

Zur Konvergenz des SSOR-Verfahrens für nichtlineare Gleichungssysteme[★]

Götz Alefeld und Peter Volkmann

Herrn Helmut Brakhage, Kaiserslautern,
anlässlich seines sechzigsten Geburtstages am 8. 7. 1986 gewidmet

Universität Karlsruhe, Fakultät für Mathematik, Postfach 6980, D-7500 Karlsruhe 1,
Bundesrepublik Deutschland

Convergence of the SSOR-Method for Nonlinear Systems of Simultaneous Equations

Summary. In this paper we prove the convergence of the symmetric successive overrelaxation method if it is applied to certain nonlinear systems of simultaneous equations. These equations are obtained, for example, by discretizing nonlinear elliptic partial differential equations.

Zusammenfassung. In der vorliegenden Arbeit wird für die Konvergenz des symmetrischen Relaxationsverfahrens (SSOR-Verfahren) bei Anwendung auf eine Klasse *nichtlinearer* Gleichungssysteme die globale Konvergenz bewiesen. Diese Gleichungssysteme treten z.B. bei der Diskretisierung nichtlinearer partieller Differentialgleichungen auf.

Subject Classifications: AMS(MOS): 65J15, CR: G1.5.

0. Einleitung

Die Konvergenz des symmetrischen Relaxationsverfahrens (SSOR-Verfahren¹) wurde in letzter Zeit bei Anwendung auf *lineare* Gleichungssysteme eingehend untersucht. Siehe dazu etwa [2, 5] und [13].

In der vorliegenden Arbeit untersuchen wir die Konvergenz dieses Verfahrens bei Anwendung auf *nichtlineare* Gleichungssysteme. Unser Hauptresultat wird gewonnen unter Verwendung des in [3] eingeführten Begriffes sogenannter *monotoner Iterationsfunktionen* (MI-Funktionen). Nachfolgend wird ausführlich auf die Ergebnisse in [3] über MI-Funktionen Bezug genommen. Für das Verständnis ist daher die Kenntnis der Darlegungen aus den Abschnitten 1–5 aus [3] notwendig.

Die vorliegende Arbeit ist wie folgt aufgebaut: Zunächst wiederholen wir die Definition des Begriffes einer zu einer Funktion f gehörigen MI-Funk-

¹ Symmetric Successive Overrelaxation Method

tion. Daran anschließend beweisen wir für Produkte von zwei MI-Funktionen die Theoreme 1–3, die für eine einzelne MI-Funktion bereits in [3] bewiesen wurden. Beweise für diese Sätze müssen durchgeführt werden, da – wie in Bemerkung 2 gezeigt wird – das Produkt zweier MI-Funktionen i.allg. keine MI-Funktion ist und man sich daher nicht auf die Theoreme 1–3 aus [3] berufen kann. Die Wiederholung des Begriffes der regulären Zerlegung einer Funktion (siehe dazu [1], [3] und [8]) sowie der Zusammenhang zwischen regulären Zerlegungen und MI-Funktionen schließt sich an (Theorem 4). Wir wiederholen dann die Definition der von Rheinboldt als Verallgemeinerung der M-Matrizen eingeführten M-Funktionen und zeigen schließlich als Hauptergebnis dieser Arbeit in Theorem 5, daß für stetige surjektive M-Funktionen das SSOR-Verfahren für $0 < \omega \leq 1$ und für jeden Startvektor gegen die eindeutige Lösung des vorliegenden Gleichungssystems konvergiert. Abschließend zeigen wir, daß Funktionen mit den angegebenen Eigenschaften bei der Diskretisierung von nichtlinearen partiellen Differentialgleichungen entstehen.

1. Monotone und globale Konvergenz für Produkte von MI-Funktionen

Es sei X eine mit der in [3], Abschnitt 1, beschriebenen Halbordnung versehene Menge, in der ein Konvergenzbegriff mit den Eigenschaften (C1)–(C6) aus Abschnitt 2 in [3] erklärt ist. Die Menge Y sei mit einer nach oben und unten gerichteten Halbordnung versehen. Weiter sei eine Funktion $f: X \rightarrow Y$ gegeben. Die Gleichung

$$f(x) = y$$

soll gelöst werden.

Definition 1 (siehe [3], Abschnitt 5). Eine Funktion

$$g: X \times Y \rightarrow X$$

heißt zu f gehörige monotone Iterationsfunktion (MI-Funktion) wenn g die folgenden Eigenschaften (M1) bis (M4) besitzt:

(M1) g ist (schwach) monoton wachsend bezüglich beider Argumente;

(M2) $g(x, f(x)) = x$, $x \in X$;

(M3) $\left\{ \begin{array}{l} x \leq g(x, y) \Rightarrow f(g(x, y)) \leq y \\ x \geq g(x, y) \Rightarrow f(g(x, y)) \geq y \end{array} \right\}$ für $x \in X$, $y \in Y$;

(M4) Falls $x_k \uparrow x$ oder $x_k \downarrow x$, so folgt $g(x_k, y) \rightarrow g(x, y)$ für alle $y \in Y$. $\square$

Die Menge der zu f gehörenden MI-Funktionen wird mit $MI(f)$ bezeichnet.

