Computing 6, 161—172 (1970)

Über die aus monoton zerlegbaren Operatoren gebildeten Iterationsverfahren

Von

G. Alefeld, Karlsruhe

(Eingegangen am 13. Juli 1969)

Zusammenfassung — Summary

Über die aus monoton zerlegbaren Operatoren gebildeten Iterationsverfahren. Vor allem von COLLATZ, SCHRÖDER und ALBRECHT wurden mit Hilfe des Begriffes des monotonen Operators verschiedene Iterationsverfahren betrachtet, welche Einschließungsmengen zum Beispiel für die Lösung eines linearen oder nichtlinearen Gleichungssystems liefern. Im folgenden wird der Zusammenhang zwischen diesen Verfahren und der Intervallrechnung hergestellt. Es zeigt sich, daß diese Verfahren als Spezialfälle in den auf Intervallbasis gebildeten Iterationsverfahren enthalten sind. Im abschließenden Teil der Arbeit wird ein allgemeiner Konvergenzsatz in einem RIESZschen Raum angegeben.

A Note on Iterative Methods Using Monoton Decomposable Operators. By means of monoton decomposable operators, COLLATZ, ALBRECHT and SCHRÖDER have considered a variety of iterative methods, by which they ge. inclusion sets f-r the solution of system of simultaneous linear or nonlinear equations for example. It is shown, that these methods are special cases of basic iterative methods for equations with interval coefficients. In the concluding part of this paper a general convergence theorem in a RIESZian-space is given.

1. Vorbemerkungen und Bezeichnungen

In der Menge $I(R)$ der reellen abgeschlossenen Intervalle $X = [x_1, x_2]$, $Y = [y_1, y_2]$, ... sind durch

$$X * Y := \{x * y \mid x \in X \wedge y \in Y\}, \quad * \in \{+, -, \cdot, :\}$$

vier Verknüpfungen definiert. $X * Y$ läßt sich allein unter Verwendung der Schranken von X und Y berechnen. Unter einem komplexen Intervall $Z = X + i\,Y$ verstehen wir hier eine komplexe Zahlenmenge der Form

$$Z = \{z = x + i\,y \mid x \in X \wedge y \in Y;\ X, Y \in I(R)\}.$$

Die Menge der komplexen Intervalle bezeichnen wir mit $I(C)$. Verknüpfungen zwischen komplexen Intervallen sind auf Verknüpfungen reeller Intervalle zurückgeführt. Die Menge der $n \times n$-Matrizen, deren Elemente aus $I(R)$ bzw. $I(C)$ sind, bezeichnen wir mit $M(I(R))$ bzw. $M(I(C))$, ihre Elemente mit großen deutschen Buchstaben: $\mathfrak{A}$, $\mathfrak{B}$, ..., die Menge der entsprechenden Vektoren mit $V(I(R))$ bzw. $V(I(C))$, ihre

Elemente mit kleinen deutschen Buchstaben: $\mathfrak{a}, \mathfrak{b}, \ldots$. Es gilt $M(I(R)) \subset \subset M(I(C))$ bzw. $V(I(R) \subset V(I(C))$. Im Gegensatz dazu bezeichnen wir die Menge der Matrizen, deren Elemente reelle bzw. komplexe Zahlen sind, mit $M(R)$ bzw. $M(C)$, die Menge der entsprechenden Vektoren mit $V(R)$ bzw. $V(C)$. Die Elemente von $M(R)$ und $M(C)$ bzw. $V(R)$ und $V_i(C)$ bezeichnen wir $\dot{\mathfrak{A}}, \dot{\mathfrak{B}}, \ldots$ bzw. $\dot{\mathfrak{a}}, \dot{\mathfrak{b}}, \ldots$. Verknüpfungen sind wie üblich definiert, also z. B. mit $\mathfrak{A} = (A_{ij})$ und $\mathfrak{b} = (B_i)$

$$\mathfrak{A} \cdot \mathfrak{b} = \left(\sum_{j=1}^{n} A_{ij} B_j \right).$$

2. Iterationsverfahren im Raum der Intervallvektoren

Gegeben sei eine Matrix $\mathfrak{A} = (A_{ij}) \in M(I(C))$ und ein Vektor $\mathfrak{b} \in V(I(C))$. Damit betrachten wir die Gleichung

$$\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}. \tag{1}$$

Kann man einen Vektor $\mathfrak{x}^* \in V(I(C))$ finden, der der Gleichung (1) genügt, so gilt für alle $\dot{\mathfrak{x}} = (\mathfrak{E} - \dot{\mathfrak{A}})^{-1}\,\dot{\mathfrak{b}}$ mit $\dot{\mathfrak{A}} \in \mathfrak{A}$ und $\dot{\mathfrak{b}} \in \mathfrak{b}$:

$$\dot{\mathfrak{x}} \in \mathfrak{x}^*,$$

d. h. $\mathfrak{x}^*$ enthält die Lösungen aller aus (1) bildbaren linearen Gleichungssysteme. Zur Bestimmung von $\mathfrak{x}^*$ sind verschiedene Iterationsverfahren betrachtet worden [2, 5, 11]:

a) $\mathfrak{x}_{m+1} := \mathfrak{A}\,\mathfrak{x}_m + \mathfrak{b}$,

b) $\mathfrak{x}_{m+1} := \mathfrak{L}\,\mathfrak{x}_{m+1} + (\mathfrak{D} + \mathfrak{R})\,\mathfrak{x}_m + \mathfrak{b}$,

c) $\mathfrak{y}_{m+1} := (1 - \omega)\,\mathfrak{y}_m + \omega\,\{\mathfrak{L}\,\mathfrak{y}_{m+1} + (\mathfrak{D} + \mathfrak{R})\,\mathfrak{y}_m + \omega\,\mathfrak{b}\}$, $\omega > 0$.

