Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

953

Iterative Solution of Nonlinear Systems of Equations

Proceedings, Oberwolfach 1982

Edited by R. Ansorge, Th. Meis, and W. Törnig

Springer-Verlag
Berlin Heidelberg New York

A DEVICE FOR THE ACCELERATION OF CONVERGENCE OF A MONOTONOUSLY ENCLOSING ITERATION METHOD

H. Cornelius and G. Alefeld

Inst. f. Angew. Mathematik
Universität Karlsruhe
Kaiserstrasse 12

7500 Karlsruhe
Germany

1. Introduction

In this paper we consider a class of iteration methods for solving simultaneous systems of nonlinear equations. These methods compute in each iteration step lower and upper bounds for all components of the unknown solution vector. Enclosing the solution repeatedly is under practical consideration advantageous since rounding outwards in a systematic manner one has guaranteed error bounds for the solution. Especially for very large systems this seems to be of great importance since one has observed that - using an arbitrary iteration method - the method comes to a rest although the iterates are still far away from the solution.
The main advantage of the methods considered in this paper consists in the fact that for certain classes of problems (which actually occur in practice) they are convergent to the solution under weaker conditions than known methods which also enclose the solution monotonously. For example, we don't have to assume convexity or similar conditions from which convexity can be derived. If these methods are applied to large systems which originate from the approximation of partial differential equations then the convergence is extremely slow. In this paper we discuss a simple device for constructing a sequence of real vectors which is faster convergent to the solution than the bounds of the enclosing vectors.

2. The method (INSI) and some theoretical results

We assume that the reader has a certain knowledge of interval-analysis to the extent one can find it, for example, in [2].All facts from interval-analysis which are only mentioned without proof can be found in [2]. In this paper we denote real numbers and real n-vectors by $x,y,\ldots$. Real matrices are denoted by $X,Y,\ldots$. Real compact intervals

and vectors, the components of which are intervals, are denoted by [x],[y],... . Similarly interval-matrices are denoted by [X],[Y],... . d([x]) denotes the width (or diameter) of the interval [x]. $|[x]| = \max_{x \in [x]} |x|$ is called absolute value of the interval [x]. For interval-vectors and interval-matrices these concepts are defined via the elements. For example, if $[A] = ([a_{ij}])$ is an n by n interval-matrix then $d([A]) = (d([a_{ij}]))$ is a real n by n matrix. If $f : \mathbb{R}^n \to \mathbb{R}^n$ has a derivative then f'([x]) denotes the so-called interval-arithmetic evaluation of the derivative over the interval-vector [x]. See, for example [2], Section 3. The interval-arithmetic evaluation of the derivative is an interval-matrix. We consider the splitting

$$f'([x]) = D([x]) - L([x]) - U([x]) \tag{1}$$

of this matrix where D([x]) denotes the diagonal part and where L([x]) and U([x]) are the parts below and above the main diagonal, respectively.

We start with an interval-vector $[x]^0$ and consider a sequence of interval-vectors $[x]^k$, which are computed by the following iteration method:

$$\left.\begin{array}{l} \left\{\begin{array}{l} \text{Choose } m([x]^k) \in [x]^k \quad (m([x]^k) \text{ a real n-vector}) \\ [y]^{k+1} = m([x]^k) - \tilde{D}([x]^k)\{L([x]^k)(m([x]^k) - [y]^{k+1}) + \\ \qquad\qquad + U([x]^k)(m([x]^k) - [x]^k) + f(m([x]^k))\} \\ [x]^{k+1} = [y]^{k+1} \cap [x]^k \end{array}\right. \\ k = 0,1,2,\ldots . \end{array}\right\} \tag{2}$$

The diagonal interval-matrix $\tilde{D}([x]^k)$ is defined in the following manner. Let $D([x]^k) = \operatorname{diag}(d_{ii}([x]^k))$. Then $\tilde{D}([x]^k) = \operatorname{diag}(1/d_{ii}([x]^k))$. (Please note that the notation $D([x]^k)^{-1}$ would not make sense since for interval-matrices no inverses exist in the ordinary sense). For clarity we stress the fact that the real n-vector $m([x]^k) \in [x]^k$ can be chosen arbitrarily in the interval-vector $[x]^k$.

