

Nullstelleneinschließung mit dem Newton-Verfahren ohne Invertierung von Intervallmatrizen

G. Alefeld und J. Herzberger

Eingegangen am 18. Dezember 1970

Including Zeros of Nonlinear Equations by Newton-method
Without Inverting Intervalmatrices

Summary. Some extensions and modifications of the well-known Newton-method, developed during the last few years by means of interval arithmetic are yielding lower and upper bounds for the solution of a set of nonlinear equations. To have superlinear convergence behaviour it is necessary, in general, to invert intervalmatrices. In this note we give a modification converging superlinear, without the necessity of inverting matrices. Two numerical examples are given.

Zusammenfassung. Verschiedene Erweiterungen und Modifikationen des Newton-Verfahrens gestatten es, eine Lösung eines nichtlinearen Gleichungssystems im R^n fortwährend einzuschließen [1, 7, 10]. Um bei diesen Verfahren überlineare Konvergenz zu erzielen, ist in jedem Iterationsschritt im allgemeinen eine Intervallmatrix zu invertieren. Es wird ein Verfahren angegeben, bei welchem diese Invertierung vermieden wird, ohne daß dabei die überlineare Konvergenz verlorenggeht.

1. Einleitung

Wir betrachten die Aufgabe, eine einfache Nullstelle einer vorgelegten Vektorfunktion f zu berechnen. Hierzu existieren Iterationsverfahren, bei deren Durchführung Folgen von Intervallvektoren bestimmt werden, welche diese Nullstelle fortwährend verbessert einschließen. Ein derartiges Verfahren wurde von Moore in [7] angegeben. Konvergenzkriterien dafür sowie zu einer Reihe verwandter Methoden sind in [1] enthalten. Die dort beschriebenen überlinear konvergenten Verfahren haben jedoch den Nachteil, daß sie in jedem Iterationsschritt die Invertierung einer Intervallmatrix erfordern. Dies bringt bei der praktischen Durchführung einen erheblichen Rechenaufwand mit sich. In der vorliegenden Arbeit entwickeln wir ein Verfahren, bei dessen Durchführung man ohne diese Berechnung von inversen Intervallmatrizen auskommt, aber trotzdem noch die gleiche Konvergenzgeschwindigkeit erhält. Diese Methode kann als Kombination des in [1] beschriebenen Verfahrens mit einer Verallgemeinerung eines in [2] angegebenen Algorithmus zur Matrixinvertierung angesehen werden. Diese Verallgemeinerungen werden zunächst in Lemma 1 und Satz 1 kurz hergeleitet. Anschließend bringen Satz 2 und Lemma 2 eine ausführliche Behandlung des Verfahrens. Abschließend sind zwei numerische Beispiele angegeben.

2. Bezeichnungen, Hilfsmittel

Es sollen hier kurz die für die folgenden Betrachtungen nötigen Begriffe und Hilfsmittel zusammengestellt werden. Vektoren des R^n bzw. reelle $n \times n$ -Matrizen bezeichnen wir mit $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ bzw. $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$; diese können als spezielle Intervallvektoren $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ aus $I(R^n)$ bzw. als spezielle Intervallmatrizen $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$ aus $M(I(R^n))$ aufgefaßt werden. Die wie üblich definierten Verknüpfungen zwischen Intervallmatrizen und -vektoren besitzen insbesondere die Einschließungseigenschaft, d.h. es gilt etwa

$$\mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{Y} \Rightarrow \mathbf{x} * \mathbf{y} \in \mathbf{X} * \mathbf{Y}, * \in \{+, -, \cdot\}. \quad (1)$$

Unter der Durchmessermatrix $d(\mathbf{X})$ einer Intervallmatrix $\mathbf{X}$ versteht man die reelle Matrix, deren Elemente $d_{ij}(\mathbf{X})$ die Durchmesser $d(X_{ij})$ der entsprechenden Elemente von $\mathbf{X}$ sind. Die Betragmatrix $|\mathbf{X}|$ einer Intervallmatrix $\mathbf{X}$ ist diejenige reelle Matrix, deren Elemente jeweils der größte vorkommende Betrag aus dem entsprechenden Element von $\mathbf{X}$ sind. Es ergeben sich in diesem Zusammenhang folgende Rechenregeln:

$$d(\mathbf{X} + \mathbf{Y}) = d(\mathbf{X}) + d(\mathbf{Y}), \quad (2)$$

$$d(\mathbf{X} \cdot \mathbf{Y}) = d(\mathbf{X}) \cdot |\mathbf{Y}|, d(\mathbf{X} \cdot \mathbf{Y}) = |\mathbf{X}| \cdot d(\mathbf{Y}), \quad (3)$$

