0.144	−0.082	−0.175	1.260	0.513	2.838
a_2				−0.276	−0.128	−1.133
a_3					0.140	
b_1	−0.888	15.996	15.967	291.236	7.912	6.342
b_2		−0.994	−0.683	0.488		
b_3			−0.408			
$\|\boldsymbol{e}_0\|^2$	4.686	3.797	3.689	4.128	2.904	2.884
s_0^2	0.048	0.040	0.039	0.043	0.031	0.030
$100\,R_0^2$	75.43	80.09	80.66	78.35	84.77	84.87

TABLE 2. Elements of the fitted spline regressions

is T-distributed with $d = n - m - p - 1 = n - r - 1$ degrees of freedom, where

$$s_j^2 = s^2 \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)_{j,j}^{-1}, \quad j = \overline{0, m+p}$$

and $\left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)_{j,j}^{-1}$ denotes the $j+1$-*th* entry of the diagonal of inverse matrix of $\boldsymbol{X}_0^\top \boldsymbol{X}_0$.

Thus, a $100\,(1-\alpha)\,\%$ confidence intervals on the regression coefficients α_i and β_i are given by

$$a_i - t_{d;1-\frac{\alpha}{2}} s_i < \alpha_i < a_i + t_{d;1-\frac{\alpha}{2}} s_i, \quad i = \overline{0, m},$$

$$b_i - t_{d;1-\frac{\alpha}{2}} s_{m+i} < \beta_i < b_i + t_{d;1-\frac{\alpha}{2}} s_{m+i}, \quad i = \overline{1, p},$$

where $t_{d;1-\frac{\alpha}{2}}$ is the $\left(1-\frac{\alpha}{2}\right)$-quantile of the T-distribution with d degrees of freedom. In the Table 3 are given 95% confidence intervals on the coefficients of the six spline regressions having the elements contained in the Table 2.

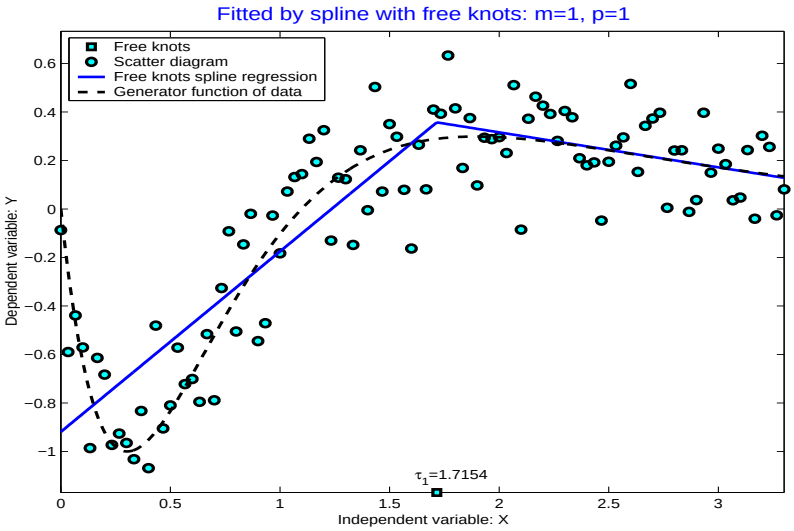


FIGURE 1. Linear spline ($m = 1$) with one knot ($p = 1$)

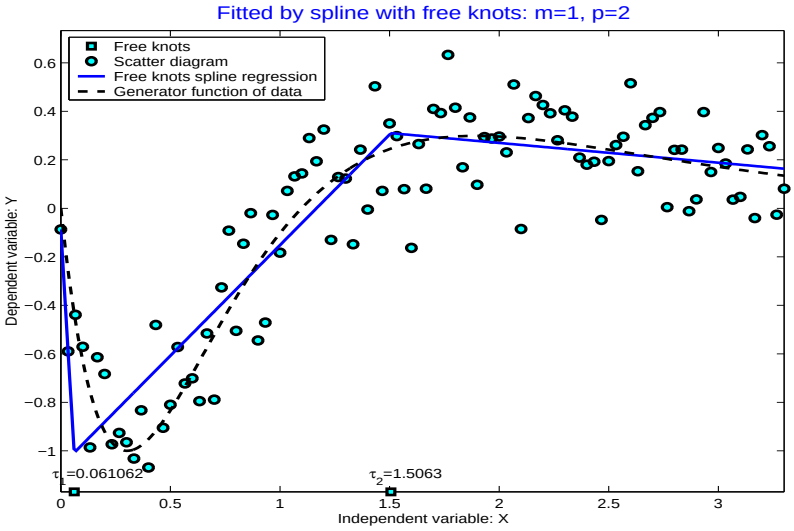


FIGURE 2. Linear spline ($m = 1$) with two knots ($p = 2$)

We have also for the classical multiple linear model (7) that the statistic

$$h^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{d\, s^2}{\sigma^2} = \frac{(n - r - 1)\, s^2}{\sigma^2},$$

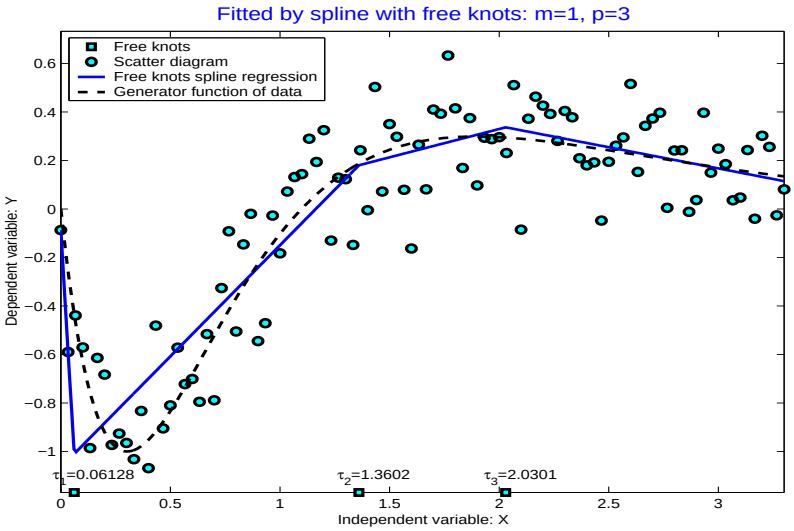


FIGURE 3. Linear spline ($m = 1$) with three knots ($p = 3$)

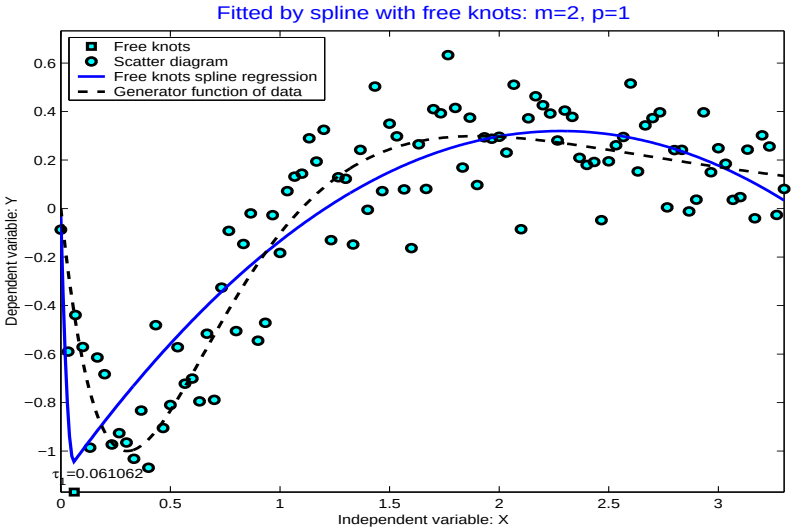


FIGURE 4. Quadratic spline ($m = 2$) with one knot ($p = 1$)

follows a χ^2-distribution with d degrees o freedom. Using this result we have a $100\,(1-\alpha)\,\%$ confidence interval on σ^2

$$\frac{d\,s^2}{\chi^2_{d;1-\frac{\alpha}{2}}} < \sigma^2 < \frac{d\,s^2}{\chi^2_{d;\frac{\alpha}{2}}},$$

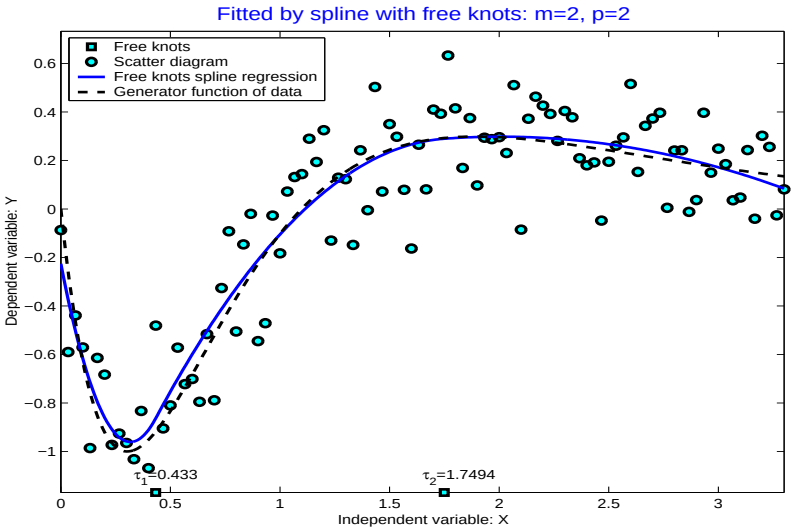


FIGURE 5. Quadratic spline ($m = 2$) with two knots ($p = 2$)

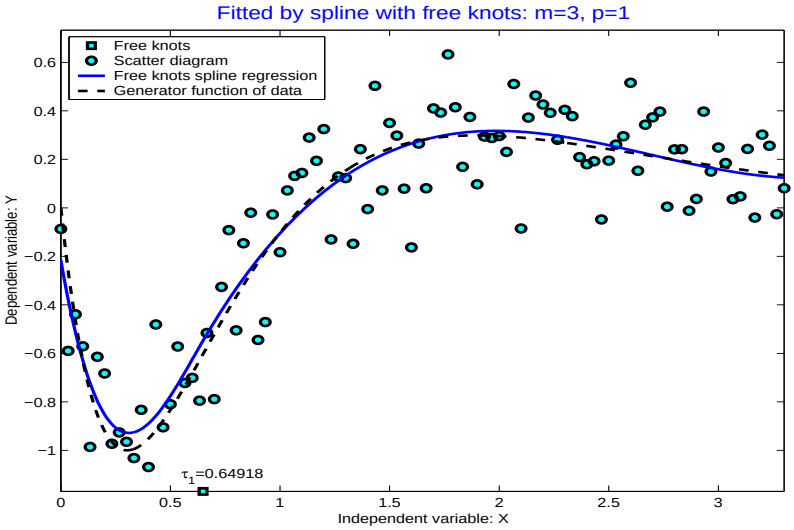


FIGURE 6. Cubic spline ($m = 3$) with one knot ($p = 1$)

where $\chi^2_{d;\gamma}$ denotes the γ-quantile of the χ^2 distribution with d degrees of freedom. The Table 3 contains also 95% confidence intervals on σ^2 of the six examples of spline regressions considered in the previous section.

	m=1		
	$p=1$	$p=2$	$p=3$
α_0	$(0.350, 0.861)$	$(0.238, 0.628)$	$(0.317, 1.066)$
α_1	$(-0.251, -0.038)$	$(-0.165, 0.002)$	$(-0.318, -0.032)$
β_1	$(-1.070, -0.707)$	$(9.696, 22.295)$	$(9.680, 22.253)$
β_2		$(-1.165, -0.822)$	$(-1.015, -0.350)$
β_3			$(-0.745, -0.072)$
σ^2	$(0.0371, 0.0654)$	$(0.0304, 0.0536)$	$(0.0298, 0.0528)$
	m=2		m=3
	$p=1$	$p=2$	$p=1$
α_0	$(-1.248, -0.990)$	$(-0.919, 0.486)$	$(-2.239, -1.663)$
α_1	$(1.083, 1.438)$	$(-0.137, 1.164)$	$(2.212, 3.465)$
α_2	$(-0.328, -0.225)$	$(-0.270, 0.014)$	$(-1.519, -0.747)$
α_3			$(0.070, 0.211)$
β_1	$(176.042, 406.431)$	$(6.066, 9.757)$	$(4.824, 7.860)$
β_2		$(-0.764, -0.211)$	
σ^2	$(0.0330, 0.0583)$	$(0.0234, 0.0415)$	$(0.0233, 0.0412)$

TABLE 3. Confidence intervals for coefficients and variation

From the construction and theoretical results on the multiple linear model (7), it results that an unbiased estimator of the mean response $E(Y \mid \boldsymbol{x})$ at a point $\boldsymbol{x}^\top = \left(1; x, \ldots, x^m; (\tau_1 - x)_+^p, \ldots, (\tau_p - x)_+^p\right)$ is given by

$$\hat{y} = \boldsymbol{x}^\top \boldsymbol{d}_0,$$

and $Var(\hat{y}) = \sigma^2 \boldsymbol{x} \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}$. For the classical multiple linear model, the statistic

$$t_{\boldsymbol{x}} = \frac{\hat{y} - E(Y \mid \boldsymbol{x})}{s\sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}} = \frac{\boldsymbol{x}^\top \boldsymbol{d}_0 - \boldsymbol{x}^\top \boldsymbol{\delta}}{s\sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}} \tag{12}$$

is T-distributed with d degrees of freedom. Using the statistic $t_{\boldsymbol{x}}$, the following $100\,(1-\alpha)\,\%$ confidence interval on the mean response $E\,(Y \mid \boldsymbol{x})$ can be obtained

$$\hat{y} - t_{d;1-\frac{\alpha}{2}} s \sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}} < E\,(Y \mid \boldsymbol{x}) < \hat{y} + t_{d;1-\frac{\alpha}{2}} s \sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}.$$

In a similar manner, to construct a confidence interval for a predicted value y of the response Y, corresponding to a new value x of the regressor X, we have the statistic

$$t_{\boldsymbol{x}} = \frac{\hat{y} - y}{s\sqrt{1 + \boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}} = \frac{\boldsymbol{x}^\top \boldsymbol{d}_0 - y}{s\sqrt{1 + \boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}}, \tag{13}$$

where again $\boldsymbol{x}^\top = \left(1; x, \ldots, x^m; (\tau_1 - x)_+^p, \ldots, (\tau_p - x)_+^p\right)$, and $t_{\boldsymbol{x}}$ is T-distributed with d degrees of freedom. Thus, a $100\,(1-\alpha)\,\%$ confidence interval on the predicted response y is given by

$$\hat{y} - t_{d;1-\frac{\alpha}{2}} s \sqrt{1 + \boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}} < y < \hat{y} + t_{d;1-\frac{\alpha}{2}} s \sqrt{1 + \boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}}.$$

We remark that $\boldsymbol{\tau}$ is the last position $\boldsymbol{\tau}_0$ obtained by iterative procedure and s^2 is the corresponding residual sum of square given by (9).

Figures (7), (8) and (9) contain plots of 95% confidence intervals on the mean response and prediction with respect to x for three of the six considered spline regressions. Each figure contains scatter diagram (circles), plot of spline regression function (solid line), plot of confidence interval on mean response (dash-dot line), plot of confidence interval on prediction (dashed line), and positions of the knots (squares).

The construction of simultaneous confidence intervals on the mean response and prediction uses the Scheffé's result. Namely, if C represents a set of points $\boldsymbol{x}^\top = \left(1; x, \ldots, x^m; (\tau_1 - x)_+^p, \ldots, (\tau_p - x)_+^p\right)$, and considering $W = \sup_{\boldsymbol{x} \in C} t_{\boldsymbol{x}}^2$, were $t_{\boldsymbol{x}}^2$ is given by (12) and (13) respectively, then the statistic $W/\,(m+p+1)$ is F-distributed with $(m+p+1, n-m-p-1) = (r+1, d)$ degrees of freedom.

In this way, we have a $100\,(1-\alpha)\,\%$ Scheffé simultaneous confidence interval on the mean response

$$\hat{y} - Ks \sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}} < E\,(Y \mid \boldsymbol{x}) < \hat{y} + Ks \sqrt{\boldsymbol{x}^\top \left(\boldsymbol{X}_0^\top \boldsymbol{X}_0\right)^{-1} \boldsymbol{x}},$$

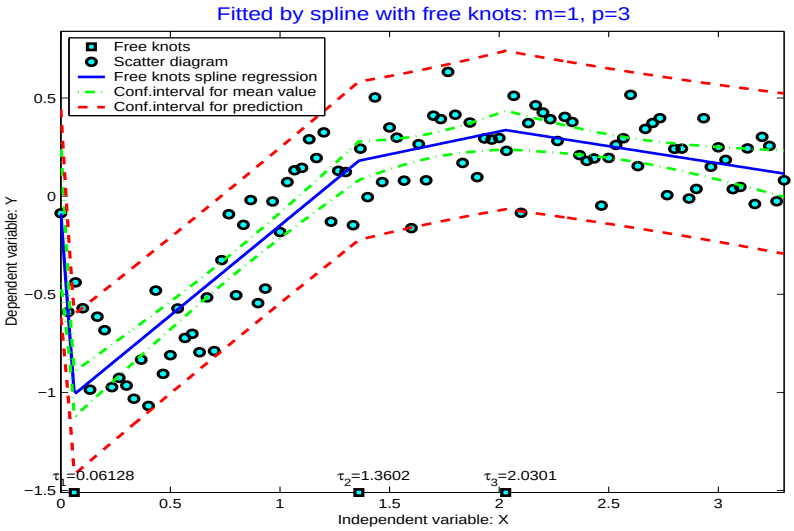


FIGURE 7. Linear spline ($m = 1$) with three knots ($p = 3$)

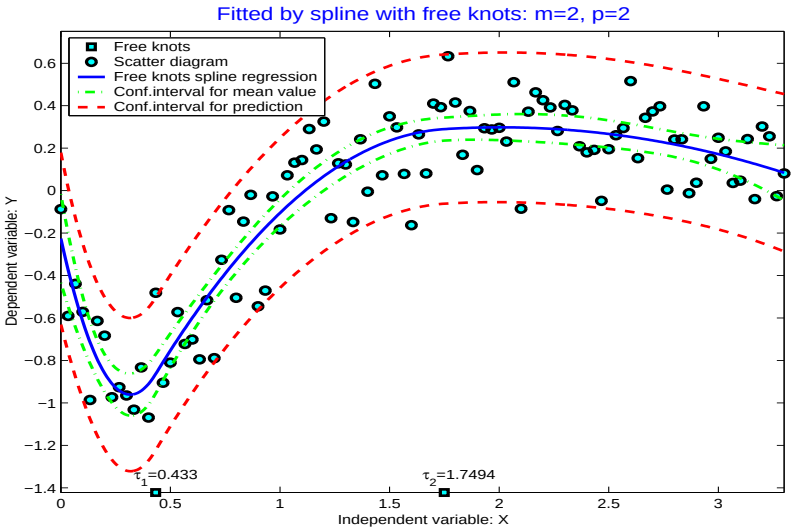


FIGURE 8. Quadratic spline ($m = 2$) with two knots ($p = 2$)

and $100\,(1-\alpha)\,\%$ Scheffé simultaneous confidence interval on the prediction

$$\hat{y} - Ks\sqrt{1 + \boldsymbol{x}^{\top}\left(\boldsymbol{X}_0^{\top}\boldsymbol{X}_0\right)^{-1}\boldsymbol{x}} < y < \hat{y} + Ks\sqrt{1 + \boldsymbol{x}^{\top}\left(\boldsymbol{X}_0^{\top}\boldsymbol{X}_0\right)^{-1}\boldsymbol{x}}.$$

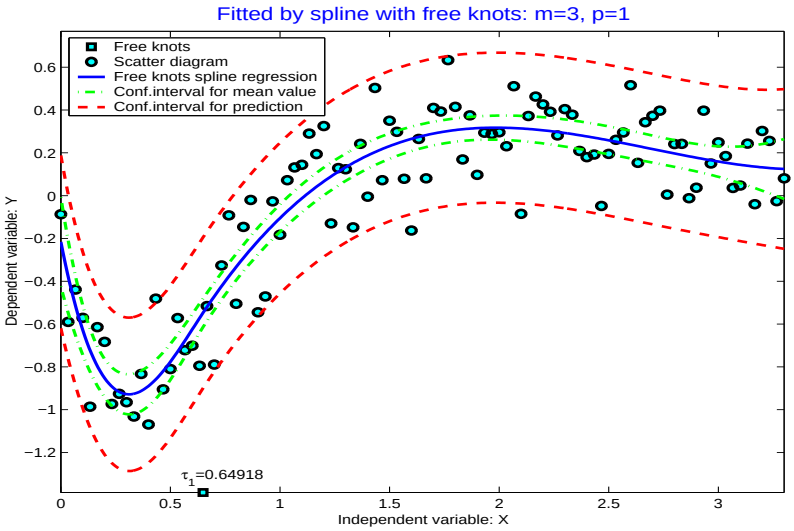


FIGURE 9. Cubic spline ($m = 3$) with one knot ($p = 1$)

Here $K = \sqrt{(r+1)\, f_{r+1,d;1-\alpha}}$, where $f_{r+1,d;1-\alpha}$ is $1-\alpha$-quantile of the F distribution with $(r+1, d)$ degrees of freedom.

Figures (10), (11) and (12) contain plots of 95% simultaneous confidence intervals on the mean response and prediction with respect to x for three of the six considered spline regressions. Each figure contains scatter diagram (circles), plot of spline regression function (solid line), plot of simultaneous confidence interval on mean response (dash-dot line), plot of simultaneous confidence interval on prediction (dashed line), and positions of knots (squares).

References

[1] Agratini, O., Blaga, P., Coman, Gh., *Lectures on Wavelets, Numerical Methods and Statistics*, Science Book House, Cluj-Napoca, 2005.

[2] Blaga, P. P., *Free knots for spline regression*, Annals of Tiberiu Popoviciu Seminar of Functional Equations, Approximation and Convexity, **3**(2005), 3–17.

[3] Box, G. E. P., Tidwell, W., *Transformation of the independent variables*, Technometrics, **4**(1962), 531–550.

[4] Eubank, R. L., *Spline smoothing and nonparametric regression*. Dekker, New York, 1988.

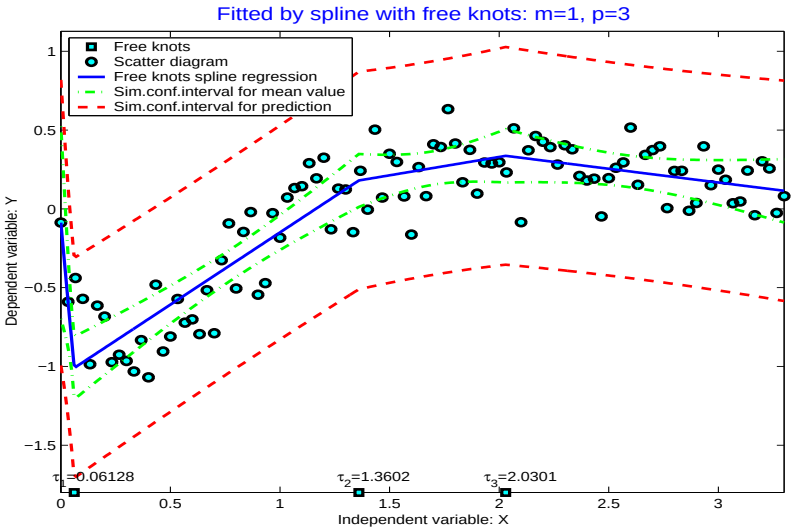


FIGURE 10. Linear spline ($m = 1$) with three knots ($p = 3$)

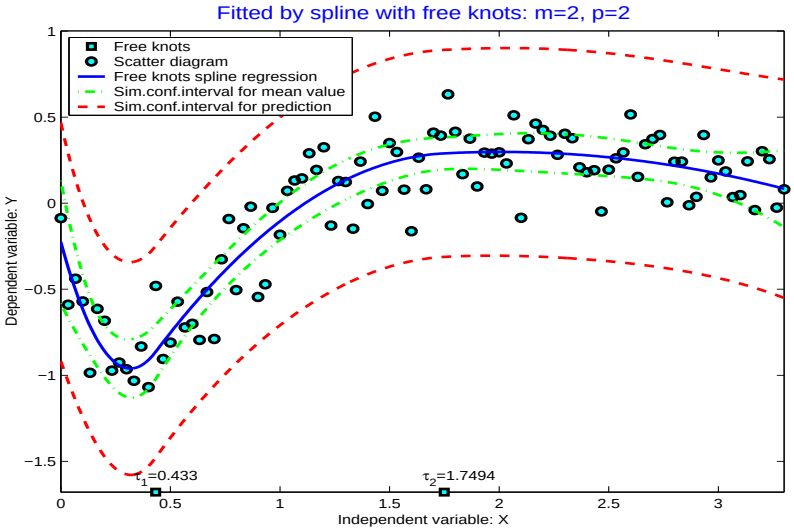


FIGURE 11. Quadratic spline ($m = 2$) with two knots ($p = 2$)

[5] Lehmann, E. L. *Testing statistical hypotheses (Second edition)*, Springer, New York-Berlin, 1997.

[6] Montgomery, D. C., Peck, E. A., Vining, G. G., *Introduction to linear regression analysis (Third edition)*, John Wiley & Sons, New York, 2001.

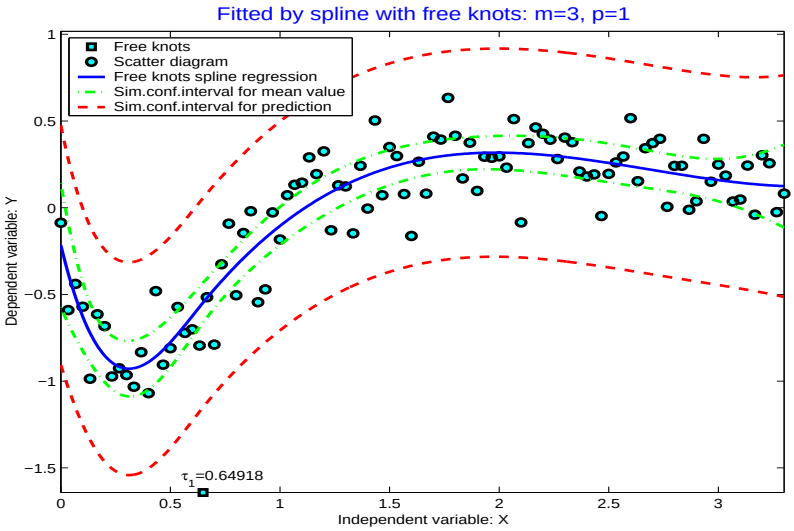


FIGURE 12. Cubic spline ($m = 3$) with one knot ($p = 1$)

[7] Scheffé, H., *The analysis of variance*, Wiley, New York, 1959.

[8] Stapleton, J. H., *Linear statistical models*, John Wiley & Sons, New York-Chichester-Brisbane, 1995.

[9] Wahba, W., *Spline models for observational data*, Society for Industrial and Applied Mathematics, Philadelphia, 1990.

[10] Wahba, G., Wold, S., *A completely automatic French curve*, Commun. Statist., **4**(1975), 1–17.

[11] Wold, S., *Spline functions in data analysis*, Technometrics, **16**(1974), 1–11.

"BABEŞ–BOLYAI" UNIVERSITY, CLUJ–NAPOCA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
STR. KOGĂLNICEANU 1, 400084 CLUJ–NAPOCA, ROMANIA
E-mail address: blaga@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

A COMBINED METHOD FOR INTERPOLATION OF SCATTERED DATA BASED ON TRIANGULATION AND LAGRANGE INTERPOLATION

TEODORA CĂTINAŞ

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. We introduce a combined method for interpolation on scattered data sets in the plane, based on the triangulation method introduced by R. Franke and G. Nielson in [10].

1. Introduction

A certain method based on triangulation, introduced by R. Franke and G. Nielson in [10], and the Shepard method, introduced in [15], are superior to other methods used in interpolation of very large scattered data sets. Both methods have very comparable fitting capabilities, but there exist situations in which one or the other may be preferable [10].

We present first the original method based on triangulation, introduced in [10].

Let f be a real-valued function defined on $X \subset \mathbb{R}^2$, and $V_i(x_i, y_i) \in X$, $i = 1, ..., N$ some distinct points. Denote by $r_i(x, y)$, the distances between a given point $(x, y) \in X$ and the points $V_i(x_i, y_i)$, $i = 1, ..., N$. The interpolant with regard to the data $V_i(x_i, y_i)$, $i = 1, ..., N$, is of the form

$$(Pf)(x, y) = \sum_{i=1}^{N} W_i(x, y) P_i(x, y), \tag{1}$$

Received by the editors: 28.06.2006.

2000 *Mathematics Subject Classification.* 41A05, 41A63.

Key words and phrases. Triangulation, scattered data, Lagrange operator.

This work has been supported by grant MEdC-ANCS no. 3233/17.10.2005.

where W_i, $i = 1, ..., N$ are the weight functions and P_i, $i = 1, ..., N$ are some local interpolation operators. We have

$$W_i(x_j, y_j) = \delta_{ij}, \quad i, j = 1, ..., N. \tag{2}$$

The method of defining the weight functions W_i, $i = 1, ..., N$ given in [10] requires the triangulation of the data $V_i(x_i, y_i)$, $i = 1, ..., N$. Each W_i will be a globally defined C^1 function with support $S_i = \cup_{jkl \in M_i} T_{jkl}$, where T_{jkl} denotes the triangle with vertices V_j, V_k, V_l and $M_i = \{jkl : T_{jkl}$ is a triangle with vertex $V_i\}$. For $(x, y) \in T_{jkl} \subset S_i$, the weight functions have the form [10]:

$$\begin{aligned} &W_i(x, y) = \\ &= b_i^2(3 - 2b_i) + 3\frac{b_i^2 b_j b_k}{b_i b_j + b_i b_k + b_j b_k} \left[b_j \frac{\|e_i\|^2 + \|e_k\|^2 - \|e_j\|^2}{\|e_k\|^2} + b_k \frac{\|e_i\|^2 + \|e_j\|^2 - \|e_k\|^2}{\|e_j\|^2} \right], \end{aligned} \tag{3}$$

where b_i, b_j, b_k are the barycentric coordinates of (x, y) with respect to the triangle T_{jkl} and $\|e_p\|$, $p = i, j, k$ represent the length of the edge opposite to V_p, $p = i, j, k$. The barycentric coordinates are given by the equations:

$$\begin{aligned} x &= b_i x_i + b_j x_j + b_k x_k, \\ y &= b_i y_i + b_j y_j + b_k y_k, \\ 1 &= b_i + b_j + b_k. \end{aligned}$$

For an arbitrary triangle T_{jkl} the only weights which are nonzero are W_i, W_j and W_k, and therefore the interpolant (1) becomes

$$(Pf)(x, y) = W_i(x, y)P_i(x, y) + W_j(x, y)P_j(x, y) + W_k(x, y)P_k(x, y). \tag{4}$$

We note that

$$W_i + W_j + W_k = 1. \tag{5}$$

As local interpolation operators P_i, $i = 1, ..., N$ R. Franke and G. Nielson considered the solution of the inverse distance weighted least squares problem at $(x, y) = (x_i, y_i)$,

i.e.,

$$P_i(x,y) = f_i + \bar{a}_{i2}(x-x_i) + \bar{a}_{i3}(y-y_i) + \bar{a}_{i4}(x-x_i)^2 + \bar{a}_{i5}(x-x_i)(y-y_i) + \bar{a}_{i6}(y-y_i)^2,$$

where $f_i = f(x_i, y_i)$. The coefficients $\bar{a}_{ik}$, $k = 2, ..., 6$ are the solutions of the system [10]

$$\min_{a_{i2},\ldots,a_{i6}} \sum_{\substack{k-1 \\ k\neq i}}^{N} \left[\frac{f_i + a_{i2}(x_k - x_i) + ... + a_{i6}(y_k - y_i)^2 - f_k}{\rho_k(x_i, y_i)} \right]^2,$$

with ρ_i given by

$$\frac{1}{\rho_i} = \frac{(R_i - r_i)_+}{R_i r_i},$$

with

$$z_+ = \begin{cases} z, & z > 0 \\ 0, & z \leq 0, \end{cases}$$

and R_i is a radius of influence about the node (x_i, y_i) and it is varying with i. The proper choice of the radius is critical to the success of the method [10]. This is taken as the distance from node i to the jth closest node to (x_i, y_i) for $j > N_w$ (N_w is a fixed value) and j as small as possible within the constraint that the jth closest node is significantly more distant than the $(j-1)$st closest node (see, e.g. [14]).

2. Main results

In this section we consider a new combined interpolation operator. It is obtained using the Lagrange interpolation operator as a local interpolation operator.

Let Λ be the set of Lagrange type information,

$$\Lambda := \Lambda_L = \{\lambda_i : \lambda_i(f) = f(x_i, y_i),\ i = 1, \ldots, N\}. \tag{6}$$

Let L_i be the bivariate Lagrange operators of degree n (associated to the node (x_i, y_i)), $i = 1, \ldots, N$, (see, e.g., [6]), that interpolates the function f, respectively, at the sets of points

$$Z_{m,i} := \{z_i, z_{i+1}, \ldots, z_{i+m-1}\}, \quad i = 1, \ldots, N,\ m < N, \tag{7}$$

with $z_{N+i} = z_i$, $i = 1, \ldots, m-1$ and $m := (n+1)(n+2)/2$ is the number of the coefficients of a bivariate polynomial of the degree n, $\sum_{i+j \leq n} a_{ij} x^i y^j$.

Remark 1. [6] *For given N, it can be considered only operators L_i^n with n such that $(n+1)(n+2)/2 < N$, i.e., for $n \in \{1, \ldots, \nu\}$, where $\nu = \text{integer}[(\sqrt{8N+1} - 3)/2]$.*

The existence and the uniqueness of the operators L_i^n are assured by the following theorem.

Theorem 2. [1] *Let $z_i := (x_i, y_i)$, $i = 1, \ldots, (n+1)(n+2)/2$ be different points in plane that do not lie on the same algebraic curve of n-th degree. Then, for every function f defined at the points z_i, $i = 1, \ldots, (n+1)(n+2)/2$ there exists a unique polynomial of degree n that interpolates f at z_i, $i = 1, \ldots, (n+1)(n+2)/2$.*

Hence, if the points z_k, $k = i, \ldots, i+m-1$ of the set (7) do not lie on an algebraic curve of n-th degree for all $i = 1, \ldots, N$ then L_i^n exists and it is unique for all $i = 1, \ldots, N$.

In what follows we suppose that the existence and the uniqueness conditions of the operators L_i^n, $i = 1, \ldots, N$, are satisfied.

We have

$$(L_i^n f)(x, y) = \sum_{k=i}^{i+m-1} l_k(x, y) f(x_k, y_k), \quad i = 1, \ldots, N,$$

where l_k are the corresponding cardinal polynomials:

$$l_k(x_j, y_j) = \delta_{kj}, \ k, j = i, \ldots, i+m-1.$$

The operators L_i^n have the following interpolatory properties:

$$(L_i^n f)(x_k, y_k) = f(x_k, y_k), \quad k = i, \ldots, i+m-1 \tag{8}$$

and the degree of exactness is

$$\text{dex}(L_i^n) = n, \tag{9}$$

for all $i = 1, \ldots, N$.

Next we use these Lagrange polynomials as local interpolation operators P_i, $i = 1, ..., N$, in (4). In this way we obtain a new interpolant of the data $V_i(x_i, y_i)$, $i =$

$1, ..., N$, with respect to the Lagrange type information, namely:

$$(Pf)(x,y) = W_i(x,y)(L_i^n f)(x,y) + W_j(x,y)(L_j^n f)(x,y) + W_k(x,y)(L_k^n f)(x,y), \quad (10)$$

with W_i, W_j, W_k given by (3).

Remark 3. *As the Lagrange operator is linear then the combined operator P is also linear.*

Theorem 4. *The combined operator P has the following interpolation properties:*

$$(Pf)(x_p, y_p) = f(x_p, y_p), \ p = 1, \ldots, N. \quad (11)$$

Proof. We have

$$\begin{aligned}(Pf)(x_p, y_p) =& W_i(x_p, y_p)(L_i^n f)(x_p, y_p) + W_j(x_p, y_p)(L_j^n f)(x_p, y_p) \\ &+ W_k(x_p, y_p)(L_k^n f)(x_p, y_p).\end{aligned}$$

Taking into account (2) and the interpolation properties of the Lagrange operators the conclusion follows. □

Theorem 5. *The degree of exactness of the combined operator P is*

$$\operatorname{dex}(P) = n \quad (12)$$

Proof. We have

$$(L_h^n e_{pq})(x,y) = e_{pq}(x,y), \ p+q \leq n, \ h = i,j,k$$

where $e_{pq}(x,y) = x^p y^q$. We have

$$\begin{aligned}(Pe_{pq})(x,y) &= W_i(x,y)e_{pq}(x,y) + W_j(x,y)e_{pq}(x,y) + W_k(x,y)e_{pq}(x,y) \\ &= e_{pq}(x,y)(W_i(x,y) + W_j(x,y) + W_k(x,y))\end{aligned}$$

and taking into account (5) it follows that

$$(Pe_{pq})(x,y) = e_{pq}(x,y), \quad p+q \leq n,$$

But $\operatorname{dex}(L_h^n) = n$ so it means that there exists a $(p,q) \in \mathbb{N}^2$ with $p+q = n+1$ such that $L_h^n e_{pq} \neq e_{pq}$, which implies that $Pe_{pq} \neq e_{pq}$. So the conclusion follows. □

The new obtained operator presents a higher degree of exactness compared to that of the Shepard operator (introduced in [15]), which is usually used in this kind of interpolation problems.

Remark 6. *For the particular case $n = 1$ we have*

$$(L_i^1 f)(x,y) = l_i(x,y)f(x_i,y_i) + l_{i+1}(x,y)f(x_{i+1},y_{i+1}) + l_{i+2}(x,y)f(x_{i+2},y_{i+2}) \quad (13)$$

for $i = 1,\ldots,N$. We have

$$l_i(x,y) = \frac{(y_{i+1}-y_{i+2})x+(x_{i+2}-x_{i+1})y+x_{i+1}y_{i+2}-x_{i+2}y_{i+1}}{(x_i-x_{i+1})(y_{i+1}-y_{i+2})-(x_{i+1}-x_{i+2})(y_i-y_{i+1})}$$

$$l_{i+1}(x,y) = \frac{(y_{i+2}-y_i)x+(x_i-x_{i+2})y+x_{i+2}y_i-x_iy_{i+2}}{(x_{i+1}-x_{i+2})(y_{i+2}-y_i)-(x_{i+2}-x_i)(y_{i+1}-y_{i+2})}$$

$$l_{i+2}(x,y) = \frac{(y_i-y_{i+1})x+(x_{i+1}-x_i)y+x_iy_{i+1}-x_{i+1}y_i}{(x_{i+2}-x_i)(y_i-y_{i+1})-(x_i-x_{i+1})(y_{i+2}-y_i)}.$$

In this case the new interpolant is of the form:

$$(Pf)(x,y) = W_i(x,y)(L_i^1 f)(x,y) + W_j(x,y)(L_j^1 f)(x,y) + W_k(x,y)(L_k^1 f)(x,y),$$

with W_i, W_j, W_k given by (3).

Remark 7. *The existence and uniqueness condition of L_i^1 is that the points (x_i,y_i), (x_{i+1},y_{i+1}), (x_{i+2},y_{i+2}) do not lie on a line $Ax+By+C=0$.*

The steps of the algorithm of computation of the previously obtained interpolant (10) are:

1. Form a triangulation of the points $V_i(x_i,y_i)$, $i = 1,...,M$.
2. Given a point (x,y) determine the triangle V_{i_0,j_0,k_0}, containing (x,y).
3. Compute the Lagrange interpolation polynomials, L_i^n, $i = i_0,j_0,k_0$, given by

$$(L_i^n f)(x,y) = \sum_{k=i}^{i+m-1} l_k(x,y)f(x_k,y_k), \quad i = i_0,j_0,k_0.$$

4. Compute the weights functions.
5. Compute the interpolant given by (10).

3. Test results

We consider several of the generally used test functions, [13], [14], [9]:

Gentle $f_1(x,y) = \exp\big[-\frac{81}{16}((x-0.5)^2+(y-0.5)^2)\big]/3,$

Saddle $f_2(x,y) = \big[1.25+\cos(5.4y)\big]/\big[6+6(3x-1)^2\big],$

Steep $f_3(x,y) = \exp\big[-\frac{81}{4}((x-0.5)^2+(y-0.5)^2)\big]/3,$

Cliff $f_4(x,y) = \big[\tanh(9y-9x)+1\big]/9.$

The following table contains mean errors for interpolation by

$$(Pf_l)(x,y) = W_i(x,y)(L_i^1 f_l)(x,y) + W_j(x,y)(L_j^1 f_l)(x,y) + W_k(x,y)(L_k^1 f_l)(x,y),$$

for $l = 1,...,4$, with W_i, W_j, W_k given by (3), L_i^1, L_j^1, L_k^1 of the form (13) and considering the consecutive interpolation nodes as the vertices of the triangle. We took 100 random generated nodes in the square $[-1,1]\times[-1,1]$. In Figures 1–4 we plot the graphics of f_i and Pf_i, $i = 1,\dots 4$.

Function	Interpolation error
f_1	0.0210
f_2	0.0340
f_3	0.0073
f_4	0.0498

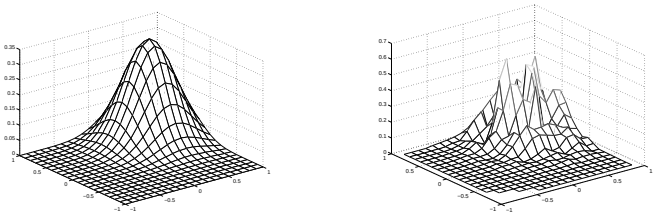


FIGURE 1. Function f_1 and interpolant Pf_1.

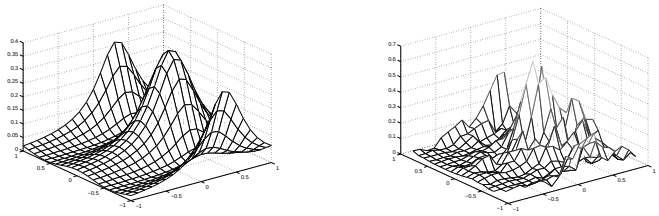


FIGURE 2. Function f_2 and interpolant Pf_2.

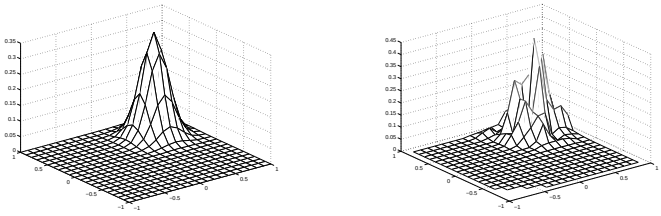


FIGURE 3. Function f_3 and interpolant Pf_3.

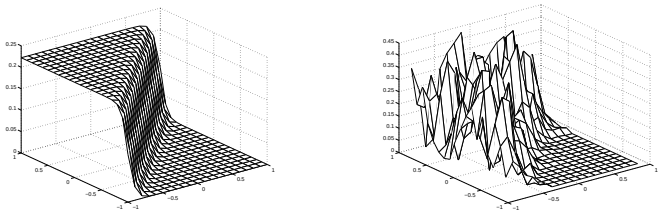


FIGURE 4. Function f_4 and interpolant Pf_4.

References

[1] Brezinski, C., *The generalizations of Newton's interpolation formula due to Mühlbach and Andoyer,* Electronic Transactions on Numerical Analysis, **2** (1994), pp. 130–137.

[2] Cătinaş, T., *The combined Shepard-Lidstone bivariate operator,* Trends and Applications in Constructive Approximation, (Eds. M.G. de Bruin, D.H. Mache, J. Szabados), International Series of Numerical Mathematics, Vol. **151**, 2005, Springer Group-Birkhäuser Verlag, pp. 77-89.

[3] Cătinaş, T., *Bounds for the remainder in the bivariate Shepard interpolation of Lidstone type,* Rev. Anal. Numér. Théor. Approx., **34** (2005), no. 1, pp. 47-53.

[4] Cătinaş, T., *Bivariate interpolation by combined Shepard operators,* Proceeding of 17^{th} IMACS World Congress, Scientific Computation, Applied Mathematics and Simulation, Paris, July 11-15, 2005, ISBN 2-915913-02-1.

[5] Coman, Gh., Cătinaş, T., Birou, M., Oprişan, A., Oşan, C., Pop, I., Somogyi, I., Todea, I., *Interpolation Operators*, Ed. Casa Cărţii de Stiinţă, Cluj Napoca, 2004.

[6] Coman, Gh., Trîmbiţaş, R., *Shepard operators of Lagrange-type*, Studia Univ. "Babeş-Bolyai", Mathematica, **42** (1997), pp. 75–83.

[7] Coman, Gh., Trîmbiţaş, R., *Bivariate Shepard interpolation,* Seminar on Numerical and Statistical Calculus, 1999, pp. 41–83.

[8] Coman, Gh., Ţâmbulea, L., *On some interpolation of scattered data*, Studia Univ. "Babeş-Bolyai", **35** (1990), no. 2, pp. 90–98.

[9] Franke, R., *Scattered data interpolation: tests of some methods,* Math. Comp., **38** (1982), no. 157, pp. 181–200.

[10] Franke, R., Nielson, G., *Smooth interpolation of large sets of scattered data*, Int. J. Numer. Meths. Engrg., **15** (1980), pp. 1691–1704.

[11] Lazzaro, D., Montefusco, L.B., *Radial basis functions for multivariate interpolation of large scattered data sets*, J. Comput. Appl. Math., **140** (2002), pp. 521–536.

[12] Micchelli, C.A., *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx., **2** (1986), no. 1, pp. 11–22.

[13] Renka, R.J., Cline, A.K., *A triangle-based C^1 interpolation method*, Rocky Mountain J. Math., **14** (1984), no. 1, pp. 223–237.

[14] Renka, R.J., *Multivariate interpolation of large sets of scattered data*, ACM Trans. Math. Software, **14** (1988), no. 2, pp. 139-148.

[15] Shepard, D., *A two dimensional interpolation function for irregularly spaced data,* Proc. 23rd Nat. Conf. ACM (1968), pp. 517-523.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: tcatinas@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

A CYCLIC ODD-EVEN REDUCTION TECHNIQUE APPLIED TO A PARALLEL EVALUATION OF AN EXPLICITE SCHEME IN MATHEMATICAL FINANCE

IOANA CHIOREAN

Dedicated to Professor Gheorghe Coman at his 70^{th} *anniversary*

Abstract. The purpose of this paper is to give a possibility of reducing the execution time involved in evaluating a financial option by means of an explicite scheme, using a cyclic odd-even reduction technique.

1. Introduction

The concept of arbitrage is largely used in the domain of mathematical finance. It allows us to establish precise relationships between prices and thence to determine them. Connected with it, the strategies of an option is very important. In the literature, the celebrated Black-Scholes differential equation for the price of the so-called European vanilla option is the best known and used.

Many papers study this equation and indicate different numerical methods in order to get the approximate solution. E.g., in [3], the finite difference method is presented. In [4], the method of radial basis is used, to avoid the mesh of discretized points.