Theorem 1. Es sei $f: X \rightarrow Y$ gegeben und g, h seien aus $MI(f)$. Es seien $x \in X$, $y \in Y$ und durch

$$\left. \begin{array}{l} a_0 = x \\ a_{k+1/2} = g(a_k, y) \\ a_{k+1} = h(a_{k+1/2}, y) \end{array} \right\} k = 0, 1, 2, \dots \quad (1)$$

sei in X eine Folge (a_k) definiert. Dann gelten

$$(a) \quad f(x) \leq y \Rightarrow f(a_k) \leq y, \quad k=0, 1, 2, \dots, a_k \uparrow, \quad (2)$$

$$(b) \quad f(x) \geq y \Rightarrow f(a_k) \geq y, \quad k=0, 1, 2, \dots, a_k \downarrow. \quad (3)$$

Beweis. Wir beweisen (a). Nach Voraussetzung gilt $f(x) = f(a_0) \leq y$. Es genügt daher zu zeigen: Für beliebiges $k \geq 0$ folgt aus $f(a_k) \leq y$

$$a_k \leq a_{k+1} \quad \text{und} \quad f(a_{k+1}) \leq y.$$

Da $g \in \text{MI}(f)$ ist, folgt unter Verwendung von (M1) und (M2)

$$a_{k+1/2} = g(a_k, y) \geq g(a_k, f(a_k)) = a_k \quad (4)$$

also $a_k \leq g(a_k, y)$ und daher wegen des ersten Teils von (M3)

$$f(a_{k+1/2}) = f(g(a_k, y)) \leq y.$$

Analog erhält man, da $h \in \text{MI}(f)$ ist und wegen (4),

$$a_{k+1} = h(a_{k+1/2}, y) \geq h(a_{k+1/2}, f(a_{k+1/2})) = a_{k+1/2} \geq a_k,$$

also $a_{k+1/2} \leq h(a_{k+1/2}, y)$ und daher wegen des ersten Teiles von (M3)

$$f(a_{k+1}) = f(h(a_{k+1/2}, y)) \leq y.$$

Damit ist (a) gezeigt. Der Beweis von (b) kann ähnlich geführt werden. $\square$

Theorem 2. Es sei $f: X \rightarrow Y$ gegeben und $g, h \in \text{MI}(f)$. Für $\underline{x}, \bar{x} \in X, y \in Y$ gelte

$$\underline{x} \leq \bar{x}, \quad f(\underline{x}) \leq y \leq f(\bar{x}). \quad (5)$$

Die Folgen $(\underline{a}_k)$ und $(\bar{a}_k)$ seien definiert durch

$$\underline{a}_0 = \underline{x}, \quad \underline{a}_{k+1/2} = g(\underline{a}_k, y), \quad \underline{a}_{k+1} = h(\underline{a}_{k+1/2}, y) \quad k=0, 1, 2, \dots \quad (6)$$

und

$$\bar{a}_0 = \bar{x}, \quad \bar{a}_{k+1/2} = g(\bar{a}_k, y), \quad \bar{a}_{k+1} = h(\bar{a}_{k+1/2}, y), \quad k=0, 1, 2, \dots \quad (7)$$

Dann gilt

$$\underline{a}_k \leq \bar{a}_k, \quad k=0, 1, 2, \dots \quad (8)$$

$$\underline{a}_k \uparrow \underline{a}, \quad \bar{a}_k \downarrow \bar{a}, \quad \underline{a} \leq \bar{a} \quad (9)$$

und

$$f(\underline{a}) = f(\bar{a}) = y. \quad (10)$$

Beweis. Wegen (5) gilt die Behauptung (8) für $k=0$. (8) gelte für ein $k \geq 0$. Dann gilt wegen (M1)

$$\underline{a}_{k+1/2} = g(\underline{a}_k, y) \leq g(\bar{a}_k, y) = \bar{a}_{k+1/2}$$

und daher wiederum wegen (M1)

$$\underline{a}_{k+1} = h(\underline{a}_{k+1/2}, y) \leq h(\bar{a}_{k+1/2}, y) = \bar{a}_{k+1}.$$

Somit gilt (8).

Nach Theorem 1 gilt $\underline{a}_k \uparrow$ und $\bar{a}_k \downarrow$. Daraus und wegen (8) folgt

$$\underline{a}_k \leq \bar{a}_l, \quad k, l = 0, 1, 2, \dots$$

Nach (C1) und (C2) aus [3] folgt somit $\underline{a}_k \uparrow \underline{a}$ und $\bar{a}_k \downarrow \bar{a}$. Aus (8) und (C6) aus [3] folgt dann $\underline{a} \leq \bar{a}$. Somit gilt (9).

