STUDIA
UNIVERSITATIS BABEȘ-BOLYAI

MATHEMATICA

1

1995

CLUJ-NAPO[redacted]

Anul XL **1995**

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI

MATHEMATICA

1

Redacţia: 3400 CLUJ-NAPOCA str. M. Kogălniceanu nr.1 ▸ Telefon : 116101

SUMAR-CONTENTS-SOMMAIRE

Gh. BANTAŞ, M. TURINICI, Hyperbolic Mean Value Theorems of Non-Differential Form ♦ Teoreme de medie de tip hiperbolic în formă nediferenţială 3
B.G. PACHPATTE, On a Certain Inequality Used in the Theory of Difference Equations ♦ Asupra unei inegalităţi folosite în teoria ecuaţiilor cu diferenţe 23
I. RAŞA, Probabilistic Positive Linear Operators ♦ Operatori liniari pozitivi probabilistici 33
A. CIUPA, A Class of Integral Favard-Szasz Type Operators ♦ O clasă de operatori integrali de tip Favard-Szasz 39
V. POPA, H.K. PATHAK, Common Fixed Points of Compatible Mappings of Type (A) ♦ Puncte fixe comune pentru funcţii compatibile de tipul (A) 49
B. FINTA, Note About a Method for Solving Nonlinear System of Equations in Finit Dimensional Spaces ♦ O metodă de rezolvare a sistemelor de ecuaţii neliniare în spaţii finit dimensionale 59
A. REVNIC, A Numerical Solution of the Differential Equation of *m*-th Order Using Spline Functions ♦ O soluţie numerică pentru ecuaţia diferenţială de ordinul *m* folosind funcţii spline 65
C. POPA, Preconditioning for the Fulfilment of the Approximation Assumption in the Algebraic Multigrid Method ♦ Precondiţionarea pentru îndeplinirea condiţiilor de aproximare în metoda multigrid algebrică 77

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

HYPERBOLIC MEAN VALUE THEOREMS OF NON-DIFFERENTIAL FORM

Gh. BANTAŞ* and M. TURINICI*

Received: June 20, 1994
AMS subject classification: 26B05

REZUMAT. - Teoreme de medie de tip hiperbolic în formă nediferențială. În lucrare sunt stabilite mai **multe teoreme** de medie pentru funcții definite pe un dreptunghi.

1. Introduction. Let I and J be nonempty intervals of the real axis (R). Denote fr simplicity $K = I \times J$. By a (standard) *rectangle* of K we mean any subset $\Delta = [a,b] \times [c,d]$ of K, where $[a,b]$, $[c,d]$ are closed sub-intervals of I and J, respectively. In this case, the points

$$A = (a,c),\ B = (b,c),\ C = (b,d),\ D = (a,d)$$

are called the *vertices* of Δ; correspondingly, the rectangle in question may be represented as [ABCD].

Let $(X,\|\ \|)$ be a normed space and $f: K \to X$, a mapping. For each rectangle Δ of K taken as above, denote

$$m_f(\Delta) = f(A) - f(B) + f(C) - f(D). \tag{1.1}$$

This will be referred to as the *hyperbolic* (Lebesque-Stieltjes) *measure* of Δ generated by this function. Note that, when $X = R$, and

$$f(t,s) = ts,\ t,s \in R$$

then, this hyperbolic measure reduces to

* *"Al. I. Cuza" University, Department of Mathematics, 6600 Iaşi, Romania*

$$m(\Delta) = ac - bc + bd - ad = (b-a)(d-c) \tag{1.2}$$

(the *usual* Lebesgue *measure* of Δ). Finally, denote

$$R_f(\Delta) = \frac{m_f(\Delta)}{m(\Delta)},\ S_f(\Delta) = \|R_f(\Delta)\|. \tag{1.3}$$

These will be referred to as the *variational quotients* of f with respect to Δ.

Now, by a mean value theorem/property for f over Δ we mean an evaluation of $R_f(\Delta)$ of $S_f(\Delta)$ with the aid of some expressions depending on the objective to be attained. More precisely, we may distinguish between

i) mean value theorems of *non-differential* (relative) *form;*

ii) mean value theorems of *differential form.*

The second class of such properties was investigated-in the bi-dimensional setting we dealt with - by Nicolescu [12, ch.19, §2], under the lines in Bögel [6,7]; see also Dobrescu [8]. The first class of such results was only tangentially discussed until now in the paper by Nicolescu [10]. It is our main aim in the present exposition to fill this gap, in a manner suggested by the one-dimensional developments in this area due to the authors [4,5]; see also Aziz and Diaz [1,2,3]. The imposed assumptions upon f are intended to be the largest possible ones; details will be given in Section 3. All preliminary facts were collected in Section 2. And, in Section 4, some aspects involving the real case ($X = R$) will be considered. Finally, it is worth noting these developments are an essential tool to get mean value theorems under differential form. A detailed account of these will be made in a future paper.

2. Preliminaries. Let again I,J be real intervals and $K = I \times J$. We also give a normed space $(X,\|\ \|)$ and take a mapping $f: K \to X$. It is our aim in the following to investigate this function by means of the associated map $\Delta \mapsto m_f(\Delta)$.

We start with an invariance property. By a *hyperbolic constant* over K we mean any map $h : K \to X$ of the form

$$h(t,s) = \varphi(t) + \psi(s),\ (t,s) \in K$$

where $\varphi : I \to X$, $\psi : J \to X$ are given functions. This term is justified by the statement below. (The proof being evident, we do not give details.)

PROPOSITION 1. *For each rectangle* Δ *of* K *and each hyperbolic constant* h *over* K, *one has*

$$m_{f+h}(\Delta) = m_f(\Delta). \tag{2.1}$$

As an immediate consequence,

$$R_{f+h}(\Delta) = R_f(\Delta) \text{ (hence } S_{f+h}(\Delta) = S_f(\Delta)).$$

In other words, any property of $R_f(\Delta)$ (or $S_f(\Delta)$) may be also tranferred to the function $f + h$ which, in principle, is no longer endowed with the properties of f. Some concrete examples in this direction will be given in Sections 3 and 4.

We are now passing to an additivity property. For any rectangle $\Delta = [a,b] \times [c,d]$ of K, denote

$$\text{int}(\Delta) =]a,b[\times]c,d[\text{ (the } interior \text{ of } \Delta)$$

This is of course related to the topological structure of the plane given, e.g., by the maximum norm. By a *division* of the rectangle Δ we mean any finite decomposition $\Delta = \bigcup_r \Delta_r$ of Δ into (standard) rectangles of K with the family $\{\text{int}(\Delta_r)\}$ being mutually disjoint. Among these, we distinguish the divisions of Δ generated by corresponding divisions of the real intervals generating Δ. Precisely, given finite decompositions

$$[a,b] = \bigcup_i [t_i, t_{i+1}],\ [c,d] = \bigcup_j [s_j, s_{j+1}]$$

of these intervals, the considered division may be written as $\Delta = \bigcup_{i,j} \Delta_{ij}$, where

$\Delta_{ij} = [t_i, t_{i+1}] \times [s_j, s_{j+1}]$, for all possible (i,j). These will be referred to in the sequel as *normal divisions* of the underlying rectangle.

PROPOSITION 2. *Let* $\{\Delta_r\}$ *be a division of the rectangle* Δ. *Then*

$$m_f(\Delta) = \sum_r m_f(\Delta_r). \tag{2.2}$$

Proof. Take any vertex, P, of an arbitrary rectangle in this decomposition, distinct from the vertices of Δ. A simple analysis shows that P belongs to either two or four rectangles in this family. (The proof being almost evident, we do not give details.) Let $\leq$ be the ordering in R^2 introduced in the usual way

$$(t_1, s_1) \leq (t_2, s_2) \text{ iff } t_1 \leq t_2,\ s_1 \leq s_2 .$$

In the first case, the point in question is extremal in one rectangle and non-extremal in another. In the second case, the considered point is two times extremal and two times non-extremal in the rectangles to which it belongs. Consequently, the contribution of $f(P)$ in $\sum_r m_f(\Delta_r)$ is zero, by the definition of these expressions. In other words, only the vertices of Δ are to be retained in this sum, and conclusion follows. ■

Remark. A different proof of this may be given along the following lines (cf. Tolstov [15, ch.2, §6]). Let $\mathcal{V}$ be the set of all vertices for the rectangles in $\{\Delta_r\}$. The projection of $\mathcal{V}$ over $[a,b]$, respectively $[c,d]$ gives finite decompositions of such intervals. Let

$$\Delta = \cup\{\Delta_{ij}; (i,j) \in \Gamma\}$$

be the normal division of Δ induced by these. It clearly follows by the described construction that a partition $\Gamma = \bigcup_r \Gamma_r$ of the index set Γ may be found so that, for each r,

$$\{\Delta_{ij}; (i,j) \in \Gamma_r\} \text{ is a normal division of } \Delta_r.$$

This, plus (2.2) being valid for normal divisions imply

$$m_f(\Delta) = \sum \{m_f(\Delta_{ij}); (i,j) \in \Gamma\} = \sum_r \sum \{m_f(\Delta_{ij}); (i,j) \in \Gamma_r\} = \sum_r m_f(\Delta_r)$$

and the assertion is proved.

Remark. Of course, the conclusion of this statement remains valid (via Proposition 1) in case f is to be replaced by $f+h$, where h is any hyperbolic constant (over K).

Now, a useful semi-continuity result will be proved. For any pair of points P,Q in the plane, we denote by $PQ = \{\lambda P + (1-\lambda)Q;\ 0 \le \lambda \le 1\}$ the *segment* between these points and by $(PQ) = \{\lambda P + (1-\lambda)Q;\ \lambda \in R\}$ the *line* passing through P and Q. Let $\Delta = [\mathrm{ABCD}]$ be a rectangle in K, given by its vertices. Denote

$$fr(\Delta) = AB \cup BC \cup CD \cup DA \text{ (the \textit{boundary} of } \Delta).$$

Let P be a point of $fr(\Delta)$, distinct from the vertices of Δ. There exists a unique line passing through P, which is orthogonal to the segment of $fr(\Delta)$, which contains P. This will be referred to as the *normal* to Δ at the considered point, and denoted $\nu_\Delta(P)$. (That P must be distinct from the vertices of Δ in this construction is a consequence of the fact that, otherwise, the normal in question would be not uniquely determined.) Now, call the underlying function $f: K \to X$, *normally continuous* at the point $P \in fr(\Delta)$ (distinct from A,B,C,D) when its restriction to $\nu_\Delta(P) \cap \Delta$ is continuous at P. We also term f, normally continuous on $fr(\Delta)$ when it is normally continuous at any point $P \in fr(\Delta)$ (distinct from the vertices of Δ).

With these conventions, let Δ be a rectangle in K. We also take a sub-rectangle Δ' of Δ in such a way that $fr(\Delta')$ has at least a segment in common with $fr(\Delta)$.

PROPOSITION 3. *Suppose that*

(H.1) *f is continuous at the vertices of Δ*

(H.2) *f is normally continuous at each vertex of Δ' (if any) lying in $fr(\Delta)$, distinct from the vertices of Δ.*

Then, for each $\eta > 0$, there exists a sub-rectangle Δ'' of Δ', interior to Δ, with

$$S_f(\Delta'') \geq (1-\eta) S_f(\Delta'). \tag{2.3}$$

Proof. Without loss, one may assume $m_f(\Delta) \neq 0$ (hence $S_f(\Delta) \neq 0$). We have several situations to discuss.

Case 1. $fr(\Delta')$ has a single segment in common with $fr(\Delta)$ This, e.g., corresponds to the choice $\Delta' = [A'B'C'D']$ where $A'B' \subset AB$ and $C', D' \in \operatorname{int}(\Delta)$; or, in other words (by the adopted notations for the rectangle Δ)

$$A' = (a', c),\ B' = (b', c),\ C' = (b', r),\ D' = (a', r)$$

with $a < a' < b' < b,\ c < r < d$. We now consider the sub-rectangle Δ_λ of K given by the vertices $A'_\lambda, B'_\lambda, C', D'$, where

$$A'_\lambda = (a', c+\lambda),\ B' = (b', c+\lambda),\ \lambda > 0 \text{ small enough.}$$

Clearly, Δ'_λ is in $\Delta' \cap \operatorname{int}(\Delta)$ for all such λ. Moreover, by (H.2),

$$f(A'_\lambda) \to f(A'),\ f(B'_\lambda) \to f(B) \text{ as } \lambda \to 0_+.$$

This, combined with

$$m(\Delta'_\lambda) \to m(\Delta') \text{ as } \lambda \to 0_+ \tag{2.4}$$

shows

$$R_f(\Delta'_\lambda) \to R_f(\Delta') \text{ (hence } S_f(\Delta'_\lambda) \to S_f(\Delta')) \text{ as } \lambda \to 0_+. \tag{2.5}$$

As a consequence, any Δ'_λ, where $\lambda > 0$ is sufficiently small, may be taken as the sub-rectangle Δ'' in the statement.

Case 2. $fr(\Delta')$ has two segments in common with $fr(\Delta)$. This, for example, may be understood as the rectangle in question being represented in the form $\Delta' = [AB'C'D']$, where

$$B' = (p, c),\ C' = (p, q),\ D' = (a, q),\ a < p < b,\ c < q < d.$$

Let us now construct a sub-rectangle Δ'_λ of Δ by the vertices $A_\lambda, B'_\lambda, C', D'_\lambda$, where

$$A_\lambda = (a+\lambda, c+\lambda),\ B'_\lambda = (p, c+\lambda),\ D'_\lambda = (a+\lambda, q).$$

(As before, $\lambda > 0$ is small enough). That $\Delta' \cap \operatorname{int}(\Delta)$ includes Δ'_λ is clear.

We also have, by (H.1) + (H.2),

$$f(A_\lambda) \to f(A),\ f(B'_\lambda) \to f(B'),\ f(D'_\lambda) \to f(D') \text{ as } \lambda \to 0_+.$$

This, in combination with (2.4) being valid in this context gives again (2.5). Hence, any Δ'_λ like before - where $\lambda > 0$ is sufficiently small - is a candidate for sub-rectangle Δ'' in the statement.

Cases 3-4. $fr(\Delta')$ has more than two segments in common with $fr(\Delta)$. (That is, either, $fr(\Delta')$ has three segments in common with $fr(\Delta)$ or else $\Delta' = \Delta$). The argument we just developed may be correspondingly modified to get a family of sub-rectangles $\{\Delta'_\lambda\}$ of Δ', interior to Δ, which in addition has the property (2.5). So, as before, it will suffice taking one of these as Δ'', to get (2.3). Having explored all possible situations, the conclusion follows. ∎

Remark. The working conditions (H.1) + (H.2) must be taken in a relative sense only. Because as results from Proposition 1, the statement above remains valid whenever f-h fulfils (H.1) + (H.2) for some hyperbolic constant h: $K \to X$ (which, in particular, may be discontinuous at any point of the rectangle Δ).

As an immediate consequence of this, we have

COROLLARY 1. *Suppose that the underlying function f satisfies* (H.1) *plus*

(H.3) *f is normally continuous over* $fr(\Delta)$.

Then, conclusion of Proposition 2 *is retainable.*

In particular, a sufficient condition for (H.1) + (H.3) is

(H.3)′ *f is continuous over* $fr(\Delta)$.

Of course, as already precised, these conditions may be put in an even more general framework, via Proposition 1; further details are not given.

Finally, a specific continuity property will be introduced for such functions. Given a pair $P_1 = (t_1, s_1)$, $P_2 = (t_2, s_2)$ of points (in K), denote by $[P_1;P_2]$ the rectangle $[a,b] \times [c,d]$, where

$$a = \min(t_1, t_2),\ b = \max(t_1, t_2);\ c = \min(s_1, s_2),\ d = \max(s_1, s_2).$$

Of course, the order of these points is not essential, here; i.e., $[P_1;P_2]$ is identical to $[P_2;P_1]$. Let P be an interior point of K. We say the function $f: K \to X$ is *hyperbolic continuous* at P whenever

$$m_f([P;Q]) \to 0 \text{ as } Q \to P;$$

or, in other words, for each $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that

$$\|m_f([P;Q])\| < \varepsilon \text{ provided } \|P - Q\| < \delta(\varepsilon).$$

Likewise, the considered function is called hyperbolic continuous over a subset of K when it is hyperbolic continuous at each point of that subset.

The relationships between this notion and the standard continuity one are precised in

PROPOSITION 4. *The following are valid*:

A) *If the function $f: K \to X$ is continuous at the point $P \in \mathrm{int}(K)$ then it is hyperbolic continuous at this point.*

B) *Suppose the function $f: K \to X$ is hyperbolic continuous at $P \in \mathrm{int}(K)$. Then, a continuous at P function $g = g_P: K \to X$ and a hyperbolic constant $h = h_P: K \to X$ may be found so that f be represented as the sum $g+h$.*

Proof. The first part is evident. For the second one note that the hyperbolic continuity of f at $P = (t_0, s_0)$ may be also written as

$$f(t_0,s_0) - f(t_0,s) - f(t,s_0) + f(t,s) \to 0, \text{ as } t \to t_0,\ s \to s_0.$$

Denote in this case

$$g(t,s) = f(t,s) - f(t_0,s) - f(t,s_0),\quad (t,s) \in K.$$

$$h(t,s) = f(t_0,s) + f(t,s_0),\quad (t,s) \in K.$$

That g,h satisfy the above requirements is clear. Hence the conclusion. ■

Remark. This result does not admit, in general, a global counterpart. In other words, if f: $K \to X$ is hyperbolic continuous over a part H of K then, a representation like $f = g+h$ where g: $K \to X$ is continuous over H and h: $K \to X$ is a hyperbolic constant (over K) is not obtainable, in general. For an example in this direction we refer to Nicolescu [12, ch.19, §2].

3. Main results (inequality form). Let the notations above be maintained. Letting I,J be real intervals, for each rectangle $\Delta = [a,b] \times [c,d]$ in $K = I \times J$, denote

diam(Δ) = max $(b-a,\ d-c)$ (the *diameter* of Δ).

This notion is related to the normed structure of the plane (given by the maximum norm). Let also $(x, \| \ \|)$, a normed space and $f : K \to X$, a mapping. As a consequence of the developments above, the first main result of the present paper is

THEOREM 1. *Let* Δ *be a (standard) rectangle in K. Then, for each* $\varepsilon > 0$, *there is a sub-rectangle* Δ_ε *of* Δ *with*

$$\operatorname{diam}(\Delta_\varepsilon) < \varepsilon,\quad S_f(\Delta) \le S_f(\Delta_\varepsilon). \tag{3.1}$$

Proof. Construct a (normal) division of Δ by

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b,\ c = s_0 < s_1 < \dots < s_{m-1} < s_m = d$$

with

$$\max(t_{i+1} - t_i, s_{j+1} - s_j) < \varepsilon,\ 0 \le i \le n-1,\ 0 \le j \le m-1.$$

Here, $n,m \geq 3$ are positive integers. Precisely, if we put

$$\Delta_{ij} = [t_i, t_{i+1}] \times [s_j, s_{j+1}],\ 0 \leq i \leq n-1,\ 0 \leq j \leq m-1,$$

the normal division in question is $\{\Delta_{ij}\}$. In addition, we have the supplementary property

$$\operatorname{diam}(\Delta_{ij}) < \varepsilon, \text{ for all possible } (i,j).$$

It is clear, via Proposition 2, that

$$R_f(\Delta) = \sum_{i,j} \lambda_{ij} R_f(\Delta_{ij})$$

where, by convention,

$$\lambda_{ij} = \frac{m(\Delta_{ij})}{m(\Delta)},\ \ 0 \leq i \leq n-1,\ 0 \leq j \leq m-1.$$

Therefore, by the triangle inequality,

$$S_f(\Delta) \leq \sum_{i,j} \lambda_{ij} S_f(\Delta_{ij}).$$

The second of this relation is a convex combination of $\{S_f(\Delta_{ij})\}$. Hence The conclusion. ■

Now, by simply adding to this the remark in Section 2 concerning the alternative proof of Proposition 2, one gets

COROLLARY 2. *Let* Δ *be a rectangle in* K *and* $\{\Delta_r\}$ *be a division of* Δ. *Then, for each* $\varepsilon > 0$, *there exists an index* $r = r(\varepsilon)$ *and a sub-rectangle* Δ_ε *of* Δ_r *so that* (3.1) *be valid.*

Note at this moment that no property is required for the function f to get the conclusion in the statement. Nevertheless, the obtained assertion is not very sharp because the possibility that $fr(\Delta_\varepsilon)$ should have a nonempty intersection with $fr(\Delta)$ cannot be avoided in general. It is natural to ask of whether is this removable. The answer is affirmative (via Proposition 3). To state it, we need a new convention. Let Δ be a rectangle in K. Take eight points systems $\{E_1, ..., E_8\}$ on the boundary of Δ, distincts from the vertices of Δ, according to the condition:

there exists a sub-rectangle Δ', interior to Δ such that $\{E_1, ..., E_8\}$ appears as the projection over $fr(\Delta)$ of the vertices of Δ'. (3.2)

Such systems will be termed *admissible* in what follows. Now, let again $(X, \| \ \|)$ be a normed space and $f: K \to X$ a mapping. As a completion of Theorem 1, the second main result of this paper is

THEOREM 2. *Suppose that f satisfies* (H.1) *plus*

(H.4) *f is normally continuous over at least one admissible eight points system* $fr(\Delta)$.

Then, for each $\varepsilon > 0$, *there is a sub-rectangle* Δ_ε *interior to* Δ, *with the property* (3.1).

Proof. Let the ambient rectangle Δ be represented as [ABCD]. Take also an admissible eight points system $\{E_1, ..., E_8\}$ in $fr(\Delta)$ (given by (H.4)). So, there exists a sub-rectangle Δ_0 = [MNPQ] interior to Δ, such that $\{E_1, ..., E_8\}$ appears as the projection of $\mathcal{V} = \{M,N,P,Q\}$ over $fr(\Delta)$. This, e.g., may be understood as

$$MQ \cap (AB \cup CD) = \{E_1, E_5\};\ NP \cap (AB \cup CD) = \{E_2, E_6\}$$

$$MN \cap (AD \cup BC) = \{E_3, E_7\};\ PQ \cap (AD \cup BC) = \{E_4, E_8\}.$$

Now, the admissible system $\{E_1, ..., E_8\}$ generates a normal division $\{\Delta_0, ..., \Delta_8\}$ of Δ. (Here, Δ_0 is the above sub-rectangle and, e.g., $\Delta_1 = [AE_1ME_7]$, $\Delta_2 = [E_1E_2NM]$, etc.) This gives at once

$$R_f(\Delta) = \sum_{i=0}^{8} \mu_i R_f(\Delta_i)$$

where, by convention,

$$\mu_i = \frac{m(\Delta_i)}{m(\Delta)},\ 0 \le i \le 8.$$

So, by the triangle inequality,

$$S_f(\Delta) \le \sum_{i=0}^{8} \mu_i S_f(\Delta_i). \tag{3.3}$$

As an immediate consequence of this,

$$S_f(\Delta) \le \max_{0\le i\le 8} S_f(\Delta_i). \tag{3.4}$$

We have two cases to discuss.

a) Relation (3.4) is holding with equality. Then, again combining with (3.3),

$$S_f(\Delta) = \sum_{i=0}^{8} \mu_i S_f(\Delta_i),$$

wherefrom

$$\sum_{i=0}^{8} \mu_i (S_f(\Delta) - S_f(\Delta_i)) = 0.$$

But, $\mu_0, \ldots, \mu_8$ are strictly positive. Therefore

$$S_f(\Delta) = S_f(\Delta_i),\ 0 \le i \le 8;$$

and from this, conclusion is clear.

b) Relation (3.4) is holding strictly (with $<$ in place of $\le$). If one happens that $S_f(\Delta) < S_f(\Delta_0)$, then we are done (by applying Theorem 1 to the same function f and the rectangle Δ_0). Otherwise,

$$S_f(\Delta) < S_f(\Delta_i),\ \text{for some } i \in \{1, \ldots, 8\}.$$

By (H.4) plus Proposition 3, we must have that for each $\eta > 0$ (small enough) there exists a sub-rectangle $\Delta_i^{(\eta)}$ of Δ_i, interior to Δ, with

$$S_f(\Delta_i^{(\eta)}) \ge (1 - \eta) S_f(\Delta_i).$$

Choose $\eta > 0$ in such a way that $(1 - \eta) S_f(\Delta_i) \ge S_f(\Delta)$. (This is possible, by the strict inequality above.) Combining these, yields

$$S_f(\Delta) \le S_f(\Delta_i^{(\eta)});$$

and this, again with Theorem 1 gives conclusion in the statement.

Now, a) + b) are the only possible situations in this discussion. Hence the result. ■

As a direct consequence of this, we have

COROLLARY 3. *Let the rectangle Δ in K and the function $f: K \to X$ be such that* *conditions* (H.1) + (H.3) *are accepted. Then, conclusion of Theorem* 2 *is retainable.*

In particular, a sufficient condition for (H.1) + (H.3) is $(H.3)'$. A natural question appearing in this context is that of determining to what extent are these statements valid when $(H.3)'$ is to be substituted by its weaker counterpart

(H.3)* *f is hyperbolic continuous over* $fr(\Delta)$.

To give a partial answer, we note that, by Proposition 4, one has at each point P in $fr(\Delta)$, the representation $f = g_P + h_P$ where $g_P: K \to X$ is continuous at P and $h_P: K \to X$ is a hyperbolic constant. Hence the functions in this representation are depending on the points in $fr(\Delta)$. But, if this dependence would be removed (i.e., the underlying functions remain unchanged when P describes $fr(\Delta)$) it follows by Proposition 1 that, in fact, $(H.3)'$ is necessarily fulfilled under (H.3)*; and so, conclusion of Theorem 2 is retainable, in view of Corollary 3. Summing up, hyperbolic continuity conditions (over $fr(\Delta)$ or, even, the all of Δ) imposed upon f are - generally - insufficient for the truth of such results. This, in particular, applied to the statement of Lemma 1 in Nicolescu [10], shows we must delete the word "hyperbolic" (as a weaker form of continuity for f) to retain its conclusion. But then, the result in question reduced to Corollary 3 above.

