SIAM J. NUMER. ANAL.
Vol. 20, No. 1, February 1983

0036-1429/83/2001-0014 $01.25/0

A QUADRATICALLY CONVERGENT KRAWCZYK-LIKE ALGORITHM*

G. ALEFELD† AND L. PLATZÖDER‡

Abstract. In this paper we introduce a method for computing a solution of a nonlinear system, which is similar to that proposed by R. Krawczyk [Computing, 4 (1969), pp. 187–201]. Our method, however, needs considerably less work per step. Starting with an interval vector, we give a criterion under which the method is convergent to the solution of the system if a solution is contained in the interval vector. If the starting vector contains no solution then the method will break down after a finite number of steps.

1. Introduction. In 1969 R. Krawczyk [2] introduced the operator

$$\mathbf{k}(\mathbf{x}, \underset{\cdot}{Y}) = m(\mathbf{x}) - \underset{\cdot}{Y}\underset{\cdot}{\mathfrak{f}}(m(\mathbf{x})) + (\underset{\cdot}{E} - \underset{\cdot}{Y}\underset{\cdot}{\mathfrak{f}}'(\mathbf{x}))(\mathbf{x} - m(\mathbf{x})), \tag{0}$$

where $\underset{\cdot}{\mathfrak{f}}\colon \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ denotes a mapping $\underset{\cdot}{\mathfrak{f}}$, $\mathbf{x} \subseteq \mathfrak{B}$ is an interval vector, $m(\mathbf{x})$ is the center of $\mathbf{x}$, $\underset{\cdot}{Y}$ is a real $n \times n$ matrix, $\underset{\cdot}{E}$ denotes the unit matrix and $\underset{\cdot}{\mathfrak{f}}'(\mathbf{x})$ is the interval arithmetic evaluation of the derivative of $\underset{\cdot}{\mathfrak{f}}$ over the interval vector $\mathbf{x}$. The result $\mathbf{k}(\mathbf{x}, \underset{\cdot}{Y})$ is an interval vector. Using this operator, Krawczyk considered the iteration method

$$\mathbf{x}^{k+1} = \mathbf{k}(\mathbf{x}^k, \underset{\cdot}{Y}) \cap \mathbf{x}^k, \qquad k \geqq 0. \tag{1}$$

If the interval vector $\mathbf{x}^0$ contains a zero $\underset{\cdot}{x}^*$ of $\underset{\cdot}{\mathfrak{f}}$, then (1) is well defined. Under some additional assumptions it holds that $\lim_{k\to\infty} \mathbf{x}^k = \underset{\cdot}{x}^*$, that is, the bounds of the sequence $\{\mathbf{x}^k\}$ converge to $\underset{\cdot}{x}^*$.

In 1977 the method (1) was reconsidered and discussed by R. E. Moore in a series of papers starting with [3].

In order to get faster than linear convergence in (1), it is necessary to modify the matrix $\underset{\cdot}{Y}$ with k. More precisely, the relation $\lim_{k\to\infty} \underset{\cdot}{Y}_k = \underset{\cdot}{\mathfrak{f}}(\underset{\cdot}{x}^*)^{-1}$ should hold. Here $\underset{\cdot}{\mathfrak{f}}'(\underset{\cdot}{x}^*)$ denotes the derivative at the point $\underset{\cdot}{x}^*$. In this case, (1) requires computing the inverse of a real matrix in each step. Taking into account $2n^3$ multiplications for forming the product $\underset{\cdot}{Y}\underset{\cdot}{\mathfrak{f}}'(\mathbf{x})$, then (1) needs $3n^3$ multiplications per step for large n.

In this paper we introduce a quadratically convergent algorithm which needs fewer multiplications for large n. The proposed algorithm needs the amount necessary for performing the triangular decomposition of a nonsingular matrix, namely $n^3/3$ multiplications. Instead of the Krawczyk operator defined by (0), we consider an operator $\mathbf{kn}(\mathbf{x}, \underset{\cdot}{A})$ defined by

$$\mathbf{kn}(\mathbf{x}, \underset{\cdot}{A}) = m(\mathbf{x}) - \mathrm{IGA}\,(\underset{\cdot}{A}, \underset{\cdot}{\mathfrak{f}}(m(\mathbf{x})) - \{\underset{\cdot}{A} - \underset{\cdot}{\mathfrak{f}}'(\mathbf{x})\} \cdot \{\mathbf{x} - m(\mathbf{x})\}). \tag{2}$$

For an arbitrary interval matrix $\mathbf{X}$ and an interval vector $\mathbf{y}$ we denote by IGA $(\mathbf{X}, \mathbf{y})$ the interval vector which results if the Gaussian algorithm is applied.

Note that in (2) the coefficient matrix is a point matrix, which implies that $\mathbf{kn}$ $(\mathbf{x}, \underset{\cdot}{A})$ can always be computed (eventually after performing row or column exchanges) if $\underset{\cdot}{A}$ is nonsingular.

Before introducing our iteration method we discuss some preliminaries.

2. Notation and preliminaries. It is assumed that the reader has a certain knowledge of the elementary concepts of interval analysis to the extent that one can find,

* Received by the editors January 20, 1982, and in final form April 16, 1982.