Es bleibt noch (10) zu zeigen.

Aus dem Beweis von Theorem 1 entnimmt man, daß nicht nur $\underline{a}_k \uparrow$ und $\bar{a}_k \downarrow$ gelten, sondern daß sogar die Ungleichungen

$$\underline{a}_0 \leq \dots \leq \underline{a}_k \leq \underline{a}_{k+1/2} \leq \underline{a}_{k+1} \leq \dots, \quad k = 0, 1, 2, \dots,$$

und

$$\bar{a}_0 \geq \dots \geq \bar{a}_k \geq \bar{a}_{k+1/2} \geq \bar{a}_{k+1} \geq \dots, \quad k = 0, 1, 2, \dots$$

bestehen. Man zeigt damit wie für die Folgen $(\underline{a}_k)$ und $(\bar{a}_k)$, daß die Folgen

$$\{\underline{a}_0, \underline{a}_{1/2}, \underline{a}_1, \dots\} \quad \text{und} \quad \{\bar{a}_0, \bar{a}_{1/2}, \bar{a}_1, \dots\}$$

konvergent sind, und zwar gegen $\underline{a}$ bzw. $\bar{a}$. Daß die Grenzwerte gleich $\underline{a}$ bzw. $\bar{a}$ sind folgt aus Remark 2 in [3]. Da die Folge $\{\underline{a}_0, \underline{a}_{1/2}, \underline{a}_1, \dots\}$ monoton wachsend gegen $\underline{a}$ konvergiert, folgt wegen (M4), daß die Folge

$$\{g(\underline{a}_0, y), g(\underline{a}_{1/2}, y), g(\underline{a}_1, y), \dots\} \quad \text{gegen} \quad g(\underline{a}, y)$$

konvergiert. Außerdem gilt wegen $\underline{a}_0 \leq \underline{a}_{1/2} \leq \underline{a}_1 \leq \dots$ und wegen (M1) auch

$$g(\underline{a}_0, y) \leq g(\underline{a}_{1/2}, y) \leq g(\underline{a}_1, y) \leq \dots$$

und daher konvergiert wiederum wegen Remark 2 in [3] auch die Teilfolge $\{g(\underline{a}_1, y), g(\underline{a}_2, y), \dots\}$ gegen $g(\underline{a}, y)$. Da die Folge $\{\underline{a}_{1/2}, \underline{a}_{3/2}, \dots\}$ wiederum wegen Remark 2 in [3] gegen $\underline{a}$ konvergiert, folgt aus $\underline{a}_{k+1/2} = g(\underline{a}_k, y)$ für $k \rightarrow \infty$ zunächst $\underline{a} = g(\underline{a}, y)$ und damit schließlich wegen (M3) $f(\underline{a}) = y$. Entsprechend zeigt man $f(\bar{a}) = y$. $\square$

Bemerkung 1. In [3], Theorem 2, haben wir (mit teilweise anderen Bezeichnungen) die folgende Aussage bewiesen:

Es sei $f: X \rightarrow Y$ gegeben und es sei $g \in \text{MI}(f)$.

Für $\underline{x}, \bar{x} \in X, y \in Y$ gelte

$$\underline{x} \leq \bar{x}, \quad f(\underline{x}) \leq y \leq f(\bar{x}).$$

Die Folgen $(\underline{b}_k)$ und $(\bar{b}_k)$ seien definiert durch

$$\underline{b}_0 = \underline{x}, \quad \underline{b}_{k+1} = g(\underline{b}_k, y), \quad k = 0, 1, 2, \dots \quad (6')$$

und

$$\bar{b}_0 = \bar{x}, \quad \bar{b}_{k+1} = g(\bar{b}_k, y), \quad k = 0, 1, 2, \dots, \quad (7')$$

dann gilt

$$\underline{b}_k \leq \bar{b}_k, \quad k = 0, 1, 2, \dots$$

$$\underline{b}_k \uparrow \underline{b}, \quad \bar{b}_k \downarrow \bar{b}, \quad \underline{b} \leq \bar{b}$$

und

$$f(\underline{b}) = f(\bar{b}) = y.$$

Ohne auf die Einzelheiten eines (durch vollständige Induktion zu führenden) Beweises einzugehen, sei mitgeteilt, daß die nach (6), (7) berechneten Folgen in dem folgenden Sinne schneller konvergieren als (6'), (7'): Es gilt

$$\underline{a}_0 = \underline{b}_0 = \underline{x} \leq \dots \leq \underline{b}_k \leq \underline{a}_k \leq \underline{a} \leq \bar{a} \leq \bar{a}_k \leq \bar{b}_k \leq \dots \leq \bar{a}_0 = \bar{b}_0 = \bar{x}.$$

Die Verwendung von (6), (7) ist also insbesondere dann sinnvoll, wenn die Durchführung nicht wesentlich mehr Aufwand erfordert als (6'), (7').