Remark. From a methodological viewpoint, the developments above may be viewed as a bi-dimensional counterpart of the contributions in this area due to Bantaş and Turinici [4]; see also Aziz and Diaz [1].

Now, it would be of interest to determine of whether or not is (H.4) removable; or, in other words, to what extent can we diminish the cardinality of an admissible system (of points in $fr(\Delta)$). The answer is affirmative: it is based on a few remarks about the associated sub-rectangles in the division of Δ. Let $\{E_1, \ldots, E_8\}$ be an admissible eight points system in $fr(\Delta)$

generated by a sub-rectangle $\Delta' = [MNPQ]$ of Δ (and interior to Δ). We associate to each vertex of Δ' its closest projections over $fr(\Delta)$. This generates a decomposition of our system into four groups of such projections. For example, under the notations encountered in the proof of Theorem 2, these groups may be depicted as

$$U_1 = \{E_1, E_7\},\ U_2 = \{E_2, E_3\},\ U_3 = \{E_4, E_6\},\ U_4 = \{E_5, E_8\}.$$

Now, let us call a four points system $\{G_1, G_2, G_3, G_4\}$ in $\{E_1, \ldots, E_8\}$, *admissible* provided

$$G_i \in U_i,\ 1 \le i \le 4.$$

There are $2^4 = 16$ such admissible four points systems generated by an admissible eight points system. However, for symmetry reasons only 4 systems from these are essential. For example, taking AB as a basis, the systems in question are

$$\{E_7, E_2, E_4, E_5\},\ \{E_7, E_3, E_6, E_5\},\ \{E_7, E_3, E_4, E_6\},\ \{E_7, E_3, E_4, E_8\}.$$

Now, given any admissible four points system $G = \{G_1, G_2, G_3, G_4\}$, there exists a division

$$\Delta = \Delta_0 \cup \Delta_1 \cup \Delta_2 \cup \Delta_3 \cup \Delta_4$$

of the rectangle Δ, where Δ_0 is the above one and (for $1 \le i \le 4$) the vertices of the sub-rectangle Δ_i lyung in $fr(\Delta)$ and distinct from those of Δ are necessarily in G. (For example, to verify this for $G = \{E_7, E_2, E_4, E_5\}$, it will suffice putting

$$\Delta_1 = [AE_2NE_7],\ \Delta_2 = [E_2BE_4P],\ \Delta_3 = [E_4CE_5Q],\ \Delta_4 = [E_7ME_5D];$$

the remaining situations are treatable in a similar way.) As a direct consequence, the argument used in Theorem 2 is also applicable to this larger setting. We thus proved

COROLLARY 4. *Suppose that f satisfies conditions* (H.1) *plus*

(H.4)′ *f is normally continuous over at least one admissible four points system of* $fr(\Delta)$.

Then, for each $\varepsilon > 0$, *there is a sub-rectangle* Δ_ε *interior to* Δ, *with the property* (3.1).

Concerning the further reduction of this number, call the two points system $\{E_1,E_2\}$ of $fr(\Delta)$ (distinct from the vertices of Δ), *admissible*, when E_1,E_2 are an opposite segments of $fr(\Delta)$ and the normals to E_1 and E_2 are identical (e.g., $E_1 \in AB$, $E_2 \in CD$ and E_1,E_2 is parallel to AD or BC). Suppose now (H.4)' is to be replaced by

(H.4)" *f is normally continuous over an admissible two points system of* $fr(\Delta)$.

Let Δ_1 and Δ_2 be the division of Δ generated by E_1E_2 (in the usual way) and assume

(H.5) $S_f(\Delta) < S_f(\Delta_i)$, *for some* $i \in \{1,2\}$.

By Proposition 3, there must be a sub-rectangle Δ_i' of Δ_i, interior to Δ, with $S_f(\Delta) < S_f(\Delta_i')$; this, plus Theorem 1 give us immediately conclusion of Theorem 2. Therefore, condition (H.4) - or its variants - has a relative character (from a cardinality viewpoint). This forces us to ask of whether or not is this condition effective in such statements. We conjecture that the answer is negative.

4. The real case. In the following, the choice $X = R$ will be considered, from an equality perspective. Precisely, let I, J be real intervals and put $K = I \times J$. Let $f: K \to R$ be a function and Δ, be a (standard) rectangle in K. As a counterpart of Theorem 2, the third main result of the paper is

THEOREM 3. *Suppose that*

(H.6) *f is continuous over* Δ.

Then, for each $\varepsilon > 0$, *there is a sub-rectangle* Δ_ε *interior to* Δ, *with the properties*

$$\operatorname{diam}(\Delta_\varepsilon) < \varepsilon,\ R_f(\Delta) = R_f(\Delta_\varepsilon). \tag{4.1}$$

Proof. Let us construct an equi-distant division of Δ by

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b,\ \ \rho = t_{i+1} - t_i < \varepsilon,\ 0 \le i \le n-1$$

$$c = s_0 < s_1 < \dots < s_{m-1} < s_m = d, \quad \sigma = s_{j+1} - s_j < \varepsilon, \ 0 \le j \le m-1.$$

(Here, $n,m \ge 3$ are fixed positive integers.) Denote for simplicity

$$\Delta(t,s) = [t, t+\rho] \times [s, s+\sigma], \quad a \le t \le t_{n-1}, \ c \le s \le s_{m-1}.$$

Of course, $\Delta(t_i, s_j)$ is, for $0 \le i \le n-1$, $0 \le j \le m-1$, nothing but Δ_{ij} alluded to in Theorem 1. Denote also

$$\phi(t,s) = R_f(\Delta(t,s)), \ a \le t \le t_{n-1}, \ c \le s \le s_{m-1}.$$

It is clear that

$$R_f(\Delta) = \sum_{i,j} \lambda_{ij} \phi(t_i, s_j) \tag{4.2}$$

where (λ_{ij}) are again as in Theorem 1. Two situations are now open before us.

Case 1. The set $\{\phi(t_i, s_j);\ 0 \le i \le n-1,\ 0 \le j \le m-1\}$ consists of exactly one element. As a consequence,

$$R_f(\Delta) = R_f(\Delta(t_1, s_1))$$

and conclusion is clear (because $\Delta(t_1, s_1)$ is interior to Δ and its diameter is inferior to ε).

Case 2. The set $\{\phi(t_i, s_j);\ 0 \le i \le n-1,\ 0 \le j \le m-1\}$ has at least two distinct elements. Hence

$$\min_{i,j} \{\phi(t_i, s_j)\} < \max_{i,j} \{\phi(t_i, s_j)\}. \tag{4.3}$$

On the other hand, by convexity arguments,

$$\min_{i,j} \{\phi(t_i, s_j)\} \le R_f(\Delta) \le \max_{i,j} \{\phi(t_i, s_j)\}. \tag{4.4}$$

Suppose one of these relations holds with equality; e.q., the second. We have, by (4.2)

$$\sum_{i,j} \lambda_{ij} (R_f(\Delta) - \phi(t_i, s_j)) = 0.$$

As $\{\lambda_{ij};\ 0 \le i \le n-1,\ 0 \le j \le m-1\}$ are stricly positive,

$$R_f(\Delta) = \phi(t_i, s_j), \ 0 \le i \le n-1, \ 0 \le j \le m-1,$$

absurd by (4.3). Hence, both inequalities in (4.4) are strict. Suppose

$$\min_{i,j} \{\phi(t_i, s_j)\} = \phi(t_p, s_q), \ \max_{i,j} \{\phi(t_i, s_j)\} = \phi(t_u, s_v)$$

for some $p, u \in \{0, ..., n-1\}$, $q, v \in \{0, ..., m-1\}$. Denote for simplicity $\Delta' = [a, t_{n-1}] \times [c, s_{m-1}]$, and let $x = x(\tau)$, $y = y(\tau)$, $0 \le \tau \le 1$ be a continuous path luing in Δ' with

(i) $(x(\tau), y(\tau)) \in \text{int}(\Delta') \subset \text{int}(\Delta), \quad 0 < \tau < 1$

(ii) $(x(0), y(0)) = (t_p, s_q), \ (x(1), y(1)) = (t_u, s_v).$

The composed function (from [0,1] to R)

$$\psi(\tau) = \phi(x(\tau), y(\tau)), \quad 0 \le \tau \le 1$$

is continuous, by (H.6); and, in view of the assumptions we just made,

$$\psi(0) < R_f(\Delta) < \psi(1).$$

Hence, by the Cauchy intersection theorem, there must be some point τ_0 in]0,1[, with $\psi(\tau_0) = R_f(\Delta)$; or in other words,

$$R_f(\Delta) = R_f(\Delta(x(\tau_0), y(\tau_0))).$$

It is now clear that $\Delta_\varepsilon = \Delta(x(\tau_0), y(\tau_0))$ has all the properties we need. This ends the argument.■

As an immediate application, the following "weak" counterpart of Theorem 2 is available. Let $(X, \| \ \|)$ be a normed space and $f: K \to X$, a mapping. Let also Δ be a rectangle in K.

COROLLARY 5. *Suppose that*

(H.6)* *f is weakly continuous over Δ.*

Then, for each $\varepsilon > 0$, *there is a sub-rectangle* Δ_ε *interior to* Δ, *such that* (3.1) *be fulfilled.*

Proof. By the Hahn-Banach theorem, we may find a linear continuous functional x^* over X, with

$$\|x^*\| = 1, \; x^*(R_f(\Delta)) = S_f(\Delta).$$

The function g: $K \to R$ given by

$$g(t,s) = x^*(f(t,s)), \; (t,s) \in K$$

fulfils, by (H.6)*, conditions of Theorem 3. So, for each $\varepsilon > 0$, there exists a sub-rectangle Δ_ε interior to Δ, with

$$\text{diam}\,(\Delta_\varepsilon) < \varepsilon, \; R_g(\Delta) = R_g(\Delta_\varepsilon).$$

But, evidently,

$$R_g(\Delta) = x^*(R_f(\Delta)) = S_f(\Delta);$$

and, moreover,

$$R_g(\Delta_\varepsilon) = |x^*(R_f(\Delta_\varepsilon))| \leq S_f(\Delta_\varepsilon).$$

Combining these facts yields the desired conclusion.■

Remark. As already precised in Section 2, the continuity condition (H.6) is relative in nature. Because, as results from Proposition 1, conclusion of the above theorem is retainable whenever (H.6) is to be admitted for some function f-h where h: $K \to R$ is a hyperbolic constant (which, in principle may be discontinuous over Δ).

Remark. These results are methodologically comparable with the statements in this direction due to Nicolescu [11]. And from a dimensional viewpoint, these may be deemed as direct extensions of the ones obtained in Bantaş and Turinici [4]; see also Aziz and Diaz [2,3]. The idea of the argument goes back to Bögel [6] and, respectively, Pompeiu [13,14]. Further aspects of the problem may be found in the survey paper by Nashed [9].

REFERENCES

1. Aziz K. and Diaz J.B., *On a mean value theorem of the differential calculus of vector-valued functions, and uniqueness theorems for ordinary differential equations in a linear-normed space*, Contrib. to Diff. Eqs., vol. 1, pp. 251-269, (1963).
2. Aziz K. and Diaz J.B., *On a mean value theorem of the weak differential calculus of vector-valued functions*, ibidem, pp. 271-273, (1963).
3. Aziz K. and Diaz J.B., *On Pompeiu's proof of the mean value theorem of the differential calculus of real-valued functions*, ibidem, pp. 467-481, (1963).
4. Bantaş Gh. and Turinici M., *Remark about mean value theorems*, Bull. Math. Soc. Sci. Math. Roum., to appear.
5. Bantaş Gh. and Turinici M., *Mean-value theorems of inequality form*, submitted.
6. Bögel K, *Mehrdimensionale Differentiation von Funktionen mehrerer reeller Veränderlichen*, J. Reine Angew. Math., vol. 170, pp. 197-217, (1934).
7. Bögel K, *Über mehrdimensionale Differentiation*, Integration und beschränkte Variation, ibidem, vol. 173, pp. 5-30, (1935).
8. Dobrescu E., *Contributions to a bidimensional infinitesimal analysis* (Romanian), Studii Cerc. Mat., vol. 8, pp. 103-130, (1957).
9. Nashed M.Z., *Differentiability and related properties of nonlinear operators. Some aspects of the role of differentials in nonlinear functional analysis*, Nonlinear Functional Analysis and Applications (L. Rall ed.), pp. 103-309, Academic Press, New York, 1971.
10. Nicolescu L.J., *On a mean value theorem* (Romanian), St. Cerc. Mat., vol. 16, pp. 987-995, (164).
11. Nicolescu L.J., *On a weak mean value theorem*, Rev. Roum. Math. Pures Appl., vol. 10, pp. 145-148, (1965).
12. Nicolescu M., *Mathematical Analysis* (Romanian), vol. 2, Editura Tehnică, Bucureşti, 1958.
13. Pompeiu D., *Sur le théorème des accroissements finis* (Première Note), Ann. Sci. Univ. Jassy, vol. 3, pp. 244-250, (1906).
14. Pompeiu D., *Sur le théorème des accroissements finis* (Deuxième Note), ibidem, vol. 4, pp. 1-12, (1906).
15. Tolstov G.P., *Measure and Integral* (Russian), Izd. Nauka, Moscow, 1976.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

ON A CERTAIN INEQUALITY USED IN THE THEORY OF DIFFERENCE EQUATIONS

B. G. PACHPATTE*

Received: April 12, 1994
AMS subject classification: 26D15, 26D20

REZUMAT. - Asupra unei inegalităţi folosite în teoria ecuaţiilor cu diferenţe. Sunt stabilite câteva noi inegalităţi cu diferenţe finite legate de o inegalitate folosită în teoria ecuaţiilor cu diferenţe.

Abstract. In the present paper we establish some new finite difference inequalities related to a certain inequality used in the theory of difference equations. The inequalities established here can be used as tools in the qualitative analysis of certain new classes of difference and sum-difference equations.

Introduction. In a recent paper [4, p.250] Mate and Navai used the following inequality while extending the well known results established by H. Poincaré in [9].

LEMMA. *Let* $u(n) \geq 0$, $p(n) \geq 0$ *be real-valued functions defined on integers and let* $c \geq 0$ *be a real constant. If*

$$u(n) \leq c + \sum_{s=n+1}^{\infty} p(s)\, u(s),$$

then

$$u(n) \leq c \exp\left(\sum_{s=n+1}^{\infty} p(s)\right).$$

Finite difference inequalities of this type are most useful in the qualitative analysis of various classes of difference equations. In the past few years, many papers on finite difference

* *Marathwada University, Department of Mathematics, Aurangabad 431004 (Maharashtra) India*

inequalities of the above type and their applications have appeared in the literature, see [1-8, 10] and the references given therein. In view of the important role played by such inequalities in the study of difference equations, it is natural to expect that some new finite difference inequalities of the type given in Lemma, would also be equally important in certain new applications. The main purpose of the present paper is to establish some new finite difference inequalities of the type given in Lemma, which can be used as tools in the analysis of certain new classes of difference and sum-difference equations for which earlier inequalities fail to apply directly. An application to obtain a bound on the solution of a certain sum-difference equation is also given.

2. Statement of results. In what follows we let $N_0 = \{0, 1, 2, \ldots\}$ and use the notations m, n, p, q to denote the elements of N_0. Let R denote the set of real numbers and $R_+ = [0,\infty)$. For $n > m$, $n,m \in N_0$ and any function $h: N_0 \to R$, we use the usual conventions $\sum_{s=n}^{m} h(s) = 0$ and $\prod_{s=n}^{m} h(s) = 1$. Throughout, without further mention, we assume that all the sums and products converge on the respective domain of their definitions.

Our main results are given in the following theorems.

THEOREM 1. *Let $u(n)$, $f(n)$, $g(n)$, $h(n)$ be functions defined on N_0 into R_+ and $c \geq 0$ be a real constant.*

(i) *If*

$$u^2(n) \leq c^2 + 2\sum_{s=n+1}^{\infty} [f(s)u^2(s) + h(s)u(s)], \quad n \in N_0, \tag{1}$$

then

$$u(n) \leq c\prod_{t=n+1}^{\infty} [1 + f(t)] + \sum_{s=n+1}^{\infty} h(s) \prod_{t=n+1}^{s-1} [1 + f(s)], \quad n \in N_0 \tag{2}$$

(ii) *If*

$$u^2(n) \le c^2 + 2\sum_{s=n+1}^{\infty} \left[f(s)u(s)\left(u(s) + \sum_{t=s+1}^{\infty} g(t)u(t)\right) + h(s)u(s) \right], \; n \in N_0, \tag{3}$$

then

$$u(n) \le c\sum_{t=n+1}^{\infty} [1 + f(t) + g(t)] + \sum_{s=n+1}^{\infty} h(s)\cdot \prod_{t=n+1}^{s-1} [1 + f(t) + g(t)], \; n \in N_0. \tag{4}$$

(iii) *If*

$$u^2(n) \le c^2 + 2\sum_{s=n+1}^{\infty} \left[f(s)u(s)\left(\sum_{t=s+1}^{\infty} g(t)u(t)\right) + h(s)u(s) \right], \; n \in N_0, \tag{5}$$

then

$$u(n) \le c\sum_{t=n+1}^{\infty} \left[1 + f(t) + \sum_{\tau=t+1}^{\infty} g(\tau)\right] + \sum_{s=n+1}^{\infty} h(s) \prod_{t=n+1}^{s-1} \left[1 + f(t) + \sum_{\tau=t+1}^{\infty} g(\tau)\right], \; n \in N_0. \tag{6}$$

THEOREM 2. *Let $u(m,n)$, $f(m,n)$, $g(m,n)$, $h(m,n)$ be functions defined for $m,n \in N_0$ into R_+ and $c \ge 0$ is a real constant.*

(iv) *If*

$$u^2(m,n) \le c^2 + 2\sum_{s=m+1}^{\infty} \sum_{t=n+1}^{\infty} [f(s,t)u^2(s,t) + h(s,t)u(s,t)], \; m,n \in N_0, \tag{7}$$

then

$$u(m,n) \le \phi(m,n) \prod_{s=m+1}^{\infty} \left[1 + \sum_{t=n+1}^{\infty} f(s,t)\right], \; m,n \in N_0, \tag{8}$$

where

$$\phi(m,n) = c + \sum_{s=m+1}^{\infty} \sum_{t=n+1}^{\infty} h(s,t). \tag{9}$$

(v) *If*

$$u^2(m,n) \le c^2 + 2\sum_{s=m+1}^{\infty} \sum_{t=n+1}^{\infty} [f(s,t)u(s,t)(u(s,t)$$

$$+ \sum_{x=s+1}^{\infty} \sum_{y=t+1}^{\infty} g(x,y)u(x,y)) + h(s,t)u(s,t)], \; m,n \in N_0, \tag{10}$$

then

$$u(m,n) \le \phi(m,n) \prod_{s=m+1}^{\infty} \left[1 + \sum_{t=n+1}^{\infty} [f(s,t) + g(s,t)]\right], \; m,n \in N_0, \tag{11}$$

where $\phi(m,n)$ *is defined as in* (9).

(vi) *If*

$$u^2(m,n) \le c^2 + 2\sum_{s=m+1}^{\infty}\sum_{t=n+1}^{\infty}\left[f(s,t)u(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\,u(x,y)\right) + h(s,t)\,u(s,t)\right],\quad m,n\in N_0, \tag{12}$$

then

$$u(m,n) \le \phi(m,n)\prod_{s=m+1}^{\infty}\left[1 + \sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\right)\right],\quad m,n\in N_0, \tag{13}$$

where $\phi(m,n)$ *is defined as in* (9).

3. Proofs of theorems 1 and 2. Since the proofs of (i)-(vi) resemble one another, we give the details for (ii) and (vi) only, the proofs of the remaining inequalities can be completed by following the proofs of (ii) and (vi).

(ii) Define a function $z(n)$ by

$$z(n) = (c+\varepsilon)^2 + 2\sum_{s=n+1}^{\infty}\left[f(s)u(s)\left(u(s) + \sum_{t=s+1}^{\infty} g(t)\,u(t)\right) + h(s)\,u(s)\right], \tag{14}$$

where $\varepsilon > 0$ is an arbitrary small constant. From (14) and using the fact that $u(n+1) \le \sqrt{z(n+1)}$, $n\in N_0$, we observe that

$$z(n) - z(n+1) \le 2\sqrt{z(n+1)}\left[f(n+1)\left(\sqrt{z(n+1)} + \sum_{t=n+2}^{\infty} g(t)\sqrt{z(t)}\right) + h(n+1)\right]. \tag{15}$$

Using the facts that $\sqrt{z(n+1)} > 0$, $\sqrt{z(n+1)} \le \sqrt{z(n)}$ for $n \in N_0$ and (15) we observe that

$$\sqrt{z(n)} - \sqrt{z(n+1)} = \frac{z(n) - z(n+1)}{\sqrt{z(n)} + \sqrt{z(n+1)}} \le \frac{z(n) - z(n+1)}{2\sqrt{z(n+1)}}$$

$$\le f(n+1)\left(\sqrt{z(n+1)} + \sum_{t=n+2}^{\infty} g(t)\sqrt{z(t)}\right) + h(n+1). \tag{16}$$

Define a function $v(n)$ by

$$v(n) = \sqrt{z(n)} + \sum_{t=n+1}^{\infty} g(t)\sqrt{z(t)}\,. \tag{17}$$

From (17) and (16) it is easy to observe that

$$v(n) - [1 + f(n+1) + g(n+1)]\, v(n+1) \leq h(n+1). \tag{18}$$

Now multiplying (18) by $\prod_{t=n+1}^{m} [1 + f(t) + g(t)]^{-1}$, for an arbitrary $m \in N_0$, then setting $n = s$ and taking the sum over $s = n, n+1, \ldots, m-1$ we obtain

$$v(n)\prod_{t=n+1}^{m} [1 + f(t) + g(t)]^{-1} \leq v(m) + \sum_{s=n+1}^{m} h(s)\prod_{t=s}^{m} [1 + f(t) + g(t)]^{-1}. \tag{19}$$

From (19) we have

$$v(n) \leq v(m)\prod_{t=n+1}^{m} [1 + f(t) + g(t)] + \sum_{s=n+1}^{m} h(s)\prod_{t=n+1}^{s-1} [1 + f(t) + g(t)]. \tag{20}$$

Noting that $\lim_{m\to\infty} v(m) = \lim_{m\to\infty} \sqrt{z(m)} = c + \varepsilon$ and letting $m \to \infty$ in (20) we get

$$v(n) \leq (c + \varepsilon)\prod_{t=n+1}^{\infty} [1 + f(t) + g(t)] + \sum_{s=n+1}^{\infty} h(s)\sum_{t=n+1}^{s-1} [1 + f(t) + g(t)]. \tag{21}$$

The required inequality in (4) now follows from (21) and using the facts that $u(n) \leq \sqrt{z(n)}$ and $\sqrt{z(n)} \leq v(n)$ and by taking $\varepsilon \to 0$.