† Institut für Angewandte Mathematik, Universität Karlsruhe, Kaiserstrasse 12, 7500 Karlsruhe, West Germany.

‡ Fachbereich 3-Mathematik, Technische Universität Berlin, 1 Berlin 12, West Germany.

for example, in [1]. Proofs of the following rules may be found in [1] or they are elementary. As already mentioned, $m(\mathbf{x})$ always denotes the center of the vector $\mathbf{x}$. If $\mathbf{A}$ is an interval matrix then $m(\mathbf{A})$ is defined in an analogous manner.

Let $\mathbf{A}$ be a point matrix, let $\mathbf{B}$ be a symmetric interval matrix ($\mathbf{B}=-\mathbf{B}$) and let $\mathbf{c}$ be a symmetric interval vector ($\mathbf{c}=-\mathbf{c}$). Then

$$(\underset{\cdot}{\mathbf{A}}\mathbf{B})\mathbf{c}=\underset{\cdot}{\mathbf{A}}(\mathbf{B}\mathbf{c}). \tag{3}$$

Let $\underset{\cdot}{\mathbf{A}}$, $\underset{\cdot}{\mathbf{B}}$ be point matrices and let $\mathbf{c}$ be an arbitrary interval vector. Then

$$(\underset{\cdot}{\mathbf{A}}\underset{\cdot}{\mathbf{B}})\mathbf{c}\subseteq\underset{\cdot}{\mathbf{A}}(\underset{\cdot}{\mathbf{B}}\mathbf{c}). \tag{4}$$

Let $\underset{\cdot}{\mathbf{A}}$ be a point matrix and let $\mathbf{B}$ and $\mathbf{C}$ be interval matrices. Then

$$\mathbf{A}(\mathbf{B}+\mathbf{C})=\underset{\cdot}{\mathbf{A}}\mathbf{B}+\underset{\cdot}{\mathbf{A}}\mathbf{C}. \tag{5}$$

Let $\mathbf{B}$ be a symmetric interval matrix ($\mathbf{B}=-\mathbf{B}$). Then for an arbitrary interval matrix $\mathbf{A}$ it holds that

$$\mathbf{A}\mathbf{B}=|\mathbf{A}|\mathbf{B} \tag{6}$$

where $|\mathbf{A}|$ denotes the absolute value of the interval matrix $\mathbf{A}$. $|\mathbf{A}|$ is defined in the following manner. For a real compact interval $A=[a_1, a_2]$ one first defines

$$|A|=\max\{|a_1|, |a_2|\}.$$

If $\mathbf{A}=(A_{ij})$ is the given interval matrix, then

$$|\mathbf{A}|=(|A_{ij}|).$$

If $A=[a_1, a_2]$ is a given interval, then one defines the width of A as

$$d(A)=a_2-a_1,$$

and using this, the width of a given interval matrix $\mathbf{A}=(A_{ij})$ is defined as

$$d(\mathbf{A})=(d(A_{ij})).$$

Subsequently we will use the following rules:

$$\mathbf{A}\subseteq\mathbf{B}\Rightarrow|\mathbf{A}|\leqq|\mathbf{B}|, \tag{7}$$

$$d(\underset{\cdot}{\mathbf{A}}\pm\mathbf{B})=d(\mathbf{B}), \tag{8}$$

$$d(\underset{\cdot}{\mathbf{A}}\mathbf{B})=|\underset{\cdot}{\mathbf{A}}|d(\mathbf{B}). \tag{9}$$

We now discuss the following problem: Let there be given a system of linear simultaneous equations. Assume that the data of the given system are known to lie in certain intervals, that is, let there be given an interval matrix $\mathbf{A}$ (the "coefficient matrix") and an interval vector $\mathbf{b}$ (the "right-hand side" of the system). The set of solutions of all linear systems described by these data can be enclosed by applying the Gaussian algorithm in a formal manner (see [1, § 15], for example). The result is an interval vector IGA $(\mathbf{A}, \mathbf{b})$ which can be represented in the following form:

$$\mathrm{IGA}\,(\mathbf{A}, \mathbf{b}):=\mathbf{D}_1(\mathbf{T}_1(\cdots\mathbf{T}_{n-1}(\mathbf{D}_n(\mathbf{G}_{n-1}(\cdots(\mathbf{G}_2(\mathbf{G}_1\mathbf{b})\cdots). \tag{10}$$

The matrices $\mathbf{D}_1, \cdots, \mathbf{D}_n, \mathbf{T}_1, \cdots, \mathbf{T}_{n-1}, \mathbf{G}_1, \cdots, \mathbf{G}_{n-1}$, are certain interval matrices which are dependent on $\mathbf{A}$ and which in the case where $\mathbf{A}$ is a point matrix are also point matrices. A precise description of these matrices is not of importance in this paper. The interested reader may find it in [4]. (H. Schwandt [4] was the first to state (10).)