Theorem 3. Es sei $f: X \rightarrow Y$ bijektiv und $g, h \in \text{MI}(f)$. Dann ist $f^{-1}: Y \rightarrow X$ monoton wachsend genau dann, wenn die durch

$$\left. \begin{aligned} a_0 &= x \\ a_{k+1/2} &= g(a_k, y) \\ a_{k+1} &= h(a_{k+1/2}, y) \end{aligned} \right\} k=0, 1, 2, \dots, \quad (11)$$

berechnete Folge (a_k) für jedes $x \in X$ und für jedes $y \in Y$ gegen $f^{-1}(y)$ konvergiert.

Beweis. a) f^{-1} sei monoton wachsend. Wir wählen $x \in X$, $y \in Y$ beliebig. Nach den Voraussetzungen über die Halbordnung in Y gibt es Elemente $\underline{y}, \bar{y} \in Y$ so daß

$$\underline{y} \leq f(x) \leq \bar{y} \quad (12)$$

und

$$\underline{y} \leq y \leq \bar{y}. \quad (13)$$

Definieren wir $\underline{x} = f^{-1}(\underline{y})$ und $\bar{x} = f^{-1}(\bar{y})$, so folgt aus (12), da f^{-1} monoton ist,

$$\underline{x} = f^{-1}(\underline{y}) \leq x \leq f^{-1}(\bar{y}) = \bar{x}. \quad (14)$$

Neben der durch (11) berechneten Folge (a_k) betrachten wir zwei Folgen $(\underline{a}_k)$ und $(\bar{a}_k)$ definiert durch

$$\underline{a}_0 = \underline{x}, \quad \underline{a}_{k+1/2} = g(\underline{a}_k, y), \quad \underline{a}_{k+1} = h(\underline{a}_{k+1/2}, y), \quad k=0, 1, 2, \dots$$

und

$$\bar{a}_0 = \bar{x}, \quad \bar{a}_{k+1/2} = g(\bar{a}_k, y), \quad \bar{a}_{k+1} = h(\bar{a}_{k+1/2}, y), \quad k=0, 1, 2, \dots$$

Wegen (13) und mit der Definition von $\underline{x}$ und $\bar{x}$ gilt $f(\underline{x}) \leq y \leq f(\bar{x})$. Zusammen mit (14) bedeutet dies, daß die Voraussetzungen von Theorem 2 über $\underline{x}$ und $\bar{x}$ gelten. Daher folgt nach Theorem 2

$$\underline{a}_k \uparrow \underline{a}, \quad \bar{a}_k \downarrow \bar{a} \quad (15)$$

sowie $f(\underline{a}) = f(\bar{a}) = y$. Da f bijektiv ist, folgt $\underline{a} = \bar{a} = f^{-1}(y)$.

Wir zeigen jetzt, daß

$$\underline{a}_k \leq a_k \leq \bar{a}_k, \quad k=0, 1, 2, \dots \quad (16)$$

gilt. Wegen (15) und (C5) aus [3] folgt dann die Behauptung. (16) ist sicherlich für $k=0$ richtig. Gilt (16) für ein $k \geq 0$, so folgt wegen (M1)

$$\underline{a}_{k+1/2} = g(\underline{a}_k, y) \leq g(a_k, y) = a_{k+1/2} \leq g(\bar{a}_k, y) = \bar{a}_{k+1/2}$$

und damit wiederum wegen (M1)

$$a_{k+1} = h(a_{k+1/2}, y) \leq h(a_{k+1/2}, y) = a_{k+1} \leq h(\bar{a}_{k+1/2}, y) = \bar{a}_{k+1}.$$

Damit ist (16) gezeigt.

b) Es gelte $a_k \rightarrow f^{-1}(y)$ für beliebige x, y für die durch (11) berechnete Folge. Es ist dann zu zeigen, daß $b_1 \leq b_2$ aus $f(b_1) \leq f(b_2)$ folgt.

Dazu betrachten wir (11) und wählen $a_0 = b_1, y = f(b_2)$. Es gilt dann

$$a_k \leq a_{k+1}, \quad f(a_k) \leq y, \quad k=0, 1, 2, \dots \quad (17)$$

Dies wird durch vollständige Induktion bewiesen:

Für $k=0$ bemerken wir zunächst, daß die Voraussetzung $f(b_1) \leq f(b_2)$ äquivalent ist mit $f(a_0) = f(b_1) \leq f(b_2) = y$. Nach (M1) und (M2) gilt

$$a_{1/2} = g(a_0, y) = g(b_1, f(b_2)) \geq g(b_1, f(b_1)) = b_1 = a_0,$$

also $a_0 \leq g(a_0, y)$. Aus (M3) folgt $f(a_{1/2}) = f(g(a_0, y)) \leq y$. Durch Wiederholung der analogen Schritte mit h anstelle von g erhält man

$$a_1 \geq a_{1/2} \geq a_0, \quad f(a_1) \leq y.$$

Für $k > 0$ werden diese Ungleichungen unter Verwendung der Induktionsvoraussetzung analog bewiesen, d.h. (17) gilt. Wegen $a_k \uparrow f^{-1}(y) = b_2$ folgt nach (C3) aus [3], daß $a_k \leq b_2, k=0, 1, 2, \dots$, insbesondere also $b_1 = a_0 \leq b_2$ gilt. $\square$