(vi) Define a function $z(m,n)$ by

$$z(m,n) = (c+\varepsilon)^2 + 2\sum_{s=m+1}^{\infty}\sum_{t=n+1}^{\infty}\left[f(s,t)\, u(s,t)\cdot\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\, u(x,y)\right) + h(s,t)\, u(s,t)\right], \tag{22}$$

where $\varepsilon > 0$ is an arbitrary small constant. From (22) and using the facts that $u(m,n) \leq \sqrt{z(m,n)}$ for $m,n \in N_0$, we observe that

$$[z(m,n) - z(m+1,n)] - [z(m,n+1) - z(m+1,n+1)]$$

$$\leq 2\sqrt{z(m+1,n+1)}\left[f(m+1,n+1)\left(\sum_{x=m+2}^{\infty}\sum_{y=n+2}^{\infty} g(x,y)\sqrt{z(x,y)}\right) + h(m+1,n+1)\right]. \tag{23}$$

Using the facts that $\sqrt{z(m,n)} > 0$, $\sqrt{z(m,n+1)} \leq \sqrt{z(m,n)}$, $\sqrt{z(m+1,n+1)} \leq \sqrt{z(m+1,n)}$,

$\sqrt{z(m+1,n+1)} \le \sqrt{z(m,n+1)}$ for $m,n \in N_0$, we observe that (see, [7, p.379])

$$\left[\sqrt{z(m,n)} - \sqrt{z(m+1,n)}\right] = \frac{[z(m,n) - z(m+1,n)]}{\left[\sqrt{z(m,n)} + \sqrt{z(m+1,n)}\right]},$$

and

$$\left[\sqrt{z(m,n)} - \sqrt{z(m+1,n)}\right] - \left[\sqrt{z(m,n+1)} - \sqrt{z(m+1,n+1)}\right]$$

$$\le \frac{[z(m,n) - z(m+1,n)] - [z(m,n+1) - z(m+1,n+1)]}{\left[\sqrt{z(m+1,n+1)} + \sqrt{z(m+1,n+1)}\right]}. \tag{24}$$

From (24) and (23) we observe that

$$\left[\sqrt{z(m,n)} - \sqrt{z(m+1,n)}\right] - \left[\sqrt{z(m,n+1)} - \sqrt{z(m+1,n+1)}\right]$$

$$\le f(m+1,n+1)\left(\sum_{x=m+2}^{\infty}\sum_{y=n+2}^{\infty} g(x,y)\sqrt{z(x,y)}\right) + h(m+1,n+1). \tag{25}$$

Now keeping m fixed in (25), set $n = t$ and sum over $t = n, n+1, \ldots, q-1$ to obtain

$$\left[\sqrt{z(m,n)} - \sqrt{z(m+1,n)}\right] - \left[\sqrt{z(m,q)} - \sqrt{z(m+1,q)}\right]$$

$$\le \sum_{t=n+1}^{q}\left[f(m+1,t)\left(\sum_{x=m+2}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\sqrt{z(x,y)}\right) + h(m+1,t)\right]. \tag{26}$$

Noting that $\lim_{q\to\infty}\sqrt{z(m,q)} = \lim_{q\to\infty}\sqrt{z(m+1,q)} = c + \varepsilon$, and by letting $q \to \infty$ in (26) we get

$$\left[\sqrt{z(m,n)} - \sqrt{z(m+1,n)}\right]$$

$$\le \sum_{t=n+1}^{\infty}\left[f(m+1,t)\left(\sum_{x=m+2}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\sqrt{z(x,y)}\right) + h(m+1,t)\right]. \tag{27}$$

Keeping n fixed in (27), set $m = s$ and sum over $s = m, m+1, \ldots, p-1$ to obtain

$$\sqrt{z(m,n)} - \sqrt{z(p,n)} \le \sum_{s=m+1}^{p}\sum_{t=n+1}^{\infty}\left[f(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\cdot\sqrt{z(x,y)}\right) + h(s,t)\right]. \tag{28}$$

Noting that $\lim\limits_{p\to\infty}\sqrt{z(p,n)} = c+\varepsilon$, and by letting $p \to \infty$ in (28) we get

$$\sqrt{z(m,n)} \le \phi_{\varepsilon}(m,n) + \sum_{s=n+1}^{\infty}\sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\sqrt{z(x,y)}\right), \tag{29}$$

where $\phi_{\varepsilon}(m,n) = c+\varepsilon+\sum\limits_{s=m+1}^{\infty}\sum\limits_{t=n+1}^{\infty} h(s,t)$. From (29) it is easy to observe that

$$\frac{\sqrt{z(m,n)}}{\phi_{\varepsilon}(m,n)} \le 1 + \sum_{s=m+1}^{\infty}\sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\frac{\sqrt{z(x,y)}}{\phi_{\varepsilon}(x,y)}\right). \tag{30}$$

Define $v(m,n)$ by

$$v(m,n) = 1 + \sum_{s=m+1}^{\infty}\sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\frac{\sqrt{z(x,y)}}{\phi_{\varepsilon}(x,y)}\right). \tag{31}$$

From (31) and (30) it is easy to observe that

$$[v(m,n)-v(m+1,n)]-[v(m,n+1)-v(m+1,n+1)]$$

$$\le f(m+1,n+1)\left(\sum_{x=m+2}^{\infty}\sum_{y=n+2}^{\infty} g(x,y)\right)v(m+1,n+1). \tag{32}$$

From the definition of $v(m,n)$ given in (31) we observe that $v(m+1,n+1) \le v(m+1,n)$ for $m,n \in N_0$. Using this in (32) we observe that

$$\frac{[v(m,n)-v(m+1,n)]}{v(m+1,n)} - \frac{[v(m,n+1)-v(m+1,n+1)]}{v(m+1,n+1)}$$

$$\le f(m+1,n+1)\sum_{x=m+2}^{\infty}\sum_{y=n+2}^{\infty} g(x,y). \tag{33}$$

Now keeping m fixed in (33), set $n = t$ and sum over $t = n, n+1, \ldots, q-1$ to obtain

$$\frac{[v(m,n)-v(m+1,n)]}{v(m+1,n)} - \frac{[v(m,q)-v(m+1,q)]}{v(m+1,q)}$$

$$\le \sum_{t=n+1}^{q} f(m+1,t)\left(\sum_{x=m+2}^{\infty}\sum_{y=t+1}^{\infty} g(x,y)\right). \tag{34}$$

Noting that $\lim\limits_{q\to\infty} v(m,q) = \lim\limits_{q\to\infty} v(m+1,q) = 1$, and by letting $q \to \infty$ in (34) we get

$$\frac{v(m,n) - v(m+1,n)}{v(m+1,n)} \le \sum_{t=n+1}^{\infty} f(m+1,t)\left(\sum_{x=m+2}^{\infty} \sum_{y=t+1}^{\infty} g(x,y)\right). \tag{35}$$

From (35) we have

$$v(m,n) \le v(m+1,n)\left[1 + \sum_{t=n+1}^{\infty} f(m+1,t)\left(\sum_{x=m+2}^{\infty} \sum_{y=t+1}^{\infty} g(x,y)\right)\right]. \tag{36}$$

Now keeping n fixed in (36), set $m = s$ and sum over $s = m, m+1, \ldots, p-1$ succesively to obtain

$$v(m,n) \le v(p,n) \prod_{s=m+1}^{p}\left[1 + \sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty} \sum_{y=t+1}^{\infty} g(x,y)\right)\right]. \tag{37}$$

Noting that as $p \to \infty$, $v(p,n) = 1$, and letting $p \to \infty$ in (37) we have

$$v(m,n) \le \prod_{s=m+1}^{\infty}\left[1 + \sum_{t=n+1}^{\infty} f(s,t)\left(\sum_{x=s+1}^{\infty} \sum_{y=t+1}^{\infty} g(x,y)\right)\right]. \tag{38}$$

The desired inequality in (13) now follows by using (38) in (30), the fact that $u(m,n) \le \sqrt{z(m,n)}$ and by taking $\varepsilon \to 0$. This completes the proof of (vi).

4. An application. In this section we present an application of our inequality given in Theorem 1 part (i) to obtain bound on the solution of the following sum-difference equation

$$y^2(n) = p(n) + \sum_{s=n+1}^{\infty} k(n,s)\, y(s) F(s, y(s)),\ n \in N_0, \tag{39}$$

where $p: N_0 \to R$, $k: N_0 \times N_0 \to R$, $F: N_0 \times R \to R$. We assume that

$$|p(n)| \le c^2,\ |k(n,s)F(s,y(s))| \le 2\,[f(s)|y(s)| + h(s)], \tag{40}$$

where f, h and c are as defined in Theorem 1. From (39) and (40) we obtain

$$|y(n)|^2 \le c^2 + 2\sum_{s=n+1}^{\infty} [f(s)\,|y(s)|^2 + h(s)|y(s)|\,]. \tag{41}$$

Now an application of the inequality given in Theorem 1 part (i) to (41) yields

$$|y(n)| \le c\prod_{t=n+1}^{\infty} [1 + f(t)] + \sum_{s=n+1}^{\infty} h(s) \prod_{t=n+1}^{s-1} [1 + f(t)],\ n \in N_0. \tag{42}$$

The inequality (42) gives the bound on the solution $y(n)$ of equation (39) in terms of the known functions.

Finally, we note that the inequalities established in Theorem 2 can be extended very easily to functions of several independent variables. We also note that there are many possible applications of the inequalities established in Theorems 1 and 2 to certain new classes of difference and sum-difference equations. However, the discussion of such applications is left to another place.

REFERENCES

1. P.R. Beesack, *"Gronwall Inequalities"*, Carleton University Mathematical Lecture Notes, No. 11, Northfield Minn., 1975.
2. A. Corduneanu, *A discrete integral inequality of the convolution type*, Bull. Inst. Politechnic Iaşi 33(1987), 35-40.
3. S.S. Dragomir, *Nonlinear discrete generalizations of Gronwall's inequality*, An. Univ. din Timişoara Ser. Sti. Mah. 29(1991), 45-50.
4. A. Mate and P. Nevai, *Sublinear perturbations of the differential equation* $y^{(n)} = 0$ *and of the analogous difference equations*, J. Differential Equations 53(1994), 234-257.
5. B.G. Pachpatte, *Finite difference inequalities and their applications*, Proc. Nat. Acad. Sci. India 43(A)(1973), 348-356.
6. B.G. Pachpatte, *On Bihari like integral and discrete inequalities*, Soochow Jour. Math. 17(1991), 213-232.
7. B.G. Pachpatte, *On certain new finite difference inequalities*, Indian J. Pure Appl. Math. 24(1993), 373-384.
8. B.G. Pachpatte, *On some new inequalities related to certain inequalities in the theory of differential equations*, J. Math. Anal. Appl., to appear.
9. H. Poincaré, *Sur les equations lineaires aux differentielles et aux differences finies*, Amer. J. Math. 7(1885), 203-258.
10. S. Sugiyama, *On the stability problems of difference equations*, Bull. Sci. Engin. Research Lab. Waswda Uni. 45(1969), 140-144.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

PROBABILISTIC POSITIVE LINEAR OPERATORS

I. RAŞA*

Received: Mars 15, 1995
AMS subject classification: 41A36

REZUMAT. - Operatori liniari pozitivi probabilistici. Pentru un şir de operatori probabilistici se indică un algoritm de tip Casteljau. Se prezintă apoi câteva aplicaţii.

1. Introduction. For every x in an interval I of the real axis let us consider a sequence of independent and identically distributed random variables $(Y_n^x)_{n\geq 1}$. Let $p_{ni} \geq 0$, $i = 1, \ldots, n$, such that $p_{n1} + \ldots + p_{nn} = 1$ for each $n \geq 1$.

For a continuous function f on the real line let us denote

$$L_n f(x) = Ef\left(\sum_{i=1}^{n} p_{ni} Y_i^x\right) \tag{1}$$

provided that the expectation is finite.

Many classical positive linear operators (in particular Bernstein, Szász, Gamma, Weierstrass and Baskakov operators) are of the form (1). The probabilistic positive linear operators have been extensively studied; see [1], [3], [7], [8] and the references therein.

Our approach is based on a recursive algorithm related to Casteljau's algorithm. It allows us to deduce some properties of L_n from those of L_1. Finally we shall generalize a result from [7] concerning monotonic convergence under convexity. Other results of this type are to be found in [4] and [13].

* *Technical University, Department of Mathematics, 3400 Cluj-Napoca, Romania*

2. The algorithm. Let f be a given continuous function on R. For $x \in I$ and $t_1, \ldots, t_n \in R$ denote

$$f_0^x(t_1, \ldots, t_n) = f(p_{n1}t_1 + \ldots + p_{nn}t_n)$$

$$f_k^x(t_1, \ldots, t_{n-k}) = Ef_{k-1}^x(t_1, \ldots, t_{n-k}, Y_{n-k+1}^x), \; k = 1, \ldots, n-1.$$

Then we have

$$L_n f(x) = Ef_0^x(Y_1^x, \ldots, Y_n^x) = Ef_1^x(Y_1^x, \ldots, Y_{n-1}^x) = \ldots =$$

$$= Ef_{n-1}^x(Y_1^x) = L_1 f_{n-1}^x(x) \tag{2}$$

Examples. (a) Let $p_{ni} = 1/n$, $n \geq 1$, $i = 1, \ldots, n$. Let $(X_k)_{k \geq 1}$ be a sequence of independent and on [0,1] uniformly distributed random variables. Let $Y_n^x = I_{(X_n \leq x)}$, $0 \leq x \leq 1$, where I_C denotes the indicator function of C. Then $L_n f(x)$ coincides with the Bernstein operator $B_n f(x)$; see [1].

For $x \in [0,1]$, $f \in C[0,1]$, $k = 1, \ldots, n-1$, $t_1, \ldots, t_n \in \{0,1\}$ we have

$$f_0^x(t_1, \ldots, t_n) = f((t_1 + \ldots + t_n)/n)$$

$$f_k^x(t_1, \ldots, t_{n-k}) = (1-x)f_{k-1}^x(t_1, \ldots, t_{n-k}, 0) + xf_{k-1}^x(t_1, \ldots, t_{n-k}, 1)$$

$$L_n f(x) = (1-x)f_{n-1}^x(0) + xf_{n-1}^x(1)$$

It follows that the computation of $L_n f(x)$ by means of (2) is equivalent to the computation of $B_n f(x)$ by means of the Casteljau algorithm [9] (see also [11] and [14]).

(b) In the case of the Szász operator (see [7]) we have for $x \geq 0$, $k = 1, \ldots, n-1$, $t_i = 0, 1, \ldots,$

$$f_0^x(t_1, \ldots, t_n) = f((t_1 + \ldots + t_n)/n)$$

$$f_k^x(t_1, \ldots, t_{n-k}) = e^{-x} \sum_{j=0}^{\infty} f_{k-1}^x(t_1, \ldots, t_{n-k}, j)x^j/j!$$

$$S_n f(x) = e^{-x} \sum_{j=0}^{\infty} f_{n-1}^x(j)x^j/j!$$

(c) Let $p_{ni} = 1/n$ and let Y_n^x be uniformly distributed on $[x-1, x+1]$. Then $L_n f(x)$ is the operator of Pečarić and Zwick [12]. We have for $k = 1, \ldots, n-1$,

$$f_0^x(t_1, \ldots, t_n) = f((t_1 + \ldots + t_n)/n)$$

$$f_k^x(t_1, \ldots, t_{n-k}) = (1/2) \int_{x-1}^{x+1} f_{k-1}^x(t_1, \ldots, t_{n-k}, t)\,dt$$

$$L_n f(x) = (1/2) \int_{x-1}^{x+1} f_{n-1}^x(t)\,dt$$

Remark 1. Let $p_{ni} = 1/n$. Denote $g_0^x = f$ and

$$g_k^x(u) = Ef((n-k)\,u/n + (Y_{n-k+1}^x + \ldots + Y_n^x)/n), \; k = 1, \ldots, n-1.$$

Then $f_k^x(t_1, \ldots, t_{n-k}) = g_k^x((t_1 + \ldots + t_{n-k})/(n-k))$.

Consider again the above example (c) and express $L_n f(x)$ by means of a divided difference (see [12]); we deduce

$$L_n f(x) = \int_R g_{n-1}^x(u)\,B_0^x(u)\,du = \int_R g_{n-2}^x(u)\,B_1^x(u)\,du = \ldots =$$
$$= \int_R g_0^x(u)\,B_{n-1}^x(u)\,du$$

where B_{j-1}^x is the B-spline function [9] of degree j-1 corresponding to the equidistant points $x-1 = t_0 < t_1 < \ldots < t_j = x+1$, $j = 1, \ldots, n$.

In particular, $L_n f(0) = \int_R f(u)\,B_{n-1}^0(u)\,du$. This means that the probability density of $(Y_1^0 + \ldots + Y_n^0)/n$ is the spline function B_{n-1}^0. The characteristic function of the same variable is

$$\varphi(t) = ((n/t)\sin(t/n))^n$$

It follows that the Fourier transform of B_{n-1}^0 is φ (see also [5]).

3. Applications. For $M > 0$ denote

$$\mathrm{Lip}\,(M; I) = \{f \in C(I) : |f(x) - f(y)| \le M|x-y|, \; x, y \in I\}.$$

The following lemma can be proved by induction and we omit the details.

LEMMA 1. (i) *If* $f \in \mathrm{Lip}\,(M;R)$ *then*

$$f_k^x(t_1, \ldots, t_{n-k-1}, \cdot) \in \mathrm{Lip}\,(Mp_{n,n-k}; R),\ k = 0, \ldots, n-1.$$

(ii) *If f is increasing, then* $f_k^x(t_1, \ldots, t_{n-k-1}, \cdot)$ *is increasing,* $k = 0, \ldots, n-1$.

THEOREM 1. *Let* $M,N > 0$. *If* L_1 *transforms the functions from* $\mathrm{Lip}(M;R)$ [*the increasing functions*] *into functions from* $\mathrm{Lip}(N;I)$ [*increasing functions*], *then the same is true for each* L_n, $n > 1$.

Proof. Let $x,y \in I$, $f \in \mathrm{Lip}(M;R)$, $n > 1$ and $q = |x-y|$. Then, by (i), $f_k^y(t_1, \ldots, t_{n-k-1}, \cdot)$ is in $\mathrm{Lip}(Mp_{n,n-k}; R)$, hence $L_1 f_k^y(t_1, \ldots, t_{n-k-1}, \cdot)$ is in $\mathrm{Lip}(Np_{n,n-k}; I)$. This means that the function $t \rightarrow Ef_k^y(t_1, \ldots, t_{n-k-1}, Y_{n-k}^t)$ is in $\mathrm{Lip}(Np_{n,n-k}; I)$ for each $k = 0, \ldots, n-1$.

Let F_x be the distribution function of Y_1^x. Since $f_0^x = f_0^y$, we have

$$L_n f(x) = Ef_0^x(Y_1^x, \ldots, Y_n^x) = Ef_0^y(Y_1^x, \ldots, Y_n^x) =$$

$$= \int_{R^{n-1}} Ef_0^y(t_1, \ldots, t_{n-1}, Y_n^x)\, dF_x(t_1) \ldots dF_x(t_{n-1}) \leq$$

$$\leq \int_{R^{n-1}} Ef_0^y(t_1, \ldots, t_{n-1}, Y_n^y)\, dF_x(t_1) \ldots dF_x(t_{n-1}) + Nqp_{nn} =$$

$$\int_{R^{n-1}} f_1^y(t_1, \ldots, t_{n-1})\, dF_x(t_1) \ldots dF_x(t_{n-1}) + Nqp_{nn} =$$

$$= Ef_1^y(Y_1^x, \ldots, Y_{n-1}^x) + Nqp_{nn}.$$

By repeating this argument we obtain finally

$$L_n f(x) \leq Ef_{n-1}^y(Y_1^x) + Nq(p_{nn} + \ldots + p_{n2}) \leq Ef_{n-1}^y(Y_1^y) + Nq.$$

By virtue of (2) we have $L_n f(x) \leq L_n f(y) + Nq$. It follows immediately that $|L_n f(x) - L_n f(y)| \leq N|x-y|$, hence $L_n f \in \mathrm{Lip}(N; I)$.

The assertion concerning increasing functions can be proved similarly.

4. Monotonic convergence. In what follows we put $p_{n,n+1} = 0$, $n \geq 1$ and we shall suppose that

$$(p_{n,1}, \dots, p_{n,n+1}) \text{ majorizes } (p_{n+1,1}, \dots, p_{n+1,n+1}) \tag{3}$$

(Concerning majorization, see [10]).

THEOREM 2. *Under the above hypothesis we have* $L_n f \geq L_{n+1} f$ *iff* f *is convex.*

Proof. Let $x \in I$. If f is convex then the function

$$(q_1, \dots, q_{n+1}) \to Ef\left(\sum_{i=1}^{n+1} q_i Y_i^x\right)$$

is convex and symmetric, hence it is Schur-convex [10; 3.C.2]. Now from (3) it follows that

$$Ef\left(\sum_{i=1}^{n+1} p_{ni} Y_i^x\right) \geq Ef\left(\sum_{i=1}^{n+1} p_{n+1,i} Y_i^x\right)$$

This means that $L_n f(x) \geq L_{n+1} f(x)$ and the proof is finished.

Remark 2. The above proof is suggested by Theorems 3.7 and 3.8 of [6]. From Theorem 2 with $p_{ni} = 1/n$ we obtain the inequality contained in [7; Theorem 3] (see also [2]) and proved there by means of a martingale-type property and the conditional version of Jensen's inequality.

REFERENCES

1. J.A. Adell and J. de la Cal, *Using stochastic processes for studying Bernstein-type operators*, Rend. Circ. Mat. Palermo (2) Suppl., **33**(1993), 125-141.
2. J.A. Adell, J. de la Cal and M. San Miguel, *On the property of monotonic convergence for multivariate Bernstein-type operators*, J. Approx. Theory, **80**(1995), 132-137.
3. F. Altomare and M. Campiti, *Korovkin-type approximation theory and its applications*, W. de Gruyter (Berlin, New York, 1994).
4. B. Della Vecchia and I. Rasa, *Bernstein-type operators, convexity and Lipschitz classes*, Approx. Theory Appl. (Beijing), To appear.
5. A. Erdélyi (Ed.), *Tables of integral transforms*, McGraw-Hill (New York, 1954).
6. S. Karlin, *Total positivity*, Stanford Univ. Press (Stanford, California, 1968).
7. R.A. Khan, *Some probabilistic methods in the theory of approximation operators*, Acta Math. Acad. Sci. Hung. **35**(1980), 193-203.
8. M.K. Khan and M.A. Peters, *Lipschitz constants for some approximation operators of a Lipschitz*

. *continuous function*, J. approx. Theory **59**(1989), 307-315.
9. F. Locher, *Numerische Mathematik für Informatiker*, Springer Verlag (Berlin - Heidelberg - New York, 1992).
10. A.W. Marshall and I. Olkin, *Inequalities: Theory of majorization and its applications*, Academic Press (New York, 1979).
11. T. Popoviciu, *Despre cea mai bună aproximaţie a funcţiilor continue prin polinoame*, Inst. Arte Grafice Ardealul (Cluj, 1937).
12. I. Rasa, *Korovkin approximation and parabolic functions*, Conf. Sem. Mat. Univ. Bari **236**(1991), 1-25.
13. I. Rasa, *Altomare projections and Lototsky-Schnabl operators*, Rend. Circ. Mat. Palermo (2) Suppl. **33**(1993), 439-451.
14. D.D. Stancu, *Folosirea interpolării liniare pentru construirea unei clase de polinoame Bernstein*, Studii Cerc. Mat. Bucureşti **28**(1976), 369-379.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

A CLASS OF INTEGRAL FAVARD-SZASZ TYPE OPERATORS

Alexandra CIUPA[*]

Received: November 6, 1994
AMS subject classification: 41A36, 41A17, 41A25

REZUMAT. - O clasă de operatori integrali de tip Favard-Szasz. În această lucrare se consideră un operator de tip integral, în sensul lui Durrmeyer [2] şi Derriennic [1], care se obţine plecând de la un operator de tip Favard-Szasz (4), introdus în 1969 de către Jakimovski şi Leviatan [4]. Autoarea dă unele estimări cantitative, exprimate cu modulele de continuitate de primele două ordine, pentru aproximarea funcţiilor cu ajutorul operatorului L_n, definit la (6).

Abstract. This paper one considers an integral type operator, in the sense of Durrmeyer [2] and Derriennic [1], which is obtained by starting from a Favard-Szasz operator (4), introduced in 1969 by Jakimovski and Leviatan [4]. The author gives some quantitative estimates, in therms of the first and the second order moduli of continuity, for the approximation of functions by means of the operator L_n, defined at (6).

1. This paper is motivated by the works of J.L. Durrmeyer [2], A. Lupaş [6] and M.M. Derriennic [1], which have obtained and studied a modified Bernstein operator

$$(B_n^* f)(x) = (n+1)\sum_{k=0}^{n} d_{n,k}(x)\int_0^1 b_{n,k}(t) f(t)\,dt,$$

$$b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}, \tag{1}$$

where f is Lebesque integrable on [0,1].

S.M. Mazhar and V. Totik [7], similarly modified the Favard-Szasz operator and they

[*] *Technical University, Department of Mathematics, 3400 Cluj-Napoca, Romania*

have defined another class of positive linear operators

$$(S_n^* f)(x) = n\sum_{k=0}^{\infty} e^{-nx} \frac{(nx)^k}{k!} \int_0^{\infty} e^{-nt} \frac{(nt)^k}{k!} f(t)\, dt \tag{2}$$

for functions $f \in L_1[0, \infty)$ ∘

By using a similar way we will modify an operator introduced by A. Jakimovski and D. Leviatan [4]. Let us remind this operators. One considers $g(z) = \sum_{n=0}^{\infty} a_n z^n$ an analytic function in the disk $|z| < R$, $R > 1$, where $g(1) \neq 0$. It is known that the Appell polynomials $p_k(x)$, $k \geq 0$ can be defined by

$$g(u)\, e^{ux} = \sum_{k=0}^{\infty} p_k(x)\, u^k, \tag{3}$$

To a function f: $[0, \infty) \to R$ one associates the Jakimovski-Leviatan operator

$$(P_n f)(x) = \frac{e^{-nx}}{g(1)} \sum_{k=0}^{\infty} p_k(nx) f\left(\frac{k}{n}\right). \tag{4}$$

The case $g(z) = 1$ yields the classical operator of Favard-Szasz

$$(S_n f)(x) = e^{-nx} \sum_{k=0}^{\infty} \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right).$$

B. Wood [9] has proved that the operator P_n is positive if and only if $\frac{a_n}{g(1)} \geq 0$, $n = 0, 1, \ldots$

Now we will modify the operator P_n, as follows: for a function f, Lebesque integrable in $[0,\infty)$, we replace $f\left(\frac{k}{n}\right)$ into P_n, by a positive linear functional

$$A_k(f) = \frac{n^{\lambda+k+1}}{\Gamma(\lambda + k + 1)} \int_0^{\infty} e^{-nt} t^{\lambda+k} f(t)\, dt, \ \lambda \geq 0 \tag{5}$$

and so we obtain the operator

$$(L_n f)(x) = \frac{e^{-nx}}{g(1)} \sum_{k=0}^{\infty} p_k(nx) \frac{n^{\lambda+k+1}}{\Gamma(\lambda + k + 1)} \int_0^{\infty} e^{-nt} t^{\lambda+k} f(t)\, dt. \tag{6}$$

For $g(z)=1$ and $\lambda=0$ the operator defined at (6) becomes the operator S_n^*.

We suppose that this operator is positive, therefore $\frac{a_n}{g(1)} \geq 0$, $n = 0, 1, \ldots$ We denote by E the class of functions of exponential type, which have the property that $|f(t)| \leq e^{At}$, for each $t \geq 0$ and some finite number A.

The following lemma is essential to study the convergence of the sequence $(L_n f)$ to the function f.