The method (2) is called interval-arithmetic version of the Newton-single-step method with forming intersections (INSI). The method was introduced in [1] where, however, only the case was considered that $m([x]^k)$ is the center of $[x]^k$.

The following results hold for (INSI). For our considerations the result about the asymptotic convergence factor is of fundamental importance.

Theorem 1. Let $f : \mathbb{D} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a mapping which has a continuous derivative on the open set $\mathbb{D}$. Assume that f has a zero x^* in $\mathbb{D}$. Furthermore we assume that for all $[x]^0 \subseteq \mathbb{D}$ with $x^* \in [x]^0$ the interval-arithmetic evaluation $f'([x]^0)$ of the derivative exists and that $f'([x]^0)$ is split according to (1). Let $0 \notin d_{ii}([x]^0)$, $1 \leq i \leq n$, where $D([x]^0) = \operatorname{diag}(d_{ii}([x]^0))$.

a) If $\rho\left(\left(|D(x^*)| - |L(x^*)|\right)^{-1} \cdot |U(x^*)|\right) < 1$ (ρ denotes the spectral-radius) where

$$f'(x^*) = D(x^*) - L(x^*) - U(x^*)$$

then $\lim_{k\to\infty} [x]^k = x^*$ for all interval-vectors which have sufficiently small width $d([x]^0)$ and for which $x^* \in [x]^0$ holds.

b) For the asymptotic convergence factor of the method (INSI) it holds that

$$R_1((\text{INSI}), x^*) \leq \rho\left(\left(|D(x^*)| - |L(x^*)|\right)^{-1} \cdot |U(x^*)|\right).$$

If $D(x^*) \geq 0$, $L(x^*) \geq 0$, $U(x^*) \geq 0$ then the equality-sign holds in this last inequality.

■

For the definition of the asymptotic convergence factor of a method which computes interval-vectors see [2], Appendix A. The very long proof of Theorem 1 can be found in [3].

The convergence result $\lim_{k\to\infty} [x]^k = x^*$ from the preceding Theorem is a local one. Under certain assumptions about the interval-arithmetic evaluation $f'([x]^0)$ of the derivative we can get explicit conditions under which the statement $\lim_{k\to\infty} [x]^k = x^*$ holds. These conditions are as follows:

The interval-matrix $[A] = ([a_{ij}])$ has property (K) iff

$$a^1_{ij}\, a^2_{ij} \geq 0 , \quad 1 \leq i,j \leq n ,$$

holds, where $[a_{ij}] = [a^1_{ij}, a^2_{ij}]$.

In order to formulate the next result we need the concept of an M-matrix. A real n by n matrix $A = (a_{ij})$ is called M-matrix iff $a_{ij} \leq 0$, $i \neq j$, and $A^{-1} \geq 0$. See [8], for example.

Theorem 2. The simultaneous system of nonlinear equations is assumed to have a zero x^* in $\mathbb{D}$. Furthermore we assume that there exists an interval-vector $[x]^0 \subseteq \mathbb{D}$ with $x^* \in [x]^0$ for which the interval-arith-

metic evaluation $f'([x]^0)$ of the derivative exists. Assume that all real matrices from $f'([x]^0)$ are M-matrices. Then the method (INSI) is well-defined and the following hold:

a) $x^* \in [x]^k$, $k \geq 0$.

b) $[x]^0 \supseteq [x]^1 \supseteq \ldots \supseteq [x]^k \supseteq [x]^{k+1} \supseteq \ldots$

c) $\lim_{k\to\infty} [x]^k = x^*$

d) $R_1\{(INSI),x^*) = \rho\left(\left(D(x^*) - L(x^*)\right)^{-1} U(x^*) \right)$. ∎

The proof may be performed by using the fact that under the assumptions of this Theorem $f'([x]^0)$ has property (K). For details see [3].

3. Application to elliptic difference equations

We are now going to demonstrate that the assumptions of the preceding Theorem can be realized with nonlinear systems which originate from elliptic boundary problems by replacing the derivatives by finite differences. We consider the partial differential equation

$$- F(x,y,u,u_x,u_y,u_{xx},u_{yy}) = 0 \quad \text{in} \quad R \subset \mathbb{R}^2$$

and the boundary condition

$$u(x,y) = \gamma(x,y) \qquad \text{on} \quad \partial R .$$

R denotes a simply connected bounded region with boundary ∂R.