Bemerkung 2. Wir zeigen an einem Beispiel, daß das Produkt zweier MI-Funktionen i.allg. keine MI-Funktion ist. Damit ist dann klar, daß die Beweise der Theoreme 1 bis 3 nicht durch Berufung auf die entsprechenden Theoreme in [3] geführt werden können. Dazu betrachten wir $X=Y=\mathbb{R}^2$ mit der komponentenweisen Halbordnung.

Ist $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ die Nullfunktion, $f=0$, so gehört eine stetige Funktion $g: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ zu $MI(0)$ genau dann, wenn sie den Regeln

(M1) $g(x, y)$ ist monoton wachsend bezüglich beider Argumente,

(M2) $g(x, \theta) = x$,

(M3) $\begin{cases} x \leq g(x, y) \Rightarrow \theta \leq y \\ x \geq g(x, y) \Rightarrow \theta \geq y \end{cases}$

genügt. (θ bezeichnet den Nullvektor im $\mathbb{R}^2$.)

Demnach sind (mit $y = (y_1, y_2)^T$)

$$g_1(x, y) = x + \begin{pmatrix} 2y_1 \\ y_2 \end{pmatrix}, \quad g_2(x, y) = x + \begin{pmatrix} y_2 \\ 2y_1 \end{pmatrix}$$

offensichtlich MI-Funktionen von $f=0$. Bildet man nun

$$g(x, y) = g_1(g_2(x, y), y),$$

so ergibt sich

$$g(x, y) = x + \begin{pmatrix} 2y_1 + y_2 \\ 2y_1 + y_2 \end{pmatrix}.$$

Offenbar ist

$$g\left(x, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right) = x + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \geq x,$$

aber es gilt *nicht* $\begin{pmatrix} 1 \\ -1 \end{pmatrix} \geq \theta$, d.h. (M3) *gilt nicht*.

Ist $f(x) = Ax$ eine lineare Abbildung mit einer Matrix A so heißt die Zerlegung $A = M - N$ mit zwei reellen Matrizen M und N regulär (s. Definition 2 in Abschnitt 2 unten), falls M^{-1} existiert und außerdem M^{-1} sowie N nur nichtnegative Elemente besitzen. Nach Theorem 5 in [3] erzeugt jede reguläre Zerlegung von A eine zu f gehörige MI-Funktion g , die die Darstellung $g(x, y) = M^{-1}Nx + M^{-1}y$ besitzt. (Siehe Theorem 4 in Abschnitt 2 unten.)

In unserem obigen Beispiel $f=0$ ist $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ und g_1 bzw. g_2 erhält man durch die regulären Zerlegungen

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix}$$

bzw.

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}$$

Sei nun allgemeiner für $x \in \mathbb{R}^2$

$$f(x) = -\varepsilon x = \begin{pmatrix} -\varepsilon & 0 \\ 0 & -\varepsilon \end{pmatrix} x \quad (\varepsilon > 0).$$

Dann ist $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ *bijektiv*. Wir betrachten die beiden regulären Zerlegungen

$$\begin{pmatrix} -\varepsilon & 0 \\ 0 & -\varepsilon \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} \varepsilon & 1/2 \\ 1 & \varepsilon \end{pmatrix} = M_1 - N_1,$$

$$\begin{pmatrix} -\varepsilon & 0 \\ 0 & -\varepsilon \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/2 + \varepsilon & 0 \\ 0 & 1 + \varepsilon \end{pmatrix} = M_2 - N_2.$$

Zu diesen gehören die MI-Funktionen von f

$$g_1(x, y) = M_1^{-1}N_1x + M_1^{-1}y = \begin{pmatrix} 1 & \varepsilon \\ 2\varepsilon & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_2 \\ 2y_1 \end{pmatrix}$$

und

$$g_2(x, y) = M_2^{-1}N_2x + M_2^{-1}y = \begin{pmatrix} 1 + 2\varepsilon & 0 \\ 0 & 1 + \varepsilon \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2y_1 \\ y_2 \end{pmatrix}.$$

Die Zusammensetzung $g(x, y) = g_1(g_2(x, y), y)$ ergibt

$$g(x, y) = \begin{pmatrix} 1 + 2\varepsilon & \varepsilon + \varepsilon^2 \\ 2\varepsilon + 4\varepsilon^2 & 1 + \varepsilon \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2y_1 + y_2 + \varepsilon y_2 \\ 2y_1 + y_2 + 4\varepsilon y_1 \end{pmatrix}.$$