Wie üblich bezeichnen wir diese Iterationsverfahren als Gesamtschritt-, Einzelschritt- und Relaxationsverfahren.

Das Gesamtschrittverfahren a) konvergiert genau dann für beliebiges $\mathfrak{x}_0 \in V(I(C))$ gegen ein eindeutiges Grenzelement, wenn der Spektralradius $\varrho(|\mathfrak{A}|) < 1$ ist. Die reelle Matrix $|\mathfrak{A}|$ ist dabei folgendermaßen definiert: In der Menge $I(C)$ wird durch $p(A, B) := q(A_1, B_1) + q(A_2, B_2)$ ein Abstand (Metrik) für zwei komplexe Intervalle definiert. Dabei ist $q(X, Y) := \max(|x_1 - y_1|, |x_2 - y_2|)$ eine Metrik in der Menge der reellen Intervalle. $|X| := q(X, 0)$ heißt Betrag des reellen Intervalles. Entsprechend heißt $|A| := p(A, 0) = |A_1| + |A_2|$ Betrag des komplexen Intervalles $A = A_1 + i\,A_2$. Die Matrix $|\mathfrak{A}|$ entsteht aus der Intervallmatrix $\mathfrak{A}$, indem man elementweise den Betrag bildet $|\mathfrak{A}| = (|A_{ij}|)$. Das Relaxationsverfahren c) konvergiert genau dann für beliebiges $\mathfrak{y}_0 \in V(I(C))$ gegen ein eindeutiges Grenzelement, wenn der Spektralradius $\varrho((\mathfrak{E} - \omega\,|\,\mathfrak{L}\,|)^{-1}\{|1 - \omega|\,\mathfrak{E} + \omega\,(|\mathfrak{D}| + |\mathfrak{R}|)\}) < 1$ ist. Für $\omega = 1$ ist c) mit b) identisch. Gesamtschrittverfahren und Einzelschrittverfahren sind entweder beide konvergent oder beide divergent. Ist a) konvergent,

so läßt sich stets ein Intervall für ω angeben, so daß für Werte von ω aus diesem Intervall auch c) konvergiert. Das Relaxationsverfahren hat jedoch den entscheidenden Nachteil, daß sein Grenzelement für $\omega > 1$ im allgemeinen eine Obermenge des durch Iteration in Gesamtschritten zu erhaltenden Fixpunktes ist. (Siehe [2].) Gesamtschrittverfahren und Einzelschrittverfahren haben jedoch das gleiche Grenzelement. Die das Gesamtschritt- und Einzelschrittverfahren betreffenden Ergebnisse sind mit Hilfe von Normen für reelle Intervallvektoren und Intervallmatrizen von O. MAYER in [11] gewonnen worden. Die entsprechenden Ergebnisse für komplexe Probleme und die Aussagen über das Relaxationsverfahren findet man in [2].

3. Die aus monoton zerlegbaren Operatoren gebildeten Iterationsverfahren

Mit Hilfe des SCHAUDERschen Fixpunktsatzes kann man folgende Aussage 1 beweisen:

Aussage 1: *In einem halbgeordneten* BANACH*-Raum B sei die Gleichung*

$$u = T\,u + r = T^*\,u$$

vorgelegt, wobei sich T als Summe aus einem isotonen und einem antitonen Operator T_1 und T_2 schreiben läßt. Ausgehend von zwei Elementen v_0, w_0 aus dem Definitionsbereich von T werde durch die Vorschrift

$$\begin{aligned} v_{m+1} &= T_1\,v_m + T_2\,w_m + r \\ w_{m+1} &= T_1\,w_m + T_2\,v_m + r \end{aligned} \qquad m = 0, 1, 2, \ldots \tag{2}$$

v_1 und w_1 bestimmt, und es gelte

$$v_0 \leqslant v_1 \leqslant w_1 \leqslant w_0.$$

Der Operator T bildet dann das Intervall $\langle v_m, w_m\rangle$ in sich ab. Die Bildmenge $T\langle v_m, w_m\rangle$ sei für irgendein m kompakt. Dann besitzt die Gleichung $u = T^*u$ eine Lösung u^* und für diese gilt

$$v_m \leqslant u^* \leqslant w_m.$$

Im folgenden sei zunächst $B := V(R)$.

Gegeben sei das reelle lineare Gleichungssystem

$$\underset{\cdot}{\mathfrak{x}} = \underset{\cdot}{\mathfrak{A}}\,\underset{\cdot}{\mathfrak{x}} + \underset{\cdot}{\mathfrak{b}} \qquad (\underset{\cdot}{\mathfrak{A}} \in M(R),\ \underset{\cdot}{\mathfrak{b}} \in V(R))$$