In this paper, considering the idea given by the cyclic odd-even reduction (see [1]), we start from an explicit scheme obtained by means of finite differences, and give an alternative of evaluating of the approximate values, using a cyclic odd-even reduction type technique, which generates a logarithmic time of execution.

Received by the editors: 20.04.2006.

2000 *Mathematics Subject Classification.* 65H05, 65N06, 91B28.

Key words and phrases. parallel calculus, finite difference, Black-Scholes equation.

2. Recalling the Black-Scholes equation

As in [3], denoting by:

- V, the value of an option
- S, the current value of the underlying asset
- t, the time
- σ, the volatility of the underlying asset
- T, the expiry
- r, the interest rate
- E, the exercise price,

We get the Black-Scholes equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \cdot s^2 \cdot \frac{\partial^2 V}{\partial s^2} + r \cdot s \cdot \frac{\partial V}{\partial s} - r \cdot V = 0 \tag{1}$$

with the boundary conditions:

$$C(0,t) = 0$$

$$C(S,T) = \max(S - E, 0).$$

3. The finite difference methods

Finite difference methods (see [2]) are a means of obtaining numerical solutions to partial differential equations (see [2], [3]). They constitute a very powerful and flexible technique and, if applied correctly, are capable of generating accurate numerical solutions to all of the mathematical finance models, also for the Black-Scholes equation (1).

So, considering a mesh of equal S-steps of size δS and equal time-steps of size δt, with $(N+1)^2$ points, central differences for S derivatives and backward differences for time derivatives, we get the explicit discretization of the Black-Scholes equation:

$$\begin{gathered} B_0 V_n^m + C_0 V_1^m = V_0^m \\ A_n V_{n-1}^m + B_n V_n^m + C_n V_{n+1}^m = V_n^{m+1}, \quad n = 1, 2, \ldots, N \end{gathered} \tag{2}$$

where m indicates the moment of time,

$$A_n = -\frac{1}{2}(\sigma^2 n^2 - (r - s_0)n)\delta t$$

$$B_n = 1 + (\sigma^2 n^2 + r)\delta t$$

$$C_n = -\frac{1}{2}(\sigma^2 n^2 + (r - s_0)n)\delta t$$

4. **A technique of the cyclic odd-even reduction type**

Relation (2) generates a system of equations of the following form:

$$\begin{bmatrix} B_0 & C_0 & 0 & \dots & 0 \\ A_1 & B_1 & C_1 & \dots & 0 \\ 0 & A_2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & C_N \\ 0 & 0 & \dots & A_N & B_N \end{bmatrix} \cdot \begin{bmatrix} V_0^m \\ V_1^m \\ \vdots \\ V_{N-1}^m \\ V_N^m \end{bmatrix} = \begin{bmatrix} V_0^{m+1} \\ V_1^{m+1} \\ \vdots \\ V_{N-1}^{m+1} \\ V_N^{m+1} \end{bmatrix} \tag{3}$$

which involves a time of execution of $O(N^2)$. Then the following theorem holds:

Theorem 1. *Using a cyclic odd-even reduction type technique, equation (2) can be computed in a time $O(N[\log_2 m])$ time.*

Proof. Rewriting (2),

$$A_n V_{n-1}^m + B_n V_n^m + C_n V_{n+1}^m = V_n^{m+1}$$

for one single value n, and replacing V_n^m using the same connection among values, we get:

$$A_n V_{n-1}^m + B_n(A_n V_{n-1}^{m-1} + B_n V_n^{m-1} + C_n V_{n+1}^{m-1}) + C_n V_{n+1}^m = V_n^{m+1}$$

or, making some computations:

$$A_n(V_{n-1}^m + B_n V_{n-1}^{m-1}) + B_n^2 V_n^{m-1} + C_n(B_n V_{n+1}^{m-1} + V_{n+1}^m) = V_n^{m+1}.$$

So, for n given, value V_n^{m+1} can be computed by means of values from two previous moments of time. Repeating the same substitution, we finally get:

$$A_n(a_m V_{n-1}^m + a_{m-1} V_{n-1}^{m-1} + \cdots + a_0 V_{n-1}^0) + B_n^m V_n^0$$

$$+C_n(a_m^1 V_{n-1}^m + a_{m-1}^1 V_{n-1}^{m-1} + \cdots + a_0^1 V_{n-1}^0) = V_n^{m+1} \tag{4}$$

where we denoted by a_i and a_i^1, $i = \overline{0, m}$ the final coefficients.

Using the double recursive technique (see [1]), in $[\log_2 m]$ parallel steps, the values in parenthesis are computed.

Finally, for $n = 1, 2, \ldots, N$, the total execution time is $O(N[\log_2 m])$. □

References

[1] Chiorean, I., *Calcul paralel*, Ed. Microinformatica, Cluj, 1994.

[2] Coman, Gh., *Analiză numerică*, Ed. Libris, Cluj, 1994.

[3] Wilmott, P., Howison, S., Dewynne, J., *The Mathematics of Financial Derivatives*, Cambridge Univ. Press, 1995.

[4] Koc, M.B., Boztsun, I., Boztsun, D., *On the Numerical Solution of Black-Scholes Equation*, International Workshop on Mesh Free Methods, 2003.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: ioana@cs.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

COMPACT OPERATORS ON SPACES WITH ASYMMETRIC NORM

S. COBZAŞ

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. The aim of the present paper is to define compact operators on asymmetric normed spaces and to study some of their properties. The dual of a bounded linear operator is defined and a Schauder type theorem is proved within this framework. The paper contains also a short discussion on various completeness notions for quasi-metric and for quasi-uniform spaces.

1. Introduction

An asymmetric norm on a real vector space X is a functional $p : X \to [0, \infty)$ satisfying the conditions

$$\text{(AN1) } p(x) = p(-x) = 0 \Rightarrow x = 0; \quad \text{(AN2) } p(\alpha x) = \alpha p(x);$$

$$\text{(AN3) } p(x + y) \leq p(x) + p(y),$$

for all $x, y \in X$ and $\alpha \geq 0$. A quasi-metric on a set X is a mapping $\rho : X \times X \to [0, \infty)$ satisfying the conditions

$$\text{(QM1) } \rho(x, y) = \rho(y, x) = 0 \iff x = y; \quad \text{(QM2) } \rho(x, z) \leq \rho(x, y) + \rho(y, z),$$

for all $x, y, z \in X$. If the mapping ρ satisfies only the conditions $\rho(x, x) = 0$, $x \in X$, and (QM2), then it is called a *quasi-pseudometric*. If p is an asymmetric norm on a vector space X, then the pair (X, p) is called an asymmetric normed space. Similarly,

Received by the editors: 28.04.2006.

2000 *Mathematics Subject Classification.* Primary: 46B28; Secondary: 47A05, 46B07, 54E15, 54E25.

Key words and phrases. Quasi-metric spaces, quasi-uniform spaces, spaces with asymmetric norm, compact operators, the conjugate operator, Schauder compactness theorem.

(X, ρ) is called a quasi-metric space. If p is an asymmetric norm on a vector space X, then $\rho(x, y) = p(y - x),\ x, y \in X$, is a quasi-metric on X. A closed, respectively open, ball in a quasi-metric space is defined by

$$B_\rho(x, r) = \{y \in X : \rho(x, y) \leq r\}, \quad B'_\rho(x, r) = \{y \in X : \rho(x, y) < r\},$$

for $x \in X$ and $r > 0$. In the case of an asymmetric norm p one denotes by $B_p(x, r), B'_p(x, r)$ the corresponding balls and by $B_p = B_p(0, 1), B'_p = B'_p(0, 1)$, the unit balls. In this case the following equalities hold

$$B_p(x, r) = x + rB_p \quad \text{and} \quad B'_p(x, r) = x + rB'_p.$$

The family of sets $B'_\rho(x, r)$, $r > 0$, is a base of neighborhoods of the point $x \in X$ for the topology τ_ρ on X generated by the quasi-metric ρ. The family $B_\rho(x, r)$, $r > 0$, of closed balls is also a neighborhood base at x for τ_ρ.

A quasi-uniformity on a set X is a filter $\mathcal{U}$ such that

(QU1) $\quad \Delta(X) \subset U,\ \forall U \in \mathcal{U}$;

(QU1) $\quad \forall U \in \mathcal{U},\ \exists V \in \mathcal{U}$, such that $V \circ V \subset U$,

where $\Delta(X) = \{(x, x) : x \in X\}$ denotes the diagonal of X and, for $M, N \subset X \times X$,

$$M \circ N = \{(x, z) \in X \times X : \exists y \in X,\ (x, y) \in M \text{ and } (y, z) \in N\}.$$

If the filter $\mathcal{U}$ satisfies also the condition

(U3) $\quad \forall U,\ U \in \mathcal{U} \ \Rightarrow\ U^{-1} \in \mathcal{U}$,

where

$$U^{-1} = \{(y, x) \in X \times X : (x, y) \in U\},$$

then $\mathcal{U}$ is called a uniformity on X. The sets in $\mathcal{U}$ are called *entourages* (or *vicinities*).

For $U \in \mathcal{U}$, $x \in X$ and $Z \subset X$ put

$$U(x) = \{y \in X : (x, y) \in U\} \quad \text{and} \quad U[Z] = \cup\{U(z) : z \in Z\}.$$

A quasi-uniformity $\mathcal{U}$ generates a topology $\tau(\mathcal{U})$ on X for which the family of sets

$$\{U(x) : U \in \mathcal{U}\}$$

is a base of neighborhoods of the point $x \in X$. A mapping f between two quasi-uniform spaces $(X,\mathcal{U})$, $(Y,\mathcal{W})$ is called *quasi-uniformly continuous* if for every $W \in \mathcal{W}$ there exists $U \in \mathcal{U}$ such that $(f(x), f(y)) \in W$ for all $(x, y) \in U$. By the definition of the topology generated by a quasi-uniformity, it is clear that a quasi-uniformly continuous mapping is continuous with respect to the topologies $\tau(\mathcal{U})$, $\tau(\mathcal{W})$.

If (X, ρ) is a quasi-metric space, then

$$B'_\epsilon = \{(x, y) \in X \times X : \rho(x, y) < \epsilon\},\ \epsilon > 0,$$

is a basis for a quasi-uniformity $\mathcal{U}_\rho$ on X. The family

$$B_\epsilon = \{(x, y) \in X \times X : \rho(x, y) \leq \epsilon\},\ \epsilon > 0,$$

generates the same quasi-uniformity. The topologies generated by the quasi-metric ρ and by the quasi-uniformity $\mathcal{U}_\rho$ agree, i.e., $\tau_\rho = \tau(\mathcal{U}_\rho)$.

The lack of the symmetry, i.e., the omission of the axiom (U3), makes the theory of quasi-uniform spaces to differ drastically from that of uniform spaces. An account of the theory up to 1982 is given in the book by Fletcher and Lindgren [21]. The survey papers by Künzi [32, 33, 34, 35] are good guides for subsequent developments. Another book on quasi-uniform spaces is [38].

On the other hand, the theory of asymmetric normed spaces has been developed in a series of papers [6], [8], [22], [23], [24], [25], [25], [26], following ideas from the theory of (symmetric) normed spaces and emphasizing similarities as well as differences between the symmetric and the asymmetric case.

Let (X, p) be an asymmetric normed space. The functional $\bar{p}(x) = p(-x)$, $x \in X$, is also an asymmetric norm on X, called the conjugate of p, $p_s(x) = \max\{p(x), \bar{p}(x)\}$, $x \in X$, is a (symmetric) norm on X and the following inequalities hold

$$|p(x) - p(y)| \leq p_s(x - y) \quad \text{and} \quad |\bar{p}(x) - \bar{p}(y)| \leq p_s(x - y),\ \forall x, y \in X.$$

For a quasi-metric space one defines similarly the conjugate of ρ by $\bar{\rho}(x,y) = \rho(y,x)$ and the associated (symmetric) metric by $\rho_s(x,y) = \max\{\rho(x,y), \rho(y,x)\}$, for $x, y \in X$.

Let (X,p), (Y,q) be two asymmetric normed space. A linear mapping $A : X \to Y$ is called *bounded*, ((p,q)-bounded if more precision is needed), or *semi-Lipschitz*, if there exists a number $\beta \geq 0$ such that

$$q(Ax) \leq \beta p(x), \tag{1.1}$$

for all $x \in X$. The number β is called a semi-Lipschitz constant for A. For properties of semi-Lipschitz functions and of spaces of semi-Lipschitz functions see [39, 40, 44, 45].

The operator A is continuous with respect to the topologies τ_p, τ_q ((τ_p, τ_q)-continuous) if and only if it is bounded and if and only if it is quasi-uniformly continuous with respect to the quasi-uniformities $\mathcal{U}_p$ and $\mathcal{U}_q$ (see [20] and [24]). Denote by $(X,Y)^\flat_{p,q}$, or simply by $(X,Y)^\flat$ when there is no danger of confusion, the set of all (p,q)-bounded linear operators. The set $(X,Y)^\flat$ need not be a linear subspace but merely a convex cone in the space $(X,Y)^\#$ of all linear operators from X to Y, i.e., $A+B \in (X,Y)^\flat$ and $\alpha A \in (X,Y)^\flat$, for any $A, B \in (X,Y)^\flat$ and $\alpha \geq 0$. Following [24], we shall call $(X,Y)^\flat$ a *semilinear space*. The functional

$$\|A| = \|A|_{p,q} = \sup\{q(Ax) : x \in B_p\} \tag{1.2}$$

is an asymmetric norm on the semilinear space $(X,Y)^\flat$, and $\|A|$ is the smallest semi-Lipschitz constant for A, i.e., the smallest number for which the inequality (1.1) holds.

Denote by $(X,Y)^*_s$ the space of all continuous linear operators from (X,p_s) to (Y,q_s), normed by

$$\|A\| = \|A\|_{p_s,q_s} = \sup\{q_s(Ax) : x \in X,\ p_s(x) \leq 1\}, \quad A \in (X,Y)^*_s. \tag{1.3}$$

It was shown in [24] that $(X,Y)^\flat_{p,q} \subset (X,Y)^*_s$, and $\|A| \leq \|A\|$ for any $A \in (X,Y)^\flat$.

Consider on $\mathbb{R}$ the asymmetric norm $u(\alpha) = \max\{\alpha, 0\}$, $\alpha \in \mathbb{R}$. Its conjugate is $\bar{u}(\alpha) = \max\{-\alpha, 0\}$ and $u_s(\alpha) = |\alpha|$ is the absolute value norm on $\mathbb{R}$. The topology

τ_u on $\mathbb{R}$ generated by u, called the upper topology of $\mathbb{R}$, has as neighborhood basis of a point $\alpha \in \mathbb{R}$ the family of intervals $(-\infty, \alpha + \epsilon)$, $\epsilon > 0$.

The space of all linear bounded functionals from an asymmetric normed space (X, p) to $(\mathbb{R}, u)$ is denoted by $X_p^\flat$. Notice that, due to the fact that p is non-negative, we have

$$\forall x \in X,\ u(\varphi(x)) \leq \beta p(x) \iff \varphi(x) \leq \beta p(x),$$

for any linear functional $\varphi : X \to \mathbb{R}$, so the asymmetric norm of a functional $\varphi \in X_p^\flat$ is given by

$$\|\varphi| = \|\varphi|_p = \sup\{\varphi(x) : x \in X,\ p(x) \leq 1\}.$$

Also, the continuity of φ from (X, τ_p) to $(\mathbb{R}, \tau_u)$ is equivalent to its upper semi-continuity from (X, τ_p) to $(\mathbb{R}, |\ |)$, (see [1, 2, 20]).

In [24] it was defined the analog of the w^*-topology on the space $X_p^\flat$, which we denote by $w^\flat$, having as a base of $w^\flat$-neighborhoods of an element $\varphi_0 \in X_p^\flat$ the sets

$$V_{x_1,...,x_n;\,\epsilon}(\varphi_0) = \{\varphi \in X_p^\flat : \varphi(x_i) - \varphi_0(x_i) \leq \epsilon,\ i = 1, ..., n\}, \tag{1.4}$$

for $n \in \mathbb{N}$, $x_1, ..., x_n \in X$, and $\epsilon > 0$.

Since

$$V_{x;\,\epsilon}(\varphi_0) \cap V_{-x;\,\epsilon}(\varphi_0) = \{\varphi \in X_p^\flat : |\varphi(x) - \varphi_0(x)| \leq \epsilon\},$$

it follows that the topology $w^\flat$ is the restriction to $X^\flat$ of the w^*-topology of $X_s^* = (X, p_s)^*$.

Some results on $w^\flat$-topology were proved in [24] as, for instance, the analog of the Alaoglu-Bourbaki theorem: the polar

$$B_p^\flat = \{\varphi \in X^\flat : \varphi(x) \leq 1,\ \forall x \in B_p\} \tag{1.5}$$

of the unit ball B_p of (X, p) is $w^\flat$-compact. Other results on asymmetric normed spaces, including separation of convex sets by closed hyperplanes and a Krein-Milman type theorem, were obtained in [6]. Asymmetric locally convex spaces were considered in [7]. Best approximation problems in asymmetric normed spaces were studied in [6] and [8].

The topology $w^\flat$ is derived from a quasi-uniformity $\mathcal{W}_p^\flat$ on $X_p^\flat$ with a basis formed of the sets

$$V_{x_1,...,x_n;\,\epsilon} = \{(\varphi_1, \varphi_2) \in X_p^\flat \times X_p^\flat : \varphi_2(x_i) - \varphi_1(x_i) \leq \epsilon,\ i = 1, ..., n\}, \tag{1.6}$$

for $n \in \mathbb{N}$, $x_1, ..., x_n \in X$ and $\epsilon > 0$. Note that, for fixed $\varphi_1 = \varphi_0$, one obtains the neighborhoods from (1.4).

On the space $(X, Y)_s^*$ we shall consider several quasi-uniformities. Namely, for $\mu \in \{p, \bar{p}, p_s\}$ and $\nu \in \{q, \bar{q}, q_s\}$ let $\mathcal{U}_{\mu,\nu}$ be the quasi-uniformity generated by the basis

$$U_{\mu,\nu;\,\epsilon} = \{(A, B); A, B \in (X, Y)_s^*,\ \nu(Bx - Ax) \leq \epsilon,\ \forall x \in B_\mu, \},\ \ \epsilon > 0, \tag{1.7}$$

where $B_\mu = \{x \in X : \mu(x) \leq 1\}$ denotes the unit ball of (X, μ). The induced quasi-uniformity on the semilinear subspace $(X, Y)_{\mu,\nu}^\flat$ of $(X, Y)_s^*$ is denoted also by $\mathcal{U}_{\mu,\nu}$ and the corresponding topologies by $\tau(\mu, \nu)$. The uniformity $\mathcal{U}_{p_s,q_s}$ and the topology $\tau(p_s, q_s)$ are those corresponding to the norm (1.3) on the space $(X, Y)_s^*$.

In the case of the dual space $X_\mu^\flat$ we shall use the notation $\mathcal{U}_\mu^\flat$ for the quasi-uniformity $\mathcal{U}_{\mu,u}$.

2. Completeness and compactness in quasi-metric and in quasi-uniform spaces

The lack of symmetry in the definition of quasi-metric and quasi-uniform spaces causes a lot of troubles, mainly concerning completeness, compactness and total boundedness in such spaces. There are a lot of completeness notions in quasi-metric and in quasi-uniform spaces, all agreeing with the usual notion of completeness in the case of metric or uniform spaces, each of them having its advantages and weaknesses.

We shall describe briefly some of these notions along with some of their properties.

The first one is that of bicompleteness. A quasi-metric space (X, ρ) is called *bicomplete* if the associated symmetric metric space (X, ρ_s) is complete. A bicomplete asymmetric normed space (X, p) is called also a *biBanach space.* The existence of a

bicompletion of an asymmetric normed space was proved in [22]. The notion can be considered also for an *extended* (i.e. taking values in $[0,\infty]$) quasi-metric, or for an extended asymmetric norm on a semilinear space.

In [24] it was defined an extended asymmetric norm on $(X,Y)_s^*$ by

$$\|A|_{p,q}^* = \sup\{q(Ax) : x \in B_p\}, \quad A \in (X,Y)_s^*. \tag{2.1}$$

The identity mapping $\mathrm{id}_{\mathbb{R}}$ is continuous from $(\mathbb{R},u)$ to $(\mathbb{R},u)$, but for $-\mathrm{id}_{\mathbb{R}}$ we have

$$\|-\mathrm{id}_{\mathbb{R}}\,|_{u,u}^* = \sup\{-\alpha : u(\alpha) \le 1\} \ge \sup\{-\alpha : \alpha \le 0\} = +\infty,$$

because $u(\alpha) = 0 \le 1$ for $\alpha \le 0$. It follows that $\|A|_{p,q}^*$ can take effectively the value $+\infty$.

If the asymmetric normed space (Y,p) is bicomplete, then the space $(X,Y)_s^*$ is complete with respect to the symmetric extended norm $(\|\,|_{p,q}^*)_s$ and $(X,Y)_{p,q}^\flat$ is a $(\|\,|_{p,q}^*)_s$-closed semilinear subspace of $(X,Y)_s^*$, so it is $\|\,|_{p,q}$-bicomplete (see [24]).

In the case of a quasi-metric space (X,ρ) there are also other completeness notions. We present them following [42], starting with the definitions of Cauchy sequences.

A sequence (x_n) in (X,ρ) is called

(a) *left (right) ρ-Cauchy* if for every $\epsilon > 0$ there exist $x \in X$ and $n_0 \in \mathbb{N}$ such that

$$\forall n \ge n_0,\ \rho(x,x_n) < \epsilon \text{ (respectively } \rho(x_n,x) < \epsilon)\,;$$

(b) *ρ-Cauchy* if for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n,k \ge n_0,\ \rho(x_n,x_k) < \epsilon\,;$$

(c) *left (right)-K-Cauchy* if for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n,k,\ n \ge k \ge n_0 \ \Rightarrow\ \rho(x_k,x_n) < \epsilon \text{ (respectively } \rho(x_n,x_k) < \epsilon)\,;$$

(d) *weakly left(right) K-Cauchy* if for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \ge n_0,\ \rho(x_{n_0},x_n) < \epsilon \text{ (respectively } \rho(x_n,x_{n_0}) < \epsilon).$$

These notions are related in the following way:

left(right) K-Cauchy $\Rightarrow$ weakly left(right) K-Cauchy $\Rightarrow$ left(right) ρ-Cauchy,

and no one of the above implications is reversible (see [42]).

Furthermore, each ρ-convergent sequence is ρ-Cauchy, but for each of the other notions there are examples of ρ-convergent sequences that are not Cauchy, which is a major inconvenience of the theory. Another one is that closed subspaces of a complete (in some sense) quasi-metric spaces need not be complete. If each convergent sequence in a regular quasi-metric space (X, ρ) admits a left K-Cauchy subsequence, then X is metrizable ([36]. This result shows that putting too many conditions on a quasi-metric, or on a quasi-uniform space, in order to obtain results similar to those in the symmetric case, there is the danger to force the quasi-metric to be a metric and the quasi-uniformity a uniformity. In fact, this is a general problem when dealing with generalizations.

For each of these notions of Cauchy sequence one obtains a notion of sequential completeness, by asking that each corresponding Cauchy sequence be convergent in (X, ρ). These notions of completeness are related in the following way:

left (right) ρ-sequentially complete $\Rightarrow$ weakly left (right) K-sequentially complete $\Rightarrow$

$\Rightarrow$ ρ-sequentially complete.

In spite of the obvious fact that left ρ-Cauchy is equivalent to right $\bar{\rho}$-Cauchy, left ρ- and right $\bar{\rho}$-completeness do not agree, due to the fact that right $\bar{\rho}$-completeness means that every left ρ-Cauchy sequence converges in $(X, \bar{\rho})$, while left ρ-completeness means the convergence of such sequences in the space (X, ρ). For concrete examples, see [42].

A subset Y of a quasi-metric space (X, ρ) is called *precompact* if for every $\epsilon > 0$ there exists a finite subset Z of X such that

$$Y \subset \cup\{B_\rho(z, \epsilon) : z \in Z\}.$$

The set Y is called *totally bounded* if for every $\epsilon > 0$, Y can be covered by a finite family of sets of diameter less that ϵ, where the diameter of a subset A of X is defined by

$$\operatorname{diam}(A) = \sup\{\rho(x, y) : x, y \in A\}.$$

As it is known, in metric spaces the precompactness and the total boundedness are equivalent notions, a result that is not longer true in quasi-metric spaces, where precompactness is strictly weaker than total boundedness, see [37] or [38].

In spite of these peculiarities there are some positive results concerning Baire theorem and compactness. For instance, any compact quasi-metric space is left K-sequentially complete and precompact. If (X, ρ) is precompact and left ρ-sequentially complete, then it is sequentially compact (see [19, 42]). Hicks [28] proved some fixed point theorems in quasi-metric spaces (see also [5, 29])

Notice also that in quasi-metric spaces compactness, countable compactness and sequential compactness are different notions (see [18] and [31]).

The considered completeness notions can be extended to quasi-uniform spaces by replacing sequences by filters or nets (for nets, see [52, 53]). Let $(X,\mathcal{U})$ be a quasi-uniform space, $\mathcal{U}^{-1} = \{U^{-1} : U \in \mathcal{U}\}$ the conjugate quasi-uniformity on X, and $\mathcal{U}_s = \mathcal{U} \vee \mathcal{U}^{-1}$ the coarsest uniformity finer than $\mathcal{U}$ and $\mathcal{U}^{-1}$. The quasi-uniform space $(X,\mathcal{U})$ is called *bicomplete* if $(X,\mathcal{U}_s)$ is a complete uniform space. This notion is useful and easy to handle, because one can appeal to results from the theory of uniform spaces which is satisfactorily accomplished.

A subset Y of a quasi-uniform space $(X,\mathcal{U})$ is called *precompact* if for every $U \in \mathcal{U}$ there exists a finite subset Z of X such that $Y \subset U[Z]$. The set Y is called *totally bounded* if for every U there exists a finite family $A_1, ..., A_n$ of subsets of X such that $A_i \times A_i \subset U$, $i = 1, ..., n,$ and $Y \subset \cup_{i=1}^n A_i$. In uniform spaces total boundedness and precompactness agree, and a set is compact if and only if it is totally bounded

and complete. A subset Y of quasi-uniform space $(X,\mathcal{U})$ is totally bounded if and only if it is totally bounded as a subset of the uniform space $(X,\mathcal{U}_s)$.

Another notion of completeness is that considered by Sieber and Pervin [49]. A filter $\mathcal{F}$ in a quasi-uniform space $(X,\mathcal{U})$ is called *$\mathcal{U}$-Cauchy* if for every $U \in \mathcal{U}$ there exists $x \in X$ such that $U(x) \in \mathcal{F}$. In terms of nets, a net $(x_\alpha, \alpha \in D)$ is called *$\mathcal{U}$-Cauchy* if for every $U \in \mathcal{U}$ there exists $x \in X$ and $\alpha_0 \in D$ such that $(x, x_\alpha) \in U$ for all $\alpha \geq \alpha_0$. The quasi-uniform space $(X,\mathcal{U})$ is called *$\mathcal{U}$-complete* if every $\mathcal{U}$-Cauchy filter (equivalently, every $\mathcal{U}$-Cauchy net) has a cluster point. If every such filter (net) is convergent, then the quasi-uniform space $(X,\mathcal{U})$ is called *$\mathcal{U}$-convergence complete*. Obviously that convergence complete implies complete, but the converse is not true. It is clear that this notion corresponds to that of ρ-completeness of a quasi-metric space. It is worth to notify that the $\mathcal{U}_\rho$-completeness of the associated quasi-uniform space $(X,\mathcal{U}_\rho)$ implies the ρ-sequential completeness of the quasi-metric space (X,ρ), but the converse is not true (see [36]). The equivalence holds for the notion of left K-completeness (which will be defined immediately): a quasi-metric space is left K-sequentially complete if and only if its induced quasi-uniformity $\mathcal{U}_\rho$ is left K-complete ([43]).

A filter $\mathcal{F}$ in a quasi-uniform space $(X,\mathcal{U})$ is called *left K-Cauchy* provided for every $U \in \mathcal{U}$ there exists $F \in \mathcal{F}$ such that $U(x) \in F$ for all $x \in F$. A net $(x_\alpha, \alpha \in D)$ in X is called *left K-Cauchy* provided for every $U \in \mathcal{U}$ there exists $\alpha_0 \in D$ such that $(x_\alpha, x_\beta) \in U$ for all $\beta \geq \alpha \geq \alpha_0$. The quasi-uniform space $(X,\mathcal{U})$ is called *left K-complete* if every left K-Cauchy filter (equivalently, every left K-Cauchy net) converges. If every left K-Cauchy filter converges with respect to the uniformity $\mathcal{U}_s$, then the quasi-uniform space $(X,\mathcal{U})$ is called *Smyth complete* (see [33] and [51]). This notion of completeness has applications to computer science, see [50]. In fact, there are a lot of applications of quasi-metric spaces, asymmetric normed spaces and quasi-uniform spaces to computer science, abstract languages, complexity, see, for instance, [23, 27, 41, 46, 47, 48].

Künzi et al [36] proved that a quasi-metric space is compact if and only if it is precompact and left K-sequentially complete, and studied the relations between completeness, compactness, precompactness, total boundedness and other related notions in quasi-uniform spaces.

Another useful notion of completeness was considered by Doitchinov [13, 14, 15, 16, 17]. A filter $\mathcal{F}$ in a quasi-uniform space $(X,\mathcal{U})$ is called *D-Cauchy* provided there exists a co-filter $\mathcal{G}$ in X such that for every $U \in \mathcal{U}$ there are $G \in \mathcal{G}$ and $F \in \mathcal{F}$ such that $F \times G \subset U$. The quasi-uniform space $(X,\mathcal{U})$ is called *D-complete* provided every D-Cauchy filter converges. A related notion of completeness was considered by Andrikopoulos [3]. For a comparative study of the completeness notions defined by pairs of filters see [10] and [4].

Notice also that these notions of completeness can be considered within the framework of bitopological spaces in the sense of Kelly [30], since a quasi-metric space is a bitopological space with respect to the topologies $\tau(\rho)$ and $\tau(\bar{\rho})$. For this approach see the papers by Deak [11, 12]. It seems that the letter K in the definition of left K-completeness comes from Kelly (see [9]).

3. **Compact operators**

Recall that a subset Z of an asymmetric normed space (X,p) is called *p-precompact* if for every $\epsilon > 0$ there exist $z_1, ..., z_n \in Z$ such that

$$\forall z \in Z,\ \exists i \in \{1, ..., n\}, \quad p(z - z_i) \leq \epsilon, \tag{3.1}$$

or, equivalently,

$$Z \subset U_\epsilon[\{z_1, ..., z_n\}],$$

where U_ϵ is the entourage

$$U_\epsilon = \{(x, x') \in X \times X : p(x' - x) \leq \epsilon\}$$

in the quasi-uniformity $\mathcal{U}_p$.

One obtains an equivalent notion taking the points z_i in X or/and $< \epsilon$ in (3.1).

Let $(X,p),(Y,q)$ be asymmetric normed spaces and, as before, let

$$\mu \in \{p, \bar{p}, p_s\} \text{ and } \nu \in \{q, \bar{q}, q_s\}. \tag{3.2}$$

A linear operator $A : X \to Y$ is called (μ,ν)-*compact* if the set $A(B_\mu)$ is ν-precompact in Y.

Some properties of compact operators are collected in the following proposition. We shall denote by $(X,Y)^k_{\mu,\nu}$ the set of all linear (μ,ν)-compact operators from X to Y. Notice that, for $\mu = p_s$ and $\nu = q_s$, the space $(X,Y)^\flat_{p_s,q_s}$ agrees with $(X,Y)^*_s$, the (p_s,q_s)-compact operators are the usual linear compact operators between the normed spaces (X,p_s) and (Y,q_s), so the proposition contains some well known results for compact operators on normed spaces.

Proposition 3.1. *Let $(X,p),(Y,q)$ be asymmetric normed spaces. The following assertions hold.*

1. *$(X,Y)^k_{\mu,\nu}$ is a semilinear subspace of $(X,Y)^\flat_{\mu,\nu}$.*
2. *$(X,Y)^k_{p,q}$ is $\tau(p,\bar{q})$-closed in $(X,Y)^\flat_{p,q}$.*

Proof. (1) We give the proof in the case $\mu = p$ and $\nu = q$. The other cases can be treated similarly.

If $A : X \to Y$ is (p,q)-compact, then there exists $x_1,...,x_n \in B_p$ such that

$$\forall x \in B_p,\ \exists i \in \{1,...,n\}, \quad q(Ax - Ax_i) \leq 1. \tag{3.3}$$

If for $x \in B_p,\ i \in \{1,...,n\}$ is chosen according to (3.3), then

$$q(Ax) \leq q(Ax - Ax_i) + q(Ax_i) \leq 1 + \max\{q(Ax_j) : 1 \leq j \leq n\},$$

showing that the operator A is (p,q)-bounded.

Suppose that $A_1, A_2 : X \to Y$ are (p,q)-compact and let $\epsilon > 0$. By the (p,q)-compactness of the operators A_1, A_2, there exist $x_1,...,x_m$ and $y_1,...,y_n$ in B_p such that

$$\forall x \in B_p,\ \exists i \in \{1,...,m\},\ \exists j \in \{1,...,n\},\ \ q(A_1x - A_1x_i) \leq \epsilon \text{ and } q(A_2x - A_2x_j) \leq \epsilon.$$

It follows that for every $x \in B_p$ there exists a pair (i,j) with $1 \le i \le m$ and $1 \le j \le n$ such that

$$q(A_1x + A_2x - A_1x_i - A_2y_j) \le q(A_1x - A_1x_i) + q(A_2x - A_2y_j) \le 2\epsilon,$$

showing that $\{A_1x_i + A_2y_j : 1 \le i \le m,\ 1 \le j \le n\}$ is a finite 2ϵ-net for $(A_1+A_2)(B_p)$.

The proof of the compactness of αA, for $\alpha > 0$ and A compact, is immediate and we omit it.

(2) *The $\tau(p,\bar{q})$-closedness of $(X,Y)^k_{p.q}$.*

Let (A_n) be a sequence in $(X,Y)^k_{p,q}$ which is $\tau(p,\bar{q})$-convergent to $A \in (X,Y)^\flat_{p,q}$.

For $\epsilon > 0$ choose $n_0 \in \mathbb{N}$ such that

$$\forall n \ge n_0,\ \forall x \in B_p,\ \ \bar{q}(A_nx - Ax) \le \epsilon\ \ (\iff\ q(Ax - A_nx) \le \epsilon). \tag{3.4}$$

Let $x_1, ..., x_m \in B_p$ such that $A_{n_0}x_i,\ 1 \le i \le m,$ is an ϵ-net for $A_{n_0}(B_p)$. Then for every $x \in B_p$ there exists $i \in \{1, ..., m\}$ such that

$$q(A_{n_0}x - A_{n_0}x_i) \le \epsilon,$$

so that, by (3.4),

$$q(Ax - Ax_i) \le q(Ax - A_{n_0}x) + q(A_{n_0}x - A_{n_0}x_i) + q(A_{n_0}x_i - Ax_i) \le 3\epsilon.$$

Consequently, $Ax_i,\ 1 \le i \le m,$ is a 3ϵ-net for $A(B_p)$, showing that $A \in (X,Y)^k_{p,q}$. □

Remark 3.2. The assertion (2) of Proposition 3.1 holds for other types of compactness too, i.e. for the spaces $(X,Y)^k_{\mu,\nu}$ with μ, ν as in (3.2), with similar proofs.

4. **The dual of a bounded linear operator**

Let (X,p), (Y,q) be asymmetric normed spaces and μ, ν as in (3.2). For $A \in (X,Y)^\flat_{\mu,\nu}$ define $A^\flat : Y^\flat_\nu \to X^\flat_\mu$ by

$$A^\flat\psi = \psi \circ A,\ \ \psi \in Y^\flat_s. \tag{4.1}$$

Obviously that $A^\flat$ is properly defined, additive and positively homogeneous. Concerning the continuity we have.

Proposition 4.1. *1. The operator $A^\flat$ is quasi-uniformly continuous with respect to the quasi-uniformities $\mathcal{U}_\nu^\flat$ and $\mathcal{U}_\mu^\flat$ on $Y_\nu^\flat$ and $X_\mu^\flat$, respectively.*

2. The operator $A^\flat$ is also quasi-uniformly continuous with respect to the $w^\flat$-quasi-uniformities on $Y_\nu^\flat$ and $X_\mu^\flat$.

Proof. (1) Take again $\mu = p$ and $\nu = q$. For $\epsilon > 0$ let

$$U_\epsilon = \{(\varphi_1, \varphi_2) \in X_p^\flat \times X_p^\flat : \varphi_2(x) - \varphi_1(x) \leq \epsilon,\ \forall x \in B_p\}.$$

If $\|A|_{p,q} = 0$, then $A = 0$, so we can suppose $\|A| = \|A|_{p,q} > 0$. Let

$$V_\epsilon = \{(\psi_1, \psi_2) \in Y_q^\flat \times Y_q^\flat : \psi_2(x) - \psi_1(x) \leq \epsilon/\|A|,\ \forall x \in B_q\}.$$

Taking into account that

$$\forall x \in B_p,\ \varphi_2(x) - \varphi_1(x) \leq \epsilon/r \iff \forall x' \in rB_p,\ \varphi_2(x') - \varphi_1(x') \leq \epsilon,$$

and

$$\forall x \in B_p, \quad q(Ax) \leq \|A|p(x) \leq \|A|,$$

it follows

$$A^\flat\psi_2(x) - A^\flat\psi_1(x) = \psi_2(Ax) - \psi_1(Ax) \leq \epsilon,$$

for all $x \in B_p$, proving the quasi-uniform continuity of A.

(2) For $x_1, ..., x_n \in X$ and $\epsilon > 0$ let

$$V = \{(\varphi_1, \varphi_2) \in X_p^\flat \times X_p^\flat : \varphi_2(x_i) - \varphi_1(x_i) \leq \epsilon,\ i = 1, ..., n\}$$

be a $w^\flat$-entourage in $X_p^\flat$. Then

$$U = \{(\psi_1, \psi_2) \in Y_q^\flat \times Y_q^\flat : \psi_2(Ax_i) - \psi_1(Ax_i) \leq \epsilon,\ i = 1, ..., n\},$$

is a $w^\flat$-entourage in $Y_q^\flat$ and $(A^\flat\psi_1, A^\flat\psi_2) \in V$ for every $(\psi_1, \psi_2) \in U$, proving the quasi-uniform continuity of $A^\flat$ with respect to the $w^\flat$-quasi-uniformities on $Y_q^\flat$ and $X_p^\flat$. □

Now we can prove the analog of the Schauder theorem for the asymmetric dual.

Theorem 4.2. *Let $(X,p),(Y,q)$ be asymmetric normed spaces. If the linear operator $A: X \to Y$ is (p,q)-compact, then $A^\flat(B_q^\flat)$ is precompact with respect to the quasi-uniformity $\mathcal{U}_p^\flat$ on $X_p^\flat$.*

Proof. For $\epsilon > 0$ let

$$U_\epsilon = \{(\varphi_1,\varphi_2) \in X_p^\flat \times X_p^\flat : \varphi_2(x) - \varphi_1(x) \le \epsilon,\ \forall x \in B_p\},$$

be an entourage in $X_p^\flat$ for the quasi-uniformity $\mathcal{U}_p^\flat$.

Since A is (p,q)-compact, there exist $x_1,...,x_n \in B_p$ such that

$$\forall x \in B_p\,,\ \exists i \in \{1,...,n\}, \quad q(Ax - Ax_i) \le \epsilon. \tag{4.2}$$

By the Alaoglu-Bourbaki theorem, [24, Theorem 4] the set $B_q^\flat$ is $w^\flat$-compact, so by the $(w^\flat,w^\flat)$-continuity of the operator $A^\flat$ (Proposition 4.1), the set $A^\flat(B_q^\flat)$ is $w^\flat$-compact in $X_p^\flat$. Consequently, the $w^\flat$-open cover

$$V_\psi = \{\varphi \in X_p^\flat : \varphi(x_i) - A^\flat\psi(x_i) < \epsilon,\ i = 1,...,n\},\ \psi \in B_q^\flat,$$

contains a finite subcover V_{ψ_k}, $1 \le k \le m$, i.e,

$$A^\flat(B_q^\flat) \subset \bigcup\{V_{\psi_k} : 1 \le k \le m\}. \tag{4.3}$$

Now let $\psi \in B_q^\flat$. By (4.3) there exists $k \in \{1,...,m\}$ such that

$$A^\flat\psi(x_i) - A^\flat\psi_k(x_i) < \epsilon,\ i = 1,...,n.$$

If $x \in B_p$, then, by (4.2), there exists $i \in \{1,...,n\}$, such that

$$q(Ax - Ax_i) \le \epsilon.$$

It follows

$$\begin{aligned}
&\psi(Ax) - \psi_k(Ax) = \\
&= \psi(Ax) - \psi(Ax_i) + \psi(Ax_i) - \psi_k(Ax_i) + \psi_k(Ax_i) - \psi(Ax_i) \\
&\le 2q(Ax - Ax_i) + \epsilon \le 3\epsilon.
\end{aligned}$$

Consequently,

$$\forall x \in B_p\,, \quad (A^\flat\psi - A^\flat\psi_k)(x) \leq 3\epsilon,$$

proving that

$$A^\flat(B_q^\flat) \subset U_{3\epsilon}[\{A^\flat\psi_1, ..., A^\flat\psi_m\}].$$

□

Comments. As a measure of precaution, we have defined the compactness of an operator A in terms of the precompactness of the image of the unit ball B_p by A, rather than by the relative compactness of $A(B_p)$, as in the case of compact operators on usual normed spaces. As can be seen from Section 2, the relations between precompactness, total boundedness and completeness are considerably more complicated in the asymmetric case than in the symmetric one. To obtain some compactness properties of the set $A(B_p)$, one needs a study of the completeness of the space $(X,Y)^\flat_{\mu,\nu}$ with respect to various quasi-uniformities and various notions of completeness, which could be the topic of a further investigation.

References

[1] Alegre, C., Ferrer, J., and Gregori, V., *Quasi-uniformities on real vector spaces*, Indian J. Pure Appl. Math. **28**(1997), no. 7, 929-937.

[2] ———, *On the Hahn-Banach theorem in certain linear quasi-uniform structures*, Acta Math. Hungar. **82**(1999), no. 4, 325-330.

[3] Andrikopoulos, A., *Completeness in quasi-uniform spaces*, Acta Math. Hungar. **105** (2004), no. 1-2, 151-173.

[4] ———, *A larger class than the Deák one for the coincidence of some notions of quasi-uniform completeness using pairs of filters*, Studia Sci. Math. Hungar. **41**(2004), no. 4, 431-436.

[5] Carlson, J. W., and Hicks, T. L., *On completeness in quasi-uniform spaces*, J. Math. Anal. Appl. **34**(1971), 618-627.

[6] Cobzaş, S., *Separation of convex sets and best approximation in spaces with asymmetric norm*, Quaestiones Math. **27**(2004), no. 3, 275-296.

[7] ———, *Asymmetric locally convex spaces,* Int. J. Math. Math. Sci. **2005:16**(2005), 2585-2608.

[8] Cobzaş, S., and Mustăţa, C., *Extension of bounded linear functionals and best approximation in spaces with asymmetric norm*, Rev. Anal. Numér. Théor. Approx. **33**(2004), no. 1, 39-50.

[9] Collins, J., and Zimmer, J., *An asymmetric Arzelà-Ascoli theorem*, Preprint 16/05 (2005), University of Bath, http://www.bath.ac.uk/math-sci/BICS.

[10] Deák, J., *On the coincidence of some notions of quasi-uniform completeness defined by filter pairs*, Studia Sci. Math. Hungar. **26**(1991), no. 4, 411-413.

[11] ———, *A bitopological view of quasi-uniform completeness. I, II*, Studia Sci. Math. Hungar. **30**(1995), no. 3-4, 389-409, 411-431.

[12] ———, *A bitopological view of quasi-uniform completeness. III*, Studia Sci. Math. Hungar. **31**(1996), no. 4, 385-404.

[13] Doitchinov, D., *Completeness and completions of quasi-metric spaces*, Rend. Circ. Mat. Palermo (2) Suppl. (1988), no. 18, 41-50, Third National Conference on Topology (Italian) (Trieste, 1986).

[14] ——— *On completeness of quasi-uniform spaces*, C. R. Acad. Bulgare Sci. **41**(1988), no. 7, 5-8.

[15] ———, *Cauchy sequences and completeness in quasi-metric spaces*, Pliska Stud. Math. Bulgar. **11**(1991), 27-34.

[16] ———, *A concept of completeness of quasi-uniform spaces*, Topology Appl. **38**(1991), no. 3, 205-217.

[17] ——— *Completeness and completion of quasi-uniform spaces*, Trudy Mat. Inst. Steklov. **193**(1992), 103-107.

[18] Ferrer, J., and Gregori, V., *A sequentially compact non-compact quasi-pseudometric space*, Monatsh. Math. **96**(1983), 269-270.

[19] ———, *Completeness and Baire spaces*, Math. Chronicle **14**(1985), 39-42.

[20] Ferrer, J., Gregori, V., and Alegre, C., *Quasi-uniform structures in linear lattices*, Rocky Mountain J. Math. **23**(1993), no. 3, 877-884.

[21] Fletcher, P., and Lindgren, W. F., *Quasi-Uniform Spaces,* M. Dekker, New York 1982.

[22] García-Raffi, L. M., Romaguera, S., and Sánchez-Pérez, E. A., *The bicompletion of an asymmetric normed linear space*, Acta Math. Hungar. **97**(2002), no. 3, 183-191.

[23] ———, *Sequence spaces and asymmetric norms in the theory of computational complexity*, Math. Comput. Modelling **36**(2002), no. 1-2, 1-11.

[24] ———, *The dual space of an asymmetric normed linear space*, Quaest. Math. **26**(2003), no. 1, 83-96.

[25] ———, *On Hausdorff asymmetric normed linear spaces*, Houston J. Math. **29**(2003), no. 3, 717-728.

[26] ———, *Metrizability of the unit ball of the dual of a quasi-normed cone*, Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8) **7**(2004), no. 2, 483-492.

[27] ———, *Weak topologies on asymmetric normed linear spaces and non-asymptotic citeria in the theory of complexity analysis*, J. Anal. Appl. **2**(2004), no. 3, 125-138.

[28] Hicks, T.L., *Fixed point theorems for quasi-metric spaces*, Math. Japonica **33**(1988), 231-236.

[29] Huffman, S. M., Hicks, T. L., and Carlson, J. W., *Complete quasi-uniform spaces*, Canad. Math. Bull. **23** (1980), no. 4, 497-498.

[30] Kelly, J.C., *Bitopological spaces*, Proc. London Math. Soc. **13**(1963), 71-89.

[31] Künzi, H.-P. A., *A note on sequentially compact quasi-pseudometric spaces*, Monatsh. Math. **95**(1983), 219-220.

[32] ———, *Quasi-uniform spaces–eleven years later*, Topology Proc. **18**(1993), 143-171.

[33] ———, *Nonsymmetric topology*, Bolyai Soc. Math. Studies, Vol. 4, Topology, Szekszárd 1993, Budapest 1995, pp. 303-338.