$$d(\mathbf{X} \cdot \mathbf{Y}) \leq d(\mathbf{X}) \cdot |\mathbf{Y}| + |\mathbf{X}| \cdot d(\mathbf{Y}), \quad (4)$$

$$d(\mathbf{X}) = \mathbf{0} \Leftrightarrow \mathbf{X} = \mathbf{X}, \quad (5)$$

$$|\mathbf{X} + \mathbf{Y}| \leq |\mathbf{X}| + |\mathbf{Y}|. \quad (6)$$

Weitere Regeln finden sich in [3]. Darüber hinaus machen wir Gebrauch von dem in [2] eingeführten Mittelpunktoperator m , welcher jeder Intervallmatrix $\mathbf{X} = (X_{ij}) = ([x_{ij}^1, x_{ij}^2])$ die reelle Matrix

$$m(\mathbf{X}) = \left(\frac{x_{ij}^1 + x_{ij}^2}{2} \right)$$

zuordnet. Er ist entsprechend für Intervallvektoren definiert. Es ergeben sich die folgenden Eigenschaften:

$$m \text{ ist ein stetiger Operator}, \quad (7)$$

$$m(\mathbf{X} \pm \mathbf{Y}) = m(\mathbf{X}) \pm m(\mathbf{Y}), \quad (8)$$

$$m(\mathbf{X} \cdot \mathbf{Y}) = \mathbf{X} \cdot m(\mathbf{Y}), \quad (9)$$

$$m(\mathbf{X} \cdot \mathbf{Y}) = m(\mathbf{X}) \cdot \mathbf{Y},$$

$$m(\mathbf{X}) = \mathbf{X}. \quad (10)$$

Eine Folge von Intervallmatrizen $\{\mathbf{X}_n\}_{n=0}^\infty$ bzw. -vektoren $\{\mathbf{x}_n\}_{n=0}^\infty$ konvergiert, falls dies komponentenweise für die Folgen der Schranken der Fall ist. Damit konvergiert jede monoton fallende Folge

$$\mathbf{x}_0 > \mathbf{x}_1 > \mathbf{x}_2 > \dots$$

bzw.

$$\mathfrak{x}_0 > \mathfrak{x}_1 > \mathfrak{x}_2 > \dots$$

stets gegen ein Grenzelement $\mathfrak{x}$ bzw. $\mathfrak{r}$. Diese Folgen konvergieren genau dann gegen eine reelle Matrix $\mathfrak{X}$ bzw. einen reellen Vektor $\mathfrak{r}$ wenn die Folge $\{d(\mathfrak{x}_k)\}_{k=0}^{\infty}$ bzw. $\{d(\mathfrak{r}_k)\}_{k=0}^{\infty}$ gegen die Nullmatrix bzw. den Nullvektor konvergiert. Bei diesem Konvergenzbegriff erweisen sich alle eingangs erwähnten Intervalloperationen als stetige Verknüpfungen. Bezüglich weiterer intervallarithmetischer Hilfsmittel verweisen wir auf die Literatur [5, 7].

3. Herleitung der Verfahren

Zur Vorbereitung beweisen wir ein Lemma, das eine Verallgemeinerung eines in [2] bewiesenen Satzes darstellt. $\mathfrak{E}$ bezeichne im folgenden die Einheitsmatrix.

Lemma 1. Gegeben sei eine Folge von Intervallmatrizen $\{\mathfrak{A}_k\}_{k=0}^{\infty}$, für die $\mathfrak{A} \in \mathfrak{A}_k$, $k \geq 0$, mit einer nichtsingulären Matrix $\mathfrak{A}$ gilt. Ferner sei $\lim_{k \rightarrow \infty} \mathfrak{A}_k = \mathfrak{A}$ und es sei $\mathfrak{x}_0$ eine Intervallmatrix mit $\mathfrak{A}^{-1} \in \mathfrak{x}_0$.

Für die Iterationsvorschrift

$$(I) \quad \mathfrak{x}_{k+1} := \{m(\mathfrak{x}_k) + \mathfrak{x}_k \cdot [\mathfrak{E} - \mathfrak{A}_k \cdot m(\mathfrak{x}_k)]\} \cap \mathfrak{x}_k, \quad k = 0, 1, 2, \dots,$$

gelten die Aussagen:

- a) $\mathfrak{A}^{-1} \in \mathfrak{x}_k$, $k \geq 0$,
- b) $(\mathfrak{x} \in \mathfrak{x}_0 \Rightarrow \exists \mathfrak{x}^{-1}) \Rightarrow \lim_{k \rightarrow \infty} \mathfrak{x}_k = \mathfrak{A}^{-1}$.