Die Matrix $\underset{\cdot}{\mathfrak{A}}$ sei zerlegt in ihre monotonen Anteile $\underset{\cdot}{\mathfrak{A}}_1$ und $\underset{\cdot}{\mathfrak{A}}_2$. Mit $\underset{\cdot}{\mathfrak{A}}_1 = (a_{ij}^1)$ und $\underset{\cdot}{\mathfrak{A}}_2 = (a_{ij}^2)$ ist also

$$a_{ij}^1 \geqslant 0 \text{ und } a_{ij}^2 \leqslant 0,\ i, j = 1\,(1)\,n.$$

Es sei $\underset{\cdot}{\mathfrak{y}}_0 \leqslant \underset{\cdot}{\mathfrak{z}}_0$. Dann lautet (2):

$$\begin{aligned} \underset{\cdot}{\mathfrak{y}}_{m+1} &= \underset{\cdot}{\mathfrak{A}}_1\,\underset{\cdot}{\mathfrak{y}}_m + \underset{\cdot}{\mathfrak{A}}_2\,\underset{\cdot}{\mathfrak{z}}_m + \underset{\cdot}{\mathfrak{b}} \\ \underset{\cdot}{\mathfrak{z}}_{m+1} &= \underset{\cdot}{\mathfrak{A}}_1\,\underset{\cdot}{\mathfrak{z}}_m + \underset{\cdot}{\mathfrak{A}}_2\,\underset{\cdot}{\mathfrak{y}}_m + \underset{\cdot}{\mathfrak{b}} \end{aligned} \qquad m = 0, 1, 2, \ldots \tag{3}$$

Daneben betrachten wir das Gesamtschrittverfahren a) aus Abschnitt 2.:

$$\mathfrak{x}_{m+1} = \mathfrak{A}\,\mathfrak{x}_m + \mathfrak{b},\ \mathfrak{x}_0 \in V(I(R)),\ m = 0, 1, 2, \ldots. \tag{4}$$

Es sei $\mathfrak{x}_m = (X_m^i)$ mit $X_m^i = [x_m^{i1}, x_m^{i2}]$, $i = 1\,(1)\,n$. Wir zeigen, daß (3) und (4) identisch sind. Für die i-te Komponente von (4) ist

$$X_{m+1}^i = \sum_{j=1}^{n} a_{ij} X_m^j + b_i.$$

Ist $a_{ij} \geqslant 0$, so gilt nach den Regeln der Intervallrechnung für alle Intervalle X_m^i: $a_{ij} X_m^i = [a_{ij} x_m^{i1},\ a_{ij} x_m^{i2}]$.
Ist dagegen $a_{ij} \leqslant 0$, so gilt $a_{ij} X_m^i = [a_{ij} x_m^{i2},\ a_{ij} x_m^{i1}]$.
Damit lautet (4):

$$\mathfrak{x}_{m+1} = ([x_{m+1}^{i1}, x_{m+1}^{i2}]) = \Bigg(\Bigg[\sum_{j=1}^{n} a_{ij}^1 x_m^{j1} + \sum_{j=1}^{n} a_{ij}^2 x_m^{j2} + b_i,\ \sum_{j=1}^{n} a_{ij}^1 x_m^{j2} + \sum_{j=1}^{n} a_{ij}^2 x_m^{j1} + b_i\Bigg]\Bigg).$$

Daraus folgt nach Definition der Gleichheit von Intervallvektoren und Intervallen

$$\begin{aligned} x_{m+1}^{i1} &= \sum_{j=1}^{n} a_{ij}^1 x_m^{j1} + \sum_{j=1}^{n} a_{ij}^2 x_m^{j2} + b_i \\ x_{m+1}^{i2} &= \sum_{j=1}^{n} a_{ij}^1 x_m^{j2} + \sum_{j=1}^{n} a_{ij}^2 x_m^{j1} + b_i \end{aligned} \qquad i = 1\,(1)\,n.$$

Dies stimmt bis auf die Bezeichnungen mit (3) überein. Umgekehrt kann man jedes System (3) in der Form (4) schreiben. Aufgrund der in Abschnitt 2. für a) angegebenen Aussagen erhalten wir damit den folgenden

Satz 1: *Genau dann, wenn der Spektralradius* $\varrho(|\mathfrak{A}|)$ *kleiner als* 1 *ist, konvergieren im Iterationsverfahren* (3) *die Folgen* $\mathfrak{y}_m$ *bzw.* $\mathfrak{z}_m$ *für beliebige Vektoren* $\mathfrak{y}_0$ *und* $\mathfrak{z}_0$ *mit* $\mathfrak{y}_0 \leqslant \mathfrak{z}_0$ *gegen den eindeutigen und gemeinsamen Grenzwert* $\mathfrak{x}^*$.

Bemerkung 1: Daß $\mathfrak{y}_m$ und $\mathfrak{z}_m$ gegen einen gemeinsamen Grenzwert konvergieren, folgt aus der Tatsache, daß für alle m $\mathfrak{y}_m = \mathfrak{z}_m$ gilt, falls man $\mathfrak{y}_0 = \mathfrak{z}_0$ wählt.

Bemerkung 2: Aufgrund der für (4) leicht nachweisbaren Eigenschaft, daß aus $\mathfrak{x}^* \in \mathfrak{x}_0$ für alle m $\mathfrak{x}^* \in \mathfrak{x}_m$ folgt, haben wir also: Gilt in (3) $\mathfrak{y}_0 \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_0$, so gilt für alle m: $\mathfrak{y}_m \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_m$. Die Folgen $\mathfrak{y}_m$ und $\mathfrak{z}_m$ schließen also allein unter dieser Voraussetzung $\mathfrak{x}^*$ stets ein. Die in *Aussage 1* verlangte zusätzliche Voraussetzung $\mathfrak{y}_0 \leqslant \mathfrak{y}_1 \leqslant \mathfrak{z}_1 \leqslant \mathfrak{z}_0$ garantiert

die Monotonie der beiden Folgen. Diese kann man jedoch auch allein unter der Voraussetzung $\mathfrak{y}_0 \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_0$ dadurch erzwingen, daß man nach jedem Iterationsschritt den Durchschnitt zweier aufeinander folgender Iterierter bildet. (Siehe dazu auch [6, 7]). Dies ist insofern von Bedeutung, da es im allgemeinen einfacher ist, zwei Vektoren $\mathfrak{y}_0$ und $\mathfrak{z}_0$ zu bestimmen, für die $\mathfrak{y}_0 \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_0$ gilt, als die stärkere Bedingung $\mathfrak{y}_0 \leqslant \mathfrak{y}_1 \leqslant \mathfrak{z}_1 \leqslant \mathfrak{z}_0$ zu erfüllen.