[34] ———, *Nonsymmetric distances and their associated topologies: about the origin of basic ideas in the area of asymmetric,* in: *Handbook of the History of General Topology,* C. E. Aull and R. Lowen (Editors), Kluwer Acad. Publ., Dordrecht 2001, pp. 853-868.

[35] ———, *Quasi-uniform spaces in the year 2001*, Recent progress in general topology, II, North-Holland, Amsterdam, 2002, pp. 313-344.

[36] Künzi, H.-P. A., Mršević, M., Reilly, I. L., and Vamanamurthy, M. K., *Convergence, precompactness and symmetry in quasi-uniform spaces*, Math. Japon. **38**(1993), no. 2, 239-253.

[37] Lambrinos, P. Th., *On precompact quasi-uniform structures*, Proc. Amer. Math. Soc. **62**(1977), 365-366.

[38] M. G. Murdeshwar and S. A. Naimpally, *Quasi-Uniform Topological Spaces*, Nordhoff, Groningen, 1966.

[39] C. Mustăţa, *Extensions of semi-Lipschitz functions on quasi-metric spaces*, Rev. Anal. Numer. Theor. Approx. **30**(2001), no. 1, 61-67.

[40] ———, *On the extremal semi-Lipschitz functions*, Rev. Anal. Numer. Theor. Approx. **31**(2002), no. 1, 103-108.

[41] Pajoohesh, M., and Schellekens, M. P., *A survey of topological work at* CEOL, Topology Atlas Invited Contributions **9**(2004), no. 2, arXiv:math. GN/0412556v1.

[42] Reilly, I. L., Subrahmanyam, P. V., Vamanamurthy, M. K., *Cauchy sequences in quasi-pseudo-metric spaces,* Monatsh. Math. **93**(1982), 127-140.

[43] Romaguera, S., *Left K-completeness in quasi-metric spaces*, Math. Nachr. **157**(1992), 15-23.

[44] Romaguera, S., and Sanchis, M., *Semi-Lipschitz functions and best approximation in quasi-metric spaces*, J. Approx. Theory **103**(2000), no. 2, 292-301.

[45] ———, *Properties of the normed cone of semi-Lipschitz functions,* Acta Math. Hungar. **108**(2005), 55-70.

[46] Romaguera, S., and Schellekens, M., *Quasi-metric properties of complexity spaces*, Topology Appl. **98**(1999), no. 1-3, 311-322, II Iberoamerican Conference on Topology and its Applications (Morelia, 1997).

[47] ———, *The quasi-metric of complexity convergence*, Quaest. Math. **23**(2000), no. 3, 359-374.

[48] ———, *Duality and quasi-normability for complexity spaces*, Appl. Gen. Topol. **3**(2002), no. 1, 91-112.

[49] Sieber, J. L., and Pervin, W. J., *Completeness in quasi-uniform spaces*, Math. Ann. **158**(1965), 79-81.

[50] Schellekens, M. P., *The Smyth completion: a common foundation for denotational semantics and complexity analysis*: in Proc. MFPS 11, Electronic Notes in Theoretical Computer Science **1**(1995), 211-232.

[51] Smyth, M. B., *Completeness of quasi-uniform and syntopological spaces*, J. London Math. Soc. **49**(1994), 385-400.

[52] Sünderhauf, Ph., *Quasi-uniform completeness in terms of Cauchy nets*, Acta Math. Hungar. **69**(1995), no. 1-2, 47-54.

[53] ———, *Smyth completeness in terms of nets: the general case*, Quaest. Math. **20**(1997), no. 4, 715-720.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: scobzas@math.ubbcluj.ro

Anul LI **2006**

S T U D I A
UNIVERSITATIS BABEŞ-BOLYAI

MATHEMATICA
4

Redacţia: 400084 Cluj-Napoca, str. M. Kogălniceanu nr. 1

Telefon: 405300

SUMAR – CONTENTS – SOMMAIRE

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

LAPLACIAN DECOMPOSITION METHOD FOR INVERSE STOKES PROBLEMS

A.E. CURTEANU, L. ELLIOTT, D.B. INGHAM AND D. LESNIC

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. This paper considers an inverse boundary value problem associated to the Stokes equations which govern the motion of slow viscous incompressible fluid flows. The solution of these equations is analyzed using a novel technique based on a Laplacian decomposition instead of the more traditional approaches based on the biharmonic streamfunction formulation or the velocity-pressure formulation. The determination of the under-specified boundary values of the normal fluid velocity is made possible by utilizing within the analysis additional pressure measurements which are available from elsewhere on the boundary. Results both on the boundary and inside the solution domain are presented and discussed for a simple benchmark test example and an application in a square geometry in order to illustrate that the Laplacian decomposition in combination with BEM provides an efficient technique, in terms of accuracy, convergence and stability to investigate numerically an inverse Stokes flow.

1. Introduction

Due to the mathematical complexity of the Navier-Stokes equations, it is well known that the general solution of these equations is not possible. Therefore in order to construct tractable mathematical models of the fluid flow systems, it is necessary to resort to a number of simplifications. One of these simplifications occurs when viscous forces are of a higher-order in magnitude as compared to the inertial

Received by the editors: 20.06.2006.

2000 *Mathematics Subject Classification.* 76D07,35J05, 34A55, 74S15.

Key words and phrases. Laplacian decomposition, inverse problems, Stokes flow, boundary element method.

forces. Consequently, one may drop the inertia terms from the steady Navier-Stokes equations to obtain:

$$\mu \overline{\nabla}^2 \, \underline{\overline{q}} = \overline{\nabla} \, \overline{P} \tag{1}$$

where $\underline{\overline{q}}$ is the fluid velocity vector, $\overline{P}$ the pressure, ρ the density and μ the fluid viscosity. Equation (1) is called the steady Stokes equation and may be regarded as the fundamental equation for the very slow motion of viscous fluids, known as creeping flows or Stokes flows. Non-dimensionalising equation (1), using typical velocity and length scales U_0 and L, respectively, and defining $\underline{\overline{r}} = L\underline{r}$, $\underline{\overline{q}} = \overline{U_0}\underline{q}$ and $\overline{P} = \frac{\mu \overline{U_0}}{L} P$, results in

$$\nabla^2 \, \underline{q} = \nabla \, P. \tag{2}$$

Since the fluid flow is assumed to be incompressible, we also have the continuity equation

$$\nabla \cdot \underline{q} = 0. \tag{3}$$

If exact data for u and v are specified at all the points on the boundary $\partial\Omega$ then the velocity and the pressure can be determined everywhere inside the solution domain Ω. However, in many practical situations it is not always possible to specify both components of the velocity at all the points on the boundary. Consequently, a part of the boundary remains under-specified and in order to compensate for this under-specification extra information is used on another part of the boundary, which gives rise to a portion of the boundary being over-specified. Such problems are called inverse problems and as Hadamard [4] pointed out their solution may not depend continuously on the input data.

In practical problems the additional information has to come from measurements and frequently it is easier to measure the pressure, in addition to the fluid velocity, rather than the vorticity. Therefore, in this paper we introduce extra information on pressure and, clearly, in this case it is more appropriate to work with the Stokes equations rather than the biharmonic equation. Nevertheless, the initial step

in obtaining a numerical solution of such an inverse and ill-posed problem is to develop a method of solution for the corresponding direct problem. The velocity-pressure formulation for direct Stokes problems, based on the Laplacian decomposition and BEM, has been described elsewhere, for example, in Curteanu *et al.* [3].

In the underlying inverse Stokes problem, we investigate the numerical solution in a domain Ω enclosed by a non-smooth boundary $\partial\Omega$, such that

$$\partial\Omega = \Gamma \cup \Gamma_0 \tag{4}$$

where Γ_0 is the under-specified boundary section and $\Gamma = \partial\Omega - \Gamma_0$. Both the normal and the tangential components of the fluid velocity, namely u and v are specified on the section of the boundary Γ, whilst only the tangential component is given on Γ_0. However, this under-specification of the boundary conditions on Γ_0 is compensated by the additional pressure measurements over $\Gamma^* \subseteq \Gamma$. This problem has been previously solved by Zeb *et al.* [8] where they used the BEM on the full Stokes equations.

Furthermore, the system of algebraic equations that results from an application of the BEM, in conjunction with the boundary conditions, is solved using the zeroth-order Tikhonov regularization method. The numerical solutions are obtained for the unspecified values of both the normal component of the fluid velocity and of the boundary pressure. Due to the ill-posed nature of the inverse Stokes problems described above, it is important to consider the stability of the numerical solution. Therefore we investigate the effect of noise on the numerical solution for the unknown values of the normal fluid velocity and the boundary pressure by adding a random error to the input data. Perturbation in the tangential component has not been considered, because, in general, this information is physically available from the no-slip condition on the solid boundary and is unlikely to contain any noise.

2. **Mathematical formulation**

For what follows, it is not restrictive to assume two-dimensional flows in a bounded domain $\Omega \subseteq \mathbb{R}^2$. Differentiating the x and y components of equation (2) with respect to x and y, respectively, then adding together and using the continuity

equation, results in

$$\nabla^2 P = 0. \tag{5}$$

In order to simplify equations (2) and (3), the following formulation for the velocity $\underline{q} = (u, v)$ components are introduced:

$$u = f + \frac{x}{2}P, \tag{6}$$

$$v = g + \frac{y}{2}P. \tag{7}$$

From (2) and (5) this results in f and g being solutions of the Laplace equation, namely

$$\nabla^2 f = 0, \tag{8}$$

$$\nabla^2 g = 0. \tag{9}$$

The above substitutions have reduced the Stokes problem to the solution of three Laplace's equations, (5), (8) and (9), interconnected through some boundary conditions involving also the continuity equation (3).

It is well known that the harmonic function P in equations (6) and (7) is unique up to an arbitrary constant, a. Moreover, if f_0, g_0 and P_0 are harmonic functions subject to the prescribed boundary conditions on u and v, then so are $f_0 - \frac{ax}{2}$, $g_0 - \frac{ay}{2}$ and $P_0 + a$. However, this non-uniqueness can be easily avoided by prescribing the value of the pressure at one arbitrary spatial point and this holds for the inverse problem considered - the pressure being prescribed on a part of the boundary.

BEM - Integral equation

In this paper, the development of the BEM for discretising the Laplace equation is the classical approach, see Brebbia *et al.* [1], and it is based on using the fundamental solution for the Laplace equation and Green's identities. Thus, for example, equation $\nabla^2 f = 0$ may be recast as follows:

$$\eta(X) f(X) = \int_\Gamma \{f'(Y) G(X,Y) - f(Y) G'(X,Y)\} d\Gamma_Y \tag{10}$$

where

(i) $X \in \Omega \cup \Gamma$ and $Y \in \partial\Omega$ and $\partial\Omega$ is the boundary of the domain Ω,

(ii) $d\Gamma_Y$ denotes the differential increment of $\partial\Omega$ at Y,

(iii) $\eta(X) = 1$ if $X \in \Omega$, and $\eta(X) =$ the internal angle between the tangents to $\partial\Omega$ on either side of X divided by 2π if $X \in \partial\Omega$,

(iv) G is the fundamental solution for the Laplace equation which in two-dimensions is given by

$$G(X,Y) = -\frac{1}{2\pi} ln|X - Y| \tag{11}$$

(v) G', f' are the outward normal derivatives of G and f, respectively.

We note that, as with the classical constant BEM, nodal points are situated only at the segment mid-points and therefore f' has precisely one value at each of these nodal points. However, with the linear BEM formulations, nodes are situated at segment end-points and therefore, if the domain has corners, at those points f' has two components, one related to each of the sides adjacent to the corner. Therefore, in order to deal with corners and singularities, discontinuous linear boundary elements are introduced in this section.

In practice, the integral equation (10) can rarely be solved analytically and thus some form of numerical approximation is necessary. Based on the BEM, we subdivide the boundary $\partial\Omega$ into a series of N elements $\partial\Omega_j$, $j = \overline{1,N}$, and approximate the functions f and f' at the collocation points of each boundary element Γ_j. In the discontinuous linear elements method (DLBEM) it is assumed that the variables in the integral equation (10) have a linear evolution along the elements. These boundary elements are segments of a straight line and the linear evolution is expressed through the values of the functions at two internal points given by

$$X^j_{\tau,1} = (1-\tau)X_{j-1} + \tau X_j \tag{12}$$

$$X^j_{\tau,2} = \tau X_{j-1} + (1-\tau)X_j \tag{13}$$

where $\tau \in (0, \frac{1}{2})$. Correspondingly, the boundary integral equation (10) becomes

$$\begin{aligned}
\eta(X)f(X) &= \sum_{j=1}^{N} f'_{2j-1}\left[\frac{1-\tau}{1-2\tau}C_j(X) - \frac{1}{1-2\tau}E_j(X)\right] \\
&+ \sum_{j=1}^{N} f'_{2j}\left[\frac{1}{1-2\tau}E_j(X) - \frac{\tau}{1-2\tau}C_j(X)\right] \\
&- \sum_{j=1}^{N} f_{2j-1}\left[\frac{1-\tau}{1-2\tau}D_j(X) - \frac{1}{1-2\tau}F_j(X)\right] \\
&- \sum_{j=1}^{N} f_{2j}\left[\frac{1}{1-2\tau}F_j(X) - \frac{\tau}{1-2\tau}D_j(X)\right] \qquad (14)
\end{aligned}$$

where C_j, D_j, E_j and F_j have the same meaning as in Mera *et al.* [6] and may be evaluated analytically. In the DLBEM, the discretised boundary integral equation (14) is applied on the boundary at each of the points $X^j_{\tau,1}$, $X^j_{\tau,2}$, $j = \overline{1,N}$, leading to a system of $2N$ equations

$$\sum_{j=1}^{2N}\left(A_{ij}f'_j - B_{ij}f_j = 0\right) \quad \text{for} \quad i = \overline{1,2N} \qquad (15)$$

where the matrices A_{ij} and B_{ij} are given by

$$\begin{aligned}
A_{i,2j-1} &= \frac{1-\tau}{1-2\tau}C_j(\underline{z}_i) - \frac{1}{1-2\tau}E_j(\underline{z}_i) \quad \text{for} \quad i = \overline{1,2N}, \quad j = \overline{1,N} \\
A_{i,2j} &= \frac{1}{1-2\tau}E_j(\underline{z}_i) - \frac{\tau}{1-2\tau}C_j(\underline{z}_i) \quad \text{for} \quad i = \overline{1,2N}, \quad j = \overline{1,N} \\
B_{i,2j-1} &= \frac{1-\tau}{1-2\tau}D_j(\underline{z}_i) - \frac{1}{1-2\tau}F_j(\underline{z}_i) \quad \text{for} \quad i = \overline{1,2N}, \quad j = \overline{1,N} \quad i \neq 2j-1 \\
B_{i,2j} &= \frac{1}{1-2\tau}F_j(\underline{z}_i) - \frac{\tau}{1-2\tau}D_j(\underline{z}_i) \quad \text{for} \quad i = \overline{1,2N}, \quad j = \overline{1,N} \quad i \neq 2j
\end{aligned}$$

and the collocation points $\underline{z}_i$, $i = 1, 2N$ are given by

$$\underline{z}_{2i-1} = X^i_{\tau,1} \quad \underline{z}_{2i} = X^i_{\tau,2} \quad i = \overline{1,N} \qquad (16)$$

Similar equations are obtained for the harmonic functions g and P.

Now the equations (8), (9) and (5) reduce to a system of $6N$ equations in $12N$ unknowns, i.e.

$$\begin{cases} A\boldsymbol{f'} - B\boldsymbol{f} = 0 \\ A\boldsymbol{g'} - B\boldsymbol{g} = 0 \\ A\boldsymbol{P'} - B\boldsymbol{P} = 0 \end{cases} \tag{17}$$

In the inverse formulation of the Stokes problem considered in this paper, both components of the fluid velocity, u and v, are specified on the section of the boundary $\Gamma = \partial\Omega - \Gamma_0$, whilst only the tangential component, e.g. u is given on Γ_0. However, this under-specification of the boundary conditions on Γ_0 is compensated by the additional pressure measurements over Γ^*. Clearly, if the velocity vector is known on $\partial\Omega$ then u' and v' can be obtained analytically by using equations $\frac{\partial u}{\partial n} = \pm\frac{\partial u}{\partial x} = \mp\frac{\partial v}{\partial y}$ and $\frac{\partial v}{\partial n} = \pm\frac{\partial v}{\partial y} = \mp\frac{\partial u}{\partial x}$, respectively. Solving the direct problem with u and v known on $\partial\Omega$, we obtain the pressure P everywhere and, in particular, over Γ^*. This numerically calculated pressure, denoted by $P^{(n)}$ is used in the inverse problem (17) and (18).

Suppose that the number of boundary elements N on $\partial\Omega$ is such that N_0 belongs to Γ_0 and N^* to Γ^*. Dividing the boundary such that $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, where $\Gamma_1 = \{(x,y)| -\frac{1}{2} \le x \le \frac{1}{2}, y = -\frac{1}{2}\}$, $\Gamma_2 = \{(x,y)|x = \frac{1}{2}, -\frac{1}{2} \le y \le \frac{1}{2}\}$, $\Gamma_3 = \{(x,y)| -\frac{1}{2} \le x \le \frac{1}{2}, y = \frac{1}{2}\}$, $\Gamma_4 = \{(x,y)|x = -\frac{1}{2}, -\frac{1}{2} \le y \le \frac{1}{2}\}$, the problem can be described mathematically by (17) and the following boundary conditions:

$$\begin{cases} \boldsymbol{f} + \frac{x}{2}\boldsymbol{P} = \boldsymbol{u}^{(n)} - \frac{x}{2}\boldsymbol{P}^{(n)} & \text{on } \partial\Omega \\ \boldsymbol{g} + \frac{y}{2}\boldsymbol{P} = \boldsymbol{v}^{(n)} - \frac{y}{2}\boldsymbol{P}^{(n)} & \text{on } \Gamma \\ \boldsymbol{f'} + \nu_1\boldsymbol{P} + \frac{x}{2}\boldsymbol{P'} = \boldsymbol{u'}^{(n)} - \nu_1\boldsymbol{P}^{(n)} & \text{on } \Gamma_2 \cup \Gamma_4 \\ \boldsymbol{g'} + \nu_2\boldsymbol{P} + \frac{y}{2}\boldsymbol{P'} = \boldsymbol{v'}^{(n)} - \nu_2\boldsymbol{P}^{(n)} & \text{on } \Gamma_1 \cup \Gamma_3 \end{cases} \tag{18}$$

where the N^* vector $\boldsymbol{P}^{(n)}$ is the given pressure on the over-specified part of the boundary $\Gamma^* \subseteq \Gamma$ and $\frac{x}{2}$, $\frac{y}{2}$, $\nu_1 = \frac{\partial(x/2)}{\partial n}$, $\nu_2 = \frac{\partial(y/2)}{\partial n}$ are the matrices $\delta_{ij}\frac{x_{(j)}}{2}$, $\delta_{ij}\frac{y_{(j)}}{2}$, $\delta_{ij}\nu_{1(j)}$ and $\delta_{ij}\nu_{2(j)}$, respectively .

In a generic form, the system of equations (17) and (18) can be rewritten as

$$\mathbb{A}\boldsymbol{x} = \boldsymbol{b} \tag{19}$$

where $\mathbb{A}$ is a known $(12N - 2N_0) \times (12N - 2N^*)$ matrix which includes the matrices A and B, $\boldsymbol{x}$ is a vector of $12N - 2N^*$ unknowns which includes the $2N$ vectors $\boldsymbol{f}|_{\partial\Omega}$, $\boldsymbol{f'}|_{\partial\Omega}$, $\boldsymbol{g}|_{\partial\Omega}$, $\boldsymbol{g'}|_{\partial\Omega}$ and $\boldsymbol{P'}|_{\partial\Omega}$, and the $2N - 2N^*$ vector $\boldsymbol{P}|_{\partial\Omega-\Gamma^*}$ and $\boldsymbol{b}$ is a vector of $12N - 2N_0$ knowns which includes $\boldsymbol{u}|_{\partial\Omega}$, $\boldsymbol{v}|_{\Gamma}$, $\boldsymbol{P}|_{\Gamma^*}$ and the derivatives of velocity. Then, using the calculated boundary data, interior solutions for the harmonic functions and the velocity can be determined explicitly using the integral equation, i.e. equation (10) for f, see Brebbia *et al.* [1].

Regularization method

The Tikhonov regularization method is an efficient method for solving inverse and ill-posed problems which arise in science and engineering. It modifies the least-squares approach and finds an approximate numerical solution which, in the case of the zero-th order regularization procedure, is given by, see Tikhonov and Arsenin [7],

$$\boldsymbol{x} = (\mathbb{A}^t\mathbb{A} + \lambda\mathbb{I})^{-1}\mathbb{A}^t\boldsymbol{b} \tag{20}$$

where $\mathbb{I}$ is the identity matrix, the superscript ^{t} denotes the transpose of a matrix and λ is the regularization parameter, which controls the degree of smoothing applied to the solution and whose choice may be based on the L-curve method, see Hansen [5]. For the zero-th order regularization procedure we plot on a log-log scale the variation of $||\boldsymbol{x}_\lambda||$ against the fitness measure, namely the residual norm $||\mathbb{A}\boldsymbol{x}_\lambda - \boldsymbol{b}||$ for a wide range of values of $\lambda > 0$. In many applications this graph results in a L-shaped curve and the choice of the optimal regularization parameter $\lambda > 0$ is based on selecting approximately the corner of this L-curve.

3. Numerical results

We investigate the solution of Stokes problem given by equations (2) and (3) in a simple two-dimensional non-smooth geometry, such as the square

$$\Omega = \left\{ (x,y)| -\frac{1}{2} < x < \frac{1}{2}, -\frac{1}{2} < y < \frac{1}{2} \right\}.$$

In order to investigate the convergence and the stability of the solution we consider first the following test example, namely the analytical expressions for the three harmonic functions f, g and P are given by:

$$f^{(an)} = -x^3/6 + y^2/2 + xy + xy^2/2 - x^2/2 \tag{21}$$

$$g^{(an)} = -y^2/2 - 3x^2y/2 + x^2/2 + y^3/2 - xy/2 \tag{22}$$

$$P^{(an)} = x^2 - y^2 + x \tag{23}$$

with the corresponding fluid velocity given by:

$$u^{(an)} = x^3/3 + y^2/2 + xy \tag{24}$$

$$v^{(an)} = -y^2/2 - x^2y + x^2/2 \tag{25}$$

For presenting the numerical results, we choose $\Gamma^* = \Gamma_2$. In order to study the effects of various locations for the under-specified boundary region we choose $\Gamma_0 = \Gamma_1$ and $\Gamma_0 = \Gamma_3$. If $\Gamma_0 = \Gamma_2$ and $\Gamma_0 = \Gamma_4$ are to be chosen then the velocity v has to be specified on Γ_0 instead of u, since on this parts of the boundary v is the tangential component.

Whilst in the direct Stokes problem we observed that the difference between the analytical solution and the numerical results for P using $N = 80$ was less than 1%, in the inverse problem, we found $N = 40$ was sufficiently large for the numerical solution to agree graphically with the corresponding numerical solution from the direct problem.

Figure 1(a) shows the numerical solution for the unspecified values of the normal component of the fluid velocity v over $\Gamma_0 = \Gamma_1$ for $\lambda = 10^{-11}$, together with its analytical value specified in the direct problem. From this figure, it is observed

that the agreement between the numerical solution and the one given in equation (25), which is specified analytically over Γ_0 in the direct problem, is excellent. In Figure 1(b), we present the numerical solution for the boundary pressure P over $\partial\Omega - \Gamma^*$ and the boundary pressure obtained from the direct problem. It can be seen in this figure that the numerical solution generated by the inverse problem agrees very well with the corresponding numerical solution obtained from the direct problem.

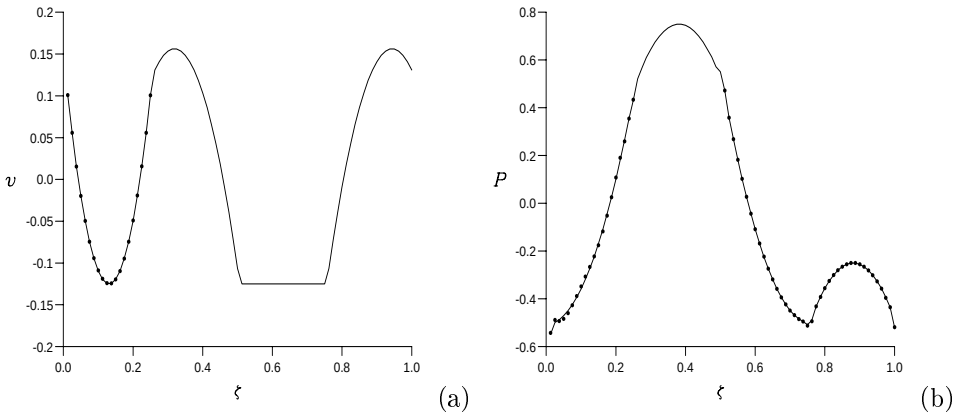


FIGURE 1. The numerical solution $(\cdots)$ for (a) the normal component of the fluid velocity $v|_{\Gamma_0}$, together with the values of v analytically specified over $\partial\Omega$, and (b) the boundary pressure $P|_{\partial\Omega-\Gamma^*}$, together with the corresponding numerical solution for P obtained in the direct problem when $\lambda = 10^{-11}$ using the BEM with 40 discontinuous linear boundary elements.

For various locations of the over-specified boundary region, i.e. $\Gamma^* = \Gamma_3$ or $\Gamma^* = \Gamma_4$, we observed that the numerical results are similar with those obtained for $\Gamma^* = \Gamma_2$. Without presenting the results graphically, we mention that when a different location of the under-specified boundary is chosen, namely $\Gamma_0 = \Gamma_3$ the agreement of the numerical results was found to be equivalent to that observed in Figure 1. Moreover, when we double or more the over-specified part of the boundary,

i.e. $\Gamma^* = \Gamma_2 \cup \Gamma_3$ or $\Gamma^* = \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, then an even better accuracy is obtained.

Effect of noise

As mentioned in the introduction, the inverse Stokes problem is ill-posed and the system of equations (19) that results is ill-conditioned, and hence the solution may not continuously depend upon the boundary data. Therefore the stability of the regularized boundary element technique is investigated by adding small amounts of noise into the input boundary data in order to simulate measurement errors which are inherently present in the data set of any practical problem. Hence, we perturb the boundary data i.e. data obtained from the direct problem, by adding random noisy perturbations ϵ to the boundary pressure $P^{(n)}$, namely

$$\tilde{P} = P^{(n)} + \epsilon \tag{26}$$

The random error ϵ is generated by using the NAG routine G05DDF, see Brent [2], and it represents a Gaussian random variable with mean zero and standard deviation σ, which is taken to be some percentage α of the maximum value of $P^{(n)}$, i.e.

$$\sigma = \max |P^{(n)}| \times \frac{\alpha}{100} \tag{27}$$

For a particular location of Γ_0, say $\Gamma_0 = \Gamma_1$, Figure 2 shows the L-curve for the inverse problem as a log-log plot of the solution norm $||\boldsymbol{x}_\lambda||$, against the residual norm $||\mathbb{A}\boldsymbol{x}_\lambda - \boldsymbol{b}||$, for various amounts of noise $\alpha = \{3, 6, 10\}$ introduced in $P^{(n)}|_{\Gamma^*}$ and for the various values of the regularization parameter λ taken from the range $[10^{-13}, 10^{-1}]$. We choose the optimal value of the regularization parameter λ, corresponding to the corner of the L-curve, as $\lambda_{opt} = \boldsymbol{O}(10^{-9})$ if $\alpha = 3$ and $\lambda_{opt} = \boldsymbol{O}(10^{-8})$ if $\alpha = \{6, 10\}$.

It was found that the numerical solution for the retrieved normal velocity $v|_{\Gamma_0}$ and pressure $P|_{\partial\Gamma - \Gamma^*}$, obtained using both the exact and noisy data, for λ_{opt} remains stable and agrees with the analytical values and the values specified in the direct problem, respectively, reasonably well according to the amount of noise introduced in the input data for pressure $P|_{\Gamma^*}$. Therefore omitting the boundary results, we present in Figure 3 (a) and (b) the lines of constant pressure and constant velocity

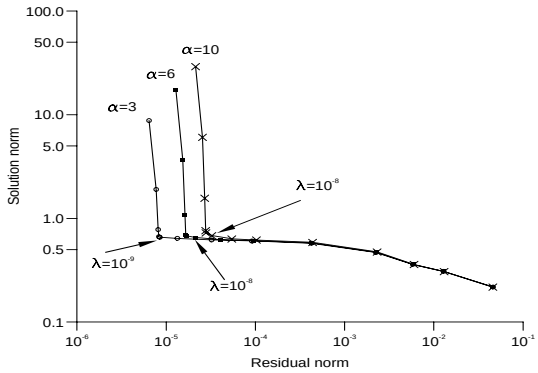


FIGURE 2. The L-curve plot of the solution norm $\| \boldsymbol{x}_\lambda \|$ as a function of the residual norm $\| \mathbb{A}\boldsymbol{x}_\lambda - \boldsymbol{b} \|$ in the inverse Stokes for $N = 40$, $\lambda = [10^{-13}, 10^{-1}]$, when various levels of noise $\alpha = \{3, 6, 10\}$ are added.

component inside the domain Ω and it can be observed that as the amount of noise decreases then the numerical solutions approximate better the solution obtained in the direct problem (for P) or analytically (for v) whilst at the same time remaining stable.

4. **Application - driven cavity**

Now we investigate an inverse problem in a square cavity filled with incompressible viscous fluid and the top lid moving with a constant velocity of unity, for which no analytical solution is available. Now, the tangential component of the fluid velocity is specified on the whole boundary, while the normal component of the fluid velocity v is unknown on e.g. the bottom side of the cavity, namely on $\Gamma_0 = \Gamma_1$ and this under-specification of the boundary conditions is compensated for by the additional pressure measurements on another part of the boundary or over the remaining

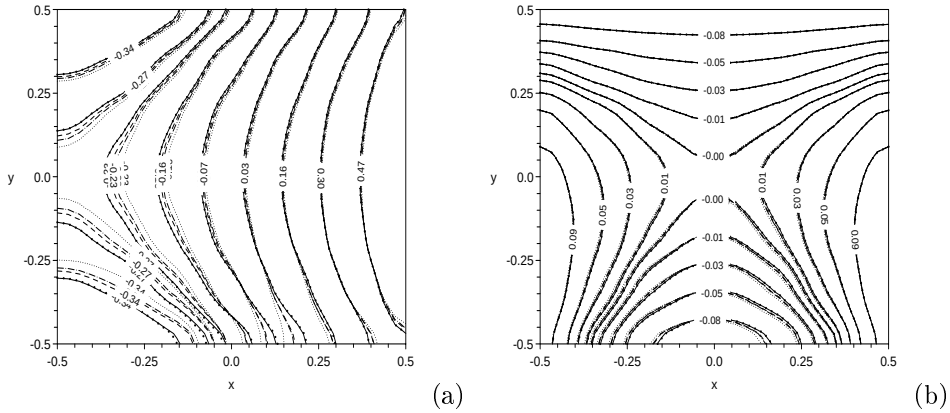


FIGURE 3. The lines of constant (a) pressure P, and (b) velocity v, inside the cavity Ω obtained with $N = 40$ discontinuous linear boundary elements when various levels of noise are introduced in $P|_{\Gamma^*=\Gamma_2}$, namely, (a) direct and (b) analytical solution (——), $\alpha = 0$(• • •), $\alpha = 3$(− − −), $\alpha = 6$(− · − · −·), and $\alpha = 10$(· · ·).

part of the boundary. The boundary conditions for the problem are as follows:

$$\begin{cases} u = v' = 0 & \text{on} \quad \Gamma_0 = \Gamma_1 \\ u = v = u' = 0 & \text{on} \quad \Gamma_2 \\ u = -1, v = v' = 0 & \text{on} \quad \Gamma_3 \\ u = v = u' = 0 & \text{on} \quad \Gamma_4 \\ P = P^{(n)} & \text{on} \quad \Gamma^* \end{cases} \tag{28}$$

When random noise, $\alpha = \{3, 6, 10\}$ is introduced in the pressure $P|_{\Gamma^*}$ the values of the regularization parameter λ given by the L-curve plots, are $\{10^{-10}, 10^{-9}, 10^{-9}\}$ when $\Gamma^* = \Gamma_2$ and $\{10^{-8}, 10^{-7}, 10^{-7}\}$ when $\Gamma^* = \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$. The boundary results were found accurate in comparison with the analytical or direct values and convergent to the exact solutions when the amount of noise decreases. Figure 4 shows the numerical results obtained for pressure inside the driven cavity when different amount of noise are introduced in $P|_{\Gamma^*}$ for two different locations of the over-specified part of the boundary. Also shown in this figure is the corresponding

solution obtained in the direct problem and it is observed that the errors between the numerical solution of the problem (17) and (28) and the values obtained in the direct problem (u and v specified on $\partial\Omega$) are comparable with the amount of noise included and the numerical solution approaches the exact solution with decreasing the amount of noise. Moreover, it can bee seen that the larger is the length of the over-specified part of the boundary, the better is the accuracy.

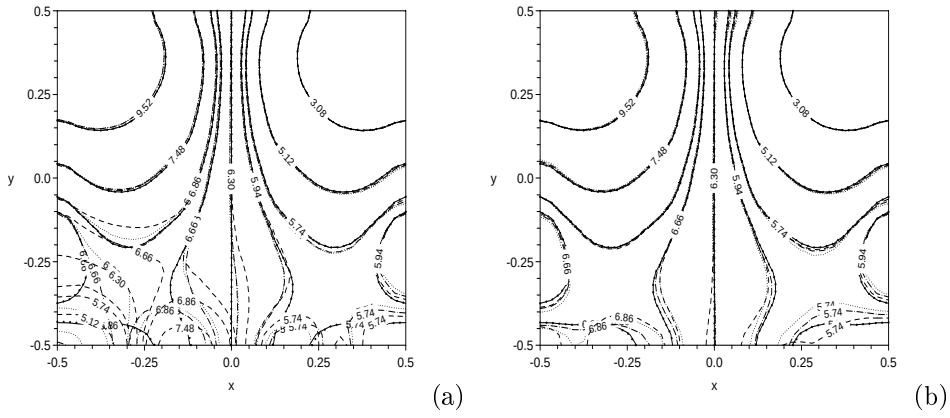


FIGURE 4. The lines of constant pressure P inside the cavity Ω using the BEM with $N = 40$ discontinuous linear boundary elements when various levels of noise are introduced in (a) $P|_{\Gamma^*=\Gamma_2}$ and (b) $P|_{\Gamma^*=\Gamma_2\cup\Gamma_3\cup\Gamma_4}$, namely, direct solution (——), $\alpha = 0$(• • •), $\alpha = 3$(– – –), $\alpha = 6$(– · – · –·), and $\alpha = 10$(· · ·) for the driven cavity problem.

In order to visualise the overall flow pattern inside the domain Ω we present in Figure 5 the velocity field at some selected interior points. The lengths of the vectors and of the arrows are proportional to the magnitude of the fluid velocity. Although not illustrated graphically, we wish to report that both the magnitude and the direction of the fluid vectors were observed to be similar to those obtained when solving both the direct and the inverse problem.

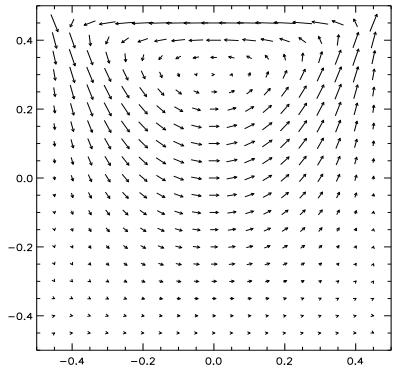


FIGURE 5. The velocity vectors at selected points inside the driven cavity obtained by solving the inverse problem with $\alpha = 0$ for $\Gamma_0 = \Gamma_1$ and $\Gamma^* = \Gamma_2$.

It is important to note that for the driven cavity problem with singularities, the solution is more sensitive to the location of the under-specified and over-specified boundaries, becoming less accurate at some points, especially on the under-specified boundary. Also these boundary errors propagate into the solution domain. However, the accuracy of the results improves by increasing the over-specified boundary.

5. **Conclusions**

In this paper the Stokes equations, subject to under-specified boundary conditions on the normal component of the fluid velocity v, but with additional pressure measurements available on another part of the boundary, have been studied. A boundary element discretisation has been applied to the resulting Laplace equations and the Tikhonov regularization method has been used to solve the resulting ill-conditioned system of linear algebraic equations. The technique has been validated for a typical benchmark test example and in a situation where no analytical solution is available in a square cavity. It has been shown that this regularized boundary element technique retrieves an accurate and stable numerical solution, both on the boundary and inside the solution domain, with respect to decreasing the amount of noise in the input

boundary data. Moreover, the numerical solutions converge for a reasonable number of boundary elements, about half the number of boundary elements used when solving the corresponding direct problem.

References

[1] Brebbia, C.A., Telles, J.C.F., and Wrobel, L.C., *Boundary Element Techniques: Theory and Applications in Engineering*, Springer-Verlag, Berlin, 1984.

[2] Brent, R.P., *Algorithm 488: A Gaussian Pseudo-Random Number Generator*, Commun. A. C. M., **17**(1974), 704-707.

[3] Curteanu, A., Ingham, D.B., Elliott, L., and Lesnic, D., *A Laplacian decomposition of the two-dimensional Stokes equation*, Fourth UK Conference on Boundary Integral Methods (Ed. S. Amini), Salford University Press, UK, 2003, 47-56.

[4] Hadamard, J., *Lectures on Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

[5] Hansen, P.C., *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., **34**(1992), 561-580.

[6] Mera, N.S., Elliott, L., Ingham, D.B., and Lesnic, D., *A comparison of boundary element method formulations for steady state anisotropic heat conduction problems*, Eng. Anal. Boundary Elem., **25**(2001), 115-128.

[7] Tikhonov, A.N., and Arsenin, V.Y., *Solutions of Ill-Posed Problems*, Winston-Wiley, Washington D.C, 1977.

[8] Zeb, A., Elliott, L., Ingham, D.B., and Lesnic, D., *Boundary element two-dimensional solution of an inverse Stokes problem*, Eng. Anal. Boundary Elem., **24**(2000), 75-88.

Department of Applied Mathematics, University of Leeds,
Leeds LS2 9JT, UK
E-mail address: amt6dbi@maths.leeds.ac.uk

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

PIECEWISE LINEAR INTERPOLATION REVISITED: BLAC-WAVELETS

H. GONSKA, D. KACSÓ, O. NEMITZ, AND P. PIŢUL

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. The central issue of the present note is the BLaC operator, a "Blending of Linear and Constant" approach. Several properties are proved, e.g., its positivity and the reproduction of constant functions. Starting from these results, error estimates in terms of ω_1 and ω_2 are given. Furthermore, we present the degree of approximation in the bivariate tensor product case. This is applicable to image compression.

1. Definitions and properties

BLaC-wavelets ("Blending of Linear and Constant wavelets") were introduced by G. P. Bonneau, S. Hahmann and G. Nielson around 1996 and constitute a tool to compromise between the perfect locality of Haar[1] wavelets and the better regularity of linear wavelets. This compromise is realized by means of a parameter $0 < \Delta \leq 1$ that will appear in the sequel. First we introduce some notations. For the real parameter $0 < \Delta \leq 1$ consider the *scaling function* $\varphi_\Delta : \mathbb{R} \to [0, 1]$ given by

Received by the editors: 15.06.2006.

2000 *Mathematics Subject Classification.* 41A15, 41A25, 41A36, 41A63.

Key words and phrases. positive linear operators, "blending of linear and constant" operator, degree of approximation, moduli of continuity, partial and total moduli of smoothness, tensor product.

[1] Alfréd Haar was born in 1885 in Budapest and died 1933 in Szeged. Until after World War I he had also a chair at the University of Cluj (then Kolozsvár). More about his biography can be found on the following site: http://www-history.mcs.st-andrews.ac.uk/Mathematicians.

$$\varphi_\Delta(x) := \begin{cases} \frac{x}{\Delta}, & 0 \le x < \Delta, \\ 1, & \Delta \le x < 1, \\ -\frac{1}{\Delta} \cdot (x - 1 - \Delta), & 1 \le x < 1 + \Delta, \\ 0, & \text{else.} \end{cases}$$

Remark 1.1. The two extreme situations are obtained for $\Delta = 1$ and $\Delta \to 0$, when φ_Δ reduces to B-spline functions of first order, also called *hat-functions*, and to *piecewise constant* functions, respectively. The gap in between can be smoothly covered by letting Δ be in the interval $(0, 1]$.

Furthermore, for $i = -1, \ldots, 2^n - 1$, $n \in \mathbb{N}$, we define by dilatation and translation of φ_Δ the following family of (fundamental) functions:

$$\varphi_i^n(x) := \varphi_\Delta(2^n x - i), \quad x \in [0, 1]. \tag{1}$$

In Figure 1 the functions φ_i^n, $i = -1, \ldots, 2^n - 1$, with a parameter $0 < \Delta < 1$ are depicted. Notice that the support of $\varphi_0^n, \ldots, \varphi_{2^n-2}^n$ is fully inside $[0, 1]$, whereas φ_{-1}^n and $\varphi_{2^n-1}^n$ can be viewed as "incomplete".

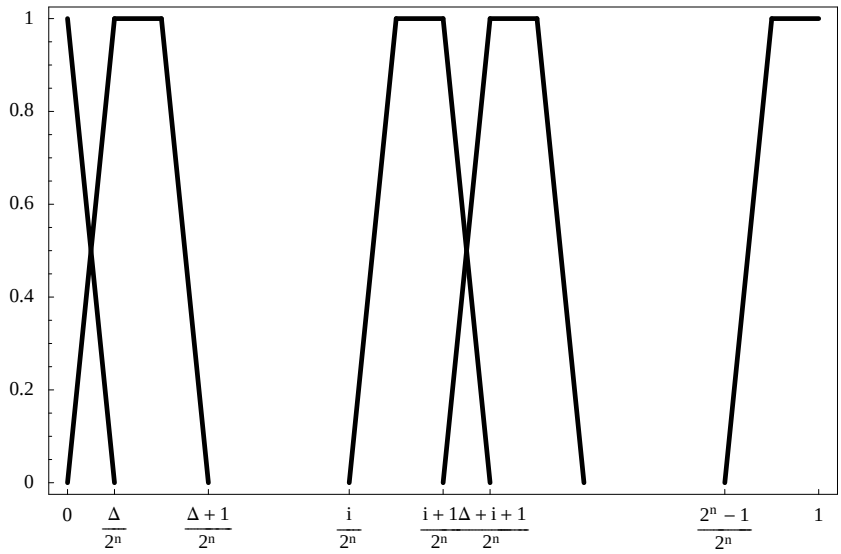


FIGURE 1

Also of great relevance are the midpoints η_i^n of the support line of each φ_i^n. Thus, for $i = 0, \ldots, 2^n - 2$, we have

$$\eta_i^n := \frac{i}{2^n} + \frac{1}{2} \cdot \frac{1+\Delta}{2^n},$$

and for $i \in \{-1, 2^n - 1\}$ we set

$$\eta_{-1}^n := 0 \text{ and } \eta_{2^n-1}^n := 1.$$

Equipped with these notations we can introduce the following operator.

Definition 1.2. For $f \in C[0,1]$ and $x \in [0,1]$ the *BLaC operator* is given by

$$BL_n(f;x) := \sum_{i=-1}^{2^n-1} f(\eta_i^n) \cdot \varphi_i^n(x). \tag{2}$$

(The abbreviation BLaC refers to "Blending of Linear and Constant".)

We first list some elementary facts.

Proposition 1.3.

(i) $BL_n : C[0,1] \to C[0,1]$ *is positive and linear;*

(ii) BL_n *interpolates* f *at the points* η_i^n, $i = -1, \ldots, 2^n - 1$ *(thus also at the endpoints* 0 *and* 1*);*

(iii) $\sum_{i=-1}^{2^n-1} \varphi_i^n(x) = 1$, *i.e.,* BL_n *reproduces constant functions.*
Hence $\|BL_n\| = 1$.

Proof. (i) This is obvious from the definition and the positivity of φ_i^n.

(ii) One can easily observe that $\varphi_i^n(\eta_j^n) = \delta_{i,j}$ (the Kronecker symbol) for $i, j = -1, \ldots, 2^n - 1$. Thus $BL_n(f;\eta_j^n) = f(\eta_j^n) \cdot \varphi_j^n(\eta_j^n) = f(\eta_j^n)$, for $j = -1, \ldots, 2^n - 1$.

(iii) For $x = 1$ we have $\sum_{i=-1}^{2^n-1} \varphi_i^n(1) = \varphi_{2^n-1}^n(1) = 1$.
Let $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$, $k \in \{0, \ldots, 2^n - 1\}$. We discuss separately:

Case 1: For $x \in \left[\frac{k}{2^n}, \frac{k+\Delta}{2^n}\right)$, we have

$$\begin{aligned}\sum_{i=-1}^{2^n-1} \varphi_i^n(x) &= \varphi_{k-1}^n(x) + \varphi_k^n(x) = \varphi_\Delta(2^n x - (k-1)) + \varphi_\Delta(2^n x - k) \\ &= -\frac{1}{\Delta}(2^n x - k - \Delta) + \frac{2^n x - k}{\Delta} = 1.\end{aligned}$$

Case 2: For $x \in \left[\frac{k+\Delta}{2^n}, \frac{k+1}{2^n}\right)$ we get $\sum_{i=-1}^{2^n-1} \varphi_i^n(x) = \varphi_k^n(x) = 1$, due to the definition of φ_Δ.

Hence $\sum_{i=-1}^{2^n-1} \varphi_i^n(x) = 1$ for all $x \in [0,1]$. $\square$

2. Degree of approximation by the BL_n operator

In the present section we investigate the degree of approximation by the BLaC-operator BL_n. The estimates are given in terms of the first and second order modulus of continuity. We use the following results given by the first author. Here and in the sequel we put $e_1(t) := t$ for $t \in [a,b]$.

Theorem 2.1. *For a positive linear operator $L : C[a,b] \to B(Y)$, $Y \subseteq [a,b]$ that reproduces constant functions the following inequality holds:*

$$|L(f;x) - f(x)| \leq \max\left\{1, \frac{1}{\delta} \cdot L(|e_1 - x|; x)\right\} \cdot \tilde{\omega}_1(f;\delta)$$

for all $f \in C[a,b]$, $x \in Y$ and $\delta > 0$.

Here $\tilde{\omega}_1(f;\cdot)$ denotes the least concave majorant of the (classical) first order modulus of continuity of $f \in C[a,b]$.

The above theorem can be formulated for general compact spaces, this version can be found in [5] (see also [6]).

We also have

Corollary 2.2. *Under the assumptions of Theorem 2.1 there holds*

$$|L(f;x) - f(x)| \leq 2 \cdot \max\left\{1, \frac{1}{\delta} \cdot L(|e_1 - x|; x)\right\} \cdot \omega_1(f;\delta),$$

where $f \in C[a,b]$, $x \in Y$ and $\delta > 0$.

We recall here also a general quantitative result involving ω_2; such estimates were first established by H. Gonska (see [6]) and later refined by R. Păltănea (see [8] or [9]). Păltănea's result reads as follows.