Beweis. Zu a). Sei $\mathfrak{A}^{-1} \in \mathfrak{x}_k$, dann folgt wegen Eigenschaft (1)

$$\begin{aligned} \mathfrak{A}^{-1} &= m(\mathfrak{x}_k) + \mathfrak{A}^{-1} \cdot [\mathfrak{E} - \mathfrak{A} \cdot m(\mathfrak{x}_k)] \\ &\in \{m(\mathfrak{x}_k) + \mathfrak{x}_k \cdot [\mathfrak{E} - \mathfrak{A}_k \cdot m(\mathfrak{x}_k)]\} \cap \mathfrak{x}_k = \mathfrak{x}_{k+1}. \end{aligned}$$

Zu b). Die monoton fallende Folge $\mathfrak{x}_0 > \mathfrak{x}_1 > \mathfrak{x}_2 > \dots$ konvergiert gegen die Matrix $\mathfrak{x}$ und damit folgt wegen (7) notwendig $\lim_{k \rightarrow \infty} m(\mathfrak{x}_k) = m(\mathfrak{x})$. Aus der Stetigkeit der Intervallverknüpfungen sowie aus (7) folgt, daß der stetige Intervallausdruck

$$\mathfrak{y}_{k+1} := m(\mathfrak{x}_k) + \mathfrak{x}_k \cdot [\mathfrak{E} - \mathfrak{A}_k \cdot m(\mathfrak{x}_k)]$$

für $k \rightarrow \infty$ gegen eine Intervallmatrix $\mathfrak{y}$ konvergiert. Somit folgt

$$\begin{aligned} \mathfrak{y} &= m(\mathfrak{x}) + \mathfrak{x} \cdot [\mathfrak{E} - \mathfrak{A} \cdot m(\mathfrak{x})], \\ \mathfrak{x} &= \mathfrak{y} \cap \mathfrak{x}, \quad \text{und damit} \quad \mathfrak{x} \subset \mathfrak{y}. \end{aligned}$$

Daraus folgt aber $m(\mathfrak{x}) \in \mathfrak{y}$. Wir betrachten jetzt die j -te Spalte der Matrix $m(\mathfrak{x})$:

$$m(\mathfrak{x})_j = m(\mathfrak{x})_j + \mathfrak{x}^{(j)} \cdot [\mathfrak{E} - \mathfrak{A} \cdot m(\mathfrak{x})]_j, \quad \text{mit einer Matrix } \mathfrak{x}^{(j)} \in \mathfrak{x}_0,$$

oder

$$\mathfrak{x}^{(j)} \cdot [\mathfrak{E} - \mathfrak{A} \cdot m(\mathfrak{x})]_j = 0.$$

Da $\mathfrak{X}^{(j)} \in \mathfrak{X}_0$ und $\mathfrak{A}$ nichtsingulär sind, folgt wegen $[\mathfrak{A} \cdot m(\mathfrak{X})]_j = \mathfrak{A} \cdot m(\mathfrak{X})_j$ für jede Spalte $m(\mathfrak{X})_j = \mathfrak{A}_j^{-1}$, also $\mathfrak{X} = \mathfrak{A}^{-1}$.

Bemerkungen. 1. Das folgende praktisch nachprüfbare Kriterium ist für das in Lemma 1 b) angegebene Kriterium hinreichend:

$$\mathfrak{X} \in \mathfrak{X}_0 \Rightarrow \varrho(|\mathfrak{E} - \mathfrak{A} \cdot \mathfrak{X}|) < 1 \quad (\varrho = \text{Spektralradius}). \quad (11)$$

2. Im Falle $\mathfrak{A}_k = \mathfrak{A}$ enthält Lemma 1 als Spezialfall ein in [2] behandeltes Verfahren zur Matrizeninvertierung.

3. Es gelte $\mathfrak{A}_0 > \mathfrak{A}_1 > \mathfrak{A}_2 > \dots$ mit $\lim_{k \rightarrow \infty} \mathfrak{A}_k = \mathfrak{A}$.