Statt nun, wie in (3), in Gesamtschritten zu iterieren, kann man versuchen, die Konvergenz mit Hilfe des Einzelschrittverfahrens zu beschleunigen. Wir betrachten hier gleich den Fall des Relaxationsverfahrens. Das Einzelschrittverfahren erhält man für $\omega = 1$. Die Matrix $\mathfrak{A}$ sei zerlegt in $\mathfrak{A} = \mathfrak{L} + \mathfrak{D} + \mathfrak{R}$, wobei $\mathfrak{L}$ eine strenge untere, $\mathfrak{R}$ eine strenge obere Dreiecksmatrix und $\mathfrak{D}$ eine Diagonalmatrix ist. Außerdem seien die Matrizen

$$\omega\,\mathfrak{L} = \omega\,\mathfrak{L}_1 + \omega\,\mathfrak{L}_2,$$
$$\mathfrak{y} = (1-\omega)\,\mathfrak{E} = \mathfrak{y}_1 + \mathfrak{y}_2,$$
$$\omega\,(\mathfrak{D} + \mathfrak{R}) = \omega\,(\mathfrak{D}_1 + \mathfrak{R}_1) + \omega\,(\mathfrak{D}_2 + \mathfrak{R}_2),$$

wie angegeben in ihre monotonen Anteile zerlegt. Dann betrachten wir das Iterationsverfahren

$$\begin{aligned}
\mathfrak{y}_{m+1} &= \omega\,\mathfrak{L}_1\,\mathfrak{y}_{m+1} + \omega\,\mathfrak{L}_2\,\mathfrak{z}_{m+1} + \mathfrak{y}_1\,\mathfrak{y}_m + \mathfrak{y}_2\,\mathfrak{z}_m + \\
&+ \omega\,(\mathfrak{D}_1 + \mathfrak{R}_1)\,\mathfrak{y}_m + \omega\,(\mathfrak{D}_2 + \mathfrak{R}_2)\,\mathfrak{z}_m + \omega\,\mathfrak{b} \\
\mathfrak{z}_{m+1} &= \omega\,\mathfrak{L}_1\,\mathfrak{z}_{m+1}\ \ \omega\,\mathfrak{L}_2\,\mathfrak{y}_{m+1} + \mathfrak{y}_1\,\mathfrak{z}_m + \mathfrak{y}_2\,\mathfrak{y}_m + \\
&+ \omega\,(\mathfrak{D}_1 + \mathfrak{R}_1)\,\mathfrak{z}_m + \omega\,(\mathfrak{D}_2 + \mathfrak{R}_2)\,\mathfrak{y}_m + \omega\,\mathfrak{b}
\end{aligned} \qquad \omega > 0, \quad (5)$$

für welches *Aussage 1* wörtlich besteht. Genauso wie für das Gesamtschrittverfahren läßt sich zeigen, daß im Spezialfall $\mathfrak{A} = \mathfrak{A} \in M\,(R)$ das im Abschnitt 2. angegebene Relaxationsverfahren c) mit (5) identisch ist. Damit haben wir also den folgenden

Satz 2. *Genau dann, wenn der Spektralradius*

$$\varrho\,((\mathfrak{E} - \omega\,|\,\mathfrak{L}\,|)^{-1}\{|\,1-\omega\,|\,\mathfrak{E} + (|\,\mathfrak{D}\,| + |\,\mathfrak{R}\,|)\})$$

kleiner als 1 ist, konvergieren in (5) die Folgen $\mathfrak{y}_m$ und $\mathfrak{z}_m$ für beliebige Vektoren $\mathfrak{y}_0$ und $\mathfrak{z}_0$ mit $\mathfrak{y}_0 \leqslant \mathfrak{z}_0$ gegen den gemeinsamen und eindeutigen Grenzwert $\mathfrak{x}^*$.

Die im Anschluß an *Satz 1* angeführten *Bemerkungen* sind auch für (5) richtig.

Für die weiteren Überlegungen benötigen wir den

Hilfssatz:

a) *Ist* $\varrho\,(|\,\mathfrak{A}\,|) < 1$, *so ist*

$$\varrho\,((\mathfrak{E} - \omega\,|\,\mathfrak{L}\,|)^{-1}\{|1-\omega\,|\,\mathfrak{E} + (|\,\mathfrak{D}\,| + |\,\mathfrak{R}\,|)\}) < 1$$
$$\text{für } 0 < \omega < \frac{2}{1 + \varrho\,(|\,\mathfrak{A}\,|)}.$$

b) $$\min_{\omega} \varrho\left((\mathfrak{E} - \omega\,|\,\dot{\mathfrak{L}}\,|)^{-1}\{|\,1-\omega\,|\,\mathfrak{E} + \omega\,(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\}\right) =$$

$$= \varrho\left((\mathfrak{E} - |\,\dot{\mathfrak{L}}\,|)^{-1}(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\right) \text{ für } 0 < \omega < \frac{2}{1 + \varrho\,(|\,\dot{\mathfrak{A}}\,|)}.$$

Der Beweis findet sich in [2].