Using the matrices occurring on the right-hand side of (10) we define the interval matrix

$$\mathrm{IGA}(\mathbf{A}) := \mathbf{D}_1(\mathbf{T}_1(\cdots \mathbf{T}_{n-1}(\mathbf{D}_n(\mathbf{G}_{n-1}\cdots(\mathbf{G}_2\mathbf{G}_1)\cdots). \tag{11}$$

If $\underset{\cdot}{A}$ is a nonsingular point matrix, then it is easy to see that

$$\mathrm{IGA}(\underset{\cdot}{A}) = \underset{\cdot}{D}_1(\underset{\cdot}{T}_1(\cdots \underset{\cdot}{T}_{n-1}(\underset{\cdot}{D}_n(\underset{\cdot}{G}_{n-1}\cdots(\underset{\cdot}{G}_2\underset{\cdot}{G}_1)\cdots) = \underset{\cdot}{A}^{-1}. \tag{12}$$

(Choose $\mathbf{b} := \underset{\cdot}{e}_i$, $1 \leqq i \leqq n$, where $\underset{\cdot}{e}_i$ denotes the ith unit vector).

After these preparations we prove the following results.

LEMMA 1. *Let $\underset{\cdot}{A}$ be a nonsingular real point matrix.*

a) *For an arbitrary interval vector $\mathbf{b}$ it holds that*

$$\underset{\cdot}{A}^{-1}\mathbf{b} \subseteq \mathrm{IGA}(\underset{\cdot}{A}, \mathbf{b}).$$

b) *If $d(\mathbf{a}) \leqq \beta d(\mathbf{b})$, $\beta \geqq 0$, then*

$$d(\mathrm{IGA}(\underset{\cdot}{A}, \mathbf{a})) \leqq \beta d(\mathrm{IGA}(\underset{\cdot}{A}, \mathbf{b})).$$

Proof. a) Using (12) and applying (4) repeatedly, it follows that

$$\begin{aligned}\underset{\cdot}{A}^{-1}\mathbf{b} = \mathrm{IGA}(\underset{\cdot}{A})\cdot\mathbf{b} &= \underset{\cdot}{D}_1(\underset{\cdot}{T}_1(\cdots \underset{\cdot}{D}_n(\underset{\cdot}{G}_{n-1}\cdots(\underset{\cdot}{G}_2\underset{\cdot}{G}_1)\cdots)\cdot\mathbf{b}\\ &\subseteq \underset{\cdot}{D}_1\{\underset{\cdot}{T}_1(\cdots \underset{\cdot}{D}_n(\underset{\cdot}{G}_{n-1}\cdots(\underset{\cdot}{G}_2\underset{\cdot}{G}_1)\cdots)\cdot\mathbf{b}\}\\ &\cdots\\ &\subseteq \underset{\cdot}{D}_1(\underset{\cdot}{T}_1(\cdots \underset{\cdot}{D}_n(\underset{\cdot}{G}_{n-1}\cdots(\underset{\cdot}{G}_2(\underset{\cdot}{G}_1\mathbf{b})\cdots)\\ &= \mathrm{IGA}(\underset{\cdot}{A}, \mathbf{b}).\end{aligned}$$

b) Using (10) and applying (9) repeatedly it follows that

$$\begin{aligned}d(\mathrm{IGA}(\underset{\cdot}{A}, \mathbf{a}) &= |\underset{\cdot}{D}_1|\,|\underset{\cdot}{T}_1|\cdots|\underset{\cdot}{G}_2|\,|\underset{\cdot}{G}_1|d(\mathbf{a})\\ &\leqq \beta|\underset{\cdot}{D}_1|\,|\underset{\cdot}{T}_1|\cdots|\underset{\cdot}{G}_2|\,|\underset{\cdot}{G}_1|d(\mathbf{b})\\ &= \beta d(\mathrm{IGA}(\underset{\cdot}{A}, \mathbf{b})).\end{aligned}$$

□

LEMMA 2. *Let there be given an interval vector $\mathbf{x}^0$ with $d(\mathbf{x}^0) > \underset{\cdot}{o}$. Assume that $\mathbf{f}: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ is a mapping for which the interval arithmetic evaluation $\mathbf{f}'(\mathbf{x}^0)$ of the derivative exists. Assume that the point matrix $\underset{\cdot}{B}_0 := m(\mathbf{f}'(\mathbf{x}^0))$ is nonsingular and that*

$$d(\mathbf{kn}(\mathbf{x}^0, \underset{\cdot}{B}_0)) \leqq \alpha d(\mathbf{x}^0), \tag{13}$$

where $0 \leqq \alpha < 1$. Then $\mathbf{f}'(\mathbf{x}^0)$ does not contain any singular point matrix.

Proof. Applying (8) and (9) it follows from (0) that

$$d(\mathbf{k}(\mathbf{x}^0, \underset{\cdot}{B}_0^{-1})) = |\underset{\cdot}{E} - \underset{\cdot}{B}_0^{-1}\mathbf{f}'(\mathbf{x}^0)|d(\mathbf{x}^0). \tag{14}$$

We show that for the Krawczyk operator $\mathbf{k}(\mathbf{x}^0, \underset{\cdot}{B}_0^{-1})$ defined by (0) and the operator $\mathbf{kn}(\mathbf{x}^0, \underset{\cdot}{B}_0)$ defined by (2), the inequality

$$d(\mathbf{k}(\mathbf{x}^0, \underset{\cdot}{B}_0^{-1})) \leqq d(\mathbf{kn}(\mathbf{x}^0, \underset{\cdot}{B}_0)) \tag{15}$$

holds. If (15) has been shown, then using (13) and (14) it follows that

$$|\underset{\cdot}{E} - \underset{\cdot}{B}_0^{-1}\mathbf{f}'(\mathbf{x}^0)|d(\mathbf{x}^0) \leqq \alpha d(\mathbf{x}^0).$$