Aus $\{\mathfrak{A}^{-1} : \mathfrak{A} \in \mathfrak{A}\} \subset \mathfrak{X}_0$ folgt bei Verfahren (I) stets $\{\mathfrak{A}^{-1} : \mathfrak{A} \in \mathfrak{A}\} \subset \mathfrak{X}_k$, $k \geq 0$. Für $\mathfrak{A}_k = \mathfrak{A}$ erhält man aus (I) ein Verfahren zur Invertierung einer Intervallmatrix $\mathfrak{A}$, welches ähnlich einem von Krückeberg [11] angegebenen ist.

Nach diesen Vorbereitungen wenden wir uns nun der eingangs formulierten Aufgabe zu. Vorgelegt sei also die reelle Vektorgleichung

$$\mathfrak{f}(\mathfrak{x}) = 0$$

mit $\mathfrak{x} = (x_1, x_2, \dots, x_n)$ und $\mathfrak{f}(\mathfrak{x}) = (f_1(\mathfrak{x}), f_2(\mathfrak{x}), \dots, f_n(\mathfrak{x}))$. Die partiellen Ableitungen $\partial f_i / \partial x_j$, $1 \leq i, j \leq n$, mögen im Quader $\mathfrak{a} = (A_1, A_2, \dots, A_n)$ existieren und stetig sein. Für festes $\mathfrak{x}$ gilt dann bekanntlich

$$f_i(\mathfrak{z}) = f_i(\mathfrak{x}) + \sum_{j=1}^n \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathfrak{x} + \theta_i \cdot (\mathfrak{z} - \mathfrak{x})} \cdot (z_j - x_j), \quad \theta_i \in (0, 1)$$

oder kurz

$$\mathfrak{f}(\mathfrak{z}) = \mathfrak{f}(\mathfrak{x}) + \mathfrak{S}(\mathfrak{x}) \cdot (\mathfrak{z} - \mathfrak{x}),$$

wobei

$$\mathfrak{S}(\mathfrak{x}) = \left(\left. \frac{\partial f_i}{\partial x_j} \right|_{\mathfrak{x} + \theta_i \cdot (\mathfrak{z} - \mathfrak{x})} \right)$$

gesetzt wird¹. Ist $\mathfrak{z}$ eine Nullstelle von $\mathfrak{f}$ und $\mathfrak{S}(\mathfrak{x})$ nichtsingulär, dann gilt:

$$\mathfrak{z} = \mathfrak{x} - \mathfrak{S}^{-1}(\mathfrak{x}) \cdot \mathfrak{f}(\mathfrak{x}).$$

Wir betrachten nun Intervallmatrizen $\mathfrak{S}(\mathfrak{x})$, die zu $\mathfrak{x} \in I(R^n)$ Abschätzungen für die Mengen $\{\mathfrak{S}(\mathfrak{x}) : \mathfrak{x} \in \mathfrak{x}\}$ sein sollen, d.h. es soll jeweils $\{\mathfrak{S}(\mathfrak{x}) : \mathfrak{x} \in \mathfrak{x}\} \subset \mathfrak{S}(\mathfrak{x})$ gelten. Solche Abschätzungen liefern etwa die Intervallmatrizen $\mathfrak{S}(\mathfrak{x})|_{\mathfrak{x}}$, welche intervallmäßige Auswertungen der Matrix $\mathfrak{S}(\mathfrak{x})$ sind. Ein Verfahren zur Bestimmung von $\mathfrak{z}$ geht aus dem folgenden Satz hervor.

Satz 1. Sei $\mathfrak{x}_0 \in I(R^n)$ und $\mathfrak{z}$ einzige Nullstelle von $\mathfrak{f}$ in $\mathfrak{x}_0$. Ferner sei $\{\mathfrak{B}_k\}_{k=0}^\infty$ eine Folge von Intervallmatrizen mit $\mathfrak{B}_0 > \mathfrak{B}_1 > \mathfrak{B}_2 > \dots$.

Dann gelten für das Iterationsverfahren

$$(II) \quad \mathfrak{x}_{k+1} = [m(\mathfrak{x}_k) - \mathfrak{B}_k \cdot \mathfrak{f}(m(\mathfrak{x}_k))] \cap \mathfrak{x}_k, \quad k = 0, 1, 2, \dots,$$

¹ Man beachte im folgenden, daß die Matrix $\mathfrak{S}(\mathfrak{x})$ stets noch von einem Vektor $\mathfrak{z}$ abhängt.