Mit diesen Aussagen kann man Angaben über die Konvergenzgeschwindigkeit der Verfahren (3) und (5) machen:

Setzen wir $\dot{\mathfrak{w}}_m := \dot{\mathfrak{z}}_m - \dot{\mathfrak{y}}_m$, so folgt aus (3) bzw. (5)

$$\dot{\mathfrak{w}}_m = |\,\dot{\mathfrak{A}}\,|^m\,\dot{\mathfrak{w}}_0 \quad \text{bzw.}$$

$$\dot{\mathfrak{w}}_m = \left((\mathfrak{E} - \omega\,|\,\dot{\mathfrak{L}}\,|)^{-1}\{|\,1-\omega\,|\,\mathfrak{E} + \omega\,(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\}\right)^m \dot{\mathfrak{w}}_0,$$

d. h. $-\ln\varrho\,(|\,\dot{\mathfrak{A}}\,|)$ bzw. $-\ln\varrho\left((\mathfrak{E} - \omega\,|\,\dot{\mathfrak{L}}\,|)^{-1}\{|\,1-\omega\,| + \omega\,(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\}\right)$ geben die asymptotische Konvergenzgeschwindigkeit an, mit der beim Verfahren (3) bzw. (5) die Folge $\dot{\mathfrak{w}}_m$ gegen den Nullvektor konvergiert. (Über die Definition der asymptotischen Konvergenzgeschwindigkeit siehe [14]). Wegen b) aus dem Hilfssatz erhalten wir also

Satz 3. *Die asymptotische Konvergenzgeschwindigkeit, mit der beim Einzelschrittverfahren ($\omega = 1$ in (5)) die Folge $\dot{\mathfrak{w}}_m$ gegen den Nullvektor konvergiert, läßt sich für den unter a) angegebenen Konvergenzbereich nicht mit dem Relaxationsverfahren beschleunigen.*

Der Spektralradius $\varrho\left((\mathfrak{E} - |\,\dot{\mathfrak{L}}\,|)^{-1}(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\right)$ ist nach bekannten Sätzen aus der Theorie der nichtnegativen Matrizen genau dann kleiner als 1, wenn $\varrho\,(|\,\dot{\mathfrak{A}}\,|) < 1$ ist, und zwar gilt dann $\varrho\left((\mathfrak{E} - |\,\dot{\mathfrak{L}}\,|)^{-1}(|\,\dot{\mathfrak{D}}\,| + |\,\dot{\mathfrak{R}}\,|)\right) \leqslant \varrho\,(|\,\dot{\mathfrak{A}}\,|)$. Damit haben wir

Satz 4. *Gesamtschrittverfahren (3) und Einzelschrittverfahren ($\omega = 1$ in (5)) sind entweder beide konvergent oder beide divergent. Im Falle der Konvergenz strebt beim Einzelschrittverfahren die Folge $\dot{\mathfrak{w}}_m$ asymptotisch mindestens ebenso schnell gegen den Nullvektor wie beim Gesamtschrittverfahren.*

Bemerkung 3. In [2] wurde gezeigt, daß man unter bestimmten Voraussetzungen über die Anfangsnäherungen Konvergenzvergleiche zwischen dem Einzelschritt- und Gesamtschrittverfahren durchführen kann, die nicht nur asymptotische Aussagen liefern, sondern für jeden Iterationsschritt bestehen. Auch dabei zeigt sich, daß das Einzelschrittverfahren dem Gesamtschrittverfahren stets überlegen ist. Diese Aussagen sind auch für die im Abschnitt 2. angegebenen Verfahren a) und b) richtig. (Siehe auch [1]).

Bemerkung 4. ALBRECHT hat in [1] die Theorie der monoton zerlegbaren Operatoren angewandt, um, wie es dort heißt, Gleichungssysteme mit fehlerhaften Koeffizienten zu behandeln. Dabei mußten sehr einschneidende Voraussetzungen über die Vorzeichenverteilung der Lösung gemacht werden (siehe auch [12]). Unter expliziter Verwendung der Rechenregeln für reelle Intervalle kann man zeigen, daß zum Beispiel die in [1]

Seite 354 durch die Formeln 3.10 bis 3.13.2 beschriebene Möglichkeit zur Behandlung solcher Systeme, mit dem Gesamtschrittverfahren a) aus Abschnitt 2. identisch ist, wenn man voraussetzt, daß man die Iteration mit einem Intervallvektor $\mathfrak{x}_0 \geqslant 0$ (d. h. $\dot{x}_0 \geqslant 0$ für alle $\dot{x}_0 \in \mathfrak{x}$) beginnt, und für alle m $\mathfrak{x}_m \geqslant 0$ gilt.

4. Bestimmung von Einschließungsmengen

Wegen des in *Bemerkung 2* angegebenen Tricks genügt es, zwei beliebige Vektoren $\mathfrak{y}_0$ und $\mathfrak{z}_0$ zu kennen, welche die Lösung der Gleichung $\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ einschließen: $\mathfrak{y}_0 \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_0$, um für alle m die Einschließung $\mathfrak{y}_m \leqslant \mathfrak{x}^* \leqslant \mathfrak{z}_m$ und die Monotonie der Folgen $\mathfrak{y}_m$ und $\mathfrak{z}_m$ zu sichern. Wir fassen die beiden reellen Vektoren $\mathfrak{y}_0$ und $\mathfrak{z}_0$ zum Intervallvektor $\mathfrak{x}_0$ zusammen.