We assume that F has derivatives with respect to u_{xx} and u_{yy} for which

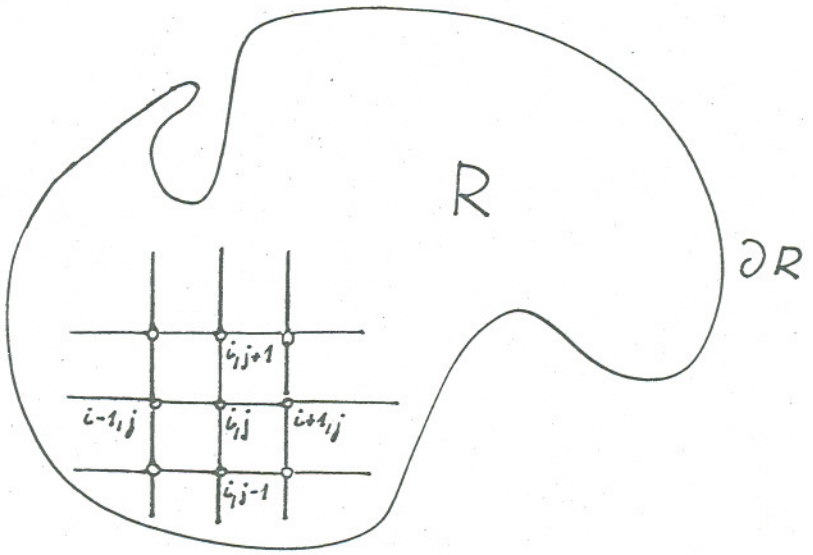
$$F_{u_{xx}} \geq m > 0, \quad F_{u_{yy}} \geq m > 0$$

hold. We choose a fixed step-size h in both directions, replace the derivatives by central difference quotients and obtain - after neglecting the discretization error - at the point (x_i,y_j) the equation

$$- g_{ij}\left(x_i,y_j,u_{ij}, \frac{u_{i+1,j} - u_{i-1,j}}{2h}, \frac{u_{i,j+1} - u_{i,j-1}}{2h}, \right.$$

$$\left. \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2}, \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{h^2} \right) = 0 .$$

$u_{i,j}$ is an approximation for the unknown function value $u(x_i,y_j)$. Setting $z_k = u_{ij}$, $z = (z_i)$, these equations may be gathered up to $G_1(z) = 0$ where the number n of equations is the same as the number



of unknowns. (We omit the details which are necessary to perform the approximation of the differential equation and the boundary conditions for points which are close to the boundary ∂R of R.)

We assume that $F_u \leq 0$ and that h is choosen such small that

$$\left|F_{u_x}\right| \frac{h}{2} < F_{u_{xx}}, \quad \left|F_{u_y}\right| \frac{h}{2} < F_{u_{yy}}$$

hold. Under these conditions it holds that for all $z \in \mathbb{R}^n$ the derivative $G_1'(z)$ is an M-matrix. See [7]. From this it follows that $G_1 : \mathbb{R}^n \to \mathbb{R}^n$ is a so-called M-function, which implies that $G_1(z) = 0$ has at most one solution. See [5] and [6]. Assume now that G_1 is surjective. (Sufficient conditions for the surjectivity of an M-function may be found in [5] and [6]).Then $G_1(z) = 0$ has exactly one solution x^*. By continuity it then follows that for sufficiently small width $d([x]^0)$ of $[x]^0$ and with $x^* \in [x]^0$ all matrices from the interval-arithmetic evaluation $G_1'([x]^0)$ of the derivative of G_1 are M-matrices. Hence the assumptions of Theorem 2 hold for these interval-vectors.