Theorem 2.3. *If Y is a subset of $[a,b]$, and if $L : C[a,b] \to B(Y)$ is a positive linear operator satisfying $L(e_0;x) = 1$ for all $x \in Y$, then for $f \in C[a,b]$, $x \in Y$ and $0 < \delta < \frac{b-a}{2}$ one has*

$$|L(f;x) - f(x)| \leq |L(e_1;x) - x| \cdot \frac{1}{\delta} \cdot \omega_1(f;\delta) + \left(1 + \frac{1}{2} \cdot \frac{1}{\delta^2} L((e_1 - x)^2;x)\right) \cdot \omega_2(f;\delta).$$

We establish next two quantitative statements, one in terms of ω_1, the second one involving both ω_1 and ω_2.

Proposition 2.4. *For any $f \in C[0,1] \to C[0,1]$ and $x \in [0,1]$ there holds*

$$|BL_n(f;x) - f(x)| \leq 2 \cdot \omega_1\left(f; \frac{1}{2^n}\right). \tag{3}$$

Proof. First we prove that

$$|BL_n(|e_1 - x|;x)| \leq \frac{1}{2^n}, \text{ for all } x \in [0,1].$$

We have $BL_n(|e_1 - x|;x) = \sum_{i=-1}^{2^n-1} |\eta_i^n - x| \cdot \varphi_i^n(x)$. We suppose that $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$, $k \in \{0, \ldots, 2^n - 1\}$. This excludes only $x = 1$ in which case we have $BL_n(|e_1 - 1|;1) = 0$.
Case 1: For $x \in \left[\frac{k}{2^n}, \frac{k+\Delta}{2^n}\right)$, we get

$$\begin{aligned} BL_n(|e_1 - x|;x) &= (x - \eta_{k-1}^n) \cdot \varphi_{k-1}^n(x) + (\eta_k^n - x) \cdot \varphi_k^n(x) \\ &= (x - \eta_{k-1}^n) \cdot \varphi_{k-1}^n(x) + (\eta_k^n - x) \cdot (1 - \varphi_{k-1}^n(x)) \\ &\leq \max\{\eta_k^n - x, x - \eta_{k-1}^n\} \leq (\eta_k^n - x + x - \eta_{k-1}^n) = \eta_k^n - \eta_{k-1}^n. \end{aligned}$$

Thus, for $k = 0$ we have $BL_n(|e_1 - x|;x) \leq \eta_0^n - \eta_{-1}^n = \frac{1}{2} \cdot \frac{1+\Delta}{2^n} \leq \frac{1}{2^n}$. For $k > 0$ we get $BL_n(|e_1 - x|;x) \leq \eta_k^n - \eta_{k-1}^n = \frac{k}{2^n} - \frac{k-1}{2^n} = \frac{1}{2^n}$.
Case 2: $x \in \left[\frac{k+\Delta}{2^n}, \frac{k+1}{2^n}\right)$. Then

$$BL_n(|e_1 - x|;x) = |\eta_k^n - x| \cdot \varphi_k^n(x) = |\eta_k^n - x| \leq \frac{1 - \Delta}{2^{n+1}} \leq \frac{1}{2^n}.$$

Thus $BL_n(|e_1 - x|;x) \leq \frac{1}{2^n}$, for all $x \in [0,1]$. Applying Corollary 2.2 with $\delta = \frac{1}{2^n}$ yields the estimate (3). □

Proposition 2.5. *For any* $f \in C[0,1] \to C[0,1]$, all $x \in [0,1]$ *and* $0 < \delta < \frac{1}{2}$ *the following inequality holds:*

$$|BL_n(f;x) - f(x)| \leq \frac{1-\Delta}{2^{n+1}} \cdot \frac{1}{\delta} \cdot \omega_1(f;\delta) + \left[1 + \frac{1}{2 \cdot \delta^2} \cdot \frac{1}{2^{2n}}\right] \cdot \omega_2(f;\delta). \tag{4}$$

Proof. In order to apply Theorem 2.3 we have to find suitable upper bounds for $BL_n(e_1 - x; x)$ and for $BL_n((e_1 - x)^2; x)$. In both cases the approach is the same as for $BL_n(|e_1 - x|; x)$. First note that $BL_n(e_1 - 1; 1) = 0$ and $BL_n((e_1 - 1)^2; 1) = 0$. We consider again two cases:

Case 1: $x \in \left[\frac{k}{2^n}, \frac{k+\Delta}{2^n}\right)$, $k \in \{0, \ldots, 2^n - 1\}$.

First we deal with the case $k = 0$. Here we have

$$BL_n(e_1 - x; x) = (\eta^n_{-1} - x) \cdot \varphi^n_{-1}(x) + (\eta^n_0(x) - x) \cdot \varphi^n_0(x)$$

and after some elementary computations we obtain in this case

$$BL_n(e_1 - x; x) = \frac{x(1-\Delta)}{2\Delta} \leq \frac{\Delta}{2^n} \cdot \frac{1-\Delta}{2\Delta} = \frac{1-\Delta}{2^{n+1}}.$$

For $1 \leq k \leq 2^n - 1$ we write successively:

$$\begin{aligned}
BL_n(e_1 - x; x) &= (\eta^n_{k-1} - x) \cdot \varphi^n_{k-1}(x) + (\eta^n_k - x) \cdot \varphi^n_k(x) \\
&= \frac{1}{2^{n+1}} \cdot \frac{1}{\Delta} \left[(2k - 1 + \Delta - 2^{n+1}x)(-2^n x + k + \Delta)\right. \\
&+ \left.(2k + 1 + \Delta - 2^{n+1}x) \cdot (2^n x - k)\right] \\
&= \frac{1}{2^{n+1}} \cdot \frac{1}{\Delta} \left[(2^n x - k) \cdot (2 - 2\Delta) + \Delta(-1 + \Delta)\right] \\
&= \frac{1}{2^{n+1}} \cdot \frac{1-\Delta}{\Delta} \left[2(2^n x - k) - \Delta\right] \\
&\leq \frac{1}{2^{n+1}} \cdot \frac{1-\Delta}{\Delta} \left[2\left(2^n \cdot \frac{k+\Delta}{2^n} - k\right) - \Delta\right] = \frac{1-\Delta}{2^{n+1}}.
\end{aligned}$$

We proceed in a similar way for the second moments. Hence we get

$$\begin{aligned}
BL_n((e_1 - x)^2; x) &= (x - \eta^n_{k-1})^2 \cdot \varphi^n_{k-1}(x) + (\eta^n_k - x)^2 \cdot \varphi^n_k(x) \\
&\leq \max\{(x - \eta^n_{k-1})^2, (\eta^n_k - x)^2\} \leq \left(\max\{(x - \eta^n_{k-1}), (\eta^n_k - x)\}\right)^2 \\
&\leq \left(\frac{1}{2^n}\right)^2 = \frac{1}{2^{2n}}.
\end{aligned}$$

Case 2: $x \in \left[\frac{k+\Delta}{2^n}, \frac{k+1}{2^n}\right)$, $k \in \{0, \ldots, 2^n - 1\}$. For the first moment we arrive at

$$|BL_n(e_1 - x; x)| \leq BL_n(|e_1 - x|; x) \leq \frac{1-\Delta}{2^{n+1}},$$

and for the second moment we have

$$BL_n((e_1 - x)^2; x) = (x - \eta_k^n)^2 \cdot \varphi_k^n(x) = (x - \eta_k^n)^2 \cdot 1 \leq \left(\frac{1-\Delta}{2^{n+1}}\right)^2 \leq \frac{1}{2^{2n}}.$$

Thus, we proved that for all $x \in [0, 1]$

$$|BL_n(e_1 - x; x)| \leq \frac{1-\Delta}{2^{n+1}} \text{ and } BL_n((e_1 - x)^2; x) \leq \frac{1}{2^{2n}}.$$

An application of Theorem 2.3 gives the statement (4). □

Proposition 2.6. *For the particular choice* $\delta = \frac{1}{2^n}$, $n \geq 1$, *the estimate (4) becomes*

$$|BL_n(f; x) - f(x)| \leq \frac{(1-\Delta)}{2} \cdot \omega_1\left(f; \frac{1}{2^n}\right) + \frac{3}{2} \cdot \omega_2\left(f; \frac{1}{2^n}\right). \tag{5}$$

Remark 2.7. BL_n is an approximation operator, i.e., $BL_n f$ converges uniformly towards f, $f \in C[0,1]$ as $n \to \infty$, see (5). For $\Delta = 1$, i.e., for *piecewise linear interpolation* at $0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n-1}{2^n}, 1$ the first term in (5) vanishes and we obtain a well-known inequality for polygonal line interpolation at the knots listed above. In fact, it was our aim to obtain for the first moments of the operator an upper bound involving the term $1 - \Delta$, in order to have it vanish for the piecewise linear interpolators.

3. Multivariate approximation

In the sequel we present statements on the degrees of approximation in the bivariate case. Only the *tensor product* case of the BL_n operators will be discussed here, but similar results can be given for Boolean sums as well. A general background on tensor products of univariate operators is provided by [2], [3] and the references cited therein. For our purposes we employ a convenient inheritance theorem that can be found in [1].

The quantitative results will be given in terms of *partial and total moduli of smoothness* of order r, $r \in \{1,2\}$, defined on compact intervals $I, J \subset \mathbb{R}$, for $f \in C(I \times J)$ and $\delta \geq 0$. We recall here their definitions.

$$\omega_r(f;\delta,0) := \sup\left\{\left|\sum_{\nu=0}^{r}(-1)^{r-\nu}\binom{r}{\nu}\cdot f(x+\nu h,y)\right| : (x,y),(x+rh,y)\in I\times J, |h|\leq\delta\right\}$$

and

$$\omega_r(f;0,\delta) := \sup\left\{\left|\sum_{\nu=0}^{r}(-1)^{r-\nu}\binom{r}{\nu}\cdot f(x,y+\nu h)\right| : (x,y),(x,y+rh)\in I\times J, |h|\leq\delta\right\}.$$

The total moduli of smoothness are

$$\begin{aligned}\omega_r(f;\delta_1,\delta_2) \quad := \quad & \sup\left\{\left|\sum_{\nu=0}^{r}(-1)^{r-\nu}\binom{r}{\nu}\cdot f(x+\nu h_1,y+\nu h_2)\right| : \right.\\ & \left.(x,y),(x+rh_1,y+rh_2)\in I\times J, |h_1|\leq\delta_1, |h_2|\leq\delta_2\right\}.\end{aligned}$$

Remark 3.1. The following relation holds between the two types of moduli

$$\{\omega_r(f;\delta_1,0),\ \omega_r(f;0,\delta_2)\} \leq \omega_r(f;\delta_1,\delta_2). \tag{6}$$

The inheritance principle mentioned involves discretely defined operators $L : C(I) \to C(I')$ and $M : C(J) \to C(J')$, where $I' \subseteq I, J' \subseteq J$ are non-trivial compact intervals of the real axis $\mathbb{R}$, and their *parametric extensions* to $C(I \times J)$. L and M are defined on finitely many, mutually distinct points x_e, $e \in E$, and $y_f, f \in F$, (with suitable index sets E and F), and have the form

$$\begin{aligned}L(g;x) &= \sum_{e\in E} g(x_e)\cdot A_e(x),\\ M(h;y) &= \sum_{f\in F} h(y_f)\cdot B_f(y),\end{aligned}$$

with $A_e \in C(I')$ and $B_f \in C(J')$ as fundamental functions. Consequently, their *parametric extensions* to $C(I \times J)$ are given by

$$\begin{aligned}{}_xL(f;x,y) &= L(f_y;x) = \sum_{e\in E} f_y(x_e)\cdot A_e(x) = \sum_{e\in E} f(x_e,y)\cdot A_e(x),\\ {}_yM(f;x,y) &= M(f_x;y) = \sum_{f\in F} f_x(y_f)\cdot B_f(y) = \sum_{f\in F} f(x,y_f)\cdot B_f(y),\end{aligned}$$

with $f \in C(I \times J)$ and $(x, y) \in I \times J$.

For discretely defined operators we have the following representation of the tensor product of L and M

$$({}_xL \circ {}_yM)(f; x, y) = \sum_{e \in E} \sum_{f \in F} f(x_e, y_f) \cdot A_e(x) \cdot B_f(y),\ f \in C(I \times J) \tag{7}$$

(and similarly for ${}_yM \circ {}_xL$).

We use the following general quantitative result regarding tensor products.

Theorem 3.2. *(see [Th. 37, 1]) Let L and M be defined as above and such that for fixed $r, s \in \mathbb{N}_0$*

$$\begin{aligned} |L(g; x) - g(x)| &\leq \sum_{\rho=0}^{r} \Gamma_{\rho,L}(x) \cdot \omega_\rho(g; \Lambda_{\rho,L}(x)),\ x \in I', g \in C(I) \text{ and} \\ |M(h; y) - h(y)| &\leq \sum_{\gamma=0}^{s} \Gamma_{\gamma,M}(y) \cdot \omega_\gamma(h; \Lambda_{\gamma,M}(y)),\ y \in J', h \in C(J). \end{aligned}$$

Here, Γ and Λ are bounded functions.

(i) Then for $(x, y) \in I' \times J'$ and $f \in C(I \times J)$ the following hold:

$$\begin{aligned} |({}_xL \circ {}_yM)f(x, y) - f(x, y)| &\leq \sum_{\rho=0}^{r} \Gamma_{\rho,L}(x) \cdot \omega_\rho(f; \Lambda_{\rho,L}(x), 0) \\ &+ \|L\| \cdot \sum_{\gamma=0}^{s} \Gamma_{\gamma,M}(y) \cdot \omega_\gamma(f; 0, \Lambda_{\gamma,M}(y)). \end{aligned}$$

(ii) A symmetric upper bound is given by

$$\sum_{\gamma=0}^{s} \Gamma_{\gamma,M}(y) \cdot \omega_\gamma(f; 0, \Lambda_{\gamma,M}(y)) + \sum_{\rho=0}^{r} \Gamma_{\rho,L}(x) \cdot \omega_\rho(f; \Lambda_{\rho,L}(x), 0).$$

From (7) we immediately get the explicit representation of the tensor product of two BLaC operators

$$({}_xBL_n \circ {}_yBL_m)(f; x, y) = \sum_{i=-1}^{2^n-1} \sum_{j=-1}^{2^m-1} f(\eta_i^n, \eta_j^m) \cdot \varphi_i^n(x) \cdot \varphi_j^m(y),$$

and can state

Theorem 3.3. *For $n, m \in \mathbb{N}$ we have*

$$\begin{aligned} \|({}_xBL_n \circ {}_yBL_m)f - f\| &\leq (1-\Delta)\omega_1\left(f; \frac{1}{2^n}, 0\right) + \frac{3}{2}\omega_2\left(f; \frac{1}{2^n}, 0\right) \\ &+ (1-\Delta)\omega_1\left(f; 0, \frac{1}{2^m}\right) + \frac{3}{2}\omega_2\left(f; 0, \frac{1}{2^m}\right) \\ &\leq 2(1-\Delta)\omega_1\left(f; \frac{1}{2^n}, \frac{1}{2^m}\right) + 3\omega_2\left(f; \frac{1}{2^n}, \frac{1}{2^m}\right). \end{aligned}$$

Proof. The proof is immediate. Take in Theorem 3.2 $r = s = 2$, $\Gamma_0(x) = 0, \Gamma_1(x) = 1 - \Delta, \Gamma_2(x) = \frac{3}{2}$ and $\Lambda_1(x) = \Lambda_2(x) = \frac{1}{2^n}$, make an analogous choice with respect to the variable y and use the relation in Proposition 2.6 twice. For the last inequality use (6). □

Remark 3.4. Similar results can be also achieved for Boolean sums of two BLaC operators, using, for example, Th. 31 from [1].

A practical application of the bivariate case is image compression. In the diploma thesis [7] of the third author a method is implemented that enables us to choose in an appropriate way the parameter Δ for a given picture (part of it). Examples are given to illustrate the fact that in most cases it is better to choose Δ not equal to 0 or 1, in order to obtain a more satisfying picture.

References

[1] Beutel, L., Gonska, H., Kacsó, D., and Tachev, G., *On variation-diminishing Schoenberg operators: new quantitative statements*, In: "Multivariate Approximation and Interpolation with Applications" (ed. by M. Gasca), Monogr. Academia Ciencias de Zaragoza **20**(2002), 9-58.

[2] Beutel, L., Gonska, H., *Quantitative inheritance properties for simultaneous approximation by tensor product operators*, Numerical Algorithms **33**(2003), 83-92.

[3] Beutel, L., and Gonska, H., *Quantitative inheritance properties for simultaneous approximation by tensor product operators II: applications*, In: "Mathematics and its Applications" (Proc. 17th Scientific Session; ed. by G.V.Orman), 1-28. Braşov: Editura Universităţii "Transilvania" 2003.

[4] Bonneau, G. P., *Multiresolution analysis with non-nested spaces*, In: "Tutorials on Multiresolution in Geometric Modelling" (Lectures European School on Principles of Multiresolution in Geometric Modelling, Munich University of Technology, August 2001; ed. by A. Iske, E. Quak and M. S. Floater), 147-163. Berlin: Springer 2002.

[5] Gonska, H., *On approximation in spaces of continuous functions*, Bull. Austral. Math. Soc. **28**(1983), no. 3, 411-432.

[6] Gonska, H., *On approximation by linear operators: improved estimates*, Anal. Numér. Théor. Approx. **14**(1985), 7-32.

[7] Nemitz, O., *BLaC-Wavelets und nicht ineinander geschachtelte Wavelets*, Diploma thesis, University of Duisburg-Essen 2003.

[8] Păltănea, R., *Best constants in estimates with second order moduli of continuity*, In: Approx. Theory, (Proc. Int. Dortmund meeting on Approx. Theory 1995, ed. by M. W. Müller et al.), Berlin, Akad. Verlag 1995, 251-275.

[9] Păltănea, R., *Approximation Theory using Positive Linear Operators*, Boston, Birkhäuser 2004.

Dept. of Mathematics, University of Duisburg-Essen, D-47048 Duisburg, Germany
E-mail address: gonska@math.uni-duisburg.de

Faculty of Mathematics, Ruhr-Universität Bochum, D-44780 Bochum, Germany
E-mail address: Daniela.Kacso@rub.de

Inst. für Numerische Simulation, University of Bonn, D-53115 Bonn, Germany
E-mail address: oliver.nemitz@ins.uni-bonn.de

Colegiul Naţional "Samuel von Brukenthal", RO-550182 Sibiu, Romania
E-mail address: pitul_paula@yahoo.com

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

THERMAL RADIATION EFFECT ON FULLY DEVELOPED FREE CONVECTION IN A VERTICAL RECTANGULAR DUCT

T. GROŞAN, T. MAHMOOD, AND I. POP

Dedicated to Professor Gheorghe Coman at his 70^{th} *anniversary*

Abstract. The effect of radiation on the steady free convection flow, i.e. the case of purely buoyancy-driven flow, in a vertical rectangular duct is investigated for laminar and fully developed regime. The Rosseland approximation is considered and temperatures of the walls are assumed constants. The governing equations are expressed in non-dimensional form and are solved both analytically and numerically. It was found that the governing parameters have a significant effect on the velocity and temperature profiles.

1. Introduction

Heat transfer in free and mixed convection in vertical channels occurs in many industrial processes and natural phenomena. It has therefore been the subject of many detailed, mostly numerical studies for different flow configurations. The fluid flow and heat transfer has been the subject of many recent books, such as, for example Bejan [1], Pop and Ingham [2], Kohr and Pop [3], etc. Most of the interest in this subject is due to its applications, for instance, in the design of cooling systems for electronic devices and in the field of solar energy collection. Some of the published papers on this topic, such as Aung [4], Aung et al.[5], Aung and Worku [6,7], Barletta [8,9], and Boulama and Galanis [10], deal with the evaluation of the temperature and velocity profiles for the vertical parallel-flow fully developed regime. As is well

Received by the editors: 25.04.2006.

2000 *Mathematics Subject Classification.* 76D05, 80A20.

Key words and phrases. porous media, non-Darcy law, boundary layer, radiation.

known, heat exchangers technology involves convective flows in vertical channels. In most cases, these flows imply conditions of uniform heating of a channel, which can be modelled either by uniform wall temperature (UWT) or uniform heat flux (UHF) thermal boundary conditions.

All the above quoted analyses of free and mixed convection flow in vertical channels are based on the hypothesis that the thermal radiation effect within the fluid is negligible. However, effects of conduction-radiation on convective flows are very important in the context of space technology and processes involving high temperatures. The inclusion of conduction-radiation effects in the energy equation however leads to a highly nonlinear partial or ordinary differential equations. The aim of the present paper is therefore to analyse the effects of thermal radiation on the steady fully developed free convection in a vertical channel such that the walls of the channels are subjected to uniform but different wall temperatures (UWT) using the Rosseland approximation model which leads to ordinary differential equations for the free convection flow of an optically dense viscous incompressible fluid that flows through the channel. The ordinary differential equations are solved both analytically and numerically using the Runge-Kutta method. Flow and heat transfer results for a range of values of the pertinent parameters have been reported. Effects of pertinent parameters, such as the radiation parameter, Rd , and the thermal parameter θ_R velocity and temperature profiles are shown graphically.

2. **Basic equations**

Consider a viscous and incompressible fluid, which steadily flows between two infinite vertical and parallel plane walls. The distance between the walls, i.e., the channel width, is L . A coordinate system is chosen such that the $x-$ axis is parallel to the gravitational acceleration vector $\mathbf{g}$, but with the opposite direction. The $y-$ axis is orthogonal to the channel walls, and the origin of the axes is such that the positions of the channel walls are $-L/2$ and $L/2$, respectively. A sketch of the system and of the coordinate axes is reported in Figure 1. The wall at $y = -L/2$ is at the given uniform temperature T_1, while the wall at $y = L/2$ is subjected to a uniform

temperature T_2, where $T_2 > T_1$. The fluid velocity $\mathbf{v}(u, v)$ is assumed to be parallel to the $x-$ axis, so that only the $x-$ component u of the velocity vector does not vanish. The Boussinesq and Rosseland approximations are employed. Fluid rises in the duct driven by buoyancy forces. Hence the flow is due to difference in temperature and the convection sets in instantaneously. Moreover the gradient of $T_2 - T_1$ is perpendicular to the gravity which we call it as Oberbeck convection and therefore there will be no pressure gradient in the basic equation. All the fluid properties except density in the buoyancy term are considered as constant. The flow being fully developed the following relations apply here

$$v = 0, \frac{\partial v}{\partial y} = 0, \frac{\partial p}{\partial x} = \frac{\partial p}{\partial y} = 0 \tag{1}$$

where p is the fluid pressure. Therefore, the continuity equation gives $\partial u/\partial x = 0$. One can thus conclude that u does not depend on x, i.e. $u = u(y)$. Under these assumptions the momentum and energy equations for the flow and heat transfer are

$$\mu \frac{\partial^2 u}{\partial y^2} + \rho_0 g \beta \left(T - T_0\right) = 0 \tag{2}$$

$$k \frac{\partial^2 T}{\partial y^2} - \frac{\partial q^r}{\partial y} = 0 \tag{3}$$

where T is the fluid temperature, g is the acceleration due to gravity, k is the thermal conductivity, β is the thermal expansion coefficient, μ is the dynamic viscosity, ρ_0 is the characteristic density, q^r is the radiative heat flux and T_0 is the characteristic temperature. We assume that q^r under the Rosseland approximation has the form

$$q^r = -\left(\frac{4\sigma}{3\chi}\right) \frac{\partial T^4}{\partial y} \tag{4}$$

where σ is the Stefan-Boltzman's constant and χ is the mean absorption coefficient. We also assume that . Equations (2) and (3) have to be solved subject to the boundary conditions

$$u\left(\mp L/2\right) = 0, T\left(-L/2\right) = T_1, T\left(L/2\right) = T_2 \tag{5}$$

In order to solve Eqs. (2) and (3), we introduce the following non-dimensional variables

$$Y = \frac{y}{L}, U(Y) = \frac{u}{U_0}, \theta(Y) = \frac{T - T_0}{T_2 - T_1} \tag{6}$$

where $U_0 = g\beta(T_2 - T_1)$ is the characteristic velocity. Substituting (6) into Eq. (2) and (3), we get the following ordinary differential equations

$$\frac{d^2U}{dY^2} + \theta = 0 \tag{7}$$

$$\frac{d}{dY}\left\{\left[1 + \frac{4}{3}Rd\,(1 + \theta_R\theta)^3\right]\frac{d\theta}{dY}\right\} = 0 \tag{8}$$

subject to the boundary conditions (5) which become

$$U\left(\mp\frac{1}{2}\right) = 0, \theta\left(-\frac{1}{2}\right) = -\frac{1}{2}, \theta\left(\frac{1}{2}\right) = \frac{1}{2} \tag{9}$$

where the radiation parameter Rd and the temperature parameter θ_R are given by

$$Rd = \frac{4\sigma T_0^3}{k\chi}, \theta_R = 2\frac{T_2 - T_1}{T_2 + T_1} \tag{10}$$

We notice that in the case when the radiation effect is absent ($Rd = 0$), Eqs. (7) and (8) reduce to those obtained by Aung [4]. The analytical solution of Eq. (7) and (8) can be expressed as

$$U = -\int_0^Y \int_0^s \theta ds dY + C_1 Y + C_2 \tag{11}$$

$$\theta + \frac{Rd}{4\theta_R}(1 + \theta_R\theta)^4 = C_3 Y + C_4 \tag{12}$$

where C_1, C_2, C_3 and C_4 are constants of integration. When $Rd = 0$ (radiation effect is absent), we get

$$\theta = Y, U = \frac{Y}{6}\left(\frac{1}{4} - Y\right) \tag{13}$$

3. Results and discussion

Equations (7) and (8), subject to the boundary conditions (9) were solved numerically using the finite-difference method for different values of the parameters $Rd = 0, 0.1, 1, 5, 10$ and $\theta_R = 1.1, 1.5, 2.0$.The velocity U and temperature θ profiles are shown on Figures 2 to 9. When the radiation is absent($Rd = 0$) one can see from Figures 4 to 9 that the numerical results are in very good agreement with the analytical solution. It means that we are confident that the present results are correct.

We notice that a reversed flow exist for small values of the radiation and temperature parameters, which is similar with the case studied by Aung [4] when the radiation is absent. The reversed flow dissapears for large values of the radiation and temperature parameters(see Figures 2, 4, 6 and 8). Further, we can see that the velocity profiles increase with the increasing of the temperature parameter θ_R (see Figure 2) and also with the increasing of the radiation parameter Rd(see Figures 4, 6 and 8).

The temperature profiles are shown in Figures 3, 5, 7 and 9. We can seen that the temperature profiles increase with the increasing of the parameters θ_R and Rd. The effect of the temperature parameter is more significant for larger values of the parameter Rd. We also notice that the radiation effects modify the simmetry of the temperature profiles, the temperature gradients are larger near the cold wall(left wall of the channel).

References

[1] Bejan, A., *Convection Heat Transfer, 2nd edition*, Wiley, New York, 1995.

[2] Pop, I. and Ingham, D.B., *Convective Heat Transfer: Mathematical and Computational Modelling of Viscous Fluids and Porous Media*, Pergamon, Oxford, 2001.

[3] Kohr, M., and Pop, I., *Viscous Incompressible Flow for Low Reynolds Numbers*, WIT Press, Southampton, 2004.

[4] Aung, W., *Fully developed laminar free convection between vertical plates heated asymmetrically*, Int. J. Heat Mass Transfer, **15**(1972), 1577-1580.

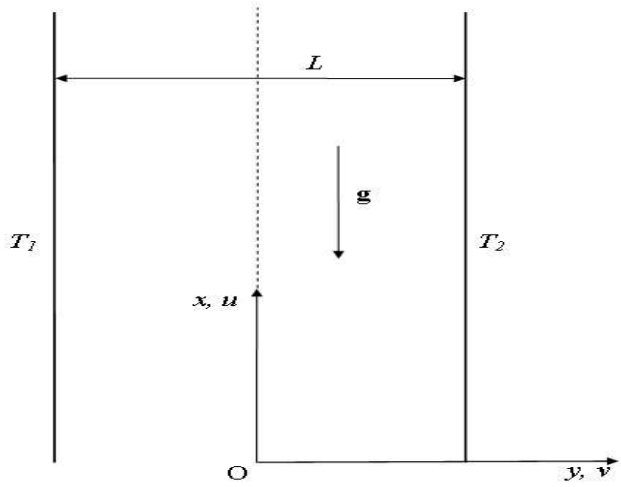


FIGURE 1. Phisycal model and co-ordinate system

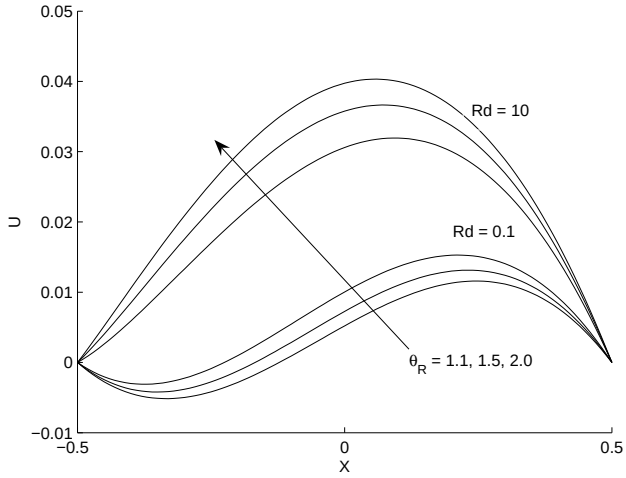


FIGURE 2. Dimensionless velocity profiles U for $Rd = 0.1$ and 10 and $\theta_R = 1.1, 1.5, 2.0$

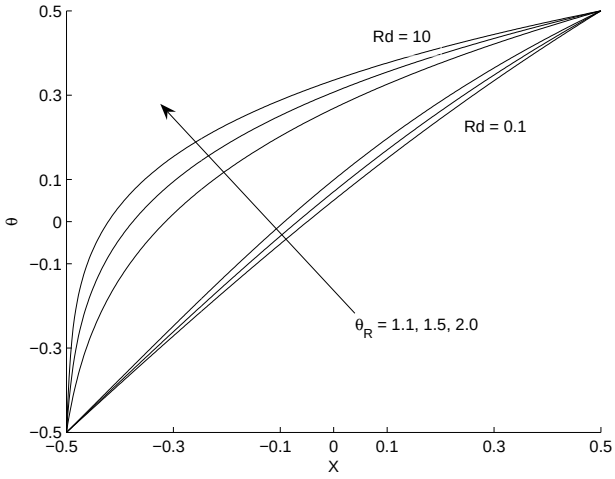


FIGURE 3. Dimensionless temperature profiles θ for $Rd = 0.1$ and 10 and $\theta_R = 1.1, 1.5, 2.0$

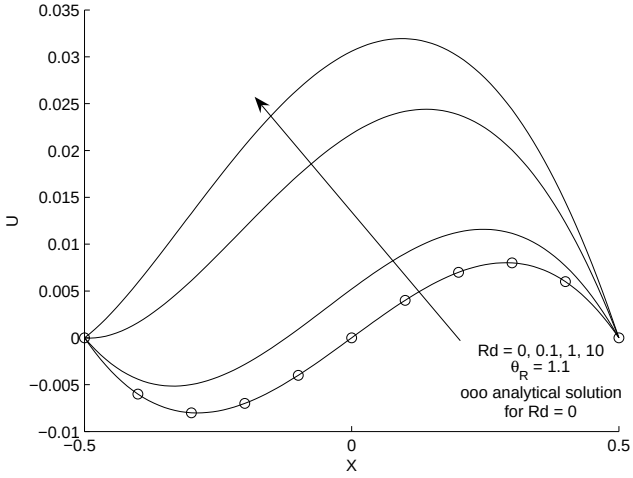


FIGURE 4. Dimensionless velocity profiles U for $Rd = 0, 0.1, 1, 10$ and $\theta_R = 1.1$

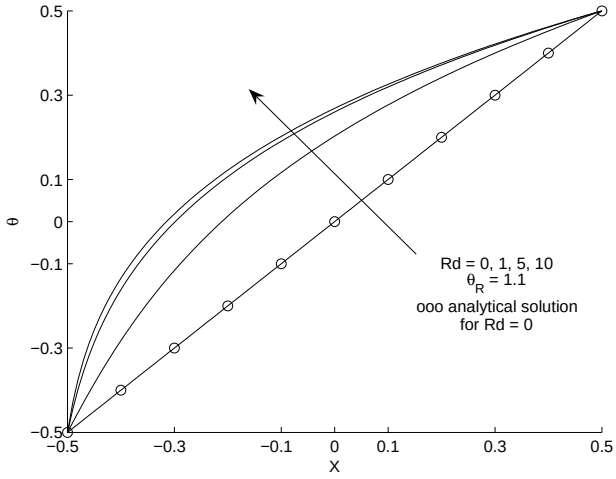


FIGURE 5. Dimensionless temperature profiles θ for $Rd = 0, 1, 5, 10$ and $\theta_R = 1.1$

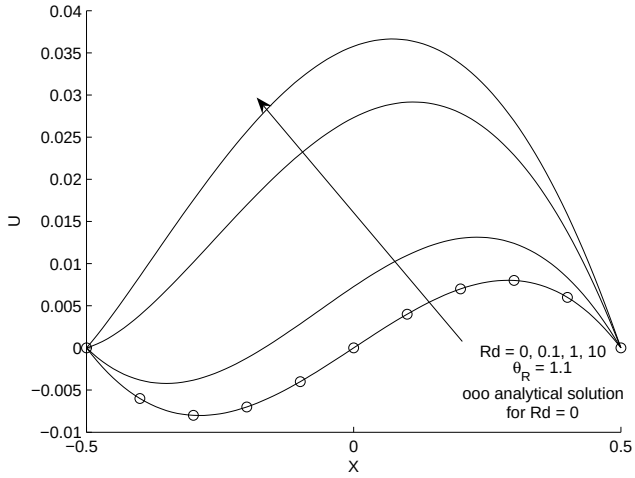


FIGURE 6. Dimensionless velocity profiles U for $Rd = 0, 0.1, 1, 10$ and $\theta_R = 1.5$

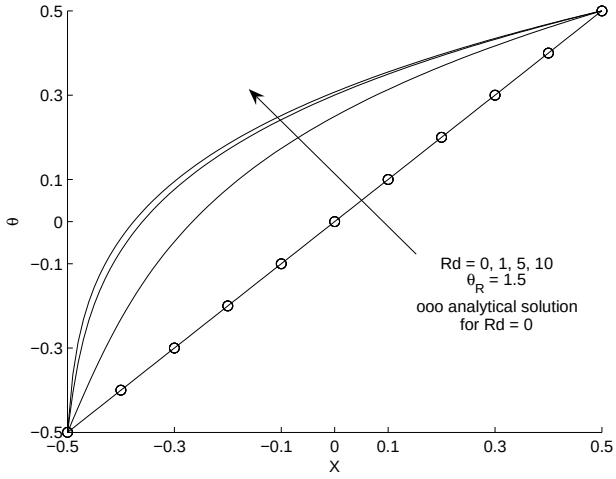


FIGURE 7. Dimensionless temperature profiles θ for $Rd = 0, 1, 5, 10$ and $\theta_R = 1.5$

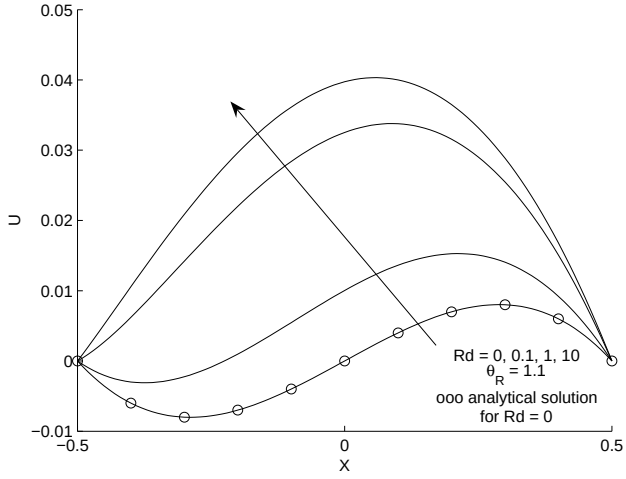


FIGURE 8. Dimensionless velocity profiles U for $Rd = 0, 0.1, 1, 10$ and $\theta_R = 2.0$

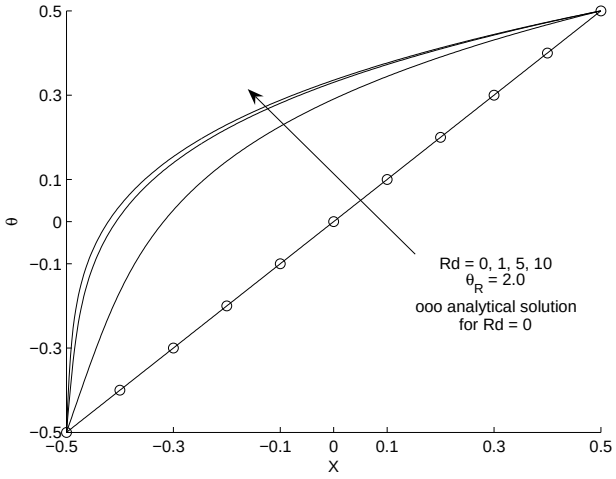


FIGURE 9. Dimensionless temperature profiles θ for $Rd = 0, 1, 5, 10$ and $\theta_R = 2.0$

[5] Aung, W., Fletcher, L.S., and Sernas, V., *Developing laminar free convection between vertical flat plates with asymmetric heating*, Int. J. Heat Mass Transfer, **15**(1972), 2293-2308.

[6] Aung, W., and Worku, G., *Developing flow and flow reversal in a vertical channel with asymmetric wall temperatures*, J. Heat Transfer, **108**(1986), 299-304.

[7] Aung, W., and Worku, G., *Theory of fully developed, combined convection including flow reversal*, J. Heat Transfer, **108**(1986), 485-488.

[8] Barletta, A., *Analysis of combined forced and free flow in a vertical channel with viscous dissipation and isothermal-isoflux boundary conditions*, J. Heat Transfer, **121**(1999), 349-356.

[9] Barletta, A., *Fully developed mixed convection and flow reversal in a vertical rectangular duct with uniform wall heat flux*, Int. J. Heat Mass Transfer, **45**(2002), 641-654.

[10] Boulama, K., and Galanis, N., *Analytical solution for fully developed mixed convection between parallel vertical plates with heat and mass transfer*, J. Heat Transfer, **126**(2004), 381-388.

Faculty of Mathematics and Computer Science,
Babeş-Bolyai University Cluj-Napoca, Romania
E-mail address: tgrosan@math.ubbcluj.ro

Department of Mathematics,
The Islamia University of Bahawal Pur, Pakistan

Faculty of Mathematics and Computer Science,
Babeş-Bolyai University Cluj-Napoca, Romania

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

PROFESSOR GHEORGHE COMAN AT HIS 70TH ANNIVERSARY

OCTAVIAN AGRATINI AND PETRU BLAGA

Gheorghe Coman was born on January 24th 1936, in Grindeni, Mureş. He attended elementary school (1944-1948) in his home-village, secondary school (1948-1951) in the town of Luduş and high-school (1951-1956) in Cluj, a rich cultural and historical city located on the Someş river banks. From 1956 to 1961 he was a student at the Faculty of Mathematics and Mechanics (nowadays the Faculty of Mathematics and Computer Science), Babeş-Bolyai University, Cluj. After graduation, he worked as an assistant professor in the Department of Numerical and Statistical Calculus. In this department, Gheorghe Coman has held a continuous academic career, being successively promoted to the positions of lecturer (1970), associate professor (1977) and full professor (1990).

In 1966 Gheorghe Coman married Mioara, a chemistry professor at the Medicine and Pharmacy University in Cluj-Napoca. They have two children: Dan, born in 1967, and Horia, born in 1970. Dan followed in his father's footsteps, becoming also a mathematician, while Horia is a medical doctor.

Under the guidance of the famous mathematician D.V. Ionescu (1901-1985), in 1970 Gheorghe Coman completed his doctoral thesis *Optimal quadrature and cubature formulas*. He also had the opportunity to participate in professional training modules abroad in Moskow (1968) and in the USA, University of Wisconsin, Madison (1973-1974).

In what follows, we briefly certify the outstanding scientific activity and teaching career of *Ghitza*, name under which professor Coman is known to those close to him. So far, twenty doctoral students have worked under his guidance: Călin Enăchescu, Sorin Pop (also supervised by professor W. Jaeger, Heidelberg), Daniela

Kacso (also supervised by professor H. Gonska, Duisburg), Andras Peter, Ioan Lazăr, Milena Solomon, Virginia Niculescu, Monica Vancea, Codruţa Vancea, Daniela Roşca, Ioana Pop, Ion Cozac, Iulia Costin, Cătălin Mitran, Teodora Gulea, Cristina Mihoc, Ioan Todea, Marius Birou, Ildiko Kovacs, Alexandra Oprişan.

Professor Coman is editor-in-chief at the Romanian journals: *Seminar on Numerical and Statistical Calculus* and *Studia Universitatis Babeş-Bolyai, Mathematica*. He is also included in the editorial board of the following journals edited by the Romanian Academy: *Mathematica* and *Revue d'Analyse Numerique et de Theorie de l'Approximation*. Since 1974 he has been member of the American Mathematical Society and reviewer for Mathematical Reviews.

Professor Coman has been invited to deliver talks at universities in France (Paris, 1994), Germany (Heidelberg, 1995; Duisburg, 1998), Hungary (Debrecen, 1991).

For many years, our colleague, a good organizer and honest judge, served the academic community of the Babeş-Bolyai University from the position of vicedean (1986-1989) and dean (1989-1996). A constant proof of his remarkable character is given by the fact that he was elected and re-elected dean, before and after 1989, the year of the Romanian Revolution.

A gifted teacher with a good understanding of students, a challenging partner for his colleagues, he succeeded in fascinating us with his spiritual youth, honesty in its purest form and, above all, his personal charm.

Coman's teaching and research activity is materialized in more than one hundred published papers and nine books (textbooks and monographs). The scientific work of professor Gheorghe Coman aims at Numerical Analysis and Approximation Theory. Regarding these contributions, we mention the following research directions: numerical integration of functions, approximation of uni and multivariate functions, optimizations of numerical methods with respect to the error, complexity and efficiency of calculus.

At the end we list the most relevant works of professor Coman.

List of Scientific Papers

[1] Agratini, O., Blaga, P., **Coman, Gh.,** *Lectures on wavelets, numerical methods and statistics*, Casa Cărţii de ştiinţă, Cluj-Napoca, 2005. iv+196 pp. MR 2006h:42001.

[2] **Coman, Gh.,** Todea, I., *On some applications of interpolation operators*, Studia Univ. Babeş-Bolyai Math., **50**(2005), no.1, 3-15. MR 2006e:41039.

[3] **Coman, Gh.,** Birou, M., Oşan, C., Somogyi, I., Cătinaş, T., Oprişan, A., Pop, I., Todea, I., *Interpolation operators. With a preface by Dimitrie D. Stancu*, Casa Cărţii de Ştiinţă, Cluj-Napoca, 2004. viii+298 pp. MR 2006i:41001.

[4] **Coman, Gh.,** Todea, I., *Bivariate Shepard operators of Abel-Gonciarov-type*, Studia Univ. Babeş-Bolyai Math., **48**(2003), no. 2, 39-48.

[5] **Coman, Gh.,** Pop, I., *Some interpolation schemes on triangle*, Studia Univ. Babeş-Bolyai Math., **48**(2003), no. 3, 57-62.

[6] **Coman, Gh.,** Birou, M., *Bivariate spline-polynomial interpolation*, Studia Univ. Babeş-Bolyai Math. **48**(2003), no. 4, 17-24. MR 2005g:41002.

[7] **Coman, Gh.,** Solomon, M., *Homogeneous numerical cubature formulas of interpolatory type*, Rev. Anal. Numér. Théor. Approx., **31**(2002), no. 1, 45-53 (2003). MR 2004j:65035.

[8] **Coman, Gh.,** Păvăloiu, I., *At the 75th birthday anniversary of academician Professor Dimitrie D. Stancu*, Rev. Anal. Numér. Théor. Approx., **31**(2002), no. 1, 5-7 (2003).

[9] **Coman, Gh.,** Mitran, C., *On some homogeneous cubature formulas. Numerical analysis and approximation theory*, (Cluj-Napoca, 2002), 148-162, Cluj Univ. Press, Cluj-Napoca, 2002. MR 2004h:65027.

[10] **Coman, Gh.,** Pop, I., Trîmbiţaş, R., *An adaptive cubature on triangle*, Studia Univ. Babeş-Bolyai Math., **47**(2002), no. 4, 27-36. MR 2004g:65024.

[11] **Coman, Gh.,** Rus, Ioan A., Ţâmbulea, L., *Professor Dimitrie D. Stancu, at his 75th birthday anniversary*, Studia Univ. Babeş-Bolyai Math., **47**(2002), no. 4, 3-12.

[12] Agratini, O., Chiorean, I., **Coman, Gh.,** Trîmbiţaş, R., *Analiză numerică şi teoria aproximării*, Vol. III. (Romanian) Presa Universitară Clujeană, Cluj-Napoca, 2002. xii+551 pp.

[13] Stancu, D. D., **Coman, Gh.,** Blaga, P., **Analiză numerică şi teoria aproximării**, Vol. II. (Romanian) Presa Universitară Clujeană, Cluj-Napoca, 2002. xii+433 pp.

[14] Gânscă, I., Bronsvoort, W. F., **Coman, Gh.,** Ţambulea, L., *Self-intersection avoidance and integral properties of generalized cylinders*, Comput. Aided Geom. Design, **19**(2002), no. 9, 695-707. MR 2003j:68142.

[15] **Coman, Gh.,** Trîmbiţaş, R., *Univariate Shepard-Birkhoff interpolation*, Dedicated to the memory of Acad. Tiberiu Popoviciu., Rev. Anal. Numér. Théor. Approx., **30**(2001), no. 1, 15-24. MR 2004m:41021.

[16] **Coman, Gh.,** Trîmbiţaş, R., *Multivariate Shepard interpolation. Symbolic and numeric algorithms on scientific computing* (Timişoara, 2001), An. Univ. Timişoara Ser. Mat.-Inform., **39**(2001), 39-48. MR 2004m:41033.

[17] Stancu, D. D., **Coman, Gh.,** Agratini, O., Trîmbiţaş, R., *Analiză numerică şi teoria aproximării*, Vol. I. (Romanian) Presa Universitară Clujeană, Cluj-Napoca, 2001. 414 pp.

[18] **Coman, Gh.,** *On D. V. Ionescu's practical numerical integration formulas*, Mathematical contributions of D. V. Ionescu, 69-76, Babeş-Bolyai Univ. Dept. Appl. Math., Cluj-Napoca, 2001.

[19] **Coman, Gh.,** Trîmbiţaş, R., *Combined Shepard univariate operators*, East J. Approx., **7**(2001), no. 4, 471-483. MR 2002j:41018.