LEMMA 1.1. *For all* $x \geq 0$, *we have*:

$$(L_n e_0)(x) = 1$$

$$(L_n e_1)(x) = x + \frac{1}{n}\left(\lambda + 1 + \frac{g'(1)}{g(1)}\right) \tag{7}$$

$$(L_n e_2)(x) = x^2 + \frac{2x}{n}\left(\lambda + 2 + \frac{g'(1)}{g(1)}\right) + \frac{1}{n^2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right],$$

where $e_i(x) = x^i$, $i \in \{0, 1, 2\}$.

Proof. We will use the properties of the gamma function and the values of the operator P_n defined at (4) for the monomials e_0, e_1, e_2:

$$(P_n e_0)(x) = 1$$

$$(P_n e_1)(x) = x + \frac{1}{n}\frac{g'(1)}{g(1)} \tag{8}$$

$$(P_n e_2)(x) = x^2 + \frac{x}{n}\left(1 + 2\frac{g'(1)}{g(1)}\right) + \frac{1}{n^2}\frac{g''(1)+g'(1)}{g(1)}.$$

For instance, let us calculate $(L_n e_1)(x)$. We have:

$$A_k(e_1) = \frac{n^{\lambda+k+1}}{\Gamma(\lambda+k+1)}\int_0^\infty e^{-nt}t^{\lambda+k+1}\,dt = \frac{1}{n}(\lambda+k+1)$$

and so we obtain

$$(L_n e_1)(x) = \frac{e^{-nx}}{g(1)}\sum_{k=0}^{\infty} p_k(nx)\frac{1}{n}(\lambda+k+1) = \frac{1}{n}(\lambda+1) + (P_n e_1)(x) =$$

$$= x + \frac{1}{n}\left(\lambda + 1 + \frac{g'(1)}{g(1)}\right)$$

THEOREM 1.2. *If* $f \in \mathbf{C}[0,\infty) \cap E$, *then* $\lim_{n\to\infty}(L_n f)(x) = f(x)$, *the convergence being uniform in each compact* $[0,a]$.

Proof. According to Lemma 1.1, we have $\lim_{n\to\infty}(L_n e_i)(x) = e_i(x)$, $i \in \{0, 1, 2\}$

uniformly on the compact $[0,a]$, so if we invoke the Bohman-Korovkin theorem, we obtain the desired result.

2. Estimate of the order of approximation. In this section we are concened with the estimate of the order of approximation of a function $f \in L_1[0,\infty)$ by means of the linear positive operator L_n. We will use the modulus of continuity defined by $\omega(f;\delta) = \sup |f(x'') - f(x')|$, where x' and x'' are points from $[0,a]$ so that $|x''-x'| < \delta$, δ being a positive number. By using a standard method we prove

THEOREM 2.1. *If $f \in L_1[0,a]$, then*

$$|(L_nf)(x)-f(x)| \le \left(1+\sqrt{2x+\frac{1}{n}\left((\lambda+1)(\lambda+2)+(2\lambda+3)\frac{g'(1)}{g(1)}+\frac{g''(1)+g'(1)}{g(1)}\right)}\right)\omega\left(f;\frac{1}{\sqrt{n}}\right)$$

Proof. Because $L_ne_0 = e_0$ and L_n is positive, we can write

$$|(L_nf)(x)-f(x)| \le \frac{e^{-nx}}{g(1)}\sum_{k=0}^{\infty}p_k(nx)\,|A_k(f)-f(x)A_k(e_0)| =$$

$$= \frac{e^{-nx}}{g(1)}\sum_{k=0}^{\infty}p_k(nx)\frac{n^{\lambda+k+1}}{\Gamma(\lambda+k+1)}\int_0^{\infty}e^{-nt}t^{\lambda+k}\,|f(t)-f(x)|\,dt \le$$

$$\le \frac{e^{-nx}}{g(1)}\sum_{k=0}^{\infty}p_k(nx)\left(1+\frac{1}{\delta}\frac{n^{\lambda+k+1}}{\Gamma(\lambda+k+1)}\int_0^{\infty}e^{-nt}t^{\lambda+k}\,|t-x|\,dt\right)\omega(f;\delta).$$

By making use of the Cauchy inequality, we obtain

$$\int_0^{\infty}e^{-nt}t^{\lambda+k}\,|t-x|\,dt \le \sqrt{\int_0^{\infty}e^{-nt}t^{\lambda+k}dt}\sqrt{\int_0^{\infty}e^{-nt}t^{\lambda+k}(t-x)^2dt} =$$

$$= \frac{\Gamma(\lambda+k+1)}{n^{\lambda+k+1}}\sqrt{x^2-2x\frac{k+\lambda+1}{n}+\frac{(k+\lambda+1)(k+\lambda+2)}{n^2}}.$$

It results that

$$|(L_n f)(x) - f(x)| \le$$

$$\left(1 + \frac{1}{\delta}\frac{e^{-nx}}{g(1)}\sum_{k=0}^{\infty} p_k(nx)\sqrt{\frac{(\lambda+1)(\lambda+2)}{n^2} + \frac{k(2\lambda+3)}{n^2} + \frac{k^2}{n^2} - 2x\frac{k+\lambda+1}{n} + x^2}\right)\omega(f;\delta).$$

We use again the Cauchy inequality and we get

$$|(L_n f)(x) - f(x)| \le \left(1 + \frac{1}{\delta}\sqrt{2\frac{x}{n} + \frac{1}{n^2}\left((\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right)}\right)\omega(f;\delta)$$

By inserting into it $\delta = \dfrac{1}{\sqrt{n}}$, we obtain the desired result. -

Next we will give some approximation theorems in different normed linear spaces. In order to establish the next results, we need some definitions.

The second order modulus of continuity of $f \in C_B[0,\infty)$ is

$$\omega_2(f;t) = \sup_{|h|\le t} \|f(\circ+2h) - 2f(\circ+h) + f(\circ)\|_{C_B},\ t \ge 0$$

where $C_B[0,\infty)$ is the class of real valued functions defined on $[0,\infty)$ which are bounded and uniformly continuous with the norm $\|f\|_{C_B} = \sup_{x\in[0,\infty)} |f(x)|$.

The Peetre K-functional of function $f \in C_B$ is defined as

$$K(f;t) = \inf_{g\in C_B^2} \left\{\|f-g\|_{C_B} + t\|g\|_{C_B^2}\right\}$$

where $C_B^2 = \left\{f \in C_B \mid f', f'' \in C_B\right\}$, with the norm $\|f\|_{C_B^2} = \|f\|_{C_B} + \|f'\|_{C_B} + \|f''\|_{C_B}$ It is known the following inequality;

$$K(f;t) \le A\left\{\omega_2(f;\sqrt{t}) + \min(1,t)\|f\|_{C_B}\right\} \tag{9}$$

for all $t \in [0,\infty)$, the constant A being independent of t and f. We will also use

LEMMA 2.2. *If $z \in C^2[0,\infty)$ and (P_n) is a sequence of linear positive operators with the property $P_n e_0 = e_0$, then*

$$|(P_n z)(x) - z(x)| \le \|z'\|\sqrt{\left(P_n(t-x)^2\right)(x)} + \frac{1}{2}\|z''\|\left(P_n(t-x)^2\right)(x)$$

The proof is analogous to the proof of theorem 2 from [3].

THEOREM 2.3. *If* $f \in C[0,a]$, *then for any* $x \in [0,a]$ *we have*

$$|(L_n f)(x) - f(x)| \le \frac{2h}{a}\|f\| + \frac{3}{4}\left(3 + \frac{a}{h}\right)\omega_2(f;h),$$

where $h = \sqrt{2\frac{x}{n} + \frac{1}{n^2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]}$.

Proof. Let f_h be the Steklov function attached to the function f. We will use the following result of V.V. Juk [5]: if $f \in C[a,b]$ and $h \in \left(0, \frac{b-a}{2}\right)$, then $\|f - f_h\| \le \frac{3}{4}\omega_2(f;h)$ and $\|f_h''\| \le \frac{3}{2}\frac{1}{h^2}\omega_2(f;h)$. Since $L_n e_0 = e_0$, we can write

$$|(L_n f)(x) - f(x)| \le |(L_n(f - f_h))(x)| + |(L_n f_h)(x) - f_h(x)| + |f_h(x) - f(x)| \le$$

$$\le 2\|f - f_h\| + |(L_n f_h)(x) - f_h(x)|$$

For the function $f_h \in C^2[0,a]$ we use lemma 2.2:

$$|(L_n f_h)(x) - f_h(x)| \le \|f_h'\|\sqrt{(L_n(t-x)^2)(x)} + \frac{1}{2}\|f_h''\|(L_n(t-x)^2)(x)$$

According to result from [3] and [5], we have

$$\|f_h'\| \le \frac{2}{a}\|f_h\| + \frac{a}{2}\|f_h''\| \le \frac{2}{a}\|f\| + \frac{a}{2}\|f_h''\| \le \frac{2}{a}\|f\| + \frac{3a}{4}\frac{1}{h^2}\omega_2(f;h).$$

By making use of this inequality and choosing $h = \sqrt{(L_n(t-x)^2(x)}$ we obtain

$$|(L_n f_h)(x) - f_h(x)| \le \frac{2}{a}\|f\|h + \frac{3a}{4}\frac{1}{h}\omega_2(f;h) + \frac{3}{4}\omega_2(f;h)$$

and therefore we get

$$|(L_n f)(x) - f(x)| \le 2\|f - f_h\| + \frac{2}{a}\|f\|h + \frac{3}{4}\left(\frac{a}{h} + 1\right)\omega_2(f;h)$$

Here we use the inequality $\|f - f_h\| \le \frac{3}{4}\omega_2(f;h)$ and we obtain the desired result.

Remark. If we consider $g(z) \equiv 1$ and $\lambda = 0$, we obtain, for the operator due to S.M. Mazhar and V. Totik [7], the estimation

$$|(S_n^* f)(x) - f(x)| \le \frac{2h}{a}\|f\| + \frac{3}{4}\left(3 + \frac{a}{h}\right)\omega_2(f;h),$$

where $h = \sqrt{2\frac{x}{n} + \frac{2}{n^2}}$.

THEOREM 2.4. *For every function* $f \in C_B^2[0,\infty)$, *we have*

$$|(L_nf)(x) - f(x)| \le \frac{1}{n}\left\{x + \frac{1}{2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]\right\}\|f\|_{C_B^2}$$

Proof. Applying the Taylor expansion to the function $f \in C_B^2$, we have

$$(L_nf)(x) - f(x) = f'(x)(L_n(t-x))(x) + \frac{1}{2}f''(\xi)(L_n(t-x)^2)(x), \text{ where } \xi \in (t,x)$$

By using lemma 1.1, we can write successively

$$|(L_nf)(x) - f(x)| \le \frac{1}{n}\left(\lambda + 1 + \frac{g'(1)}{g(1)}\right)\|f'\|_{C_B} +$$

$$+\frac{1}{2n}\left\{2x + \frac{1}{n}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]\right\}\|f''\|_{C_B} \le \frac{1}{n}\left(\lambda+1+\frac{g'(1)}{g(1)}\right)\|f'\|_{C_B} +$$

$$+\frac{1}{n}\left\{x + \frac{1}{2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]\right\}\|f''\|_{C_B} \le$$

$$\le \frac{1}{n}\left\{x + \frac{1}{2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]\right\}\left(\|f'\|_{C_B} + \|f''\|_{C_B}\right)$$

Remark. If we take into it $g(z) = 1$ and $\lambda = 0$, we obtain

$$|(S_n^*f)(x) - f(x)| \le \frac{1}{n}(x+1)\|f\|_{C_B^2}$$

result obtained by S.P. Singh and M.K. Tiwari [8].

THEOREM 2.5. *If* $f \in C_B[0,\infty)$, *then we have*

$$|(L_nf)(x) - f(x)| \le 2A\left(\omega_2(f;h) + \lambda_n(x)\|f\|_{C_B}\right),$$

where $h = \sqrt{\frac{1}{2n}\left\{x + \frac{1}{2}\left[(\lambda+1)(\lambda+2) + (2\lambda+3)\frac{g'(1)}{g(1)} + \frac{g''(1)+g'(1)}{g(1)}\right]\right\}}$,

$\lambda_n(x) = \min(1, h^2)$ *and A is a constant independent of h and f.*

Proof. We will use the theorem 2.4 and the *K*-functional. For $f \in C_B[0,\infty)$ and

$z \in C_B^2[0,\infty)$, we have

$$|(L_nf)(x)-f(x)| \le |(L_nf)(x)-(L_nz)(x)| + |(L_nz)(x)-z(x)| + |z(x)-f(x)| \le$$

$$\le 2\|f-z\|_{C_B} + \frac{1}{n}\left\{x+\frac{1}{2}\left[(\lambda+1)(\lambda+2)+(2\lambda+3)\frac{g'(1)}{g(1)}+\frac{g''(1)+g'(1)}{g(1)}\right]\right\}\|z\|_{C_B^2},$$

Because the left side of this inequality does not depend of the function $z \in C_B^2$, it result that

$$|(L_nf)(x)-f(x)| \le 2K(f;A(x,n)),$$

where

$$A(x,n) = \frac{1}{2n}\left\{x+\frac{1}{2}\left[(\lambda+1)(\lambda+2)+(2\lambda+3)\frac{g'(1)}{g(1)}+\frac{g''(1)+g'(1)}{g(1)}\right]\right\}$$

By making use (9), we obtain

$$|(L_nf)(x)-f(x)| \le 2A\left\{\omega_2(f;\sqrt{A(x,n)}\,)+\min(1,A(x,n))\|f\|_{C_B}\right\} =$$

$$= 2A\left(\omega_2(f;h)+\min(1,h^2)\|f\|_{C_B}\right)$$

Remark. For $g(z) = 1$ and $\lambda = 0$, we have $A(x,n) = \dfrac{x+1}{2n}$ and we obtain a result due to S.P. Singh and M.K. Tiwari [8]:

$$|(S_n^*f)(x)-f(x)| \le 2A\left\{\omega_2\left(f;\sqrt{\frac{x+1}{2n}}\right)+\min\left(1,\frac{x+1}{2n}\right)\|f\|_{C_B}\right\}$$

REFERENCES

1. M.M. Derriennic, *Sur l'approximation des functions integrables sur* [0,1] *par des polynomes de Bernstein modifiés*, J. Approx. Theory, 31(1981), 325-343.
2. J.L. Durrmeyer, *Une formule d'inversion de la transformée de Laplace: Applications á la theorie des moments*, These de 3e cycle, Faculté de Sciences de l'Université de Paris, 1967.
3. I. Gavrea, I. Raşa, *Remarks on some quantitative Korovkin type results* (to appear).
4. A. Jakimovski, D. Leviatan, *Generalized Szasz operators for the approximation in infinite interval*, Mathematica (Cluj), 34(1969), 97-103.
5. V.V. Juk, *On S.N. Bernstein polynomials and the Lip1 class functions* (russian), Vestnik Leningradskovo

Universiteta, Seria 1, Matematika, Astronomia, no. 1, 1989, 25-30.
6. A. Lupaş, *Die Folge der Betaoperatoren*, Universität Stuttgart, Doktor Dissertation, 1972.
7. S.M. Mazhar, V. Totik, *Approximation by modified Szasz operators*, Acta Sci. Math, 49(1985), 257-269.
8. S.P. Singh, M.K. Tiwari, *Quantitative estimates on the integral Szasz operators in the different normed spaces*, Proc. Math. Soc., B.H.U., Vol. 2(1986), 133-144.
9. B. Wood, *Generalized Szasz operators for the approximation in the complex domain*, SIAM J. Appl. Math., 17, Nr. 4, 1969, 790-801.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

COMMON FIXED POINTS OF COMPATIBLE MAPPINGS OF TYPE (A)

Valeriu POPA* and Hemant Kumar PATHAK**

Received: May 31, 1994
AMS subject classification: 54H25

REZUMAT. - Puncte fixe comune pentru funcţii compatibile de tipul (A). În această lucrare vom da unele teoreme de punct fix pentru funcţii compatibile de tipul (A) extinzând unele rezultate din [4]-[7].

Abstract. In this paper, we give some common fixed point theorems for compatible mappings of type (A) extinding some results from [4] - [7].

Rhoades [8] summarized contractive mappings of some types and discussed on fixed points. Wang, Li Gao and Iseki [10] proved some fixed point theorems on expansion mappings, which correspond some contractive mappings. Rhoades [9] generalized the results for pairs of mappings. Recently, Popa [4] -[7] proved some theorems on unique fixed point for expansion mappings.

The purpose of this paper is to prove some fixed point theorems on expansion mappings extending some results from [4], [5], [6] and [7] for compatible mappings of type (A).

Let R_+ be the set all non-negative reals numbers and $\psi: R_+^3 \rightarrow R_+$ be a real function. Throughout this paper, (X,d) denotes a metric space.

* *University of Bacău, Department of Mathematics, 5500 Bacău, Romania*

** *Kalyan Mahavidyalaya, Department of Mathematics, Bhilai Nagar (M.P.), 490006 India*

DEFINITION 1. $\psi: R_+^3 \to R_+$ satisfies property (h) if there exists $h \geq 1$ such that for every $u,v \in R_+$ with $u \geq \psi(v,u,v)$ or $u \geq \psi(v,v,u)$, we have $u \geq hv$.

DEFINITION 2 ([6]). $\psi: R_+^3 \to R_+$ satisfies property (u) if $\psi(u,0,0) > 0$, $u > 0$.

DEFINITION 3 ([1]). Let S,T: $(X,d) \to (X,d)$ be mappings, S and T are said to be compatible if

$$\lim_{n\to\infty} d(STx_n, TSx_n) = 0$$

whenever $\{x_n\}$ is a sequence in X such that $\lim_{n\to\infty} Sx_n = \lim_{n\to\infty} Tx_n = t$ for some t in X.

DEFINITION 4 ([2]). Let S,T: $(X,d) \to (X,d)$ be mappings, S and T are said to be compatible of type (A) if

$$\lim_{n\to\infty} d(TSx_n, SSx_n) = 0 \text{ and } \lim_{n\to\infty} d(STx_n, TTx_n) = 0$$

whenever $\{x_n\}$ is a sequence in X such that $\lim_{n\to\infty} Sx_n = \lim_{n\to\infty} Tx_n = t$ for some t in X.

Remark. By ex. 2.1 of [2], it follows that the notions of "compatible mappings" and "compatible mappings of type (A)" are independent.

LEMMA 1 ([2)]. *Let S, T: $(X,d) \to (X,d)$ be compatible mappings of type* (A). *If one of S and T is continuous, then S and T are compatible.*

LEMMA 2 ([1]). *Let S and T be compatible mappings from a metric space (X,d) into itself. Suppose that $\lim_{n\to\infty} Sx_n = \lim_{n\to\infty} Tx_n = t$ for some t in X. Then $\lim_{n\to\infty} TSx_n = St$ if S is continuous.*

LEMMA 3 ([2]). *Let S, T: $(X,d) \to (X,d)$ be mappings. If S and T are compatible of type* (A) *and $S(t) = T(t)$ for some $t \in X$, then $ST(t) = TT(t) = SS(t) = TS(t)$.*

LEMMA 4 ([7]). *Let (X,d) be a metric space, A, B, S, T four mappings of X satisfying the inequality*

$$d(Ax, By) \geq \psi(d(Sx, Ty), d(Ax, Sx), d(By, Ty)) \tag{1}$$

for all x,y in X, where ψ *satisfies property* (u). *Then A, B, S, T have at most one common fixed point.*

THEOREM 1. *Let A, B, S and T be mappings from a complete metric space* (X,d) *into itself satisfying the conditions:*

(1°) *A and B are surjective,*

(2°) *One of A, B, S, T is continuous,*

(3°) *A and S as well B and T are compatible of type* (*A*),

(4°) *The inequality* (1) *holds for all x,y in X, where* ψ *satisfied property* (h) *with* $h > 1$.

If property (u) *holds and* ψ *is continuous, then A, B, S and T have a unique common fixed point.*

Proof. Let $x_0 \in X$ be arbitrary. By (1°) we choose a point x_1 in X such that $Ax_1 = Tx_0 = y_0$ and for this point x_1, there exists a point x_2 in X such that $Bx_2 = Sx_1 = y_1$. Inductively, we can define a sequence $\{y_n\}$ in X such that

$$Ax_{2n+1} = Tx_{2n} = y_{2n} \text{ and } Bx_{2n+2} = Sx_{2n+1} = y_{2n+1}. \tag{2}$$

By (1) and (2) we have

$$d(y_0,y_1) = d(Ax_1,Bx_2) \geq \psi(d(Sx_1,Tx_2),d(Sx_1,Ax_1),d(Tx_2,Bx_2))$$
$$= \psi(d(y_1,y_2),d(y_1,y_0),d(y_2,y_1)).$$

Then by property (h), we have

$$d(y_0,y_1) \geq h \cdot d(y_2,y_1), \text{ where } h > 1.$$

Thus $d(y_2,y_1) \leq \frac{1}{h} d(y_0,y_1)$. Similarly, we have

$$d(y_n,y_{n+1}) \leq \left(\frac{1}{h}\right)^n \cdot d(y_0,y_1).$$

Then by a routine calculation we can show that $\{y_n\}$ is a Cauchy sequence and since X is complete, there is a $z \in X$ such that $\lim y_n = z$. Consequently, the subsequences $\{Ax_{2n+1}\}$, $\{Bx_{2n}\}$, $\{Sx_{2n+1}\}$ and $\{Tx_{2n}\}$ converges to z.

Now, suppose that A is continuous. Since A and S are compatible of type (A) and A is continuous by Lemma 1 A and S are compatible. Lemma 2 implies $A^2x_{2n+1} \to Az$ and $SAx_{2n+1} \to Az$ as $n \to \infty$. By (1), we have

$$d(A^2x_{2n+1}, Bx_{2n}) \geq \psi(d(SAx_{2n+1}, Tx_{2n}), d(SAx_{2n+1}, A^2x_{2n+1}), d(Tx_{2n}, Bx_{2n})).$$

Letting n tend to infinity we have by continuity of ψ

$$d(Az, z) \geq \psi(d(Az, z), 0, 0).$$

By property (u) follows $d(Az,z) > d(Az,z)$ if $Az \neq z$. Thus $z = Az$. By (1) we have

$$d(Az, Bx_{2n}) \geq \psi(d(Sx, Tx_{2n}), d(Sz, Az), d(Tx_{2n}, Bx_{2n})).$$

Letting n tend to infinity we have by continuity of ψ

$$0 \geq d(Az, z) \geq \psi(d(Sz, z), d(Sz, z), 0).$$

By definition (1) we have $0 \geq h\, d(Sz,z)$ which implies $z = Sz$. Let $z = Bu$ for some $u \in X$. Then we have by (1)

$$d(A^2x_{2n+1}, Bu) \geq \psi(d(SAx_{2n+1}, Tu), d(SAx_{2n+1}, A^2x_{2n+1}), d(Tu, Bu)).$$

Letting n tend to infinity we have by continuity of ψ

$$0 = d(Az, Bu) \geq \psi(d(Az, Tu), 0, d(Tu, Bu)) = \psi(d(z, Tu), 0, d(z, Tu)).$$

By definition (1) we have $0 \geq h \cdot d(z, Tu)$ which implies $z = Tu$. Since B and T are compatible of type (A) and $Bu = Tu = z$ Lemma 3 $Bz = BTu = TBu = Tz$, moreover by (1), we have

$$d(Ax_{2n+1}, Bz) \geq \psi(d(Sx_{2n+1}, Tz), d(Sx_{2n+1}, Ax_{2n+1}), d(Tz, Bz)).$$

Letting n tend to infinity we have by continuity of ψ

$$d(z, Tz) \geq \psi(d(z, Tz), 0, 0).$$

From property (u) it follows that $d(z, Tz) > d(z, Tz)$ if $z \neq Tz$. Thus $z = Tz$. Therefore, z is a common fixed point of A, B, S, T. Similarly, we can complete the proof in the case of the continuity of B.

Next, suppose that S is continuous. Since A and S are compatibly of type (A) and S is continuous by Lemma 1 A and S are compatible. Lemma 2 implies $S^2x_{2n+1} \to Sz$ and $ASx_{2n+1} \to Sz$ as $n \to \infty$. By (1), we have

$$d(ASx_{2n+1}, Bx_{2n}) \geq \psi(d(S^2x_{2n+1}, Tx_{2n}), d(S^2x_{2n+1}, ASx_{2n+1}), d(Tx_{2n}, Bx_{2n})).$$

Letting n tend to infinity we have by continuity of ψ

$$d(Sz, z) \geq \psi(d(Sz, z), 0, 0).$$

By property (u) we have $d(Sz, z) > d(Sz, z)$ if $z \neq Sz$. Thus $z = Sz$. Let $z = Av$ and $z = Bw$ for some v and w in X, respectively. Then by (1) we have

$$d(ASx_{2n+1}, Bw) \geq \psi(d(S^2x_{2n+1}, Tw), d(S^2x_{2n+1}, ASx_{2n+1}), d(Tw, Bw))$$

Letting n tend to infinity we have by continuity of ψ

$$0 = d(Sz, z) \geq \psi(d(Sz, Tw), 0, d(Bw, Tw)) = \psi(d(z, Tw), 0, d(z, Tw)).$$

By Definition (1) we have $0 \geq h \cdot d(z, Tw)$ which implies $z = Tw$. Since B and T are compatible of type (A) and $Bw = Tw = Tz$ by Lemma 3 $Bz = BTw = TBw = Tz$. Moreover, by (1), we have

$$d(Ax_{2n+1}, Bz) \geq \psi(d(Sx_{2n+1}, Tz), d(Sx_{2n+1}, Ax_{2n+1}), d(Bz, Tz)).$$

Letting n tend to infinity we have by continuity of ψ

$$d(z, Tz) = d(z, Bz) \geq \psi(d(z, Tz), 0, 0).$$

By property (u), it follows that $d(z, Tz) > d(z, Tz)$ if $z \neq Tz$. Thus $z = Tz$. Further, we have by (1)

$$d(Av, Bz) \geq \psi(d(Sv, Tz), d(Av, Sv), d(Tz, Bz)) \text{ and}$$

$0 = d(z,z) \geq \psi(d(Sv,z), d(z,Sv), 0)$. By Definition 1 we have $0 \geq h \cdot d(Sv,z)$ and thus $Sv = z$. Since A and S are compatible of type (A) and $Av = Sv = z$ by Lemma 3 $Az = ASv = SAv = Sz$. Therefore, z is a common fixed point of A, B, S and T. Similarly, we can complete the proof in the case of continuity of T.