Applying [6, Cor. 3 of Thm. 1.5], it follows that the spectral radius ρ of the real matrix $|\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{f}'(\mathbf{x}^0)|$ is less than one. If $\mathbf{X}\in\mathbf{f}'(\mathbf{x}^0)$, then (7) implies

$$|\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{X}|\leqq|\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{f}'(\mathbf{x}^0)|.$$

From the Perron–Frobenius theory for nonnegative matrices it follows that $\rho(\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{X})<1$. Hence $(\mathbf{E}-(\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{X}))^{-1}=\mathbf{X}^{-1}\mathbf{B}_0$ exists.

We are now going to prove (15). We apply Lemma 1a) with

$$\mathbf{A}:=\mathbf{B}_0,$$
$$\mathbf{b}:=\mathbf{f}(m(\mathbf{x}^0))-\{\mathbf{B}_0-\mathbf{f}'(\mathbf{x}^0)\}\cdot\{\mathbf{x}^0-m(\mathbf{x}^0)\}$$

and get the inclusion relation

$$\mathbf{B}_0^{-1}\{\mathbf{f}(\mathbf{m}(\mathbf{x}^0))-\{\mathbf{B}_0-\mathbf{f}'(\mathbf{x}^0)\}\cdot\{\mathbf{x}^0-m(\mathbf{x}^0)\}\}$$
$$\subseteq \mathrm{IGA}\,(\mathbf{B}_0,\mathbf{f}(m(\mathbf{x}^0))-\{\mathbf{B}_0-\mathbf{f}'(\mathbf{x}^0)\}\cdot\{\mathbf{x}^0-m(\mathbf{x}^0)\}).$$

Using (2), the right-hand side of this relation may be written as $m(\mathbf{x}^0)-\mathbf{kn}(\mathbf{x}^0,\mathbf{B}_0)$. We now show that the left-hand side is equal to $m(\mathbf{x}^0)-\mathbf{k}(\mathbf{x}^0,\mathbf{B}_0^{-1})$. Then

$$\mathbf{k}(\mathbf{x}^0,\mathbf{B}_0^{-1})\subseteq\mathbf{kn}(\mathbf{x}^0,\mathbf{B}_0) \tag{16}$$

holds, from which (15) follows.

Using (5) and (3) the left-hand side of the above inclusion relation may in fact be written as

$$\mathbf{B}_0^{-1}\{\mathbf{f}(m(\mathbf{x}^0))-\{\mathbf{B}_0-\mathbf{f}'(\mathbf{x}^0)\}\cdot\{\mathbf{x}^0-m(\mathbf{x}^0)\}\}$$
$$=\mathbf{B}_0^{-1}\mathbf{f}(m(\mathbf{x}^0))-\{\mathbf{B}_0^{-1}(\mathbf{B}_0-\mathbf{f}'(\mathbf{x}^0))\}(\mathbf{x}^0-m(\mathbf{x}^0))$$
$$=\mathbf{B}_0^{-1}\mathbf{f}(m(\mathbf{x}^0))-\{\mathbf{E}-\mathbf{B}_0^{-1}\mathbf{f}'(\mathbf{x}^0)\}\cdot\{\mathbf{x}^0-m(\mathbf{x}^0)\}$$
$$=m(\mathbf{x}^0)-\mathbf{k}(\mathbf{x}^0,\mathbf{B}_0^{-1}).$$ □

As shown in the proof of Lemma 2, formula (15) is a consequence of (16). (16) means that the result of the Krawczyk operator is always a subset of the corresponding result of the operator (2). However, as we already have pointed out, the new operator is much cheaper to evaluate.

E. Kaucher and S. M. Rump [9] have considered an iterative method using (2), which is similar to (1). Their experience is—probably as a consequence of (16)—that the accuracy obtained for $k\to\infty$ is better with (1).

3. The algorithm. Let there be given an interval vector $\mathbf{x}^0\subseteq\mathfrak{B}$ and a mapping $\mathbf{f}:\mathfrak{B}\subseteq V_n(\mathbb{R})\to V_n(\mathbb{R})$ for which the interval arithmetic evaluation of the derivative exists. We then consider Algorithm A below.

For large n this algorithm needs $n^3/3$ multiplications per step (besides computing $\mathbf{f}(m(\mathbf{x}^k))$ and $\mathbf{f}'(\mathbf{x}^k)$). This is the number of operations necessary for performing the triangular decomposition of $\mathbf{A}_k$. Of course, we assume that the statement $\mathbf{B}_k:=\mathbf{A}_k$ is performed by storing the triangular decomposition of $\mathbf{A}_k$. The same holds for the statement $\mathbf{B}_k:=\mathbf{B}_{k-1}$. The use of n^2 additional storage places might be considered as a disadvantage of this algorithm.