[20] **Coman, Gh.,** Somogyi, I., *Homogeneous numerical integration formulas*, Analysis, functional equations, approximation and convexity (Cluj-Napoca, 1999), 45-49, Carpatica, Cluj-Napoca, 1999. (41A55)

[21] Gânscă, I., **Coman, Gh.,** Ţâmbulea, L., *Rational Bézier curves and surfaces with independent coordinate weights*, Studia Univ. Babeş-Bolyai Math., **43**(1998), no. 2, 29-38. MR 2002e:65026.

[22] Gânscă, Ioan, **Coman, Gh.,** Ţâmbulea, L., *Remodelling given Bézier spline curves and surfaces*, Studia Univ. Babeş-Bolyai Math., **43**(1998), no. 1, 29-38. MR 2002e:65025.

[23] **Coman, Gh.,** Purdea, I., *On the remainder term in multivariate approximation*, Studia Univ. Babeş-Bolyai Math., **43**(1998), no. 1, 7-14.

[24] **Coman, Gh.,** *Shepard operators of Birkhoff-type*, Calcolo, **35**(1998), no. 4, 197-203. MR 2001b:41026.

[25] **Coman, Gh.,** Trîmbiţaş, R., *On complexity of Romberg and adaptive-recursive numerical integration methods*, Mathematica, **40**(63)(1998), no. 1, 63-70. MR 2000h:65199.

[26] **Coman, Gh.,** *Hermite-type Shepard operators*, Rev. Anal. Numér. Théor. Approx., **26**(1997), no. 1-2, 33-38.

[27] Gânscă, I., **Coman, Gh.,** Ţâmbulea, L., *Contributions to rational Bézier curves and surfaces*, Approximation and optimization, Vol. II (Cluj-Napoca, 1996), 107-116, Transilvania, Cluj-Napoca, 1997. MR 99a:65023.

[28] **Coman, Gh.,** Breckner, W. W., Blaga, P., *Approximation and optimization*, Vol. II. Proceedings of the International Conference (ICAOR) held at Babeş-Bolyai University,

Cluj-Napoca, July 29-August 1, 1996, Edited by Dimitrie D. Stancu, Transilvania Press, Cluj-Napoca, 1997. viii+252 pp. MR 98g:41002.

[29] **Coman, Gh.,** Iancu, C., *Spline interpolation of Birkhoff type*, Approximation and optimization, Vol. I (Cluj-Napoca, 1996), 233-240, Transilvania, Cluj-Napoca, 1997. MR 98k:41014.

[30] **Coman, Gh.,** Breckner, W. W., Blaga, P., *Approximation and optimization*, Vol. I, Proceedings of the International Conference (ICAOR) held at Babeş-Bolyai University, Cluj-Napoca, July 29-August 1, 1996, Edited by Dimitrie D. Stancu, Transilvania Press, Cluj-Napoca, 1997. xiv+374 pp. MR 98g:41001.

[31] **Coman, Gh.,** Ţâmbulea, L., Gânscă, I., *Multivariate approximation*, Seminar on Numerical and Statistical Calculus, 29-60, Preprint, 96-1, Babeş-Bolyai Univ., Cluj-Napoca, 1996.

[32] **Coman, Gh.,** Iancu, C., *Lacunary interpolation by cubic spline*, Studia Univ. Babeş-Bolyai Math., **40**(1995), no. 4, 77-84. MR 98a:41015.

[33] Rus, I. A., Both, N., **Coman, Gh.,** Mihoc, I., Mihoc, M., Purdea, I., Ţarină, M., *Matematica şi aplicaţiile sale*, (Romanian) Editura Ştiinţifică, Bucharest, 1995. 360 pp. MR 97c:00003.

[34] **Coman, Gh.,** Gânscă, I., Ţâmbulea, L., *Blending approximation*, Romanian Symposium on Computer Science (Iaşi, 1993), 126-139, A. I. Cuza Univ. Iasi, Iaşi, 1994.

[35] Gânscă, I., **Coman, Gh.,** Ţâmbulea, L., *Generalizations of Bézier curves and surfaces*, Curves and surfaces in geometric design (Chamonix-Mont-Blanc, 1993), 169-176. MR 95g:65023.

[36] **Coman, Gh.,** *Homogeneous cubature formulas*, Studia Univ. Babeş-Bolyai Math., **38**(1993), no. 2, 91-101.

[37] **Coman, Gh.,** Gânscă, I., Ţâmbulea, L., *Surfaces generated by blending interpolation*, Studia Univ. Babeş-Bolyai Math., **38**(1993), no. 3, 39-48.

[38] **Coman, Gh.,** *Professor Dimitrie D. Stancu at his 65th birthday*, Studia Univ. Babeş-Bolyai Math., **37**(1992), no. 1, 113-128. MR 95k:01025.

[39] **Coman, Gh.,** Gânscă, I., Ţâmbulea, L., *New interpolation procedure in triangles*, Studia Univ. Babeş-Bolyai Math. **37**(1992), no. 1, 37-45.

[40] **Coman, Gh.,** Ţâmbulea, L., *Bivariate Birkhoff interpolation of scattered data*, Studia Univ. Babeş-Bolyai Math. **36**(1991), no. 2, 77-86.

[41] **Coman, Gh.**, Gânscă, I., Ţâmbulea, L., *Some new roof-surfaces generated by blending interpolation technique*, Studia Univ. Babeş-Bolyai Math., **36**(1991), no. 1, 119-130. MR 95a:65034.

[42] **Coman, Gh.,** *On some parallel methods in linear algebra*, Studia Univ. Babeş-Bolyai Math., **36**(1991), no. 3, 17-33. MR 94c:65174.

[43] **Coman, Gh.,** Ţâmbulea, L., *On the complexity of some scattered data interpolation procedures*, Seminar on Complexity of Algorithms, 50-63, Preprint, 89-10, Babeş-Bolyai Univ., Cluj-Napoca, 1991.

[44] **Coman, Gh.,** Chiorean, I., *On the efficiency of some parallel simultaneous methods for the approximation of polynomial zeros*, Seminar on Complexity of Algorithms, 39-49, Preprint, 89-10, Babeş-Bolyai Univ., Cluj-Napoca, 1991.

[45] **Coman, Gh.,** *Inverse interpolation methods for nonlinear equations*, Research Seminar on Numerical and Statistical Calculus, 45-52, Preprint, 94-1, Babeş-Bolyai Univ., Cluj-Napoca, 1994. MR 98g:65046.

[46] **Coman, Gh.,** Ţâmbulea, L., *On some interpolation procedure of scattered data*, Studia Univ. Babeş-Bolyai Math., **35**(1990), no. 2, 90-98. MR 94h:41069.

[47] Gânscă, I., **Coman, Gh.,** Ţâmbulea, L., *On the shape of Bézier surfaces*, Studia Univ. Babeş-Bolyai Math., **35**(1990), no. 3, 37-42.

[48] **Coman, Gh.,** Gânscă, I., Ţâmbulea, L., *Some practical application of blending interpolation*, Itinerant Seminar on Functional Equations, Approximation and Convexity (Cluj-Napoca, 1989), 5-22, Preprint, 89-6, Univ. Babeş-Bolyai , Cluj-Napoca, 1989.

[49] **Coman, Gh.,** Ţâmbulea, L., *A Shepard-Taylor approximation formula*, Studia Univ. Babeş-Bolyai Math., **33**(1988), no. 3, 65-73. MR 90i:41003.

[50] **Coman, Gh.,** *Shepard-Taylor interpolation*, Itinerant Seminar on Functional Equations, Approximation and Convexity (Cluj-Napoca, 1988), 5-14, Preprint, 88-6, Univ. Babeş-Bolyai , Cluj-Napoca, 1988.

[51] **Coman, Gh.,** *On the parallel complexity of some numerical algorithms for solving linear systems*, Itinerant Seminar on Functional Equations, Approximation and Convexity (Cluj-Napoca, 1987), 7-16, Preprint, 87-6, Univ. Babeş-Bolyai , Cluj-Napoca, 1987.

[52] **Coman, Gh.,** *The remainder of certain Shepard type interpolation formulas*, Studia Univ. Babeş-Bolyai Math., **32**(1987), no. 4, 24-32. MR 89j:41003.

[53] **Coman, Gh.,** *On the efficiency of parallel computation in numerical quadrature*, Seminar on Numerical and Statistical Calculus (Cluj-Napoca, 1987), 49-56, Preprint, 87-9, Univ. Babeş-Bolyai, Cluj-Napoca, 1987.

[54] **Coman, Gh.,** Rus, I. A., *On the sixtieth birthday of Professor D. D. Stancu*, Seminar on Numerical and Statistical Calculus (Cluj-Napoca, 1987), 1-20, Preprint, 87-9, Univ. Babeş-Bolyai, Cluj-Napoca, 1987. MR 89i:01070.

[55] **Coman, Gh.,** *Optimal algorithms with respect to the complexity*, Studia Univ. Babeş-Bolyai Math., **32**(1987), no. 1, 41-52. MR 90c:65083.

[56] **Coman, Gh.,** *Optimal quadratures with regard to the efficiency*, Calcolo, **24**(1987), no. 1, 85-100. MR 89k:65022.

[57] **Coman, Gh.,** *Optimal algorithms for the solution of nonlinear equation with regard to the efficiency*, Studia Univ. Babeş-Bolyai Math., **32**(1987), no. 3, 46-52. MR 90c:65084.

[58] Gânscă, I., **Coman, Gh.,** *On blending interpolation and cubature formulas*, (Romanian) Bul. Ştiinţ. Inst. Politehn. Cluj-Napoca Ser. Mat. Mec. Apl. Construc. Maş., **28**(1985), 35-40.

[59] **Coman, Gh.,** *Homogeneous multivariate approximation formulas with applications in numerical integration*, Seminar of numerical and statistical calculus (Cluj-Napoca, 1984-1985), 46-63, Preprint, 85-4, Univ. Babeş-Bolyai , Cluj-Napoca, 1985. MR 88b:65027.

[60] **Coman, Gh.,** *Some efficient methods for nonlinear equations*, Itinerant seminar on functional equations, approximation and convexity (Cluj-Napoca, 1985), 53-58, Preprint, 85-6, Univ. Babeş-Bolyai , Cluj-Napoca, 1985.

[61] **Coman, Gh.,** *On the complexity of some numerical algorithms*, Seminar on computer science, 1-33, Preprint, 84-4, Univ. Babeş-Bolyai , Cluj-Napoca, 1984.

[62] **Coman, Gh.,** *The optimal quadrature formulas from efficiency point of view*, Mathematica (Cluj) **26**(49)(1984), no. 2, 101-108. MR 86i:65014.

[63] **Coman, Gh.,** *Solution of boundary value problems by interpolating procedure*, Itinerant seminar on functional equations, approximation and convexity (Cluj-Napoca, 1984), 27-32, Preprint, 84-6, Univ. Babeş-Bolyai , Cluj-Napoca, 1984.

[64] **Coman, Gh.,** *Some practical approximation methods for nonlinear equations*, Anal. Numér. Théor. Approx., **11**(1982), no. 1-2, 41-48. MR 84m:65063.

[65] Gânscă, I., **Coman, Gh.,** *Blending approximation with applications to constructions*, (Romanian) Bul. Ştiinţ. Inst. Politehn. Cluj-Napoca Ser. Electrotehn.-Energet.-Inform., **24**(1981), 35-40.

[66] **Coman, Gh.,** *The complexity of the quadrature formulas*, Mathematica (Cluj), **23**(46)(1981), no. 2, 183-192. MR 84b:65022.

[67] Blaga, P., **Coman, Gh.,** *Multivariate interpolation formulas of Birkhoff type*, Studia Univ. Babeş-Bolyai Math., **26**(1981), no. 2, 14-22. MR 83j:41002.

[68] Böhmer, K., **Coman, Gh.,** *On some approximation schemes on triangle*, Mathematica (Cluj), **22**(45)(1980), no. 2, 231-235. MR 83g:41035.

[69] **Coman, Gh.,** *On some practical quadrature and cubature formulas*, Studia Univ. Babeş-Bolyai Math., **25**(1980), no. 3, 40-47. MR 83g:65029.

[70] **Coman, Gh.,** *On some Hermite-type interpolation formulas with applications to the numerical integration of functions*, (Romanian) Stud. Cerc. Mat., **32**(1980), no. 3, 291-307. MR 81k:41018.

[71] Blaga, P., **Coman, Gh.,** *On some bivariate spline operators*, Anal. Numér. Théor. Approx., **8**(1979), no. 2, 143-153. MR 81g:41014.

[72] **Coman, Gh.,** *Minimal monosplines in L_2 and optimal cubature formulae*, Anal. Numér. Théor. Approx., **7**(1978), no. 2, 147-155. MR 80h:41003.

[73] Böhmer, K., **Coman, Gh.,** *Blending interpolation schemes on triangles with error bounds*, Constructive theory of functions of several variables (Proc. Conf., Math. Res. Inst., Oberwolfach, 1976), pp. 14-37. Lecture Notes in Math., Vol. 571, Springer, Berlin, 1977. MR 57#14331.

[74] Böhmer, K., **Coman, Gh.,** *Smooth interpolation schemes in triangles with error bounds*, Mathematica (Cluj), **18**(41)(1976), no. 1, 15-27. MR 57#6991.

[75] **Coman, Gh.,** *Multivariate approximation schemes and the approximation of linear functionals*, Mathematica (Cluj), **16**(39)(1974), no. 2, 229-249. MR 58#29711.

[76] **Coman, Gh.,** Gânscă, I., *On a certain two-dimensional monospline with minimal deviation in L_1 and a certain optimal cubature formula*, (Romanian), Stud. Cerc. Mat., **26**(1974), 367-374. MR 52#3837.

[77] **Coman, Gh.,** Frenţiu, M., *Bivariate spline approximation*, Studia Univ. Babeş-Bolyai Ser. Math.-Mech., **19**(1974), no. 1, 59-64. MR 49#7655.

[78] **Coman, Gh.,** *Generalized monosplines and optimal quadrature formulas*, (Romanian) Stud. Cerc. Mat., **25**(1973), 495-503. MR 57#6989.

[79] **Coman, Gh.,** *Two-dimensional monosplines and optimal cubature formulae*, Studia Univ. Babeş-Bolyai Ser. Math.-Mech., **18**(1973), no. 1, 41-53. MR 49#6562.

[80] **Coman, Gh.,** *Applications of spline functions to the construction of optimal quadrature formulas*, (Romanian) Stud. Cerc. Mat., **24**(1972), 329-334. MR 52#7089.

[81] **Coman, Gh.,** *Quadrature formulas of Sard type*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Mech., **17**(1972), no. 2, 73-77. MR 51#14532.

[82] **Coman, Gh.,** *Monosplines and optimal quadrature formulae in L_p*, Rend. Mat., **5**(6)(1972), 567-577. MR 51#9438.

[83] **Coman, Gh.,** *Optimal cubature formulas for certain classes of functions*, (Russian) Rev. Roumaine Math. Pures Appl., **17**(1972), 1025-1036. MR 48#3226.

[84] **Coman, Gh.,** *Monosplines and optimal quadrature formulae*, Collection of articles dedicated to G. Călugăreanu on his seventieth birthday, Rev. Roumaine Math. Pures Appl., **17**(1972), 1323-1327. MR 48#764.

[85] **Coman, Gh.,** *On certain practical quadrature formulae*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Mech. **17**(1972), no. 1, 61-65. MR 46#10174.

[86] **Coman, Gh.,** Micula, Gh., *Optimal cubature formulae*, Rend. Mat. **4**(6)(1971), 303-311. MR 46#10175.

[87] **Coman, Gh.,** *A practical quadrature formula, optimal for a class of functions*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Mech., **16**(1971), fasc. 1, 73-79. MR 44#4898.

[88] **Coman, Gh.,** *Nouvelles formules de quadrature à coefficients égaux*, (French) Mathematica (Cluj) **12**(35)(1970), 253-264. MR 48#12781.

[89] **Coman, Gh.,** *On best cubature formulas for certain classes of functions*, (Romanian) Stud. Cerc. Mat., **22**(1970), 551-561. MR 48#11868.

[90] **Coman, Gh.,** *On certain optimal quadrature formulas*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Mech., **15**(1970), fasc. 2, 39-54. MR 43#6640.

[91] **Coman, Gh.,** Micula, Gh., *Optimal cubature formulas for certain classes of functions*, An. Şti. Univ. Al. I. Cuza Iaşi Secţ. I a Mat. (N.S.) **16**(1970), 345-356. MR 43#5226.

[92] **Coman, Gh.,** *A generalization of the trapezon quadrature formula and of Simpson's formula*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Phys., **14**(1969), no. 2, 53-58. MR 41#6391.

[93] **Coman, Gh.,** *On some cubature formulas with fixed nodes*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Phys. **13**(1968), no. 2, 51-54. MR 39#3699.

[94] Orbán, B., Groze, V., **Coman, Gh.,** *On the projective transformation of a nomogram with rectilinear scales*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Phys., **12**(1967), no. 1, 15-23. MR 35#6391.

[95] Groze, V., **Coman, Gh.,** *An optimal nomogram in a class of nomograms with aligned points of order 3*, (Romanian) Studia Univ. Babeş-Bolyai Ser. Math.-Phys., **10**(1965), no. 1, 13-22. MR 32#3300.

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

HOMOGENIZATION WITH MULTIPLE SCALE EXPANSION ON SELFSIMILAR STRUCTURES

J. KOLUMBÁN AND A. SOÓS

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. The homogenization theory is devoted to analysis of partial differential equations with rapidly oscillating coefficients. Let $\mathcal{A}^k$ be a given partial differential operator and we consider the equation $\mathcal{A}^k u^k = f$, together with the appropriate boundary initial conditions. Here $k \in \mathbb{N}$ and $f \in H^1(\mathbb{R}^n)$. We are interested in studying the solutions of this system in the limit as $k \to \infty$.

The homogenization theory is devoted to analysis of partial differential equations with rapidly oscillating coefficients. Let $\mathcal{A}^\epsilon$ be a given partial differential operator and we consider the equation

$$\mathcal{A}^\epsilon u^\epsilon = f,$$

together with the appropriate boundary initial conditions. Here ϵ is a small parameter $\epsilon << 1$, associated with the oscillations. We are interested in studying the solutions of this system in the limit as $\epsilon \to 0$. The homogenization theory study the following issues:

- Convergence to a limit.
- Characterization of the limiting process.

$$\mathcal{A}u = f$$

- Explicit analytical construction of $\mathcal{A}$.

Received by the editors: 01.07.2006.

2000 *Mathematics Subject Classification.* 28A80, 35B27.

Key words and phrases. homogenization, invariant measures, PDS, two-scale convergence.

- Properties of the limiting equation.

This type of equation models various physical problems. As examples we mention composite materials, flow in porous media, atmospheric turbulence. A common feature of all these problems is that phenomena occur at various length and times scales. In the classical homogenization theory the structure are periodic. This implies that the coefficients of the corresponding PDE which model the physical phenomenon under investigation are periodic.

In this article we will quit the periodicity assumption. The coefficients are generated by an iterated function system. We will give a generalization of classical homogenization theory. We will use some basic notions from the fractal geometry as the invariant set and the Hutchinson's invariant measure regarding iterating function systems. Usually the solutions of the limiting equations live on selfsimilar fractals.

1. **Setting of the problem**

Consider the similarities $\varphi_1, ..., \varphi_m : \mathbb{R}^n \to \mathbb{R}^n$ with the scale factors $r_1, ..., r_n \in]0, 1[$, respectively. Suppose there exists an open, bounded set $O \subset \mathbb{R}^n$ such that $\varphi_i(O) \subset O$ and $\varphi_i(O) \cap \varphi_j(O) = \emptyset \, (i \neq j)$. For $i_1, ..., i_k \in \{1, ..., m\}$ denote σ the word $i_1...i_k$, and let $|\sigma| = k$ be the lenght of σ. Let $\varphi_\sigma = \varphi_{i_1} \circ ... \circ \varphi_{i_k}$ and $r_\sigma = r_{i_1}...r_{i_k}$.

For $A \subseteq \mathbb{R}^n$ put

$$F(A) := \varphi_1(A) \cup ... \cup \varphi_m(A),$$

$$F^1 := F, \, F^k := F \circ F^{k-1} \, (k \in \mathbb{N}), \text{and } K := \cap_{k \geq 1} F^k(\overline{O}).$$

Then we have

$$F(K) = K \subseteq \overline{O}$$

and

$$\lim_{k \to \infty} F^k(A) = K,$$

for every compact subset $A \subset \mathbb{R}^n$, where the limit is understood in sense of the Hausdorff metric. K is the unique nonempty compact set which is invariant under F.

Moreover, the Hausdorff dimension of K is the solution of

$$\sum_{i=1}^{m} r_i^s = 1.$$

If $\mathcal{H}^s(K)$ denotes the s-dimensional Hausdorff measure of K, then $0 < \mathcal{H}^s(K) < +\infty$.

Define $\mathcal{M}$ to be the set of positive, Borel regular measures μ on $\mathbb{R}^n$ having bounded support and finite mass. Put "spt" for support.

Define

$$\mathcal{M}^1 := \{\mu \in \mathcal{M} : \mu(\mathbb{R}^n) = 1\}.$$

Let

$$C(\mathbb{R}^n) := \{f : \mathbb{R}^n \to \mathbb{R} : \ f \text{ is continuous}\}.$$

For $\mu \in \mathcal{M}$, $\psi \in C(\mathbb{R}^n)$, define

$$\mu(\psi) := \int_{\mathbb{R}^n} \psi d\mu.$$

If $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, then we define the push forward measure $\varphi_\# : \mathcal{M} \to \mathcal{M}$ by

$$\varphi_\#\mu(A) := \mu(\varphi^{-1}(A)), \ A \subseteq \mathbb{R}^n,$$

equivalently

$$\varphi_\#\mu(\psi) := \mu(\psi \circ \varphi), \ \psi \in \mathbb{C}(\mathbb{R}^n).$$

We define the weak topology on $\mathcal{M}$ by taking as a sub-basis all sets of the form $\{\mu : \ a < \mu(\varphi) < b\}$, for arbitrary real $a < b$ and arbitrary $\varphi \in C(\mathbb{R}^n)$. We have $\mu_i \to \mu$ in the weak topology iff $\mu_i(\varphi) \to \mu(\varphi)$ for all $\varphi \in C(\mathbb{R}^n)$.

For $\mu, \nu \in \mathcal{M}^1$ let

$$L(\mu, \nu) := \sup\{\mu(\varphi) - \nu(\varphi) : \varphi \in C(\mathbb{R}^n), \ Lip\varphi \leq 1\}.$$

Then L is a metric on $\mathcal{M}^1$ and the metric topology coincide with the weak topology on $\mathcal{M}^1$. Moreover, the metric space $(\mathcal{M}^1, L)$ is complete (see [6]).

If $\nu \in \mathcal{M}^1$, let

$$G(\nu) := \sum_{i=1}^{m} r_i^s \varphi_{i\#}\nu.$$

Thus

$$G(\nu)(\psi) = \sum_{i=1}^{m} r_i^s \nu(\psi \circ \varphi_i),\ \psi \in C(\mathbb{R}^n),$$

and $G : \mathcal{M}^1 \to \mathcal{M}^1$ is a contraction map. Consequently, there exists a unique measure $\overline{\mu} \in \mathcal{M}^1$, the *invariant measure on* K, such that

$$G(\overline{\mu}) = \overline{\mu} \quad \text{and} \quad spt\overline{\mu} = K.$$

For $\nu \in \mathcal{M}^1$ put

$$G^1(\nu) := G(\nu),\ G^k(\nu) = G \circ G^{k-1}(\nu),\ k \in \mathbb{N}.$$

It is easy to see that

$$G^k(\nu) = \sum_{|\sigma|=k} r_\sigma^s \varphi_{\sigma\#}(\nu),$$

and

$$\overline{\mu} = \lim_{k\to\infty} G^k(\nu)$$

in the sense of L metric.

Moreover,

$$\overline{\mu} = (\mathcal{H}^s(K))^{-1}\mathcal{H}^s|_K.$$

For all these properties we refer to Hutchinson [6].

Let $\Omega \subset \mathbb{R}^n$ be a nonempty bounded open domain with smooth boundary. Let ν be the Lebesgue measure on Ω divided by the Lebesque measure of Ω.

Let us suppose that the functions φ_i are composed by homotheties and translations, and $\Omega \subseteq O$. For the word σ consider the set

$$\Omega_\sigma := \varphi_\sigma(\Omega).$$

In this paper we will develop the method of homogenization through multiple scales for the Dirichlet problem

$$\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij}\left(\varphi_\sigma^{-1}(x)\right)\frac{\partial u^k(x)}{\partial x_j}\right) = f(x), \qquad \text{for } x \in \Omega_\sigma,\ \text{for all } \sigma \text{ with } |\sigma| = k, \tag{1.1}$$

$$u^k(x) = 0, \qquad \text{for } x \in \partial\Omega\sigma. \tag{1.2}$$

We are interested in studying the solutions u^k of (1.1) in the limit as $k \to \infty$. In particular, we would like to understand if the limit exists and what kind of properties does the limit u satisfy.

We will assume that the coefficients $a_{ij} : \mathbb{R}^n \to \mathbb{R}$ are smooth and uniformly elliptic on Ω, i.e. there exists $\alpha > 0$ such that

$$\sum_{i,j=1}^{n} a_{ij}(y)\xi_i\xi_j \geq \alpha|\xi|^2, \, \forall y \in \mathbb{R}^n, \, \xi \in \mathbb{R}^n. \tag{1.3}$$

It is to accentuate, that the coefficients a_{ij} have not to be periodic. We will also assume that the function f is smooth and independent of k.

Example 1.1. *(Homogenization for periodic structures) Let $Y = [0,1]^n$ be the unit cell and suppose that the coefficients a_{ij} are 1-periodic, i.e.*

$$a_{ij}(y + e_k) = a_{ij}(y), \;\; i, j, k = 1, ..., n, \; y \in Y.$$

Define $\varphi_1(x) = \frac{x}{2}$. For $m = 2^n$ and $i \in \{2.,,,.m\}$ we define the functions $\varphi_i(x)$ as translations of φ_1 by sumes of unit vectors such that $Y = \cup_{i=1}^{n}\varphi_i(Y)$. We can choose $O = [0,1]^n$. In this case, for $|\sigma| = k$ and $\epsilon = \frac{1}{2^k}$, we have $a_{ij}(\varphi_\sigma^{-1}(x)) = a_{ij}(\frac{x}{\epsilon})$. Therefore, in this case, problem (1.1) redeces to the homogenization problem for periodic structures (see [4]). Insteed of $]0,1[^n$ we can choose O as an open cub containing Ω. In this case the functions φ_i will be translated by a corresponding constant vector.

Example 1.2. *(Homogenization on Sierpinski gasket) Let $n = 2$, $m = 3$, and q_1, q_2, q_3 be the vertices of an equilateral triangle. The functions φ_i are defined by*

$$\varphi_i(x) = \frac{1}{2}(x - q_i) + q_i, \; i = 1, 2, 3.$$

The Sierpinski gasket is the unique nonempty compact set $K \subset \mathbb{R}^2$ such that

$$K = \varphi_1(K) \cup \varphi_2(K) \cup \varphi_3(K).$$

The set K is one of the simplest examples of a self-similar fractal. Its Hausdorff dimension is $\frac{\log 3}{\log 2}$. In this case $r_1 = r_2 = r_3 = \frac{1}{2}$ and O can be choosen as the interior

of the triangle $q_1q_2q_3$. If we take $\Omega = O$, then the problem of homogenization reduces to the study of the limit of u^k on the Sierpinsku gasket.

2. **The multiple scales expansion**

By the iterated functions system properties, for $\sigma \neq \sigma'$, $|\sigma| = |\sigma'|$ we have:

$$\Omega_\sigma \cap \Omega_{\sigma'} = \emptyset.$$

The idea behind the method of multiple scales is to assume that the solution u^k is of the form

$$u^k(x) = u_0(x, \varphi_\sigma^{-1}(x)) + r_\sigma u_1(x, \varphi_\sigma^{-1}(x)) + r_\sigma^2 u_2(x, \varphi_\sigma^{-1}(x)) + ..., \text{ for } x \in \Omega_\sigma \quad (2.4)$$

Using the two-scale convergence method, the validity of this expansion will be justified later. Anyway, since $\lim_{|\sigma|\to\infty} r_\sigma = 0$, from physical point of view it is reasonable to expect that the solution of (1.1) is of the (2.4) form, since there are different lenght scales in our problem and the above expansion takes this fact explicitly into account. The variables x and $y = \varphi_\sigma^{-1}(x)$ represent the "slow" (macroscopic) and "fast" (microscopic) scales of the problem respectively. For big $|\sigma|$ the variable y changes much more rapidly then x. We can think of x as being a constant, when looking at the problem at the microscopic scale. So we can treat x and y as independent variables.

The fact that $y = \varphi_\sigma^{-1}(x)$ implies that the partial derivatives with respect to x_j become:

$$\frac{\partial}{\partial x_j} \to \frac{\partial}{\partial x_j} + \frac{1}{r_\sigma}\frac{\partial}{\partial y_j},\ j = 1, ..., n.$$

Using this we can write the differential operator

$$\mathcal{A}^\sigma = -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij} \circ \varphi_\sigma^{-1}\frac{\partial}{\partial x_j}\right) \quad (2.5)$$

in the form

$$\mathcal{A}^\sigma = r_\sigma^{-2}\mathcal{A}_0 + r_\sigma^{-1}\mathcal{A}_1 + \mathcal{A}_2, \quad (2.6)$$

where

$$\mathcal{A}_0 : = -\sum_{i,j=1}^{n} \frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right), \tag{2.7}$$

$$\mathcal{A}_1 : = -\sum_{i,j=1}^{n} \left[\frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial x_j}\right) + \frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right)\right], \tag{2.8}$$

$$\mathcal{A}_2 : = -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial}{\partial x_j}\right). \tag{2.9}$$

Now, equation (1.1), on account of (2.6), becomes

$$\left(r_\sigma^{-2}\mathcal{A}_0 + r_\sigma^{-1}\mathcal{A}_1 + \mathcal{A}_2\right) u^k(x) = f(x), \text{ for } x \in \Omega\sigma \text{ and } |\sigma| = k. \tag{2.10}$$

If we substitute (2.4) into (2.10) we obtain the following sequence of problems:

$$\mathcal{A}_0 u_0 = 0 \tag{2.11}$$

$$\mathcal{A}_0 u_1 = -\mathcal{A}_1 u_0 \tag{2.12}$$

$$\mathcal{A}_0 u_2 = -\mathcal{A}_1 u_1 - \mathcal{A}_2 u_0. \tag{2.13}$$

These equations are of the form:

$$\mathcal{A}_0 u = h \tag{2.14}$$

Here $u = u(x, y)$ and $h = h(x, y)$. However x enters merely as a parameter since $\mathcal{A}_0$ is a uniformly elliptic partial differential operator with respect to y. This equation admits a unique, smooth solution if and only if the right hand side averages to 0 on Ω:

$$\int_\Omega h(x, y)d\nu(y) = 0, \tag{2.15}$$

This solvability condition is a consequence of the Fredholm alternative (see [5]).

By (1.3) the only solutions of the homogenous equation $\mathcal{A}_0 u_0 = 0$ are constants in y: $u_0(x, y) = u(x)$. This means that the first term in the multiple scales expansion is independent of the fast scales represented by y. Consequently, we can derive a homogenized equation for $u(x)$ which is independent of the microscopic scales.

In view of relation $u_0(x,y) = u(x)$ the equation (2.12) becomes:

$$\mathcal{A}_0 u_1 = -\sum_{i,j=1}^{n} \frac{\partial a_{ij}}{\partial y_i}\frac{\partial u}{\partial x_j}. \tag{2.16}$$

Suppose

$$\sum_{i=1}^{n}\int_{\Omega}\frac{\partial a_{ij}(y)}{\partial y_i}d\nu(y) = 0, \text{ for all } j = 1, ..., n. \tag{2.17}$$

In this case, the solvability condition is satisfied, and equation (2.16) is well posed: it admits a unique solution, up to constants in y. To solve this equation we will use the separation of variable tehnique. To this end, we look for a solution of the form:

$$u_1(x,y) = -\sum_{j=1}^{n}\chi^j(y)\frac{\partial u(x)}{\partial x_j} + \overline{u}_1(x). \tag{2.18}$$

Substituting (2.18) into (2.16) we obtain:

$$\mathcal{A}_0\chi^j(y) = -\sum_{i=1}^{n}\frac{\partial a_{ij}(y)}{\partial y_i}, \text{ for } j = 1, ..., n. \tag{2.19}$$

This is called the *cell problem* and $\chi^j(y)$ is the *first order corrector field*. Condition (2.17) implies that the problem is well posed. We remark that at this moment $\overline{u}_1$ is undetermined.

Now we consider equation (2.13). The function f being independent of y, then the solvability condition implies:

$$\int_{\Omega}\left[\mathcal{A}_1 u_1(x,y) + \mathcal{A}_2 u(x)\right]d\nu(y) = f(x). \tag{2.20}$$

We have:

$$\begin{aligned}
&\int_{\Omega}\left[\mathcal{A}_1 u_1(x,y) + \mathcal{A}_2 u(x)\right]d\nu(y)\\
&= \sum_{i,j=1}^{n}\int_{\Omega}\left[-\frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial u_1}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial u_1}{\partial y_j}\right)\right]d\nu(y)\\
&\qquad + \sum_{i,j=1}^{n}\int_{\Omega} -\frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial u(x)}{\partial x_j}\right)d\nu(y)\\
&= \sum_{i,j,l=1}^{n}\int_{\Omega}\frac{\partial}{\partial y_i}\left[a_{ij}(y)\chi^l(y)\frac{\partial^2 u(x)}{\partial x_l \partial x_j}\right]d\nu(y)
\end{aligned}$$

$$+\sum_{i,j,l=1}^{n}\frac{\partial^2 u(x)}{\partial x_i\partial x_l}\int_\Omega a_{ij}(y)\frac{\partial\chi^l(y)}{\partial y_j}d\nu(y)-\sum_{i,j=1}^{n}\frac{\partial^2 u(x)}{\partial x_i\partial x_j}\int_\Omega a_{ij}(y)d\nu(y)$$

$$=\sum_{i,l=1}^{n}\frac{\partial^2 u(x)}{\partial x_i\partial x_l}\int_\Omega\left\{\sum_{j=1}^{n}\left[a_{ij}(y)\frac{\partial\chi^l(y)}{\partial y_j}+\chi^l(y)\frac{\partial a_{ji}(y)}{\partial y_j}+a_{ji}(y)\frac{\partial\chi^l(y)}{\partial y_j}\right]-a_{il}(y)\right\}d\nu(y)$$

Denote

$$\overline{a}_{il}:=-\int_\Omega\left\{\sum_{j=1}^{n}\left[(a_{ij}(z)+a_{ji}(z))\frac{\partial\chi^l(z)}{\partial z_j}+\chi^l(z)\frac{\partial a_{ji}(z)}{\partial y_j}\right]-a_{il}(z)\right\}d\nu(z).$$

By this notation the homogenized equation will be the following:

$$-\sum_{i,l=1}^{n}\overline{a}_{il}\frac{\partial^2 u(x)}{\partial x_i\partial x_l} = f(x),\ \text{for } x\in\Omega_\sigma \tag{2.21}$$

$$u(x) = 0,\ \text{for } x\in\partial\Omega_\sigma, \tag{2.22}$$

and for all σ.

3. **Two scale convergence**

In this section we will recall the homogenization procedure given by [7]. The notion of two-scale convergence introduced by Nguetseng [8],[9] and developed futher by Allaire [1],[2] was modified by Kolumbán for iterating function system inthe following way:

Denote $C_b(\mathbb{R}^n)$ the set of bounded and continuous real functions defined on $\mathbb{R}^n$. We will also use the space $L^2_\sigma(\Omega,C_b(\mathbb{R}^n))$, which is the set of all measurable functions $u:\Omega\to C_b(\mathbb{R}^n)$ such that $\parallel u\parallel\in L^2(\Omega)$. The norm of this space is

$$\parallel u\parallel_{L^2(\Omega,C_b(\mathbb{R}^n))}=\left[\int_\Omega|\sup_{y\in\mathbb{R}^n}u(x,y)|^2dx\right]^{\frac{1}{2}}.$$

Theorem 3.1. *(Oscillations lemma, [7]) Let ν be the Lebesgue measure restricted to Ω and let $\Phi\in L^2(O,C_b(\mathbb{R}^n))$. Then the following convergence result holds:*

$$\lim_{k\to\infty}\sum_{|\sigma|=k}r_\sigma^s\int_\Omega\Phi(\varphi_\sigma(z),z)d\nu(z)=\int_K\left[\int_\Omega\Phi(x,y)d\nu(y)\right]d\overline{\mu}(x), \tag{3.23}$$

where K is the invariant set and $\overline{\mu}$ is the invariant measure of the iterated function system $\{\varphi_1,...,\varphi_m\}$.

Let u^k be a sequence in $L^2(O)$. We say that u^k two-scale converges weakly to $u_0 \in L^2(K \times \Omega)$ and write $u^k \overset{2}{\rightharpoonup} u_0$ if for every function $\Phi \in L^2(O, C_b(\mathbb{R}^n))$ we have

$$\lim_{k\to\infty} \sum_{|\sigma|=k} r_\sigma^s \int_\Omega u^k(\varphi_\sigma(x))\Phi(\varphi_\sigma(x), x)d\nu = \int_K \int_\Omega u_0(x,y)\Phi(x,y)d\nu(y)d\overline{\mu}(x). \quad (3.24)$$

Two-scale convergence implies a kind of weak convergence in $L^2(\Omega)$. In fact we have the following lemma:

Lemma 3.1. *([7]) Let u^k be in $L^2(\cup_{|\sigma|=k}\Omega_\sigma)$ which two-scale converges weakly to u_0. Then, for all $\Phi \in L^2(\Omega)$,*

$$\lim_{k\to\infty} \sum_{|\sigma|=k} r_\sigma^s \int_\Omega u^k(\varphi_\sigma(x))\Phi(\varphi_\sigma(x)d\nu(x) = \int_K \left(\int_\Omega u_0(x,y)d\nu(y)\right)\Phi(x)d\overline{\mu}(x). \quad (3.25)$$

The first result over the two scale expansion is the following lema:

Lemma 3.2. *Let $u^k \in L^2(\Omega)$ be a function which admits the two-scale expansion*

$$u^k(x) = u_0(x, \varphi_\sigma^{-1}(x)) + r_\sigma u_1(x, \varphi_\sigma^{-1}(x)) + ..., \text{ for } x \in \Omega_\sigma,$$

where $u_j \in L^2(\Omega, C_b(\mathbb{R}^n))$, $j \in \{0,1\}$. Then $u^k \overset{2}{\rightharpoonup} u_0$.

Proof. For $\Phi \in L^2(\Omega, C_b(\mathbb{R}^n))$ we have:

$$\begin{aligned}
&\sum_{|\sigma|=k} r_\sigma^s \int_{\Omega_\sigma} u^k(x)\Phi(x, \varphi_\sigma^{-1}(x))d\nu_\sigma(x) = \\
= &\sum_{|\sigma|=k} r_\sigma^s \int_\Omega u^k(\varphi_\sigma(x))\Phi(\varphi_\sigma(x), x)d\nu(x) = \\
= &\sum_{|\sigma|=k} r_\sigma^s \int_\Omega u_0(\varphi_\sigma(x), x)\Phi(\varphi_\sigma(x), x)d\nu(x) + \\
+ &\sum_{|\sigma|=k} r_\sigma^{s+1} \int_\Omega u_1(\varphi_\sigma(x), x)\Phi(\varphi_\sigma(x), x)d\nu(x) + ...
\end{aligned}$$

The first term converges to $\int_K \int_\Omega u_0(x,y)d\nu(y)d\overline{\mu}(x)$.

Using Cauchy-Schwartz inequality, we obtain:

$$\begin{aligned} & |\sum_{|\sigma|=k} r_\sigma^{s+1} \int_\Omega u_1(\varphi_\sigma(x),x)\Phi(\varphi_\sigma(x),x))d\nu(x)| \leq \\ \leq & \sum_{|\sigma|=k} r_\sigma^{s+1} \parallel u_1(\varphi(\cdot),\cdot) \parallel\parallel \Phi(\varphi(\cdot),\cdot) \parallel_{L^2(\Omega\times\Omega)} \leq \\ \leq & \parallel u_1 \parallel_{L^2(\Omega\times\Omega)} \parallel \Phi \parallel_{L^2(\Omega\times\Omega)} \sum_{|\sigma|=k} r_\sigma^{s+1} \to 0 \end{aligned}$$

So

$$\int_\Omega u^k(x)\Phi(x,\varphi_\sigma^{-1}(x))dx \to \int_K \int_\Omega u_0(x,y)\Phi(x,y)d\nu(y)d\overline{\mu}(x).$$

Hence u^k two-scale converges weakly to u_0. □

We will use the following compactness result as a criteria which enable to conclude that a given sequence is weakly two-scale convergent.

Theorem 3.2. *([7]) Let $u^k \in L^2(\cup_{|\sigma|=k}\Omega_\sigma)$, $k \in \mathbb{N}$, and let*

$$a_k := \left[\sum_{|\sigma|=k} r_\sigma^s \int_\Omega [u^k \circ \varphi_\sigma(x)]^2 d\nu(x)\right]^{\frac{1}{2}}.$$

If the sequence (a_k) is bounded then there exists a subsequence of (u^k) which two-scale converges weakly to a function $u_0 \in L^2_{\overline{\mu}\otimes\nu}(K\times\Omega)$.

The weakly two-scale convergence defined is still a weak type of convergence, since it is defined in terms of the product of a sequence u^k with an appropriate test function. We also define a notion of strong two-scale convergence.

Let u^k be a sequence in $L^2(\Omega)$. We say that u^k two-scale converges strongly to $u_0 \in L^2_{O\times\Omega}$ and write $u^k \xrightarrow{2} u_0$ if

$$\lim_{k\to\infty} \sum_{|\sigma|=k} r_\sigma^s \int_\Omega |u^k(\varphi_\sigma(x))|^2 d\nu(x) = \int_K \int_\Omega |u_0(x,y)|^2 d\nu(y)d\overline{\mu}(x) \tag{3.26}$$

Although every strongly two-scale convergent sequence is also weakly two-scale convergent, the converse is not true.

As it is always the case with weak convergence, the limit of the product of two-scale convergent sequences is not in general the product of the limits. However, we can pass to the limit when we one of the two sequence is strongly two-scale convergent.

The next theorem can be proved as the similar result in the classical homogenization theory (see [2]).

Theorem 3.3. *Suppose* $u^k \xrightarrow{2} u_0$ *and* $v^k \xrightarrow{2} v_0$. *Then* $u^k v^k \xrightarrow{2} u_0 v_0$.

For simplicity in the following we suppose $n = 1$ and $O = \Omega =]a, b[$, $[a, b] \subseteq \mathbb{R}$. Let μ be an atomless finite Borel measure on $[a, b]$. Further let $K := spt\mu$ with $a, b \in K$ and $L_2 := L^2_\mu(K)$ be the separable Hilbert space with scalar product $\langle f, g\rangle = \int_a^b f g d\mu$. We define

$$\mathcal{D}_1^\mu := \{f \in L_2 : \exists f' \in L_2 \text{ with } f(x) = f(a) + \int_a^x f'(y)d\mu(y),\ x \in spt\mu\}.$$

Proposition 3.1. $\mathcal{D}_1^\mu \subset C(K)$, *i.e. every function in* $\mathcal{D}_1^\mu$ *is continuous on* K, *and the function* f' *defined above is unique in* L_2.

So we can introduce the $\mu-$ derivative of f. The μ-derivative of f on $\mathcal{D}_1^\mu$ is

$$\nabla^\mu f = \frac{df}{d\mu}.$$

In the case $\mu = \nu$, where ν denotes Lebesgue measure on $\mathbb{R}$, $\mathcal{D}_1^\mu$ coincides with the Sobolev space W_2^1. As in the classical Lebesgue case the $\mu-$Dirichlet form on $\mathcal{D}_1^\mu$ is defined as

$$\mathcal{E}^\mu(f, g) := \langle \nabla^\mu f, \nabla^\mu g\rangle$$

Denote

$$\mathcal{D}_2^\mu := \{f \in \mathcal{D}_1^\mu :\ \nabla^\mu f \in \mathcal{D}_1^\mu\}.$$

The $\mu-$Laplace operator from $\mathcal{D}_2^\mu$ is given by

$$\Delta^\mu f := \nabla^\mu(\nabla^\mu f) = f'.$$

Remark 3.1.

$$\mathcal{D}_2^\mu = \left\{ f \in L_2 :\ \exists f', f'' \in L_2 \text{ with } f(x) = f(a) + \int_a^x f'(y)d\mu(y),\ x \in K,\right.$$
$$\left. f'(y) = f'(a) + \int_a^y f''(z)d\mu(z),\ y \in K \right\}.$$

Using Fubini's theorem we have the following representation of $f \in \mathcal{D}_2^\mu$*:*

$$f(x) = f(a) + \nabla^\mu f(a)\mu([a, x]) + \int_a^x \mu([y, x])\Delta^\mu f(y)d\mu(y),\ x \in K.$$

Proposition 3.2. *For any $c, d \in K$ with $c \leq d$ and $f, g \in \mathcal{D}_1^\mu$ we have*

$$\int_c^d (\nabla^\mu f) g d\mu = fg|_c^d - \int_c^d f(\nabla^\mu g) d\mu$$

Proof. By definition of ∇^μ and Fubini theorem it follows that

$$\begin{aligned}
& \int_c^d (\nabla^\mu f)(x) g(x) d\mu(x) \\
= & \int_c^d (\nabla^\mu f)(x) \left[g(c) + \int_c^x (\nabla^\mu g)(y) d\mu(y) \right] d\mu(x) \\
= & \; g(c) \left[f(d) - f(c) \right] + \int_c^d (\nabla^\mu g)(y) \left[\int_y^d (\nabla^\mu f)(x) d\mu(x) \right] d\mu(y) \\
= & \; g(c) \left[f(d) - f(c) \right] + \int_c^d (\nabla^\mu g)(y) \left[f(d) - f(y) \right] d\mu(y) \\
= & \; g(c) \left[f(d) - f(c) \right] + f(d)[g(d) - g(c)] - \int_c^d f(y)(\nabla^\mu g)(y) d\mu(y) \\
= & \; f(d)g(d) - f(c)g(c) - \int_c^d f(y)(\nabla^\mu g)(y) d\mu(y). \qquad (3.27)
\end{aligned}$$

□

In the similar way we can prove the following proposition:

Proposition 3.3. *For any $c, d \in K$ with $c \leq d$ and $f, g \in \mathcal{D}_2^\mu$ we have*

$$\int_c^d (\Delta^\mu f) g d\mu = (\nabla^\mu f) g|_c^d - \int_c^d (\nabla^\mu f)(\nabla^\mu g) d\mu$$

$$\int_c^d \left[(\Delta^\mu f) g - f(\Delta^\mu g) \right] d\mu = (\nabla^\mu f) g - f(\nabla^\mu g)|_c^d$$

These are analogues of the classical Gauss Green formulae.