From Lemma 4 it follows that z is the unique common fixed point of A, B, S and T.

DEFINITION 5 ([3]). $\psi: R_+^3 \to R_+$ satisfies property (B) if for every $u,v \in R_+$ such $u \geq \psi(v,u,v)$ we have $u \geq hv$, where $\psi(1,1,1) = h \geq 1$.

DEFINITION 6 ([7]). $\psi: R_+^3 \to R_+$ satisfies property (B^*) if for every $u,v \in R_+$ such that $u \geq \psi(v,v,u)$, we have $u \geq hv$, where $\psi(1,1,1) = h \geq 1$.

COROLLARY 1. *Let A,B,S and T be mappings from a complete metric space (X,d) into itself satisfying:*

(1) *the conditions* (1°), (2°), (3°) *of Theorem* 1,

(2) *The inequality* (1) *holds for all x,y in X where ψ satisfies property* (B) *and* (B^*) *with $h \geq 1$.*

If property (u) *holds and ψ is continuous, then A,B,S and T have a unique common fixed point.*

THEOREM 2. *Let A,B,S and T be mappings from a complete metric space (X,d) into itself satisfying conditions* (1°), (2°) *and* (3°) *of Theorem* 1. *If there exist non negative reals a,b,c,d with $a+b+c+d > 1$ such that*

$$d^k(Ax,By) \geq a \cdot d^k(Sx,Ty) + b \cdot d^m(Ax,Sx) \cdot d^{k-m}(By,Ty) + c \cdot d^{k-p}(Sx,Ty) \cdot d^p(Ax,Sx) + d \cdot d^q(By,Ty) \cdot d^{k-q}(Sx,Ty) \tag{3}$$

where $k \geq 1$, $q \geq 0$, $m \geq 0$, $p \geq 0$ and $q \leq k$, $p \leq k$, $m \leq k$ hold for all x and y in X, then A,B,S and T have a common unique fixed point if $a > 1$.

Proof. Let

$$\psi(t_1,t_2,t_3) = \left[a\cdot t_1^k + b\cdot t_2^m\cdot t_3^{k-m} + c\cdot t_2^p\cdot t_1^{k-p} + d\cdot t_3^q\cdot t_1^{k-q}\right]^{1/k}.$$

Let u,v such that $u \geq \psi(v,u,v)$, then

$$a \geq \left[a, v^k + bu^m v^{k-m+\text{cup},\,k-p} + dv^k\right]^{1/k} \text{ and}$$

$$a_k \geq av^k + bu^m v^{k-m} + cu^p\cdot v^{k-p} + dv^k.$$

Thus $(a+d)\cdot t^k + b\cdot t^m + c\cdot t^p - 1 \leq 0$ where $t = v/u$.

Let $g_1(t)$: $[0,\infty) \to R$ be the function $g_1(t) = (a+d)t^k + bt^m + ct^p - 1$. Then $g_1'(t) > 0$ for $t > 0$, $g_1(0) < 0$ and $g_1(1) = a+b+c+d-1 > 0$. Let $r_1 \in (0,1)$ be the root of the equation $g_1(t) = 0$, then $g_1(t) < 0$ for $t < r_1$. Let u,v be such that $u \geq \psi(v,v,u)$, then

$$u \geq \left[av^k + bv^{k-m}u^m + cv^k + du^{q,\,k-q}\right]^{1/k}.$$

Similarly, we have

$$(a+c)t^k + bt^{k-m} + dt^{k-q} - 1 \leq 0$$

where $t = v/u$. Let g_2: $[0,\infty) \to R$ be the function $g_2(t) = (a+c)t^k + b\cdot t^{k-m} + dt^{k-q} - 1$. Let $r_2 \in (0,1)$ the root of the equation $g_2(t) = 0$, then $g_2(t) < 0$ for $t < r_2$. Thus $g_1(t) < 0$ and $g_2(t) < 0$ for $t < \min\{r_1,r_2\} = r$, $r \in (0,1)$. Then $(v/u) < r$ and $u > (1/r)v$. Thus $h = (1/r) > 1$ and $u \geq hv$ with $h > 1$.

On the other hand we have $\psi(u,0,0) = a^{1/k}u > u$. By Theorem 1, it follows that A,B,S and T have a unique common fixed point.

COROLLARY 2 ([4]). *Let (X,d) be a complete metric space and f: $(X,d) \to (X,d)$ a surjective mapping. If there exist non-negative reals a,b,c,d with $a+b+c+d > 1$ such that*

$$d^k(fx,fy) \geq a\cdot d^q(x,fx)\cdot d^{k-q}(x,y) + b\cdot d^m(y,fy)\cdot d^{k-m}(x,y) +$$
$$c\cdot d^p(x,fx)\cdot d^{k-p}(y,fy) + d\cdot d^k(y), \tag{4}$$

where $k \geq 1$, $q \geq 0$, $m \geq 0$, $p \geq 0$ and $q \leq k$, $m \leq k$, $p \leq k$ for each x,y in X with $x \neq y$, and *if $d > 1$, then f has a unique fixed point.*

CORROLARY 3 ([5]). *Let (X,d) be a complete metric space and $f: (X,d) \to (X,d)$ a surjective mapping. If there exist non-negative a,b,c with $a < 1$ and $c > 1$, then f has a unique fixed point.*

THEOREM 3. *Let S, T and $\{f_i\}_{i \in N}$ be mappings from a complete metric space (X,d) into itself satisfying the conditions:*

(1°) *$\{f_i\}_{i \in N}$ are surjective,*

(2°) *S or T or f_1 is continuous,*

(3°) *S and $\{f_i\}_{i \in N}$ are compatible of type (A) and T and $\{f_i\}_{i \in N}$ are compatible of type (A).*

(4°) *The inequality*

$$d(f_i x, f_{i+1} y) \geq \psi(d(Sx, Ty), d(f_i x, Sx), d(f_{i+1} y, Ty)) \tag{5}$$

hold for all x and y in X, $\forall\ i \in N$, where ψ is continuous, satisfies property (h) with $h > 1$ and property (u), then $\{f_i\}_{i \in N}$, A and B have a unique common fixed point.

Proof. It is similar to the proof of [7, Theorem 4].

COROLLARY 4. *Let S, T and $\{f_i\}_{i \in N}$ be mappings from a complete metric space (X,d) into itself satisfying the conditions (1°), (2°), (3°) of Theorem 3 and*

$$d^k(f_i x, f_{i+1} y) \geq a \cdot d^k(Sx, Ty) + b \cdot d^k(f_i x, Sx) + c \cdot d^k(f_{i+1} y, Ty), \tag{6}$$

where $k \geq 1$, $0 \leq b$, $c < 1$, $a > 1$ hold for all x and y in X, $\forall\ i \in N$, then S, T and $\{f_i\}_{i \in N}$ have a unique common fixed point.

We conclude this paper with the following example, which shows that "surjectivity of A and B" is a necessary condition in Theorem 1.

Example 1. Let $X = [0,\infty)$. Define A,S,B and T: $X \to X$ given by $Ax = kx + 1$, Sx

$x + 1$, $Bx = Tx = 1$ for x in X and $2 \geq k > 1$. Note that the following mapping satisfies properties (h) and (u):

$$\psi(t_1, t_2, t_3) = k \cdot \max\{t_1, t_2, t_3\}, \text{ where } k > 1.$$

Now, $d(Ax, By) = kx = k \cdot \max\{x, (k-1)x, 0\} = k \cdot \max\{d(Sx, Ty), d(Ax, Sx), d(By, Ty)\} = \psi(d(Sx, Ty), d(Ax, Sx), d(By, Ty))$, for all x,y in X, where $2 \geq k > 1$. Consider a sequence $\{x_n\} \subset X$ such that $x_n \to 0$. Then it is to see, by routine calculation, that A,S and B,T are compatible of type (A). Moreover, A,B,S and T are all continuous. Therefore, we see that all the hypothesis of Theorem 1 are satisfied except surjectivity of A and B, but the mappings A,B,S and T have no fixed point in X.

REFERENCES

1. Jungck, G., *Compatible mappings and fixed points*, J. Math. and Math. Sci. 9(1986), 771-779.
2. Jungck, G., Murthy, P.P. and Cho. J.J., *Compatible mappings of type* (A) *and common fixed points*, Math. Japonica, **38**, 2(1993), 381-390.
3. Khan M.A., Khan, S. and Sessa, S., *Some theorems on expansion mapping and their fixed points*, Demonstratio Mathematica 19(1986), 673-683.
4. Popa, V. *Some fixed point theorems of expansion mappings*, Demonstratio Mathematica 19, 3(1986), 699-702.
5. Popa, V., *Fixed point theorems for expansion mappings*, Babeş-Bolyai Univ. Fac. Math. Phys., Research Semin., Seminar on fixed point theory, Preprint nr. 3(1987), 25-30.
6. Popa, V., *Theorems of unique fixed point for expansion mappings*, Demonstratio Mathematica 23(1990), 213-218.
7. Popa, V., *Common fixed points of compatible mappings*, Demonstratio Mathematica **26**, 3-4(1993), 803-809.
8. Rhoades, B.E., *A comparison of various definitions of contractive mappings*, Trans. Amer. Math. Soc. 226(1977), 257-290.
9. Rhoades, B.E., *Some fixed point theorems for pairs of mappings*, Jñānābh 15(1985), 151-158.
10. Wang, S.Z., Li, B.Y., Gao, Z.M. and Iséki, K., *Some fixed point theorems on expansion mappings*, math. Japonica 29(1984), 631-636.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

NOTE ABOUT A METHOD FOR SOLVING NONLINEAR SYSTEM OF EQUATIONS IN FINIT DIMENSIONAL SPACES

Béla FINTA*

Received: December 18, 1994
AMS subject classification: 65H10

REZUMAT. - O metodă de rezolvare a sistemelor de ecuaţii neliniare în spaţii finit dimensionale. În această lucrare se aplică ideea lui Seidel şi metoda SOR pentru forma iterativă a unui sistem de ecuaţii neliniare şi se dau condiţii suficiente, care asigură convergenţa şirului iterativ.

It is well known to obtain solutions for a linear and nonlinear system of equations one kind are the iterative methods (see [1] pages 177-188, [2] pages 40-49 and 127-166, or [3] pages 82-106 and 322-363). For the linear case the simplest method of such type is known as Jacobi's method. For the nonlinear case this method appears, too as Jacobi's theorem.

Let us consider the function $f: D \subset R^n \to R^n$, where $D \neq \emptyset$, and let us transform the system of equations $f(x) = \theta_{R^n}$ in the iterative form $x = \phi(x)$ (see [4] pages 21-22).

THEOREM (Jacobi). *Let us suppose that D' is a domain, and $\phi : D' \subset R^n \to R^n$ a Fréchet differentiable map. If $A \subset D'$ is a closed convex subset such that $\phi(A) \subset A$, and there exists $\alpha \in (0,1)$ with property: $\sum_{j=1}^{n} \left| \frac{\partial \phi_i}{\partial x_j}(x) \right| \leq \alpha$ for every $x \in A$ and for $i = \overline{1,n}$, then the system $\phi(x) = x$ has a unique solution in A* (see [5] page 81).

For the linear system of equations transformed in the iterative form there exist other methods, which increase the rapidity of convergence for the iterative sequence obtained by the Jacobi's iterative method, like the Gauss-Seidel, and more, the successive overrelaxation

* *Technical University, Str. N. Iorga, 1, Târgu-Mureş, Romania*

methods.

In the case of nonlinear systems of the form $f(x) = \theta_{R^n}$, it is known the so called Seidel-SOR method, which combines Seidel's idea with the successive overrelaxation. The existing result in this direction is the following:

Let us consider the function $f: D \subset R^n \to R^n$, $f = (f_1, \ldots, f_n)$. We suppose it is known the k-th term x^k of the iterative sequence, and we want to find x^{k+1}. If we suppose that the first i-1 components of x^{k+1} are determined, than let's consider x_i' to be the solution of the equation: $f_i(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \ldots, x_n^k) = 0$. We can calculate this approximative value x_i' by one of the methods of solving nonlinear equations in one variable.

Then we obtain the i-th component like: $x_i^{k+1} = x_i^k + \omega \cdot (x_i' - x_k')$, where $\omega \in R^*$ is a factor of relaxation. We consider the decomposition $f(x) = D(x) - L(x) - U(x)$ for the Jacobian of f, where $D(x)$ is the diagonal matrix formed by the diagonal elements, $L(x)$ is the lower triangular matrix, and $U(x)$ is the upper triangular matrix. We note:

$$B(x) = \omega^{-1} \cdot [D(x) - \omega \cdot L(x)],$$

$$C(x) = \omega^{-1} \cdot [(1 - \omega) \cdot D(x) + \omega \cdot U(x)],\ H(x) = B^{-1} C.$$

Now we are ready to announce the theorem obtained by the local linearization around the solution x^* of the nonlinear system of equations:

THEOREM (Seidel-SOR). *Let us consider the function* $f: D \subset R^n \to R^n$ *and let us suppose that* x^* *is a solution of the system* $f(x) = \theta_{R^n}$. *If we suppose the following conditions hold:* i) f *is continuously differentiable in a neighborhood* $V(x^*) \subset D$ *of* x^*, ii) $D(x^*)$ *is not singular,* iii) *the spectral radius* $\rho(H(x^*)) < 1$, *then there exists a sphere* $S(x^*, r) = \{x \in R^n \mid \|x - x^*\| \le r\} \subset V(x^*)$ *such that for every* $x \in S(x^*, r)$ *the iterative sequence* $\{x^k\}_{k \in N}$ *generated by the Seidel-SOR method is unique defined and converges to* x^*, *i.e.* x^* *is an attractive point for f* (see [6] pages 89-95).

The purpose of this work is to apply the Siedel's idea and the SOR method for the iterative form $\phi(x) = x$ of the nonlinear system of equations and to find sufficient conditions which assure us the convergence of the iterative sequence. First we rewrite the system $f(x) = \theta_{R^n}$ in the nonlinear relaxation form:

We suppose that we can form the function $\phi^*: D' \subset R^n \to R^n$ in the following way:

$$\phi_1^*(x) = x_1 + \omega\cdot(\phi_1(x) - x_1),$$

$$\phi_2^*(x) = x_2 + \omega\cdot(\phi_2(\phi_1^*(x), x_2, \dots, x_n) - x_2), \dots,$$

$$\phi_i^*(x) = x_i + \omega\cdot(\phi_i(\phi_1^*(x), \dots, \phi_{i-1}^*(x), x_i, \dots, x_n) - x_i), \dots,$$

$$\phi_n^*(x) = x_n + \omega\cdot(\phi_n(\phi_1^*(x), \dots, \phi_{n-1}^*(x), x_n) - x_n)$$

for every $x \in D'$, where $\omega \in R^*$ is the factor of relaxation. For $A \neq \emptyset$, $A \subset D'$ closed, convex set, let us consider the numbers:

for $i = \overline{1,n}$:

$$a_{ii} = \sup\left\{\left|1 - \omega + \omega\cdot\frac{\partial \phi_i}{\partial x_i}(x)\right| \;\middle|\; x \in A\right\},$$

for $i,j = \overline{1,n}$ and $i \neq j$

$$a_{ij} = \sup\left\{\left|\omega\cdot\frac{\partial \phi_i}{\partial x_j}(x)\right| \;\middle|\; x \in A\right\}.$$

For $i = \overline{1,n}$ we generate the numbers:

$$M_i = \sum_{j=1}^{i-1} a_{ij}\cdot M_j + \sum_{j=i}^{n} a_{ij},$$

with convention: $\sum_{j=1}^{0} a_{1j}\cdot M_j = 0$.

THEOREM 1. *If we suppose that:* i) $\phi: D' \subset R^n \to R^n$ *is a Fréchet differentiable function,* ii) *there exists a closed convex subset* $A \neq \emptyset$, $A \subset D'$ *such that* $\phi^*(A) \subset A$, iii) *for this set* A, $\max\{M_i | i = \overline{1,n}\} < 1$, *then for every* $x^0 \in A$ *the iterative sequence* $x^{k+1} = \phi^*(x^k)$ *exists and converges to the unique fixed point of the function* ϕ.

Proof. We consider the norm:

$$\|x\|_\infty = \max\{|x_i| \mid i = \overline{1,n}\}$$

on R^n, and we apply the Banach fixed point theorem to the function ϕ^*. For every $i = \overline{1,n}$ we obtain:

$$|\phi_i^*(y) - \phi_i^*(x)| =$$

$$= |y_i - x_i + \omega\cdot\{\phi_i(\phi_1^*(y),\dots,\phi_{i-1}^*(y),y_i,\dots,y_n) - \phi_i(\phi_1^*(x),\dots,\phi_{i-1}^*(x),x_i,\dots,x_n) - (y_i - x_i)\}| =$$

$$= |(y_i - x_i)(1-\omega) + \omega\cdot\{\phi_i(\phi_1^*(y),\dots,\phi_{i-1}^*(y),y_i,\dots,y_n) - \phi_i(\phi_1^*(x),\dots,\phi_{i-1}^*(x),x_i,\dots,x_n)\}| =$$

$$= |(y_i - x_i)(1-\omega) + \omega\cdot d\phi_i(u)|,$$

where

$$u = (\phi_1^*(x),\dots,\phi_{i-1}^*(x),x_i,\dots,x_n) + \xi\cdot(\phi_1^*(y) - \phi_1^*(x),\dots,\phi_{i-1}^*(y) - \phi_{i-1}^*(x), y_i - x_i,\dots,y_n - x_n)),$$

with $\xi \in (0,1)$, and

$$|(y_i - x_i)(1-\omega) + \omega\sum_{j=1}^{i-1}\frac{\partial\phi_i}{\partial x_j}(u)(\phi_j^*(y) - \phi_j^*(x)) + \omega\sum_{j=i}^{n}\frac{\partial\phi_i}{\partial x_j}(u)(y_j - x_j)| \le$$

$$\le \sum_{j=1}^{i-1}|\omega\cdot\frac{\partial\phi_i}{\partial x_j}(u)|\cdot|\phi_j^*(y) - \phi_j^*(x)| + |(1-\omega) + \omega\frac{\partial\phi_i}{\partial x_j}(u)|\cdot|y_i - x_i| +$$

$$+ \sum_{j=i+1}^{n}|\omega\frac{\partial\phi_i}{\partial x_j}(u)|\;|(y_i - x_j)| \le$$

$$\le (\sum_{j=1}^{i-1}a_{ij}\cdot M_j + a_{ii} + \sum_{j=i+1}^{n}a_{ij})\|y - x\|_\infty$$

Consequently

$$\|\phi^*(y) - \phi^*(x)\|_\infty = \max\{|\phi_i^*(y) - \phi_i^*(x)\| \, i = \overline{1,n}\} \le$$

$$\le \max\{M_i \mid i = \overline{1,n}\}\cdot\|y - x\|_\infty,$$

with $\max\{M_i \mid i = \overline{1,n}\} < 1$. So ϕ^* is a contraction and we can easily see that the fixed point of ϕ^* will be a fixed point for ϕ, too.

For $\omega = 1$ we obtain the Seidel's method for the system of nonlinear equations in the iterative form. In this case we define the function $\phi^*: D' \subset R^n \to R^n$ in the following way:

$$\phi_1^*(x) = \phi_1(x),$$

$$\phi_2^*(x) = \phi_2(\phi_1^*(x), x_2, \ldots, x_n), \ldots,$$

$$\phi_i^*(x) = \phi_i(\phi_1^*(x), \ldots, \phi_{i-1}^*(x), x_i, \ldots, x_n), \ldots,$$

$$\phi_n^*(x) = \phi_n(\phi_1^*(x), \ldots, \phi_{n-1}^*(x), x_n),$$

and for $A \neq \emptyset$, $A \subset D'$ closed, convex set, we consider the numbers:

$$a_{ij} = \sup\left\{ \left| \frac{\partial \phi_i}{\partial x_j}(x) \right| \,\middle|\, x \in A \right\}, \text{ for } i,j = \overline{1,n}$$

and we generate the numbers:

$$M_i = \sum_{j=1}^{i-1} a_{ij} \cdot M_j + \sum_{j=i}^{n} a_{ij}, \text{ for } i = \overline{1,n}$$

with convention: $\sum_{j=1}^{0} a_{1j} \cdot M_j = 0$.

THEOREM 2. *If we suppose that:* i) $\phi: D' \subset R^n \to R^n$ *is a Fréchet differentiable function,* ii) *there exists a closed convex subset* $A \neq \emptyset, A \subset D'$ *such that* $\phi^*(A) \subset A$, iii) *for this set* A, $\max\{M_i | i = \overline{1,n}\} < 1$, *then for every* $x^0 \in A$ *the iterative sequence* $x^{k+1} = \phi^*(x^k)$ *exists and converges to the unique fixed point of the function* ϕ.

Example. We solve the following nonlinear system of equations:

$$7\sin x = x^2 + yz + \cos z$$

$$9\sin y = xz^2 + y\cos(xyz) + 1$$

$$8\sin z = x\sin z^2 + y^2\cos(xy)$$

for $x,y,z \in [-1,1]$. We transform the system in the following iterative form:

$$x = \arcsin[(x^2 + yz + \cos z)|7]$$

$$y = \arcsin[(xz^2 + y\cos(xyz) + 1)|9]$$

$$z = \arcsin[(x\sin z^2 + y^2\cos(xy))|8]$$

and we solve it using Jacobi's theorem, Theorem 2 and Theorem 1 with $\omega = 1.1$. If we consider the initial point $x^0 = (0.5, 0.5, 0.5)$ then we obtain the solution with accurate to two, three, four, five decimal places by making 4, 3, 3; and 5, 4, 4; and 6, 5, 5; and 7, 6, 5 iterations, respectively.

Remark. We can obtain theorems like Jacobi's theorem, Theorem 1 and Theorem 2 by using other norms on R^n. One problem is to find such a norm, for that the conditions on the system are larger.

REFERENCES

1. I.Gh. Şabac etc. *Matematici speciale*, Ed. Didactică şi Pedagogică, Bucureşti, 1983.
2. Popper György, Csizmás Ferenc, *Numerikus módszerek mérnököknek*, Akadémiai Kiadó, Typotex, Budapest, 1993.
3. Stoyan Gisbert, Takó Galina, *Numerikus módszerek*, ELTE, Typotex, Budapest, 1993.
4. Eberhard Zeidler, *Nonlinear functional analysis and its application*, Springer Verlag, Berlin, Tokio, 1985.
5. S. Rădulescu, M. Rădulescu, *Teoreme şi probleme de analiză matametică*, Ed. Didactică şi Pedagogică, Bucureşti., 1982.
6. Şt. Măruşter, *Metode numerice în rezolvarea ecuaţiilor neliniare*, Ed. Tehnică, Bucureşti, 1981.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

A NUMERICAL SOLUTION OF THE DIFFERENTIAL EQUATION OF m-TH ORDER USING SPLINE FUNCTIONS

Adrian REVNIC*

Received: August 27, 1993
AMS subject classification: 65L10, 65Q05

REZUMAT. - O soluţie numerică pentru ecuaţia diferenţială de ordinul *m* folosind funcţii spline. Se construieşte un procedeu numeric folosind funcţii spline polinomiale pentru rezolvarea unei clase de ecuaţii diferenţiale neliniare de ordin m cu condiţii iniţiale. Se estimează eroarea şi se investighează stabilitatea metodei propuse.

I. Introduction. In the last years, the problem of approximating the solution of non linear differential equations by spline functions has been of growing interest. Many authors [1]-[6] have proposed various methods to approximate the solution by means of spline.

Recently, J. Gyorvari and Cs. Mihalyko [3] gave a spline algorithm to solve numerically a differential equation with initial conditions. In this paper, using the idea of *T.* Fawzy in [1], [2] an improved algorithm is constructed using spline functions and in addition, the stability of the proposed method is given.

Consider the differential equation with initial condition

$$z^{(m)}(x) = f(x, z(x), z'(x), \dots, z^{(m-1)}(x)),\ x \in [0,b],\ b > 0 \tag{1.1}$$

$$z^{(j)}(0) = z_0^{(j)}, \quad j = \overline{0, m-1}$$

where $f \in C^r([0, b] \times \mathbf{R}^r)$ and $r \in N$.