die folgenden Aussagen:

a) Aus $\eta \in \mathfrak{X}_k$ folge² stets $\{\mathfrak{S}^{-1}(\mathfrak{x}) : \mathfrak{x} \in \mathfrak{X}_k\} \subset \mathfrak{B}_k$; dann gilt:

$$\mathfrak{z} \in \mathfrak{X}_k, \quad k \geq 0.$$

b) Existiert für alle $\mathfrak{B} \in \mathfrak{B}_k$, $k \geq 0$, stets $\mathfrak{B}^{-1}$ dann gilt:

$$\lim_{k \rightarrow \infty} \mathfrak{X}_k = \mathfrak{z}.$$

Beweis. Zu a). Nach Voraussetzung ist $m(\mathfrak{x}_0) \in \mathfrak{X}_0$ und somit folgt, wie oben gezeigt wurde,

$$\begin{aligned} \mathfrak{z} &= m(\mathfrak{x}_0) - \mathfrak{S}^{-1}(m(\mathfrak{x}_0)) \cdot \mathfrak{f}(m(\mathfrak{x}_0)) \\ &\in [m(\mathfrak{x}_0) - \mathfrak{B}_0 \cdot \mathfrak{f}(m(\mathfrak{x}_0))] \cap \mathfrak{X}_0 \\ &= \mathfrak{X}_1. \end{aligned}$$

Entsprechend folgt dann $\mathfrak{z} \in \mathfrak{X}_k$, $k \geq 2$.

Zu b). Die monoton fallende Folge $\mathfrak{x}_0 > \mathfrak{x}_1 > \mathfrak{x}_2 > \dots$ konvergiert gegen $\mathfrak{x}$ und entsprechend konvergiert $\mathfrak{B}_0 > \mathfrak{B}_1 > \mathfrak{B}_2 > \dots$ gegen $\mathfrak{B}$. Für $\mathfrak{x}$ gilt:

$$\begin{aligned} \eta &= m(\mathfrak{x}) - \mathfrak{B} \cdot \mathfrak{f}(m(\mathfrak{x})), \\ \mathfrak{x} &= \eta \cap \mathfrak{x}, \quad \text{d.h. } m(\mathfrak{x}) \in \eta. \end{aligned}$$

Wir zeigen, daß $d(\mathfrak{x}) = 0$ gilt, woraus mit a) dann $\lim_{k \rightarrow \infty} \mathfrak{X}_k = \mathfrak{z}$ folgt. Wäre etwa $m(\mathfrak{x}) \neq \mathfrak{z}$, dann folgte $m(\mathfrak{x}) = m(\mathfrak{x}) - \mathfrak{B} \cdot \mathfrak{f}(m(\mathfrak{x}))$ und damit $\mathfrak{f}(m(\mathfrak{x})) = 0$ oder der Widerspruch $m(\mathfrak{x}) = \mathfrak{z}$.

Bemerkung. Bei der praktischen Anwendung von Verfahren (II) besteht die Notwendigkeit, in jedem Schritt eine Einschließung $\mathfrak{B}_k$ neu zu berechnen. Dies könnte etwa durch Invertierung von $\mathfrak{S}(\mathfrak{x}_k)$ mit Hilfe des in Lemma 1, Bemerkung 3, erwähnten Verfahrens geschehen. Wir wollen nun Verfahren (II) so abändern, daß diese Invertierung nicht mehr nötig ist.

Satz 2. Sei $\mathfrak{x}_0 \in I(R^n)$ und $\mathfrak{z}$ einzige Nullstelle von $\mathfrak{f}(\mathfrak{x})$ in $\mathfrak{X}_0$. Ferner sei $\mathfrak{X}_0$ eine Intervallmatrix, welche die Inversen $\mathfrak{S}^{-1}(\mathfrak{x})$, $\mathfrak{x} \in \mathfrak{X}_0$, enthält. Für die Iterationsvorschrift

$$\begin{aligned} \text{(III)} \quad \mathfrak{X}_{k+1} &= [m(\mathfrak{x}_k) - \mathfrak{X}_k \cdot \mathfrak{f}(m(\mathfrak{x}_k))] \cap \mathfrak{X}_k \\ \mathfrak{X}_{k+1} &= \{m(\mathfrak{x}_k) + \mathfrak{X}_k \cdot [\mathfrak{E} - \mathfrak{S}(\mathfrak{x}_{k+1}) \cdot m(\mathfrak{x}_k)]\} \cap \mathfrak{X}_k \quad k = 0, 1, 2, \dots, \end{aligned}$$

mit

$$\mathfrak{S}(\mathfrak{x}_k) = \mathfrak{S}(\mathfrak{x})|_{\mathfrak{X}_k}$$

gelten dann folgende Aussagen:

a) Für alle $k \geq 0$ gilt $\mathfrak{z} \in \mathfrak{X}_k$.

b) Existiert für alle $\mathfrak{X} \in \mathfrak{X}_0$ die Inverse $\mathfrak{X}^{-1}$, dann gilt $\lim_{k \rightarrow \infty} \mathfrak{X}_k = \mathfrak{z}$.