Für hinreichend kleine positive $\varepsilon > 0$ ist dieses g kein Element von $MI(f)$ mit $f(x) = -\varepsilon x$. Andernfalls wäre nach (M3) für $x = \theta$ die Implikation

$$\theta \leq \begin{pmatrix} 2y_1 + y_2 + \varepsilon y_2 \\ 2y_1 + y_2 + 4\varepsilon y_1 \end{pmatrix} \Rightarrow -\varepsilon \begin{pmatrix} 2y_1 + y_2 + \varepsilon y_2 \\ 2y_1 + y_2 + 4\varepsilon y_1 \end{pmatrix} \leq \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

richtig. Mit $y_1 = 1, y_2 = -1$ erhält man sofort einen Widerspruch.

2. Reguläre Zerlegungen und MI-Funktionen

In [3] wurde gezeigt, daß jede reguläre Zerlegung einer Funktion f eine zu f gehörige MI-Funktion erzeugt. Wir wiederholen zunächst die Definition einer regulären Zerlegung und geben dann den Zusammenhang mit MI-Funktionen an.

Definition 2 (siehe [11, 12, 8, 1, 3]). Gegeben sei eine Funktion $f: X \rightarrow Y$. Eine Funktion $r: X \times X \rightarrow Y$ heißt dann reguläre Zerlegung von f , wenn die folgenden Eigenschaften (R1)–(R5) gelten:

(R1) $r(x, x) = f(x), x \in X$;

(R2) $r(a, x) \leq r(b, x) \Rightarrow a \leq b$;

(R3) Die Funktion $r(\cdot, x): X \rightarrow Y$ ist für jedes $x \in X$ bijektiv;

(R4) $a \leq b \Rightarrow r(x, a) \geq r(x, b)$;

(R5) Gilt für die Folgen (a_k) und (b_k) aus X neben $r(a_k, b_k) = y, k = 0, 1, 2, \dots$, entweder $a_k \uparrow a, b_k \uparrow b$ oder $a_k \downarrow a, b_k \downarrow b$, so gilt $r(a, b) = y$.

Theorem 4 (siehe [3], Theorem 5). Es sei $r: X \times X \rightarrow Y$ eine reguläre Zerlegung einer gegebenen Funktion $f: X \rightarrow Y$. Wird $g: X \times Y \rightarrow X$ definiert durch

$$g(x, y) = a \quad \text{für } r(a, x) = y,$$

so ist $g \in MI(f)$. $\square$

3. Anwendung

In diesem Abschnitt ist $X = Y = \mathbb{R}^n$. Die Halbordnung ist komponentenweise definiert.

Eine Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ sei gegeben durch

$$f(x) = (f_i(x)) = (f_i(x_1, \dots, x_n)).$$

Definition 3 (siehe [7] und [8]). Die Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ heißt außendiagonal antiton genau dann, wenn für jedes $x \in \mathbb{R}^n$ die Funktionen

$$\psi_{ij}: \mathbb{R}^1 \rightarrow \mathbb{R}^1, \quad \psi_{ij}(t) = f_i(x + t e^j), \quad i, j = 1(1)n, \quad i \neq j,$$

monoton fallend sind. (e^i bezeichnet den i -ten Einheitsvektor.) $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ heißt surjektiv diagonal isoton, wenn für jedes $x \in \mathbb{R}^n$ die Abbildungen

$$\psi_{ii}: \mathbb{R}^1 \rightarrow \mathbb{R}^1, \quad \psi_{ii}(t) = f_i(x + t e^i), \quad i = 1(1)n,$$

surjektiv und streng isoton sind. $\square$

Für $\omega > 0$ betrachten wir die folgenden Funktionen $r, s: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$:

$$a) \quad r(x, z) = (f_i(x_1, \dots, x_{i-1}, 1/\omega[x_i - (1-\omega)z_i], z_{i+1}, \dots, z_n)) \quad (18)$$

$$b) \quad s(x, z) = (f_i(z_1, \dots, z_{i-1}, 1/\omega[x_i - (1-\omega)z_i], x_{i+1}, \dots, x_n)). \quad (19)$$

Es gilt dann das folgende

Lemma 1. Ist $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig, außendiagonal antiton und surjektiv diagonal isoton, so sind r und s für $0 < \omega \leq 1$ reguläre Zerlegungen von f . $\square$

Der Beweis der Behauptung für r wurde bereits in [8] und in [1], Lemma 1, angegeben. Die Behauptung für s zeigt man genauso.

Nach Theorem 4 werden unter den Voraussetzungen von Lemma 1 durch r und s zwei zu f gehörige MI-Funktionen, etwa g und h definiert. Das durch g (d.h. das durch die Zerlegung r) definierte Iterationsverfahren zur Auflösung der Gleichung $f(x) = y$ lautet (ohne Verwendung von g geschrieben):

Wähle $x^0 = (x_i^0) \in \mathbb{R}^n$ („Startvektor“)

Für $k=0, 1, 2, \dots$:

Für $i=1(1)n$:

Löse $f_i(x_1^{k+1}, \dots, x_{i-1}^{k+1}, z_i^{k+1}, x_{i+1}^k, \dots, x_n^k) = y_i$ nach z_i^{k+1} auf.