Now we introduce the Dirichlet boundary condition

$$\mathcal{D}_{2,D}^\mu := \{ f \in \mathcal{D}_2^\mu : \; f(a) = f(b) = 0 \}.$$

From the last proposition we obtain the following

Corollary 3.1. *$-\Delta^\mu$ is a positive symmetric operator on $\mathcal{D}_{2,D}^\mu$.*

Theorem 3.4. *Let $u^k \in H_0^1(\cup_{|\sigma|=k}\Omega_\sigma)$. If*

$$a_k := \left[\sum_{|\sigma|=k} r_\sigma^s \int_\Omega [u^k \circ \varphi_\sigma(x)]^2 d\nu(x)\right]^{\frac{1}{2}}$$

and

$$b_k := \left[\sum_{|\sigma|=k} r_\sigma^s \int_\Omega [\nabla u^k \circ \varphi_\sigma(x)]^2 d\nu(x)\right]^{\frac{1}{2}}$$

are bounded then there exist a subsequence of (u^k) and functions $u_0(\cdot), u(\cdot) \in L^2_{\overline{\mu}}(K)$ with $\nabla^{\overline{\mu}} u \in L^2_{\overline{\mu}}(K)$, such that

$$\begin{aligned}\lim_{k\to\infty} \sum_{|\sigma|=k} r_\sigma^s \int_{\Omega_\sigma} (\nabla u^k)(x)\varphi(x) d\nu_\sigma(x) &= \int_K \nabla^{\overline{\mu}} u(x)\Phi(x) d\overline{\mu}(x) = \\ &= -\int_K u_0(x)\nabla\Phi(x) d\overline{\mu}(x),\ \forall \Phi \in H_0^1(O).\end{aligned}$$

Proof. By Theorem 3.2 we can choose a subsequence denoted by u^k too, a function $u_0 \in L^2(K\times\Omega)$ and a function $v \in L^2(K\times\Omega)$ such that $u^k \overset{2}{\rightharpoonup} u_0$ and $\nabla u^k \overset{2}{\rightharpoonup} v$.

First we prove the independence of u_0 of the second variable y. To this end let $\Phi \in C^1(O\times O)$. We have

$$\begin{aligned}&\sum_{|\sigma|=k} r_\sigma^s \int_\Omega (\nabla u^k)(\varphi_\sigma(z))\Phi(\varphi_\sigma(z), z) d\nu(z) = \\ = &-\sum_{|\sigma|=k} r_\sigma^s \int_\Omega u^k(\varphi_\sigma(z))[r_\sigma \nabla_x \Phi(\varphi_\sigma(z), z) + \nabla_y \Phi(\varphi_\sigma(z), z))] d\nu(z)\end{aligned}$$

Since $u^k(\varphi_\sigma(z))\nabla_x\Phi(\varphi_\sigma(z), z)$ is bounded in $L^2(\Omega)$ it follows that

$$\sum_{|\sigma|=k} r_\sigma^{s+1} \int_\Omega u^k(\varphi_\sigma(z))\nabla_x\Phi(\varphi_\sigma(z), z) d\nu(z) \to 0.$$

This implies

$$\sum_{|\sigma|=k} r_\sigma^s \int_\Omega (\nabla u^k)(\varphi_\sigma(z))\Phi(\varphi_\sigma(z), z) d\nu(z) \to -\int_K \int_\Omega u_0(x,y)\nabla_y\Phi(x,y) d\overline{\mu}(x) d\nu(y). \tag{3.28}$$

On the other hand

$$\frac{1}{r_\sigma}\nabla u^k(\varphi_\sigma(z))\Phi(\varphi_\sigma(z),z))$$

is bounded in $L^2(\Omega)$ which implies the convergence to 0. Consequently

$$-\int_K\int_\Omega u_0(x,y)\nabla_y\Phi(x,y)d\overline{\mu}(x)d\nu(y)=0$$

for all Φ. Hence the two scale limit is a.e. independent of y, i.e. $u_0(x,y)=u_0(x)$.

Let us now suppose Φ is independent of y. We compute

$$\begin{aligned}&\sum_{|\sigma|=k} r_\sigma^s\int_\Omega(\nabla u^k)(\varphi_\sigma(z))\Phi(\varphi_\sigma(z))d\nu(z)=\\ =\;& -\sum_{|\sigma|=k} r_\sigma^s\int_\Omega u^k(\varphi_\sigma(z))\nabla\Phi(\varphi_\sigma(z))d\nu(z)\to\\ \to\;& -\int_K\int_\Omega u_0(x)\nabla\Phi(x)d\nu(y)d\overline{\mu}(x)=-\int_K u_0(x)\nabla\Phi(x)d\overline{\mu}(x)\end{aligned}$$

According to the two-scale convergence of ∇u^k, we have

$$\begin{aligned}&\sum_{|\sigma|=k} r_\sigma^s\int_\Omega(\nabla u^k)(\varphi_\sigma(z))\Phi(\varphi_\sigma(z))d\nu(z)\to\\ \to\;& \int_K\int_\Omega v(x,y)\Phi(x)d\nu(y)d\overline{\mu}(x).\end{aligned}$$

The last two relations implies that

$$-\int_K u_0(x)\nabla\Phi(x)d\overline{\mu}(x)=\int_K\int_\Omega v(x,y)\Phi(x)d\nu(y)d\overline{\mu}(x)\tag{3.29}$$

for all $\Phi\in C^1(\Omega)$. Denote

$$V(x)=\int_\Omega v(x,y)d\nu(y)\;\text{ and }u(x)=\int_a^x V(x)d\overline{\mu}(x).$$

Then

$$V(x)=\nabla^{\overline{\mu}}u(x).$$

By a density argument on Φ the assertion follows. □

References

[1] Allaire, G., *Homogeneization et convergence a deux echelle, application a un probleme de convection diffusion*, C.R. Acad Sci. Paris, **312**(1991), 581-586.

[2] Allaire, G., *Homogenization and two-scale convergence*, SIAM J. Math. Anal. **23**(6) (1992), 1482-1518.

[3] Bensoussan, A., Lions, J.L., Papanicolau, G., *Asymptotic analysis of periodic structures,* North Holland, Amsterdam, 1978.

[4] Cioranescu, D., Donato, P., *An Introduction to Homogenization*, Oxford University Press, New York, 1999.

[5] Evans, L.C., *Partial Differential Equations*, AMS, Providence, Rhode Island, 1998.

[6] Hutchinson, J.E., *Fractals and Self Similarity,* Indiana University Mathematics Journal, **30**(1981), no.5, 713-747.

[7] Kolumbán, J., *Two scale convergence of measures* (to appear).

[8] Nguetseng, G., *A general convergence result for a functional related to the theory of homogenization,* SIAM J. Math. Anal., **20**(3)(1989), 608-623.

[9] Nguetseng, G., *Asymptotic analysis for a still variational problem arising in mechanics*, SIAM J. Math. Anal. **21**(6)(1989), 1394-1414.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: asoos@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

ON SUPERCONVERGENT SPLINE COLLOCATION METHODS FOR THE RADIOSITY EQUATION

SANDA MICULA

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. In this paper we study collocation methods based on piecewise polynomial interpolation for the radiosity equation. We give a brief outline of this equation and its properties. With a special choice of interior nodes, we show that interpolation of degree r of the solution leads to an error in the collocation method of $O\left(h^{r+1}\right)$, where h is the mesh size of the triangulation. We conclude the paper by giving superconvergence results, considering separately the case where r is odd and the case where r is even.

1. The radiosity equation

Radiosity is a method of describing illumination based on a detailed analysis of light reflections off diffuse surfaces. It is typically used to render images of the interior of buildings. In computer graphics, the computation of lighting can be done via radiosity.

1.1. Definition. Properties

Radiosity is defined as being the energy per unit solid angle that leaves a surface. The *radiosity equation* is a mathematical model for the brightness of a collection of one or more surfaces. The equation is

$$u(P) - \frac{\rho(P)}{\pi} \int\limits_S u(Q)G(P,Q)V(P,Q)dS_Q = E(P), \quad P \in S \tag{1}$$

Received by the editors: 15.08.2006.

2000 *Mathematics Subject Classification.* 31A10, 31A15, 41A15, 41A55, 45B05, 65R20, 65M70.

Key words and phrases. integral equations, radiosity equation, collocation, spline, superconvergence.

where $u(P)$ is the *radiosity*, or the brightness, at $P \in S$. $E(P)$ is the *emissivity* at $P \in S$, the energy per unit area emitted by the surface.

The function $\rho(P)$ gives the *reflectivity* at $P \in S$, i. e. the bidirectional reflection distribution function. We have that $0 \leq \rho(P) < 1$, with $\rho(P)$ being 0 where there is no reflection at all at P. The radiosity equation is derived from the rendering equation under the *radiosity assumption*: all surfaces in the environment are *Lambertian diffuse reflectors*. What this means is that the reflectivity $\rho(P)$ is independent of the incoming and outgoing directions and, hence, of the angle at which the reflection takes place. Thus, $\rho(P)$ can be taken out from under the integral of a more general formulation (the rendering equation, see Cohen and Wallace [5]), leading to (1).

The function G, a geometric term, is given by

$$\begin{aligned} G(P,Q) &= \frac{[(Q-P)\cdot \mathbf{n}_P]\,[(P-Q)\cdot \mathbf{n}_Q]}{|P-Q|^4} \\ &= \frac{\cos\theta_P \cdot \cos\theta_Q}{|P-Q|^2} \end{aligned} \tag{2}$$

where $\mathbf{n}_P$ is the inner unit normal to S at P, θ_P is the angle between $\mathbf{n}_P$ and $Q-P$, and $\mathbf{n}_Q$ and θ_Q are defined analogously.

The function $V(P,Q)$ is a *visibility* function. It is 1 if the points P and Q are "mutually visible" (meaning they can "see each other" along a straight line segment which does not intersect S at any other point), and 0 otherwise. Surfaces S for which $V \equiv 1$ on S are called *unoccluded*, and this is the case that we will consider here. More about the radiosity equation can be found in Cohen and Wallace [5].

We can write (1) in the form

$$u(P) - \int_S K(P,Q)u(Q)dS_Q = E(P), \quad P \in S \tag{3}$$

with

$$K(P,Q) = \frac{\rho(P)}{\pi} G(P,Q)V(P,Q), \quad P,Q \in S \tag{4}$$

or, in operator form

$$(I-K)u = E \tag{5}$$

Let S be a smooth surface, although not necessarily connected. Later on, more assumptions on the surface S will be made.

The function $G(P,Q)$ given in (2) has a singularity at $P=Q$ and is smooth otherwise. Since this function plays an important role in the study of the solvability of equation (1), we give in the next lemma some of its properties.

Lemma 1. *Let S be a smooth C^{i+1} surface to which the Divergence Theorem can be applied. Let $P \in S$. Then*

a) $|G(P,Q)| \le c_1, \quad P,Q \in S, \quad P \neq Q;$

b) $G(P,Q) \ge 0,$ for $Q \in S;$

c) $\int\limits_S G(P,Q)\, dS_Q = \pi;$

d) if S is the unit sphere, then $G(P,Q) \equiv \dfrac{1}{4};$

e) $\left|D_Q^i G(P,Q)\right| \le \dfrac{c_2}{|P-Q|^i}, \quad P \neq Q,$ *c_2 independent of P and Q.*

For the proof, see [10].

Since the surface S is smooth and by Lemma 1, it is relatively easy to prove that the integral operator $\mathcal{K}$ of (5) is compact as an operator on either $C(S)$ or $L^2(S)$ into itself (see Mikhlin [13] pp. 160-162).

1.2. Solvability and Regularity of the Radiosity Equation

The solvability theory for the radiosity equation (1) is relatively straightforward, being based on the Geometric Series Theorem.

Let S be a smooth unoccluded surface (not necessarily connected). Thus the normal $\mathbf{n}_P$ is to be a continuous function of $P \in S$. In addition to the *radiosity assumption* (discussed in Section 1.1., we will also assume that the reflectivity function $\rho(P) \in C(S)$ and that it satisfies

$$\|\rho\|_\infty < 1 \tag{6}$$

From the physical point of view, what (6) means is that the surface does not reflect 100% of all the light that it receives, which is a reasonable assumption.

For the regularity of the solution of (1), we have

Lemma 2. *Let $m \geq 0$ be an integer, S a smooth unoccluded surface. Assume the reflectivity function $\rho \in C^{m+1}(S)$ and it satisfies (6). Then*

$$u \in C^m(S) \Rightarrow \mathcal{K}u \in C^{m+1}(S) \tag{7}$$

Theorem 3. *Let $m \geq 0$ be an integer. Let $\hat{S}$ be the boundary of a convex open set Ω, and assume $\hat{S}$ is a surface to which the Divergence Theorem can be applied. Assume S is a smooth (possibly disconnected) unoccluded surface $S \subset \hat{S}$. Also, assume $\rho, E \in C^m(S)$. Then*

(a) *The equation (1) is uniquely solvable for each E, with the solution $u(P)$ satisfying*

$$\|u\|_\infty \leq \frac{\|E\|_\infty}{1 - \|\mathcal{K}\|} \tag{8}$$

(b) *The solution $u \in C^m(S)$.*

For the proof, see [10].

2. **Preliminaries for Collocation Methods**

Let S be a smooth unoccluded surface in $\mathbb{R}^3$, which can be written as

$$S = S_1 \cup S_2 \cup ... \cup S_J \tag{9}$$

with each S_j the continuous image of a polygonal region in the plane

$$F_j : R_j \xrightarrow[onto]{1-1} S_j, \quad j = 1, ..., J \tag{10}$$

Generally, we will need to assume that the mappings F_j are several times continuously differentiable.

To create triangulations for S, we first triangulate each R_j and then map this triangulation onto S_j. Let $\{\widehat{\Delta}_{n,k}^j \mid k = 1, ..., n_j\}$ be a triangulation of R_j, and then define

$$\Delta_{n,k}^j = F_j(\widehat{\Delta}_{n,k}^j)$$

This yields a triangulation of S, which we refer to collectively as $\mathcal{T}_n = \{\Delta_1, ..., \Delta_n\}$. Let

$$h \equiv h_n = \max_{1\le j\le J} \max_{1\le k\le n_j} \textit{diameter}\left(\widehat{\Delta}^j_{n,k}\right) \tag{11}$$

be the mesh size of this triangulation. (The number of triangles n is to be understood implicitly; from now on, we dispense with it.)

We make the following assumptions concerning this triangulation:

T1. The set of all vertices of the surface S is a subset of the set of all vertices of the triangulation $\mathcal{T}_n$.

T2. The union of all edges of S is contained in the union of all edges of all triangles in $\mathcal{T}_n$.

T3. If two triangles in $\mathcal{T}_n$ have a nonempty intersection, then that intersection consists either of (i) a single common vertex, or (ii) all of a common edge.

We call triangulations satisfying T1 - T3 *conforming triangulations*.

Let Δ_k be some element from $\mathcal{T}_n$, and let it correspond to some $\widehat{\Delta}_k$, say $\widehat{\Delta}_k \subset R_j$ and $\Delta_k = F_j(\widehat{\Delta}_k)$. Let $\{\hat{v}_{k,1}, \hat{v}_{k,2}, \hat{v}_{k,3}\}$ denote the vertices of $\widehat{\Delta}_k$. Define $m_k : \sigma \xrightarrow[onto]{1-1} \Delta_k$ by

$$m_k(s,t) = F_j(u\hat{v}_{k,1} + t\hat{v}_{k,2} + s\hat{v}_{k,3}), \quad (s,t) \in \sigma, \quad u = 1 - s - t \tag{12}$$

(an affine mapping), where σ is the unit simplex $\sigma = \{(s,t) | 0 \le s, t, s + t \le 1\}$. Now we can define interpolation and numerical integration over a triangular surface element Δ by means of a similar formula over σ.

Let α be a given constant with $0 \le \alpha \le \dfrac{1}{3}$. Define the interpolation nodes by

$$q_{i,j} = \left(\frac{i + (r-3i)\alpha}{r}, \frac{j + (r-3j)\alpha}{r}\right), \quad i, j \ge 0, \quad i + j \le r \tag{13}$$

These $f_r = \dfrac{(r+1)(r+2)}{2}$ nodes form a uniform grid over σ. If $\alpha = 0$, some of these points are on the edges of σ. If $\alpha > 0$, then they are symmetrically placed points in the interior of σ. To avoid problems with the unit normal and with the nonsmoothness of the kernel, throughout this paper we want to consider only nodes that are interior to the triangular elements, so we will work with $0 < \alpha < \dfrac{1}{3}$.

Denote by $l_{i,j}(s,t)$ the corresponding Lagrange interpolation basis functions. Then for a given $g \in C(\sigma)$, the formula

$$p_r(s,t) = \sum_{0 \le i+j \le r} g(q_{i,j}) l_{i,j}(s,t) \tag{14}$$

is the unique polynomial of degree r that interpolates $g(s,t)$ at the nodes $\{q_{i,j} \mid i,j \ge 0, \quad i+j \le r\}$.

Denote the nodes and the basis functions collectively by $\{q_1, ..., q_{f_r}\}$ and $\{l_1, ..., l_{f_r}\}$. So, now we have the interpolation formula

$$g(s,t) \approx \sum_{j=1}^{f_r} g(q_j) l_j(s,t), \quad g \in C(S) \tag{15}$$

Integrating (15) over σ, we obtain the quadrature formula

$$\int_\sigma g(s,t) d\sigma \approx \sum_{j=1}^{f_r} \omega_j g(q_{i,j}) \tag{16}$$

where $\omega_j = \int_\sigma l_j(s,t) d\sigma$. Since the formula (15) is exact for all polynomials of degree $\le r$, formula (16) has degree of precision at least r.

Let

$$\mathcal{P}_n g(m_k(s,t)) = \sum_{j=1}^{f_r} g(m_k(q_j)) l_j(s,t), \quad P = m_k(s,t) \in \Delta_k \tag{17}$$

Define a collocation method using (17) (the collocation nodes coincide with the interpolation nodes). Substitute

$$\begin{aligned} u_n(P) &= \sum_{j=1}^{f_r} u_n(v_{k,j}) l_j(s,t), \quad P \in m_k(s,t) \in \Delta_k \\ v_{k,j} &= m_k(q_j), \quad k = 1, ..., n \end{aligned} \tag{18}$$

into (1). This leads to the linear system

$$\begin{aligned} u_n(v_i) \quad - \quad & \frac{\rho(P)}{\pi} \sum_{k=1}^{n} \sum_{j=1}^{nf_r} u_n(v_{k,j}) \int_\sigma G\left(v_i, m_k(s,t)\right) l_j(s,t) \\ & \cdot \quad \left|\left(D_s m_k \times D_t m_k\right)(s,t)\right| d\sigma = E(v_i), \quad i = 1, ..., nf_r \end{aligned} \tag{19}$$

This can be written abstractly as

$$(\mathcal{I} - P_n\mathcal{K})u_n = \mathcal{P}_n E \tag{20}$$

Also, introduce the iterated collocation solution

$$\hat{u}_n = E + \mathcal{K}u_n \tag{21}$$

We will give an error analysis based on standard projection operator theory (e. g. see Atkinson [2] Section 4.2). We have

Theorem 4. *Assume S is a smooth unoccluded surface in $\mathbb{R}^3$, and assume $S \subset \hat{S}$, with $\hat{S}$ the type of surface required in Lemma 1. Assume S satisfies (9) and (10) with each $F_j \in C^{r+2}$. Then for all sufficiently large n, say $n \geq n_0$, the operators $\mathcal{I} - P_n\mathcal{K}$ are invertible on $C(S)$ and have uniformly bounded inverses. Moreover, for the true solution u of (1) and the solution u_n of (20)*

$$\|u - u_n\|_\infty \leq \left\|(\mathcal{I} - P_n\mathcal{K})^{-1}\right\| \|(u - \mathcal{P}_n u)\|_\infty, \quad n \geq n_0 \tag{22}$$

Furthermore, if the emissivity $E \in C^{r+1}(S)$, then

$$\|u - u_n\|_\infty \leq O(h^{r+1}), \quad n \geq n_0 \tag{23}$$

3. **Superconvergent Collocation Methods**

So we know that under suitable assumptions, interpolation of degree r leads to an error of order $O(h^{r+1})$ in the collocation method associated with it. Sometimes at the collocation node points, the collocation method converges more rapidly than over all S, in which case

$$\lim_{n\to\infty} \frac{\max\limits_{1\leq i\leq nf_r} |u(v_i) - \hat{u}_n(v_i)|}{\|u - u_n\|_\infty} = 0 \tag{24}$$

Such methods are *superconvergent* at the collocation node points.

Let us examine more carefully the terms in (24). For simplicity, we work with the solution $\hat{u}_n$ of the iterated collocation equation (21). This should cause no problems, since we know that the convergence of $\hat{u}_n$ to u is at least as rapid as that

of the solution of the collocation equation (20) to u. Moreover, $\hat{u}(v_i) = u_n(v_i)$ at all collocation nodes.

By looking at the linear system associated with

$$(\mathcal{I} - \mathcal{K}P_n)(u - \hat{u}_n) = \mathcal{K}(u - \mathcal{P}_n u) \tag{25}$$

we have

$$\max_{1\le i\le nf_r} |u(v_i) - \hat{u}_n(v_i)| \le c \max_{1\le i\le nf_r} |\mathcal{K}(\mathcal{I} - P_n)u(v_i)| \tag{26}$$

(see Atkinson [2] p. 449). So, to find superconvergent methods, now we focus on finding errors for $\mathcal{K}(I - P_n)u(v_i)$.

Let $\tau \subset \mathbb{R}^2$ be a planar triangle with vertices $\{v_1, v_2, v_3\}$ and define the mapping $m_\tau : \sigma \longrightarrow \tau$ as in (12). For $g \in C(\tau)$, define

$$\mathcal{L}_\tau g(x, y) = \sum_{j=1}^{f_r} g(m_\tau(q_j))l_j(s, t) \tag{27}$$

which is a polynomial of degree r in the parametrization variables s and t, interpolating g at the nodes $\{m_\tau(q_1), ..., m_\tau(q_{f_r})\}$.

Define a numerical integration formula over τ by

$$\int_\tau g(x, y)d\tau \approx \int_\tau \mathcal{L}_\tau g(x, y)d\tau \tag{28}$$

which has degree of precision at least r. In what follows, for differentiable functions g, we will use the notation

$$|D^k g(x, y)| = \max_{0\le i\le k} \left| \frac{\partial^k g(x, y)}{\partial x^i \partial y^{k-i}} \right| \tag{29}$$

In investigating superconvergent collocation methods based on interpolation r, we have to distinguish two cases: where r is odd and where r is even.

3.1. Interpolation of Odd Degree

Consider the quadrature formula (28), based on interpolation of degree r, an odd number. It has degree of precision at least r. Suppose we can find a value $0 < \alpha_0 < \dfrac{1}{3}$, such that for $\alpha = \alpha_0$, formula (28) has degree of precision $r + 1$. Then, if we extend it to a rectangle, it will have degree of precision $r + 2$. We have the following result.

Lemma 5. *Let τ_1 and τ_2 be planar right triangles that form a square R of length h on a side. Let $g \in C^{r+3}(R)$. Let $\Phi \in L^1(R)$ two times differentiable with derivatives of order 1 and 2 in $L^1(R)$. Assume $\alpha = \alpha_0$. Then*

$$\left| \int_R \Phi(x,y)(I - \mathcal{L}_\tau)g(x,y)d\tau \right| \leq ch^{r+3} \left[\int_R (|\Phi| + |D\Phi| + |D^2\Phi|)d\tau \right] \max_{\substack{R \\ i=r+1,r+2,r+3}} \left\{ |D^i g| \right\} \tag{30}$$

with $\mathcal{L}_\tau g(x,y) \equiv \mathcal{L}_{\tau_i} g(x,y)$, where $(x,y) \in \tau_i$, $i = 1, 2$.

If integrating over a single triangle, the bound is given by

Lemma 6. *Let τ be a planar right triangle and assume the two sides which form the right angle have length h. Assume $\alpha = \alpha_0$. Let $g \in C^{r+2}(\tau), \Phi \in L^1(\tau)$ differentiable with first derivatives in $L^1(\tau)$. Then*

$$\left| \int_\tau \Phi(x,y)(\mathcal{I} - L_\tau)g(x,y)d\tau \right| \leq ch^{r+2} \left[\int_\tau (|\Phi| + |D\Phi|)d\tau \right] \cdot \max_\tau \left\{ |D^{r+1}g|, |D^{r+2}g| \right\} \tag{31}$$

where c denotes a generic constant.

For the proofs, see [10].

Remark. These results can be extended to general triangles, but then the derivatives of g and Φ will involve the mapping m_τ from (12). Let $h(\tau)$ denote the diameter of τ and $h^*(\tau)$ the radius of the circle inscribed in τ and tangent to its sides. Define

$$r(\tau) = \frac{h(\tau)}{h^*(\tau)} \tag{32}$$

Assume that for our triangulations $\mathcal{T}_n = \{\Delta_{n,k}\}$, $n \geq 1$, we have

$$\sup_n \left[\max_{\Delta_{n,k} \in \mathcal{T}_n} r(\Delta_{n,k}) \right] < \infty \tag{33}$$

Condition (33) prevents the triangles $\Delta_{n,k}$ from having angles which approach 0 as $n \to \infty$.

Now, we want to apply these results to the individual subintegrals in

$$\begin{aligned} \mathcal{K}u(v_i) &= \frac{\rho(v_i)}{\pi} \sum_{k=1}^{n} \int_\sigma G\left(v_i, m_k(s,t)\right) u\left(m_k(s,t)\right) \\ &\cdot \left| \left(D_s m_k \times D_t m_t\right)(s,t) \right| d\sigma, \quad i = 1, ..., 6n \end{aligned} \tag{34}$$

with

$$\begin{aligned} g(s,t) &= u\left(m_k(s,t)\right)\left|\left(D_s m_k \times D_t m_t\right)(s,t)\right| \\ \Phi(s,t) &= G\left(v_i, m_k(s,t)\right) \end{aligned} \tag{35}$$

Theorem 7. *Assume the hypotheses of Theorem 4, with each $F_j \in C^{r+2}$. Assume $u \in C^{r+2}(S)$. Assume the triangulation $\mathcal{T}_n$ of S satisfies (33) and that it is symmetric. For those integrals in (34) for which $v_i \in \Delta_k$, assume that all such integrals are evaluated with an error of $O(h^{r+3})$. Assume $\alpha = \alpha_0$. Then*

$$\max_{1\le i\le nf_r} |u(v_i) - \hat{u}_n(v_i)| \le ch^{r+3}\log h \tag{36}$$

Proof. We bound

$$\max_{1\le i\le nf_r} |\mathcal{K}(I - P_n)u(v_i))|$$

By our assumption, the error in evaluating the integral of (34) over Δ^* will be $O(h^{r+3})$.

Partition $\mathcal{T}_n^*$ into parallelograms to the maximum extent possible. Denote by $\mathcal{T}_n^{(1)}$ the set of all triangles making up such parallelograms and let $\mathcal{T}_n^{(2)}$ contain the remaining triangles. Then

$$\mathcal{T}_n^* = \mathcal{T}_n^{(1)} \cup \mathcal{T}_n^{(2)}$$

It is easy to show that the number of triangles in $\mathcal{T}_n^{(1)}$ is $O(n) = O(h^{-2})$, and the number of triangles in $\mathcal{T}_n^{(2)}$ is $O(\sqrt{n}) = O(h^{-1})$.

It can be shown that all but a finite number of the triangles in $\mathcal{T}_n^{(2)}$, bounded independent of n, will be at a minimum distance from v_i. That means that the triangles in $\mathcal{T}_n^{(2)}$ are "far enough" from v_i, so that the function $G(v_i, Q)$ is uniformly bounded for Q being in a triangle in $\mathcal{T}_n^{(2)}$.

By Lemma 6, the contribution to the error coming from the triangles in $\mathcal{T}_n^{(2)}$ will be $O\left(h^{r+3}\|D^{r+2}u\|_\infty\right)$.

Using Lemma 5 we have that the contribution to the error coming from triangles in $\mathcal{T}_n^{(1)}$ is of order

$$ch^{r+3} \int\limits_{S-\Delta^*} \sum_{j=0}^{2} \frac{1}{|v_i - Q|^j} dS_Q \tag{37}$$

Using a local representation of the surface and then using polar coordinates, the expression in (37) is of order

$$ch^{r+3} \left(h^2 + h + \log h\right) = O(h^{r+3} \log h)$$

Combining the errors arising from the integrals over Δ^*, $\mathcal{T}_n^{(1)}$, and $\mathcal{T}_n^{(2)}$, we have (36).

3.2. Interpolation of Even Degree

Analogously, consider the quadrature formula (28), based on interpolation of degree r, an even number, which has degree of precision at least r. Considered over a rectangle formed by two symmetric triangles, it has degree of precision $r+1$, since r is an even number. Define a collocation method with it as before. We have:

Lemma 8. *Let τ_1 and τ_2 be planar right triangles that form a square R of length h on a side. Let $g \in C^{r+2}(R)$. Let $\Phi \in L^1(R)$ differentiable with first order derivatives in $L^1(R)$. Then*

$$\left| \int\limits_R \Phi(x,y)(I - \mathcal{L}_\tau) g(x,y) d\tau \right| \leq ch^{r+2} \left[\int\limits_\tau (|\Phi| + |D\Phi|) d\tau \right] \cdot \max_{\substack{R \\ i=r+1, r+2}} \left\{ |D^i g| \right\} \tag{38}$$

with $\mathcal{L}_\tau g(x,y) \equiv \mathcal{L}_{\tau_i} g(x,y)$, where $(x,y) \in \tau_i$, $i = 1, 2$.

For integration over one triangle only, the term in h in (38) is only h^{r+1}. We use these results to prove the following superconvergence result.

Theorem 9. *Assume the hypotheses of Theorem 4, with each $F_j \in C^{r+2}$. Assume $u \in C^{r+2}(S)$. Assume the triangulation $\mathcal{T}_n$ of S satisfies (33) and that it is symmetric. For those integrals in (34) for which $v_i \in \Delta_k$, assume that all such integrals are evaluated with an error of $O(h^{r+2})$. Then*

$$\max_{1 \leq i \leq nf_r} |u(v_i) - \hat{u}_n(v_i)| \leq ch^{r+2} \tag{39}$$

The proof of Theorem 9 is similar to that of Theorem 7.

References

[1] Atkinson, K., *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind*, SIAM Publications, 1976.

[2] Atkinson, K., *The Numerical Solution of Integral Equations of the Second Kind*, Cambridge University Press, 1997.

[3] Atkinson, K., Chien, D., *The collocation method for solving the radiosity equation for unoccluded surfaces*, J. Integral Equations Appl., **10**(1998), 253-289.

[4] Atkinson, K., Chien, D., Seol, J., *Numerical analysis of the radiosity equation using the collocation method* , Electron. Trans. Numer. Anal., **11**(2000), 94-120.

[5] Cohen, M., Wallace, J., *Radiosity and Realistic Image Synthesis*, Academic Press, New York, 1993.

[6] Colton, D., *Partial Differential Equations: An Introduction*, Random House, New York, 1988.

[7] Hansen, O., *On the stability of the collocation method for the radiosity equation on polyhedral domains*, IMA Journal of Numerical Analysis, **22**(2002), 463-479.

[8] Günter, N., *Potential Theory*, Ungar Publishing Company, New York, 1967.

[9] Kress, R., *Linear Integral Equations*, Springer-Verlag, New York, 1989.

[10] Micula, S., *Numerical Methods for the Radiosity Equation and Related Problems*, (Ph. D. Thesis), University of Iowa, Iowa City, 1997.

[11] Micula, Gh., and Micula, S., *Handbook of Splines*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1999, 620 pp.

[12] Micula, Gh., and Micula, S., *On the Superconvergent Spline Collocation Methods for the Fredholm Integral Equations on Surfaces*, Mathematica Balkanica, New Series Vol. 19, 2005, Fasc. 1-2, 155-166.

[13] Mikhlin, S., *Mathematical Physics: An Advanced Course*, North-Holland Publishing, 1970.

[14] Rudin, W., *Real and Complex Analysis*, McGraw-Hill, 1986.

[15] Stroud, A., *Approximate Calculation of Multiple Integrals*, Prentice Hall, New Jersey, 1971.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: smicula@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

ON POSITIVE DEFINITENESS OF SOME LINEAR FUNCTIONALS

G.V. MILOVANOVIĆ, A.S. CVETKOVIĆ AND M.M. MATEJIĆ

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. In this paper we investigate the positive definiteness of linear functionals $\mathcal{L}$ defined on the space of all algebraic polynomials $\mathcal{P}$ by

$$\mathcal{L}(p) = \sum_{k\in\mathbb{N}} w_k p(z_k), \quad p \in \mathcal{P}.$$

1. Introduction

Let $\mathcal{P}$ be the space of all algebraic polynomials. In this paper we investigate linear functionals $\mathcal{L}$ defined by

$$\mathcal{L}(p) = \sum_{k\in\mathbb{N}} w_k p(z_k), \quad p \in \mathcal{P}. \tag{1}$$

In general, we investigate functionals for which $w_k, z_k \in \mathbb{C}\backslash\{0\}$, but with the following restrictions. First, we assume that $w_k \neq 0$, $k \in \mathbb{N}$. This condition is rather natural, since, assuming $w_k = 0$, for some $k \in \mathbb{N}$, simply produces a linear functional where summation is performed over $\mathbb{N}\backslash\{k\}$. Additionally, we will not loose any generality if we assume that $z_i \neq z_j$, $i, j \in \mathbb{N}$, since, for example, we may skip summation over j and use $w_i' = w_i + w_j$ at point z_i.

For the set of nodes z_i, $i \in \mathbb{N}$, we introduce the notation $\mathcal{Z} = \{z_k \mid k \in \mathbb{N}\}$.

Second we are going to assume that

$$\lim_{k\to+\infty} z_k = 0 \tag{2}$$

Received by the editors: 15.08.2006.

2000 *Mathematics Subject Classification.* 54C50, 32A05.

Key words and phrases. linear functionals, algebraic polynomials.

and, in order to have absolute integrability of all polynomials $p \in \mathcal{P}$, we assume that

$$\sum_{k\in\mathbb{N}} |w_k| \leq M < +\infty. \tag{3}$$

We assume in the sequel that the sequence z_k, $k \in \mathbb{N}$, is ordered in such a way that $|z_{k+1}| \leq |z_k|$, $k \in \mathbb{N}$.

Note that the linear functional $\mathcal{L}$ can be interpreted as the linear functional acting on the space of all bounded complex sequences ℓ_∞. Namely, according to the condition (3) we have that the sequence w_k, $k \in \mathbb{N}$, belongs to the space ℓ_1, the space of all absolutely summable complex sequences (see [3, p. 30], [2, p. 39]). As is known $\ell_1 \subset \ell'_\infty$, where ℓ'_∞ denotes dual of ℓ_∞.

Create now a linear mapping $\mathcal{I} : \mathcal{P} \mapsto \ell_\infty$ in the following way

$$\mathcal{I}(p) = (p(z_1), p(z_2), \ldots, p(z_n), \ldots).$$

The linear space $\mathcal{P}$ can be normed as

$$||p|| = \sup_{k\in\mathbb{N}} |p(z_k)|, \quad p \in \mathcal{P}.$$

Lemma 1.1. *The linear mapping $\mathcal{I} : \mathcal{P} \mapsto \ell_\infty$ is an bounded embedding of $\mathcal{P}$ into ℓ_∞.*

Proof. Given $\mathcal{L}$, any polynomial $p \in \mathcal{P}$ achieves its maximum on the compact set $\overline{\mathcal{Z}}$, hence any sequence $p(z_k)$, $k \in \mathbb{N}$, $p \in \mathcal{P}$, is uniformly bounded in k and belongs to ℓ_∞.

Norm preserving property is easily established. We note that if two polynomials satisfy $\mathcal{I}(p_1 - p_2) = 0$, we have that $p_1 = p_2$, since those are two analytic functions equal on the set $\mathcal{Z}$ which has one accumulation point. Hence, $\mathcal{I}(\mathcal{P}) \subset \ell_\infty$ is an embedding.

It is easily seen that $||\mathcal{I}|| = 1$. □

Now, define the linear functional $\mathcal{L}' : \ell_\infty \mapsto \mathbb{C}$ in the following way

$$\mathcal{L}'(u) = \sum_{k\in\mathbb{N}} w_k u_k, \quad u = (u_1, u_2, \ldots) \in \ell_\infty.$$

Obviously $\mathcal{L}'$ is bounded, since

$$|\mathcal{L}'(u)| \leq \sum_{k\in\mathbb{N}} |w_k||u_k| \leq ||u|| \sum_{k\in\mathbb{N}} |w_k|, \quad u \in \ell_\infty,$$

and $\mathcal{L}' \circ \mathcal{I} = \mathcal{L}$ on $\mathcal{P}$. Hence, for the certain extent we can identify $\mathcal{L}'$ and $\mathcal{L}$ and we may consider $\mathcal{L}'$ as a bounded linear extension of $\mathcal{L}$ to the whole of ℓ_∞.

Define $\mathcal{P}_+$ to be the set of all polynomials $p \in \mathcal{P}\backslash\{0\}$ which are nonnegative on the real line and denote by $\mathcal{P}_\mathbb{R}$ the set of all real algebraic polynomials.

We recall that linear functional $\mathcal{L} : \mathcal{P} \mapsto \mathbb{C}$ is called positive definite provided for every polynomial $p \in \mathcal{P}_+$ we have $\mathcal{L}(p) > 0$ (see [1, p. 13]). As a direct consequence of positive definiteness we have:

Lemma 1.2. *If the linear functional $\mathcal{L} : \mathcal{P} \mapsto \mathbb{C}$ is positive definite, then*

$$\mathcal{L}(x^{2n}) > 0, \quad \mathcal{L}(x^{2n+1}) \in \mathbb{R}, \quad \mathcal{L}(p) \in \mathbb{R}, \quad p \in \mathcal{P}_\mathbb{R}, \quad n \in \mathbb{N}_0. \tag{4}$$

Proof. Since $x^{2n} \in \mathcal{P}_+$, we have directly $\mathcal{L}(x^{2n}) > 0$. For the odd powers we have

$$\mathcal{L}(x-1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} (-1)^{2n-k} \mathcal{L}(x^k) > 0,$$

and using induction over $n \in 2\mathbb{N}$, we have

$$\mathcal{L}(x^{2n-1}) < \frac{1}{2n} \sum_{k=0, k\neq 2n-1}^{2n} \binom{2n}{k} (-1)^k \mathcal{L}(x^k).$$

Finally, we have according to linearity of $\mathcal{L}$ the last statement. □

The question we answer is summarized in the following theorem.

Theorem 1.1. *A linear functional $\mathcal{L}$ given by* (1) *is positive definite if and only if $w_k > 0$ and $z_k \in \mathbb{R}$, $k \in \mathbb{N}$.*

Finally, we introduce the following notation

$$e_n = (0, \ldots, 0, 1, 0, \ldots) \in \ell_\infty, \quad n \in \mathbb{N},$$

where number 1 occupies n-th position with zeros on all other positions.

2. Auxiliary results

We give first, the following auxiliary lemmas.

Lemma 2.1. *Choose $z_n \in \mathcal{Z}$ and assume that $\overline{z}_n \notin \mathcal{Z}$. Then there exists $p^n \in \mathbb{C}$, $|p^n| = 1$, such that for every $r^n \in \mathcal{P}_{\mathbb{R}}$ we have $p^n r^n(z_n) e_n \in \overline{\mathcal{I}(\mathcal{P}_{\mathbb{R}})}$. If $z_n \in \mathbb{R}\backslash\{0\}$ then $p^n = 1$.*

Proof. We are going to construct the sequence $p_k^n \in \mathcal{P}_+$, $k \in \mathbb{N}$, $n \in \mathbb{N}$, such that $\lim_{k\to+\infty} \mathcal{I}(p_k^n) = \alpha_n e_n$ for some complex number $\alpha_n \in \mathbb{C}\backslash\{0\}$.

Choose some fixed $z_n \in \mathcal{Z}$ and assume that $\overline{z}_n \notin \mathcal{Z}$. Then choose some polynomial $r^n \in \mathcal{P}_{\mathbb{R}}$. We define

$$p_k^n(z) = r^n(z) \prod_{i=1, i\neq n}^{k} \frac{(z - z_i)(z - \overline{z}_i)}{\lambda_i^n}, \quad k \in \mathbb{N},$$

where we denote

$$\lambda_i^n = |z_n - z_i||z_n - \overline{z}_i|, \quad i \neq n.$$

Obviously we have $p_k^n \in \mathcal{P}_+$, $k, n \in \mathbb{N}$.

Since r^n is an algebraic polynomial it is uniformly bounded on the compact set $\overline{\mathcal{Z}}$. Hence, for some $M > 0$ we have $|r^n(z_\nu)| < M$, $\nu \in \mathbb{N}$.

According to the property (2), we can choose some $i_{01} \in \mathbb{N}$ such that

$$|z_n|/2 < |z_\nu - z_i|, \quad |z_n|/2 < |z_\nu - \overline{z}_i|, \quad i > i_{01}, \quad \nu = 1, \ldots, n.$$

Fix some $q \in (0, 1)$. We can choose some $i_{02} \in \mathbb{N}$ such that

$$|z_i| < |z_n| q/4, \quad i > i_{02}.$$

Now, define $i_0 = \max\{i_{01}, i_{02}\}$. For $k > i_0$ and $\nu > k$, we have

$$\begin{aligned}
|p_k^n(z_\nu)| &= |r^n(z_\nu)| \prod_{i=1, i\neq n}^{i_0} \frac{|z_\nu - z_i||z_\nu - \overline{z}_i|}{\lambda_i^n} \prod_{i=i_0+1}^{k} \frac{|z_\nu - z_i||z_\nu - \overline{z}_i|}{|z_n - z_i||z_n - \overline{z}_i|} \\
&\leq M \prod_{i=1, i\neq n}^{i_0} \frac{|z_\nu - z_i||z_\nu - \overline{z}_i|}{\lambda_i^n} \prod_{i=i_0+1}^{k} \frac{|z_n| q/2 |z_n| q/2}{|z_n|/2 |z_n|/2} \\
&\leq M \left(\frac{2|z_1|}{m}\right)^{2i_0-2} q^{2(k-i_0)},
\end{aligned}$$

where $m = \min\limits_{i=1,\ldots,i_0,\ i\neq n} \{|z_n - z_i|, |z_n - \overline{z}_i|\} > 0$. We note that $p_k^n(z_\nu) = 0$ for $\nu < k$, $\nu \neq n$. From here it can be easily seen that we have uniform convergence in $\nu \neq n$ of $p_k^n(z_\nu)$ to zero for $k \to +\infty$, i.e., given $\varepsilon > 0$, for

$$k > k_{01} = i_0 + \frac{1}{2\log q} \log \frac{\varepsilon}{M} \left(\frac{m}{2|z_1|}\right)^{2i_0-2},$$

we have $|p_k^n(z_\nu) - 0| < \varepsilon$, $\nu \in \mathbb{N}\backslash\{n\}$.

Now, we consider $p_k^n(z_n)$, we have

$$|p_k^n(z_n)| = |r^n(z_n)|,$$

according to the definition of λ_i^n. This means that $p_k^n(z_n)$ has constant norm as $k \to +\infty$.

The product

$$\prod_{i=1,i\neq n}^{k} \frac{(z_n - z_i)(z_n - \overline{z}_i)}{\lambda_i^n}, \quad k \in \mathbb{N},$$

is just product of the complex numbers having modulus 1, hence, represent the sequence on the unit circle in the complex plane. According to the compactness of the unit circle in $\mathbb{C}$, we easily conclude that there exists some subsequence of the products which converge to some p^n which norm is one.

Denote set of indices for convergent subsequence as N_1. Then according to the convergence, given $\varepsilon > 0$, we can choose some $k_{02} \in N_1$, such that for $k > k_{02}$, $k \in N_1$, we have

$$|p_k^n(z_n) - r^n(z_n)p^n| < \varepsilon.$$

Now consider the vector $r^n(z_n)p^n e_n$, we have

$$||\mathcal{I}(p_k^n) - r^n(z_n)p^n e_n|| = \sup_{\nu\in\mathbb{N}} |p_k^n(z_\nu) - r^n(z_n)p^n e_n| < \varepsilon,$$

for $k > \max\{k_{01}, k_{02}\}$, $k \in N_1$.

Hence, if we enumerate, again the sequence p_k^n using only indexes $k \in N_1$, we have the sequence $p_k^n \in \mathcal{P}_{\mathbb{R}}$, such that

$$\lim_{k\to+\infty} \mathcal{I}(p_k^n) = p^n e_n.$$

Finally, if $z_n \in \mathbb{R}\backslash\{0\}$ we see that since $r^n \in \mathcal{P}_{\mathbb{R}}$, we have $r^n(z_n) \in \mathbb{R}$ and

$$p_k^n(z_n) = r^n(z_n) \prod_{i=1, i\neq n}^{k} \frac{|(z_n - z_i)(z_n - \overline{z}_i)|}{\lambda_i^n} \in \mathbb{R}$$

and also the terms of the product are simply equal to 1, hence, $p^n = 1$.

We can repeat construction for every $n \in \mathbb{N}$, i.e., every point $z_n \in \mathcal{Z}$ for which $\overline{z}_n \notin \mathcal{Z}$. $\square$

In the case $r^n \in \mathcal{P}_+$, we easily see that the sequence p_k^n also belongs to $\mathcal{P}_+$, so that we have the following result.

Lemma 2.2. *Assume that $\overline{z}_n \notin \mathcal{Z}$. Then there exists $p^n \in \mathbb{C}$, $|p^n| = 1$, such that for every $r^n \in \mathcal{P}_+$ we have $p^n r^n(z_n) e_n \in \overline{\mathcal{I}(\mathcal{P}_+)}$. If $z_n \in \mathbb{R}\backslash\{0\}$ we have $p^n = 1$.*

Next we consider the case when $\overline{z}_n \in \mathcal{Z}$. Without loss of generality we may assume that $z_{n+1} = \overline{z}_n$, since this can be achieved by the simple renumeration of the sequence z_n, $n \in \mathbb{N}$.

Lemma 2.3. *Let $z_{n+1} = \overline{z}_n$ for some $n \in \mathbb{N}$. Then there exist some $p^n \in \mathbb{C}$, $|p^n| = 1$, such that for every $r^n \in \mathcal{P}_{\mathbb{R}}$ we have*

$$p^n r^n(z_n) e_n + \overline{p^n r^n(z_n)} e_{n+1} \in \overline{\mathcal{I}(\mathcal{P}_{\mathbb{R}})}.$$

Proof. We consider the sequence of the polynomials

$$p_k^n(z) = r^n(z) \prod_{i=1, i\neq n, n+1}^{k} \frac{(z - z_i)(z - \overline{z}_i)}{\lambda_i^n},$$

where all notation is from the proof of Lemma 2.1. The only problem is definition of the sequence λ_i^n, but luckily we have

$$|z_n - z_i||z_n - \overline{z}_i| = |z_{n+1} - z_i||z_{n+1} - \overline{z}_i|,$$

since $z_{n+1} = \overline{z}_n$. Hence, we can apply safely the same definition.

It can proved using the same arguments that

$$|p_k^n(z_\nu) - 0| < \varepsilon,$$

provided

$$k > k_{01} = i_0 + \frac{1}{2\log q}\log\frac{\varepsilon}{M}\left(\frac{m}{2|z_1|}\right)^{2i_0-4}.$$

Also we have $p_k^n(z_n) = \overline{p_k^n(z_{n+1})}$, which gives the convergence for some sequence of $k \in N_1$ to mutually conjugated values. $\square$

It is clear that we may choose $r^n \in \mathcal{P}_+$ to get the following immediate consequence.