Neben dem gegebenen Problem betrachten wir die Gleichung

$$\mathfrak{y} = \mathfrak{B}\,\mathfrak{y} + \mathfrak{b} \tag{6}$$

mit $\mathfrak{B} = (B_{ij}) \in M\,(I\,(R))$. Dabei sei $B_{ij} = [-|a_{ij}|, |a_{ij}|]$, $i, j = 1\,(1)\,n$, also $|\mathfrak{B}| = |\mathfrak{A}|$. Da $\varrho\,(|\mathfrak{B}|) = \varrho\,(|\mathfrak{A}|) < 1$ ist, besitzt die Gleichung (6) einen eindeutig bestimmten Fixpunkt $\mathfrak{y}^* \in V\,(I\,(R))$, den man zum Beispiel durch Iteration bestimmen kann. Wegen der speziellen Gestalt von $\mathfrak{B}$ läßt sich $\mathfrak{y}^*$ explizit angeben: $\mathfrak{y}^*$ ist symmetrisch zu $\mathfrak{b}$. Beginnt man nämlich die Iteration $\mathfrak{y}_{m+1} = \mathfrak{B}\,\mathfrak{y}_m + \mathfrak{b}$ mit einem beliebigen Vektor aus $V\,(I\,(R))$, so ist $\mathfrak{y}_1$ (und damit alle Iterierten) symmetrisch zu $\mathfrak{b}$, wie man leicht zeigen kann. Wir können daher für $\mathfrak{y}^*$ den Ansatz

$$\mathfrak{y}^* = ([b_i - \xi_i, b_i + \xi_i]),\ \xi_i \geqslant 0,\ i = 1\,(1)\,n$$

machen. Dies führt nach einfacher Rechnung auf

$$(\xi_i) = (\mathfrak{E} - |\mathfrak{A}|)^{-1}\,|\mathfrak{A}|\,|\mathfrak{b}|. \tag{7}$$

Damit ist $\mathfrak{y}^*$ bekannt. Wegen $\mathfrak{A} \in \mathfrak{B}$ gilt $\mathfrak{x}^* \in \mathfrak{y}^*$, d. h. $\mathfrak{y}^*$ und jeder Intervallvektor $\mathfrak{x}_0$ der $\mathfrak{y}^*$ enthält, liefert eine Einschließungsmenge für $\mathfrak{x}^*$. (7) aufzulösen ist nicht einfacher als die ursprüngliche Aufgabe. Jedoch läßt sich in gewissen Fällen (ξ_i) einfach abschätzen:

1. Die Matrix $|\mathfrak{A}|$ erfülle das Zeilensummenkriterium.

Gilt für einen Vektor $(\tilde{\xi}_i)$ $(\tilde{\xi}_i) \geqslant (\mathfrak{E} - |\mathfrak{A}|)^{-1}\,|\mathfrak{A}|\,|\mathfrak{b}|$, so gilt wegen (7) $\xi_i \leqslant \tilde{\xi}_i$, $i = 1\,(1)\,n$. Insbesondere kann man alle $\tilde{\xi}_i$ gleich wählen $\tilde{\xi}_i = \xi$, $i = 1\,(1)\,n$. Dies führt auf

$$\xi_i \leqslant \tilde{\xi} = \max_i \frac{\sum_{j=1}^{n} |a_{ij}| \cdot |b_j|}{1 - \sum_{j=1}^{n} |a_{ij}|}. \tag{8}$$

Unmittelbarer Übergang zur Vektornorm des maximalen Komponentenbetrages in (7) liefert im allgemeinen eine wesentlich schlechtere Abschätzung als (8).

2. Die Matrix $|\mathfrak{A}|$ erfülle das Spaltensummenkriterium.

Dann folgt aus (7) unmittelbar

$$\xi_i \leqslant \tilde{\tilde{\xi}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}| \cdot |b_j|}{1 - \max_j \sum_{i=1}^{n} |a_{ij}|}. \tag{9}$$

Im ersten Fall schließt also der Intervallvektor

$$\mathfrak{x}_0 := ([b_i - \tilde{\xi},\ b_i + \tilde{\xi}])$$

$\mathfrak{x}^*$ ein. Im zweiten Falle liefert

$$\mathfrak{x}_0 := ([b_i - \tilde{\tilde{\xi}},\ b_i + \tilde{\tilde{\xi}}])$$

die Einschließung.

Bemerkung 5. Die beiden hier angegebenen Möglichkeiten zur Bestimmung von Einschließungsmengen lassen sich entsprechend zur Bestimmung einer Obermenge des Fixpunktes $\mathfrak{x}^*$ der allgemeineren Gleichung

$$\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b} \quad \text{mit} \quad \mathfrak{A} \in M\,(I\,(C)),\ \mathfrak{b} \in V\,(I\,(C))$$

anwenden. Dazu ist in (8) bzw. (9) $|a_{ij}|$ durch $|A_{ij}|$ und $|b_i|$ durch $|b_i| + \frac{1}{2} d\,(B_i)$ zu ersetzen, wobei b_i der Mittelpunkt und $d\,(B_i)$ der Durchmesser der i-ten Komponente des Intervallvektors $\mathfrak{b} = (B_i)$ ist. Der Nachweis ergibt sich wie oben.