Setze

$$x_i^{k+1} = x_i^k + \omega(z_i^{k+1} - x_i^k).$$

Dieses Verfahren wird als (nichtlineares) Relaxationsverfahren (SOR-Verfahren) bezeichnet.

Definition 4 ([7, 8]). Eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ heißt M-Funktion, wenn f außendiagonal antiton und f^{-1} monoton wachsend ist. $\square$

In [7], Theorem 2.10 hat Rheinboldt gezeigt, daß eine stetige, surjektive M-Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ surjektiv diagonal isoton ist.

Für $0 < \omega \leq 1$ und eine stetige surjektive M-Funktion wurde die Konvergenz des SOR-Verfahrens für jedes $x^0 \in \mathbb{R}^n$ und jedes $y \in \mathbb{R}^n$ gegen die eindeutige Lösung von $f(x) = y$ in [7], Theorem 3.3 gezeigt. Siehe auch [1], Abschnitt V.

Unter Verwendung der zu f gehörigen MI-Funktionen g und h , welche durch (18) und (19) unter den Voraussetzungen von Lemma 1 definiert sind, betrachten wir nun das Iterationsverfahren (11) aus Theorem 3. Dieses läßt sich in diesem Falle (ohne explizite Verwendung von g und h) schreiben als:

Wähle $x^0 = (x_i^0) \in \mathbb{R}^n$ („Startvektor“)

Für $k=0, 1, 2, \dots$:

Für $i=1(1)n$:

Löse $f_i(x_1^{k+1/2}, \dots, x_{i-1}^{k+1/2}, z_i^{k+1/2}, x_{i+1}^k, \dots, x_n^k) = y_i$ nach $z_i^{k+1/2}$ auf.

Setze $x_i^{k+1/2} = x_i^k + \omega(z_i^{k+1/2} - x_i^k)$.

Für $i=n(-1)1$:

Löse $f_i(x_1^{k+1/2}, \dots, x_{i-1}^{k+1/2}, w_i^{k+1}, x_{i+1}^{k+1}, \dots, x_n^{k+1}) = y_i$ nach w_i^{k+1} auf.

Setze $x_i^{k+1} = x_i^{k+1/2} + \omega(w_i^{k+1} - x_i^{k+1/2})$.

Dieses Verfahren wird als (nichtlineares) *symmetrisches Relaxationsverfahren* (SSOR-Verfahren) bezeichnet. Für lineare Gleichungssysteme hat dieses Verfahren in letzter Zeit als „preconditioner“ an Bedeutung gewonnen. Siehe dazu etwa [4]. Im nichtlinearen Fall haben wir das folgende Resultat.

Theorem 5. Ist $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine stetige surjektive M-Funktion, so konvergiert das SSOR-Verfahren für $0 < \omega \leq 1$ für jedes $y \in \mathbb{R}^n$ und für jeden Anfangsvektor $x^0 \in \mathbb{R}^n$ gegen die eindeutige Lösung der Gleichung $f(x) = y$. $\square$

Zum Beweis bemerken wir, daß eine M-Funktion notwendig injektiv ist (s. dazu [7], Lemma 2.3). Zusammen mit der vorausgesetzten Surjektivität folgt, daß f bijektiv ist. Nach Definition 4, Lemma 1, Theorem 4 und Theorem 3 folgt sofort die Behauptung.

4. Beispiel: Nichtlineare Differenzengleichungen

Wir betrachten im $\mathbb{R}^2$ das Dirichletproblem

$$\left. \begin{aligned} Tu &= -\Phi(x, y, u, u_x, u_y, u_{xx}, u_{yy}) = 0 & \text{in } G \subset \mathbb{R}^2 \\ u &= g & \text{auf } \partial G. \end{aligned} \right\} \quad (20)$$

Unter Verwendung eines in beiden Richtungen äquidistanten Gitters G_h mit dem Rand ∂G_h wird eine Diskretisierung vorgenommen. Diese liefert ein System von Gleichungen für die Komponenten der Gitterfunktion $v_h = (v_{kl}^h)$. Die Werte auf ∂G_h sind i.allg. ebenfalls durch Approximation entstanden. Die Ersetzung der zweiten partiellen Ableitungen durch die Differenzenquotienten zweiter Ordnung und der ersten partiellen Ableitungen durch zentrale Differenzenquotienten führt auf das nichtlineare Gleichungssystem

$$\begin{aligned} f_{kl}^h(v_h) = -\Phi \left(x_k, y_l, v_{kl}^h, \frac{v_{k+1,l}^h - v_{k-1,l}^h}{2h}, \frac{v_{k,l+1}^h - v_{k,l-1}^h}{2h}, \right. \\ \left. \frac{v_{k+1,l}^h - 2v_{k,l}^h + v_{k-1,l}^h}{h^2}, \frac{v_{k,l+1}^h - 2v_{k,l}^h + v_{k,l-1}^h}{h^2} \right) = 0 \end{aligned} \quad (21)$$

mit $(x_k, y_l) \in G_h$. Siehe dazu [9] und [10].