ALGORITHM A

$\mathfrak{B}_0 := m(\mathfrak{f}'(\mathbf{x}^0))$
$\mathbf{x}^1 := \mathbf{kn}(\mathbf{x}^0, \mathfrak{B}_0) \cap \mathbf{x}^0$
For $k \geq 1$ **do**
begin
  $\mathfrak{A}_k := m(\mathfrak{f}'(\mathbf{x}^k))$
  Compute $\mathbf{kn}(\mathbf{x}^k, \mathfrak{A}_k)$
  If $d(\mathbf{kn}(\mathbf{x}^k, \mathfrak{A}_k)) \leq \alpha d(\mathbf{x}^k)$ **then**
  begin
    $\mathbf{x}^{k+1} := \mathbf{kn}(\mathbf{x}^k, \mathfrak{A}_k) \cap \mathbf{x}^k$ (*)
    $\mathfrak{B}_k := \mathfrak{A}_k$
  end
  else
  begin
    $\mathfrak{B}_k := \mathfrak{B}_{k-1}$
    Compute $\mathbf{kn}(\mathbf{x}^k, \mathfrak{B}_k)$
    $\mathbf{x}^{k+1} := \mathbf{kn}(\mathbf{x}^k, \mathfrak{A}_k) \cap \mathbf{kn}(\mathbf{x}^k, \mathfrak{B}_k) \cap \mathbf{x}^k$ (**)
  end
end

We now prove the following.

THEOREM. *Let there be given a mapping $\mathfrak{f}: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ and an interval vector $\mathbf{x}^0 \subseteq \mathfrak{B}$ with $d(\mathbf{x}^0) > \mathfrak{o}$ for which $\mathfrak{f}'(\mathbf{x}^0)$ exists. Assume that $\mathfrak{B}_0 := m(\mathfrak{f}'(\mathbf{x}^0))$ is nonsingular and that*

$$d(\mathbf{kn}(\mathbf{x}^0, \mathfrak{B}_0)) \leq \alpha d(\mathbf{x}^0) \tag{17}$$

where $0 \leq \alpha < 1$.

a) *If $\mathbf{x}^0$ contains a zero $\mathfrak{x}^*$ of $\mathfrak{f}$, then Algorithm* A *is well defined and* $\lim_{k\to\infty} \mathbf{x}^k = \mathfrak{x}^*$ holds. *If for some norm*

$$d((\mathfrak{f}'(\mathbf{x}))_{ij}) \leq c\|d(\mathbf{x})\| \tag{18}$$

holds, then Algorithm A *is at least quadratically convergent.*

b) *If $\mathbf{x}^0$ contains no zero of $\mathfrak{f}$, then after a finite number of steps Algorithm* A *breaks down because of empty intersection.*

Proof. a) Because of Lemma 2, none of the matrices $\mathfrak{A}_k$, $k \geq 1$, is singular if (17) holds. Therefore A is well defined as long as no empty intersection occurs. For the following reasons the intersection is not empty: If $\mathbf{x}^k$ contains a zero $\mathfrak{x}^*$ which is assumed for $k = 0$, then for an arbitrary nonsingular point matrix $\mathfrak{X}$ the relation

$$\mathfrak{x}^* \in \mathbf{k}(\mathbf{x}^k, \mathfrak{X}^{-1})$$

holds. This was already proven in [2]. In a similar manner the relation

$$\mathfrak{x}^* \in \mathbf{kn}(\mathbf{x}^k, \mathfrak{X})$$

can be shown. We omit the details. Hence it is clear that in performing Algorithm A, neither at (*) nor at (**) can the intersection be empty. Furthermore $\mathfrak{x}^* \in \mathbf{x}^{k+1}$ holds.

We now prove that

$$d(\mathbf{x}^k) \leqq \alpha^k d(\mathbf{x}^0), \qquad k \geqq 1,$$

holds. Because of $0 \leqq \alpha < 1$ and $\underset{\cdot}{x}^* \in \mathbf{x}^k$, $k \geqq 0$, it then clearly follows that $\lim_{k \to \infty} \mathbf{x}^k = \underset{\cdot}{x}^*$. From the assumption (17) the assertion holds true for $k = 1$. Assume now that

(19) $$d(\mathbf{x}^k) \leqq \alpha^k d(\mathbf{x}^0)$$

for $1 \leqq k \leqq k_0$, which we have just shown to be true for $k_0 = 1$ by assumption. If now $\mathbf{x}^{k_0+1}$ is computed by using the statement (*), that is, if

$$\mathbf{x}^{k_0+1} = \mathbf{kn}(\mathbf{x}^{k_0}, \underset{\cdot}{A}_{k_0}) \cap \mathbf{x}^{k_0}$$

holds, then it follows that

(20) $$d(\mathbf{kn}(\mathbf{x}^{k_0}, \underset{\cdot}{A}_{k_0})) \leqq \alpha d(\mathbf{x}^{k_0}),$$

and because of the induction hypotheses (19) we finally have

$$d(\mathbf{x}^{k_0+1}) \leqq \alpha^{k_0+1} d(\mathbf{x}^0).$$

If on the other hand $\mathbf{x}^{k_0+1}$ is computed by the statement (**), then

(21) $$d(\mathbf{x}^{k_0+1}) \leqq d(\mathbf{kn}(\mathbf{x}^{k_0}, \underset{\cdot}{B}_{k_0})).$$

The way $\underset{\cdot}{B}_{k_0}$ is formed shows that there is a $\hat{k}$ where $0 < \hat{k} \leqq k_0 - 1$ such that

$$\underset{\cdot}{B}_{k_0} = \underset{\cdot}{A}_{\hat{k}} = m(\underset{\cdot}{f}'(\mathbf{x}^{\hat{k}})).$$