Lemma 2.4. *Let $z_{n+1} = \overline{z}_n$ for some $n \in \mathbb{N}$. Then there exist some $p^n \in \mathbb{C}$, $|p^n| = 1$, such that for every $r^n \in \mathcal{P}_+$ we have*

$$p^n r^n(z^n)e_n + \overline{p^n r^n(z_n)}e_{n+1} \in \overline{\mathcal{I}(\mathcal{P}_+)}.$$

3. Proof of the main result

Now we are ready to prove the main result.

Proof of Theorem 1.1. It can be easily seen that if $w_n > 0$ and $z_n \in \mathbb{R}$, $n \in \mathbb{N}$, for some $p \in \mathcal{P}_+$, we have

$$\mathcal{L}(p) = \sum_{k\in\mathbb{N}} w_k p(z_k) > 0,$$

according to the simple fact that $p(z_k) \geq 0$, $k \in \mathbb{N}$.

Now, assume that $\mathcal{L}$ is positive definite. Choose some $n \in \mathbb{N}$ and suppose that $\overline{z}_n \notin \mathcal{Z}$. Then, according to Lemma (2.1), we have

$$\lim_{k\to+\infty} \mathcal{L}(p_k^n) = \lim_{k\to+\infty} (\mathcal{L}' \circ \mathcal{I})(p_k^n) = \mathcal{L}'(p^n r^n(z_n)e_n),$$

where we have used the fact that $\mathcal{L}'$ is continuous on ℓ_∞. But then

$$\mathcal{L}'(p^n r^n(z_n)e_n) = w_n r^n(z_n)p^n.$$

Choose $r^n(z) = 1$, $r^n \in \mathcal{P}_+$, and $r^n(z) = z$, $r^n \in \mathcal{P}_\mathbb{R}$, to get

$$\mathcal{L}'(p^n e_n) = w_n p^n \geq 0 \quad \text{and} \quad \mathcal{L}'(p^n z_n e_n) = w_n z_n p^n \in \mathbb{R}.$$

Since $z_n \neq 0$ and according to the construction $p^n \neq 0$, we have that $\mathcal{L}'(p^n e^n) = w_n p^n > 0$. Then we have

$$z_n = \frac{\mathcal{L}'(p^n z_n e_n)}{\mathcal{L}'(p^n e_n)} \in \mathbb{R}$$

and also $w_n > 0$, according to the fact that $p^n = 1$ for $z_n \in \mathbb{R}\backslash\{0\}$.

Now let $\overline{z}_n = z_{n+1}$. Note that in this case we cannot have $z_n \in \mathbb{R}$, since in that case we would have $z_n = z_{n+1}$, which is impossible according to the conditions imposed on the set $\mathcal{Z}$. Then, according to Lemma (2.3) and positive definiteness of $\mathcal{L}$ for $r^n(z) = 1$ and $r^n(z) = z$, we have

$$\mathcal{L}'(p^n e_n + \overline{p^n} e_{n+1}) = w_n p^n + w_{n+1}\overline{p^n} = \alpha \geq 0$$

and

$$\mathcal{L}'(p^n z_n e_n + \overline{p^n z_n} e_{n+1}) = w_n z_n p^n + w_{n+1}\overline{z}_n = \beta \in \mathbb{R}.$$

We can rewrite these equations as the linear system in p^n and $\overline{p^n}$, which has the unique solution

$$p^n = \frac{\alpha\overline{z}_n - \beta}{w_n(\overline{z}_n - z_n)}, \quad \overline{p^n} = \frac{\alpha z_n - \beta}{w_{n+1}(z_n - \overline{z}_n)}.$$

Using these expressions we readily get $w_{n+1} = \overline{w}_n$ and also we see that we cannot have $\alpha^2 + \beta^2 = 0$, since it would imply $p^n = 0$, which is impossible.

Now, choose $r^n(z) = z^{2\nu}$, $\nu \in \mathbb{N}$. We have

$$\mathcal{L}'(p^n z_n^{2\nu} e_n + \overline{p^n z_n^{2\nu}} e_{n+1}) = w_n z_n^{2\nu} p^n + \overline{w_n z_n^{2\nu} p^n} = \mathrm{Re}(w_n z_n^{2\nu} p^n) \geq 0, \quad \nu \in \mathbb{N}_0.$$

If we denote $\alpha_n = \arg(w_n)$, $\beta_n = \arg(p^n)$ and $\varphi_n = \arg(z_n)$, where $\varphi_n \neq 0$ and $\varphi_n \neq \pi$, we get

$$|w_n z_n^{2\nu} p^n| \cos(\alpha_n + \beta_n + 2\nu\varphi_n) \geq 0, \quad \nu \in \mathbb{N}_0.$$

We want to show that there exist some $\nu \in \mathbb{N}_0$ such that cos function is negative which will produce a contradiction.

The cos-function is negative provided 2ν is an element of some interval

$$J_k = \left(\frac{(4k+1)\pi - 2(\alpha_n + \beta_n)}{2\varphi_n}, \frac{(4k+3)\pi - 2(\alpha_n + \beta_n)}{2\varphi_n} \right), \quad k \in \mathbb{Z}.$$

The interval J_k has length $\pi/|\varphi_n| > 1$, hence, there is at least one integer inside every interval J_k. If $\pi/|\varphi_n| > 2$ then there are at least two consecutive integers inside every J_k and at least one of them is even. Choosing 2ν to be equal to such an integer produces a contradiction. So, we assume $\pi/|\varphi_n| \leq 2$.

The intervals

$$G_k = \left[\frac{(4k+3)\pi - 2(\alpha_n+\beta_n)}{2\varphi_n}, \frac{(4k+5)\pi - 2(\alpha_n+\beta_n)}{2\varphi_n}\right], \quad k \in \mathbb{Z},$$

we are going to call gaps, obviously $\mathbb{R} = \cup_{k\in\mathbb{Z}}(J_k \cup G_k)$.

If $\pi/|\varphi_n| = 2$, we have $\varphi_n = \pm\pi/2$, which means that if

$$\cos(\alpha_n + \beta_n \pm 2\cdot 0\cdot \pi/2) > 0,$$

we have

$$\cos(\alpha_n + \beta_n \pm 2\cdot 1\cdot \pi/2) = -\cos(\alpha+\beta_n) < 0,$$

which produces a contradiction. If $\cos(\alpha_n + \beta_n \pm 2\cdot 0\cdot \pi/2) = 0$, then we have

$$\mathrm{Re}(w_n z_n^{2\nu} p^n) = |w_n z_n^{2\nu} p^n| \cos(\alpha_n + \beta_n \pm \nu\pi) = 0, \quad \nu \in \mathbb{N},$$

and, choosing $r^n(z) = z^{2\nu+1}$, $\nu \in \mathbb{N}_0$, we have

$$\begin{aligned}\mathrm{Re}(w_n z^{2\nu+1} p^n) &= |w_n z_n^{2\nu+1} p^n| \cos(\alpha_n+\beta_n \pm (2\nu+1)\pi/2)\\ &= \pm(-1)^\nu |w_n z_n^{2\nu+1} p^n| \sin(\alpha_n+\beta_n) \neq 0, \quad \nu\in\mathbb{N}_0.\end{aligned}$$

According to the fact $\cos(\alpha_n+\beta_n) = 0$, we have $\sin(\alpha_n+\beta_n) = \pm 1$, therefore, the expression cannot be equal zero. Consider now polynomials $r^n(z) = z^{2\nu}(z-1)^2$, $\nu \in \mathbb{N}_0$. Obviously $r^n \in \mathcal{P}_+$, so that it must be

$$\begin{aligned}\lim_{k\to+\infty} \mathcal{L}(p_k^n) &= \lim_{k\to+\infty} (\mathcal{L}'\circ\mathcal{I})(p_k^n) = w_n z_n^{2\nu}(z_n-1)^2 p^n + \overline{w_n z_n^{2\nu}(z_n-1)^2 p^n}\\ &= \mathrm{Re}(w_n z_n^{2\nu}(z_n-1)^2 p^n) \geq 0.\end{aligned}$$

According to linearity we must have

$$\begin{aligned}\mathrm{Re}(w_n z_n^{2\nu}(z_n-1)^2 p^n) &= \mathrm{Re}(-2w_n z_n^{2\nu+1} p^n)\\ &= \mp 2(-1)^\nu |w_n z_n^{2\nu+1} p^n| \sin(\alpha_n+\beta_n) > 0, \quad \nu\in\mathbb{N}_0.\end{aligned}$$

This is, of course, a contradiction.

Finally, it must be $1 < \pi/|\varphi_n| < 2$. Assume that in some interval J_k we have an integer $2m+1$. Then we can always choose some $\nu \in \mathbb{N}$, such that

$$\frac{\pi}{|\varphi_n|} > \frac{2\nu - 2m - 1}{2\nu - 2m - 2} > 1.$$

Then counting from $2m+1$ and finishing with 2ν there are exactly $2\nu-2m$ integers and those are covered with

$$2\nu-2m-2+1>\frac{2\nu-2m-1}{\pi/|\varphi_n|}+1,$$

intervals and gaps. Since we are starting and ending with an interval there are $\nu-m-1$ gaps and $\nu-m$ intervals. According to pigeon-hole principle there is at least one either interval or gap which contains at least two consecutive integers. If some interval contains two consecutive integers we are done. So assume that it is some gap. If gap contains even and odd integer, then next interval holds an even integer and we are done. If gap holds odd and even integer, then interval in front of it holds an even integer, and we are done.

We conclude that it cannot be $z_n, \overline{z}_n \in \mathcal{Z}$. We have seen also that if $z_n \in \mathcal{Z}$, then $z_n \in \mathbb{R}$ and $w_n > 0$, which finishes the proof. □

References

[1] Chihara, T.S., *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York, 1978.

[2] Lax, P.D., *Functional Analysis*, Wiley-Interscience, 2002.

[3] Rakočević, V., *Functional Analysis*, Naučna knjiga, Beograd, 1994.

University of Niš, Faculty of Electronic Engineering,
Department of Mathematics, P.O. Box 73, 18000 Niš, Serbia
E-mail address: glade@junis.ni.ac.yu

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

NUMERICAL SOLUTIONS OF LOTKA-VOLTERRA SYSTEM WITH DELAY BY SPLINE FUNCTIONS OF EVEN DEGREE

DIANA OTROCOL

Dedicated to Professor Gheorghe Coman at his 70^{th} *anniversary*

Abstract. This paper presents a numerical method for the approximate solution of a Lotka-Volterra system with delay. This method is essentially based on the natural spline functions of even degree introduced by using the derivative-interpolating conditions on simple knots.

1. Introduction

In recent years many papers were devoted to the problem of approximate integration of system of differential equation by spline functions. The theory of spline functions presents a special interest and advantage in obtaining numerical solutions of differential equations.

The splines functions of even degree are defined in a similar manner with that for odd degree spline functions, but using the derivative-interpolating conditions. These spline functions preserve all the remarkable extremal and convergence properties of the odd degree splines, and are very suitable for the numerical solutions of the differential equation problems, especially for the delay differential equations with initial conditions.

In this paper we consider a spline approximation method for the numerical solution of a Lotka-Volterra system with delay. The purpose of the present study is to extend the results of [1], [2], [3], [5] from the delay differential equations to the delay

Received by the editors: 02.02.2006.

2000 *Mathematics Subject Classification.* 34A34, 41A15.

Key words and phrases. natural spline function, derivative-interpolating function, delay differential system.

This work has been supported by MEdC-ANCS under grant ET 3233/17.10.2005.

differential system. In the same manner we shall develop some theory and algorithms for the numerical solutions of a class of delay Lotka-Volterra system.

2. Basic definitions and properties of even degree splines

Let Δ_n be the following partition of the real axis

$$\Delta_n : -\infty = t_0 < a = t_1 < ... < t_n = b < t_{n+1} = +\infty$$

and let m, n be two given natural numbers, satisfying the conditions $n \geq 1,\ m \leq n+1$. One denotes by I_k the following subintervals

$$I_k := [t_k, t_{k+1}[,\ k = \overline{1,n},\ I_0 :=]t_0, t_1[.$$

Definition 1. [3] *For the couple (m, Δ_n) a function $s : \mathbb{R} \to \mathbb{R}$ is called a natural spline function of even degree $2m$ if the following conditions are satisfied:*

1^0 $s \in C^{2m-1}(\mathbb{R})$,

2^0 $s\,|_{I_k} \in \mathcal{P}_{2m},\ k = \overline{1,n}$,

3^0 $s\,|_{I_0} \in \mathcal{P}_m,\ s\,|_{I_n} \in \mathcal{P}_m$,

where $\mathcal{P}_k$ represents the set of algebraic polynomials of degree $\leq k$.

We denote by $\mathcal{S}_{2m}(\Delta_n)$ the linear space of natural polynomial splines of even degree $2m$ with the simple knots $t_1, ..., t_n$.

We now show that $\mathcal{S}_{2m}(\Delta_m)$ is a finite dimensional linear space of functions and we give a basis of it.

Theorem 1. [3] *Any element $s \in \mathcal{S}_{2m}(\Delta_n)$ has the following representation*

$$s(t) = \sum_{i=0}^{m} A_i t^i + \sum_{k=1}^{n} a_k (t - t_k)_+^{2m},$$

where the real coefficients $(A_i)_0^m$ are arbitrary, and the coefficients $(a_k)_1^n$ satisfy the conditions

$$\sum_{k=1}^{n} a_k t_k^i = 0,\ i = \overline{0, m-1}.$$

Remark 1. [3] *If $n + 1 = m$, then $a_k = 0,\ k = \overline{1,n}$.*

Theorem 2. [3] *Suppose that* $n+1 \geq m$, *and let* $f : [t_1, t_n] \to \mathbb{R}$ *be a given function such that* $f'(t_k) = y'_k,\ k = \overline{1,n}$, *and* $f(t_1) = y_1$, *where* $y'_k,\ k = \overline{1,n}$, *and* y_1 *are given real numbers. Then there exists a unique spline function* $s_f \in \mathcal{S}_{2m}(\Delta_n)$, *such that the following derivative-interpolating conditions*

$$s_f(t_1) = y_1, \tag{2.1}$$

$$s'_f(t_k) = y'_k,\ k = \overline{1,n}, \tag{2.2}$$

hold.

Corollary 1. [3] *There exists a unique set of* $n+1$ *fundamental natural polynomial spline functions* $S_k \in \mathcal{S}_{2m}(\Delta_n),\ k = \overline{1,n}$, *and* $s_0 \in \mathcal{S}_{2m}(\Delta_n)$ *satisfying the conditions:*

$$\begin{aligned} &s_0(t_1) = 1, \quad s'_0(t_k) = 0, \qquad k = \overline{1,n}, \\ &S_k(t_1) = 0, \quad S'_k(t_i) = \delta_{ik}, \quad i,k = \overline{1,n}. \end{aligned}$$

It is clear that the functions $\{s_0, S_k,\ k = \overline{1,n}\}$, form a basis of the linear space $\mathcal{S}_{2m}(\Delta_n)$, and for s_f we obtain the representation

$$s_f(t) = s_0(t)f(t_1) + \sum_{k=1}^{n} S_k(t)f'(t_k).$$

But because $s_0(t) = 1$, it follows that

$$s_f(t) = f(t_1) + \sum_{k=1}^{n} S_k(t)f'(t_k).$$

Let us introduce the following sets of functions

$$\begin{aligned} &W_2^{m+1}(\Delta_n) := \{g : [a,b] \to \mathbb{R} \mid g^{(m)}\text{abs.cont.on } I_k \text{ and } g^{(m+1)} \in L_2[a,b]\}, \\ &W_2^{m+1}[a,b] := \{g : [a,b] \to \mathbb{R} \mid g^{(m)}\text{abs.cont.on } [a,b] \text{ and } g^{(m+1)} \in L_2[a,b]\}, \\ &W_{\substack{0\\2,f}}^{m+1}(\Delta_n) := \{g \in W_2^{m+1}(\Delta_n) \mid g'(t_k) = f'(t_k),\ k = \overline{1,n}\}, \\ &W_{2,f}^{m+1}(\Delta_n) := \{g \in W_2^{m+1}(\Delta_n) \mid g(t_0) = f(t_0)\}. \end{aligned}$$

Theorem 3. [3] *(Minimal norm property). If* $s \in \mathcal{S}_{2m}(\Delta_n) \cap W_{\substack{0\\2,f}}^{m+1}(\Delta_n)$, *then*

$$\left\| s^{(m+1)} \right\|_2 \leq \left\| g^{(m+1)} \right\|_2, \ \forall g \in W_{\substack{0\\2,f}}^{m+1}(\Delta_n),$$

holds, $\|\cdot\|_2$ *being the usual* L_2*-norm.*

For any function $f \in W_2^{m+1}(\Delta_n)$, we have the following corollaries.

Corollary 2. [3] $\left\| f^{(m+1)} \right\|_2^2 = \left\| s_f^{(m+1)} \right\|_2^2 + \left\| f^{(m+1)} - s_f^{(m+1)} \right\|_2^2$.

Corollary 3. [3] $\left\| s_f^{(m+1)} \right\|_2 \leq \left\| f^{(m+1)} \right\|_2$.

Corollary 4. [3] $\left\| f^{(m+1)} - s_f^{(m+1)} \right\|_2 \leq \left\| f^{(m+1)} \right\|_2$.

Remark 2. [3] *If* $\widetilde{s} := s_f + p_m$, *where* $p_m \in \mathcal{P}_m$, *it follows* $\left\| \widetilde{s}^{(m+1)} \right\|_2 \leq \left\| f^{(m+1)} \right\|_2$.

Theorem 4. [3] *(Best approximation property). If* $f \in W_2^{m+1}(\Delta_n)$ *and* $s_f \in \mathcal{S}_{2m}(\Delta_n)$ *is the derivative-interpolating spline function of even degree, then, for any* $s \in \mathcal{S}_{2m}(\Delta_n)$ *the relation*

$$\left\| s_f^{(m+1)} - f^{(m+1)} \right\|_2 \leq \left\| s^{(m+1)} - f^{(m+1)} \right\|_2$$

holds.

Remark 3. [3] *If* $s_f - s \in \mathcal{P}_m$ *then*

$$\left\| s_f^{(m+1)} - f^{(m+1)} \right\|_2 = \left\| s^{(m+1)} - f^{(m+1)} \right\|_2.$$

3. The numerical solutions of Lotka-Volterra system with delay by spline functions of even degree

Let us consider the following delay differential system with a constant delay $\omega > 0$

$$\frac{dy^u}{dt} = f^u(t, y^1(t), y^2(t), y^1(t-\omega), y^2(t-\omega)), \ a \leq t \leq b, \ u = 1, 2 \tag{3.1}$$

with initial conditions

$$y^u(t) = \varphi^u(t), \ t \in [a-\omega, a], \ u = 1, 2 \tag{3.2}$$

and we suppose that $f^u : D \subset \mathbb{R}^4 \to \mathbb{R}$, satisfies all the conditions assuring the existence and uniqueness of the solutions $y^u : [a, b] \to \mathbb{R}$ of the problem (3.1)+(3.2).

We propose an algorithm to approximate the solutions y^u of the problem (3.1)+(3.2) by spline functions of even degree $s^u \in \mathcal{S}_{2m}(\Delta_n)$, where Δ_n is a partition of $[a,b]$ and m,n are two integers satisfying the conditions $n \geq 1$ and $m \leq n+1$.

For $t \in [a, a+\omega]$, the problem (3.1)+(3.2) reduces to the following usual initial value problems:

$$\begin{cases} \dfrac{dy^u}{dt} = f^u(t, y^1(t), y^2(t), y^1(t-\omega), y^2(t-\omega)), \ a \leq t \leq a+\omega \\ y^u(t) = \varphi^u(a) = y_1^u \ , \ u = 1,2 \end{cases}$$

Theorem 5. *If y^u are the exact solutions of the problem (3.1)+(3.2), then, there exists some unique spline functions $s_{y^u} \in \mathcal{S}_{2m}(\Delta_n)$ such that:*

$$\begin{aligned} & s_{y^u}(t_1) = y^u(t_1) = \varphi^u(t_1), \\ & \frac{ds_{y^u}}{dt}(t_k) = \frac{dy^u}{dt}(t_k), \ k = \overline{1,n}, \ u = 1,2 \end{aligned} \tag{3.3}$$

The assertion of this theorem is a direct consequence of Theorem 2 by substituting t_1 by a and f by y^u.

Denoting $y_k^u := y^u(t_k)$, $\overline{y}_k^u := y^u(t_k - \omega)$, $k = \overline{1,n}$, $u = 1,2$, we have

$$\begin{aligned} & s_{y^u}(t_1) = y_1^u \\ & \frac{ds_{y^u}}{dt}(t_k) = f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2), \ k = \overline{1,n}, \ u = 1,2. \end{aligned}$$

Corollary 5. *If the functions $\{s_0, S_k, k = \overline{1,n}\}$ are the fundamental spline functions in $\mathcal{S}_{2m}(\Delta_n)$, then we can write*

$$s_{y^u}(t) = \varphi^u(a) + \sum_{k=1}^{n} S_k(t) f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2), \ u = 1,2, \tag{3.4}$$

where y_k^1, y_k^2, $k = \overline{2,n}$, are unknown, and

$$\overline{y}_k^u = \begin{cases} \varphi^u(t_k - \omega), & \text{if } t_k \leq a+\omega, \text{ are known,} \\ y^u(t_k - \omega), & \text{if } t_k > a+\omega, \text{ are unknown.} \end{cases}$$

We shall call the function $s_{y^u}(t)$, the approximating solution of the problem (3.1)+(3.2) and it can be written as follows

$$\begin{aligned} s_{y^u}(t) &= \varphi^u(a)+ \\ &+ \sum_{t_k \le a+\omega} S_k(t) f^u(t_k, y_k^1, y_k^2, \varphi^1(t_k-\omega), \varphi^2(t_k-\omega)) \\ &+ \sum_{t_k > a+\omega} S_k(t) f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2). \end{aligned} \tag{3.5}$$

For simplicity, in writing (3.5), let us use the following index sets:

$J_1 := \{j \in \mathbb{N} \mid t_j > a+\omega, \exists i : t_j - \omega = t_i\} =: \{j_1, j_2, ..., j_q\}$,

$J_0 := \{i \in \mathbb{N} \mid \exists j \in J_1 : t_j - \omega = t_i\} =: \{i_1, i_2, ..., i_q\}$,

$I := \{j \in \mathbb{N} \mid t_j > a+\omega, \nexists i : t_j - \omega = t_i\} =: \{d_1, d_2, ..., d_p\}$.

Thus, we can write (3.5) in the form

$$\begin{aligned} s_{y^u}(t) &= \varphi^u(a)+ \\ &+ \sum_{t_k \le a+\omega} S_k(t) f^u(t_k, y_k^1, y_k^2, \varphi^1(t_k-\omega), \varphi^2(t_k-\omega)) \\ &+ \sum_{k=1}^{q} S_{j_k}(t) f^u(t_{j_k}, y_{j_k}^1, y_{j_k}^2, y_{i_k}^1, y_{i_k}^2) \\ &+ \sum_{k=1}^{p} S_{d_k}(t) f^u(t_{d_k}, y_{d_k}^1, y_{d_k}^2, \overline{y}_{d_k}^1, \overline{y}_{d_k}^2), \end{aligned} \tag{3.6}$$

where the values $y_k^u,\ k = \overline{2,n}$, and $\overline{y}_k^u,\ k = \overline{1,p},\ u = 1,2$ are unknown.

Before giving an algorithm to determine these values, we shall give the following estimation error and convergence theorem.

Theorem 6. [3]*If $y^u \in W_2^{m+1}[a,b],\ u = 1,2$ are the exact solutions of the problem (3.1)+(3.2) and s_{y^u} is the spline approximating solution for y^u, the following estimations hold:*

$$\left\| y^{u(k)} - s_{y^u}^{(k)} \right\|_\infty \le \sqrt{m}(m-1)(m-2)...k\Delta_n^{m-k+\frac{1}{n}} \left\| y^{u(m+1)} \right\|_2,$$

for $k = 1, 2, ..., m$, where $\|\Delta_n\| := \max\limits_{i=\overline{2,n}} \{t_i - t_{i-1}\},\ u = 1,2$.

Corollary 6. [3]*If $y^u \in W_2^{m+1}[a,b]$, we have*

$$\|y^u - s_{y^u}\|_\infty \le (b-a)\sqrt{m}(m-1)! \left\| y^{u(m+1)} \right\|_2 \|\Delta_n\|^{m-\frac{1}{2}},\ u = 1,2.$$

Corollary 7. [3] $\lim\limits_{\|\Delta_n\| \to 0} \left\| y^{u(k)} - s_{y^u}^{(k)} \right\|_\infty = 0,\ k = \overline{1,m},\ u = 1,2.$

4. **Effective development of the algorithm**

For any $t \in [a,b]$, we suppose that $y^u(t) \approx s_{y^u}(t),\ u = 1,2.$

If we denote, as usual, $e^u(t) := y^u(t) - s_{y^u}(t),\ t \in [a,b]$,

we have

$$|e^u(t)| \leq \sqrt{m}(m-1)!\,\|\Delta_n\|^{m-\frac{1}{2}} \left\| y^{u(m+1)} \right\|_2,$$

or

$$|e^u(t)| = O(\|\Delta_n\|^{m-\frac{1}{2}}),\ \forall t \in [a,b].$$

If we denote

$$\begin{aligned} w_i^u &:= s_{y^u}(t_i), & e_i^u &:= e^u(t_i) = y^u(t_i) - s_{y^u}(t_i), i = \overline{1,n}, \\ \overline{w}_i^u &:= s_{y^u}(t_i - \omega), & \overline{e}_i^u &:= e^u(t_i - \omega) = y^u(t_i - \omega) - s_{y^u}(t_i - \omega), i = \overline{1,n}, \end{aligned}$$

then we have $y_i^u = w_i^u + e_i^u,\ \overline{y}_i^u = \overline{w}_i^u + \overline{e}_i^u$, where

$$w_i^u = y_1^u + \sum_{k=1}^{n} S_k(t_i) f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2),\ i = \overline{1,n},\ u = 1,2, \tag{4.1}$$

$$\overline{w}_i^u = y_1^u + \sum_{k=1}^{n} S_k(t_i - \omega) f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2),\ i = \overline{1,n},\ u = 1,2.$$

In what follows, we suppose that in (3.1)+(3.2) the functions

$$\begin{aligned} & f^u : D \subset \mathbb{R}^5 \to \mathbb{R}\ (D \subset [a,b] \times \mathbb{R}^4), \\ & \frac{\partial f^u(t,u_1,u_2,u_3,u_4)}{\partial u_1},\ \frac{\partial f^u(t,u_1,u_2,u_3,u_4)}{\partial u_2}, \\ & \frac{\partial f^u(t,u_1,u_2,u_3,u_4)}{\partial u_3},\ \frac{\partial f^u(t,u_1,u_2,u_3,u_4)}{\partial u_4} \end{aligned}$$

are continuous. Thus,

$$\begin{aligned} f^u(t_k, y_k^1, y_k^2, \overline{y}_k^1, \overline{y}_k^2) &= f^u(t_k, w_k^1 + e_k^1, w_k^2 + e_k^2, \overline{w}_k^1 + \overline{e}_k^1, \overline{w}_k^2 + \overline{e}_k^2) \\ &= f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2) + e_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_1} \\ &+ e_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_2} + \overline{e}_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_3} \\ &+ \overline{e}_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_4} \end{aligned}$$

where

$$\min(w_k^u, w_k^u + e_k^u) < \xi_k^u < \max(w_k^u, w_k^u + e_k^u),$$

$$\min(\overline{w}_k^u, \overline{w}_k^u + \overline{e}_k^u) < \eta_k^u < \max(\overline{w}_k^u, \overline{w}_k^u + \overline{e}_k^u),\ u = 1, 2.$$

We can write the system (4.1) in the form

$$w_i^u = y_1^u + \sum_{k=1}^{n} S_k(t_i) f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2) + E_i^u,\ i = \overline{1, n},\ u = 1, 2$$

$$\overline{w}_i^u = y_1^u + \sum_{k=1}^{n} S_k(t_i - \omega) f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2) + \overline{E}_i^u,\ i = \overline{1, n},\ u = 1, 2$$

where

$$\begin{aligned} E_i^u &= \sum_{k=1}^{n} S_k(t_i) e_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_1} + \sum_{k=1}^{n} S_k(t_i) \overline{e}_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_2} \\ &+ \sum_{k=1}^{n} S_k(t_i) e_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_3} + \sum_{k=1}^{n} S_k(t_i) \overline{e}_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_4} \\ &= O(\|\Delta_n\|^{m-\frac{1}{2}}), \end{aligned}$$

$$\begin{aligned} \overline{E}_i^u &= \sum_{k=1}^{n} S_k(t_i - \omega) e_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_1} + \\ &+ \sum_{k=1}^{n} S_k(t_i - \omega) \overline{e}_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_2} + \\ &+ \sum_{k=1}^{n} S_k(t_i - \omega) e_k^1 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_3} + \\ &+ \sum_{k=1}^{n} S_k(t_i - \omega) \overline{e}_k^2 \frac{\partial f^u(t_k, \xi_k^1, \xi_k^2, \eta_k^1, \eta_k^2)}{\partial u_4} = \\ &= O(\|\Delta_n\|^{m-\frac{1}{2}}),\ i = \overline{1, n},\ u = 1, 2, \end{aligned}$$

supposing that

$$\begin{aligned} &\left| \frac{\partial f^u(t, u_1, u_2, u_3, u_4)}{\partial u_1} \right| \le M_1, \left| \frac{\partial f^u(t, u_1, u_2, u_3, u_4)}{\partial u_2} \right| \le M_2, \\ &\left| \frac{\partial f^u(t, u_1, u_2, u_3, u_4)}{\partial u_3} \right| \le M_3, \left| \frac{\partial f^u(t, u_1, u_2, u_3, u_4)}{\partial u_4} \right| \le M_4, \end{aligned} \tag{4.2}$$

on D. Obviously, $E_i^u \to 0$ and $\overline{E}_i^u \to 0$ for $\|\Delta_n\| \to 0,\ u = 1, 2$.

Now, we have to solve the following nonlinear system:

$$\begin{cases} w_i^u = y_1^u + \sum\limits_{k=1}^{n} S_k(t_i) f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2),\ i = \overline{1,n}, \\ \overline{w}_i^u = y_1^u + \sum\limits_{k=1}^{n} S_k(t_i - \omega) f^j(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2),\ i = \overline{1,n}\,. \end{cases} \tag{4.3}$$

Let us denote:

$$w^u := (w_1^u, ..., w_n^u),\ \overline{w}^u := (\overline{w}_1^u, ..., \overline{w}_n^u),\ W^u = (w^u, \overline{w}^u),$$

$$H_i^u(w, \overline{w}) := y_1^u + \sum_{k=1}^{n} S_k(t_i) f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2),\ i = \overline{1,n},$$

$$\overline{H}_i^u(w, \overline{w}) := y_1^u + \sum_{k=1}^{n} S_k(t_i - \omega) f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2),\ i = \overline{1,n},$$

$$H^u(W^u) := H^u(w^u, \overline{w}^u)$$

$$:= (H_1^u(w^u, \overline{w}^u), ..., H_n^u(w^u, \overline{w}^u), \overline{H}_1^u(w^u, \overline{w}^u), ..., \overline{H}_n^u(w^u, \overline{w}^u))$$

and

$$A^u = \begin{pmatrix} \frac{\partial H_1^u(w^u, \overline{w}^u)}{\partial w_1^u} & \cdots & \frac{\partial H_1^u(w^u, \overline{w}^u)}{\partial w_n^u} & \frac{\partial H_1^u(w^u, \overline{w}^u)}{\partial \overline{w}_1^u} & \cdots & \frac{\partial H_1^u(w^u, \overline{w}^u)}{\partial \overline{w}_n^u} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial H_n^u(w^u, \overline{w}^u)}{\partial w_1^u} & \cdots & \frac{\partial H_n^{uj}(w^u, \overline{w}^u)}{\partial w_n^u} & \frac{\partial H_n^u(w^u, \overline{w}^u)}{\partial \overline{w}_1^u} & \cdots & \frac{\partial H_n^u(w^u, \overline{w}^u)}{\partial \overline{w}_n^u} \\ \frac{\partial \overline{H}_1^u(w^u, \overline{w}^u)}{\partial w_1^u} & \cdots & \frac{\partial \overline{H}_1^u(w^u, \overline{w}^u)}{\partial w_n^u} & \frac{\partial \overline{H}_1^u(w^u, \overline{w}^u)}{\partial \overline{w}_1^u} & \cdots & \frac{\partial \overline{H}_1^u(w^u, \overline{w}^u)}{\partial \overline{w}_n^u} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \overline{H}_n^u(w^u, \overline{w}^u)}{\partial w_1^u} & \cdots & \frac{\partial \overline{H}_n^u(w^u, \overline{w}^u)}{\partial w_n^u} & \frac{\partial \overline{H}_n^u(w^u, \overline{w}^u)}{\partial \overline{w}_1^u} & \cdots & \frac{\partial \overline{H}_n^u(w^u, \overline{w}^u)}{\partial \overline{w}_n^u} \end{pmatrix}$$

Shortly, we write the system (4.3) by

$$W^u = H^u(W^u) \tag{4.4}$$

In order to investigate the solvability of the nonlinear system (4.4) we shall use a classical theorem.

Theorem 7. [6] *Let $\Omega \subset \mathbb{R}^{2n+2}$ be a bounded domain and let $H^u : \Omega \to \Omega$ be a vector function defined by*

$$\begin{aligned} W^u &= (w^u, \overline{w}^u) \longmapsto \\ &(H_1^u(w^u, \overline{w}^u), ..., H_n^u(w^u, \overline{w}^u), \overline{H}_1^u(w^u, \overline{w}^u), ..., \overline{H}_n^u(w^u, \overline{w}^u)) \\ &= H^u(W^u). \end{aligned}$$

If the functions H^u, and $\dfrac{\partial H^u}{\partial W^u}$, are continuous in Ω, then there exists in Ω a fixed point W^{u*} of H^u, i.e. $W^{u*} = H^u(W^{u*})$, which can be found by iterations. $W^{u*} = \lim\limits_{n\to\infty} W^{u(n)}$, $W^{u(k)} := H^u(W^{u(k-1)})$, $k = 1, 2, ...$, $W^{u(0)} \in \Omega$ (arbitrary). If in addition $\|A\| \leq L < 1$, for any iteration $W^{u(k)}$, the following estimation holds:

$$\left\| W^u - W^{u(k)} \right\| \leq \frac{L^k}{1-L} \left\| W^{u(1)} - W^{u(0)} \right\|.$$

Taking in consideration the expression of H^u, the matrix A^u is $A^u = SF^u$, where

$$S = \begin{pmatrix} S_1(t_1) & \cdots & S_n(t_1) & S_1(t_1) & \cdots & S_n(t_1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ S_1(t_n) & \cdots & S_n(t_n) & S_1(t_n) & \cdots & S_n(t_n) \\ S_1(t_1-\omega) & \cdots & S_n(t_1-\omega) & S_1(t_1-\omega) & \cdots & S_n(t_1-\omega) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ S_1(t_n-\omega) & \cdots & S_n(t_n-\omega) & S_1(t_n-\omega) & \cdots & S_n(t_n-\omega) \end{pmatrix}$$

and F is the diagonal matrix with the following elements:

$$\frac{\partial f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2)}{\partial w_k^u}, \ \frac{\partial f^u(t_k, w_k^1, w_k^2, \overline{w}_k^1, \overline{w}_k^2)}{\partial \overline{w}_k^u}, k = \overline{1, n}, \ u = 1, 2.$$

Theorem 8. *Suppose that there exists the constants M, N such that (4.2) holds and*

$$|f^u(t, u_1, u_2, u_3, u_4)| \leq N_u, \ \forall (t, u_1, u_2, u_3, u_4) \in D, \ u = 1, 2.$$

If $M_u \leq \|S\|^{-1}$, then the system (4.3) has a solution which can be found by iterations.

5. Numerical example

Example 1. *Consider the following Lotka-Volterra delay differential system*

$$\begin{cases} \dfrac{dy^1}{dt} = y^1\left[y^1(t-1) + y^2(t-1) + 1 - e^{t-1} - e^{2t-2}\right] \\ \dfrac{dy^2}{dt} = y^2\left[y^2(t-1) + 2 - e^{2t-2}\right] \end{cases}, t \in [0,b],$$

with initial conditions

$$\begin{cases} y^1(t) = \varphi^1(t) = e^t,\ t \in [-1,0] \\ y^2(t) = \varphi^2(t) = e^{2t},\ t \in [-1,0] \end{cases},$$

and the corresponding exact solutions

$$(y^1(t), y^2(t)) = (e^t, e^{2t}).$$

In the below table are given the actual errors for the considered examples. The table list

$$\max\left\{|w_i^u - y^u(t_i)|,\ i = \overline{1,n};\ |\overline{w}_i^u - y^u(t_j - \omega)|,\ j \in I;\right.$$
$$\left.|s_{y^u}(a + 0.1i) - y(a + 0.1i)|,\ i = \overline{1, 10(b-a)}\right\},$$

for $m = 1, 2, 3$ and the interval $[a,b]$ is $[0,2]$.

$[a,b]$	$[0,2]$		
$n \setminus m$	1	2	3
6	65.6521	5.4291	6.4198
9	12.2874	0.75975	0.25095
11	7.0645	0.39634	0.072303

For $a = 0,\ b = 2,\ \omega = 1,\ m = 1,\ n = 6$ we obtain $r = 3$ (the number of the nodes at the left of $a + \omega$), $p = 3,\ q = 0$. The approximating solution $\widetilde{s}^u$ and the exact solution y^u , $u = 1, 2$, in this case,are plotted in FIGURE 1 and FIGURE 2. For $a = 0,\ b = 2,\ \omega = 1,\ m = 2,\ n = 9$ we obtain $r = 5,\ p = 0,\ q = 4$. The approximating solution $\widetilde{s}^u$ and the exact solution y^u , $u = 1, 2$, in this case, are plotted in FIGURE 3 and FIGURE 4.

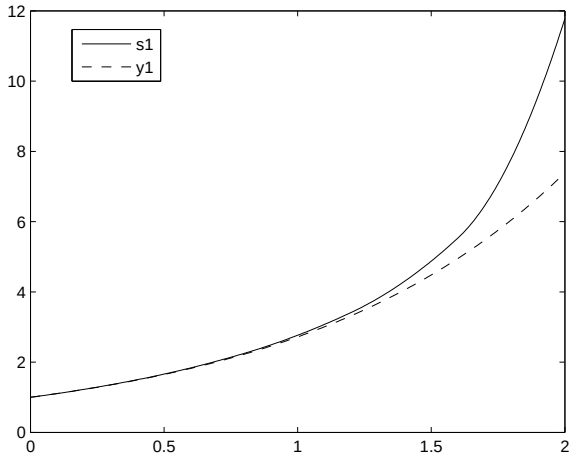


FIGURE 1. Comparison between the approximation solution $\widetilde{s}^1$ and the exact solution y^1 in the first case.

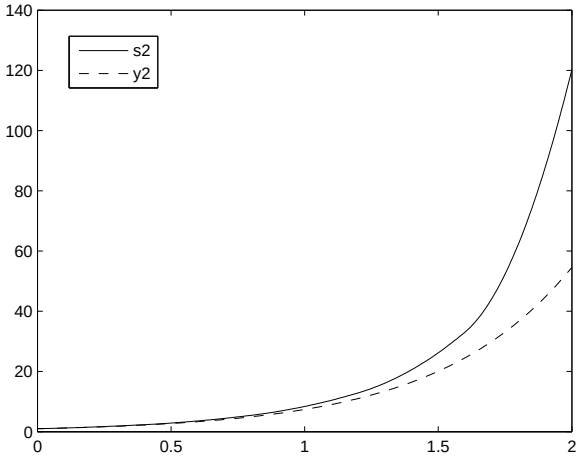


FIGURE 2. Comparison between the approximation solution $\widetilde{s}^2$ and the exact solution y^2 in the first case.

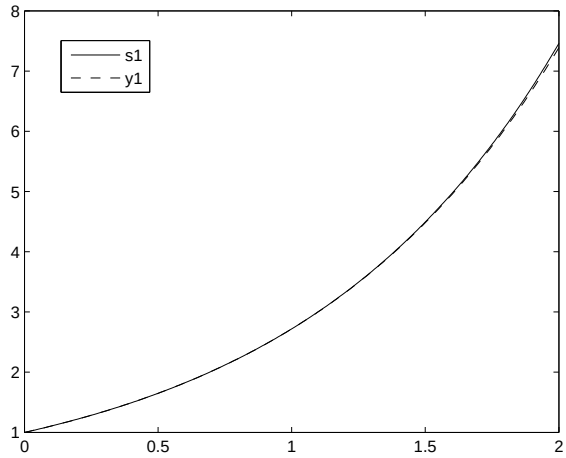


FIGURE 3. Comparison between the approximation solution $\widetilde{s}^1$ and the exact solution y^1 in the second case.

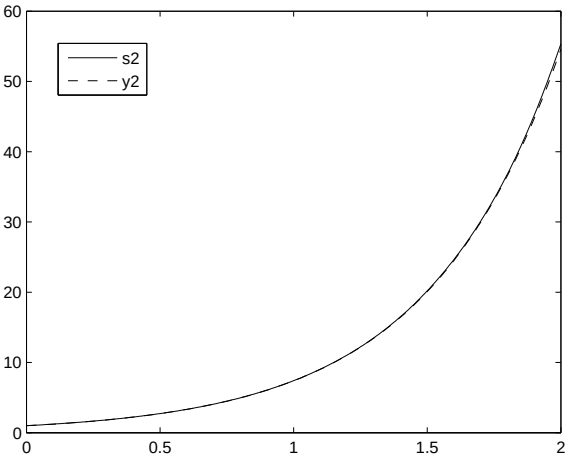


FIGURE 4. Comparison between the approximation solution $\widetilde{s}^2$ and the exact solution y^2 in the second case.

References

[1] Akça, H., Micula, Gh., *Numerical solutions of system of differential equation with deviating argument by spline functions*, Itinerant seminar of functional equations approximation and convexity, Cluj-Napoca, 1990.

[2] Blaga, P., *Some even degree spline interpolation*, Studia Univ. "Babes-Bolyai", Mathematica, **37**, 1(1992), 65-72.

[3] Blaga, P., Micula, Gh., *Polynomial natural spline functions of even degree*, Studia Univ. Babeş-Bolyai, Mathematica **38**, 2(1993), 31-40.

[4] Blaga, P., Micula, Gh., *Polynomial spline functions of even degree approximating the solutions of differential equations*, Analele Universităţii din Timişoara, Vol. **36**(1998), fasc. 2.

[5] Blaga, P., Micula, Gh., Akça, H., *On the use of spline functions of even degree for the numerical solution of the delay differential equations*, Calcolo **32**, no. 1-2(1996), 83-101.

[6] Coman, Gh., Pavel, G., Rus, I., Rus, I.A., *Introducere în teoria ecuaţiilor operatoriale*, Editura Dacia, Cluj-Napoca, 1976.

[7] Micula, Gh., *Funcţii spline şi aplicaţii*, Editura Tehnică, Bucureşti, 1978.

"Babeş-Bolyai" University, Department of Applied Mathematics,
Str. M. Kogalniceanu 1, RO-400084 Cluj-Napoca, Romania
E-mail address: dotrocol@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

A NOTE ON MULTIVALUED MEIR-KEELER TYPE OPERATORS

ADRIAN PETRUŞEL AND GABRIELA PETRUŞEL

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. The purpose of this work is to present some fixed point results for multivalued generalized Meir-Keeler type operators.

1. Introduction

Throughout this paper, the standard notations and terminologies in nonlinear analysis (see [12], [13], [8]) are used. For the convenience of the reader we recall some of them.

Let (X,d) be a metric space. By $\tilde{B}(x_0;r)$ we denote the closed ball centered in $x_0 \in X$ with radius $r>0$.

Also, we will use the following symbols:

$P(X) := \{Y \subset X |\ Y \text{ is nonempty}\}$, $P_{cl}(X) := \{Y \in P(X) |\ Y \text{ is closed}\}$,

$P_b(X) := \{Y \in P(X) |\ Y \text{ is bounded }\}$, $P_{b,cl}(X) := P_{cl}(X) \bigcap P_b(X)$.

Let A and B be nonempty subsets of the metric space (X,d). The gap between these sets is

$$D(A,B) = \inf\{d(a,b) |\ a \in A,\ b \in B\}.$$

In particular, $D(x_0,B) = D(\{x_0\},B)$ (where $x_0 \in X$) is called the distance from the point x_0 to the set B.

Also, if $A, B \in P_b(X)$, then one denote

$$\delta(A,B) := \sup\{d(a,b) |\ a \in A,\ b \in B\}.$$

Received by the editors: 15.08.2006.

2000 *Mathematics Subject Classification.* 47H10, 54H25.

Key words and phrases. metric space, fixed point, multivalued operator, Meir-Keeler operator, generalized contraction.

The Pompeiu-Hausdorff generalized distance between the nonempty closed subsets A and B of the metric space (X,d) is defined by the following formula:

$$H(A,B) := \max\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(a,b)\}.$$

The symbol $T : X \multimap Y$ means $T : X \to P(Y)$, i. e. T is a set-valued operator from X to Y. We will denote by $Graf(T) := \{(x,y) \in X \times Y | y \in T(x)\}$ the graph of T.

For $T : X \to P(X)$ the symbol $Fix(T) := \{x \in X |\ x \in T(x)\}$ denotes the fixed point set of the set-valued operator T. Also, for $x \in X$, we denote $F^n(x) := F(F^{n-1}(x))$, $n \in \mathbb{N}^*$, where $F^0(x) := x$.

Definition 1.1. If $f : X \to X$ is an single-valued operator, let us consider the following conditions:

i) α-contraction condition:

(1) $\alpha \in [0,1[$ and for $x,y \in X \Rightarrow d(f(x),f(y)) \leq \alpha d(x,y)$;

ii) contractive condition:

(2) $x,y \in X,\ x \neq y \Rightarrow d(f(x),f(y)) < d(x,y)$;

iii) Meir-Keeler type condition:

(3) for each $\eta > 0$ there exists $\delta > 0$ such that $x,y \in X$, $\eta \leq d(x,y) < \eta+\delta \Rightarrow d(f(x),f(y)) < \eta$;

iv) ϵ-locally Meir-Keeler type condition (where $\epsilon > 0$)

(4) for each $0 < \eta < \epsilon$ there is $\delta > 0$ such that $x,y \in X$, $\eta \leq d(x,y) < \eta+\delta \Rightarrow d(f(x),f(y)) < \eta$.

Let us observe that, condition (iii) implies (ii), (iii) implies (iv) and each of these conditions implies the continuity of f.

Definition 1.2. If $F : X \to P_{cl}(X)$ is a multi-valued operator then F is said to be:

i) α-contraction if:

(5) $\alpha \in [0,1[$ and for $x,y \in X \Rightarrow H(F(x),F(y)) \leq \alpha d(x,y)$;

ii) contractive if:

(6) $x,y \in X,\ x \neq y \Rightarrow H(F(x),F(y)) < d(x,y)$;

iii) Meir-Keeler type operator if:

(7) for each $\eta > 0$ there exists $\delta > 0$ such that $x, y \in X$, $\eta \leq d(x, y) < \eta + \delta \Rightarrow H(F(x), F(y)) < \eta$;

iv) ϵ-locally Meir-Keeler type operator (where $\epsilon > 0$) if:

(8) for each $0 < \eta < \epsilon$ there is $\delta > 0$ such that $x, y \in X$, $\eta \leq d(x, y) < \eta + \delta \Rightarrow H(F(x), F(y)) < \eta$.