We assume that f satisfies the following Lipschitz conditions

$$|f^{(q)}(x,u) - f^{(q)}(x,v)| \le L_f \|u - v\| \tag{1.2}$$

* "Babeş-Bolyai" University, Faculty of Mathematics and Computer Science, 3400 Cluj-Napoca, Romania

$x \in [0,b], \ u,v \in \mathbb{R}^r, \ q = \overline{0,r}$

The differential equation (1.1) can be reduced to a system of m differential equations of first degree as follows:

One denote: $y_0(x) := z(x), y_1(x) := z'(x), \dots, y_{m-1}(x) := z^{(m-1)}(x)$

Then (1.1) is equivalent to

$$y'(x) = F(x,y(x)), \quad x \in [0,b] \tag{1.3}$$

$$y = (y_0, \dots, y_{m-1}) : [0,b] \to \mathbb{R} \text{ and}$$

$$F(x,y(x)) = (y_1(x), \dots, y_{m-1}(x), f(x,y(x))).$$

One have $F^{(q)}(x,y(x)) = (y_1^{(q)}(x), \dots, y_{m-1}^{(q)}(x), f^{(q)}(x,y(x)))$ so the Lipschitz conditions for f holds for F too:

$$\| F^{(q)}(x,u) - F^{(q)}(x,v) \| \le L \| u - v \| \tag{1.4}$$

$x \in [0,b], \ u,v \in \mathbb{R}^r, \ q = \overline{0,r}.$

One consider for the system (1.3) the initial conditions

$$y(0) = y_0$$

On $[0,b]$ we define an uniform partition by the knots

$$\Delta : 0 = x_0 < x_1 < \dots < x_{n-1} < x_n = b \quad n \in \mathbb{N}^*$$

with the step $h = x_{k+1} - x_k, \ k = \overline{0,n-1}$ and one denote $y_k^{(j)} = y^{(j)}(x_k), \ k = \overline{0,n}, \ j = \overline{0,r}.$

II. The first approximation process. Let y be the exact solution of Cauchy problem for the system (1.3). By integrating from x_k to x we get

$$y(x) = y_k + \int_{x_k}^{x} F(t,y(t))dt, \quad x \in [x_k, x_{k+1}] \tag{2.1}$$

and for $x := x_{k+1}$ we get

$$y_{k+1} = y_k + \int_{x_k}^{x_{k+1}} F(t,y(t))dt \tag{2.2}$$

This equality may be approximated with

$$\bar{y}_{k+1} = \bar{y}_k + \int_{x_k}^{x_{k+1}} F(t, y_k^*(t))dt \tag{2.3}$$

where

$$y_k^*(t) = \sum_{j=0}^{r+1} (t - x_k)^j \frac{\bar{y}_k^{(j)}}{j!}, \; t \in [x_k, x_{k+1}] \tag{2.3}$$

which corresponds to the Taylor expansion:

$$y(t) = \sum_{j=0}^{r} (t - x_k)^j \frac{y_k^{(j)}}{j!} + \frac{y^{(r+1)}(\xi_k)}{(r+1)!}(t - x_k)^{r+1}, \tag{2.4}$$

$t \in [x_k, x_{k+1}]$, $x_k < \xi_k < x_{k+1}$.

Now, we assume that the function f has the modulus of continuity $\omega_r(h)$ associated to the above defined mesh of points.

One will also use: $\bar{y}_0 = y_0$, $\bar{y}_0' = y_0'$, ..., $\bar{y}_0^{(r+1)} = y_0^{(r+1)}$.

LEMMA 2.1 *The inequality*

$$\| y_{k+1} - \bar{y}_{k+1} \| \le \| y_k - \bar{y}_k \| (1 + c_0 h) + c_1 \omega_r(h) h$$

holds for $k = \overline{0, n-1}$, *where* c_0 *and* c_1 *are positive and independent of* h.

Proof.

$$\| y_{k+1} - \bar{y}_{k+1} \| \le \| y_k - \bar{y}_k \| + L \int_{x_k}^{x_{k+1}} \| y(t) - y_k^*(t) \| \, dt \le$$

$$\le \| y_k - \bar{y}_k \| + L \int_{x_k}^{x_{k+1}} \Big\| \sum_{j=0}^{r} \frac{y_k^{(j)}}{j!}(t - x_k)^j + \frac{y^{(r+1)}(\xi_n)}{(r+1)!}(t - x_k)^{r+1} + \sum_{j=0}^{r+1} \frac{y_k^{(j)}}{j!}(t - x_k)^j \Big\| dt \le$$

$$\le \| y_k - \bar{y}_k \| + L \sum_{j=0}^{r+1} \frac{\| y_k^{(j)} - \bar{y}_k^{(j)} \|}{(j+1)!} h^{j+1} + L \frac{h^{r+2}}{(r+2)!} \omega_r(h) = \| y_k - \bar{y}_k \| + Lh \| y_k - \bar{y}_k \| +$$

$$+ L \sum_{q=0}^{r} \frac{\| F^{(q)}(x_k, y_k) - F^{(q)}(x_k, \bar{y}_k) \|}{(q+2)!} h^{q+2} + L \frac{h^{r+2}}{(r+2)!} \omega_r(h) =$$

$$= \| y_k - \bar{y}_k \| + \| y_k - \bar{y}_k \| \cdot Lh + \| y_k - \bar{y}_k \| \cdot L^q \sum_{q=0}^{r} \frac{h^{q+2}}{(q+2)!} + \frac{L}{(r+2)!} \omega_r(h) h^{r+2} \le$$

$$\leq \|y_k - \bar{y}_k\|(1 + c_0 h) + c_1 \omega_r(h) h^{r+2}.$$

THEOREM 2.2 *The convergence of the approximate value* $\bar{y}_{k+1}$ *to the exact value* y_{k+1} *is given by the inequality*

$$\|y_{k+1} - \bar{y}_{k+1}\| \leq c_3 \omega_r(h) h^{r+1}.$$

Proof. One apply succesively Lemma 2.1:

$$\|y_{k+1} - \bar{y}_{k+1}\| \leq \|y_k - \bar{y}_k\| \cdot (1 + c_0 h) + c_1 \omega_r(h) h^{r+2}$$

$$\|y_{k+1} - \bar{y}_{k+1}\|(1 + c_0 h) \leq \|y_k - \bar{y}_k\| \cdot (1 + c_0 h)^2 + c_1 \omega_r(h) h^{r+2}(1 + c_0 h)$$

$$\|y_{k+1} - \bar{y}_{k+1}\|(1 + c_0 h)^k \leq \|y_k - \bar{y}_k\| \cdot (1 + c_0 h)^{k+1} + c_1 \omega_r(h) h^{r+2}(1 + c h)^k.$$

Adding the inequalities above one obtain

$$\|y_{k+1} - \bar{y}_{k+1}\| \leq c_1 \omega_r(h) h^{r+2} \sum_{q=0}^{k} (1 + c_0 h)^2 = c_1 \omega_r(h) h^{r+2} \frac{(1 + c_0 h)^{k+1} - 1}{c_0 h}.$$

Because $(1 + c_0 h)^{k+1} = \left(1 + \frac{b c_0}{n}\right)^{k+1} \leq \left(1 + \frac{b c_0}{n}\right)^n \leq e^{b c_0} =$ cosnstant, $(1 + c_0 h)^{k+1}$ is bounded, so $\|y_{k+1} - \bar{y}_{k+1}\| \leq c_3 \omega_r(h) h^{r+1}$.

THEOREM 2.3 *The error for* $\bar{y}_{k+1}^{(q+1)}$ *is given by the inequality*

$$\|y_{k+1}^{(q+1)} - \bar{y}_{k+1}^{(q+1)}\| c_4 \omega_r(h) h^{r+1}, \quad q = \overline{0, r}$$

Proof. $\|y_{k+1}^{(q+1)} - \bar{y}_{k+1}^{(q+1)}\| = \|F^{(q)}(x_{k+1}, y_{k+1}) - F^{(q)}(x_{k+1}, \bar{y}_{k+1})\| \leq$

$$\leq L \cdot \|y_{k+1} - \bar{y}_{k+1}\| \leq c_4 \omega_r(h) h^{r+1}.$$

So, one obtained the approximative values $\bar{y}_0, \bar{y}_1, \ldots, \bar{y}_n \in \mathbb{R}^r$ corresponding to the mesh of points $0 = x_0 < x_1 < \ldots < x_n = b$.

In x_k one obtained the following approximations for the solution of (1.3): $y_k^{(q)} = (y_{k,1}^{(q)}, y_{k,2}^{(q)}, \ldots, y_{k,m}^{(q)})$ for $y_k^{(q)}$, $q = \overline{0, r+1}$ which correspond in (1.1) to $(z, z', \ldots, z^{(m-1)})$.

One denote: $\bar{z}_k := \bar{y}_{k,1}$, $\bar{z}'_k := \bar{y}_{k,2}, \dots,$ $\bar{z}_k^{(m-1)} := \bar{y}_{k,m}$, $\bar{z}_k^{(m)} := \bar{y}'_{k,m}$, $\bar{z}_k^{(r+m+1)} := \bar{y}_{k,m}^{(r+1)}$

THEOREM 2.4 *The convergence of the approximative value* $\bar{z}_{k+1}^{(j)}$ *to the exact value* $z_{k+1}^{(j)}$ *is given by the inequality*

$$|z_{k+1}^{(j)} - \bar{z}_{k+1}^{(j)}| \le c_5 \omega_r(h) h^{r+1}, \; j = \overline{0, r+m+1}$$

Proof. This is a direct consequence of Theorems 2.1 and 2.3.

III. The second approximation process. One obtain the following sets of approximate values:

$$\bar{Z}^{(q)}: \bar{z}_0^{(q)}, \dots, \bar{z}_n^{(q)}, \; q = \overline{0, r+m}$$

which correspond respectively to

$$Z^{(q)}: z_0^{(q)}, z_1^{(q)}, \dots, z_n^{(q)}, \; q = \overline{0, r+m}$$

We are going to construct a spline function S_Δ interpolated to the set $\bar{Z}$ on the mash Δ and approximating the solution of (1.1).

THEOREM 3.1 *For a given mesh of points*

$$\Delta : 0 = x_0 < x_1 < \dots < x_k < x_{k+1} < \dots < x_n = b, \; x_{k+1} - x_k = h, \; k = \overline{0, n-1}$$

and for the given sets of values $\bar{Z}^{(q)}: \bar{z}_0^{(q)}, \bar{z}_1^{(q)}, \dots, \bar{z}_n^{(q)}$, $q = \overline{0, r+m}$ *there is a unique spline function* S_Δ *interpolated to the set* $\bar{Z}$ *on the mesh and satisfying the following conditions:*

(i) $\quad S_\Delta(\bar{z}, x) = S_\Delta(x) \in C^{r+m}[0, b]$.

(ii) $\quad S_k^{(q)}(x_k) = \bar{z}_k^{(q)}$ *for* $q = \overline{0, r+m}$, $k = \overline{0, n}$

(iii) $\quad$ For $x_k \le x \le x_{k+1}$, $k = \overline{0, n-1}$

$$S_\Delta(x) = \sum_{j=1}^{r+m} \frac{\bar{z}_k^{(j)}}{j!}(x - x_k)^j + \sum_{p=1}^{r+m+1} a_p^{(k)} (x - x_k)^{p+r+m}.$$

Proof. From the continuity condition (i), for $x = x_{k+1}$, using (ii) we get

$$S_k^{(i)}(x_{k+1}) = S_{k+1}^{(i)}(x_{k+1}) = \bar{z}_{k+1}^{(i)}. \tag{3.1}$$

Substituting from (3.1) in (iii) we get the following linear system of equations:

$$\sum_{p=1}^{r+m+1} t!\,C^{t}_{r+m+p}\,a_p^{(k)}\,h^{p-1} = h^{t-r-m-1}\left(\bar{z}^{(t)}_{k+1} - \sum_{j=0}^{r+m-t}\frac{\bar{z}_n^{(j+t)}}{j!}h^{j}\right),\quad t = \overline{0,r+m} \tag{3.2}$$

for the unknowns $a_p^{(k)}$, $p = \overline{1,r+m+1}$. One denote

$$F_t^{(k)} = h^{t-r-m-1}\left(\bar{z}^{(t)}_{k+1} - \sum_{j=0}^{r+m-1}\frac{\bar{z}_k^{(j+t)}}{j!}h\right). \tag{3.3}$$

The system (3.2) has always (for $h = 0$) a unique solution because its determinant is

$$D_r = \begin{vmatrix} 1 & & h^{p-1} & & h^{r+m} \\ C^{1}_{r+m+1}\cdot 1! & & C^{1}_{r+m+p}\cdot 1!\,h^{p-1} & & C_{2r+2m+1}\cdot 1!\,h^{r+m} \\ C^{2}_{r+m+1}\cdot 2! & & C^{2}_{r+m+p}\cdot 2!\,h^{p-1} & & C^{2}_{2r+2m+1}\cdot 2!\,h^{r+m} \\ & & & & \\ C^{r+m}_{r+m+1}(r+m)! & \dots & C^{r+m}_{r+m+p}(r+m)!\,h^{p-1} & \dots & C^{r+m}_{2r+2m+1}(r+m)!\,h^{r+m} \end{vmatrix}$$

$$\prod_{t=0}^{r+m} t!\,h^{1+2+\dots+(r+m)}\,1 = h^{\frac{1}{2}(r+m)(r+m+1)}\prod_{t=0}^{r+m} t! \neq 0.$$

So $D_r \neq 0$ and the system (3.2) has always a unique solution for $h > 0$ i.e. the spline function approximating the solution of (1.1) exists and is unique determined.

The coefficients are determined as follows.

One replace the column p in D_r by the column

$$(F_0^{(k)}, F_1^{(k)}, \dots, F_{r+m}^{(k)})$$

and we denote the determinant obtained by D_r^P. Then, the solution of system (3.2) will be
$a_p^{(k)} = \dfrac{D_r^p}{D_r}$, $p = \overline{1,r+m+1}$.

By factorising D_r^P in terms of $F_0^{(k)}, \dots, F_{r+m}^{(k)}$ we get

$$a_p^{(k)} = \frac{1}{h^{p-1}}\sum_{i=0}^{r+m} cp_i F_i^{(k)} \tag{3.4}$$

where $1/h^{p-1}$ is a factor put in front of the sum so the coefficients c_{pi} be independent of h.

Now we shall discuss the convergence of the spline function to the solution.

LEMMA 3.2 *The inequalities* $|a_p^{(k)}| \le \dfrac{A_p}{h^p}\omega_r(h)$ *hold* $p = \overline{1,r+m+1}$ *where* A_p *are* *constants independent of h.*

Proof. One estimate

$$|F_t^{(k)}| = h^{t-r-m-1}\,|\,\bar{z}_{k+1}^{(t)} - \sum_{j=0}^{r+m-t} \frac{\bar{z}_k^{(j+t)}}{j!} h^j\,|\,.$$

One have the following Taylor expansion for $z^{(t)}(x)$, for $x_k \le x \le x_{k+1}$.

$$z^{(t)}(x) = \sum_{j=0}^{r+m-1-t} \frac{z_k^{(j+t)}}{j!}(x-x_k)^j + \frac{z^{(r+m)}(\xi_{kt})}{(r+m-t)!}(x-x_k)^{r+m-t},\ t = \overline{0,r+m}.$$

and for $x = x_{k+1}$:

$$z_{k+1}^{(t)} = \sum_{j=0}^{r+m+1-t} \frac{z_k^{(j+t)}}{j!} h^j + \frac{z^{(r+m)}(\xi_{kt})}{(r+m-t)!} h^{r+m-t},\ t = \overline{0,r+m}.$$

Using (3.5) and the t-th equation in the system (3.2) we get

$$|F_k^{(t)}| \le h^{t-r-m-1}\left[\,|\,z_{k+1}^{(t)} - \bar{z}_{k+1}^{(t)}| + \sum_{j=0}^{r+m-t} \frac{|\,z_k^{(j+t)} - \bar{z}_k^{(j+t)}|}{j!} h^j\right.$$

$$+ \frac{|\,z^{(r+m)}(\xi_{k,t}) - z^{(r+m)}|}{(r+m+t)!} \le h^{t-r-m-1}\,[c_t^{*}\,\omega_r(h)\,h^{r+m-t}],$$

with $c_t^{*} > 0,\ t = \overline{0,r+m}$, independent of h, so

$$F_t^{(k)} \le c_t^{*}\,\frac{\omega_r(h)}{h},\ t = \overline{0,r+m}$$

One substitute (3.6) in (3.4) and one obtain

$$a_p^{(k)} = \frac{1}{h^{p-1}} \sum_{i=0}^{r+m} c_i F_i^{(k)} \le \frac{1}{h^{p-1}} \sum_{i=0}^{r+m} c_i c_i^{*}\,\omega_r(h)\,\frac{1}{h} =$$

$$= \frac{1}{h^p}\,\omega_r(h) \sum_{i=0}^{r+m} c_i\ c_i^{*} = A_p \frac{\omega_r(h)}{h^p},$$ where $A_p := \sum_{i=0}^{r+m} c_i\ c_i^{*}$ is a cosstant independent of h.

THEOREM 3.3 *Let z be the exact solution of* (1.1). *If S_Δ is the spline function constructed in Theorem* 3.1 *then there exists a constant E independent of h for which the inequalities*

$$|\,z^{(q)}(x) - S_\Delta^{(q)}(x)| \le E\,\omega_r(h)\,h^{r+m-q},\ q = \overline{0,r+m}$$

hold for any $x \in [0,b]$.

Proof. Using the Taylor expansion previously constructed for $z^{(q)}(x)$ and condition (iii) in Theorem 3.1 we get

$$|z^{(q)}(x) - S_\Delta^{(q)}(x)| = |\sum_{j=0}^{r+m+1-q} \frac{z_k^{(j+q)}}{j!}(x-x_k)^j + \frac{z^{(r+m)}(\xi_{kq})}{(r+m-q)!}(x-x_k)^{r+m-q} -$$

$$-\sum_{j=0}^{r+m+1-q} \frac{\bar{z}_k^{(j+q)}}{j!}(x-x_k)^j - \frac{\bar{z}_k^{(r+m)}}{(r+m-q)!}(x-x_k)^{r+m-q} - \sum_{p=1}^{r+m+1-q} q!\, c_{p+r+m}^{q}\, a_p^{(k)}(x-x_k)^{(p+r+m-q}|$$

$$\le \sum_{j=0}^{r+m+1-q} \frac{|z_k^{(j+q)} - \bar{z}_k^{(j+q)}|}{j!} h^j + \frac{|z^{(r+m)}(\xi_{k,q} - \bar{z}_k^{(r+m)}|}{(r+m-2)!} h^{r+m-q} + \sum_{p=1}^{r+m-q-1} q!\, c_{p+r+m}^{q}\, a_p^{(q)} h^{p+r+m-q} \le$$

$$\le c_q^{**}\,\omega_r(h)\, h^{r+m-q}.$$

Taking $E = \max\{c_q^{**};\ q = \overline{0, m+r}\}$, the theorem is proved.

THEOREM 3.4 *If we denote by* $S_\Delta^{(m)}$ *the function* $s_\Delta^{(m)}(x) = f(x, S_\Delta(x), S_\Delta'(x), \ldots, S_\Delta^{(m-1)}(x))$, $x \in [0, b]$ *and if* $\bar{S}_\Delta$ *is the spline function defined in Theorem* 3.1 *then for any* $x \in [0,b]$

$$|\bar{S}_\Delta^{(m)}(x) - S_\Delta^{(m)}(x)| \le M\omega_r(h)\, h^r.$$

where M is a positive constant independent of h (i.e. *the spline function verifies the equation while* $n \to \infty$ *or* $h \to 0$).

Proof. $|\bar{S}_\Delta^{(m)}(x) - S_\Delta^{(m)}(x)| \le |\bar{S}_\Delta^{(m)}(x) - z^{(m)}(x)| + |x^{(m)}(x) - S_\Delta^{(m)}(x)| =$

$$= |f(x, S_\Delta(x), \ldots, S_\Delta^{(m-1)}(x)) - f(x, z(x), \ldots, z^{(m-1)}(x))| + |z^{(m)}(x) - S_\Delta^{(m)}(x)| \le$$

$$\le LK|S_\Delta(x) - z(x)| + LK|S'(x) - z'(x)| + \ldots +$$

$$+ LK|S_\Delta^{(m-1)}(x) - z^{(m-1)}(x)| + |z^{(m)}(x) - S_\Delta^{(m)}(x)| \le$$

$$\le LKE\omega_r(h)h^{r+m} + LKE\omega_r(h)h^{r+m-1} + \ldots + LKE\omega_r(h)h^{r+1} + E\omega_r(h)h^r =$$

$$= (LKEh^m + LKEh^{m-1} + \ldots + LKEh + E)h^r\omega_r(h) \le M\omega_r(h)h^r,$$

where $M > 0$ is independent of h.

Remark. If $f \in C^\infty([0, b] \times \mathbb{R}^m)$, as the error is $O(h^{r+m})$ we may choose $r \in \mathbf{N}$

suitable so that the method is available.

IV. The stability of the method. A change in one of the calculated values from $\bar{y}_k$ to $\bar{u}_k$ will lead us to solve

$$\bar{u}_{i+1} = \bar{u}_i + \int_{x_i}^{x_{i+1}} F(t, u_i^*(t))\, dt. \tag{4.1}$$

Let $\varepsilon_k := \|\bar{u}_k - \bar{y}_k\|$, the introduced error.

THEOREM 4.1 *If any of the calculated values $\bar{y}_k$ is changed into $\bar{u}_k$ then the inequality $\|u_i^{(t)} - y_i^{(t)}\| \le c_8 \varepsilon_k$ holds for any $i = \overline{k+1, n}$ and $t = \overline{0, r+1}$.*

Proof. Substracting (2.3) from (4.1) and proceeding as in the proof of Lemma 2.1 we get

$$\varepsilon_{i+1} \le \varepsilon_i (1 + c_6 h) \le (1 + c_6 h)^{i-k} \varepsilon_k \le e^{c_6 b} \varepsilon_k \le c \varepsilon_k$$

where c is independent of h. Also, for $q = \overline{0, r}$ we get

$$\|\bar{u}_i^{(q+1)} - \bar{y}_i^{(q+1)}\| = \|F^{(q)}(x_i, \bar{u}_i) - F^{(q)}(x_i, \bar{y}_i)\| \le L\|\bar{u}_i - \bar{y}_i\| \le Lc\varepsilon_k \le c_7 \varepsilon_k$$

so

$$\|u_i^{(t)} - y_i^{(t)}\| \le c_8 \varepsilon_k, \quad t = \overline{0, r+1}.$$

As we did in paragrapf II., we shall denote

$$\bar{v}_k := \bar{u}_{k,1},\ \bar{v}_k' := \bar{u}_{k,2}, \ldots, \bar{v}_k^{(m-1)} := \bar{u}_{k,m},\ \bar{v}_k^{(m)} := \bar{u}_{k,m}',\ \bar{v}_k^{(r+m+1)} := \bar{u}_{k,m}^{(r+m)}$$

So

$$|\bar{v}_i^{(t)} - \bar{z}_i^{(t)}| \le \|\bar{u}_i - \bar{y}_i\| \le c_8 \varepsilon_k \quad \text{for} \quad t = \overline{0, m-1}$$

$$|\bar{v}_i^{(t)} - \bar{z}_i^{(t)}| \le \|\bar{u}_i^{(m-1-t)} - \bar{y}_i^{(m-1-t)}\| \le c_8 \varepsilon_k \quad \text{for} \quad t = \overline{m, m+r+1}.$$

and thus the theorem is proved.

THEOREM 4.2 *If any of the calculated values $\bar{y}_k$ is changed into $\bar{u}_k$ and consequently, the spline function approximating the solution of (1.1) is changed from S into* *s, then for any* $x \in [x_i, x_{i+1}]$, $i = \overline{k, n-1}$, *the inequality*

$$|s_i(x) - S_i(x)| \leq c_{10}\varepsilon_k \text{ holds.}$$

Proof. Consider the interval $[x_i, x_{i+1}]$ where $i = \overline{k, n+1}$. Then, analogously to the spline function S_Δ introduced Theorem 3.1, the new spline function due to the variation of $\bar{y}_k$ to $\bar{u}_k$ will be

$$s_i(x) = \sum_{j=0}^{r+m} \frac{\bar{v}_i^{(j)}}{j!}(x - x_i)^j + \sum_{p=1}^{r+m+1} b_p^{(i)}(x - x_i)^{p+r+m} \tag{4.3}$$

and will satisfy the conditions

$$s_i^{(t)}(x_{i+1}) = s_{i+1}^{(t)}(x_{i+1}) = \bar{v}_{i+1}^{(t)},\ s_{n-1}^{(t)}(x_n) = \bar{v}_n^{(t)} \tag{4.4}$$

for $i = \overline{k, n-2}$.

Then the linear system corresponding to (3.2) will be

$$\sum_{p=1}^{r+m+1} t!\,C_{r+m+p}^{t}\, b_p^{(i)} h^{p-1} = G_t^{(i)},\ t = \overline{0, r+m} \tag{4.5}$$

where

$$G_t^{(i)} = h^{t-r-m-1}\left(\bar{v}_{i+1}^{(t)} - \sum_{j=0}^{r+m-t} \frac{\bar{v}_i^{(j+t)}}{j!} h^j\right),\ t = \overline{0, r+m} \tag{4.6}$$

and corresponding to (3.4) we get

$$b_p^{(i)} = \frac{1}{h^{p-1}} c_{pt} G_t^{(i)}. \tag{4.7}$$

$$|s_i(x) - S_i(x)| = |\sum_{j=0}^{r+m} \frac{\bar{v}_i^{(j)}}{j!}(x - x_i)^j + \sum_{p=1}^{r+m+1} b_p^{(i)}(x - x_i)^{p+r+m} -$$

$$\sum_{j=0}^{r+m} \frac{\bar{z}_i^{(j)}}{j!}(x - x_i)^j - \sum_{p=1}^{r+m+1} a_p^{(i)}(x - x_i)^{p+r+m}| \leq$$

$$\leq \sum_{j=0}^{r+m} \frac{|\bar{v}_i^{(j)} - \bar{z}_i^{(j)}|}{j!} h^j + \sum_{p=1}^{r+m+1} |b_p^{(i)} - a_p^{(i)}|\, h^{p+r+m}.$$

From (3.4) and (4.7) we get

$$|b_p^{(i)} - a_p^{(i)}| \leq \frac{1}{h^{p-1}} \sum_{t=0}^{r+m} c_{pt}\, |G_t^{(i)} - F_t^{(i)}|$$

From (3.3) and (4.6) we get

$$|G_i^{(t)} - F_i^{(t)}| = h^{t-r-m-1} \, |\bar{v}_{i+1}^{(t)} - \sum_{j=0}^{r+m-t} \frac{\bar{v}_i^{(j+t)}}{j!} h^j - \bar{z}_{i+1}^{(t)} + \sum_{j=0}^{r+m-t} \frac{\bar{z}_i^{(j+t)}}{j!} h^j| \le$$

$$\le h^{t-r-m-1} \, |\bar{v}_{i+1}^{(t)} - z_{i+1}^{(t)}| + h^{t-r-m-1} \sum_{j=0}^{r+m-t} \frac{|\bar{v}_i^{(j+t)} - \bar{z}_i^{(j+t)}|}{j!} h^j \le$$

$$\le h^{t-r-m-1} \left(c_8 \varepsilon_k + \sum_{j=0}^{r+m-t} c_8 \varepsilon_k \frac{h^j}{j!} \right) \le c_9 \varepsilon_k h^{t-r-m-1}$$

and so we get

$$|b_p^{(i)} - a_p^{(i)}| \le \frac{1}{h^{p-1}} \sum_{t=0}^{r+m} c_9 c_{pt} \varepsilon_k h^{t-r-m-1} .$$

Using Theorem 4.1 we get

$$|s_i(x) - S_i(x)| \le \sum_{j=0}^{r+m} c_8 \varepsilon_k \frac{h^j}{j!} + \sum_{p=1}^{r+m+1} h^{p+r+m} \frac{1}{h^{p-1}} \sum_{t=0}^{r+1} c_9 c_{pt} \varepsilon_k h^{t-r-m-1} =$$

$$= c_8 \varepsilon_k \frac{h^j}{j!} + c_9 \varepsilon_k \sum_{p=1}^{r+m+1} \sum_{t=0}^{r+1} c_{pt} h^t \le c_{10} \varepsilon_k$$

which is a bounded multiple of the introduced error.