Because of (*) we have for this $\hat{k}$

(22) $$d(\mathbf{kn}(\mathbf{x}^{\hat{k}}, \underset{\cdot}{A}_{\hat{k}})) \leqq \alpha d(\mathbf{x}^{\hat{k}})$$

and therefore

(23) $$d(\mathbf{x}^{\hat{k}+1}) \leqq \alpha d(\mathbf{x}^{\hat{k}}).$$

Using mathematical induction we now prove the following:

(24) $$d(\mathbf{x}^{\hat{k}+j+1}) \leqq \alpha^{j+1} d(\mathbf{x}^{\hat{k}}), \qquad 0 \leqq j \leqq k_0 - \hat{k}.$$

For $j = 0$, (24) is identical to (23). If (24) holds for some $j \geqq 0$ ($j < k_0 - \hat{k}$) then applying (7) and (24) it follows that

$$\begin{aligned} d(\underset{\cdot}{f}(m(\mathbf{x}^{\hat{k}+j+1})) &+ \{\underset{\cdot}{B}_{k_0} - \underset{\cdot}{f}'(\mathbf{x}^{\hat{k}+j+1})\} \cdot \{\mathbf{x}^{\hat{k}+j+1} - m(\mathbf{x}^{\hat{k}+j+1})\}) \\ &= |\underset{\cdot}{B}_{k_0} - \underset{\cdot}{f}'(\mathbf{x}^{\hat{k}+j+1})| d(\mathbf{x}^{\hat{k}+j+1}) \\ &\leqq \alpha^{j+1} |\underset{\cdot}{B}_{k_0} - \underset{\cdot}{f}'(\mathbf{x}^{\hat{k}})| d(\mathbf{x}^{\hat{k}}) \\ &= \alpha^{j+1} d(\underset{\cdot}{f}(m(\mathbf{x}^{\hat{k}})) + \{\underset{\cdot}{B}_{k_0} - \underset{\cdot}{f}'(\mathbf{x}^{\hat{k}})\} \cdot \{\mathbf{x}^{\hat{k}} - m(\mathbf{x}^{\hat{k}})\}). \end{aligned}$$

Taking into account (22) and applying Lemma1b), it therefore follows that

$$d(\mathbf{kn}(\mathbf{x}^{\hat{k}+j+1}, \underset{\cdot}{B}_{k_0})) \leqq \alpha^{j+1} d(\mathbf{kn}(\mathbf{x}^{\hat{k}}, \underset{\cdot}{B}_{k_0})) \leqq \alpha^{j+2} d(\mathbf{x}^{\hat{k}}).$$

Therefore it also holds that

$$d(x^{\hat{k}+j+2}) \leqq d(\mathbf{kn}(\mathbf{x}^{\hat{k}+j+1}, \underset{\cdot}{B}_{k_0})) \leqq \alpha^{j+2} d(\mathbf{x}^{\hat{k}}),$$

which completes the proof of (24). For $j = k_0 = \hat{k}$, (24) reads

$$d(\mathbf{x}^{k_0+1}) \leqq \alpha^{k_0 - \hat{k}+1} d(\mathbf{x}^{\hat{k}}).$$

Using (19) with $k := \hat{k}$ we finally have the assertion

$$d(\mathbf{x}^{k_0+1}) \leqq \alpha^{k_0+1} d(\mathbf{x}^0).$$

We are now going to prove the statement concerning the order of convergence No matter whether $\mathbf{x}^{k+1}$ is computed using (*) or (**), we have—because of forming intersections—

$$d(\mathbf{x}^{k+1}) \leqq d(\mathrm{kn}(\mathbf{x}^k, \mathbf{A}_k)) \tag{25}$$

where $\mathbf{A}_k = m(\mathbf{f}'(\mathbf{x}^k))$. Using (2), the representation (10) and the formulas (6), (8) and (9) we get

$$\begin{aligned} d(\mathrm{kn}(\mathbf{x}^k, \mathbf{A}_k)) &\leqq |\mathbf{D}_1|\,|\mathbf{T}_1| \cdots |\mathbf{G}_{n-1}| \cdots |\mathbf{G}_2|\,|\mathbf{G}_1|\, d(\{\mathbf{A}_k - \mathbf{f}'(\mathbf{x}^k)\} \cdot \{\mathbf{x}^k - m(\mathbf{x}^k)\}) \\ &\leqq \mathbf{H} \cdot \tfrac{1}{2} \cdot d(\mathbf{f}'(\mathbf{x}^k)) \cdot d(\mathbf{x}^k) \end{aligned} \tag{26}$$

where

$$|\mathbf{D}_1|\,|\mathbf{T}_1| \cdots |\mathbf{G}_{n-1}| \cdots |\mathbf{G}_2|\,|\mathbf{G}_1| \leqq \mathbf{H}. \tag{27}$$