5. Verallgemeinerungen

Es sei jetzt zusätzlich vorausgesetzt, daß der durch einen Kegel halbgeordnete BANACH-Raum B ein RIESZscher Raum ist, d. h. mit zwei Elementen x und y ist auch stets das Supremum und Infimum vorhanden. Dann kann man den Betrag eines Elementes $x \in B$ durch

$$|x| := \sup(-x, x) \in B$$

einführen. Folgt aus $0 \leqslant x_n \leqslant y_n$ und $\|y_n\| \to 0$ stets $\|x_n\| \to 0$ (dies ist z. B. erfüllt, falls aus $0 \leqslant x \leqslant y$ stets $\|x\| \leqslant \|y\|$ folgt), so wird B durch

$$q(x, y) := |x - y| \in B$$

zu einem pseudometrischen Raum (B, q, B); besitzt der Ordnungskegel einen inneren Punkt, so ist der dadurch in B gegebene neue Konvergenzbegriff mit dem durch die Norm gegebenen identisch. Neben B betrachten

wir die Menge der Intervalle $X = [x_1, x_2]$, $Y = [y_1, y_2]$ ... Wir bezeichnen sie mit $I(B)$. Es ist $B \subset I(B)$. Durch die Festsetzung

$$p(X, Y) := \sup (q(x_1, y_1), q(x_2, y_2)) \in B$$

wird $I(B)$ zu einem pseudometrischen Raum $(I(B), p, (B, q, B))$.

Wir betrachten nun wieder die Gleichung

$$u = T u + r = T^* u.$$

Der Operator T^* bildet B in sich ab. T sei zerlegbar in $T = T_1 + T_2$. T_1 sei monoton wachsend und T_2 monoton fallend. T_1 und T_2 seien beschränkt, d. h. es existieren zwei lineare positive Operatoren $|T_1|$ und $|T_2|$, die B in sich abbilden und es gilt

$$|T_i v - T_i w|| \leqslant |T_i| \cdot |v - w|, \; i = 1, 2.$$

Durch

$$\hat{T}[v, w] := [T_1 v + T_2 w + r, T_1 w + T_2 v + r]$$

ist ein Operator $\hat{T}$ gegeben, der $I(B)$ in sich abbildet. Es seien $X = [x_1, x_2]$, $Y = [y_1, y_2]$ zwei beliebige Elemente aus dem Definitionsbereich $\tilde{D} \subseteq I(B)$. Dann gilt

$$p(\hat{T}[x_1, x_2], \hat{T}[y_1, y_2]) =$$

$$= p([T_1 x_1 + T_2 x_2 + r, T_1 x_2 + T_2 x_1 + r], [T_1 y_1 + T_2 y_2 + r, T_1 y_2 + T_1 y_1 + r]) =$$

$$= \sup (|T_1 x_1 + T_2 x_2 - T_1 y_1 - T_2 y_2|, |T_1 x_2 + T_2 x_1 - T_1 y_2 - T_1 y_1|) \leqslant$$

$$\leqslant \sup (|T_1| \cdot |x_1 - y_1| + |T_2| \cdot |x_2 - y_2|, |T_1| \cdot |x_2 - y_2| + |T_2| \cdot |x_1 - y_1|) \leqslant$$

$$\leqslant (|T_1| + |T_2|) \sup (|x_1 - y_1|, |x_2 - y_2|) = (|T_1| + |T_2|) p(X, Y).$$

Durch Anwendung des Fixpunktsatzes für pseudometrische Räume erhalten wir damit den folgenden

Satz 5. *Der Operator $\hat{T}$ besitze den Definitionsbereich $\tilde{D} \subseteq I(B)$. Liegen alle Elemente $\hat{T}[d_1, d_2]$ ($[d_1, d_2] \in \tilde{D}$) in $\tilde{D}$ und konvergiert die Reihe $\sum_{n=0}^{\infty} (|T_1| + |T_2|)^n x$ für jedes $x \in B$, so besitzt die Gleichung*

$$u = T u + r$$

eine eindeutig bestimmte Lösung $u^* \in B$. Die Folgen v_m und w_m mit

$$v_{m+1} = T_1 v_m + T_2 w_m + r,$$
$$w_{m+1} = T_1 w_m + T_2 v_m + r,$$

konvergieren für beliebige Elemente v_0, w_0 mit $v_0 \leqslant w_0$ und $[v_0, w_0] \in \tilde{D}$ gegen u^*.

Daß v_m und w_m gegen den gleichen Grenzwert konvergieren, folgt aus der Tatsache, daß $v_m = w_m$ für alle m gilt, falls man speziell $v_0 = w_0$ wählt.

Anwendungen dieses Satzes liegen nahe, wie der in dieser Arbeit ausführlich behandelte Fall $B := V(R)$ zeigt. Daneben liegt bei praktischen Aufgaben häufig der Raum $B := C^\circ$ der stetigen Funktionen vor. Dabei ist die Halbordnung punktweise erklärt, und die Definition des Betrages lautet für ein Element $x \in C^\circ$: $|x| := |x(t)|$. Der oben angegebene Konvergenzbegriff ist hier mit der gleichmäßigen Konvergenz einer Funktionenfolge identisch.