If F has the special form

$$-F(x,y,u,u_x,u_y,u_{xx},u_{yy}) =$$
$$-\left(A(x,y)u_x\right)_x - \left(C(x,y)u_y\right)_y + f(x,y,u)$$

where $A \geq m > 0$, $C \geq m > 0$, $f_u \geq 0$, then it is advantageous to use

the following approximations:

$$(Au_x)_x \approx \frac{1}{h^2}\{A(x+\tfrac{h}{2},y)[u(x+h,y) - u(x,y)] -$$
$$- A(x-\tfrac{h}{2},y)[u(x,y) - u(x-h,y)]\} ,$$

$$(Cu_y)_y \approx \frac{1}{h^2}\{C(x,y+\tfrac{h}{2})[u(x,y+h) - u(x,y)] -$$
$$- C(x,y-\tfrac{h}{2})[u(x,y) - u(x,y-h)]\} .$$

This leads to a nonlinear system of the form

$$G_2(z) = Az + \Phi(z) = 0, \quad \Phi(z) = (\phi_i(z_i)) ,$$

where A is a symmetric, positive definite M-matrix and where the derivative of $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is isotone. $G_2(z)$ is for arbitrarily choosen step-size h a surjective M-function, that is $G_2(z) = 0$ has exactly one solution x^*.

Furthermore this solution is enclosed by the real n-vectors $z^1 = -A^{-1} \cdot |\Phi(0)|$ and $z^2 = -z^1$:

$$z^1 \leq x^* \leq z^2 .$$

See [4], p. 460. Since Φ is isotone we can set the lower bounds of the diagonal elements of the interval-arithmetic evaluation $\Phi'([x]^0)$ equal to zero if these bounds are negative. Since A is an M-matrix it then follows that all point-matrices which are contained in $G_2'([x]^0)$ are M-matrices, that is the assumptions of Theorem 2 hold. We stress the fact that for the system $G_2(z) = 0$ no assumptions about the width of $[x]^0$ are needed in order that the assumptions of Theorem 2 hold. The only really important assumption about $[x]^0$ is the inclusion $x^* \in [x]^0$ of the solution x^*. As shown above such an $[x]^0$ can be computed by solving a linear system of simultaneous equations.

If one applies (INSI) to systems of the form $G_1(z) = 0$ or $G_2(z) = 0$ then one observes that the convergence is extremely show. This is especially the case if the number of unknowns becomes larger and larger. The reason is that the spectral-radius $\rho\left((D(x^*)-L(x^*))^{-1} U(x^*)\right)$ approaches one if h goes to zero. By part d) of Theorem 2 this implies that the same is the case for the asymptotic convergence factor $R_1((\text{INSI}),x^*)$. For this reason (INSI) can not be considered to be a realistic method for computing the solution of $G_1(z) = 0$ or $G_2(z) = 0$.

On the other hand we consider the fact that (INSI) applied to $G_2(z)$ is convergent to x^* (without important additional assumptions) for all

starting interval-vectors which enclose x^* as an advantageous property which we would not like to give up.

4. The proposed modification of (INSI)

In order to compute sequences of vectors which are converging faster to the solution than the bounds of the interval-vectors computed using (INSI) we now use the fact that $m([x]^k) \in [x]^k$ can be choosen arbitrarily in each step of (INSI). Therefore we introduce an instruction for choosing $m([x]^k)$, which uses only data which are already known from (INSI) and for which there is a chance that the sequence $\{m([x]^k)\}$ is faster convergent to x^* than the bounds of $\{[x]^k\}$.

In order to formulate this instruction we need a so-called cut-off function κ which is defined in the following manner:
Let $w \in \mathbb{R}$ and the interval $[x] = [x_1, x_2]$ be given. Then

$$\kappa(w,[x]) = \begin{cases} x_1 & , \quad w < x_1 \\ w & , \quad w \in [x] \\ x_2 & , \quad w > x_2 \end{cases} .$$

For $[x] = ([x]_i)$, $u = (u_i)$ we define

$$p(u,[x]) = \left(\kappa(u_i,[x]_i)\right) .$$

Using p we consider the following method, called (INSI) + (SOR), which differs from (INSI) only by adding an explicit rule for the selection of $m([x]^k)$.