In [9], Satz 4.1 und der daran sich anschließenden Folgerung wurde die folgende Aussage bewiesen.

Theorem 6. Für $(u, p, q, r, t) \in \mathbb{R}^5$ sei $\Phi(x, y, u, p, q, r, t)$ bezüglich u, p, q, r, t einmal stetig partiell differenzierbar und es gelte

$$\frac{\partial}{\partial r} \Phi > 0, \quad \frac{\partial}{\partial t} \Phi > 0, \quad \frac{\partial}{\partial u} \Phi \leq 0.$$

Dann ist die durch (21) definierte Funktion $f^h(v_h) = (f_{kl}^h(v_h))$ für $h \leq h_0$ eine M-Funktion, wenn h_0 so gewählt wird, daß

$$\frac{h_0}{2} \left| \frac{\partial}{\partial p} \Phi \right| < \frac{\partial}{\partial r} \Phi, \quad \frac{h_0}{2} \left| \frac{\partial}{\partial q} \Phi \right| < \frac{\partial}{\partial t} \Phi$$

für alle $(x, y) \in G \cup \partial G$ gilt. $\square$

Falls Φ stetig ist, so ist auch $f^h(v_h)$ stetig. Zur Anwendung von Theorem 5 auf das nichtlineare Gleichungssystem (21) benötigt man noch die Surjektivität von $f^h(v_h)$. Rheinboldt hat in [7] und [8] hinreichende Bedingungen für die Surjektivität von M-Funktionen angegeben. Beispielsweise erhält man für sogenannte halblinare Aufgaben

$$\begin{aligned} Tu &= -\Delta u + \varphi(x, y, u, u_x, u_y) = 0, & G \subset \mathbb{R}^2 \\ u &= g & \text{auf } \partial G \end{aligned}$$

mit der oben angegebenen Diskretisierung unter den Voraussetzungen von Theorem 6 bei stetigem φ eine stetige surjektive M-Funktion so daß Theorem 5 anwendbar ist.

Literatur

1. Alefeld, G.: Über reguläre Zerlegungen bei nichtlinearen Abbildungen. Z. Angew. Math. Mech. **52**, 233–238 (1972)
2. Alefeld, G., Varga, R.S.: Zur Konvergenz des symmetrischen Relaxationsverfahrens. Numer. Math. **25**, 291–295 (1976)
3. Alefeld, G., Volkman, P.: Regular Splittings and Monotone Iteration Functions. Numer. Math. **46**, 213–228 (1985)
4. Axelsson, O., Barker, V.A.: Finite element solution of boundary value problems. New York: Academic Press (1984)
5. Neumaier, A., Varga, R.S.: Exact Convergence and Divergence Domains for the Symmetric Successive Overrelaxation Iterative (SSOR) Method Applied to H -Matrices. Linear Algebra Appl. **58**, 261–272 (1984)
6. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press 1970
7. Rheinboldt, W.C.: On M-Functions and Their Application to Nonlinear Gauss-Seidel Iterations and to Network Flows. J. Math. Anal. Appl. **32**, 274–307 (1970)
8. Rheinboldt, W.C.: On Classes of n -Dimensional Nonlinear Mappings Generalizing Several Types of Matrices. In: Numerical Solution of Partial Differential Equations – II, SYNPADE 1970 (B. Hubbard, ed.), pp. 501–546. New York: Academic Press 1971
9. Törnig, W.: Monoton konvergente Iterationsverfahren zur Lösung nichtlinearer Differenzen-Randwertprobleme. Beitr. Numer. Math. **4**, 245–257 (1975)
10. Törnig, W.: Monoton einschließend konvergente Iterationsprozesse vom Gauß-Seidel-Typ zur Lösung nichtlinearer Gleichungssysteme im $\mathbb{R}^N$ und Anwendungen. Math. Methods Appl. Sci. **2**, 489–503 (1980)
11. Varga, R.S.: Factorization and normalized iterative methods. In: Boundary Problems in Differential Equations. (R.E. Langer, ed.) pp. 121–142. Madison: University of Wisconsin Press 1960
12. Varga, R.S.: Matrix Iterative Analysis. Englewood Cliffs, N.J.: Prentice Hall 1962
13. Varga, R.S., Niethammer, W., Cai, D.-Y.: P -cyclic matrices and the symmetric successive overrelaxation method. Linear Algebra Appl. **58**, 425–439 (1984)

Received May 6, 1986