Beweis. Wir führen die obigen Aussagen auf die entsprechenden Aussagen von Satz 1 zurück. Dazu setzen wir $\mathfrak{B}_k = \mathfrak{X}_k$ und erhalten dann die dort unter

² Die $\mathfrak{S}(\mathfrak{x})$ hängen hier von η ab.

a) aufgeführte Zusatzvoraussetzung wie folgt: Wie im Beweis zu Satz 1 a) folgt unmittelbar $\beta \in \mathfrak{X}_1$. Sei nun $\mathfrak{x} \in \mathfrak{X}_1 \subset \mathfrak{X}_0$, dann ist $\mathfrak{S}^{-1}(\mathfrak{x}) \in \mathfrak{X}_0$ und $\mathfrak{S}(\mathfrak{x}) \in \mathfrak{H}(\mathfrak{X}_1)$ und mit (1) folgt

$$\begin{aligned}\mathfrak{S}^{-1}(\mathfrak{x}) &= m(\mathfrak{x}_0) + \mathfrak{S}^{-1}(\mathfrak{x}) \cdot [\mathfrak{E} - \mathfrak{S}(\mathfrak{x}) \cdot m(\mathfrak{x}_0)] \\ &\in \{m(\mathfrak{x}_0) + \mathfrak{x}_0 \cdot [\mathfrak{E} - \mathfrak{H}(\mathfrak{X}_1) \cdot m(\mathfrak{x}_0)]\} \cap \mathfrak{X}_0 \\ &= \mathfrak{X}_1.\end{aligned}$$

Entsprechend läßt sich dieser Schritt für beliebige k durchführen. Somit gilt a).

Die Folge $\{\mathfrak{X}_k\}_{k=0}^{\infty}$ erfüllt die Voraussetzungen an die Folge $\{\mathfrak{B}_k\}_{k=0}^{\infty}$ in Satz 1 und wir erhalten somit die obige Aussage b).

Nach dieser Herleitung von Verfahren (III) wollen wir nun noch kurz Aussagen über das Konvergenzverhalten machen und dabei einen Vergleich mit dem ursprünglichen Vorgehen (II) ziehen. Als Maß für die Konvergenzgeschwindigkeit dieser Iterationsverfahren betrachten wir das Konvergenzverhalten der Durchmesser $\{d(\mathfrak{X}_k)\}_{k=0}^{\infty}$ der Iterierten.

Wir fassen dies in folgendem Lemma zusammen.

Lemma 2. a) *Unter den Voraussetzungen von Satz 1 gilt für die nach (II) berechnete Iterationsfolge mit einer monotonen Norm $\|\cdot\|$:*

$$\|d(\mathfrak{X}_{k+1})\| \leq c \cdot \|d(\mathfrak{X}_k)\|.$$

Gilt darüberhinaus in Satz 1 noch $\lim_{k \rightarrow \infty} d(\mathfrak{B}_k) = 0$, dann gilt:

$$\|d(\mathfrak{X}_{k+1})\| \leq c_k \cdot \|d(\mathfrak{X}_k)\| \quad \text{mit} \quad \lim_{k \rightarrow \infty} c_k = 0,$$

d.h. es liegt überlineare Konvergenz vor.

b) *Unter den Voraussetzungen von Satz 2 gilt für die nach (III) berechnete Iterationsfolge mit einer monotonen Norm $\|\cdot\|$:*

$$\|d(\mathfrak{X}_{k+1})\| \leq c'_k \cdot \|d(\mathfrak{X}_k)\| \quad \text{mit} \quad \lim_{k \rightarrow \infty} c'_k = 0,$$

d.h. es liegt überlineare Konvergenz vor.