THEOREM 4.3 *Under the assumptions of Theorem* 4.2 *the inequalities*

$$|s_i^{(t)}(x) - S_i^{(t)}(x)| \le c_{11} \varepsilon_k$$

hold for any $t = \overline{0,m}$ *and* $i = \overline{k,n-1}$.

Proof. Following the same procedure as in Theorem 4.2 one obtain the requested inequalities.

Conclusion. As any variation of the calculated error is a bounded multiple of the introduced error, the method is stable.

REFERENCES

1. Th. Fawzy, *Spline functions and the Cauchy problem III*, Annales Univ. Sci. Budapest. Sectio Computatorica, 1(1978), 25-34.
2. Th. Fawzy, *Spline functions and the Cauchy problem IV*, Acta Mathematica Academiae Scientiarum Hungaricae, Tomus 30(3-4)(1977), pp. 219-226.
3. J. Gyorvari, Cs. Mihalyko, *The numerical solution of nonlinear differential equations by spline functions*, Acta Mathematica Hungarica, 59(1-2)(1992), 39-48.
4. Gh. Micula, *Spline Functions and Applications* (Romanian), Editura Tehnică, Bucureşti, 1978.
5. Gh. Micula, *The numerical solution of nonlinear differential equations*, Z.A.M.M., 55(1975), 254-255.
6. S. Sallam, W. Ameen, *Numerical solution of general n-th order differential equations via spline*, Applied Numerical Mathematics, 6(1989/90), 225-238.

STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 1, 1995

PRECONDITIONING FOR THE FULFILMENT OF THE APPROXIMATION ASSUMPTION IN THE ALGEBRAIC MULTIGRID METHOD

Constantin POPA*

Received: November 15, 1994
AMS subject classification: 65M55, 65M60, 65N55, 65F10

REZUMAT. - Precondiţionarea pentru îndeplinirea condiţiilor de aproximare în metoda multigrid algebrică. Se prezintă o metodă de precondiţionare pentru sisteme liniare simetrice şi pozitiv definite. Folosind un operator de interpolare se dovedeşte că se realizează îndeplinirea condiţiilor de aproximare, care de obicei cauzează cele mai multe dificultăţi în utilizarea algoritmilor algebrici multigrid [4], [17]. Astfel se obţine convergenţa V-cicluri de tip multigrid pentru sistemele simetrice generale pozitiv definite. Lucrarea se încheie cu prezentarea mai multor exemple numerice pentru ecuaţiile Dirichlet, precum şi Poisson şi Helmholtz anisotropice.

Abstract. In the last years a lot of papers ([1], [2], [3], [15], [20]) presented various preconditioning techniques for the improvement of the condition number of symmetric and positive definite M-matrices arrising from the discretization of elliptic partial differential equations. All of these techniques essentially use the "geometric" information offerend by the continuous problem ("good" properties of the partial differential operator, special types of regular discretizations etc.). Thus, even the ideas are quite general we cannot apply these methods for arbitrary systems.

In this paper we present a method of preconditioning for arbitrary symetric and positive definite linear systems. We don't obtain an improvement of the condition number of the system matrix (which is very hard in this general approach) but using a special construction of the interpolation operator we prove the fulfilment of the approximation

* *University of Constanţa, Department of Mathematics, 8700 Constanţa, Romania*

assumption (which usually causes the most troubles in the algebraic multigrid algorithms, see [4], [17]). Thus we obtain the convergence of the V-cycle type algebraic multigrid for general symetric and positive definite systems.

At the end of the paper we present numerical examples on Dirichlet, anisotropic Poisson and Helmholtz equations.

1. Introduction. In this section we shall use the notations, definitions and results from [17]. Let A be an n by n symmetric and positive definite matrix. For $b \in \mathbb{R}^n$ we consider the system

$$Au = b, \tag{1}$$

with the (unique) exact solution $u \in \mathbb{R}^n$. Let $q \geq 2$ be an integer and $C_1, C_2, \ldots, C_q$ a sequence of nonvoid subsets of $\{1, \ldots, n\}$ such that

$$\{1, \ldots, n\} = C_1 \supset C_2 \supset \ldots \supset C_q, \tag{2}$$

$$|C_m| = n_m,\ m = 1, \ldots, q, \tag{3}$$

$$n = n_1 > n_2 > \ldots > n_q > 1, \tag{4}$$

where by $|C_m|$ we denoted the number of elements in the set C_m. Furthermore for $m = 1, 2, \ldots, q-1$ we consider the linear operators

$$I^m_{m+1} : \mathbb{R}^{n_{m+1}} \to \mathbb{R}^{n_m},\ I^{m+1}_m : \mathbb{R}^{n_m} \to \mathbb{R}^{n_{m+1}} \tag{5}$$

and the matrices A^{m+1} with the properties

$$I^{m+1}_m = (I^m_{m+1})^t,\ A^1 = A,\ A^{m+1} = I^{m+1}_m A^m I^m_{m+1}. \tag{6}$$

For $m = 1, \ldots, q$-1 we define the coarse grid correction operators T^m by

$$T^m = I_m - I^m_{m+1} (A^{m+1})^{-1} I^{m+1}_m A^m \tag{7}$$

and the smoothing process of the form

$$u^m_{new} = G^m u^m_{old} + (I_m - G^m)(A^m)^{-1} b^m, \tag{8}$$

where I_m is the identity and

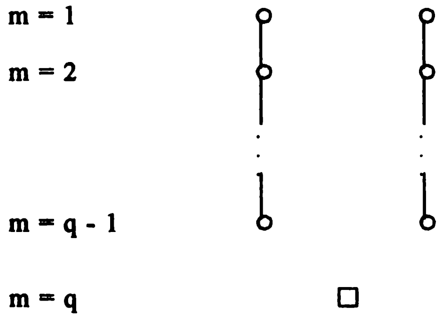
$$A^m u^m = b^m \tag{9}$$

are the systems corresponding to the coarse levels.

Remarks 1. The sets C_m, $m = 1, \ldots, q$ formally play the same role as coarse grids in the classical geometric multigrid ([3]), I^m_{m+1}, I^{m+1}_m are the interpolation and restriction operators, respectively and A^m the coarse grids matrices.

2. The form (8) of the smoothing process includes the classical relaxation schemes (ω-Jacobi, Gauss-Seidel, S O R, I L U - decomposition).

3. With all the above defined elements we consider a classical V - cycle type algorithm (with at least one smoothing step performed after each coarse grid correction step) looking like (e.g. [18])



(10)

where we suppose that on the last grid ($m = q$) the system (9) is solved exactly.

We introduce the matrix

$$D_m = \mathrm{diag}(A^m), \ m = 1, \ldots, q-1 \tag{11}$$

and define on each level the inner products

$$<u^m, v^m>_0 = <D_m u^m, v^m>, \; <u^m, v^m>_1 = <A^m u^m, v^m>, \tag{12}$$

$$<u^m, v^m>_2 = <D_m^{-1} A^m u^m, A^m v^m>,$$

along with their corresponding norms $\|\cdot\|_i$, $i = 0, 1, 2$, where $<\cdot, \cdot>$ is the Euclidean inner product and $\|\cdot\|$ the Euclidean norm (on the spaces $\mathbb{R}^{n_m}$). We shall denote by $e^m = v^m - u^m$ the error on each level $m = 1, ..., q$-1. We know the following result concerning the convergence of the above defined V-cycle.

THEOREM 1. ([17]) *Assume that the interpolations* I^m_{m+1}, $m = 1, \ldots, q-1$ *have full rank and that there exists a constant* $\delta > 0$ *independently on m and* e^m *such that*

$$\|G^m e^m\|_1^2 \leq \|e^m\|_1^2 - \delta \|T^m e^m\|_1^2, \; m = 1, \ldots, q-1. \tag{13}$$

Then $\delta \leq 1$ *and the* V *- cycle* (10) *to solve* (1) *has a convergence factor* (*in the energy norm* $\|\cdot\|_1$) *bounded above by* $\sqrt{1-\delta}$.

COROLLARY 1. ([17]) *If there exists constants* $\alpha, \beta > 0$ *independently of m and* e^m *such that*

$$\|G^m e^m\|_1^2 \leq \|e^m\|_1^2 - \alpha \|e^m\|_2^2, \tag{14}$$

$$\|T^m e^m\|_1^2 \leq \beta \|e^m\|_2^2, \tag{15}$$

for every $m = 1, ..., q$-1 *then we have* (13) *with*

$$\delta = \alpha/\beta \tag{16}$$

Remarks 1. Properties (14), (15) are called the smoothing assumption (SA) and the approximation assumption (AA), respectively ([17]).

2. (SA) is fulfilled by the classical relaxation schemes (see [4], [12], [17]).

3. The condition (AA) causes the most troubles. There are two weaker forms, namely

$$\text{(AA}_1\text{)} \qquad \|T^m e^m\|_1^2 \le \beta_1 \|T^m e^m\|_2^2, \tag{17}$$

$$\text{(AA}_2\text{)} \qquad \min\{\|e^m - I_{m+1}^m e^{m+1}\| + O^2,\ e^{m+1} \in \mathbf{R}^{n_{m+1}}\} \le \beta_2 \|e^m\|_1^2, \tag{18}$$

where the positive constants β_1 and β_2 are also independently on m and e^m. Following the result from [17] (AA$_2$) implies (AA$_1$) with $\beta_1 = \beta_2$ and one of them with the smoothing assumption (14) ensures the convergence of the two grid algorithm $(m, m+1)$. For the multilevel case ($q \ge 3$) it is necesary that (15) holds. This is, in fact; our principal aim in the present paper.

2. Preconditioning - the two level case. We present in this section the method of preconditioning for a pair of two consecutive grids $(m, m+1)$ where $m \in \{1, ..., q-1\}$ is arbitrary fixed. In order to simplify the notations we shall write $n, p, I_p^n, I_n^p, C_p, A, A_p$ instead of $n_m, n_{m+1}, I_{m+1}^m, I_m^{m+1}, C_m, A^m, A^{m+1}$ respectively. We shall suppose also that the coarse grid C_p satisfies

$$C_p = \{n-p+1, n-p+2, ..., n\} \tag{19}$$

Accordingly to (19) we consider the block decomposition of A

$$A = \begin{bmatrix} A_1 & B \\ B^t & A_2 \end{bmatrix} \tag{20}$$

where A_1, A_2 are symmetric invertible matrices of dimension n-p and p, respectively, with A_1 positive definite. Let $\bar{A}_1$ be another symmetric and positive definite matrix of dimension n-p.

We consider the Cholesky decompositions of A_1 and $\bar{A}_1$

$$A_1 = L_1 L_1^t,\ \bar{A}_1 = \bar{L}_1 \bar{L}_1^t \tag{21}$$

and we define the matrix $\bar{\Delta}_1$ (of dimension n) by

$$\bar{\Delta}_1 = \begin{bmatrix} \bar{L}_1 \bar{L}_1^{-1} & O \\ O & I_2 \end{bmatrix} \tag{22}$$

where I_2 is the identity on $\mathbf{R}^p$. We shall also denote by I_1 the identity on $\mathbf{R}^{n-p}$ and by $u = [u_1,u_2]$ a vector $u \in \mathbf{R}^n$ for the descomposition

$$\mathbf{R}^n = \mathbf{R}^{n-p} \oplus \mathbf{R}^p \tag{23}$$

We precondition the system (1) in the following way

$$(\bar{\Delta}_1 A \bar{\Delta}_1^t)((\bar{\Delta}_1^t)^{-1} u) = \bar{\Delta}_1 b. \tag{24}$$

Thus the system (1) becomes

$$\bar{A}\bar{u} = \bar{b}, \tag{25}$$

where

$$\bar{u} = (\bar{\Delta}_1^t)^{-1} u,\ \bar{b} = \bar{\Delta}_1 b, \tag{26}$$

and

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & \bar{B} \\ \bar{B}^t & A_2 \end{bmatrix} = \bar{\Delta}_1 A \bar{\Delta}_1^t\ , \tag{27}$$

with the $(n-p) \times p$ matrix $\bar{B}$ given by

$$\bar{B} = \bar{L}_1 L_1^{-1} B. \tag{28}$$

Remark. It is clear that u is the solution of (1) if and only if $\bar{u}$ from (26) is the solution of (25). We also observe that the preconditioned matrix $\bar{A}$ from (27) is symmetric and positive definite. Thus we can define for $\bar{A}$ the inner products from (12) and the associated norms. These norms will be denoted by $|||\cdot|||_i$, $i = 0, 1, 2$. Accordingly with [13] and [14] we shall define the interpolation I_p^n by

$$I_p^n = \begin{bmatrix} -\bar{A}_1^{-1}\ \bar{B} \\ I_2 \end{bmatrix}. \tag{29}$$

Then I_p^n has full rank and from (21) and (28) we obtain

$$I_p^n = \begin{bmatrix} -(\bar{L}_1^t)^{-1}\ L_1^{-1}B \\ I_2 \end{bmatrix} \tag{30}$$

PROPOSITION 1. (i) *The coarse grid matrix A_p is given by*

$$A_p = A_2 - B^t A_1^{-1} B \tag{31}$$

and is independent on the matrix $\bar{A}_1$ of the preconditioning.

(ii) *The coarse grid correction operator, T, is given by*

$$T = \begin{bmatrix} I_1 & \bar{A}_1^{-1}\bar{B} \\ O & O \end{bmatrix}. \tag{32}$$

(iii) *If $\bar{e} = [\bar{e}_1, \bar{e}_2] \in \mathbb{R}^n$ is the error after the correction step then*

$$|||\bar{e}|||_1^2 = <\bar{A}\bar{e}, \bar{e}> = <\bar{A}_1\bar{e}_1, \bar{e}_1> \geq \lambda_{min}(\bar{A}_1)\,\|\bar{e}\|^2, \tag{33}$$

where $\lambda_{min}(\bar{A}_1)$ is the smallest eigenvalue of $\bar{A}_1$.

Proof. (i) Firstly we observe that from (29), (6) and (27) we obtain

$$I_m^p \bar{A} = [O \,\vdots\, A_2 - \bar{B}^t \bar{A}_1^{-1} \bar{B}] = [O \,\vdots\, \tilde{A}_2], \tag{34}$$

with $\tilde{A}_2$ given by

$$\tilde{A}_2 = A_2 - \bar{B}^t \bar{A}_1^{-1} \bar{B}.$$

Then

$$A_p = I_n^p \bar{A} I_p^n = [O \vdots \bar{A}_2] \begin{bmatrix} -\bar{A}_1^{-1}\bar{B} \\ I_2 \end{bmatrix} = \tilde{A}_2$$

But, using (21), (35) and (28) we have

$$\tilde{A}_2 = A_2 - B^t(L_1^{-1})^t \bar{L}_1^t (\bar{L}_1^t)^{-1} (\bar{L}_1)^{-1} \bar{L}_1 \bar{L}_1^{-1} B = A_2 - B^t (L_1^{-1})^t L_1^{-1} B = A_2 - B^t A_1^{-1} B,$$

which gives us (31).

(ii) Using (7), (34) and (31) we obtain

$$T = I - I_p^n A_p^{-1} I_n^p \bar{A} = I - I_p^n A_p^{-1} [O \vdots A_p] = I - \begin{bmatrix} -\bar{A}_1^{-1}\bar{B} \\ I_2 \end{bmatrix} [O \vdots I_2] =$$

$$= \begin{bmatrix} I_1 & O \\ O & I_2 \end{bmatrix} - \begin{bmatrix} O & -\bar{A}_1^{-1}\bar{B} \\ O & I_2 \end{bmatrix} = \begin{bmatrix} I_1 & \bar{A}_1^{-1}\bar{B} \\ O & O \end{bmatrix}$$

which is exactly (32).

(iii) If $\bar{e} = [\bar{e}_1, \bar{e}_2] \in \mathbb{R}^n$ is the error after the correction step we have ([8])

$$I_n^p \bar{A} \bar{e} = 0.$$

From (39), (36) we obtain

$$[O \vdots A_p] \begin{bmatrix} \bar{e}_1 \\ \bar{e}_2 \end{bmatrix} = 0$$

thus

$$A_p \bar{e}_2 = 0 \Rightarrow \bar{e}_2 = 0,$$

because A_p is invertible. Then (33) is obvious.

We shall make now the following assumption: *there exists a constant $c > 0$ independently of the dimension n of the matrix A such that*

$$\|\bar{A}_1^{-1}\bar{B}\| \le c. \tag{42}$$

Then we obtain the following result concerning the fulfilment of (15) and (17).

THEOREM 2. *For every vector $e = [e_1, e_2] \in \mathbb{R}^n$ we have*

$$|||Te|||_1^2 \le \frac{\min\{\bar{a}_{ii}, 1 \le i \le n-p\}}{\lambda_{\min}(\bar{A}_1)} |||Te|||_2^2 \tag{43}$$

and

$$|||Te|||_1^2 \le c^2 \frac{\max\{\bar{a}_{ii}, 1 \le i \le n-p\} \cdot \min\{\bar{a}_{ii}, 1 \le i \le n-p\}}{\min\{a_{ii}, n-p+1 \le i \le n\} \cdot \lambda_{\min}(\bar{A}_1)} |||e|||_2^2, \tag{44}$$

where we denoted the elements of the matrix $\bar{A}_1$ by $\bar{a}_{ij}$.

Proof. We denote the vector

$$\bar{e} = Te \tag{45}$$

by $\bar{e} = [\bar{e}_1, \bar{e}_2]$ i.e. the error after the correction step. From Proposition 1 (iv)

$$\bar{e}_2 = 0, \tag{46}$$

thus, using (33) and (46),

$$|||\bar{e}|||_1^2 \ge \lambda_{\min}(\bar{A}_1)\|e_1\|^2 = \lambda_{\min}(\bar{A}_1)\|\bar{e}\|^2 \ge \frac{\lambda_{\min}(\bar{A}_1)}{\min\{\bar{a}_{ii}, 1 \le i \le n-p\}} |||\bar{e}|||_0^2. \tag{47}$$

But a simple calculation using Cauchy-Schwarz inequality (see also [4], [17]) yields for $\bar{e}$, using also (46),

$$|||\bar{e}|||_1^2 \le |||\bar{e}|||_2 \, |||\bar{e}|||_0. \tag{48}$$

Combining (47) with (48) we get (43).

For the second assertion, (44), we firstly observe that

$$\bar{A}T = T'\bar{A}, \tag{49}$$

which follows from (7) and the symmetry of $\bar{A}$. Then we have

$$|||\bar{e}|||_2^2 = |||Te|||_2^2 = <\bar{D}^{-1}\bar{A}Te, \bar{A}Te> \leq \|\bar{D}^{\frac{1}{2}}T\bar{D}^{-1}T'\bar{D}^{\frac{1}{2}}\|\cdot|||e|||_2^2 = \rho(EE')|||e|||_2^2, \tag{50}$$

where $\bar{D} = \mathrm{diag}(\bar{A})$ and E is the matrix given by

$$E = \bar{D}^{\frac{1}{2}}T(\bar{D})^{-\frac{1}{2}}. \tag{51}$$

But from (32) and (51) we obtain

$$E = \begin{bmatrix} I_1 & \bar{D}^{1/2}\bar{A}_1^{-1}\bar{B}(\bar{D})^{-1/2} \\ O & O \end{bmatrix} \tag{52}$$

then

$$EE' = \begin{bmatrix} I_1 + KK' & O \\ O & O \end{bmatrix} \tag{53}$$

where K is the matrix

$$K = \bar{D}_1^{\frac{1}{2}}\bar{A}_1^{-1}\bar{B}(\bar{D}_2)^{-\frac{1}{2}} \tag{54}$$

Then, using (42), it results that

$$\rho(EE') = \rho(I_1 + KK') = 1 + \rho(KK') \leq 1 + \|K\|^2 \leq 1 + \|\bar{D}_1\|\cdot\|\bar{A}_1^{-1}\bar{B}\|\cdot\|\bar{D}_2^{-1}\| \leq$$
$$\leq \frac{\max\{\bar{a}_{ii}, 1 \leq i \leq n-p\}}{\min\{a_{ii}, n-p+1 \leq i \leq n\}}\cdot c^2 \tag{55}$$

and from (50) and (55) we obtain

$$|||Te|||_2^2 = |||\bar{e}|||_2^2 \leq \frac{\max\{\bar{a}_{ii}, 1 \leq i \leq n-p\}}{\min\{a_{ii}, n-p+1 \leq i \leq n\}} c^2\cdot|||e|||_2^2. \tag{56}$$

Now, using (43), (44) is obvious.

It remains now to see under what assumptions $\lambda_{\min}(\bar{A}_1)$ from (33) and c from (42) are constants which not depend on the dimension of the matrices A or $\bar{A}$. In that sense we have the following result.

PROPOSITION 2. *Suppose that there exists a constant* $\gamma > 0$, *independently on the dimension n of A such that*

$$\lambda_{\min}(A_1) \geq \gamma, \; \lambda_{\min}(\bar{A}_1) \geq \gamma. \tag{57}$$

Then (42) holds with $c > 0$ *given by*

$$c = \frac{\|A\|_\infty}{\gamma} \tag{58}$$

where by $\|S\|_\infty$ *we denoted the number*

$$\|S\|_\infty = \max_i \sum_j |s_{ij}| \tag{59}$$

for an arbitrary matrix $S = (s_{ij})$.

Proof. From (30) we have

$$\bar{A}_1^{-1}\bar{B} = (\bar{L}_1')^{-1}(L_1^{-1})B$$

Thus

$$\|\bar{A}_1^{-1}\bar{B}\| \leq \|(\bar{L}_1')^{-1}\| \cdot \|L_1^{-1}\| \cdot \|B\| \tag{60}$$

But, because $\bar{L}_1$ and L_1 are Cholesky factors, we obtain

$$\|(\bar{L}_1')^{-1}\| = \sqrt{\rho(\bar{A}_1^{-1})} \leq \frac{1}{\sqrt{\gamma}} \tag{61}$$

and

$$\|L_1^{-1}\| = \sqrt{\rho(A_1^{-1})} \leq \frac{1}{\sqrt{\gamma}} \tag{62}$$

For $\|B\|$ we can write (using the symmetry of A)

$$\|B\| = \sqrt{\rho(B'B)} \leq \sqrt{\|B'B\|_\infty} \leq \sqrt{\|B'\|_\infty \cdot \|B\|_\infty} \leq \|A\|_\infty \tag{63}$$

Then, introducing (61)-(63) in (60) we obtain (58).

We shall denote by β_m the positive constant

$$\beta_m = \frac{1}{\gamma^3} \frac{\max\{\bar{a}_{ii}, 1 \leq i \leq n-p\} \cdot \{\bar{a}_{ii}, 1 \leq i \leq n-p\}}{\min\{a_{ii}, n-p < i \leq n\}} \cdot \| A \|_\infty^2 \tag{64}$$

where $m \in \{1, ..., q-1\}$ is the arbitrary level considered at the begining of this section. Accordingly to (44), (57), (58) and (64) we obtain

$$|||Te|||_1^2 \leq \beta_m |||e|||_2^2, \tag{65}$$

i.e. the approximation assumption (15) (on the level m). Defining $\beta > 0$ by

$$\beta = \max\{\beta_m, 1 \leq m \leq q-1\}, \tag{66}$$

from (65) it results for every $e^m \in \mathbb{R}^{n_m}$,

$$|||T^m e^m|||_1^2 \leq \beta |||e^m|||_2^2, \ (\forall)\, m = 1, ..., q-1, \tag{67}$$

where T^m is the same matrix with T from (38) (on the level m).

3. The smoothing assumption for the preconditioned system. We obtained in (67) the approximation assumption for the preconditioned system with respect to the norms $|||\cdot|||_i$, $i = 1, 2$ defined with the inner products from (12) for the preconditioned matrix $\bar{A}$. Thus, it is necessary that the smoothing assumption be also fulfilled with respect to these norms. This is the aim of the present section.