$$\left.\begin{array}{l} \text{Choose } m([x]^0) \text{ to be center of } [x]^0. \\ \omega_{-1} := 1. \\ \left\{\begin{array}{l} [y]^{k+1} = m([x]^k) - \tilde{D}([x]^k)\{L([x]^k)\left(m([x]^k)-[y]^{k+1}\right) + \\ \qquad\qquad + U([x]^k)\left(m([x]^k)-[x]^k\right)+f\left(m([x]^k)\right)\} \\ [x]^{k+1} = [y]^{k+1} \cap [x]^k \\ \gamma_k = \dfrac{\|d([x]^{k+1})\|_\infty}{\|d([x]^k)\|_\infty} \quad (\text{if } d([x]^k) \neq 0) \\ \omega_k = \begin{cases} \dfrac{2}{1+\sqrt{1-\gamma_k}} & \text{if } \gamma_k \neq 1 \\ \omega_{k-1} & \text{otherwise} \end{cases} \end{array}\right. \\ \vdots \end{array}\right\} \qquad (3)$$

$$\left.\begin{array}{l} \vdots \\ u^{k+1} = m([x]^k) - \omega_k \left\{ m(D([x]^k)) - \omega_k \, m(L([x]^k)) \right\}^{-1} \cdot f(m[x]^k) \\ m([x]^{k+1}) = p(u^{k+1}, [x]^{k+1}) \\ k = 0,1,2,\ldots \; . \end{array} \right\} \vdots$$

$\{m(D([x]^k))$ and $m(L([x]^k))\}$ are arbitrary real matrices which are taken from $D([x]^k)$ and $L([x]^k)$ respectively. Naturally one can choose the centers of these matrices. In [3],p. 38 ff,it is demonstrated in detail why with the given choice of $m([x]^k)$ one can be rather sure, that the sequence $\{m([x]^k)\}$ converges considerably faster to x^* than the bounds of the sequence $\{[x]^k\}$.

We finally remark that the instruction which is used for computing u^{k+1} (and therefore also for computing $m([x]^{k+1})$) may be considered to be an approximate step of the Newton-SOR-method applied to $f(x) = 0$. (Concerning the Newton-SOR-method see [4], p.217 ff).

In passing we note that instead of performing one approximate step of the Newton-SOR-method one can do the same using any other iteration method which promises to converge faster to x^* than the bounds of the sequence $\{[x]^k\}$. In [3] the use of the ADI-method was discussed in some detail. The numerical results are even much better than with the results listet subsequently for (INSI) +(SOR). However, no theoretical foundation can be given in this case because of the fact that certain matrices do not commute.

5. Numerical examples

Example 1.

As a first example we consider the equation

$$\Delta u = \frac{u^3}{1+x^2+y^2} \quad \text{in} \quad (0,1) \times (0,1)$$

with the boundary conditions

$$u(x,0) = 1 \quad \text{and} \quad u(x,1) = 2-e^x \quad \text{for} \quad x \in [0,1]$$

$$u(0,y) = 1 \quad \text{and} \quad u(1,y) = 2-e^y \quad \text{for} \quad y \in [0,1] \; .$$

Example 2.

$$\Delta u = e^u \quad \text{in} \quad (0,1) \times (0,1) = R$$

$$u(x,y) = x + 2y \quad \text{on} \quad \partial R$$

(Please note that the results of this paper are not limited to rectangular regions. Numerical examples for which the boundary is curvilinear are given in [3]).

In the tables given subsequently we have compared our results with those from the paper "Aspects of Nonlinear Block-Successive Overrelaxation" by L.A.Hagemann and T.A.Porsching (SIAM J. Numer.Anal., 12, 316-335 (1975)). In that paper a very lengthy instruction is given which forces the normally only local convergent nonlinear block-successive overrelaxation method to converge to the solution x^*. Therefore this modification is comparable to our method (INSI) + (SOR) where by (INSI) convergence is guaranteed for all interval-vectors which enclose the solution.

In both examples the following termination criteria were used:

$$\|d([x]^k)\|_\infty \leq 2 \cdot 10^{-6} \quad \text{for} \quad \text{(INSI)}$$

$$\|u^{k+1} - m([x]^k)\|_\infty \leq 10^{-6} \quad \text{for} \quad \text{(INSI) + (SOR)}$$

$$\|x^{k+1} - x^k\|_\infty \leq 10^{-6} \quad \text{for} \quad \text{(H-P)}.$$

In order to make a fair judgement on the proposed method (INSI) + (SOR) one has to take into account that the interval-operations necessary for performing this method have been programmed using subroutines. If there would be available a realization of the interval-operations which - concerning the execution time - is comparable to the usual floating-point operations, then the proposed method would compare even more favorable.