Beweis. Zu a). Es gilt unter Verwendung von (2)–(6):

$$\begin{aligned}d(\mathfrak{X}_{k+1}) &\leq d(\mathfrak{B}_k) \cdot |\mathfrak{f}(m(\mathfrak{x}_k))| \\ &= d(\mathfrak{B}_k) \cdot |\mathfrak{f}(m(\mathfrak{x}_k)) - \mathfrak{f}(\beta)| \\ &= d(\mathfrak{B}_k) \cdot |\mathfrak{S}(m(\mathfrak{x}_k)) \cdot [m(\mathfrak{x}_k) - \beta]| \\ &\leq d(\mathfrak{B}_k) \cdot |\mathfrak{S}(m(\mathfrak{x}_k))| \cdot |m(\mathfrak{x}_k) - \beta| \\ &\leq d(\mathfrak{B}_k) \cdot \|\mathfrak{H}(\mathfrak{X}_k)\| \cdot \frac{1}{2} \cdot d(\mathfrak{X}_k).\end{aligned}$$

Daraus folgt

$$\|d(\mathfrak{X}_{k+1})\| \leq \frac{1}{2} \cdot \|d(\mathfrak{B}_k)\| \cdot \|\mathfrak{H}(\mathfrak{X}_0)\| \cdot \|d(\mathfrak{X}_k)\|$$

mit

$$\begin{aligned}c_k &= \frac{1}{2} \|d(\mathfrak{B}_k)\| \cdot \|\mathfrak{H}(\mathfrak{X}_0)\| \\ &\leq \frac{1}{2} \cdot \|d(\mathfrak{B}_0)\| \cdot \|\mathfrak{H}(\mathfrak{X}_0)\| = c.\end{aligned}$$

Zu b). Wir folgern wie unter a) die Beziehung

$$\|d(\mathfrak{x}_{k+1})\| \leq c'_k \cdot \|d(\mathfrak{x}_k)\| \quad \text{mit} \quad c'_k = \frac{1}{2} \cdot \|d(\mathfrak{x}_k)\| \cdot \|\mathfrak{S}(\mathfrak{x}_k)\|.$$

Da mit $\lim_{k \rightarrow \infty} \mathfrak{x}_k = \mathfrak{z}$ aber $\lim_{k \rightarrow \infty} \mathfrak{S}(\mathfrak{x}_k) = \mathfrak{S}(\mathfrak{z})$ und $\mathfrak{S}(\mathfrak{z}) \in \mathfrak{S}(\mathfrak{x}_k)$ gelten, sind die Voraussetzungen von Lemma 1 erfüllt, woraus dann $\lim_{k \rightarrow \infty} \mathfrak{x}_k = \mathfrak{S}^{-1}(\mathfrak{z})$ und damit $\lim_{k \rightarrow \infty} \|d(\mathfrak{x}_k)\| = \lim_{k \rightarrow \infty} c'_k = 0$ folgt.

Bemerkung. Lemma 2 zeigt, daß Verfahren (III) ein ebenso gutes Konvergenzverhalten besitzt wie Verfahren (II), wo zur Erfüllung von $\lim_{k \rightarrow \infty} d(\mathfrak{S}_k) = \mathfrak{D}$ etwa jeweils die Invertierung von $\mathfrak{S}(\mathfrak{x}_k)$ notwendig war.

4. Numerische Beispiele

a) Gegeben sei das System

$$\mathfrak{f}(\mathfrak{x}) = \begin{pmatrix} x_1^4 + x_2^4 - 16 \\ x_2 - x_1^2 + 1 \end{pmatrix}, \quad (\text{s. [1]}).$$

Wir wählen

$$\mathfrak{x}_0 = \begin{pmatrix} [1, 3] \\ [0.25, 2] \end{pmatrix}$$

und erhalten für diesen Vektor

$$\mathfrak{S}^{-1}(\mathfrak{x}) \in \mathfrak{x}_0 = \left(\begin{array}{cc} \frac{1}{4X_1^3 + 8X_1 \cdot X_2^3} - \frac{X_2^3}{X_1^3 + 2X_1 \cdot X_2^3} & \\ \frac{1}{2X_1^2 + 4X_2^2} - \frac{X_1^2}{X_1^2 + 2X_2^2} & \end{array} \right) \Big|_{\mathfrak{x} = \mathfrak{x}_0}.$$

Man sieht unmittelbar, daß alle $\mathfrak{x} \in \mathfrak{x}_0$ invertierbar sind. Die nach Verfahren (III) berechneten Iterierten sind in Tabelle 1 angegeben:

Tabelle 1

n	X_1	X_2
0	[1, 3]	[0.25, 2]
1	[1, 1.99427001]	[0.423366477, 2]
2	[1.29590404, 1.99427002]	[1.38928294, 2]
3	[1.55598577, 1.99427002]	[1.70368002, 2]
4	[1.55598577, 1.6937756]	[1.70368002, 1.78875532]
5	[1.61060832, 1.69377556]	[1.70368002, 1.73659971]
6	[1.62601463, 1.65150179]	[1.70368002, 1.73302946]
7	[1.63899586, 1.65150179]	[1.70368002, 1.72048369]
8	[1.64551965, 1.64943361]	[1.71264757, 1.71579412]
9	[1.64745548, 1.64747143]	[1.71413391, 1.71414651]
10	[1.64746445, 1.64746448]	[1.71413915, 1.71413917]
11	[1.64746446, 1.64746447]	[1.71413916, 1.71413917]