with a point matrix $\mathbf{H}$ which is independent of k. (Note that the matrices occurring on the left-hand side of (27) are dependent on k. Since $\mathbf{A}_k \in \mathbf{f}'(\mathbf{x}^k) \subseteq \mathbf{f}'(\mathbf{x}^0)$, since $\mathbf{f}'(\mathbf{x}^0)$ may be considered to be a compact set of matrices and since finally $\mathbf{f}'(\mathbf{x}^0)$ does not contain any singular point matrix, the product on the left-hand side of (27) may be bounded independently of k). Using (18) and some monotone vector norm and finally using the norm equivalence theorem, we get from (25) and (26) the assertion

$$\|d(\mathbf{x}^{k+1})\| \leqq \gamma \|d(\mathbf{x}^k)\|^2.$$

b) If the intersection at (*) or (**) is never empty, then in the same manner as before one shows

$$d(\mathbf{x}^{k+1}) \leqq \alpha^{k+1} d(\mathbf{x}^0).$$

Because of

$$\mathbf{x}^0 \supseteq \mathbf{x}^1 \supseteq \mathbf{x}^2 \supseteq \cdots,$$

it follows $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$ where $\mathbf{x}^*$ is some vector. From this fact it follows that also $\lim_{k\to\infty} m(\mathbf{f}'(\mathbf{x}^k)) = \mathbf{f}'(\mathbf{x}^*)$. Because of (*) and (**), we have

$$\mathbf{x}^{k+1} \subseteq \mathrm{kn}(\mathbf{x}^k, \mathbf{A}_k).$$

Since $\mathrm{kn}(\mathbf{x}, \mathbf{A})$ depends continuously on $\mathbf{x}$ and $\mathbf{A}$, we have

$$\mathbf{x}^* = \mathrm{kn}(\mathbf{x}^*, \mathbf{f}'(\mathbf{x}^*)) = \mathbf{x}^* - \mathrm{IGA}\,(\mathbf{f}'(\mathbf{x}^*), \mathbf{f}(\mathbf{x}^*)) = \mathbf{x}^* - \mathbf{f}'(\mathbf{x}^*)^{-1}\mathbf{f}(\mathbf{x}^*)$$

for $k \to \infty$ and hence the contradiction $\mathbf{f}(\mathbf{x}^*) = \mathbf{o}$ follows. □

Let us close this section with some remarks on the condition (17) which is the central assumption in our theorem. This condition holds, for example, if $\mathrm{kn}(\mathbf{x}^0, \mathbf{B}_0) \subset \mathbf{x}^0$ and if $\mathrm{kn}(\mathbf{x}^0, \mathbf{B}_0)$ contains no boundary point of $\mathbf{x}^0$. (In the more general case where $\mathrm{kn}(\mathbf{x}^0, \mathbf{B}_0) \subseteq \mathbf{x}^0$, it can be shown—using Brouwer's fixed point theorem—that in $\mathbf{x}^0$ a zero $\mathbf{x}^*$ of $\mathbf{f}(\mathbf{x}) = \mathbf{o}$ exists.) However, (17) can hold without $\mathrm{kn}(\mathbf{x}^0, \mathbf{B}_0) \subset \mathbf{x}^0$ being true.

In passing we note that the assumption (18) which we have used for proving the statement about the order of convergence is not a very strong condition. See [1, § 3], for example.

4. Numerical examples.

Example 1. Let there be given the nonlinear system

$$\mathfrak{f}(\mathfrak{x}) = \begin{pmatrix} x_1^2 + x_2^2 - 1 \\ x_1^2 + x_2 \end{pmatrix} = \mathfrak{o}$$

where $\mathfrak{x} = (x_1, x_2)^T$. If we choose the interval vector

$$\mathbf{x}^0 = \begin{pmatrix} [0, 1] \\ [0, 1] \end{pmatrix},$$

then the condition (17) does not hold. Therefore a bisection procedure is used until this condition holds. One of the two interval vectors appearing during this process is stored in a last-in-first-out array. The bisection procedure is continued using the other one until either the condition $d(\mathbf{kn}(\mathbf{x}^0, \mathfrak{B}_0)) < d(\mathbf{x}^0)$ holds or a specific criterion for nonexistence of a solution holds.

We have used the same bisection process as it was discussed in [7, pp. 1058–1059]. After four bisection steps we get the interval vector

$$\mathbf{x}^0 = \begin{pmatrix} [0.5, 0.75] \\ [0.5, 0.75] \end{pmatrix}$$

for which $d(\mathbf{kn}(\mathbf{x}^0, \mathfrak{B}_0)) \leqq \alpha d(\mathbf{x}^0)$, $\alpha < 1$, holds. After the first step of Algorithm A we obtain

$$\mathbf{x}^1 \cap \mathbf{kn}(\mathbf{x}^1, \mathfrak{A}_1) = \emptyset$$

from which we conclude that this last $\mathbf{x}^0$ does not contain a solution of $\mathfrak{f}(\mathfrak{x}) = \mathfrak{o}$.

The next interval vector for which the condition $d(\mathbf{kn}(\mathbf{x}^0, \mathfrak{B}_0)) \leqq \alpha d(\mathbf{x}^0)$, $\alpha < 1$, holds is .

$$\mathbf{x}^0 = \begin{pmatrix} [0.75, 1] \\ [0.5, 0.75] \end{pmatrix}.$$

After 5 steps of Algorithm A where the statement (*) was always performed, the interval vector could no longer be improved using the machine under consideration. Table 1 contains $\|d(\mathbf{x}^k)\|_\infty$.