6. Beispiel

Zur Erläuterung von *Satz 5* betrachten wir das folgende nichtlineare System $\underset{\cdot}{x} = T\,\underset{\cdot}{x} + r$ mit $\underset{\cdot}{x} = (x_1, x_2, x_3)'$ und

$$T\,\underset{\cdot}{x} + r := \begin{Bmatrix} -\frac{1}{8}x_1 x_2 + \frac{1}{2\pi}\sin\frac{\pi x_2}{4} - \frac{1}{8}\sqrt{5(1+x_2 x_3)} + 2 \\ -\frac{\sqrt{2}}{\pi}\cos\frac{\pi x_1}{8} + \frac{1}{4}\ln(1+x_2) - \frac{1}{16}x_1 x_3 + \frac{3}{2} \\ \frac{1}{32}x_1^3 - \frac{1}{32}x_1 x_2 + \frac{1}{2e^2}e^{x_3} + 1 \end{Bmatrix};$$

Für $0 \leqq x_i \leqq 2$, $i = 1\,(1)\,3$, läßt sich T in die Summe aus einem monoton wachsenden Operator T_1 und einem monoton fallenden Operator T_2 zerlegen:

$$T_1\,\underset{\cdot}{x} := \begin{Bmatrix} \frac{1}{2\pi}\sin\frac{\pi x_2}{4} \\ -\frac{\sqrt{2}}{\pi}\cos\frac{\pi x_1}{8} + \frac{1}{4}\ln(1+x_2) \\ \frac{1}{32}x_1^3 - \frac{1}{32}x_1 x_2 + \frac{1}{2e^2}e^{x_3} \end{Bmatrix};$$

$$T_2\,\underset{\cdot}{x} := \begin{Bmatrix} -\frac{1}{8}x_1 x_2 - \frac{1}{8}\sqrt{5(1+x_2 x_3)} \\ -\frac{1}{16}x_1 x_3 \\ -\frac{1}{32}x_1 x_2 \end{Bmatrix}.$$

Für diese läßt sich nach einfacher Rechnung $|T_i\,\underset{\cdot}{x} - T_i\,\underset{\cdot}{y}| \leqq |T_i| \cdot |\underset{\cdot}{x} - \underset{\cdot}{y}|$, $i = 1, 2$, zeigen, wobei für die Summe der Matrizen $|T_1|$ und $|T_2|$ der Spektralradius $\varrho(|T_1| + |T_2|)$ kleiner als 1 ist. Außerdem kann man unmittelbar nachprüfen, daß der angegebene Definitionsbereich in sich abgebildet wird. Nach Satz 5 konvergieren daher die Folgen v_m und w_m gegen die eindeutige Lösung des gegebenen Systems.

Bemerkung 6: Die Konvergenz von v_m und w_m gegen den gemeinsamen und eindeutigen Grenzwert läßt sich mit *Aussage 1* aus Abschnitt 3 allein nicht folgern.

Tabelle 1. *Ergebnisse*

	Iterationsschritt	Untere Schranke	Obere Schranke
x_1	0	0	2
x_2		0	2
x_3		0	2
x_1	1	0.875000	1.87972
x_2		0.799613	1.45605
x_3		0.942667	1.75000
x_1	5	1.43178	1.49370
x_2		1.17858	1.22090
x_3		1.25976	1.32343
x_1	10	1.46197	1.46410
x_2		1.20057	1.20203
x_3		1.28728	1.28945
x_1	20	1.46303	1.46304
x_2		1.20130	1.20131
x_3		1.28835	1.28835

Literatur

[1] Albrecht, J.: Monotone Iterationsfolgen und ihre Verwendung zur Lösung linearer Gleichungssysteme. Num. Math. **3**, 345–358 (1961).

[2] Alefeld, G.: Intervallrechnung über den komplexen Zahlen und einige Anwendungen. Diss., Universität Karlsruhe. 1968.

[3] Apostolatos, N., und U. Kulisch: Über die Konvergenz des Relaxationsverfahrens bei nichtnegativen und diagonaldominanten Matrizen. Comp. **2**, 17–24 (1967).

[4] Apostolatos, N., und U. Kulisch: Grundlagen einer Maschinenintervallrechnung. Comp. **2**, 89–104 (1967).

[5] Apostolatos, N., und U. Kulisch: Grundlagen einer Intervallrechnung für Matrizen und einige Anwendungen. Elektron. Rechenanlagen **10**, H 2, 73–83 (1968).

[6] Bohl, E.: Iterationsverfahren mit Fehlerabschätzung für lineare Operatorgleichungen. Arch. Rational Mech. Anal. **30**, 285–296 (1968).

[7] Braess, D.: Die Konstruktion monotoner Iterationsfolgen zur Lösungseinschließung bei linearen Gleichungssystemen. Arch. Rational Mech. Anal. **9**, 97–106 (1962).

[8] Collatz, L.: Funktionalanalysis und Numerische Mathematik. Berlin-Göttingen-Heidelberg: Springer-Verlag. 1964.

[9] Herzberger, J.: Metrische Eigenschaften von Mengensystemen und einige Anwendungen. Diss., Universität Karlsruhe. 1969.

[10] Kulisch, U.: Grundzüge der Intervallrechnung. Erscheint im Bibliographischen Institut Mannheim.

[11] Mayer, O.: Über die in der Intervallrechnung auftretenden Räume und einige Anwendungen. Diss., Universität Karlsruhe. 1968.

[12] Schmidt, J. W.: Ausgangsvektoren für monotone Iteration bei linearen Gleichungssystemen. Num. Math. **6**, 78–88 (1964).

[13] Schmidt, J. W.: Konvergenzuntersuchungen und Fehlerabschätzungen für ein verallgemeinertes Iterationsverfahren. Arch. Rational Mech. Anal. **6**, 261–276 (1960).

[14] VARGA, R. S.: Matrix Iterative Analysis. Prentice Hall. 1962.
[15] SCHRÖDER, J.: Computing error bounds in solving linear systems. Math. Comp. **16**, 323—337 (1962).

Dr. Götz Alefeld
Institut für Angewandte Mathematik
und Rechenzentrum
Universität Karlsruhe
Englerstraße 2, D-75 Karlsruhe
Bundesrepublik Deutschland

Printed in Austria
Druck: Paul Gerin, A-1021 Wien