Alle weiteren Näherungen stimmen mit $\mathfrak{x}_{11}$ überein.

b) Gegeben sei das System

$$2x_1 - \frac{1}{2}x_2 + \sin x_1 = 0 \quad -\frac{1}{2}x_1 + x_2 + e^{x_2} = 0.$$

Man kann zeigen, daß das System genau eine Lösung besitzt, welche im Vektor $\mathfrak{x}_0 = \begin{pmatrix} [-2, 2] \\ [-2, 2] \end{pmatrix}$ enthalten ist. Es gilt

$$\mathfrak{J}^{-1}(\mathfrak{x}) \in \mathfrak{x}_0 = \frac{1}{(1 + \cos(X_1))(1 + e^{X_2})^{\frac{1}{4}}} \begin{pmatrix} 1 + e^{X_2} & \frac{1}{2} \\ \frac{1}{2} & 2 + \cos(X_1) \end{pmatrix} \bigg|_{\mathfrak{x} = \mathfrak{x}_0}$$

Es sind alle Voraussetzungen von Satz 3 erfüllt. Die nach Verfahren III berechneten Näherungen sind in Tabelle 2 enthalten.

Tabelle 2

n	X_1	X_2
0	$[-2, 2]$	$[-2, 2]$
1	$[-0.322954777, -0.020066485]$	$[-1.93772867, -0.063564733]$
2	$[-0.159920105, -0.041102434]$	$[-0.965592645, -0.450837015]$
3	$[-0.159920105, -0.078450960]$	$[-0.636720800, -0.569599135]$
4	$[-0.117167309, -0.082322925]$	$[-0.604520562, -0.592923774]$
5	$[-0.099928144, -0.099918318]$	$[-0.599210581, -0.599205199]$
6	$[-0.099923420, -0.099923419]$	$[-0.599208115, -0.599208114]$

Alle weiteren Näherungen stimmen mit $\mathfrak{x}_6$ überein.

Die angegebenen Beispiele wurden auf der elektronischen Rechananlage Electrologica X8 am Rechenzentrum der Universität Karlsruhe gerechnet. Die angegebenen Zahlenwerte sind gerundet.

Literatur

1. Alefeld, G., Herzberger, J.: Über das Newton-Verfahren bei nichtlinearen Gleichungssystemen. Z. Angew. Math. Mech. **50**, 773—774 (1970).
2. Alefeld, G., Herzberger, J.: Verfahren höherer Ordnung zur iterativen Einschließung der inversen Matrix. Z. Angew. Math. Phys. **21**, 819—824 (1970).
3. Apostolatos, N., Kulisch, U.: Grundzüge einer Intervallrechnung für Matrizen und einige Anwendungen. Elektronische Rechananlagen **10**, 73—83 (1968).
4. Christ, H.: Realisierung einer Maschinenintervallarithmetic mit beliebigen ALGOL-60 Compilern. Elektronische Rechananlagen **10**, 217—222 (1968).
5. Kulisch, U.: Grundzüge der Intervallrechnung. Überblicke Mathematik **2**, 51—98, Bibliographisches Institut Mannheim (1969).
6. Mayer, O.: Algebraische und metrische Strukturen in der Intervallrechnung und einige Anwendungen. Computing **5**, 144—162 (1970).
7. Moore, R. E.: Interval analysis. Englewood Cliffs, N. J.: Prentice Hall Inc. 1966.
8. Schmidt, J. W., Leder, D.: Ableitungsfreie Verfahren ohne Auflösung linearer Gleichungen. Computing **5**, 71—81 (1970).

9. Ulm, S.: Über Iterationsverfahren mit sukzessiver Approximation des inversen Operators [Russ.]. Izr. Akad. Nauk Est. SSR **16**, 403—411 (1967).
10. Hansen, E. R.: On solving systems of equations using interval arithmetic. Math. Computation **22**, 374—384 (1968).
11. Krückeberg, F.: Inversion von Matrizen mit Fehlererfassung. Z. Ang. Math. Phys. **46**, T69 (1966).

Dr. Götz Alefeld
Dr. Jürgen Herzberger
Universität Karlsruhe
Institut für Angewandte Mathematik
D-7500 Karlsruhe, Englerstraße 2
Deutschland