It is easily to see that condition (iii) implies (ii), (iii) implies (iv) and each of these conditions implies the upper semi-continuity of F.

The following theorems are fundamental in the theory of Meir-Keeler type operators.

The first result is known as Meir-Keeler fixed point principle for self single-valued operators.

Theorem 1.3. (Meir-Keeler [5]) *Let (X, d) be a complete metric space and f an operator from X into itself. If f satisfies the Meir-Keeler type condition (3) then f has a unique fixed point, i.e. $F_f = \{x^*\}$. Moreover, for any $x \in X$, we have $\lim_{n\to\infty} f^n(x) = x^*$.*

For the multivalued case, a similar result was proved by S. Reich, as follows.

Theorem 1.4. (Reich [9]) *Let (X, d) be a complete metric space and $F : X \to P_{cp}(X)$ be a multivalued operator. If F satisfies the Meir-Keeler type condition (7), then F has at least one fixed point.*

For the case of a multivalued contractive operator Smithson proved:

Theorem 1.5. (Smithson [14]) *Let (X, d) be a compact metric space and $F : X \to P_{cl}(X)$ be a multivalued contractive operator. Then F has at least one fixed point.*

The purpose of this work is to consider a generalized Meir-Keeler type multivalued operator and to discuss some connections with the classical one. Some fixed point results are also given. Two open problems are pointed out. Our results are in connections with some theorems given in S. Reich [9], R. P. Agarwal, D. O'Regan, N.

Shahzad [1], S. Leader [4], S. Park, W. K. Kim [7], T. Cardinali, P. Rubbioni [2], I. A. Rus [10], [11], etc.

2. **Main Results**

Let (X,d) be a metric space and $F : X \to P_{cl}(X)$ be a multivalued operator. For $x, y \in X$, let us denote

$$M(x,y) := \max\{d(x,y), D(x,F(x)), D(y,F(y)), \frac{1}{2}[D(x,F(y)) + D(y,F(x))]\}.$$

Consider the following two Meir-Keeler type conditions on F:

(9) for each $\eta > 0$ there exists $\delta > 0$ such that $x, y \in X$, $\eta \leq M(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$;

(10) for each $\eta > 0$ there exists $\delta > 0$ such that $x, y \in X$, $M(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$.

Our first remark is that (9) $\Leftrightarrow$ (10).

This follows from the following two lemmas.

Lemma 2.1. *If (X,d) is a metric space and $F : X \to P_{cl}(X)$ satisfies* (9)*, then $H(F(x),F(y)) \leq M(x,y)$, for each $x, y \in X$.*

Proof. We discuss two cases:

1) $M(x,y) = 0$; Then $x = y$ and we are done.

2) $M(x,y) > 0$; Let $\eta > 0$ and $\delta > 0$ such that (9) holds. Suppose, by contradiction, that $H(F(x),F(y)) > M(x,y)$. Then $H(F(x),F(y)) > M(x,y) \geq \eta$, a contradiction with (9). □

Lemma 2.2. *Let (X,d) be a metric space and $F : X \to P_{cl}(X)$. Then* (9) $\Leftrightarrow$ (10)*.*

Proof. (10) $\Rightarrow$ (9) is obviously. For the reverse implication, let us consider $\eta > 0$ and $x, y \in X$ such that (9) holds. We have the following two situations:

1) $M(x,y) < \eta$; Then from Lemma 2.1 we get that $H(F(x),F(y)) < \eta$.

1) $M(x,y) \geq \eta$; Then from (9) we have $H(F(x),F(y)) < \eta$. □.

Also, we have:

Lemma 2.3. *If (X,d) is a metric space and $F: X \to P_{cl}(X)$ satisfies* (10)*, then $H(F(x),F(y)) < M(x,y)$, for each $x,y \in X$, with $x \neq y$.*

Proof. If there exist $x \neq y \in X$ such that $H(F(x),F(y)) \geq M(x,y)$, then we contradict (9). □

Lemma 2.4. *Let (X,d) be a metric space and $F: X \to P_{cl}(X)$ be a multivalued operator such that* (9) *(or equivalently* (10)*) holds. Then:*

(11) *for each $\eta > 0$ there exists $\delta > 0$ such that $(x,y) \in GrafF$, $\eta \leq d(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$.*

Proof. For $\eta > 0$ let $\delta > 0$ be such that (9) holds. Let $y \in F(x)$ be arbitrary. Then $M(x,y) = \max\{d(x,y), D(x,F(x)), D(y,F(y)), \frac{1}{2}D(x,F(y))\}$.

Since $D(x,F(x)) \leq d(x,y)$ and $D(x,F(y)) \leq d(x,y) + D(y,F(y))$ it follows that $M(x,y) = \max\{d(x,y), D(y,F(y))\}$.

If $M(x,y) = D(y,F(y))$, then from (9) we have the following contradiction: $\eta \leq D(y,F(y)) \leq H(F(x),F(y)) < \eta$. So $M(x,y) = d(x,y)$. □

In a similar way as above, we have:

Lemma 2.5. *Let (X,d) be a metric space and $F: X \to P_{cl}(X)$ be a multivalued operator. Consider the following condition:*

(12) *for each $\eta > 0$ there exists $\delta > 0$ such that $(x,y) \in GrafF$, $d(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$.*

Then (11) $\Leftrightarrow$ (12).

The following result is an easy consequence of the above lemmas.

Lemma 2.6. *If (X,d) is a metric space and $F: X \to P_{cl}(X)$ satisfies*

(9') for each $\eta > 0$ there exists $\delta > 0$ such that $(x,y) \in GrafF$, $\eta \leq M(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$

or

(10') for each $\eta > 0$ there exists $\delta > 0$ such that $(x,y) \in GrafF$, $M(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$,

then $H(F(x),F(y)) < d(x,y)$, for each $(x,y) \in GrafF$, with $x \neq y$.

Open Problem A. *Establish fixed point results for (generalized) Meir-Keeler multivalued operators on graphic, i. e. satisfying the condition* $(9')$ *or* $(10')$.

For example, from the above results and Theorem 1.5. we immediately obtain:

Theorem 2.7. *Let* (X,d) *be a compact metric space and* $F: X \to P_{cp}(X)$ *be a multivalued operator, such that it satisfies the contractive condition* (6) *for each* $(x,y) \in (X \times X) \setminus GrafF$. *Suppose that* F *satisfies the following generalized Meir-Keeler type condition:*

(13) *for each* $\eta > 0$ *there exists* $\delta > 0$ *such that* $(x,y) \in GrafF$, $M(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$,

then F *has at least one fixed point.*

Proof. The assumption (13) implies (12) which is equivalent to (11). From the above remark, F satisfies the contractive condition (6) for each $(x,y) \in GrafF$. Hence, F is contractive on X. The rest of the proof follows from Theorem 1.5. □

Open Problem B. *Establish results of the above type for the case of a locally (generalized) Meir-Keeler multivalued operator, see also the conditions* (4) *and* (8).

The following theorem is a slight modification of a result established by R. P. Agarwal, D. O'Regan, N. Shahzad in [1].

Theorem 2.8. *Let* (X,d) *be a complete metric space,* $x_0 \in X$, $r > 0$ *and* $f: \widetilde{B}(x_0;r) \to X$ *an operator. Suppose that:*

i) for each $\eta > 0$ *there exists* $\delta > 0$ *such that* $x,y \in \widetilde{B}(x_0;r)$, $\eta \leq d(x,y) < \eta + \delta \Rightarrow d(f(x),f(y)) < \eta$.

ii) $d(x_0, f^n(x_0)) < r$, *for each* $n \in \mathbb{N}^*$.

Then $Fixf = \{x^*\}$.

An extension to the multivalued case is the following:

Theorem 2.9. *Let* (X,d) *be a complete metric space,* $x_0 \in X$, $r > 0$ *and* $F: \widetilde{B}(x_0;r) \to P_{cp}(X)$ *be a multivalued operator. Suppose that:*

i) for each $\eta > 0$ *there exists* $\delta > 0$ *such that* $x,y \in \widetilde{B}(x_0;r)$, $\eta \leq d(x,y) < \eta + \delta \Rightarrow H(F(x),F(y)) < \eta$.

ii) $\delta(x_0, F^n(x_0)) < r$, *for each* $n \in \mathbb{N}^*$.

Then $FixF \neq \emptyset$.

Sketch of the proof. Let us consider the operator $G : \widetilde{B}(\{x_0\}; r) \subset (P_{cp}(X), H) \to (P_{cp}(X), H)$, given by $G(Y) := \bigcup_{x \in Y} F(x)$. Then G satisfies all the hypothesis of Theorem 2.8. Hence, there exists $Y^* \in P_{cp}(X)$ such that $Y^* = G(Y^*)$. Define $h : Y^* \to \mathbb{R}_+$, by $h(a) := D(a, F(a))$. Since F is contractive on $\widetilde{B}(x_0; r)$ it follows that F is upper semicontinuous on $\widetilde{B}(x_0; r)$. Thus h is lower semicontinuous on Y^*. Since Y^* is compact, there exists $b \in Y^*$ and $c \in F(b)$ such that $\inf_{a \in Y^*} h(a) = d(b, c)$. If we suppose that $d(b, c) > 0$ then we get a contradiction: $h(c) = D(c, F(c)) \leq H(F(b), F(c)) < d(b, c)$. Hence $b = c$ and so the conclusion follows. □

References

[1] Agarwal, R. P., O'Regan, D., Shahzad, N., *Fixed point theory for generalized contractive maps of Meir-Keeler type*, Math. Nachr., **276**(2004), 3-22.

[2] Cardinali, T., Rubbioni, P., *An extension to multifunctions of the Keeler-Meir's fixed point theorem*, Fixed Point Theory, **7**(2006), 23-36.

[3] Covitz, H., Nadler Jr., S.B., *Multi-valued contraction mapping in generalized metric spaces*, Israel J. Math. **8**(1970), 5-11.

[4] Leader, S., *Equivalent Cuachy sequences and contractive fixed points in metric spaces*, Studia Math., **76**(1983), 63-67.

[5] Meir, A., Keeler, E., *A theorem on contraction mappings*, J. Math. Anal. Appl., **28**(1969), 326-329.

[6] Nadler Jr., S.B., *Multivalued contraction mappings*, Pacific J. Math., **30**(1969), 475-488.

[7] Park, S., Kim, W. K., *Extensions of the weak contractions of Dugundji and Granas*, J. Korean Math. Soc., **21**(1984), 1-7.

[8] Petruşel, A., *Generalized multivalued contractions*, Nonlinear Analysis, **47**(2001), 649-659.

[9] Reich, S., *Fixed points of contractive functions*, Boll. U.M.I. (4), **5**(1972), 26-42.

[10] Rus, I.A., *Fixed point theorems for multivalued mappings in complete metric spaces*, Math. Japonica, **20**(1975), 21-24.

[11] Rus, I.A., *Generalized contractions and applications*, Transilvania Press Cluj-Napoca, 2001.

[12] Rus, I.A., Petruşel, A., Petruşel, G., *Fixed point theory 1950-2000 : Romanian contributions*, House of the Book of Science, Cluj-Napoca, 2002.

[13] Rus, I.A., Petruşel, A., Sîntămărian, A., *Data dependence of the fixed point set of some multivalued weakly Picard operators*, Nonlinear Analysis, **52**(2003), 1947-1959.

[14] Smithson, R.E., *Fixed points for contractive multifunctions*, Proc. A.M.S., **27**(1971), 192-194.

Department of Applied Mathematics,
Babeş-Bolyai University Cluj-Napoca,
Kogălniceanu 1, 400084, Cluj-Napoca, Romania
E-mail address: petrusel@math.ubbcluj.ro, gabip@math.ubbcluj.ro

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

FIXED POINT STRUCTURES WITH THE COMMON FIXED POINT PROPERTY: MULTIVALUED OPERATORS

IOAN A. RUS

Dedicated to Professor Gheorghe Coman at his 70^{th} anniversary

Abstract. The concept of fixed point structure with the common fixed point property is extended to multivalued operators. In the terms of this concept some common fixed point theorems are given.

1. Introduction

In this paper we follow the notations and terminologies in I.A. Rus [10] and [12].

Let X be a nonempty set and $T, Q : X \to P(X)$ two multivalued operators.

In the present paper we shall consider the following problems:

Problem A. In which conditions we have that:

$$F_T \neq \emptyset, \quad F_Q \neq \emptyset, \quad T \circ Q = Q \circ T \ \Rightarrow \ F_T \cap F_Q \neq \emptyset?$$

Problem B. In which conditions we have that:

$$(SF)_T \neq \emptyset, \quad (SF)_Q \neq \emptyset, \quad T \circ Q = Q \circ T \ \Rightarrow \ (SF)_T \cap (SF)_Q \neq \emptyset?$$

The aim of this paper is to study these problems in terms of the fixed point structures ([10]).

We recall that if $T : X \to P(X)$ is a multivalued operator then we shall denote:

$$F_T := \{x \in X \mid x \in T(x)\};$$

Received by the editors: 01.07.2006.

2000 *Mathematics Subject Classification.* 47H10.

Key words and phrases. fixed point structure, common fixed point property, multivalued operator, commuting pair, (θ, φ)-contraction pair, θ-condensing pair, open problem.

$$(SF)_T := \{x \in X \mid T(x) = \{x\}\};$$

$$I(T) := \{A \subset X \mid T(A) \subset A\}.$$

2. Fixed point structures with the common fixed point property

Definition 2.1. A fixed point structure $(X, S(X), M^0)$ on a set X (see [10]) is with the common fixed point property iff:

$$Y \in S(X), \quad T, Q \in M^0(Y), \quad T \circ Q = Q \circ T \ \Rightarrow \ F_T \cap F_Q \neq \emptyset.$$

Definition 2.2. A strict fixed point structure $(X, S(X), M^0)$ on a set X (see [10]) is with the common strict fixed point property iff:

$$Y \in S(X), \quad T, Q \in M^0(Y), \quad T \circ Q = Q \circ T \ \Rightarrow \ (SF)_T \cap (SF)_Q \neq \emptyset.$$

Remark 2.1. For the case of singlevalued operators see I.A. Rus [11].

Remark 2.2. For the common fixed point theorems in terms of the fixed point structures see A. Muntean [8] and A. Sîntămărian [14].

Remark 2.3. For the common fixed point theorems for the generalized commuting operators (weakly commuting, R-weakly commuting, compatible, δ-compatible,...) see G.F. Jungck [5], O. Hadzic [3], O. Hadzic and Lj. Gajic [4], B.E. Rhoades [9], A. Ahmad and M. Imdad [1], M.A. Ahmed [2], T. Kamran [6], H. Kaneko [7],...

Example 2.1. The trivial fixed point structure is a fixed point structure with the common fixed point property.

Example 2.2. Let (X, d) be a complete metric space, $S(X) := P_{cl}(X)$ and $M^0(Y) := \{T : Y \to P_{cl}(Y) \mid T$ is a multivalued contraction with $(SF)_T \neq \emptyset\}$. The triple $(X, P_{cl}(X), M^0)$ is a strict fixed point structure with the common strict fixed point property.

Indeed, from the Theorem 3.2 in [12] it follows that $(X, P_{cl}(X), M^0)$ is a strict fixed point structure. Let $Y \in P_{cl}(X)$, $T, Q \in M^0(Y)$ such that $T \circ Q = Q \circ T$. We have $F_T = (SF)_T = \{x^*\}$ and $F_Q = (SF)_Q = \{y^*\}$. From $T \circ Q = Q \circ T$ it follows that $x^* = y^*$.

Remark 2.4. For other examples see I.A. Rus [11], A. Muntean [8] and A. Sîntămărian [14].

Remark 2.5. To give examples of fixed point structures with the common fixed point property is one of the basic open problem of the common fixed point theory.

3. (θ, φ)-contraction pairs

Let X be a nonempty set, $Y \subset X$, $Z \subset P(X)$ and $\theta : Z \to \mathbb{R}_+$.

Definition 3.1. A pair of operators $T, Q : Y \to P(Y)$ is a (θ, φ)-contraction pair iff:

(i) $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is a comparison function;

(ii) $A \in P(Y) \cap Z$ implies that $T(A) \cup Q(A) \in Z$;

(iii) $\theta(T(A) \cup Q(A)) \leq \varphi(\theta(A)),\ \forall\ A \in I(T) \cap I(Q) \cap Z$.

We have the following general common fixed point principles.

Theorem 3.1. *Let $(X, S(X), M^0)$ be a fixed point structure with the common fixed point property and (θ, η) a compatible pair with this fixed point structure. Let $Y \in \eta(Z)$ and $T, Q \in M^0(Y)$. We suppose that:*

(i) $\theta|_{\eta(Z)}$ has the intersection property;

(ii) $T \circ Q = Q \circ T$;

(iii) the pair (T, Q) is a (θ, φ)-contraction pair.

Then, $F_T \cap F_Q \neq \emptyset$.

Proof. Let $Y_1 := \eta(T(Y) \cup Q(Y)), \ldots, Y_{n+1} = \eta(T(Y_n) \cup Q(Y_n)),\ n \in \mathbb{N}$. First of all we remark that $Y_n \in I(T) \cap I(Q),\ \forall\ n \in \mathbb{N}$. From the conditions (ii) and (iii) we have that

$$\theta(Y_{n+1}) = \theta(\eta(T(Y_n) \cup Q(Y_n))) = \theta(T(Y_n) \cup Q(Y_n))$$

$$\leq \varphi(\theta(Y_n)) \leq \cdots \leq \varphi^{n+1}(\theta(Y)) \to 0 \text{ as } n \to \infty.$$

From the condition (i) it follows that

$$Y_\infty := \bigcap_{n \in \mathbb{N}} Y_n \neq \emptyset \quad \text{and} \quad \theta(Y_\infty) = 0.$$

Now, we remark that $\eta(Y_\infty) = Y_\infty$ and $Y_\infty \in I(T) \cap I(Q)$. From the definition of the fixed point structure it follows that $Y_\infty \in S(X)$ and from Definition 2.1 the operators $T|_{Y_\infty}$ and $Q|_{Y_\infty}$ have a common fixed point. So, $F_T \cap F_Q \neq \emptyset$.

In a similar way we have

Theorem 3.2. *Let $(X, S(X), M^0)$ be a strict fixed point structure with the common fixed point property and (θ, η) a compatible pair with $(X, S(X), M^0)$. Let $Y \in \eta(Z)$ and $T, Q \in M^0(Y)$. We suppose that:*

(i) $\theta|_{\eta(Z)}$ has the intersection property;

(ii) $T \circ Q = Q \circ T$;

(iii) the pair (T, Q) is a (θ, φ)-contraction pair.

Then, $(SF)_T \cap (SF)_Q \neq \emptyset$.

4. θ-condensing pairs

Let X be a nonempty set, $Y \subset X$, $\theta : Z \to \mathbb{R}_+$ and $Z \subset P(X)$.

Definition 4.1. A pair $T, Q : Y \to P(Y)$ is a θ-condensing pair iff:

(i) $A_i \in Z$, $i \in I$, $\bigcap_{i \in I} A_i \neq \emptyset \ \Rightarrow \ \bigcap_{i \in I} A_i \in Z$;

(ii) $A \in P(Y) \cap Z \ \Rightarrow \ T(A) \cup Q(A) \in Z$;

(iii) $\theta(T(A) \cup Q(A)) < \theta(A)$, for all $A \in I(T) \cap I(Q) \cap Z$ such that $\theta(A) \neq \emptyset$.

We have

Theorem 4.1. *Let $(X, S(X), M^0)$ be a f.p.s. with the common fixed point property and (θ, η) a compatible pair with this fixed point structure. Let $Y \in \eta(Z)$ and $T, Q \in M^0(Y)$.*

We suppose that:

(i) $x \in Y$, $A \in Z$ imply $A \cup \{x\} \in Z$ and $\theta(A \cup \{x\}) = \theta(A)$;

(ii) $T \circ Q = Q \circ T$;

(iii) the pair (T, Q) is θ-condensing pair.

Then, $F_T \cap F_Q \neq \emptyset$.

Proof. Let $x_0 \in Y$. By Lemma 2.3 in [14] there exists $A_0 \subset Y$ such that $x_0 \in A_0$, $A_0 \in F_\eta \cap I(T) \cap I(Q)$ and $\eta(T(A_0) \cup Q(A_0) \cup \{x_0\}) = A_0$. From the

condition (iii) it follows that $\theta(A_0) = 0$. But $\eta(A_0) = A_0$ and $\theta(A_0) = 0$ imply that $A_0 \in S(X)$. From the Definition 2.1 the operators $T|_{A_0}$ and $Q|_{A_0}$ have a common fixed point. So, $F_T \cap F_Q \neq \emptyset$.

In a similar way we have

Theorem 4.2. *Let $(X, S(X), M^0)$ be a strict fixed point structure with the common strict fixed point property and (θ, η) a compatible pair with this fixed point structure. Let $Y \in \eta(Z)$ and $T, Q \in M^0(Y)$. We suppose that:*

(i) $x \in Y$, $A \in Z$ imply $A \cup \{x\} \in Z$ and $\theta(A \cup \{x\}) = \theta(A)$;

(ii) $T \circ Q = Q \circ T$;

(iii) the pair (T, Q) is θ-condensing pair.

Then, $(SF)_T \cap (SF)_Q \neq \emptyset$.

References

[1] Ahmad, A., and Imdad, M., *Some common fixed point theorems for mappings and multivalued mappings*, J. Math. Anal. Appl., 218(1998), 546-560.

[2] Ahmed, M.A., *Common fixed point theorems for weakly compatible mappings*, Rocky Mountain J. Math., 33(2003), No.4, 1189-1203.

[3] Hadzic, O., *On coincidence points in convex metric spaces*, Zb. Rad. Univ. Novom Sadu, 19(1989), 233-240.

[4] Hadzic, O. and Gajic, Lj., *Coincidence points for set-valued mappings in convex metric spaces*, Rev. Res., 16(1986), 13-25.

[5] Jungck, G.F., *Common fixed point theorems for compatible self-maps of Hausdorff topological spaces*, Fixed Point Theory and Appl., 2005, No.3, 355-363.

[6] Kamran, T., *Fixed points of asymptotically regular noncompatible maps*, Demonstr. Math., 38(2005), No.2, 485-494.

[7] Kaneko, H., *A common fixed point of weakly commuting multivalued mappings*, Math. Jap., 33(1988), No.5, 741-744.

[8] Muntean, A., *Fixed Point Principles and Applications to Mathematical Economics*, Cluj University Press, Cluj-Napoca, 2002.

[9] Rhoades, B.E., *Common fixed points of compatible set-valued mappings*, Publ. Math. Debrecen, 48(1996), no.3-4, 237-240.

[10] Rus, I.A., *Technique of the fixed point structure for multivalued mappings*, Math. Japonica, 38(1993), 289-296.

[11] Rus, I.A., *Fixed point structures with the common fixed point property*, Mathematica, 38(1996), 181-187.

[12] Rus, I.A., *Strict fixed point theory*, Fixed Point Theory, 4(2003), No.2, 177-183.

[13] Singh, S.L., and Mishra, S.N., *Coincidence points, hybrid fixed and stationary points of orbitally weakly dissipative maps*, Math. Japonica, 39(1994), No.3, 451-459.

[14] Sîtămărian, A., *Common fixed point structures for multivalued operators*, Scientiae Mathematicae Japonicae, 63(2006), No.1, 37-46.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania

STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 4, December 2006

THE STUDY OF AN ADAPTIVE ALGORITHM FOR SOME CUBATURE FORMULAS ON TRIANGLE

ILDIKÓ SOMOGYI AND RADU TRÎMBIŢAŞ

Dedicated to Professor Gheorghe Coman at his 70^{th} *anniversary*

Abstract. We study two nonproduct quadrature formulas of algebraic degree 2 and 3, respectively. The second is then turned into an adaptive quadrature algorithm. A MATLAB implementation and some examples are given.

1. The formulas

The purpose of this paper is to give some practical cubature formulas when the integration domain is a triangle and also to study an adaptive algorithm for these cubature formulas in approximation of the integral

$$I = \int_{T_h} f(x,y)\, dxdy, \tag{1.1}$$

where T_h is a triangular domain, $T_h = \{(x,y)/x \geq 0, y \geq 0, x+y \leq h\}$, and $f : T_h \to \mathbb{R}$ is an integrable function on T_h. We shall consider two cubature formulas from [8]. One of them is a cubature formula which satisfy the minimal condition of Stroud regarding the minimal number of knots of a cubature formula [9]. The degree of exactness of this formula is equal to 2. The other one is a cubature formula which has more knots, but a greater degree of exactness. We consider the following practical cubature formula:

$$\int_{T_h}\int f(x,y)dxdy = \frac{h^2}{6}\left[f(0,\frac{h}{2}) + f(\frac{h}{2},0) + f(\frac{h}{2},\frac{h}{2})\right] + R_2(f). \tag{1.2}$$

Received by the editors: 08.06.2006.

2000 *Mathematics Subject Classification.* 65D30, 65D32.

Key words and phrases. numerical integration, adaptive cubature, MATLAB.

The degree of exactness of this formula is 2, therefore we can use the Peano theorem for the representation of the error, an we can give the following delimitation of the approximation error:

Theorem 1. *If* $f^{(3,0)}(\cdot,0) \in C[0,h], f^{(2,1)}(\cdot,0) \in C[0,h], f^{(0,3)}(0,\cdot) \in C[0,h]$ *and* $f^{(1,2)}(s,t) \in C(T_h)$ *than we have*

$$|R_2(f)| \leq M_{30}f\frac{h^5}{720} + M_{21}f\frac{h^5}{364} + M_{03}f\frac{h^5}{720} + M_{12}f\frac{h^5}{24} \tag{1.3}$$

where

$$M_{30}f = \max_{s\in[0,h]} \left|f^{(3,0)}(s,0)\right|, M_{21}f = \max_{s\in[0,h]} \left|f^{(2,1)}(s,0)\right|,$$
$$M_{03}f = \max_{t\in[0,h]} \left|f^{(0,3)}(0,t)\right|, M_{12}f = \max_{T_h} \left|f^{(1,2)}(s,t)\right|.$$

Remark 1. *The cubature formula (2) has an optimal character, because it satisfies the condition established by Stroud in* [9] *regarding the minimal number of knots for a cubature formula. If the degree of exactness of a cubature formula is equal to* 2*, then the minimal number of knots is* $N = n + 1$*, where* n *is the dimension number. The cubature formula (1.2) with the degree of exactness* 2 *and three knots, has a minimal number of knots.*

Let us now consider a cubature formula with a higher degree of exactness:

$$I = \int_{T_h}\int f(x,y)dxdy = \frac{h^2}{120}\left[3f(0,0) + 3f(h,0) + 3f(0,h) + 8f(\frac{h}{2},0) + \right.$$
$$\left. + 8f(\frac{h}{2},\frac{h}{2}) + 8f(0,\frac{h}{2}) + 27f(\frac{h}{3},\frac{h}{3})\right] + R_3(f). \tag{1.4}$$

Because the degree of exactness of this formula is equal to 3, we can give the following theorem for the delimitation of the absolute error:

Theorem 2. *If* $f^{(4,0)}(s,0) \in C[0,h]$, $f^{(3,1)}(s,0) \in C[0,h]$, $f^{(0,4)}(0,t) \in C[0,h]$, $f^{(1,3)}(0,t) \in C[0,h]$, *and* $f^{(2,2)}(s,t) \in C(T_h)$ *then*

$$|R_3(f)| \leq M_{40}f\frac{h^6}{8640} + M_{31}f\frac{7h^6}{1440} + M_{13}f\frac{7h^6}{1440} + M_{04}f\frac{h^6}{8640} + M_{22}f\frac{h^6}{768},$$

where

$$M_{40}f = \max_{s\in[0,h]} \left|f^{(4,0)}(s,0)\right|, M_{31}f = \max_{s\in[0,h]} \left|f^{(3,1)}(s,0)\right|,$$
$$M_{13}f = \max_{t\in[0,h]} \left|f^{(1,3)}(0,t)\right|, M_{04}f = \max_{t\in[0,h]} \left|f^{(0,4)}(0,t)\right|,$$
$$M_{22}f = \max_{(s,t)\in T_h} \left|f^{(2,2)}(s,t)\right|.$$

We shall use an affine transformation to transform these cubature formulas from the standard triangle T_h to an arbitrary triangle Δ with the vertices $V_i(x_i, y_i), i = 1, 2, 3$.

Let $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ denote the affine transformation from T_h to Δ,

$$\varphi(\overline{x}, \overline{y}) = A(\overline{x}, \overline{y}) + b \tag{1.5}$$

where

$$A = \begin{pmatrix} \dfrac{x_2 - x_1}{h} & \dfrac{x_3 - x_1}{h} \\ \dfrac{y_2 - y_1}{h} & \dfrac{y_3 - y_1}{h} \end{pmatrix} \tag{1.6}$$

and

$$b = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{1.7}$$

Let J be the Jacobian matrix of the transformation, in this case J is independent of $(\overline{x}, \overline{y})$, $\det J(\overline{x}, \overline{y}) = \det A$, and the transformation rule is

$$\int_{T_h} f(x,y)dxdy = |\det A| \int_{\Delta} f(\varphi(\overline{x}, \overline{y}))d\overline{x}d\overline{y}.$$

2. **Implementation**

For a detailed description of an adaptive numerical integration algorithm see [10, 7].

In this section we focus our attention to an adaptive algorithm, based on formula (1.4). For implementation details on an adaptive algorithm based on formula (1.2), see [3].

Now, using the transform given by (1.5), (1.6), and (1.7), we rewrite (1.4) in the form

$$I \approx \frac{\text{area}(\Delta)}{60}\left(3\sum_{i=1}^{3} f(V_i) + 8\sum_{i=1}^{3} f(P_i) + 27f(G)\right), \tag{2.1}$$

where

$$P_i = \frac{1}{2}\left(P_i + P_j\right), \qquad \{i,j,k\} = \{1,2,3\},$$

are the midpoints of the edges of Δ, and $G = \frac{1}{3}(V_1 + V_2 + V_3)$ is the barycenter of Δ (see Figure 1).

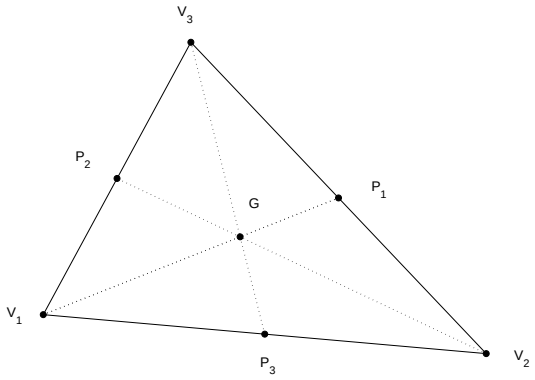


FIGURE 1. The elementary third degree formula

The intial triangle, Δ, is decomposed into four triangles, Δ_1, Δ_2, Δ_3, and Δ_4, determined by verices and the middle points (see Figure 2).

In the first step we apply the formula given by (2.1) to Δ. Then we apply the same formula to each of the triangles Δ_i, $i = 1, \ldots, 4$. Let I_1 be the value provided by (2.1), and I_2 the value obtained by summing the four valued obtained applying (2.1) to each of the four triangle of the subdivision. A possible stopping criterion is

$$|I_1 - I_2| < \varepsilon,$$

where ε is the desired tolerance. If the criterion is not fulfilled, then we apply the same procedure recursively to each triangle of the subdivision. A detailed description is given in Algorithm 1.

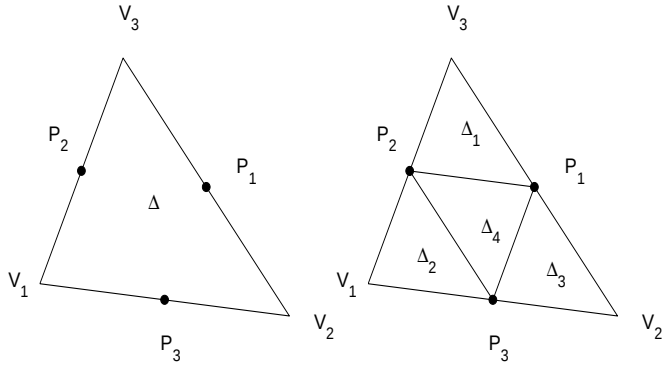


FIGURE 2. The initial triangle and the subdivision

Algorithm 1 An adaptive cubature algorithm on triangle; call *result* := adapt(f, Δ, ε), where f is the integrand, Δ is the triangle, and ε is the desired tolerance; elem_formula implemntes the elementary cubature given by (2.1).

Let Δ_1, Δ_2, Δ_3, Δ_4 be the triangles determined by vertices and midpoints;
$I1$:= elem_formula(f, Δ);
$I2$:= elem_formula(f, Δ_1) + elem_formula(f, Δ_2)+
elem_formula(f, Δ_3) + elem_formula(f, Δ_4);
if $|I_1 - I_2| < \varepsilon$ **then**
result := I_2;
else
result := adapt$(f, \Delta_1, \varepsilon)$ + adapt$(f, \Delta_2, \varepsilon)$+
adapt$(f, \Delta_3, \varepsilon)$ + adapt$(f, \Delta_4, \varepsilon)$;
end if

The papers [6, 5] give useful guidelines for implementation of adaptive cubatures on triangle. We have implemented this algorithm in MATLAB[1]. The implementation follows the description given by Algorithm 1. The optional input parameter `trace`, when it is nonzero, allows us to represent graphically the process of computing. The optional output parameter `stat` gives us the number of function evaluation and

[1] MATLAB© is a trademark of the MathWorks Inc., Natick, MA 01760-2098

the number of triangles. Some optimizations which save several function evaluations are possible. Since the value I_1 is the value of the integral on Δ, we compute it once and provide it further as an input parameter. We do the same thing with the values of function at midpoints and barycenter.

In the sequel we give the MATLAB code.

```
function [vi,stat]=mpcubatd3mb(f,V,err,trace)
%MPCUBATD3MB - cubature with midpoints and barycenter,
% exact for P_3^2
%call [vi,stat]=mpcubatd3mb(f,V,err,trace,...)
%f function
%V - coordinates of vertices
%err - error
%trace - tracing indicator

global FEN TRIN sfl
if nargin <4, trace=0;
else
    if trace
        clf
    end
end
if nargin < 3, err=1e-3; end
sfl=[nargout==2];
if sfl
    FEN=0; TRIN=0;
end
P=midpoints(V); G=sum(V,2)/3;
fv=feval(f,V); fp=feval(f,P); fg=feval(f,G);
area = 1/2*abs(det([V',ones(3,1)]));
```

```
I1=elform(area,fv,fp,fg);
if trace, tracefun([V,P,G]); hold on; end
vi=quadrg3(f,V,P,G,fv,fp,fg,err,area,I1,trace);
if sfl
    FEN = FEN+7;
    stat=struct('nev',FEN,'ntri',TRIN);
end

function vi=quadrg3(f,V,P,G,fv,fp,fg,err,area,I1,trace)
%QUADRG3 - cubature with midpoints and barycenter, internal use
%call vi=quadrg3(f,V,P,G,fv,fp,fg,err,area,I1,trace)
%f - function
%V - coordinates of vertices
%P - midpoints coordinates
%G - barycenter coordinates
%fv - values of f at verices
%fp - values of f at midpoints
%fg - values at barycenter
%err - error
%area - area of triangle
%I1 - the first estimation (elementary formula)
%trace - tracing indicator

global FEN TRIN sfl
area=area/4;
V1=[V(:,1),P(:,[2,3])]; fv1=[fv(1),fp([2,3])];
P1=midpoints(V1); fp1=feval(f,P1);
G1=sum(V1,2)/3; fg1=feval(f,G1);
I11=elform(area,fv1,fp1,fg1);
V2=[V(:,2),P(:,[1,3])]; fv2=[fv(2),fp([1,3])];
```

```
P2=midpoints(V2); fp2=feval(f,P2);
G2=sum(V2,2)/3;   fg2=feval(f,G2);
I12=elform(area,fv2,fp2,fg2);
V3=[V(:,3),P(:,[1,2])]; fv3=[fv(3),fp([1,2])];
P3=midpoints(V3); fp3=feval(f,P3);
G3=sum(V3,2)/3; fg3=feval(f,G3);
I13=elform(area,fv3,fp3,fg3);
V4=P; fv4=fp; P4=[P1(:,1),P2(:,1),P3(:,1)];
fp4=[fp1(1),fp2(1),fp3(1)]; G4=G; fg4=fg;
I14=elform(area,fv4,fp4,fg4);
I2=I11+I12+I13+I14;
if sfl
    FEN=FEN+12;
    TRIN=TRIN+4;
end
if trace, tracefun([P1,P2,P3,G1,G2,G3]); end
if abs(I2-I1)<err
    vi=I2;
else
    vi=quadrg3(f,V1,P1,G1,fv1,fp1,fg1,err,area,I11,trace)+...
        quadrg3(f,V2,P2,G2,fv2,fp2,fg2,err,area,I12,trace)+...
        quadrg3(f,V3,P3,G3,fv3,fp3,fg3,err,area,I13,trace)+...
        quadrg3(f,V4,P4,G4,fv4,fp4,fg4,err,area,I14,trace);
end %if

function v=elform(area,fv,fp,fg)
v=area/60*(3*sum(fv)+8*sum(fp)+27*fg);

function P=midpoints(V);
P(:,1)=(V(:,2)+V(:,3))/2; P(:,2)=(V(:,1)+V(:,3))/2;
```

```
P(:,3)=(V(:,1)+V(:,2))/2;


function tracefun(L)
%TRACEFUN - represents points where function is evaluated
plot(L(1,:),L(2,:),'.k','Markersize',4);
```

3. Numerical examples

Consider the triangle $T_1 = \{(x,y)/x \geq 0, y \geq 0, x+y \leq 1\}$, and the function $f : T_1 \to \mathbb{R},\ f(x,y) = \text{humps}(x)\text{humps}(y)$, where

$$\text{humps}(x) = \frac{1}{(x-0.3)^2+0.01} + \frac{1}{(x-0.9)^2+0.04} - 6.$$

The graph of f is given in Figure 3 as surface and as contour. First, we approximate

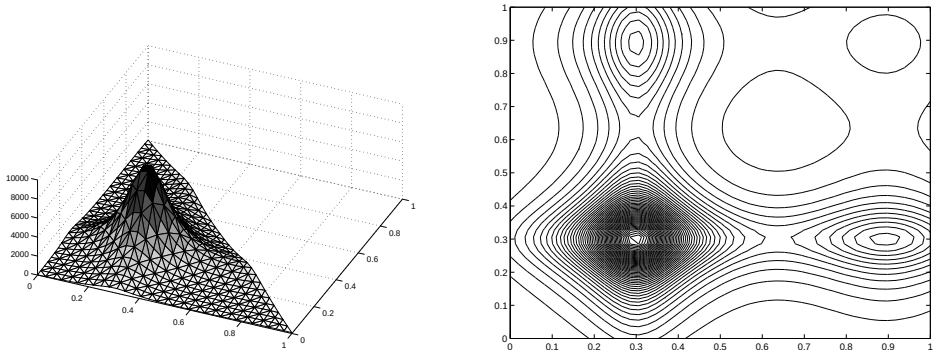


FIGURE 3. The graph of f, as surface (left) and as contour

the integral for a tolerance $\varepsilon = 10^{-5}$ using the adaptive quadrature based on (1.2) and (2.1), respectively. The `trace` flag is set.

```
>> [vib,statb]=mpcubatd2mb(@humps2dv,V,1e-5,1)
vib =
    5.997039610414015e+002
statb =
     nev: 57684
    ntri: 25636
```

```
>> [vib3,statb3]=mpcubatd3mb(@humps2dv,V,1e-5,1)
vib3 =
    5.997039668483903e+002
statb3 =
     nev: 30499
    ntri: 10164
```

The figure 4 shows the points where the MATLAB functions evaluate the integrands. Now, for a higher accuracy (10^{-9}) and a timer included we got the following results

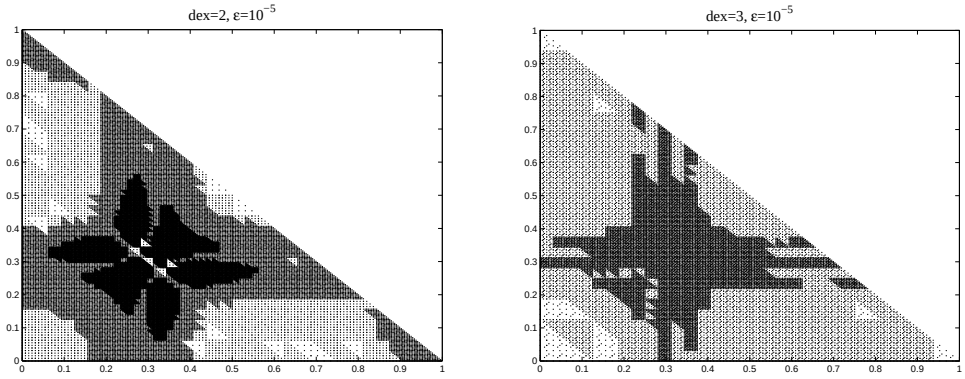


FIGURE 4. Evaluation points for the second order formula (left) and for the third order formula, for function f, $\varepsilon = 10^{-5}$

```
>> tic,[vib,statb]=mpcubatd2mb(@humps2dv,V,1e-9); toc
Elapsed time is 43.444590 seconds.
>> tic,[vib3,statb3]=mpcubatd3mb(@humps2dv,V,1e-9);toc
Elapsed time is 19.121086 seconds.
>> vib,statb
vib =
    5.997039625857019e+002
statb =
     nev: 2219736
    ntri: 986548
```

```
>> vib3,statb3
vib3 =
    5.997039625817022e+002
statb3 =
     nev: 640915
    ntri: 213636
```

Thus, for this function the third degree formula is faster.

For functions that do not exhibit high oscillations and for modest accuracy requirements, the second degree formula requires fewer function evaluation. Consider the function $f(x,y) = y \sin x$ (implemented by MATLAB function `fintegrv`) to be integrated on T_1. For $\varepsilon = 10^{-4}$, one obtains:

```
>> tic,[vib,statb]=mpcubatd2mb(@fintegrv,V,1e-4); toc
Elapsed time is 0.001425 seconds.
>> tic,[vib3,statb3]=mpcubatd3mb(@fintegrv,V,1e-4); toc
Elapsed time is 0.012022 seconds.
>> vib,statb
vib =
   0.04030110314738
statb =
     nev: 48
    ntri: 20
>> vib3,statb3
vib3 =
   0.04030317282902
statb3 =
     nev: 67
    ntri: 20
```

References

[1] Barnhill, R.E., Gordon, W.J., and Thomas, D.H., *The method of successive decomposition for multiple integration*, Research Rep. GMR-1281, General Motors, Warren, Mich., 1972.

[2] Coman, Gh., *Analiza numerică*, Libris, Cluj, 1995.

[3] Coman, Gh., Pop, I., Trîmbiţaş, R., *An adaptive cubature on triangle*, Studia UBB, Mathematica, vol. XLVII, No.4, pp. 27-36, 2002.

[4] Coman, Gh., Stancu, D.D., Blaga, P., *Analiză numerică şi teoria aproximării*, vol II, Presa Universitară Clujeană, Cluj-Napoca, 2002.

[5] Cools, R., Laurie, D.P., Pluym, L., *Cubpack++: A C++ package for Authomatic Two-dimensional Cubature*, ACM TOMS, vol. 23, No. 1, 1997, pp. 1-15.

[6] Laurie, D.P., *Algorithm 584: CUBTRI: Automatic Cubature over a triangle*, ACM TOMS, vol. 8, No. 2, 1982, 210-218.

[7] Rice, J., *A metaalgorithm for adaptive quadrature*, JACM, vol. 22, No. 1, 1975, 61-82.

[8] Somogyi, I., *Practical cubature formulas in triangles with error bounds*, Seminar on Numerical and Statistical Calculus, 2004, 131-137.

[9] Stroud, A.H., *Approximate Calculation of Multiple Integrals*, Englewood Cliffs, N.J. Prentice-Hall, Inc. 1971.

[10] Überhuber, Cr., *Computer Numerik*, Band II, Springer, 1995.

[11] Welfert, B., *Numerical Analysis*, Lecture Notes, University of Arizona.

Babeş-Bolyai University,
Faculty of Mathematics and Computer Science,
Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania
E-mail address: tradu@math.ubbcluj.ro

În cel de al LI-lea an (2006) *STUDIA UNIVERSITATIS BABEŞ-BOLYAI* apare în următoarele serii:

matematică (trimestrial)
informatică (semestrial)
fizică (trimestrial)
chimie (semestrial)
geologie (trimestrial)
geografie (semestrial)
biologie (semestrial)
filosofie (semestrial)
sociologie (semestrial)
politică (anual)
efemeride (semestrial)
studii europene (trei apariții pe an)
dramatica (semestrial)
business (semestrial)
psihologie-pedagogie (anual)
ştiințe economice (semestrial)
ştiințe juridice (trimestrial)
istorie (trei apariții pe an)
filologie (trimestrial)
teologie ortodoxă (semestrial)
teologie catolică (trei apariții pe an)
teologie greco-catolică - Oradea (semestrial)
teologie catolică - Latina (anual)
teologie reformată (semestrial)
educație fizică (semestrial)

In the LI-th year of its publication (2006) *STUDIA UNIVERSITATIS BABEŞ-BOLYAI* is issued in the following series:

mathematics (quarterly)
computer science (semesterily)
physics (quarterly)
chemistry (semesterily)
geology (quarterly)
geography (semesterily)
biology (semesterily)
philosophy (semesterily)
sociology (semesterily)
politics (yearly)
ephemerides (semesterily)
European studies (three issues / year)
business (semesterily)
dramatica (semestrial)
psychology - pedagogy (yearly)
economic sciences (semesterily)
juridical sciences (quarterly)
history (three issues / year)
philology (quarterly)
orthodox theology (semesterily)
catholic theology (three issues / year)
greek-catholic theology - Varadiensis (semesterily)
catholic theology - Latina (yearly)
reformed theology (semesterily)
physical training (semesterily)

Dans sa LI-ème année (2006) *STUDIA UNIVERSITATIS BABEŞ-BOLYAI* paraît dans les séries suivantes:

mathématiques (trimestriellement)
informatiques (semestriellement)
physique (trimestriellement)
chimie (semestriellement)
géologie (trimestriellement)
géographie (semestriellement)
biologie (semestriellement)
philosophie (semestriellement)
sociologie (semestriellement)
politique (annuellement)
éphémérides (semestriellement)
études européennes (trois apparitions / année)
dramatica (semestrial)
affaires (semestriellement)
psychologie - pédagogie (annuellement)
études économiques (semestriellement)
études juridiques (trimestriellement)
histoire (trois apparitions / année)
philologie (trimestriellement)
théologie orthodoxe (semestriellement)
théologie catholique (trois apparitions / année)
théologie greco-catholique - Varadiensis (semestriellement)
théologie catholique - Latina (annuellement)
théologie réformée - (semestriellement)
éducation physique (semestriellement)