We shall mentain the notational conventions from the above section. Firstly we observe that a relaxation step of the type (8) can be written in the form

$$u_{new} = M^{-1} N u_{old} + M^{-1} b, \tag{68}$$

where

$$A = M - N \tag{69}$$

is a splitting of the matrix A with M invertible and

$$\rho(M^{-1}N) < 1, \tag{70}$$

(indeed, it is sufficient to define $G = M^{1}N$ and from (68) we get (8)). Suppose that relaxation (68) satisfies the smoothing assumption (14) (on the level m) with a constant $\alpha_m > 0$, i.e. $(\forall) e \in \mathbb{R}^r$

$$\|M^{-1}Ne\|_1^2 \leq \|e\|_1^2 - \alpha_m \|e\|_2^2 \tag{71}$$

We shall define now (only for theoretical purpose!) for the preconditioned system (25) a similar relaxation, i.e.

$$\bar{u}_{new} = \bar{M}^{-1}\bar{N}\bar{u}_{old} + \bar{M}^{-1}\bar{b}, \tag{72}$$

where the matrices $\bar{M}$ and $\bar{N}$ are given by

$$\bar{M} = \Delta_1 M \Delta_1^t,\ \bar{N} = \bar{\Delta}_1 N \bar{\Delta}_1^t. \tag{73}$$

We denote by e, $\bar{e}$ respectively the errors

$$e = u_{old} - u \tag{74}$$

and

$$\bar{e} = \bar{u}_{old} - \bar{u} \tag{75}$$

THEOREM 3. (i) *$\bar{M}$ is invertible, $\bar{A} = \bar{M} - \bar{N}$ and*

$$\rho(\bar{M}^{-1}\bar{N}) < 1. \tag{76}$$

(ii) *If*

$$\bar{u}_{old} = (\bar{\Delta}_1^t)^{-1} u_{old} \tag{77}$$

then

$$\bar{u}_{new} = (\bar{\Delta}_1^t)^{-1} u_{new}. \tag{78}$$

(iii) *The relaxation* (72) *satisfies the smoothing assumption with the same constant* α_m, i.e.

$$|||\bar{M}^{-1}\bar{N}\bar{e}|||_1^2 \le |||\bar{e}|||_1^2 - \alpha_m |||\bar{e}|||_2^2. \tag{79}$$

Proof. (i) The first two statements are obvious. For the third, using the well known equality $\rho(AB) = \rho(BA)$ (see e.g. [19]) we obtain

$$\rho(\bar{M}^{-1}\bar{N}) = \rho((\bar{\Delta}_1^t)^{-1} M^{-1} N (\bar{\Delta}_1^t)) = \rho(M^{-1}N) < 1$$

(ii) It results by simple computations using (68), (72), (73), (25) and (26).

(iii) From (26) and (77) we have

$$\bar{e}_{old} = (\bar{\Delta}_1^t)^{-1} e_{old} \tag{80}$$

Then, it is sufficient to observe, using (73), that

$$<\bar{A}\bar{M}^{-1}\bar{N}\bar{e}_{old}, \bar{M}^{-1}\bar{N}\bar{e}_{old}> = <AM^{-1}Ne_{old}, M^{-1}Ne_{old}>,$$

$$<\bar{A}\bar{e}_{old}, \bar{e}_{old}> = <Ae_{old}, e_{old}>,$$

$$<\bar{D}^{-1}\bar{A}\bar{e}_{old}, \bar{A}\bar{e}_{old}> = <D^{-1}Ae_{old}, Ae_{old}>$$

and the prof is complete.

Remark. From the assertion (ii) of the above theorem we obtain the following usefull fact: computing $\bar{u}_{new}$ with (72) and a given approximation $\bar{u}_{old}$ is the same as computing u_{new} with (68) and u_{old} given by

$$u_{old} = \bar{\Delta}_1^t \bar{u}_{old} \tag{81}$$

and calculate

$$\bar{u}_{new} = (\bar{\Delta}_1^t)^{-1} u_{new}, \tag{82}$$

In this way, the relaxation proces (72), for the preconditioned system (25), can be carried out using a classical relaxation of the type (68) for the initial system and the relations (81)-(82).

Like in the previous section we can now define

$$\alpha = \min\{\alpha_m, 1 \le m \le q-1\} \tag{83}$$

Then, over denoting $\bar{G} = \bar{M}^{-1}\bar{N}$ from (79) by G^m and $\bar{e}$ by e^m we obtain

$$|||G^m e^m|||_1^2 \le |||e^m|||_1^2 - \alpha \cdot |||e^m|||_2^2, \; (\forall) m = 1, \ldots, q-1, \tag{84}$$

i.e. the smoothing assumption (14).

4. The convergence of the algebraic multigrid algorithm. Accordingly to the Theorem 1 we obtain that the V - cycle type multigrid algorithm defined in section 2 converges to the exact solution u of (1) and the convergence factor, in the energy norm of the preconditioned matrix A is bounded above by

$$\bar{\rho} = \sqrt{1 - \alpha/\beta} \tag{85}$$

with α and β from (83) and (66) respectively.

We have the possibility (see the next section) to obtain γ from (57) independently of the dimension and the spectrum of the matrices A and $\bar{A}$. Thus, the constants α_m and β_m from (64) and (79) will depend only on the coefficients of the matrices A^m and $\bar{A}^m$ ($\bar{A}^m$ is $\bar{A}$ on the level m). But, unfortunately, in the general case, α and β, and so $\bar{\rho}$ from (85), will depend on the number of levels used in the V - cycle. It is very hard, even in particular cases, to find a theoretical value of the factor $\bar{\rho}$. The only way is to use an accurate coarsening process and to define an efficient interpolation such that the coarse grid matrices keep the properties of the initial matrix.

In our case an encouraging aspect comes to helps us. Indeed, from the relation (31) it results that the coarse grid matrix A_p, obtain with the Galerkin approach (6) and I_p^n from (29), don't depend on the preconditioning. More than that, A_p is the Schur complement of A_1 obtained with Gaussian elimination. But there exist results (see e.g. [9]) which say that, for example, A is (weakly) diagonally dominant, A_p keeps this property a.s.o. In this way we can controle the coefficients of A_p, their signs, absolute values, positions (i.e. the sparsity of the matrix). Thus, defining interpolations like in (29) the only problem is 'to properly choose' A_1 (and $\bar{A}_1$) for that the 'extra work' and the computational costs be not too expensive.

Remark. Choosing A_1 means, from (19) and (20), choosing the coarse grid C_p. Some facts related to this aspect can be found in the papers [4], [17], [16]. Concerning the (spectral) condition of the preconditioned matrix $\bar{A}$ denoted by $k(\bar{A})$, we can easly obtain some precise informations. Indeed, from (27) we have

$$k(\bar{A}) = \|\bar{A}\|\cdot\|\bar{A}^{-1}\| \le k(A)\cdot\|\bar{\Delta}_1^t\bar{\Delta}_1\|\cdot\|(\bar{\Delta}_1^t)^{-1}\bar{\Delta}_1^{-1}\|. \tag{86}$$

From (20) and (57) we obtain

$$\|\bar{\Delta}_1^t\bar{\Delta}_1\| \le \frac{\|\bar{A}_1\|}{\gamma},\ \|(\bar{\Delta}_1^t)^{-1}\bar{\Delta}_1^{-1}\| \le \frac{\|A_1\|}{\gamma}. \tag{87}$$

Then, (86), (87) and similar arguments with A instead $\bar{A}$ get

$$k(A)\frac{\gamma^2}{\|A_1\|\cdot\|\bar{A}_1\|} \le k(\bar{A}) \le \frac{\|A_1\|\cdot\|\bar{A}_1\|}{\gamma^2}k(A), \tag{88}$$

with $k(\bar{A})$ the spectral condition number of A.

Then, for an accurate and realistic γ in (57), $k(A)$ is of the same order with $k(A)$ and the convergence in the norm $|||\cdot|||_1$ will not deteriorate the results.

5. Some particular cases.

I. $\bar{A}_1 = A_1$. Then $\bar{A} = A$ thus no preconditioning occurs. Condition (57) will hold if, for example A_1 is strictly diagonally dominant, i.e.

$$\nu_i = a_{ii} - \sum_{j=1, j\neq i}^{n-p} |a_{ij}| > 0, i = 1, \dots, n-p. \tag{89}$$

Then, we can take γ from (57) to be (from Gershgorin's theorem, [19])

$$\gamma = \min\{\nu_i, i = 1, \dots, n-p\}. \tag{90}$$

The interpolation operator will be given by (see also [13]).

$$I_p^n = \begin{bmatrix} -A_1^{-1}B \\ I_2 \end{bmatrix}. \tag{91}$$

The following result gives us a way for constructing I_p^n without inverting the matrix A_1. Firstly we have to observe that, the matrix A being positive definite (and symmetric) we can perform the Gaussian elimination algorithm without pivoting ([16]) and making 1 on the diagonal of A_1, for the first n-p columns. After that we obtain a matrix $\tilde{A}$ of the form (in block notation)

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{B} \\ O & \bar{A}_2 \end{bmatrix}. \tag{92}$$

or elementwise

$$\tilde{A} = \begin{bmatrix} 1 & \tilde{a}_{12} & \tilde{a}_{13} & \dots & \tilde{a}_{1,n-p} & \tilde{a}_{1,n-p+1} & \dots & \tilde{a}_{1n} \\ o & 1 & \tilde{a}_{23} & \dots & \tilde{a}_{2,n-p} & \tilde{a}_{2,n-p+1} & \dots & \tilde{a}_{2n} \\ \dots & \dots & & & & & & \\ 0 & 0 & 0 & \dots & 1 & \tilde{a}_{n-p,n-p+1} & \dots & \tilde{a}_{n-p,n} \\ 0 & 0 & 0 & \dots & 0 & \tilde{a}_{n-p+1,n-p+1} & \dots & \tilde{a}_{n-p+1,n} \\ \dots & \dots & & & & & & \\ 0 & 0 & 0 & \dots & 0 & \tilde{a}_{n,n-p+1} & \dots & \tilde{a}_{nn} \end{bmatrix} \tag{93}$$

For $k = 1, \dots, n\text{-}p$ we define the matrices H_k of dimension $(n\text{-}k) \times (n\text{-}k+1)$ and H of dimension $p \times n$ by

$$H_k = \begin{bmatrix} -\tilde{a}_{k,k+1} & 1 & 0 & \dots & 0 \\ -\tilde{a}_{k,k+2} & 0 & 1 & \dots & 0 \\ -\tilde{a}_{k,n} & 0 & 0 & \dots & 1 \end{bmatrix} \tag{94}$$

and

$$H = H_{n-p} H_{n-p-1} \dots H_1. \tag{95}$$

Observation. The first column of H_k (without minus sign) is the k - row of the matrix $[\tilde{A}_1 \vdots \tilde{B}]$ from (93) without the 1 on the diagonal.

THEOREM 4. *With the above considerations we have*

$$I_n^p = H. \tag{96}$$

Proof. It results from (94) and (95) that the matrix H has the structure

$$H = [\tilde{H} \vdots I_2], \tag{97}$$

where I_2 is the identity on $\mathbb{R}^p$ and $\tilde{H}$ is a $p \times (n\text{-}p)$ real matrix. We observe that the first column of H_1 is given by

$$\tilde{a}_{1k} = a_{1k}/a_{11},\ k = 2, \tag{98}$$

Thus, in block notation,

$$H_1 A = [O \,\vdots\, A^{(1)}], \tag{99}$$

or elementwise

$$H_1 A = \begin{bmatrix} 0 & a_{22} - (a_{21}a_{12})/a_{11} & \cdots & a_{2n} - (a_{21}a_{1n})/a_{11} \\ 0 & a_{32} - (a_{31}a_{12})/a_{11} & \cdots & a_{3n} - (a_{31}a_{1n})/a_{11} \\ 0 & a_{n2} - (a_{n1}a_{12})/a_{11} & \cdots & a_{nn} - (a_{n1}a_{1n})/a_{11} \end{bmatrix}. \tag{100}$$

From the symmetry of A it results that the matrix $A^{(1)}$ from (100) is the same with the square matrix of dimension $(n-1) \times (n-1)$ obtained after the first step of the Gaussian elimination by neglecting the first row and column. Recursively we obtain that

$$H_{n-p} H_{n-p-1} \ldots H_1 A = [O \,\vdots\, \tilde{A}_1]. \tag{101}$$

But, from (101), (20) and (97) it results

$$\tilde{H} A_1 + B^t = O, \tag{102}$$

which gives us

$$\tilde{H} = -B^t A_1^{-1}, \tag{103}$$

COROLLARY 2. *The interpolation operator* I_p^n *and the coarse grid matrix* A_p *are given by*

$$I_p^n = H_1^t H_2^t \ldots H_{n-p}^t, \tag{104}$$

$$A_p = H_{n-p} \ldots H_2 H_1 A H_1^t H_2^t \ldots H_{n-p}^t. \tag{105}$$

Remark. We observe that for the construction of I_p^n (or I_n^p) we must perform the Gaussian elimination only on the matrix A_1 (i.e. only for the n-p rows of A).

II. $\bar{A}_1 = \text{diag}(d_1, d_2, \dots, d_{n-p})$ where we suppose that

$$d_i > 0,\ i = 1, \dots, n-p. \tag{106}$$

Then

$$\bar{L}_1 = \bar{L}_1^t = \text{diag}\,(d_1^{\frac{1}{2}}, d_2^{\frac{1}{2}}, \dots, d_{n-p}^{\frac{1}{2}}) \tag{107}$$

and the interpolation I_p^n is given by

$$I_p^n = \begin{bmatrix} -\bar{L}_1^{-1} L_1^{-1} B \\ I_2 \end{bmatrix}. \tag{108}$$

In order to obtain the product $L_1^{-1}B$ (with L_1 the Cholesky factor of A_1, from (21)) we make a Gaussian elimination (Without pisoting and making 1 on the diagonal) on the first n-p rows of A. In this way we obtain the matrix

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{B} \\ B^t & A_2 \end{bmatrix} \tag{109}$$

where

$$A_1 = \tilde{L}_1 \tilde{A}_1 \tag{110}$$

is an (LU) - decomposition of A_1 ($\tilde{A}_1$ is upper triangular with 1 on his diagonal) and

$$\tilde{B} = \tilde{L}_1^{-1} B \tag{111}$$

Then, if $\tilde{D}_1 = \text{diag}\,(\tilde{L}_1) = \text{diag}\,(\tilde{l}_{11}, \tilde{l}_{22}, \dots, \tilde{l}_{n-p,n-p})$ it is obvious that

$$L_1^t = \tilde{D}_1^{1/2} \tilde{A}_1 \tag{112}$$

Then elements of the matrix $\tilde{D}_1$ can be recursively obtained by the formulas

$$a_{11} = \tilde{l}_{11},\ a_{ii} = \tilde{l}_{ii} + \sum_{k=1}^{i-1} \tilde{l}_{kk} \cdot \tilde{a}_{ki}^2,\ i = 2, \dots, n-p \tag{113}$$

(where $\tilde{a}_{ij}$ are the elements of $\tilde{A}_1$). Then we have

$$L_1^{-1} B = \tilde{D}_1^{1/2} \cdot \tilde{B} \tag{114}$$

The constant γ from (57) can be taken as

$$\gamma = \min \{\nu_i, d_i,\ i = 1, \dots, n-p\} \tag{115}$$

III. $\bar{A}_1 = A_1 + R_1$ where

$$A_1 = \bar{A}_1 - R_1 \tag{116}$$

is an incomplete Cholesky decomposition of A_1 (if A_1 is supposed to be an M - matrix, cf. [11]). The factor $\bar{L}_1$ is obtained during this decomposition. We know from [11] that

$$\rho = \rho(\bar{A}_1^{-1} R_1) < 1 \tag{117}$$

From (116) we obtain

$$\bar{A}_1^{-1} A_1 = I - \bar{A}_1^{-1} R_1 \tag{118}$$

Thus, if $\lambda \in \mathbf{C}$ is an eigenvalue of $\bar{A}_1^{-1} A_1$, $1-\lambda$ will be an eigenvalue for $\bar{A}_1^{-1} R_1$ and (using (117))

$$|1 - |\lambda|| \le |1 - \lambda| \le \rho \tag{119}$$

From (119) it results that for every eigenvalue λ of $\bar{A}_1^{-1} A_1$

$$1 - \rho \le |\lambda| \le 1 + \rho \tag{120}$$

In particular

$$1-\rho \le \rho(\bar{A}_1^{-1} A_1) \le \|\bar{A}_1^{-1} A_1\| \le \|\bar{A}_1^{-1}\| \cdot \|A_1\| \tag{121}$$

and

$$\lambda_{\min}(\bar{A}_1) = \frac{1}{\|\bar{A}_1^{-1}\|} \leq \frac{\|A_1\|}{1-\rho} \leq \frac{\|A_1\|_\infty}{1-\rho} \tag{122}$$

Thus, γ from (57) can be taken as

$$\gamma = \min\left\{\min\{\nu_i, i = 1, \dots, n-p\}, \frac{\|A_1\|_\infty}{1-\rho}\right\} \tag{123}$$

Remark. Relation (123) tells us that the number 1-ρ must not depend on the dimension of the matrix A_1. Thus, the ILU-decomposition (116) must not be 'too incomplete', i.e. the matrix R_1 must not have too much nonempty entries, the 'ideal' case being

$$R_1 = 0, \tag{124}$$

i.e. our particular case I.

6. Nummerical examples. We considered the following plane problems:

Dirichlet
$$\begin{cases} -\Delta u = f \text{ in } \Omega \\ u = 0 \textit{ on } \partial\Omega \end{cases}$$

Anisotropic Poisson
$$\begin{cases} -\varepsilon \cdot \dfrac{\partial^2 u}{\partial x^2} - \dfrac{\partial^2 u}{\partial y^2} = f \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega \end{cases}$$

Helmholtz
$$\begin{cases} \Delta u + k^2 u = f \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega \end{cases}$$

with $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$, discretized by a classical 5-point stencil finite differences (see e.g. [8]). We used two different initial (finest grid) discretizations (corresponding to meshsizes $h = 1/14$ and $h = 1/32$) and a 5 - grids V - cycle algebraic multigrid (see section 1). We applied the preconditioning methods from cases I and II (section 5). As relexation we used the classical Gauss - Seidel method ([19]). The stopping criterion of the multigrid algorithm was

$$|||u^N - u|||_1 \leq 10^{-6} \tag{125}$$

where u is the exact solution and u^N the corresponding approximation (N is the minimum number of iteration such that (125) holds).

In tables 1-4 we indicated the worst norm reduction factor per iteration step, ρ, computed with the formula

$$\rho = \sup\left\{ \frac{|||e^{j+1}|||_1}{|||e^j|||_1},\ j = 1, \ldots, N-1 \right\} \tag{126}$$

for Dirichlet and anisotropic Poisson problems and

$$\rho = \sup\left\{ \frac{\|e^{j+1}\|}{\|e^j\|},\ j = 1, \ldots, N-1 \right\} \tag{127}$$

for Helmholtz equation ($e^j = u^j - u$ is the error at the j-th iteration of the multigrid algorithm).

Remarks 1. For coarsening we used the algorithm presented in the paper [16].

2. In the case of Helmoltz equation the algebraic system is symmetric but not more positive definite. But following the results of Mandel ([10]), the condition (33), with $\lambda_{min}(\bar{A}_1)$ not depending on the dimension of the initial matrix A, ensures the convergence of the two grid algorithm even in the indefinite case.

3. Some improvements in order to avoid the fill - in process appearing sometimes in

the coarses grids matrices were presented in [7].

4. The values of ε (table 2) and k^2 (tables 3 and 4) were selected accordingly to similar examples solved in papers [17] and [5] respectively.

Acknowledgement. The author thanks to Mr. Gabriel Golubovici, Departement of Mathematics, Univ. of Constanţa, for his help preparation of numerical examples.

h	1/14	1/32
ρ for case I	0.051	0.078
ρ for case II	0,19	0.4

Table 1. The Dirichlet problem

h		1/14	1/32
ρ for case I	$\varepsilon = 10^{-1}$	0.052	0.078
	$\varepsilon = 10^{-2}$	0.052	0.078
	$\varepsilon = 10^{-6}$	0.054	0.079
ρ for case II	$\varepsilon = 10^{-1}$	0.19	0.41
	$\varepsilon = 10^{-2}$	0.2	0.41
	$\varepsilon = 10^{-6}$	0,23	0.42

Table 2. The anisotropic Poisson problem

ρ for case I	$k^2 = 4$	0.054
	$k^2 = 19$	0.058
	$k^2 = 25$	0.09
	$k^2 = 30$	0.37
ρ for case II	$k^2 = 4$	0.21
	$k^2 = 10$	0.27
	$k^2 = 25$	0.48
	$k^2 = 30$	0.74

Table 3. The Helmmoltz problem, $h = 1/14$.

ρ for case I	$k^2 = 19$	0.077
	$k^2 = 55$	0.34
	$k^2 = 100$	0.83
ρ for case II	$k^2 = 19$	0.56
	$k^2 = 55$	0.8
	$k^2 = 100$	0.97

Table 4. The Helmholtz problem, h = 1/32.

REFERENCES

1. D. Adam, *Mesh independence of Galerkin approach by preconditioning*. Intern. J. Computer Math., vol. 28, pp. 227-236, 1989.
2. O. Axelsson, P.S. Vassilevski, *Algebraic multilevel preconditioning methods* I. Numer. Math, 56, pp. 157-177, 1989.
3. O. Axelsson, M. Neytcheva, *Algebraic multilevel iteration method for Stieltjes matrices*. Report 9102, Dept. of Math., Catholic Univ. Toemooiveld, 1991.
4. A. Brandt, *Algebraic multigrid theory: the symmetric case*. Preprint, Weizmann Inst. Of Science, Rehovot, Israel, 1983.
5. A. Brandt, S. Ta'assan, *Multigrid method for nearly singular and slighty indefinite problems*. Lecture Notes in Math., vol. 1228, Springer, Berlin, 1986.
6. G.H. Golub, C.F. van Loan, *Matrix computations*. John's Hopkins Univ. Press, Baltimore, 1983.
7. G. Golubovici, C. Popa, *Interpolation and related coarsening techniques for the algebraic multigrid method*. Paper presented at the 4th European Conference on Multigrid Methods, 6-9 July 1993, Amsterdam.
8. W. Hackbusch, *Multigrid methods and applications*. Springer, Berlin, 1985.
9. A.S. Honseholder, The Theory of matrices in numerical analysis. Blaisdell Publ. Comp., New York, 1964.
10. J. Mandel, *On multigrid and iterative aggregation methods for nonsymmetric problems*. Lecture Notes in Math., vol. 1228, Springer, Berlin, 1986.
11. J.A. Meijerink, H.A. van der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*. Math. of Comput., vol. 31, nr. 157, pp. 148-162, 1977.
12. C. Popa, *On smoothing properties of SOR relaxation for algebraic multigrid method*. Studii şi Cerc. Mat., vol. 5, pp. 399-406, 1989.
13. C. Popa, *ILU decomposition for coarse grid correction step on algebraic multigrid*. In GMD - Studien, nr. 189, 1991.
14. C. Popa, *The construction of the interpolation operator with ILU decomposition for algebraic positive definite systems*. In Proceed. of the Eight GAMM - Seminar, Notes on Numerical Fluid Mech., vol. 41, Vieweg, Braunschweig, 1993).
15. C. Popa, *Mest independence of the condition number of discrete Galerkin systems by preconditioning*. To appear in Intern, J. Computer Math., London, 1993.

16. C. Popa, *A coarsening algorithm for the algebraic method.* To appear in Studii şi Cerc. Matem., Romanian Academy of Sciences, 1994.
17. J. Ruge, K. Stuben, *Algebraic multigrid (AMG).* In "Multigrid Methods", S. McCormick eed., SIAM, 1987.
18. K. Stuben, U. Trottenberg, *Multigrid methods. Fundamental algorithms, model problem analysis and apllications.* In Lecture Notes in Math., vol. 960, Springer, Berlin, 1982.
19. R. Varga, *Matrix iterative analysis.* Pretince - Hall, New Jersey, 1962.
20. H. Yserentant, *Hierarchical bases of finite - elements spaces in the discretization of nonsymmetric elliptic b.v.p.* Computing, vol. 35, pp. 39-49, 1985.

Tehnoredactare computerizată: Lucia Săcelean, Marcela Topliceanu

următoarele serii:

matematică (trimestrial)
fizică (semestrial)
chimie (semestrial)
geologie (semestrial)
geografie (semestrial)
biologie (semestrial)
filosofie (semestrial)
sociologie-politologie (semestrial)
psihologie-pedagogie (semestrial)
științe economice (semestrial)
științe juridice (semestrial)
istorie (semestrial)
filologie (trimestrial)
teologie ortodoxă (semestrial)
educație fizică (semestrial)

In the XL-th year of its publication (1995) *Studia Universitatis Babeș-Bolyai* is issued in the following series:

mathematics (quarterly)
physics (semesterily)
chemistry (semesterily)
geology (semesterily)
geography (semesterily)
biology (semesterily)
philosophy (semesterily)
sociology-politology (semesterily)
psychology-pedagogy (semesterily)
economic sciences (semesterily)
juridical sciences (semesterily)
history (semesterily)
philology (quarterly)
orthodox theologie (semesterily)
physical training (semesterily)

Dans sa XL-e année (1995) *Studia Universitatis Babeș-Bolyai* paraît dans les séries suivants:

mathématiques (trimestriellement)
physique (semestriellement)
chimie (semestriellement)
geologie (semestriellement)
géographie (semestriellement)
biologie (semestriellement)
philosophie (semestriellement)
sociologie-politologie (semestriellement)
psychologie-pédagogie (semestriellement)
sciences économiques (semestriellement)
sciences juridiques (semestriellement)
histoire (semestriellement)
philologie (trimestriellement)
théologie orthodoxe (semestriellement)
éducation physique (semestriellement)