TABLE 1

k	$\|d(\mathbf{x}^k)\|_\infty$
0	0.25×10^0
1	0.1062×10^0
2	0.2149×10^{-1}
3	0.7897×10^{-3}
4	0.1073×10^{-5}
5	0.1994×10^{-11}
6	0.1066×10^{-13}

Example 2. We consider the nonlinear boundary value problem

$$y'' = y + \sin y, \quad y(0) = 0, \quad y(1) = 1.$$

If one chooses n points in the interval $(0, 1)$ in equal distance from each other and approximates the derivative by the central differences of the second order, then one gets the nonlinear system of simultaneous equations

$$2x_1 - x_2 + h^2(x_1 + \sin x_1) = 0,$$

$$-x_{i-1} + 2x_i - x_{i+1} + h^2(x_i + \sin x_i) = 0, \qquad i = 2(1)n-1,$$

$$-x_{n-1} + 2x_n + h^2(x_n + \sin x_n) - 1 = 0,$$

where $h = 1/(n+1)$.

We choose $n = 25$ and compute an interval vector $\mathbf{x}^0$ which includes the unique solution of this system by applying [8, 13.4.6c]. For this interval vector we have that $d(\mathrm{kn}(\mathbf{x}^0, \mathbf{B}_0)) \leqq \alpha d(\mathbf{x}^0)$, $\alpha < 1$. After 8 iteration steps of Algorithm A the enclosing interval vector could no longer be improved using the machine under consideration. Subsequently, we tabulate the iterates X_{13}^k which enclose the approximation for $y(\frac{1}{2})$. The iterates $\mathbf{x}^k$, $k = 2, 3$, have been computed by the statement (*), whereas (**) was used for $k = 4(1)8$. All rounding errors have been taken into account during the computation process.

TABLE 2

k	X_{13}^k
0	[−.500 000 000 000 45, .500 000 000 000 45]
1	[.395 999 883 156 03, .405 269 489 411 98]
2	[.398 687 073 619 27, .398 688 824 530 66]
3	[.398 688 025 543 73, .398 688 025 544 26]
4	[.398 688 025 543 75, .398 688 025 544 22]
5	[.398 688 025 543 76, .398 688 025 544 22]
6	[.398 688 025 543 77, .398 688 025 544 22]
7	[.398 688 025 543 78, .398 688 025 544 22]
8	[.398 688 025 543 79, .398 688 025 544 22]

The influence of rounding errors on Algorithm A was not discussed in § 3 of this paper. However, as was pointed out by an (anonymous to us) referee, these rounding errors can have a great influence on the practical behaviour of this algorithm. This can also be seen from the results given in Table 2. Theoretically the statement (*) of the algorithm will be used infinitely often, as soon as the region in which the algorithm is quadratically convergent is reached. Using a computer with fixed mantissa length for floating point numbers, this behaviour will be violated as soon as the limiting accuracy of the computer under consideration has been reached. The possible improvements from one step to the next can only be very small (if an improvement is possible at all). Therefore the following iterates are all computed using (**). Actually, this means a waste of computer time. This can be avoided, for example, by modifying the Algorithm A in the following manner: As soon as $d(\mathrm{kn}(\mathbf{x}^{\tilde{k}}, \mathbf{A}_{\tilde{k}})) > \alpha d(\mathbf{x}^{\tilde{k}})$ holds for some $\tilde{k} \geqq 1$, where α is defined by (17), we don't compute new values of $\mathbf{A}_k$ in the following steps, that is, we simply replace the statements of Algorithm A beginning with line 4 by

$$\mathbf{x}^{k+1} = \mathrm{kn}(\mathbf{x}^k, \mathbf{A}_{\tilde{k}}) \cap \mathbf{x}^k, \qquad k \geqq \tilde{k}.$$

Of course, one can also use more sophisticated modifications of the algorithm.

Both examples were computed using a Cyber 170 at the Zentraleinheit Rechenzentrum of the Technical University of Berlin, West Germany.

REFERENCES

[1] G. ALEFELD AND J. HERZBERGER, *Einführung in die Intervallrechnung*, Reihe Informatik 12, Bibliographisches Institut Mannheim, 1974.

[2] R. KRAWCZYK, *Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerschranken*, Computing, 4 (1969), pp. 187–201.

[3] R. E. MOORE, *A test for existence of solutions to nonlinear systems*, this Journal, 14 (1977), pp. 611–615.

[4] H. SCHWANDT, *Schnelle fast global konvergente Verfahren für die Fünf-Punkte-Diskretisierung der Poisson-Gleichung mit Dirichletschen Randbedingungen auf Rechteckgebieten*, Dissertation, Fachbereich Mathematik der TU Berlin, 1981.

[5] W. TOST, *Über einige Eigenschaften des Intervall-Gauss-Algorithmus und einige Anwendungen auf den Krawczyk-Algorithmus*, Diplomarbeit, Fachbereich Mathematik der TU Berlin, 1980 (not available).

[6] R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[7] R. E. MOORE AND S. T. JONES, *Safe starting regions for iterative methods*, this Journal, 14 (1977), pp. 1051–1065.

[8] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, San Francisco, London, 1970.

[9] E. KAUCHER AND S. M. RUMP, Private communication, 1981.