The examples have been computed using a CYBER 175 of the Wissenschaftliches Rechenzentrum Berlin (WRB).

Example 1

	h	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{91}$	$\frac{1}{128}$
	$n=(\frac{1}{h}-1)^2$	9	48	225	361	961	3969		16129
(INSI)	Steps	21	90	366	572	1466	5856*	---	23424*
	Time (sec)	0.067	1.624	30.448	76.368	521.922	2.39*hours	---	38.85*hours
(INSI)+ (SOR)	Steps	11	22	47	61	105	248	400	---
	Time (sec)	0.041	0.473	4.605	9.548	45.3	431.8	1396	1.17*hours
(H-P)**	Steps	22	39	69	88	123	237	339	
	Time (sec)	0.04	0.403	2.950	6.081	24.5	183.9	518	

Both for (INSI) and for (INSI) + (SOR) all components of the starting vector $[x]^0$ have been chosen to be the interval [-1,2]. For (H-P) we chose $x^0 = (x_i^0)$, $x_i^0 = 2$ for all i.

*Estimated values.

**Means the point overrelaxation method using the strategy given by Hagemann and Porsching in the paper mentioned in the text.

Example 2

	h	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{91}$	$\frac{1}{128}$
	$n=(\frac{1}{h}-1)^2$	9	49	225	361	961	3969		16129
(INSI)	Steps	19	81	324	507	1298	-	-	-
	Time(sec)	0.043	0.943	17.54	44.2	301	-	-	-
(INSI)+ (SOR)	Steps	10	21	46	59	102	248	393	-
	Time(sec)	0.029	0.311	3.164	6.554	30.4	298.128	987	49*min
(H-P)**	Steps	21	39	77	74	122	251	342	
	Time(sec)	0.051	0.507	4.718	7.911	33.09	258.1	692.5	-

Both for (INSI) and for (INSI) + (SOR) all components of the starting vector $[x]^0$ have been chosen to be the interval [0,3]. For (H-P) we chose $x^0 = (x_i^0)$, $x_i^0 = 3$.

*Estimated value.

**Means the point overrelaxation method using the strategy given by Hagemann and Porsching in the paper mentioned in the text.

References

[1] G.Alefeld : Über die Existenz einer eindeutigen Lösung bei einer Klasse nichtlinearer Gleichungssysteme und deren Berechnung mit Iterationsverfahren.
Aplikace Matematiky 17, 267-294 (1972).

[2] G.Alefeld, J.Herzberger : Einführung in die Intervallrechnung.
Bibliographisches Institut, Reihe Informatik 12, Mannheim 1974.

[3] H.Cornelius : Untersuchungen zu einem intervallarithmetischen Iterationsverfahren mit Anwendungen auf eine Klasse nichtlinearer Gleichungssysteme.
Dissertation. Fachbereich Mathematik der TU Berlin, Berlin 1981.

[4] J.M.Ortega, W.C.Rheinboldt : Iterative Solution of Nonlinear Equations in Several Variables.
Academic Press, New York - London 1970.

[5] W.C.Rheinboldt : On M-Functions and their Application to Nonlinear Gauss-Seidel Iterations and Network Flows. Ges.f.Math.u.Datenverarbeitung, Birlinghoven/Germany. Tech.Rep. BMwF-GMD-23. (1969).

[6] W.C.Rheinboldt : On classes of n-dimensional nonlinear mappings generalizing several types of matrices. Numerical Solution of Partial Differential Equations - II. Synspade 1970. B.Hubbard (Ed.).Academic Press, New York - London 1971.

[7] W.Törnig : Monoton einschließend konvergente Iterationsprozesse vom Gauss-Seidel Typ zur Lösung nichtlinearer Gleichungssysteme im $\mathbb{R}^N$ und Anwendungen. Technische Hochschule Darmstadt.
Preprint-Nr. 517, Dezember 1979.

[8] R.S.Varga : Matrix Iterative Analysis. Prentice-Hall Inc. Englewood Cliffs, N.J. 1962.